frontiers | Research Topics

Community series-extremophiles: Microbial genomics and taxogenomics, volume II

Edited by
Rafael R. de la Haba, André Antunes, Brian P. Hedlund and Mohamed Jebbar

Published in
Frontiers in Microbiology

ISSN 1664-8714
ISBN 978-2-8325-4458-7
DOI 10.3389/978-2-8325-4458-7

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

Community series-extremophiles: Microbial genomics and taxogenomics, volume II

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
Mohamed Jebbar — Université de Bretagne Occidentale, France

Citation

de la Haba, R. R., Antunes, A., Hedlund, B. P., Jebbar, M., eds. (2024). *Community series-extremophiles: Microbial genomics and taxogenomics, volume II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-4458-7

Table of contents

frontiers | Frontiers in Microbiology

TYPE Editorial
PUBLISHED 31 January 2024
DOI 10.3389/fmicb.2024.1371210

OPEN ACCESS

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

*CORRESPONDENCE
André Antunes
aglantunes@must.edu.mo
Rafael R. de la Haba
rrh@us.es
Mohamed Jebbar
mohamed.jebbar@univ-brest.fr
Brian P. Hedlund
brian.hedlund@unlv.edu

RECEIVED 16 January 2024
ACCEPTED 16 January 2024
PUBLISHED 31 January 2024

CITATION
Antunes A, de la Haba RR, Jebbar M and Hedlund BP (2024) Editorial: Community series-extremophiles: microbial genomics and taxogenomics, volume II.
Front. Microbiol. 15:1371210.
doi: 10.3389/fmicb.2024.1371210

Editorial: Community series-extremophiles: microbial genomics and taxogenomics, volume II

André Antunes[1,2,3]*, Rafael R. de la Haba[4]*, Mohamed Jebbar[5]* and Brian P. Hedlund[6,7]*

[1]State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology, Taipa, Macau SAR, China, [2]Macau Center for Space Exploration and Science, China National Space Administration (CNSA), Taipa, Macau SAR, China, [3]China-Portugal Belt and Road Joint Laboratory on Space and Sea Technology Advanced Research, Virtual Lab, [4]Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [5]Univ Brest, CNRS, Ifremer, EMR 6002 BIOMEX, Unité Biologie et Écologie des Écosystèmes Marins Profonds BEEP, Plouzané, France, [6]School of Life Sciences, University of Nevada, Las Vegas, NV, United States, [7]Nevada Institute of Personalized Medicine, University of Nevada, Las Vegas, NV, United States

KEYWORDS

extremophiles, genomics, taxonomy, metagenomics, astrobiology, space microbiology, biodiversity

Editorial on the Research Topic
Community series-extremophiles: microbial genomics and taxogenomics, volume II

Introduction

The importance of extreme environments and their microbial inhabitants has been gaining visibility. Such environments account for most of Earth's habitable zone by volume (Gold, 1992; Charette and Smith, 2010) and they are sources of valuable new strains and bio-products that will fuel the bioeconomy and will provide new solutions to the current challenges we face in the in areas of human health, energy, environment, and agriculture (Antunes et al., 2016; Corral et al., 2019). Developments in extremophile microbiology are increasingly linked with technological improvements in sequencing, bioinformatics, and other upcoming technologies that are opening new paths for scientific breakthroughs and promises to change our understanding of life on Earth. The evolution and physiology of extremophiles remain relevant to investigate the universal ancestor and the limits of life (Merino et al., 2019). The study of extremophiles is also vital to both Astrobiology and Space Microbiology and will support humanity's next steps into space (e.g., Antunes and Meyer-Dombard, 2023; Simões et al., 2023).

Within this setting, we hosted the special Research Topic entitled *Extremophiles: Microbial genomics and taxogenomics* (de la Haba et al., 2022). Based on its success, we then launched a second volume as the start of a community series dedicated to this topic. Our new volume includes 14 manuscripts, covering a wide range of areas within the field, with contributions on: (i) the discovery of new taxa and genomic-based analysis of their properties; (ii) new community-based studies in previously overlooked extreme and poly-extreme environments; and (iii) new insights into high taxonomic ranks of extremophiles.

The volume was well received, with a total of 77 contributing authors from institutions in 14 different countries spread across the globe, namely Australia, Belgium, Botswana, China, Ethiopia, Germany, India, Italy, Poland, Russia, Spain, Sweden, United Kingdom, and United States of America. At the time of writing, over 190,000 views and downloads have been recorded for both volumes of the series.

New taxa and their properties

The isolation of new extremophiles is vital within this field as highlighted by several contributions to this volume describing new taxa and the use of genomics to uncover some of their remarkable properties.

Two articles on hypersaline soils (Odiel Saltmarshes Natural Area, Spain) describe new taxa from this previously overlooked type of environment. In the first article, Galisteo et al. (a) describe the novel species *Fodinibius salsisoli*. Comparative genomics revealed that the species of the genera *Aliifodinibius* and *Fodinibius* belong to the same genus, so the authors propose the reclassification of the species of the former into the single genus *Fodinibius*. Further analysis highlighted abundant but still uncultured representatives of the family *Balneolacea* in this environment and their potential role as a source of biotin for other organisms. In the second publication from this group, Galisteo et al. (b) describe the new genus *Terrihalobacillus*, and the new species *Terrihalobacillus insolitus* and *Aquibacillus salsiterrae*, both based on several isolates. The genomic features of these new taxa were also analyzed, and both were suggested to belong to the rare biosphere of these environments.

An additional study by García-Roldan et al., also focusing on high-salinity environments in Spain (a solar saltern from Isla Cristina, Huelva), describes a new species of *Natronomonas*, named *Natronomonas aquatica*. Further genome-based phylogenetic and metabolic analysis of this genus confirmed its phylogenetic coherence, indicated a heterotrophic lifestyle and versatile nitrogen metabolisms, and suggested its ubiquity in mid- to high-salinity environments.

The contribution by Zavarzina et al. focuses on haloalkaliphiles inhabiting soda lakes in Mongolia and Kenya. In this article, two new strains of *Dethiobacter alkaliphilus*, initially described as sulfur/thiosulfate-reducing and iron-reducing bacteria, respectively, were both found to be capable of reducing iron and thiosulfate. Distinct sets of multiheme cytochromes that probably mediate these reactions were identified and their high variation was suggested to be an effective adaptive strategy for occupying geochemically diverse extreme environments. Analysis of the distribution of this genus in natural environments revealed strong connection to alkaline, sulfur- or iron-enriched ecotopes (soda lakes and serpentinite-associated sediments). Soda lakes are geologically ancient ecosystems that are still widespread on Earth and that are thought to harbor archaic microbial communities, but they might constitute secondary habitats for *D. alkaliphilus* when compared with Fe-rich serpentinites. These results increase our understanding of the metabolic diversity of extremophiles and point to evolutionary traits that could have occurred in prokaryotes during the early stages of the biosphere's evolution, namely when the sulfur biogeochemical cycle overtook the iron cycle as a predominant biogeochemical process.

The inspection of the extensive repositories in culture collections can also yield new taxa. The contribution by Montero-Calasanz et al. looks at isolates deposited at the Leibniz Institute DSMZ in the 1990s that originated from soils in the Atacama Desert, Chile, and landfill leachate from Vancouver, Canada. Analysis of these isolates, which belong to the *Geodermatophilaceae*, used new genomic data to determine genotype-phenotype correlations within this family. Based on their results, the authors describe and name four novel species in the genus *Blastococcus*, introduce the new genera *Trujillonella*, *Pleomorpha*, and *Goekera*, and reclassify *Blastococcus endophyticus* as *Trujillonella endophytica*, *Geodermatophilus daqingensis* as *Pleomorpha daqingensis*, and *Modestobacter deserti* as *Goekera deserti*. The authors also emphasize how genomics can effectively replace and/or complement some of the routine phenotypic analyses in microbial systematics.

Additionally, two articles from the Daroch group describe new genera of thermophilic cyanobacteria isolated from terrestrial hot springs and name them under the botanical code *Trichothermofontia sichuanensis* (Tang et al.) and *Thermocoleostomius sinensis* (Jiang et al.). Both articles also investigate carbon-concentrating mechanisms based on annotations of bicarbonate transporters, carbonic anhydrases, and carboxysome-associated genes. For *Thermocoleostomius sinensis*, these genes were also shown to be expressed under low-CO_2 conditions.

Community-based studies

The use of molecular methods to analyze microbial community structure and functional diversity in different environments continues to be a formidable tool. This volume includes new community-based studies in previously overlooked or understudied extreme and poly-extreme environments in Africa, Asia, Europe, and the Americas.

In one of the contributions, Gawas and Kerkar study sediments from three salt pans, adjoining different prominent estuaries in Goa, India. Marked differences in the unique genera found for each site led the authors to suggest that the different estuaries select for unique bacterial diversity.

In another publication, an amplicon sequencing study in the East African Rift Valley analyzes prokaryotic and eukaryotic diversity in three soda lakes, Lake Abijata, Lake Chitu, and Lake Shala, and compares the results with enrichment culture methods (Jeilu et al.). The culture-independent approach captured higher diversity than previously detected in these lakes, provided a more reliable estimate of microbial diversity than traditional culture methods, pointed to differences in microbial community composition among the three lakes, and identified the most commonly found prokaryotic taxa—*Pseudomonadota* (*Halomonas*), *Bacillota* (*Bacillus*, *Clostridia*), *Bacteroidota* (*Bacteroides*), *Euryarchaeota* (*Thermoplasmata*, *Thermococci*, *Methanomicrobia*, *Halobacter*), and *Nanoarchaeota* (*Woesearchaeia*)—and eukaryotic taxa—*Ascomycota* and *Basidiomycota*.

In a different poly-extreme setting, Zhang et al. assess the prokaryotic taxonomy and potential functional diversity of three hyper-arid soil samples from the Gobi Desert. Their study compares two different next-generation sequencing platforms and identifies 36 bacterial phyla, including *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, *Acidobacteriota*, *Nitrososphaerota*, and *Planctomycetota*. The authors suggest that environmental factors such as total dissolved salts, available potassium, total nitrogen, and organic matter are positively correlated with the abundance of most groups, and that the community structure was mainly controlled by stochastic processes.

Another focus of this section is on metagenomes from geothermal systems. The study by Ugwuanyi et al. compares metagenomes from different substrates—water, mud, and fumarolic deposits—from Solfatara and Pisciarelli hydrothermal systems in Italy, and reveals a high abundance of pathways for carbon fixation and sulfur oxidation. A separate metagenomic study recovered and interpreted *Caldarchaeales* metagenome-assembled genomes from geothermal features in Hawai'i and Chile and proposes both chemoorganotrophic and chemolithotrophic lifestyles (Balbay et al.).

New insights into high taxonomic ranks of extremophiles

The final highlights of this volume center on studies targeting high ranks in the taxonomy of extremophiles and the new insights that such studies can provide.

Looking into new frontiers within this field, Wu et al. delve into the archaeal class *Halobacteria* and provide an overview of astrobiology-relevant exposure experiments, identifying key knowledge gaps and research opportunities within this model group of microorganisms. The authors stress that further testing is still needed to fully cover the taxonomic diversity present in this archaeal group both at high and low ranks. This should be prioritized given the relevance of the group for the search for life on Mars and on the exooceans of the icy moons, for planetary protection, and for the expected future needs of the space biotechnology sector.

The final contribution looks at the *Halomonadaceae*, the most diverse family of halophilic bacteria, and a group in dire need of an in-depth taxogenomic analysis. The study by de la Haba et al. delivers on this daunting task by sequencing the genomes of 17 type strains and comparing them with all other publicly available genome sequences within this family. Based on comparative genomic, phylogenomic, and clade-specific signature gene analyses, the authors suggest that the genus *Halovibrio* is misplaced within the family, propose a division of the genus *Halomonas* into seven separate genera with reclassification of some of its current species to the genus *Modicisalibacter*, and identify various synonymous species names within the family.

Final remarks

The study of extreme environments and their microbial inhabitants is the subject of intense interest and activity by researchers across different expertise and spread across the globe. The next few years of research, supported by the increased use and further developments in genomics and taxogenomics, promise to continue to deliver exciting new advances on this disciplinary field with implications on different aspects of our lives on Earth and beyond.

Author contributions

AA: Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. RRH: Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. MJ: Data curation, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing. BPH: Conceptualization, Data curation, Project administration, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. AA acknowledges funding from the National Key R&D Program of China (Grant No. 2022YFE0204600), the Science and Technology Development Fund, Macau SAR, China [File No. SKL-LPS(MUST)-2021-2023], and from a Faculty Research Grant of Macau University of Science and Technology's (FRG grant No. FRG-22-079-LPS). RRH was supported by grants PID2020-118136GB-I00 funded by MCIN/AEI/10.13039/501100011033 and from the Junta de Andalucía (P20_01066 and BIO-213), both with FEDER funds, and was a recipient of a short-stay grant (PRX21/00598) from the Spanish Ministry of Universities. MJ gratefully acknowledges the support from the University of Brest, Ifremer and CNRS.

Acknowledgments

We would like to thank all contributing authors for their submissions to this volume of the community series, external handling editors, as well as all the reviewers for their valuable comments and inputs.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Antunes, A., and Meyer-Dombard, D. R. (2023). Rising stars in space microbiology: 2022. *Front. Microbiol.* 14, 1322924. doi: 10.3389/fmicb.2023.1322924

Antunes, A., Stackebrandt, E., and Lima, N. (2016). Fueling the bio-economy: European culture collections and microbiology education and training. *Trends Microbiol.* 24, 77–79. doi: 10.1016/j.tim.2015.11.010

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

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., Antunes, A., and Hedlund, B. P. (2022). Extremophiles: Microbial genomics and taxogenomics. *Front. Microbiol.* 13, 984632. doi: 10.3389/fmicb.2022.984632

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

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10, 780. doi: 10.3389/fmicb.2019.00780

Simões, M. F., Cortesão, M., Azua-Bustos, A., Bai, F. Y., Canini, F., Casadevall, A., et al. (2023). The relevance of fungi in astrobiology research–Astromycology. *Mycosphere* 14, 1190–1253. doi: 10.5943/mycosphere/14/1/13

frontiers | Frontiers in Microbiology

TYPE Perspective
PUBLISHED 13 October 2022
DOI 10.3389/fmicb.2022.1023625

OPEN ACCESS

EDITED BY
Mohammad Ali Amoozegar,
University of Tehran, Iran

REVIEWED BY
Ricardo Amils,
Autonomous University of Madrid, Spain
Bonnie K. Baxter,
Westminster College, United States

*CORRESPONDENCE
André Antunes
aglantunes@must.edu.mo

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

RECEIVED 20 August 2022
ACCEPTED 07 September 2022
PUBLISHED 13 October 2022

CITATION
Wu J-H, McGenity TJ, Rettberg P, Simões MF, Li W-J and Antunes A (2022) The archaeal class *Halobacteria* and astrobiology: Knowledge gaps and research opportunities. *Front. Microbiol.* 13:1023625. doi: 10.3389/fmicb.2022.1023625

The archaeal class *Halobacteria* and astrobiology: Knowledge gaps and research opportunities

Jia-Hui Wu[1,2], Terry J. McGenity[3], Petra Rettberg[4], Marta F. Simões[1,2], Wen-Jun Li[5] and André Antunes[1,2]*

[1]State Key Laboratory of Lunar and Planetary Sciences, Macau University of Science and Technology (MUST), Taipa, Macau SAR, China, [2]China National Space Administration (CNSA), Macau Center for Space Exploration and Science, Taipa, Macau SAR, China, [3]School of Life Sciences, University of Essex, Colchester, United Kingdom, [4]German Aerospace Center (DLR), Institute of Aerospace Medicine, Köln, Germany, [5]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

Water bodies on Mars and the icy moons of the outer solar system are now recognized as likely being associated with high levels of salt. Therefore, the study of high salinity environments and their inhabitants has become increasingly relevant for Astrobiology. Members of the archaeal class *Halobacteria* are the most successful microbial group living in hypersaline conditions and are recognized as key model organisms for exposure experiments. Despite this, data for the class is uneven across taxa and widely dispersed across the literature, which has made it difficult to properly assess the potential for species of *Halobacteria* to survive under the polyextreme conditions found beyond Earth. Here we provide an overview of published data on astrobiology-linked exposure experiments performed with members of the *Halobacteria*, identifying clear knowledge gaps and research opportunities.

KEYWORDS

astrobiology, archaea, *Halobacteria*, extremophiles, Mars, icy moons

Introduction

Astrobiology is a relatively recent but increasingly relevant cross-disciplinary field of research, which focuses on the origin, distribution, and evolution of Life in the universe (Des Marais et al., 2008; Horneck et al., 2016; Hallsworth et al., 2021b). Fittingly for such a wide scope of study, astrobiology combines techniques and expertise from fields as diverse as biology, chemistry, geology, or planetary sciences (Dartnell and Burchell, 2009; Race et al., 2012; Taşkın and Aydinoglu, 2015). With the increase in frequency and broadening of scope of space exploration, including human space travel, it has become vital to better understand the survival and reproduction of microbial life under multiple extreme conditions, and their impact on space exploration missions and on human travelers who they may encounter or with whom they may cohabit

(Cavicchioli, 2002; Moissl-Eichinger, 2011; Moissl-Eichinger et al., 2016; Cottin et al., 2017; Checinska Sielaff et al., 2019; Merino et al., 2019; Nickerson et al., 2022).

Terrestrial analogs

One of the main pillars of astrobiology is the investigation of specific terrestrial analog environments and the study of their microbial inhabitants. Such analogs exhibit characteristics found in other locations in the Solar System (Preston and Dartnell, 2014; Martins et al., 2017). Thus far, most terrestrial analog sites used for astrobiology have been selected based on conditions found on Mars (Fairén et al., 2010; Martins et al., 2017), although a few suitable analogs for the exooceans of the icy moons of the outer solar system have also been proposed (Marion et al., 2003; Kereszturi and Keszthelyi, 2013; Antunes et al., 2020).

The study of terrestrial analog sites is seen as essential for: (i) studying the limits of life (e.g., Azua-Bustos et al., 2012; DasSarma et al., 2017; Koschnitzki et al., 2021), (ii) obtaining new microbes for astrobiological exposure experiments (e.g., Musilova et al., 2015; Beblo-Vranesevic et al., 2020), (iii) analyzing long-term viability and preservation of microbes and biomolecules (e.g., Gramain et al., 2011; Leuko et al., 2017; Schreder-Gomes et al., 2022), (iv) technology development and testing for life-detection in space missions (e.g., Harris et al., 2015; García-Descalzo et al., 2019; Lantz et al., 2020; Dypvik et al., 2021; Rull et al., 2022), and (v) defining and refining planetary protection measures (e.g., Rettberg et al., 2019).

When analyzing the potential viability of microbes in other parts of our Solar System, several environmental factors need to be taken into consideration (Horneck et al., 2010; Moissl-Eichinger et al., 2016; Cottin et al., 2017). The limited information on the physical-chemical conditions present in the exooceans of the icy moons partly restrains more in-depth analyses on the survivability of microbes, despite some very interesting studies on potential microbial metabolisms (e.g., Taubner et al., 2018; Goordial, 2021; Sherwood Lollar et al., 2021). This is in stark contrast with the large amount of information already available about Mars and its apparently unwelcoming conditions. These are particularly harsh to mesophiles, combining high levels of radiation (particularly ultra-violet UVC and UVB) (Cockell et al., 2000; Vicente-Retortillo et al., 2020; Zhang et al., 2022), desiccation/low water activity (Knoll and Grotzinger, 2006; Stevenson et al., 2015b; Toner and Catling, 2018; Rivera-Valentín et al., 2020; Hallsworth et al., 2021a), high concentrations of salts and volatile oxidants induced by radiation (Quinn et al., 2013; Lasne et al., 2016; Wu et al., 2018), locations with acidic conditions (Benison and Bowen, 2006; Bibring et al., 2006; Noe Dobrea and Swayze, 2010; Ehlmann et al., 2016), solar particle events (Lillis and Brain, 2013; Ramstad et al., 2018), low temperatures, low pressures (Pätzold et al., 2016; Spiga et al., 2018), and low nutrient availability (Tomkins et al., 2019; Fackrell et al., 2021; Tarnas et al., 2021). Although the extreme setting of Mars has traditionally been seen as biocidal, this is not necessarily the case, as recently highlighted by Hallsworth (2021).

Water is essential for the existence and development of all known life forms, so the search for life in other parts of our solar system and beyond is directly linked with the existence of this vital substance. Luckily for our pursuits, water seems to be more abundant than previously thought (Nimmo and Pappalardo, 2016; Grasset et al., 2017; Greenwood et al., 2018; Schörghofer et al., 2021); there is now evidence that water deposits exist on Mars (Ojha et al., 2014; Martín-Torres et al., 2015; Stillman et al., 2017; Rivera-Valentín et al., 2020; Lauro et al., 2021; Deutsch et al., 2022) as well as on several icy moons (Spencer and Nimmo, 2013; Vance et al., 2014; Hayes, 2016; Jia et al., 2018; Genova et al., 2022). Despite several unknowns, these water bodies outside our planet appear to contain high levels of salt, including perchlorate (Navarro-González et al., 2010; Oren et al., 2014; Martín-Torres et al., 2015; Lauro et al., 2021). Therefore, the study of Earth's wide range of high salinity environments (from salterns, to polar brine aquifers, deep-sea brines or halite and other salt deposits), as well as their microbial inhabitants, has become increasingly relevant in the context of astrobiology.

Halophiles as models for astrobiology

Organisms living in high salinity environments are known as halophiles (i.e., salt loving), with representatives spread across the three domains of Life. Microbes living specifically at the higher range of salinities are almost exclusively composed of members of the archaeal class *Halobacteria*, a phylogenetically coherent group of extreme halophiles (Edbeib et al., 2016; DasSarma and DasSarma, 2017). Among other properties, members of this class are known for their resilience and ability to thrive under multiple environmental extremes (polyextremophilic) including high salinity, high levels of radiation, vacuum, extreme cold, utilization or growth in the presence of caustic materials such as perchlorate, oxygen deprivation, low-nutrient availability, and desiccation, all of which are relevant for astrobiology and particularly important in a Martian context (e.g., Rittmann et al., 2004; McCready et al., 2005; Crowley et al., 2006; Baxter et al., 2007; Coker et al., 2007; Oren et al., 2014; Stan-Lotter and Fendrihan, 2015; Stevenson et al., 2015a; Laye et al., 2017; Sorokin et al., 2017; Thombre R. et al., 2017; Laye and DasSarma, 2018; Leuko et al., 2020; further details provided in Section "The archaeal class Halobacteria and astrobiology: Astrobiological testing," and **Table 1**).

TABLE 1 Overview of *Halobacteria* tested for astrobiology-relevant features (number of species/total species, genera/total genera in each taxon).

Strain information						Astrobiology tests																		
						Simulated single factors tests																Simulated multiple factors tests	Exposure	
C	O	F	Species/Strain	T	Origin	Ionizing radiation				UV	Vacuum	Freeze-thaw	Martian chemistry			Martian environment		Desiccation	Gravity			Stratosphere	Orbital
						R1	R2	R3	R4				M1	M2	M3	M4	M5		G1	G2			
Halobacteria (27/309), (14/72)			*Har. argentinensis* RR10		Saltern														x[1]				
	Halobacteriales (12/123), (6/35)	*Haloarculaceae* (4/43), (1/11)	*Har. japonica*	T	Saltern (soil)					x[2]													
			Har. marismortui	T	Saline lake									x[3]									
			Har. vallismortis	T	Salt pool									x[3]									
		Halobacteriaceae (5/70), (4/23)	*Hbt. bonnevillei*	T	Salt flat (crust)								x[4,5]										
			Hbt. salinarum	T	Salted fish		x[6,7]			x[6,7]								x[8] #					
			Hbt salinarum 63-R2		Salted hide					x[9]													
			Hbt. salinarum R1		Lab strain					x[9]				x[3]									
			Hbt. salinarum DSM 670		Salted fish	x[10]	x[10]	x[10]		x[11] #								x[8] #			x[11] a		
			Hbt. salinarum NRC-1		Saltern		x[12-16]	x[17]	x[18,19]	x[2,19-22]	x[16] #	x[23-25]		x[3,26]		x[27]		x[16] # x[28]		x[29]	x[28] b x[30] c	x[23]	
			Hbt. salinarum NCCB 33035		Salted fish																x[31] d x[32,33] e x[34] f		
			Hvn. carboxidivorans	T	Saline soil										x[4]								
			Hvr. luteus	T	Salt lake					x[35] #	x[35] #							x[35] #					
			Nmn. pharaonis	T	Soda lake					x[36]													
		Haloferacaceae (3/10), (1/1)	*Hcc. dombrowskii*	T	Salt deposit					x[2] #						x[37]			x[38]		x[30] c		x[39]
			Hcc. hamelinensis	T	Stromatolites	x[10]	x[10]	x[10]		x[11] # x[40]													
			Hcc. morrhuae	T	Saline lake	x[10]	x[10]	x[10]		x[11] # x[41]	x[41]	x[41]				x[41]		x[41]			x[11] a x[41] n x[41] o		x[41] n x[41] o

(Continued)

10.3389/fmicb.2022.1023625

TABLE 1 (Continued)

C	O	F	Species/ strain	T	Origin	Ionizing radiation				UV	Vacuum	Freeze-thaw	Martian chemistry			Martian environment		Desiccation	Gravity		Simulated multiple factors tests	Stratosphere	Orbital
						R1	R2	R3	R4				M1	M2	M3	M4	M5		G1	G2			
Halobacteria (27/309), (14/73)	*Haloferacales* (9/115), (3/19)	*Haloferacaceae* (5/49), (1/10)	*Hfx. denitrificans*	T	Saltern									x[3,42]									
			Hfx. gibbonsii	T	Saltern									x[3]									
			Hfx. mediterranei	T	Saltern									x[3,43]					x[38]				
			Hfx. namakaokahaiae	T	Saline cinder										x[44]								
			Hfx. volcanii	T	Saline lake (sediment)					x[9]				x[3]									
		Halobacteriaceae (4/66), (2/9)	*Hbl. saliterrae*	T	Saline soil								x[4,45]										
			Hrr. ezzemoulense Halo-G*		Sea salt													x[28,46]			x[28] b x[47] j,k x[46] i,n,p,q,r		x[46] n,p,q,r x[48] n,q x[49]
			Hrr. lacusprofundi	T	Saline lake (Antarctic)							x[24] x[23]		x[26]								x[23]	
			Hrr. saccharovorum H3a		Salt lake (water and silt)							x[50,51]		x[51]									
			Hrr. str. BV1		Salt flat (sediment)										x[52]								
			Hrr. str. EDT3, EDT6		Saltern					x[53]													
			Hrr. str. H2a, H4a, H7a, H11a		Salt lake							x[50,51]											
			Hrr. str. H13a		Salt lake							x[50,51]		x[51]									
			Hrr. Str. "salsolis"		Salt					x[54]													
	Natrialbales (6/71), (5/18)	*Natrialbaceae* (6/71), (5/18)	*Htg. jeotgali* MS2		Saline river		x[55]											x[55]					
			Htg. str. SGH1		Halite (endolithic consortia)									x[56]									
			Nab. segyptia MS17		Saline river		x[55]											x[55]					
			Nab. magadii	T	Soda lake					x[57-59]	x[60]										x[60] h		
			Ncc. jeotgali RR17		Laterite rock															x[61]			
			Nnm. hispanicum	T	Salt lake/saltern	x[62]				x[63]		x[63]						x[63]					
			Nnm. pallidum	T	Salted fish					x[64]													
			Nnm. pallidum MS50		Saline river		x[55]											x[55]					
			Nrr. str. HG-1		Salt lake (sediment)					x[65]		x[65]					x[65]				x[65] l		

C = Class; O = Order; F = Family, T = Type strain; R1 = HZE (highly charged and energetic particles); R2 = γ-rays; R3 = x-rays; R4 = electron-beam irradiation; UV = ultra-violet; M1 = effect of [ClO_4^-] coupled with CO oxidation; M2 = effect of [ClO_4^-]; M3 = use of CO as carbon source; M4 = simulated Martian atmosphere (95% CO_2, 2.7% N_2, 1.6% Ar, 0.15% O_2 and 370 ppm H_2O, 10^3 Pa); M5 = Martian soil analog; G1 = microgravity; G2 = hypergravity. # = tests performed with cells within halite crystals.

Results are classified according to the original publication and respectively marked as x = growth, x = survival, x = no survival, x = data not available.

Multiple factors combinations: a = UV + low temperature, b = low temperature + desiccation, c = M4 + Martian average temperature (−60°C), d = vacuum + temperature range (100K~500K) + proton irradiation, e = vacuum + low temperature + proton irradiation, f = UV + simulated Martian temperature + M4 + proton irradiation, h = UV + vacuum, i = vacuum + desiccation, j = UV + TechShot Mars Simulation Chamber[47] + M5, k = dark + TechShot Mars Simulation Chamber[47] + M5, l = UV + M5, n = solar radiation + space vacuum + space temperature + desiccation, o = solar radiation + M4 + space temperature + desiccation, p = solar radiation + 10^5 Pa argon + space temperature + desiccation, q = dark + space vacuum + space temperature + desiccation, r = dark + 10^5 Pa argon + space temperature + desiccation.

Genus abbreviation (14 genera in total): *Har.* = *Haloarcula*, *Hbt.* = *Halobacterium*, *Hvr.* = *Halovarius*, *Hvn.* = *Halovenus*, *Nmn.* = *Natronomonas*, *Hcc.* = *Halococcus*, *Hfx.* = *Haloferax*, *Hbl.* = *Halobaculum*, *Hrr.* = *Halorubrum*, *Htg.* = *Haloterrigena*, *Nab.* = *Natrialba*, *Ncc.* = *Natrononoccus*, *Nnm.* = *Natrinema*, *Nrr.* = *Natronorubrum*.

References: 1. Thombre R. et al., 2017; 2. Fendrihan et al., 2009; 3. Oren et al., 2014; 4. Myers and King, 2020; 5. Myers and King, 2019; 6. Shahmohammadi et al., 1997; 7. Shahmohammadi et al., 1998; 8. Rittmann et al., 2004; 9. McCready, 1996; 10. Leuko and Rettberg, 2017; 11. Kish et al., 2009; 12. Webb et al., 2013; 13. Confalonieri and Sommer, 2011; 14. Whitehead et al., 2006; 15. Kottemann et al., 2005; 16. Karan et al., 2014; 17. DeVeaux et al., 2007; 18. Evans et al., 2018; 19. Leuko et al., 2015; 20. Baliga et al., 2004; 21. McCready et al., 2005; 22. Boubriak et al., 2008; 23. DasSarma et al., 2017; 24. Reid et al., 2006; 25. Weidler et al., 2004; 26. Laye and DasSarma, 2018; 27. Weidler et al., 2002; 28. Mancinelli et al., 2004; 29. Kish et al., 2012; 30. Stan-Lotter et al., 2003; 31. Koike, 1991; 32. Koike et al., 1992; 33. Koike and Oshima, 1993; 34. Koike et al., 1995; 35. Feshangsaz et al., 2020; 36. Moeller et al., 2010; 37. Leuko et al., 2002; 38. Dornmayr-Pfaffenhuemer et al., 2011; 39. Fendrihan et al., 2010; 40. Leuko et al., 2011; 41. Leuko et al., 2020; 42. Okeke et al., 2002; 43. Martínez-Espinosa et al., 2015; 44. McDuff et al., 2016; 45. Myers and King, 2017; 46. Mancinelli, 2015; 47. Johnson et al., 2011; 48. Mancinelli et al., 1998; 49. Horneck, 1999; 50. Bryanskaya et al., 2020; 51. Bryanskaya et al., 2013; 52. King, 2015; 53. Trigui et al., 2011; 54. Baxter et al., 2007; 55. Shirsalimian et al., 2017; 56. Flores et al., 2020; 57. Abrevaya et al., 2008; 58. Abrevaya et al., 2011a; 59. Abrevaya et al., 2010; 60. Abrevaya et al., 2011b; 61. Thombre R. S. et al., 2017; 62. Romano et al., 2022; 63. Mastascusa et al., 2014; 64. Martin et al., 2000; 65. Peeters et al.,

10.3389/fmicb.2022.1023625

As a result of their unique combination of adaptations and capabilities, these polyextremophilic microbes have long been hailed as models for astrobiology and flagged as priority groups for testing (DasSarma, 2006; Reid et al., 2006; Baxter et al., 2007; DasSarma et al., 2020). Accordingly, several authors have followed this research thread and data has been accumulating. However, a systematic approach for thoroughly testing *Halobacteria* or a good overview of its current status seems to be missing. We aimed to help to address this growing issue by collecting all available data from primary literature on exposure-based testing with strains of this group, mapping the results as reported in their original publications, and discussing the snapshot that we have recorded.

The archaeal class *Halobacteria* and astrobiology: Taxonomic diversity

Currently, the class *Halobacteria* is grouped into six families and three orders (Gupta et al., 2016). The order *Halobacteriales* presently has the highest number of species within the group and is split into three families: *Haloarculaceae*, *Halobacteriaceae*, and *Halococcaceae*. *Halobacteriaceae* is the largest family within the *Halobacteria*, contrasting with the *Halococcaceae* which is the smallest one and comprises a single genus (and only 10 species). To date, the order containing the second-highest number of species is *Haloferacales*, spread across two families: *Haloferacaceae*, and *Halorubraceae*, followed by the order *Natrialbales*, which includes the single family *Natrialbaceae*.

As outlined in **Table 1** and in **Supplementary Figure 1**, only a relatively small number of species within the class *Halobacteria* has been tested for their capacity to grow or survive under astrobiological- or Martian-relevant conditions, with a slightly better performance when looking at genus-based coverage. This is reflected both when looking at numbers for each class and when breaking down into intermediate taxa level such as order or family. The tested species are usually below 10% of the total number within each taxon, and between 10 and 20% for tested genera. The most noteworthy exception to these ranges can be found within the *Halococcaceae*, as it is composed of a single genus and has a small number of total species. These low numbers occur despite the widely recognized importance of the *Halobacteria* for astrobiology (e.g., Kunte et al., 2002; DasSarma, 2006; Reid et al., 2006; Baxter et al., 2007; Leuko et al., 2014; Oren, 2014a; Rettberg et al., 2019; DasSarma et al., 2020).

Another aspect worth highlighting is that most astrobiology-based studies have focused on a very narrow range of genera and species within the class *Halobacteria*, mostly associated with members of the order *Halobacteriales*, and particularly focused on the genus *Halobacterium* (**Table 1**). Indeed, the *Halobacteriales* represent more than half of the tested species within the class, followed by the *Haloferacales* (nine tested species), and by the almost unexplored *Natrialbales* (six species). Furthermore, within several families of the class we find a relatively common trend: tests are usually focused on a single genus (**Table 1**). This can be seen in the order *Halobacteriales* for family *Haloarculaceae* (genus *Haloarcula*) and family *Halococcaceae* (genus *Halococcus*), but also in the order *Haloferacales* for the family *Haloferacaceae* (genus *Haloferax*).

A better representation of diversity is found within the family *Halobacteriaceae* with four different genera (*Halobacterium*, *Halovarius*, *Halovenus*, and *Natronomonas*), despite the clear emphasis on *Halobacterium*.

Regarding the strains selected for astrobiology-related testing, most of them are type strains for their respective species (**Table 1**). The use of such type strains is usually recommended, as this contributes to more readily available data on the strains, and their accessibility facilitates experimental reproducibility and comparative studies (Antunes et al., 2017). One very clear and notable exception to this predominance of type strains is the use of *Halobacterium* sp. NRC-1. Although not a type strain, it has been deposited in microbial biological resource centers and is readily available; it has also been very extensively studied and has long been considered a model system for studies in Archaea, namely due to ease of culturing and genetic manipulation (Ng et al., 2000; McCready et al., 2005; Boubriak et al., 2008; Anderson et al., 2016; Dassarma et al., 2016). The status of this strain explains the very extensive number of astrobiological-based tests performed with it (**Table 1**).

The archaeal class *Halobacteria* and astrobiology: Astrobiological testing

Another insight of our analysis is the realization of how restricted and limited datasets on testing currently are. Indeed, most exposure experiments performed with the class *Halobacteria* focus exclusively on the effects of a single environmental factor (with a noticeable predominance of radiation-exposure tests; **Table 1** and **Supplementary Figure 2**). While the testing of one variable at a time is consistent with classical experimental techniques, the low number of tests with multiple variables is limiting. This gap in our understanding of the effect of simultaneous exposure to multiple extreme conditions is particularly apparent given that astrobiology-relevant settings beyond our planet include a combination of multiple environmental extremes (e.g., Merino et al., 2019; Simões and Antunes, 2021).

Space-exposure testing provides the ideal platform for obtaining a better perspective on this, but the number of tested

Halobacteria species remains low as most experiments have made use of simulated conditions only. However, it should be noted that these issues are not exclusive to *Halobacteria* but are rather major general limitations that apply to most microbial taxa. Several upcoming projects and recent publications have flagged members of the *Halobacteria* as key target organisms (Sgambati et al., 2020; Beblo-Vranesevic et al., 2022), so we can expect relevant new results in the near future.

Despite the generalized sparsity of data across the board, we have some good examples worth highlighting regarding completeness of the datasets. It should be relatively easy to complete a few datasets and cover some of the most obvious gaps. For example, testing the effects of exposure to UV currently has been performed for at least one representative of each family of the class *Halobacteria* except for the *Halorubraceae* (**Table 1** and **Supplementary Figure 2**). Likewise, the family *Haloarculaceae* is currently the only one that has not been subjected to exposure to multiple combined extremes (**Table 1** and **Supplementary Figure 2**).

Looking at the individual species/strain level (**Table 1**), we also have some good examples that are close to complete. *Halobacterium salinarum* NRC-1 has the most complete datasets, with very few gaps (12 out of 18 listed astrobiology tests). As previously mentioned, the latter is widely recognized as a model organism within the *Halobacteria* and an extensive amount of testing has been done with this strain. Nonetheless, regarding simulated single-factor astrobiology tests, to date there are no reported results for exposure to microgravity and several Martian chemistry and environment factors (M1 and M3–M5; **Table 1** and **Supplementary Figure 2**). Furthermore, this strain has been included in stratospheric exposure experiments, but has not been subjected to orbital testing. Given its status as model organism (McCready et al., 2005; Boubriak et al., 2008; Dassarma et al., 2016), this strain should be flagged as a priority for future tests.

The second good example we would like to highlight regarding completeness is *Halococcus morrhuae*. Although currently having less complete datasets than *Halobaculum salinarum* NRC-1 (10 out of 18 listed astrobiology tests; **Table 1**), it could be flagged as a further target for future completion. This species has the advantage of having already been used in orbital exposure tests, which are arguably the most challenging from a logistic and access point of view.

Natriabales is the only order of *Halobacteria* that does not have a representative already subjected to orbital testing. This order has the most incomplete datasets, as it is also absent from testing for most radiation types (with the notable exception of UV), most Martian chemistry and environmental factors (apart from testing with Martian soil analogs), desiccation, microgravity, and stratospheric testing (**Table 1** and **Supplementary Figure 2**).

Other important factors, not included on the overview table are exposure to low temperatures and low pH. Both conditions are seen as relevant from a Martian perspective but have not been directly listed as both are part of the standard characterization of new species. The number of species of *Halobacteria* capable of growth below 15°C or below pH 5 is notably small (currently consisting of 3 and 9 species, respectively). Nonetheless, within those, only *Halorubrum lacusprofundi* has been targeted for astrobiology testing.

On Earth, sites which are simultaneously acidic and highly saline are relatively rare (Mormile et al., 2009; Conner and Benison, 2013; Escudero et al., 2018), and studies on cold hypersaline sites remain largely underexplored (Oren, 2014b; Lamarche-Gagnon et al., 2015; Sapers et al., 2017). This might explain the reduced number of species within the *Halobacteria* that could be highlighted here. Further efforts in exploring these sites should be prioritized and could prove quite beneficial.

Looking at coverage based on the type of stressor also uncovered some unexpected areas where further work needs to be done (**Supplementary Figure 2**). Selected highlights of these gaps include exposure to single factors such as microgravity (missing data for *Halobacteriaceae*, *Halorubraceae*, and *Natrialbaceae*), simulated Martian atmosphere (with no representation for *Natrialbales*, *Haloferacales*, or *Haloarculaceae*) or Martian soil analogs/regolith (only tested for the *Natrialbaceae*). On the more complete datasets, tests looking at exposure to desiccation or to perchlorate are still needed to ensure that all families are represented.

Future work: Knowledge gaps and research opportunities

As outlined above, we can see that there are currently clear and significant gaps in the study of the class *Halobacteria* under the scope of Astrobiology. More in-depth testing is needed to cover the taxonomic diversity present in this archaeal group both at higher- and lower-ranking taxonomy levels (particularly in families with single-genus testing or with few tested species). The fact that species of this class have large pangenomes, such that even closely related taxa may be functionally quite distinct (Gaba et al., 2020; de la Haba et al., 2021), further emphasizes the need for more extensive testing.

Likewise, there is a need for more completed datasets on astrobiology-relevant stressors on members of this class, both as single factors and combined, particularly making use of space-exposure experiments. A notable (and surprising) gap here is on the study of the effects of different salts. Although their relationship with salt is at the core of the definition of halophiles, studies have been almost exclusively restricted to sodium chloride (NaCl). The astrobiological relevance of studying the interaction with different salts (including perchlorates), as well as looking into chaotropic and kosmotropic effects, is clear and has been well noted (Hallsworth et al., 2007; Ball and Hallsworth, 2015; Stevenson et al., 2015b; dC Rubin et al., 2017). Yet, most

species of *Halobacteria* still remain untested. Another noticeably neglected field here is the study of exposure to high pressure. Although technically more challenging, these settings would be very relevant in the context of the Martian subsurface or of the exooceans of the icy moons (Ueta and Sasaki, 2013; Jebbar et al., 2020; Olson et al., 2020).

From a physiological diversity perspective, new opportunities might also be opening up. Although members of the *Halobacteria* were typically seen as aerobic heterotrophs (Andrei et al., 2012) and reliant on the "salt-in" strategy (e.g., Kunte et al., 2002), they have consistently proven to be a lot more diverse (Müller et al., 2005; Oren, 2008; Gunde-Cimerman et al., 2018). An increasing number of species has been reported to use alternative anaerobic metabolic pathways, or even to show preference for anaerobic conditions or phototrophy (e.g., Antunes et al., 2008; Sorokin et al., 2017, 2018; Weber et al., 2017; Miralles-Robledillo et al., 2021).

There might be possible implications of this versatility in overall resilience and adaptability to astrobiology-relevant conditions. Microorganisms with "less typical" features could thus be interesting candidates for targeted future exposure studies and comparisons. An over-reliance on quickly growing or easier strains in detriment of those with higher relevance for astrobiology, such as anaerobic ones (Beblo-Vranesevic et al., 2020) or psychrotolerant/psychrophilic (Kish et al., 2012), should also be avoided despite the additional challenges and technical issues they might bring (Hamm et al., 2019; Cui and Dyall-Smith, 2021; Wang et al., 2021).

From an ecological perspective, it is relevant to note that the vast majority of tested species were not isolated from environments traditionally, or officially, flagged as terrestrial analog sites, despite some notable exceptions (e.g., *Hrr. lacusprofundi*). It is true that the identification of terrestrial analog sites can be seen as somewhat subjective and, in some ways, all hypersaline sites could potentially be seen as Martian analog sites (e.g., Mancinelli et al., 2004; Parro et al., 2011; Murray et al., 2012; Xiao et al., 2017). Nonetheless, it would be a useful priority to place further stress on: (i) testing members of the class *Halobacteria* from well-established analog sites and (ii) further identification and flagging of additional hypersaline sites as suitable terrestrial analogs. Likewise, the increase in cultivation-based studies specifically aimed toward isolating new strains and describing new species from such sites, namely polar, deep-sea, or magnesium-sulfate brines (Franzmann et al., 1988; Antunes et al., 2008; Albuquerque et al., 2012; Mou et al., 2012), would also prove beneficial and provide excellent new targets for exposure tests (DasSarma et al., 2017; Laye and DasSarma, 2018).

Furthermore, while single-species experiments are valuable, we should note that microbes do not generally live in isolation (e.g., Van Der Wielen and Heijs, 2007; Vikram et al., 2016; Abed et al., 2018; Uritskiy et al., 2021). Therefore, parallel experiments involving constructed consortia (as well as characterized natural consortia) of halophiles, could provide a useful complementary approach (e.g., Wadsworth et al., 2019; Leuko et al., 2020). Other approaches could (and should) focus on the distinction between survival and growth/multiplication in exposure experiments or the issue of long-term viability of microbes trapped in salt, a very relevant topic for halophilic research (Gramain et al., 2011; Jaakkola et al., 2014; Stan-Lotter and Fendrihan, 2015), but still under-explored from an astrobiological perspective.

Access to launching opportunities has been rather restricted in the past and has severely limited capability to perform experiments in space, but this sector is rapidly changing. The on-going rise of the space private sector, the increased use of micro-satellites, or even the recent launch of the Chinese Space Station, all provide further launch and testing opportunities that should be harnessed to address at least some of the gaps that we have identified. Perhaps there is also room for technical improvements on reducing some historical resistance toward high-salt experiments in space (due to increased risks of corrosion and clogging), allowing for, e.g., the testing of more live strains in liquid medium rather than just in lyophilized, freeze-dried form (Barravecchia et al., 2018; Rcheulishvili et al., 2020).

Finally, we highlight the increasing relevance of data and its analysis in the context of *Halobacteria* and astrobiology. Obviously, the selection and identification of model organisms for astrobiology should make use of a variety of data on their capabilities and characteristics. The increased relevance of the class *Halobacteria*, together with its significant on-going increase in the description of new species and genera, as well as in genomic and phenotypic data (e.g., DasSarma et al., 2010, 2019; Oren, 2014b; Antunes et al., 2017; Cui and Dyall-Smith, 2021; Lach et al., 2021), make it notoriously difficult to navigate this ocean of relevant yet dispersed information. Despite some efforts targeting the collection of data from this specific taxon (e.g., Pfeiffer et al., 2008; DasSarma et al., 2010; Ukani et al., 2011; Sharma et al., 2014; Loukas et al., 2018; Reimer et al., 2019), the absence of dedicated tools to quickly compare species or even to quickly identify strains with a combination of selected properties can be seen as a significant bottleneck in a range of fields of applied research with this group of archaea. Further community-wide efforts and the development of such tools should be thus seen as a key priority for researchers working with this taxon in the context of astrobiology and beyond.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Author contributions

AA conceived this study. J-HW conducted the data collection and processing. J-HW and AA conducted the core analysis and interpretation of the data and wrote the manuscript. All authors contributed to the analysis and interpretation of the data, revised the manuscript, and have read and approved the final manuscript.

Funding

This work of J-HW, MS, and AA is funded by the Science and Technology Development Fund, Macao SAR, China. AA and MS are supported by Faculty Research Grants of Macau University of Science and Technology's (FRG, AA: grant No. FRG-22-079-LPS; MS: grant No. FRG-22-RG-22-080-LPS).

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1023625/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Proportion of species and genera subjected to testing of astrobiology-relevant features within the class *Halobacteria* and each of its orders and families (see **Table 1** for full information).

SUPPLEMENTARY FIGURE 2
Overview of total number of reports for different types of tests of astrobiological relevance performed for each family of the *Halobacteria* (see **Table 1** for full information). R1, HZE (highly-charged and energetic particles); R2, γ-rays; R3, x-rays; R4, electron-beam irradiation; UV, ultra-violet radiation; vac, vacuum; temp, low temperature; M1, effect of $ClO4^-$ coupled with CO oxidation; M2, effect of $ClO4^-$; M3, use of CO as carbon source; M4, simulated Martian atmosphere (95% CO2, 2.7% N2, 1.6% Ar, 0.15% O2, and 370 ppm H_2O, 10^3 Pa); M5, Martian soil analogue; des, desiccation; G1, microgravity; G2, hypergravity; mult, multiple conditions; str, stratosphere testing; orb, orbital testing.

References

Abed, R. M. M., Kohls, K., Leloup, J., and de Beer, D. (2018). Abundance and diversity of aerobic heterotrophic microorganisms and their interaction with cyanobacteria in the oxic layer of an intertidal hypersaline cyanobacterial mat. *FEMS Microbiol. Ecol.* 94:fix183. doi: 10.1093/femsec/fix183

Abrevaya, X. C., Adamo, H. P., Cortó, N. E., and Mauas, P. J. D. (2008). The UV limits of life on extrasolar planets: An experiment with halophile *Archaea* bacteria at different UV doses. *Bol. de la Asociacion Argent. de Astronomia La Plata Argent.* 51, 3–6.

Abrevaya, X. C., Cortón, E., and Mauas, P. J. D. (2010). UV habitability and dM stars: An approach for evaluation of biological survival. *Proc. Int. Astron. Union* 5, 443–445. doi: 10.1017/S1743921309993073

Abrevaya, X. C., Cortón, E., and Mauas, P. J. D. (2011a). Flares and habitability. *Proc. Int. Astron. Union* 7, 405–409. doi: 10.1017/S1743921312005145

Abrevaya, X. C., Paulino-Lima, I. G., Galante, D., Rodrigues, F., Mauas, P. J., Cortón, E., et al. (2011b). Comparative survival analysis of *Deinococcus radiodurans* and the *Haloarchaea Natrialba magadii* and *Haloferax volcanii* exposed to vacuum ultraviolet irradiation. *Astrobiology* 11, 1034–1040. doi: 10.1089/ast.2011.0607

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

Anderson, I. J., DasSarma, P., Lucas, S., Copeland, A., Lapidus, A., Del Rio, T. G., et al. (2016). Complete genome sequence of the Antarctic *Halorubrum lacusprofundi* type strain ACAM 34. *Stand. Genom. Sci.* 11:70. doi: 10.1186/s40793-016-0194-2

Andrei, A. -Ş, Banciu, H. L., and Oren, A. (2012). Living with salt: Metabolic and phylogenetic diversity of *Archaea* inhabiting saline ecosystems. *FEMS Microbiol. Lett.* 330, 1–9. doi: 10.1111/j.1574-6968.2012.02526.x

Antunes, A., Olsson-Francis, K., and McGenity, T. J. (2020). Exploring deep-sea brines as potential terrestrial analogues of oceans in the icy moons of the outer solar system. *Curr. Issues Mol. Biol.* 38, 123–162. doi: 10.21775/cimb.038.123

Antunes, A., Simões, M. F., Grötzinger, S. W., Eppinger, J., Bragança, J., and Bajic, V. B. (2017). "Bioprospecting *Archaea*: Focus on Extreme Halophiles," in *Bioprospecting: Success, Potential and Constraints*, eds R. Paterson and N. Lima (Berlin: Springer International Publishing), 81–112.

Antunes, A., Taborda, M., Huber, R., Moissl, C., Nobre, M. F., and da Costa, M. S. (2008). Halorhabdus tiamatea sp. nov., a non-pigmented, extremely halophilic *Archaeon* from a deep-sea, hypersaline anoxic basin of the Red Sea, and emended description of the genus *Halorhabdus*. *Int. J. Syst. Evol. Microbiol.* 58, 215–220. doi: 10.1099/ijs.0.65316-0

Azua-Bustos, A., Urrejola, C., and Vicuña, R. (2012). Life at the dry edge: Microorganisms of the Atacama Desert. *FEBS Lett.* 586, 2939–2945. doi: 10.1016/j.febslet.2012.07.025

Baliga, N. S., Bjork, S. J., Bonneau, R., Pan, M., Iloanusi, C., Kottemann, M. C., et al. (2004). Systems level insights into the stress response to UV radiation in the halophilic *Archaeon Halobacterium* NRC-1. *Genome Res.* 14, 1025–1035. doi: 10.1101/gr.1993504

Ball, P., and Hallsworth, J. E. (2015). Water structure and chaotropicity: Their uses, abuses and biological implications. *Phys. Chem. Chem. Phys.* 17, 8297–8305. doi: 10.1039/C4CP04564E

Barravecchia, I., Cesari, C. D., Pyankova, O. V., Scebba, F., Mascherpa, M. C., Vecchione, A., et al. (2018). Pitting corrosion within bioreactors for space cell-culture contaminated by *Paenibacillus glucanolyticus*, a case report. *Microgravity Sci. Technol.* 30, 309–319. doi: 10.1007/s12217-018-9601-1

Baxter, B. K., Eddington, B., Riddle, M. R., Webster, T. N., and Avery, B. (2007). Great Salt Lake halophilic microorganisms as models for astrobiology: Evidence for desiccation tolerance and ultraviolet irradiation resistance. *Instrum. Methods Missions Astrobiol.* 6694:669415. doi: 10.1117/12.732621

Beblo-Vranesevic, K., Bohmeier, M., Schleumer, S., Rabbow, E., Perras, A. K., Moissl-Eichinger, C., et al. (2020). Impact of simulated martian conditions on (facultatively) anaerobic bacterial strains from different mars analogue sites. *Curr. Issues Mol. Biol.* 38, 103–122. doi: 10.21775/cimb.038.103

Beblo-Vranesevic, K., Piepjohn, J., Antunes, A., and Rettberg, P. (2022). Surviving Mars: New insights into the persistence of facultative anaerobic microbes from analogue sites. *Int. J. Astrobiol.* 21, 110–127. doi: 10.1017/S1473550422000064

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

Bibring, J.-P., Langevin, Y., Mustard, J. F., Poulet, F., Arvidson, R., Gendrin, A., et al. (2006). Global mineralogical and aqueous mars history derived from OMEGA/mars express data. *Science* 312, 400–404. doi: 10.1126/science.1122659

Boubriak, I., Ng, W. L., DasSarma, P., DasSarma, S., Crowley, D. J., and McCready, S. J. (2008). Transcriptional responses to biologically relevant doses of UV-B radiation in the model *Archaeon*, *Halobacterium* sp. NRC-1. *Saline Syst.* 4:13. doi: 10.1186/1746-1448-4-13

Bryanskaya, A. V., Berezhnoy, A. A., Rozanov, A. S., Peltek, S. E., and Pavlov, A. K. (2013). Adaptive capabilities of microorganisms of salt lakes of the Altai Region under conditions of early Mars. *Paleontol. J.* 47, 1089–1092. doi: 10.1134/S0031030113090050

Bryanskaya, A. V., Berezhnoy, A. A., Rozanov, A. S., Serdyukov, D. S., Malup, T. K., and Peltek, S. E. (2020). Survival of halophiles of Altai lakes under extreme environmental conditions: Implications for the search for Martian life. *Int. J. Astrobiol.* 19, 1–15. doi: 10.1017/S1473550419000077

Cavicchioli, R. (2002). Extremophiles and the Search for Extraterrestrial Life. *Astrobiology* 2, 281–292. doi: 10.1089/153110702762027862

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

Cockell, C. S., Catling, D. C., Davis, W. L., Snook, K., Kepner, R. L., Lee, P., et al. (2000). The ultraviolet environment of mars: Biological implications past, present, and future. *Icarus* 146, 343–359. doi: 10.1006/icar.2000.6393

Coker, J. A., DasSarma, P., Kumar, J., Müller, J. A., and DasSarma, S. (2007). Transcriptional profiling of the model *Archaeon Halobacterium* sp. NRC-1: Responses to changes in salinity and temperature. *Saline Syst.* 3:6. doi: 10.1186/1746-1448-3-6

Confalonieri, F., and Sommer, S. (2011). Bacterial and *Archaea*l resistance to ionizing radiation. *J. Phys.* 261:012005. doi: 10.1088/1742-6596/261/1/012005

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

Cottin, H., Kotler, J. M., Billi, D., Cockell, C., Demets, R., Ehrenfreund, P., et al. (2017). Space as a tool for astrobiology: Review and recommendations for experimentations in earth orbit and beyond. *Space Sci. Rev.* 209, 83–181. doi: 10.1007/s11214-017-0365-5

Crowley, D. J., Boubriak, I., Berquist, B. R., Clark, M., Richard, E., Sullivan, L., et al. (2006). The uvrA, uvrB and uvrC genes are required for repair of ultraviolet light induced DNA photoproducts in *Halobacterium* sp. NRC-1. *Saline Syst.* 2:11. doi: 10.1186/1746-1448-2-11

Cui, H.-L., and Dyall-Smith, M. L. (2021). Cultivation of halophilic *Archaea* (class *Halobacteria*) from thalassohaline and athalassohaline environments. *Mar. Life Sci. Technol.* 3, 243–251. doi: 10.1007/s42995-020-00087-3

Dartnell, L. R., and Burchell, M. J. (2009). Survey on astrobiology research and teaching activities within the United Kingdom. *Astrobiology* 9, 717–730. doi: 10.1089/ast.2009.0348

DasSarma, P., Capes, M. D., and DasSarma, S. (2019). "Comparative Genomics of Halobacterium Strains From Diverse Locations," in *Microbial Diversity in the Genomic Era*, eds S. Das and H. R. Dash (Cambridge: Academic Press), 285–322.

DasSarma, P., Laye, V., Harvey, J., Reid, C., Shultz, J., Yarborough, A., et al. (2017). Survival of halophilic *Archaea* in Earth's cold stratosphere. *Int. J. Astrobiol.* 16, 321–327.

Dassarma, P., Tuel, K., Nierenberg, S. D., Phillips, T., Pecher, W. T., and Dassarma, S. (2016). Inquiry-driven teaching & learning using the *Archaeal* microorganism *Halobacterium* NRC-1. *Am. Biol. Teach.* 78, 7–13. doi: 10.1525/abt.2016.78.1.7

DasSarma, S. (2006). Extreme halophiles are models for astrobiology. *Microbe Mag.* 1, 120–126. doi: 10.1128/microbe.1.120.1

DasSarma, S., and DasSarma, P. (2017). "Halophiles," in *eLS*, ed. John Wiley & Sons, Ltd (Hoboken, NJ: John Wiley & Sons, Ltd), 1–13.

DasSarma, S., DasSarma, P., Laye, V. J., and Schwieterman, E. W. (2020). Extremophilic models for astrobiology: *Haloarchaea*l survival strategies and pigments for remote sensing. *Extremophiles* 24, 31–41. doi: 10.1007/s00792-019-01126-3

DasSarma, S. L., Capes, M. D., DasSarma, P., and DasSarma, S. (2010). HaloWeb: The *Haloarchaea*l genomes database. *Saline Syst.* 6:12. doi: 10.1186/1746-1448-6-12

dC Rubin, S. S., Marín, I., Gómez, M. J., Morales, E. A., Zekker, I., San Martín-Uriz, P., et al. (2017). Prokaryotic diversity and community composition in the Salar de Uyuni, a large scale, chaotropic salt flat. *Environ. Microbiol.* 19, 3745–3754. doi: 10.1111/1462-2920.13876

de la Haba, R. R., 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

Des Marais, D. J., Nuth, J. A., Allamandola, L. J., Boss, A. P., Farmer, J. D., Hoehler, T. M., et al. (2008). The NASA astrobiology roadmap. *Astrobiology* 8, 715–730. doi: 10.1089/ast.2008.0819

Deutsch, A. N., Levy, J. S., Dickson, J. L., and Head, J. W. (2022). Daily and seasonal processes shape hydrological activity and detectability of moisture in Antarctic, Mars-analog soils. *Icarus* 380:114990. doi: 10.1016/j.icarus.2022.114990

DeVeaux, L. C., Müller, J. A., Smith, J., Petrisko, J., Wells, D. P., and DasSarma, S. (2007). Extremely radiation-resistant mutants of a halophilic *Archaeon* with increased single-stranded DNA-binding protein (RPA) gene expression. *Radiat. Res.* 168, 507–514. doi: 10.1667/rr0935.1

Dornmayr-Pfaffenhuemer, M., Legat, A., Schwimbersky, K., Fendrihan, S., and Stan-Lotter, H. (2011). Responses of *Haloarchaea* to simulated microgravity. *Astrobiology* 11, 199–205. doi: 10.1089/ast.2010.0536

Dypvik, H., Hellevang, H., Krzesińska, A., Sætre, C., Viennet, J.-C., Bultel, B., et al. (2021). The planetary terrestrial analogues library (PTAL) – An exclusive lithological selection of possible martian earth analogues. *Planet. Space Sci.* 208:105339. doi: 10.1016/j.pss.2021.105339

Edbeib, M. F., Wahab, R. A., and Huyop, F. (2016). Halophiles: Biology, adaptation, and their role in decontamination of hypersaline environments. *World J. Microbiol. Biotechnol.* 32:135. doi: 10.1007/s11274-016-2081-9

Ehlmann, B. L., Swayze, G. A., Milliken, R. E., Mustard, J. F., Clark, R. N., Murchie, S. L., et al. (2016). Discovery of alunite in cross crater, terra sirenum, mars: Evidence for acidic, sulfurous waters. *Am. Mineral.* 101, 1527–1542. doi: 10.2138/am-2016-5574

Escudero, L., Oetiker, N., Gallardo, K., Tebes-Cayo, C., Guajardo, M., Nuñez, C., et al. (2018). A thiotrophic microbial community in an acidic brine lake in Northern Chile. *Antonie Leeuwenhoek* 111, 1403–1419. doi: 10.1007/s10482-018-1087-8

Evans, J. J., Gygli, P. E., McCaskill, J., and DeVeaux, L. C. (2018). Divergent roles of RPA homologs of the model *Archaeon Halobacterium salinarum* in survival of DNA damage. *Genes* 9:223. doi: 10.3390/genes9040223

Fackrell, L. E., Schroeder, P. A., Thompson, A., Stockstill-Cahill, K., and Hibbitts, C. A. (2021). Development of martian regolith and bedrock simulants: Potential and limitations of martian regolith as an in-situ resource. *Icarus* 354:114055. doi: 10.1016/j.icarus.2020.114055

Fairén, A. G., Davila, A. F., Lim, D., Bramall, N., Bonaccorsi, R., Zavaleta, J., et al. (2010). Astrobiology through the ages of Mars: The study of terrestrial analogues to understand the habitability of Mars. *Astrobiology* 10, 821–843. doi: 10.1089/ast.2009.0440

Fendrihan, S., Bérces, A., Lammer, H., Musso, M., Rontó, G., Polacsek, T. K., et al. (2009). Investigating the effects of simulated martian ultraviolet radiation on *Halococcus dombrowskii* and other extremely halophilic *Archaebacteria*. *Astrobiology* 9, 104–112. doi: 10.1089/ast.2007.0234

Fendrihan, S., Grosbacher, M., and Stan-Lotter, H. (2010). *Response of the Extremely Halophilic Halococcus Dombrowskii Strain H4 to UV Radiation and Space Conditions in the EXPOSE -ADAPT Project on the International Space Station*. Munich: EGU General Assembly.

Feshangsaz, N., Semsarha, F., Tackallou, S. H., Nazmi, K., Monaghan, E. P., Riedo, A., et al. (2020). Survival of the halophilic *Archaeon Halovarius luteus* after desiccation, simulated martian UV radiation and vacuum in comparison to *Bacillus atrophaeus*. *Orig. Life Evol. Biosph.* 50, 157–173. doi: 10.1007/s11084-020-09597-7

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.* 11:342. doi: 10.3389/fmicb.2020.00324

Franzmann, P. D., Stackebrandt, E., Sanderson, K., Volkman, J. K., Cameron, D. E., Stevenson, P. L., et al. (1988). *Halobacterium lacusprofundi* sp. nov., a Halophilic bacterium isolated from deep lake, Antarctica. *Syst. Appl. Microbiol.* 11, 20–27. doi: 10.1016/S0723-2020(88)80044-4

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

García-Descalzo, L., Parro, V., García-Villadangos, M., Cockell, C. S., Moissl-Eichinger, C., Perras, A., et al. (2019). Microbial markers profile in anaerobic Mars analogue environments using the LDChip (Life Detector Chip) antibody microarray core of the SOLID (Signs of Life Detector) platform. *Microorganisms* 7:365. doi: 10.3390/microorganisms7090365

Genova, A., Smith, D. E., Canup, R., Hurford, T., Goossens, S., Mazarico, E., et al. (2022). Geodetic investigations of the mission concept MAGIC to reveal Callisto's internal structure. *Acta Astronaut.* 195, 68–76. doi: 10.1016/j.actaastro.2022.02.013

Goordial, J. (2021). Cryomicrobial ecology: Still much to learn about life left out in the cold. *mSystems* 6:e0085221. doi: 10.1128/mSystems.00852-21

Gramain, A., Díaz, G. C., Demergasso, C., Lowenstein, T. K., and McGenity, T. J. (2011). *Archaea*l diversity along a subterranean salt core from the Salar Grande (Chile). *Environ. Microbiol.* 13, 2105–2121. doi: 10.1111/j.1462-2920.2011.02435.x

Grasset, O., Castillo-Rogez, J., Guillot, T., Fletcher, L. N., and Tosi, F. (2017). Water and volatiles in the outer solar system. *Space Sci. Rev.* 212, 835–875. doi: 10.1007/s11214-017-0407-z

Greenwood, J. P., Karato, S.-I., Vander Kaaden, K. E., Pahlevan, K., and Usui, T. (2018). Water and volatile inventories of mercury, venus, the moon, and Mars. *Space Sci. Rev.* 214:92. doi: 10.1007/s11214-018-0526-1

Gunde-Cimerman, N., Plemenitaš, A., and Oren, 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

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 Leeuwenhoek* 109, 565–587. doi: 10.1007/s10482-016-0660-2

Hallsworth, J. E. (2021). Mars' surface is not universally biocidal. *Environ. Microbiol.* 23, 3345–3350. doi: 10.1111/1462-2920.15494

Hallsworth, J. E., Mancinelli, R. L., Conley, C. A., Dallas, T. D., Rinaldi, T., Davila, A. F., et al. (2021b). Astrobiology of life on Earth. *Environ. Microbiol.* 23, 3335–3344. doi: 10.1111/1462-2920.15499

Hallsworth, J. E., Koop, T., Dallas, T. D., Zorzano, M.-P., Burkhardt, J., Golyshina, O. V., et al. (2021a). Water activity in Venus's uninhabitable clouds and other planetary atmospheres. *Nat. Astron.* 5, 665–675. doi: 10.1038/s41550-021-01391-3

Hallsworth, J. E., Yakimov, M. M., Golyshin, P. N., Gillion, J. L. M., D'Auria, G., De Lima Alves, F., et al. (2007). Limits of life in $MgCl_2$-containing environments: Chaotropicity defines the window. *Environ. Microbiol.* 9, 801–813. doi: 10.1111/j.1462-2920.2006.01212.x

Hamm, J. N., Erdmann, S., Eloe-Fadrosh, E. A., Angeloni, A., Zhong, L., Brownlee, C., et al. (2019). Unexpected host dependency of Antarctic *Nanohaloarchaeota*. *Proc. Natl. Acad. Sci. U.S.A.* 116, 14661–14670. doi: 10.1073/pnas.1905179116

Harris, J. K., Cousins, C. R., Gunn, M., Grindrod, P. M., Barnes, D., Crawford, I. A., et al. (2015). Remote detection of past habitability at Mars-analogue hydrothermal alteration terrains using an ExoMars Panoramic Camera emulator. *Icarus* 252, 284–300. doi: 10.1016/j.icarus.2015.02.004

Hayes, A. G. (2016). The Lakes and Seas of Titan. *Annu. Rev. Earth Planet. Sci.* 44, 57–83. doi: 10.1146/annurev-earth-060115-012247

Horneck, G. (1999). European activities in exobiology in Earth orbit: Results and perspectives. *Adv. Space Res.* 23, 381–386. doi: 10.1016/S0273-1177(99)00061-7

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

Horneck, G., Walter, N., Westall, F., Grenfell, J. L., Martin, W. F., Gomez, F., et al. (2016). AstRoMap European astrobiology roadmap. *Astrobiology* 16, 201–243. doi: 10.1089/ast.2015.1441

Jaakkola, S. T., Zerulla, K., Guo, Q., Liu, Y., Ma, H., Yang, C., et al. (2014). Halophilic *Archaea* cultivated from surface sterilized Middle-Late Eocene rock salt Are Polyploid. *PLoS One* 9:e110533. doi: 10.1371/journal.pone.0110533

Jebbar, M., Hickman-Lewis, K., Cavalazzi, B., Taubner, R.-S., Rittmann, S. 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

Jia, X., Kivelson, M. G., Khurana, K. K., and Kurth, W. S. (2018). Evidence of a plume on Europa from Galileo magnetic and plasma wave signatures. *Nat. Astron.* 2, 459–464. doi: 10.1038/s41550-018-0450-z

Johnson, A. P., Pratt, L. M., Vishnivetskaya, T., Pfiffner, S., Bryan, R. A., Dadachova, E., et al. (2011). Extended survival of several organisms and amino acids under simulated martian surface conditions. *Icarus* 211, 1162–1178.

Karan, R., DasSarma, P., Balcer-Kubiczek, E., Weng, R. R., Liao, C.-C., Goodlett, D. R., et al. (2014). Bioengineering radioresistance by overproduction of RPA, a mammalian-type single-stranded DNA-binding protein, in a halophilic *Archaeon*. *Appl. Microbiol. Biotechnol.* 98, 1737–1747. doi: 10.1007/s00253-013-5368-x

Kereszturi, A., and Keszthelyi, Z. (2013). Astrobiological implications of chaos terrains on Europa to help targeting future missions. *Planet. Space Sci.* 77, 74–90. doi: 10.1016/j.pss.2012.08.028

King, G. M. (2015). Carbon monoxide as a metabolic energy source for extremely halophilic microbes: Implications for microbial activity in Mars regolith. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4465–4470. doi: 10.1073/pnas.1424989112

Kish, A., Griffin, P. L., Rogers, K. L., Fogel, M. L., Hemley, R. J., and Steele, A. (2012). High-pressure tolerance in *Halobacterium salinarum* NRC-1 and other non-piezophilic prokaryotes. *Extremophiles* 16, 355–361. doi: 10.1007/s00792-011-0418-8

Kish, A., Kirkali, G., Robinson, C., Rosenblatt, R., Jaruga, P., Dizdaroglu, M., et al. (2009). Salt shield: Intracellular salts provide cellular protection against ionizing radiation in the halophilic *Archaeon*, *Halobacterium salinarum* NRC-1. *Environ. Microbiol.* 11, 1066–1078. doi: 10.1111/j.1462-2920.2008.01828.x

Knoll, A. H., and Grotzinger, J. (2006). Water on Mars and the prospect of martian life. *Elements* 2, 169–173. doi: 10.2113/gselements.2.3.169

Koike, J. (1991). Fundamental questions concerning the contamination of other planets with terrestrial micro-organisms carried by space-probes. *J. Space Technol. Sci.* 7, 9–14. doi: 10.11230/jsts.7.2_9

Koike, J., and Oshima, T. (1993). Planetary quarantine in the solar system. survival rates of some terrestrial organisms under simulated space conditions by proton irradiation. *Acta Astronaut.* 29, 629–632. doi: 10.1016/0094-5765(93)90080-G

Koike, J., Oshima, T., Kobayashi, K., and Kawasaki, Y. (1995). Studies in the search for life on Mars. *Adv. Space Res.* 15, 211–214. doi: 10.1016/S0273-1177(99)80086-6

Koike, J., Oshima, T., Koike, K. A., Taguchi, H., Tanaka, R., Nishimura, K., et al. (1992). Survival rates of some terrestrial microorganisms under simulated space conditions. *Adv. Space Res.* 12, 271–274. doi: 10.1016/0273-1177(92)90182-W

Koschnitzki, D., Moeller, R., Leuko, S., Przybyla, B., Beblo-Vranesevic, K., Wirth, R., et al. (2021). Questioning the radiation limits of life: Ignicoccus hospitalis between replication and VBNC. *Arch. Microbiol.* 203, 1299–1308. doi: 10.1007/s00203-020-02125-1

Kottemann, M., Kish, A., Iloanusi, C., Bjork, S., and DiRuggiero, J. (2005). Physiological responses of the halophilic *Archaeon Halobacterium* sp. strain NRC1 to desiccation and gamma irradiation. *Extremophiles* 9, 219–227. doi: 10.1007/s00792-005-0437-4

Kunte, H. J., Trüper, H. G., and Stan-Lotter, H. (2002). "Halophilic Microorganisms," in *Astrobiology: The Quest for the Conditions of Life*, eds G. Horneck and C. Baumstark-Khan (Berlin: Springer), 185–200.

Lach, J., Jêcz, P., Strapagiel, D., Matera-Witkiewicz, A., and Stączek, P. (2021). The methods of digging for "gold" within the salt: Characterization of Halophilic prokaryotes and identification of their valuable biological products using sequencing and genome mining tools. *Genes* 12:1756. doi: 10.3390/genes12111756

Lamarche-Gagnon, G., Comery, R., Greer, C. W., and Whyte, L. G. (2015). Evidence of in situ microbial activity and sulphidogenesis in perennially sub-0 °C and hypersaline sediments of a high Arctic permafrost spring. *Extremophiles* 19, 1–15. doi: 10.1007/s00792-014-0703-4

Lantz, C., Poulet, F., Loizeau, D., Riu, L., Pilorget, C., Carter, J., et al. (2020). Planetary Terrestrial Analogues Library project: 1. characterization of samples by

near-infrared point spectrometer. *Planet. Space Sci.* 189:104989. doi: 10.1016/j.pss.2020.104989

Lasne, J., Noblet, A., Szopa, C., Navarro-González, R., Cabane, M., Poch, O., et al. (2016). Oxidants at the surface of Mars: A review in light of recent exploration results. *Astrobiology* 16, 977–996. doi: 10.1089/ast.2016.1502

Lauro, S. E., Pettinelli, E., Caprarelli, G., Guallini, L., Rossi, A. P., Mattei, E., et al. (2021). Multiple subglacial water bodies below the south pole of Mars unveiled by new MARSIS data. *Nat. Astron.* 5, 63–70. doi: 10.1038/s41550-020-1200-6

Laye, V. J., and DasSarma, S. (2018). An antarctic extreme halophile and its polyextremophilic enzyme: Effects of perchlorate salts. *Astrobiology* 18, 412–418. doi: 10.1089/ast.2017.1766

Laye, V. J., Karan, R., Kim, J.-M., Pecher, W. T., DasSarma, P., and DasSarma, S. (2017). Key amino acid residues conferring enhanced enzyme activity at cold temperatures in an Antarctic polyextremophilic β-galactosidase. *Proc. Natl. Acad. Sci. U.S.A.* 114, 12530–12535. doi: 10.1073/pnas.1711542114

Leuko, S., Bohmeier, M., Hanke, F., Böettger, U., Rabbow, E., Parpart, A., et al. (2017). On the stability of deinoxanthin exposed to mars conditions during a long-term space mission and implications for biomarker detection on other planets. *Front. Microbiol.* 8:1680. doi: 10.3389/fmicb.2017.01680

Leuko, S., Domingos, C., Parpart, A., Reitz, G., and Rettberg, P. (2015). The survival and resistance of *Halobacterium salinarum* NRC-1, *Halococcus hamelinensis*, and *Halococcus morrhuae* to simulated outer space solar radiation. *Astrobiology* 15, 987–997. doi: 10.1089/ast.2015.1310

Leuko, S., Neilan, B. A., Burns, B. P., Walter, M. R., and Rothschild, L. J. (2011). Molecular assessment of UVC radiation-induced DNA damage repair in the stromatolitic halophilic *Archaeon*, *Halococcus hamelinensis*. *J. Photochem. Photobiol. B* 102, 140–145. doi: 10.1016/j.jphotobiol.2010.10.002

Leuko, S., and Rettberg, P. (2017). The Effects of HZE Particles, γ and X-ray Radiation on the Survival and Genetic Integrity of *Halobacterium salinarum* NRC-1, *Halococcus hamelinensis*, and *Halococcus morrhuae*. *Astrobiology* 17, 110–117. doi: 10.1089/ast.2015.1458

Leuko, S., Rettberg, P., Pontifex, A. L., and Burns, B. P. (2014). On the response of halophilic *Archaea* to space conditions. *Life* 4, 66–76. doi: 10.3390/life4010066

Leuko, S., Stan-Lotter, H., Lamers, G., Sjöström, S., Rabbow, E., Parpart, A., et al. (2020). "Resistance of the *Archaeon* Halococcus morrhuae and the Biofilm-Forming Bacterium Halomonas muralis to Exposure to Low Earth Orbit for 534 Days," in *Extremophiles as Astrobiological Models*, eds J. Seckbach and H. Stan-Lotter (Hoboken: Wiley-Scrivener), 221–236.

Leuko, S., Weidler, G., Radax, C., Legat, A., Kömle, N., Kargl, G., et al. (2002). "Examining the physico-chemical resistance of Halobacteria with the live-dead kit, following exposure to simulated Martian atmospheric conditions and heat," in *Proceedings of the First European Workshop on Exo-Astrobiology*, ed. H. Lacoste (Paris: ESA), 473–474.

Lillis, R. J., and Brain, D. A. (2013). Nightside electron precipitation at Mars: Geographic variability and dependence on solar wind conditions. *J. Geophys. Res.* 118, 3546–3556. doi: 10.1002/jgra.50171

Loukas, A., Kappas, I., and Abatzopoulos, T. J. (2018). HaloDom: A new database of halophiles across all life domains. *J. Biol. Res. Thessaloniki* 25:2. doi: 10.1186/s40709-017-0072-0

Mancinelli, R. L. (2015). The affect of the space environment on the survival of *Halorubrum chaoviator* and *Synechococcus* (Nägeli): Data from the Space Experiment OSMO on EXPOSE-R. *Int. J. Astrobiol.* 14, 123–128. doi: 10.1017/S147355041400055X

Mancinelli, R. L., Fahlen, T. F., Landheim, R., and Klovstad, M. R. (2004). Brines and evaporites: Analogs for martian life. *Adv. Space Res.* 33, 1244–1246. doi: 10.1016/j.asr.2003.08.034

Mancinelli, R. L., White, M. R., and Rothschild, L. J. (1998). Biopan-survival I: Exposure of the osmophiles *Synechococcus* sp.(Nageli) and *Haloarcula* sp. to the space environment. *Adv. Space Res.* 22, 327–334. doi: 10.1016/S0273-1177(98)00189-6

Marion, G. M., Fritsen, C. H., Eicken, H., and Payne, M. C. (2003). The search for life on Europa: Limiting environmental factors, potential habitats, and earth analogues. *Astrobiology* 3, 785–811. doi: 10.1089/153110703322736105

Martin, E., Reinhardt, R., Baum, L., Becker, M., Shaffer, J., and Kokjohn, T. (2000). The effects of ultraviolet radiation on the moderate halophile *Halomonas elongata* and the extreme halophile *Halobacterium salinarum*. *Can. J. Microbiol.* 46, 180–187.

Martínez-Espinosa, R. M., Richardson, D. J., and Bonete, M. J. (2015). Characterisation of chlorate reduction in the *haloarchaeon Haloferax mediterranei*. *Biochim. Biophys. Acta Gen. Subj.* 1850, 587–594. doi: 10.1016/j.bbagen.2014.12.011

Martins, Z., Cottin, H., Kotler, J. M., Carrasco, N., Cockell, C. S., de la Torre Noetzel, R., et al. (2017). Earth as a tool for astrobiology—a european perspective. *Space Sci. Rev.* 209, 43–81. doi: 10.1007/s11214-017-0369-1

Martín-Torres, F. J., Zorzano, M.-P., Valentín-Serrano, P., Harri, A.-M., Genzer, M., Kemppinen, O., et al. (2015). Transient liquid water and water activity at Gale crater on Mars. *Nat. Geosci.* 8, 357–361. doi: 10.1038/ngeo2412

Mastascusa, V., Romano, I., Di Donato, P., Poli, A., Della Corte, V., Rotundi, A., et al. (2014). Extremophiles survival to simulated space conditions: An astrobiology model study. *Orig. Life Evol. Biosph.* 44, 231–237. doi: 10.1007/s11084-014-9397-y

McCready, S. (1996). The repair of ultraviolet light-induced DNA damage in the halophilic archaebacteria, *Halobacterium cutirubrum*, *Halobacterium halobium* and *Haloferax volcanii*. *Mutat. Res. DNA Repair* 364, 25–32. doi: 10.1016/0921-8777(96)00018-3

McCready, S., Müller, J. A., Boubriak, I., Berquist, B. R., Ng, W. L., and DasSarma, S. (2005). UV irradiation induces homologous recombination genes in the model *Archaeon*. Halobacterium sp. NRC-1. *Saline Syst.* 1:3. doi: 10.1186/1746-1448-1-3

McDuff, S., King, G., Neupane, S., and Myers, M. (2016). Isolation and characterization of extremely halophilic CO-oxidizing *Euryarchaeota* from hypersaline cinders, sediments and soils and description of a novel CO oxidizer, *Haloferax namakaokahaiae* Mke2. 3T, sp. nov. *FEMS Microbiol. Ecol.* 92:fiw028. doi: 10.1093/femsec/fiw028

Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780

Miralles-Robledillo, J. M., Bernabeu, E., Giani, M., Martínez-Serna, E., Martínez-Espinosa, R. M., and Pire, C. (2021). Distribution of denitrification among *Haloarchaea*: A Comprehensive Study. *Microorganisms* 9:1669. doi: 10.3390/microorganisms9081669

Moeller, R., Reitz, G., Douki, T., Cadet, J., Horneck, G., and Stan-Lotter, H. (2010). UV photoreactions of the extremely haloalkaliphilic *Euryarchaeon Natronomonas pharaonis*. *FEMS Microbiol. Ecol.* 73, 271–277. doi: 10.1111/j.1574-6941.2010.00893.x

Moissl-Eichinger, C. (2011). *Archaea* in artificial environments: Their presence in global spacecraft clean rooms and impact on planetary protection. *ISME J.* 5, 209–219. doi: 10.1038/ismej.2010.124

Moissl-Eichinger, C., Cockell, C., and Rettberg, P. (2016). Venturing into new realms? microorganisms in space. *FEMS Microbiol. Rev.* 40, 722–737. doi: 10.1093/femsre/fuw015

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

Mou, Y.-Z., Qiu, X.-X., Zhao, M.-L., Cui, H.-L., Oh, D., and Dyall-Smith, M. L. (2012). *Halohasta litorea* gen. nov. sp. nov., and *Halohasta litchfieldiae* sp. nov., isolated from the Daliang aquaculture farm, China and from Deep Lake, Antarctica, respectively. *Extremophiles* 16, 895–901. doi: 10.1007/s00792-012-0485-5

Müller, V., Spanheimer, R., and Santos, H. (2005). Stress response by solute accumulation in *Archaea*. *Curr. Opin. Microbiol.* 8, 729–736. doi: 10.1016/j.mib.2005.10.011

Murray, A. E., Kenig, F., Fritsen, C. H., McKay, C. P., Cawley, K. M., Edwards, R., et al. (2012). Microbial life at −13 °C in the brine of an ice-sealed Antarctic lake. *Proc. Natl. Acad. Sci.U.S.A.* 109, 20626–20631. doi: 10.1073/pnas.1208607109

Musilova, M., Wright, G., Ward, J. M., and Dartnell, L. R. (2015). Isolation of radiation-resistant bacteria from Mars analog Antarctic Dry valleys by preselection, and the correlation between radiation and desiccation resistance. *Astrobiology* 15, 1076–1090. doi: 10.1089/ast.2014.1278

Myers, M. R., and King, G. (2020). *Halobacterium bonnevillei* sp. nov., *Halobaculum saliterrae* sp. nov. and *Halovenus carboxidivorans* sp. nov., three novel carbon monoxide-oxidizing *Halobacteria* from saline crusts and soils. *Int. J. Syst. Evol. Microbiol.* 70, 4261–4268. doi: 10.1099/ijsem.0.004282

Myers, M. R., and King, G. M. (2017). Perchlorate-coupled carbon monoxide (CO) oxidation: Evidence for a plausible microbe-mediated reaction in martian brines. *Front. Microbiol.* 8:2571. doi: 10.3389/fmicb.2017.02571

Myers, M. R., and King, G. M. (2019). Addendum: Perchlorate-coupled carbon monoxide (CO) Oxidation: Evidence for a plausible microbe-mediated reaction in martian brines. *Front. Microbiol.* 10:1158. doi: 10.3389/fmicb.2019.01158

Navarro-González, R., Vargas, E., de la Rosa, J., Raga, A. C., and McKay, C. P. (2010). Reanalysis of the viking results suggests perchlorate and organics at midlatitudes on Mars. *J. Geophys. Res.* 115:E12010. doi: 10.1029/2010JE003599

Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., et al. (2000). Genome sequence of *Halobacterium* species NRC-1. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12176–12181. doi: 10.1073/pnas.190337797

Nickerson, C. A., Medina-Colorado, A. A., Barrila, J., Poste, G., and Ott, C. M. (2022). A vision for spaceflight microbiology to enable human health and habitat sustainability. *Nat. Microbiol.* 7, 471–474. doi: 10.1038/s41564-021-01015-6

Nimmo, F., and Pappalardo, R. T. (2016). Ocean worlds in the outer solar system. *J. Geophys. Res.* 121, 1378–1399. doi: 10.1002/2016JE005081

Noe Dobrea, E. Z., and Swayze, G. (2010). "Acid Pedogenesis on Mars? Evidence for Top-Down Alteration on Mars from CRISM and HiRISE Data," in *41st Annual Lunar and Planetary Science Conference LPI Contribution No. 1533*, (Woodlands, TX), 2620.

Ojha, L., McEwen, A., Dundas, C., Byrne, S., Mattson, S., Wray, J., et al. (2014). HiRISE observations of Recurring Slope Lineae (RSL) during southern summer on Mars. *Icarus* 231, 365–376. doi: 10.1016/j.icarus.2013.12.021

Okeke, B. C., Giblin, T., and Frankenberger, W. T. (2002). Reduction of perchlorate and nitrate by salt tolerant bacteria. *Environ. Pollut.* 118, 357–363. doi: 10.1016/S0269-7491(01)00288-3

Olson, S. L., Jansen, M., and Abbot, D. S. (2020). Oceanographic considerations for exoplanet life detection. *Astrophys. J.* 895:19. doi: 10.3847/1538-4357/ab88c9

Oren, A. (2008). Microbial life at high salt concentrations: Phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2

Oren, A. (2014a). Halophilic *Archaea* on earth and in space: Growth and survival under extreme conditions. *Philos. Trans. R. Soc. A* 372:20140194. doi: 10.1098/rsta.2014.0194

Oren, A. (2014b). Taxonomy of halophilic *Archaea*: Current status and future challenges. *Extremophiles* 18, 825–834. doi: 10.1007/s00792-014-0654-9

Oren, A., Elevi Bardavid, R., and Mana, L. (2014). Perchlorate and halophilic prokaryotes: Implications for possible halophilic life on Mars. *Extremophiles* 18, 75–80. doi: 10.1007/s00792-013-0594-9

Parro, V., de Diego-Castilla, G., Moreno-Paz, M., Blanco, Y., Cruz-Gil, P., Rodríguez-Manfredi, J. A., et al. (2011). A microbial Oasis in the Hypersaline Atacama subsurface discovered by a life detector chip: Implications for the search for life on Mars. *Astrobiology* 11, 969–996. doi: 10.1089/ast.2011.0654

Pätzold, M., Häusler, B., Tyler, G. L., Andert, T., Asmar, S. W., Bird, M. K., et al. (2016). Mars express 10 years at Mars: Observations by the Mars express radio science experiment (MaRS). *Planet. Space Sci.* 127, 44–90. doi: 10.1016/j.pss.2016.02.013

Peeters, Z., Vos, D., Ten Kate, I., Selch, F., Van Sluis, C., Sorokin, D. Y., et al. (2010). Survival and death of the *Haloarchaeon Natronorubrum* strain HG-1 in a simulated martian environment. *Adv. Space Res.* 46, 1149–1155. doi: 10.1016/j.asr.2010.05.025

Pfeiffer, F., Broicher, A., Gillich, T., Klee, K., Mejía, J., Rampp, M., et al. (2008). Genome information management and integrated data analysis with HaloLex. *Arch. Microbiol.* 190, 281–299. doi: 10.1007/s00203-008-0389-z

Preston, L. J., and Dartnell, L. R. (2014). Planetary habitability: Lessons learned from terrestrial analogues. *Int. J. Astrobiol.* 13, 81–98. doi: 10.1017/S1473550413000396

Quinn, R. C., Martucci, H. F. H., Miller, S. R., Bryson, C. E., Grunthaner, F. J., and Grunthaner, P. J. (2013). Perchlorate radiolysis on Mars and the origin of martian soil reactivity. *Astrobiology* 13, 515–520. doi: 10.1089/ast.2013.0999

Race, M., Denning, K., Bertka, C. M., Dick, S. J., Harrison, A. A., Impey, C., et al. (2012). Astrobiology and society: Building an interdisciplinary research community. *Astrobiology* 12, 958–965. doi: 10.1089/ast.2011.0723

Ramstad, R., Barabash, S., Futaana, Y., Nilsson, H., and Holmström, M. (2018). Ion escape from Mars through time: An extrapolation of atmospheric loss based on 10 years of Mars express measurements. *J. Geophys. Res.* 123, 3051–3060. doi: 10.1029/2018JE005727

Rcheulishvili, N., Zhang, Y., Papukashvili, D., and Deng, Y.-L. (2020). Survey and evaluation of spacecraft-associated aluminum-degrading microbes and their rapid identification methods. *Astrobiology* 20, 925–934. doi: 10.1089/ast.2019.2078

Reid, I. N., Sparks, W. B., Lubow, S., McGrath, M., Livio, M., Valenti, J., et al. (2006). Terrestrial models for extraterrestrial life: Methanogens and halophiles at Martian temperatures. *Int. J. Astrobiol.* 5, 89–97. doi: 10.1017/S1473550406002916

Reimer, L. C., Vetcininova, A., Carbasse, J. S., Söhngen, C., Gleim, D., Ebeling, C., et al. (2019). BacDive in 2019: Bacterial phenotypic data for High-throughput biodiversity analysis. *Nucleic Acids Res.* 47:D631–D636. doi: 10.1093/nar/gky879

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

Rittmann, S., Legat, A., Fendrihan, S., and Stan-Lotter, H. (2004). "Viability and morphology of *Halobacterium* species following desiccation - implications for contaminants on Mars," in *Proceedings of the Third European Workshop on Exo-Astrobiology*, eds R. A. Harris and L. Ouwehand (Paris: ESA), 275–276.

Rivera-Valentín, E. G., Chevrier, V. F., Soto, A., and Martínez, G. (2020). Distribution and habitability of (meta)stable brines on present-day Mars. *Nat. Astron.* 4, 756–761. doi: 10.1038/s41550-020-1080-9

Romano, I., Camerlingo, C., Vaccari, L., Birarda, G., Poli, A., Fujimori, A., et al. (2022). Effects of ionizing radiation and long-term storage on hydrated vs. dried cell samples of Extremophilic microorganisms. *Microorganisms* 10:190. doi: 10.3390/microorganisms10010190

Rull, F., Veneranda, M., Manrique-Martinez, J. A., Sanz-Arranz, A., Saiz, J., Medina, J., et al. (2022). Spectroscopic study of terrestrial analogues to support rover missions to Mars – a raman-centred review. *Anal. Chim. Acta* 1209:339003. doi: 10.1016/j.aca.2021.339003

Sapers, H. M., Ronholm, J., Raymond-Bouchard, I., Comrey, R., Osinski, G. R., and Whyte, L. G. (2017). Biological characterization of microenvironments in a hypersaline cold spring Mars analog. *Front. Microbiol.* 8:2527. doi: 10.3389/fmicb.2017.02527

Schörghofer, N., Benna, M., Berezhnoy, A. A., Greenhagen, B., Jones, B. M., Li, S., et al. (2021). Water group exospheres and surface interactions on the moon. mercury, and Ceres. *Space Sci. Rev.* 217:74. doi: 10.1007/s11214-021-00846-3

Schreder-Gomes, S. I., Benison, K. C., and Bernau, J. A. (2022). 830-million-year-old microorganisms in primary fluid inclusions in halite. *Geology* 50, 918–922. doi: 10.1130/G49957.1

Sgambati, A., Deiml, M., Stettner, A., Kahrs, J., Brozek, P., Kapoun, P., et al. (2020). SPECTROModule: A modular in-situ spectroscopy platform for exobiology and space sciences. *Acta Astronaut.* 166, 377–390. doi: 10.1016/j.actaastro.2019.10.010

Shahmohammadi, H. R., Asgarani, E., Terato, H., Ide, H., and Yamamoto, O. (1997). Effects of ^{60}Co gamma-rays, ultraviolet light, and mitomycin C on *Halobacterium salinarium* and *Thiobacillus intermedius*. *J. Radiat. Res.* 38, 37–43. doi: 10.1269/jrr.38.37

Shahmohammadi, H. R., Asgarani, E., Terato, H., Saito, T., Ohyama, Y., Gekko, K., et al. (1998). Protective roles of bacterioruberin and intracellular KCl in the resistance of *Halobacterium salinarium* against DNA-damaging agents. *J. Radiat. Res.* 39, 251–262. doi: 10.1269/jrr.39.251

Sharma, N., Farooqi, M. S., Chaturvedi, K. K., Lal, S. B., Grover, M., Rai, A., et al. (2014). The halophile protein database. *Database* 2014:bau114. doi: 10.1093/database/bau114

Sherwood Lollar, B., Heuer, V. B., McDermott, J., Tille, S., Warr, O., Moran, J. J., et al. (2021). A window into the abiotic carbon cycle – Acetate and formate in fracture waters in 2.7 billion year-old host rocks of the Canadian Shield. *Geochim. Cosmochim. Acta* 294, 295–314. doi: 10.1016/j.gca.2020.11.026

Shirsalimian, M. S., Amoozegar, M. A., Sepahy, A. A., Kalantar, S. M., and Dabbagh, R. (2017). Isolation of extremely halophilic *Archaea* from a saline river in the Lut Desert of Iran, moderately resistant to desiccation and gamma radiation. *Microbiology* 86, 403–411. doi: 10.1134/S0026261717030158

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

Sorokin, D. Y., Messina, E., La Cono, V., Ferrer, M., Ciordia, S., Mena, M. C., et al. (2018). Sulfur respiration in a group of facultatively Anaerobic Natrono *Archaea* Ubiquitous in Hypersaline Soda Lakes. *Front. Microbiol.* 9:2359. doi: 10.3389/fmicb.2018.02359

Sorokin, D. Y., Messina, E., Smedile, F., Roman, P., Damsté, J. S. S., Ciordia, S., et al. (2017). Discovery of anaerobic lithoheterotrophic *Haloarchaea*, ubiquitous in hypersaline habitats. *ISME J.* 11, 1245–1260. doi: 10.1038/ismej.2016.203

Spencer, J. R., and Nimmo, F. (2013). Enceladus: An active ice world in the Saturn System. *Annu. Rev. Earth Planet. Sci.* 41, 693–717. doi: 10.1146/annurev-earth-050212-124025

Spiga, A., Banfield, D., Teanby, N. A., Forget, F., Lucas, A., Kenda, B., et al. (2018). Atmospheric Science with InSight. *Space Sci. Rev.* 214:109. doi: 10.1007/s11214-018-0543-0

Stan-Lotter, H., and Fendrihan, S. (2015). Halophilic *Archaea*: Life with desiccation, radiation and oligotrophy over geological times. *Life* 5, 1487–1496. doi: 10.3390/life5031487

Stan-Lotter, H., Radax, C., Gruber, C., Legat, A., Pfaffenhuemer, M., Wieland, H., et al. (2003). Astrobiology with *Haloarchaea* from permo-triassic rock salt. *Int. J. Astrobiol.* 1, 271–284. doi: 10.1017/S1473550403001307

Stevenson, A., Cray, J. A., Williams, J. P., Santos, R., Sahay, R., Neuenkirchen, N., et al. (2015b). Is there a common water-activity limit

for the three domains of life? *ISME J.* 9, 1333–1351. doi: 10.1038/ismej.2014.219

Stevenson, A., Burkhardt, J., Cockell, C. S., Cray, J. A., Dijksterhuis, J., Fox-Powell, M., et al. (2015a). Multiplication of microbes below 0.690 water activity: Implications for terrestrial and extraterrestrial life. *Environ. Microbiol.* 17, 257–277. doi: 10.1111/1462-2920.12598

Stillman, D. E., Michaels, T. I., and Grimm, R. E. (2017). Characteristics of the numerous and widespread recurring slope lineae (RSL) in Valles Marineris. *Mars. Icarus* 285, 195–210. doi: 10.1016/j.icarus.2016.10.025

Tarnas, J. D., Mustard, J. F., Sherwood Lollar, B., Stamenkoviæ, V., Cannon, K. M., Lorand, J. P., et al. (2021). Earth-like habitable environments in the subsurface of Mars. *Astrobiology* 21, 741–756. doi: 10.1089/ast.2020.2386

Taşkın, Z., and Aydinoglu, A. U. (2015). Collaborative interdisciplinary astrobiology research: A bibliometric study of the NASA Astrobiology Institute. *Scientometrics* 103, 1003–1022. doi: 10.1007/s11192-015-1576-8

Taubner, R.-S., Pappenreiter, P., Zwicker, J., Smrzka, D., Pruckner, C., Kolar, P., et al. (2018). Biological methane production under putative Enceladus-like conditions. *Nat. Commun.* 9:748. doi: 10.1038/s41467-018-02876-y

Thombre, R., Shinde, V., Dixit, J., Jagtap, S., and Vidyasagar, P. B. (2017). Response of extreme *Haloarchaeon Haloarcula argentinensis* RR10 to simulated microgravity in clinorotation. *Biotech* 7:30. doi: 10.1007/s13205-016-0596-2

Thombre, R. S., Bhalerao, A. R., Shinde, V. D., Dhar, S. K., and Shouche, Y. S. (2017). Response of haloalkaliphilic *Archaeon Natronococcus jeotgali* RR17 to hypergravity. *Microgravity Sci. Technol.* 29, 191–200. doi: 10.1007/s12217-017-9538-9

Tomkins, A. G., Genge, M. J., Tait, A. W., Alkemade, S. L., Langendam, A. D., Perry, P. P., et al. (2019). High Survivability of Micrometeorites on Mars: Sites With Enhanced Availability of Limiting Nutrients. *J. Geophys. Res.* 124, 1802–1818. doi: 10.1029/2019JE006005

Toner, J. D., and Catling, D. C. (2018). Chlorate brines on Mars: Implications for the occurrence of liquid water and deliquescence. *Earth Planet. Sci. Lett.* 497, 161–168. doi: 10.1016/j.epsl.2018.06.011

Trigui, H., Masmoudi, S., Brochier-Armanet, C., Maalej, S., and Dukan, S. (2011). Survival of extremely and moderately halophilic isolates of Tunisian solar salterns after UV-B or oxidative stress. *Can. J. Microbiol.* 57, 923–933. doi: 10.1139/w11-087

Ueta, S., and Sasaki, T. (2013). The Structures of Surface H_2O Layers of Ice-Covered Planets with High-Pressure Ice. *Astrophys. J.* 775:96. doi: 10.1088/0004-637x/775/2/96

Ukani, H., Purohit, M. K., Georrge, J. J., Paul, S., and Singh, S. (2011). HaloBase: Development of database system for halophilic bacteria and *Archaea* with respect to proteomics, genomics & other molecular traits. *J. Sci. Ind.* 70, 976–981. doi: 10.5281/zenodo.4398512

Uritskiy, G., Tisza, M. J., Gelsinger, D. R., Munn, A., Taylor, J., and DiRuggiero, J. (2021). Cellular life from the three domains and viruses are transcriptionally active in a hypersaline desert community. *Environ. Microbiol.* 23, 3401–3417. doi: 10.1111/1462-2920.15023

Van Der Wielen, P. W. J. J., and Heijs, S. K. (2007). Sulfate-reducing prokaryotic communities in two deep hypersaline anoxic basins in the Eastern Mediterranean deep sea. *Environ. Microbiol.* 9, 1335–1340. doi: 10.1111/j.1462-2920.2006.01210.x

Vance, S., Bouffard, M., Choukroun, M., and Sotin, C. (2014). Ganymede×s internal structure including thermodynamics of magnesium sulfate oceans in contact with ice. *Planet. Space Sci.* 96, 62–70. doi: 10.1016/j.pss.2014.03.011

Vicente-Retortillo, A., Martínez, G. M., Rennó, N. O., Lemmon, M. T., de la Torre-Juárez, M., and Gómez-Elvira, J. (2020). In Situ UV Measurements by MSL/REMS: Dust deposition and angular response corrections. *Space Sci. Rev.* 216:97. doi: 10.1007/s11214-020-00722-6

Vikram, S., Guerrero, L. D., Makhalanyane, T. P., Le, P. T., Seely, M., and Cowan, D. A. (2016). Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.* 18, 1875–1888. doi: 10.1111/1462-2920.13088

Wadsworth, J., Rettberg, P., and Cockell, C. S. (2019). Aggregated Cell masses provide protection against space extremes and a microhabitat for hitchhiking co-inhabitants. *Astrobiology* 19, 995–1007. doi: 10.1089/ast.2018.1924

Wang, F., Li, M., Huang, L., and Zhang, X.-H. (2021). Cultivation of uncultured marine microorganisms. *Mar. Life sci. Technol.* 3, 117–120. doi: 10.1007/s42995-021-00093-z

Webb, K. M., Yu, J., Robinson, C. K., Noboru, T., Lee, Y. C., and DiRuggiero, J. (2013). Effects of intracellular Mn on the radiation resistance of the halophilic *Archaeon Halobacterium salinarum*. *Extremophiles* 17, 485–497. doi: 10.1007/s00792-013-0533-9

Weber, H. S., Habicht, K. S., and Thamdrup, B. (2017). Anaerobic Methanotrophic *Archaea* of the ANME-2d Cluster Are Active in a Low-sulfate, Iron-rich Freshwater Sediment. *Front. Microbiol.* 8:619. doi: 10.3389/fmicb.2017.00619

Weidler, G., Leuko, S., Radax, C., Kargl, G., Kömle, N., and Stan-Lotter, H. (2002). "Viability and DNA damage of *Halobacteria* under physical stress condition including a simulated Martian atmosphere," in *Proceedings of the First European Workshop on Exo-Astrobiology*, ed. H. Lacoste (Paris: ESA), 491–492.

Weidler, G., Leunko, S., and Stan-Lotter, H. (2004). "Survival and growth of *Halobacterium* sp. NRC-1 following incubation at −15°C, freezing or freeze-drying, and the protective effect of cations," in *Proceedings of the Third European Workshop on Exo-Astrobiology*, eds R. A. Harris and L. Ouwehand (Paris: ESA), 311–312.

Whitehead, K., Kish, A., Pan, M., Kaur, A., Reiss, D. J., King, N., et al. (2006). An integrated systems approach for understanding cellular responses to gamma radiation. *Mol. Syst. Biol.* 2:47. doi: 10.1038/msb4100091

Wu, Z., Wang, A., Farrell, W. M., Yan, Y., Wang, K., Houghton, J., et al. (2018). Forming perchlorates on Mars through plasma chemistry during dust events. *Earth Planet. Sci. Lett.* 504, 94–105. doi: 10.1016/j.epsl.2018.08.040

Xiao, L., Wang, J., Dang, Y., Cheng, Z., Huang, T., Zhao, J., et al. (2017). A new terrestrial analogue site for Mars research: The Qaidam Basin, Tibetan Plateau (NW China). *Earth Sci. Rev.* 164, 84–101. doi: 10.1016/j.earscirev.2016.11.003

Zhang, J., Guo, J., Dobynde, M. I., Wang, Y., and Wimmer-Schweingruber, R. F. (2022). From the Top of Martian Olympus to deep craters and beneath: Mars radiation environment under different atmospheric and regolith depths. *J. Geophys. Res.* 127:e2021JE007157. doi: 10.1029/2021JE007157

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 10 November 2022
DOI 10.3389/fmicb.2022.975365

OPEN ACCESS

EDITED BY
André Antunes,
Macau University of Science and Technology, China

REVIEWED BY
Javier Pascual,
Darwin Bioprospecting Excellence, Spain
Bärbel Ulrike Fösel,
Helmholtz Center München, Helmholtz Association of German Research Centres (HZ), Germany

*CORRESPONDENCE
Maria del Carmen Montero-Calasanz
mariac.montero.calasanz@juntadeandalucia.es

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

RECEIVED 22 June 2022
ACCEPTED 11 October 2022
PUBLISHED 10 November 2022

CITATION
Montero-Calasanz MdC, Yaramis A, Rohde M, Schumann P, Klenk H-P and Meier-Kolthoff JP (2022) Genotype–phenotype correlations within the *Geodermatophilaceae*. *Front. Microbiol.* 13:975365. doi: 10.3389/fmicb.2022.975365

Genotype–phenotype correlations within the *Geodermatophilaceae*

Maria del Carmen Montero-Calasanz[1,2]*, Adnan Yaramis[2], Manfred Rohde[3], Peter Schumann[4], Hans-Peter Klenk[2] and Jan P. Meier-Kolthoff[5]

[1]IFAPA Las Torres-Andalusian Institute of Agricultural and Fisheries Research and Training, Junta de Andalucía, Seville, Spain, [2]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [3]Central Facility for Microscopy, HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany, [4]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [5]Department Bioinformatics and Databases, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

The integration of genomic information into microbial systematics along with physiological and chemotaxonomic parameters provides for a reliable classification of prokaryotes. *In silico* analysis of chemotaxonomic traits is now being introduced to replace characteristics traditionally determined in the laboratory with the dual goal of both increasing the speed of the description of taxa and the accuracy and consistency of taxonomic reports. Genomics has already successfully been applied in the taxonomic rearrangement of *Geodermatophilaceae (Actinomycetota)* but in the light of new genomic data the taxonomy of the family needs to be revisited. In conjunction with the taxonomic characterisation of four strains phylogenetically located within the family, we conducted a phylogenetic analysis of the whole proteomes of the sequenced type strains and established genotype–phenotype correlations for traits related to chemotaxonomy, cell morphology and metabolism. Results indicated that the four isolates under study represent four novel species within the genus *Blastococcus*. Additionally, the genera *Blastococcus*, *Geodermatophilus* and *Modestobacter* were shown to be paraphyletic. Consequently, the new genera *Trujillonella*, *Pleomorpha* and *Goekera* were proposed within the *Geodermatophilaceae* and *Blastococcus endophyticus* was reclassified as *Trujillonella endophytica* comb. nov., *Geodermatophilus daqingensis* as *Pleomorpha daqingensis* comb. nov. and *Modestobacter deserti* as *Goekera deserti* comb. nov. Accordingly, we also proposed emended descriptions of *Blastococcus aggregatus, Blastococcus jejuensis, Blastococcus saxobsidens* and *Blastococcus xanthilyniticus*. *In silico* chemotaxonomic results were overall consistent with wet-lab results. Even though *in silico* discriminatory levels varied depending on the respective chemotaxonomic trait, this approach is promising for effectively replacing and/or complementing chemotaxonomic analyses at taxonomic ranks above the species level. Finally, interesting but previously overlooked insights regarding morphology and ecology were revealed by the presence of a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly in all representatives of

the family. A rich carbon metabolism including four different CO_2 fixation pathways and a battery of enzymes able to degrade complex carbohydrates were also identified in *Blastococcus* genomes.

KEYWORDS

***Trujillonella*, *Pleomorpha*, *Goekera*, phylogenetic systematics, *in silico* chemotaxonomy**

Introduction

The family *Geodermatophilaceae,* belonging to order *Geodermatophilales* (Sen et al., 2014), was initially proposed by Normand et al. (1996), confirmed by Stackebrandt et al. (1997), formally described by Normand (2006) and later emended by Zhi et al. (2009) and Montero-Calasanz et al. (2017). The family accommodates the genera *Blastococcus* (Ahrens and Moll, 1970; Urzì et al., 2004; Lee, 2006; Hezbri et al., 2016), *Geodermatophilus*, the type genus (Luedemann, 1968; Montero-Calasanz et al., 2017; Montero-Calasanz, 2020a), *Klenkia* (Montero-Calasanz et al., 2017), and *Modestobacter* (Mevs et al., 2000; Reddy et al., 2007; Xiao et al., 2011; Qin et al., 2013; Montero-Calasanz et al., 2019).

The genus *Blastococcus* consists of Gram-reaction-positive aerobic bacteria that can be distinguished by the presence of individual motile rod-shaped cells and/or non-motile cocci (Lee, 2006) with a tendency to form aggregates (Urzì et al., 2004; Hezbri et al., 2016). At the time of writing, the genus comprises 10 validly named species including *B. aggregatus* (type species), isolated from sediment from Baltic Sea (Ahrens and Moll, 1970), *B. atacamensis*, isolated from an extreme hyper-arid Atacama Desert soil (Castro et al., 2018), *B. capsensis,* isolated from an ancient Roman pool in Tunisia (Hezbri et al., 2016), *B. colisei,* isolated from the ruins of a Roman amphitheatre in Tunisia (Hezbri et al., 2017), *B. deserti,* isolated from Gurbantunggut Desert in China (Yang et al., 2019), *B. endophyticus,* isolated from Chinese medicinal plant leaves (Zhu et al., 2013; Hezbri et al., 2016), *B. jejuensis,* isolated from sand in Korea (Lee, 2006; Hezbri et al., 2016), *B. litoris,* isolated from sea-tidal flat sediment (Lee et al., 2018), *B. saxobsidens,* isolated from an archaeological site in Greece (Urzì et al., 2004), and *B. xanthinilyticus,* isolated from a marble sample collected from the Bulla Regia monument in Tunisia (Hezbri et al., 2018).

The integration of genomic information into microbial systematics (Klenk and Göker, 2010) along with physiological and chemotaxonomic parameters enables a reliable classification of prokaryotes (Chun et al., 2018). Laboratory-based DNA–DNA hybridisation methods have already been routinely replaced by overall genome relatedness indices (OGRI; Chun et al., 2018) of which digital DNA:DNA hybridization was shown to be the best-performing method (Meier-Kolthoff et al., 2013a). The use of *in silico* analysis for chemotaxonomic traits is now being introduced to replace characteristics traditionally determined in the laboratory (Amaral et al., 2014; Fotedar et al., 2020; Lawson et al., 2020) with the dual goal of both increasing the speed of the description of taxa and the accuracy and consistency of taxonomic reports (Chun and Rainey, 2014; Whitman, 2015).

Genomics has already successfully been applied in the taxonomic rearrangement of *Geodermatophilaceae* (Montero-Calasanz et al., 2017) although other research studies based on genomic data are still scarce in the family. In particular, only two papers were published so far about the genus *Blastococcus* including genomic and proteomic mining of *B. saxobsidens* DD2 and genome mining of *B. atacamensis* which revealed numerous genes involved in stress response and adaptation to harsh habitats (Chouaia et al., 2012; Sghaier et al., 2016) apart from an exceptional potential to produce novel natural compounds (Castro et al., 2018).

Based on genome-scale data and phenotypic evidence according to the principles of phylogenetic systematics (Hennig, 1965; Wiley and Lieberman, 2011), this study characterises four novel species in the genus *Blastococcus,* introduces the new genera *Trujillonella, Pleomorpha,* and *Goekera* in *Geodermatophilaceae* and reclassifies *Blastococcus endophyticus* as *Trujillonella endophytica* comb. nov., *Geodermatophilus daqingensis* as *Pleomorpha daqingensis* comb. nov. and *Modestobacter deserti* as *Goekera deserti* comb. nov., all representing type species of the new genera. Accordingly, we also propose emended descriptions within *Blastococcus*. Additionally, we emphasise how genomics could effectively replace and/or complement some of the phenotypic analyses routinely carried out in microbial systematics.

Materials and methods

Isolation

Strains AT 7-1^T, AT 7–8, AT 7-14^T, and AT 7(-2)-11^T were isolated during a screening of microorganisms from soil samples from the Atacama Desert (Chile) in the 90s and deposited by Prof. Fred A. Rainey (University of Alaska Anchorage, United States) at Leibniz Institute DSMZ. On the other hand, strain G1S^T was isolated by Eppard et al. (1996) from leachate of a landfill in Vancouver (Canada). All strains were subsequently accessed in the

DSMZ open collection as DSM 44268^T, DSM 44269, DSM 44270^T, DSM 44272^T, and DSM 44205^T, respectively.

Genotypic analysis

Genomic DNA extraction, PCR-mediated amplification of the 16S rRNA gene, and purification of the PCR product were carried out as previously described by Rainey et al. (1996). For genome sequencing, strains were cultivated in GYM *Streptomyces* broth (DSMZ medium 65, pH 6.8 ± 0.2; https://www.dsmz.de/collection/catalogue/microorganisms/culture-technology/list-of-media-for-microorganisms) at 28°C and genomic DNA was extracted by using JetFlex™ Genomic DNA Purification Kit (Thermo Fisher Scientific) following the manufacturer's instructions. The project information is available through the Genomes Online Database (Mukherjee et al., 2017). The draught genomes were generated at the DOE Joint Genome Institute (JGI) as part of the Genomic Encyclopaedia of Archaeal and Bacterial Type Strains, Phase II (KMG-II): "From individual species to whole genera" (Kyrpides et al., 2014) and ACTINO 1000: "Exploiting the genomes of the *Actinobacteria*: plant growth promoters and producers of natural products and energy relevant enzymes united in a taxonomically unresolved phylum," following the same protocol as in Nouioui et al. (2017). All genomes were annotated using the DOE-JGI annotation pipeline (Huntemann et al., 2015; Chen et al., 2016) and released through the Integrated Microbial Genomes system (Chen et al., 2017). In addition, genomes were uploaded to RAST (Aziz et al., 2008; Brettin et al., 2015) and analysed through the SEED viewer (Overbeek et al., 2014). CRISPRFinder (Grissa et al., 2007) was used to identify CRISPR elements. Further details about sequencing projects are summarised in the Supplementary Table S1.

Phylogenetic analysis of the 16S rRNA gene sequences from the type strains of all species with validated names in *Geodermatophilaceae*, as well as those from *Antricoccus suffuscus* DSM 100065^T, *Cryptosporangium arvum* DSM 44712^T, *Cryptosporangium aurantiacum* DSM 46144^T, *Epidermidibacterium keratini* JCM 31644^T, *Longivirga aurantiaca* NBRC 112237^T and *Sporichthya polymorpha* DSM 43042^T for use as outgroup, was conducted as previously described (Göker et al., 2011; Montero-Calasanz et al., 2014). Pairwise 16S rRNA gene sequence similarities were calculated as recommended by Meier-Kolthoff et al. (2013b) to determine strains with ≥99.0% similarity, between which (digital) DNA:DNA hybridization experiments were conducted for clarifying species affiliation. Genome-scale phylogenies were inferred from the available *Geodermatophilaceae* (and outgroup) whole proteome sequences using the high-throughput version (Meier-Kolthoff et al., 2014a) of the Genome BLAST Distance Phylogeny (GBDP) approach (Auch et al., 2010) in conjunction with FastME (Lefort et al., 2015). The tree was rooted according to Nouioui et al. (2018). The entire approach is implemented in the Type (Strain) Genome Server (TYGS; Meier-Kolthoff and Göker, 2019; Meier-Kolthoff et al., 2022). The GBDP tree restricted to the well-supported branches (≥95% pseudo-bootstrap support) was used as a backbone constraint in a subsequent 16S rRNA gene sequence analysis to integrate information from genome-scale data (Hahnke et al., 2016). Digital DNA:DNA hybridisations were conducted using the recommended settings of the Genome-To-Genome Distance Calculator (GGDC) version 3.0 (Meier-Kolthoff et al., 2013a) as implemented in the TYGS (Meier-Kolthoff and Göker, 2019; Meier-Kolthoff et al., 2022). The G + C content was calculated from the genome sequences as described by Meier-Kolthoff et al. (2014b). The genomic homogeneity of strains DSM 44268^T and DSM 44269 was examined by ribotyping and matrix-assisted laser-desorption/ionisation time-of-flight (MALDI-TOF) mass spectra (MS). Automated ribotyping of PvuII-digested samples was carried out using the RiboPrinter microbial characterisation system (Qualicon, DuPont) as described in Bruce (1996) and Stackebrandt et al. (2002). MALDI-TOF analysis was performed according to Tóth et al. (2008).

Phenotypic analysis

Morphological and physiological tests

Morphological features of colonies were examined by stereomicroscope on GYM *Streptomyces* agar at 28°C and observed periodically during 7 days. Bacterial cell morphology and motility was determined by phase contrast microscopy with a magnification of 100x and with a field-emission scanning electron microscope (FE-SEM Merlin, Zeiss, Germany). Catalase production was tested on slides with the addition of some drops of 3% H_2O_2 solution into a loop of bacterial biomass. It was considered positive when the production of bubbles was observed after mixing. Oxidase activity was determined by obtaining blue-purple colour after addition of some drops of 1% (w/v) solution of *N,N,N′,N′*-tetramethyl-*p*-phenylenediamine (Sigma-Aldrich) to a loop of bacterial biomass spread on filter paper. Gram reaction was performed according to the procedure developed by Gregersen (1978). API ZYM system (bioMérieux) was used to determine enzymatic activities according to the manufacturer's protocol at 28°C after 24 h incubation. Homogeneous bacterial suspensions of strains DSM 44205^T, DSM 44268^T, DSM 44269, DSM 44270^T, and DSM 44272^T along with the reference type strains in the genus *Blastococcus* were prepared in viscous inoculating fluid (IF C) at 86% T (Transmittance) for *B. atacamensis* P6^T, 83% T for *B. capsensis* DSM 46835^T, 89% T for DSM 44270^T, *B. colisei* DSM 46837^T, and *B. xanthinilyticus* DSM 46842^T, 90% T for strains DSM 44268^T, DSM 44269, and DSM 44205^T, 91% T for DSM 44272^T, and 95% T for all the others reference strains. Then those were inoculated in GEN III Microplates per duplicate and incubated for 10 days at 28°C in an Omnilog device (BIOLOG Inc., Hayward, CA, United States). Data were analysed using the *opm* package R v.1.3.72 (Vaas et al., 2012, 2013). Reactions that reflected different behaviour between replicates were regarded as ambiguous. Temperature ranges were evaluated on GYM *Streptomyces* medium for 15 days at 4, 10, 15,

20, 25, 30, 37, 40 and 45°C. pH ranges were determined using modified ISP2 medium (Montero-Calasanz et al., 2015) from 4.5 to 12.5 in increments of 0.5 pH unit adjusting with NaOH or HCl since the use of a buffer system inhibited the growth of strains (Montero-Calasanz et al., 2014). pH of solidified plates was measured before inoculation and when recording results (15 days post-inoculation). Hydrolysis of specific substrates such as tyrosine (Gordon and Smith, 1955), casein, starch, xanthine as well as hypoxanthine was also tested as outlined by Montero-Calasanz et al. (2015).

Chemotaxonomic tests

For chemotaxonomic analysis, strains were grown in GYM *Streptomyces* broth with shaking (120 rpm) at 28°C for 7 days. Bacterial biomass was collected and subsequently freeze dried. Extraction of whole-cell amino acids and sugars was carried out as developed by Lechevalier and Lechevalier (1970) followed by thin layer chromatography (TLC) analysis and identification (Staneck and Roberts, 1974). Analysis of peptidoglycan hydrolysates (6N HCl, 100°C for 16 h) was performed on cellulose TLC plates as described by Schleifer and Kandler (1972). Menaquinones (MK) analyses (Kroppenstedt, 1982) were accomplished as indicated by Collins et al. (1977) following the identification of MK by high-performance liquid chromatography (HPLC; Kroppenstedt, 1982). Polar lipids were extracted and identified by 2D TLC as outlined by Minnikin et al. (1984) with modifications proposed by Kroppenstedt and Goodfellow (2006). Dragendorff's reagent (Merck Millipore, 102035) was additionally sprayed to identify choline-containing lipids (Tindall, 1990). For fatty acid analysis, cell biomass grown on PYGV (DSMZ 621) agar plates for 16 days at 20°C was harvested and fatty acid methyl esters were extracted according to Sasser (2001). Microbial Identification System (MIDI) Sherlock Version 6.1 (method TSBA40, ACTIN6 database) was used for performing data analysis. All the chemotaxonomic analyses for strains were carried out under standardised conditions.

Results and discussion

Basic genome statistics

Standard draught genomes were obtained for strains DSM 44205^T, DSM 44268^T, DSM 44270^T, and DSM 44272^T and type strains of species *B. aggregatus, B. endophyticus* and *B. xanthinilyticus.* An improved-high-quality genome draught according to GOLD sequencing quality standards (Mukherjee et al., 2017) was also obtained for both *B. colisei* DSM 46837^T and *B. saxobsidens* DSM 44509^T (Supplementary Table S1). The genome sizes ranged from 4.0 Mbp (strain DSM 44205^T) to 5.1 Mbp (strains DSM 44272^T and DSM 46837^T), with an average genome size for all sequenced *Blastococcus* strains of 4.6 ± 0.4 Mbp (Supplementary Table S2). The genomic G + C content of the sole chromosome in the type strains varied between 72.5–74.6%. Deviations from the published G + C contents of reference strains were all below the 1% threshold determined within species (Meier-Kolthoff et al., 2014b). In accordance with the genome sizes, the number of protein-coding genes ranged from 3,939 for strain DSM 44205^T to 5,007 for strain DSM 44272^T, not observing *a priori* any association between genome size and habitat of isolation. Both tRNA and rRNA genes represented between 1.3% (47 tRNA and five rRNA genes) in *B. xanthinilyticus* DSM 46842^T and 2.2% (70 tRNA and nine rRNA genes) in DSM 44205^T of the whole genome sequences. A single copy of the 16S rRNA gene was identified for strains DSM 44270^T, DSM 44272^T, *B. aggregatus* DSM 4725^T, and *B. xanthinilyticus* DSM 46842^T. Three almost identical 16S rRNA genes copies were annotated in the genome sequences of strain DSM 44205^T, *B. colisei* DSM 46837^T, *B. endophyticus* DSM 45413^T, and *B. saxobsidens* DSM 44509^T, and four almost identical ones in the genome sequence of strain DSM 44268^T (intragenomic heterogeneity of 16S rRNA genes is below the 1% level for species delimitation; Meier-Kolthoff et al., 2013a). Previous studies already reported low copy numbers (2–4) for slow-growing actinobacteria (Sun et al., 2013; Vetrovsky and Baldrian, 2013) isolated from oligotrophic environments (Klappenbach et al., 2000). The number of rRNA operon copies is suggested to be related to the life strategy of bacteria and how quickly they respond to favourable changes in environmental conditions and nutrient availability (Klappenbach et al., 2000). Hence, rRNA copy number appears to be correlated with growth speed. The number of identified tRNAs ranged from 47 to 49 in all the *Blastococcus* representatives except in the genome sequence of strain DSM 44205^T that contained 70 tRNAs. The 20 standard tRNA genes were present in all genomes but in a different distribution. For example, four tRNA-Arg gene copies (six gene copies for strain DSM 44205^T) versus one tRNA-Tyr gene copy. Such uneven distribution supports the co-adaptation between tRNA abundance and specific codon usage of each organism as a higher ratio of optimal codons may facilitate the transcription efficiency (Du et al., 2017). Out of the 70 (DSM 44205^T) and 48 (*B. endophyticus*) annotated tRNAs, seven (DSM 44205^T) and one (*B. endophyticus*), respectively, were identified as pseudogene(s). The latter are known to be dysfunctional in translation but to be potentially involved in modulating other processes such as antibiotic and cell wall biosynthesis and gene expression (Rogers et al., 2012). One copy of nonsense suppressor tRNA (tRNA amber gene) was also identified in the repertoires of DSM 44270^T, DSM 44272^T, and *B. colisei,* which could prevent the premature termination of translation caused by nonsense mutations (Ko et al., 2013). The number of pseudogenes was within the expected values for free-living prokaryotes (0%–5%; Liu et al., 2004) and varied from zero (for example: DSM 44205^T, DSM 44272^T, *B. endophyticus, B. xanthinilyticus* or *B. atacamensis*, amongst others) to 123 in *B. colisei* DSM 46837^T (2.5%). Each genome contained a largely unique set of pseudogenes, with over half of those annotated as "hypothetical" proteins, suggesting that pseudogenes in *Blastococcus* are formed and eliminated relatively rapidly as previously reported by Kuo and Ochman (2010) for

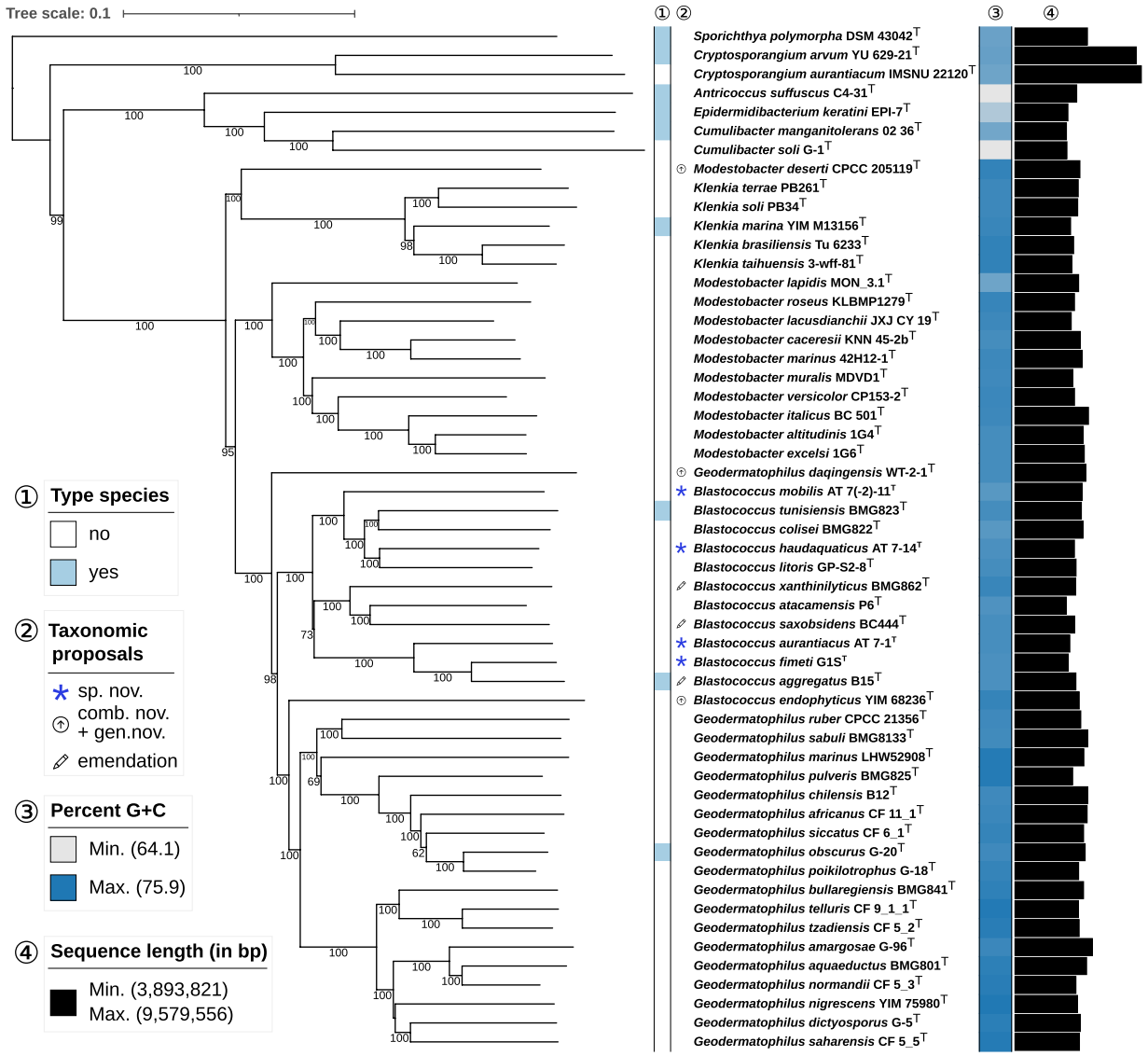


FIGURE 1
Phylogenomic tree inferred with GBDP. The tree was inferred with FastME from GBDP distances calculated from whole proteomes. The numbers above branches are GBDP pseudo-bootstrap support values from 100 replications. Tip colours indicate genome sizes and the exact G+C content as calculated from the genome sequences (see the embedded legend for details). *A., Antricoccus; B., Blastococcus; C., Cryptosporangium; G., Geodermatophilus; K., Klenkia; M., Modestobacter; S., Sporichthya.*

Salmonella genomes. In addition, the number of paralogous genes varied from 437 (11%) in strain DSM 44205^T to 1,601 (32%) in *B. colisei* DSM 46837^T, the signal peptide percentage was 3.0%–5.8% whilst the percentage of transmembrane proteins ranged from 19.7% to 24.7%. According to CRISPRFinder, the number of "questionable" Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) ranged from zero to seven (Supplementary Table S2). Additional analyses to test the genetic environment often associated with CRISPR structures should nevertheless be carried out to confirm the presence of a naturally occurring genome editing system in *Blastococcus* representatives.

The average percentage of genes with a predicted function in the Clusters of Orthologous Groups (COGs) database was 63.3 ± 2.3% (ranged from 2,558 to 3,117 genes; Supplementary Table S2). As expected, functional COG profiles of *Blastococcus* strains are similar (Supplementary Figure S1; Supplementary Table S3).

Phylogenetic analysis

The phylogenetic tree inferred by whole proteomes of the sequenced type strains and strains under study revealed the affiliation of DSM 44268^T, DSM 44270^T, DSM 44272^T and DSM 44205^T to the genus *Blastococcus* (Figure 1). The subtree containing the genera *Geodermatophilus, Modestobacter* and *Blastococcus* was

not monophyletic which was caused by three misclassified species. First, *B. endophyticus* DSM 45413^T appeared with maximum support as a sister group of *Geodermatophilus* and was not phylogenetically placed within the maximally supported subtree harbouring all the other *Blastococcus* type strains. Second, *G. daqingensis* DSM 104001^T represented a maximally supported sister group to the large subtree comprising all other *Geodermatophilus* and *Blastococcus* type strains. Third, *M. deserti* CPCC 205119^T formed a deeply-branching and maximally supported sister group of *Klenkia*.

The constrained 16S rRNA phylogenetic tree (Figure 2) based on gene sequences of all validly published species in the family and strains under study, including strain DSM 44269, confirmed the observations from the proteome-based GBDP analysis. Moreover, strains DSM 44268^T and DSM 44269 were placed with high support in a clade together with strains DSM 44205^T and *B. aggregatus* DSM 4725^T.

The 16S rRNA gene sequence similarity values between (i) DSM 44268^T and DSM 44269, (ii) between DSM 44205^T, DSM 44270^T and *G. aggregatus* DSM 4725^T, and, (iii) between DSM 44270^T and *B. litoris* GP-S2-8^T, ranged above the *Actinobacteria*-specific 16S rRNA threshold of 99.0% (accepting an error probability of 1%; Meier-Kolthoff et al., 2013b).

It was thus necessary to investigate the species status *via* digital DDH (dDDH). The dDDH values between strain DSM 44268^T and strains DSM 44205^T (30.8%), DSM 44270^T (23.8%), and *B. aggregatus* DSM 4725^T (30.3%) and between *B. litoris* GP-S2-8^T and strains DSM 44270^T (26.9%) resulted in values below the 70% threshold throughout, thus indicating distinct species (Wayne et al., 1987).

In addition, the MALDI-TOF dendrogram (Supplementary Figure S2) confirmed the affiliation of strains DSM 44268^T and DSM 44269 to the same species showing a 100% similarity regarding mass spectra patterns of proteins and separated from those under study at a distance level > 900 arbitrary units. The riboprinting patterns of the DSM 44268^T and DSM 44269 isolates were also highly similar (98.3% similarity; Supplementary Figure S3) but differentiated them as clonal variants of each other.

Phenotypic-genotypic correlations

Morphology and physiology

All target strains were aerobic, Gram-reaction positive and showed coccoid cells with a tendency to form aggregates as previously described by Urzì et al. (2004) for the genus *Blastococcus* (Supplementary Figure S4). In addition, under studied culture conditions, reproduction by budding was predominant although bacterial binary fission was also observed (Supplementary Figure S4). It was already indicated for strain DSM 44205^T by Eppard et al. (1996). Effectively, the presence of the cell division protein FtsZ and related genes such as the actin-like protein MreB were annotated in all genomes but no known mechanisms involved in division *via* budding were identified (Supplementary Table S4). Whilst FtsZ is an essential structural homologue of the eukaryotic cytoskeletal element tubulin (Lowe and Amos, 1998) and involved in forming a cytokinetic ring at the division site (Wang and Lutkenhaus, 1993), MreB is mostly essential for cell elongation and maintenance of the cell (Rivas-Marin et al., 2020). Mechanisms of bacterial division by budding are still poorly understood in evolutionary cell biology (Rivas-Marin et al., 2020). *Geodermatophilaceae* representatives are described as aerobic bacteria (Montero-Calasanz, 2020b). Nevertheless, some microaerophilic representatives isolated from the oxic/anoxic interface of marine sediments were also reported (Ahrens and Moll, 1970). To date no anaerobic isolate was described, although the presence of reductases using anaerobic terminal electron acceptors such as arsenate, ferredoxin or flavodoxins as well as a repertoire of genes involved in the fermentation of propanoate, butanoate and pyruvate, amongst others, in the *Blastococcus* genomes could indicate that anaerobic respiration may be possible. It was already suggested by Sghaier et al. (2016).

In vivo motility was not observed in any of the studied strains. The presence of homologues involved in flagellar biosynthesis and function and chemotaxis in all the genome sequences, nevertheless, suggests that *Blastococcus* strains may be flagellated chemotactic bacteria and motility may have been influenced by culture conditions. In fact, Ahrens and Moll (1970) suggested that the motile rod stage in *B. aggregatus* could be favoured under microaerophilic conditions. Later studies carried out by Holt and Chaubal (1997) for *Salmonella pullorum* strains, confirmed that flagellar synthesis is affected by agar concentration, carbohydrate concentration and type, and incubation temperature. The presence of flagellar gene clusters is consistent with observed motility by type strains of *B. jejuensis* and *B. saxobsidens* (Urzì et al., 2004; Lee 2006) and other strains in the family (Montero-Calasanz et al., 2017). In fact, flagellin synthesis was identified as the most highly expressed protein in the proteogenome of *Blastococcus* sp. DD2 (Sghaier et al., 2016). The core set of flagellar structural genes for *Geodermatophilaceae* representatives seems to be largely consistent with that described by Liu and Ochman (2007) for the bacterial domain. Twenty-eight different genes comprise specific proteins that form the filament (*fliC*, two copies for strain DSM 44205^T, DSM 44270^T, and *B. xanthinilyticus*), the hook-filament junction (*flgL* and *flgK*), the hook (*flgE*), the rod (*flgB* and *flgC*), the MS ring (*fliF*; absent in DSM 44272^T), the C ring (*fliG, fliM* and *fliN*), the motor (*motA* and *motB*), and the export apparatus (*flhA, flhB, fliP, fliQ,* and *fliR*). In addition, *flgD* encoding for the hook-capping protein, which is required for flagellar assembly (Liu and Ochman, 2007), and other broadly distributed flagellar structural genes such as *fliD* (filament cap), *fliE, fliK* and *fliH* (hook) as well as, *fliZ, fliS* and *fliO* are also part of the core set in *Geodermatophilaceae* (Supplementary Table S4). The L and P ring proteins *FlgH* and *FlhI* are absent in Gram-positive bacteria. Chemotaxis, on the other hand, may be controlled by the universal two-component system. Methyl-accepting chemotaxis proteins (MCP) would control autophosphorylation of a cytoplasmic histidine kinase (*CheA*), *via* a coupling protein (*CheW*), which would activate the response regulator *CheY* and, consequently, interact with the switch

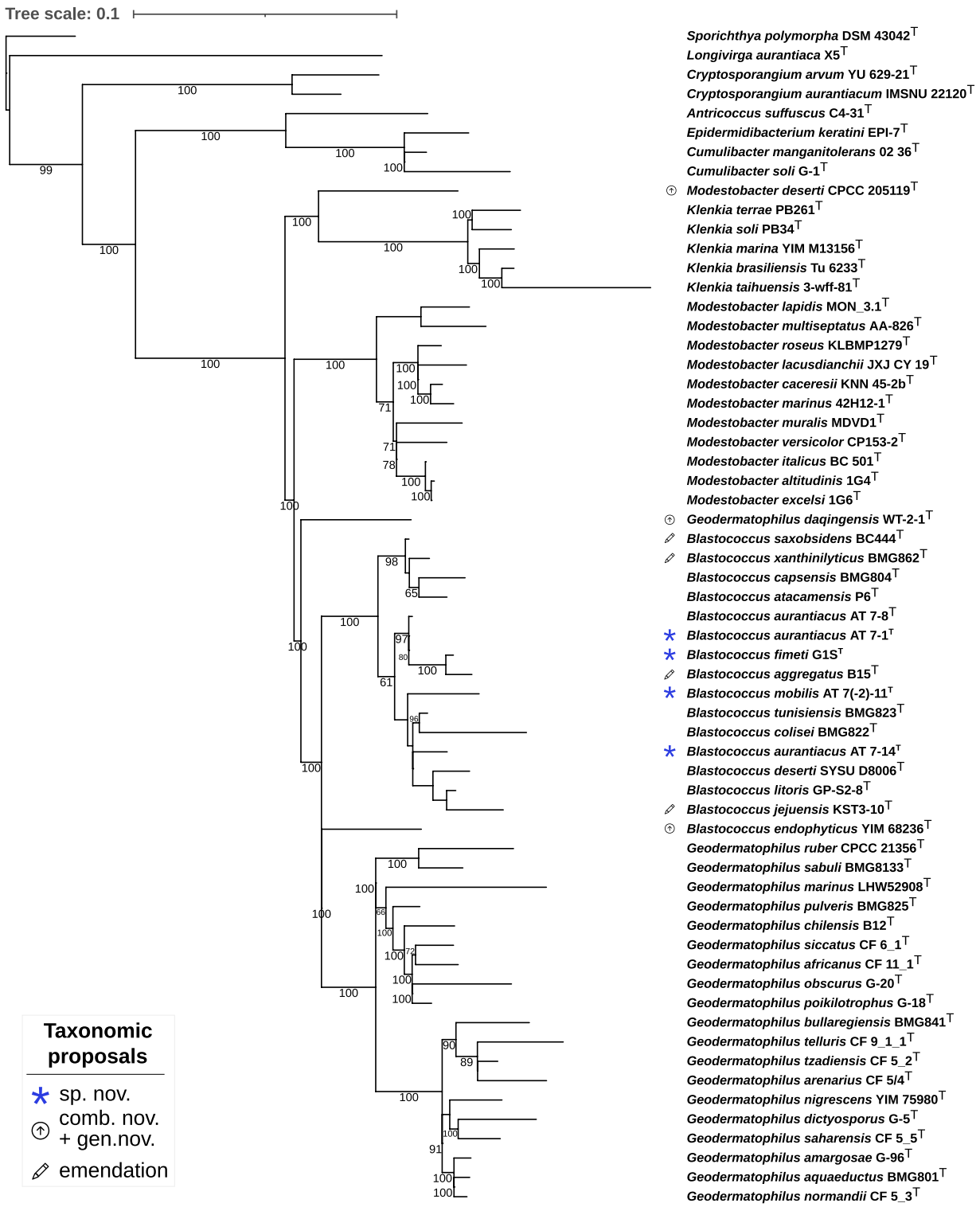


FIGURE 2
Maximum likelihood phylogenetic tree inferred from 16S rRNA gene sequences, showing the phylogenetic position of the strains DSM 44205^T, DSM 44268^T, DSM 44269, DSM 44270^T and DSM 44272^T relative to the type strains within *Geodermatophilaceae*. The branches are scaled in terms of the expected number of substitutions per site (see size bar). Support values from maximum-likelihood constrained bootstrapping are shown above the branches if equal to or larger than 60%.

mechanism in the flagellar motor (Szurmant and Ordal, 2004). Surprisingly, a dedicated ribose-binding protein (ribose transport system substrate-binding protein, *RbsB*) was annotated in the genome sequences of DSM 44205^T, DSM 44268^T, and DSM 44270^T suggesting a key role of ribose as chemoeffector in those strains at low concentrations (Galloway and Furlong, 1977).

In addition, *tad* (tight adherence) genes were unambiguously annotated in all the studied genomes (Supplementary Table S4). The *tad* genes encode the machinery required for the assembly of adhesive *Flp* (fimbrial low-molecular-weight protein) pili (Tomich et al., 2007). Those are essential for biofilm formation, colonisation, DNA transfer and pathogenesis and have implications in functions such as phage binding, rough colony morphology, and twitching motility (Proft and Baker, 2009). The presence of Tad complexes in *Geodermatophilaceae* genomes could be important for successful colonisation of altered stones and patina formation (Eppard et al., 1996; Urzì et al., 2001). Howbeit, the presence and abundance of *tad* genes varied amongst strains suggesting that strains may present different adherence-related phenotypes as previously demonstrated by Kachlany et al. (2000) and Perez et al. (2006).

Production of spores was not observed in any of the isolates under the given culture conditions. Nevertheless, the unambiguous presence of single copies of *SsgA*-like proteins (SALPs) in all studied genomes suggested that sporulation may be a typical feature previously overlooked in *Blastococcus* spp. and that, similar to *Geodermatophilus* spp., the presence of SALPs may be affected by growth conditions (Ishiguro and Wolfe, 1970). SALPs occur exclusively in morphologically complex actinomycetes where they play an important role in morphogenesis and control of cell division (Jakimowics and van Wezel, 2012) and are essential for sporulation (van Wezel et al., 2000). Furthermore, the presence of single SALPs (presumably and invariably *SsgB*) in *Blastococcus saxobsidens* DD2 and other *Geodermatophilaceae* representatives was already described by Girard et al. (2013) and is suggested as a link between the number of SALPs and the complexity of the developmental process (Traag and van Wezel, 2008), suggesting that these actinomycetes would produce single spores.

Colonies were opaque with a moist surface and regular margin and varied from coral (strain DSM 44268^T and DSM 44205^T) to brownish-orange in colour (Strains DSM 44270^T and DSM 44272^T and DSM 44269). Accordingly, phytoene synthase (*crtB;* EC 2.51.32) and phytoene desaturase (*crtI*; 4 step enzyme; EC 1.3.99.28, EC 1.3.99.29, EC 11.3.99.31), the key enzymes involved in the biosynthesis of the colourful carotenoids ζ-carotene, neurosporene, and lycopene (Paniagua-Michel et al., 2012) were identified in all the genomes. Yet, the putative operon mentioned by Sghaier et al. (2016) for *B.* sp. DD2 consisting of *hopC* [squalene-associated FAD-dependent desaturase (fragment)], *ispA* (generaylgeranyl pyrophosphate synthase), *shc* (squalene-hopene cyclise), *hpnH* (hopanoid biosynthetic associated radical SAM protein HpnH), was not identified in these genomes, except for crtB and ilvC (ketol-acid reductoisomerase; Supplementary Table S4).

Strains grew well on GYM *Streptomyces* and PYGV agar media. They tolerated temperature ranging from 15°C to 37°C, except strain DSM 44269, which only tolerated a more restricted range from 20°C to 37°C. All strains showed optimal growth at 25°C–30°C. In addition, growth was observed at pH 6.0–12.0 for strains DSM 44268^T and DSM 44269, at pH 5.0–11.5 for strain DSM 44205^T, at 5.0–11.0 for strain DSM 44270^T, and at 6.5–11.0 for strain DSM 44272^T. All strains had an optimal pH range from 6.5 to 8.0. Regarding tolerance to salt, strains DSM 44269 and DSM 44272^T were the least tolerant (0%–1% NaCl, although at 4% NaCl the Biolog system determined an ambiguous result for DSM 44272^T), followed by both DSM 44268^T and DSM 44205^T tolerating 0%–4% NaCl, and strain DSM 44270^T, which tolerated up to 8% NaCl. A summary of selected differential phenotypic characteristics is presented in Table 1 (for an overview of phenotypic profiles in *Blastococcus* see Supplementary Figure S5; Supplementary Table S5). Details about core metabolism will be provided later on in this manuscript.

Chemotaxonomy

Whole-cell hydrolysates of all strains showed *meso*-diaminopimelic (DAP) acid (Cell wall type III; Lechevalier and Lechevalier, 1970) as the diagnostic diamino acid being consistent with the family *Geodermatophilaceae* (Montero-Calasanz, 2020b) and genomic data (presence of *MurE* encoding for UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase [6.3.2.13] and absence of UDP-N-acetylmuramoyl-L-alanyl-D-glutamate—L-lysine ligase [6.3.2.7]; Supplementary Table S4). In addition, D-alanyl-D-alanine carboxypeptidase [3.4.16.4] was automatically annotated in all *Blastococcus* representative genomes except in *B. atacamensis* P6^T and in most *Geodermatophilaceae* representatives. The presence of diaminopimelate epimerase (*dapF*), responsible for the interconversion of the LL- and *meso*-isomers of DAP (Antia et al., 1957), was also annotated in all genomes including in the genome of *Sporichthya polymorpha* DSM 43042^T, whose cell wall is characterised by containing large amounts of LL-DAP (Trujillo and Normand, 2019). In this way, genomic data could easily be used to distinguish taxa with DAP-type peptidoglycan from those with Lys-type pepdioglycan but fail to differentiate those incorporating LL-DAP (e.g., *Sporichthyales*) from those incorporating *meso*-DAP (e.g., *Geodermatophilales*). Although the evolutionary explanation for the incorporation of one or another stereoisomer is still unclear, studies in *E. coli* suggest that the differentiation of DAP stereoisomers might be related to dimerization of the DAP epimerase (Mengin-Lecreulx et al., 1988; Hor et al., 2013).

As suggested by Hezbri et al. (2016) for the emendation of the genus *Blastococcus*, whole-cell sugar analysis displayed a basic pattern consisting of ribose, glucose, and mannose. Additionally, galactose was identified in all strains under study except in the profile of strain DSM 44205^T. Similarly, arabinose was found as part of the sugar cell wall composition in all studied strains excluding strain DSM 44270^T. Xylose was detected in strain DSM 44205^T.

The predominant menaquinone (MK) for strains DSM 44205^T, DSM 44268^T, DSM 44269, DSM 44270^T, and DSM 44272^T was MK-9(H_4; 75.9%, 92.6%, 56.7%, 78%, and 62.9%, respectively) in agreement with what was described for *Geodermatophilaceae*

TABLE 1 Phenotypic characteristics of strains DSM 44205^T, DSM 44268^T, DSM 44269, DSM 44270^T and DSM 44272^T in comparison with the validly named species in the genus *Blastococcus*.

Characteristics	1	2	3	4	5	6	7	8	9	10	11	12	13
Motility	–	–	–	–	+	+	–	–	–		+	+	–
Oxidation of:	+/–	+	+	+/–	+/–	+	+/–	–	+/–	+	+	+	–
Stachyose													
α-D-Lactose	–	+	+	–	–	+	+	+	+	–	–	+	+
D-Melibiose	+/–	+/–	+	+/–	–	+	–	+	+/–	+/–	+/–	+	+/–
N-Acetyl-D-Galactosamine	–	+	+/–	–	–	+/–	+/–	+/–	+/–	+/–	–	+/–	+/–
D-Mannose	–	–	+	+/–	+	+/–	–	+	–	+/–	+/–	+	+
L-Rhamnose	–	–	–	–	–	+/–	–	+/–	+	+/–	–	+	–
L-Aspartic acid	+/–	+/–	–	+	+/–	–	–	+	+/–	–	+/–	+/–	–
Quinic acid	–	–	+/–	–	+	+/–	+	+/–	+	+/–	+	+/–	–
L-Lactic acid	–	–	–	+	+	–	+/–	–	+/–	–	+	–	–
Citric acid	+	–	+	+	–	+	+/–	+/–	+/–	–	+/–	–	–
α-Keto-Glutaric acid	+	+	+	–	+/–	+	+	+	+	+	+	+	+/–
D-Malic acid	+/–	+	+	–	–	+	+	+	+/–	+	+	–	+
L-Malic acid	–	+	+	+	–	+/–	+/–	+	+/–	+	+	–	–
Polar lipids	DPG, PC, PE, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PG, PI	DPG, PC, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PG, PI	DPG, PC, PE, PI	DPG, PC, PE, PI	DPG, PC, PE, PI	DPG, PC, PE, PG, PI
Sugars	Ribose Xylose Arabinose Mannose Glucose	Rhamnose Ribose Arabinose Mannose Glucose Galactose	Ribose Arabinose Mannose Glucose Galactose	Ribose Mannose Glucose Galactose	Rhamnose Ribose Arabinose Glucose Galactose	Ribose Arabinose Mannose Glucose	Ribose Glucose Galactose	Rhamnose Ribose Glucose	Ribose Glucose Galactose	Rhamnose Xylose Ribose Mannose Glucose	Rhamnose Ribose Arabinose Mannose Glucose	Ribose Mannose Glucose Galactose	Ribose Glucose Galactose
Menaquinones[a#]	MK-9(H_4) MK-8(H_4)	MK-9(H_4) MK-9(H_0) MK-10(H_4)	MK-9(H_4) MK-9(H_2) MK-9(H_0) MK-10(H_4)	MK-9(H_4) MK-9(H_2) MK-8(H_4) MK-9(H_0) MK-9(H_6)	MK-9(H_4) MK-9(H_2) MK-9(H_0)	MK-9(H_4) MK-8(H_4)	MK-9(H_4) MK-9(H_2)	MK-9(H_4) MK-9(H_2) MK-9(H_0)	MK-9(H_4) MK-8(H_4) MK-9(H_0)	MK-9(H_4) MK-8(H_4) MK-9(H_0) MK-8(H_0)	MK-9(H_4) MK-9(H_0)	MK-9(H_4) MK-9(H_2) MK-8(H_4) MK-9(H_0) MK-9(H_6)	MK-9(H_4) MK-9(H_2) MK-8(H_4)
Fatty acids[b]	iso-$C_{16:0}$ anteiso-$C_{17:0}$ anteiso-$C_{17:1}$ iso-$C_{16:1}$ $C_{18:1}\omega9c$	$C_{18:1}\omega9c$ iso-$C_{16:0}$ iso-$C_{15:0}$ iso-$C_{16:1}$	$C_{18:1}\omega9c$ iso-$C_{16:0}$ iso-$C_{15:0}$ $C_{16:1}\omega7c$ $C_{16:0}$	iso-$C_{16:0}$ $C_{17:1}\omega8c$ iso-$C_{16:1}$ iso-$C_{15:0}$ $C_{18:1}\omega9c$	iso-$C_{16:0}$ $C_{18:1}\omega9c$ $C_{17:1}\omega8c$ iso-$C_{15:0}$ iso-$C_{16:1}$	iso-$C_{16:0}$ iso-$C_{16:1}$ $C_{18:1}\omega9c$	iso-$C_{16:0}$ iso-$C_{16:1}$ iso-$C_{15:0}$ 9-methyl-$C_{16:0}$	$C_{17:1}\omega8c$ $C_{16:1}\omega7c$ iso-$C_{15:0}$ iso-$C_{16:0}$ iso-$C_{16:1}$	iso-$C_{16:0}$ iso-$C_{16:1}$ $C_{17:1}\omega8c$ $C_{18:1}\omega9c$ iso-$C_{15:0}$	iso-$C_{16:0}$ iso-$C_{15:0}$ iso-$C_{16:1}$ $C_{16:1}\omega7c$	iso-$C_{16:0}$ $C_{18:1}\omega9c$ $C_{17:1}\omega8c$ iso-C15:0	iso-$C_{16:0}$ $C_{17:1}\omega8c$ iso-$C_{16:1}$ iso-$C_{15:0}$	iso-$C_{16:0}$ $C_{17:1}\omega8c$ iso-$C_{15:0}$ $C_{18:1}\omega9c$ iso-$C_{16:1}$

Strains: 1, strain DSM 44205^T; 2, strain DSM 44268^T; 3, strain DSM 44269; 4, strain DSM 44270^T; 5, strain DSM 44272^T; 6, *B. aggregatus* DSM 4725^T; 7, *B. atacamensis* NRRL B-65468^T; 8, *B. capsensis* DSM 46835^T; 9, *B. colisei* DSM 46837^T; 10, *B. endophyticus* DSM 45413^T; 11, *B. jejuensis* DSM 19597^T; 12, *B. saxobsidens* DSM 44509^T; 13, *B. xanthinilyticus* DSM 46842^T. +, positive reaction; –, negative reaction; +/–, ambiguous; DPG, diphosphatidylglycerol; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; PI, phosphatidylinositol; PC, phosphatidylcholine; MK, menaquinones.

10.3389/fmicb.2022.975365

(Montero-Calasanz et al., 2017). MK-9(H_0) was however also found in minor levels in the strains isolated from Atacama soils (DSM 44268^T, 0.7%; DSM 44269, 17.4%; DSM 44270^T, 7.4%; DSM 44272^T, 20.2%). Minor amounts of MK-9(H_2) were also observed in the MK profiles of strains DSM 44269, DSM 44270^T, and DSM 44272^T (1.6, 1.3, and 15.0%, respectively). MK-8(H_4) was also identified in strains DSM 44205^T (5.5%) and DSM 44270^T (2.9%). But MK-10(H_4) was only found in strains DSM 44268^T (4.7%) and DSM 44269 (0.6%) and MK-9(H_6) in strain DSM 44270^T (1.1%). The presence of quinones other than MK-9(H_4) is commonly observed in other representatives in *Geodermatophilaceae* (Montero-Calasanz et al., 2017; Montero-Calasanz, 2020b, 2021), although it is worth mentioning that this is the first report of MK-10(H_4) in members of the genus *Blastococcus*. The biosynthesis of menaquinones, *via* chorismate derived from Shikimate pathway (genes: *MenA-G*), was confirmed by genomic data in the order *Geodermatophilales*. Curiously, the futalosine pathway, the other major type of menaquinone biosynthetic pathway (Seto et al., 2008) was also annotated in genomes of *Cryptosporangium arvum* and *Crytosporangium aurantiaca* type strains (Supplementary Table S4). *MenJ* encoding for menaquinone-9-β-reductase [EC:1.3.99.38] and implicated in menaquinone side chain saturation (Upadhyay et al., 2015) was misidentified by databases as a geranylgeranyl reductase family protein but was verified in all genomes after BLAST using UniProtKB:P9WNY8 (MENJ_MYCTO). Moreover, the presence of 1,4-dihydroxy-2-naphthoate polyprenyltransferase [EC:2.5.1.74] and demethylmenaquinone methyltransferase [EC:2.1.1.163] in all genomes confirmed the potential of *Geodermatophilales* representatives to synthesise MK 8, 9, and 10. Regarding the elongation of isoprenoid side chains, genes *uppS* and *hepST*, encoding for ditrans,polycis-undecaprenyl-diphosphate synthase ((2E,6E)-farnesyl-diphosphate specific) and polyprenyl diphosphate synthase (annotated here as geranylgeranyl pyrophosphate synthase), respectively, were annotated in all genomes. But the length and the degree of saturation of the C-3 isoprenyl side chains were impossible to predict. MKs consist of repeated isoprene subunits and the exact number is determined by the synthase encoded by the particular microbe and a microbe-dependent molecular ruler mechanism that involves bulky amino acid residues blocking the enzyme active sites to stop chain elongation (Han et al., 2015). The inability to predict the isoprenyl chain length of menaquinones was already reported by Baek et al. (2018).

Regarding fatty acid profiles, the five strains qualitatively showed similar patterns to those found in other representatives in *Blastococcus* (Table 1), but the predominant fatty acids significantly varied amongst them. Dominant fatty acids for strains DSM 44268^T and DSM 44269 (>5%) were $C_{18:1}\omega9c$ (36.2% and 37.2%, respectively), iso-$C_{16:0}$ (19.3% and 11.3%, respectively), iso-$C_{15:0}$ (6.2% and 8.6%, respectively), and $C_{16:1}\omega7c$ (7.8% and 7.2%, respectively). Additionally, DSM 44268^T showed iso-$C_{16:1}$ H (5.4%) and DSM 44269 displayed some minor amounts of $C_{16:0}$ (5.6%) and $C_{17:1}\omega8c$ (6.8%). Those latter fatty acids were also present in DSM 44268^T and iso-$C_{16:1}$ H in DSM 44269 but at levels below 5%. On the other hand, strain DSM 44270^T had the major fatty acids (>5%) iso-$C_{16:0}$ (32.0%), $C_{17:1}\omega8c$ (27.0%), iso-$C_{16:1}$ H (8.7%), iso-$C_{15:0}$ (7.3%), and $C_{18:1}\omega9c$ (6.0%). A similar composition was observed for strain DSM 44272^T although those quantitatively varied as follows: iso-$C_{16:0}$ (24.0%), $C_{18:1}\omega9c$ (23.0%), $C_{17:1}\omega8c$ (13.7%), iso-$C_{15:0}$ (10.8%), and iso-$C_{16:1}$ H (8.1%). Finally, the major fatty acids of strain DSM 44205^T (>5%) consisted of iso-$C_{16:0}$ (23.4%), anteiso-$C_{17:0}$ (19.8%), and anteiso-$C_{17:1}$ C (11.9%), iso-$C_{16:1}$ H (11.1%), and $C_{18:1}\omega9c$ (6.7%). The complete gene set (*Fab* cluster) related to fatty acids metabolism (type II fatty acid synthase (FAS II) system) was present in all the studied genomes. Similar to the biosynthesis of MKs, fatty acid synthesis occurs *via* recurring reactions. It is thus not possible to predict just by using genomic data modifications that result in desaturations and/or elongations (López-Lara and Geiger, 2019).

The major polar lipids were diphosphatidylglycerol (DPG), phosphatidylethanolamine (PE), phosphatidylcholine (PC), and phosphatidylinositol (PI; Figure 3). It was consistent with those already described by Hezbri et al. (2016). Significant amounts of phosphatidylglycerol (PG) were also identified in all the isolates except in strain DSM 44205^T. The inconsistent presence of PG was already noted for other species in the family (Montero-Calasanz et al., 2014, 2015, 2017; Hezbri et al., 2015a,b, 2016). Similarly, the characteristic glycophosphoinositol (GPI; annotated as GPL in Hezbri et al., 2016) as well as the reproducible presence of two unidentified polar lipids (PL2 and PL3) was detected in strains DSM 44268^T, DSM 44269, DSM 44270^T, DSM 44272^T but not in DSM 44205^T. However, a minor amount of an unidentified glycolipid (GL4) was found in the polar lipids pattern of strain DSM 44205^T. Three different ones (GL1, GL2, and GL3) were also identified in the profile of strain DSM 44268^T but not in strain DSM 44269. In contrast, an unidentified aminolipid (AL1) was displayed by strain DSM 44269 but not by DSM 44268^T. One having a similar chromatographic mobility was present in strain DSM 44270^T, too. Genomic data were consistent with those obtained experimentally with some minor discrepancies. In particular, the enzymes phosphatidylcholine synthase and cardiolipin synthase [EC 2.7.8.41] required in the synthesis of PC and cardiolipin (DPG), respectively, were present in all the genomes. As expected, phosphatidylcholine synthase was missing in *Modestobacter* and *Klenkia* type species. CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase [EC 2.7.8.50] involved in the synthesis of phosphatidyl-glycerophosphate, and also in fatty acids metabolism, was present in all species although phosphatidylglycerophosphatase [EC 3.1.3.27] required to the final conversion to PG was just annotated in DSM 44272^T and in *B. litoris*. Phosphatidylserine descarboxylase [EC 4.1.1.65] required to synthesise phosphatidylethanolamine was not annotated in *M. multiseptatus*. CDP-diacylglycerol—inositol 3-phosphatidyltransferase [EC 2.7.8.11] needed for the synthesis of PI was annotated in all genomes. Finally, some other enzymes involved in the biosynthesis of other polar lipids were also annotated in the studied genomes suggesting that those may

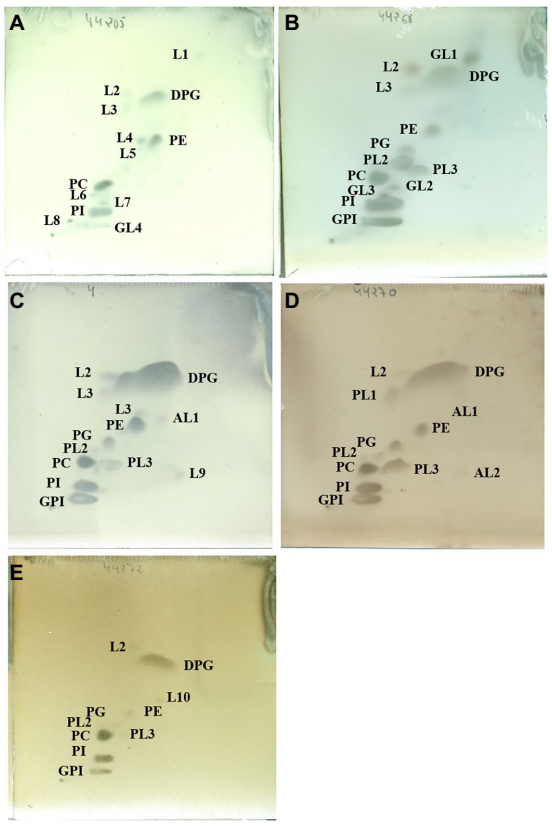


FIGURE 3
Polar lipids profile of strains DSM 44205^T **(A)**, DSM 44268^T **(B)**, DSM 44269 **(C)**, DSM 44270^T **(D)** and DSM 44272^T **(E)** after separation by two-dimensional TLC using the solvents chloroform:methanol:water (65:25:4; v:v:v) in the first dimension and chloroform:methanol:acetic acid:water (80:12:15:4; v:v:v:v) in the second one. Plates were sprayed with molybdatophosphoric acid (3.5%; Merck™) for detection of the total polar lipids. DPG, diphosphatidylglycerol; PG, phosphatidylglycerol; PE, phosphatidylethanolamine; PI, phosphatidylinositol; PC, phosphatidylcholine; GPI, glycophosphatidylinositol; PL1-3, Phospholipid; GL1-4, glycolipid; AL1-2, aminolipid; L1-10, unidentified lipids. All data are from this study.

correspond to those minor unidentified lipids shown on the TLC plate.

Carbon metabolism

As expected, and in agreement with results determined by GEN III Biolog System (Supplementary Table S5), *Blastococcus* spp. show aerobic microbial respiration potentially driven by glycolysis, the tricarboxylic acid (TCA) cycle, and oxidative phosphorylation. In addition, CO_2 could be assimilated *via* C4-dicarboxylic acid cycle by phosphoenolpyruvate carboxylase [EC 4.1.1.31], which converts phosphonolpyruvate coming from pyruvate metabolism, glucolysis or gluconeogenesis into oxaloacetate, and is present in all species except DSM 44268^T and *B. aggregatus*. A second mechanism to fixate CO_2 based on the reductive citric acid is also observed in all species (i.e., a molecule of CO_2 is incorporated along succinyl-CoA, *via* 2-oxoglucarate/2-oxoacid ferredoxin oxidoreductase [EC 1.2.7.3], which in turn will be converted into isocitrate *via* isocitrate dehydrogenase [EC 1.1.1.42], using another molecule of CO_2). The presence of carbonic anhydrases (CAs, EC 4.2.1.1), which are involved in the catalysis of CO_2 to bicarbonate and vice versa (Supuran and Capasso, 2017) is found in all the genomes. Finally, the Wood-Ljungdahl pathway, which assimilates CO_2 into Acetyl-CoA *via* the complex carbon monoxide deshydrogenase/acetyl-CoA synthase (CODH/ACS), may potentially be present. A complete tetrahydrofolate (THF) methyl-branch along with the presence of carbon monoxide dehydrogenases, required for the carbonyl-branch, were annotated in all genomes. These microorganisms could also be able to produce acetate as end product (acetogenesis). But no acetyl-coA synthase [2.3.1.169] homologue was identified in this study. And neither an autotroph nor an acetogen was described in the family so far. Overall, the presence of a number of mechanisms related to chemolithoautotrophy denotes that *Blastococcus* representatives could play a key role in the global carbon cycle. Microorganisms able to efficiently capture and store CO_2 are prime candidates to be used as biological approaches in the reduction of the atmospheric CO_2 concentration and as production platforms for a wide range of bioproducts from CO_2 (Katsyv and Müller, 2020).

Apart from D-glucose as carbon source, genomics data identified other potential carbon sources such as fructose (fructokinase mediated: EC 2.7.1.4); α-galactose (*galk*, galactokinase); D-galactose (*galM*, Aldose 1-epimerase; EC 5.1.3.3); which is absent in DSM 44205^T; L-gulose (*gnl*, RGN gluconolactonase; EC 3.1.1.17); 1-butanol (alcohol deshydrogenase; EC 1.1.1-), only present in *B. aggregatus*; D-mannose-6P/1P (*manB* phosphomannomutase; EC 5.4.2.8 /*manC* mannose-1-phosphate guanylyltransferase; EC 2.7.7.13), absent in *B. aggregates*; glycine *via* GDC glycine dehydrogenase and glycine hydroxymethyltransferase (*glyA*, EC 2.1.2.1); L-glutamate (*gdhA*, glutamate dehydrogenase; EC 1.4.1.3); L-serine (glycine hydroxymethyltransferase; *glyA*, EC 2.1.2.1), isocitrate (isocitrate dehydrogenase), absent in DSM 44272^T; formamide (formamidase; EC 3.5.1.49) present in DSM 44205^T, DSM 44272^T, and *B. aggregatus*; trehalose (*treS*, maltose α-D-glucosyltransferase; EC 4.99.16 and D-glucose α,α-trehalose phosphorylase; EC 2.4.1.65); L-lactate (*lldD*, L-lactate dehydrogenase; EC 1.1.2.3), just present in *B. endophyticus*; sucrose (*malZ*, α-glucosidase; EC 3.2.1.20); maltose (maltokinase; EC 2.7.1.175), absent in *B. xanthinilyticus*; isomaltose (oligo-1,6-glucosidase), amongst others, that may also be utilised by *Blastococcus* strains. These results supported the presence of a rich carbohydrate metabolism in the genus and were largely in correlation with those showed by GEN III Biolog and API System (See Supplementary Figure S5; Supplementary Tables S4, S5).

Additionally, *Blastoccocus* genomes revealed extracellular enzymes involved in the degradation of complex carbohydrates, which are closely linked to the decay of organic residues, transformation of native soil organic matter, mineralisation of plant nutrients, and soil aggregation (Balezentiene, 2012). For example,

α-glucosidases [EC 3.2.1.20], involved in the cellulose hydrolysis, were observed in all studied genomes (three copies in *B. xanthinilyticus*); β-glucosidases [EC 3.2.1.21] were also annotated in all of them except in *B. colisei* and *B. xanthinilyticus,* having three gene copies in DSM 44205^T, DSM 44268^T, and *B. saxobsidens;* Moreover, all studied genomes contained α-amylases [EC 3.2.1.1], extracellular enzymes handling the almost complete starch saccharification (Tu and Miles, 1976), but enzymatic activity was not recorded in any of them according to the APIM ZYM system; the presence of glycosidase (glycoside hydrolase; annotated as α-amylase by Pfam database) was, nevertheless, only annotated in *B. endophyticus;* Finally, chitinases [EC 3.2.1.14] were annotated in *B. endophyticus* and *B. saxobsidens.* Cellulases (β-1,4-endoglucan hydrolase; EC 3.2.1.4) and 1,4- β- cellobiohydrolase [EC 3.2.1.91] were absent. On the other hand, lipolytic enzymes are an important group of biotechnologically relevant enzymes presenting a key role in regulating the levels of hydrocarbons in soil (Gupta et al., 2004). Extracellular lipases such as the well-known triacylglycerol acylhydrolase (here as tryacylglycerol lipase, EC 3.1.13) was only annotated in DSM 44268^T. In addition, lysophospholipases [EC 3.1.1.5] and monoglyceride lipases (EC 3.1.1.23) were present in all the genomes analysed except DSM 44272^T and *B. endophyticus*. Polyhydroxyalkanoic acid (PHA) synthases (EC 2.3.1.-), the key enzyme in the biosynthesis of PHAs, a class of aliphatic polyesters that are generally regarded as a carbon and energy reserve material in bacteria and archaea (McCool and Cannon, 2001), were also identified in all the genomes in a number of copies that ranged from one to six with the exception of DSM 44272^T. Again, large correlations but also discrepancies were observed between genomic and experimental data. For example, the hydrolysis of xanthine by *B. xanthinilyticus,* previously described by Hezbri et al. (2018), was confirmed by the presence of xanthine deshydrogenase but experimentally it could not be proven for strains DSM 44205^T and DSM 44268^T and *B. endophyticus* although such enzyme was also annotated in their genomes. Similar inconsistencies were also identified for other unambiguously annotated enzymes such as α-glucosidase which was not expressed under standardised experimental conditions in strains DSM 44205^T, DSM 44268^T, DSM 44270^T, DSM 44272^T, and *B. aggregatus* or β-glucosidase which was only expressed by *B. endophyticus* and *B. saxobsidens* (for more examples see Supplementary Figure S5; Supplementary Tables S4, S5).

Final remarks and taxonomic consequences

The integration of genomic information into microbial systematics have proven to be a reproducible, reliable and highly informative approach to reveal phylogenetic relationships amongst prokaryotes (Chun and Rainey, 2014; Chun et al., 2018). Montero-Calasanz et al. (2017) already applied phylogenomics in combination with the identification of diagnostic features in the taxonomic rearrangement of *Geodermatophilaceae.* In light of new genomic information now available, the taxonomic status of the family was revisited in this study. The whole proteome-based phylogenomic tree of the sequenced type strains and strains under study along with phenotypic tests showed with strong support that DSM 44268^T, DSM 44270^T, DSM 44272^T, and DSM 44205^T represent novel species within the genus *Blastococcus.* Additionally, the non-monophyly of *Blastococcus, Geodermatophilus* and *Modestobacter* was revealed, i.e., *B. endophyticus, G. daqingensis* and *M. deserti* formed separate lineages within *Geodermatophilaceae*. Based on principles of phylogenetic systematics and applying taxonomic conservatism, we propose the following reclassifications: *Blastococcus endophyticus* as the type species of the new genus *Trujillonella* gen. nov., which the name *Trujillonella endothytica* comb. nov is given; *Geodermatophilus daqingensis* as the type species of the new genus *Pleomorpha* gen. nov., which the name *Pleomorpha daqingensis* comb. nov is given; *Modestobacter deserti* gen. nov., as the type species of the new genus *Goekera* which the name *Goekera deserti* comb. nov is given. Finally, based on the additional taxonomic data now available we propose the emendation of *B. aggregatus, B. jejuensis, B. saxobsidens* and *B. Xanthinilyticus* to include accession numbers for the whole genome sequences, genome size and genomic G + C content and insights. The use of genomic data for the characterisation of novel species in *Geodermatophilaceae* also revealed interesting insights regarding their metabolism and ecology. In particular, the number of 16S rRNA and tRNA gene copies was suggested to be related to their slow growth speed and their life strategy supporting their prevalence in oligotrophic environments such as desert soils and decayed monuments (Neilson et al., 2012; Giongo et al., 2013). In addition, the presence of a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly in all representatives of the family highlighted potential yet unobserved features that would enable us to better understand their metabolism, morphogenesis and ecology. Combining high-throughput phenotype analyses with genomics also supported the uncovering of a rich carbon metabolism in *Blastococcus* spp. Results obtained from GEN III microplates and API System were largely correlated with what was observed in the genomes. In this way, *Blastococcus* spp. were described as aerobic bacteria with the potential to present anaerobiosis and fixate CO_2 by using four potential pathways and to degrade a range of complex carbohydrates. Such metabolic versatility would support their success to colonise a variety of biotopes and a potential key role in the global carbon cycle and as biological tools to reduce the atmospheric CO2 concentration and produce bioproducts from CO_2.

Even though chemotaxonomic procedures marked a turning point in prokaryotic systematics (Goodfellow et al., 2012), results can be influenced by cultivation conditions (i.e., fatty acid analysis) or by inexperienced staff (i.e., regarding the interpretation of polar lipids patterns on 2-D TLC plates; Sutcliffe et al., 2013). In favour of a revitalization of bacterial taxonomy, *in silico* chemotaxonomic analyses were already used in some studies to complement and, in some cases, replace characteristics traditionally determined in the laboratory (Amaral et al., 2014; Fotedar et al., 2020;

Lawson et al., 2020). Here, we attempted to determine *in silico* cell-wall peptidoglycan, respiratory quinone, fatty acids and polar lipid patterns of *Geodermatophilaceae*. Overall, genomic data were consistent with the results obtained from the laboratory. Nevertheless, the amount of information that can be predicted based on genomic data varied depending on the chemotaxonomic trait. For instance, the *in silico* polar lipid analyses were largely in line with those obtained experimentally and allowed us to distinguish at the genus level by the presence/absence of the gene encoding for phosphatidylcholine synthase. On the other hand, the presence of the DAP as the diagnostic diamino acid over L-lysine in the whole-cell hydrolysates was confirmed but the *in silico* procedure failed in differentiating microorganisms incorporating LL-DAP rather than *meso*-DAP (discrimination at order level). Regarding both MK and fatty acids analysis, final products could not be predicted due to the nature of biosynthetic pathways. Nevertheless, sufficient information would, for example, be gathered from Shikimate pathway along with the absence of futalosine pathway to establish certain taxonomic discrimination at order level. The use of genomic data for species delineation and the impractical value of chemotaxonomic data when genomic data are available are widely proven (Nouioui et al., 2018). Nevertheless, as argued by Vandamme and Sutcliffe (2021), knowledge of cellular components can be valuable to establish discriminative thresholds when applied at taxonomic ranks above the species level. However, improvements in analytical methods leading to the use of automated and higher-resolution technologies such as mass spectrometry lipidomic methods (Rashid et al., 2017) should first be introduced and deployed. Our results here proved the use of *in silico* chemotaxonomic analysis as a simple and fast approach to complement phylogenomic results and to provide sufficient resolution to support the affiliation of novel isolates into appropriate taxonomic groups. Its routine application would undoubtedly increase the speed of taxon descriptions and the accuracy and consistency of taxonomic reports.

In summary, the integration of genomics into the systematics of *Geodermatophilaceae* not only allowed for more stable and reliable taxonomic arrangements but also showed the feasibility for a potential replacement of wet-lab chemotaxonomy. Second, it provided deeper insights into the molecular mechanisms behind phenotypic features and revealed potentially overlooked ones that could be key to an understanding of their evolution, ecology and biotechnological potential.

Description of *Trujillonella* gen. nov.

Tru.jil.lo.nel'la. N.L. fem. n. *Trujillonella,* named in honour of Martha E. Trujillo in recognition of her contributions to microbial systematics, mainly on Actinobacteria, on Bergey's Manual trust, and as the Editor-in-chief of the International Journal of Systematic and Evolutionary Microbiology.

Cells are aerobic, non-motile, non-spore-forming, Gram-stain positive, catalase-positive and oxidase-negative. Cells occur singly, in pairs or in tetrads, often tending to form aggregates. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4), with MK-8 and MK-9(H_6) as minor components. The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylcholine, phosphati dylethanolamine, and phosphatidylinositol. The major fatty acids are iso-$C_{16:0}$, iso-$C_{15:0}$ and $C_{18:1}\omega9c$. The basic whole-cell sugar pattern includes arabinose and galactose. The genomic G + C content is 71%–72%.

The type species of *Trujillella* is *Trujillella endophytica*, sp. nov.

Description of *Trujillella endophytica* comb. nov.

T. en.do.phy'ti.ca (Gr. pref. *Endo-,* within; Gr. neut. n. *phyton,* plant; L. fem. Adj. suff. *–ica,* adjectival suffix used with the sense of belonging to; N.L. fem. Adj. *endophytica,* within plant, endophytic, pertaining to the isolation from plant tissues).

Basonym: *Blastococcus endophyticus* Zhu et al. (2013) emend. Hezbri et al. (2016).

The properties are as given in the species description by Zhu et al. (2013) and emendation by Hezbri et al. (2016) with the following modification. The genomic G + C content is 74.6%. The genome size is 4.9 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain DSM 45413^T is FOEE00000000.

The type strain YIM 68236^T = CCTCC AA 209045^T = DSM 45413^T = KCTC 19998^T was isolated from healthy leaves of *Camptotheca acuminate* collected in Yunnan Province, south-west China.

Description of *Pleomorpha* gen. nov.

Ple.o.mor'pha. Gr. adv. *Pleon* more; Gr. fem. n. *morphe*, shape or form; N.L. fem. n. *Pleomorpha*, organism showing multiple forms.

Pleiomorphic, motile, spore-forming, aerobic, Gram-stain positive cells. Those occurs singly or associated in aggregates. The peptidoglycan in the cellwall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4). The basic polar lipid profile contains diphosphatidylglycerol, phosphati dylcholine, phosphatidylglycerol, phosphatidylethanolamine, and phosphatidylinositol. The major fatty acids are iso-$C_{16:0}$ and iso-$C_{15:0}$. The basic whole-cell sugar pattern includes galactose, glucose and xylose. The genomic G + C content is 73%–74%. The type species of *Pleomorpha* is *Pleomorpha daqingensis*, sp. nov.

Description of *Pleomorpha daqingensis* comb. nov.

P. da.qing.en'sis (N.L. masc./fem. Adj. *daqingensis*, pertaining to Daqing city, China, where the type strain was isolated).

Basonym: *Geodermatophilus daqingensis* Wang et al. (2017).

The properties are as given in the species description by Wang et al. (2017) with the following modification. The genomic G + C content is 73.6%. The genome size is 5.4 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain DSM 104001^T is JACBZT000000000.

The type strain WT-2-1^T = CGMCC 4.7381^T = DSM 104001^T was isolated from pretroleum-contaminated soil in Daqing city, China.

Description of *Goekera* gen. nov.

Goe'ke.ra. N.L. fem. n. Goekera, named in honour of Markus Göker in recognition of his contributions to microbial systematics, including work on *Actinobacteria*, on the List of Prokaryotic Names with Standing in Nomenclature (LPSN), and as a member of the Judicial Commission.

Cells are motile, non-spore-forming, aerobic, Gram-stain positive, catalase and oxidase positive cocci and/short rods. Bud-like structure was observed for some cells. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4), with MK-8(H_4) as a minor component. The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylinositol, phosphatidylmethylethanolamine and phosphatidylinositol mannoside. The major fatty acids are $C_{18:1}\omega9c$, iso-$C_{16:0}$, $C_{16:0}$, iso-$C_{15:0}$, and $C_{16:1}\omega7c$. The basic whole-cell sugar pattern includes arabinose, glucose and ribose. The genomic G + C content is 74%–75%. The type species of *Goekera* is *Goekera deserti*, sp. nov.

Description of *Goekera deserti* comb. nov.

G. de.ser'ti (L. gen. Neut. n. *deserti*, of a desert, where the organisms were acquired).

Basonym: *Modestobacter deserti* Jiang et al. (2021).

The properties are as given in the species description by Jiang et al. (2021) with the following modification. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain CPCC 205119^T is JAAGWK000000000.

The type strain CPCC 205119^T = I12A-02624^T = KCTC 49201^T = NBRC 113528^T was isolated from moss-dominated soil crusts collected from Shapotou NDER in Tengger Desert, China.

Emended description of *Blastococcus aggregatus*

The properties are as given in the species description by Ahrens and Moll (1970) and emendations by Urzì et al. (2004) and Hezbri et al. (2016) with the following emendation. The genomic G + C content is 73.3%. The genome size is 4.6 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle, and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 4725^T is OBQI00000000.

The type strain B15^T = ATCC 25902^T = DSM 4725^T = JCM 12602^T = NCIMB 1849^T was isolated from Baltic Sea.

Emended description of *Blastococcus jejuensis*

The properties are as given in the species description by Lee (2006) and emendation by Hezbri et al. (2016) with the following emendation. The presence of phosphatidylmethylethanolamine listed by Lee (2006) and confirmed by Hezbri et al. (2016) was incorrectly annotated by the latter and it should be recognised as hydroxiphosphatidylethanolamine accordingly to its chromatographic mobility.

The type strain KST3-10^T = DSM 19597^T = JCM 15614^T = KCCM 42251^T = NRRL B-24440^T was isolated from sand sediment of a beach in Jeju, Korea.

Emended description of *Blastococcus saxobsidens*

The properties are as given in the species description by Urzì et al. (2004) and emendation by Hezbri et al. (2016) with the following emendation. The genomic G + C content is

73.5%. The genome size is 4.5 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 44509^T is SHKV01000001.

The type strain BC444^T = DSM 44509^T = JCM 13239^T = NRRL B-24246^T was isolated from from the surface of marble and calcareous stones in Italy.

Emended description of *Blastococcus xanthilyniticus*

The properties are as given in the species description by Hezbri et al. (2018) with the following emendation. The genomic G + C content is 74.4%. The genome size is 4.6 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 46842^T is VNHW00000000.

The type strain BMG 862^T = DSM 46842^T = CECT 8884^T was isolated from a marble sample collected from the Bulla Regia monument, Northern Tunisia.

Description of *Blastococcus aurantiacus* sp. nov.

Blastococcus aurantiacus (au.ran.ti.a'cus. N.L. masc. Adj. *aurantiacus* orange-coloured, referring to the orange colour of the colonies).

Colonies are bright red orange-coloured, opaque with a dry surface and regular margin. Cells are Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 µm in diameter) with a tendency to form aggregates. It grows in aerobiosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and 6.0–12.0 (optimum 6.5–8.0), respectively. NaCl is not needed for growth. It can grow between 0% and 4% NaCl (w/v; optimal range) but not at 8%. It grows well on GYM *Streptomyces*, R2A (DSMZ Medium 830), tryticase soy agar (TSA), Luedemann (DMSZ medium 877), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase lipase (C8), lipase (C14), leucine arylamidase, valine acrylamidase, cysteine acrylamidase, trypsin, and α-galatosidase. According to the Biolog System, it oxidises: dextrin, D-maltose, D-trehalose, D-cellobiose, *β*-gentiobiose, sucrose, stachyose, *α*-D-lactose, *β*-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, N-acetyl-*β*-D-mannosamine, N-acetyl-D-galactosamine, 3-*O*-methyl-D-glucose, D-fucose, L-fucose, glycerol, D-glucose-6-phosphate, rifamycin SV, D-galacturonic acid, D-gluconic acid, D-glucuronic acid, D-saccharic acid, *α*-keto-glutaric acid, D-malic acid, L-malic acid, nalidixic acid, lithium chloride, potassium tellurite, tween 40, *γ*-amino-n-butyric acid, acetoacetic acid, propionic acid, acetic acid, aztreonam, butyric acid, and sodium bromate but not D-mannose, D-galactose, L-rhamnose, sodium lactate, fusidic acid, D-mannitol, D-arabitol, D-aspartic acid, D-serine, troleandomycin, minocycline, glycine-proline, L-alanine, L-arginine, L-glutamic acid, L-histidine, L-pyroglutamic acid, L-serine, lincomycin, guanidine hydrochloride, niaproof, pectin, L-galactonic acid-*γ*-lactone, glucuronamide, mucic acid, quinic acid, vancomycin, tetrazolium violet, tetrazolium blue, *p*-hydroxy-phenylacetic acid, methyl pyruvate, L-lactic acid, citric acid, bromo-succinic acid, *β*-hydroxy-butyric acid, *α*-keto-butyric acid, and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are $C_{18:1}\omega 8c$, iso-$C_{16:0}$, iso-$C_{16:1}$ H, iso-$C_{15:0}$, and iso-$C_{16:1}$ H. The cell wall peptidoglycan contains *meso*-diaminopimelic acid. Galactose, glucose, and arabinose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol, glycophosphatidylinositol, two unidentified phospholipids, and three unidentified gylcolipids. The high-quality draught genome of strain DSM 44268^T was resolved to 18 scaffolds consisting of 4,163,046 bp, with a G + C content of 73.6%, 4,044 candidate protein-coding genes, 47 tRNA genes, and eight rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain AT 7-1^T (=DSM 44268^T = JCM 18931^T) isolated from soil in the Atacama Desert (Chile) is MH479060. The IMG accession number for the whole genome sequence of strain DSM 44268^T is 2599185358.

Description of *Blastococcus fimeti* sp. nov.

Blastococcus fimeti (fi.me'ti. L. neut. Gen. n. *fimeti*, of or from a dunghill).

Colonies are pink red-coloured, opaque with a greasy surface and irregular margin. Cells are Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 μm in diameter) with a tendency to form aggregates. It grows in aerobiosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative and aesculin positive. Temperature and pH ranges are 15°C–37°C (optimal range) and 6.0–12.0 (optimum 7.0–9.0), respectively. NaCl is not needed for growth. It can grow between 0–4% NaCl (w/v; optimal range) but not at 8%. It grows well on GYM *Streptomyces*, R2A (DSMZ Medium 830), tryticase soy agar (TSA), Luedemann (DMSZ medium 877), GPHF (DSMZ medium 553), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase lipase (C8), lipase (C14), leucine arylamidase, valine arylamidase, *α*-fucosidase. According to the Biolog System, it oxidises: dextrin, D-maltose, D-trehalose, D-cellobiose, sucrose, turanose, *β*-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, N-acetyl-*β*-D-mannosamine, D-glucose, 3-*O*-methyl-D-glucose, L-fucose, glycerol, D-glucose-6-phosphate, rifamycin SV, L-glutamic acid, methyl pyruvate, D-lactic acid methyl ester, citric acid, *α*-keto-glutaric acid, nalidixic acid, lithium chloride, potassium tellurite, tween 40, γ-amino-*n*-butyric acid, *β*-hydroxy-butyric acid, *α*-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, aztreonam, butyric acid, and sodium bromate but not D-raffinose, *α*-D-lactose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, D-mannose, L-rhamnose, fusidic acid, D-serine (inhibitory concentration), D-arabitol, myo-inositol, D-fructose-6-phosphate, D-aspartic acid, D-serine, troleandomycin, minocycline, glycine-proline, L-alanine, L-arginine, L-histidine, L-pyroglutamic acid, L-serine, lincomycin, guanidine, hydrochloride, niaproof, L-galactonic acid-*γ*-lactone, mucic acid, quinic acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, L-lactic acid, L-malic, acid, bromo-succinic acid and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-$C_{16:0}$, anteiso-$C_{17:0}$, anteiso-$C_{17:1}$ C, iso-$C_{16:1}$ H and $C_{18:1}\omega$9c. The cell wall peptidoglycan contains *meso*-diaminopimelic acid. Ribose, xylose, arabinose, mannose, and glucose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylcholine and phosphatidylinositol. The high-quality draught genome of strain DSM 44205^T was resolved to 30 scaffolds consisting of 4,044,261 bp, with a G + C content of 73.4%, 3,853 candidate protein-coding genes, 70 tRNA genes, and nine rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain G1S^T (=DSM 44205^T = CECT 8406^T) isolated from seepage water of a dumping ground in Vancouver (Canada) is MH479059. The IMG accession number for the whole genome sequences of strain DSM 44205^T is 2599185193.

Description of *Blastococcus haudaquaticus* sp. nov.

Blastococcus haudaquaticus (hau.da.qua'ti.cus. L. adv. *Haud*, not at all, by no means; L. masc. Adj. *aquaticus*, living, growing, or found in or by the water, aquatic; N.L. masc. Adj. *haudaquaticus*, growing far away from or without any water).

Colonies are reddish-brown coloured, opaque with a mucoid surface and regular margin. Cells are aerobic, Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 μm in diameter) with a tendency to form aggregates. It grows in aerobiosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and pH 5.0–11.0 (optimum 6.5–8.0), respectively. NaCl is not needed for growth. It can grow between 0% and 8% NaCl (w/v; optimal range). It grows well on GYM *Streptomyces*, R2A (DSMZ Medium 830), tryticase soy agar (TSA), Luedemann (DMSZ medium 877), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase lipase (C8), lipase (C14), leucine arylamidase, valine acrylamidase, trypsin, and cystine acrylamidase. According to the Biolog System, it oxidises: dextrin, D-maltose, D-trehalose, D-cellobiose, *β*-gentiobiose, sucrose, turanose, *β*-methyl-D-glucoside, D-salicin, N-acetyl-D-glucosamine, D-glucose, D-fructose, sodium lactate, D-sorbitol, glycerol, rifamycin SV, L-aspartic acid, L-glutamic acid, L-histidine, mucic acid, L-lactic acid, citric acid, L-malic acid, lithium chloride, potassium tellurite, *γ*-amino-*n*-butyric acid, propionic acid, acetic acid, sodium formate, aztreonam, butyric acid, and sodium bromate but not *α*-D-lactose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, 3-*O*-methyl-D-glucose, D-fucose, L-fucose, L-rhamnose, inosine, fusidic acid, D-serine (inhibitory concentration), D-arabitol, myo-inositol, D-aspartic acid, D-serine, troleandomycin, minocycline, glycine-proline,

L-arginine, L-pyroglutamic acid, lincomycin, guanidine hydrochloride, niaproof, L-galactonic acid-γ-lactone, D-glucuronic acid, glucuronamide, quinic acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, D-lactic acid methyl ester, α-keto-glutaric acid, D-malic acid, bromo-succinic acid, nalidixic acid, α-keto-butyric acid, and acetoacetic acid. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-$C_{15:0}$, iso-$C_{16:1}$ H, iso-$C_{16:0}$, $C_{17:1}\omega8c$, $C_{18:1}\omega9c$, and $C_{16:1}\omega7c$. The cell wall peptidoglycan contains *meso*-diaminopimelic acid. Ribose, mannose, galactose, and glucose are the whole-cell sugars. Polar lipids consist of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylcholine, and phosphatidylinositol. The high-quality draught genome of strain DSM 44270^T was resolved to 11 scaffolds consisting of 4,512,672 bp, with a G + C content of 73.29%, 4,410 candidate protein-coding genes, 48 tRNA genes, and six rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain AT 7-14^T (=DSM 44270^T = JCM 18932^T) isolated from soil in the Atacama Desert (Chile) is MH479062. The IMG accession number for the whole genome sequences of strain DSM 44270^T is 2728369258.

Description of *Blastococcus mobilis* sp. nov.

Blastococcus mobilis (mo'bi.lis. L. masc. Adj. *mobilis* movable, motile).

Colonies are reddish-brown coloured, opaque with a mucoid surface and regular margin. Cells are aerobic, Gram-reaction-positive and catalase and oxidase negative cocci (0.6–1.6 µm in diameter) with a tendency to form aggregates. It grows in aerobiosis but, according to genomic data, anaerobiosis and acetogenesis may occur. Reproduction by budding is predominant but binary fission is also observed. Cells are non-motile and non-spore-forming but a repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Degradation for casein, tyrosine, starch, xanthine, and hypoxanthine are negative. Temperature and pH ranges are 15–37°C (optimal range) and pH 6.5–11.0 (optimum 6.0–10.0), respectively. NaCl is not needed for growth. It can grow between 0% and 1% (w/v; optimal range) but not at 4%. It grows well on GYM *Streptomyces*, R2A (DSMZ Medium 830), Luedemann (DMSZ medium 877), and PYGV (DSMZ medium 621) media. According to API ZYM strips, the following enzymatic activities are present: esterase (C4), esterase lipase (C8), lipase (C14), leucine arylamidase, and valine arylamidase. According to the Biolog System, it oxidises: dextrin, D-maltose, sucrose, turanose, β-methyl-D-glucoside, D-glucose, D-mannose, D-fructose, D-galactose, sodium lactate, D-sorbitol, D-mannitol, D-arabitol, myo-inositol, glycerol, L-pyroglutamic acid, pectin, D-glucuronic acid, quinic acid, methyl pyruvate, D-lactic acid methyl ester, L-lactic acid, nalidixic acid, potassium tellurite, tween 40, α-hydroxy-butyric acid, β-hydroxy-butyric acid, β-keto-butyric acid, acetoacetic acid, propionic acid, acetic acid, aztreonam, butyric acid, and sodium bromated but not D-raffinose, α-D-lactose, D-melibiose, N-acetyl-D-galactosamine, N-acetyl-neuraminic acid, L-fucose, L-rhamnose, inosine, fusidic acid, D-serine (inhibitory concentration), D-aspartic acid, D-serine, troleandomycin, rifamycin SV, minocycline, gelatin, glycine-proline, L-histidine, L-serine, lincomycin, guanidine hydrochloride, niaproof, L-galactonic acid-γ-lactone, mucic acid, D-saccharic acid, vancomycin, tetrazolium violet, tetrazolium blue, p-hydroxy-phenylacetic acid, citric acid, D-malic acid, L-malic acid, bromo-succinic acid, γ-amino-*n*-butyric acid, and sodium formate. In correlation, a range of genes involved in the degradation of complex carbohydrates were identified (chemoheterotrophy). Three different autotrophic mechanisms including the Wood-Ljungdahl pathway, reductive citric acid cycle and carbonic anhydrases were besides annotated (chemolithoautotrophy). Predominant fatty acids are iso-$C_{15:0}$, iso-$C_{16:1}$ H, iso-$C_{16:0}$, $C_{17:1}$ $\omega8c$, $C_{18:1}$ $\omega9c$, and $C_{16:1}$ $\omega7c$. The cell wall peptidoglycan contains *meso*-diaminopimelic acid. Rhamnose, ribose, mannose, arabinose, galactose, and glucose are the whole-cell sugars. Polar lipid profile consists of diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylcholine, and phosphatidylinositol. The high-quality draught genome of strain DSM 44272^T was resolved to 88 scaffolds consisting of 5,094,633 bp, with a G + C content of 72.5%, 4,938 candidate protein-coding genes, 49 tRNA genes, and six rRNA regions.

The INSDC accession number for the 16S rRNA gene sequences of the type strain AT 7(−2)-11^T (=DSM 44272^T = JCM 18933^T) isolated from soil in the Atacama Desert (Chile) is MH479063. The IMG accession number for the whole genome sequences of strain DSM 44272^T is 2724679778.

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

MdCM-C designed the study. MdCM-C, AY, MR, and PS performed experiments. MdCM-C, AY, and JM-K performed bioinformatics analyses and wrote the manuscript. H-PK funded

the study. All authors contributed to the article and approved the submitted version.

Funding

MdCM-C was the recipient of a DSMZ postdoctoral fellowship 2013-2015. MdCM-C is grateful for funding received from the Ramón y Cajal Research Grant (RYC2019-028468-I) from the Spanish Ministry of Economy, Industry and Competitiveness (MINECO). The work conducted by the Joint Genome Institute, a U.S. Department of Energy Office of Science User Facility, is supported under contract no. DE-AC02-05CH11231.

Acknowledgments

We would like to gratefully acknowledge the help of Cathrin Spröer and Bettina Sträubler (Both at DSMZ, Braunschweig) for preliminary DNA:DNA hybridization analysis and Prof. Aharon Oren for taxonomic nomenclature. Special thanks go out to the editor Brian Hedlund for a detailed and very constructive review of our 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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.975365/full#supplementary-material

References

Ahrens, R., and Moll, G. (1970). Ein neues knospendes Bakterium aus der Ostsee. *Arch. Mikrobiol.* 70, 243–265. doi: 10.1007/BF00407714

Amaral, G. R. S., Dias, G. M., Wellington-Oguri, M., Chimetto, L., Campeão, M. E., Thompson, F. L., et al. (2014). Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. *Int. J. Syst. Evol. Microbiol.* 64, 357–365. doi: 10.1099/ijs.0.057927-0

Antia, M., Hoare, D. S., and Work, E. (1957). The stereoisomers of α,ε-diaminopimelic acid. III. Properties and distribution of diaminopimelic acid racemase, an enzyme causing interconversion of the LL and meso isomers. *Biochem. J.* 65, 448–459. doi: 10.1042/bj0650448

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

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

Baek, I., Kim, M., Lee, I., Na, S.-I., Goodfelow, M., and Chun, J. (2018). Phylogeny trumps chemotaxonomy: a case study involving Turicella otitidis. *Front. Microbiol.* 9:834. doi: 10.3389/fmicb.2018.00834

Balezentiene, L. (2012). Hydrolases related to C and N cycles and soil fertility amendment: responses to different management styles of agro-ecosystems. *Polish J. Environ. St.* 21, 1153–1159.

Brettin, T., Davs, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365. doi: 10.1038/srep08365

Bruce, J. (1996). Automated system rapidly identifies and characterizes microorganisms in food. *Food Technol.* 50, 77–81.

Castro, J. F., Nouioui, I., Sangal, V., Choi, S., Yang, S.-J., Kim, B.-Y., et al. (2018). Blastococcus atacamensis sp. nov., a novel strain adapted to life in the Yungay core region of the Atacama Desert. *Int. J. Syst. Evol. Microbiol.* 68, 2712–2721. doi: 10.1099/ijsem.0.002828

Chen, I. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, D507–D516. doi: 10.1093/nar/gkw929

Chen, I. M., Markowitz, V. M., Palaniappan, K., Szeto, E., Chu, K., Huang, J., et al. (2016). Supporting community annotation and user collaboration in the integrated microbial genomes (IMG) system. *BMC Genomics* 17:307. doi: 10.1186/s12864-016-2629-y

Chouaia, B., Crotti, E., Brusetti, L., Daffonchio, D., Essoussi, I., Nouioui, I., et al. (2012). Genome sequence of Blastococcus saxobsidens DD2, a stone-inhabiting bacterium. *J. Bacteriol.* 194, 2752–2753. doi: 10.1128/JB.00320-12

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

Chun, J., Oren, A., Ventosa, A., Christensen, H., Ruiz-Arahal, D., 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

Collins, M. D., Pirouz, T., Goodfellow, M., and Minnikin, D. E. (1977). Distribution of menaquinones in actinomycetes and corynebacteria. *J. Gen. Microbiol.* 100, 221–230. doi: 10.1099/00221287-100-2-221

Du, M.-Z., Wei, W., Qin, L., Liu, S., Zhang, A.-Y., Zhang, Y., et al. (2017). Co-adaptation of tRNA gene copy number and amino acid usage influences translation rates in three life domains. *DNA Res.* 24, 623–633. doi: 10.1093/dnares/dsx030

Eppard, M., Krumbein, W. E., Koch, C., Rhiel, E., Staley, J. T., and Stackebrandt, E. (1996). Morphological, physiological, and molecular characterization of actinomycetes isolated from dry soil, rocks, and monument surfaces. *Arch. Microbiol.* 166, 12–22. doi: 10.1007/s002030050350

Fotedar, R., Caldwell, M. E., Sankaranarayanan, K., Al Zeyara, A., Al Malki, A., Kaul, R., et al. (2020). Ningiella ruwaisensis gen. Nov., sp. nov., a member of the family Alteromonadaceae isolated from marine water of the Arabian gulf. *Int. J. Syst. Evol. Microbiol.* 70, 4130–4138. doi: 10.1099/ijsem.0.004256

Galloway, D. R., and Furlong, C. E. (1977). The role of ribose-binding protein in transport and chemotaxis in Escherichia coli K12. *Arch. Biochem. Biophys.* 184, 496–504. doi: 10.1016/0003-9861(77)90459-3

Giongo, A., Favet, J., Lapanje, A., Gano, K. A., Kennedy, S., Davis-Richardson, A. G., et al. (2013). Microbial hitchhikers on intercontinental dust: high-throughput sequencing to catalogue microbes in small sand samples. *Aerobiologia* 29, 71–84. doi: 10.1007/s10453-012-9264-0

Girard, G., Traag, B. A., Sangal, V., Mascini, N., Hoskisson, P. A., Goodfellow, M., et al. (2013). A novel taxonomic marker that discriminates between morphologically complex actinomycetes. *Open Biol.* 3:130073. doi: 10.1098/rsob.130073

Goodfellow, M., Kampfer, P., Busse, H.-J., Trujillo, M. E., Suzuki, K.-I., Ludwig, W., et al. (2012). *Bergey's Manual of Systematics Bacteriology*. New York, NY: Springer.

Göker, M., Cleland, D., Saunders, E., Lapidus, A., Nolan, M., Lucas, S., et al. (2011). Complete genome sequence of Isosphaera pallida type strain (IS1B^{T}). *Stand. Genomic Sci.* 4, 63–71. doi: 10.4056/sigs.1533840

Gordon, R. E., and Smith, M. M. (1955). Proposed group of characters for the separation of Streptomyces and Nocardia. *J. Bacteriol.* 69, 147–150. doi: 10.1128/jb.69.2.147-150.1955

Gregersen, T. (1978). Rapid method for distinction of gram-negative from gram-positive bacteria. *Eur. J. Appl. Microbiol. Biotechnol.* 5, 123–127. doi: 10.1007/BF00498806

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res.* 35, W52–W57. doi: 10.1093/nar/gkm360

Gupta, R., Gupta, N., and Rathi, P. (2004). Bacterial lipases: an overview of production, purification and biochemical properties. *Appl. Microbiol. Biotechnol.* 64, 763–781. doi: 10.1007/s00253-004-1568-8

Han, X., Chen, C.-C., Kuo, C.-J., Huang, C.-H., Zheng, Y., Ko, T.-P., et al. (2015). Crystal structures of ligand-bound octaprenyl pyrophosphate synthase from Escherichia coli reveal the catalytic and chain-length determining mechanisms. *Proteins* 83, 37–45. doi: 10.1002/prot.24618

Hahnke, R. L., Meier-Kolthoff, J. P., García-López, M., Mukherjee, S., Huntemann, M., Ivanova, N. N., et al. (2016). Genome-based taxonomic classification of Bacteroidetes. *Front. Microbiol.* 7:2003. doi: 10.3389/fmicb.2016.02003

Hennig, W. (1965). Phylogenetic systematics. *Annu. Rev. Ecol. Syst.* 10, 97–116. doi: 10.1146/annurev.en.10.010165.000525

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. D. C., Sghaier, H., Rohde, M., Schumann, P., et al. (2015a). *Geodermatophilus* sabuli sp. nov., a γ-radiation-resistant actinobacterium isolated from desert limestone in Sahara Desert. *Int. J. Syst. Evol. Microbiol.* 65, 3365–3372. doi: 10.1099/ijsem.0.000422

Hezbri, K., Ghodhbane-Gtari, F., Montero-Calasanz, M. D. C., Sghaier, H., Rohde, M., Spröer, C., et al. (2015b). Description of *Geodermatophilus* bullaregiensis sp. nov. *Antonie Van Leeuwenhoek* 108, 415–425. doi: 10.1007/s10482-015-0494-3

Hezbri, K., Louati, M., Nouioui, I., Gtari, M., Rohde, M., Spröer, C., et al. (2016). Blastococcus capsensis sp. nov., isolated from an archaeological Roman pool and emended description of the genus *Blastococcus*, *B. aggregatus*, *B. saxobsidens*, *B. jejuensis* and *B. endophyticus*. *Int. J. Syst. Evol. Microbiol* 66, 4864–4872. doi: 10.1099/ijsem.0.001443

Hezbri, K., Nouioui, I., Rohde, M., Schumann, P., Gtari, M., Klenk, H.-P., et al. (2017). Blastococcus colisei sp. nov, isolated from an archaeological amphitheatre. *Antonie Van Leeuwenhoek* 110, 339–346. doi: 10.1007/s10482-016-0804-4

Hezbri, K., Nouioui, I., Rohde, M., Sproer, C., Schumann, P., Gtari, M., et al. (2018). Blastococcus xanthinilyticus sp. nov., isolated from monument. *Int. J. Syst. Evol. Microbiol.* 68, 1177–1183. doi: 10.1099/ijsem.0.002646

Holt, P. S., and Chaubal, L. H. (1997). Detection of motility and putative synthesis of flagellar proteins in salmonella pullorum cultures. *J. Clin. Microbiol.* 35, 1016–1020.

Hor, L., Dobson, R. C. J., Downton, M. T., Wagner, J., Hutton, C. A., and Perugini, M. A. (2013). Dimerization of bacterial diamonopimelate epimerase is essential for catalysis. *J. Biol. Chem.* 288, 9238–9248. doi: 10.1074/jbc.M113.450148

Huntemann, M., Ivanova, N. N., Mavromatis, K., Tripp, H. J., Paez-Espino, D., Palaniappan, K., et al. (2015). The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). *Stand. Genomic Sci.* 10:86. doi: 10.1186/s40793-015-0077-y

Ishiguro, E. E., and Wolfe, R. S. (1970). Control of morphogenesis in *Geodermatophilus*: ultrastructural studies. *J. Bacteriol.* 104, 566–580. doi: 10.1128/jb.104.1.566-580.1970

Jakimowics, D., and van Wezel, G. P. (2012). Cell division and DNA segregation in Streptomyces: how to build a septum in the middle of nowhere? *Mol. Microbiol.* 85, 393–404. doi: 10.1111/j.1365-2958.2012.08107.x

Jiang, Z.-M., Zhang, B.-H., Sun, H.-M., Zhang, T., Yu, L.-Y., and Zhang, Y.-Q. (2021). Properties of Modestobacter deserti sp. Nov., a kind of novel phosphate-solubilising actinobacteria inhabited in the desert biological soil crust. *Front. Microbiol.* 12:742798. doi: 10.3389/fmicb.2021.742798

Kachlany, S. C., Planet, P. J., Bhattacherjee, M. K., Kollia, E., DeSalle, R., Fine, D. H., et al. (2000). Nonspecific adherence by Actinobacillus actinomycetemcomitans requires genes widespread in bacteria and archaea. *J. Bacteriol.* 182, 6169–6176. doi: 10.1128/JB.182.21.6169-6176.2000

Katsyv, A., and Müller, V. (2020). Overcoming energetic barriers in acetogenic C1 conversion. *Front. Bioeng. Biotechnol.* 8:621166. doi: 10.3389/fbioe.2020.621166

Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. S. (2000). rRNA operon copy number reflects ecological strategies of bacteria. *App. Environ. Microbiol.* 66, 1328–1333. doi: 10.1128/AEM.66.4.1328-1333.2000

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

Ko, J.-H., Montero Llopis, P., Heinritz, J., Jacobs-Wagner, C., and Sll, D. (2013). Suppression of amber codons in Caulobacter crescentus by the orthogonal Escherichia coli Histidyl-tRNA synthetase/tRNA^{His} pair. *PLoS One* 8:e83630. doi: 10.1371/journal.pone.0083630

Kroppenstedt, R., and Goodfellow, M. (2006). "The family thermomonosporaceae: actinocorallia, actinomadura, spirillospora and thermomonospora" in *The Prokaryotes*. eds. M. Dworkin and S. Falkow (New York, NY: Springer)

Kroppenstedt, R. M. (1982). Separation of bacterial menaquinones by HPLC using reverse phase (RP18) and a silver loaded ion exchanger as stationary phases. *J. Liquid Chromatography* 5, 2359–2367. doi: 10.1080/01483918208067640

Kuo, C.-H., and Ochman, H. (2010). The extinction dynamics of bacterial pseudogenes. *PLoS Genet.* 6:e1001050. doi: 10.1371/journal.pgen.1001050

Kyrpides, N. C., Hugenholtz, P., Eisen, J. A., Woyke, T., Göker, M., Parker, C. T., et al. (2014). Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains. *PLoS Biol.* 12:e1001920. doi: 10.1371/journal.pbio.1001920

Lawson, P. A., Patel, N. B., Mohammed, A., Moore, E. R. B., Lo, A. S., Sardi, A., et al. (2020). Parapseudoflavitalea muciniphila gen. Nov., sp. nov., a member of the family Chitinophagaceae isolated from a human peritoneal tumour and reclassification of Pseudobacter ginsenosidimutans as Pseudoflavitalea ginsenosidimutans comb. nov. *Int. J. Syst. Evol. Microbiol.* 70, 3639–3646. doi: 10.1099/ijsem.0.004204

Lechevalier, M. P. L. H., and Lechevalier, H. (1970). Chemical composition as a criterion in the classification of aerobic actinomycetes. *Int. J. Syst. Evol. Microbiol.* 20, 435–443. doi: 10.1099/00207713-20-4-435

Lee, S. D. (2006). Blastococcus jejuensis sp. nov., an actinomycete from beach sediment, and emended description of the genus Blastococcus Ahrens and Moll 1970. *Int. J. Syst. Evol. Microbiol.* 56, 2391–2396. doi: 10.1099/ijs.0.64268-0

Lee, D. W., Lee, H., Kwon, B. O., Khim, J. S., Yim, U. H., Kim, B. S., et al. (2018). Blastococcus litoris sp. nov., isolated from sea-tidal flat sediment. *Int. J. Syst. Evol. Microbiol.* 68, 3435–3440. doi: 10.1099/ijsem.0.003004

Liu, Y., Harrison, P. M., Kunin, V., and Gerstein, M. (2004). Comprehensive analysis of pseudogenes in prokaryotes: widespread gene decay and failure of putative horizontally transferred genes. *Genome Biol.* 5:1715

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

Liu, R., and Ochman, H. (2007). Stepwise formation of the bacterial flagellar system. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7116–7121. doi: 10.1073/pnas.0700266104

López-Lara, I. M., and Geiger, O. (2019). "Chapter 3: formation of fatty acids," in *Biogenesis of Fatty Acids, Lipids and Membranes*. ed. O. Geiger (Springer).

Lowe, J., and Amos, L. A. (1998). Crystal structure of the bacterial cell-division protein FtsZ. *Nature* 391, 203–206. doi: 10.1038/34472

Luedemann, G. M. (1968). *Geodermatophilus*, a new genus the Dermatophilaceae (Actinomycetales). *J. Bacteriol.* 96, 1848–1858. doi: 10.1128/jb.96.5.1848-1858.1968

McCool, G. J., and Cannon, M. C. (2001). PhaC and PhaR are required for polyhydroxyalkanoic acid synthase activity in bacillus megaterium. *J. Bacteriol.* 183, 4235–4243. doi: 10.1128/JB.183.14.4235-4243.2001

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013a). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinf.* 14:60. doi: 10.1186/1471-2105-14-60

Meier-Kolthoff, J. P., Göker, M., Spröer, C., and Klenk, H. P. (2013b). When should a DDH experiment be mandatory in microbial taxonomy? *Arch. Microbiol.* 195, 413–418. doi: 10.1007/s00203-013-0888-4

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2014a). Highly parallelized inference of large genome-based phylogenies. *Concurr. Comput. Pract. Exp.* 26, 1715–1729. doi: 10.1002/cpe.3112

Meier-Kolthoff, J. P., Klenk, H.-P., and Göker, M. (2014b). Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int. J. Syst. Evol. Microbiol.* 64, 352–356. doi: 10.1099/ijs.0.056994-0

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/s41467-019-10210-3

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

Mengin-Lecreulx, D., Michau, C., Richaud, C., Blanot, D., and Heijenoort, J. V. (1988). Incorporation of LL-Diaminopimelic acid into peptidoglycan of Escherichia coli mutants lacking diaminopimelate epimerase encoded by dapF. *J. Bacteriol.* 170, 2031–2039. doi: 10.1128/jb.170.5.2031-2039.1988

Mevs, U., Stackebrandt, E., Schumann, P., Gallikowski, C. A., and Hirsch, P. (2000). Modestobacter multiseptatus gen. Nov., sp. nov., a budding actinomycete from soils of the Asgard range (Transantarctic Mountains). *Int. J. Syst. Evol. Microbiol.* 50, 337–346. doi: 10.1099/00207713-50-1-337

Minnikin, D. E., O'Donnell, A. G., Goodfellow, M., Alderson, G., Athalye, M., Schaal, A., et al. (1984). An integrated procedure for the extraction of bacterial isoprenoid quinones and polar lipids. *J. Microbiol. Methods* 2, 233–241. doi: 10.1016/0167-7012(84)90018-6

Montero-Calasanz, M. D. C. (2021). *Klenkia in Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd: Chichester, UK.

Montero-Calasanz, M. D. C. (2020a). *Geodermatophilus in Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd: Chichester, UK.

Montero-Calasanz, M. D. C. (2020b). *Geodermatophilaceae in Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd: Chichester, UK.

Montero-Calasanz, M. D. C., Hofner, B., Göker, M., Rohde, M., Spröer, C., Hezbri, K., et al. (2014). *Geodermatophilus* poikilotrophi sp. nov., a multi-tolerant actinomycete isolated from dolomitic marble. *Biomed. Res. Int.* 2014:914767. doi: 10.1155/2014/914767

Montero-Calasanz, M. D. C., Hezbri, K., Göker, M., Sghaier, H., Rohde, M., Spröer, C., et al. (2015). Description of gamma radiation-resistant *Geodermatophilus* dictyosporus sp. nov. to accommodate the not validly named *Geodermatophilus* obscurus subsp. dictyosporus (Luedemann, 1968). *Extremophiles* 19, 77–85. doi: 10.1007/s00792-014-0708-z

Montero-Calasanz, M. D. C., Meier-Kolthoff, J. P., Zhang, D.-F., Yaramis, A., Rohde, M., Woyke, T., et al. (2017). Genome-scale data call for a taxonomic rearrangement of *Geodermatophilaceae*. *Front. Microbiol.* 8:2501. doi: 10.3389/fmicb.2017.02501

Montero-Calasanz, M. D. C., Yaramis, A., Nouioui, I., Igual, J. M., Spröer, C., Castro, J. F., et al. (2019). Modestobacter italicus sp. nov., isolated from Carrara marble quarry and emended descriptions of the genus Modestobacter and the species Modestobacter marinus, Modestobacter multiseptatus, Modestobacter roseus and Modestobacter versicolor. *Int. J. Syst. Evol. Microbiol.* 69, 1537–1545. doi: 10.1099/ijsem.0.003282

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2017). Genomes online database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi: 10.1093/nar/gkw992

Neilson, J. W., Quade, J., Ortiz, M., Nelson, W. M., Legatzki, A., Tian, F., et al. (2012). Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* 16, 553–566. doi: 10.1007/s00792-012-0454-z

Normand, P., Orso, S., Cournoyer, B., Jeannin, P., Chapelon, C., Dawson, J., et al. (1996). Molecular phylogeny of the genus Frankia and related genera and emendation of the family Frankiaceae. *Int. J. Syst. Bacteriol.* 46, 1–9. doi: 10.1099/00207713-46-1-1

Normand, P. (2006). *Geodermatophilaceae* fam. Nov., a formal description. International. *Int. J. Syst. Evol. Microbiol.* 56, 2277–2278. doi: 10.1099/ijs.0.64298-0

Nouioui, I., Göker, M., Carro, L., Montero-Calasanz, M. D. C., Rohde, M., Woyke, T., et al. (2017). High quality draft genome of Nakamurella lactea type strain, a rock actinobacterium, and emended description of Nakamurella lactea. *Stand. Genomic Sci.* 12:4. doi: 10.1186/s40793-016-0216-0

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

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

Paniagua-Michel, J., Olmos-Soto, J., and Ruiz, M. A. (2012). "Pathways of carotenoid biosynthesis in bacteria and microalgae" in *Microbial carotenoids from bacteria and microalgae. Methods in molecular biology (methods and protocols)*. ed. J. L. Barredo, vol. *892* (Totowa, NJ: Humana Press)

Perez, B. A., Planet, P. J., Kachlany, S. C., Tomich, M., Fine, D. H., and Figurski, D. H. (2006). Genetic analysis of the requirement for flp-2, tadV and rcpB in Actinobacillus actinomycetemcomitans biofilm formation. *J. Bacteriol.* 188, 6361–6375. doi: 10.1128/JB.00496-06

Proft, T., and Baker, E. N. (2009). Pili in gram-negative and gram-positive bacteria – structure, assembly and their role in disease. *Cell. Mol. Life Sci.* 66, 613–635. doi: 10.1007/s00018-008-8477-4

Qin, S., Bian, G. K., Zhang, Y. J., Xing, K., Cao, C. L., Liu, C. H., et al. (2013). Modestobacter roseus sp. nov., an endophytic actinomycete isolated from the coastal halophyte Salicornia europaea Linn., and emended description of the genus Modestobacter. *Int. J. Syst. Evol. Microbiol.* 63, 2197–2202. doi: 10.1099/ijs.0.044412-0

Rainey, F. A., Ward-Rainey, N., Kroppenstedt, R. M., and Stackebrandt, E. (1996). The genus Nocardiopsis represents a phylogenetically coherent taxon and a distinct actinomycete lineage: proposal of Nocardiopsaceae fam. Nov. *Int. J. Syst. Bacteriol.* 46, 1088–1092. doi: 10.1099/00207713-46-4-1088

Rashid, R., Cazenave-Gassiot, A., Gao, I. H., Nair, Z. J., Kumar, J. K., Gao, L., et al. (2017). Comprehensive analysis of phospholipids and glycolipids in the opportunistic pathogen enterococcus faecalis. *PLoS One* 12:e0175886. doi: 10.1371/journal.pone.0175886

Reddy, G. S., Potrafka, R. M., and Garcia-Pichel, F. (2007). Modestobacter versicolor sp. nov., an actinobacterium from biological soil crusts that produces melanins under oligotrophy, with emended descriptions of the genus Modestobacter and Modestobacter multiseptatus Mevs et al. 2000. *Int. J. Syst. Evol. Microbiol.* 57, 2014–2020. doi: 10.1099/ijs.0.64932-0

Rivas-Marin, E., Peeters, S. H., Claret Fernández, L., Jogler, C., van Niftrik, L., Wiegan, S., et al. (2020). Non-essentiality of canonical cell division genes in the planctomycete Planctopirus limnophila. *Sci. Rep.* 10:66. doi: 10.1038/s41598-019-56978-8

Rogers, T. E., Ataide, S. F., Dare, K., Katz, A., Seveau, S., Roy, H., et al. (2012). A pseudo-tRNA modulates antibiotic resistance in Bacillus cereus. *PLoS One* 7:e41248. doi: 10.1371/journal.pone.0041248

Sasser, M. (2001). Identification of bacteria by gas chromatography of cellular fatty acids. *Technical. Note* 101, 1–6.

Schleifer, K. H., and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36, 407–477. doi: 10.1128/br.36.4.407-477.1972

Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A. M., and Normand, P. (2014). Phylogeny of the class Actinobacteria revisited in the light of complete genomes. The orders 'Frankiales' and Micrococcales should be split into coherent entities: proposal of Frankiales Ord. Nov., Geodermatophilales Ord. Nov., Acidothermales Ord. Nov. an. *Int. J. Syst. Evol. Microbiol.* 64, 3821–3832. doi: 10.1099/ijs.0.063966-0

Seto, H., Jinnai, Y., Hiratsuka, T., Fukawa, M., Furihata, K., Itoh, N., et al. (2008). Studies on a new biosynthetic pathway for menaquinone. *J. Am. Chem. Soc.* 160, 5614–5615. doi: 10.1021/ja710207s

Sghaier, H., Hezbri, K., Ghodhbane-Gtari, F., Pujic, P., Sen, A., Daffonchio, D., et al. (2016). Stone-dwelling actinobacteria Blastococcus saxobsidens, Modestobacter marinus and *Geodermatophilus* obscurus proteogenomes. *ISME J.* 10, 21–29. doi: 10.1038/ismej.2015.108

Stackebrandt, E., Rainey, F. A., and Ward-Rainey, N. L. (1997). Proposal for a new hierarchic classification system, Actinobacteria classis nov. *Int. J. Syst. Bacteriol.* 47, 479–491. doi: 10.1099/00207713-47-2-479

Stackebrandt, E., Breymann, S., Steiner, U., Prauser, H., Weiss, N., Schumann, P., et al. (2002). Re-evaluation of the status of the genus Oerskovia, reclassification of Promicromonospora enterophila (Jáger et al. 1983) as *Oerskovia enterophila* comb. nov. and description of *Oerskovia jenensis* sp. nov. and *Oerskovia paurometabola* sp. nov. *Int. J. Syst. Evol. Microbiol.* 52, 1105–1111. doi: 10.1099/00207713-52-4-1105

Staneck, J. L., and Roberts, G. D. (1974). Simplified approach to identification of aerobic actinomycetes by thin-layer chromatography. *Appl. Microbiol.* 28, 226–231. doi: 10.1128/am.28.2.226-231.1974

Sun, D.-L., Jiang, X., Wu, Q. L., and Zhou, N.-Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* 79, 5962–5969. doi: 10.1128/AEM.01282-13

Supuran, C. T., and Capasso, C. (2017). An overview of the bacterial carbonic anhydrases. *Meta* 7:56. doi: 10.3390/metabo7040056

Sutcliffe, I. C., Trujillo, M. E., Whitman, W. B., and Goodfellow, M. (2013). A call to action for the international committee on systematics of prokaryotes. *Trends Microbiol.* 21, 51–52. doi: 10.1016/j.tim.2012.11.004

Szurmant, H., and Ordal, G. W. (2004). Diversity in chemotaxis mechanisms among the bacteria and archaea. *Microbiol. Mol. Biol. Rev.* 68, 301–319. doi: 10.1128/MMBR.68.2.301-319.2004

Tindall, B. J. (1990). A comparative study of the lipid composition of Halobacterium saccharovorum from various sources. *Syst. Appl. Microbiol.* 13, 128–130. doi: 10.1016/S0723-2020(11)80158-X

Tu, C. M., and Miles, J. R. W. (1976). Interactions between insecticides and soil microbes. *Res. Rev.* 64, 17–65.

Tomich, M., Planet, P. J., and Figurski, D. H. (2007). The tad locus: postcards from the widespread colonization island. *Nat. Rev.* 5, 363–375. doi: 10.1038/nrmicro1636

Tóth, E. M., Schumann, P., Borsodi, A. K., Kéki, Z., Kovács, A. L., and Márialigeti, K. (2008). Wohlfahrtiimonas chitiniclastica gen. Nov., sp. nov., a new

gammaproteobacterium isolated from *Wohlfahrtia magnifica* (Diptera: Sarcophagidae). *Int. J. Syst. Evol. Microbiol.* 58, 976–981. doi: 10.1099/ijs.0.65324-0

Traag, B. A., and van Wezel, G. P. (2008). The SsgA-like proteins in actinomycetes: small proteins up to a big task. *Antonie Van Leeuwenhoek* 94, 85–97. doi: 10.1007/s10482-008-9225-3

Trujillo, M. E., and Normand, P. (2019). *Sporitchthya in Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley and Sons, Ltd: Chichester.

Upadhyay, A., Fontes, F. L., Gonzalez-Jarero, M., NcNeil, M. R., Crans, D. C., Kacson, M., et al. (2015). Partial saturation of menaquinone in mycobacterium tuberculosis: function and essentiality of a novel reductase. *MenJ. ACS Cent. Sci.* 1, 292–302. doi: 10.1021/acscentsci.5b00212

Urzì, C., Brusetti, L., Salamone, P., Sorlini, C., Stackebrandt, E., and Daffonchio, D. (2001). Biodiversity of geodermatophilaceae isolated from altered stones and monuments in the mediterranean basin. *Environ. Microbiol.* 3, 471–479. doi: 10.1046/j.1462-2920.2001.00217.x.o

Urzì, C., Salamone, P., Schumann, P., Rohde, M., and Stackebrandt, E. (2004). Blastococcus saxobsidens sp. nov., and emended descriptions of the genus Blastococcus Ahrens and Moll 1970 and Blastococcus aggregatus Ahrens and Moll 1970. *Int. J. Syst. Evol. Microbiol.* 54, 253–259. doi: 10.1099/ijs.0.02745-0

Vaas, L. A. I., Sikorski, J., Hofner, B., Fiebig, A., Buddruhs, N., Klenk, H. P., et al. (2013). Opm: an R package for analysing OmniLog® phenotype microarray data. *Bioinformatics* 29, 1823–1824. doi: 10.1093/bioinformatics/btt291

Vaas, L. A. I., Sikorski, J., Michael, V., Göker, M., and Klenk, H. P. (2012). Visualization and curve-parameter estimation strategies for efficient exploration of phenotype microarray kinetics. *PLoS One* 7:e34846. doi: 10.1371/journal.pone.0034846

Van Wezel, G. P., van der Meulen, J., Kwamoto, S., Luiten, R. G., Koerten, H. K., and Kraal, B. (2000). SsgA is essential for sporulation of Streptomyces coelicolor A3(2) and affects hyphal development by stimulating septum formation. *J. Bacteriol.* 182, 5653–5662. doi: 10.1128/JB.182.20.5653-5662.2000

Vandamme, P., and Sutcliffe, I. (2021). Out with the old and in with the new: time to rethink twentieth century chemotaxonomic practices in bacterial taxonomy. *Int. J. Syst. Evol. Microbiol.* 71:005127. doi: 10.1099/ijsem.0.005127

Vetrovsky, 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

Wang, X., and Lutkenhaus, J. (1993). The FtsZ protein of Bacillus subtilis is localized at the division site and has GTPase activity that is dependent upon FtsZ concentration. *Mol. Microbiol.* 9, 435–442. doi: 10.1111/j.1365-2958.1993.tb01705.x

Wang, Y., Zhang, L., Zhang, X., Huang, J., Zhao, Y., Zhao, Y., et al. (2017). *Geodermatophilus* daqingensis sp. Nov., isolated from petroleum-contaminated soil. *Antonie Van Leeuwenhoek* 110, 803–809. doi: 10.1007/s10482-017-0853-3

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. Bacteriol.* 37, 463–464. doi: 10.1099/00207713-37-4-463

Whitman, W. B. (2015). Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst. Appl. Microbiol.* 38, 217–222. doi: 10.1016/j.syapm.2015.02.003

Wiley, E. O., and Lieberman, B. B. S. (2011). *Phylogenetics: Theory and Practice of Phylogenetic Systematics*, *2nd*. Hoboken, NJ: John Wiley and Sons.

Xiao, J., Luo, Y., Xu, J., Xie, S., and Xu, J. (2011). Modestobacter marinus sp. nov., a psychrotolerant actinobacterium from deep-sea sediment, and emended description of the genus Modestobacter. *Int. J. Syst. Evol. Microbiol.* 61, 1710–1714. doi: 10.1099/ijs.0.023085-0

Yang, Z. W., Asem, M. D., Li, X., Li, L. Y., Salam, N., Alkhalifah, D. H. M., et al. (2019). Blastococcus deserti sp. nov., isolated from a desert sample. *Arch. Microbiol.* 201, 193–198. doi: 10.1007/s00203-018-1604-1

Zhu, W. Y., Zhang, J. L., Qin, Y. L., Xiong, Z. J., Zhang, D. F., Klenk, H. P., et al. (2013). Blastococcus endophyticus sp. nov., an actinobacterium isolated from Camptotheca acuminata. *Int. J. Syst. Evol. Microbiol.* 63, 3269–3273. doi: 10.1099/ijs.0.049239-0

Zhi, X. Y., Li, W. J., and Stackebrandt, E. (2009). An update of the structure and 16S rRNA gene sequence-based definition of higher ranks of the class Actinobacteria, with the proposal of two new suborders and four new families and emended descriptions of the existing higher taxa. *Int. J. Syst. Evol. Microbiol.* 59, 589–608. doi: 10.1099/ijs.0.65780-0

frontiers | Frontiers in Microbiology

TYPE Correction
PUBLISHED 20 January 2023
DOI 10.3389/fmicb.2022.1100319

OPEN ACCESS

EDITED AND REVIEWED BY
André Antunes,
Macau University of Science and Technology, China

*CORRESPONDENCE
Maria del Carmen Montero-Calasanz
✉ mariac.montero.calasanz@juntadeandalucia.es

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

RECEIVED 17 November 2022
ACCEPTED 16 December 2022
PUBLISHED 20 January 2023

CITATION
Montero-Calasanz MdC, Yaramis A, Rohde M, Schumann P, Klenk H-P and Meier-Kolthoff JP (2023) Corrigendum: Genotype–phenotype correlations within the *Geodermatophilaceae*. *Front. Microbiol.* 13:1100319. doi: 10.3389/fmicb.2022.1100319

Corrigendum: Genotype–phenotype correlations within the *Geodermatophilaceae*

Maria del Carmen Montero-Calasanz[1,2]*, Adnan Yaramis[2], Manfred Rohde[3], Peter Schumann[4], Hans-Peter Klenk[2] and Jan P. Meier-Kolthoff[5]

[1]IFAPA Las Torres-Andalusian Institute of Agricultural and Fisheries Research and Training, Junta de Andalucía, Seville, Spain, [2]School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [3]Central Facility for Microscopy, HZI – Helmholtz Centre for Infection Research, Braunschweig, Germany, [4]Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany, [5]Department Bioinformatics and Databases, Leibniz Institute DSMZ – German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

KEYWORDS

***Trujillonella*, *Pleomorpha*, *Goekera*, phylogenetic systematics, *in silico* chemotaxonomy**

A corrigendum on
Genotype–phenotype correlations within the *Geodermatophilaceae*

by Montero-Calasanz, M. d. C., Yaramis, A., Rohde, M., Schumann, P., Klenk, H.-P., and Meier-Kolthoff, J. P. (2022). *Front. Microbiol.* 13:975365. doi: 10.3389/fmicb.2022.975365

In the published article, there were some nomenclatural errors in the protologues of *Trujillonella* and *Trujillonella endophytica*, *Pleomorpha* and *Pleomorpha daqingensis*, *Goekera* and *Goekera deserti* and *Blastococcus xantinilythicus*.

A correction has been made to the section **Final remarks and taxonomic consequences**, as follows:

Description of *Trujillonella* gen. nov.

Tru.jil.lo.nel'la. N.L. fem. dim. n. *Trujillonella* named in honour of Martha E. Trujillo in recognition of her contributions to microbial systematics, mainly on Actinobacteria, on Bergey's Manual trust, and as the Editor-in-chief of the International Journal of Systematic and Evolutionary Microbiology.

Cells are aerobic, non-motile, non-spore-forming, Gram-stain positive, catalase-positive and oxidase-negative. Cells occur singly, in pairs or in tetrads, often tending to form aggregates. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4), with MK-8 and MK-9(H_6) as minor components. The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylcholine, phosphatidylethanolamine, and phosphatidylinositol. The major fatty acids are iso-$C_{16:0}$, iso-$C_{15:0}$ and $C_{18:1}\omega9c$. The basic whole-cell sugar pattern

includes arabinose and galactose. The genomic G + C content is 71–72%. The type species of *Trujillonella* is *Trujillonella endophytica* comb. nov.

Description of *Trujillonella endophytica* comb. nov.

T. en.do.phy'ti.ca (Gr. pref. *endo-*, within; Gr. neut. n. *phyton,* plant; L. fem. adj. suff. *–ica,* adjectival suffix used with the sense of belonging to; N.L. fem. adj. *endophytica,* within plant, endophytic, pertaining to the isolation from plant tissues).

Basonym: *Blastococcus endophyticus* Zhu et al. 2013 emend. Hezbri et al. 2016.

The properties are as given in the species description by Zhu et al. (2013) and emendation by Hezbri et al. (2016) with the following modification. The genomic G + C content is 74.6%. The genome size is 4.9 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain DSM 45413^T is FOEE00000000.

The type strain YIM 68236^T = CCTCC AA 209045^T = DSM 45413^T = KCTC 19998^T was isolated from healthy leaves of *Camptotheca acuminata* collected in Yunnan Province, south-west China.

Description of *Pleomorpha* gen. nov.

Ple.o.mor'pha. Gr. adv. *pleon* more; Gr. fem. n. *morphe*, shape or form; N.L. fem. n. *Pleomorpha*, organism showing multiple forms.

Pleiomorphic, motile, spore-forming, aerobic, Gram-stain positive cells. Those occurs singly or associated in aggregates. The peptidoglycan in the cellwall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4). The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylcholine, phosphatidylglycerol, phosphatidylethanolamine, and phosphatidylinositol. The major fatty acids are iso-$C_{16:0}$ and iso-$C_{15:0}$. The basic whole-cell sugar pattern includes galactose, glucose and xylose. The genomic G + C content is 73–74%. The type species of *Pleomorpha* is *Pleomorpha daqingensis* comb. nov.

Description of *Pleomorpha daqingensis* comb. nov.

P. da.qing.en'sis (N.L. fem. adj. *daqingensis*, pertaining to Daqing city, China, where the type strain was isolated).

Basonym: *Geodermatophilus daqingensis* Wang et al. 2017.

The properties are as given in the species description by Wang et al. (2017) with the following modification. The genomic G + C content is 73.6%. The genome size is 5.4 Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain DSM 104001^T is JACBZT000000000.

The type strain WT-2-1^T = CGMCC 4.7381^T = DSM 104001^T was isolated from petroleum- contaminated soil in Daqing city, China.

Description of *Goekera* gen. nov.

Goe'ke.ra. N.L. fem. n. *Goekera*, named in honour of Markus Göker in recognition of his contributions to microbial systematics, including work on *Actinobacteria*, on the List of Prokaryotic Names with Standing in Nomenclature (LPSN), and as a member of the Judicial Commission.

Cells are motile, non-spore-forming, aerobic, Gram-stain positive, catalase and oxidase positive cocci and/short rods. Bud-like structure was observed for some cells. The peptidoglycan in the cell-wall contains *meso*-diaminopimelic acid. The predominant menaquinone is MK-9(H_4), with MK-8(H_4) as a minor component. The basic polar lipid profile contains diphosphatidylglycerol, phosphatidylethanolamine, phosphatidylglycerol, phosphatidylinositol, phosphatidylmethylethanolamine and phosphatidylinositol mannoside. The major fatty acids are $C_{18:1}\omega9c$, iso-$C_{16:0}$, $C_{16:0}$, iso-$C_{15:0}$, and $C_{16:1}\omega7c$.The basic whole-cell sugar pattern includes arabinose, glucose and ribose. The genomic G + C content is 74–75%. The type species of *Goekera* is *Goekera deserti* comb. nov.

Description of *Goekera deserti* comb. nov.

G. de.ser'ti (L. gen. neut. n. *deserti*, of a desert, where the organisms were acquired).

Basonym: *Modestobacter deserti* Jiang et al. 2023.

The properties are as given in the species description by Jiang et al. (2021) with the following modification. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of strain CPCC 205119^T is JAAGWK000000000.

The type strain CPCC 205119^T = I12A-02624^T = KCTC 49201^T = NBRC 113528^T was isolated from moss-dominated soil crusts collected from Shapotou NDER in Tengger Desert, China.

Emended description of *Blastococcus xanthinilyticus* Hezbri et al. (2018)

The properties are as given in the species description by Hezbri et al. (2018) with the following emendation. The genomic G + C content is 74.4%. The genome size is 4.6

Mbp. According to genomic data, anaerobiosis and acetogenesis may occur. A repertoire of genes related to flagellum synthesis, chemotaxis, spore production and pilus assembly were annotated. Four different autotrophic mechanisms including the Wood-Ljungdahl pathway, C4-dicarboxylic acid and reductive citric acid cycles and carbonic anhydrases as well as a range of genes involved in the degradation of complex carbohydrates were also identified.

The accession number for the whole genome sequence of the type strain DSM 46842^T is VNHW00000000.

The type strain BMG 862^T = DSM 46842^T = CECT 8884^T was isolated from a marble sample collected from the Bulla Regia monument, Northern Tunisia.

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way.

Publisher's note

References

Hezbri, K., Louati, M., Nouioui, I., Gtari, M., Rohde, M., Spröer, C., et al. (2016). *Blastococcus capsensis* sp. nov., isolated from an archaeological Roman pool and emended description of the genus *Blastococcus*, *B. aggregatus*, *B. saxobsidens*, *B. jejuensis* and *B. endophyticus*. *Int. J. Syst. Evol. Microbiol* 66, 4864–4872. doi: 10.1099/ijsem.0.001443

Hezbri, K., Nouioui, I., Rohde, M., Sproer, C., Schumann, P., Gtari, M., et al. (2018). *Blastococcus xanthinilyticus* sp. nov., isolated from monument. *Int. J. Syst. Evol. Microbiol.* 68, 1177–1183. doi: 10.1099/ijsem.0.002646

Jiang, Z.-M., Zhang, B.-H., Sun, H.-M., Zhang, T., Yu, L.-Y., and Zhang, Y.-Q. (2021). Properties of *Modestobacter deserti* sp. Nov., a kind of novel phosphate-solubilising actinobacteria inhabited in the desert biological soil crust. *Front. Microbiol.* 12, 742798. doi: 10.3389/fmicb.2021.742798

Wang, Y., Zhang, L., Zhang, X., Huang, J., Zhao, Y., Zhao, Y., et al. (2017). *Geodermatophilus daqingensis* sp. Nov., isolated from petroleum-contaminated soil. *Antonie Van Leeuwenhoek* 110, 803–809. doi: 10.1007/s10482-017-0853-3

Zhu, W. Y., Zhang, J. L., Qin, Y. L., Xiong, Z. J., Zhang, D. F., Klenk, H. P., et al. (2013). *Blastococcus endophyticus* sp. nov., an actinobacterium isolated from *Camptotheca acuminata*. *Int. J. Syst. Evol. Microbiol.* 63, 3269–3273. doi: 10.1099/ijs.0.049239-0

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 08 December 2022
DOI 10.3389/fmicb.2022.999876

OPEN ACCESS

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

REVIEWED BY
Noha M. Mesbah,
Suez Canal University,
Egypt
Jose Felix Aguirre Garrido,
Autonomous Metropolitan University,
Mexico

*CORRESPONDENCE
Oliyad Jeilu
Oliyad.jeilu.oumer@slu.se

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

RECEIVED 21 July 2022
ACCEPTED 11 November 2022
PUBLISHED 08 December 2022

CITATION
Jeilu O, Gessesse A, Simachew A, Johansson E and Alexandersson E (2022) Prokaryotic and eukaryotic microbial diversity from three soda lakes in the East African Rift Valley determined by amplicon sequencing.
Front. Microbiol. 13:999876.
doi: 10.3389/fmicb.2022.999876

Prokaryotic and eukaryotic microbial diversity from three soda lakes in the East African Rift Valley determined by amplicon sequencing

Oliyad Jeilu[1,2]*, Amare Gessesse[1,3], Addis Simachew[1], Eva Johansson[2] and Erik Alexandersson[4]

[1]Institute of Biotechnology, Addis Ababa University, Addis Ababa, Ethiopia, [2]Department of Plant Breeding, Swedish University of Agricultural Sciences, Lomma, Sweden, [3]Department of Biological Sciences and Biotechnology, Botswana International University of Science and Technology, Palapye, Botswana, [4]Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Lomma, Sweden

Soda lakes are unique poly-extreme environments with high alkalinity and salinity that support diverse microbial communities despite their extreme nature. In this study, prokaryotic and eukaryotic microbial diversity in samples of the three soda lakes, Lake Abijata, Lake Chitu and Lake Shala in the East African Rift Valley, were determined using amplicon sequencing. Culture-independent analysis showed higher diversity of prokaryotic and eukaryotic microbial communities in all three soda lakes than previously reported. A total of 3,603 prokaryotic and 898 eukaryotic operational taxonomic units (OTUs) were found through culture-independent amplicon sequencing, whereas only 134 bacterial OTUs, which correspond to 3%, were obtained by enrichment cultures. This shows that only a fraction of the microorganisms from these habitats can be cultured under laboratory conditions. Of the three soda lakes, samples from Lake Chitu showed the highest prokaryotic diversity, while samples from Lake Shala showed the lowest diversity. *Pseudomonadota* (*Halomonas*), *Bacillota* (*Bacillus*, *Clostridia*), *Bacteroidota* (*Bacteroides*), *Euryarchaeota* (*Thermoplasmata*, *Thermococci*, *Methanomicrobia*, *Halobacter*), and *Nanoarchaeota* (*Woesearchaeia*) were the most common prokaryotic microbes in the three soda lakes. A high diversity of eukaryotic organisms were identified, primarily represented by *Ascomycota* and *Basidiomycota*. Compared to the other two lakes, a higher number of eukaryotic OTUs were found in Lake Abijata. The present study showed that these unique habitats harbour diverse microbial genetic resources with possible use in biotechnological applications, which should be further investigated by functional metagenomics.

KEYWORDS

microbial diversity, extremophiles, haloalkalophiles, operational taxonomic units, soda lakes

Introduction

Extremophiles are organisms that are able to thrive in environments considered extreme, at least from the human perspective. These environments include extreme physical factors such as pressure, radiation, and temperature and geochemical factors such as desiccation, pH, and salinity (Shrestha et al., 2018; Zhu et al., 2020). While most extremophiles are able to exist and grow under a single extreme environmental condition, a few can grow under multiple extreme conditions. Soda lakes, with both high salinity and alkalinity, are good examples of ecosystems with double extremes. Though such polyextreme environments were expected to have limited biodiversity, recent studies showed that soda lakes have huge microbial biodiversity (Sorokin et al., 2014; Ersoy Omeroglu et al., 2021).

Soda lakes have worldwide distribution (Andreote et al., 2018), with many such ecosystems found along the East African Rift Valley (Schagerl, 2016). Investigations on the microbial diversity of Ethiopian soda lakes have resulted in the identification of diverse groups of (halo)alkaliphilic bacterial and archaeal phyla (Martins et al., 2001; Lanzén et al., 2013; Simachew et al., 2016; Islam et al., 2021). Most studies of extremophiles originating from the Ethiopian soda lakes have so far been based on culture-dependent systems with the aim to selectively isolating pure culture strains to understand their physiology or to evaluate their potential in biotechnological applications (Gessesse and Gashe, 1997; Martins et al., 2001; Haile and Gessesse, 2012). A major drawback of this approach is that only a small fraction of the microorganisms are able to grow in standard culture media (Bodor et al., 2020). This makes a culture-dependent approach, in addition to being laborious and time-consuming (Maukonen et al., 2003; Kambura et al., 2016; Rosenthal et al., 2017), not reliable for studying microbial diversity. Therefore, a better and more reliable estimate of microbial diversity in any natural habitat can be obtained using culture-independent methods using molecular techniques.

Although the microbial diversity from the three Ethiopian soda lakes was recently studied using culture-independent methods, these were based on the analysis of clone libraries (Tafesse, 2014) or sequencing platforms that produce low read depths (Lanzén et al., 2013; Simachew et al., 2016). Due to the low sequence capture in the cloning step, while creating DNA libraries, and the low number of amplicon sequence reads generated, these techniques have inherent limitations in accurately reflecting the true picture of microbial diversity (Burke and Darling, 2016; Besser et al., 2018). Moreover, in previous culture-independent studies, little attention was given to studying the microbial diversity of eukaryotes from soda lakes (e.g., Lanzén et al., 2013). On the other hand, studies of other extreme environments have demonstrated the presence of extremophilic fungi that grow and reproduce at elevated temperatures or under alkaline and high salinity (Orwa et al., 2020). Such extremophilic eukaryotes, apart from their ecological function (de Oliveira and Rodrigues, 2019), may serve as potential sources of new biotechnological products (Ali et al., 2019).

The Ethiopian soda lakes are known for their high primary productivity, which in fact, represents one of the highest for any natural environment (Wood et al., 1984). Studies based on culture-independent and high-throughput next-generation sequencing have the potential to increase our understanding of the diversity of the microbiota in these ecosystems and to further elucidate the possible uses in biotechnological applications. The main aim of this study was to estimate the diversity of prokaryotic and eukaryotic microorganisms in samples of three soda lakes of the East African Rift Valley by a culture-independent analysis utilizing Illumina HiSeq sequencing. In addition, we sequenced the microbial communities after the samples were cultured in the laboratory to explore whether some of these organisms could be maintained in controlled conditions.

Materials and methods

Sampling sites, sample collection and preparation

Water and sediment samples from three soda lakes in the East African Rift Valley, Lakes Abijata, Chitu, and Shala, were collected using sterile Niskin bottles (Ocean Scientific International Ltd.) and polyethylene bags, respectively (Figure 1). Samples from 19 sites were collected in triplicates (Supplementary Table 1). The sampling sites were selected randomly based on the accessibility of the lakes. Geographical coordinates, depths, pH, and salinity were measured using GPS, Ekman grab attached with meter, pH meter (OAKTON-pH110), and refractometer (HHTEC; Supplementary Table 1). The water samples were divided into two parts, with one part filtered within 24 h after sample collection, using a polycarbonate filter membrane (22 μm pore size, 47 mm diameter; GE) to collect the microbes for DNA extraction.

For the enrichment culture study, the second part of the water samples was again divided into two halves, with one half filtered as mentioned above and the other half left unfiltered. The filtered lake water (brine) samples were used as a culture media and with unfiltered water samples as inoculum (5% v/v). The sediment samples were suspended in distilled water first, and once the sediment particles settled, the water was used as an inoculum. The inoculated brine was then incubated on a shaker (INFORS HT Ecotron Incubator; 120 rpm) for up to 7 days at room temperature. Until the seventh day, aliquots of each sample were taken every 24 h. The aliquots were centrifuged for 12 min at 4500 × g, and the cell pellets were pooled and saved for DNA extraction.

DNA extraction

DNA was extracted from the water and enrichment culture samples according to Øvreås et al. (2003) with some

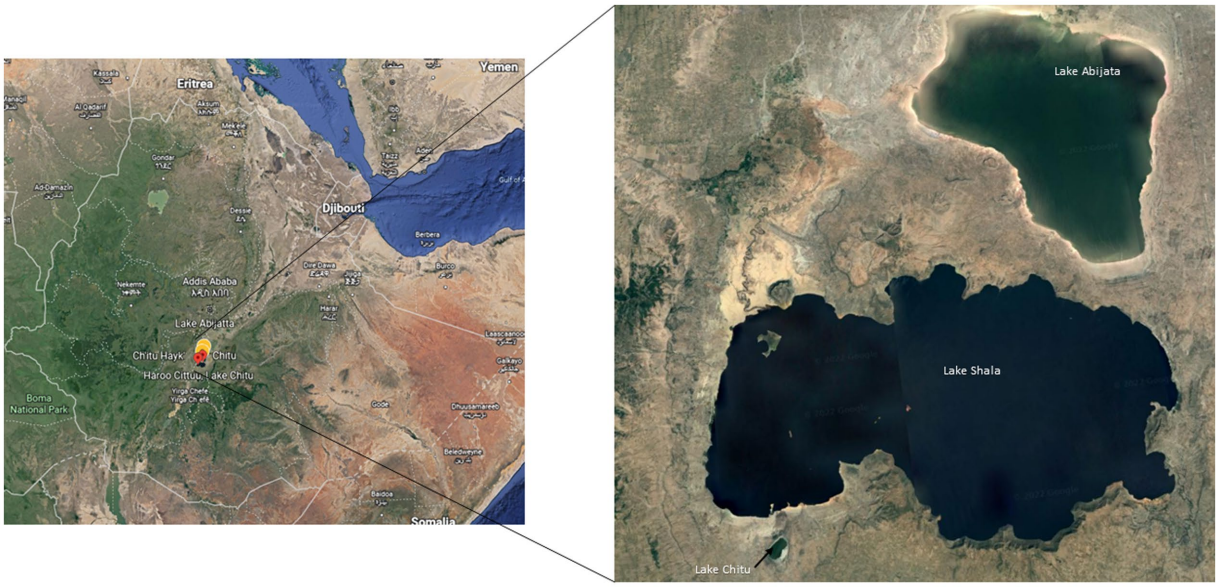


FIGURE 1
Sampling locations of the three soda lakes in the Ethiopian Rift Valley (image adapted from Google map 7th September 2021).

modifications as described below. Briefly, 250 µl of Lysozyme (1 mg/ml)/RNase solution (0.5 mg/ml, Thermo Scientific) was added to the microbial biomass on polycarbonate filter membrane followed by incubation for 15 min at 37°C. After that, 10 µl of Proteinase K (1 mg/ml, Thermo Scientific) was added to the filter, which again was incubated for 15 min at 37°C, followed by the addition of 250 µl pre-heated sodium dodecyl sulfate (SDS) solution (10% w/v) before further incubation for 15 min at 55°C. Finally, 80 µl of NaCl (5 M) and 100 µl of cetyltrimethylammonium bromide (CTAB; 1%) were added and incubated for 10 min at 65°C followed by the addition of 750 µl chloroform/isoamyl alcohol (24:1). The mixture was centrifuged at 12,000 × g for 15 min, and the aqueous layer was transferred into Eppendorf tubes and precipitated with 0.6 volume of isopropanol. After centrifugation at 12,000 × g for 15 min, the pellet was washed with 70% ethanol, dried at room temperature, and dissolved in TE buffer (pH 8.0).

DNA from the sediment samples was extracted according to Verma and Satyanarayana (2011) with some modifications as described below. About 10 g of sediment placed in sterile Falcon tubes were suspended in 13.5 ml of extraction buffer [1% CTAB, (w/v), 100 mM Tris (pH 8.0), 100 mM NaH_2PO_4 (pH 8.0), 100 mM EDTA and 1.5 M NaCl]. Thereafter, 50 µl of proteinase K (10 mg/ml, Thermo Scientific) was added, followed by incubation for 30 min at 37°C. After that, 1.5 ml of 20% SDS was added, followed by incubation at 65°C for 2 h, with gentle inversion every 15 min. The samples were then centrifuged at 4,000 × g for 20 min at room temperature, to separate the sediment residues from the cell lysates, which were transferred to new sterile centrifuge tubes. An equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) was added to the cell lysates, and samples were centrifuged at 16,000 × g for 5 min. The aqueous supernatant was again transferred to sterile centrifuge tubes, and an equal volume of chloroform was added to the tubes. After mixing, the tubes were centrifuged at 16,000 × g for 5 min. The aqueous layer was transferred to new sterile tubes, and the DNA was precipitated by adding about 0.6 volumes of isopropanol and recovered by centrifuging at 16,000 × g for 10 min. The pellet was washed with 70% ethanol and centrifuged at 16,000 × g for 5 min. Finally, the pellet was air-dried and dissolved in TE buffer (pH 8.0).

DNA purification and pooling

The extracted DNA was further purified using the DNeasy PowerSoil DNA extraction kit (QIAGEN), according to the manufacturer's instructions. About 250 µl aluminium chloride solution was added to each 50 µl DNA sample, which were then incubated for 5 min at room temperature. Thereafter, the samples were centrifuged at 15,000 × g for 2 min, and the supernatant was mixed with binding solution and loaded into the spin column, which was centrifuged at 10,000 × g for 1 min to discard the flow-through. The bound DNA was washed with 70% ethanol a few times and eluted with 50 µl of TE buffer (pH 8.0). Quantity and quality of DNA purified were checked using NanoDrop (260/280 ratio > 1.7; Thermo Scientific) and agarose gel electrophoresis (Thermo Scientific). The DNA of each biological triplicate were pooled before amplification of 16S rRNA and ITS genes.

Amplification of 16S rRNA and ITS gene sequences and sequencing

The V4 region of the 16S rRNA gene of the prokaryotic community was amplified using the primer sets 515F (5'-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5'-GGACTAC HVGGGTWTCTAAT-3′; Walters et al., 2016). In addition, ITS primer sets ITS1F (5'-CTTGGTCATTTAGAGGAAGTAA-3′) and ITS2R (5′- GCTGCGTTCTTCATCGATGC-3′) were used to amplify the ITS region of the fungal community (Op De Beeck et al., 2014). The 16S rRNA and ITS amplicon products were sequenced at BGI in Hong Kong using Illumina HiSeq Sequencing with a 2 × 50 pair end approach.

16S rRNA and ITS sequence quality control and filtering

The 16S rRNA and ITS sequences were analyzed using the Nextflow computational pipeline ampliseq v1.1.2.[1] Briefly, raw sequencing reads were quality checked using FastQC (Andrews, 2010), followed by trimming adaptor sequences from the reads using cutadapt v2.7 (Martin, 2011). Quality distribution of trimmed reads was then analyzed using tools available in the QIIME2 software package v2019.10 (Bolyen et al., 2019). After quality filtering, the sequences were denoised, dereplicated, and filtered for chimeric sequences using pair-ended DADA2 (Callahan et al., 2016), resulting in exact amplicon sequence variants (ASVs) tables.

OTU clustering and taxonomic classification

Amplicon sequence variants (ASVs) identified in 16S rRNA gene sequences were taxonomically classified from phylum to species level, and clustered with 99% similarity using the SILVA v132 database (Quast et al., 2013). Following the taxonomic classification of ASVs to OTUs (Operational Taxonomic Units), the OTUs classified as Mitochondria or Chloroplast were removed. The taxonomic classification of ASVs inferred in ITS sequences was performed using UNITE v8.99 database of ITS sequences, with the same parameters used for 16S rRNA gene sequences (Nilsson et al., 2019).

Diversity analysis and visualization

Analyses and visualizations of amplicon sequence variants (ASVs) were carried out *via* R (v4.0.4; R Core Team, 2014). The sequence reads were normalized to a minimal library size using the rarefying method (Weiss et al., 2017). The analysis of diversity measures such as alpha diversity, beta diversity, and ordination and visualization were performed using the R packages phyloseq v1.34.0 (McMurdie and Holmes, 2013), microbiome v1.12.0 (Lahti et al., 2017), and vegan v2.5.7 (Dixon, 2003).

1 https://github.com/nf-core/ampliseq

Results

Physico-chemical properties of the soda lakes

Salinity and pH differed between the lakes at the time of sampling, with the lowest salinity level of 3% in Lake Shala and the highest level of 15% in Lake Abijata, while the lowest pH of 9.3 was found in Lake Shala and the highest pH of 10.0 in Lake Chitu (Table 1; Supplementary Table 1).

Prokaryotic microbial sequence reads and OTUs

About 1,859,230 prokaryotic sequence reads (between 40,413 and 397,096 per sample) were captured in samples from the three investigated lakes using 16S rRNA sequencing. From these sequence reads, a total of 3,603 prokaryotic OTUs were identified (Table 1).

About 80% (2,872) of the OTUs were unique to samples of one lake, while the remaining 20% (731 OTUs) were detected in samples from two or all three lakes (Table 1). Samples of Lake Chitu showed the highest OTU abundance among the lakes, accounting for 75% of all the OTUs detected in the samples of all the three lakes, while samples of Lake Shala showed the lowest OTU abundance (14%) with only 512 OTUs (Table 1).

The total sequence reads of water and sediment samples ranged from 266,209 to 988,054 in the three lakes (Table 1). For Lake Abijata, only 36% of the OTUs were detected in the water samples, while the highest proportion of OTUs (92%) were detected from the sediment samples. In contrast, 99% of the OTUs from, Lake Chitu were detected in the water samples and only 13% were detected in the sediment samples (Table 1). For Lake Shala, the division of OTUs between water and sediment samples were more equal, with 55% of the OTUs detected in the water samples and 70% of the OTUs from sediment.

Samples of Lake Chitu showed the highest alpha diversity among the lakes, while samples of Lake Shala had the lowest diversity (Figure 2A). Species diversity differed, with the highest species diversity in the Lake Chitu samples, CH33334, CH33421, and CH339, which were all water samples, although high species diversity was also noted for sediment samples from Lake Abijata (Figure 2B).

Shared and unique prokaryotic OTUs within and between the soda lakes

As presented in Figure 3A, a relatively low proportion (28, 12 and 25%, respectively) of the OTUs were shared between water (W)

10.3389/fmicb.2022.999876

TABLE 1 General features of the soda lakes, the detected prokaryotic OTUs and sequence reads of the environmental samples.

Lake	pH	Salinity (%)	OTUs detected			Number of reads			Unique OTUs*	Shared OTUs	% OTU Occurrence***	Domain		
			Water	Sediment	Total	Water	Sediment	Total	(%)	(%)		Bacteria	Archaea	Unidentified
Abijata	9.5	15	459	1,168	1,276	251,101	353,866	604,967	18.2	17.2	35.4	1,166	110	0
Chitu	10	6	2,659	355	2,692	721,627	266,427	988,054	55.1	19.7	74.7	2,285	406	1
Shala	9.3	3	280	359	512	215,467	50,742	266,209	6.4	7.8	14.2	503	9	0
Total				3,603	3,603		1,859,230	1,859,230	79.7	20.3	100		3,603	

*Unique OTUs refers to those that are detected only in samples of one lake.
**Shared OTU refers to those that occur in samples of two or all three lakes.
***OTU occurrence refers to the proportion of OTUs detected in samples of each lake out of the total OTU detected.

and sediment (S) samples in the Abijata (AB), Chitu (CH) and Shala (SH) lakes. Similarly, a rather low proportion of the OTUs were shared among the different lakes (Figure 3B), e.g., only 146 OTUs (about 4% of the total) were shared between samples of the three lakes, and only a total of 451 OTUs (about 12%) were shared between samples of Lake Abijata and Lake Chitu. As pointed out above, samples of Lake Chitu had the highest number of OTUs among samples of the lakes, but also, the proportion of unique OTUs (74%) was the highest (Figure 3B). Analysis by one-way ANOVA showed that the OTUs between the lakes were significantly different (F-value = 10.1 $p < 0.05$). A Principal Coordinates Analysis (PCoA) displayed a spread of the samples' microbial composition from the different lakes along the first principal coordinate (explaining 25.9% of the variation), visualizing the variation in OTUs between samples also within each of the three lakes. The second principal coordinate divided the Lake Shala and Chitu samples from the Lake Abijata samples, indicating a higher similarity among Lake Shala and Lake Chitu OTUs than for Lake Abijata OTUs, although the second coordinate had a relatively low level of explanation of variation (18.2%; Figure 3C).

Prokaryotic taxonomic distribution

With 99% sequence similarity, 87% of OTUs (3,120) were identified as bacteria, 13% (482) as archaea, while a single OTU was not classified as either Bacteria or Archaea using the SILVA database (Table 1).

The OTUs were further classified into 58 phyla (50 identified bacterial phyla, 7 identified archaeal phyla and 1 unidentified phylum; Figure 4), where three phyla, *Pseudomonadota, Bacillota,* and *Bacteroidota,* were the most abundant and represented 49% of the OTUs (Figure 4) whereas *GN01, Rokubacteria, Asgardaeota, Fibrobacterota, MAT-CR-M4-B07, FCPU426, Chrysiogenota*, PAUC34f, *WS2,* and *Modulibacteria* had the lowest abundance (Supplementary Table 2).

Bacterial community composition

For Lake Abijata and Lake Shala, the phylum *Pseudomonadota* was the most dominant in both water and sediment samples, followed by *Bacillota*, while for Lake Chitu, the phylum *Bacillota* was represented by slightly more OTUs in both water and sediment samples than *Pseudomonadota. Bacteroidota* was the third most abundant phylum in samples in all of the lakes (Table 2; Figure 4; Supplementary Table 2).

In addition, *Cyanobacteria* were one of the most prevalent phyla in the three lakes studied (Figure 4; Supplementary Table 2). *Cyanobiaceae* (*Cyanobium*), *Phormidiaceae* (*Arthrospira*), *Cyanobacteriaceae* (*Cyanobacterium*), *Nostocaceae* (*Nodularia*), and *Paraspirulinaceae* (*Spirulina*) were the prevalent taxa in the lakes within the phylum *cyanobacteria* (Supplementary Table 2).

10.3389/fmicb.2022.999876

TABLE 2 Distribution of bacterial phylum of the environmental samples identified through cuture-independent investigation.

Phylum	All		Abijata		Chitu		Shala	
	OTUs	RA%	OTUs	RA%	OTUs	RA %	OTUs	RA%
Pseudomonadota	605	19.4	266	22.8	370	16.2	337	19.8
Bacillota	519	16.6	222	19	393	17.2	330	19.4
Bacteroidota	447	14.3	179	15.4	305	13.4	254	14.9
Planctomycetota	203	6.5	51	4.4	162	7.1	84	4.9
Patescibacteria	119	3.8	37	3.2	90	3.9	48	2.8
Spirochaetota	116	3.7	38	3.3	96	4.2	62	3.6
Actinomycetota	113	3.6	61	5.2	67	2.9	66	3.9
Chloroflexota	104	3.3	41	3.5	84	3.7	62	3.6
Kiritimatiellaeota	98	3.1	25	2.1	84	3.7	59	3.5
Mycoplasmatota	87	2.8	31	2.7	60	2.6	46	2.7
Unidentified	81	2.6	14	1.2	68	3	22	1.3
Others	628	20.1	201	17.2	502	22	330	19.4
Grand Total	3,120	100	1,166	100	2,281	100	1700	100

Others: Verrucomicrobiota, Halanaerobiaeota, Gemmatimonadota, BHI80-139, Omnitrophicaeota, Cloacimonetes, Acidobacteria, Elusimicrobiota, Hydrogenedentes, Chlamydiota, Deinococcota, Lentisphaerota, uncultured, BRC1, Cyanobacteria, WS1, Dependentiae, Fibrobacterota, Thermotogota, Armatimonadota, Fusobacteriota, Acetothermia, Synergistota, Margulisbacteria, Atribacterotaia, Epsilonbacteraeota, LCP-89, TA06, Caldisericota, CK-2C2-2, Latescibacteria, MAT-CR-M4-B07, Modulibacteria, WOR-1, Chrysiogenota, FCPU426, PAUC34f, Rokubacteria, GN01, WS2. *RA: Relative abundance as proportion out of the total.

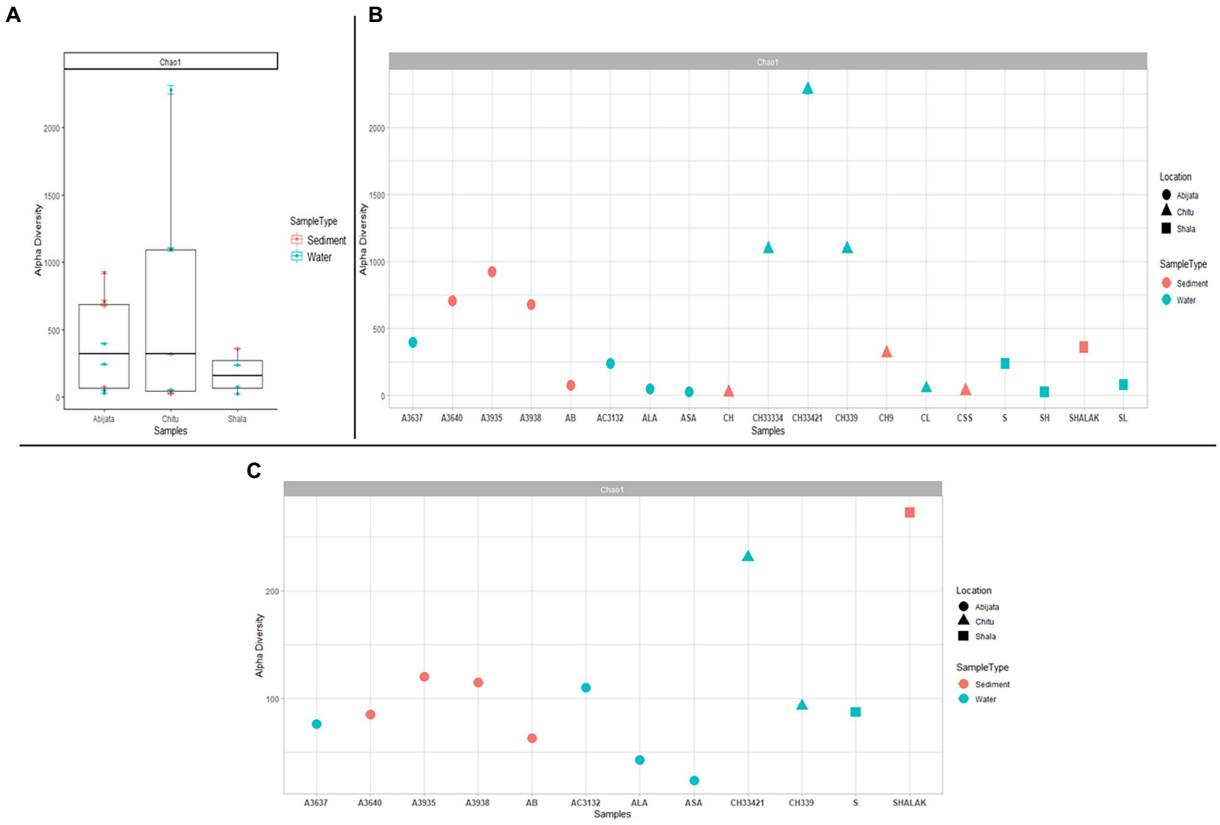


FIGURE 2
Alpha diversity for prokaryotes in the lakes **(A)** either comparing between the three lakes **(B)** or as a diversity between all samples. **(C)** Alpha diversity measures between samples for eukaryotes. The sample names are on the x-axis and the estimated diversity on the y-axis.

Halomonadaceae, *Ectothiorhodospiraceae*, and *Idiomarinaceae* were the most abundant families of phylum *Pseudomonadota*, with *Halomonadaceae* as the most abundant family in lakes Abijata and Shala. At the same time, *Syntrophomonadaceae* were the most abundant family in lake Chitu. Furthermore, *Bacillota* (*Bacillaceae*), *Rhodothermia* (*Bacteroidota*), and *Nitriliruptoria*

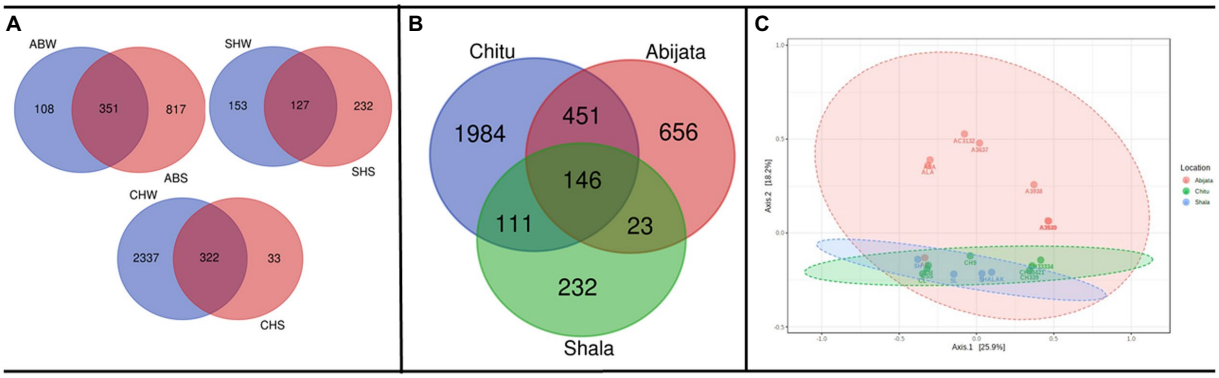
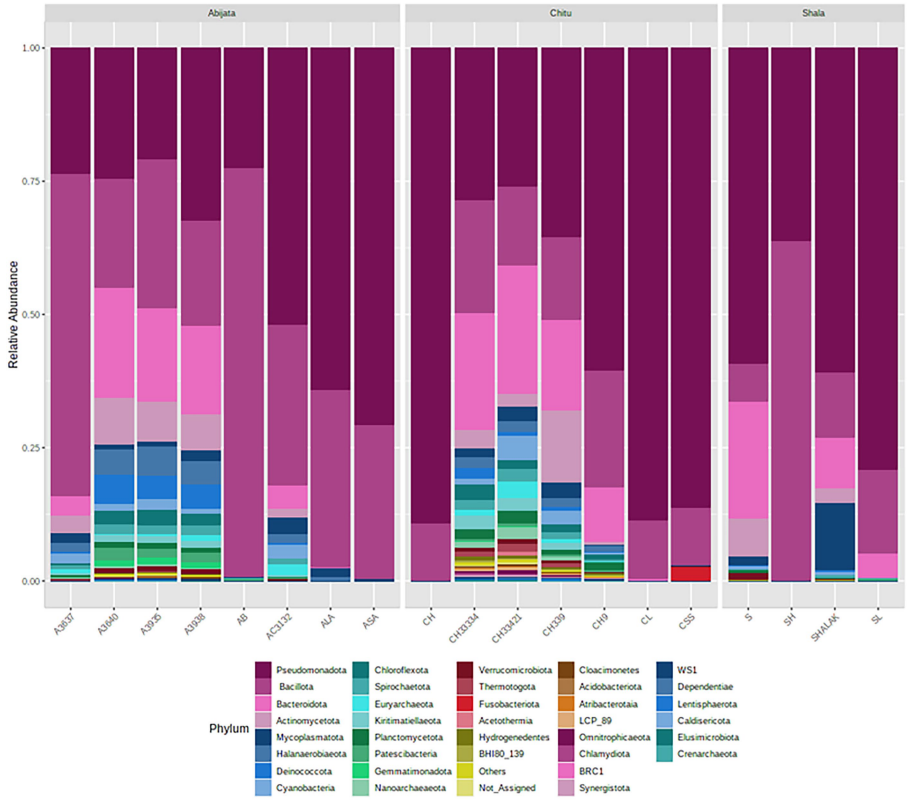


FIGURE 3
Shared and unique prokaryotic OTUs **(A)** between sample types in the three soda lakes; ABW: Abijata water samples, ABS: Abijata Sediment samples, CHW: Chitu water samples, CHS: Chitu sediment samples, SHW: Shala water samples, SHS: Shala sediment samples **(B)** between the lakes **(C)** PCoA using Bray distance; Statistical significance was found out using Analysis of group Similarities [ANOSIM] R: 0.32947; value of p <0.005. Sample names (Water samples: A3637, AB, AC3132, ALA, ASA, CH33334, CH33421, CH339, CL, S, SH, SL; Sediment samples: A3640, A3935, A3938, CH, CH9, CSS, SHALAK).



Sample	Location	Sample Type
A3637	Abijata	Water
A3640	Abijata	Sediment
A3935	Abijata	Sediment
A3938	Abijata	Sediment
AB	Abijata	Sediment
AC3132	Abijata	Water
ALA	Abijata	Water
ASA	Abijata	Water
CH	Chitu	Sediment
CH33334	Chitu	Water
CH33421	Chitu	Water
CH339	Chitu	Water
CH9	Chitu	Sediment
CL	Chitu	Water
CSS	Chitu	Sediment
S	Shala	Water
SH	Shala	Water
SHALAK	Shala	Sediment
SL	Shala	Water

FIGURE 4
Stacked plot showing the taxonomic composition of the community at the phylum level (small taxa with read counts <100 merged as others).

(*Actinomycetota*) were among the most abundant taxa (Supplementary Table 3).

About 37.4% of OTUs were unidentified in the Silva database at the species level. Most of the identified species are listed as uncultivable bacterial species (Supplementary Table 2).

Archaeal community composition

In samples from Lake Chitu, seven archaeal phyla were present, with two phyla, *Euryarchaeota* and *Nanoarchaeaeota*, accounting for 88% of the OTUs (Table 3). In samples from

10.3389/fmicb.2022.999876

Lake Abijata, five archaeal phyla were detected. Again, *Euryarchaeota* and *Nanoarchaeaeota* were dominant, like in Lake Chitu, accounting for 91% of the OTUs. In samples of Lake Shala, the phylum *Nanoarchaeaeota* was the only archaea detected (Table 3).

The *Thermoplasmata*, *Thermococci*, *Methanomicrobia*, and *Halobacteria* were the most abundant groups of *Euryarchaeota*. More than 96% of the *Nanoarcheota* belonged to the class *Woesearchaeia*, and the remaining 4% was unclassified. In samples of Lake Abijata, 60% of *Euryarchaeota* belonged to *Halobacteria,* whereas in samples of Lake Chitu, 78% of OTUs of *Euryarchaeota* belonged to *Thermoplasmata* and *Methanomicrobia* (Supplementary Table 3). Around 39% of the archaeal sequences were unidentified archaeal species, and more than 60% of the sequences were listed as uncultured (Supplementary Table 2).

Eukaryotic community composition

By ITS sequencing, a total of 1,0250,22 eukaryotic reads were identified. These reads were categorized into 898 OTUs (Table 4). About 51% of OTUs were represented in samples of Lake Abijata, followed by 37% in samples of Lake Shala. No eukaryotic sequences were detected in Lake Chitu sediment samples (Table 4). The alpha diversity showed that the Lake Shala sediment sample (SHALAK) had the highest species diversity (Figure 2C). All eukaryotic reads were identified to be fungal.

TABLE 3 Archeael phyla distribution among the lakes identified from environmental samples through culture-independent investigation.

Phylum	Number Of OTUs							
	All three lakes		Abijata		Chitu		Shala	
	No.	RA (%)	No.	RA (%)	No.	RA (%)	No.	RA (%)
Altiarchaeota	4	0.8	0	0	4	1	0	0
Asgardaeota	9	1.9	0	0	9	2.2	0	0
Crenarchaeota	24	5	6	5.5	20	4.9	0	0
Diapherotrites	8	1.7	2	1.8	8	2	0	0
Euryarchaeota	114	23.7	53	48.2	74	18.2	0	0
Hadesarchaeaeota	9	1.9	2	1.8	7	1.7	0	0
Nanoarchaeaeota	313	64.9	47	42.7	283	69.7	9	100
Unidentified	1	0.2	0	0	1	0.3	0	0
Grand Total	**482**	**100**	**110**	**100**	**406**	**100**	**9**	**100**

*RA: Relative abundance (Proportion out of the total).

Shared and unique fungal OTUs within and between samples of the lakes

Sediment samples of Lake Abijata and Lake Shala had a higher number of unique fungal OTUs than water samples from the same lakes (Figure 5A). Only 46 OTUs were shared between samples from all three lakes, with samples of Lake Chitu and Lake Shala sharing 23 OTUs (Figure 5B). One-way ANOVA analysis showed a significant difference between OTUs of the lakes (F-value = 9.0, value of p <0.05). The PCoA showed a similar variation of samples from Lake Shala and Lake Chitu, with positive values on both principal coordinates, while some of the Lake Abijata samples clustered with positive first principal coordinate values and negative second coordinate values (Figure 5C).

Fungal community taxonomic distribution

The ITS sequence reads of the fungal communities were categorized into four phyla, where *Ascomycota* and *Basidiomycota* were the most abundant in all lake samples. All the identified eukaryotic phyla were found in samples of Lakes Abijata and Shala. Lake Shala had the highest proportion of unidentified sequence reads (>17%; Table 5).

All the fungal communities were categorized into 19 classes. The *Dothideomycetes*, *Saccharomycetes,* and *Sordariomycetes* from phylum *Ascomycota*; *Agaricomycetes*, *Tremellomycetes,* and *Malasseziomycetes* from phylum *Basidiomycota* were the abundant groups, respectively. More than 50% of fungal OTUs were not identified at the species level (Supplementary Table 4).

TABLE 4 The obtained eukaryotic OTUs and sequence of the environmental samples identified through culture-independent investigation.

Lake	OTUs detected		Number of reads		Total OTUs	Total Reads	% Reads	Unique OTUs*	Shared OTUs	% OUT Occurrence***
	Water	Sediment	Water	Sediment				(%)	(%)	
Abijata	199	320	467,937	521,659	460	9,89,596	71.6	37.9	13.4	51.2
Chitu	293	0	152,889	0	293	1,52,889	11	21.3	11.4	32.6
Shala	87	271	94,098	146,021	334	2,40,119	17.4	25	12.5	37.2
Total	898		1,025,022		898	1,025,022	100	84.1	15.9	100

*Unique OTUs refer to those that are detected only in one lake.
**Shared OTUs refer to those that occur in two or all three lakes.
***OTUs occurrence refers to the proportion of OTUs detected in each lake out of the total OUT detected.

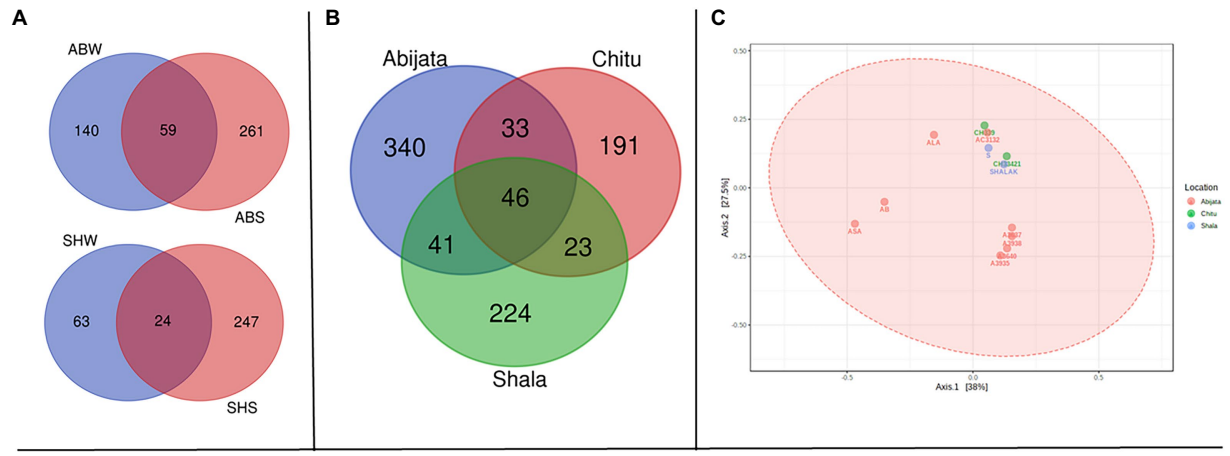


FIGURE 5
Shared OTUs between **(A)** sample types; ABW: Abijata water samples, ABS: Abijata Sediment samples, SHW: Shala water samples, SHS: Shala sediment samples **(B)** between the lakes **(C)** PCoA plot using Bray distance. The explained variances are shown in brackets (F-value: 1.1663; R-squared: 0.20583; p>0.05; Sediment Samples: A3640, A3935, A3938, AB, SHALAK; Water Samples: A3637, AC3132, ALA, ASA, CH33421, CH339, S).

TABLE 5 Fungal phylum composition of environmental samples identified through culture-independent investigation.

Phylum	All of the lakes		Abijata		Chitu		Shala	
	OTUs	RA (%)	OTUs	RA (%)	OTUs	RA (%)	OTUs	RA (%)
Ascomycota	454	50.56	225	48.9	161	55	180	53.9
Basidiomycota	283	31.51	154	33.5	102	34.8	92	27.5
Chytridiomycota	7	0.78	6	1.3	0	0	1	0.3
Rozellomycota	3	0.33	1	0.2	0	0	2	0.6
Unidentified	151	16.82	74	16.1	30	10.2	59	17.7
Total	898	100	460	100	293	100	334	100

Enrichment cultures

Enrichment cultures yielded a total of 134 OTUs, which belonged to five phyla. Of these, *Pseudomonadota*, *Bacillota* and *Bacteroidota* accounted for more than 95% of the identified OTUs. No archaeal or eukaryotic groups were identified in the enrichment samples (Table 6). *Pseudomonadota* were represented by a higher number of OTUs in Lake Chitu and Shala samples, whereas Bacillota was the most abundant phylum in Lake Abijata (Table 6).

In total, only 3.0% of the OTUs identified in the culture-independent (environmental) samples, were detected in the enrichment samples, all of which were bacterial OTUs. Thus, out of the 50 bacterial phyla that were identified from environmental samples, only 5 phyla were detected in the enrichment cultures (Table 7). Three phyla, *Pseudomonadota*, *Bacillota*, and *Bacteriodetes* with high abundance in the environmental samples were also the most abundant in the enrichment cultures (Table 7).

Further analysis of taxa levels found that more than 83% of the *Pseudomonadota* were *Gammaproteobacteria*, in which *Idiomarinaceae* and *Halomonadaceae* accounted for 60%. In the *Bacillota* phylum, more than 80% of the OTUs belonged to the *Bacilli* class, where the remaining was *Clostridia*. Furthermore, at the species level 46.5 and 16.2%, belonged to unidentified and ambiguous taxa, respectively (Supplementary Table 5).

Discussion

We observed high prokaryotic and eukaryotic microbial diversity from the three Ethiopian soda lakes using amplicon sequencing. The current study also demonstrated the challenges of capturing these unique microbes, which are adapted to extreme environments, using typical enrichment cultures. Although previous studies have also reported high biodiversity of microorganisms in the Ethiopian soda lakes (Lanzén et al., 2013; Simachew et al., 2016), significantly higher levels were found in the present study. We identified a total of 3,603 prokaryotic and 898 eukaryotic OTUs in the samples from the three soda lakes, which is a significantly higher number than the 2,704 OTUs reported in a previous study from five Ethiopian soda lakes, including the three lakes studied here (Lanzén et al.,

TABLE 6 The Phylum distributions identified through enrichment culture study.

Phyla	All		Abijata		Chitu		Shala	
	OTUs	OTUs (%)	OTUs	OTUs (%)	OTUs	OTUs (%)	OTUs	OTUs (%)
Actinomycetota	1	0.75	1	1.1	1	1.7	1	1.6
Bacteroidota	33	24.6	23	24.5	8	13.8	16	25.0
BacillotaBacillota	47	35.1	34	36.2	16	27.6	23	35.9
Pseudomonadota	48	35.8	32	34.0	32	55.2	24	37.5
Mycoplasmatota	5	3.7	4	4.3	1	1.7		0.00
Grand Total	134	100	94	100	58	100	64	100

2013). In fact, the OTU richness observed in this study was also higher than that reported previously for other hypersaline lakes (Dimitriu et al., 2008b; Zorz et al., 2019). Here, we used high-abundance short-read sequencing for the estimation of the microbial diversity to circumvent the shortcomings of sequencing methods with low coverage and sequence depth used in previous studies of the Ethiopian soda lakes. However, the three investigated soda lakes differed in microbial biodiversity; 75% of all OTUs detected in the three lakes were present in samples of Lake Chitu, while only 35 and 14% of the OTUs were present in samples of Lake Abijata and Lake Shala, respectively. This can be explained by the differences in the primary production of the lakes (Ogato et al., 2015). The East African Rift Valley soda lakes are among the world's most productive aquatic systems with primary production rates exceeding 10 g C m^{-2} day^{-1} (Grant and Jones, 2016), providing abundant organic matter supporting a diverse group of microorganisms. Lake Chitu is the most productive among the three lakes, having thick biomass of the blue-green algae, *Arthrospira fusiformis* (Faris, 2017).

In the present study, samples of Lake Chitu and Lake Abijata shared the highest number of prokaryotic OTUs, while in previous studies, a higher overlap of OTUs was observed between Lake Shala and Lake Abijata (Lanzén et al., 2013). This discrepancy in overlaps might result from the fourfold salinity increment in Lake Abijata, which is a higher increase than in the other Ethiopian soda lakes over the last three decades (Ayenew and Legesse, 2007; Klemperer and Cash, 2007; Faris, 2017). During the last three decades, the surface area of Lake Abijata has decreased by more than 50% and the average depth by 5 m as a result of the abstraction of water for soda ash production by the Abijata-Shala Soda Ash Factory and the diversion of the lake's feeder rivers, Horakelo and Bulbula, for irrigation (Ayenew and Legesse, 2007; Simachew et al., 2016). Several studies have shown that a change in salinity is an important environmental factor affecting the abundance and diversity of microorganisms, eliminating some and benefiting other groups (Lozupone et al., 2007; Xiong et al., 2012). A change in the microbial composition of Lake Abijata over time has been reported in previous studies (Simachew et al., 2016). Thus, salinity might be a critical factor in structuring the prokaryotic communities of haloalkaline environments, as also indicated by the present study.

Bacteria accounted for the vast majority of the prokaryotic OTUs identified in this study (87%), with Archaea accounting for only a small fraction (13%). The primary explanation for this might be that these lakes are moderately saline soda lakes with brine salinities ranging from 50 to 250 g/l (Sorokin et al., 2014). Previous research on microbial diversity in moderately saline soda lakes found that these ecological niches have more diverse archaea and bacteria communities than less and high saline environments, with bacteria dominating (Felföldi, 2020).

In line with previous studies from low and moderate saline soda lakes, *Gammaproteobacteria* (including the genus *Halomonas*), *Bacillota* (*Bacillus, Clostridia*), *Bacteroidota* (*Cytophaga*, *Flavobacterium*, *Bacteroides*), and *Rhodobacteraceae* (Humayoun et al., 2003; Dimitriu et al., 2008a; Asao et al., 2011; Lanzén et al., 2013; Sorokin et al., 2014) were the prevailing prokaryotes in this study as well. The groups *Halomonas*, *Bacillota*, and other heterotrophic bacteria are primarily responsible for immediate degradation of organic matter produced by autotrophic bacteria like Cyanobacteria (Jones and Grant, 1999; Sorokin et al., 2014). *Halomonas* have been reported to have biotechnological potential in the production of exopolysaccharides, enzymes, and compatible solutes such as ectoine (Ye and Chen, 2021). In addition, *Halomonas* has an active role in denitrification and the degradation of aromatic compounds (Tahrioui et al., 2013). *Rhodobacteraceae*, another abundant microbial taxon in the soda lakes, comprises aerobic photo- and chemoheterotrophs and purple non-sulfur bacteria, which perform photosynthesis in anaerobic environments. They are deeply involved in sulfur and carbon biogeochemical cycling and are in symbiosis with aquatic micro- and macroorganisms (Pujalte et al., 2014).

Environmental heterogeneity with more ecological niches has been pointed out as an important variable for high microbial biodiversity (Wobus et al., 2003). Such heterogeneity has been reported as the reason for higher biodiversity in sediment than in corresponding water samples (Zinger et al., 2011; Banda et al., 2020), which was also the case in the present study. In this study, the sediment samples from Lake Abijata and Lake Shala showed higher levels of prokaryote and eukaryote OTUs than the corresponding water samples. Furthermore, sediment samples from Lake Abijata and Lake Shala were taken as lake surface sediment samples. Lake

10.3389/fmicb.2022.999876

TABLE 7 Comparison of OTUs detected in through culture-independent and enrichement culture studies.

Phylum	Total OTU detected		% grown in culture
	Environmental samples	Enrichement samples	
Bacteria			
Pseudomonadota	605	48	7.9
Bacillota	519	47	9.1
Bacteroidota	447	33	7.4
Planctomycetota	203	0	0
Patescibacteria	119	0	0
Spirochaetota	116	0	0
Actinomycetota	113	1	0.9
Chloroflexota	104	0	0
Kiritimatiellaeota	98	0	0
Mycoplasmatota	87	5	5.7
Unidentified	81	0	0
Others	628	0	0
Total Bacteria	**3,120**	**134**	**4.3**
Archaea			
Altiarchaeota	4	0	0
Asgardaeota	9	0	0
Crenarchaeota	24	0	0
Diapherotrites	8	0	0
Euryarchaeota	114	0	0
Hadesarchaeaeota	9	0	0
Nanoarchaeaeota	313	0	0
Unidentified	1	0	0
Total Archaea	**482**	**0**	**0**
Unassigned prokaryote 1 0 0			
Eukaryotes			
Ascomycota	454	0	0
Basidiomycota	283	0	0
Chytridiomycota	7	0	0
Rozellomycota	3	0	0
Unidentified	151	0	0
Total Eukaryotes	**898**	**0**	**0**
TOTAL	**4,501**	**134**	**3**

surface sediments contain abundant and diverse microbial populations (Yang et al., 2016), and microorganisms in lake surface sediments play vital roles in regulating nutrient dynamics and biogeochemical cycles (Zhang et al., 2020). In contrast to Lake Abijata and Lake Shala, and to previous results (Tafesse, 2014), Lake Chitu showed a higher number of prokaryotic OTUs in the water samples than the sediment samples. This might be attributed to abundant algal biomass and other organic components in the water of Lake Chitu that sustain diverse microbial communities (Ogato et al., 2015). In samples of Lake Abijata, *Halobacteria* was the most abundant class, possibly because this lake had higher salinity compared to the other two lakes during our sampling. Hypersaline lakes have in previous studies, been reported to be dominated by halophilic archaea belonging to the class *Halobacteria* (Ochsenreiter et al., 2002; Mesbah et al., 2007).

The archaeal OTUs in the samples from Lake Shala and Chitu were associated to the *Nanoarchaeaeota* phylum that belongs to the class of *Woesearchaeia*. These are likely anaerobic, fermentative, and syntrophic archaea and are often found in marine environments with high organic matter content (Gründger et al., 2019). However, it is unknown how geochemical circumstances shape the distribution pattern of *Woesearchaeia* and their ecological role, particularly on a global scale (Xiao et al., 2020). *Woesearchaeia* can enable carbon and hydrogen metabolism in anoxic circumstances, but they are also associated with symbiotic or parasitic lifestyles, which is reflected in their small genome sizes and limited metabolic capabilities. Furthermore, certain reports imply that this archaeal group plays an essential role in iron and methane biogeochemical cycles (Liu et al., 2018; Gründger et al., 2019).

Thermoplasmata, *Methanomicrobia*, and *Thermococci* were dominant microbial prokaryotes in phylum *Euryarchaeota* of these lakes. Previous studies have explored the methane cycle in soda lakes as an essential part of the microbial carbon cycle (Sorokin et al., 2014), and *Thermoplasmata* (Poulsen et al., 2013), *Thermococci* (Evans et al., 2019), *Methanomicrobia* (Lee et al., 2014), and *Halobacteria* (Sorokin et al., 2017) have been evaluated for their role in methane cycling. The abundance of microorganisms involved in methane metabolism identified together with the presence of cyanobacteria points at high concentrations of dissolved biogenic methane in these lakes. In anoxic lake sediments, the abundant photosynthetic biomass eventually undergoes methanogenic degradation (Fazi et al., 2021).

While archaea and bacteria are known for their ability to adapt to extreme environments, fungi are generally not expected in saline soda lakes, since they generally prefer acidic or neutral pH (Grum-Grzhimaylo et al., 2016). However, recent studies have shown that eukaryotic microorganisms, including fungi, can endure or even thrive in extreme environments. The fungal diversity of the Ethiopian soda lakes has been rarely studied previously except from the report of Lanzén et al., 2013. In the present study, fungi were found in samples of all three lakes, with *Ascomycota* as the most abundant phylum in all three lakes. The Ascomycota orders *Capnodiales*, *Dothideales*, *Cladosporium*, *Alternaria*, and *Eurotiales* have previously been found in different extreme environments globally, including saline lakes, soda lakes, and in Arctic glacier ice (Cantrell et al., 2006; Anne et al., 2016; Fotedar et al., 2018; Orwa et al., 2020). Specifically, the detection of *Malassezia* in the soda lake samples studied here extends previous culture-independent studies of fungi that have shown that *Malassezia* is exceedingly widespread and ecologically diverse (Boekhout et al., 2010). Recent studies in little-characterized marine environments point to extensive diversification of Malassezia-like organisms, suggesting further

opportunities to explore this group's ecology, evolution, and diversity (Amend, 2014).

The present study showed that culturing captures only a small proportion of naturally occurring microbial biodiversity; out of a total of 4,500 OTUs detected by sequencing of environmental samples, only 134 OTUs (representing 3% of the total) grew in enrichment cultures. Previous studies have indicated that since most microorganisms are impossible or difficult to culture, standard culture-dependent techniques provide information on 1% or less of the microbial diversity in a given environmental sample (Coughlan et al., 2015). One reason for this is that culturing cannot reproduce the natural condition of the ecological niches where samples are collected (Carraro et al., 2011). For example, the media selected for microbial cultivation have been shown to selectively favour a small fraction of the microbial diversity (Carraro et al., 2011; Al-Awadhi et al., 2013). To minimize the selectivity of the culture media due to, for example, lack of critical trace elements, we used lake water from the three lakes to mimic the natural environment in our enrichment culture study. Nevertheless, even under such conditions, 97% of the OTUs detected in the environmental samples could not be detected after cultivation. Since we used a mixed liquid culture, quorum sensing is probably not responsible for the failure of most organisms to grow (Bodor et al., 2020). In their natural habitat, these microorganisms are exposed to different concentrations of nutrients and oxygen due to, for example, the depth of the water and diurnal fluctuations, which are impossible to reproduce in the laboratory. Thus, our results clearly showed the importance of culture-independent methods in studying microbial diversity.

High microbial diversity under polyextremophilic conditions (high salt and alkaline pH), as observed here, is important for possible industrial applications. Enzymes that are active and stable under alkaline conditions and in the presence of high salt concentration potential applicability in different industrial processes, as shown by previous microbial cultivation and screening studies that have resulted in the production of novel enzymes from Ethiopian soda lakes (Gessesse and Gashe, 1997; Martins et al., 2001; Haile and Gessesse, 2012). Targeted sequencing of enzyme groups or whole genome sequencing will be powerful tools to secure unique enzymes with specific properties.

Conclusion

This study showed that the amplicon sequencing utilized was highly effective in detecting a high degree of microbial diversity in three samples from East African Rift Valley soda lakes, suggesting that increased sampling covering all microhabitats of these soda lakes would detect considerably more organisms. This study also showed the presence of eukaryotic microorganisms, such as fungi, in soda lakes. In addition, despite our attempts to replicate the natural environment for microbe cultivation, only a small fraction of microorganisms detected by sequencing could be cultured in the laboratory. Additional culture-independent studies are needed to successfully exploit the full potential of microorganisms, including the unculturable microbes, in soda lakes.

Data availability statement

The data presented in the study are deposited in the NCBI repository with accessions PRJNA816843 and PRJNA817405. This data can be found at: https://www.ncbi.nlm.nih.gov/sra/PRJNA816843 and https://www.ncbi.nlm.nih.gov/sra/PRJNA817405.

Author contributions

All authors contributed to the study conception and design. Experiment was designed by OJ, AG, AS, EJ, and EA. Data collection and analysis were performed by OJ. The first draft of the manuscript was written by OJ. All authors commented on previous versions of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was financed by the Swedish International Development Cooperation Agency (SIDA) through the research and training grant awarded to Addis Ababa University and the Swedish University of Agricultural Sciences (AAU-SLU Biotech)[2] and FORMAS (2019–00527).

Acknowledgments

The authors would like to thank the Wildlife Conservation Authority for providing an access permit to samples from the three Soda lakes. We would also like to thank the Biodiversity Institute and National Soil Testing Center, Ministry of Agriculture of Ethiopia, for granting permission for genetic material transfer.

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

2 https://sida.aau.edu.et/index.php/biotechnology-phdprogram

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.999876/full#supplementary-material

References

Al-Awadhi, H., Dashti, N., Khanafer, M., Al-Mailem, D., Ali, N., and Radwan, S. (2013). Bias problems in culture-independent analysis of environmental bacterial communities: a representative study on hydrocarbonoclastic bacteria. *Springerplus* 2, 1–11. doi: 10.1186/2193-1801-2-369

Ali, I., Khaliq, S., Sajid, S., and Akbar, A. (2019). "Biotechnological applications of Halophilic fungi: past, present, and future," in *Fungi in extreme environments: Ecological role and biotechnological significance*. eds. S. Tiquia-Arashiro and M. Grube (Cham: Springer), 291–306.

Amend, A. (2014). From dandruff to Deep-Sea vents: Malassezia-like fungi are ecologically hyper-diverse. *PLoS Pathog.* 10:277. doi: 10.1371/journal.ppat.1004277

Andreote, A. P. D., Dini-Andreote, F., Rigonato, J., Machineski, G. S., Souza, B. C. E., Barbiero, L., et al. (2018). Contrasting the genetic patterns of microbial communities in soda lakes with and without cyanobacterial bloom. *Front. Microbiol.* 9:244. doi: 10.3389/fmicb.2018.00244

Andrews, S. (2010). FastQC - a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Babraham Bioinforma.

Anne, K. K., Romano, K. M., Remmy, W. K., Edward, N. K., Huxley, M. M., and Hamadi, I. B. (2016). Diversity of fungi in sediments and water sampled from the hot springs of Lake Magadi and little Magadi in Kenya. *African J Microbiol Res* 10:879. doi: 10.5897/AJMR2015.7879

Asao, M., Pinkart, H. C., and Madigan, M. T. (2011). Diversity of extremophilic purple phototrophic bacteria in Soap Lake, a Central Washington (USA) Soda Lake. *Environ. Microbiol.* 13, 2146–2157. doi: 10.1111/j.1462-2920.2011.02449.x

Ayenew, T., and Legesse, D. (2007). The changing face of the Ethiopian rift lakes and their environs: call of the time. *Lakes Reserv. Res. Manag.* 12:332. doi: 10.1111/j.1440-1770.2007.00332.x

Banda, J. F., Lu, Y., Hao, C., Pei, L., Du, Z., Zhang, Y., et al. (2020). The effects of salinity and pH on microbial community diversity and distribution pattern in the brines of Soda Lakes in Badain Jaran Desert, China. *Geomicrobiol J.* 37:568. doi: 10.1080/01490451.2019.1654568

Besser, J., Carleton, H. A., Gerner-Smidt, P., Lindsey, R. L., and Trees, E. (2018). Next-generation sequencing technologies and their application to the study and control of bacterial infections. *Clin. Microbiol. Infect.* 24, 335–341. doi: 10.1016/j.cmi.2017.10.013

Bodor, A., Bounedjoum, N., Vincze, G. E., Erdeiné Kis, Á., Laczi, K., Bende, G., et al. (2020). Challenges of unculturable bacteria: environmental perspectives. *Rev Environ Sci Bio/Technology* 19, 1–22. doi: 10.1007/s11157-020-09522-4

Boekhout, T., Guého-Kellermann, E., Mayser, P., and Velegraki, A. (2010) in *Malassezia and the skin*. eds. T. Boekhout, P. Mayser, E. Guého-Kellermann and A. V. Berlin (Heidelberg: Springer Berlin Heidelberg)

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

Burke, C. M., and Darling, A. E. (2016). A method for high precision sequencing of near full-length 16S rRNA genes on an Illumina MiSeq. *PeerJ* 2016:492. doi: 10.7717/peerj.2492

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* 13:869. doi: 10.1038/nmeth.3869

Cantrell, S. A., Casillas-Martinez, L., and Martinez, M. (2006). Characterization of fungi from hypersaline environments of solar salterns using morphological and molecular techniques. *Mycol. Res.* 110, 962–970. doi: 10.1016/J.MYCRES.2006.06.005

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.* 162:002. doi: 10.1016/j.resmic.2011.01.002

Coughlan, L. M., Cotter, P. D., Hill, C., and Alvarez-Ordóñez, A. (2015). Biotechnological applications of functional metagenomics in the food and pharmaceutical industries. *Front. Microbiol.* 6:672. doi: 10.3389/fmicb.2015.00672

de Oliveira, T. B., and Rodrigues, A. (2019). "Ecology of thermophilic fungi," in *Fungi in extreme environments: Ecological role and biotechnological significance*. eds. S. Tiquia-Arashiro and M. Grube (Cham: Springer), 39–57.

Dimitriu, P. A., Pinkart, H. C., Peyton, B. M., and Mormile, M. R. (2008). Spatial and temporal patterns in the microbial diversity of a meromictic soda Lake in Washington state. *Appl. Environ. Microbiol.* 74, 4877–4888. doi: 10.1128/AEM.00455-08

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

Ersoy Omeroglu, E., Sudagidan, M., Yurt, M. N. Z., Tasbasi, B. B., Acar, E. E., and Ozalp, V. C. (2021). Microbial community of soda Lake Van as obtained from direct and enriched water, sediment and fish samples. *Sci. Rep.* 11:18364. doi: 10.1038/s41598-021-97980-3

Evans, P. N., Boyd, J. A., Leu, A. O., Woodcroft, B. J., Parks, D. H., Hugenholtz, P., et al. (2019). An evolving view of methane metabolism in the Archaea. *Nat. Rev. Microbiol.* 17, 219–232. doi: 10.1038/s41579-018-0136-7

Faris, G. (2017). Threats, use and management interventions for restoration of Lake Chitu west Arsi. *Ethiopia. Am J Biol Environ Stat* 3:11. doi: 10.11648/j.ajbes.20170301.11

Fazi, S., Amalfitano, S., Venturi, S., Pacini, N., Vazquez, E., Olaka, L. A., et al. (2021). High concentrations of dissolved biogenic methane associated with cyanobacterial blooms in east African lake surface water. *Commun Biol* 4:365. doi: 10.1038/s42003-021-02365-x

Felföldi, T. (2020). Microbial communities of soda lakes and pans in the Carpathian Basin: a review. *Biol Futur* 71, 393–404. doi: 10.1007/s42977-020-00034-4

Fotedar, R., Kolecka, A., Boekhout, T., Fell, J. W., Al-Malki, A., Zeyara, A., et al. (2018). Fungal diversity of the hypersaline Inland Sea in Qatar. *Bot. Mar.* 61:048. doi: 10.1515/bot-2018-0048

Gessesse, A., and Gashe, B. A. (1997). Production of alkaline protease by an alkaliphilic bacteria isolated from an alkaline soda lake. *Biotechnol. Lett.* 19:853. doi: 10.1023/A:1018308513853

Grant, W. D., and Jones, B. E. (2016). "Bacteria, archaea and viruses of soda lakes," in *Soda Lakes of East Africa*. ed. M. Schagerl (Cham: Springer).

Grum-Grzhimaylo, A. A., Georgieva, M. L., Bondarenko, S. A. M., Debets, A. J., Bilanenko, E. N., Bondarenko, , et al. (2016). On the diversity of fungi from soda soils. *Fungal Divers.* 76, 27–74. doi: 10.1007/s13225-015-0320-2

Gründger, F., Carrier, V., Svenning, M. M., Panieri, G., Vonnahme, T. R., Klasek, S., et al. (2019). Methane-fuelled biofilms predominantly composed of methanotrophic ANME-1 in Arctic gas hydrate-related sediments. *Sci. Rep.* 9:209. doi: 10.1038/s41598-019-46209-5

Haile, G., and Gessesse, A. (2012). Properties of alkaline protease C45 produced by alkaliphilic bacillus Sp. isolated from Chitu, Ethiopian Soda Lake. *J Biotechnol Biomater* 02:136. doi: 10.4172/2155-952x.1000136

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

Islam, T., Hernández, M., Gessesse, A., Colin Murrell, J., and Øvreås, L. (2021). A novel moderately thermophilic facultative methylotroph within the class alphaproteobacteria. *Microorganisms* 9:477. doi: 10.3390/microorganisms9030477

Jones, B. E., and Grant, W. D. (1999). Microbial diversity and ecology of the Soda Lakes of East Africa, in *Proceedings of the 8th International Symposium on Microbial Ecology*.

Kambura, A. K., Mwirichia, R. K., Kasili, R. W., Karanja, E. N., Makonde, H. M., and Boga, H. I. (2016). Bacteria and Archaea diversity within the hot springs of Lake Magadi and little Magadi in Kenya. *BMC Microbiol.* 16:748. doi: 10.1186/s12866-016-0748-x

Klemperer, S. L., and Cash, M. D. (2007). Temporal geochemical variation in Ethiopian lakes Shala, Arenguade, Awasa, and Beseka: possible environmental impacts from underwater and borehole detonations. *J. African Earth Sci.* 48:006. doi: 10.1016/j.jafrearsci.2006.10.006

Lahti, L., Shetty, S., and Blake, T. (2017). Tools for microbiome analysis in R. *Microbiome Packag Version* 099:88.

Lanzén, A., Simachew, A., Gessesse, A., Chmolowska, D., Jonassen, I., and Øvreås, L. (2013). Surprising prokaryotic and eukaryotic diversity, community structure and biogeography of Ethiopian soda lakes. *PLoS One* 8:577. doi: 10.1371/journal.pone.0072577

Lee, H. J., Kim, S. Y., Kim, P. J., Madsen, E. L., and Jeon, C. O. (2014). Methane emission and dynamics of methanotrophic and methanogenic communities in a flooded rice field ecosystem. *FEMS Microbiol. Ecol.* 88:282. doi: 10.1111/1574-6941.12282

Liu, X., Li, M., Castelle, C. J., Probst, A. J., Zhou, Z., Pan, J., et al. (2018). Insights into the ecology, evolution, and metabolism of the widespread Woesearchaeotal lineages. *Microbiome* 6:488. doi: 10.1186/s40168-018-0488-2

Lozupone, C. A., Hamady, M., Kelley, S. T., and Knight, R. (2007). Quantitative and qualitative diversity measures Lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* 73, 1576–1585. doi: 10.1128/AEM.01996-06

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17:200. doi: 10.14806/ej.17.1.200

Martins, R. F., Davids, W., Abu Al-Soud, W., Levander, F., Rådström, P., and Hatti-Kaul, R. (2001). Starch-hydrolyzing bacteria from Ethiopian soda lakes. *Extremophiles* 5:183. doi: 10.1007/s007920100183

Maukonen, J., Mättö, J., Wirtanen, G., Raaska, L., Mattila-Sandholm, T., and Saarela, M. (2003). Methodologies for the characterization of microbes in industrial environments: a review. *J. Ind. Microbiol. Biotechnol.* 30, 327–356. doi: 10.1007/s10295-003-0056-y

McMurdie, P. J., and Holmes, S. (2013). Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:217. doi: 10.1371/journal.pone.0061217

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

Nilsson, R. H., Larsson, K. H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., et al. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic Acids Res.* 47:1022. doi: 10.1093/nar/gky1022

Ochsenreiter, T., Pfeifer, F., and Schleper, C. (2002). Diversity of Archaea in hypersaline environments characterized by molecular-phylogenetic and cultivation studies. *Extremophiles* 6:253. doi: 10.1007/s00792-001-0253-4

Ogato, T., Kifle, D., and Lemma, B. (2015). Underwater light climate, thermal and chemical characteristics of the tropical soda lake Chitu, Ethiopia: Spatio-temporal variations. *Limnologica* 52:003. doi: 10.1016/j.limno.2015.02.003

Op De Beeck, M., Lievens, B., Busschaert, P., Declerck, S., Vangronsveld, J., and Colpaert, J. V. (2014). Comparison and validation of some ITS primer pairs useful for fungal metabarcoding studies. *PLoS One* 9:629. doi: 10.1371/journal.pone.0097629

Orwa, P., Mugambi, G., Wekesa, V., and Mwirichia, R. (2020). Isolation of haloalkaliphilic fungi from Lake Magadi in Kenya. *Heliyon* 6:823. doi: 10.1016/j.heliyon.2019.e02823

Øvreås, L., Daae, F. L., Torsvik, V., and Rodríguez-Valera, F. (2003). Characterization of microbial diversity in hypersaline environments by melting profiles and reassociation kinetics in combination with terminal restriction fragment length polymorphism (T-RFLP). *Microb. Ecol.* 46, 291–301. doi: 10.1007/s00248-003-3006-3

Poulsen, M., Schwab, C., Borg Jensen, B., Engberg, R. M., Spang, A., Canibe, N., et al. (2013). Methylotrophic methanogenic Thermoplasmata implicated in reduced methane emissions from bovine rumen. *Nat. Commun.* 4:432. doi: 10.1038/ncomms2432

Pujalte, M. J., Lucena, T., Ruvira, M. A., Arahal, D. R., and Macián, M. C. (2014). "The family Rhodobacteraceae," in *The prokaryotes*. eds. E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt and F. Thompson (Berlin, Heidelberg: Springer Berlin Heidelberg), 439–512.

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.* 41:219. doi: 10.1093/nar/gks1219

R Core Team. (2014). *R: A language and environment for statistical computing*. R Found Stat Comput Vienna, Austria.

Rosenthal, K., Oehling, V., Dusny, C., and Schmid, A. (2017). Beyond the bulk: disclosing the life of single microbial cells. *FEMS Microbiol. Rev.* 41, 751–780. doi: 10.1093/femsre/fux044

Schagerl, M. (2016). *Soda lakes of East Africa*. New York: Springer International Publishing.

Shrestha, N., Chilkoor, G., Vemuri, B., Rathinam, N., Sani, R. K., and Gadhamshetty, V. (2018). Extremophiles for microbial-electrochemistry applications: a critical review. *Bioresour. Technol.* 255, 318–330. doi: 10.1016/j.biortech.2018.01.151

Simachew, A., Lanzén, A., Gessesse, A., and Øvreås, L. (2016). Prokaryotic community diversity along an increasing salt gradient in a soda ash concentration pond. *Microb. Ecol.* 71:675. doi: 10.1007/s00248-015-0675-7

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

Sorokin, D. Y., Makarova, K. S., Abbas, B., Ferrer, M., Golyshin, P. N., Galinski, E. A., et al. (2017). Discovery of extremely halophilic, methyl-reducing euryarchaea provides insights into the evolutionary origin of methanogenesis. *Nat. Microbiol.* 2:81. doi: 10.1038/nmicrobiol.2017.81

Tafesse, M. (2014). Metagenomic exploration of Ethiopian soda Lake sediments: microbial diversity and community structure metagenomic exploration of Ethiopian soda Lake. Available at: http://etd.aau.edu.et/handle/123456789/8663.

Tahrioui, A., Schwab, M., Quesada, E., and Llamas, I. (2013). Quorum sensing in some representative species of Halomonadaceae. *Life* 3, 260–275. doi: 10.3390/life3010260

Verma, D., and Satyanarayana, T. (2011). An improved protocol for DNA extraction from alkaline soil and sediment samples for constructing metagenomic libraries. *Appl. Biochem. Biotechnol.* 165:454. doi: 10.1007/s12010-011-9264-5

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., et al. (2016). Improved bacterial 16S rRNA gene (V4 and V4-5) and fungal internal transcribed spacer marker gene primers for microbial community surveys. *mSystems* 1:009. doi: 10.1128/mSystems.00009-15

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5:237. doi: 10.1186/s40168-017-0237-y

Wobus, A., Bleul, C., Maassen, S., Scheerer, C., Schuppler, M., Jacobs, E., et al. (2003). Microbial diversity and functional characterization of sediments from reservoirs of different trophic state. *FEMS Microbiol. Ecol.* 46, 331–347. doi: 10.1016/S0168-6496(03)00249-6

Wood, R. B., Baxter, R. M., and Prosser, M. V. (1984). Seasonal and comparative aspects of chemical stratification in some tropical crater lakes, Ethiopia. *Freshw Biol* 14:176. doi: 10.1111/j.1365-2427.1984.tb00176.x

Xiao, J., Zhang, Y. Y. Y., Chen, W., Xu, Y., Zhao, R., Tao, L., et al. (2020). Diversity and biogeography of Woesearchaeota: a comprehensive analysis of multi-environment data. *bioRxiv*. 1:16. doi: 10.1101/2020.08.09.243345

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., et al. (2012). Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ. Microbiol.* 14, 2457–2466. doi: 10.1111/j.1462-2920.2012.02799.x

Yang, J., Ma, L., Jiang, H., Wu, G., and Dong, H. (2016). Salinity shapes microbial diversity and community structure in surface sediments of the Qinghai-Tibetan Lakes. *Sci. Rep.* 6:078. doi: 10.1038/srep25078

Ye, J.-W., and Chen, G.-Q. (2021). Halomonas as a chassis. *Essays Biochem.* 65, 393–403. doi: 10.1042/EBC20200159

Zhang, B., Li, Y., Xiang, S. Z., Yan, Y., Yang, R., Lin, M. P., et al. (2020). Sediment microbial communities and their potential role as environmental pollution indicators in Xuande atoll, South China Sea. *Front Microbiol* 11:011. doi: 10.3389/fmicb.2020.01011

Zhu, D., Adebisi, W. A., Ahmad, F., Sethupathy, S., Danso, B., and Sun, J. (2020). Recent development of extremophilic bacteria and their application in biorefinery. *Front. Bioeng. Biotechnol.* 8:483. doi: 10.3389/fbioe.2020.00483

Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., et al. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS One* 6, –e24570. doi: 10.1371/journal.pone.0024570

Zorz, J. K., Sharp, C., Kleiner, M., Gordon, P. M. K. K., Pon, R. T., Dong, X., et al. (2019). A shared core microbiome in soda lakes separated by large distances. *Nat. Commun.* 10:195. doi: 10.1038/s41467-019-12195-5

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 January 2023
DOI 10.3389/fmicb.2023.1109549

OPEN ACCESS

EDITED BY
Jesse G. Dillon,
California State University, Long Beach,
United States

REVIEWED BY
Aharon Oren,
Hebrew University of Jerusalem, Israel
Ryosuke Nakai,
National Institute of Advanced Industrial
Science and Technology (AIST), Japan

*CORRESPONDENCE
Cristina Sánchez-Porro
✉ sanpor@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 27 November 2022
ACCEPTED 03 January 2023
PUBLISHED 20 January 2023

CITATION
García-Roldán A, Durán-Viseras A, de la Haba RR, Corral P, Sánchez-Porro C and Ventosa A (2023) Genomic-based phylogenetic and metabolic analyses of the genus *Natronomonas*, and description of *Natronomonas aquatica* sp. nov.. *Front. Microbiol.* 14:1109549. doi: 10.3389/fmicb.2023.1109549

Genomic-based phylogenetic and metabolic analyses of the genus *Natronomonas*, and description of *Natronomonas aquatica* sp. nov.

Alicia García-Roldán[1], Ana Durán-Viseras[1], Rafael R. de la Haba[1], Paulina Corral[2], Cristina Sánchez-Porro[1]*and Antonio Ventosa[1]*

[1]Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [2]Department of Biology, University of Naples Federico II, Naples, Italy

The genus *Natronomonas* is classified on the family *Haloarculaceae*, within the class *Halobacteria* and currently includes six species isolated from salterns, saline or soda lakes, and salt mines. All are extremely halophilic (optimal growth at 20–25% [w/v] NaCl) and neutrophilic, except *Natronomonas pharaonis*, the type species of the genus, that is haloalkaliphilic (showing optimal growth at pH 9.0) and possesses distinct phenotypic features, such as a different polar lipid profile than the rest of species of the genus. We have carried out a genome-based study in order to determine the phylogenetic structure of the genus *Natronomonas* and elucidate its current taxonomic status. Overall genomic relatedness indexes, i.e., OrthoANI (Average Nucleotide Identity), dDDH (digital DNA–DNA hybridization), and AAI (Average Amino acid Identity), were determined with respect to the species of *Natronomonas* and other representative taxa of the class *Halobacteria*. Our data show that the six species of *Natronomonas* constitute a coherent cluster at the genus level. Besides, we have characterized a new haloarchaeon, strain F2-12^T, isolated from the brine of a pond of a saltern in Isla Cristina, Huelva, Spain, and we determined that it constitutes a new species of *Natronomonas*, for which we propose the name *Natronomonas aquatica* sp. nov. Besides, the metabolic analysis revealed a heterotrophic lifestyle and a versatile nitrogen metabolism for members of this genus. Finally, metagenomic fragment recruitments from a subset of hypersaline habitats, indicated that the species of *Natronomonas* are widely distributed in saline lakes and salterns as well as on saline soils. Species of this haloarchaeal genus can be considered as ubiquitous in intermediate to high salinity habitats.

KEYWORDS

***Natronomonas*, hypersaline environments, salterns, extremophiles, haloarchaea, metabolism, phylogenomics, polar lipids**

1. Introduction

Haloarchaea are extremely halophilic archaea well adapted to hypersaline environments such as saline lakes, salterns, saline soils, or salted products (Ventosa, 2006; Oren, 2011). They are "*Euryarchaeota*" included into the class *Halobacteria*, with three orders: *Halobacteriales, Haloferacales*, and *Natrialbales,* that include three (*Halobacteriaceae*, *Halococcaceae*, and *Haloarculaceae*), two (*Haloferacaceae* and *Halorubraceae*), and one (*Natrialbaceae*) families, respectively, and more than 70 genera (Amoozegar et al., 2017; Oren et al., 2017; Parte et al., 2020).

The genus *Natronomonas* belongs to the family *Haloarculaceae*, and currently it includes six species: *Natronomonas pharaonis* (type species; Soliman and Trüper, 1982; Kamekura et al., 1997),

Natronomonas moolapensis (Burns et al., 2009), *Natronomonas gomsonensis* (Kim et al., 2013), *Natronomonas halophila* (Yin et al., 2020), *Natronomonas salina* (Yin et al., 2020), and *Natronomonas salsuginis* (Durán-Viseras et al., 2020). The most peculiar feature of the genus *Natronomonas* is that it comprises the first alkaliphilic and extremely halophilic archaeon so far described (*Nmn. pharaonis*), while the rest of the species of this genus are non-alkaliphilic. Although only six species have been described to date, several studies have shown that they are found in a variety of environments, such as hypersaline and alkaline lakes or soils, salterns or sediments, supporting the widespread distribution of this genus (Oren and Ventosa, 2017; Minegishi and Kamekura, 2018).

The members of the genus *Natronomonas* are characterized by a diverse cell morphology: short rods or pleomorphic shapes (coccoid, flat, and tetragonal shapes). They are motile and Gram-stain-negative. In solid culture media they produce red pigmented colonies. They are aerobic microorganisms and have a chemoorganotrophic metabolism. They are obligately halophilic (requiring at least 15% [w/v] NaCl for growth and growing optimally at 20–25% [w/v] NaCl). Alkaliphilic species grow at the pH range 8.0 to 11.0 (optimal growth at pH 8.5–9.0), but neutrophilic species grow at pH 5.5–8.5, and optimally at pH 7.0–7.5 (Minegishi and Kamekura, 2018). Membrane polar lipids have been shown to be a relevant feature for the taxonomic characterization of haloarchaea (Torreblanca et al., 1986; Oren et al., 2009; de la Haba et al., 2018). All members of the genus *Natronomonas* have phosphatidylglycerol and phosphatidylglycerol phosphate methyl ester. However, unidentified phospholipids or glycolipids may exist. The polar lipid profile of the haloalkaliphilic species *Nmn. pharaonis* is different from those of the neutrophilic species, without membrane glycolipids, and thus, it has been suggested that this species could be placed into a separated genus from the neutrophilic species of this genus (Minegishi and Kamekura, 2018).

Our studies on the prokaryotic diversity of several hypersaline environments of South-West Spain permitted us to isolate a new strain, designated as F2-12^T, from the brine of a pond of one of those salterns. The aim of this paper is firstly to elucidate the phylogenomic and metabolic structure of the genus *Natronomonas*, based on current genomic-based analysis as well as phenotypic features, to clarify if *Nmn. pharaonis* and the remaining neutrophilic species of the genus should be separated into two different genera. Besides, we describe the taxonomic features of the new archaeon F2-12^T which we propose as a new species of the genus *Natronomonas*, *Natronomonas aquatica* sp. nov. Finally, we have carried out an in-depth analysis, based on metagenomics fragment recruitments, to show the ecological distribution and abundance of species of the genus *Natronomonas* in different hypersaline habitats.

2. Materials and methods

2.1. Isolation and culture of the new haloarchaeal strain

Strain F2-12^T was isolated from the brine of a crystallizer pond of a marine saltern located in Isla Cristina, Huelva, South-West Spain by Durán-Viseras (2020). It was isolated in pure culture on R2A medium supplemented with 25% (w/v) seawater salt solution. The composition of this salt mixture is (g l^{-1}): NaCl, 195; $MgCl_2{\cdot}6H_2O$, 32.5; $MgSO_4{\cdot}7H_2O$, 50.8; $CaCl_2$, 0.83; KCl, 5.0; $NaHCO_3$, 0.21; NaBr, 0.58 (Durán-Viseras, 2020). The pH of the medium was adjusted to 7.5 with 1 M KOH. For the preparation of the solid media, purified agar was added at 1.8% (w/v). The culture was routinely incubated aerobically at 37°C for 14 days, both in liquid and solid medium. For comparative purposes, the type strains of species of the genus *Natronomonas*, *Natronomonas pharaonis* DSM 2160^T, *Natronomonas gomsonensis* JCM 17867^T, *Natronomonas moolapensis* CECT 7526^T, and *Natronomonas salsuginis* F20-122^T were used as culture collection reference strains.

2.2. DNA extraction, purification, and sequencing

The DNA extraction and purification were carried out by the method described by Marmur (1961). DNA quantification and purity were determined by fluorometry (Qubit 3.0 Fluorometer, Thermofisher Scientific, USA) and spectrophotometry (NanoDrop One, Thermofisher Scientific, USA), respectively. The 16S rRNA and *rpoB'* genes were amplified by PCR using the primers ArchF (TTC CGG TTG ATC CTG CCG GA) and ArchR (GGT TAC CTT GTT ACG ACT T) as well as rpoBF (TGT AAA ACG ACG GCC AGT TCG AAG AGC CGG ACG ACA TGG) and rpoBR (CAG GAA ACA GCT ATG ACC GGT CAG CAC CTG BAC CGG NCC), respectively (Durán-Viseras, 2020). The PCR products were sequenced using the Sanger method by StabVida (Caparica, Portugal), and primers 16RB36 (GGA CTA CCA GGG TAT CTA) and 16RD34 (GGT CTC GCT CGT TGC CTG) were also used in order to obtain the complete sequence of 16S rRNA gene. The draft genome sequence of strain F2-12^T was also determined in this study by Illumina NovaSeq 2×150 PE (Novogene Co., United Kingdom).

2.3. Phylogenetic analyses

The 16S rRNA and *rpoB'* gene sequences of the new isolate, strain F2-12^T, were analyzed, corrected, and assembled with ChromasPro version 1.5 (Technelysium Pty Ltd.). The 16S rRNA gene sequence was compared with EzBioCloud database.[1] Using ARB software package (Ludwig et al., 2004) and LTPs_106_SSU database, alignments and phylogenetic trees were generated by three different methods: maximum-parsimony (Fitch, 1971), neighbor-joining (Saitou and Nei, 1987) and maximum-likelihood (Felsenstein, 1981) algorithms. A bootstrap analysis (1,000 pseudo-replicates) was performed to evaluate the robustness of the phylogenetic trees (Felsenstein, 1985). The *rpoB'* gene sequence was compared with other sequences available in NCBI database by BLAST search (http://www.ncbi.nlm.nih.gov/BLAST/; Altschul et al., 1990). The alignment and phylogenetic trees were obtained with BioEdit 3.3.19.0 version (Hall, 1999) and MEGA 6.11 software (Tamura et al., 2013),[2] respectively.

The 16S rRNA and *rpoB'* gene sequences of strain F2-12^T were deposited in GenBank/EMBL/DDBJ, under the accession numbers MZ318646 and MZ327708, respectively.

1 https://www.ezbiocloud.net/

2 https://doi.org/10.1093/molbev/msab120

2.4. Genome assembly, annotation, and determination of genomic parameters

The raw reads resulting from genome of strain F2-12^T were quality checked with FastQC 0.11.9 version (Andrews, 2010) and further trimmed (settings: qtrim = rl trimq = 18) and screened for adaptor and vector contamination (settings: k = 21 tbo ordered cardinality) using BBDuk from BBTools suite (Bushnell, 2021). Filtered reads were assembled with Spades v.3.12.0 (settings: --isolate -k 21, 33, 55, 77, 99, 127; Bankevich et al., 2012) and the quality of the assembly was quantified with QUAST v.2.3 (Gurevich et al., 2013). Finally, genome completeness and contamination were estimated with CheckM v.1.0.5 software (settings: lineage_wf --reduced_tree; Parks et al., 2015).

The genome sequence of strain F2-12^T was initially annotated using Prokka (default settings) and further following the NCBI Prokaryotic Genome Annotation Pipeline (PGAP; Tatusova et al., 2016) and was deposited in GenBank/EMBL/DDBJ, under the accession number GCA_024449025.1.

Several genomic indexes were used for circumscribing species and genera. The Orthologous Average Nucleotide Identity (OrthoANI) was determined with OAT v.0.93.1 (default settings; Lee et al., 2016), the digital DNA–DNA hybridization (dDDH) was calculated by the Genome-to-Genome Distance Calculator (GGDC) v.2.1 (Meier-Kolthoff et al., 2013) website, using BLAST+ as the local alignment tool and the formula 2 for distance calculations (Auch et al., 2019) and the Average Amino acid Identity (AAI) was calculated by CompareM (default settings; https://github.com/dparks1134/CompareM).

2.5. Phylogenomic comparative analysis

The phylogenomic comparative analysis was carried out using the genomes of the type strains of species of *Natronomonas*, as well as that of strain F2-12^T and other genomes of related haloarchaeal species available from GenBank database. The features and the accession numbers of the genomes of the type strains of the species of the genus *Natronomonas* used in this study are shown in Table 1. The quality of the genome sequences used in this study was in agreement with the recommended minimal standards for the use of genome data for the taxonomy of prokaryotes (Chun et al., 2018).

To determine the core-genome, the Enveomics tool (Rodríguez-R and Konstantinidis, 2016) was used. To identify clusters of orthologous genes (OGs), we carried out an all-versus-all BLASTp search based on the translated protein-coding gene sequences of the type strains of species of *Natronomonas* and of strain F2-12^T as well as those of other genera of the family *Haloarculaceae* available in databases. The OGs shared among all taxa and present in a single copy per genome were selected for further analysis. They were aligned with MUSCLE v. 3.8.31 (Edgar, 2002) and subsequently concatenated. An approximately maximum-likelihood phylogenomic tree was obtained using FastTree v. 2.1.9 (Price et al., 2010) with the JTT replacement matrix (Jones et al., 1992) under the CAT approximation (single rate for each site) with 20 rate categories. Local support values were determined with the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999). An alternative phylogenomic tree based on selected 53 markers as recommended by the Genome Taxonomy Database (GTDB) was constructed using GTDB-Tk v. 2.1.0+ (Chaumeil et al., 2022) using the reference database R207_v2.

2.6. Phenotypic characterization

Phenotypic tests were determined for all species of the genus *Natronomonas* and they were carried out following the proposed minimal standards for the description of novel taxa in the class *Halobacteria* (Oren et al., 1997). Morphology and motility were studied by light optical under a phase-contrast microscope (Olympus BX41), in cultures incubated for 14 days at 37°C. The range and optimal growth conditions of salinity, pH, and temperature were determined on liquid medium R2A 25%, with the same composition but only modifying each of these factors. The medium R2A was supplemented with 0.9, 3, 5, 10, 15, 20, 25, and 30% (w/v) seawater salt solution (see section 2.1) to determine the growth under different salinity conditions. The growth at different pH values was determined in this medium supplemented with suitable buffers (MOPS, pH 6.0–7.0; Tris, pH 7.5–8.5; CHES, pH 9.0–10.0) to maintain stable the pH values to 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, and 10.0. The growth at 4, 15, 20, 25, 30, 37, 40, 45, 50, and 55°C permitted to determine the range and optimal growth at different temperatures.

Anaerobic growth was determined using R2A 25% plates supplemented with 1% L-arginine, 1% dimethyl sulfoxide (DMSO), and 1% KNO_3, respectively, that were inoculated and incubated at 37°C in an anaerobic chamber (Oxoid) for 21 days. All biochemical tests were carried out in R2A 25% medium, pH 7.5 and 37°C. Catalase activity was observed by adding 3% (w/v) H_2O_2 to a colony of strain F2-12^T (Cowan and Steel, 1993). Oxidase activity was determined by adding 1% (w/v) tetramethyl-p-phenylenediamine (Kovács, 1956). Hydrolysis of starch, gelatin, aesculin or Tween 80 were determined as previously described (Durán-Viseras et al., 2020). Indole production was tested as described by Kovács (1928). Methyl red and Voges-Proskauer tests were performed by two methods (Werkman, 1930; Cowan and Steel, 1993). Nitrate and nitrite reduction was determined by using sulfanilic acid and α-naphthylamine (Skerman, 1967). The formation of H_2S was determined as described by Ventosa et al. (1982). Acid production from carbohydrates was determined by adding a solution of the carbohydrate (1%, w/v, final carbohydrate concentration) to the basal medium with phenol red and supplemented with 0.05% (w/v) of yeast extract. Finally, the utilization of different compounds as sole carbon and energy source was determined by adding a filtered-sterilized solution of 1% (w/v) of different carbohydrates, alcohols, organic acids or amino acids to the medium SW25 supplemented with 0.05% (w/v) of yeast extract (Ventosa et al., 1982).

2.7. Chemotaxonomic analysis

The polar lipids of *Natronomonas moolapensis* CECT 7526^T, *Natronomonas salsuginis* F20-122^T, *Natronomonas gomsonensis* JCM 17867^T and strain F2-12^T were determined from cells cultured in medium R2A 25% at 37°C for 14 days, and the biomass from *Natronomonas pharaonis* CECT 4578^T was obtained in the alkaline medium (Soliman and Trüper, 1982) at 37°C for 14 days. The species *Halobacterium salinarum* DSM 3754^T and *Halorubrum saccharovorum* DSM 1137^T were used as reference for the determination of the polar lipids profiles. The total polar lipids were extracted as follows: the cell biomass was washed by adding 25% (w/v) NaCl sterile solution and centrifuged for 1 min at 6,000 *g*; the pellet was then resuspended in 0.8 ml of 25% (w/v) NaCl solution until obtaining a cell suspension. To 0.8 ml of cell suspension, 2 ml of methanol and 1 ml of chloroform were added to create a monophasic mixture. After gently mixing by inversion for 1 h and later centrifuged for 1 min at 6,000 *g*, the supernatant was collected from the colorless pellet. The supernatant

TABLE 1 General features of the genome sequences of *Natronomonas aquatica* F2-12^T and the type strains of the species of the genus *Natronomonas* used in this study.

Feature	1	2	3	4	5	6	7
Size (Mb)	3.21	2.90	2.91	2.75	3.21	3.75	3.38
Contigs	82	12	1	3	1	1	2
Genome coverage	508.3x	352.1x	60.1x	5.8x	100.0x	100.0x	100.0x
G + C (mol%)	62.7	63.2	64.5	63.1	64.4	67.5	64
N50 (bp)	84,628	675,769	2,912,573	2,595,221	3,211,682	3,746,575	1,800,000
Total genes	3,333	3,010	2,922	2,864	3,413	3,921	3,612
Protein coding genes	3,189	2,863	2,806	2,785	3,347	3,809	3,435
rRNA	3	3	3	3	3	3	3
tRNA	47	45	45	46	45	45	73
Assembly accession number	GCA_024449025.1	GCA_005239135.1	GCA_000591055.1	GCA_000026045.1	GCA_013391085.1	GCA_013391105.1	GCA_013391635.1

1, *Natronomonas aquatica* F2-12^T, **2**, *Natronomonas salsuginis* F20-122^T, **3**, *Natronomonas moolapensis* 8.8.11^T, **4**, *Natronomonas pharaonis* DSM 2160^T, **5**, *Natronomonas halophila* C90^T, **6**, *Natronomonas salina* YPL13^T, **7**, *Natronomonas gomsonensis* JCM 17867^T.

in monophase was disrupted adding 500 µl of KCl 0.2 M and 1 ml of chloroform, followed by centrifugation during 1 min at 6,000 *g*. A bilayer phase was formed and the lower pigmented phase corresponding to the chloroform fraction was recovered and reduced to ~500 µl by evaporation under a fume hood or vacuum concentrator. The extract was transferred to a weighted empty glass vial of 2 ml, dried, weighted and stored at −20°C. To perform the chromatography, the total lipid extract was dissolved in a final concentration of 100 mg ml^{-1}. A volume of 10 µl of total lipid extract (100 mg ml^{-1}) were analyzed by high-performance thin layer chromatography (HPTLC), using HPTLC silica gel 60 plates crystal back (10 × 20 cm; Merck art. 5626); the plates were developed in the solvent system chloroform/methanol/90% (v/v) acetic acid (65:4:35) as previously described (Angelini et al., 2012). To detect all polar lipids, the plate was sprayed with 5% (v/v) sulfuric acid in water and charred by heating at 160°C.

2.8. Metagenomic fragment recruitment analyses

In order to determine the presence in different saline habitats of the six species of the genus *Natronomonas* and strain F2-12^T, fragment recruitments with different environmental metagenomic datasets were performed. The genome contigs were concatenated and then all the rRNA gene sequences obtained were masked. BLASTn search (with the cut-offs: alignment length ≥30 nt, e-value ≤1 × 10^{-5}, identity >95%) was used to align the metagenomic quality-filtered shotgun reads against the concatenated contigs of the type strains of all six species of the genus *Natronomonas*, as well as those of strain F2-12^T. Recruitment plot representations were performed in R using the library "Hmisc."

3. Results and discussion

3.1. The genus *Natronomonas* is a phylogenomically coherent taxonomic group

Previous studies have suggested that the type species of the genus *Natronomonas*, *Nmn. pharaonis*, possesses, in fact, substantial differences as to be considered as a member of a separated genus from the remaining five species of *Natronomonas* (Minegishi and Kamekura, 2018). Those phenotypic differences are, essentially, the alkaliphilic behavior of *Nmn. pharaonis* versus the neutrophilic nature of the other five species, and the lack of glycolipids in the polar lipid profile of the haloalkaliphilic species *Nmn. pharaonis*. In order to shed light on this matter, a genome-based phylogeny was performed including all the validly described species names of the genus *Natronomonas* and closest relatives. This phylogenomic tree was obtained after the alignment and concatenation of the translated sequences of 870 core, orthologous, single-copy genes from the genomes under study (Figure 1). As it can be observed, all the species of the genus *Natronomonas*, including the conflicting species *Nmn. pharaonis*, formed a robust branch (100% bootstraps) that demonstrates that they constitute a monophyletic group of species. This study was complemented with the phylogenomic tree based on the comparison of the 53 concatenated conserved single-copy proteins recommended by the GTDB (Supplementary Figure S1), which also showed the monophyletic topology of the species of the genus *Natronomonas*, including *Nmn. pharaonis* and the new strain F2-12^T.

On the other hand, the percentages of AAI among all species of *Natronomonas* were 68.7–75.3%, while the values between the species of *Natronomonas* and the other species of related haloarchaeal genera were equal to or lower than 61.4%. These percentages show unequivocally that all species of *Natronomonas* constitute a coherent genus, clearly separated from the other related genera (Figure 2). Thus, even considering that the species *Nmn. pharaonis* shows some differential phenotypic data, such as the optimal pH and range supporting growth or a different polar lipid profile than the other species of *Natronomonas*, there is no reason to taxonomically separate it from the remaining species of the genus *Natronomonas*.

3.2. A putative new species of the genus *Natronomonas*

During our studies on the culturable diversity of a Isla Cristina saltern, located in Huelva, Spain, the strain F2-12^T was isolated (Durán-Viseras, 2020). The almost-complete 16S rRNA gene sequence of this strain was amplified, sequenced, and analyzed (1,400 bp). The results

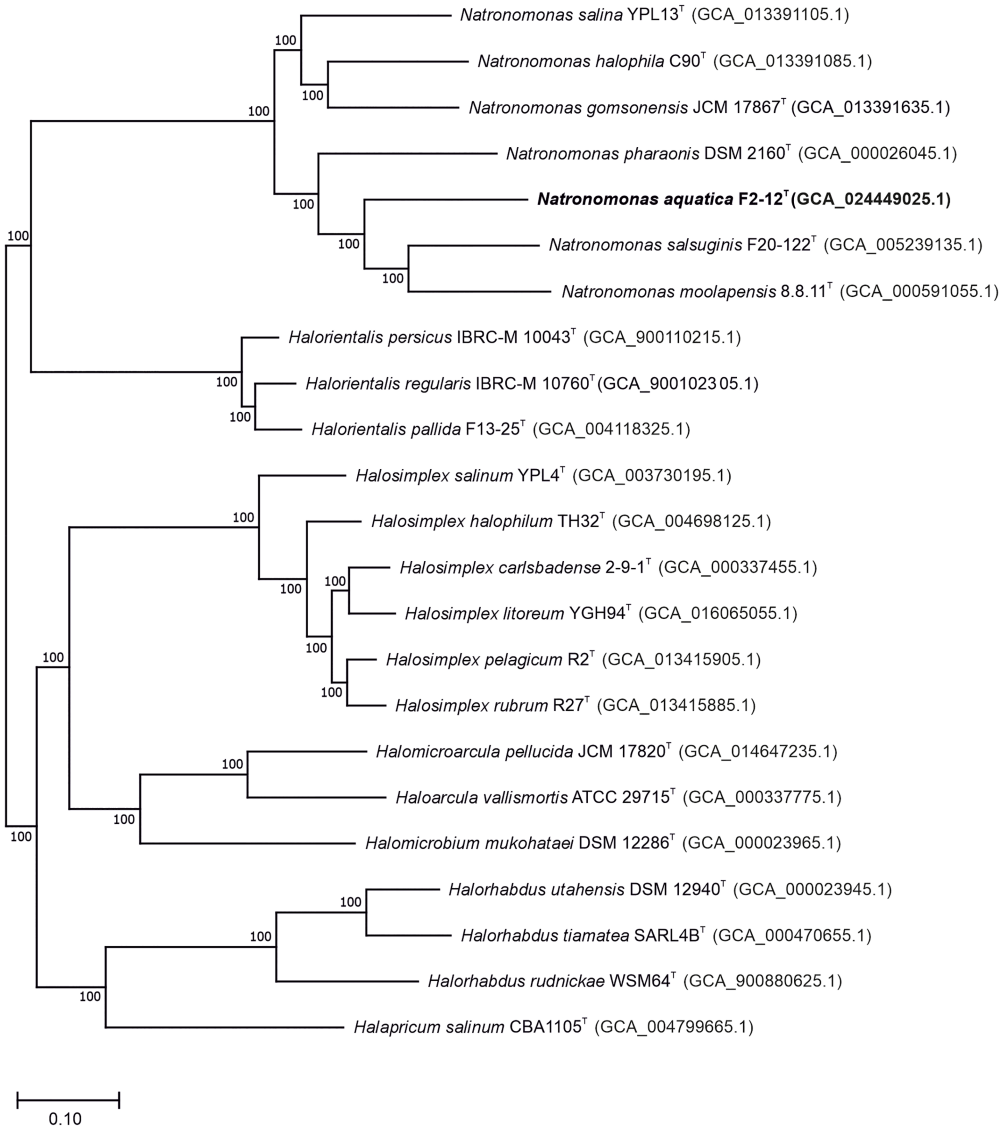


FIGURE 1
Approximately maximum-likelihood phylogenomic tree reconstruction based on the translated core orthologous genes of members of the genus *Natronomonas*, including strain F2-12^T and related species. This tree was obtained after the alignment and concatenation of the translated sequences of 870 shared orthologous single-copy genes of these genomes. Bootstrap values higher than 70% are indicated at branch points. Bar, 0.1 substitutions per nucleotide position.

generated by the EzBioCloud tool indicated that strain F2-12^T is a member of *Natronomonas,* showing the highest percentages of identity with the type strains of the species *Natronomonas moolapensis* 8.8.11^T (98.0%), *Natronomonas salsuginis* F20-122^T (97.3%), *Natronomonas pharaonis* DSM 2160^T (96.8%), *Natronomonas salina* YPL13^T (96.8%), *Natronomonas halophila* C90^T (96.7%) and *Natronomonas gomsonensis* SA3^T (95.8%), while the species of other related genera, such as *Halocatena* or *Salinirubellus,* showed percentages of similarity lower than 93.8%. The percentages of similarity between strain F2-12^T and all species of *Natronomonas* are lower than 98.7%, considered the cutoff delineation of prokaryotic species (Konstantinidis et al., 2017).

The 16S rRNA phylogenetic tree, constructed by the maximum-parsimony algorithm (Figure 3), showed that strain F2-12^T clustered with the species of the genus *Natronomonas,* but it constituted an independent branch. The phylogenetic position of strain F2-12^T was also confirmed in trees obtained using the maximum-likelihood and neighbor-joining algorithms. These results suggest that strain F2-12^T could be a new member of the genus *Natronomonas.*

Several copies of the 16S rRNA gene with divergent sequences have been reported in species of haloarchaea, such as *Halosimplex*, *Haloarcula, Halomicrobium*, or *Halorientalis* (Vreeland et al., 2002; Boucher et al., 2004; Cui et al., 2009; Sun et al., 2013; Durán-Viseras et al., 2019). In order to avoid the limitations caused by the intragenomic heterogeneity among haloarchaeal 16S rRNA genes, it has been claimed the use of the *rpoB'* gene as an alternative phylogenetic marker in haloarchaea (Walsh et al., 2004; Enache et al., 2007; Minegishi et al., 2010). Thus, we also determined the partial *rpoB'* sequence (527 bp) of strain F2-12^T and used it for comparative studies between this strain and the already described species of *Natronomonas*. The analysis showed that strain F2-12^T was also related to species of the genus *Natronomonas*, showing percentages of similarity between 90.5 and 87.1% for the species of *Natronomonas* and lower percentages regarding to other haloarchaeal genera.

AAI

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	100.0													
2	70.4	100.0												
3	68.7	69.9	100.0											
4	69.3	70.8	74.1	100.0										
5	72.8	70.7	69.9	69.4	100.0									
6	68.8	70.6	73.5	73.2	69.2	100.0								
7	72.5	70.9	70.0	69.9	75.3	69.4	100.0							
8	59.1	59.7	59.7	60.0	60.1	59.7	59.8	100.0						
9	59.5	59.8	60.1	60.0	59.8	60.0	59.9	62.1	100.0					
10	59.4	60.2	60.6	60.3	60.2	60.5	60.0	63.0	73.9	100.0				
11	59.6	60.3	60.6	60.5	60.2	60.8	60.2	62.9	66.9	67.1	100.0			
12	57.7	58.7	58.4	58.5	58.4	58.8	58.2	63.1	60.4	60.8	61.1	100.0		
13	60.5	61.0	61.1	61.4	60.9	61.4	60.9	62.6	62.3	62.6	62.6	60.5	100.0	
14	58.9	59.5	59.5	59.7	59.4	59.7	59.4	62.0	62.5	63.1	63.3	60.9	62.1	100.0

FIGURE 2
Average Amino acid Identity (AAI) percentages among the species of the genus *Natronomonas*, including strain F2-12^T, and other related haloarchaeal genera. Strains: 1, *Natronomonas aquatica* F2-12^T (GCA_024449025.1); 2, *Natronomonas pharaonis* DSM 2160^T (GCA_000026045.1); 3, *Natronomonas gomsonensis* JCM 17867^T (GCA_013391635.1); 4, *Natronomonas halophila* C90^T (GCA_013391085.1); 5, *Natronomonas moolapensis* 8.8.11^T (GCA_000591055.1); 6, *Natronomonas salina* YPL13^T (GCA_013391105.1); 7; *Natronomonas salsuginis* F20-122^T (GCA_005239135.1); 8, *Halapricum salinum* CBA1105^T (GCA_004799665.1); 9, *Haloarcula vallismortis* ATCC 29715^T (GCA_000337775.1); 10, *Halomicroarcula pellucida* JCM 17820^T (GCA_014647235.1); 11, *Halomicrobium mukohataei* DSM 12286^T (GCA_000023965.1); 12, *Halorhabdus utahensis* DSM 12940^T (GCA_000023945.1); 13, *Halorientalis regularis* IBRC-M 10760^T (GCA_9001023 05.1); 14, *Halosimplex carlsbadense* 2-9-1^T (GCA_000337455.1).

A maximum-parsimony *rpoB*'-based phylogenetic tree was also obtained (Figure 4). This tree shows that strain F2-12^T belongs to the genus *Natronomonas*, but it is far enough away from species of this genus, and also supports the possibility that it could constitute a new species. On the other hand, both 16S rRNA and *rpoB*' gene-based phylogenetic trees show that all species of *Natronomonas* cluster together and formed a separate phylogenetic group with respect to species of the related haloarchaeal genera investigated, also supporting our aforementioned conclusion that no arrangements are necessary in the genus *Natronomonas*.

3.3. Genome-based analyses confirm the new species of *Natronomonas*

The draft genome sequence of strain F2-12^T was successfully assembled in 82 contigs, with a sequencing depth of 508.3x of the entire genome. The genome size of this strain was 3,214,353 bp and the DNA G + C content was 62.7 mol%. The range of G + C content for species of *Natronomonas* is from 63.1 to 67.5 mol%, with an intermediate value of 63.1 mol% for the type species of the genus, *Nmn. pharaonis*. The genome size ranged from 2.75 to 3.75 Mb, which are within those described for species of haloarchaea. Additional genomic characteristics are shown in Table 1. The comparison of this genome and those of the type strains of the species of *Natronomonas* and other closely related haloarchaea permitted us the reconstruction of the phylogenomic core-genome tree (Figure 1), as stated before. Strain F2-12^T belongs to the same clade of *Natronomonas moolapensis* 8.8.11^T, *Natronomonas salsuginis* F20-122^T, *Natronomonas pharaonis* DSM 2160^T, *Natronomonas salina* YPL13^T, *Natronomonas gomsonensis* JCM 17867^T and *Natronomonas halophila* C90^T, but clustered in a different branch, separated from the rest of species of *Natronomonas*, with a bootstrap percentage of 100%, reinforcing the hypothesis of being a new species of the genus *Natronomonas*.

The previous findings were confirmed by calculating the overall genome relatedness indexes, namely OrthoANI (Average Nucleotide Identity), dDDH (digital DNA–DNA hybridization), and AAI (Average Amino acid Identity), which are currently used as genomic thresholds for delineation of new prokaryotic taxa (Meier-Kolthoff et al., 2013; Lee et al., 2016; Auch et al., 2019).The percentages of OrthoANI between the strain F2-12^T and the species of the genus *Natronomonas* ranged from 76.2 to 79.6% (Figure 5). These values are lower than the threshold percentage currently accepted for species delineation (95%; Konstantinidis et al., 2017), confirming that strain F2-12^T should be considered as a different species of the genus *Natronomonas*. On the other hand, percentages of dDDH higher or equal to 70% indicate that the strains can constitute the same species, while values lower than 70% show that strains belong to different prokaryotic species (Konstantinidis and Tiedje, 2004; Goris et al., 2007; Kim et al., 2014). Figure 5 shows that the percentages of dDDH were equal or lower than 23.5% between strain F2-12^T and the species of the genus *Natronomonas*, supporting that this strain constitutes a separate species of this genus. Finally, AAI was calculated to confirm that strain F2-12^T is well assigned to the genus *Natronomonas*. The AAI percentages of strain F2-12^T with respect to the species of *Natronomonas* were in the range of 68.8 to 72.8, and 60.5% or lower with respect to species of other related haloarchaeal genera (Figure 2). Since the AAI threshold established for species assigned to the same genus is 65% (Goris et al., 2007;

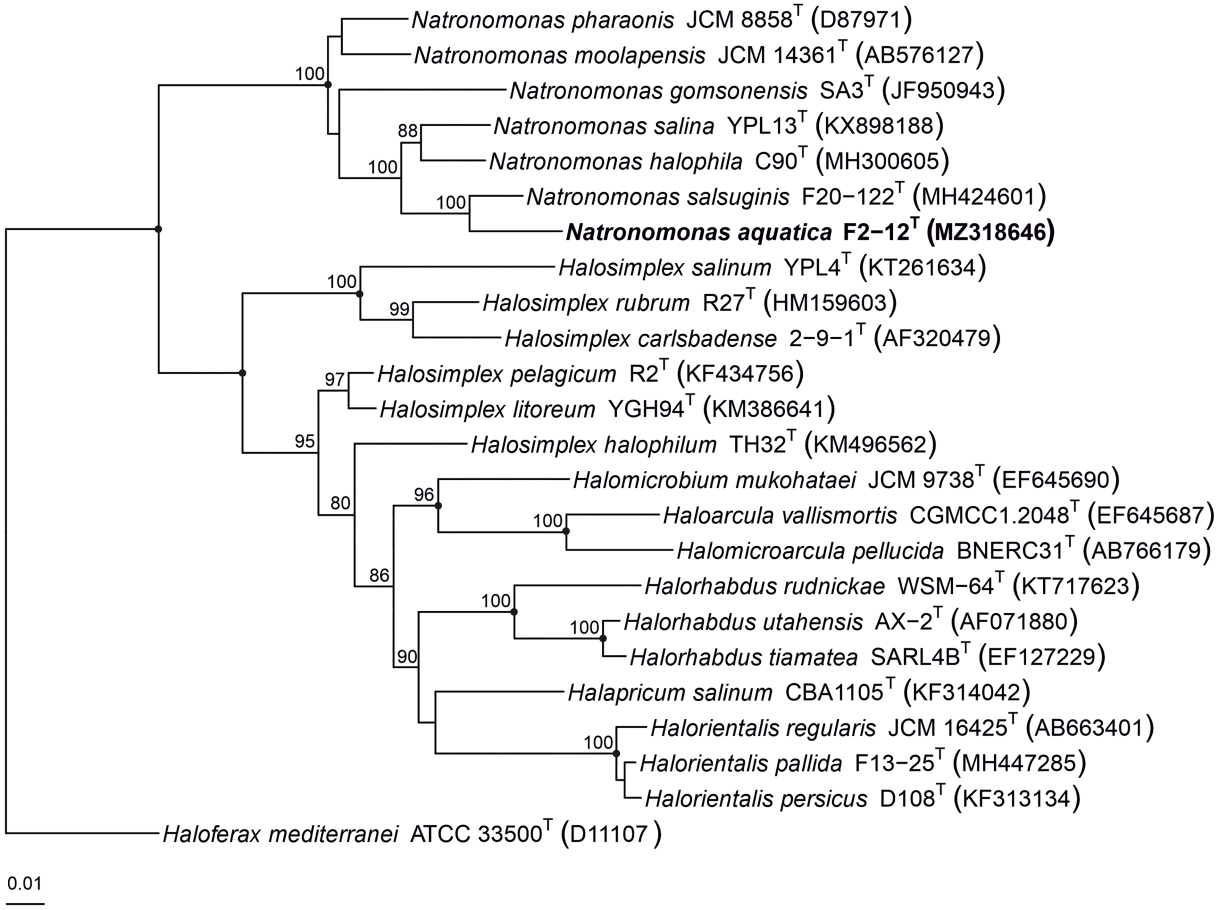


FIGURE 3
Maximum-parsimony phylogenetic tree based on the comparison of the 16S rRNA gene sequences showing the relationship between strain F2-12^T, the species of the genus *Natronomonas* and other related haloarchaeal genera. The sequence accession numbers are shown in parentheses. Bootstrap values equal or higher than 70% are indicated above the nodes. Black circles indicate that the corresponding nodes were also obtained in the trees generated with the maximum-likelihood and neighbor-joining algorithms. *Haloferax mediterranei* ATCC 33500^T was used as outgroup. Bar, 0.01 substitutions per nucleotide position.

Konstantinidis et al., 2017), we can confirm that strain F2-12^T certainly belong to the genus *Natronomonas*.

3.4. Phenotypic characterization of the new proposed species

For the phenotypic characterization, we compared strain F2-12^T with the most closely related species under the same laboratory conditions, as recommended by Oren et al. (1997). Cells of strain F2-12^T were motile rods of different sizes (0.3 µm width by 1–3 µm length), and produced circular, pink pigmented colonies with 1 mm in diameter on R2A 25% medium after 14 days of incubation at 37°C (Table 2). Strain F2-12^T is an extremely halophilic archaeon. It is not able to grow in media without NaCl. It can grow at 15–30% (w/v) total salts and its optimal growth is at 25% (w/v) salts. It is a neutrophilic haloarchaeon, growing at a pH range of 7.0–8.5, with optimal growth at pH 7.5–8.0. Concerning the temperature, strain F2-12^T is able to grow from 20 to 50°C and optimally at 37°C (Table 2). Catalase and oxidase activity are negative. Able to reduce nitrate and nitrite and to produce acids from D-glucose, as well as to utilize D-glucose, glycerol and pyruvate as sole carbon and energy source. The results for other biochemical and nutritional tests are described in Table 2, showing the differential characteristics with respect to the most closely related species of *Natronomonas*.

3.5. Chemotaxonomic characterization of the new proposed species

The total polar lipids of strain F2-12^T were extracted and compared with those from the closely related neighbors, *Natronomonas salsuginis* F20-122^T, *Natronomonas moolapensis* CECT 7526^T, and *Natronomonas pharaonis* CECT 4578^T, as well as with the reference strains *Halorubrum saccharovorum* DSM 1137^T and *Halobacterium salinarum* DSM 3754^T, which permitted an identification of the major polar lipids of the new isolate.

The HPTLC (Supplementary Figure S2) revealed that the polar lipids profile of strain F2-12^T consisted of phosphatidylglycerol (PG) and phosphatidylglycerol phosphate methyl ester (PGP-Me), both derived from $C_{20}C_{20}$ and $C_{20}C_{25}$ archaeol, and phosphatidylglycerol sulfate (PGS) as major polar lipids. Traces of biphosphatidylglycerol (BPG), minor phospholipids, and an unidentified glycolipid were also detected. The profile of polar lipids of the new strain F2-12^T shares the major polar lipids described for all species of the genus *Natronomonas*, except for the

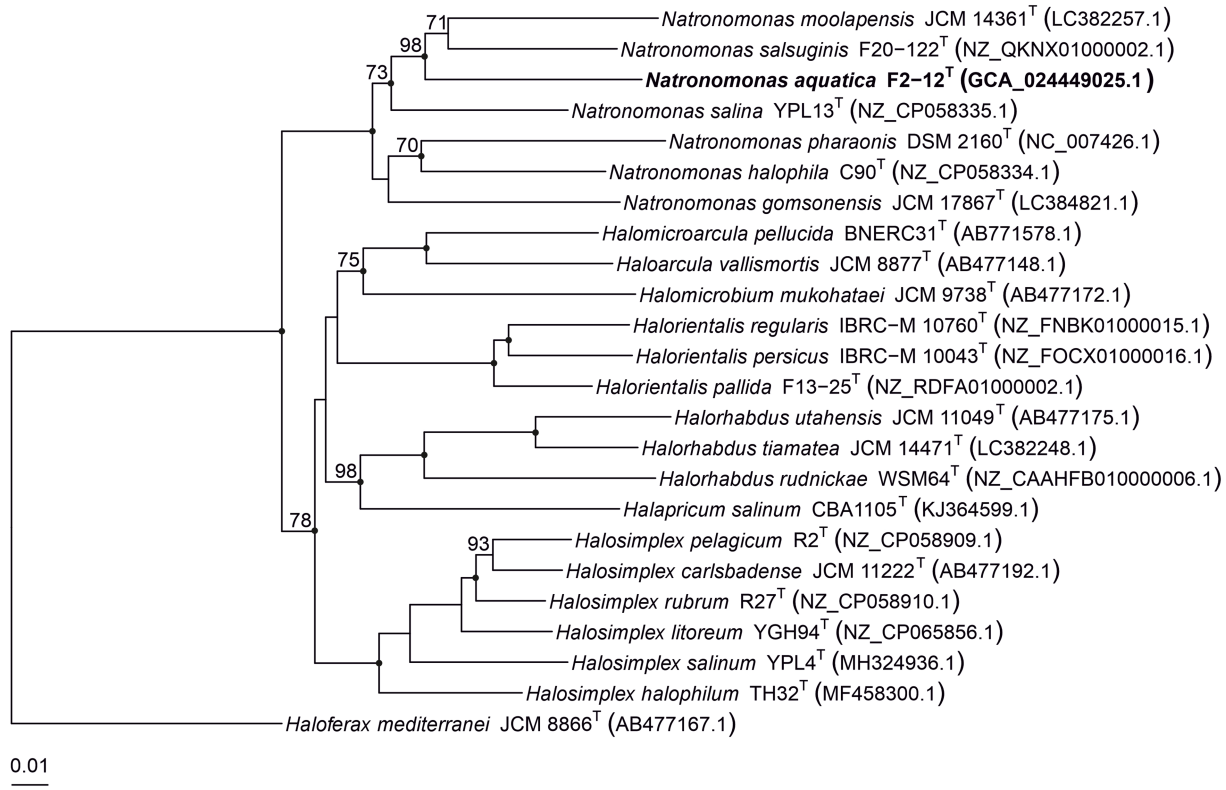


FIGURE 4
Maximum-parsimony phylogenetic tree based on the comparison of *rpoB'* gene sequences showing the phylogenetic relationship between strain F2-12^T, the species of the genus *Natronomonas* and other related haloarchaeal genera. The sequences accession numbers are shown in parentheses. Bootstrap values equal or higher than 70% are indicated above the nodes. Black circles indicate that the corresponding nodes were also obtained in the trees generated with the maximum-likelihood and neighbor-joining algorithms. *Haloferax mediterranei* ATCC 33500^T was used as outgroup. Bar, 0.01 substitutions per nucleotide position.

haloalkaliphilic species *Nmn. pharaonis* CECT 4578^T, that presents a different lipid profile, possibly due to its adaptation to alkaline environments (Soliman and Trüper, 1982). This feature has already been determined for alkaliphilic species of other haloarchaeal genera. Although the lipid pattern of the haloarchaeal species may vary according to the culture conditions, the alkaliphilic species of *Halorubrum* lack phosphatidylglycerol sulfate and glycolipids, that are present on neutrophilic species of this genus (Oren et al., 2009; Corral et al., 2016; de la Haba et al., 2018). A similar situation has been reported for the haloalkaliphilic species *Halostagnicola alkaliphila* (Nagaoka et al., 2011) and *Halostagnicola bangensis* (Corral et al., 2015), in which glycolipids have not been detected, in contrast to the neutrophilic species of the genus *Halostagnicola* (Castillo et al., 2006). A recent study of the polar lipid composition of extremely haloalkaliphilic strains from hypersaline lakes (Bale et al., 2019) also showed the absence of glycolipids in these isolates, and determined the presence of new cardiolipins, that were related to glycocardiolipins previously described in two alkaliphilic species of haloarchaea (Angelini et al., 2012). It has been postulated that this feature may be characteristic of alkaliphilic strains growing at low Mg^{2+} concentrations (Bale et al., 2019).

3.6. Metabolism of the genus *Natronomonas*

Based on the genome annotation from representative species on the genus *Natronomonas* and strain F2-12^T, the major metabolic pathways of this group of microorganisms were inferred. Regarding the carbohydrate metabolism, the complete pathways for glycolysis, gluconeogenesis citrate and glyoxylate cycles were identified along all analyzed genomes, confirming the heterotrophic metabolism for members of the genus *Natronomonas*.

Concerning the nitrogen metabolism, it was very diverse. Genes involved in assimilatory nitrate reduction (*nasAB* and *nirA*) were present in the genomes of *Nmn. pharaonis*, *Nmn. halophila*, and *Nmn. salina*, while this pathway was incomplete on the genomes analyzed for the other studied strains (*Nmn. moolapensis*, *Nmn. salsuginis*, strain F2-12^T, and *Nmn. gomsonensis*) which exclusively presented the gene *nirA*. The species *Nmn. pharaonis* also exhibited a complete dissimilatory nitrate reduction pathway to ammonia, while *Nmn. salsuginis* and *Nmn. moolapensis* presented the *nirBD* gene involved in the same pathway for nitrite reduction to ammonia. Additionally, different genes involved in the denitrification pathway were found in the genomes of *Nmn. pharaonis*, *Nmn. gomsonensis*, *Nmn. halophila*, and *Nmn. salina*, indicating that some of the species of this genus could also carry out modular steps of denitrification.

On the other side, strain F2-12^T and *Nmn. salina* exhibited the almost complete assimilatory sulfate reduction pathway which converts inorganic sulfate to sulfide, which is further incorporated into carbon skeletons of amino acids to form cysteine. The complete pathways for the biosynthesis of the amino acids threonine, cysteine, valine, isoleucine, lysine, arginine, proline and tryptophan were encountered in

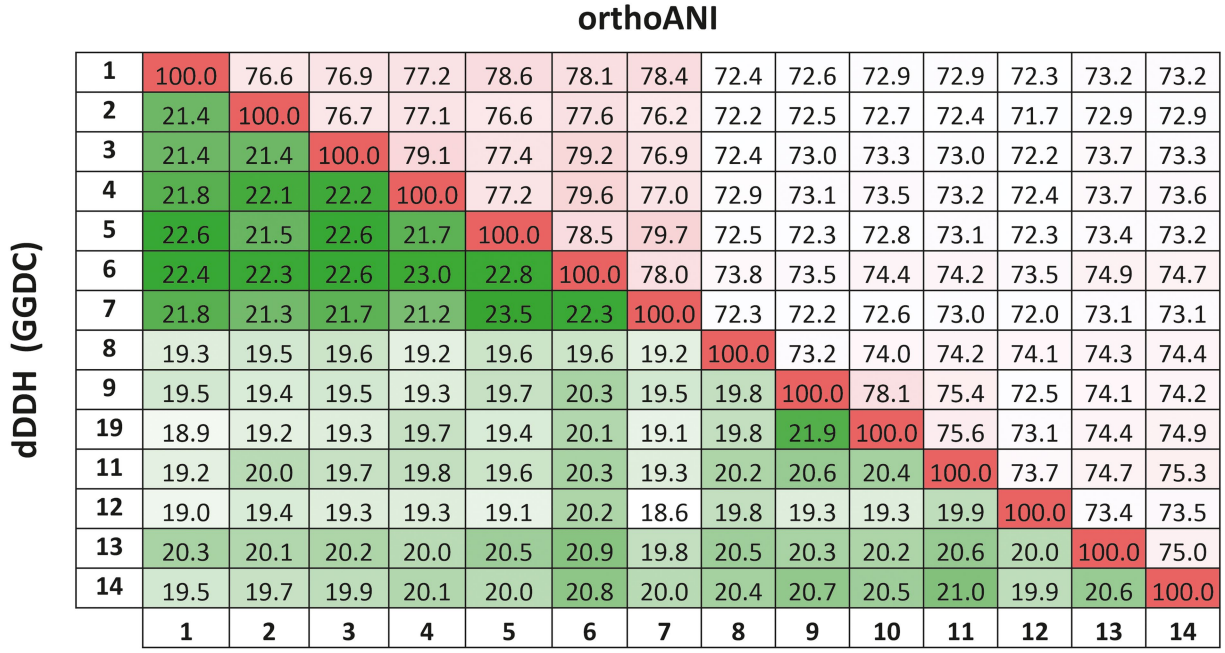


FIGURE 5
Average Nucleotide Identities (orthoANI) and digital DNA–DNA hybridization (dDDH) calculated by the Genome-to-Genome Distance Calculator (GGDC) percentages of strain F2-12^T, the species of the genus *Natronomonas* and other related genera. Strains: 1, *Natronomonas aquatica* F2-12^T (GCA_024449025.1); 2, *Natronomonas pharaonis* DSM 2160^T (GCA_000026045.1); 3, *Natronomonas gomsonensis* JCM 17867^T (GCA_013391635.1); 4, *Natronomonas halophila* C90^T (GCA_013391085.1); 5, *Natronomonas moolapensis* 8.8.11^T (GCA_000591055.1); 6, *Natronomonas salina* YPL13^T (GCA_013391105.1); 7; *Natronomonas salsuginis* F20-122^T (GCA_005239135.1); 8, *Halapricum salinum* CBA1105^T (GCA_004799665.1); 9, *Haloarcula vallismortis* ATCC 29715^T (GCA_000337775.1); 10, *Halomicroarcula pellucida* JCM 17820^T (GCA_014647235.1); 11, *Halomicrobium mukohataei* DSM 12286^T (GCA_000023965.1); 12, *Halorhabdus utahensis* DSM 12940^T (GCA_000023945.1); 13, *Halorientalis regularis* IBRC-M 10760^T (GCA_9001023 05.1); 14, *Halosimplex carlsbadense* 2-9-1^T (GCA_000337455.1).

all the genomes under study, while the leucine biosynthesis pathway was also found in strain F2-12^T, *Nmn. gomsonensis*, *Nmn. halophila*, and *Nmn. salina* and the histidine biosynthesis pathway in *Nmn. gomsonensis*, *Nmn. halophila*, and *Nmn. salina*.

Overall, the current species of the genus *Natronomonas* and strain F2-12^T exhibited similar metabolic reconstructions, even considering the alkaliphilic nature of *Nmn. pharaonis* and the neutrophilic nature of other species of the genus, which further support the taxonomic coherence of this group of archaea as a single genus.

3.7. Ecological importance and environmental abundance of *Natronomonas*

To determine the presence of the new species of *Natronomonas*, *Nmn. aquatica* F2-12^T, as well as of the other six species of the genus *Natronomonas* in several aquatic and terrestrial hypersaline environments, fragment recruitments with different environmental metagenomic datasets were performed. Figure 6 shows the recruitments of strain F2-12^T against nine metagenomic databases from Santa Pola salterns, Alicante, Spain (brines from ponds with salinities of 13, 19, 33, and 37%), Lake Meyghan in Iran (salinities of 18 and 30%), Isla Cristina saltern, Huelva, Spain (21% salinity) and hypersaline soils located in Odiel Saltmarshes, Huelva, Spain (conductivities of 24 and 54 mS/cm). In general, recruitments were evident in saline habitats with intermediate to high salinity, especially in the two metagenomes from Lake Meyghan and that from Isla Cristina saltern (hypersaline habitat from which this new species was isolated).

With respect to the other six species of the genus *Natronomonas*, the recruitment studies show that they are also more abundant in the Lake Meyghan (at intermediate and high salinities) in comparison to the other hypersaline environments tested (Supplementary Figures S3–S8). These results are consistent with previous data by Naghoni et al. (2017), indicating that the genus *Natronomonas* constituted one of the most dominant genera inhabiting that lake. It is noteworthy that they are also detected on the two metagenomic datasets from hypersaline soils, indicating that the members of the genus *Natronomonas* are ubiquitous in hypersaline habitats, being present in aquatic as well as in terrestrial hypersaline environments. Finally, the abundance of metagenomic reads recruited below 95% similarity might suggests the existence of additional species belonging to the genus *Natronomonas* present in these habitats.

4. Conclusion

The 16S rRNA gene, *rpoB'* gene, phylogenomic analysis, genomic indexes OrthoANI, dDDH and AAI, and metabolic analysis demonstrate that the current species of the genus *Natronomonas* constitute a coherent taxonomic group, even considering that the type species of the genus, *Nmn. pharaonis,* is a haloalkaliphilic archaeon that shows some specific phenotypic characteristics, such as a different polar lipid profile, probably due to its adaptive features to alkaline environments. The new strain F2-12^T, isolated in pure culture from the pond of a saltern located in Isla Cristina, Spain, was studied and characterized using genomic and classical taxonomic methods. This study shows that strain F2-12^T constitutes a new species of the genus *Natronomonas*, and thus,

TABLE 2 Differential characteristics of *Natronomonas aquatica* F2-12^T, *Natronomonas salsuginis* F20-122^T, *Natronomonas moolapensis* CECT 7526^T, *Natronomonas pharaonis* DSM 2160^T, *Natronomonas gomsonensis* SA3^T, *Natronomonas halophila* C90^T, and *Natronomonas salina* YPL13^T.

Characteristics	1	2	3	4	5	6	7
Morphology	Rods	Coccoid[a]	Pleomorphic[b]	Rods[c]	Coccoid[d]	Pleomorphic[e]	Pleomorphic[e]
Motility	+	–	+	+	+[d]	+[e]	+[e]
Cell size (μm)	0.3 × 1–3	1.0 × 1.2–2.5[a]	0.7 × 1.7[b]	0.8 × 1–3[c]	0.8–0.9 × 1.2[d]	ND	ND
Colony size (mm)	1	0.2–0.3	0.5–1.0	0.5	1.0–2.0[d]	ND	ND
Colony pigmentation	Pink	Pink	Red	Red	Red[d]	Red[e]	Red[e]
NaCl range (optimum) (%, w/v)	15–30(25)	10–30(25)[a]	14–36(18–20)[b]	12–30(20)[c]	18–30(24)[d]	5–28(25)[e]	5–28(20)[e]
Temperature range for growth (optimum) (°C)	20–50(37)	25–50(37)[a]	25–45(37–45)[b]	20–50(45)[c]	20–45(40)[d]	30–60(40)[e]	25–50(37)[e]
pH range (optimum)	7.0–8.5(7.5–8.0)	6.5–9.0(8.0)[a]	5.5–8.5(7.0–7.5)[b]	8.0–11.00(8.5–9.0)[c]	5.5–8.0(7.0)[d]	6.5–9.5(8.0)[e]	5.0–8.0(6.5)[e]
Oxidase	–	+	–	+	+[d]	–[e]	+[e]
Indole production	–	–	–	+	–[d]	–[e]	–[e]
Hydrolysis of gelatin	–	–	–	+	–[d]	–[e]	–[e]
Reduction of nitrate	+	+	+	–	–[d]	+[e]	+[e]
Reduction of nitrite	+	+	–	–	–[d]	+[e]	+[e]
Production of acids from carbohydrates:							
D-fructose	–	+	–	ND	+[d]	ND	ND
D-glucose	+	+	–	ND	+[d]	ND	ND
Utilization as sole carbon and energy source of:							
D-glucose	+	–	+	ND	+[d]	–[e]	–[e]
Glycerol	+	–	+	ND	+[d]	–[e]	–[e]
Fumarate	–	–	+	ND	+[d]	+[e]	+[e]
Pyruvate	+	–	+	ND	+[d]	+[e]	+[e]

Strains: **1**, *Natronomonas aquatica* F2-12^T, **2**, *Natronomonas salsuginis* F20-122^T, **3**, *Natronomonas moolapensis* CECT 7526^T, **4**, *Natronomonas pharaonis* DSM 2160^T, **5**, *Natronomonas gomsonensis* SA3^T, **6**, *Natronomonas halophila* C90^T, **7**, *Natronomonas salina* YPL13^T. All data from this study, except [a]Durán-Viseras et al. (2020), [b]Burns et al. (2009), [c]Soliman and Trüper (1982) and Burns et al. (2009), [d]Kim et al. (2013) and [e]Yin et al. (2020). +, Positive; –, negative; ND, not determined.

we propose to name it as *Natronomonas aquatica* sp. nov. The formal description of this new species is given below. Finally, the metagenomic recruitment studies show that members of the genus *Natronomonas* are widely distributed on aquatic (salterns, lake) and also on terrestrial hypersaline habitats.

Description of *Natronomonas aquatica* sp. nov.

Natronomonas aquatica (a.qua'ti.*ca.* L. fem. adj. *aquatica*, lives or grows in the water, aquatic, it is referred to the place where it was isolated).

Cells are Gram-stain-negative, motile rods with a size of 0.3 μm × 1–3 μm. Colonies are circular, pink pigmented and 1 mm in diameter on R2A medium supplemented with 25% salts after 14 days of incubation at 37°C. Chemoorganotrophic and aerobic. Extremely halophilic and neutrophilic archaeon, growing in media with 15–30% (w/v) NaCl and with an optimal growth at 25% (w/v) NaCl. Not able to grow in the absence of NaCl. The pH range for growth is 7.0–8.5 and the optimal growth is at pH 7.5–8.0. Able to grow from 20 to 50°C and optimally at 37°C. Does not grow anaerobically with potassium nitrate, L-arginine or DMSO. Catalase and oxidase activity are negative. Aesculin, gelatin, starch and Tween 80 are not hydrolyzed. Nitrate and nitrite are reduced, H_2S is not produced from thiosulfate. Simmons' citrate and Voges-Proskauer tests are negative. Methyl red test is positive. Indole is not produced from tryptophan. Acid is produced from D-glucose and D-ribose but not from D-amygdalin, D, L-ethionine, D-arabinose, arbutin, D-cellobiose, L-citrulline, D-fructose, D-galactose, glycerol, inulin, D-maltose, D-mannitol, D-mannose, D-melezitose, D-raffinose, D-sucrose, sorbitol, D-trehalose, L-xylitol, and D-xylose. The following compounds are used as sole source of carbon and energy: D-arabinose, D-glucose, glycerol, xylitol, L-alanine, glutamine, isoleucine and pyruvate, but not fructose, lactose, maltose, L-raffinose, sucrose, D-trehalose, D-xylose, salicin, butanol, D-mannitol, propanol, L-arginine, L-cysteine, L-methionine, L-glycine, L-lysine, valine, benzoate, citrate, formate, fumarate, propionate, valerate, hippurate, malate, and tartrate. The major polar lipids are phosphatidylglycerol (PG), phosphatidylglycerol phosphate methyl ester (PGP-Me), and phosphatidylglycerol sulfate (PGS). Traces of biphosphatidylglycerol (BPG) and other minor phospholipids and an unidentified glycolipid may also be present.

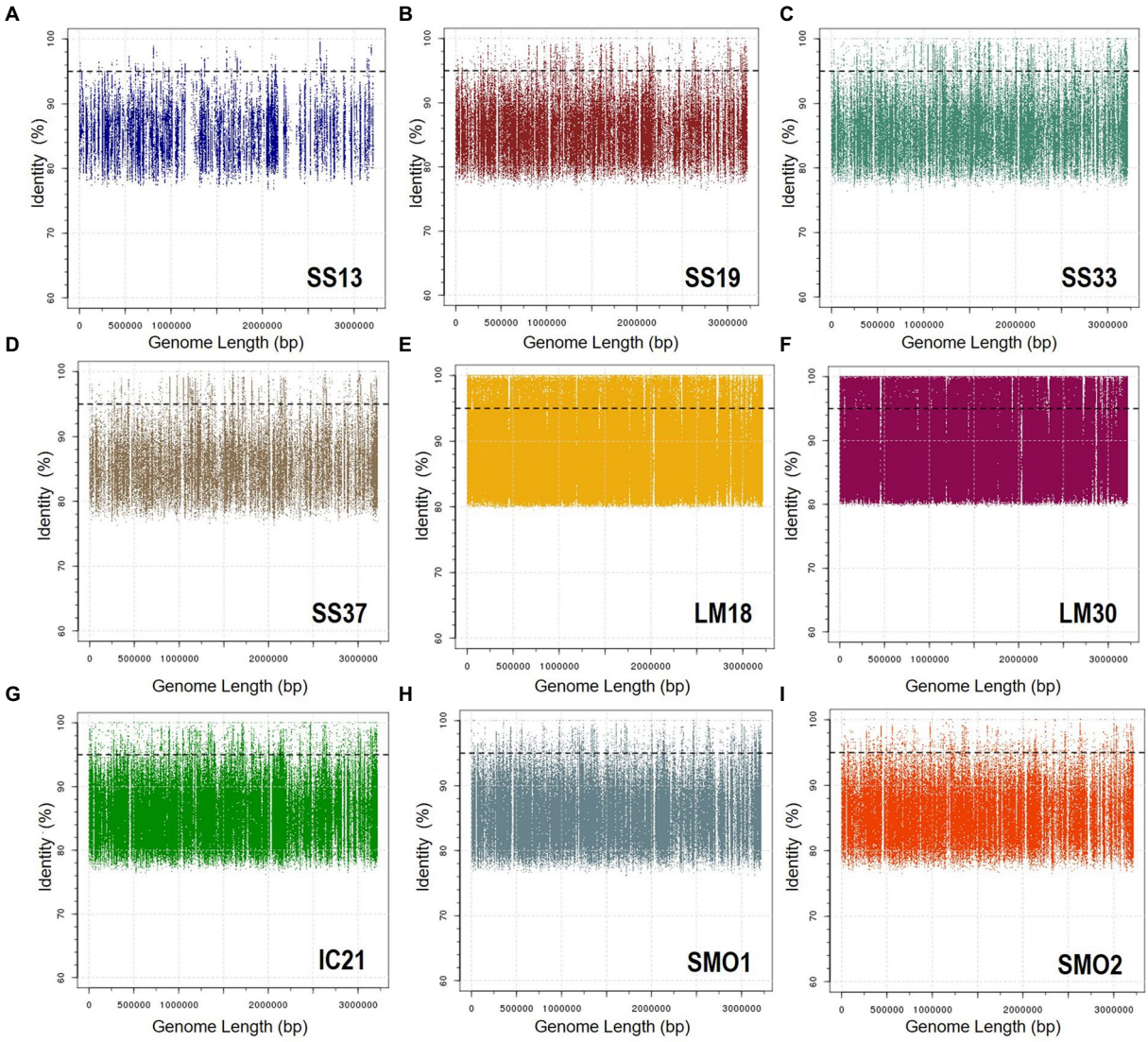


FIGURE 6
Recruitment plots of *Natronomonas aquatica* F2-12^T against different metagenomic datasets: **(A)** SS13, **(B)** SS19, **(C)** SS33, **(D)** SS37, **(E)** LM18, **(F)** LM30, **(G)** IC21, **(H)** SMO1, and **(I)** SMO2. In each panel the Y axis represents the identity percentage and X axis represents the genome length. The black dashed line shows the threshold for the presence of same species (95% identity). Abbreviations: SS13: metagenome from Santa Pola saltern (Spain) with 13% salinity (SRX328504; Fernández et al., 2014a), SS19: metagenome from Santa Pola saltern (Spain) with 19% salinity (SRX090228; Ghai et al., 2011), SS33: metagenome from Santa Pola saltern (Spain) with 33% salinity (SRX347883; Fernández et al., 2013), SS37: metagenome from Santa Pola saltern (Spain) with 37% salinity (SRX090229; Ghai et al., 2011), LM18: metagenome from Lake Meyghan (Iran) with 18% salinity (ERS1455390; Naghoni et al., 2017), LM30: metagenome from Lake Meyghan (Iran) with 30% salinity (ERS1455391; Naghoni et al., 2017), IC21: metagenome from Isla Cristina saltern (Spain) with 21% salinity (Fernández et al., 2014b,c), SMO1: metagenome from Odiel saltmarshes hypersaline soil, 24 mS/cm conductivity (SRR5753725; Vera-Gargallo and Ventosa, 2018), SMO2: metagenome from Odiel saltmarshes hypersaline soil, 54 mS/cm conductivity (SRR5753724; Vera-Gargallo and Ventosa, 2018).

The type strain is F2-12^T (= CCM 9001^T = CECT 9970^T = JCM 33798^T), and was isolated from a marine saltern located in Isla Cristina, Huelva, Spain. The DNA G+C content of the type strain F2-12^T is 62.7 mol%, as determined from its genome.

The GenBank/EMBL/DDBJ accession number for the 16S rRNA and *rpoB'* gene sequences of strain F2-12^T are MZ318646 and MZ327708, respectively, and that of the genome sequence is GCA_024449025.1.

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/, MZ318646, MZ327708, and JAHLKM000000000.

Author contributions

AD-V, CS-P, and AV did the conceptualization. AD-V performed the isolation of the strain. AG-R carried out the phenotypic characterization. AG-R and AD-V carried out the genomic analyses, RRH contributed to the phylogenomic analyses, PC to the polar lipids composition and CS-P to the recruitment experiments. AG-R, AD-V, CS-P, and AV prepared the draft manuscript and the tables and figures. AV and CS-P did the funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants PID2020-118136GB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ESF Investing in your future and by Junta de Andalucía (Spain; P20_01066 and BIO-213), all including European (FEDER) funds. AG-R was the recipient of a predoctoral fellowship from the Spanish Ministry of Universities. AD-V was supported by a Margarita Salas postdoctoral research fellowship from the Spanish Ministry of Universities (financed by the European Union, under the NextGenerationEU funds). RRH was the recipient of a short-stay grant "Salvador Madariaga" from the Spanish Ministry of Science, Innovation and Universities (Programa Estatal de Promoción del Talento y su Empleabilidad en I+D+I, Subprograma Estatal de Movilidad, del Plan Estatal de Investigación Científica y Técnica y de Innovación 2017–2020; PRX21/00598).

Acknowledgments

We thank C. Galisteo, D. Straková, and M. J. León for their helpful discussion and A. Oren for his help on the nomenclature of the new species.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be constructed 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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1109549/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

Andrews, S. (2010). Fast QC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (Accessed September 21, 2021).

Angelini, R., Corral, P., Lopalco, P., Ventosa, A., and Corcelli, A. (2012). Novel ether lipid cardiolipins in archaeal membranes of extreme haloalkaliphiles. *Biochim. Biophys. Acta* 1818, 1365–1373. doi: 10.1016/j.bbamem.2012.02.014

Auch, A. F., Klenk, H. P., and Göker, M. (2019). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Bale, N. J., Sorokin, D. Y., Hopmans, E. C., Koenen, M., Rijpstra, W. I., Villanueva, L., et al. (2019). New insights into the polar lipid composition of extremely halo(alkali)philic euryarchaea from hypersaline lakes. *Front. Microbiol.* 10:377. doi: 10.3389/fmicb.2019.00377

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

Boucher, Y. F., Douady, C. J., Sharma, A. K., Kamekura, M., and Doolittle, W. F. (2004). Intragenomic heterogeneity and intergenomic recombination among haloarchaeal rRNA genes. *J. Bacteriol.* 186, 3980–3990. doi: 10.1128/JB.186.12.3980-3990.2004

Burns, D. G., Janssen, P. H., Itoh, T., Minegishi, H., Usami, R., Kamekura, M., et al. (2009). *Natronomonas moolapensis* sp. nov., non-alkaliphilic isolates recovered from a solar saltern crystallizer pond, and emended description of the genus *Natronomonas*. *Int. J. Syst. Evol. Microbiol.* 60, 1173–1176. doi: 10.1099/ijs.0.010132-0

Bushnell, B. (2021). BBMap short read aligner, and other bioinformatic tools. Available at: sourceforge.net/projects/bbmap/ [Accessed September 21, 2021].

Castillo, A. M., Gutiérrez, M. C., Kamekura, M., Xue, Y., Ma, Y., Cowan, D. A., et al. (2006). *Halostagnicola larsenii* gen. nov., sp. nov., an extremely halophilic archaeon from a saline lake in Inner Mongolia, China. *Int. J. Syst. Evol. Microbiol.* 56, 1519–1524. doi: 10.1099/ijs.0.64286-0

Chaumeil, P.-A., Mussig, A. J., Hugenholthz, P., and Parks, D. H. (2022). GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38, 5315–5316. doi: 10.1093/bioinformatics/btac672

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

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., Sánchez-Porro, C., Amoozegar, M. A., Papke, R. T., and Ventosa, A. (2016). *Halorubrum halodurans* sp. nov., an extremely halophilic archaeon isolated from a hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 66, 435–444. doi: 10.1099/ijsem.0.000738

Cowan, S. T., and Steel, K. J. (1993). *Manual for the Identification of Medical Bacteria. 3rd Edn.*, Cambridge: Cambridge University Press.

Cui, H.-L., Zhou, P.-J., Oren, A., and Liu, S.-J. (2009). Intraspecific polymorphism of 16S rRNA genes in two halophilic archaeal genera, *Haloarcula* and *Halomicrobium*. *Extremophiles* 13, 31–37. doi: 10.1007/s00792-008-0194-2

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

Durán-Viseras, A. (2020). Overcoming the limits of cultivation: Taxogenomics and comparative genomics of new haloarchaea. PhD Thesis, University of Sevilla, Sevilla.

Durán-Viseras, A., Sánchez-Porro, C., and Ventosa, A. (2019). *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., 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

Edgar, R. C. (2002). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 5, 1792–1797. 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 *rpo*B' 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

Fernández, A. B., Ghai, R., Martin-Cuadrado, A.-B., Sánchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2013). Metagenome sequencing of prokaryotic microbiota from two hypersaline ponds of a marine saltern in Santa Pola, Spain. *Genome Announc.* 1:e00933-13. doi: 10.1128/genomeA.00933-13

Fernández, A. B., Ghai, R., Martin-Cuadrado, A.-B., Sánchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2014a). Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* 88, 623–635. doi: 10.1111/1574-6941.12329

Fernández, A. B., León, M. J., Vera, B., Sánchez-Porro, C., and Ventosa, A. (2014b). Metagenomic sequence of prokaryotic microbiota from an intermediate-salinity pond of a saltern in Isla Cristina, Spain. *Genome Announc.* 2:e00045-14. doi: 10.1128/genomeA.00045-14

Fernández, A. B., Vera-Gargallo, B., Sánchez-Porro, C., Ghai, R., Papke, R. T., Rodriguez-Valera, F., et al. (2014c). 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

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

Ghai, R., Pašić, L., Fernández, A. B., Martin-Cuadrado, A.-B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135

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

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

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–98.

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

Kamekura, M., Dyall-Smith, M. L., Upasani, V., Ventosa, A., and Kates, M. (1997). Diversity of alkaliphilic halobacteria: proposals for transfer of *Natronobacterium vacuolatum, Natronobacterium magadii*, and *Natronobacterium pharaonis* to *Halorubrum*, *Natrialba*, and *Natronomonas* gen. nov., respectively, as *Halorubrum vacuolatum* comb. nov., *Natrialba magadii* comb. nov., and *Natronomonas pharaonis* comb. nov., respectively. *Int. J. Syst. Bacteriol.* 47, 853–857. doi: 10.1099/00207713-47-3-853

Kim, T. Y., Kim, S. J., Park, S. J., Kim, J. G., Cha, I. T., Jung, M. Y., et al. (2013). *Natronomonas gomsonensis* sp. nov., isolated from a solar saltern. *Antonie van Leeuwenhoek* 104, 627–635. doi: 10.1007/s10482-013-9970-9

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

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. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3160–3165. doi: 10.1073/pnas.0409727102

Kovács, N. (1928). Eine vereinfachte Methode zum Nachweis der Indolbildung durch Bakterien. *Z. Immunitätsforsch.* 55, 311–315.

Kovács, N. (1956). Identification of *Pseudomonas pyocyanea* by the oxidase reaction. *Nature* 178:703. doi: 10.1038/178703a0

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

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, , et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from micro-organisms. *J. Mol. Biol.* 3, 208–IN1. 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 Bioinform.* 14:60. doi: 10.1186/1471-2105-14-60

Minegishi, H., and Kamekura, M. (2018). "Genus *Natronomonas*" in *Bergey's Manual of Systematics of Archaea and Bacteria*. ed. W. D. Whitman (New Jersey: Whiley & Sons), 1–8.

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' (*rpo*B') gene. *Int. J. Syst. Evol. Microbiol.* 60, 2398–2408. doi: 10.1099/ijs.0.017160-0

Nagaoka, S., Minegishi, H., Echigo, A., Shimane, Y., Kamekura, M., and Usami, R. (2011). *Halostagnicola alkaliphila* sp. nov., an alkaliphilic haloarchaeon from commercial rock salt. *Int. J. Syst. Evol. Microbiol.* 61, 1149–1152. doi: 10.1099/ijs.0.023119-0

Naghoni, A., Emtiazi, G., Amoozegar, M. A., Cretoiu, M. S., Stal, L. J., Etemadifar, Z., et al. (2017). Microbial diversity in the hypersaline Lake Meyghan, Iran. *Sci. Rep.* 7:11522. doi: 10.1038/s41598-017-11585-3

Oren, A. (2011). "Ecology of halophiles" in *Extremophiles Handbook*. ed. K. Horikoshi (Tokyo: Springer), 343–361.

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. (2017). "Family *Haloarculaceae*" in *Bergey's Manual of Systematics of Archaea and Bacteria*. ed. W. D. Whitman (New Jersey: Whiley & Sons), 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

Oren, A., Ventosa, A., and Kamekura, M. (2017). "Class *Halobacteria*" in *Bergey's Manual of Systematics of Archaea and Bacteria*. ed. W. D. Whitman (New Jersey: Whiley & Sons), 1–5.

Parks, D. H., Imelfort, M., Skennerton, C., 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

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2-approximately maximum-likelihood trees for large alignments. *PLoS One* 5:3. doi: 10.1371/journal.pone.0009490

Rodríguez-R, L. M., and Konstantinidis, K. T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

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 loglikelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114–1116. doi: 10.1093/oxfordjournals.molbev.a026201

Skerman, V. B. D. A. (1967). *Guide to the Identification of the Genera of Bacteria, with Methods and Digests of Generic Characteristics. 1st Ed.*, Baltimore: Williams & Wilkins.

Soliman, G. S., and Trüper, H. G. (1982). *Halobacterium pharaonis* sp. nov., a new, extremely haloalkaliphilic archaebacterium with low magnesium requirement. *Zentralbl. Bakteriol. Mikrobiol. Hyg. I. Abt. Orig. C* 3, 318–329.

Sun, D. L., Jiang, X., Wu, Q. L., and Zhou, N. Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Appl. Environ. Microbiol.* 79, 5962–5969. doi: 10.1128/AEM.01282-13

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

Torreblanca, M., Rodriguez-Valera, F., Juez, G., Ventosa, A., Kamekura, M., and Kates, M. (1986). Classification of non-alkaliphilic halobacteria based on numerical taxonomy and polar lipid composition, and description of *Haloarcula* gen. nov. and *Haloferax* gen. nov. *Syst. Appl. Microbiol.* 8, 89–99. doi: 10.1016/S0723-2020(86)80155-2

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

Ventosa, A., Quesada, E., Rodríguez-Valera, F., Ruiz-Berraquero, F., and Ramos Cormenzana, A. (1982). Numerical taxonomy of moderately halophilic Gram-negative rods. *J. Gen. Microbiol.* 128, 1959–1968. doi: 10.1099/00221287-128-9-1959

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

Vreeland, R. H., Straighr, S., Krammes, J., Dougherty, K., Rosenzweig, W. D., and Kamekura, M. (2002). *Halosimplex carlsbadense* gen. nov., sp. nov., a unique halophilic archaeon, with three 16S rRNA genes, that grows only in defined medium with glycerol and acetate or pyruvate. *Extremophiles* 6, 445–452. doi: 10.1007/s00792-002-0278-3

Walsh, D. A., Bapteste, E., Kamekura, M., and Doolittle, W. F. (2004). Evolution of the RNA polymerase B' subunit gene (*rpo*B') in *Halobacteriales*: a complementary molecular marker to the SSU rRNA gene. *Mol. Biol. Evol.* 21, 2340–2351. doi: 10.1093/molbev/msh248

Werkman, C. H. (1930). An improved technic for the Voges-Proskauer test. *J. Bacteriol.* 20, 121–125. doi: 10.1128/jb.20.2.121-125.1930

Yin, X. M., Yang, X. Y., Hou, J., Zhu, L., and Cui, H. L. (2020). *Natronomonas halophila* sp. nov. and *Natronomonas salina* sp. nov., two novel halophilic archaea. *Int. J. Syst. Evol. Microbiol.* 70, 5686–5692. doi: 10.1099/ijsem.0.004463

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fmicb.2022.1101464

OPEN ACCESS

EDITED BY
Sumit Kumar,
Amity Institute of Biotechnology,
Amity University,
India

REVIEWED BY
Aharon Oren,
Hebrew University of Jerusalem,
Israel
XueWei Xu,
Ministry of Natural Resources,
China

*CORRESPONDENCE
Cristina Sánchez-Porro
✉ sanpor@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 17 November 2022
ACCEPTED 22 December 2022
PUBLISHED 26 January 2023

CITATION
Galisteo C, de la Haba RR, Sánchez-Porro C and Ventosa A (2023) Biotin pathway in novel *Fodinibius salsisoli* sp. nov., isolated from hypersaline soils and reclassification of the genus *Aliifodinibius* as *Fodinibius*. *Front. Microbiol.* 13:1101464. doi: 10.3389/fmicb.2022.1101464

Biotin pathway in novel *Fodinibius salsisoli* sp. nov., isolated from hypersaline soils and reclassification of the genus *Aliifodinibius* as *Fodinibius*

Cristina Galisteo, Rafael R. de la Haba, Cristina Sánchez-Porro* and Antonio Ventosa*

Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain

Hypersaline soils are extreme environments that have received little attention until the last few years. Their halophilic prokaryotic population seems to be more diverse than those of well-known aquatic systems. Among those inhabitants, representatives of the family *Balneolaceae* (phylum *Balneolota*) have been described to be abundant, but very few members have been isolated and characterized to date. This family comprises the genera *Aliifodinibius* and *Fodinibius* along with four others. A novel strain, designated 1BSP15-2V2^T, has been isolated from hypersaline soils located in the Odiel Saltmarshes Natural Area (Southwest Spain), which appears to represent a new species related to the genus *Aliifodinibius*. However, comparative genomic analyses of members of the family *Balneolaceae* have revealed that the genera *Aliifodinibius* and *Fodinibius* belong to a single genus, hence we propose the reclassification of the species of the genus *Aliifodinibius* into the genus *Fodinibius,* which was first described. The novel strain is thus described as *Fodinibius salsisoli* sp. nov., with 1BSP15-2V2^T (=CCM 9117^T=CECT 30246^T) as the designated type strain. This species and other closely related ones show abundant genomic recruitment within 80–90% identity range when searched against several hypersaline soil metagenomic databases investigated. This might suggest that there are still uncultured, yet abundant closely related representatives to this family present in these environments. In-depth *in-silico* analysis of the metabolism of *Fodinibius* showed that the biotin biosynthesis pathway was present in the genomes of strain 1BSP15-2V2^T and other species of the family *Balneolaceae*, which could entail major implications in their community role providing this vitamin to other organisms that depend on an exogenous source of this nutrient.

KEYWORDS

***Balneolaceae*, hypersaline soils, phylogenomics, biotin biosynthesis, genomic analysis, moderate halophile, taxonomic reclassification**

1. Introduction

The family *Balneolaceae* (Xia et al., 2016) belongs to the phylum *Balneolota*, class *Balneolia* (Munoz et al., 2016), and includes six genera: *Aliifodinibius*, *Balneola*, *Fodinibius*, *Gracilimonas*, *Halalkalibaculum*, and *Rhodohalobacter* (Parte et al., 2020), which are found in marine and hypersaline environments, such as deep-sea sediments, salt-mines, salterns, and saline soils (Urios et al., 2006; Choi et al., 2009; Wang et al., 2012, 2013; Cho et al., 2013; Xia et al., 2017; Wu et al., 2022a,b). The

genus *Fodinibius* was proposed in 2012 based on a single species, *Fodinibius salinus* (Wang et al., 2012) and the genus *Aliifodinibius* was firstly described only one year later (Wang et al., 2013), including at that time two species, *Aliifodinibius roseus* (type species) and *Aliifodinibius sediminis*. These taxa shared low 16S rRNA gene sequence identities (<92.4%) with *F. salinus,* the most closely related species. However, no further genomic comparison was performed at that time. The aforementioned identity values, along with their differential fatty acids and polar lipids profiles, as well as some phenotypic features, were considered enough evidence to propose that *A. roseus* and *A. sediminis* represented a novel genus, different from *Fodinibius*. At the time of writing, five species names have been validly published within the genus *Aliifodinibius*: *Aliifodinibius roseus* and *Aliifodinibius sediminis* (Wang et al., 2013), *Aliifodinibius halophilus* (Xia et al., 2016), *Aliifodinibius salicampi* (Cho et al., 2017, 2018), and *Aliifodinibius saliphilus* (Cho and Whang, 2020). Besides, the species "*Aliifodinibius salipaludis*" (Zhao et al., 2020) has been described but its name has not been validly published so far. All these species were isolated from hypersaline environments, such as salt mines, salterns, and saline soils. They are Gram-stain-negative, rod-shaped, non-motile, and moderately halophilic bacteria, with a NaCl requirement for growing between 4 and 25% (w/v) and an optimum between 8 and 15%, except for "*A. salipaludis*," that can grow optimally at 25% (w/v) NaCl (Zhao et al., 2020). The pH values supporting their growth range between pH 6.0 and 11.0, with an optimum at pH 7.0–8.0. They are able to produce catalase and oxidase but unable to hydrolyze starch and DNA (Wang et al., 2013; Xia et al., 2016; Cho et al., 2017; Cho and Whang, 2020; Zhao et al., 2020).

The Odiel Saltmarshes Natural Area is located in the estuary of the Odiel and Tinto rivers (Southwest Spain), and its hypersaline soils have been previously explored by Vera-Gargallo and Ventosa (2018). They obtained two metagenomes database of these area named SMO1 and SMO2. Their shotgun metagenomic analysis taxonomically annotated almost a 20% of the prokaryotic diversity as representatives of the phylum *Balneolota*. Genus-level taxonomic affiliation of 16S rRNA genes detected in the reads above 95% identity assigned 9.0 and 3.4% of the abundance to the genera *Fodinibius* and *Gracilimonas*, respectively, for the SMO1 metagenomic database, and 3.0 and 1.6% for the SMO2 metagenomic database. Other studies have also revealed the presence of this taxonomic group on different metagenomic datasets. The abundance of the uncultured members of the class *Balneolia* was up to 8% in the hypersaline soda lake sediment studied by Vavourakis et al. (2018), being one of the three predominant bacterial groups along with the phylum *Bacillota* and the class *Gammaproteobacteria*. Both studies recovered MAGs related to uncultured members of the family *Balneolaceae* (Vavourakis et al., 2018; Vera-Gargallo and Ventosa, 2018). Kimbrel et al. (2018) obtained a medium-quality MAG related to *Gracilimonas tropica*, with a 0.26% of abundance, from sediments with salinity of 7.5%. The thalassohaline brines located in Maras (Peru) showed a considerable abundance of up to 6% of *Balneolota*, mainly associated (5% out of the total) to the species *A. roseus* (Castelán-Sánchez et al., 2019). These previous studies indicate that the members of the family *Balneolaceae* are fairly abundant in hypersaline ecosystems, although the number of cultivated species is still scarce.

In the present study, we describe the isolation of a new strain, following a culturomic strategy, and we have carried out its exhaustive description as a novel species within the family *Balneolaceae*. Besides, we completed the phylogenomic analyses of the whole family, which revealed the need for merging two of its genera. Furthermore, we performed genome-based functional analyses of these bacteria, including the vitamin B_7 pathway, to elucidate their potential ecological role and adaptative mechanisms to survive under harsh conditions of salinity and heavy metals, along with its distribution in hypersaline habitats.

2. Materials and methods

2.1. Strain isolation and physico-chemical features of the sampling site

The samples were collected from hypersaline soils located at the saltmarshes of the Odiel Natural Area in Huelva (Southwest Spain) as previously described (Vera-Gargallo and Ventosa, 2018; Vera-Gargallo et al., 2019). The pH and electrical conductivity of the saline soil was measured with the pHmeter CRISON BASIC 20 and the conductometer CRISON 35+, respectively, after a 1:5 dilution. Arsenic and cadmium concentrations were measured by Inductively Coupled Plasma Mass Spectrometry (ICP-MS); copper and zinc concentrations, by Inductively Coupled Plasma Optical Emission Spectroscopy (ICP-OES); and lead concentration, by atomic absorption spectroscopy. All those spectrometric assays were carried out by Innoagral Laboratories (Mairena del Aljarafe, Sevilla, Spain).

The strain 1BSP15-2V2^T was isolated using the standard dilution-plating technique and grew on a low-nutrient medium enriched with pyruvate (SMM medium) at 28°C after 3 months of incubation. The colonies were subcultured until a pure culture was obtained. The composition of the low nutrients SMM medium was the following ($g\,L^{-1}$): casein digest, 5.0; sodium pyruvate, 1.1; supplemented with up to 15% (w/v) salt concentration from a seawater (SW) stock ([$g\,L^{-1}$] NaCl, 234.0; $MgCl_2{\cdot}6H_2O$, 39.0; $MgSO_4{\cdot}7H_2O$, 61.0; $CaCl_2$, 1.0; KCl, 6.0; $NaHCO_3$, 0.2; NaBr, 0.7); the pH was adjusted to 7.5; solidified with 2.0% (w/v) agar when needed. This medium was first used by León et al. (2016) to isolate the species of the genus *Spiribacter*, which is an abundant halophilic bacterium of the salterns community at intermediate salinities. The culture was preserved at −80°C in the same liquid medium with 40% (v/v) glycerol. Additionally, Marine Agar (MA, Difco 2216) medium, supplemented with 10% (w/v) NaCl, was used to determine the fatty acid profile of this strain for better comparative purposes with other species of the genus under study. In addition, reference type strains of the species of the genus *Aliifodinibius* were used for comparative phenotypic analyses. The following strains were obtained from the corresponding culture collections: *A. halophilus* KCTC 42497^T, *A. roseus* DSM 21986^T, *A. salicampi* KACC 19060^T, "*A. salipaludis*" KCTC 52855^T, *A. saliphilus* KACC 19126^T, and *A. sediminis* DSM 21194^T. These strains were grown and preserved in R2A medium supplemented with 7.5% (w/v) SW. Additionally, DNA of strain *A. salicampi* KACC 19060^T was extracted and sequenced as detailed below as it was the only species of the genus *Aliifodinibius* whose genome was not publicly available.

2.2. DNA extraction and PCR amplification

Genomic DNA of strain 1BSP15-2V2^T was extracted using the Marmur (1961) method modified for small volumes. The almost complete 16S rRNA gene sequence was amplified by PCR with universal primers 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R

(5′-GGT TAC CTT GTT ACG ACT T3′) (Lane, 1991). Both genomic DNA and PCR products were purified using MEGAquick-spin™ plus (iNtRON Biotechnology), following the manufacturer's recommendations. The 16S rRNA gene was sequenced using Sanger technology by StabVida (Caparica, Portugal). The same protocol was followed for the DNA extraction of strain *A. salicampi* KACC 19060^T.

2.3. Phylogenetic analysis

Identification of phylogenetic neighbors of strain 1BSP15-2V2^T according to the 16S rRNA gene sequence identity was carried out searching the EzBioCloud database[1] (Yoon et al., 2017). For further phylogenetic analysis, the type strains of the identified closely related taxa of the family *Balneolaceae* were used. The 16S rRNA gene sequences of the strains included in the analysis were obtained from SILVA and GenBank databases. Alignment of the primary and secondary structures of the selected 16S rRNA sequences was performed using the integrated fast aligner implemented in the ARB package (Ludwig et al., 2004). For phylogenetic tree reconstruction, maximum-likelihood (Felsenstein, 1981), maximum-parsimony (Felsenstein, 1983), and neighbor-joining (Saitou and Nei, 1987) algorithms integrated in the ARB software (Ludwig et al., 2004) were used, and the Jukes-Cantor model (Jukes and Cantor, 1969) was selected to correct the distance matrix. The robustness of the branch topology was checked by bootstrap analysis (1,000 pseudoreplicates) (Felsenstein, 1985). The script "gitana"[2] was employed for formatting and imaging of the tree visualization.

2.4. Comparative genomic analyses

Whole shotgun sequencing of the genome of strain 1BSP15-2V2^T was performed by Novogene Europe (Cambridge, United Kingdom) on the Illumina NovaSeq PE150 platform. Paired end reads were assembled with SPAdes v3.13.0 (Prjibelski et al., 2020). Quality control of the assembly was performed with QUAST v2.3 (Gurevich et al., 2013), which also calculated some genomic statistics, such as the DNA G + C content. The length-filtered scaffolds were screened by using Prodigal v2.60 (Hyatt et al., 2010), to extract the nucleotide coding sequences as well as their protein translation, and Prokka v1.12 (Seemann, 2014) for the standard GenBank annotation. CheckM v1.0.5 (Parks et al., 2015) allow us to verify genome completeness and contamination. The same pipeline was followed for the genome sequencing and analysis of strain *A. salicampi* KACC 19060^T. Additionally, the available draft genomes of the type strains of species of the genera *Aliifodinibius*, *Balneola*, *Fodinibius*, *Gracilimonas, Halalkalibaculum,* and *Rhodohalobacter* were used for comparative genomic analyses (Supplementary Table S1).

Isoelectric point calculation of predicted proteins was carried out using the "iep" program implemented in the EMBOSS package v6.5.7.0 (Rice et al., 2000), and plotted with the R package "ggplot2" v3.3.3 (Wickham, 2009). Meanwhile, the R package "fmsb" v0.7.3 (Nakazawa, 2022) was used for displaying the spider (or radar) graph of amino acid frequency. In both cases, colors were chosen from the wide selection of the R package "paletteer" v.1.4.0 (Hvitfeldt, 2021). Functional annotation was performed with the online tool BlastKOALA (Kanehisa et al., 2016). The KEGG Orthology numbers (KO), corresponding to KEGG pathways and KEGG modules, were assigned to the identified coding sequences, and then plotted in a heatmap to identify the most abundant functions. KO definitions were manually inspected and compared among the species of the family *Balneolaceae*.

Following the proposed minimal standards for prokaryotic taxonomy (Chun et al., 2018), we used different Overall Genome Relatedness Indexes (OGRIs) to determine the status of strain 1BSP15-2V2^T within the family *Balneolaceae*. Average Amino acid Identity (AAI) and Average Nucleotide Identity for orthologous sequences (orthoANI) were calculated using the Enveomics toolbox (Rodriguez-R and Konstantinidis, 2016) and OAU software v1.2 (Lee et al., 2016), respectively. For digital DNA–DNA hybridization (dDDH), the Genome-to-Genome Distance Calculator bioinformatic tool (GGDC v3.0) available from the Leibniz Institute DSMZ (Meier-Kolthoff et al., 2021) was used. For a more reliable determination of the relationship among the members of the *Balneolaceae*, a phylogenomic tree based on the core-proteome sequences of those strains was built. The orthologous protein clusters were identified after BLASTp v2.2.28+ search and extracted from the proteome by using the Markov Cluster Algorithm, as implemented in the Enveomics toolbox (Rodriguez-R and Konstantinidis, 2016). Muscle v3.8.31 (Edgar, 2004) was employed to align the orthologous sequences. The phylogeny was calculated based on 1,049 concatenated protein sequences with the approximately maximum-likelihood algorithm implemented in FastTreeMP v2.1.8 (Price et al., 2010), considering the Jones-Taylor-Thornton model of amino acid evolution (Jones et al., 1992). Furthermore, the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999) was used to check the reliability of each node. The visualization of the final topology was performed with the online tool iTOL v6 (Letunic and Bork, 2021), and the intersection of the shared orthologous proteins among the species under study was calculated with the R package "UpSetR" v1.4.0 (Conway et al., 2017).

2.5. Ecological distribution

With the aim to determine the presence and ecological distribution in hypersaline environments of strain 1BSP15-2V2^T, as well as those of the previously described members of the genera *Aliifodinibius*, *Balneola*, *Fodinibius*, *Gracilimonas*, *Halalkalibaculum*, and *Rhodohalobacter*, a fragment recruitment analysis was performed against different environmental metagenomic datasets (Supplementary Table S2) showing similar metadata features to avoid bias due to them. The rRNA gene sequences of each of the analyzed genomes were masked due to the highly conserved nature of those genomic regions. Subsequently, the ≥30 bp high-quality metagenomic reads were BLASTn v2.2.28+ searched against each of the query genomes.

2.6. Fatty acids composition and phenotypic characterization

For the determination of the fatty acid profile, strain 1BSP15-2V2^T was grown on MA medium supplemented with 10% (w/v) NaCl at 37°C for 5 days. Fatty acids were determined by gas chromatography with an Agilent 6850 system at the Spanish Type Culture Collection (CECT), Valencia, Spain. The protocol recommended by the MIDI Microbial Identification System was followed (Sasser, 1990), using the TSBA6 library (MIDI, 2008).

1 https://www.ezbiocloud.net/

2 https://github.com/cristinagalisteo/gitana

Colonial morphology and pigmentation of strain 1BSP15-2V2^T were observed in the same SMM medium used for isolation supplemented with 7.5% (w/v) SW solution at pH 7.5, after 5 days of incubation at 37°C. Cell morphology and motility were examined by phase contrast microscopy (Olympus CX41). Anaerobic growth was determined after incubation on the above-mentioned medium, conditions, and time, using the AnaeroGen™ system (Oxoid).

The range and optimal salt concentration, pH, and temperature requirements for strain 1BSP15-2V2^T were determined by monitoring the optical density. Absorbance at 600 nm was measured every 2 h for 72 h using an Infinite M Nano (Tecan, Grödig, Austria) microplate reader. The incubation was set at 37°C with linear shaking. For determination of the salinity requirements, the SMM liquid medium was supplemented with SW stock to give a final salt concentration of 0 (in this case only distilled water was added), 3, 5, 6, 7.5, 9, 10, 12, 14, 15, 16, 17, 19, 20, 22, and 25% (w/v). To determine pH values supporting growth, the pH was adjusted to 3.0, 4.0, 5.0, 6.0, 7.0, 7.5, 8.0, 9.0, and 10.0 using a buffered system. Optimal temperature and range were ascertained in the same SMM liquid medium at the optimal salinity and pH values for the growth of strain 1BSP15-2V2^T at different temperatures: 4, 13, 14, 15, 20, 28, 34, 37, 40, 42, 43, 44, and 45°C. In this case, the optical density was measured in a Spectronic 20D+ (ThermoSpectronics, Cambridge, United Kingdom).

In order to phenotypically characterize the new isolate, strain 1BSP15-2V2^T was routinely grown in SMM medium supplemented with 7.5% (w/v) salts at pH 7.5 and 37°C for 5 days. However, the reference strains *A. halophilus* KCTC 42497^T, *A. roseus* DSM 21986^T, *A. salicampi* KACC 19060^T, "*A. salipaludis*" KCTC 52855^T, *A. saliphilus* KACC 19126^T, and *A. sediminis* DSM 21194^T were routinely grown in the medium recommended by the culture collections, that is, R2A medium supplemented with 7.5% (w/v) SW solution at the same pH, temperature, and time than those for strain 1BSP15-2V2^T. All biochemical tests were accomplished under these culture conditions. Catalase activity was examined by adding a 3% (w/v) H_2O_2 solution to colonies (Cowan and Steel, 1965), and oxidase activity was determined with 1% (v/v) tetramethyl-p-phenylenediamine (Kovacs, 1956). The following tests were determined as described by Cowan and Steel (1965): hydrolysis of gelatin, starch, Tween 80, DNA, casein, and aesculin, production of indole, methyl red and Voges–Proskauer tests, Simmons' citrate, nitrate, and nitrite reduction, H_2S production, urease, and phenylalanine deaminase. Acid production from carbohydrates was determined using a modified phenol red base medium prepared with 7.5% (w/v) salts supplemented with 0.05% (w/v) yeast extract and 1% (w/v) of the filter-sterilized carbohydrate to be checked (Cowan and Steel, 1965; Ventosa et al., 1982). The medium described by Koser (1923), as modified by Ventosa et al. (1982), was used to assess the utilization of different substrates as sole carbon and energy sources or as sole carbon, nitrogen, and energy sources. The studied carbohydrates, alcohols, organic acids, and amino acids were added as filter-sterilized solutions to give a final concentration of 1 g L^{-1} for all of them, except for carbohydrates, whose final concentration was 2 g L^{-1}.

3. Results and discussion

3.1. Heavy metal contamination of the studied samples

To our best knowledge there are no reference studies on the concentration of heavy metals in the hypersaline soils from the Odiel Saltmarshes National Area, but it is known that the Odiel river waters and sediments contain high concentrations of arsenic, cadmium, copper, lead, and zinc due to unregulated industrial and mining activities in the past (Sainz et al., 2002, 2004). The sampled hypersaline soil presented a pH value of 8.4 and an electrical conductivity of 24.57 mS/cm. The concentrations of arsenic, cadmium, copper, lead, and zinc were (mg kg^{-1}): 14.86, 0.47, 106.40, 24.81, and 129.30, respectively. Arsenic, copper, and zinc presented values above the reference ranges for non-contaminated soils, according to the criteria of the Government of the region of Andalucía, where the Odiel Saltmarshes National Area is located (Consejería de Medio Ambiente, 1999), while cadmium and lead concentrations were within the reference ranges (Table 1).

3.2. The new isolated strain constitutes a novel species of the genus *Fodinibius*

The strain 1BSP15-2V2^T was cultured in pure culture after an extensive screening of the prokaryotic diversity carried out in hypersaline soil samples obtained from the Odiel Saltmarshes Natural Area (Huelva, Southwest Spain), where more than 4,000 isolates were investigated. The 1,455 bp-long 16S rRNA gene sequence of this strain showed the closest relationship with *A. roseus* DSM 21986^T and *A. saliphilus* ECH52^T (95.02% sequence identity, in both cases). Reconstruction of the 16S rRNA gene-based phylogeny by the maximum-likelihood algorithm clustered strain 1BSP15-2V2^T together to the species of the genus *Aliifodinibius* (Figure 1). The species *Fodinibius salinus* YIM D15^T was also clustered within the other species of the genus *Aliifodinibius*, which suggests that the two genera are closely related. However, the bootstrap value for the node harboring strain 1BSP15-2V2^T is below 70% and the topology differs sightly depending on the algorithms used (maximum-likelihood, maximum-parsimony, or neighbor-joining). A similar issue can be observed for the branch that accommodates *Fodinibius salinus* YIM D15^T. Therefore, the robustness of this tree is not strong enough as to infer reliable relationship between the species of the genera *Aliifodinibius* and *Fodinibius*, and strain 1BSP15-2V2^T, but offers an approach to their evolutionary history.

The draft genome sequence of strain 1BSP15-2V2^T (GCF_026229185.1) was assembled into 43 contigs, with a total size of 4,850,852 bp, and a DNA G+C content of 44.47 mol%. We also obtained the genome sequence of the closely related species *A. salicampi* KACC 19060^T (GCF_026228885.1), not previously available in public databases, which contains 13 contigs, slightly shorter size (3,937,193 bp), and a genomic G+C content of 42.75 mol%. The members of the genus *Aliifodinibius* possess genome sizes ranging from 3,583,276 bp

TABLE 1 Concentration of the most recurring heavy metal contaminants in the sampling area.

Heavy metal	Sample concentration (mg kg^{-1})	Reference range for non-contaminated soils (mg kg^{-1})
Arsenic	14.86	2–5
Cadmium	0.47	0.4–0.8
Copper	106.40	17–100
Lead	24.81	10–50
Zinc	129.30	10–70

The reference ranges refer to soils categorized as not contaminated in Andalucía (Spain), where the Odiel Saltmarshes National Area is located.

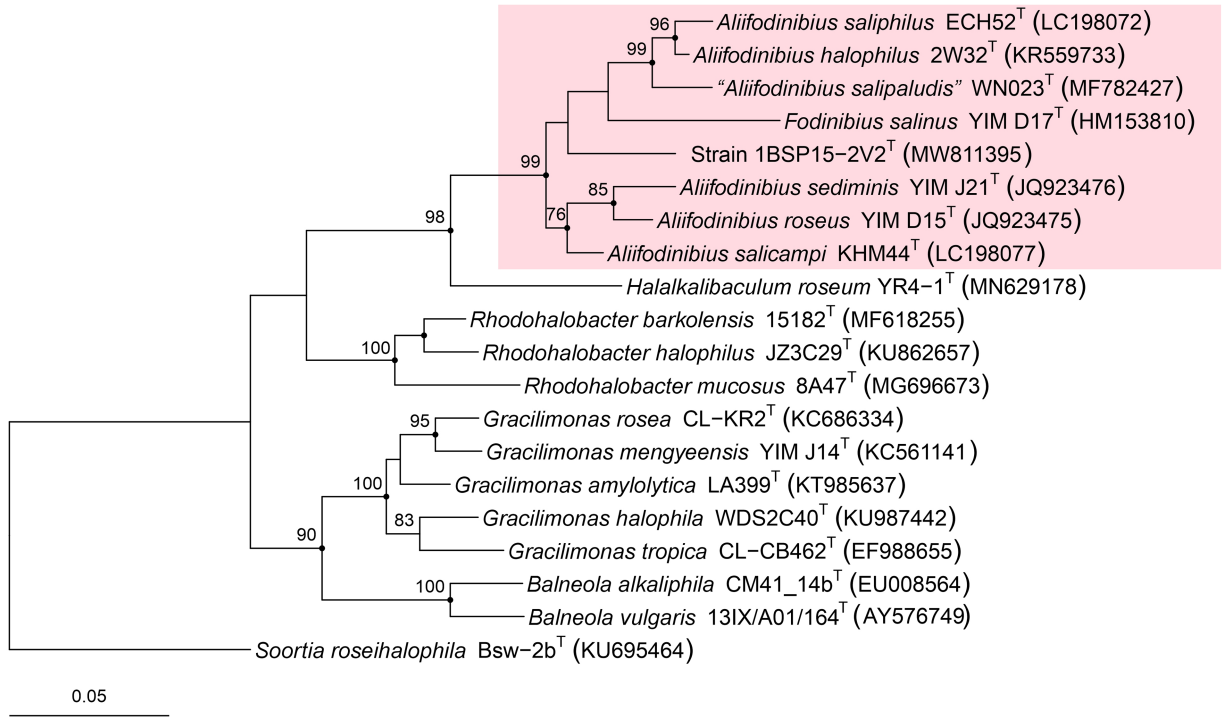


FIGURE 1
Maximum-likelihood phylogenetic tree based on the comparison of the 16S rRNA gene sequences showing the relationships among members of the family *Balneolaceae*. Bootstrap values equal or higher than 70% based on 1,000 pseudoreplicates are indicated at selected branch nodes. Filled circles indicate that the corresponding node was also obtained in the trees generated with the maximum-parsimony and neighbor-joining algorithms. The species *Soortia roseihalophila*, the closest relative to the family *Balneolaceae*, was selected as an outgroup. Bar, 0.05 substitutions per nucleotide position.

("*A. salipaludis*" WN023^T) to 5,082,244 bp (*A. roseus* DSM 21986^T) (Figure 2A; Supplementary Table S1). *Fodinibius salinus* DSM 21935^T has one of the smallest genomes of the members within the family *Balneolaceae* (2,861,751 bp), but with similar G+C percentage than those of the genera *Aliifodinibius*, *Gracilimonas*, *Halalkalibaculum*, and *Rhodohalobacter*. On the contrary, the species of the genus *Balneola*, the type genus of the family *Balneolaceae*, have smaller average genome size than the rest of members of the family (except the genus *Fodinibius*), as well as the lowest G+C content (<40 mol%) (Figures 2A,B; Supplementary Table S1).

The approximated maximum-likelihood phylogenomic tree, based on the concatenation of 1,049 translated core genes, showed that the strain 1BSP15-2V2^T clustered again with the species of the genus *Aliifodinibius* (Figure 3). This clustering was supported by a 100% bootstrap value, indicating the robustness of the group, which was unclear in the 16S rRNA gene-based phylogeny (Figure 1). Furthermore, *Fodinibius salinus* DSM 21935^T, the only described species of this genus, also clustered with 100% branch support together to the species of the genus *Aliifodinibius*, rather than forming a separated and independent branch. This fact seems to indicate that members of the genera *Aliifodinibius* and *Fodinibius* could be actually merged into the single genus *Fodinibius*.

Concerning the OGRIs analysis, AAI values between strain 1BSP15-2V2^T and the other members of the genus *Aliifodinibius* ranged from 72.1% (with *A. sediminis* DSM 21194^T) to 68.1% (with *A. halophilus* 2W32^T) (Figure 4). These values were above the 65–72% threshold suggested to circumscribe members of the same genus (Konstantinidis and Tiedje, 2007; Konstantinidis et al., 2017). Similar AAI values were obtained for *Fodinibius salinus* DSM 21935^T, which varied from 72.9% (with *A. halophilus* 2W32^T) to 68.1% (with *A. roseus* DSM 21986^T) with respect to the species of the genus *Aliifodinibius*, but were equal or lower than 64.9% for *Halalkalibaculum roseum* YR4-1^T and the remaining species of the genera of the family *Balneolaceae* (Figure 4). Furthermore, orthoANI percentages among the analyzed genomes within the *Balneolaceae* were all below the 95% threshold established for species differentiation (Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014), including strain 1BSP15-2V2^T, which shares the highest similarity with *A. roseus* DSM 21986^T and *A. sediminis* DSM 21194^T (72.7% for both species) (Figure 5). These data undoubtedly indicate that all strains under study represented different species. The same conclusion can be inferred from dDDH results, where all values are far below the 70% cutoff defined for species delineation (Stackebrandt and Goebel, 1994; Auch et al., 2010) (Figure 5).

The distribution of core, soft-core, and accessory genomic content among the 20 genomes of the family *Balneolaceae* investigated here are shown in Figure 6. All the studied genomes of the species of the family *Balneolaceae* share a total of 1,049 core genes. The species "*Aliifodinibius salipaludis*" WN023^T possesses the highest number of unique genes (270), although it has the smaller genome size (3.58 Mb) and the lower number of CDS (3,143) (Supplementary Table S1). This could be related with the habitat where "*Aliifodinibius salipaludis*" WN023^T dwells, since to date it is the only member of the family isolated from saline-alkaline soil instead of hypersaline sediments or salterns (Zhao et al., 2020). However, our new strain 1BSP15-2V2^T was also isolated from saline soils and its exclusive genetic material is not high enough (12 unique genes) as to be displayed among the most frequent intersections (Figure 6). Strain 1BSP15-2V2^T shares a total of 61 genes exclusively with the type species of the genus *Aliifodinibius*, *A. roseus* DSM 21986^T, and 50 additional genes if the species

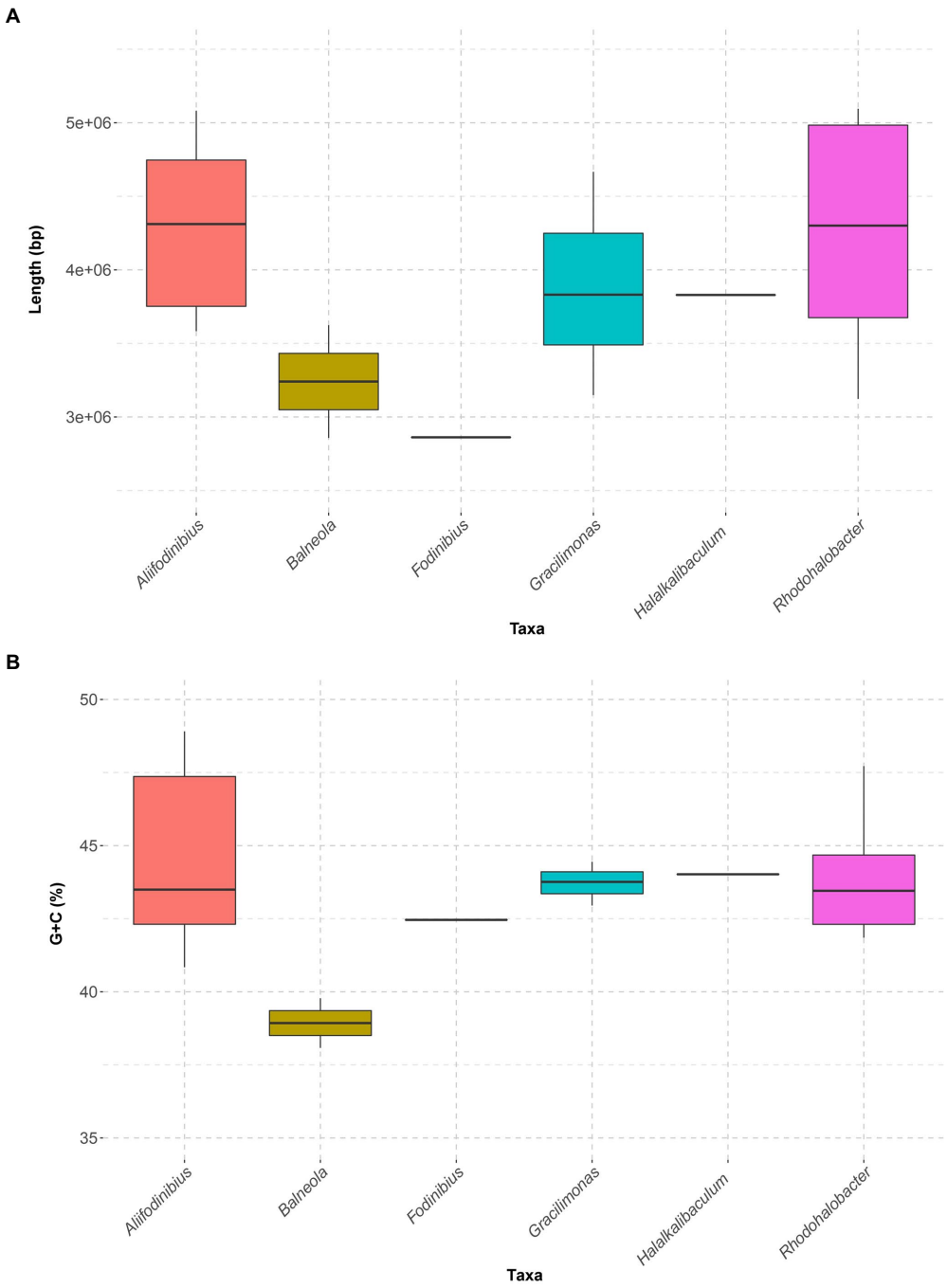


FIGURE 2
Boxplots summarizing genome size **(A)** and G+C content **(B)** of the genera of the family *Balneolaceae*, sorted alphabetically. The family *Balneolaceae* includes genera with very different genomes sizes, but their G+C percentages are mostly stable around 43 mol%, except for the type genus *Balneola*, below 40 mol%.

A. sediminis DSM 21194^T is also considered. In any case, strain 1BSP15-2V2^T shares a higher number of genes with the species of the genus *Aliifodinibius* than with any other of the remaining genera of the family *Balneolaceae*. Moreover, a large cluster of 88 genes that are present solely within the genera *Aliifodinibius* and *Fodinibius*, as well as the genus *Halalkalibaculum*, points to their closed evolutionary relationships (Figure 6). Similarly, 97 genes are exclusively shared between the members of the genus *Rhodohalobacter*. In summary, we observed that the most closely related the species are the highest number of genes they share. Further differences in the protein functions are described in Section 3.4.

3.3. Chemotaxonomic and phenotypic characterization support the placement of the new isolate within the genus *Fodinibius*

The major fatty acids determined for strain 1BSP15-2V2^T were iso-$C_{15:0}$ (36.4%), $C_{16:1}$ ω7c and/or $C_{16:1}$ ω6c (24.7%), and iso-$C_{17:1}$ ω9c and/or 10-methyl $C_{16:0}$ (10.8%). The profile of this strain is similar to those of the type strains of the species of the genera *Aliifodinibius* and *Fodinibius* (Supplementary Table S3). No clear differential profiles were

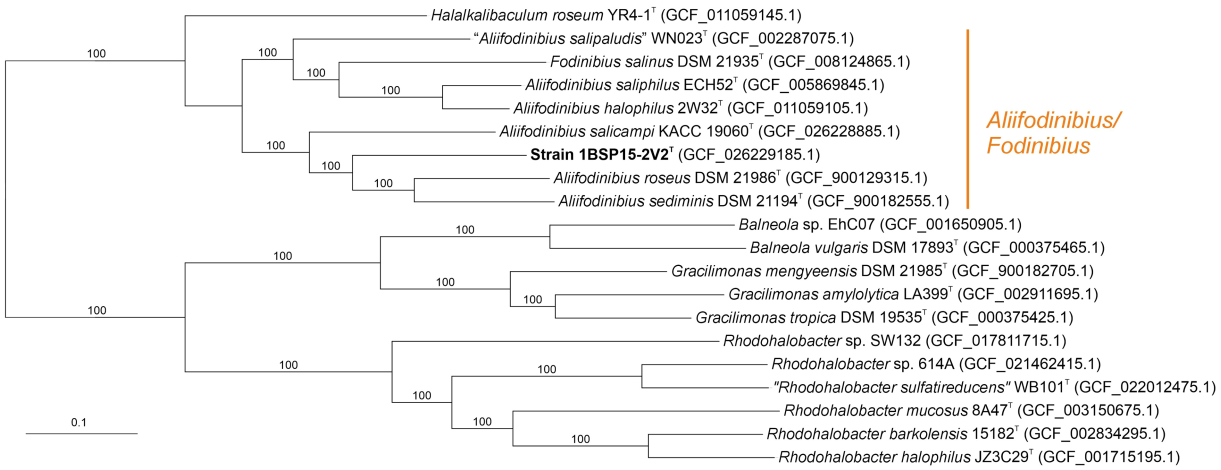


FIGURE 3
Approximately maximum-likelihood phylogenomic tree based on 1,049 concatenated core protein sequences of members of the family *Balneolaceae*. Bootstrap values ≥70% are indicated above the branches. Bar, 0.1 substitutions per nucleotide position.

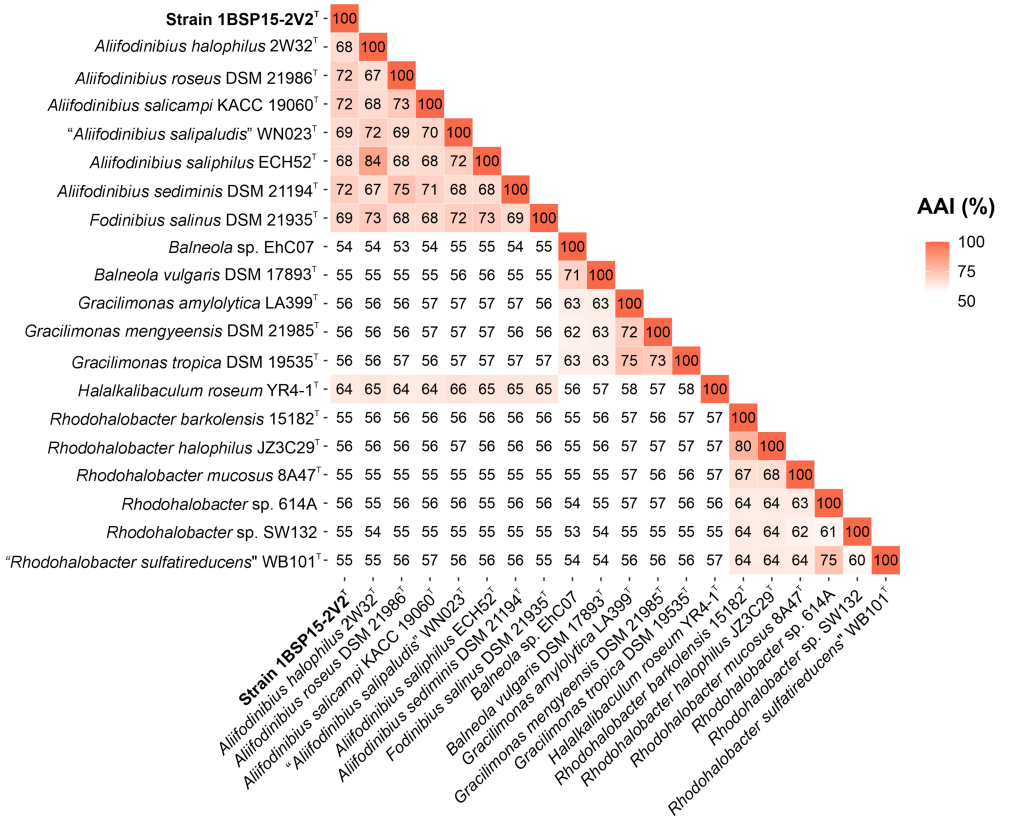


FIGURE 4
AAI matrix among the 20 studied genomes of the family *Balneolaceae*. AAI values among species of the cluster *Fodinibius-Aliifodinibius* were all higher than the threshold (65%) accepted for genus delineation. Isolate 1BSP15-2V2^T shared values above 68% with species of the genera *Aliifodinibius* and *Fodinibius*, and equal or lower than 64% with other members of the family *Balneolaceae*, hinting at the affiliation of the novel strain to the genera *Aliifodinibius*/*Fodinibius*.

obtained for species of the genera *Aliifodinibius* and *Fodinibius*, supporting the hypothesis that they all belong to the same genus.

The strain 1BSP15-2V2^T formed irregular, opaque, and orange-red pigmented colonies. Cells were Gram-stain-negative, non-motile, non-endospore-forming rods, with a size of 0.05–0.10 × 0.20–0.30 μm. The detailed phenotypic characteristics of strain 1BSP15-2V2^T are included in the new species description. Besides, additional phenotypic data for the previously described species of the genera *Aliifodinibius* and *Fodinibius* included in this study are shown in Supplementary Table S4. The morphological and many of the biochemical characteristics of the isolate 1BSP15-2V2^T agree with those reported for the other members of the genera *Aliifodinibius* and *Fodinibius*, and thus, support the

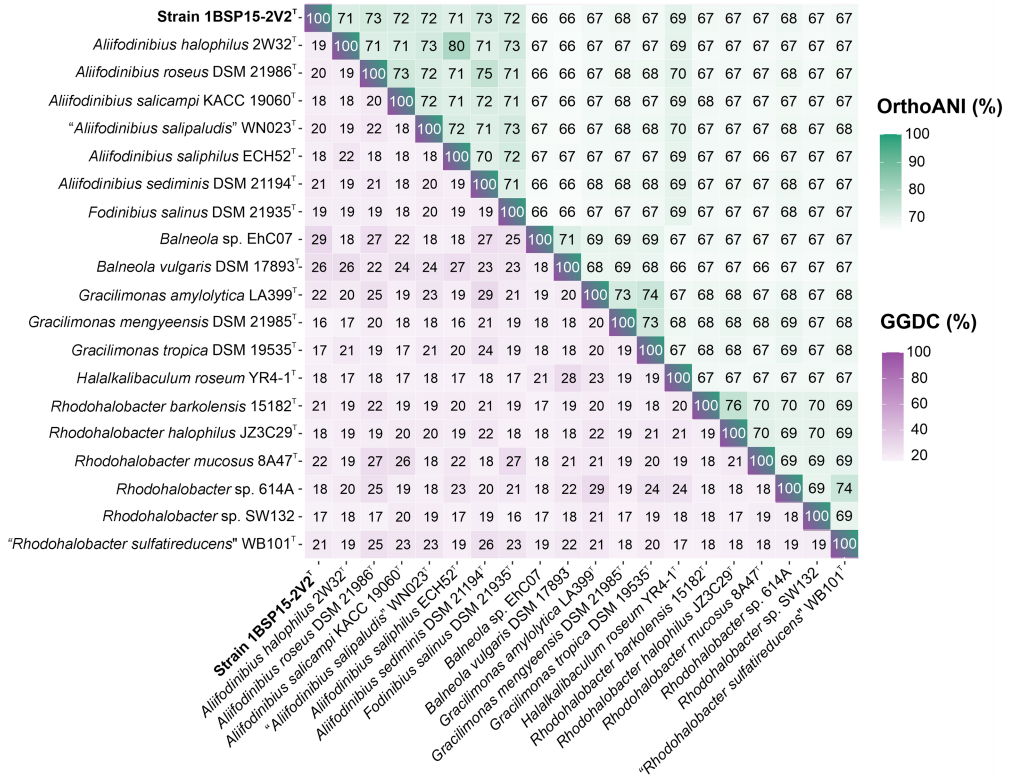


FIGURE 5
OrthoANI (upper triangle) and GGDC (lower triangle) matrix among the 20 studied genomes of the family *Balneolaceae*. All GGDC (purple) and orthoANI (green) values were below the cutoff for species delineation, indicating that all the genomes belong to different species.

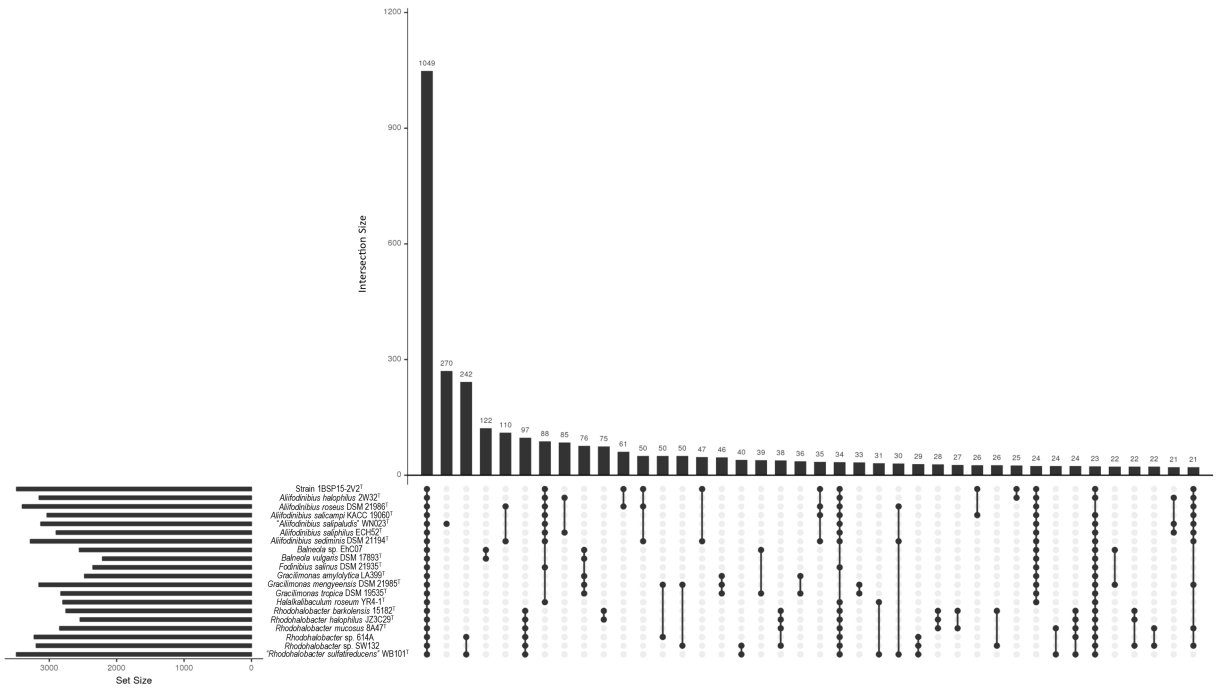


FIGURE 6
Upset plot showing the proteins shared among the 20 *Balneolaceae* genomes under study. Only the 40 most frequent intersecting sets are displayed. From a pan genome of 8,113 translated genes, 1,049 were encoded in all the genomes (core genome). *"Aliifodinibius salipaludis"* WN023^{T} harbored the highest number of unique translated genes (270). The closely related genera *Aliifodinibius*, *Fodinibius*, and *Halalkalibaculum* shared 88 translated genes.

FIGURE 7
KO numbers grouped by KEGG categories annotated for the representative genomes of the family *Balneolaceae*. KEGG pathways are sorted according to their abundance in the strain 1BSP15-2V2^T and only the 50 most abundant are displayed. The most prominent pathways detected in those genomes belonged the following KEGG global categories: signal transduction (dark blue), energy metabolism (blue), translation (orange), nucleotide metabolism (yellow), carbohydrate metabolism (black), membrane transport (dark red), cell motility (light blue), amino acid metabolism (light red), cellular community (light green), glycan biosynthesis and metabolism (dark green), lipid metabolism (green), metabolism of cofactors and vitamins (pink), and biosynthesis of other secondary metabolites (purple). 1. Strain 1BSP15-2V2^T; 2. *Aliifodinibius halophilus* 2W32^T; 3. *Aliifodinibius roseus* DSM 21986^T; 4. *Aliifodinibius salicampi* KACC 19060^T; 5. "*Aliifodinibius salipaludis*" WN023^T; 6. *Aliifodinibius saliphilus* ECH52^T; 7. *Aliifodinibius sediminis* DSM 21194^T; 8. *Fodinibius salinus* DSM 21935^T; 9. *Balneola vulgaris* DSM 17893^T; 10. *Gracilimonas amylolytica* LA399^T; 11. *Gracilimonas mengyeensis* DSM 21985^T; 12. *Gracilimonas tropica* DSM 19535^T; 13. *Halalkalibaculum roseum* YR4-1^T; 14. *Rhodohalobacter barkolensis* 15182^T; 15. *Rhodohalobacter halophilus* JZ3C29^T; 16. *Rhodohalobacter mucosus* 8A47^T.

placement of strain 1BSP15-2V2^T within the genus *Fodinibius* (which should also include the described species of *Aliifodinibius*). Besides, strain 1BSP15-2V2^T possesses several features that allow to distinguish it from the close-related species of the genera *Aliifodinibius* and *Fodinibius*, namely, the ability to use several amino acids as sole carbon, nitrogen, and energy sources (Supplementary Table S4).

3.4. Functional genomic analysis reveals the potential ecological role of the new strain as a biotin producer

BlastKOALA annotation identified a total of 1,986 different KO numbers across the 16 available genomes of the type strains of species of the family *Balneolaceae*. Those KO numbers were classified according to KEGG pathway database, and the 50 most abundant pathways are shown in Figure 7. Strain 1BSP15-2V2^T seems to exhibit a functional profile like that of the already described species of the genera *Aliifodinibius*/*Fodinibius* and the rest of the genera within the family *Balneolaceae*. Nevertheless, it must be noted that 22 KO numbers were only present in strain 1BSP15-2V2^T and not in any other member of the family. Among those unique functional orthologs, we detected *frdA*, *frdB*, *frdC*, and *frdD* genes in the genome of strain 1BSP15-2V2^T, involved in the conversion of fumarate to succinate (Supplementary Table S5).

The most prevalent pathways of strain 1BSP15-2V2^T were included into energy, carbohydrate, and amino acid metabolism KEGG global categories. All the species of the genera *Aliifodinibius* and *Fodinibius* harbored genes involved in pyruvate metabolism, including the pyruvate oxidation pathway comprising KO numbers K00161 (*pdhA*, pyruvate dehydrogenase E1 component α-subunit), K00162 (*pdhB*, pyruvate

dehydrogenase E1 component β-subunit), K00627 [*pdhC*, pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)], and K00382 (*pdhD*, dihydrolipoamide dehydrogenase) (Supplementary Table S5). However, only strain 1BSP15-2V2^T, *A. halophilus* KCTC 42497^T, and *A. roseus* DSM 21986^T were able to use pyruvate as sole carbon and energy source when tested experimentally (Supplementary Table S4). The other genera of the family *Balneolaceae* showed the same *in-silico* metabolism for pyruvate, but it was not confirmed in this study under laboratory conditions.

Strain 1BSP15-2V2^T and the species of the genus *Aliifodinibius* possessed genes linked to flagellar assembly (Figure 7; Supplementary Table S5), in contrast to the genera *Fodinibius*, *Balneola*, *Halalkalibaculum*, and *Rhodohalobacter*, as well as the species *Gracilimonas amylolytica* LA399^T and *G. tropica* DSM 19535^T, although the route was incomplete in all the cases. This finding agrees with the fact that none of the descriptions of the species of the genus *Aliifodinibius* reported the cellular motility (Wang et al., 2013; Xia et al., 2016; Cho et al., 2017; Cho and Whang, 2020; Zhao et al., 2020). Cells of 1BSP15-2V2^T did not show motility either in the present study. The presence of a broken pathway for flagellar motility seems to indicate that the ancestor of the family *Balneolaceae* may have had that feature, which was gradually lost as it was not required for their survival in the nature. Major differences between the functional profile of species of the genera *Aliifodinibius* and *Fodinibius* are related to bacterial motility, that is flagellar assembly and flagellar motor switch (two component system). This supports the theory proposed above about the loss of motility-related genes, which might have happened quicker in *F. salinus*. Other than those two pathways, *F. salinus* presents a mostly similar profile to the species of the genus *Aliifodinibius*. We can find minor differences between the genomes of the studied species of *Aliifodinibius/Fodinibius*, such as less abundance of functions related to oxidative phosphorylation and purine metabolism in the species "*Aliifodinibius salipaludis*" and *F. salinus*, though similar between them, and an oscillation in the number of genes encoding enzimatic activities in the pentose phosphate pathway. In any case, all the studied genomes are uncompleted, and their range size is broad among the members of the *Aliifodinibius/Fodinibius*. Hence, we can only speculate about the presence or lack of functions, as it may be due to a deficient annotation or assembly with the current bioinformatic tools, or to an adaptation of the bacteria to their particular environment.

The biosynthesis pathway for biotin (vitamin B_7) appears to be present in the species of the family *Balneolaceae*. Though the understanding of biotin biosynthesis is still fragmentary, previous studies (Cronan, 2018; Manandhar and Cronan, 2018) propose two stages: the synthesis of a pimelate moiety, and the assembly of the bicyclic rings of the biotin molecule. The latter is mostly conserved, with four associated genes: *bioF*, *bioA*, *bioD*, and *bioB*, all of them present in strain 1BSP15-2V2^T (Figure 8A; Supplementary Table S5). These genes were also detected in *A. halophilus*, *A. saliphilus*, *Fodinibius salinus, Halalkalibaculum roseum*, and *Balneola vulgaris* (Figure 8B; Supplementary Table S5). Only *bioB* and *bioF* were present in *A. salicampi* and "*A. salipaludis*." However, the *bioF* gene, which performs the first step of the assembly of the bicyclic rings, can be found in all the studied species of the family *Balneolaceae*. On the contrary, there are at least three known pathways to accomplish the first stage of biotin biosynthesis: the BioC-BioH and the BioI-BioW pathways, studied in *Escherichia coli* and *Bacillus subtilis*, respectively (Lin and Cronan, 2011), and the most recently proposed BioZ pathway (Zhang et al., 2021). However, the BioC-BioH pathway seems to be the most prevalent of these three (Feng et al., 2014; Wei et al., 2019). Genes *bioC*, *fabF*, *fabG*, *fabZ*, and *fabI* from the BioC-BioH pathway are found in all the species of the genus *Aliifodinibius*, except the *bioC* gene lacking in *A. roseus*. These genes are also present in strain 1BSP15-2V2^T and *Fodinibius salinus*, as well as in most species of the family *Balneolaceae*. Noteworthy, all studied genomes of the family seem to lack the *bioH* gene, which performs the last step of the BioC-BioH pathway. However, homologous genes have been found in other organisms as replacements (Rodionov et al., 2002), such as *bioG* in *Haemophilus* sp. (Shi et al., 2016), *bioK* in *Synechococcus* sp. (Shapiro et al., 2012), *bioJ* in *Francisella* sp. (Wei et al., 2019), and *bioV* in *Helicobacter* sp. (Bi et al., 2016). None of these alternative genes were found in the genomes under study, but we cannot discard the possibility that the family *Balneolaceae* harbors a different homologous gene that hydrolyzes the pimeloyl-ACP methyl ester into pimeloyl-ACP in the last step of the BioC-BioH pathway (Lin and Cronan, 2011). On the other hand, all the species of the family *Balneolaceae* present *birA, accB*, and *accC* genes in their genomes. The protein BirA has a dual function: attach biotin to AccB protein and down-regulate biotin biosynthesis in the absence of free AccB acceptor protein. AccB is a carrier protein that forms a complex with the AccC subunit (Choi-Rhee and Cronan, 2003), so both macromolecules indirectly regulate the transcription of the biotin operon. Low levels of available AccB protein, that is when AccB protein is biotinylated or unbiotinylated but bonded to AccC, avoid biotin transfer from BirA to AccB, resulting in higher levels of BirA ligand bound to biotin. The accumulation of BirA-bound biotin represses biotin operon transcription (Abdel-Hamid and Cronan, 2007). In summary, the genome of 1BSP15-2V2^T, as well as those of the rest of the species of the family *Balneolaceae*, contain genes involved in the regulation of biotin biosynthesis. With a single exception, they also have the complete set of *bioC*, *fabF*, *fabG*, *fabZ*, and *fabI* genes that catalize the first steps of biotin biosynthesis, albeit all of them lack the *bioH* gene, which encodes the last enzyme of this first stage. Since several *bioH* homologs have been reported in the literature, the species of the family *Balneolaceae* are hypothesized to harbor one of those not-yet described homologous genes. Considering that the more conserved second stage (*bioF*, *bioA*, *bioD*, and *bioB* genes) is only present in the genomes of strain 1BSP15-2V2^T and other five members of the family, the complete (or almost complete) biotin biosynthesis pathway was only identified in the genomes of strain 1BSP15-2V2^T, *A. halophilus*, *A. saliphilus*, *Fodinibius salinus*, *Halalkalibaculum roseum*, and the type species of the family, *Balneola vulgaris*, whereas it was missing from the genera *Gracilimonas* and *Rhodohalobacter*. Since the biotin biosynthetic pathway is well conserved across biotin producers (Lin and Cronan, 2011), it may be explained the presence of at least some of those genes within all described species of the family *Balneolaceae*, even if the pathway is uncompleted. Biotin plays an important role in fatty acid synthesis, amino acid metabolism, and gluconeogenesis in bacteria, archaea, and eukaryotes (Knowles, 1989; Jitrapakdee and Wallace, 2003). The biosynthesis of biotin is exclusive of some prokaryotes, fungi, and plants but it is an essential cofactor for mammals too, which depend on the exogenous supply from the diet or the microbiota (Cronan et al., 2014). For example, vitamin B_7 deficiency in humans is associated with neurological diseases (León-Del-Río, 2019). It is a high-costing metabolic pathway, especially the steps carried out by the proteins coded by *bioC*, *bioA* and *bioB* genes. Hence, most organisms prefer an exogenous uptake of this cofactor (Sirithanakorn and Cronan, 2021). The isolated species could play a role in its ecosystem by suppling biotin to those organisms that lack the mechanisms for *de novo* synthesis. Besides, it might be further explored as an interesting source of biotin for human deficiency with biotechnological application. Nowadays, biotin for pharmacological purposes is chemically synthetize since bacterial biosynthesis has given unsuccessful results to date (Sirithanakorn and Cronan, 2021). This is, partly, due to the uncomplete understanding of the biotin pathway in prokaryotes. Thus, the genomic identification of this route in strain 1BSP15-2V2^T will allow its targeting as a possible candidate for further studies of the biotin biosynthesis.

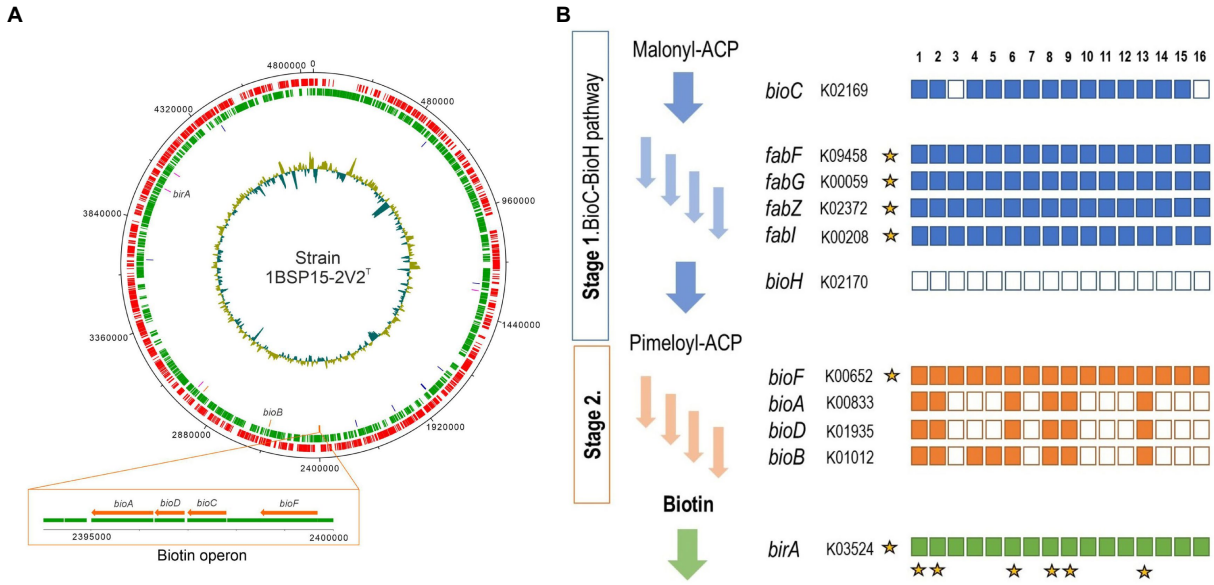


FIGURE 8
(A) Circular genome map of strain 1BSP15-2V2^T annotated by Prokka. The graph was generated with "DNAPlotter" (Carver et al., 2009). Circles indicate, from outside inwards: coding sequences in the leading strand (red); coding sequences in the lagging strand (green); coding sequences related to the biotin biosynthesis; G+C content. Biotin pathway-related genes are displayed in colors: stage 1 of biosynthesis, blue; stage 2 of biosynthesis, orange; regulation and transport, pink. A diagram of the biotin operon arrangement is shown in the box. Genes *bioA*, *bioD*, *bioC*, and *bioF* are located in the same region of the genome with a hypothetical protein encoded between *bioF* and the rest of the *bioADC* genes. The organization in form of operon allows the cotranscript of their proteins in the correct stoichiometry (Sirithanakorn and Cronan, 2021). *bioB* and *birA* genes are placed in a different region of the chromosome. **(B)** Biotin biosynthesis pathway for members of the family *Balneolaceae*. Although almost the complete BioC-BioH pathway was detected in most of the genomes, *bioH* gene was missing for all of them. Stars indicate either that the gene is present in all the genomes or that the pathway (except *bioH* gene) is complete for the corresponding genome. 1. Strain 1BSP15-2V2^T; 2. *Aliifodinibius halophilus* 2W32^T; 3. *Aliifodinibius roseus* DSM 21986^T; 4. *Aliifodinibius salicampi* KACC 19060^T; 5. "*Aliifodinibius salipaludis*" WN023^T; 6. *Aliifodinibius saliphilus* ECH52^T; 7. *Aliifodinibius sediminis* DSM 21194^T; 8. *Fodinibius salinus* DSM 21935^T; 9. *Balneola vulgaris* DSM 17893^T; 10. *Gracilimonas amylolytica* LA399^T; 11. *Gracilimonas mengyeensis* DSM 21985^T; 12. *Gracilimonas tropica* DSM 19535^T; 13. *Halalkalibaculum roseum* YR4-1^T; 14. *Rhodohalobacter barkolensis* 15182^T; 15. *Rhodohalobacter halophilus* JZ3C29^T; 16. *Rhodohalobacter mucosus* 8A47^T.

In addition to the biotin biosynthesis, strain 1BSP15-2V2^T also presents functional orthologs related to other cofactors and vitamins, such as folate, pantothenate and CoA biosynthesis, and metabolism of thiamine, nicotinate, and nicotinamide (Figure 7; Supplementary Table S5). On the other hand, only the species of the genera *Aliifodinibius* and *Fodinibius*, including the strain 1BSP15-2V2^T, showed genes involved in the cobalamin (vitamin B_{12}) transport system (*btuC*, *btuD*, and *btuF*), and all members of the family *Balneolaceae* harbors *btuB*, a vitamin B_{12} transporter. Moreover, a previous study has reported the presence of cobalamin-dependent enzymes in the metagenomic dataset SMO2, obtained from hypersaline soils at the Odiel Saltmarshes National Area (Durán-Viseras et al., 2019). These enzymes corresponded to the KO numbers K00525 (*nrdA*, ribonucleoside-diphosphate reductase α-chain), K01848 (*mcmA1*, methylmalonyl-CoA mutase, N-terminal domain), and K00548 (*metH*, 5-methyltetrahydrofolate–homocysteine methyltransferase). Two of them, K00525 and K00548, along with K01847 (*MUT*, methylmalonyl-CoA mutase) and K11942 (*icmF*, isobutyryl-CoA mutase), were found in the genomes of strain 1BSP15-2V2^T and members of the family *Balneolaceae*. However, none of them encoded the cobalamin biosynthesis pathway, which is not surprising, as it is a high-cost metabolic mechanism (Morris, 2012; Doxey et al., 2015). Thus, it seems to indicate that strain 1BSP15-2V2^T as well as the rest of the family *Balneolaceae* depend on an external source of vitamin B_{12}, such as species of the haloarchaeal genera *Halobacterium* (Woodson et al., 2003) and *Halonotius* (Durán-Viseras et al., 2019).

3.5. Metalloresistant abilities to bear up against contaminated soils

As stated above, cadmium and lead concentrations in our sampled hypersaline soils were within the reference ranges for a non-contaminated soil, whereas arsenic, copper, and zinc concentrations were above the maximum thresholds (Table 1). Therefore, we can classify our samples as arsenic, copper, and zinc contaminated soils.

Several genes related to heavy metal resistance were found in the genome of strain 1BSP15-2V2^T. The *ars* operon (KO numbers K03325, K03741, and K03892) (Supplementary Table S5) is responsible for reduction of arsenate, As (V), to arsenite, As (III), and its excretion from the cell (Chauhan et al., 2019; Islam et al., 2022). Arsenic resistance is believed to be a widespread mechanism among prokaryotes due to the high concentration of this metalloid in the early period of the Earth (Chen et al., 2020). On the other hand, the CzcCBA efflux system (K15725, K15726, K15727, and K16264) (Supplementary Table S5) provides resistance to cadmium, copper, and zinc (Legatzki et al., 2003). This pump has been found in many metallotolerant Gram-negative bacteria (Manara et al., 2012; Khan et al., 2015; Alquethamy et al., 2021). Furthermore, the *zntA* gene (K01534) (Supplementary Table S5) for Zn^{2+}/Cd^{2+}-exporting (Rensing et al., 1997; Noll and Lutsenko, 2000) was also detected in the genome of strain 1BSP15-2V2^T. Previous studies have demonstrated the ability of CzcCBA efflux system to efficiently remove zinc and cadmium from the cytoplasm (Legatzki et al., 2003). Thus, the existence of an additional *zntA* gene-encoding pump in strain 1BSP15-2V2^T seems to

be an adaptative mechanism to cope with the harsh metal conditions of the Odiel Saltmarshes National Area hypersaline soils. Nevertheless, it must be noted that our predictions are entirely based on *in-silico* functional annotation analysis and, therefore, further laboratory research is necessary to confirm the heavy metal resistance of this bacterium.

3.6. Genomic information unveils a *salt-out* osmoregulation strategy

Two main osmoregulatory mechanisms allow prokaryotes to dwell in high-salt concentration environments. The *salt-in* strategy is typical of extreme halophiles, mostly haloarchaea (Youssef et al., 2014) and some bacteria, such as those of the widely studied genus *Salinibacter* (Antón et al., 2002). To maintain the structure and activity of their proteins under high cytoplasmic ion concentrations these microorganisms possess a more acidic proteome (Oren, 2008). However, this kind of proteome has also been described in microorganisms with a *salt-out* mechanism (Elevi Bardavid and Oren, 2012; Oren, 2013), which is the most extended strategy across prokaryotes allowing adaptation to wider variations in the surrounding osmotic pressure (Oren, 2008).

All genomes under study have a similar isoelectric profile (Figure 9A) and amino acid frequency distribution (Figure 9B), showing a less acidic proteome than those of the extreme haloarchaeon *Haloarcula vallismortis* and the bacterium *Salinibacter ruber*. Concerning the genome of strain 1BSP15-2V2^T, it harbors Ktr potassium importer genes (K03498, K03499) that may contribute to a fast response against osmotic stress, as it has been previously observed in *Bacillus subtilis* (Holtmann et al., 2003; Hoffmann and Bremer, 2016) and *Synechocystis* sp. (Zulkifli et al., 2010). Another ion homeostasis-related protein annotated in the genome of strain 1BSP15-2V2^T (K03313) is the NhaA Na^+/H^+ antiporter, which plays a relevant role to avoid the toxicity of high cytoplasmatic Na^+ concentrations (Krulwich et al., 2011; Patiño-Ruiz et al., 2022). However, potassium and sodium fluxes are impractical mechanisms to deal with sustained high osmolarity (Hoffmann and Bremer, 2017). For that reason, strain 1BSP15-2V2^T also presents compatible solute transporters of the Opu family (K05845, K05846, K05847), involved in the ATP-dependent uptake of a wide variety of compatible solutes into the cytoplasm, such as glycine betaine, choline, and proline betaine (Hoffmann and Bremer, 2017; Teichmann et al., 2018). Sudden decrease of osmolarity could also trigger stress in bacteria. Accordingly, strain 1BSP15-2V2^T encodes Msc mechanosensitive channels (K03282, K03442) which release ions and organic molecules if an osmotic down shock occurs (Booth and Blount, 2012; Booth, 2014). The different strategies to balance the osmotic pressure encoded in the genome of strain 1BSP15-2V2^T (Supplementary Table S5) indicate that this microorganism can thrive on a range of salt concentrations, which is consistent with the physiological traits tested under laboratory conditions (Supplementary Table S4). Consequently, strain 1BSP15-2V2^T can be defined as a moderately halophilic bacterium with a *salt-out* osmoregulation mechanism.

3.7. Members of the family *Balneolaceae* prefer saline soil versus aquatic hypersaline environments

Genomic fragment recruitment analysis was performed against nine metagenomic datasets (Supplementary Table S2) to assess the environmental distribution of strain 1BSP15-2V2^T and the other species of the family *Balneolaceae*. SMO1 and SMO2 metagenomes come from the same hypersaline soils used in this study. The other seven metagenomic databases originated from hypersaline aquatic environments (salterns and lakes) were included in this study to determine the distribution of the new taxon in different saline habitats (soils and water). The selected datasets ranged from 13 to 37% (w/v) salt concentration. Of these, SS13 metagenome has the closest salinity (13% [w/v] salts) to the optimal salt concentration defined for strain 1BSP15-2V2^T (9% [w/v] salts). Our results show few reads (<0.01%) recruited at ≥95% identity from all species of the genera *Aliifodinibius*/*Fodinibius* for all metagenomic datasets. As nucleotide sequences are being compared, the 95% ANI threshold is considered for species delineation. Hence, the isolated species seems to be not highly abundant in the studied environments. However, soil and aquatic environments seem to follow different patterns, with a huge read recruitment at 80–90% identity for SMO1 and SMO2 (soils) datasets (Figure 10) not observed in water metagenomes (Figure 11). Barco et al. (2020) proposed an ANI cutoff value of 73.98% for genus demarcation, and thus, on this basis, it would suggest the existence of a taxon or group of taxa very closely related to the genera *Aliifodinibius* and *Fodinibius* (that could represent different species or even genera) abundant in those hypersaline soil environments. On the contrary, the scarce recruitment at 80–90% identity for the aquatic hypersaline databases might indicate that closed relatives of the genera *Aliifodinibius*/*Fodinibius* have a preference for hypersaline soils over their aquatic counterparts.

4. Conclusion

Phylogenetic and genomic evidence, fatty acid profiles, and phenotypic characteristics of the microorganisms under study show unequivocally that the species of the genera *Aliifodinibius* and *Fodinibius* should be placed into a single genus. We propose the reclassification of the members of the genus *Aliifodinibius* into the genus *Fodinibius*, since the genus name *Fodinibius* was validly published first and, thus, it has priority over the genus name *Aliifodinibius*, according to the International Code of Nomenclature of Prokaryotes (Parker et al., 2019). Additionally, strain 1BSP15-2V2^T represents a new species of the genus *Fodinibius*, for which the name *Fodinibius salsisoli* sp. nov. is proposed. The description of the new species, as well as the reclassification of the species of *Aliifodinibius* into the genus *Fodinibius* and, consequently, the emended description of the genus *Fodinibius* are enclosed below.

Description of *Fodinibius salsisoli* sp. nov.

Fodinibius salsisoli (sal.si.so'li. L. masc. adj. *salsus* salty; L.; neut. n. *solum* soil; N.L. gen. n. *salsisoli* of salty soil).

Cells are Gram-stain-negative, non-endospore-forming, non-motile rods with a size of 0.05–0.10 × 0.20–0.30 µm. Colonies are irregular, opaque, and orange-red pigmented in SMM medium supplemented with 7.5% (w/v) salts after 5 days of incubation at 37°C. Grows between 3 and 20% (w/v) salt concentration (optimally at 9% [w/v]), pH 5.0–8.0 (optimally at pH 6.0), and 14–43°C (optimally at 37°C). Growth is not observed under anaerobic conditions. Catalase positive and oxidase negative. Reduces nitrate and nitrite. Positive for methyl red test but negative for Voges-Proskauer test, indicating that it uses the mixed acid pathway for glucose fermentation. Simmons' citrate and phenylalanine deaminase are positive. Aesculin is hydrolyzed, but casein, DNA, gelatin, Tween 80, starch, and

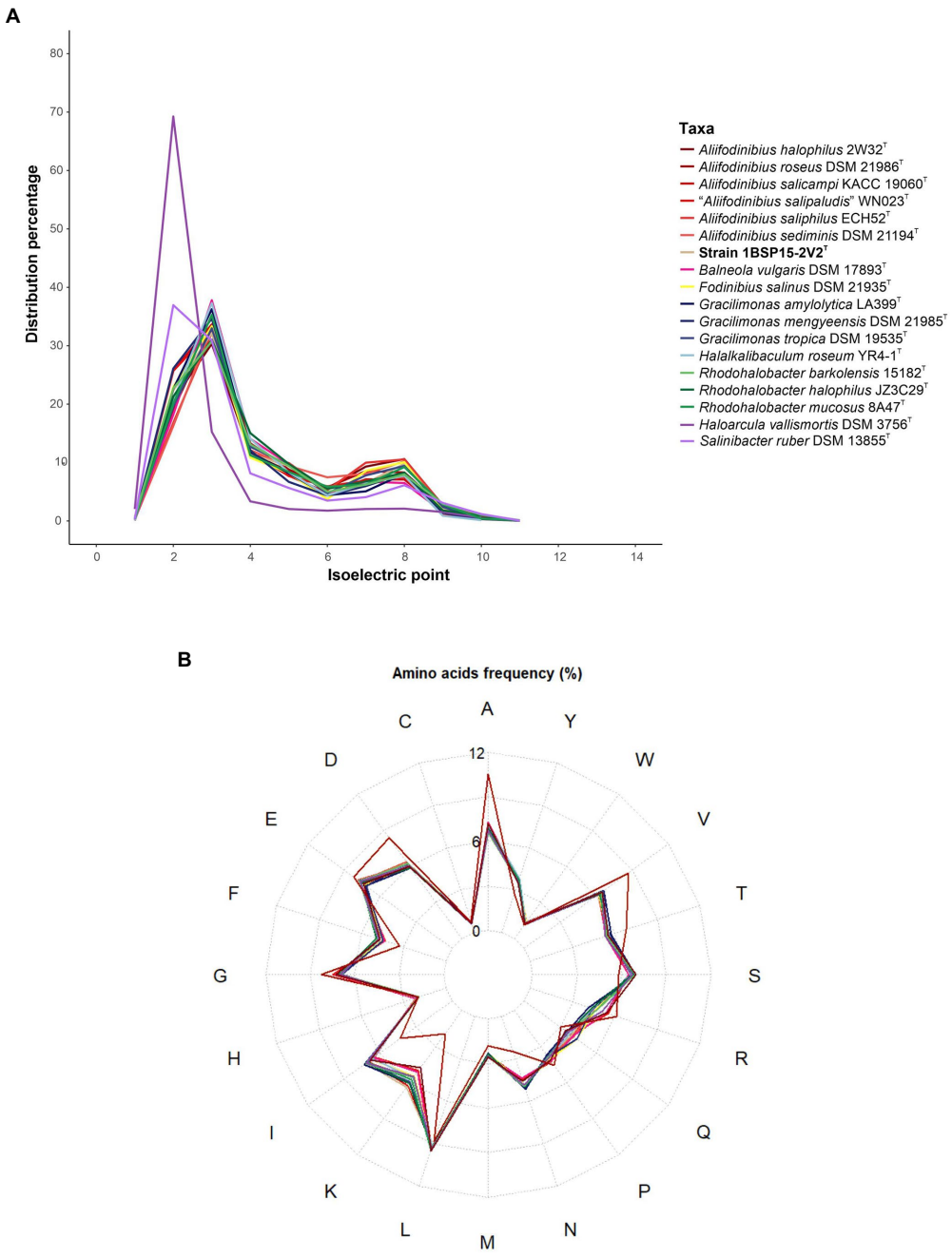


FIGURE 9
Isoelectric profile **(A)** and amino acid frequency **(B)** of the type strains of each species of the family *Balneolaceae*. The haloarchaeon *Haloarcula vallismortis* DSM 3756^T (GCF_900106715.1) and the bacterium *Salinibacter ruber* DSM 13855^T (GCF_000013045.1) have been included to compare our results with those of extreme halophiles with *salt-in* osmoregulatory mechanism. The genomes under study possessed a less acidic proteome, suggesting a *salt-out* strategy, though further *in vivo* analyses are necessary to confirm this hypothesis.

urea are not. H_2S is produced, but not indole. Acids are produced from D-arabinose, D-fructose, D-glucose, maltose, sucrose, and D-xylose but not from D-galactose, glycerol, lactose, mannitol, and D-trehalose. Utilizes amygdaline, D-cellobiose, D-fructose, D-galactose, D-glucose, D-lactose, D-maltose, D-mannose, D-melezitose, ribose, D-raffinose, salicin, sucrose, D-trehalose, D-xylose, dulcitol, ethanol, glycerol, mannitol, D-sorbitol, xylitol, fumarate, hippurate, malate, propionate, and pyruvate as sole sources of carbon and energy, but not L-arabinose, starch, butanol, methanol, propranolol, acetate, benzoate, butyrate, citrate, formate, and valerate. Utilizes L-asparagine, aspartic acid, cysteine, L-glutamine, L-methionine, ornithine, L-phenylalanine, and L-serine as the sole sources of carbon, nitrogen, and energy, but not L-alanine, arginine, L-glutamate, glycine, L-isoleucine, lysine, L-threonine, tryptophan, and valine. The major fatty acids are iso-$C_{15:0}$, $C_{16:1}$ ω7c and/or $C_{16:1}$ ω6c, and iso-$C_{17:1}$ ω9c and/or 10-methyl $C_{16:0}$. The genome of the type strain has an approximate size of 4.85 Mb and its G + C content is 44.5 mol%.

The type strain, 1BSP15-2V2^T (=CCM 9117^T = CECT 30246^T), was isolated from hypersaline soils at the Odiel Saltmarshes Natural Area in Huelva (Southwest Spain). The accession number for the 16S rRNA gene sequence is MW811395 and that of the genome sequence GCF_026229185.1.

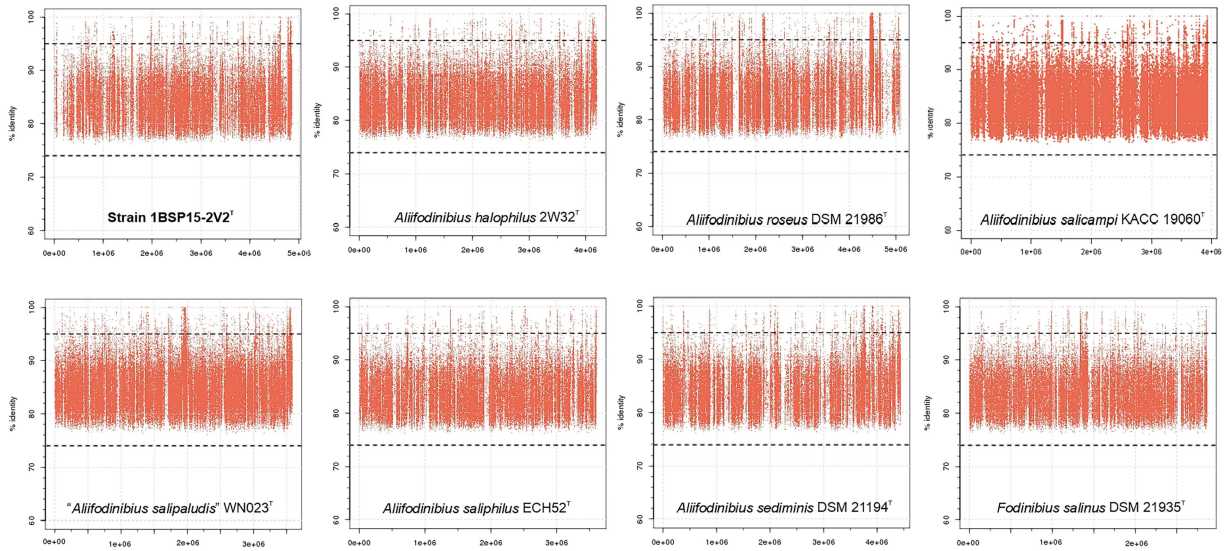


FIGURE 10
Genomic recruitment plots for the isolated strain 1BSP15-2V2^T and other type strains of species of the genera *Aliifodinibius* and *Fodinibius* against SMO1 hypersaline soil metagenomic reads. X-axis represents the genome length and Y-axis represents the percentage of identity of the reads matching the genome. Red dots above the 95% threshold line indicate read assignment at the species level. As it can be seen, most of the reads matches the genome at 80%–90% identity (above the proposed 73.98% for genus delineation), meaning they cannot be affiliated to any of the described *Aliifodinibius*/*Fodinibius* species, but they probably belong to a closed relative well represented in the SMO1 hypersaline soil.

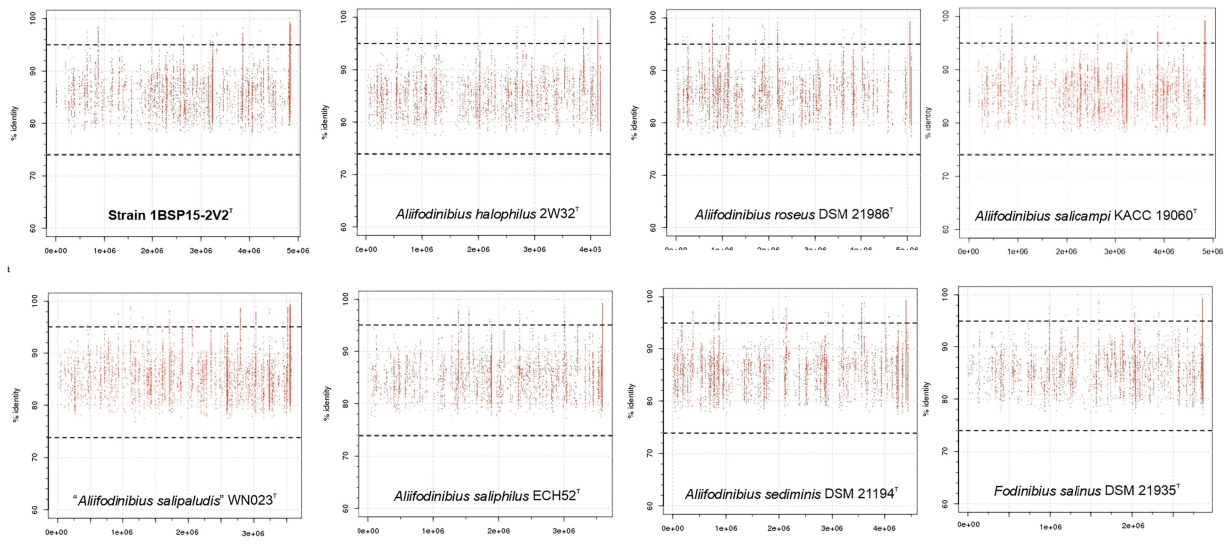


FIGURE 11
Genomic recruitment plots for the isolated strain 1BSP15-2V2^T and other type strains of species of the genera *Aliifodinibius* and *Fodinibius* against SS13 saltern metagenomic reads. X-axis represents the genome length and Y-axis represents the percentage of identity of the reads matching the genome. Red dots above the 95% threshold line indicate read assignment at the species level. A genus delineation line is also shown at 73.98% identity. The scarce read recruitment by all the species analyzed demonstrates their low abundance in this aquatic environment.

Description of *Fodinibius halophilus* comb. nov.

Basonym: *Aliifodinibius halophilus* Xia et al. 2016.

The description is the same as for *Aliifodinibius halophilus* (Xia et al., 2016). The genome of the type strain has an approximate size of 4.19 Mb and its G + C content is 42.5 mol%. The accession number for the 16S rRNA gene sequence is KR559733 and that of the genome sequence GCF_011059105.1.

Type strain: 2W32^T (=CICC 23869^T = KCTC 42497^T).

Description of *Fodinibius roseus* comb. nov.

Basonym: *Aliifodinibius roseus* Wang et al. 2013.

The description is the same as for *Aliifodinibius roseus* (Wang et al., 2013). The genome of the type strain has an approximate size of 5.08 Mb and

its G + C content is 48.3 mol%. The accession number for its 16S rRNA gene sequence is JQ923475 and that of the genome sequence GCF_900129315.1.

Type strain: YIM D15^T (=ACCC 10715^T = DSM 21986^T = KCTC 23442^T).

Description of *Fodinibius salicampi* comb. nov.

Basonym: *Aliifodinibius salicampi* Cho et al. 2017.

The description is the same as for *Aliifodinibius salicampi* (Cho et al., 2017, 2018). The genome of the type strain has an approximate size of 3.94 Mb and its G + C content is 42.8 mol%. The accession number for its 16S rRNA gene sequence is LC198077 and that of the genome sequence GCF_026228885.1.

Type strain: NBRC 112531^T (=KACC 19060^T = KHM44^T).

Description of *Fodinibius saliphilus* comb. nov.

Basonym: *Aliifodinibius saliphilus* Cho and Whang 2020.

The description is the same as for *Aliifodinibius saliphilus* (Cho and Whang, 2020). The genome of the type strain has an approximate size of 3.60 Mb and its G + C content is 40.8 mol%. The accession number for its 16S rRNA gene sequence is LC198072 and that of the genome sequence GCF_005869845.1.

Type strain: ECH52^T (=KACC 19126^T = NBRC 112664^T).

Description of *Fodinibius sediminis* comb. nov.

Basonym: *Aliifodinibius sediminis* Wang et al. 2013.

The description is the same as for *Aliifodinibius sediminis* (Wang et al., 2013). The genome of the type strain has an approximate size of 4.43 Mb and its G + C content is 48.1 mol%. The accession number for its 16S rRNA gene sequence is JQ923475 and that of the genome sequence GCF_900182555.1.

Type strain: YIM J21^T (=ACCC 10714^T = DSM 21194^T).

Emended description of the genus *Fodinibius* Wang et al. 2012

Cells are Gram-stain-negative, non-endospore-forming, non-motile rods, 0.05–1.5 µm in width and 0.2–5.5 µm in length (Wang et al., 2012). Colonies are salmon pink, pink, rose red, and reddish pigmented, circular, convex, and opaque with regular margins, and sometimes viscid. Some species are strictly aerobic bacteria, while others possess a facultatively anaerobic metabolism. Temperature range for growth between 14 and 45°C (optimum 37°C). Growth occurs at 3–25% (w/v) NaCl (optimum at 8–10%) and at pH values 5.0–10.0 (optimum at 6.0–8.0). All species produce catalase and most of them also oxidase. All species are unable to hydrolyze starch and DNA, and they cannot produce indole from tryptophane. Major fatty acids are iso-$C_{15:0}$, iso-$C_{17:1}$ ω9c, anteiso-$C_{15:0}$, $C_{16:1}$ ω7c and/or $C_{16:1}$ ω6c, $C_{16:1}$ ω7c and/or iso-$C_{15:0}$ 2-OH, and iso-$C_{17:1}$ ω9c and/or 10-methyl $C_{16:0}$. The DNA G + C content ranges between 40.8–48.3 mol% (genome).

The type species is *Fodinibius salinus*. The genome of the type strain of this type species has an approximate size of 2.86 Mb and its G + C content is 42.5 mol%. The genus is member of the family *Balneolaceae*, order *Balneolales*, class *Balneolia*, phylum *Balneolota*.

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 at: https://www.ncbi.nlm.nih.gov/genbank/, JAGGJA000000000 https://www.ncbi.nlm.nih.gov/genbank/, JAJNDC000000000 https://www.ncbi.nlm.nih.gov/genbank/, MW811395.

Author contributions

AV, CS-P, and RRH conceived the study. CG, CS-P, and AV acquired the environmental samples. CG performed the laboratory experiments and bioinformatic analysis, supported by CS-P and RRH, respectively. CG drafted the manuscript. CG, CS-P, RRH, and AV revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grants PID2020-118136GB-I00 funded by MCIN/AEI/10.13039/501100011033 and ERDF A way of making Europe, and from the Junta de Andalucía (P20_01066 and BIO-213), both with FEDER funds. CG was a recipient of a predoctoral fellowship (PRE2018-083242) from the Spanish Ministry of Science and Innovation. RRH was a recipient of a short-stay grant (PRX21/00598) from the Spanish Ministry of Universities.

Acknowledgments

We thank A. Oren (The Hebrew University of Jerusalem) for his help on the nomenclature of the new species.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022.1101464/full#supplementary-material

References

Abdel-Hamid, A. M., and Cronan, J. E. (2007). Coordinate expression of the acetyl coenzyme a carboxylase genes, *accB* and *accC*, is necessary for normal regulation of biotin synthesis in *Escherichia coli*. *J. Bacteriol.* 189, 369–376. doi: 10.1128/JB.01373-06

Alquethamy, S. F., Adams, F. G., Maharjan, R., Delgado, N. N., Zang, M., Ganio, K., et al. (2021). The molecular basis of *Acinetobacter baumannii* cadmium toxicity and resistance. *Appl. Environ. Microbiol.* 87:e0171821. doi: 10.1128/AEM.01718-21

Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the bacteria from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1099/00207713-52-2-485

Auch, A. F., Klenk, H. P., and Göker, M. (2010). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.* 2, 142–148. doi: 10.4056/sigs.541628

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

Bi, H., Zhu, L., Jia, J., and Cronan, J. E. (2016). A biotin biosynthesis gene restricted to *Helicobacter*. *Sci. Rep.* 6:21162. doi: 10.1038/srep21162

Booth, I. R. (2014). Bacterial mechanosensitive channels: progress towards an understanding of their roles in cell physiology. *Curr. Opin. Microbiol.* 18, 16–22. doi: 10.1016/j.mib.2014.01.005

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

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119–120. doi: 10.1093/bioinformatics/btn578

Castelán-Sánchez, H. G., Elorrieta, P., Romoacca, P., Liñan-Torres, A., Sierra, J. L., Vera, I., et al. (2019). Intermediate-salinity systems at high altitudes in the Peruvian Andes unveil a high diversity and abundance of bacteria and viruses. *Genes (Basel)* 10:891. doi: 10.3390/genes10110891

Chauhan, D., Srivastava, P. A., Agnihotri, V., Yennamalli, R. M., and Priyadarshini, R. (2019). Structure and function prediction of arsenate reductase from *Deinococcus indicus* DR1. *J. Mol. Model.* 25:15. doi: 10.1007/s00894-018-3885-3

Chen, S. C., Sun, G. X., Yan, Y., Konstantinidis, K. T., Zhang, S. Y., Deng, Y., et al. (2020). The great oxidation event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc. Natl. Acad. Sci. U. S. A.* 117, 10414–10421. doi: 10.1073/pnas.2001063117

Cho, Y., Chung, H., Jang, G. I., Choi, D. H., Noh, J. H., and Cho, B. C. (2013). *Gracilimonas rosea* sp. nov., isolated from tropical seawater, and emended description of the genus *Gracilimonas*. *Int. J. Syst. Evol. Microbiol.* 63, 4006–4011. doi: 10.1099/ijs.0.052340-0

Cho, G. Y., Lee, J. C., and Whang, K. S. (2017). *Aliifodinibius salicampi* sp. nov., a moderately halophilic bacterium isolated from a grey saltern. *Int. J. Syst. Evol. Microbiol.* 67, 2598–2603. doi: 10.1099/ijsem.0.001981

Cho, G. Y., Lee, J. C., and Whang, K. S. (2018). Erratum: *Aliifodinibius salicampi* sp. nov., a moderately halophilic bacterium isolated from a grey saltern. *Int. J. Syst. Evol. Microbiol.* 68:692. doi: 10.1099/ijsem.0.002572

Cho, G. Y., and Whang, K. S. (2020). *Aliifodinibius saliphilus* sp. nov., a moderately halophilic bacterium isolated from sediment of a crystallizing pond of a saltern. *Int. J. Syst. Evol. Microbiol.* 70, 358–363. doi: 10.1099/ijsem.0.003765

Choi, D. H., Zhang, G. I., Noh, J. H., Kim, W.-S., and Cho, B. C. (2009). *Gracilimonas tropica* gen. nov., sp. nov., isolated from a *Synechococcus* culture. *Int. J. Syst. Evol. Microbiol.* 59, 1167–1172. doi: 10.1099/ijs.0.005512-0

Choi-Rhee, E., and Cronan, J. E. (2003). The biotin carboxylase-biotin carboxyl carrier protein complex of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* 278, 30806–30812. doi: 10.1074/jbc.M302507200

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

Consejería de Medio Ambiente (1999). *Los criterios y estándares para declarar un suelo contaminado en Andalucía y la metodología y técnicas de toma de muestra y análisis para su investigación*. Sevilla: Junta de Andalucía.

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

Cowan, S. T., and Steel, K. J. (1965). *Manual for the Identification of Medical Bacteria*. Cambridge: University Press.

Cronan, J. E. (2018). Advances in synthesis of biotin and assembly of lipoic acid. *Curr. Opin. Chem. Biol.* 47, 60–66. doi: 10.1016/j.cbpa.2018.08.004

Cronan, J. E., Stewart, V., and Begley, T. (2014). Biotin and lipoic acid: synthesis, attachment, and regulation. *EcoSal Plus* 6:2012. doi: 10.1128/ecosalplus.ESP-0001-2012

Doxey, A. C., Kurtz, D. A., Lynch, M. D. J., Sauder, L. A., and Neufeld, J. D. (2015). Aquatic metagenomes implicate *Thaumarchaeota* in global cobalamin production. *ISME J.* 9, 461–471. doi: 10.1038/ismej.2014.142

Durán-Viseras, A., Andrei, A.-S., Ghai, R., Sánchez-Porro, C., and Ventosa, A. (2019). New *Halonotius* species provide genomics-based insights into cobalamin synthesis in haloarchaea. *Front. Microbiol.* 10:1928. doi: 10.3389/fmicb.2019.01928

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

Elevi Bardavid, R., and Oren, A. (2012). Acid-shifted isoelectric point profiles of the proteins in a hypersaline microbial mat: an adaptation to life at high salt concentrations? *Extremophiles* 16, 787–792. doi: 10.1007/s00792-012-0476-6

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

Felsenstein, J. (1983). Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* 14, 313–333. doi: 10.1146/annurev.es.14.110183.001525

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

Feng, Y., Napier, B. A., Manandhar, M., Henke, S. K., Weiss, D. S., and Cronan, J. E. (2014). A *Francisella* virulence factor catalyses an essential reaction of biotin synthesis. *Mol. Microbiol.* 91, 300–314. doi: 10.1111/mmi.12460

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

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

Hoffmann, T., and Bremer, E. (2016). "Management of osmotic stress by *Bacillus subtilis*: genetics and physiology" in *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*. ed. F. J. de Bruijn (New York: John Wiley & Sons), 657–676.

Hoffmann, T., and Bremer, E. (2017). Guardians in a stressful world: the Opu family of compatible solute transporters from *Bacillus subtilis*. *Biol. Chem.* 398, 193–214. doi: 10.1515/hsz-2016-0265

Holtmann, G., Bakker, E. P., Uozumi, N., and Bremer, E. (2003). KtrAB and KtrCD: two K^+ uptake systems in *Bacillus subtilis* and their role in adaptation to hypertonicity. *Society* 185, 1289–1298. doi: 10.1128/JB.185.4.1289

Hvitfeldt, E. (2021). Paletteer: Comprehensive Collection of Color Palettes. Available at: https://github.com/EmilHvitfeldt/paletteer

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Islam, M. N., Suzauddula, M., Ahamed, Z., Rabby, M. G., Hossen, M. M., Biswas, M., et al. (2022). Phylogenetic analysis and characterization of arsenic (As) transforming bacterial marker proteins following isolation of As-tolerant indigenous bacteria. *Arch. Microbiol.* 204:660. doi: 10.1007/s00203-022-03270-5

Jitrapakdee, S., and Wallace, J. C. (2003). The biotin enzyme family: conserved structural motifs and domain rearrangements. *Curr. Protein Pept. Sci.* 4, 217–229. doi: 10.2174/1389203033487199

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*. ed. H. N. Munro (London: Academic Press), 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

Khan, Z., Nisar, M. A., Hussain, S. Z., Arshad, M. N., and Rehman, A. (2015). Cadmium resistance mechanism in *Escherichia coli* P4 and its potential use to bioremediate environmental cadmium. *Appl. Microbiol. Biotechnol.* 99, 10745–10757. doi: 10.1007/s00253-015-6901-x

Kimbrel, J. A., Ballor, N., Wu, Y. W., David, M. M., Hazen, T. C., Simmons, B. A., et al. (2018). Microbial community structure and functional potential along a hypersaline gradient. *Front. Microbiol.* 9:1492. doi: 10.3389/fmicb.2018.01492

Knowles, J. R. (1989). The mechanism of biotin-dependent enzymes. *Annu. Rev. Biochem.* 58, 195–221. doi: 10.1146/annurev.bi.58.070189.001211

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. (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

Koser, S. A. (1923). Utilization of the salts of organic acids by the colon-aerogenes group. *J. Bacteriol.* 8, 493–520. doi: 10.1128/jb.8.5.493-520.1923

Kovacs, N. (1956). Identification of *Pseudomonas pyocyanea* by the oxidase reaction. *Nature* 178:703. doi: 10.1038/178703a0

Krulwich, T. A., Sachs, G., and Padan, E. (2011). Molecular aspects of bacterial pH sensing and homeostasis. *Nat. Rev. Microbiol.* 9, 330–343. doi: 10.1038/nrmicro2549

Lane, D. (1991). "16S/23S rRNA sequencing" in *Nucleic acid Techniques in Bacterial Systematics*. ed. E. Stackebrandt (New York: Wiley), 115–175.

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

Legatzki, A., Grass, G., Anton, A., Rensing, C., and Nies, D. H. (2003). Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. *J. Bacteriol.* 185, 4354–4361. doi: 10.1128/JB.185.15.4354-4361.2003

León, M. J., Vera-Gargallo, B., Sánchez-Porro, C., and Ventosa, A. (2016). *Spiribacter roseus* sp. nov., a moderately halophilic species of the genus *Spiribacter* from salterns. *Int. J. Syst. Evol. Microbiol.* 66, 4218–4224. doi: 10.1099/ijsem.0.001338

León-Del-Río, A. (2019). Biotin in metabolism, gene expression, and human disease. *J. Inherit. Metab. Dis.* 42, 647–654. doi: 10.1002/jimd.12073

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

Lin, S., and Cronan, J. E. (2011). Closing in on complete pathways of biotin biosynthesis. *Mol. BioSyst.* 7, 1811–1821. doi: 10.1039/c1mb05022b

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, , et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

Manandhar, M., and Cronan, J. E. (2018). A canonical biotin synthesis enzyme, 8-amino-7-oxononanoate synthase (BioF), utilizes different acyl chain donors in *Bacillus subtilis* and *Escherichia coli*. *Appl. Environ. Microbiol.* 84, e02084–e02017. doi: 10.1128/AEM.02084-17

Manara, A., DalCorso, G., Baliardini, C., Farinati, S., Cecconi, D., and Furini, A. (2012). *Pseudomonas putida* response to cadmium: changes in membrane and cytosolic proteomes. *J. Proteome Res.* 11, 4169–4179. doi: 10.1021/pr300281f

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., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2021). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

MIDI (2008). *Sherlock Microbial Identification System Operating Manual, version 6.1*. Newark, DE: MIDI Inc.

Morris, M. S. (2012). The role of B vitamins in preventing and treating cognitive impairment and decline. *Adv. Nutr.* 3, 801–812. doi: 10.3945/an.112.002535

Munoz, R., Rosselló-Móra, R., and Amann, R. (2016). Revised phylogeny of *Bacteroidetes* and proposal of sixteen new taxa and two new combinations including *Rhodothermaeota* phyl. nov. *Syt. Appl. Microbiol.* 39, 281–296. doi: 10.1016/j.syapm.2016.04.004

Nakazawa, M. (2022). Fmsb: Functions for Medical Statistics Book with Some Demographic Data. R Package Version 0.3. Available at: https://cran.r-project.org/package=fmsb

Noll, M., and Lutsenko, S. (2000). Expression of ZntA, a zinc-transporting P1-type ATPase, is specifically regulated by zinc and cadmium. *IUBMB Life* 49, 297–302. doi: 10.1080/15216540050033168

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2

Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315

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

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

Patiño-Ruiz, M., Ganea, C., and Călinescu, O. (2022). Prokaryotic Na^+/H^+ exchangers – transport mechanism and essential residues. *Int. J. Mol. Sci.* 23:9156. doi: 10.3390/ijms23169156

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

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* 70:e102. doi: 10.1002/cpbi.102

Rensing, C., Mitra, B., and Rosen, B. P. (1997). The *zntA* gene of *Escherichia coli* encodes a Zn(II)-translocating P-type ATPase. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14326–14331. doi: 10.1073/pnas.94.26.14326

Rice, P., Longden, I., 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

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

Rodionov, D. A., Mironov, A. A., and Gelfand, M. S. (2002). Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea. *Genome Res.* 12, 1507–1516. doi: 10.1101/gr.314502

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

Sainz, A., Grande, J. A., and de la Torre, M. L. (2004). Characterization of heavy metal discharge into the ria of Huelva. *Environ. Int.* 30, 557–566. doi: 10.1016/j.envint.2003.10.013

Sainz, A., Grande, J. A., de La Torre, M. L., and Sánchez-Rodas, D. (2002). Characterization of sequential leachate discharges of mining waste rock dumps in the Tinto and Odiel rivers. *J. Environ. Manag.* 64, 345–353. doi: 10.1006/jema.2001.0497

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

Sasser, M. (1990). Identification of bacteria by gas chromatography of cellular fatty acids. *Technical Note* 101, 1–6.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shapiro, M. M., Chakravartty, V., and Cronan, J. E. (2012). Remarkable diversity in the enzymes catalyzing the last step in synthesis of the pimelate moiety of biotin. *PLoS One* 7:e49440. doi: 10.1371/journal.pone.0049440

Shi, J., Cao, X., Chen, Y., Cronan, J. E., and Guo, Z. (2016). An atypical α/β-hydrolase fold revealed in the crystal structure of pimeloyl-acyl carrier protein methyl esterase BioG from *Haemophilus influenzae*. *Biochemistry* 55, 6705–6717. doi: 10.1021/acs.biochem.6b00818

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

Sirithanakorn, C., and Cronan, J. E. (2021). Biotin, a universal and essential cofactor: synthesis, ligation and regulation. *FEMS Microbiol. Rev.* 45:fuab003. doi: 10.1093/femsre/fuab003

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

Teichmann, L., Kümmel, H., Warmbold, B., and Bremer, E. (2018). OpuF, a new *Bacillus* compatible solute ABC transporter with a substrate-binding protein fused to the transmembrane domain. *Appl. Environ. Microbiol.* 84:01718. doi: 10.1128/AEM.01728-18

Urios, L., Agogué, H., Lesongeur, F., Stackebrandt, E., and Lebaron, P. (2006). *Balneola vulgaris* gen. nov., sp. nov., a member of the phylum *Bacteroidetes* from the North-Western Mediterranean Sea. *Int. J. Syst. Evol. Microbiol.* 56, 1883–1887. doi: 10.1099/ijs.0.64285-0

Vavourakis, C. D., Andrei, A. S., Mehrshad, M., Ghai, R., Sorokin, D. Y., and Muyzer, G. (2018). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* 6:168. doi: 10.1186/s40168-018-0548-7

Ventosa, A., Quesada, E., Rodriguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1982). Numerical taxonomy of moderately halophilic Gram-negative rods. *Microbiology* 128, 1959–1968. doi: 10.1099/00221287-128-9-1959

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 (Basel)* 9:152. doi: 10.3390/genes9030152

Wang, Y. X., Liu, J. H., Xiao, W., Ma, X. L., Lai, Y. H., Li, Z. Y., et al. (2013). *Aliifodinibius roseus* gen. nov., sp. nov., and *Aliifodinibius sediminis* sp. nov., two moderately halophilic bacteria isolated from salt mine samples. *Int. J. Syst. Evol. Microbiol.* 63, 2907–2913. doi: 10.1099/ijs.0.043869-0

Wang, Y. X., Liu, J. H., Xiao, W., Zhang, X. X., Li, Y. Q., Lai, Y. H., et al. (2012). *Fodinibius salinus* gen. nov., sp. nov., a moderately halophilic bacterium isolated from a salt mine. *Int. J. Syst. Evol. Microbiol.* 62, 390–396. doi: 10.1099/ijs.0.025502-0

Wei, W., Guan, H., Zhu, T., Zhang, S., Fan, C., Ouyang, S., et al. (2019). Molecular basis of BioJ, a unique gatekeeper in bacterial biotin synthesis. *iScience.* 19, 796–808. doi: 10.1016/j.isci.2019.08.028

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

Woodson, J. D., Peck, R. F., Krebs, M. P., and Escalante-Semerena, J. C. (2003). The *cobY* gene of the archaeon *Halobacterium* sp. strain NRC-1 is required for de novo cobamide synthesis. *J. Bacteriol.* 185, 311–316. doi: 10.1128/JB.185.1.311-316.2003

Wu, S., Wang, J., Wang, J., Du, X., Ran, Q., Chen, Q., et al. (2022a). *Halalkalibacterium roseum* gen. nov., sp. nov., a new member of the family *Balneolaceae* isolated from soil. *Int. J. Syst. Evol. Microbiol.* 72:005339. doi: 10.1099/ijsem.0.005339

Wu, S., Wang, J., Wang, J., Du, X., Ran, Q., Chen, Q., et al. (2022b). Corrigendum: *Halalkalibacterium roseum* gen. nov., sp. nov., a new member of the family *Balneolaceae* isolated from soil. *Int. J. Syst. Evol. Microbiol.* 72:005451. doi: 10.1099/ijsem.0.005451

Xia, J., Ling, S. K., Wang, X. Q., Chen, G. J., and Du, Z. J. (2016). *Aliifodinibius halophilus* sp. nov., a moderately halophilic member of the genus *Aliifodinibius*, and proposal of *Balneolaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* 66, 2225–2233. doi: 10.1099/ijsem.0.001012

Xia, J., Xie, Z.-H., Dunlap, C. A., Rooney, A. P., and Du, Z.-J. (2017). *Rhodohalobacter halophilus* gen. nov., sp. nov., a moderately halophilic member of the family *Balneolaceae*. *Int. J. Syst. Evol. Microbiol.* 67, 1281–1287. doi: 10.1099/ijsem.0.001806

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, S., Xu, Y., Guan, H., Cui, T., Liao, Y., Wei, W., et al. (2021). Biochemical and structural characterization of the BioZ enzyme engaged in bacterial biotin synthesis pathway. *Nat. Commun.* 12:2056. doi: 10.1038/s41467-021-22360-4

Zhao, X., Miao, S., Sun, Y., Gong, Q., Zhao, J., Wang, J., et al. (2020). *Aliifodinibius salipaludis* sp. nov., isolated from saline-alkaline soil. *Curr. Microbiol.* 77, 1328–1333. doi: 10.1007/s00284-019-01863-w

Zulkifli, L., Akai, M., Yoshikawa, A., Shimojima, M., Ohta, H., Guy, H. R., et al. (2010). The KtrA and KtrE subunits are required for Na^+–dependent K^+ uptake by KtrB across the plasma membrane in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 192, 5063–5070. doi: 10.1128/JB.00569-10

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fmicb.2023.1066406

OPEN ACCESS

EDITED BY
Brian P. Hedlund,
University of Nevada,
Las Vegas,
United States

REVIEWED BY
Daniel Colman,
Montana State University,
United States
Melody Lindsay,
Bigelow Laboratory for Ocean Sciences,
United States

*CORRESPONDENCE
Mihaela Glamoclija
✉ m.glamoclija@rutgers.edu
Ifeoma R. Ugwuanyi
✉ ifeoma.ugwuanyi@rutgers.edu

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

RECEIVED 10 October 2022
ACCEPTED 06 January 2023
PUBLISHED 01 February 2023

CITATION
Ugwuanyi IR, Fogel ML, Bowden R, Steele A, De Natale G, Troise C, Somma R, Piochi M, Mormone A and Glamoclija M (2023) Comparative metagenomics at Solfatara and Pisciarelli hydrothermal systems in Italy reveal that ecological differences across substrates are not ubiquitous.
Front. Microbiol. 14:1066406.
doi: 10.3389/fmicb.2023.1066406

Comparative metagenomics at Solfatara and Pisciarelli hydrothermal systems in Italy reveal that ecological differences across substrates are not ubiquitous

Ifeoma R. Ugwuanyi[1]*, Marilyn L. Fogel[2], Roxane Bowden[3], Andrew Steele[3], Giuseppe De Natale[4,5], Claudia Troise[4,5], Renato Somma[4,6], Monica Piochi[4], Angela Mormone[4] and Mihaela Glamoclija[1]*

[1]Department of Earth and Environmental Sciences, Rutgers University, Newark, NJ, United States, [2]EDGE Institute, University of California, Riverside, Riverside, CA, United States, [3]Earth and Planets Laboratory, Carnegie Institution for Science, Washington, DC, United States, [4]Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Naples, Italy, [5]Consiglio Nazionale delle Ricerche INO, Naples, Italy, [6]Consiglio Nazionale delle Ricerche IRISS, Naples, Italy

Introduction: Continental hydrothermal systems (CHSs) are geochemically complex, and they support microbial communities that vary across substrates. However, our understanding of these variations across the complete range of substrates in CHS is limited because many previous studies have focused predominantly on aqueous settings.

Methods: Here we used metagenomes in the context of their environmental geochemistry to investigate the ecology of different substrates (i.e., water, mud and fumarolic deposits) from Solfatara and Pisciarelli.

Results and Discussion: Results indicate that both locations are lithologically similar with distinct fluid geochemistry. In particular, all substrates from Solfatara have similar chemistry whereas Pisciarelli substrates have varying chemistry; with water and mud from bubbling pools exhibiting high SO_4^{2-} and NH_4^+ concentrations. Species alpha diversity was found to be different between locations but not across substrates, and pH was shown to be the most important driver of both diversity and microbial community composition. Based on cluster analysis, microbial community structure differed significantly between Pisciarelli substrates but not between Solfatara substrates. Pisciarelli mud pools, were dominated by (hyper)thermophilic archaea, and on average, bacteria dominated Pisciarelli fumarolic deposits and all investigated Solfatara environments. Carbon fixation and sulfur oxidation were the most important metabolic pathways fueled by volcanic outgassing at both locations. Together, results demonstrate that ecological differences across substrates are not a widespread phenomenon but specific to the system. Therefore, this study demonstrates the importance of analyzing different substrates of a CHS to understand the full range of microbial ecology to avoid biased ecological assessments.

KEYWORDS

hydrothermal system, Solfatara, Pisciarelli, microbial diversity, metagenomics, MAGs

Introduction

Continental hydrothermal systems (CHSs) are usually formed within diverse igneous lithologies (i.e., mafic, andesitic, and felsic) and the hydrothermal fluids derived from deep subsurface sources may be mixed with marine, brine, or meteoritic water resulting in geochemically diverse environments. These diverse environments exhibit substantial differences in the availability and abundance of electron acceptors and donors (Fournier, 1989; Shock et al., 2010; Lowenstern et al., 2015; Lindsay et al., 2018; Amenabar and Boyd, 2019), which microorganisms exploit as sources of energy (Amenabar et al., 2017; Lindsay et al., 2019; Puopolo et al., 2020; Aulitto et al., 2021).

CHSs have been the subject of many studies as they are a surface manifestation of hydrothermal activities, which are easily accessible for diverse microbiological studies (Inskeep et al., 2013; Menzel et al., 2015; Colman et al., 2016; Mardanov et al., 2018; Oliverio et al., 2018; Power et al., 2018; Boyer et al., 2020; Podar et al., 2020; Crognale et al., 2022 as a few examples). These studies have revealed that hot springs support microbial communities that are exceptionally diverse and vary in their response to physical and geochemical parameters (Inskeep et al., 2013; Sharp et al., 2014; Lindsay et al., 2018; Power et al., 2018; Colman et al., 2019a,b; Podar et al., 2020). Some studies have found pH to be the primary physical parameter that influences the microbial community composition of hot springs in Yellowstone National Park (Boyd et al., 2010; Inskeep et al., 2013; Colman et al., 2016, 2019a,b), Tengchong, China (Hou et al., 2013; Xie et al., 2015), Iceland (Moreras-Marti et al., 2021), and volcanic provinces in New Zealand (Power et al., 2018). In contrast, other studies have identified temperature as the most important driver of microbial community composition in hot springs (Miller et al., 2009; Cole et al., 2013; Sharp et al., 2014; Podar et al., 2020). Further, at Yellowstone National Park, subsurface processes including phase separation and mixing with meteoritic fluids were shown to shape the ecology of hot spring communities through their influence on the availability of nutrients that support microbial metabolism (Lindsay et al., 2018; Colman et al., 2019a). In addition to temperature, pH, phase separation and mixing of fluids, Fullerton et al. (2021) found that microbial diversity analyzed from fluids and associated sediments reflects the subsurface geological structures that fluids traverse, which in turn influences carbon cycling within the subduction zone in Costa Rica.

While we have learned a great deal about hot springs, their geochemical complexity and the observed ecological differences between water and hot spring deposits (Cole et al., 2013; Colman et al., 2016) highlight the need for studies analyzing different substrates (e.g., fumarolic deposits) to understand the full range of microbial ecology in CHSs. Some studies have analyzed microbial communities from different available substrates (Glamoclija et al., 2004; Ellis et al., 2008; Benson et al., 2011; Sharp et al., 2014; Wall et al., 2015; Medrano-Santillana et al., 2017; Crognale et al., 2018; Marlow et al., 2020; Arif et al., 2021); however, to the best of our knowledge, no CHSs study has used as many different substrates from the same hydrothermal system and correlated their environmental settings and metagenomes to characterize microbial ecology. The lack of comparison among different substrates (fumarolic deposits, water and mud from mud pools) may derive from the fact that not all substrates are present at all CHSs. In this study, we use metagenomes to investigate the microbial ecology in different substrates (i.e., water, mud and fumarolic deposits) from two CHSs in the context of their environmental geochemistry. The objectives of this study were to: (1) assess the microbial diversity of different CHSs and a variety of available substrates (2) identify the metabolic potential of these microbial communities in connection to environmental geochemistry.

The study areas are Solfatara and Pisciarelli CHSs (Figure 1) located within Campi Flegrei Caldera (CFC); an 8 km nested caldera in Naples (Italy) formed by the Campania Ignimbrite (39 ka) and Neapolitan Yellow Tuff (15 ka) eruptions (De Natale et al., 2016; Rolandi et al., 2020a,b). Present-day activities at CFC are characterized by large-scale hydrothermal circulation, gaseous emissions, and intense ground deformation (Troise et al., 2019; Moretti et al., 2020; Chiodini et al., 2021). According to Caliro et al. (2007), fumaroles and mud pools at Solfatara and Pisciarelli are fed by fluids of mixed magmatic-meteoritic origin. The Solfatara crater is 0.6 km in diameter and maintains hydrothermal activity at fumaroles where gas emissions reach temperatures from 145°C up to 165°C, while mud pools have an average temperature of approximately 45°C (Glamoclija et al., 2004; Chiodini et al., 2011). Furthermore, the Solfatara environment is extremely acidic (mud pool: 1.9–2.1 pH and fumarolic deposits: 1.3–2.2 pH; Crognale et al., 2018). The second location, Pisciarelli, sits on the outside northeastern wall of the Solfatara crater. Pisciarelli holds an unstable fumarole field characterized by consistent deposit

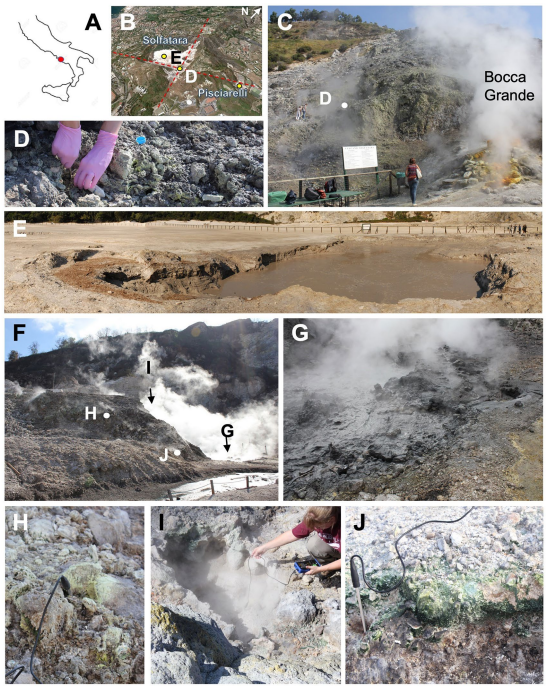


FIGURE 1
Sampling points at Solfatara-Pisciarelli hydrothermal systems. **(A)** Location map of Solfatara-Pisciarelli hydrothermal systems on contoured map of Italy (red dot). **(B)** Oblique view of Solfatara and Pisciarelli systems; red dashed lines represent faults; yellow dots are sampling points (E and D stand for Solfatara mud pool and fumaroles, respectively). **(C)** Sampling point of Solfatara fumarolic deposits (SF) next to the main fumarole Bocca Grande. **(D)** Close-up view of deposits to show heterogenicity of the material in regard to grain size and mineralogy (note different colors of deposits). **(E)** Solfatara mud pool (SMP) with bubbling water and warm muddy substrate near the pool. **(F)** Overview of the Pisciarelli location with marked sampling points. **(G)** Pisciarelli large pool (PLP) with bubbling hot mud (temp. 84.1°C). **(H)** Pisciarelli fumarolic deposits (PF) with sulfur crystals precipitate. **(I)** Pisciarelli small pool (PSP) sampling point. **(J)** Pisciarelli epilithic microbial layer (PLP-E) observed and sampled on the wall near the outflow channel of the large pool.

degassing, fluid emission from ephemeral vents, and boiling mud pools. This degassing activity is episodically accompanied by seismic swarms and macroscopic morphology changes such as the appearance of vigorously degassing vents, collapsing landslides, and bubbling mud (Fedele et al., 2021). Fumarolic deposits and mud pools at Pisciarelli are also very acidic (pH: 0.5–3) and temperatures of fumaroles ranges from 95°C to 110°C, while mud pools have temperatures between 84°C and 95°C (Ciniglia, 2005; Troiano et al., 2014; Poichi et al., 2019). In Solfatara-Pisciarelli CHSs, the major gases in fumarole vents are H_2O and CO_2 while the minor gases include H_2S, N_2, H_2, CH_4, He, Ar, and CO (Chiodini et al., 2010; Aiuppa et al., 2013). In both systems, fumaroles have a similar content of H_2O, CO_2, Ar, He, and N_2, however, they differ in their concentration of H_2S, H_2, and CO (Chiodini et al., 2001). Solfatara and Pisciarelli also differ in their fluid geochemistry; fluid in mud pools of Pisciarelli have been reported to have higher NH_4^+ concentrations (508–1,026 mg L^{-1}) compared to Solfatara (<1 mg L^{-1}; Martini et al., 1991; Valentino and Stanzione, 2003; Glamoclija et al., 2004; Poichi et al., 2019) resulting in the precipitation of the minerals mascagnite [$(NH_4)_2SO_4$] and tschermigite [$(NH_4)Al(SO_4)_2{\cdot}12(H_2O)$] (Poichi et al., 2019). The geochemical diversity of substrates in Solfatara-Pisciarelli CHSs provides an excellent location for evaluating the composition, structure, and functional potential of thermophilic microorganisms in different hydrothermal substrates and locations.

Materials and methods

To investigate the near-surface microbial community composition, structure and function at two CHSs, water, mud, and fumarolic deposits samples were collected from Solfatara and Pisciarelli (Figures 1A,B), in October 2012 (Supplementary Method). All samples were collected in triplicates, aseptically, using sterile Falcon tubes, Nalgene bottles, scoops, and gloves. After the collection, the samples were stored at −20°C until further processing; samples for long term storage were stored at −80°C. Temperature, pH, and redox potential (Eh) were measured *in situ* using a portable probe (Table 1). Gas readings were taken from the continuous Istituto Nazionale di Geofisica e Volcanologia (INGV) gas monitoring station at Solfatara and Pisciarelli (Table 1).

Geochemical characterization

About 1 g of solid sample (fumarolic deposits and mud) was added to 20 ml of MilliQ water and extracted as described in Lezcano et al. (2019). The concentration of water-soluble anions (F^-, Cl^-, NO_2^-, Br^-, NO_3^-, PO_4^{3-}, and SO_4^{2-}) in extracts and water samples was measured in triplicate using an 881 compact IC pro ion chromatography system (Metrohm, Switzerland) with a Metrosep A Supp 5-250/4.0 column. The concentration of NH_4^+ in extracts and water samples was measured using the alkaline hypochlorite/phenol nitroprusside method, after adding sodium citrate to prevent the precipitation of calcium and magnesium salts (Solorzano, 1969). Before measuring concentrations of NH_4^+ in water samples, samples were diluted 10-fold. Ammonium sulfate $(NH_4)_2SO_4$ solutions (0, 660, 1,320, 3,300, 6,600, and 13,200 ppm) were prepared and used as standard. The absorbance of each sample was measured in triplicate using an Evolution 60S UV-Vis Spectrophotometer at 640 nm wavelength.

X-ray fluorescence (XRF) was used to examine the chemical composition of mud and fumarolic deposits samples. Samples were analyzed in triplicate using a Horiba XGT-1000WR X-ray Fluorescence with an Rh tube X-ray source and elemental wt.% was determined using the XGT-1000WR software's quantification (Supplementary Method).

Stable isotopic analysis of carbon, nitrogen, and sulfur

Carbon, nitrogen, and sulfur stable isotope and elemental concentration analyses were performed at the Earth and Planets Laboratory, Carnegie Institution for Science. For organic carbon (TOC), and organic $\delta^{13}C$ measurements, samples were weighed into silver boats and fumed with 12 N HCl for 12–14 h. $\delta^{13}C$ and $\delta^{15}N$ isotopes were measured using a Thermo Scientific Delta VPlus isotope ratio gas-source mass spectrometer connected to a Carlo Erba (NA 2500) elemental analyzer (EA/IRMS) *via* a Conflo III interface. $\delta^{34}S$ was analyzed by the same gas-source mass spectrometer but connected to an Elementar Americas vario Micro CUBE elemental analyzer (EA-IRMS) *via* a Conflo III interface. Stable isotope values are reported in standard delta notation as ‰ variations relative to: Pee Dee Belemnite (PDB) for $\delta^{13}C$, atmospheric N_2 gas (AIR) for $\delta^{15}N$,

TABLE 1 Field measurements for water (W), mud (M), mud outlet (MO), epilith from dry mud wall (E), and fumarolic deposits (D) samples collected from Solfatara-Pisciarelli hydrothermal systems.

Location	Sample	T (°C)	pH	Eh (mV)	CO_2 (%)	CH_4 (ppm)	He (ppm)	H_2 (ppm)	N_2 (ppm)	Ar (ppm)	H_2S (ppm)
Pisciarelli	PLP	84.1	2.5	−485							
	PLP-MO	78.9	2.5	−486							
	PLP-E	74	–	−501	98.43	259	9	870.28	7,686	110	4,496
	PSP	88.8	1.5	−628							
	PSP-M	94.6	2	−622							
	PF-D	93.3	1.5	−622							
Solfatara	SMP-W	42.5	1	330							
	SMP-M	68.8	1	330	24.1	13	2.35	490	469	0.39	1,250
	SF-D	88.7	1	198							

The gas readings were taken close to the fumarolic emission sites; at Solfatara, the gas monitoring station was near our sampling site for SF-S, and at Pisciarelli, near sampling site for PF-S2. Pisciarelli large pool (PLP), Pisciarelli small pool (PSP), Pisciarelli fumarole (PF), Solfatara mud pool (SMP), and Solfatara fumarole (SF).

Vienna Canyon Diablo Troctolite (V-CDT) for $\delta^{34}S$, and with an analytical error of ±0.1‰. In-house $\delta^{13}C$ and $\delta^{15}N$ standards were also used and calibrated against international and certified standards as well. Additional $\delta^{34}S$ standards include the International Atomic Energy Agency reference materials IAEA-S-1 ($\delta^{34}S = -0.3$‰), and IAEA S3 ($\delta^{34}S = -32.3$‰), as well as NBS-123 ($\delta^{34}S = -17.09$‰), NBS-127 ($\delta^{34}S = -21.17$‰), and USGS-42 ($\delta^{34}S = -7.84$‰). A subset of the samples was analyzed in 2021 at the EDGE Stable Isotope Laboratory at the University of California Riverside to confirm the very negative nitrogen isotope values in some of these samples. USGS25 ($\delta^{15}N = -30.41$‰) and USGS40 ($\delta^{15}N = -4.52$‰) were used as two of our calibrating standards and confirmed the values measured earlier at the Earth and Planets Laboratory.

DNA extraction, metagenomic sequencing and sequence processing

About 0.25 g of mud and fumarolic deposits was used for DNA extraction. Before DNA extraction, fumarolic deposits were powdered using a sterilized agate mortar and pestle. Water samples were filtered in the laboratory using a 0.2 µm VWR black polycarbonate filter. Filters were cut into small pieces with a sterile scalpel and used for DNA extraction. DNA was extracted using DNeasy PowerSoil kit (Qiagen Inc., Valencia, CA, United States) with modifications to the manufacturer's instructions (Supplementary Method) and were stored at −80°C until further processing. Sequencing libraries were prepared using the Accel-NGS 2S Plus DNA Library Kit (Swift Biosciences Inc., Ann Arbor, MI, United States) according to the manufacturer's protocols. The prepared library was sequenced on the Illumina Novaseq platform at the Oklahoma Medical Research Foundation. Metagenome sequence quality was assessed using FastQC v0.11.5[1] and reads with a mean quality score less than 25 were discarded. Sequences were quality trimmed and Illumina sequencing adapters removed using Trimmomatic v0.36 (Bolger et al., 2014) with the following parameters: SLIDINGWINDOW:4:15, LEADING:3, TRAILING:3, and MINLEN:36. After quality control, taxonomic annotation of metagenomes was performed using Kaiju v1.8.2 (Menzel et al., 2016) against the NCBI nr + euk database (accessed on February 16, 2022) with the following parameters (run mode: greedy, minimum match length: 11, minimum match score: 75, allowed mismatches: 5). To compare the effects of taxonomic profiling using metagenome vs. 16S rRNA gene sequences, the trimmed metagenome sequences were uploaded to the MG-RAST online server (Meyer et al., 2008) and then passed through the MG-RAST QC pipeline. Following the quality control on MG-RAST, 16S rRNA gene sequences present in the metagenome (metagenome-derived 16S rRNA gene sequences) were identified and taxonomy assigned. Taxonomic classification of the metagenome-derived 16S rRNA gene sequences was done using the Greengenes rRNA database (DeSantis et al., 2006) hosted in MG-RAST with a minimum cut-off identity of 60% and e-value of 5. Since contamination leads to an overestimation of diversity, especially in low biomass samples like that of Solfatara and Pisciarelli (Karstens et al., 2019), after assigning taxonomy, we removed molecular biology kit and laboratory contaminants (which hereinafter are referred to as "kitome"), low abundance taxa (<0.01%) and sequences that were unclassified at the domain level from the taxonomy table before performing downstream analyses (Supplementary Method).

Co-assembly, metagenomic binning, and MAG quality assessment

Bins were generated from 10 Solfatara and 17 Pisciarelli metagenomes. To maximize genome recovery, we co-assembled several samples based on their taxonomic composition. All 10 Solfatara metagenomes were co-assembled and all 17 Pisciarelli metagenomes were co-assembled using MEGAHIT v1.2.9 (Li et al., 2015). Contigs longer than 1,000 bp were then binned using the metaWRAP binning module (Uritskiy et al., 2018) which incorporates three binning methods: CONCOCT v1.1.0 (Alneberg et al., 2014), MaxBin2 v2.2.6 (Wu et al., 2016), and metaBAT2 v2.12.1 (Kang et al., 2019). The metaWRAP refinement module (Uritskiy et al., 2018) was used to merge results from the three binning methods using the -c 50 and -x 10 options to obtain bins with over 50% completeness and less than 10% contamination according to the CheckM tool v1.0.12 (Parks et al., 2015). Bins with >50% completeness and <10% contamination were then reassembled with SPAdes v3.13.0 (Bankevich et al., 2012) to improve the assembly quality. The contamination and completeness of resulting Metagenome Assembled Genomes (MAGs) were reassessed with CheckM (Parks et al., 2015).

Taxonomic and functional annotation of MAGs

Taxonomy was assigned to MAGs using GTDB-TK v1.7.0 (Chaumeil et al., 2020). To assess whether MAGs belong to the same species (species ANI ≥ 95%), average nucleotide identity (ANI) was calculated for each possible pair of MAGs using FastANI v1.33 (Jain et al., 2018). MAGs with taxonomic assignment similar to previously identified kitome (e.g., *Corynebacterium* and *Staphylococcus*; Sheik et al., 2018; Weyrich et al., 2019) were not included in downstream analyses. Functional annotation of MAGs, including gene prediction was done using Prokka v1.14.5 (Seemann, 2014). Predicted genes were compared against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using BlastKOALA server (Kanehisa et al., 2016) to obtain KEGG Orthology (KO) annotations.

Abundance of MAGs in metagenomes

To assess the abundance of recovered MAGs, quality trimmed sequences from each metagenome were mapped against each MAG using BBMap v38.96. Sequence counts were normalized as the number of sequences recruited per kilobase of MAG and gigabase of metagenome (RPKG). The normalized sequence counts allowed for direct comparison of genome abundance between metagenome of different depths.

Statistical analyses

All statistical analyses were performed in R v4.0.2 (R Core Team, 2020). Principal Component Analysis (PCA) was performed with

1 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

environmental parameters using the "*prcomp*" function in the stat package with scaling enabled. PCA results were graphed using the "*ggbiplot*" function from the ggbiplot v0.55 package. Before using the geochemical data for statistical analyses, variables below detection limit in all samples were excluded and the dataset transformed using z-score. After removing kitome and all potential contaminants, we calculated alpha diversity using the species abundance table obtained from metagenome-derived 16S rRNA gene (Greengenes) and metagenome (NCBI nr + euk) taxonomic assignment. Alpha diversity (Shannon, observed richness and Pielou's evenness) was calculated using the "*estimate_richness*" function from the phyloseq v1.32.0 package in R. Nonpareil v3.0 (Rodriguez-R et al., 2018) was then used to estimate coverage and also calculate diversity with kmer kernel and default parameters. Kruskal-Wallis test was used to investigate whether alpha diversity varied significantly between locations and substrates. The "*lm()*" function in R was used to perform linear regression to evaluate potential effects of environmental variables on alpha diversity. Bray–Curtis dissimilarity based on genus and species abundance was used to determine differences in microbial community composition. The Bray-Curtis distance calculated was visualized using a Non-metric Multidimensional Scaling (NMDS) plot. Analysis of Similarity (ANOSIM) was performed on the Bray–Curtis dissimilarity matrix using the vegan v2.5-6 package to evaluate the significance of microbial compositional differences between the two locations and different types of substrates. Similarity Percentage (SIMPER) analysis was performed using the vegan v2.5-6 package to identify genera that contributed to the dissimilarity between two locations and different types of substrates. Mantel test using Spearman's correlation coefficient with 999 permutations was performed to evaluate the significance of correlation between community composition (Bray–Curtis dissimilarity) and environmental parameters (Euclidean distance).

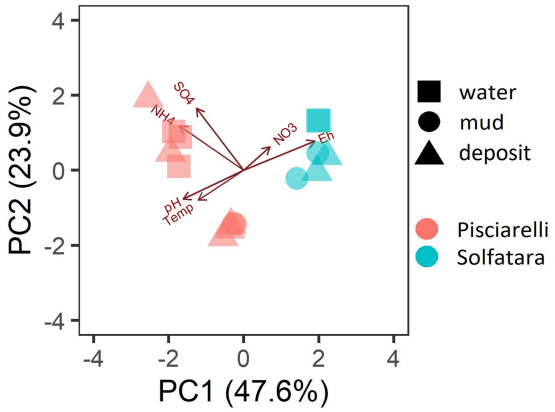


FIGURE 2
Biplot based on Principal Component Analysis (PCA) using temperature, pH, Eh, water-soluble anions and cations. Variables that had values below detection limit in most samples (F^-, Cl^-, and PO_4^{3-}) were excluded from PCA.

Results and discussion

Geochemical context of Solfatara-Pisciarelli hydrothermal systems

At Solfatara-Pisciarelli CHSs (Figure 1), the crater structure that forms the rock substrate (alkaline potassic tephra and lava ranging from trachybasalt to phonotrachyte) for both locations was produced and shaped by the same volcanic events (Piochi et al., 2015). Solid substrates from both locations have similar bulk elemental composition [Supplementary Figure 1; Supplementary Table 1 and Piochi et al. (2014), Piochi et al. (2015)]. In addition, the same magma chamber provides a heat source for the fumaroles and mud pools (Chiodini et al., 2001; Valentino and Stanzione, 2003) of both systems. However, tectonic features diverge the fluids that feed Pisciarelli mud pools to pass through old marine sediments with strong organic imprints where the fluids become enriched in NH_4^+ (Poichi et al., 2019).

PCA based on temperature, pH, Eh, and water-soluble nutrients showed further distinction between Solfatara and Pisciarelli (Figure 2; Supplementary Table 2). Pisciarelli an extremely acidic (pH: 1.5–2.5) and reducing (Eh: −628 to −485) environment has higher temperatures (74°C to 95°C) and gas concentrations including H_2S and CH_4 (Table 1) compared to Solfatara (pH: ~1; temperature: 42°C to 88.7°C; Eh: 198 and 330 mV), making Pisciarelli a more active and extreme hydrothermal environment than Solfatara. These higher temperatures, which inhibit the solubility of oxygen (Boyer et al., 2020), together with higher concentrations of reducing gases may contribute to the more reducing conditions we observed at Pisciarelli. The H_2S gas released in CHSs may be oxidized abiotically by oxygenated meteoritic waters (Nordstrom et al., 2005, 2009) or biotically by chemosynthetic microorganisms leading to the higher concentration of SO_4^{2-} we observed in mud pools (Pisciarelli: 3,326–6,208 ppm; Solfatara: 206–2,888 ppm) compared to fumarolic deposits (Pisciarelli: 118–439 ppm; Solfatara: 567–610 ppm). The concentrations of SO_4^{2-} in Pisciarelli mud pools were two times that of Solfatara mud pool possibly due to the higher concentration of H_2S gas measured at Pisciarelli. The concentration of NH_4^+ was also higher in mud pools, particularly in Pisciarelli mud pools (Pisciarelli: 1,130–1,998 ppm; Solfatara: 1.1–36 ppm) than in fumarolic deposits (Pisciarelli: 0.8 to 38 ppm; Solfatara: 0.3 ppm) which corresponds with other studies that report higher concentrations of NH_4^+ in Pisciarelli mud pools (Martini et al., 1991; Valentino and Stanzione, 2003; Glamoclija et al., 2004; Poichi et al., 2019). Temperature and geochemical measurements from Pisciarelli large mud pool revealed an environmental gradient; values were lower at the discharge channel (temperature: 79°C; SO_4^{2-}: 134.4–193.6 ppm; NH_4^+: 46.5–67.8 ppm) compared to the main bubbling pool (temperature: 84°C; SO_4^{2-}: 3326–6,208 ppm; NH_4^+: 1,130–1,998 ppm). The low concentrations of SO_4^{2-} and NH_4^+ along the discharge channel may result from the removal of H_2S and NH_4^+ possibly due to oxidation and volatilization as fluid flows into the channel.

Discrepancies between metagenome-derived 16S rRNA gene and metagenomic profiles

In this study, a total of 7 Solfatara and 10 Pisciarelli samples were analyzed. Solfatara samples had between 54,932,932 to 210,855,916 (average = 126,261,781) quality filtered metagenome sequences whereas Pisciarelli had between 119,199,732 to 230,316,988 (average = 179,177,105) quality filtered metagenome sequences. We compared taxonomic profiles generated from metagenome against metagenome-derived 16S rRNA gene sequences and observed that metagenome sequences identified more taxa than metagenome-derived

16S rRNA gene sequences, which corresponds to other studies that detected an increased number of taxa with whole genome shotgun sequencing compared to the 16S amplicon method (Ranjan et al., 2016; Brumfield et al., 2020). For example, Ranjan et al. (2016) reported that with the same number of sequences, whole genome shotgun sequencing identified twice as many species as the 16S method. We observed that metagenome sequences identified 56 phyla while metagenome-derived 16S rRNA gene sequences identified 26 phyla. Only 3 out of the 26 phyla detected by metagenome-derived 16S rRNA gene profiling were not detected by metagenomic profiling. Metagenomic profiling, on the other hand, identified 35 phyla not identified by metagenome-derived 16S rRNA gene profiling with 12 of them being viral and eukaryotic phyla. Overall, the dominant bacterial and archaeal phyla detected were similar irrespective of profiling method. Predominant bacterial and archaeal phyla detected by both profiling methods include *Crenarchaeota*, *Proteobacteria*, *Firmicutes*, and *Actinobacteria* (Figures 3A,B). At the genus level, metagenomic profiling also identified more genera than metagenome-derived 16S rRNA gene profiling (Figures 4A,B). Of the 245 genera identified across all samples by metagenomic profiling methods only 48 genera (20%) were identified with both profiling methods. Metagenomic profiling identified 197 genera not identified by metagenome-derived 16S rRNA gene profiling including *Acidilobus*, *Acidibacillus*, *Thermogymnomonas*, *Ampullavirus*, and *Bicaudavirus* while genera including *Acetobacterium* and *Caldococcus* were unique to metagenome-derived 16S rRNA gene profiling. We observed that both methods identified dominant genera including *Acidianus*, *Pyrobaculum*, and *Sulfobacillus*.

Next, we analyzed the effect of both taxonomic profiling methods on diversity (Shannon diversity, observed richness, Pielou's evenness). Before calculating the diversity, we identified and removed kitome (Supplementary Figure 3; Supplementary Tables 4, 5), low abundance taxa and sequences that were unclassified at the domain level from the taxonomy table to avoid an overestimation of diversity and a misrepresentation of community composition (Karstens et al., 2019). Our results showed that Shannon diversity and observed richness were consistently higher when the taxonomy table from metagenomic profiling was used as input compared to when we used the taxonomy table from metagenome-derived 16S rRNA gene profiling (Figures 5A,B). Overall, our results reflect that the 16S amplicon approach, which has been the most employed approach for studying CHSs' microbiome, identifies a significantly lower number of bacterial species and completely excludes viruses and fungi which is similar to reports from other studies (Ranjan et al., 2016; Brumfield et al., 2020). Although metagenome sequencing permits the identification of more taxa, the choice between 16S rRNA gene and metagenome sequencing ultimately depends on the ecological questions and objectives of any given study. Since metagenome sequences encompass all members of the microbiome and provide high-resolution diversity analysis, which agrees with our objective to understand the full range of microbial ecology in available substrates from Solfatara-Pisciarelli CHSs, the further analyses and results hereinafter are based on the taxonomy table generated from metagenomic profiling.

Microbial diversity of Solfatara-Pisciarelli hydrothermal systems

Shannon diversity in Solfatara-Pisciarelli CHSs ranged from 1.20 to 4.26 (Figure 5B; Supplementary Table 8). We observed significant variations in Shannon diversity (Kruskal–Wallis: $p < 0.05$) between

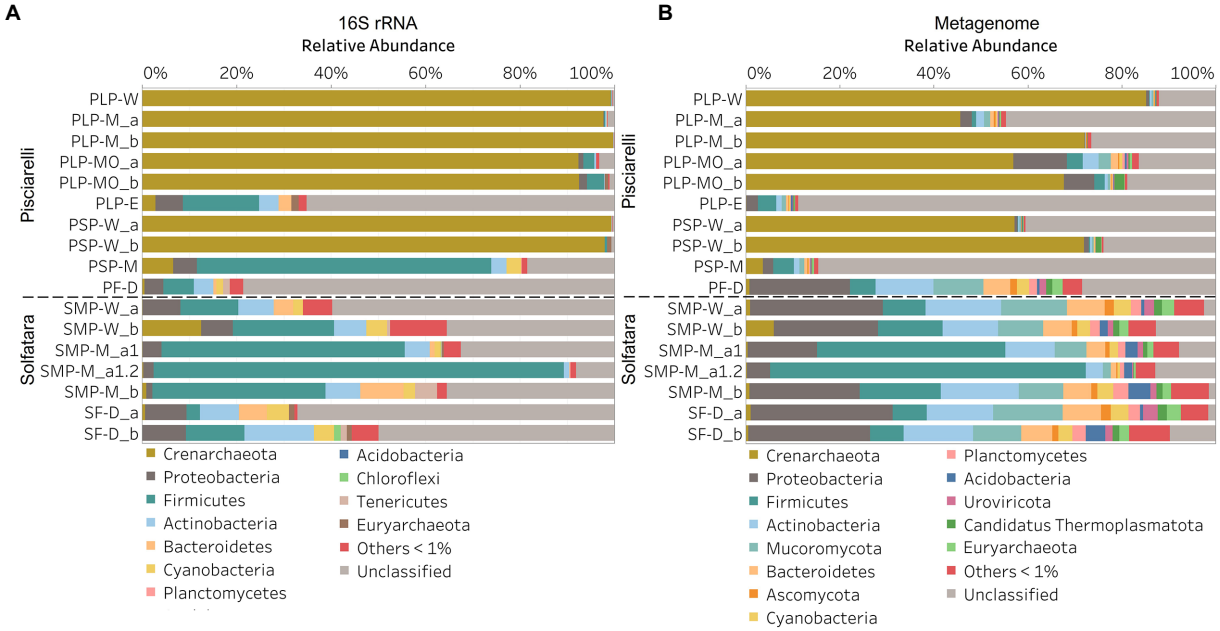


FIGURE 3
Relative abundance of phyla based on **(A)** metagenome-derived 16S rRNA gene and **(B)** metagenomic profiling. Relative abundance was calculated after removing sequences that were unclassified at the domain level, kitome, singleton, species present in only one sample and low abundance species. Phyla with average abundance of <1% were grouped into Others <1%. W represents water, M for mud, MO for mud outlet, E for epilithic microbial layer from dry mud, and D for fumarolic deposits samples. Samples labeled a and b are replicates sampled from the same spot. Sample SMP-M_a1 and SMP-M_a1.2 are sequencing replicates.

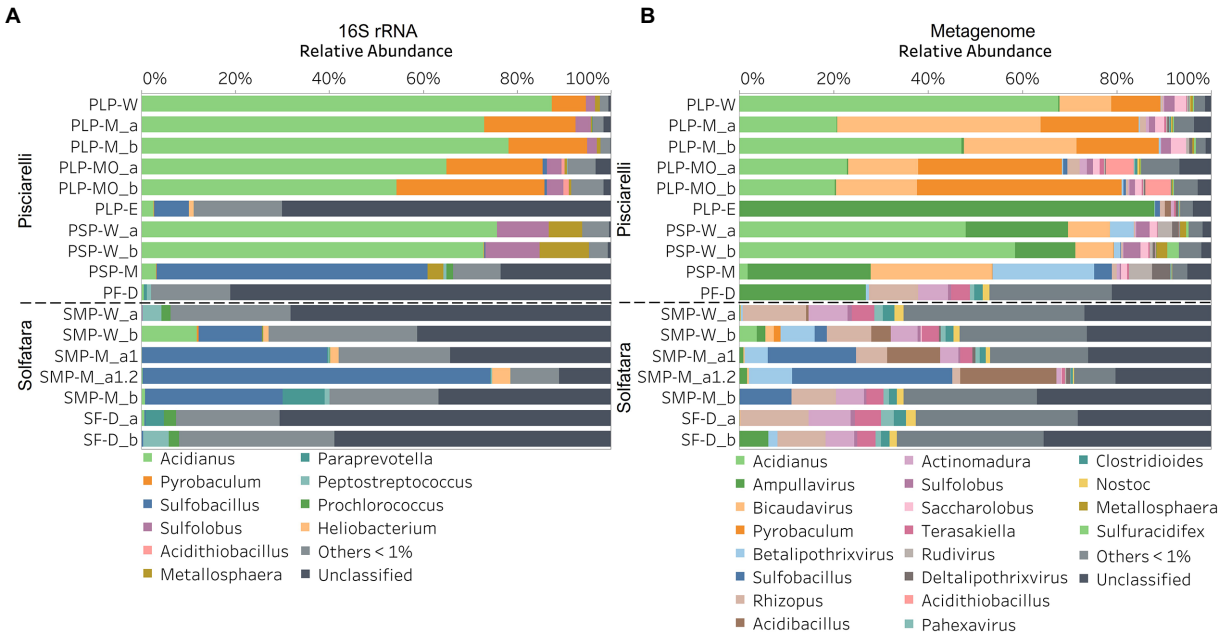


FIGURE 4
Relative abundance of genera based on **(A)** metagenome-derived 16S rRNA gene and **(B)** metagenomic profiling. Relative abundance was calculated after removing sequences that were unclassified at the domain level, kitome, singleton, species present in only one sample and low abundance species. Genera with average abundance of <1% were grouped into Others <1%. W represents water, M for mud, MO for mud outlet, E for epilithic microbial layer from dry mud, and D for fumarolic deposits samples. Samples labeled a and b are replicates sampled from the same spot. Sample SMP-M_a1 and SMP-M_a1.2 are sequencing replicates.

Solfatara and Pisciarelli which were supported by Nonpareil diversity values (Supplementary Figure 4; Supplementary Table 8); on average, Shannon diversity was lower at Pisciarelli (Shannon: 1.20–3.52) than at Solfatara (Shannon: 3.55–4.26). Additionally, microbial communities were more uneven (Kruskal–Wallis: $p < 0.05$) at Pisciarelli (Pielou's evenness: 0.26–0.74) than at Solfatara (Pielou's evenness: 0.67–0.84). However, there was no significant difference in observed richness between these locations (Solfatara: 114–201 species; Pisciarelli: 102–166 species; Kruskal–Wallis: $p = 0.24$). The decreased Shannon diversity at Pisciarelli may have resulted from the combined effects of reducing conditions, extremely low pH, and high temperature. The addition of highly reducing conditions at Pisciarelli to an already extreme environment (high temperature and low pH) may cause further selection against certain microbial groups by hindering their growth and metabolism, thus leading to the lower Shannon diversity observed (Kimbrough et al., 2006; Seo and DeLaune, 2010; Husson, 2013; Zhang et al., 2015). Alpha diversity in substrates showed some variation (e.g., on average, mud and fumarolic deposits had higher Shannon diversity than water, and richness showed less variability in water and mud); however, these differences were not statistically significant (Kruskal–Wallis, Shannon: $p = 0.97$; Pielou's evenness: $p = 0.83$; observed genus richness: $p = 0.09$) even when substrates were analyzed separately for each location which implies that species alpha diversity does not differ across substrates in the Solfatara-Pisciarelli CHSs. This result contrast with other studies comparing microbial diversity in water and solid substrates from hot springs, which showed differences in richness (Colman et al., 2016) and evenness (Cole et al., 2013) across substrates.

A comparison of diversity and physical parameters showed pH (Shannon: $R^2 = 0.73$; Nonpareil: $R^2 = 0.43$, $p < 0.05$) to be the primary driver of microbial diversity which is consistent with several studies that report pH as the major predictor of microbial diversity in some hydrothermal systems (Oliverio et al., 2018; Power et al., 2018). Results also showed that Eh had a significant correlation (Shannon: $R^2 = 0.62$; Nonpareil: $R^2 = 0.43$, $p < 0.05$) with diversity. Temperature on the other hand had minimal effect on diversity (Shannon: $R^2 = 0.12$, $p = 0.09$; Nonpareil: $R^2 = 0.02$, $p = 0.25$). This is in contrast to previous studies that reported a strong relationship between temperature and microbial diversity in hot springs (Miller et al., 2009; Sharp et al., 2014). Comparably, minimal effect of temperature on alpha diversity has been reported for hot spring communities in Yellowstone, United States, Fludir, Iceland and Tibetan Plateau, China (Wang et al., 2013; Podar et al., 2020).

Microbial community composition of Solfatara-Pisciarelli hydrothermal systems

After removing kitome and all potential contaminants, a total of 490 species (245 genera) were identified across all samples corresponding to 11 archaeal, 31 bacterial, 10 eukaryotic, and four viral phyla (Supplementary Table 6). Our results showed that overall, 33% (133) of the species detected were shared across the different types of substrates but each substrate contained unique species among which water had the least number of unique taxa and mud contained the highest (Figure 6B). NMDS plot (Figure 7) generated based on Bray–Curtis dissimilarity matrix revealed that the microbial community structure of Pisciarelli was distinct from Solfatara (ANOSIM: $R = 0.63$ for species; $R = 0.67$ for genus, $p < 0.05$). On average, the most abundant phylum at Pisciarelli was *Crenarchaeota* (46%) followed by *Proteobacteria* (5%) and *Actinobacteria* (2%), whereas Solfatara was dominated by *Firmicutes* (23%), *Proteobacteria* (21%), and *Actinobacteria* (12%; Figure 3B). In addition, Pisciarelli substrates had a higher number of viral sequences

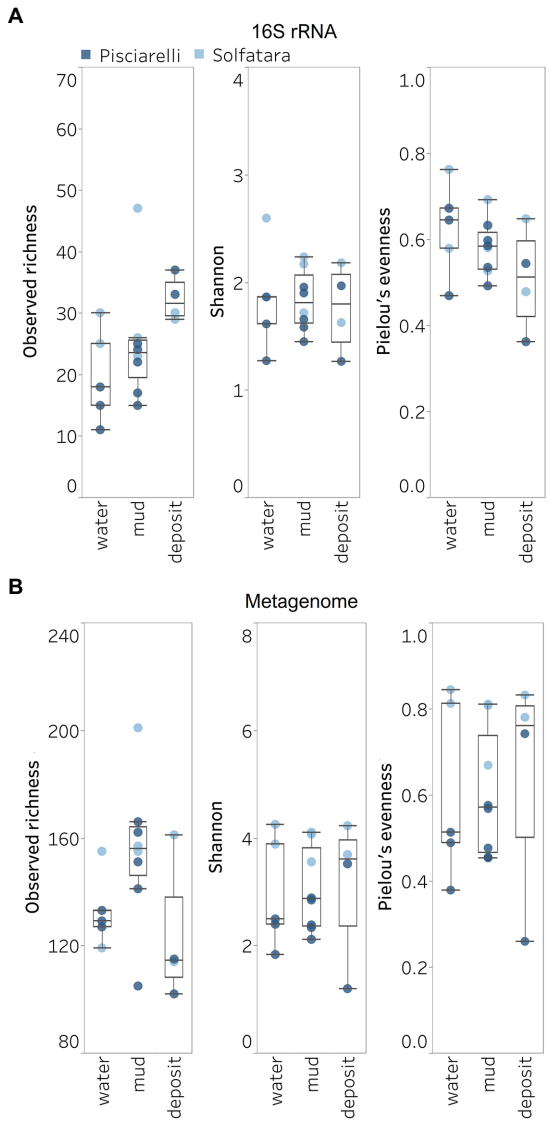


FIGURE 5
Alpha diversity across locations and substrate types calculated using species abundance from **(A)** metagenome-derived 16S rRNA gene and **(B)** metagenomic profiling.

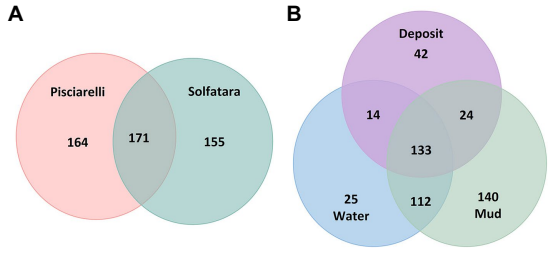


FIGURE 6
Venn diagram showing the number of unique and shared species between **(A)** two locations and **(B)** different types of substrates.

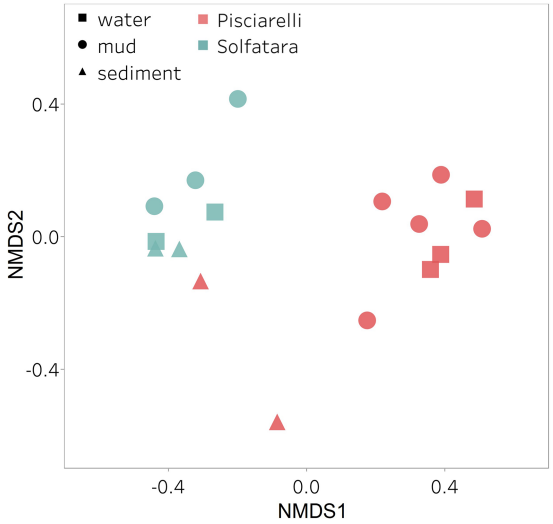


FIGURE 7
Nonmetric Dimensional Scaling (NMDS) ordination based on Bray–Curtis dissimilarity matrix. Bray–Curtis dissimilarity was calculated using species abundance from metagenomic profiling. The shapes define substrate types (water, mud, and deposits) and the colors represent locations (Pisciarelli and Solfatara).

(4–49%) than Solfatara (1–4%; Supplementary Figure 2). Interestingly, we detected *Cyanobacteria* (0.5 to 4%) especially at Solfatara even in samples (SF-D_and SF-D_b) with a temperature (89°C) that exceeds the known temperature limit for photosynthesis in acidic environments (Brock, 1985; Cox et al., 2011; Hamilton et al., 2012). The *Cyanobacteria* we detected may have been introduced into the hydrothermal system from surrounding cooler environments (by wind or fluid circulations). However, we cannot decipher with certainty if they were environmentally introduced or if they are true members of the Solfatara-Pisciarelli hydrothermal community. We observed that at the species level, both locations shared 171 species (35%), however, there were over 100 species unique to each location (Figure 6A). Species unique to Pisciarelli include thermophiles such as *Pyrolobus fumarii*, *Pyrobaculum aerophilum* and *Metallosphaera sedula*. Unique species identified at Solfatara include *Thermaerobacter* sp. *FW80*, *Methylacidiphilum kamchatkense* and *Ilumatobacter coccineus*.

Our results showed that the microbial community structure in the high temperature environments (74–95°C) of Pisciarelli was different. The source of Pisciarelli large mud pool (84°C) was dominated by *Acidianus* (21–68%), *Pyrobaculum* (11–21%) and *Bicaudavirus* (11–43%) but as temperature decreased away from the source (79°C), we detected a decrease in the abundance of *Acidianus* (20–23%) and *Bicaudavirus* (15–17%) but an increase in the abundance of *Pyrobaculum* (30–43%) in the discharge channel. Water from Pisciarelli small mud pool (89°C) was also dominated by *Acidianus* (48–59%), but *Pyrobaculum* accounted for <1% of total sequences in this mud pool. Interestingly, the viral genus, *Ampullavirus* accounted for over 80% of sequences in Pisciarelli epilithic microbial layer (74°C) and for 26% of sequences in both Pisciarelli fumarolic deposit (93°C) and mud from Pisciarelli small mud pool (88°C). Most of the viruses detected in Pisciarelli fumarolic deposits, epilithic microbial layer and mud from the small mud pool were known archaeal viruses (e.g., *Ampullavirus*, *Bicaudavirus*, and *Betalipothrixvirus*) even though archaea accounted for <3% of the total sequences in these samples. This difference in relative abundance between archaeal viruses and their known host suggests that each archaeal species may host more than one virus type or the archaeal viruses present in Solfatara-Pisciarelli CHSs may have a

broader host range that includes bacteria or/and eukaryotes (Munson-McGee et al., 2018). When substrates were compared at each location, we observed that Pisciarelli substrates harbored significantly different (ANOSIM species: $R = 0.64$; genus: $R = 0.62$, $p < 0.05$) microbial communities. SIMPER analysis showed that the sulfur-oxidizing archaeal genus *Acidianus* was significantly ($p < 0.05$) more abundant in Pisciarelli water communities, the archaeal genus *Pyrobaculum* showed a higher relative abundance in mud communities, and *Ampullavirus* showed a higher relative abundance in fumarole deposits communities (Supplementary Table 9; Figure 4B). Overall, Pisciarelli water and mud communities were similar but distinct from communities in Pisciarelli fumarolic deposits. The distinction between communities in the mud pools and fumarolic deposits at Pisciarelli may be linked to fluids and/or may reflect the availability of oxygen. Although the entire Pisciarelli environment has high temperatures and is highly reducing, the vigorous mixing of hydrothermal gas with air may create pockets of atmospheric O_2 in pore spaces of fumarolic deposits which support aerobic communities that may be absent or in low abundance in mud pools. In contrast, high temperatures of mud pools inhibit the dissolution of oxygen thereby creating microaerobic environments that may favor thermophilic facultative and obligate anaerobes such as *Acidianus* and *Pyrobaculum*. Furthermore, the presence of *Pyrobaculum* specifically *Pyrobaculum arsenaticum* which is a strict anaerobe (Huber et al., 2000) and members of the genus *Sulfuracidifex* who are all obligate aerobes in Pisciarelli large pool may be an indication that conditions in this pool fluctuate between microaerobic and completely anaerobic, possibly in congruence with the intensity of hydrothermal activity.

At Solfatara, we observed variability in genus abundance across samples even between replicates (Figure 4B). For example, *Acidibacillus* accounted for 11% of sequences in SMP-M_a1 but was <1% in replicate mud sample, SMP-M_b. The difference between replicates may derive from the heterogeneous nature of Solfatara-Pisciarelli CHSs; however, seeing that genus abundance also varied in our sequencing replicates (SMP-M_a1 and SMP-M_a1.2), the variability in genus abundance may likely be bias introduced from sequencing. Results also showed that a large percentage of Solfatara sequences (20–36%) were unclassified at the genus level, suggesting a considerable amount of potentially novel genera at this environment. Interestingly, at Solfatara, we detected several genera that are not known thermophiles and not commonly found at pH < 2 (e.g., *Rhizopus*; 1.7–14%, *Actinomadura*; 1–9%, *Terasakiella*; 0.7–5.6%, and *Clostridioides*; 0.3–2.5%; Wang et al., 2014; Strazzulli et al., 2020). These genera were present in Solfatara water (43°C), mud (69°C), and fumarolic deposits (89°C). Furthermore, we detected these genera in most Pisciarelli samples but to a lesser degree. These microorganisms may not be native to the hydrothermal systems but likely introduced into the habitat from extensive human activity around the crater. In addition, the fluids in mud pools are a mixture of hydrothermal fluids and meteoric water which could deliver genetic material from non-native microorganisms to the hydrothermal systems. This possible introduction of non-native microorganisms into CHSs may be a potential problem for microbial ecology studies since the very nature of the CHSs makes it difficult to completely exclude all environmental contaminants. In contrast to Solfatara water and fumarolic deposits communities, Solfatara mud communities was dominated by *Sulfobacillus* (10–34%) and *Acidibacillus* (4–20%). However, unlike Pisciarelli, the differences across Solfatara substrates were not significant (ANOSIM species: $R = 0.03$, $p = 0.50$; genus: $R = 0.03$, $p = 0.43$) suggesting that variations in microbial community structure across substrates is not a general phenomenon but is specific to hydrothermal systems and possibly linked to the physical and geochemical conditions of individual systems (Wang et al., 2014). The microbial community structure observed in our study is in contrast with studies that investigated the microbial ecology of Solfatara-Pisciarelli CHSs and identified *Acidithiobacillus* and *Metallosphaera* as the dominant genus in Solfatara and Pisciarelli, respectively (Iacono et al., 2020; Strazzulli et al., 2020). Furthermore, the dominance of archaea which we observed in Pisciarelli mud pools is in contrast with a study that reported a low archaea/bacteria ratio in high temperature waters (>85°C) of Pisciarelli (Crognale et al., 2022). This contrast between our study and other Solfatara-Pisciarelli studies is likely the consequence of a microbial diversity that reflects differences in sampling year/periods from active CHSs with variable geochemistry. Additionally, the dissimilarity may be related to differences in sampling points (10 cm depth vs. 1 cm depth), sample handling, DNA extraction protocol, bias introduced by 16S rRNA gene primers or sequencing technology. An assessment of the effects of environmental variables on microbial community composition showed that pH and Eh (Mantel: $\rho = 0.53$ and $\rho = 0.47$, respectively, $p < 0.05$) had the strongest correlation to beta diversity which is consistent with reports from other studies that pH and redox potential shape microbial community composition (Alsop et al., 2014; Power et al., 2018). The concentrations of NH_4^+ and SO_4^{2-} (Mantel: $\rho = 0.38$ and $\rho = 0.23$, respectively, $p < 0.05$) also had significant correlations to beta diversity. However, temperature had no significant (Mantel: $\rho = 0.1$, $p = 0.14$) effect on the microbial community composition of Solfatara-Pisciarelli CHSs, which coincides with temperature trends observed for alpha diversity.

Diversity, abundance and metabolic potential of Solfatara-Pisciarelli MAGs

To identify the metabolic potential of members of the microbial community and their connection to geochemistry, we generated MAGs. After co-assembly and binning, a total of 22 MAGs (10 from Solfatara and 12 from Pisciarelli) with >50% completeness and <10% contamination were recovered. Out of the 22 MAGs recovered, 16 had completeness >90% and contamination <5%. Taxonomic classification revealed that MAGs belonged to three bacterial (*Firmicutes*, *Proteobacteria*, and *Aquificota*) and 2 archaeal (*Thermoproteota* also known as *Crenarchaeota* and *Thermoplasmatota*) phyla (Supplementary Tables 10, 11), with over 40% (9 out of 22) of unclassified genomes at the genus level suggesting that our MAGs represent novel genera. MAGs belonging to the phylum *Thermoproteota* were most represented in Pisciarelli water and mud while *Firmicutes* were more abundant in all Solfatara substrates and Pisciarelli fumarolic deposits (Figure 8A), which is consistent with our results from metagenome fragments. On average, the most abundant MAGs in Pisciarelli water and mud belong to species of the *Acidianus* and *Pyrobaculum* genera and an unknown species of the family *Sulfolobaceae* whereas the most abundant MAGs in fumarolic deposits belong to 2 unknown species: one species of the *Alicyclobacillaceae* family and the other species of the *Alicyclobacillales* order and a third species belonging to the *Acidianus* genus (Supplementary Table 11). In contrast, the most abundant MAG in all Solfatara samples belong to species of the *Acidianus* genus. When MAGs were dereplicated based on FastANI of ≥95% we found that only 2 species (*Acidianus infernus*; and an unknown species of the *Alicyclobacillaceae* family) were shared between locations further confirming that although an overlap exists between both

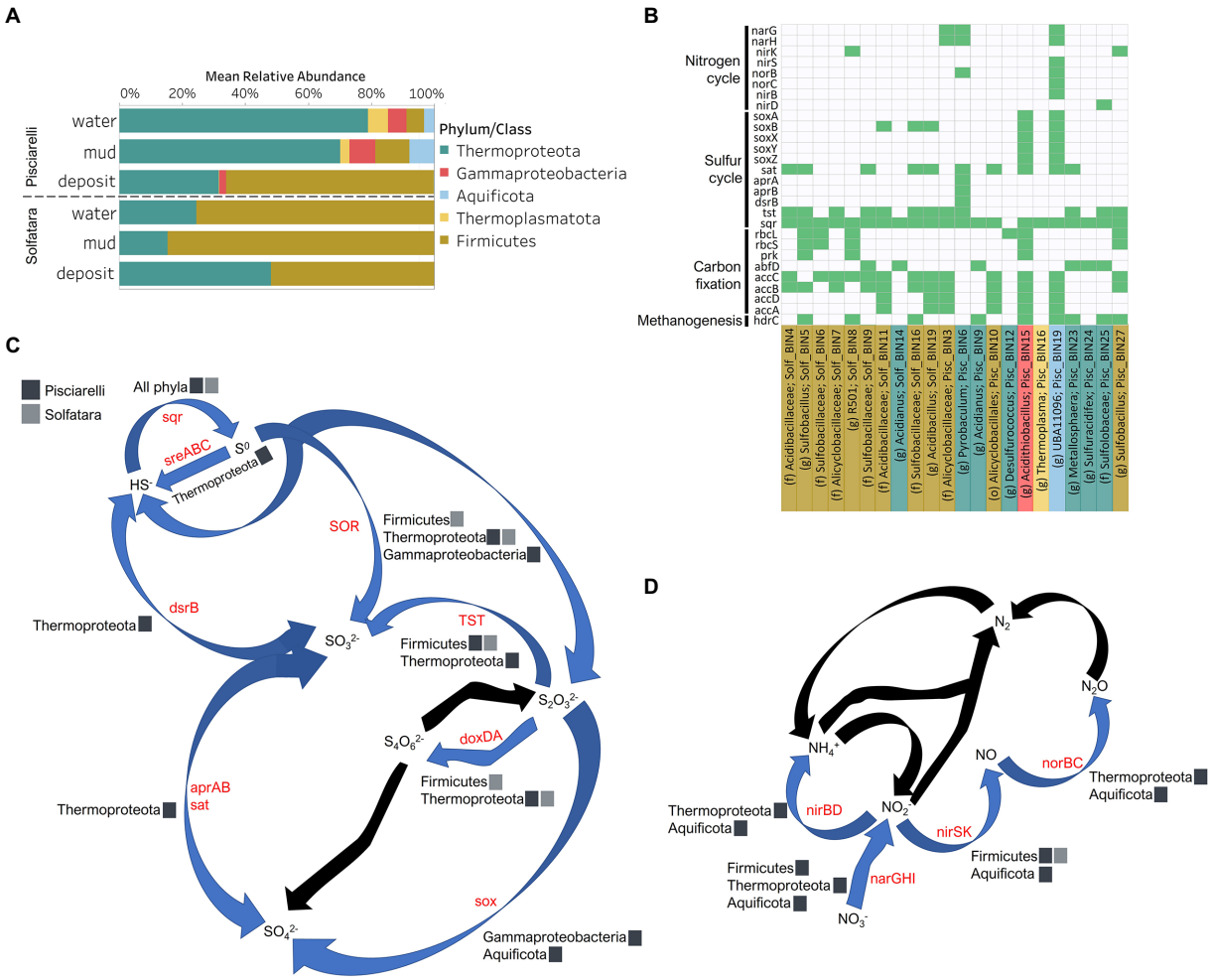


FIGURE 8
Taxonomy and functional potential of Metagenome Assembled Genomes (MAGs) in Solfatara-Pisciarelli hydrothermal systems. **(A)** Normalized mean (relative) abundance of sequences assigned to each MAG at the phylum/class level in each substrate. **(B)** Tile plot of selected marker genes (*y*-axis) present in each MAG (*x*-axis). Green color represents presence of gene in MAG. MAGs on the *x*-axis are colored based on phylum/class. Each MAG is labeled either by genus level assignment (g) or by the lowest taxonomic level assigned (f; family and o; order) to the MAG. **(C)** Sulfur cycle and **(D)** Nitrogen cycle. Phyla/Class from the different locations participating in each process are indicated. Gene participating in each process are indicated in red text. Black arrows denote processes that were not found in MAGs.

locations; they have different microbial community structures. Since acidic hydrothermal environments are dominated by chemolithoautotrophic communities that derive energy from inorganic compounds, we analyzed MAGs for marker genes involved in carbon fixation, sulfur, nitrogen, and methane metabolism.

Carbon Metabolism

Organic and inorganic carbon isotope (wt% and $\delta^{13}C$) values (Figures 9A,C) were very similar indicating that carbon measured at both locations is mostly organic and falls within the range typical for microorganisms with varied carbon fixation pathways (Sharp, 2007; Havig et al., 2011), which is also indicated by the functional potential of our MAGs. Genes encoding the key enzymes for Calvin-Benson-Bassham (CBB) cycle, ribulose-biphosphate carboxylase (*rbc*) and phosphoribulokinase (*prk*) were found in *Firmicutes* MAGs (*Sulfobacillus* and *R501*) as well as *Gammaproteobacteria* MAG (*Acidithiobacillus*; Supplementary Table 12; Figure 8B) similar to what has been reported in other studies (Caldwell et al., 2007; Cerqueira et al., 2018). Owing to the abundance of *Firmicutes* in Pisciarelli fumarolic deposits and Solfatara microbial communities, CBB may be the primary mode of carbon fixation in these environments. The *rbc* gene was also detected in an archaeal MAG (*Desulfurococcus*); however, in archaea, this gene encodes enzymes reported to be involved in the reductive hexulose-phosphate pathway (Kono et al., 2017). The marker gene encoding 4-hydroxybutyryl-CoA dehydratase (*abfD*), an enzyme in 3-hydroxypropionate/4-hydroxybutyrate (3-HP/4-HB) cycle, was present in all MAGs of the *Sulfolobaceae* family (*Acidianus*, *Metallosphaera*, *Sulfuracidifex*) which is consistent with the presence of the 3-HP/4-HB cycle in the crenarchaeal order *Sulfolobales* (Berg et al., 2010). *Crenarchaeota* especially the order *Sulfolobales* were abundant in Pisciarelli water and mud; hence, 3-HP/4-HB cycle may be the primary mode of carbon fixation in Pisciarelli mud pools. The *abfD* gene was also detected in one *Firmicutes* MAG (unknown species of the *Sulfobacillaceae* family) from Solfatara. The 3-HP/4-HB cycle to the best of our knowledge has not been reported in *Sulfobacillaceae* and the absence of other genes encoding enzymes involved in the 3-HP/4-HB cycle implies that this cycle may be absent in this *Sulfobacillaceae* MAG. One possible explanation is that the enzyme encoded by the *abfD*

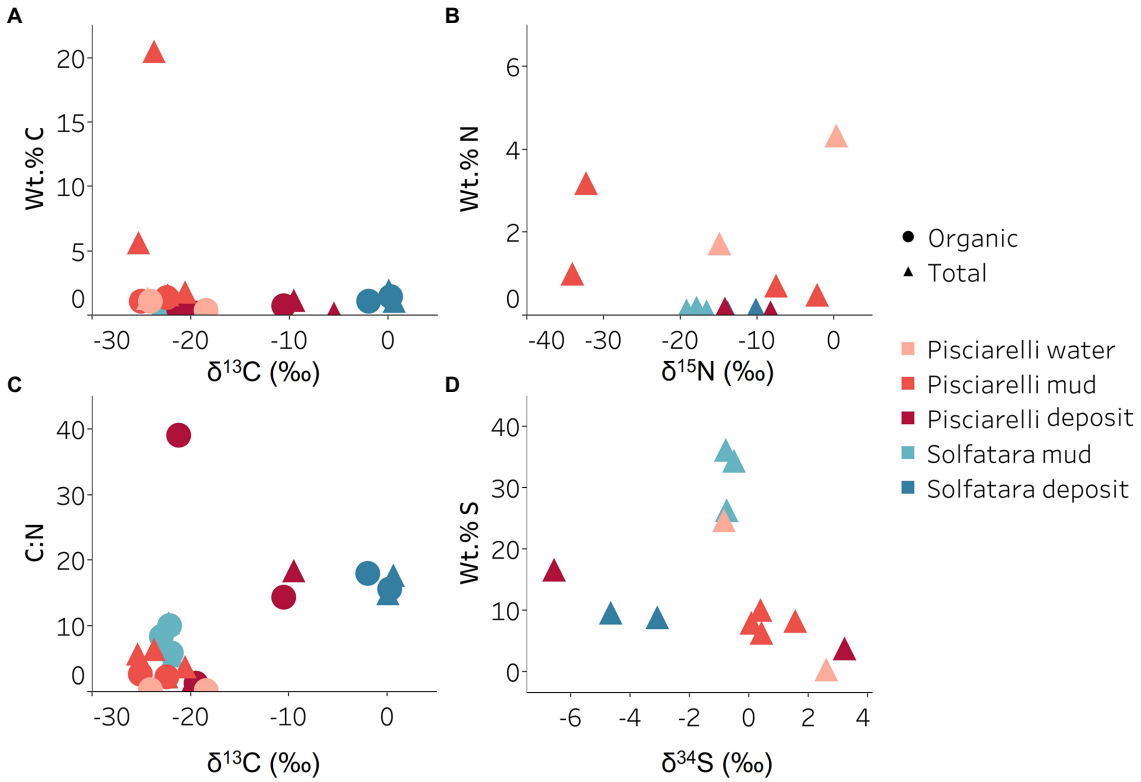


FIGURE 9
Carbon, nitrogen, and sulfur isotope values measured in Solfatara and Pisciarelli hydrothermal systems. **(A)** Total Carbon (TC) wt.%, Total Organic Carbon (TOC) wt.% and $\delta^{13}C$ values. The wt.% values of TC and TOC are very similar indicating that carbon measured in the samples is mostly organic. The $\delta^{13}C$ values exhibit a wide range of values. **(B)** Total Nitrogen wt.% and $\delta^{15}N$ values. Pisciarelli mud and water have the highest concentrations of N. The $\delta^{15}N$ values exhibit a large variation with values ranging from 0.41 to −34.04‰. **(C)** C:N ratio and $\delta^{13}C$ values both point out that most of the carbon measured shows values typical for microorganisms with various carbon fixation pathways. **(D)** Total Sulfur wt.% and $\delta^{34}S$ exhibit values characteristic for oxidized fumarolic H_2S and sulfur with sedimentary origin.

gene in this *Sulfobacillaceae* MAG may be involved in an alternative metabolic pathway like has been reported for *Clostridium* (Scherf and Buckel, 1993; Scherf et al., 1994). Another possible explanation is that the *Sulfobacillaceae* MAG may have acquired the *abfD* gene *via* horizontal gene transfer.

Genes encoding the key enzymes in Wood-Ljungdahl (WL) pathway and 3-hydroxyprioponate (3-HP) bicycle were absent in all MAGs and in the entire metagenome dataset, suggesting that both pathways may not be important for carbon fixation at both locations. The marker gene encoding ATP citrate lyase (*aclAB*), an enzyme involved in reductive tricarboxylic acid (rTCA) cycle, was also absent in all MAGs. ATP citrate lyase catalyzes citrate cleavage in rTCA cycle but in *Aquificaceae*, this reaction is catalyzed by the combined action of citryl-CoA synthetase (*ccsAB*) and citryl-CoA lyase (*ccl*) (Aoshima et al., 2004a,b) which were found in our *Aquificota* MAG (*UBA11096*) from Pisciarelli metagenome indicating potential for rTCA.

Potential for methanogenesis was absent in MAGs from both locations. Although the gene encoding hydrogen disulfide reductase (*hdrABC*) involved in methanogenesis was found in *Firmicutes* (*Sulfobacillus*, *R501*, and unknown species of the *Sulfobacillaceae* family), *Gammaproteobacteria* (*Acidithiobacillus*), *Aquificota* (*UBA11096*), and *Thermoproteota* (*Metallosphaera*) MAGs, the gene encoding key enzyme, methyl-CoM reductase (*mcr*) in methanogenesis was absent. Given the absence of *mcr* gene, the absence of genes encoding other enzymes involved in methanogenesis especially in the *Metallosphaera* MAG that was 100% complete, and the absence of $\delta^{13}C$ values lower than −30‰, it is likely that the *hdrABC genes* in these MAGS are not used in methanogenesis but are most likely used in sulfur metabolism (Mander et al., 2004; Wang et al., 2019). Potential for methane oxidation (methanol dehydrogenase; *mdh1* and particulate methane monooxygenase; *pmo*) was also absent in all MAGs.

Sulfur metabolism

Sulfur isotope values ($\delta^{34}S$) measured in deposits from both locations (Figure 9D) are characteristic of sulfur with sedimentary origin, but $\delta^{34}S$ values (−0.83 to 2.62‰) measured in water and mud where oxidation of H_2S is extensive, fall within the range for SO_4^{2-} formed from sulfide oxidation (−2.5 to 2.4‰; Allard et al., 1991). In acidic environments like Solfatara and Pisciarelli, abiotic oxidation of H_2S-rich gas produced from the disproportionation of volcanic SO_2 is inhibited (D'Imperio et al., 2008), thus providing a great habitat for sulfur-oxidizing microorganisms that utilize H_2S for energy generation. The abundance in H_2S is consistent with the presence of genes encoding enzymes involved in the oxidation of a variety of reduced sulfur compounds (H_2S, S^0 or $S_2O_3^{2-}$) at both locations. Genes encoding the enzyme, sulfide:quinone oxidoreductases (*sqr*) for H_2S oxidation to elemental sulfur (S^0) were detected in all MAGs except one *Thermoproteota* MAG (*Desulfurococcus*; Figures 8B,C). The presence of *sqr* gene in most MAGs is indicative of the importance of H_2S oxidation to microbial communities at both locations. S^0 formed can be reduced

to H_2S by the enzyme molybdopterin sulfur reductase (*sreABC*) or disproportionated to H_2S, thiosulfate ($S_2O_3^{2-}$) and sulfite (SO_3^{2-}) by sulfur oxygenase reductase (*SOR*). The *sreABC* gene was found in *Thermoproteota* MAG (*Acidianus*) while *SOR* gene was found in *Firmicutes* (*Sulfobacillus*, *R501*, and *Acidibacillus*) *Gammaproteobacteria* (*Acidithiobacillus*) and *Thermoproteota* (*Acidianus* and *Sulfuracidifex*) MAGs.

Thiosulfate ($S_2O_3^{2-}$) may not be available for microbial metabolism since it disproportionates to S^0 and SO_3^{2-} at pH < 4 (Nordstrom et al., 2004); however, microbial communities exhibited potential to utilize $S_2O_3^{2-}$. Thiosulfate can be oxidized to SO_4^{2-} *via* the *Sox* system (*soxXYZABCD*) which we detected in *Gammaproteobacteria* (*Acidithiobacillus*) and *Aquificota* (*UBA11096*) MAGs. In both MAGs, *soxCD* was absent suggesting that they oxidize $S_2O_3^{2-}$ to S^0 instead of SO_4^{2-} which is consistent with studies that identified the *sox* cluster without *soxCD* in *Acidithiobacillus* (Wang et al., 2019). Thiosulfate can also be oxidized to tetrathionate ($S_4O_6^{2-}$) by thiosulfate:quinol oxidoreductase (*doxDA*) or reduced to SO_3^{2-} by thiosulfate:cyanide sulfur transferase (*tst*). The *doxDA* gene was found in all *Sulfolobaceae* MAGs (*Acidianus*, *Metallosphaera*, *Sulfuracidifex*) and in the *Firmicutes* (*R501*) MAG, while *tst* gene was found in 2 *Thermoproteota* (*Pyrobaculum*, *Metallosphaera*) MAGs and most *Firmicutes* MAGs. H_2S can be regenerated by reducing SO_4^{2-} *via* dissimilatory sulfate reduction (DSR). Genes encoding enzymes (dissimilatory sulfite reductase; *dsrAB*, adenylylsulfate reductase; *aprAB*, and sulfate adenylyltransferase; *sat*) involved in DSR were detected in one *Thermoproteota* (*Pyrobaculum*) MAG. The presence of both genes encoding enzymes involved in sulfur oxidation and reduction in some MAGs implies that microbial communities in Solfatara-Pisciarelli systems may take advantage of the high energy yield that results from coupling sulfur oxidation to sulfur reduction, which has been previously observed in other microbial community studies (Glamoclija et al., 2004; Crognale et al., 2022). Our results are consistent with studies that found the abundance of organisms capable of utilizing sulfur (Inskeep et al., 2013) and an enrichment of genes (*sqr*, *doxDA*, and *SOR*) involved in metabolizing reduced sulfur compounds in acidic hot springs (Colman et al., 2019a).

Nitrogen metabolism

Our results showed that nitrogen metabolism may not be an important energy-yielding pathway for microbial communities even in Pisciarelli mud pools where NH_4^+ was abundant (Figures 8B,D). Exceptionally low $\delta^{15}N$ values were observed at both locations (Figure 9B) and the lowest $\delta^{15}N$ values ever measured were found in the Pisciarelli epilithic microbial layer. These extremely low $\delta^{15}N$ values are difficult to explain in the context of metabolic potential present in our MAGs or metagenome fragments. Low and negative $\delta^{15}N$ values are sometimes linked to microbial ammonia oxidation and/or nitrogen fixation (Tozer et al., 2005; Havig et al., 2011). Although Crognale et al. (2022) reported ammonia oxidation to be the most likely metabolic pathway in Pisciarelli mud pools due to its high energy yield, the absence of genes encoding enzymes involved in ammonia oxidation (ammonia monooxygenase; *amoCAB* and hydroxylamine dehydrogenase; *hao*) or nitrogen fixation (*nifD*, *nifK*, and *nifH*) in our MAGs and metagenome dataset indicates that microbial nitrogen metabolism cannot explain the geochemistry we observed. Previously observed negative $\delta^{15}N$ values (about −10‰) have been associated with the dissolution of magmatic nitrogen in water at elevated temperatures and pressures (Labidi et al., 2020). While it is very likely, the mechanism by which magmatic nitrogen, high temperature and pressure would contribute to the extremely low $\delta^{15}N$ values measured at Solfatara and Pisciarelli is unclear.

Evidence of denitrification which has not been linked to low and negative $\delta^{15}N$ values was found in five MAGs (*Sulfobacillus*, *R501*, unknown species of the *Alicyclobacillaceae* family, *Pyrobaculum*, and *UBA11096*). However, none of the MAGs carried all the genes encoding enzymes for the complete denitrification pathway (nitrate reductase, *narGH* or *napA*; nitrite reductase, *nirK*; nitric oxide reductase, *norB*; nitrous oxide reductase, *nosZ*). The complete set of genes for denitrification were also absent in metagenome fragments. One Pisciarelli MAG (*UBA11096*) belonging to the family *Aquificaceae*, carried most of the denitrification genes except *nosZ* which is consistent with studies that found nitrous oxide (N_2O) to be the end-product of denitrification in members of *Aquificaceae* (Nakagawa et al., 2004). Dissimilatory nitrate reduction to ammonia *via* nitrite reductases (*nirBD*) was detected only in the *Aquificaceae* MAG from Pisciarelli.

Conclusion

Results from this study showed that Solfatara and Pisciarelli are lithologically similar but geochemically distinct. At Pisciarelli, we observed varying geochemistry among mud pools and fumarolic deposits, which results from subsurface fluid–rock interactions, whereas Solfatara substrates were geochemically similar. Shannon diversity was significantly different between locations but showed no significant difference across substrates. We found pH to be the most important driver of alpha diversity and microbial community composition. NMDS plot and ANOSIM showed that Solfatara's microbial community structure exhibited no significant difference across substrates which also coincides with observed trends in geochemistry. In contrast, at Pisciarelli, microbial community structure followed trends in fluid availability rather than geochemical trends. Geochemically, water and mud from Pisciarelli large mud pool and water from Pisciarelli small mud pool clustered distinctly from fumarolic deposits, discharge channel mud and mud from Pisciarelli small pool. However, based on microbial communities, all water and mud samples clustered distinctly from fumarolic deposits that were the driest substrates analyzed; results showed that Pisciarelli fumarolic deposits were more similar to Solfatara samples than samples from Pisciarelli mud pools. Overall, the genus *Acidianus* dominated Pisciarelli water and *Pyrobaculum* was more abundant in Pisciarelli mud. Interestingly, Pisciarelli fumarolic deposits, epilithic microbial layer, and mud from small mud pool had a high number of viral sequences. Solfatara and Pisciarelli had distinct microbial communities; on average, *Acidianus* and *Pyrobaculum* dominated Pisciarelli while *Sulfobacillus* and *Acidibacillus* were dominant at Solfatara. Although microbial communities were distinct between the two locations and between Pisciarelli substrates, MAGs indicated functional redundancy. For example, the potential to reduce H_2S to S^0 was observed in almost all MAGs. Further, MAGs showed that microbial communities inhabiting both locations took advantage of the major volcanic gases (CO_2 and H_2S) *via* carbon fixation and sulfur oxidation. The 3HP/4-HB pathway found in most archaeal MAGs is most likely the primary mode of carbon fixation in Pisciarelli mud pools dominated by

(hyper)thermophilic archaea whereas CBB cycle is most likely the primary carbon fixation pathway in all other environments where bacteria was abundant. At both locations, sulfur cycling is represented by oxidation of H_2S-rich volcanic gas and other reduced sulfur compounds. Comparative geochemical and metagenomic analyses demonstrate that ecological differences across substrates are not a widespread phenomenon but specific to the system. Therefore, this study demonstrates the importance of analyzing different substrates of CHSs to understand the full range of microbial ecology to avoid biased ecological assessment.

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/, Bioproject accession PRJNA889931.

Author contributions

MG, AS, MF, GN, and CT designed and organized the project. MG, RS, MP, AM collected the samples. IU, RB, MF, and MG collected the data. IU, MG, and MF contributed to the data processing and analyses. IU and MG wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was enabled through the Alfred P. Sloan Foundation's support of the Deep Carbon Observatory Deep Earth Carbon Degassing program (DECADE) to MG and Rutgers Faculty Program Start Up. IU was supported by Rutgers University Transform Graduate Fellowship.

Acknowledgments

We would like to thank Ying Lin at UCR for re-running samples to confirm $\delta^{15}N$ values and Ashley E. Murphy for running samples on XRF. We thank the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and the associated resources that have contributed to the results reported in this manuscript. We thank two reviewers for their thoughtful and constructive comments which have significantly improved this 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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1066406/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Ballot based on Principal Component Analysis (PCA) using bulk elemental composition of mud and fumarolic deposits.

SUPPLEMENTARY FIGURE 2
Taxonomic profile of Solfatara and Pisciarelli microorganisms based on metagenomic profiling. Sequences not classified at the domain level and species with relative abundance less than 0.01% were grouped into unclassified and low abundance species. All kitome were grouped into contaminant. Samples collected from Pisciarelli large pool are denoted by PLP, Pisciarelli small pool by PSP, Pisciarelli fumarole by PF, Solfatara mud pool by SMP, and Solfatara fumarole by SF. W represents water, M for mud, MO for mud outlet, E for epilithic microbial layer from dry mud wall, and D for deposits samples. Samples labeled a and b are replicates sampled from the same spot. Sample SMP-M_a1 and SMP-M_a1.2 are sequencing replicates.

SUPPLEMENTARY FIGURE 3
Relative abundance of kitome families. Kitome families with average abundance of <1% were grouped into Others <1%.

SUPPLEMENTARY FIGURE 4
Estimation of the coverage of microbial communities in **(A)** Solfatara and **(B)** Pisciarelli. The empty circles on the curves represent community coverage estimate at the sequencing effort applied. The top and bottom horizontal red dashed indicate 100 and 95% average community coverage, respectively. The arrows are the Nonpareil diversity estimates.

References

Aiuppa, A., Tamburello, G., Di Napoli, R., Cardellini, C., Chiodini, G., Giudice, G., et al. (2013). First observations of the fumarolic gas output from a restless caldera: implications for the current period of unrest (2005–2013) at Campi Flegrei. *Geochem. Geophys.* 14, 4153–4169. doi: 10.1002/ggge.20261

Allard, P., Maiorani, A., Tedesco, D., Cortecci, G., and Turi, B. (1991). Isotopic study of the origin of sulfur and carbon in Solfatara fumaroles, Campi Flegrei caldera. *J. Volcanol. Geotherm. Res.* 48, 139–159. doi: 10.1016/0377-0273(91)90039-3

Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103

Alsop, E. B., Boyd, E. S., and Raymond, J. (2014). Merging metagenomics and geochemistry reveals environmental controls on biological diversity and evolution. *BMC Ecol.* 14:16. doi: 10.1186/1472-6785-14-16

Amenabar, M. J., and Boyd, E. S. (2019). A review of the mechanisms of mineral-based metabolism in early earth analog rock-hosted hydrothermal ecosystems. *World J. Microbiol. Biotechnol.* 35:29. doi: 10.1007/s11274-019-2604-2

Amenabar, M. J., Shock, E. L., Roden, E. E., Peters, J. W., and Boyd, E. S. (2017). Microbial substrate preference dictated by energy demand rather than supply. *Nat. Geosci.* 10, 577–581. doi: 10.1038/ngeo2978

Aoshima, M., Ishii, M., and Igarashi, Y. (2004a). A novel enzyme, citryl-CoA lyase, catalysing the second step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. *Mol. Microbiol.* 52, 763–770. doi: 10.1111/j.1365-2958.2004.04010.x

Aoshima, M., Ishii, M., and Igarashi, Y. (2004b). A novel enzyme, citryl-CoA synthetase, catalysing the first step of the citrate cleavage reaction in Hydrogenobacter thermophilus TK-6. *Mol. Microbiol.* 52, 751–761. doi: 10.1111/j.1365-2958.2004.04009.x

Arif, S., Willenberg, C., Dreyer, A., Nacke, H., and Hoppert, M. (2021). Sasso Pisano geothermal field environment Harbours diverse Ktedonobacteria representatives and illustrates habitat-specific adaptations. *Microorganisms* 9:1402. doi: 10.3390/microorganisms9071402

Aulitto, M., Gallo, G., Puopolo, R., Mormone, A., Limauro, D., Contursi, P., et al. (2021). Genomic insight of Alicyclobacillus Mali FL18 isolated from an arsenic-rich hot spring. *Front. Microbiol.* 12:669. doi: 10.3389/fmicb.2021.639697

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

Benson, C. A., Bizzoco, R. W., Lipson, D. A., and Kelley, S. T. (2011). Microbial diversity in nonsulfur, sulfur and iron geothermal steam vents. *FEMS Microbiol. Ecol.* 76, 74–88. doi: 10.1111/j.1574-6941.2011.01047.x

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* 8, 447–460. doi: 10.1038/nrmicro2365

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

Boyd, E. S., Hamilton, T. L., Spear, J. R., Lavin, M., and Peters, J. W. (2010). [FeFe]-hydrogenase in Yellowstone National Park: evidence for dispersal limitation and phylogenetic niche conservatism. *ISME J.* 4, 1485–1495. doi: 10.1038/ismej.2010.76

Boyer, G. M., Schubotz, F., Summons, R. E., Woods, J., and Shock, E. L. (2020). Carbon oxidation state in microbial polar lipids suggests adaptation to hot spring temperature and redox gradients. *Front. Microbiol.* 11:229. doi: 10.3389/fmicb.2020.00229

Brock, T. D. (1985). Life at high temperatures. *Science* 230, 132–138. doi: 10.1126/science.230.4722.132

Brumfield, K. D., Huq, A., Colwell, R. R., Olds, J. L., and Leddy, M. B. (2020). Microbial resolution of whole genome shotgun and 16S amplicon metagenomic sequencing using publicly available NEON data. *PLoS One* 15:e0228899. doi: 10.1371/journal.pone.0228899

Caldwell, P. E., MacLean, M. R., and Norris, P. R. (2007). Ribulose bisphosphate carboxylase activity and a Calvin cycle gene cluster in *Sulfobacillus* species. *Microbiology* 153, 2231–2240. doi: 10.1099/mic.0.2007/006262-0

Caliro, S., Chiodini, G., Moretti, R., Avino, R., Granieri, D., Russo, M., et al. (2007). The origin of the fumaroles of La Solfatara (Campi Flegrei, South Italy). *Geochim. Cosmochim. Acta* 71, 3040–3055. doi: 10.1016/j.gca.2007.04.007

Cerqueira, T., Barroso, C., Froufe, H., Egas, C., and Bettencourt, R. (2018). Metagenomic signatures of microbial communities in deep-sea hydrothermal sediments of Azores vent fields. *Microb. Ecol.* 76, 387–403. doi: 10.1007/s00248-018-1144-x

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. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Chiodini, G., Avino, R., Caliro, S., and Minopoli, C. (2011). Temperature and pressure gas geoindicators at the Solfatara fumaroles (Campi Flegrei). *Ann. Geophys.* 54:107245. doi: 10.4401/ag-5002

Chiodini, G., Caliro, S., Avino, R., Bini, G., Giudicepietro, F., De Cesare, W., et al. (2021). Hydrothermal pressure-temperature control on CO_2 emissions and seismicity at Campi Flegrei (Italy). *J. Volcanol. Geotherm. Res.* 414:107245. doi: 10.1016/j.jvolgeores.2021.107245

Chiodini, G., Caliro, S., Cardellini, C., Granieri, D., Avino, R., Baldini, A., et al. (2010). Long-term variations of the Campi Flegrei, Italy, volcanic system as revealed by the monitoring of hydrothermal activity. *J. Geophys. Res. Solid Earth* 115:B03205. doi: 10.1029/2008JB006258

Chiodini, G., Frondini, F., Cardellini, C., Granieri, D., Marini, L., and Ventura, G. (2001). CO_2 degassing and energy release at Solfatara volcano, Campi Flegrei, Italy. *J. Geophys. Res. Solid Earth* 106, 16213–16221. doi: 10.1029/2001JB000246

Ciniglia, C. V. (2005). Influences of geochemical and mineralogical constraints on algal distribution in acidic hydrothermal environments: Pisciarelli (Naples, Italy) as a model site. *Arch. Hydrobiol.* 162, 121–142. doi: 10.1127/0003-9136/2005/0162-0121

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., Feyhl-Buska, J., Robinson, K. J., Fecteau, K. M., Xu, H., Shock, E. L., et al. (2016). Ecological differentiation in planktonic and sediment-associated chemotrophic microbial populations in Yellowstone hot springs. *FEMS Microbiol. Ecol.* 92:137. doi: 10.1093/femsec/fiw137

Colman, D. R., Lindsay, M. R., Amenabar, M. J., and Boyd, E. S. (2019a). The intersection of geology, geochemistry, and microbiology in continental hydrothermal systems. *Astrobiology* 19, 1505–1522. doi: 10.1089/ast.2018.2016

Colman, D. R., Lindsay, M. R., and Boyd, E. S. (2019b). Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat. Commun.* 10:681. doi: 10.1038/s41467-019-08499-1

Cox, A., Shock, E. L., and Havig, J. R. (2011). The transition to microbial photosynthesis in hot spring ecosystems. *Chem. Geol.* 280, 344–351. doi: 10.1016/j.chemgeo.2010.11.022

Crognale, S., Venturi, S., Tassi, F., Rossetti, S., Cabassi, J., Capecchiacci, F., et al. (2022). Geochemical and microbiological profiles in hydrothermal extreme acidic environments (Pisciarelli Spring, Campi Flegrei, Italy). *FEMS Microbiol. Ecol.* 98:fiac088. doi: 10.1093/femsec/fiac088

Crognale, S., Venturi, S., Tassi, F., Rossetti, S., Rashed, H., Cabassi, J., et al. (2018). Microbiome profiling in extremely acidic soils affected by hydrothermal fluids: the case of the Solfatara crater (Campi Flegrei, southern Italy). *FEMS Microbiol. Ecol.* 94:fiy190. doi: 10.1093/femsec/fiy190

De Natale, G., Troise, C., Mark, D., Mormone, A., Piochi, M., Di Vito, M. A., et al. (2016). The Campi Flegrei deep drilling project (CFDDP): new insight on caldera structure, evolution and hazard implications for the Naples area (Southern Italy). *Geochem. Geophys.* 17, 4836–4847. doi: 10.1002/2015GC006183

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05

D'Imperio, S., Lehr, C. R., Oduro, H., Druschel, G., Kühl, M., and McDermott, T. R. (2008). Relative importance of H2 and H2S as energy sources for primary production in geothermal springs. *Appl. Environ. Microbiol.* 74, 5802–5808. doi: 10.1128/AEM.00852-08

Ellis, D. G., Bizzoco, R. W., and Kelley, S. T. (2008). Halophilic archaea determined from geothermal steam vent aerosols. *Environ. Microbiol.* 10, 1582–1590. doi: 10.1111/j.1462-2920.2008.01574.x

Fedele, A., Somma, R., Troise, C., Holmberg, K., De Natale, G., and Matano, F. (2021). Time-lapse landform monitoring in the Pisciarelli (Campi Flegrei-Italy) fumarole field using UAV photogrammetry. *Remote Sens.* 13:118. doi: 10.3390/rs13010118

Fournier, R. O. (1989). Geochemistry and dynamics of the Yellowstone National Park hydrothermal system. *Annu. Rev. Earth Planet. Sci.* 17, 13–53. doi: 10.1146/annurev.ea.17.050189.000305

Fullerton, K. M., Schrenk, M. O., Yücel, M., Manini, E., Basili, M., Rogers, T. J., et al. (2021). Effect of tectonic processes on biosphere–geosphere feedbacks across a convergent margin. *Nat. Geosci.* 14, 301–306. doi: 10.1038/s41561-021-00725-0

Glamoclija, M., Garrel, L., Berthon, J., and López-García, P. (2004). Biosignatures and bacterial diversity in hydrothermal deposits of Solfatara Crater, Italy. *Geomicrobiol J.* 21, 529–541. doi: 10.1080/01490450490888235

Hamilton, T., Vogl, K., Bryant, D., Boyd, E., and Peters, J. (2012). Environmental constraints defining the distribution, composition, and evolution of chlorophototrophs in thermal features of Yellowstone National Park. *Geobiology* 10, 236–249. doi: 10.1111/j.1472-4669.2011.00296.x

Havig, J. R., Raymond, J., Meyer-Dombard, D. A. R., Zolotova, N., and Shock, E. L. (2011). Merging isotopes and community genomics in a siliceous sinter-depositing hot spring. *J. Geophys. Res.* 116:G01005. doi: 10.1029/2010JG001415

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

Huber, R., Sacher, M., Vollmann, A., Huber, H., and Rose, D. (2000). Respiration of arsenate and selenate by hyperthermophilic archaea. *Syst. Appl. Microbiol.* 23, 305–314. doi: 10.1016/S0723-2020(00)80058-2

Husson, O. (2013). Redox potential (eh) and pH as drivers of soil/plant/microorganism systems: a transdisciplinary overview pointing to integrative opportunities for agronomy. *Plant Soil* 362, 389–417. doi: 10.1007/s11104-012-1429-7

Iacono, R., Cobucci-Ponzano, B., De Lise, F., Curci, N., Maurelli, L., Moracci, M., et al. (2020). Spatial metagenomics of three geothermal sites in Pisciarelli hot spring focusing on the biochemical resources of the microbial consortia. *Molecules* 25:4023. doi: 10.3390/molecules25174023

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

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

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., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359

Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W. T., Wolfe, A. J., et al. (2019). Controlling for contaminants in low-biomass 16S rRNA gene sequencing experiments. *mSystems* 4:e00290-19. doi: 10.1128/mSystems.00290-19

Kimbrough, D. E., Kouame, Y., Moheban, P., and Springthorpe, S. (2006). The effect of electrolysis and oxidation–reduction potential on microbial survival, growth, and disinfection. *Int. J. Environ. Pollut.* 27, 211–221. doi: 10.1504/IJEP.2006.010464

Kono, T., Mehrotra, S., Endo, C., Kizu, N., Matusda, M., Kimura, H., et al. (2017). A RuBisCO-mediated carbon metabolic pathway in methanogenic archaea. *Nat. Commun.* 8:14007. doi: 10.1038/ncomms14007

Labidi, J., Barry, P. H., Bekaert, D. V., Broadley, M. W., Marty, B., Giunta, T., et al. (2020). Hydrothermal 15 N 15 N abundances constrain the origins of mantle nitrogen. *Nature* 580, 367–371. doi: 10.1038/s41586-020-2173-4

Lezcano, M. Á., Moreno-Paz, M., Carrizo, D., Prieto-Ballesteros, O., Fernández-Martínez, M. Á., Sánchez-García, L., et al. (2019). Biomarker profiling of microbial mats in the geothermal band of Cerro Caliente, Deception Island (Antarctica): life at the edge of heat and cold. *Astrobiology* 19, 1490–1504. doi: 10.1089/ast.2018.2004

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Lindsay, M. R., Amenabar, M. J., Fecteau, K. M., Debes, R. V., Fernandes Martins, M. C., Fristad, K. E., et al. (2018). Subsurface processes influence oxidant availability and chemoautotrophic hydrogen metabolism in Yellowstone hot springs. *Geobiology* 16, 674–692. doi: 10.1111/gbi.12308

Lindsay, M. R., Colman, D. R., Amenabar, M. J., Fristad, K. E., Fecteau, K. M., Debes, R. V., et al. (2019). Probing the geological source and biological fate of hydrogen in Yellowstone hot springs. *Environ. Microbiol.* 21, 3816–3830. doi: 10.1111/1462-2920.14730

Lowenstern, J. B., Bergfeld, D., Evans, W. C., and Hunt, A. G. (2015). Origins of geothermal gases at Yellowstone. *J. Volcanol. Geotherm. Res.* 302, 87–101. doi: 10.1016/j.jvolgeores.2015.06.010

Mander, G. J., Pierik, A. J., Huber, H., and Hedderich, R. (2004). Two distinct heterodisulfide reductase-like enzymes in the sulfate-reducing archaeon Archaeoglobus profundus. *Eur. J. Biochem.* 271, 1106–1116. doi: 10.1111/j.1432-1033.2004.04013.x

Mardanov, A. V., Gumerov, V. M., Beletsky, A. V., and Ravin, N. V. (2018). Microbial diversity in acidic thermal pools in the Uzon caldera, Kamchatka. *Antonie Van Leeuwenhoek* 111, 35–43. doi: 10.1007/s10482-017-0924-5

Marlow, J. J., Colocci, I., Jungbluth, S. P., Weber, N. M., Gartman, A., and Kallmeyer, J. (2020). Mapping metabolic activity at single cell resolution in intact volcanic fumarole sediment. *FEMS Microbiol. Lett.* 367:fnaa031. doi: 10.1093/femsle/fnaa031

Martini, M., Giannini, L., Buccianti, A., Prati, F., Legittimo, P. C., Iozzelli, P., et al. (1991). 1980–1990: ten years of geochemical investigation at Phlegrean fields (Italy). *J. Volcanol. Geotherm. Res.* 48, 161–171. doi: 10.1016/0377-0273(91)90040-7

Medrano-Santillana, M., Souza-Brito, E. M., Duran, R., Gutierrez-Corona, F., and Reyna-López, G. E. (2017). Bacterial diversity in fumarole environments of the Paricutín volcano, Michoacán (Mexico). *Extremophiles* 21, 499–511. doi: 10.1007/s00792-017-0920-8

Menzel, P., Gudbergsdóttir, S. R., Rike, A. G., Lin, L., Zhang, Q., Contursi, P., et al. (2015). Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.* 70, 411–424. doi: 10.1007/s00248-015-0576-9

Menzel, P., Ng, K. L., and Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with kaiju. *Nat. Commun.* 7:11257. doi: 10.1038/ncomms11257

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server–a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386

Miller, S. R., Strong, A. L., Jones, K. L., and Ungerer, M. C. (2009). Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone National Park. *Appl. Environ. Microbiol.* 75, 4565–4572. doi: 10.1128/AEM.02792-08

Moreras-Marti, A., Fox-Powell, M., Zerkle, A. L., Stueeken, E., Gazquez, F., Brand, H. E., et al. (2021). Volcanic controls on the microbial habitability of Mars-analogue hydrothermal environments. *Geobiology* 19, 489–509. doi: 10.1111/gbi.12459

Moretti, R., De Natale, G., and Troise, C. (2020). *Hydrothermal Versus Magmatic: Geochemical Views and Clues Into the Unrest Dilemma at Campi FlegreiVesuvius, Campi Flegrei, and Campanian Volcanism*. Amsterdam, Netherlands: Elsevier, 371–406.

Munson-McGee, J. H., Peng, S., Dewerff, S., Stepanauskas, R., Whitaker, R. J., Weitz, J. S., et al. (2018). A virus or more in (nearly) every cell: ubiquitous networks of virus–host interactions in extreme environments. *ISME J.* 12, 1706–1714. doi: 10.1038/s41396-018-0071-7

Nakagawa, S., Nakamura, S., Inagaki, F., Takai, K., Shirai, N., and Sako, Y. (2004). Hydrogenivirga caldilitoris gen. Nov., sp. nov., a novel extremely thermophilic, hydrogen- and sulfur-oxidizing bacterium from a coastal hydrothermal field. *Int. J. Syst. Evol. Microbiol.* 54, 2079–2084. doi: 10.1099/ijs.0.03031-0

Nordstrom, D., Ball, J., and McCleskey, R. (2004). "Oxidation reactions for reduced Fe, As, and S in thermal outflows of" in *Water-Rock Interaction*. eds. R. Wanty and R. Seal (London, UK: Taylor and Francis Group), 59–62.

Nordstrom, D. K., Ball, J. W., and McCleskey, R. B. (2005). "Ground water to surface water: geochemistry of thermal outflows in Yellowstone National Park" in *Geothermal Biology and Geochemistry in Yellowstone National Park*. eds. W. Inskeep and T. R. McDermott (Bozeman, MT: Montana State University), 73–94.

Nordstrom, D. K., McCleskey, R. B., 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

Oliverio, A. M., Power, J. F., Washburne, A., Cary, S. C., Stott, M. B., and Fierer, N. (2018). The ecology and diversity of microbial eukaryotes in geothermal springs. *ISME J.* 12, 1918–1928. doi: 10.1038/s41396-018-0104-2

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

Piochi, M., Kilburn, C., Di Vito, M., Mormone, A., Tramelli, A., Troise, C., et al. (2014). The volcanic and geothermally active Campi Flegrei caldera: an integrated multidisciplinary image of its buried structure. *Int. J. Earth Sci.* 103, 401–421. doi: 10.1007/s00531-013-0972-7

Piochi, M., Mormone, A., Balassone, G., Strauss, H., Troise, C., and De Natale, G. (2015). Native sulfur, sulfates and sulfides from the active Campi Flegrei volcano (southern Italy): genetic environments and degassing dynamics revealed by mineralogy and isotope geochemistry. *J. Volcanol. Geotherm. Res.* 304, 180–193. doi: 10.1016/j.jvolgeores.2015.08.017

Podar, P. T., Yang, Z., Björnsdóttir, S. H., and Podar, M. (2020). Comparative analysis of microbial diversity across temperature gradients in hot springs from yellowstone and Iceland. *Front. Microbiol.* 11:1625. doi: 10.3389/fmicb.2020.01625

Poichi, M. M. A., Strauss, H., and Balassone, G. (2019). The acid sulfate zone and the mineral alteration styles of the Roman Puteoli (Neapolitan area, Italy): clues on fluid fracturing progression at the Campi Flegrei volcano. *Solid Earth* 10, 1809–1831. doi: 10.5194/se-10-1809-2019

Power, J. F., Carere, C. R., Lee, C. K., Wakerley, G. L., Evans, D. W., Button, M., et al. (2018). Microbial biogeography of 925 geothermal springs in New Zealand. *Nat. Commun.* 9:2876. doi: 10.1038/s41467-018-05020-y

Puopolo, R., Gallo, G., Mormone, A., Limauro, D., Contursi, P., Piochi, M., et al. (2020). Identification of a new heavy-metal-resistant strain of Geobacillus stearothermophilus isolated from a hydrothermally active volcanic area in southern Italy. *Int. J. Environ. Res. Public Health* 17:2678. doi: 10.3390/ijerph17082678

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Available at: https://www.R-project.org/

Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. (2016). Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem. Biophys. Res. Commun.* 469, 967–977. doi: 10.1016/j.bbrc.2015.12.083

Rodriguez-R, L. M., Gunturu, S., Tiedje, J. M., Cole, J. R., and Konstantinidis, K. T. (2018). Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. *mSystems* 3:e00039-18. doi: 10.1128/mSystems.00039-18

Rolandi, G., De Natale, G., Kilburn, C. R., Troise, C., Somma, R., Di Lascio, M., et al. (2020a). *The 39 ka Campanian Ignimbrite Eruption: New Data on Source Area in the Campanian Plain Vesuvius, Campi Flegrei, and Campanian Volcanism*. Amsterdam, Netherlands: Elsevier, 175–205.

Rolandi, G., Di Lascio, M., and Rolandi, R. (2020b). *The Neapolitan Yellow Tuff Eruption as the Source of the Campi Flegrei Caldera Vesuvius, Campi Flegrei, and Campanian Volcanism*. Amsterdam, Netherlands: Elsevier, 273–296.

Scherf, U., and Buckel, W. (1993). Purification and properties of an iron-sulfur and FAD-containing 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA delta 3-delta 2-isomerase from clostridium aminobutyricum. *Eur. J. Biochem.* 215, 421–429. doi: 10.1111/j.1432-1033.1993.tb18049.x

Scherf, U., Söhling, B., Gottschalk, G., Linder, D., and Buckel, W. (1994). Succinate-ethanol fermentation in clostridium kluyveri: purification and characterisation of 4-hydroxybutyryl-CoA dehydratase/vinylacetyl-CoA Δ3-Δ2-isomerase. *Arch. Microbiol.* 161, 239–245. doi: 10.1007/BF00248699

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Seo, D. C., and DeLaune, R. D. (2010). Effect of redox conditions on bacterial and fungal biomass and carbon dioxide production in Louisiana coastal swamp forest sediment. *Sci. Total Environ.* 408, 3623–3631. doi: 10.1016/j.scitotenv.2010.04.043

Sharp, Z. (2007). *Principles of Stable Isotope Geochemistry*. Pearson Education Inc., Upper Saddle River, NJ

Sharp, C. E., Brady, A. L., Sharp, G. H., Grasby, S. E., Stott, M. B., and Dunfield, P. F. (2014). Humboldt's spa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8, 1166–1174. doi: 10.1038/ismej.2013.237

Sheik, C. S., Reese, B. K., Twing, K. I., Sylvan, J. B., Grim, S. L., Schrenk, M. O., et al. (2018). Identification and removal of contaminant sequences from ribosomal gene databases: lessons from the census of deep life. *Front. Microbiol.* 9:840. doi: 10.3389/fmicb.2018.00840

Shock, E. L., Holland, M., Amend, J. P., Osburn, G., and Fischer, T. P. (2010). Quantifying inorganic sources of geochemical energy in hydrothermal ecosystems, Yellowstone National Park, USA. *Geochim. Cosmochim. Acta* 74, 4005–4043. doi: 10.1016/j.gca.2009.08.036

Solorzano, L. (1969). Determination of ammonia in natural waters by the phenol hypochlorite method. *Limnol. Oceanogr.* 14, 799–801.

Strazzulli, A., Cobucci-Ponzano, B., Iacono, R., Giglio, R., Maurelli, L., Curci, N., et al. (2020). Discovery of hyperstable carbohydrate-active enzymes through metagenomics of extreme environments. *FEBS J.* 287, 1116–1137. doi: 10.1111/febs.15080

Tozer, W. C., Hackell, D., Miers, D., and Silvester, W. (2005). Extreme isotopic depletion of nitrogen in New Zealand lithophytes and epiphytes; the result of diffusive uptake of atmospheric ammonia? *Oecologia* 144, 628–635. doi: 10.1007/s00442-005-0098-0

Troiano, A., Di Giuseppe, M. G., Patella, D., Troise, C., and De Natale, G. (2014). Electromagnetic outline of the Solfatara–Pisciarelli hydrothermal system, Campi Flegrei (southern Italy). *J. Volcanol. Geotherm. Res.* 277, 9–21. doi: 10.1016/j.jvolgeores.2014.03.005

Troise, C., De Natale, G., Schiavone, R., Somma, R., and Moretti, R. (2019). The Campi Flegrei caldera unrest: discriminating magma intrusions from hydrothermal effects and implications for possible evolution. *Earth-Sci. Rev.* 188, 108–122. doi: 10.1016/j.earscirev.2018.11.007

Uritskiy, G. V., DiRuggiero, J., and Taylor, J. (2018). MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* 6:158. doi: 10.1186/s40168-018-0541-1

Valentino, G., and Stanzione, D. (2003). Source processes of the thermal waters from the Phlegraean fields (Naples, Italy) by means of the study of selected minor and trace elements distribution. *Chem. Geol.* 194, 245–274. doi: 10.1016/S0009-2541(02)00196-1

Wall, K., Cornell, J., Bizzoco, R. W., and Kelley, S. T. (2015). Biodiversity hot spot on a hot spot: novel extremophile diversity in Hawaiian fumaroles. *MicrobiologyOpen* 4, 267–281. doi: 10.1002/mbo3.236

Wang, S., Dong, H., Hou, W., Jiang, H., Huang, Q., Briggs, B. R., et al. (2014). Greater temporal changes of sediment microbial community than its waterborne counterpart in Tengchong hot springs, Yunnan Province, China. *Sci. Rep.* 4:7479. doi: 10.1038/srep07479

Wang, S., Hou, W., Dong, H., Jiang, H., Huang, L., Wu, G., et al. (2013). Control of temperature on microbial community structure in hot springs of the Tibetan plateau. *PLoS One* 8:e62901. doi: 10.1371/journal.pone.0062901

Wang, R., Lin, J.-Q., Liu, X.-M., Pang, X., Zhang, C.-J., Yang, C.-L., et al. (2019). Sulfur oxidation in the acidophilic autotrophic Acidithiobacillus spp. *Front. Microbiol.* 9:3290. doi: 10.3389/fmicb.2018.03290

Weyrich, L. S., Farrer, A. G., Eisenhofer, R., Arriola, L. A., Young, J., Selway, C. A., et al. (2019). Laboratory contamination over time during low-biomass sample analysis. *Mol. Ecol. Resour.* 19, 982–996. doi: 10.1111/1755-0998.13011

Wu, Y.-W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638

Xie, W., Zhang, C. L., Wang, J., Chen, Y., Zhu, Y., de la Torre, J. R., et al. (2015). Distribution of ether lipids and composition of the archaeal community in terrestrial geothermal springs: impact of environmental variables. *Environ. Microbiol.* 17, 1600–1614. doi: 10.1111/1462-2920.12595

Zhang, Z., Wang, H., Zhou, J., Li, H., He, Z., Van Nostrand, J. D., et al. (2015). Redox potential and microbial functional gene diversity in wetland sediments under simulated warming conditions: implications for phosphorus mobilization. *Hydrobiologia* 743, 221–235. doi: 10.1007/s10750-014-2039-6

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 27 April 2023
DOI 10.3389/fmicb.2023.1111809

OPEN ACCESS

EDITED BY
Brian P. Hedlund,
University of Nevada,
Las Vegas, United States

REVIEWED BY
Paulina Corral,
Sevilla University, Spain
Nicole Pietrasiak,
New Mexico State University, United States
Peerada Prommeenate,
National Science and Technology Development Agency (NSTDA), Thailand

*CORRESPONDENCE
Maurycy Daroch
✉ m.daroch@pkusz.edu.cn

RECEIVED 30 November 2022
ACCEPTED 06 April 2023
PUBLISHED 27 April 2023

CITATION
Tang J, Zhou H, Jiang Y, Yao D, Waleron KF, Du L-M and Daroch M (2023) Characterization of a novel thermophilic cyanobacterium within *Trichocoleusaceae*, *Trichothermofontia sichuanensis* gen. et sp. nov., and its CO_2-concentrating mechanism.
Front. Microbiol. 14:1111809.
doi: 10.3389/fmicb.2023.1111809

Characterization of a novel thermophilic cyanobacterium within *Trichocoleusaceae*, *Trichothermofontia sichuanensis* gen. et sp. nov., and its CO_2-concentrating mechanism

Jie Tang[1], Huizhen Zhou[1], Ying Jiang[2], Dan Yao[1], Krzysztof F. Waleron[3], Lian-Ming Du[1] and Maurycy Daroch[2]*

[1]School of Food and Bioengineering, Chengdu University, Chengdu, Sichuan, 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

Thermophiles from extreme thermal environments have shown tremendous potential regarding ecological and biotechnological applications. Nevertheless, thermophilic cyanobacteria remain largely untapped and are rarely characterized. Herein, a polyphasic approach was used to characterize a thermophilic strain, PKUAC-SCTB231 (hereafter B231), isolated from a hot spring (pH 6.62, 55.5°C) in Zhonggu village, China. The analyses of 16S rRNA phylogeny, secondary structures of 16S-23S ITS and morphology strongly supported strain B231 as a novel genus within *Trichocoleusaceae*. Phylogenomic inference and three genome-based indices further verified the genus delineation. Based on the botanical code, the isolate is herein delineated as *Trichothermofontia sichuanensis* gen. et sp. nov., a genus closely related to a validly described genus *Trichocoleus*. In addition, our results suggest that *Pinocchia* currently classified to belong to the family *Leptolyngbyaceae* may require revision and assignment to the family *Trichocoleusaceae*. Furthermore, the complete genome of *Trichothermofontia* B231 facilitated the elucidation of the genetic basis regarding genes related to its carbon-concentrating mechanism (CCM). The strain belongs to β-cyanobacteria according to its β-carboxysome shell protein and 1B form of Ribulose bisphosphate Carboxylase-Oxygenase (RubisCO). Compared to other thermophilic strains, strain B231contains a relatively low diversity of bicarbonate transporters (only BicA for HCO_3^- transport) but a higher abundance of different types of carbonic anhydrase (CA), β-CA (*ccaA*) and γ-CA (*ccmM*). The BCT1 transporter consistently possessed by freshwater cyanobacteria was absent in strain B231. Similar situation was occasionally observed in freshwater thermal *Thermoleptolyngbya* and *Thermosynechococcus* strains. Moreover, strain B231 shows a similar composition of carboxysome shell proteins (*ccmK1-4*, *ccmL*, *-M*, *-N*, *-O*, and *-P*) to mesophilic cyanobacteria, the diversity of which was higher than many thermophilic strains lacking at least one of the four *ccmK* genes. The genomic distribution of CCM-related genes suggests that the expression of some components is regulated as an operon and others in an independently controlled satellite locus. The current study also offers fundamental information for future taxogenomics, ecogenomics and geogenomic studies on distribution and significance of thermophilic cyanobacteria in the global ecosystem.

KEYWORDS

16S rRNA, 16S-23S ITS, CO2-concentrating mechanism, thermophilic cyanobacterium, *Trichocoleusaceae*, *Leptolyngbyaceae*, *Pinocchia*, *Trichothermofontia*

Introduction

Thermophilic cyanobacteria are widely distributed in hot spring ecological niches (Miller et al., 2007; Alcorta et al., 2020). Moreover, many studies have demonstrated the importance of thermophilic cyanobacteria as primary photosynthetic producers of geothermal ecosystems, accounting for a large part of those ecosystems' biomass and productivity (Esteves-Ferreira et al., 2018; Chen et al., 2021). Besides, high-value-added products can be harvested from thermophilic cyanobacteria and have been applied to numerous industries, e.g., agricultural, pharmaceutical and nutraceutical (Patel et al., 2019).

Isolation of thermophilic cyanobacteria from diverse ecosystems is critical for multidisciplinary studies regarding their morphology, genetics, physiology, biochemistry, and for providing potential strains for biotechnology and industrial applications (Cordeiro et al., 2020). It is, however, challenging to assign the taxonomy of thermophilic cyanobacteria due to their simple morphology. As a result, the diversity of thermophilic cyanobacterial genera might be severely underestimated, and community efforts are being made to rectify it. *Synechococcus*-like strains (Tang et al., 2022c) and *Leptolyngbya*-like strains (Mai et al., 2018; Yao et al., 2021) have undergone an extensive reevaluation based on multi-locus sequence analysis and genomic data, providing new insights into genetic diversity of the two genera. Furthermore, polyphasic taxonomic classification approaches have been widely utilized for cyanobacterial identification, particularly for understudied or unresolved polyphyletic families/genera/species and identification of novel families and genera (Raabova et al., 2019; Shalygin et al., 2020). Consequently, these reassignments will better taxonomically resolve cyanobacterial genera and families. Especially, the increased number of genomic sequences will facilitate the development of taxogenomics and complement the traditional 16S rRNA-based taxa identification.

The aquatic environments where thermophilic cyanobacteria live are characterized by low availability of CO_2, primarily due to external factors, e.g., temperature, pH and gas exchange (Durall and Lindblad, 2015). Meanwhile, the availability of dissolved inorganic carbon in form of carbonates is highly variable. Therefore, to survive in the hostile aquatic habitat, the thermophilic cyanobacteria utilize CO_2-concentrating mechanisms (CCM) to ensure that the Ribulose bisphosphate Carboxylase-Oxygenase (RubisCO) with low affinity for CO_2 is surrounded by high CO_2 levels and functions regardless of thermal stress (Galmés et al., 2015, 2016). Therefore, investigating the molecular component at the genomic level is a prerequisite for understanding the thermophilic cyanobacterial CCM and its relationship with their niches. In addition, under the current scenario of global warming, the studies on the molecular components of cyanobacterial CCM in relation to their specific habitats may shed light on the evolution of hot spring genomes as an example of selective pressure in warmer environments.

In our previous study, we isolated a *Leptolyngbya*-like strain, B231, from a green microbial mat of a hot spring in Zhonggu village, Sichuan, China, which can grow at 47°C and/or at the concentration of 0.1 M $NaHCO_3$ (Tang et al., 2018a). Herein, thorough polyphasic characterization for strain B231, including 16S rRNA phylogeny, the secondary structure of 16S-23S ITS, and morphology description has been achieved. According to the botanical code, a new genus name *Trichothermofontia sichuanensis* gen. et sp. nov. has been proposed for strain B231 as the first representative of the genus *Trichothermofontia*. Moreover, based on our research interests, the molecular basis of CCM has been investigated for strain B231 by computational identification, and the CCM component has been further related to its adaptation. The current study lays a foundation for future taxogenomic, ecogenomic and geogenomic studies on the distribution and significance of thermophilic cyanobacteria.

Materials and methods

Origins, cultivation, and basic physiological assessment of *Trichothermofontia sichuanensis* B231

The strain B231, capable of forming mats, used in the present study was initially isolated from a hot spring around Zhonggu village in Ganzi Prefecture of Sichuan Province, China. Sample collection was done on 12 May 2016, with the humidity close to 71%. The ambient temperature at the time of collection was 15°C, and the light intensity was around 1,000 lux. The temperature of the hot spring, its pH, and the concentration of total dissolved solids were 55.5°C, 6.62, and 492 mmol L^{-1}, respectively. Information about the sampling site and preliminary taxonomic assignment of the strain was detailed in our previous studies (Tang et al., 2018a,b). A unicyanobacterial culture of B231, recovered from 10% DMSO stocks maintained at −80°C for over 2 years, was used to establish experimental cultures as described previously (Tang et al., 2018a). Briefly, the recovered strain was cultivated at 45°C in 150 mL BG-11 medium in 500 mL Erlenmeyer flasks agitated at 100 rpm under 16l:8D photoperiod at 45 µmol m^{-2} s^{-1} provided by fluorescent tubes unless stated otherwise. The strain initially denoted and maintained in Peking University Algae Collection as PKUAC-SCTB231 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession number FACHB-3573. The strain was assessed for the capacity to utilize nitrite and nitrate using the modifications of BG-11, essentially as described earlier (Liang et al., 2019). Briefly, BG-11 medium without nitrogen was supplemented with 0, 0.075, 0.5, 1.5, 5, 5.7 g L^{-1} $NaNO_3$; 1.218 g L^{-1} $NaNO_2$. The effect of bicarbonate has been tested in regular BG-11 medium supplemented with 0, 0.1, 0.3, 0.5, 0.7, 1.0 M sodium bicarbonate. The growth of the strain has been assessed qualitatively due to the mat-forming character of the strain.

Genome sequencing, assembly, and annotation

Integrated sequencing strategies employing MGISEQ-2000 PE150 and Oxford Nanopore Technologies PromethION sequencing platforms were applied for whole-genome sequencing of the strain B231. The sequencing was performed by a commercial provider BGI-TECH (Wuhan, China). The genomic DNA was isolated using FastDNA™ SPIN Kit for Soil (MP Biomedicals, Irvine, CA, United States), and its integrity verified with agarose electrophoresis. After sequencing, a total of 806,280,450 bp of short read data and 1,717,621,070 bp of the long read data of an average length of 14,717 bp were used for the assembly of the genome yielding a single circular chromosome and no plasmids. The assembly has been performed from long-read data using Flye v2.7 (Kolmogorov et al., 2019) module integrated into the commercial Geneious Prime 2022.2 package (Kearse et al., 2012) and subsequently refined with short-read data using Geneious mapper on default settings. The genome of strain B231 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 (O'Leary et al., 2016), and further using the RAST annotation system (Brettin et al., 2015) to minimize poor calls. The genome annotation of strain B231 was summarized in Supplementary Table 1. The complete genome has been deposited in GenBank under accession number CP110848.

Phylogenetic inference of 16S rRNA

The full-length 16S rRNA gene sequence of strain B231 was extracted from its genome sequence. Additional 97 16S rRNA gene sequences of cyanobacterial references were retrieved from NCBI through BLASTN search. Muscle complemented in Mega7 (Kumar et al., 2016) was employed to generate multiple sequence alignments, and manual editing comprising trimming to the same length and adjusting poorly aligned regions were carried out where necessary. Sequences of the alignment were trimmed to the same length (1,013 bp). The alignment of 16S rRNA gene sequences was subjected to Bayesian Inference using MrBayes v3.2.7 (Ronquist et al., 2012). The following parameters were applied: NST = 6, Rates = equal, MCMC Ngen = 10,000,000. Default settings were used for all the other parameters. The Bayesian analysis had a mean estimated sample size (ESS) exceeding 4,700 for all parameters, far above the average of 200 typically accepted as sufficient by phylogeneticists (Drummond et al., 2006). The final average standard deviation of split frequencies was 0.002624. The potential scale reduction factor (PSRF) value for all the estimated parameters in the Bayesian analysis was 1.00, indicating that convergence of the MCMC chains was statistically achieved (Gelman and Rubin, 1992).

Analysis of 16S-23S ITS

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 using the structure, draw mode untangle with loop fix, default conditions in Mfold were used in all cases.

Whole-genome comparisons

A genome dataset was compiled for whole-genome comparisons, including genomes of strain B231 and 17 representative focus taxa (one representative from the family *Trichocoleusaceae*, eight from the family *Oculatellaceae* and nine from the family *Leptolyngbyaceae* as references). The quality of each genome was assessed using CheckM (Parks et al., 2015) to ensure a high-quality dataset with near completeness (≥95%) and low contamination (<5%). Moreover, the corresponding protein sequences of each genome were downloaded from the NCBI database.

Three indices useful for genus delineation were calculated to summarize the similarity or distance between genomes. 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] The percentages of conserved proteins (POCP) between genomes were determined according to the method described previously (Qin et al., 2014).

Phylogenomic reconstruction

The phylogenomic relationship between strain B231 and focal taxa was elucidated by the concatenated protein sequences from 647 single-copy genes shared by all the genomes. The homologous gene clusters identified by OrthoMCL (Li et al., 2003) were used to refine single-copy genes shared by all the genomes, which were concatenated employing a 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). A total of 546 protein models were used to select the optimal substitution model for phylogenomic analysis using ModelFinder implemented in IQ-TREE. Bootstrap tests (1,000 replicates) were carried out for the assessment of tree topologies using UltraFast Bootstrap (Hoang et al., 2018).

Investigation on CCM in strain B231

Orthologous proteins involved in CCM of strain B231 were identified as previously described (Tang et al., 2022d). Briefly, amino acid sequences of 28 proteins involved in CCM of *Synechocystis* sp. PCC 6803 were retrieved as a reference protein set. Orthologous genes in strain B231 were identified with the bidirectional best hit (BBH) criterion using BLASTP with the following thresholds: *E*-value cut-off of 1E-6, ≥30% identity and 70% coverage, and manually curated. Amino acid sequences of RubisCO large subunit (*rbcL*) were collected for strain B231 and reference cyanobacteria to infer the protein

1 http://enve-omics.ce.gatech.edu/g-matrix/

function and classification. Protein sequences of genes *ccmK*, *-L*, *-M*, *-N*, *-O*, and *-P* encoding carboxysome shell proteins were also collected for phylogenetic reconstruction. All the phylogenetic analyses were performed using Maximum-Likelihood (ML) algorithm as previously described (Tang et al., 2022c).

Morphology investigation

All microscopic operations were performed essentially as described by Tang et al. (2021). In short, strain B231 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 using transmission electron microscopy (TEM; HT7800, HITACHI, Japan). Cell measurements have been performed on 100 representative cells selected from six independent micrographs and presented as a range.

Taxonomic evaluation

The classification system in this study was applied according to Komárek et al. (2014). Briefly, taxonomy assignment was determined based on phylogenetic position of the corresponding entity, as well as its morphological and ecological characters. 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).

Results and discussion

General genomic characteristics of strain B231

The complete genome of strain B231 has been achieved by integrating Oxford Nanopore and DNBSEQ sequencing systems yielding genome coverage of 387× and 181×, respectively. The B231 genome comprises a single circular chromosome with a size of 4,436,989 bp and a GC content of 53.9%. Bioinformatic annotation indicates that two ribosomal RNA (rrn) operons, 45 tRNA genes and 4,352 protein-coding sequences (CDS), were present in the B231 genome (Supplementary Table 1). The two ribosomal RNA operons differed in length by an 11 bp insertion in the ITS region, two single bp insertions in the 23S rRNA gene region and 2 bp difference between the two variants of 16S rRNA gene, indicating their 99.98% identity. None of the above-mentioned differences fundamentally impacted the phylograms or predicted secondary structures of the ITS region, as the 11 bp insertion was outside the regions of interest. Approximately 48.2% of protein-coding genes (2,098 out of 4,352) protein-coding genes were predicted to be hypothetical proteins. It was not surprising to identify such a high percentage of hypothetical protein in the B231 genome since similar observations are typical in the genomes of other thermophilic cyanobacteria (Cheng et al., 2020; Kono et al., 2022; Tang et al., 2022a,b).

Phylogeny reconstruction of 16S rRNA gene

The Bayesian phylograms (Figure 1; Supplementary Figure 1) inferred by 16S rRNA gene sequences categorize the cyanobacterial strains into three well-defined families. Genera were also taxonomically resolved within families with the support of high posterior probabilities. The focal strain B231 is closely grouped with five thermophilic strains isolated from various hot springs in Zhonggu village (Table 1). Moreover, five other uncultured cyanobacteria were also closely related to strain B231. These 11 strains formed a well-supported clade distinct from the two described genera within the family *Trichocoleusaceae*. This result suggests the phylogenetically novel clade and a new genus within the family *Trichocoleusaceae*. In fact, the family *Trichocoleusaceae* is a monophyletic family that has been recently proposed by dividing the *Leptolyngbyaceae* into family-level clades based on molecularly-supported data (Mai et al., 2018). To date, the genus *Trichocoleus* (Mühlsteinova et al., 2014) was the only described member of the family *Trichocoleusaceae*. The phylogram generated in this study suggests the existence of two more genera in this family. The genus *Pinocchia* (Dvorak et al., 2015) is still classified within the family *Leptolyngbyaceae*, which contradicts the 16S-based phylogeny obtained in this study. Therefore, we propose that the genus *Pinocchia* should be reclassified into the family *Trichocoleusaceae*.

Furthermore, the helices within 16S rRNA have been studied. Only helices 18, 23 and 27 were investigated as they were considered the most informative for family distinction (Mai et al., 2018). As shown in Table 2, one distinctive nucleotide was found in helix 18 between *Trichocoleusaceae* and *Oculatellaceae*/*Leptolyngbyaceae*. Within helix 23, two types of distinctive nucleotides were present in *Trichocoleusaceae*, both capable of differentiating them from *Oculatellaceae* and *Leptolyngbyaceae*. However, the distinctive nucleotides in helix 27 were shared among families. Taken together, several nucleotide positions in different helices were considered informative indicators of family-level classification. Again, these results confirm the recognition of the novel clade and *Pinocchia* to join *Trichocoleus* as members of the family *Trichocoleusaceae*.

Within the novel clade, all the strains show high 16S rRNA sequence identities (97.6%–99.2%) to the strain B231 except for clone Mat-CYANO-S19 (Table 1). Therefore, according to the recommended threshold of 16S rRNA gene identity for bacterial species (98%–99%) or genera (94.5%–95%) demarcation (Yarza et al., 2014), clone Mat-CYANO-S19 can be proposed to be new species and the remaining 10 strains to be another new species. Such delineation was also supported by the 16S rRNA phylogeny (Figure 1). Besides, the novel clade shows 92.8%–93.5% of 16S rRNA sequence identities to *Trichocoleus* strains and 91.3%–92.9% to *Pinocchia* strains. This result was consistent with the taxonomic delineation of this novel clade as a new genus member within the family *Trichocoleusaceae*.

Intriguingly, the habitat niches were quite distinct among strains from the novel clade (Table 1). The six strains isolated from China were originally recovered from three separate freshwater hot springs located at a high altitude (3,200 m; Tang et al., 2018a,b). The other five strains, for which only their molecular signatures were available, also mostly show thermal origin. Three were from marine hot springs in Indonesia, one from a mesophilic microbial mat in the United States and one from a hot spring sediment in Thailand. The distinct habitat niches suggested that these strains might be classified into at least

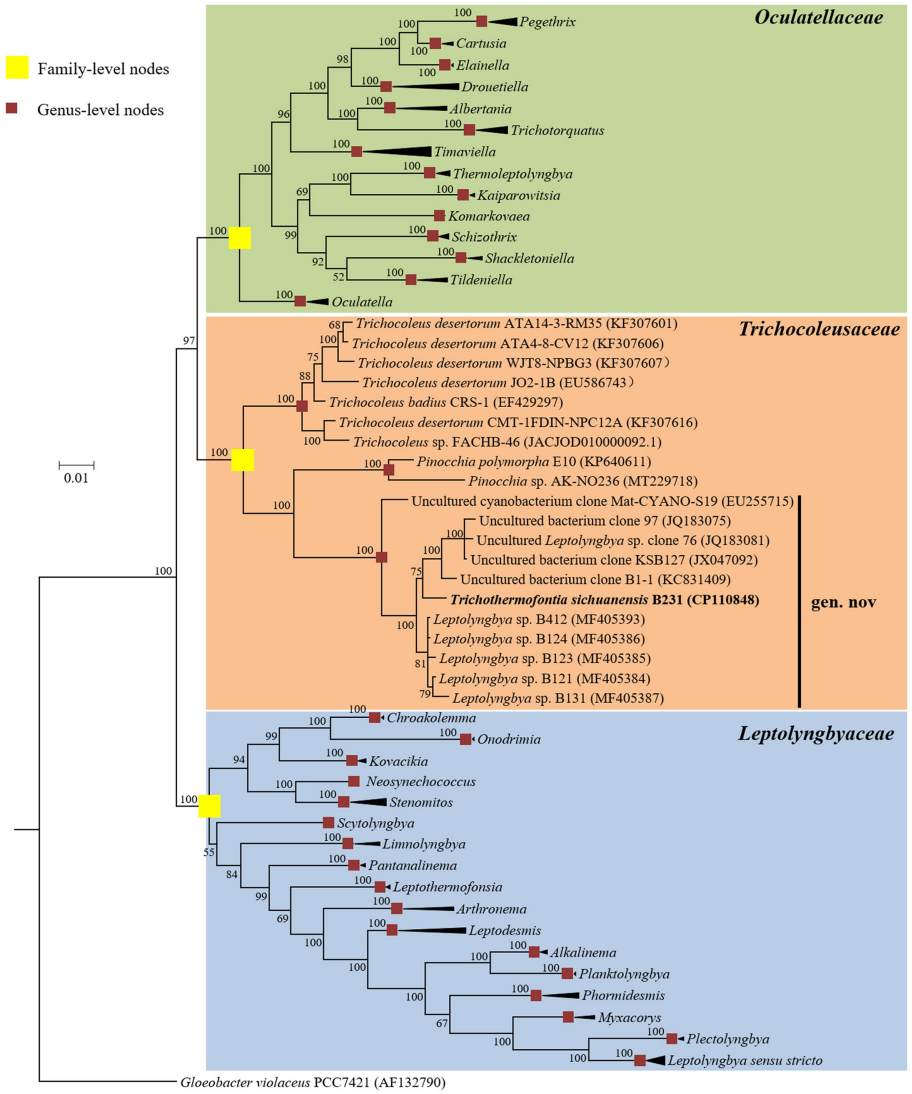


FIGURE 1
Bayesian inference of 16S rRNA gene sequences representing 98 cyanobacterial strains. Collapsed genera are indicated by black polygons, with a length corresponding to the distance from the most basal sequence to the most diverged sequence of the genus. The complete phylogram refers to Supplementary Figure 1. Posterior probabilities (%) are given above the nodes.

three ecotypes. The ecotype determination of uncultured bacterium clone B1-1 appears impractical in light of the minimal ecological information. In addition, the acclimation of these strains to diverse habitats strongly suggests the underlying genetic diversity and a wide distribution among representatives of this genus. Verifying this speculation will be an interesting topic that could be explored using phylogenomic and ecogenomic approaches providing that more isolates and genome sequences are available. Unfortunately, the data on uncultured strains were restricted to molecular signatures isolated from environments, hindering further detailed comparisons (e.g., physiological, ecological and genomic studies) that would require isolates.

Secondary structures of 16S-23S ITS

In the present study, the phylogeny of 16S-23S ITS has not been reconstructed mainly for two reasons. First, the 16S-23S ITS of important reference cyanobacteria, *Pinocchia* strains, only contained conserved $tRNA^{Ile}$, while another conserved $tRNA^{Ala}$ was missing (Dvorak et al., 2015). Second, the 16S-23S ITS of strain B231 lacks the V2 region (Table 3); indicating significant operon heterogeneity in these strains. Taken together, the current dataset of 16S-23S ITS sequences was extraordinarily divergent and would most likely result in a misleading taxonomic assignment (Johansen et al., 2011). However, sequence comparisons of 16S-23S ITS were still performed among representatives from the three families. Excluding two highly conserved tRNAs from ITS sequences, the length of the remaining ITS sequences tremendously varied from 230 to 535 bp (Table 3). Such an immense discrepancy was primarily attributed to the length differences in D1-D1' (51–141 bp), V2 (0–218 bp), boxB (33–70 bp), and V3 (21–161 bp). Within family *Trichocoleusaceae*, enormous variations were also observed among genera (Table 3). As a result, the nucleotide differences of these domains result in divergences of the secondary structures among representative strains from family *Trichocoleusaceae* (Supplementary Figure 2).

TABLE 1 The sequence identities of 16S rRNA gene between strain B231 and other closely related strains.

Strain	Sequence identity with B231 (%)	Isolation source	Temperature and pH	References
Trichothermofontia sichuanensis B231	100 (1,501)	Hot spring in Zhonggu village, Ganzi Prefecture, China, green microbial mat	55.5°C, 6.62	Tang et al. (2018b)
Leptolyngbya sp. B131	99.2 (1,355)	Hot spring in Zhonggu village, Ganzi Prefecture, China, green microbial mat, 3 m from B231 spring	53.1°C, 6.35	Tang et al. (2018b)
Leptolyngbya sp. B121	98.8 (1,355)	Hot spring in Zhonggu village, Ganzi Prefecture, China, green microbial mat, 3 m from B231 spring	53.1°C, 6.35	Tang et al. (2018b)
Leptolyngbya sp. B124	98.8 (1,355)	Hot spring in Zhonggu village, Ganzi Prefecture, China, green microbial mat, 3 m from B231 spring	53.1°C, 6.35	Tang et al. (2018b)
Leptolyngbya sp. B412	98.8 (1,355)	Hot spring in Zhonggu village, Ganzi Prefecture, China, sediment from cooler external part of the spring, 500 m from B231 spring	85.0°C, 8.50	Tang et al. (2018b)
Leptolyngbya sp. B123	98.7 (1,355)	Hot spring in Zhonggu village, Ganzi Prefecture, China, green microbial mat, 3 m from B231 spring	53.1°C, 6.35	Tang et al. (2018b)
Uncultured bacterium clone KSB127	98.0 (1,457)	Marine hot spring, Kalianda Island, Indonesia	NA	NA
Uncultured *Leptolyngbya* sp. clone 76	97.9 (1,453)	Marine hot spring, Kalianda Island, Indonesia	NA	NA
Uncultured bacterium clone 97	97.8 (1,457)	Marine hot spring, Kalianda Island, Indonesia	NA	NA
Uncultured bacterium clone B1-1	97.6 (1,455)	Sediment in Betong hot spring, Yala province, Thailand	NA	NA
Uncultured cyanobacterium clone Mat-CYANO-S19	96.7 (1,416)	Freshwater mesophilic microbial mat, United States	NA	NA

The number in brackets indicates the pairwise alignment length. Strains are sorted by order of identity from high to low. NA, not available.

TABLE 2 Nucleotide comparisons of focal helices within 16S rRNA among families.

Family	Helix 18	Helix 23	Helix 27
Trichocoleusaceae	TGCCAGCAGCCGCGGTAA**G**A	ATCGGGAAGAACACC**A**G**T**G ATCGGGAAGAACACC**A**G**A**G (*Pinocchia*)	GGGAGTACGC**T**CGCAAG**A**GTGAAACTC GGGAGTACGC**A**CGCAAG**T**GTGAAACTC (*Pinocchia*)
Oculatellaceae	TGCCAGCAGCCGCGGTAA**T**A	ATTRGRAAGAACAYC**G**G**T**G	GGGAGTACGC**T**CGCAAG**A**GTGAAACTC
Leptolyngbyaceae	TGCCAGCAGCCGCGGTAA**T**A	ATTGGGAAGAACACC**A**G**C**G	GGGAGTAYGC**A**CGCAAG**T**GTGAAACTC

Nucleotides variable between families but consistent within families are in bold. IUPAC code letters are assigned for nucleotides that vary within the consensus sequences.

The inferred D1-D1' helix of strain B231 differs from the other two inferred structures (Supplementary Figure 2A). The most similar structure to D1-D1' helix of strain B231 was that of *T. desertorum*. Both D1-D1' helices vary in the overall length and topology while retaining the basal stem structures of five base pairs followed by a right asymmetrical loop.

TABLE 3 The Length summary (bp) of corresponding regions within 16S-23S ITS of strain B231 and focal taxa.

Family	Species	ITS length (tRNA removed)	D1-D1′ helix	D2	D3	boxA	D4	V2 helix	boxB helix	V3 helix
Trichocoleusaceae	*Trichothermofontia sichuanensis* B231	269	63	12	5	12	7	0	70	75
	Trichocoleus desertorum ATA4-8-CV12	244	62	12	5	12	7	19	37	63
	Pinocchia polymorpha E10	372	119	12	5	12	7	10	46	162
Oculatellaceae	*Albertania skiophila* SA373	320	64	12	5	12	7	46	47	22
	Drouetiella hepatica Uher 2000/2452	281	64	12	5	12	7	21	34	50
	Elainella saxicola E1	296	62	12	5	12	7	58	33	19
	Kaiparowitsia implicata GSE-TBC-09CA2	394	142	12	5	12	7	9	36	121
	Komarkovaea angustata EY01-AM2	325	64	12	5	12	7	20	41	94
	Oculatella sp. FACHB-28	265	64	12	5	12	7	17	34	52
	Pegethrix bostrychoides GSE-TBD-MK4-15B	308	87	12	5	12	7	24	36	94
	Thermoleptolyngbya sichuanensis A183	535	64	12	5	12	7	218	48	74
	Tildeniella torsiva UHER1998/13D	262	66	12	5	12	7	7	49	92
	Timaviella obliquedivisa GSE-PSE-MK23-08B	319	63	12	5	12	7	29	49	59
	Trichotorquatus coquimbo ATA2-1-KO25A	314	100	12	5	12	7	23	36	118
Leptolyngbyaceae	*Alkalinema pantanalense* CENA528	296	64	12	5	12	7	24	48	54
	Chroakolemma pellucida 719	268	61	12	5	12	7	16	41	53
	Kovacikia muscicola HA7619-LM3 clone 41A	345	63	12	5	12	7	90	41	95
	Leptodesmis sichuanensis A121	325	63	12	5	12	7	81	33	98
	Leptolyngbya boryanum PCC 73110	275	51	12	5	12	7	10	33	21
	Leptothermofonsia sichuanensis E412	380	121	12	5	12	7	76	45	98
	Limnolyngbya circumcreta CHAB5667	388	98	12	5	12	7	83	63	76
	Myxacorys californica WJT24-NPBG12B	258	86	12	5	12	7	9	33	71
	Neosynechococcus sphagnicola sy1	230	63	12	5	12	7	11	39	95
	Onodrimia javanensis 28	280	105	12	5	12	7	7	44	47
	Phormidesmis priestleyi ANT.L52.4	329	113	12	5	12	7	12	56	77
	Plectolyngbya hodgsonii ANT.LPR2.2	306	55	12	5	12	7	29	44	19
	Scytolyngbya timoleontis XSP2	276	64	12	5	12	7	14	40	94
	Stenomitos rutilans HA7619-LM2	258	65	12	5	12	7	7	34	92

V2 helix was absent in the 16S-23S ITS of strain B231. A similar structure of the V2 helix was shared by the other two representative strains from the family *Trichocoleusaceae*, comprising one stem and a hairpin loop (Supplementary Figure 2B). However, the V2 helices of the two strains differ in the stem length and residues of hairpins.

All three strains share a basal stem structure (AGCA-TGCT) in boxB helices (Supplementary Figure 2C). Strain B231 shows a much

longer residue length than the other two strains (Table 3), resulting in a divergent boxB helix structure.

The V3 helix of strain B231 consists of an asymmetrical loop and a 4-residue hairpin loop, and two stems (Supplementary Figure 2D). The V3 helix of strain B231 was different from those of the other two strains, while a basal stem structure (TGTC-GACA) was shared by all the strains.

Conclusively, the phylogeny of 16S rRNA and the result of 16S-23S ITS secondary structure analysis supports the delineation of strain B231 as a novel genus within this family.

Genome comparisons

To our best knowledge, there are no genomic-level comparisons within the family *Trichocoleusaceae* or between members of the family *Trichocoleusaceae* and genera from other different families. Therefore, it was crucial to elucidate a snapshot of genomic divergences between Strain B231 and focal taxa. Herein, three indices of whole genome comparisons between strain B231 and representative species from *Trichocoleusaceae*, *Oculatellaceae* and *Leptolyngbyaceae* were presented (Table 4).

Considerable divergences in genomes were observed between strain B231 and the other 18 strains, as revealed by the ANI and AAI values (Table 4). The ANI and AAI values were less than 82 and 63%, respectively. This result conforms to the suggested values for genus (ANI < 83%, AAI ≤ 70%) delimitation (Walter et al., 2017; Jain and Rodriguez, 2018). However, in some cases, misleading results might be achieved for the classification of the prokaryotic genus using ANI or AAI (Pannekoek et al., 2016). Therefore, the POCP specific for genus delineation was also calculated between strain B231 and the other focal taxa. The POCP values range from 38.78% to 49.23% (Table 4), all within the threshold (<50%) for the definition of a prokaryotic genus (Qin et al., 2014). Taken together, all the results verify the genus demarcation of strain B231 as a novel genus within the family *Trichocoleusaceae*. This conclusion was in accordance with the results of molecular phylogeny (Figure 1).

TABLE 4 Indices values (%) of whole genome comparisons between strain B231 and representative species from *Trichocoleusaceae*, *Oculatellaceae* and *Leptolyngbyaceae*.

Family	Species	ANI	AAI	POCP
Trichocoleusaceae	*Trichocoleus desertorum* ATA4-8-CV12	71.97	61.53	47.60
Oculatellaceae	*Drouetiella hepatica* Uher 20,002,452	72.67	59.85	45.22
	Elainella saxicola E1	75.41	60.36	38.78
	Kaiparowitsia implicata GSE-PSE-MK54-09C	72.34	58.94	43.57
	Oculatella sp. FACHB-28	74.05	60.68	43.94
	Pegethrix bostrychoides GSE-TBD4-15B	72.83	59.77	49.01
	Thermoleptolyngbya sichuanensis A183	76.44	61.43	49.19
	Tildeniella torsiva UHER_199813D	73.45	57.84	45.65
	Timaviella obliquedivisa GSE-PSE-MK23-08B	71.25	59.96	49.57
Leptolyngbyaceae	*Alkalinema* sp. FACHB-956	74.39	58.51	49.21
	Leptodesmis sichuanensis A121	77.66	62.17	49.23
	Leptolyngbya boryana dg5	78.97	58.92	44.76
	Leptothermofonsia sichuanensis E412	81.51	62.07	46.79
	Myxacorys almedinensis A	74.17	59.78	49.23
	Neosynechococcus sphagnicola sy1	75.85	59.72	44.28
	Pantanalinema sp. GBBB05	74.03	61.23	48.76
	Phormidesmis priestleyi BC1401	75.61	59.87	46.91
	Stenomitos frigidus ULC18	72.21	59.95	45.06

Phylogenomic analysis

Analysis of homologous gene clusters generated 647 single-copy genes shared by the genomes studied. The concatenated alignments of these genes possess 216,534 aligned amino acid sites. Using the optimal substitution model (LG + F + R5), the ML inference of the supergene alignment provides a phylogeny with strong bootstrap support for all branches (Figure 2), defining representative species from each described genus. Within the *Trichocoleusaceae* clade, strain B231 is quite divergent from the described genus *Trichocoleus* and *Neosynechococcus*. Strain B231 is well-separated by the long branches from other described genera in the family, suggesting the taxon as a novel genus.

The phylogenomic topology was almost identical to that of the 16S rRNA gene (Figure 1). However, one exception was present in the phylogram. The phylogenetic analysis of the 16S rRNA gene suggests the clear affiliation of *Neosynechococcus sphagnicola* sy1 to the family *Leptolyngbyaceae* (Figure 1; Dvorak et al., 2014), whereas sy1 was located in a position within the genome-scale phylogram, forming a clade with two genera from family *Trichocoleusaceae* (Figure 2). This position of the strain was consistent with the phylogenomic results reported in previous studies (Tang et al., 2022a,b). Taken together, the discordant two trees suggest that currently used approaches cannot determine whether the taxonomic position of sy1 is accurate. The inconsistent phylogenies imply that an extensive study of this organism encompassing phenotypical, chemotaxonomical, physiological and genotypical studies should be carried out.

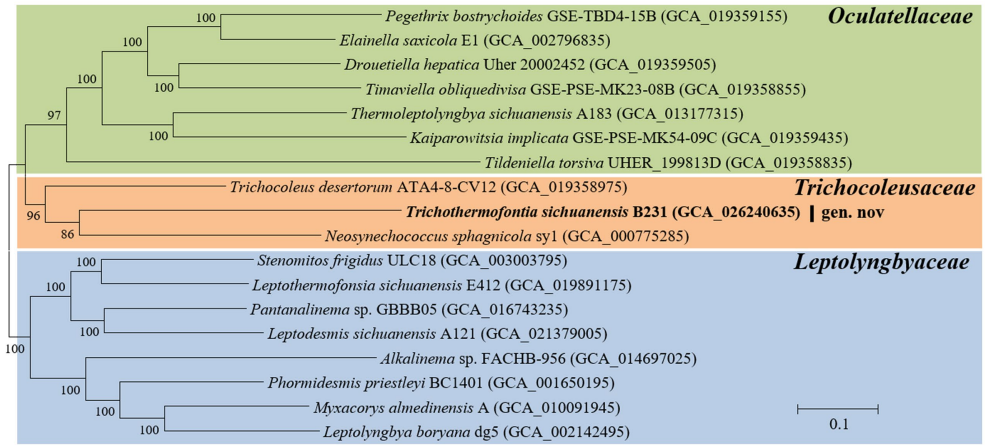


FIGURE 2
ML phylogenomic inference of concatenated protein alignment of 647 single-copy genes shared by all genomes. Strain no. in bold represent the strain identified in this study. Bootstrap values are indicated at nodes. Scale bar=10% substitutions per site.

Classification of CCM in strain B231

The aquatic environments where most thermophilic cyanobacteria live usually suffer from low availability of CO_2, primarily due to external factors, e.g., temperature, pH and gas exchange. On the other hand the availability of soluble carbonates is highly variable depending on the geochemistry of the habitat. Therefore, it was crucial to investigate the molecular component and organization of CCM of thermophilic strains in relation to their habitats.

Phylogenetic analysis of the molecular marker, *rbcL*, was frequently used to indicate the carboxysome type of cyanobacteria (Klanchui et al., 2017). Herein, the ML phylogram of *rbcL* positions strain B231 within the category of β-cyanobacteria (Figure 3), indicating the presence of RubisCO 1B form in this thermophile. This result was in accordance with a previous study that surveyed 17 thermophilic cyanobacteria and allocated them to the β-cyanobacteria (Tang et al., 2022d). Nevertheless, strain B231 clusters with none of these thermophilic cyanobacteria but is uniquely positioned among mesophilic cyanobacteria (Figure 3), suggesting considerable genetic diversity of *rbcL* amino acid sequences among these thermophiles. Furthermore, the evolutionary relationship based on habitats or morphology cannot be elucidated from the present rbcL phylogram, suggesting that the phylogenetic inference of *rbcL* may be unsuitable for elaborating the relationship between evolution and cyanobacterial habitat and environments (Komárek, 2016).

Genes encoding Ci uptake systems in strain B231

In general, the cyanobacterial CCM comprises two primary components: Ci uptake systems and carboxysomes. Cyanobacteria have been reported to have up to five different systems to actively acquire and transport Ci into the cells. Two for the uptake of CO_2 and three for the transport of HCO_3^- (Pronina et al., 2017). Herein, three different transport systems (Figure 4) have been identified in strain B231, including two CO_2 uptake systems, NDH-1_3 and NDH-1_4 complex, and one HCO_3^- transport system, BicA. Existence of both CO_2 uptake systems in strain B231 aligns with the previous studies indicating that these CO_2 transporters were present in β-cyanobacteria living in freshwater, brackish or eutrophic lakes (Durall and Lindblad, 2015). This is in sharp contrast to oceanic α-cyanobacteria (e.g., *Prochlorococcus* species) and marine β-cyanobacteria (e.g., *Trichodesmium* species) that contain only one or even lack them entirely (Pronina et al., 2017). Taken together, the presence of NDH-1_3 and NDH-1_4 might be relevant to the environments where these cyanobacteria inhabit. Indeed, a low-CO_2 inducible high-affinity CO_2 uptake system (NDH-1_3 complex) and a constitutive low-affinity CO_2 uptake system (NDH-1_4 complex), may allow the thermophilic cyanobacteria more alternative strategies to survive in environments with significant CO_2 fluctuation, particularly in hot springs. In addition, the protein sequences of five genes (*ndhD4*, *ndhF4*, *cupB*, *ndhD3*, *ndhF3*) encoding the two CO_2 uptake systems exhibit different identities with the sequences of non-thermophilic reference cyanobacteria (*Synechocystis* PCC 6803), ranging from 60.4 to 68.6%, but a high degree of homology (84.4% identity) to *cupA* gene of NDH-1_3 complex.

Two homologs of BicA were present in strain B231 (Figure 4). However, the protein sequences of the two homologs show distinct identities to that of *Synechocystis* PCC 6803, namely 66.5% and 36.7%. Such discrepancy in CDS of *bicA* genes in strain B231 requires future studies to elucidate the possible evolutionary processes and functional differences. In addition, the HCO_3^- transport systems, *sbt* regulator and BCT1, were not present in strain B231. The absence of *sbt* regulator in strain B231 was consistent with the previous finding that the thermophilic cyanobacteria typically have *bicA* rather than *sbt* as suggested by their dominant presence. At the same time, a lack of BCT1 was observed in several thermophilic cyanobacteria, e.g., *Thermoleptolyngbya* and *Thermosynechococcus* strains (Tang et al., 2022d). Meanwhile, thermophiles without both *sbt* regulator and BCT1 were rare (Tang et al., 2022d). This result suggests that the BicA in strain B231 may be sufficient to manage the transport of dissolved inorganic carbon in its habitat.

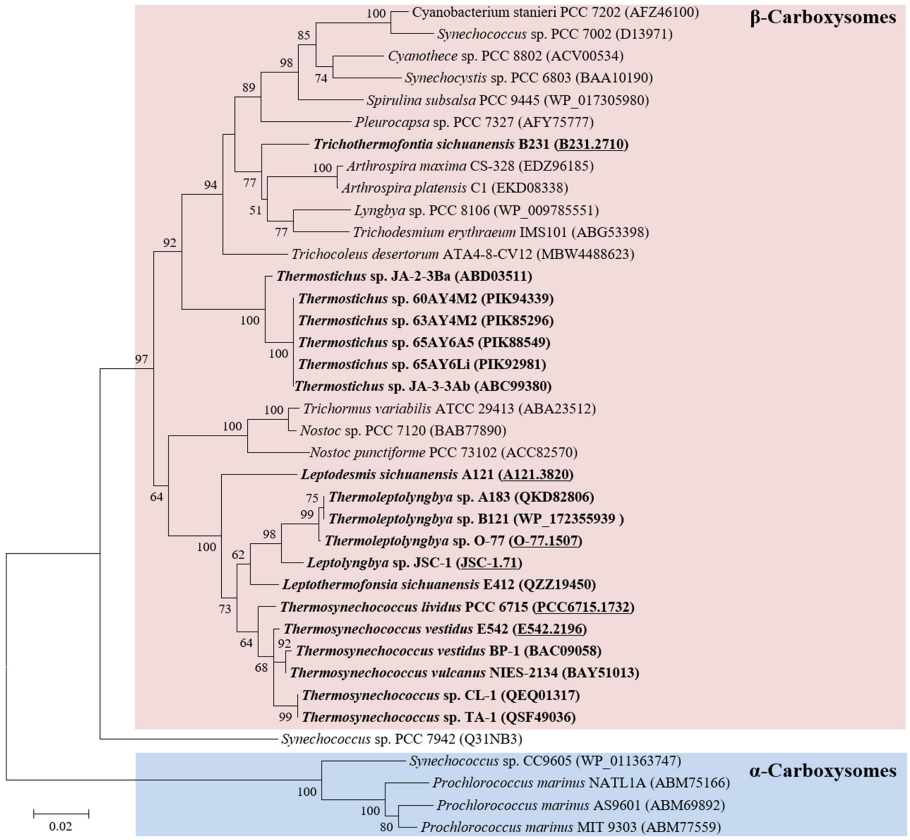


FIGURE 3
ML phylogenetic inference of RubisCO large subunit protein sequences. The thermophilic cyanobacterium investigated in this study and previously reported thermophiles are indicated in bold. The accession numbers underscored refer to the gene IDs in Supplementary Table 1 or the reference (Tang et al., 2022d).

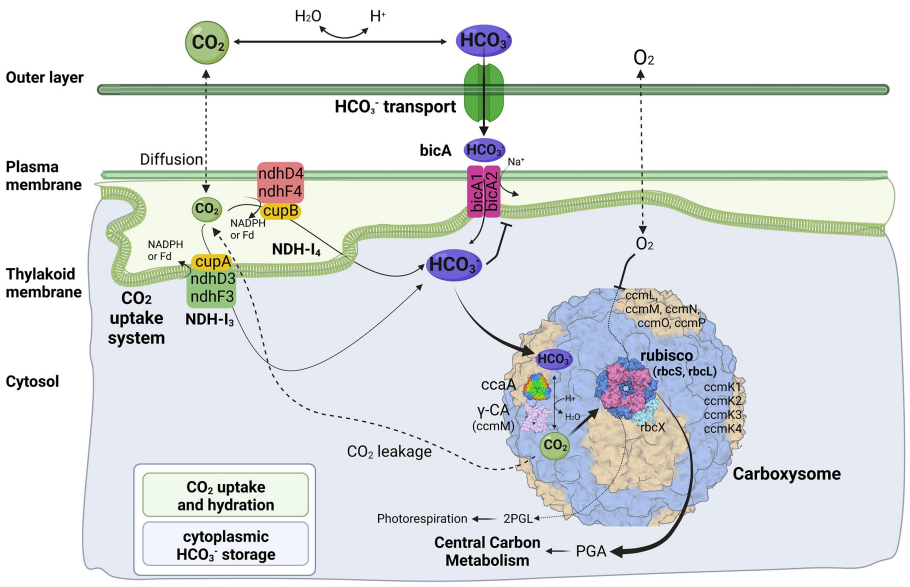


FIGURE 4
Molecular components of CCM for *Trichothermofontia sichuanensis* B231. 2PGL, 2-phosphoglycolate; PGA, phosphoglyceric acid. Three-dimensional structures of proteins visualized using related structures of 2YBV, 6OWF, 5SWC, 5BS1, 6KI1. Figure created with BioRender.com.

Genes encoding carboxysomes in strain B231

The cyanobacterial carboxysomes comprise protein shells and two encapsulated enzymes, RubisCO and carbonic anhydrase (CA; Kerfeld and Melnicki, 2016). For β-cyanobacteria, shell proteins are normally encoded by *ccmKLMNO* operon and *ccmP* (Melnicki et al., 2021). Strain B231 contains *ccmK1-4* (Figure 4), a typical gene set in the β-cyanobacterial genome (MacCready et al., 2020). It is known that *ccmK1* and *ccmK2* are the main structural proteins of the carboxysome shell and share high sequence conservation (Cai et al., 2016). Phylogenetic analysis suggests that the two homologs of *ccmK1/2* in strain B231 form a separate cluster, and both group into the clade of *ccmK2* (Figure 5A). Although 10 amino acid-long C-terminal extension of *ccmK1* can distinguish *ccmK1* and *ccmK2* (Kerfeld et al., 2005), the remaining part of *ccmK1/2* protein sequences share a similarity as high as 92.2%, leading to their assembly into one cluster. The structural and functional specificity of the two *ccmK* genes requires future careful investigation. The *ccmK3* and *ccmK4* of strain B231 were relatively divergent from other cyanobacteria, as suggested by long branches and substitution rates (Figures 5B,C). The *ccmK3* and *ccmK4* of strain B231 show a moderate degree of homology to that of *Synechocystis* PCC 6803, 61.5% and 74.5%, respectively. The *ccmK3/K4* present in the strain B231 may function as adjusting the properties of carboxysome for rapid adaptation to environmental changes in thermal regions through, e.g., expanding the range of permeability properties of metabolite channels in carboxysome shells (Sommer et al., 2019).

Apart from *ccmK*, other genes encoding carboxysome shell proteins, *ccmL*, *-M*, *-N*, *-O*, and *-P*, were also present in strain B231 (Figure 4). Only *ccmM* and *ccmN* show weak homologs (46.2, 32.1%) to that of *Synechocystis* PCC 6803, while the other three genes exhibit an identity of 62.2%, 61.4%, and 72.0%, respectively. Phylogenetic analysis of these genes suggests extensive genetic divergence between strain B231 and reference cyanobacteria, as suggested by the assignments of strain B231 into separate branches (Supplementary Figure 3).

The carboxysomal CA catalyzes the conversion of HCO_3^- into CO_2, a substrate for RubisCO. The subunits of RubisCO, *rbcL* and *rbcS*, and RubisCO assembly chaperone, *rbcX*, were present in strain B231 (Figure 4). The *rbcL* of strain B231 shows high sequence conservation to *Synechocystis* PCC 6803 (89.7% identity), whereas moderate conservation (66.4% identity) was observed in *rbcS*. The protein sequence of *rbcX* was more divergent, showing 58% identity with the reference protein. As for CA, only one type of β-CA, carboxysomal *ccaA* (Figure 4), was present in strain B231, with a similarity of 56.4% to the reference protein. Moreover, Strain B231 shows a similar primary structure (Supplementary Figure 4) of amino acid residues in γ-CA-like domain of *ccmM* to that with CA activity in *Thermosynechococcus* BP-1 (Peña et al., 2010) and *Nostoc* PCC 7120 (de Araujo et al., 2014), indicating that the *ccmM* protein of strain B231 may possess CA activity. Thus, the CA activities of *ccmM* and *ccaA* might confer strain B231 with more alternative strategies for regulating carboxysome function. Intriguingly, previous studies (Peña et al., 2010; de Araujo et al., 2014; Tang et al., 2022d) suggest that the CA activity of *ccmM* (γ-CA) was usually present in cyanobacteria lacking *ccaA*, while cyanobacteria comprising both *ccmM* and *ccaA* are likely to have non-functional γ-CA domain and only *ccaA* function as CA activity. Future experimental studies should be performed to elucidate the function and relationship of the two identified proteins with potential CA activity in strain B231.

Genomic organization of CCM-related genes in strain B231

Investigation of the genomic organization of CCM-related genes in strain B231 may provide insights into the function and evolution of these genes. As shown in Figure 6, genes encoding each NDH-1 complex separately cluster together. The two homologs of *bicA* genes are distantly located on the genome (Figure 6). Regarding the gene organization of the carboxysomal shell proteins, a typical main carboxysome locus (MCL) was present. It contained *ccmK1/2* genes, followed by *ccmL*, *-M*, *-N*, and *-O* (Figure 6). The sequential arrangement of *ccmK2* and *ccmK1* in the MCL may facilitate protein complex assembly or balancing of the shell protein stoichiometry during the translation of MCL genes (Cai et al., 2016), while *ccmK*, *-L*, *-M*, *-N*, and *-O* may be co-regulated in an MCL as an operon (Rae et al., 2013; Billis et al., 2014; Sommer et al., 2017). In addition to the MCL, the other three shell proteins form two satellite loci, *ccmP* and an operon with *ccmK3/4* genes. The separated organization of *K1/K2* and *K3/K4* paralogs may be associated with the structural segregation of the two groups (Sommer et al., 2019). Moreover, the expression of *ccmK3/4* from a satellite locus may increase the flexibility of carboxysome shell assembly and permeability (Sommer et al., 2019), and may provide differing metabolite selectivities (Sommer et al., 2017). The localization of genes encoding *rbcL*, *rbcS* and *rbcX* for RubisCO was consistent with the previous findings that the three genes always cluster in an operon in β-cyanobacteria (Badger et al., 2002). In addition, the RubisCO gene cluster was remote from the MCL, suggesting that this gene cluster was a satellite locus to MCL and may conduct independent expression regulation. The role of *rbcX* in thermophilic cyanobacteria is still uncertain and requires future investigations.

Comparative analysis indicates that strain B231 shows a distinct molecular component to the phylogenetically closely related neighbor, *Trichocoleus* ATA4-8-CV12 (Supplementary Table 2). For the CO_2 uptake systems, strain B231 has one set of genes encoding NDH-1_4 complex, while *Trichocoleus* ATA4-8-CV12 has two sets of homologs. The constitutive low-affinity CO_2 uptake system (NDH-1_4 complex) with an extra set of genes may be indispensable for the strain to survive in its terrestrial origin (desert) where gaseous CO_2 is the primary carbon source (Mühlsteinova et al., 2014) and bicarbonate is less available, or simply the aftermath of strain evolutionary history. More studies are needed on the topic of multiple paralogs of carbon uptake systems in cyanobacteria. Both strains have only one type of HCO_3^- transport system, namely BicA in strain B231 and BCT1 in strain ATA4-8-CV12. A different type of HCO_3^- transport system could be attributed to the distinct traits of the two transporters and the habitat of the two strains. First, BicA exhibits low affinity and high flux rate to HCO_3^- and the genes encoding BicA were found to be primarily constitutively expressed (Price et al., 2004). Therefore, the BicA transport system might be necessary for strain B231 to acclimatize to the exogenous HCO_3^- in its habitat (pH 6.62) where HCO_3^- remains dominant dissolved inorganic specie between pH 6 and 10. Second, BCT1 encoded by *cmpABCD* operon shows a high affinity for

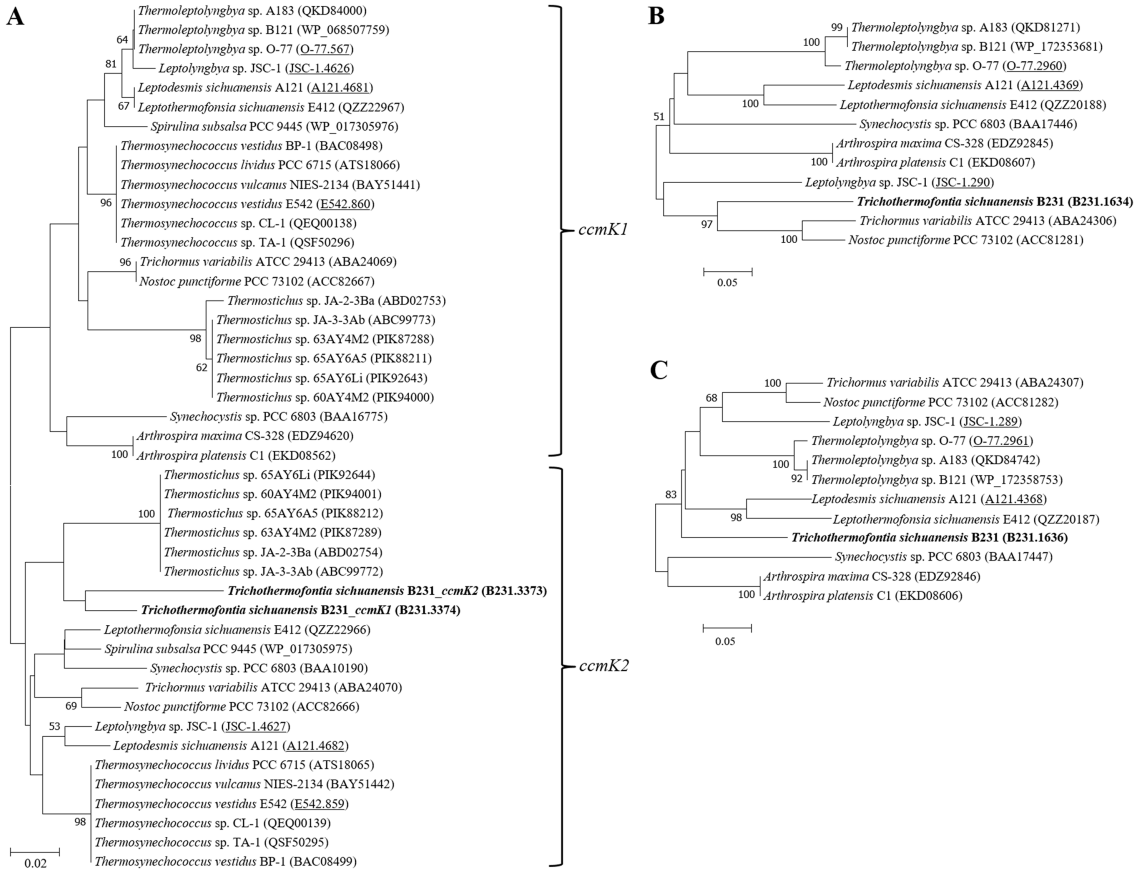


FIGURE 5
ML phylogenetic inference of protein sequences of *ccmK1/2* **(A)**, *ccmK3* **(B)**, and *ccmK4* **(C)**. The thermophilic cyanobacterium investigated in this study is indicated in bold. The accession numbers underscored refer to the gene IDs in Supplementary Table 1 or the reference (Tang et al., 2022d).

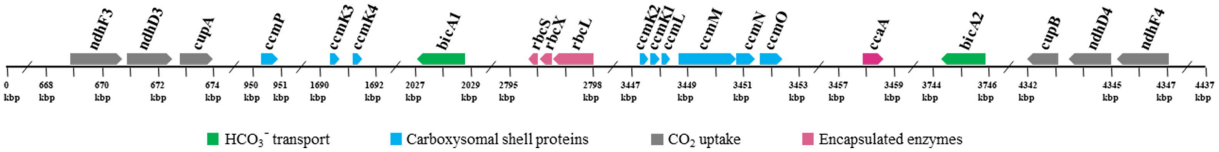


FIGURE 6
Genomic organization of CCM-related genes for *Trichothermofontia sichuanensis* B231. Solid arrow boxes refer to genes and the direction of transcription.

HCO_3^- and was induced under low levels of Ci and enhanced by high light conditions (Omata et al., 2002). Thus, strain ATA4-8-CV12 may use BCT1 to compensate for the functions of constitutive genes (e.g., BicA) for adaptation to high-light conditions in deserts. In addition, strain B231 has one type of β-CA (*ccaA*) and *ccmM* (γ-CA) with CA activity, whereas *Trichocoleus* ATA4-8-CV12 contains two types of β-CA (*ccaA* and *ecaB*) and its *ccmM* appears to have non-functional γ-CA domain (Supplementary Figure 4). The presences of other genes related to CCM are consistent with the two strains. However, the phylogenetic analysis suggests that the protein sequences are divergent (Figures 3, 5; Supplementary Figure 3). Though the genome of *Trichocoleus* ATA4-8-CV12 was only assembled as a draft, preliminary analysis suggests that its clustering pattern of CCM-related genes in the genome was in accordance with that of strain B231.

Overall, the molecular component and organization of strain B231 were similar to that of other reported thermophilic cyanobacteria (Supplementary Table 2; Tang et al., 2022d). However, distinct characteristics were evident in strain B231. A relatively low diversity of bicarbonate transporters distinguishes strain B231 from other thermophilic strains, each possessing at least two HCO_3^- transport systems (Tang et al., 2022d). The low diversity may be related to the habitat of strain B231 where a higher variety of transporters with similar functionalities is not required. In addition, strain B231 contains a higher abundance of different types of carboxysomal CAs, β-CA (*ccaA*) and γ-CA (*ccmM*), compared to other thermophilic strains (β-CA or γ-CA), but still below those of other freshwater cyanobacteria that possess the abundance of CAs, α-/β-/γ-CAs (Tang et al., 2022d). The CA activities of *ccmM* and *ccaA* may provide strain

B231 with alternative strategies to ensure that the RubisCO with low affinity for CO_2 was surrounded by high CO_2 levels and regularly function regardless of thermal stress in this habitat. Strain B231 with freshwater origin lacks the BCT1 transporter, which is near-ubiquitous in freshwater cyanobacteria. Moreover, strain B231 shows a similar composition of carboxysome shell proteins to mesophilic cyanobacteria, the diversity of which was higher than many thermophilic strains lacking at least one of the four *ccmK* genes (Tang et al., 2022d).

Morphological and physiological features of strain B231

The primary physiological characteristics of the B231 strain were assessed by monitoring its growth at the temperature of 47°C in derivatives of BG-11 medium; the unmodified medium, supplemented with bicarbonate up to the concentration of 1.0 M, and nitrogen-free. The strain could grow in all tested nitrogen conditions, showing optimal growth with moderate (a third of the regular BG-11 medium nitrate content) to high nitrogen availability (five times the regular BG-11 medium nitrate content) and in the presence of 17.65 mM nitrite. Interestingly, despite the lack of heterocytes the strain could grow in a nitrogen-free medium, which is consistent with the presence of the *nif*HDK gene cluster and associated genes required for molecular nitrogen fixation. The strain grew optimally from ambient up to the 0.1 M bicarbonate concentration. Higher loads of dissolved inorganic carbon had a deleterious effect on the strain, concurrent with its highly restricted bicarbonate transport system revealed earlier. The results are summarized in Table 5.

TABLE 5 Physiological characteristics of *Trichothermofontia* B231 grown in 47°C, 150μmolm^{-2} s^{-1} under varying carbonate and nitrate concentrations.

Supplemented inorganic carbon		Supplemented nitrogen	
HCO_3^- (mol/L)	Growth	NO_3^- (mmol/L)	Growth
0 (ambient)	++	0	+
0.1	+	0.88	+
0.3	–	5.88	++
0.5	–	17.65	++
0.7	–	58.83	++
1.0	–	88.24	++

"–" no growth; "+" poor growth; "++" baseline growth in BG-11.

Light, scanning, and transmission microscopy were utilized for the morphological description and delineation of strain B231 as the representative of a new genus and to illustrate its key characteristics. The morphology of the strain is presented in Figure 7. The unicyanobacterial culture consists of blue-green filaments (Figures 7A,B), sometimes forming clumps or flat mats when grown in stationary liquid cultures. Trichomes of the cells grown in the liquid medium are solitary or in colony-like mats, straight or bent, moderately entangled or curved, and contain elongated cylindrical barrel-shaped cells terminated with mostly sharply pointed, conical apical cells (Figures 7A,C). Cells grown on a solid medium had multiple trichomes in one sheath (Figure 7B).

Cells of the strain were longer than they were wide (2.0–2.4 μm wide, 3.4–5.0 μm long), with a length-to-width ratio ranging from 1.4–2.0, and with an irregular ratio in apical cells. The trichomes of this strain were thinner than those of other members of *Trichocoleus* (Table 6), which typically have length-to-width below 1.0 and *Pinocchia* with very diverse measurements for both cell lengths and widths. The sheath was not evident in liquid cultures but more apparent in solidified medium, cells were connected with hyaline bridges, slightly constricted at poorly visible cross-walls, with the content mostly homogeneous (Figures 7A,B). The number of thylakoids (Figures 7E,F) varies from 5 to 8 layers, and peripheral thylakoids were arrayed layer by layer, parallel to the long axis of the filament, occasionally with an irregular arrangement. The presence of large cyanophycin granules, often in the center of the cell, characterized the cytoplasm (Figures 7D,E). The relevant biosynthetic genes encoding cyanophycin synthase (*cphA*) and cyanophycinase (*cphB*) were identified in the genome. The carboxysomes were observed in the cytoplasm in small numbers, similar to related *Pinocchia* strains (Dvorak et al., 2015; Kim et al., 2021). Gas vesicles and polyphosphate granules were also observed in the cytoplasm (Figure 7D). Gas vesicles were composed of cylindrical structures, typically two proteins GvpA forming the vesicle core and GvpC providing structural support and were responsible for the shape of gas vesicles (Miklaszewska et al., 2012). Analysis of the B231 genome confirms the presence of the *gvpA*, *gvpC* genes. The other genes involved in gas vesicle formation were also identified.

As shown in Table 6, the overall morphology of strain B231 was more similar to *Pinocchia* spp. (Dvorak et al., 2015; Kim et al., 2021) than to other members of the *Trichocoleusaceae* family (Lange et al., 1992; Flechtner et al., 2009; Alwathnani and Johansen, 2011; Komárek and Kovacik, 2013; Roncero-Ramos et al., 2019; Mehda et al., 2021). This difference can be linked to the habitat. The family appears to have two morphologies, probably driven by environmental conditions. The primarily terrestrial strains, such as *T. desertorum,* have a very thick sheath capable of maintaining multiple trichomes. Meanwhile, freshwater strains, such as *Pinocchia* spp. and the B231, exhibit phenotypic plasticity showing marginal sheath when grown in liquid cultures and thick sheath when grown on a solid medium.

Conclusion

In this manuscript, the results of 16S rRNA phylogeny, secondary structures of 16S-23S ITS and strain's morphology strongly supported the B231 as a novel genus within *Trichocoleusaceae* and its delineation as a representative of this new taxon. Phylogenomic inference and three genome-based indices (ANI, AAI, and POCP) also support the delineation at the genus level. Consequently, we have proposed a new genus *Trichothermofontia sichuanensis* gen. et sp. nov. Moreover, based on the results, we suggest family-level revision of *Pinocchia* from *Leptolyngbyaceae* to *Trichocoleusaceae*. The delineation was strongly supported at the molecular level, and its morphological distinction from *Trichocoleus* sp. can be attributed to the different habitats (terrestrial vs. freshwater). In addition, the obtained complete genome of the newly delineated *Trichothermofontia* B231 facilitates the elucidation of genetic basis regarding genes related to CCM. The strain belongs to β-cyanobacteria according to its β-carboxysome shell

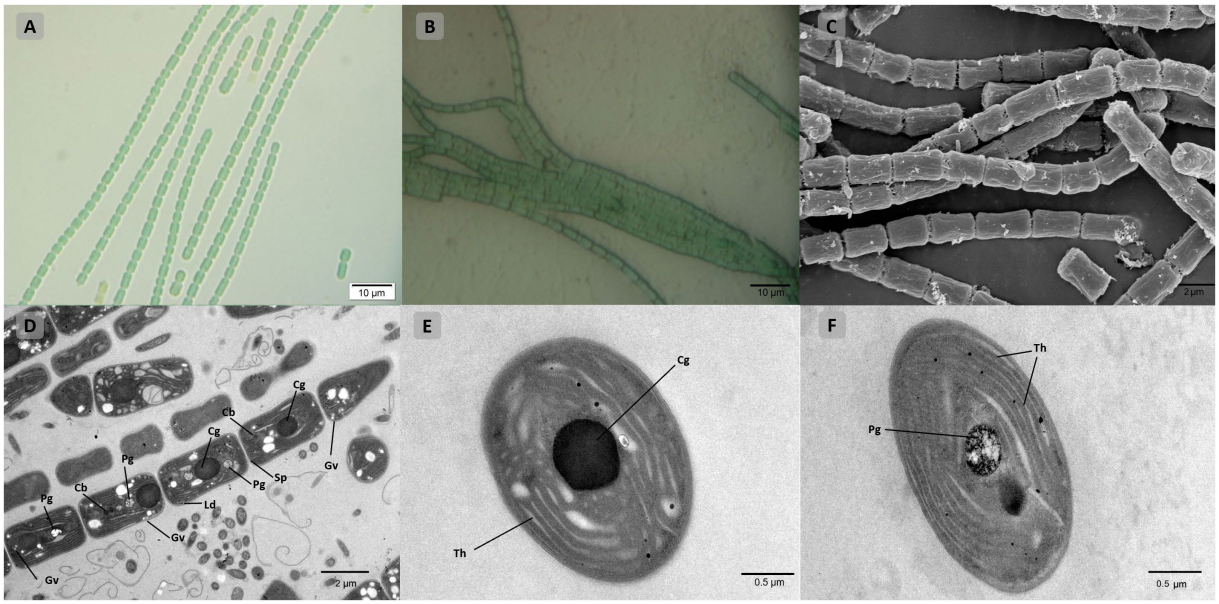


FIGURE 7
Microscopic morphology of *Trichothermofontia* B231. **(A)** Light microscopy image of liquid-medium grown cultures; **(B)** light microscopy image of solid-medium grown cultures; **(C)** SEM image; **(D–F)** TEM images. Cb, carboxysome; Cg, cyanophycin granule; Ld, lipid droplets; Pg, polyphosphate granule; Sp, septum; Th, thylakoid membrane, Gv, gas vesicles. Magnifications were 1,000× **(A,B)**; 5,000× **(C,D)**; 20,000× **(E,F)**.

protein with 1B form of RubisCO. Compared to other thermophilic strains, strain B231 contains a relatively low diversity of bicarbonate transporters (only BicA for HCO_3^- transport) but a higher abundance of different types of carbonic anhydrase (CA), β-CA (*ccaA*) and γ-CA (*ccmM*). Strain B231 with freshwater origin lacks the BCT1 transporter, which is consistently possessed by freshwater cyanobacteria. Furthermore, strain B231 shows a similar composition of carboxysome shell proteins (*ccmK1-4*, *ccmL*, *-M*, *-N*, *-O*, and *-P*) to mesophilic cyanobacteria, the diversity of which was higher than many thermophilic strains lacking at least one of the four *ccmK* genes. The genomic distribution of CCM-related genes suggests various regulations of expression. Overall, the first complete genome of a new genus representative obtained in this study provides insight into the genomic features, CCM components, and the fundamental information for future global taxogenomic, ecogenomic and geogenomic studies.

Taxonomic treatment and description of *Trichothermofontia sichuanensis* Daroch, Tang, and Zhou et al. gen. et sp. nov.

Phylum: Cyanobacteria
Order: Synechococcales
Family: *Trichocoleusaceae*
Genus: Trichothermofontia, gen. nov.

Description: Filaments solitary, entangled, lacking false branching, forming small or short fragments in a nutritional deficiency state. Sheath colorless, thin. Trichomes straight or curved, terminated with apical, rounded cells, with invisible crosswalls that separate each cell. Cells mostly longer than wide, with multilayer peripheral thylakoids.

Etymology: Gr. fem. n. *thrix (gen. trichos)*, hair; Gr. masc. Adj. *thermos*, hot; L. masc. n. *fons (gen. fontis)*. "*Tricho*" referring to Greek word for hair, representing exhibiting typical thin morphology of trichome, "thermo" referring to Greek word for heat and representing thermophilic (high temperature tolerant) character of the strain, "fontia"—a genus epithet derived from the Latin word *fons* meaning spring, since all representative of the genus to date were hot spring isolates.

Type species: *Trichothermofontia sichuanensis* Daroch, Tang, and Zhou et al. sp. nov.

Trichothermofontia sichuanensis Daroch, Tang, and Zhou et al. gen. et. sp. nov.

Diagnosis: Differing from other species of the genus based on the 16S rRNA sequence identity.

Description Colony bright green, flat and tightly packed in a recognizable interleaved or tangled reticular morphology on agar plates. Filaments long or short, blue-green, no branching, typically solitary straight or bent, moderately entangled or curved (Figures 7A,B). Trichomes contained elongated cylindrical barrel-shaped cells terminated with conical, sharply pointed apical cells, mat-forming in stationary liquid culture, yielding solitary filaments when shaken. Cells rectangular in cross-section, connected with hyaline bridges. Intracellular connections between vegetative cells not present. Cells typically longer (3.4–4.0 μm), than wide (2.0–2.4 μm), with a length-to-width ratio ranging from 1.4 to 2.0 (Figures 7A–D), with peripheral thylakoids arrayed in 5 to 8 layers (Figures 7E,F). The sheath minimal in liquid cultures (Figure 7A), and visible in solid-medium grown cultures (Figure 7B). Carboxysomes observed in small numbers. Large centrally positioned cyanophycin granule

TABLE 6 Comparison of morphological features of *Trichothermofontia* and closely related strains.

Species	Morphology	Cell width (um)	Cell length (um)	Sheaths	Thylakoids No.	Color	References
Trichothermofontia sichuanensis B231	Straight solitary unbranched filaments, multiple filaments in one sheath when grown on solid medium	2.0–2.4	2.0–5.0	Thin, colorless	5–7	Blue-green	This study
Trichocoleus desertorum LSB90	Straight, solitary, unbranched filaments, multiple filaments in one sheath	3.9 ± 0.3	3.0 ± 0.5	Thick, colorless	NA	Green	Mehda et al. (2021)
Trichocoleus desertorum CAU7	Straight, solitary, unbranched filaments, multiple filaments in one sheath	2.8 ± 0.4	2.0 ± 0.5	Colorless	NA	Green	Roncero-Ramos et al. (2019)
Trichocoleus desertorum ATA4-8-CV2	Entangled, solitary, unbranched filaments, multiple filaments in one sheath	2.5–3.8	1.5–5.5	Colorless	NA	Blue-green	Mühlsteinova et al. (2014)
Trichocoleus abiscoensis	Solitary, 1–4 to more trichomes in one sheath	1.0–2.0	NA	Colorless	NA	Blue-green	Komárek and Kovacik (2013)
Trichocoleus sociatus LSB16	Solitary, unbranched filaments, multiple filaments in one sheath	5.2 ± 0.2	3.3 ± 0.4	Colorless	NA	Green	Mehda et al. (2021)
Trichocoleus sociatus SAG 26.92	30 trichomes in one sheath	2.0–2.8	NA	Colorless	NA	Blue-green	Lange et al. (1992)
Trichocoleus cf. delicatulus (W. et G.S. West) Anag.	Flexuous, spreading, ropy fascicles, unbranched filaments	1.2–1.6	1.6–3.5	Colorless	NA	Olive	Flechtner et al. (2009)
Trichocoleus sp. 1	Slightly flexuous, tapering, unbranched filaments, 1–2 trichomes per sheath	2.5–3.0	1.5–3.5	Thin, colorless	NA	Blue-green	Alwathnani and Johansen (2011)
Pinocchia polymorpha E5	Straight solitary unbranched filaments	1.09–2.86	1.28–8.63	Thin, colorless	NA	Blue-green	Dvorak et al. (2015)
Pinocchia daecheonga FBCC-A230	Straight or bent filaments, constricted at cross-walls	1.04–1.87	1.57–5.99	Thin, colorless	NA	Blue-green	Kim et al. (2021)

NA, not available.

(Figures 7D,E), and membrane-delimited vesicles (polyphosphate granules and gas vesicles) in the cytoplasm. Development of heterocytes not observed, genetic toolkit for molecular nitrogen fixation detected.

Etymology: "*sichuanensis*" species epithet derives from the name of the collection province is B231 (=FACHB-3573).

***Type locality*:** Thermal spring, Zhonggu village in Ganzi Prefecture of Sichuan Province, China.

Ecology of type locality: the sample occurred as a macroscopic dark green mat attached to the rock.

***Habitat*:** Thermal springs in Ganzi Prefecture of Sichuan Province, China (30°36′39″ N, 101°41′9″ E; Supplementary Figure 5).

Holotype here designated: The dried inactive holotype was deposited in the Herbarium of North Minzu University with the voucher number: NMU00231 (contact: Lei Zhang, zhangsanshi-0319@163.com).

Reference strain: The culture of *Trichothermofontia sichuanensis* Daroch, Tang, and Zhou et al. gen. et sp. nov. was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTB231 has also been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with accession

number FACHB-3573 as *Trichocoleusaceae* sp. 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-3573.

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

JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing-original draft, writing-review and editing, visualization, supervision, project administration, and funding acquisition. HZ: formal analysis, investigation, data curation, and writing-original draft. YJ: formal analysis, investigation, data curation, and writing-review and editing. DY: formal analysis, investigation, and data curation. KW: methodology, validation, data curation, and writing-review and editing. L-MD: methodology, software, and data curation. 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, 32071480, and 3221101094) and Tenure-Track Fund to MD. Funding bodies had no influence over the design and execution of this research.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1111809/full#supplementary-material

References

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

Alwathnani, H., and Johansen, J. R. (2011). Cyanobacteria in soils from a mojave desert ecosystem. *Monogr. West. N. Am. Nat.* 5:19. doi: 10.3398/042.005.0103

Badger, M. R., Hanson, D., and Price, G. D. (2002). Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria. *Funct. Plant Biol.* 29, 161–173. doi: 10.1071/PP01213

Billis, K., Billini, M., Tripp, H. J., Kyrpides, N. C., and Mavromatis, K. (2014). Comparative transcriptomics between *Synechococcus* PCC 7942 and *Synechocystis* PCC 6803 provide insights into mechanisms of stress acclimation. *PLoS One* 9:e109738. doi: 10.1371/journal.pone.0109738

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5:8365. doi: 10.1038/srep08365

Cai, F., Bernstein, S. L., and Wilson, S. C. (2016). Production and characterization of synthetic carboxysome shells with incorporated luminal proteins. *Plant Physiol.* 170, 1868–1877. doi: 10.1104/pp.15.01822

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

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

de Araujo, C., Arefeen, D., Tadesse, Y., Long, B. M., Price, G. D., Rowlett, R. S., et al. (2014). Identification and characterization of a carboxysomal γ-carbonic anhydrase from the cyanobacterium *Nostoc* sp. PCC 7120. *Photosynth. Res.* 121, 135–150. doi: 10.1007/s11120-014-0018-4

Drummond, A. J., Ho, S. Y., Phillips, M. J., and Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. *PLoS Biol.* 4:e88. doi: 10.1371/journal.pbio.0040088

Durall, C., and Lindblad, P. (2015). Mechanisms of carbon fixation and engineering for increased carbon fixation in cyanobacteria. *Algal Res.* 11, 263–270. doi: 10.1016/j.algal.2015.07.002

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, 024–034. doi: 10.11646/phytotaxa.170.1.3

Dvorak, P., Jahodářová, E., Hasler, P., Gusev, E., and Poulíčková, A. (2015). A new tropical cyanobacterium *Pinocchia polymorpha* gen. Et sp. nov. derived from the genus *Pseudanabaena*. *Fottea* 15, 113–120. doi: 10.5507/fot.2015.010

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

Flechtner, V., Johansen, J., and Belnap, J. (2009). The biological soil crusts of the san nicolas island: enigmatic algae from a geographically isolated ecosystem. *West. N. Am. Nat.* 68, 405–436. doi: 10.3398/1527-0904-68.4.405

Galmés, J., Hermida-Carrera, C., Laanisto, L., and Niinemets, Ü. (2016). A compendium of temperature responses of Rubisco kinetic traits: variability among and within photosynthetic groups and impacts on photosynthesis modeling. *J. Exp. Bot.* 67, 5067–5091. doi: 10.1093/jxb/erw267

Galmés, J., Kapralov, M. V., Copolovici, L. O., Hermida-Carrera, C., and Niinemets, Ü. (2015). Temperature responses of the Rubisco maximum carboxylase activity across domains of life: phylogenetic signals, trade-offs, and importance for carbon gain. *Photosynth. Res.* 123, 183–201. doi: 10.1007/s11120-014-0067-8

Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472.

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

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

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kerfeld, C. A., and Melnicki, M. R. (2016). Assembly, function and evolution of cyanobacterial carboxysomes. *Curr. Opin. Plant Biol.* 31, 66–75. doi: 10.1016/j.pbi.2016.03.009

Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M., et al. (2005). Protein structures forming the shell of primitive bacterial organelles. *Science* 309, 936–938. doi: 10.1126/science.1113397

Kim, D.-H., Choi, H. J., Ki, J.-S., and Lee, O.-M. (2021). *Pinocchia daecheonga* sp. nov. (Synechococcales, cyanobacteria) isolated from a Daecheong Lake in Geum River, Republic of Korea. *Phytotaxa* 510, 135–147. doi: 10.11646/phytotaxa.510.2.2

Klanchui, A., Cheevadhanarak, S., Prommeenate, P., and Meechai, A. (2017). Exploring components of the CO_2-concentrating mechanism in alkaliphilic cyanobacteria through genome-based analysis. *Comput. Struct. Biotechnol. J.* 15, 340–350. doi: 10.1016/j.csbj.2017.05.001

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

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., 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 Kovacik, L. (2013). Schizotrichacean cyanobacteria from Central Spitsbergen (Svalbard). *Polar Biol.* 36, 1811–1822. doi: 10.1007/s00300-013-1402-9

Kono, M., Martinez, J., Sato, T., and Haruta, S. (2022). Draft genome sequence of the thermophilic unicellular cyanobacterium Synechococcus sp. strain C9. *Microbiol. Resour. Announc.* 11:e0029422. doi: 10.1128/mra.00294-22

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

Lange, O. L., Kidron, G. J., Budel, B., Meyer, A., and Abeliovich, E. K. (1992). Taxonomic composition and photosynthetic characteristics of the biological soil crusts' covering sand dunes in the western Negev Desert. *Funct. Ecol.* 6, 519–527. doi: 10.2307/2390048

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 CO_2 utilisation. *Bioresour. Technol.* 278, 255–265.

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

MacCready, J. S., Basalla, J. L., and Vecchiarelli, A. G. (2020). Origin and evolution of carboxysome positioning systems in cyanobacteria. *Mol. Biol. Evol.* 37, 1434–1451. doi: 10.1093/molbev/msz308

Mai, T., Johansen, J. R., Pietrasiak, N., Bohunicka, M., and Martin, M. P. (2018). Revision of the Synechococcales (cyanobacteria) through recognition of four families including Oculatellaceae fam. Nov. and *Trichocoleaceae* fam nov. and seven new genera containing 14 species. *Phytotaxa* 365, 1–59. doi: 10.11646/phytotaxa.365.1.1

Mehda, S., Muñoz-Martín, M., Oustani, M., Hamdi-Aïssa, B., Perona, E., and Mateo, P. (2021). Microenvironmental conditions drive the differential cyanobacterial community composition of biocrusts from the sahara desert. *Microorganisms* 9:487. doi: 10.3390/microorganisms9030487

Melnicki, M. R., Sutter, M., and Kerfeld, C. A. (2021). Evolutionary relationships among shell proteins of carboxysomes and metabolosomes. *Curr. Opin. Microbiol.* 63, 1–9. doi: 10.1016/j.mib.2021.05.011

Miklaszewska, M., Waleron, M., Morin, N., Calusinska, M., Wilmotte, A., Tandeau De Marsac, N., et al. (2012). Elucidation of the gas vesicle gene clusters in cyanobacteria of the genus *Arthrospira* (Oscillatoriales, Cyanophyta) and correlation with ITS phylogeny. *Eur. J. Phycol.* 47, 233–244. doi: 10.1080/09670262.2012.692817

Miller, S. R., Castenholz, R. W., and Pedersen, D. (2007). Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Appl. Environ. Microbiol.* 73, 4751–4759. doi: 10.1128/AEM.02945-06

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

Mühlsteinova, R., Johansen, J. R., Pietrasiak, N., Martin, M. P., Osorio-Santos, K., and Warren, S. D. (2014). Polyphasic characterization of *Trichocoleus desertorum* sp. nov. (Pseudanabaenales, cyanobacteria) from desert soils and phylogenetic placement of the genus *Trichocoleus*. *Phytotaxa* 163, 241–261. doi: 10.11646/phytotaxa.163.5.1

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, 733–745. doi: 10.1093/nar/gkv1189

Omata, T., Takahashi, Y., Yamaguchi, O., and Nishimura, T. (2002). Structure, function and regulation of the cyanobacterial high-affinity bicarbonate transporter, BCT1. *Funct. Plant Biol.* 29, 151–159. doi: 10.1071/PP01215

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

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

Peña, K. L., Castel, S. E., de Araujo, C., Espie, G. S., and Kimber, M. S. (2010). Structural basis of the oxidative activation of the carboxysomal γ-carbonic anhydrase, CcmM. *Proc. Natl. Acad. Sci.* 107, 2455–2460. doi: 10.1073/pnas.0910866107

Price, G. D., Woodger, F. J., Badger, M. R., Howitt, S. M., and Tucker, L. (2004). Identification of a SulP-type bicarbonate transporter in marine cyanobacteria. *Proc. Natl. Acad. Sci.* 101, 18228–18233. doi: 10.1073/pnas.0405211101

Pronina, N. A., Kupriyanova, E. V., and Igamberdiev, A. U. (2017). "Photosynthetic carbon metabolism and CO_2-concentrating mechanism of cyanobacteria" in *Modern topics in the phototrophic prokaryotes: Metabolism, bioenergetics, and omics*. ed. P. C. Hallenbeck (Cham: Springer International Publishing), 271–303.

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. doi: 10.1128/JB.01688-14

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

Rae, B. D., Long, B. M., Badger, M. R., and Price, G. D. (2013). Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO_2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.* 77, 357–379. doi: 10.1128/MMBR.00061-12

Roncero-Ramos, B., Muñoz-Martín, M., Chamizo, S., Fernández-Valbuena, L., Mendoza, D., Perona, E., et al. (2019). Polyphasic evaluation of key cyanobacteria in biocrusts from the most arid region in Europe. *PeerJ* 7:e6169. doi: 10.7717/peerj.6169

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

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

Sommer, M., Cai, F., Melnicki, M., and Kerfeld, C. A. (2017). β-Carboxysome bioinformatics: identification and evolution of new bacterial microcompartment protein gene classes and core locus constraints. *J. Exp. Bot.* 68, 3841–3855. doi: 10.1093/jxb/erx115

Sommer, M., Sutter, M., Gupta, S., Kirst, H., Turmo, A., Lechno-Yossef, S., et al. (2019). Heterohexamers formed by CcmK3 and CcmK4 increase the complexity of Beta carboxysome shells. *Plant Physiol.* 179, 156–167. doi: 10.1104/pp.18.01190

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

Tang, J., Du, L., Li, M., Yao, D., Waleron, M., Waleron, K. F., et al. (2022a). 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., 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., 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

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., Shah, M. R., Yao, D., Du, L., Zhao, K., Li, L., et al. (2022b). 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

Tang, J., Yao, D., Zhou, H., Du, L., and Daroch, M. (2022c). Reevaluation of *Parasynechococcus*-like strains and genomic analysis of their microsatellites and compound microsatellites. *Plan. Theory* 11:1060. doi: 10.3390/plants11081060

Tang, J., Zhou, H., Yao, D., Riaz, S., You, D., Klepacz-Smółka, A., et al. (2022d). Comparative genomic analysis revealed distinct molecular components and organization of CO_2-concentrating mechanism in thermophilic cyanobacteria. *Front. Microbiol.* 13:876272. doi: 10.3389/fmicb.2022.876272

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. In: *International code of nomenclature for algae, fungi, and plants (Shenzhen code) adopted by the nineteenth international botanical congress Shenzhen, China, July 2017*.

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

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

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415. doi: 10.1093/nar/gkg595

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 09 May 2023
DOI 10.3389/fmicb.2023.1192059

OPEN ACCESS

EDITED BY
Ram Karan,
King Abdullah University of Science and Technology, Saudi Arabia

REVIEWED BY
Aharon Oren,
Hebrew University of Jerusalem, Israel
Horia Leonard Banciu,
Babeș-Bolyai University, Romania

*CORRESPONDENCE
Cristina Sánchez-Porro
✉ sanpor@us.es
Antonio Ventosa
✉ ventosa@us.es

RECEIVED 22 March 2023
ACCEPTED 12 April 2023
PUBLISHED 09 May 2023

CITATION
Galisteo C, de la Haba RR, Sánchez-Porro C and Ventosa A (2023) A step into the rare biosphere: genomic features of the new genus *Terrihalobacillus* and the new species *Aquibacillus salsiterrae* from hypersaline soils. *Front. Microbiol.* 14:1192059. doi: 10.3389/fmicb.2023.1192059

A step into the rare biosphere: genomic features of the new genus *Terrihalobacillus* and the new species *Aquibacillus salsiterrae* from hypersaline soils

Cristina Galisteo, Rafael R. de la Haba, Cristina Sánchez-Porro* and Antonio Ventosa*

Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain

Hypersaline soils are a source of prokaryotic diversity that has been overlooked until very recently. The phylum *Bacillota*, which includes the genus *Aquibacillus*, is one of the 26 phyla that inhabit the heavy metal contaminated soils of the Odiel Saltmarshers Natural Area (Southwest Spain), according to previous research. In this study, we isolated a total of 32 strains closely related to the genus *Aquibacillus* by the traditional dilution-plating technique. Phylogenetic studies clustered them into two groups, and comparative genomic analyses revealed that one of them represents a new species within the genus *Aquibacillus*, whereas the other cluster constitutes a novel genus of the family *Bacillaceae*. We propose the designations *Aquibacillus salsiterrae* sp. nov. and *Terrihalobacillus insolitus* gen. nov., sp. nov., respectively, for these two new taxa. Genome mining analysis revealed dissimilitude in the metabolic traits of the isolates and their closest related genera, remarkably the distinctive presence of the well-conserved pathway for the biosynthesis of molybdenum cofactor in the species of the genera *Aquibacillus* and *Terrihalobacillus*, along with genes that encode molybdoenzymes and molybdate transporters, scarcely found in metagenomic dataset from this area. In-silico studies of the osmoregulatory strategy revealed a *salt-out* mechanism in the new species, which harbor the genes for biosynthesis and transport of the compatible solutes ectoine and glycine betaine. Comparative genomics showed genes related to heavy metal resistance, which seem required due to the contamination in the sampling area. The low values in the genome recruitment analysis indicate that the new species of the two genera, *Terrihalobacillus* and *Aquibacillus*, belong to the rare biosphere of representative hypersaline environments.

KEYWORDS

***Terrihalobacillus*, *Aquibacillus*, *Bacillota*, hypersaline soils, osmoregulation mechanism, phylogenomics, genome mining, rare biosphere**

1. Introduction

The genus *Aquibacillus*, first described in 2014, is one of the more than 100 genera of the family *Bacillaceae* within the phylum *Bacillota*. At the time of writing, it comprises a total of seven species (Parte et al., 2020), two of them being a reclassification of previously described *Virgibacillus* species (Amoozegar et al., 2014): *Aquibacillus halophilus* (Amoozegar et al., 2014),

Aquibacillus koreensis (Lee et al., 2006; Amoozegar et al., 2014), *Aquibacillus albus* (Zhang et al., 2012; Amoozegar et al., 2014), *Aquibacillus salifodinae* (Zhang et al., 2015), *Aquibacillus sediminis* (Lee and Whang, 2019), *Aquibacillus kalidii* (Wang et al., 2021), and *Aquibacillus saliphilus* (Cho and Whang, 2022). These species have been isolated from hypersaline environments, such as salt mine (Zhang et al., 2015), saltern soils (Lee et al., 2006; Lee and Whang, 2019), salt lakes (Zhang et al., 2012; Amoozegar et al., 2014), grey salterns (Cho and Whang, 2022), and *Kalidium cuspidatum* plants from saltern lands (Wang et al., 2021). The moderately halophilic species of the genus *Aquibacillus* are Gram-stain-positive endospore-forming rods with optimum growth between 4 and 10% (w/v) NaCl, at pH 7–8, and 25–37°C. They are motile and strictly aerobic, although the species *A. saliphilus* presents a facultatively anaerobic metabolism (Cho and Whang, 2022). The pigmentation of the colonies is cream to white color, their major fatty acid is anteiso-$C_{15:0}$ and their most predominant polar lipids are phosphatidylglycerol and diphosphatidylglycerol (Lee et al., 2006; Zhang et al., 2012, 2015; Amoozegar et al., 2014; Lee and Whang, 2019; Wang et al., 2021; Cho and Whang, 2022). In 2019, a new species with a very close relationship with the genus *Aquibacillus* was described as a new genus, *Radiobacillus*, based on its lack of motility and the presence of an aminophospholipid as one of the major polar lipids, among other characteristics (Li et al., 2020).

The Odiel Saltmarshes Natural Area represents a saline environment in Huelva, Southwest Spain, specifically between the Odiel and Tinto rivers. The area has suffered from industrial and mining activity for years, and some studies have determined it as contaminated by heavy metals (i.e., arsenic, cadmium, copper, lead, and zinc) (Sainz et al., 2002, 2004). The prokaryotic diversity of its hypersaline soils has been previously studied by metagenomic techniques (Vera-Gargallo and Ventosa, 2018; Vera-Gargallo et al., 2019), which detected the phylum *Bacillota* as a minor fraction among the 26 different retrieved phyla. To our best knowledge, there are no reference studies on the ecological distribution of *Aquibacillus* other than its presence in table salt, determined by metataxonomic and culturomic approaches (Satari et al., 2021).

The present study reports the isolation and characterization of 32 novel strains closely related to the genus *Aquibacillus*, within the family *Bacillaceae*. In order to determine their taxonomic position, an in-depth phylogenomic analysis of three selected strains was carried out along with supporting phylogenetic, chemotaxonomic, and phenotypic comparative studies. Additionally, we investigated the functional annotation of the genomes to perceive similarities and differences between the isolates and several genera of the family *Bacillaceae*, in particular, *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*. Besides, we further dug into the genome sequences to unveil possible mechanisms of adaptation of the isolates to the extreme habitat where they inhabit (i.e., osmoregulation and heavy metal resistance strategies).

2. Materials and methods

2.1. Odiel Saltmarshes Natural Area sampling and isolation of strains

Sampling on the hypersaline soils located at the saltmarshes of the Odiel Natural Area, in Huelva, Southwest Spain (37°12′26.6″N 6°57′52.5″W), was carried out in Whirl-Pak bags as indicated by Vera-Gargallo et al. (2019). The pH, electrical conductivity, and the concentration of arsenic, cadmium, copper, lead, and zinc were measured as described by Galisteo et al. (2023). Dilution-plating technique was used for the isolation of the strains on R2A medium supplemented with 7.5% (w/v) salts, after 3 months of incubation at 28°C, and then the colonies were subcultured on the same medium until pure cultures were obtained. The composition of the R2A medium is ($g\,L^{-1}$): yeast extract, 0.5; proteose peptone no. 3, 0.5; casamino acids, 0.5; dextrose, 0.5; starch, 0.5; sodium pyruvate, 0.3; K_2HPO_4, 0.3; $MgSO_4$, 0.05. This medium was supplemented with a concentrated seawater (SW) stock diluted to a final salt concentration of 7.5% (w/v), and the pH was adjusted to 7.5. For solid medium, commercial R2A agar (Difco) was prepared with the aforementioned pH and salt concentration and supplemented with 2.0% (w/v) agar. The composition of the SW stock was the following: ($g\,L^{-1}$): NaCl, 234.0; $MgCl_2{\cdot}6H_2O$, 39.0; $MgSO_4{\cdot}7H_2O$, 61.0; $CaCl_2$, 1.0; KCl, 6.0; $NaHCO_3$, 0.2; NaBr, 0.7. For long-term preservation, the liquid culture was mixed with 40% (v/v) glycerol and stored at -80°C. Besides, Marine Agar (MA, Difco 2216) supplemented with 6% (w/v) NaCl was prepared for better comparative purposes of the fatty acid composition with species of the genera *Aquibacillus* and *Radiobacillus*. The following type strains of the genus *Aquibacillus* were obtained from culture collections and used as reference strains for phenotypic comparative analysis: *A. albus* JCM 17364^T, *A. koreensis* JCM 12387^T, and *A. salifodinae* JCM 19761^T. The same medium and conditions stated above were used for their routinary growth. The genomic material of strain *A. koreensis* JCM 12387^T was extracted, purified, and sequenced, as explained below, for phylogenomic comparative purposes.

2.2. Phylogenetic analyses

The method of Marmur (1961) modified for small volumes was used for genomic DNA extraction of the isolates. The universal primers used for 16S rRNA gene amplification were 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R (5′-GGT TAC CTT GTT ACG ACT T-3′) (Lane, 1991). The PCR product was sequenced using Sanger methodology by StabVida (Caparica, Portugal). Library preparation of genomic material from strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 was performed using Novogene NGS DNA Library Prep Set (Cat. No. PT004), followed by whole shotgun sequencing of the genomes on an Illumina NovaSeq PE150 platform by Novogene Europe (Cambridge, United Kingdom). The same protocol was carried out for *A. koreensis* JCM 12387^T, whose genome was not previously available.

Identification of the new isolates was achieved by comparing their partial or almost complete 16S rRNA gene sequences against the EzBioCloud database for prokaryotes[1] (Yoon et al., 2017). The identity shared among the strains isolated in this study was calculated by BLASTn v2.2.28+.[2] For phylogenetic tree reconstructions, the 16S rRNA gene sequences from the closest related species to the new

1 https://www.ezbiocloud.net

2 https://blast.ncbi.nlm.nih.gov

strains were obtained from SILVA (Quast et al., 2013) and GenBank databases (Clark et al., 2016). The fast aligner tool integrated in the ARB package (Ludwig et al., 2004) was employed to align the sequences at the primary and secondary structure level. Maximum-likelihood (Felsenstein, 1981), maximum-parsimony (Felsenstein, 1983), and neighbor-joining (Saitou and Nei, 1987) algorithms, implemented in the ARB package software (Ludwig et al., 2004), were used for tree inferences, and the Jukes-Cantor was selected as the nucleotide substitution model (Jukes and Cantor, 1969) to correct the distance matrix. Bootstrap analysis with 1,000 pseudoreplicates was carried out in order to validate the robustness of the branches. The script "gitana"[3] performed the visual editing of the tree.

2.3. Comparative genome analyses and ecological distribution

SPAdes v3.15.2 (Prjibelski et al., 2020) was utilized to assemble the quality filtered paired-end reads (options "--careful -k 21, 33, 55, 77, 99, 127"). Contigs shorter than 500 bp or SPAdes coverage below 20 were removed. QUAST v2.3 (Gurevich et al., 2013) allowed us to calculate the assembly statistics and CheckM v1.0.5 (Parks et al., 2015) to evaluate the completeness and contamination of the assembled genomes. In order to sort the contigs of the draft genomes, they were aligned against the closest related strain with available complete genome, i.e., *Radiobacillus deserti* TKL69^T, by using nucmer, integrated in MUMmer v4.0.0rc1 compilation of utilities and scripts (Marçais et al., 2018). Coding sequences (CDS) were extracted with Prodigal v2.60 (Hyatt et al., 2010) and annotated with Prokka v1.12 (Seemann, 2014) to generate the standard GenBank files. The online tool BlastKOALA (Kanehisa et al., 2016) was employed to perform a detailed functional annotation of the predicted translated CDS, by assigning KEGG Orthology (KO) identifiers and KEGG pathways. The "iep" program from EMBOSS package v6.5.7.0 (Rice et al., 2000) was utilized to determine the isoelectric point of the predicted proteins.

In-depth placement of the three sequenced isolates (strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286) within the family *Bacillaceae* was carried out by phylogenomic reconstruction based on the concatenation of the translated single-copy core genes from 79 members of this family whose genome sequence was available in RefSeq database. BLASTp v2.2.28+ and Markov Cluster Algorithm, implemented in the Enveomics toolbox (Rodriguez-R and Konstantinidis, 2016), were used to search and to extract the translated orthologous genes, which were further aligned with Muscle v3.8.31 (Edgar, 2004). FastTreeMP v2.1.8 (Price et al., 2010) was employed to infer the approximately maximum-likelihood phylogeny based on 739 concatenated protein sequences, considering the Jones-Taylor-Thornton model of amino acid evolution (Jones et al., 1992). The robustness of the obtained nodes was checked by the Shimodaira-Hasegawa test (Shimodaira and Hasegawa, 1999). Tree image was edited and visualized with the script "gitana" (see text footnote 3). "UpSetR" v1.4.0 package for R (Conway et al., 2017) allowed us to visualize the intersection of the 18,524 orthologous genes identified after BLASTp search of their translated sequences. The proposed minimal standards for prokaryotic taxonomy (Chun et al., 2018) suggest the use of Overall Genome Relatedness Indexes (OGRIs) for a reliable determination of the taxonomic status of new taxa, such as the digital DNA–DNA hybridization (dDDH), the Average Amino acid Identity (AAI), and the Average Nucleotide Identity for orthologous sequences (orthoANI). The Genome-to-Genome Distance Calculator (GGDC v3.0) from the Leibniz Institute DSMZ (Meier-Kolthoff et al., 2021) was utilized to obtain the dDDH relatedness values, whereas the Enveomics toolbox (Rodriguez-R and Konstantinidis, 2016) and OAU software v1.2 (Lee et al., 2016) were selected for AAI and orthoANI calculations, respectively.

Metagenomic dataset SMO1 (Supplementary Table S1) from a hypersaline soil of the Odiel Saltmarshes Natural Area (Vera-Gargallo et al., 2018) was selected for the screening of functional genes in the environment under study. Raw reads with length ≥ 30 bp were assembled with Megahit v1.2.9 (Li et al., 2015, 2016). Contigs were examined to extract translated CDS by Prodigal v2.60 (Hyatt et al., 2010), and KO identifiers were assigned to them by GhostKOALA (Kanehisa et al., 2016). Then, functions of interest were manually selected.

The ecological distribution of the new strains in hypersaline environments was determined by fragment recruitment analysis against 16 environmental metagenomic datasets (Supplementary Table S1). The 16S rRNA gene sequences from the genomes were masked due to their highly conserved nature. Metagenomic reads above ≥30 bp were BLASTn v2.2.28+ searched, independently, against each genome. BLASTn results with identity values <95%, alignment length < 50 bp, and e-value >10^{-5} were filtered out, as recommended by Mehrshad et al. (2018). In order to normalize the relative abundance values, we computed the RPKG (reads recruited per kilobase of genome per gigabase of metagenome), proposed by Nayfach and Pollard (2015). Besides, the genomes of *Haloquadratum walsbyi* C23^T (GCF_000237865.1), *Salinibacter ruber* DSM 13855^T (GCF_000013045.1), and *Spiribacter salinus* M19-40^T (GCF_000319575.2) were included in the analysis as references for comparison.

Plots generated in this study were created with the following R packages: "aplot" v0.1.8 (Guangchuang et al., 2022), "gghighlight" v0.3.2 (Yutani, 2021), "ggplot2" v3.3.3 (Wickham, 2009), "ggpubr" v0.4.0 (Kassambara, 2020), "ggtext" v0.1.2 (Wilke and Wiernik, 2022), "gridExtra" v2.3 (Auguie, 2017), and "paletteer" v1.4.0 (Hvitfeldt, 2021). R packages "phytools" v1.2-0 (Revell, 2012), "reshape2" v1.4.4 (Wickham, 2007), "scale" v.1.1.1 (Wickham and Seidel, 2020), and "seqinr" v4.2-16 (Charif and Lobry, 2007) were requested to reformat input data. DNAplotter application (Carver et al., 2009) was used to generate the circular representation of the genomes.

2.4. Fatty acids composition and phenotypic features

The fatty acid profile of the type strains of the two proposed species, 3ASR75-54^T and 3ASR75-11^T, was determined by gas chromatography with an Agilent 6850 system at the Spanish Type Culture Collection (CECT), Valencia, Spain. For that purpose, strains 3ASR75-54^T and 3ASR75-11^T were grown in MA medium supplemented with 6% (w/v) NaCl at 35°C for 3 days. TSBA6 library (MIDI, 2008) was used for the determination of fatty acids following the protocol suggested by MIDI Microbial Identification System (Sasser, 1990).

3 https://github.com/cristinagalisteo/gitana

Colonial morphology and pigmentation of strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 were observed after 3 days of growth on R2A medium supplemented with 7.5% (w/v) salts and the pH adjusted to 7.5, at 37°C. Cell morphology and motility were determined by phase contrast microscopy (Olympus CX41). To detect their ability to grow anaerobically, strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 were incubated using the AnaeroGen™ system (Oxoid) under the aforementioned conditions. Optical density measures allowed us to determine the range and optimum salt concentration and pH values supporting growth for type strains 3ASR75-54^T and 3ASR75-11^T. Infinite M Nano microplate reader (Tecan, Grödig, Austria) adjusted at 37°C, with linear shaking, was utilized to measure the absorbance at 600 nm every 2 h for 3 days. R2A broth was supplemented with SW stock to obtain a final salt concentration of 0.5, 2, 4, 5, 6, 7, 7.5, 8, 9, 10, 12, 15, 17, 20, 22, and 25% (w/v) in order to determine the salinity range and optimum. R2A liquid medium supplemented with optimum salt concentration was also employed to test growth at pH values of 3.0, 4.0, 5.0, 6.0, 7.0, 7.5, 8.0, 9.0, and 10.0, using a buffered system to maintain pH conditions (Sánchez-Porro et al., 2009). The temperature range for growth was measured in R2A broth adjusted to the optimal salinity and pH, and incubated at 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 28, 37, 40, 42, 43, 44, 45, 46, and 48°C, and the optical density was assessed in a Spectronic 20D+ (ThermoSpectronics, Cambridge, United Kingdom).

The proposed minimal standards for describing new taxa of aerobic endospore-forming bacteria (Logan et al., 2009) were followed for the phenotypic characterization of strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286. R2A medium supplemented with 7.5% (w/v) salts, and pH adjusted at 7.5 was used for routine growth with an incubation period of 3 days at 37°C. The same methodology was used for reference strains *A. albus* JCM 17364^T, *A. koreensis* JCM 12387^T, and *A. salifodinae* JCM 19761^T. These incubation conditions were used for all the biochemical tests. The protocols for determination of catalase, hydrolysis of gelatin, starch, Tween 80, DNA, casein, and aesculin, production of indole, methyl red and Voges–Proskauer tests, Simmons' citrate, nitrate and nitrite reduction, H_2S production, urease, and phenylalanine deaminase are described by Cowan and Steel (1965). A drop of 1% (v/v) tetramethyl-p-phenylenediamine (Kovacs, 1956) was employed to test the oxidase activity in young cultures. Modified phenol red base medium with 0.05% (w/v) yeast extract and supplemented with 7.5% (w/v) salts allowed the determination of acid production. Carbohydrates were filter-sterilized and added to a final concentration of 1% (Cowan and Steel, 1965; Ventosa et al., 1982). In order to test the use of a wide variety of substrates as sole carbon and energy source, or as sole carbon, nitrogen and energy sources, strains were inoculated in the medium described by Koser (1923), as modified by Ventosa et al. (1982). Amino acids, alcohols, and organic acids were supplied to give a final concentration of 1 g L^{-1}, and carbohydrates of 2 g L^{-1}. All the substrates were added after filter-sterilization.

3. Results and discussion

3.1. Saline soils sampled from Odiel Saltmarshes Natural Area are heavily contaminated

The heavy metals concentration of the Odiel river waters and sediments have been previously studied due to the past industrial and mining activities in its surroundings, showing high concentrations of arsenic, cadmium, copper, lead, and zinc (Sainz et al., 2002, 2004). The Government of the region of Andalucía, where the area under study is located, sets the following reference criteria for noncontaminated soils (mg kg^{-1}): arsenic, 2–5; cadmium, 0.4–0.8; copper, 17–100; lead, 10–50; and zinc, 10–70 (Consejería de Medio Ambiente, 1999). The soils of the Odiel Saltmarshes Natural Area studied here presented values substantially above those ranges (mg kg^{-1}): arsenic, 124.3; cadmium, 2.0; copper, 1,853.0; lead, 257.5; and zinc, 443.8, which indicate heavy metal contamination in the sampled soils. This sampling area represents the most contaminated region among the hypersaline soils in the Odiel Saltmarshes Natural Area studied so far (Vera-Gargallo et al., 2019; Galisteo et al., 2023), probably related to its close location to the mouth of the Canal del Burro Grande into the Odiel river. The pH of the sample was 7.04 whereas the electrical conductivity (EC) was 18.49 mS cm^{-1} at 25°C, which is above the 4 mS cm^{-1} at 25°C threshold for defining saline soils (Richards, 1954).

3.2. Isolated strains can be split into two groups

In a previous study, more than 4,000 strains were isolated in an extensive screening carried out in the hypersaline soils located in the Odiel Saltmarshes Natural Area (Huelva, Southwest Spain) (Galisteo et al., 2023). Out of them, 32 strains showed a close relationship with the genus *Aquibacillus,* according to their partial or almost complete 16S rRNA gene sequence comparison. All the isolates presented a percentage of identity below the 98.65% cutoff for species delineation (Figure 1). Their top hits were either *A. koreensis* BH30097^T (97.84–95.51%) for 24 strains or *A. albus* YIM 93624^T (97.63–95.51%) for the other eight strains, which seem to indicate that they might be clustered into two groups, denoted as group 1 and group 2. The identity values among themselves exhibited the same pattern for clustering (Figure 2). Within the groups, the identity varied from 100 to 99%, except for the strains 3ASR75-118 and 3ASR75-2, sharing 96.45–96.39% identity with strains of the group 1. In any case, these two strains exhibited higher identity values with respect to members of the group 1 than to members of the group 2, and thus, they both were initially affiliated to the *A. koreensis*-like group 1. Between clusters, the percentage of identity dropped below 97%, indicating that they may constitute two different species. Strains 3ASR75-54^T and 3ASR75-11^T were selected as type strain of groups 1 and 2, respectively, as they grew well under laboratory conditions and their 16S rRNA genes were sequenced at high quality and long length (1,432 bp and 1,479 bp, respectively).

The phylogenetic tree based on the 16S rRNA gene sequences (Figure 3) provides an enhanced view of the proposed clusters, considering the most discriminative power of phylogenetic methods over identity matrixes. Members of the genus *Aquibacillus* and of the closely related genera *Amphibacillus*, *Radiobacillus*, and *Sediminibacillus,* as well as some representative species of the genus *Virgibacillus*, were included in the phylogenetic analysis. The 24 strains most closely related to *A. koreensis* BH30097^T (group 1) clustered together once again, including strains 3ASR75-118 and 3ASR75-2. The eight strains that showed their highest 16S rRNA sequence identity with *A. albus* YIM 93624^T also conformed a clear single cluster (group 2). In this latter case, the closest neighbor was not *A. albus* YIM 93624^T as expected, but *A. sediminis* BH258^T. In addition, the branch supporting group 2

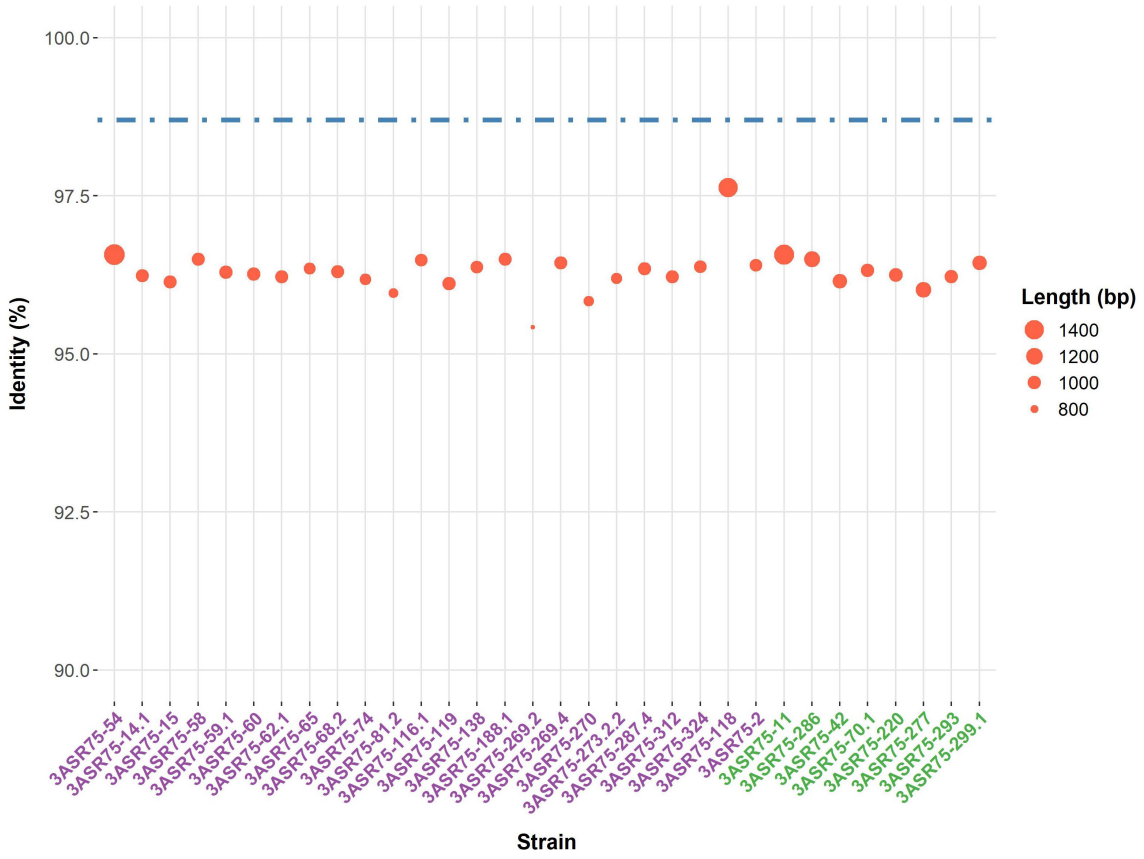


FIGURE 1
Top hit identity values (%) of the 32 isolates against EzBioCloud database. Best hit was either the species *Aquibacillus koreensis* BH30097^T (strain names purple-colored) or *Aquibacillus albus* YIM 93624^T (strain names green-colored). Dot size is proportional to the length of the sequenced 16S rRNA gene. Dashed line indicates the 98.7% identity cutoff for species delineation.

displayed a close relationship with species of the genus *Amphibacillus*, turning the genus into polyphyletic. Further incongruences were observed in the tree reconstruction, such as the clustering of the only member of the genus *Radiobacillus, R. deserti,* together with four species of the genus *Aquibacillus,* giving rise to an *Aquibacillus-Radiobacillus* group. Moreover, the species of the genus *Sediminibacillus* formed a monophyletic group neighbor to the *Aquibacillus-Radiobacillus* cluster, constituting a branch separated from other members of the genus *Aquibacillus* and from strains of the groups 1 and 2. Only the genus *Virgibacillus* formed a clearly independent monophyletic branch including all the species within this genus. Clearly, the 16S rRNA gene sequence analysis demonstrated unstable tree topologies since only a few nodes were conserved for all the three tree-constructing algorithms (maximum-likelihood, maximum-parsimony, and neighbor-joining) and bootstrap values were, in general, below 70%. Further analyses considering the whole genome sequences are indispensable to elucidate the correct taxonomic position of the new isolates.

3.3. Comparative genomic analyses shed light on the taxonomic status of the new isolates

The draft genome sequence of the selected type strains 3ASR75-54^T (GCF_028416595.1) and 3ASR75-11^T (GCF_028416575.1) were *de novo* obtained. Moreover, we sequenced an additional reference strain from group 2 (i.e., strain 3ASR75-286, GCF_028416555.1), given the placement of this group next to the genus *Amphibacillus*, which made us suspicious of group 2 forming a new separated genus. The genomes of the mentioned strains were assembled in 70, 71, and 67 contigs, respectively. Their total genome size and G+C content ranged 3.59–3.70 Mb and 38.0–38.1 mol%, although the first parameter was slightly higher for strain 3ASR75-54^T (Supplementary Table S2). Comparisons with closely related genera showed that genome size (Figure 4A) and G+C content (Figure 4B) of the new strains were more similar to those of the genus *Radiobacillus*. On the contrary, the species of the genus *Aquibacillus* possessed larger genomes, 4.22–4.41 Mb, and a lower G+C content, 35.7–37.4 mol%, which reinforces the idea of the new isolates not belonging to the previously described species of this genus (Supplementary Table S2; Figures 4A,B). The genome of strain 3ASR75-54^T encoded 3,535 CDS, 100 tRNA, and 15 rRNA, whereas the genomes of strains 3ASR75-11^T and 3ASR75-286 encoded 3,590 and 3,617 CDS, respectively, and harbored less RNA sequences than strain 3ASR75-54^T (66 tRNA and 9 rRNA) (Supplementary Table S2). The DNA of the reference species *A. koreeensis* JCM 12387^T was likewise sequenced (GCF_028416535.1) as it was not available at the beginning of this study. This genome sequence was assembled into 56 contigs with a total size of 4.33 Mb and a G+C content of 36.8 mol% (Supplementary Table S2), in agreement with the other six genomes

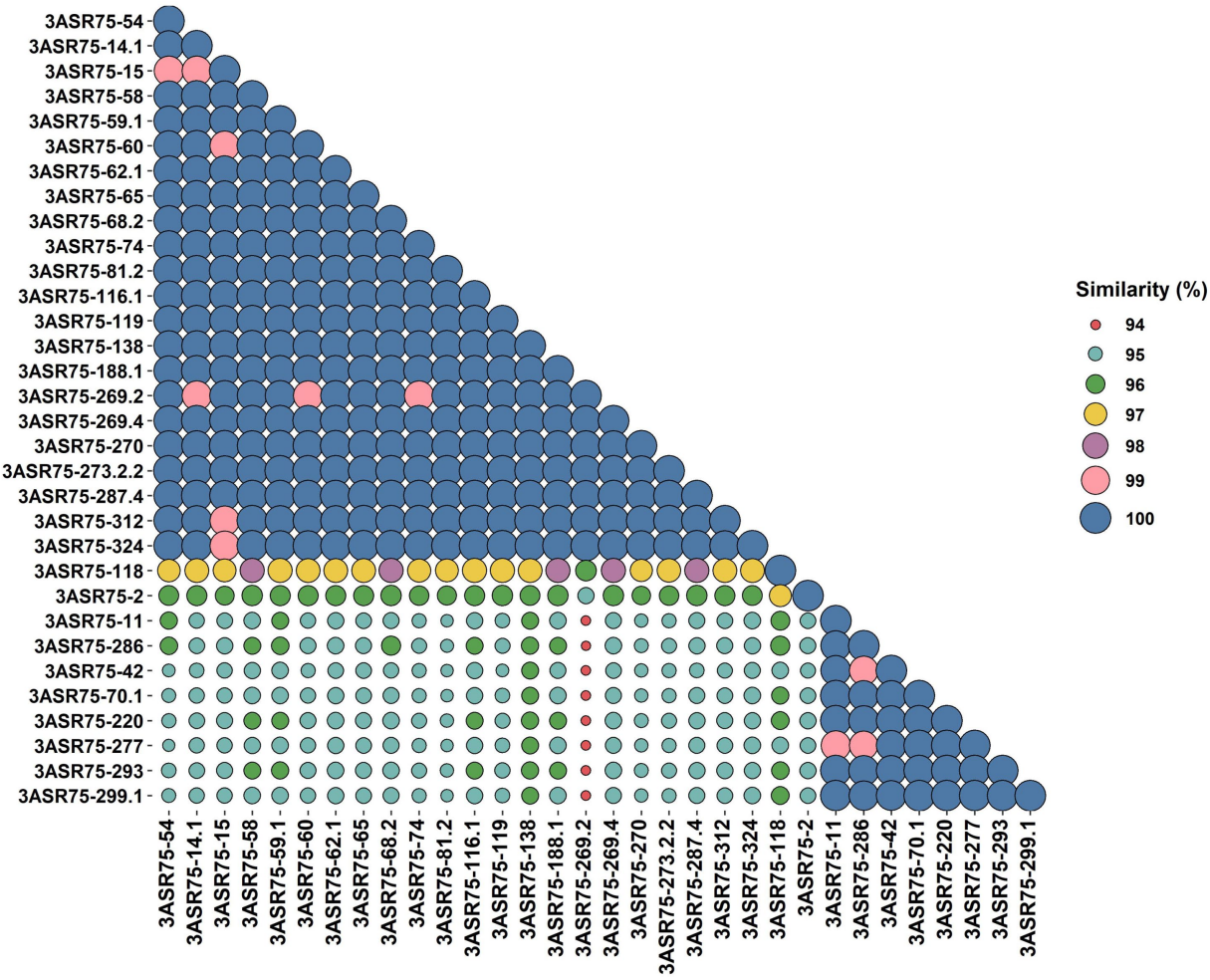


FIGURE 2
BLASTn identity matrix among the partial or almost complete 16S rRNA gene sequences of the 32 strains isolated in this study.

of type strains available for the genus *Aquibacillus*. Further genomic features are shown in Supplementary Table S2.

The genus *Aquibacillus* belongs to a large family, *Bacillaceae*, along with other 116 genera with validly published names (Parte et al., 2020; last consulted on 10/02/2023). Considering the weak robustness of the 16S rRNA gene-based phylogenetic tree as stated above, a more reliable phylogenomic tree was constructed based on a large set of genomes from the genus *Aquibacillus* and neighbor genera. The 739 protein-based approximately maximum-likelihood inference included a total of 76 species from nine genera, as well as the three new strains and *A. koreensis* JCM 12387^T, whose genomes were sequenced in this study (Figure 5). Unlike the single 16S rRNA gene phylogeny (Figure 3), most branches were now supported by a 100% bootstrap value, providing a consistent topology to elucidate the evolutionary relationship between the new isolates and the closely related taxa. All the described species of the genus *Aquibacillus* clustered together in a monophyletic group, including the recently sequenced genomes of *A. koreensis* JCM 12387^T and the new strain 3ASR75-54^T. Phylogenomic analysis displayed the closest relationship of strain 3ASR75-54^T with *A. albus*, even if its 16S rRNA gene sequence showed a higher percentage of identity with *A. koreensis* (Figures 1, 3). On the other hand, strains 3ASR75-11^T and 3ASR75-286, whose BLASTn top hit was *A. albus* (Figure 1) and the closest neighbor according to the 16S rRNA gene sequence phylogeny was *A. sediminis* (Figure 3), now constitute a single branch related to the single species of the genus *Radiobacillus*. A first glimpse might affiliate strains 3ASR75-11^T and 3ASR75-286 with a novel species of the genus *Radiobacillus*. However, the length of branch connecting both strains to the node shared with *R. deserti* is sufficiently large to consider strains 3ASR75-11^T and 3ASR75-286 as members of a different genus. Further analyses are needed in order to confirm this hypothesis.

Concerning OGRIs analyses, dDDH and orthoANI values were estimated between the three new isolates and the species of the genera *Aquibacillus*, *Amphibacillus*, *Radiobacillus*, and *Sediminibacillus*, which are the closest related genera within the family *Bacillaceae* (Figure 6A). The highest dDDH percentage obtained was 24.2%, which is far below the 70% cutoff for species delineation (Stackebrandt and Goebel, 1994; Auch et al., 2010). Similarly, orthoANI results were all equal or lower than 72%, again lower that the 95% threshold established for species differentiation (Goris et al., 2007; Richter and Rosselló-Móra, 2009; Chun and Rainey, 2014). Nevertheless, the outcome between strains 3ASR75-11^T and 3ASR75-286 exceeded both limits, with values of 96.5 and 100% for dDDH and orthoANI, respectively. Thus, we can conclude that our isolates constitute two

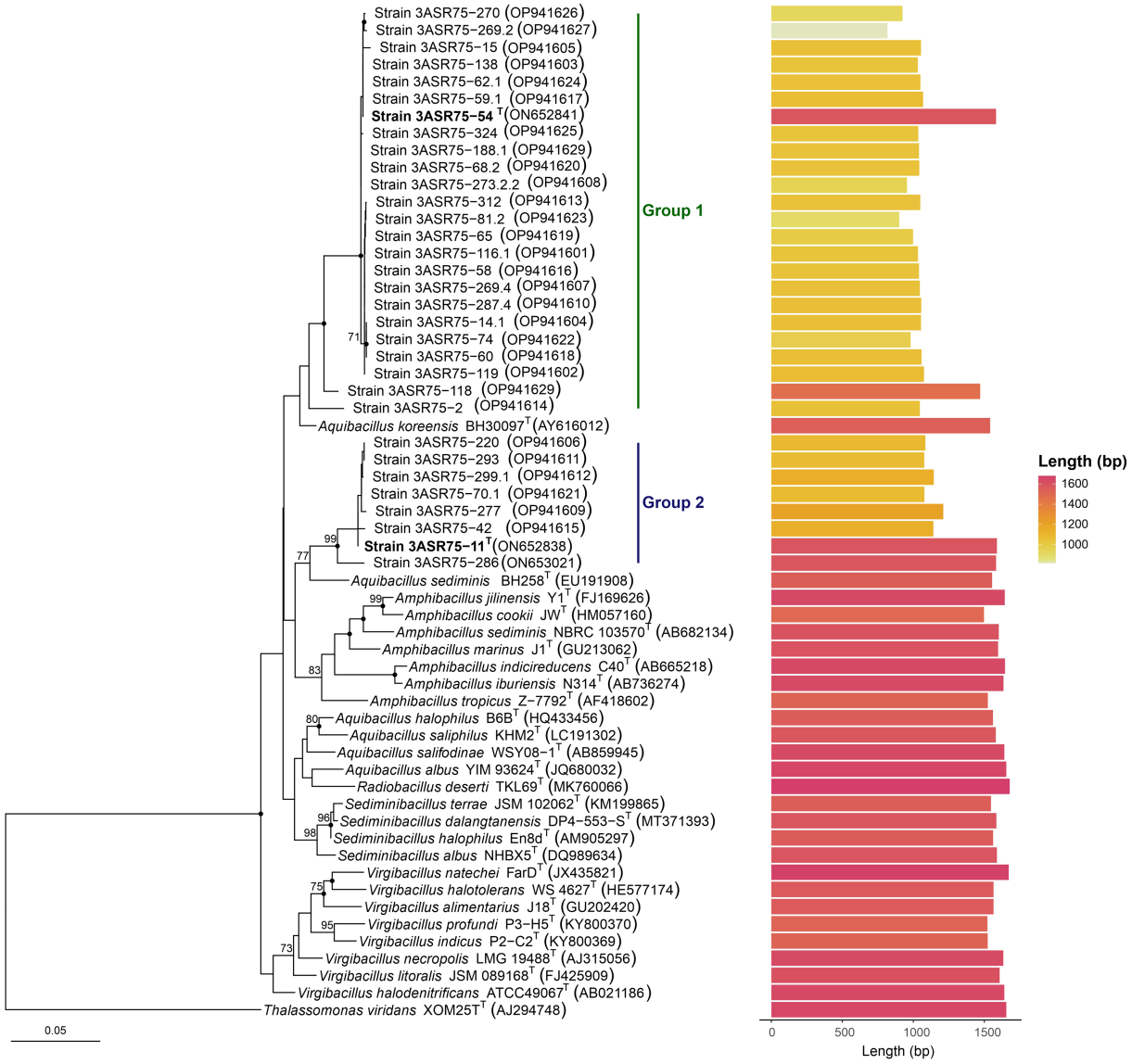


FIGURE 3
Neighbor-joining phylogenetic tree based on the comparison of the 16S rRNA gene sequences showing the relationships among the new strains and species of the closely related genera *Amphibacillus*, *Radiobacillus*, *Sediminibacillus*, and *Virgibacillus*. Bootstrap values ≥70%, based on 1,000 pseudoreplicates, are indicated above branches. Nodes conserved across the three tree-constructing methods are marked with a filled circle. The species *Thalassomonas viridans* was used as an outgroup. Bar, 0.05 substitutions per nucleotide position.

novel species, one represented by strain 3ASR75-54^T and the other comprising strains 3ASR75-11^T and 3ASR75-286.

Another widely used OGRI is AAI, which considers that pairs of genomes with values lower than 72–65% belong to species of different genera (Konstantinidis and Tiedje, 2007; Konstantinidis et al., 2017). In particular, we can observe that species within the genus *Aquibacillus* shared AAI percentages between 67.6 and 69.7% (Figure 6B). The new species represented by strain 3ASR75-54^T exhibited an AAI range of 71.3-67.1% with species of *Aquibacillus*, while the values were fairly distant with other closely related genera (the highest being 66.2% with *Sediminibacillus albus*). These data, along with the robust topology of the phylogenomic tree, clearly indicate that the strain 3ASR75-54^T belongs to a non-yet described species of the genus *Aquibacillus*. On the other hand, the AAI values of strains 3ASR75-11^T and 3ASR75-286 with respect to the species of the genus *Aquibacillus* varied between 65.9–64.2%, lower than the current intrageneric range for *Aquibacillus* and in the lower bound or below the accepted 72–65% cutoff for genus delineation. The AAI comparisons between strains 3ASR75-11^T and 3ASR75-286 and the other closely related genera within the *Bacillaceae* showed the highest values for the genus *Sediminibacillus* (66.6–65.8%), which might suggest their affiliation to this genus. However, the phylogenomic tree (Figure 5) allows us to discard this thesis considering their polyphyly. As stated before, genomic-based inference groups strains 3ASR75-11^T and 3ASR75-286 with *R. deserti*, supporting their placement into the genus *Radiobacillus*. This assumption is not well supported because AAI values between the two taxa (64.9–64.8%) contravenes the AAI threshold for genus delineation. Nevertheless, it must be noted that only one species of the genus *Radiobacillus* is described to date, turning impossible to

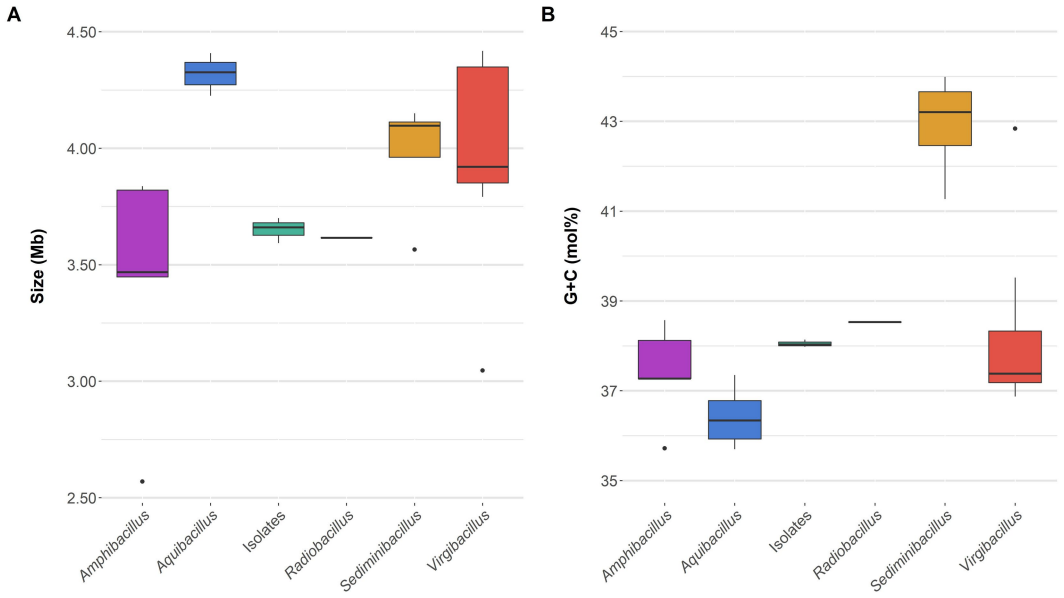


FIGURE 4
Genome size **(A)** and G+C content **(B)** boxplots of the genome sequences belonging to the family *Bacillaceae* included into this study. The three new isolates exhibited a very similar value among them and with respect to the single species of the genus *Radiobacillus*.

predict if future descriptions of *Radiobacillus* species will entail the rise in the upper bound AAI range, making feasible the grouping of strains 3ASR75-11^T and 3ASR75-286 as members of the genus *Radiobacillus*. Therefore, we do not have a sound scientific evidence to assert whether those two strains are part of the genus *Radiobacillus* or, on the contrary, they constitute a novel genus within the family *Bacillaceae*. In order to elucidate this issue, we plotted the AAI-orthoANI pairs of values for the studied genomes within and between genera (Figure 7). When strains 3ASR75-11^T and 3ASR75-286 were considered as a separated genus, the inter-and intra-genus results do not overlap (Figure 7A). However, the clustering of strains 3ASR75-11^T and 3ASR75-286 with the genus *Radiobacillus* gave rise to an intra-genus spot located within the inter-genus point cloud (Figure 7B). Consequently, our results suggest that the species constituted by strains 3ASR75-11^T and 3ASR75-286 does not belong to any of the currently described genera within the *Bacillaceae* and should be accommodated in a novel genus.

The genomes of strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, together to those of the type strains of the species of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus* shared a total of 1,451 core genes. Part of the accessory genome is also common for some of the studied species; however, a great number of strain-specific genes were detected. Excluding the core genome, the larger cluster of genes (458) was that shared by the isolated strains 3ASR75-11^T and 3ASR75-286. The former harbored 153 strain-exclusive genes while the latter 112. Out of our three isolates, strain 3ASR75-54^T was the one containing the higher number of singletons (197). Furthermore, those three strains shared more orthologous genes between them (82) than with any of the other genomes under study. Remarkably, strains 3ASR75-11^T and 3ASR75-286 did not display a significant number of common genes with any particular genus, even with their closest relative, the genus *Radiobacillus* (Figure 8). A further insight about the functions encoded in the accessory genome is described in section 3.5. Besides, the observed variability between species and between strains of the same species may indicate their specialization either for adaptation to a specific habitat or to carry out an ecological role in the community.

3.4. Chemotaxonomic and phenotypic analyses support the new taxa descriptions as members of the family *Bacillaceae*

Strain 3ASR75-54^T exhibited a chemotaxonomic profile similar to that of the other members of the genus *Aquibacillus* (Supplementary Table S3), with anteiso-$C_{15:0}$ as the most abundant fatty acid (47.9%), followed by iso-$C_{15:0}$ (11.6%), and anteiso-$C_{17:0}$ (10.7%). Strain 3ASR75-11^T showed a high predominance of anteiso-$C_{15:0}$ (66.4%), whereas other minor fatty acids present were anteiso-$C_{17:0}$ (9.9%) and iso-$C_{16:0}$ (6.8%). This fatty acid composition of strain 3ASR75-11^T was comparable to that of the species of the closely related genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus* (Lee et al., 2006; Carrasco et al., 2008; Zhang et al., 2012, 2015; Amoozegar et al., 2014; Lee and Whang, 2019; Li et al., 2020; Wang et al., 2021; Cho and Whang, 2022).

Major morphological and physiological characteristics of the new strains (Table 1) were consistent with those described for members of the family *Bacillaceae*. Cells were rod-shaped, endospore-forming, and motile. Colonies were circular and white pigmented. Strain 3ASR75-54^T displayed an optimal NaCl concentration supporting growth similar to other members of the genus *Aquibacillus*. On the other hand, strain 3ASR75-11^T grew optimally at lower NaCl concentrations than species of the genera *Aquibacillus* and *Radiobacillus*. Further biochemical characteristics are detailed in the

10.3389/fmicb.2023.1192059

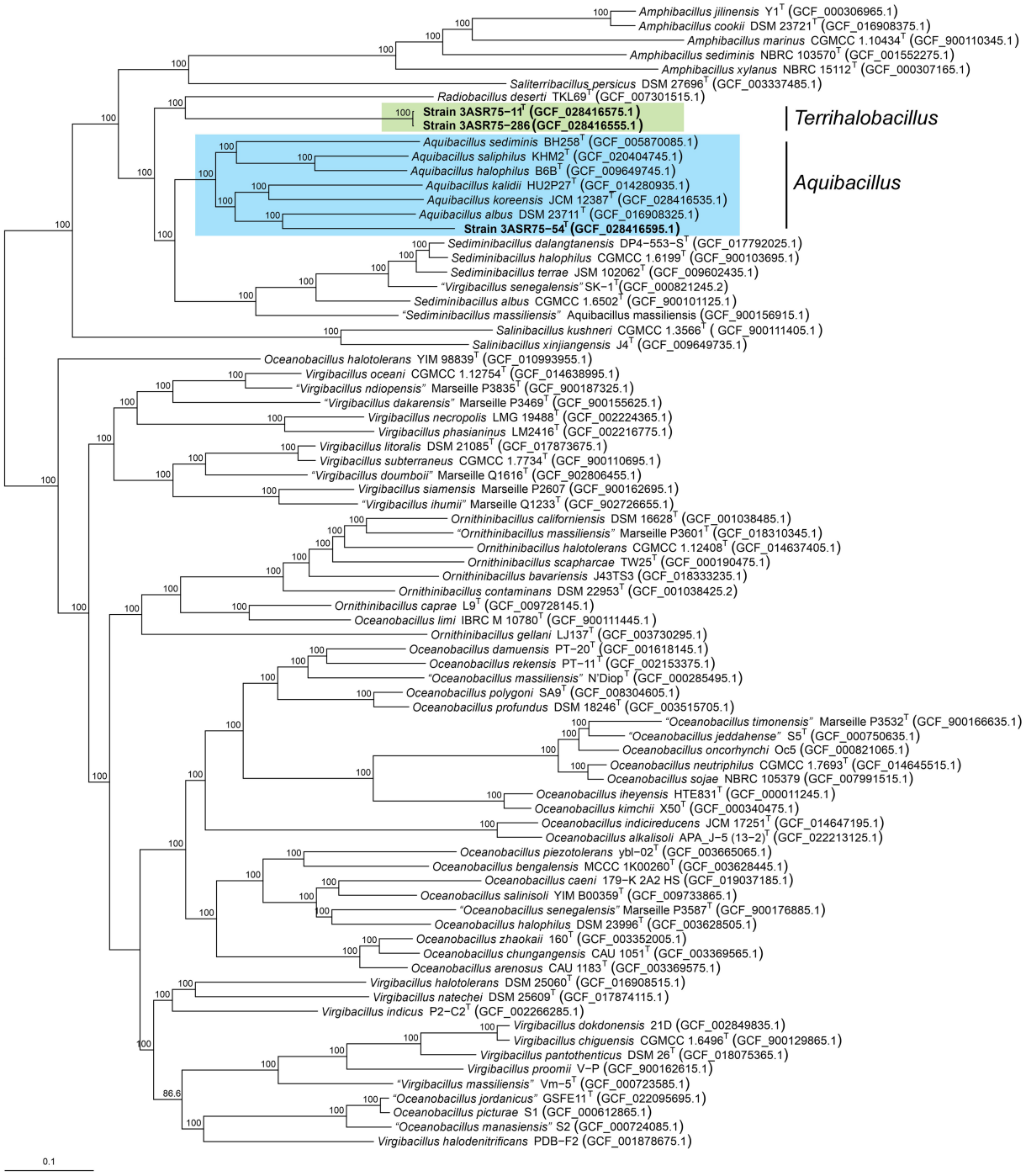


FIGURE 5
Approximately maximum-likelihood phylogenomic tree based on 739 concatenated core protein sequences showing the relationships between the novel isolates and 79 closely related species of the family *Bacillaceae*. Bootstrap values ≥70% are indicated above the respective branch. Bar, 0.1 substitutions per nucleotide position.

new species descriptions and in Supplementary Table S4. The phenotypic similarities found between the strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 and the studied members of the family *Bacillaceae* reinforce the proposal for their placement within this family. At the same time, features such as motility and optimal NaCl concentration for growth allow their differentiation from the closely related species.

3.5. Functional genomic analysis correlates with the phenotype and reveals the existence of a variety of cell membrane transporters

A total of 1,584 KO were identified for strain 3ASR75-54^T. Out of them, 21 were exclusive of this strain, that is, not present in the genome

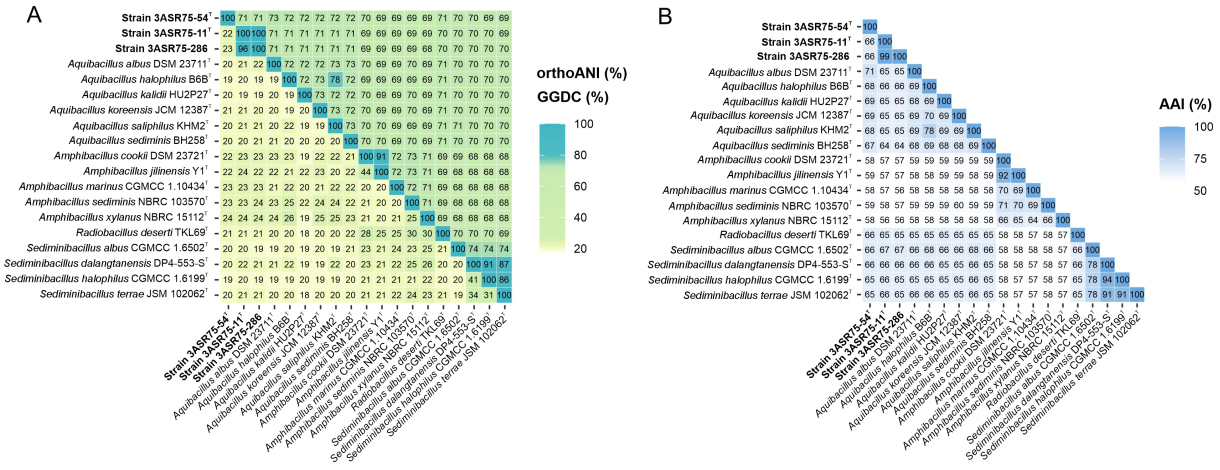


FIGURE 6
OrthoANI (upper triangle)/GGDC (lower triangle) **(A)** and AAI **(B)** values (%) among strains 3ASR75-54^T, 3ASR75-11^T, 3ASR75-286, and the closest related members of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*. OrthoANI/GGDC results confirmed that strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 cannot be affiliated to any of the currently described species of these genera and also proved that strains 3ASR75-11^T and 3ASR75-286 constituted a single species. AAI data unequivocally assigned strain 3ASR75-54^T to the genus *Aquibacillus* and suggested the placement of strains 3ASR75-11^T and 3ASR75-286 into a new genus within the family *Bacillaceae*.

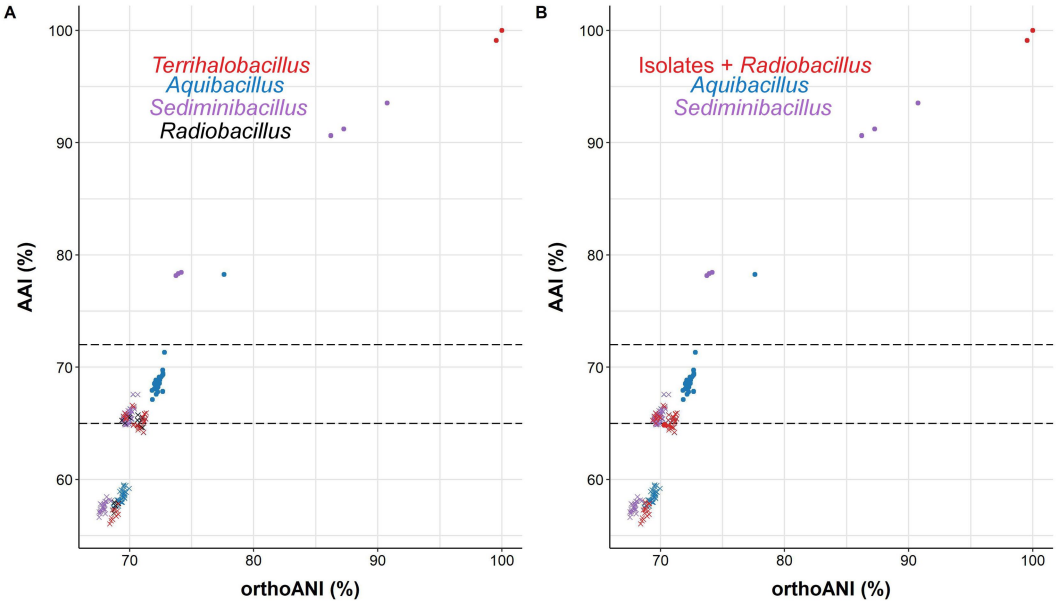


FIGURE 7
Scatter plot displaying AAI-orthoANI pairs of values for the studied genomes within and between genera assuming the placement of strains 3ASR75-11^T and 3ASR75-286 into the new genus *Terrihalobacillus* **(A)** and into the genus *Radiobacillus* **(B)**. Dotted lines denote the 65–72% AAI cutoff for genus delineation. Strain 3ASR75-54^T has been included among the species of the genus *Aquibacillus*. Circles and crosses indicate intra-and inter-genus values, respectively.

of any of the studied species of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*, nor in that of the strains 3ASR75-11^T and 3ASR75-286. Among these functions, strain 3ASR75-54^T exhibited a putative zinc/manganese transport system (K02074, K02075, and K02077); the large (*nirB* gene) and small (*nirD* gene) subunits of the nitrite reductase enzyme turning nitrite into ammonia (K00362 and K00363); and CRISPR-associated protein-coding genes, such as *cas3*, *cas4*, *csd1*, *csd2*, and *cas5d* (K07012, K07464, K19117, K19118, and K19119).

The genomes of strains 3ASR75-11^T and 3ASR75-286 were annotated with 1,604 and 1,573 KO, respectively. Strain 3ASR75-11^T possessed 61 KO that were not present in strain 3ASR75-286, while 3ASR75-286 only presented 30 KO not detected in strain 3ASR75-11^T. Those 61 strain-specific functions encoded mostly saccharide transporters (multisugar, ribose/autoinducer 2/D-xylose, and rhamnose), as well as the ability to transform sorbitol into sorbitol 6-phosphate (K02781, K02782, and K02783) in strain 3ASR75-11^T. On

10.3389/fmicb.2023.1192059

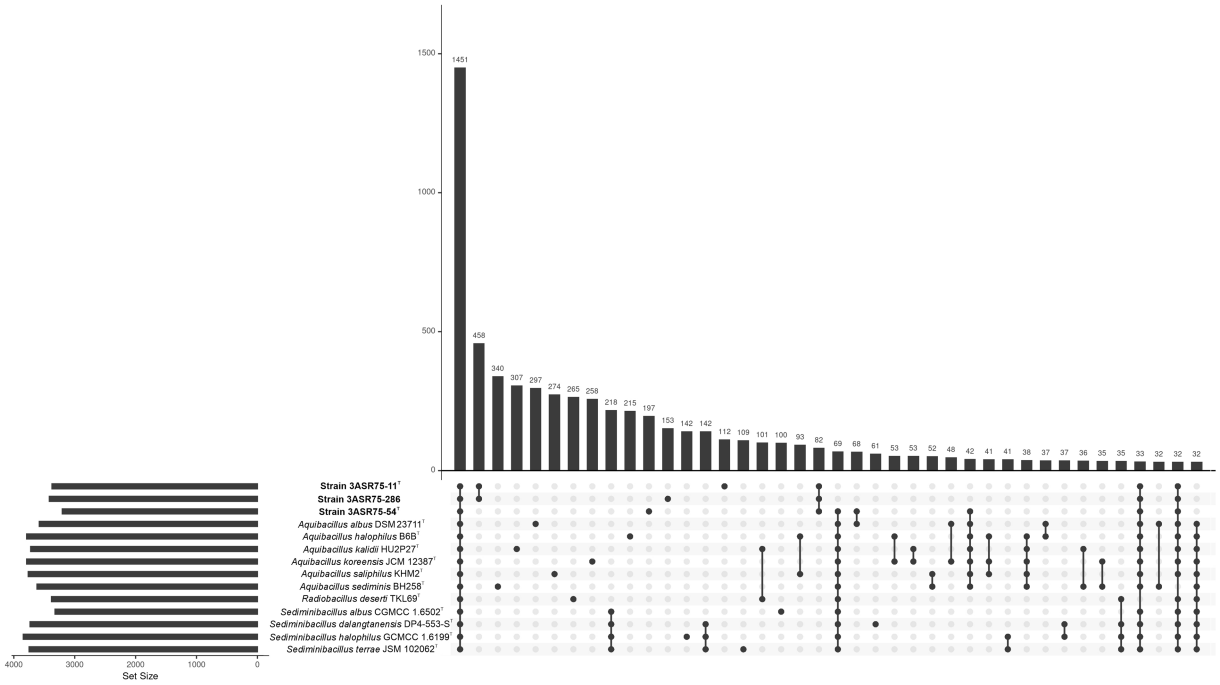


FIGURE 8
Upset plot showing the intersecting orthologous genes of the isolated strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, and the species of the closely related genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*. The core genome encoded a total of 1,451 genes out of a pangenome of 18,524 genes. Isolated strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 harbored an accessory genome with 197, 112, and 153 exclusive genes, respectively. A total of 458 genes were solely present in strains 3ASR75-11^T and 3ASR75-286, while 82 genes where exclusively shared between the three isolates.

the other hand, strain 3ASR75-286 harbored singular metal related functions, such as copper efflux regulator (K19591), cadmium/lead responsive transcriptional repressor (K21885), and iron-siderophore transporter system permease protein (K25111). Besides their differences, both strains displayed 35 common functions that were not present in any other of the studied genomes of the closest related genera. Among them, we could identify transporters, such as those for arginine/lysine/histidine (*artPQM* genes; K17077, K23059, and K23060), chitobiose (*chiEFG* operon; K17244, K17245, and K17246), and zinc and cadmium (ZIPB; K16267), which passively uptakes these ions into the cytoplasm (Lin et al., 2010); and Ca^{2+}/H^+ antiporter (*chaA*; K07300), working as a K^+ extrusion system to maintain K^+ homeostasis under salt stress conditions (Radchenko et al., 2006). Enzyme-coding genes were also identified among the 35 shared KO, such as those for N-methylhydrantoinase (*hyuAB*; K01473 and K01474), glutaconate CoA-transferase subunits A and B (*gctAB*; K01039 and K01040), and ferritin (K02217). Additionally, the genome of strain 3ASR75-11^T was annotated with 10 KO not present in any of the other analyzed genomes, although these functions do not seem to provide any relevant feature. On the other hand, strain 3ASR75-286 harbored two KO related to heavy metal resistance among its 10 exclusive KEGG Orthology identifiers, one being the abovementioned copper efflux regulator (K19591) and the second an alkylmercury lyase (K00221). Further heavy metal tolerance mechanisms are discussed in section 3.8.

All the three strains under study, 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, contained between 39 and 41 KO related to the sporulation process, like the species of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus* (Supplementary Figure S1). This in-silico analysis agrees with the morphological observation revealing the formation of terminal endospores at the poles of cells after incubation at 37°C for 7 days. Additionally, the three strains harbored the response regulator for oxygen limitation (K07651, K07775, and K02259), which correlates with their anaerobic growth in laboratory conditions (Table 1). Moreover, although the information encoded into their genome sequences suggests the existence of mechanisms for low-temperature tolerance, the strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, as well as the other closely related species of the genera *Aquibacillus* and *Radiobacillus* have only shown growth above 10°C. On the contrary, the high-temperature tolerance deducted from the genome sequence has also been tested *in vitro*, with the bacteria being able to survive up to 50°C.

3.6. A putatively functional molybdenum cofactor biosynthetic pathway identified in the new isolates and species of the genus *Aquibacillus*

Since the closest described evolutionary relative of strains 3ASR75-11^T and 3ASR75-286 was the sole species of the genus *Radiobacillus* (Figure 5), an in-depth comparison of their genome-inferred metabolisms was explored in order to unveil their main similarities and differences. BlastKOALA annotation yielded 1,601 KO for the genome sequence of *R. deserti*, of which 284 were absent in the genomes of strains 3ASR75-11^T and 3ASR75-286, whereas 248 KO identified for the new isolated strains were missing in *R. deserti*. The single species of the genus *Radiobacillus* harbored metal transporters, such as those for iron (III) (K02010, K02011, and K02012),

TABLE 1 Differential features of the new isolated strains and closely related species.

Characteristics	1	2	3	4[a]	5[b]	6[c]	7[d]	8[e]	9[f]	10[g]	11[h]
Cell size	0.7–1.0 × 2.8-3.0	0.7–1.0 × 3.0–5.0	0.5–0.7 × 3.0–5.0	0.3–0.5 × 2.0–6.0	0.5–0.7 × 2.0–4.0	0.5–0.7 × 2.0–4.0	0.3–0.5 × 4.0–6.0	0.4–0.6 × 5.0–8.1	0.7–0.8 × 1.2–2.9	0.7–1.2 × 3.5–5.0	0.2–0.6 × 1.4–5.6
Colony pigmentation	White	White	White	White	White	Milk	White	Cream	White	Cream	White
Motility	+	+	+	NA	+	+	+	+	+	+	–
NaCl range (%, w/v)	0.5–17	0.5–20	NA	1–17	0–14	0–14	0.5–20	0–10	1–20	0.5–20	0–12
NaCl optimum (%, w/v)	7	2	NA	5–10	5–8	5–8	10	4	10	7–10	0–8
pH range	6–8	4–9	NA	4–9	5–12	6–9	6.5–9	6–9	6–10	5.5–9	6–9
pH optimum	6	5	NA	7	7	7	7	7	8	7	6–8.5
Temperature range (° C)	11–45	10–45	NA	15–45	10–40	10–40	20–40	25–45	10–45	15–45	20–50
Temperature optimum (° C)	37	37	NA	25–30	30	30	35	37	37	35	35
Anaerobic growth	+	+	+	–	–	–	–	–	+	NA	–
Aesculin hydrolysis	+	–	–	+*	+	+	+*	+*	+	+	NA
Urease	–	–	–	+*	–	+	+*	+*	NA	+	–
Nitrite reduction	+	–	–	+*	–	+	–*	–*	NA	+	NA
G + C content (mol%, genome)	38.0	38.0	38.1	36.7	35.9	36.0	36.8	36.9**	35.7	37.4	38.5

1. Strain 3ASR75-54^T; 2. Strain 3ASR75-11^T; 3. Strain 3ASR75-286; 4. *Aquibacillus albus* YIM 93624^T (G + C content from *A. albus* DSM 23711^T); 5. *Aquibacillus halophilus* B6B^T; 6. *Aquibacillus kalidii* HU2P27^T; 7. *Aquibacillus koreensis* BH30097^T (G + C content from *A. koreensis* JCM 12387^T); 8. *Aquibacillus salifodinae* WSY08-1^T; 9. *Aquibacillus saliphilus* KHM2^T; 10. *Aquibacillus sediminis* BH258^T; 11. *Radiobacillus deserti* TKL69^T. NA, not available. *, data from this study. **, determined by HPLC. [a]Zhang et al. (2012); [b]Amoozegar et al. (2014); [c]Wang et al. (2021); [d]Lee et al. (2006); [e]Zhang et al. (2015); [f]Cho and Whang (2022); [g]Lee and Whang (2019); [h]Li et al. (2020).

iron-siderophore (K23185, K23186, K23187, and K23188), manganese (K19975, K19976, and K19973), and zinc (K09815, K09816, and K09817). Besides, we found a mechanism for acid tolerance and Na^+ transporters encoded in the genome of *R. deserti,* but not in those of strains 3ASR75-11^T and 3ASR75-286. Conversely, strains 3ASR75-11^T and 3ASR75-286 possessed heme transporters (K02193, K02194, and K02195), as well as iron (III) citrate transport systems (K23181, K23182, K23183, and K23184). More significantly, strains 3ASR75-11^T and 3ASR75-286 presented ABC-type molybdate transporters (*modABC* genes; K02020, K02018, and K02017) and the biosynthetic pathway for the molybdenum cofactor (Moco), a relevant molecule in all domains of life (Hille, 1996; Leimkühler, 2020). Molybdoenzymes, or enzymes in which the active metal is molybdenum (Mo), are widespread in prokaryotes and eukaryotes, and more than 60 different molecules have been described to date (Hille et al., 2014; Peng et al., 2018). Organisms encoding them also harbors Moco biosynthetic and transport pathways in their genome (Peng et al., 2018). Those enzymes are mostly involved in redox reactions (Hille, 1996), with a key role in the metabolism of nitrogen, sulfur, and carbon compounds (Schoepp-Cothenet et al., 2012; Hille et al., 2014; Magalon and Mendel, 2015), and are related to anaerobic respiration in bacteria (Zupok et al., 2019). According to their Mo centers, they are divided into the xanthine oxidase (XO) family, the sulfite oxidase (SO) family, and the DMSO reductase family (Hille, 1996).

Several genes organized into five operons (*moaABCDE*, *mobAB*, *mocA*, *moeAB*, and *mogA*) (Shanmugam et al., 1992) had been related to the biosynthesis of Moco and one operon (*modABCD*) to the molybdate uptake system (Walkenhorst et al., 1995). Most of them were identified in the genomes of strains 3ASR75-54^T, 3ASR75-11^T, 3ASR75-286, and all the species of the genus *Aquibacillus*, but not in those of the genera *Sediminibacillus* and *Radiobacillus* (Supplementary Table S5). Moco biosynthetic pathway is highly conserved and there is evidence that it was once encoded by the last universal common ancestor (Schoepp-Cothenet et al., 2012; Magalon and Mendel, 2015). Though complex, Moco biosynthesis is well understood in prokaryotes (Zupok et al., 2019; Leimkühler, 2020) and it can be divided into four steps: (a) 5′-GTP (5′-guanosine triphosphate) is transformed into cPMP (cyclic pyranopterin monophosphate) by MoaA [a two (4Fe-4S)-cluster-containing enzyme] and MoaC (Wuebbens and Rajagopalan, 1993); (b) two sulfur atoms are inserted into cPMP by MoaD and MoaE (Pitterle et al., 1993), with IscS and TusA proteins involved in the sulfur-transfer process (Leimkühler, 2020), leading to molybdopterin (MPT); (c) MogA and MoeA add Mo to the molecule obtaining Mo-MTP (=Moco) (Joshi et al., 1996); (d) Mo-MTP can now be used as a cofactor by proteins of the SO family (Brokx et al., 2005; Havelius et al., 2011) or can be further modified to MCD (MPT cytosine dinucleotide) or bis-MGD (MPT guanosine dinucleotide) by MocA and MobA, respectively (Neumann et al., 2011; Reschke et al., 2013) (Figure 9). None of the studied genomes encoded the *mogA* gene, involved in step 3 of Moco biosynthesis. Nevertheless, when the surrounding medium possesses a high concentration of molybdate (>1 mM), the reaction catalyzed by the ATP-dependent MogA has been demonstrated not to be essential (Neumann and Leimkühler, 2008). Furthermore, the *moeB* gene, whose function is not-yet known, was only detected in strain 3ASR75-54^T and some species of *Aquibacillus*. However, considering that this study was based on draft (not complete) genomes (Supplementary Table S2) and that Moco biosynthetic pathway is highly conserved in organisms with Mo-dependent enzymes (Zhang et al., 2011; Hille et al., 2014), we could safely assume that the new isolates and all the species of the genus *Aquibacillus* possess a putatively functional biosynthetic route

for this cofactor. No molybdoenzymes were found in the proteomes of species of *Radiobacillus* and *Sediminibacillus*, in agreement with the lack of Moco biosynthetic pathway, but some Mo-requiring enzymes were encoded by strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, especially from the DMSO reductase family, the most prevalent family of molybdoenzymes in both bacteria and archaea (Zhang et al., 2011). Among them, we could highlight the presence of nitrate reductase *narGHI* (K00370, K00371, K00374) in strains 3ASR75-11^T, 3ASR75-286, *A. kalidii* and *A. sediminis*, and arsenite oxidase *aoxA* (K08355) in strain 3ASR75-54^T and the studied species of *Aquibacillus*, except for *A. sediminis*, in both cases with activities in the respiratory chain (Hille et al., 2014; Miralles-Robledillo et al., 2019).

The biosynthesis of Moco is positively regulated by ModE (Anderson et al., 2000), although this protein was not encoded in any of the studied genomes. However, the carbon storage regulator CsrA has been proved to enhance the Moco synthesis under conditions of high demand (Zupok et al., 2019), and its coding gene was present in all the analyzed genomes, including those from the genera *Sediminibacillus* and *Radiobacillus*.

The functional annotation of the metagenomic dataset SMO1 showed the presence of KO identifiers involved in the Moco biosynthesis among the 317,837 total assigned KO numbers. However, only 4 to 6 copies have been annotated for the proteins constituting the ModABC molybdate transporter. Therefore, the uptake of Mo from the media is not a widespread feature among the prokaryotic inhabitants of the hypersaline soils from the Odiel Saltmarshes Natural Area. The presence of this transporter in the genome of our isolates may indicate that their metabolisms highly rely on molybdenum related proteins and ensure the Mo uptake by specific transporters, whereas other molybdenum-requiring microorganisms acquiree Mo using other more unspecific mechanisms. Besides, molybdoenzyme *aoxA*, present in the genomes of 3ASR75-54^T and most members of the genus *Aquibacillus*, was not found in the SMO1 proteome.

To sum up, Moco biosynthesis is a well conserved pathway in prokaryotes. Its absence usually involves the lack of Mo-dependent proteins. Our genome analysis revealed the presence of a supposedly operational route for Moco synthesis, as well as of molybdate transporters and molybdoenzymes in the species of the genus *Aquibacillus* and the strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, whereas they were missing for the genera *Radiobacillus* and *Sediminibacillus*. Besides, strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, and possible other members of their genera, are among the few prokaryotes that harbor the molybdate transporter ModABC in the hypersaline soils from Odiel Saltmarshes Natural Area, according to the low genomic information annotated for this transporter in the SMO1 metagenomic dataset.

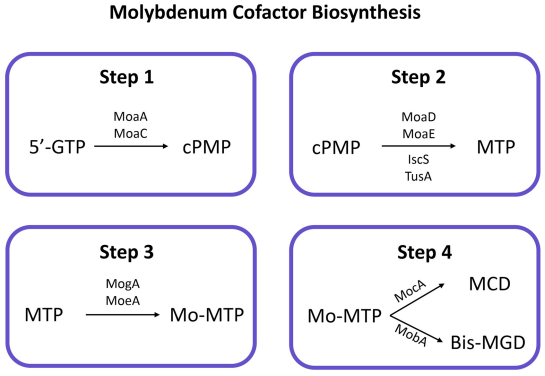


FIGURE 9
Conserved biosynthesis pathway for molybdenum cofactor in prokaryotes. Enzymes catalyzing each reaction are indicated next to the arrow. 5′-GTP, 5′-guanosine triphosphate; cPMP, cyclic pyranopterin monophosphate; MPT, molybdopterin; Mo-MTP, molybdenum inserted in MTP; MCD, MPT cytosine dinucleotide; bis-MGD, MPT guanosine dinucleotide.

3.7. Mechanisms detected for survival in hypersaline environments

Salt-in and *salt-out* strategies are the two main mechanisms in prokaryotes for osmoregulation under salt stress conditions. Haloarchaea (Youssef et al., 2014) and other extremely halophilic bacteria, such as species of the well-known genus *Salinibacter* (Antón et al., 2002), use the *salt-in* mechanism. These organisms commonly present an acidic proteome to avoid the denaturalization of their proteins under high salt concentrations inside the cell (Oren, 2008), but this adaptation can also be found in microorganisms with *salt-out* strategy (Elevi Bardavid and Oren, 2012; Oren, 2013), which is the most extended mechanism in prokaryotes as it allows survival under a wider range of osmotic conditions (Oren, 2008).

The isoelectric profile of strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 (Figure 10A) is similar to that of the species of the genera *Aquibacillus*, *Radiobacillus*, *Sediminibacillus*, and *Amphibacillus*, and differs from the proteome of the extremely halophilic archaeon *Haloarcula vallismortis* and the bacterium *Salinibacter ruber*. This result can be seen as the first evidence that the new isolates could have adopted the *salt-out* mechanism. An insight into the functional annotation of the analyzed genomes of the family *Bacillaceae* pointed out that a sudden increase in the osmotic pressure can be balanced with a cellular uptake of K^+ through Ktr potassium importers (K03498 and K03499), as it has been previously observed in *Bacillus subtilis* and *Synechocystis* sp. (Holtmann et al., 2003; Zulkifli et al., 2010; Hoffmann and Bremer, 2016). Efflux of Na^+ is crucial for cell survival due to the toxicity produced by high cytoplasmatic concentration of this ion (Patiño-Ruiz et al., 2022). Mrp multisubunit Na^+/H^+ exchangers (K05565, K05566, K05567, K05568, K05569, K05570, and K05571), encoded by the *mrpABCDEFG* operon and detected in our genome dataset, have been shown to provide salt tolerance in *Bacillus subtilis* (Ito et al., 1999), the slight halophile *Halomonas zhaodongensis* (Meng et al., 2014), and the cyanobacteria *Anabaena* sp. and *Synechococcus elongatus* (Blanco-Rivero et al., 2005; Cui et al., 2020), among others. Besides, Mrp antiporters (also known as Sha/Mnh/Pha) seem to contribute to the sporulation process (Kosono et al., 2000; Sun and Shi, 2001; Yoshinaka et al., 2003), which has been observed in the morphology of seven-day-old cells of strains 3ASR75-54^T, 3ASR75-11^T, 3ASR75-286, and the species of the genus *Aquibacillus* and *Radiobacillus* (Lee et al., 2006; Zhang et al., 2012, 2015; Amoozegar et al., 2014; Lee and Whang, 2019; Li et al., 2020; Wang et al., 2021; Cho and Whang, 2022). Additionally, the ChaA antiporter (K07300), which has a relevant role in the extrusion of Na^+ in *Escherichia coli* (Ohyama et al., 1994), has been found in the annotated genome of

strains 3ASR75-11^T and 3ASR75-286, but not in any of the other studied strains.

Although ion exchange is useful as a first barrier of defense against osmotic pressure, it is more convenient to store compatible solutes into the cytoplasm for prolonged periods of stress (Hoffmann and Bremer, 2016, 2017). There is a diverse spectrum of compatible solutes, but a common characteristic of all of them is their low molecular weight, which allows their accumulation resulting in an increased cytoplasmatic water content without altering the biochemical processes of the prokaryotic cells (Gregory and Boyd, 2021). All the analyzed genomes, except that of strain 3ASR75-54^T, harbored the osmoprotectant uptake (Opu) system (*opuABC* genes; K05845, K05846, K05847), an ABC transporter for acquisition of different compatible solutes (especially choline) from the environment (Hoffmann and Bremer, 2017; Teichmann et al., 2018) (Figures 10B,C). Besides, all the studied strains presented OpuD from the BCCT (betaine-choline-carnitine-transporter) family, with high affinity for glycine betaine (Kappes et al., 1996), but only 3ASR75-11^T together with the species of *Sediminibacillus* and some species of *Aquibacillus* exhibited OpuE from the SSS (sodium-solute-symporter) family of transporters (von Blohn et al., 1997). Another ABC-type transporter, namely ProVWX (K02000, K02001, and K02002), for glycine betaine uptake (Gregory and Boyd, 2021) was present in all genomes under study (Figure 10B).

TeaABC transporter (encoded by *teaABC* genes) (Figure 10C) is a member of the TRAP transporters family with high affinity for ectoine and, to a lesser extent, for its derivative 5-hydroxyectoine (Grammann et al., 2002; Kuhlmann et al., 2008b). It allows the reuptake of secreted ectoine in *Halomonas elongata* and, additionally, plays a role as an effective salvage system for ectoine leaking through the membrane (Kuhlmann et al., 2008b). Previous studies have demonstrated that the presence of the three genes (*teaABC*) is mandatory for the correct function of this ectoine-specific transporter (Grammann et al., 2002). Whereas *teaC* was present in all the studied genomes, *teaA* and *teaB* were only detected in some of them, so, theoretically only the species *A. albus*, *A. kalidii*, *A. saliphilus*, *R. deserti*, *S. halophilus*, and *S. terrae* might have this transporter available. None of the three new isolates encoded the complete set of *teaABC* genes (Figure 10C). UehABC is a second TRAP transporter, previously studied in *Silicibacter pomeroyi* DSS-3, that imports ectoine and hydroxyectoine (Figure 10B). One main difference with TeaABC is that the *uehABC* genes are coregulated with other genes for ectoine degradation. It must be noted that *S. pomeroyi* uses ectoine as sole carbon and nitrogen source, whereas *H. elongata* mainly utilizes it for osmoprotection purposes (Lecher et al., 2009). Only *A. sediminis* harbored the three *uehABC* genes, while they were quite scarce in the other studied genomes (Figure 10C). The higher abundance of TeaABC-related genes over UehABC-related ones might indicate a prevalence of the osmoprotective activity of ectoine over its carbon and nitrogen source utilization in the analyzed strains of the family *Bacillaceae*.

Transportation of ions and compatible solutes into the cytoplasm to balance osmotic pressure must be paired with the ability of the cell to secrete them when a drop in the environmental salt concentration occurs. Msc mechanosensitive channels, which were identified in our genome dataset (K16053, for strain 3ASR75-54^T; K03282, for the remaining strains), play an important role in releasing ions and organic molecules under osmotic down shock stress (Booth and Blount, 2012; Booth, 2014).

In addition to the compatible solute uptake, *de novo* biosynthetic potential for the osmolytes glycine betaine and ectoine has been detected in the analyzed genome sequences (Figure 10C). Choline is transformed into glycine betaine in two oxidative steps carried out by BetA (K00108) and BetB (K00130) (Boch et al., 1994), under aerobic conditions (Czech and Bremer, 2018) (Figure 10B). This reaction is widely present in halophilic bacteria and archaea (Gregory and Boyd, 2021) as well as in eukaryotic cells. On the contrary, biosynthesis of ectoine and its derivative is specific to prokaryotes (Czech and Bremer, 2018). Ectoine is obtained from L-aspartate in five steps mediated by aspartate kinase (*lysC*; K00928), aspartate semialdehyde dehydrogenase (*asd*; K00133), diaminobutyrate-2-oxoglutarate transaminase (*ectB*; K00836), L-2,4-diaminobutyric acid acetyltransferase (*ectA*; K06718), and L-ectoine synthase (*ectC*; K06720) (Ono et al., 1999; Figure 10B). All the analyzed genomes possessed the machinery to *de novo* synthesize ectoine. The *ectABC* genes are usually arranged into a single operon (Louis and Galinski, 1997; Kuhlmann et al., 2008a; van Thuoc et al., 2020), as is the case with the studied strains (Figure 11), whereas in other bacterial genomes the synteny and chromosomal placement of the genes is not conserved (León et al., 2018). Some bacteria possess the *ectD* gene that enable the conversion of ectoine into 5-hydroxyectoine, but among the analyzed strains only *A. albus* and *A. halophilus* harbored that gene (Figure 10C). Both osmolytes, ectoine and 5-hydroxyectoine, can be catabolized to be used as carbon and energy sources, albeit this mechanism has not been completely understood. Regardless, DoeA and DoeB proteins seem to be relevant in the degradation (Schwibbert et al., 2011; Reshetnikov et al., 2020). Ectoine producers do not usually catabolize it (Schulz et al., 2017), although some microorganisms can synthesize and consume ectoine, such as *Halomonas elongata* (Schwibbert et al., 2011) and *Sinobaca* sp. (Chen et al., 2022). All the studied genome sequences encoded the ectoine hydrolase gene (*doeA*) and, additionally, *doeB* gene was present in the genome of strains 3ASR75-11^T and 3ASR75-286. Thus, organisms that are, at the same time, ectoine producers and consumers may not be as uncommon as previously thought.

After having carried out a comparative genomic analysis we could assert that ectoine plays a significant role in the physiology of strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, as well as in that of their closely related species. Ectoine biosynthetic pathway was completely conserved in the 14 studied genomes. Moreover, two ectoine-targeted transporters were found, one of them (TeaABC) present entirely or with only one gene missing in 11 strains, and the other (UehABC) mostly incomplete for the vast majority of genomes. Besides, only one gene involved in ectoine degradation (*doeA*) was found across all taxa (Figure 10C). Thus, it seems likely that ectoine is mainly utilized as an osmolyte rather than as a source of carbon and nitrogen in our new isolates.

Ectoine is a valuable molecule for biotechnological purposes due to its ability to protect cell components under stressful conditions, such as freezing, high temperature, and drying (Ma et al., 2020; Vandrich et al., 2020). The moderately halophilic bacterium "*Halomonas bluephagenesis*" TD01 has been demonstrated to yield 28 g L^{-1} of ectoine during a 28-h fed-batch growth process (Ma et al., 2020). Besides, ectoine biosynthetic pathway shares the first two steps with the synthesis of one of the

10.3389/fmicb.2023.1192059

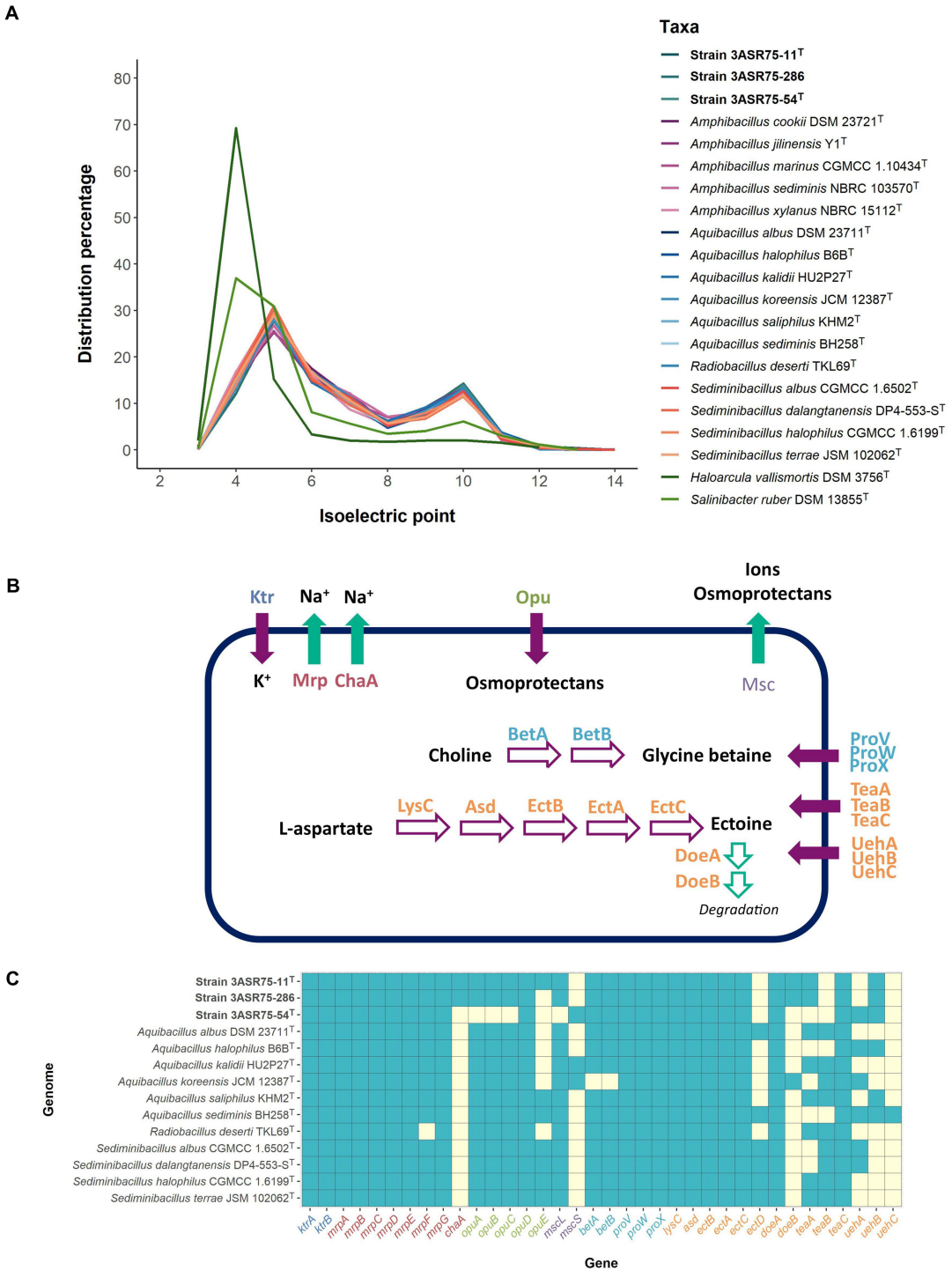


FIGURE 10
(A) Distribution of the isoelectric point of the proteomes under study. *Haloarcula vallismortis* DSM 3756^T (GCF_900106715.1) and *Salinibacter ruber* DSM 13855^T (GCF_000013045.1) were also included as representatives of extreme halophiles with *salt-in* osmoregulation strategy. The proteomes of members of the family *Bacillaceae* were less acidic than those of the extreme halophiles, suggesting a *salt-out* mechanism. **(B)** Reconstruction of the different strategies employed for the studied strains to cope with osmotic stress. Purple filled arrows indicate uptake, green filled arrows indicate extrusion, purple empty arrows indicate biosynthesis, and green empty arrows indicate degradation. Enzymes and transporters are colored key as follows: K+ uptake (dark blue), Na+ extrusion (red), unspecific osmoprotectant uptake (green), unspecific extrusion (light purple), glycine betaine biosynthesis and uptake (light blue), ectoine metabolism and uptake (orange). **(C)** Heatmap of presence (light blue)/absence (light yellow) of osmoregulation-related genes in the new strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, and the closely related species of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*.

most industrially produced amino acids, threonine (Dong et al., 2012). Actually, the aforementioned strain "*H. bluephagenesis*" TD01 is able to produce 33 g L^{-1} of threonine in a 7-liter bioreactor (Du et al., 2020). Strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286, as well as the species of the genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus*, also harbored the remaining genes that

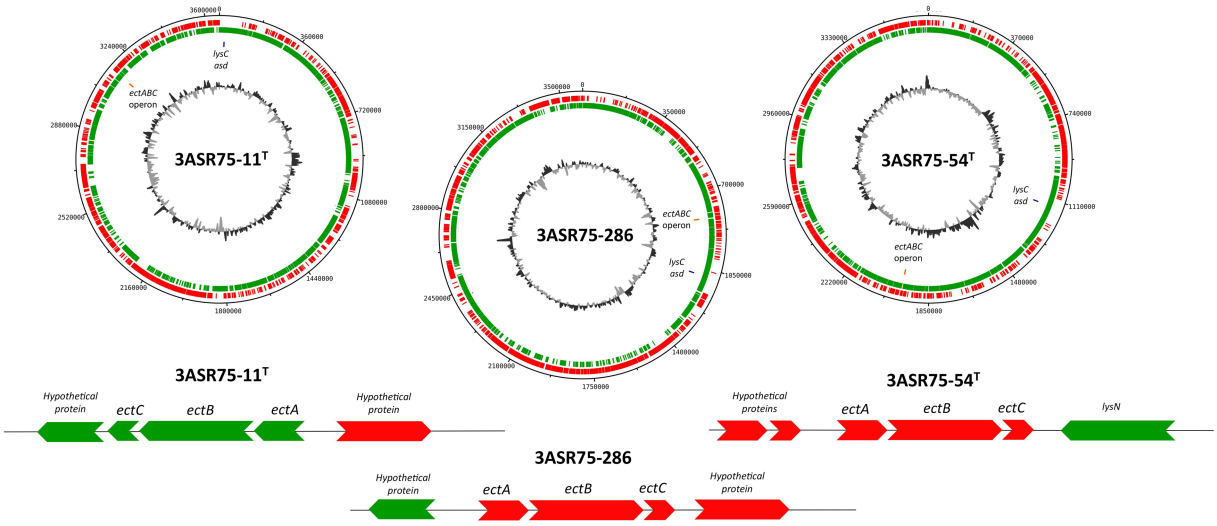


FIGURE 11
Circular genomic DNA map of strains 3ASR75-11^T, 3ASR75-286, and 3ASR75-54^T. Circles indicate, from the inside outward: G+C content; location of the *ectABC* operon (orange) and *lysC* and *asd* genes (blue); coding sequences in the lagging strand (green); coding sequence in the leading strand (red). A closer insight of the *ectABC* operon context is displayed below each genome map.

encode the last three steps of threonine biosynthesis: homoserine dehydrogenase (*hom*; K00003), homoserine kinase (*thrB*; K00872), and threonine synthase (*thrC*; K01733). Due to the slow growth rate of the novel isolates, they might not be the preferred source for the biotechnological production of those two molecules. However, they could be of importance to increase their yield after an in-depth study of their ectoine and threonine pathways.

3.8. Strategies to thrive in heavy metal contaminated soils

Considering the high concentration of heavy metals detected in the sampled soils, we decided to explore the exporting mechanisms that might be present in the genomes of the new strains and their closest relatives (Supplementary Table S6). The *ars* operon (K03325, K03741, and K03892) can accomplish the reduction of arsenate to arsenite and its posterior expulsion from the cytoplasm (Chauhan et al., 2019; Islam et al., 2022). The presence of those genes in our dataset was expected due to its wide spreading among prokaryotes (Chen et al., 2020). The *zntA* gene (K01534) encoding Zn^{2+}/Cd^{2+}/Pb^{2+} pumping out coupled to ATP hydrolysis (Rensing et al., 1997; Noll and Lutsenko, 2000) was also identified, but the efficient CzcCBA efflux system for zinc and cadmium was not found (Legatzki et al., 2003). CopA and CopB proteins (K17686 and K01533) are P-type copper efflux transporters that confers resistance to this metal (Rensing et al., 2000; Mana-Capelli et al., 2003) and they were encoded in most of the studied genomes, including the strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286. Furthermore, MerB alkylmercury lyase (K00221) was solely found in strain 3ASR75-286. Bacteria can cope with methylmercury contamination thanks to the sequential activity of MerB and MerA proteins, whose expression is controlled by the regulatory protein MerR (Schaefer et al., 2004). However, neither MerA nor MerR were identified in the genome of this strain, which is consistent with the fact that disturbingly high concentrations of methylmercury have not been perceived in the sampled area before (Sainz et al., 2002, 2004).

Multiple copies of *copA*, *copB*, and *zntA* genes have been annotated for the translated CDS of the SMO1 reference metagenomic dataset from a hypersaline soil of Odiel Saltmarshes Natural Area (Supplementary Table S6). Those three genes are among the five more identified KO in the SMO1 dataset, along with a putative transposase (K07496) and a putative ABC transport system permease protein (K02004). The frequency of these genes in the SMO1 metagenome are 2,106, 1,642, and 1,729 copies from a total of 317,837 KO annotated CDS. Therefore, we can assume the importance of the activities of those divalent cation transporters in the metabolism of the prokaryotic population of Odiel soils, including the novel isolates 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286. On the other hand, the arsenic resistant activities seem to be less spread among the prokaryotes inhabiting the soils under study, and none of the annotated functions in the reference SMO1 dataset corresponded to *merB* gene (Supplementary Table S6) which is present exclusively in isolate 3ASR75-286.

3.9. Overlooked inhabitants in sampled soils

Since the ecological distribution of species of *Aquibacillus* and related genera has not been investigated in depth, a fragment recruitment analysis of the new isolated strains was performed from a total of 16 metagenomic libraries originated from hypersaline environments (i.e., saltern ponds with different salt concentrations, hypersaline lakes, saline soils, desert soils, salt crust, microbialites and arctic spring sediments). In order to compare the recruitment results, three representative halophilic microorganisms known to be significantly abundant in saline

habitats (*Haloquadratum walsbyi*, *Salinibacter ruber*, and *Spiribacter salinus*) were also included into the analysis. Reads recruitment normalized against the size of the genomes and the database, denoted as RPKG, was low for all the three isolates in the studied metagenomic datasets (Figure 12A). Their abundance was especially rare in environments with (almost-)saturated salt concentration, such as the salterns ponds from Chile (Cáhuil) (Plominsky et al., 2014), Isla Cristina (IC21) (Fernández et al., 2014b), and Santa Pola (SS33 and SS37) (Ghai et al., 2011; Fernández et al., 2014a) in Spain, and Puerto Rico (Cabo Rojo) (Couto-Rodríguez and Montalvo-Rodríguez, 2019), the hypersaline lakes from Australia (Tyrrell 0.1 and Tyrrell 0.8) (Podell et al., 2014) and Iran (Urmia) (Kheiri et al., 2023), and the salt crust from the Qi Jiao Jing Lake in China (Xinjiang) (Xie et al., 2022), and slightly higher at intermediate salinities [SS13 and SS19 from Santa Pola salterns (Ghai et al., 2011; Fernández et al., 2014a), and microbialites from Campo Naranja in Argentina (Perez et al., 2020)]. Recruitment plot from SMO1 and SMO2 metagenomes (Vera-Gargallo et al., 2018), corresponding to samples collected a few years ago from the same hypersaline soils (Odiel) than those analyzed in the present study, displayed similar abundance of the new isolates to that found for intermediate salterns ponds (SS13 and SS19). Furthermore, previous taxonomic annotation of SMO1 and SMO2 databases registered that the phylum *Bacillota* (to which the new strains belong) represented a small fraction of the microbial population (Vera-Gargallo et al., 2019). Indeed, relative abundances for strains 3ASR75-11^T and 3ASR75-286 varied between 0.02–0.0182% whereas strain 3ASR75-54^T was even less abundant, with 0.0163–0.0157%. Soils from Gujarat desert, which are mostly dominated by *Pseudomonadota* (Patel et al., 2015), and hypersaline Arctic Spring sediments, whose microbial life has been hypothesized as survival organisms on Mars (Magnuson et al., 2022), harbored the highest abundance for our isolates (Figure 12A), but with values ranging from 0.0416–0.0364% and 0.048–0.042%, respectively. In all cases, these values were lower than the 0.1% threshold commonly used to label the so-called "rare biosphere" (Pedrós-Alió, 2012). Therefore, strains 3ASR75-54^T, 3ASR75-11^T, and 3ASR75-286 could be considered part of the low-abundant prokaryotic fraction inhabiting soils from Odiel Saltmarshes Natural Area. A closer look at the recruitments to visualize up to what extent the genome of the new strains is covered in SMO1 and Gujarat metagenomes showed many coverage gaps over the 95% identity, a widely accepted cutoff for species delineation (Figure 12B). Again, this finding suggests that the new species represented by strain 3ASR75-54^T and by strains 3ASR75-11^T and 3ASR75-286, respectively, are scarce inhabitants of the studied soils, although they have demonstrated to be relatively easy to cultivate and manipulate in laboratory conditions. Our work emphasizes the relevance of the traditional isolation and characterization methodology to explore the rare biosphere. Previous research to uncover the culturable diversity of the sampled environment succeeded to isolate a new member of the phylum *Balneolota* (Galisteo et al., 2023), which has been listed as one of the major phyla in the hypersaline soils of Odiel Saltmarhes Natural Area (Vera-Gargallo et al., 2019). In the authors' opinion, culture-dependent studies are indispensable for a better knowledge of the microbial diversity unveiled by metagenomic approaches but also for discovering taxa that cannot be detected by high throughput sequencing.

4. Conclusion

A large set of strains affiliated to the family *Bacillaceae* was isolated in this study after a long incubation period and further divided into two separate groups. Representative strains of these groups were selected for phylogenomic, comparative genomic, phenotypic, and chemotaxonomic analyses, which confirmed their placement as new bacterial species. Besides, AAI values between one of these taxa and the species of the closely related genera *Aquibacillus*, *Radiobacillus*, and *Sediminibacillus* acknowledged its description as a new genus. Thus, we propose the classification of strain 3ASR75-54^T within the genus *Aquibacillus,* as a new species, for which the name *Aquibacillus salsiterrae* sp. nov. is proposed, and the placement of strains 3ASR75-11^T and 3ASR75-286 into a new separate genus and species, for which the new name *Terrihalobacillus insolitus* gen. nov., sp. nov. is proposed. The descriptions of these new taxa are shown below.

The new species of the genus *Aquibacillus* and the new genus *Terrihalobacillus* encoded the well-conserved molybdenum cofactor biosynthetic pathway and the molybdenum-dependent enzymes, differentiating them from other closely related member of the family *Bacillaceae*, such as the genera *Radiobacillus* and *Sediminibacillus*. Besides, in-depth in-silico analysis of their genome sequences revealed strategies to deal with high salt concentration and heavy metal contamination of the soils from Odiel Saltmarshes Natural Area. The *salt-out* mechanism of osmoregulation seems to be prevailing in the three new isolates, harboring proteins for uptake and *de novo* biosynthesis of ectoine and glycine betaine, two of the most frequent compatible solutes in prokaryotes. The genes *arsC*, *arsR*, *arsB*, *zntA*, *copA*, and *copB*, among others, have been identified in the genomes of the isolates, pointing at their tolerance to heavy metals, such as arsenic, zinc, cadmium, lead, and copper. The abundance of the new described species from the genera *Aquibacillus* and *Terrihalobacillus* is extremely low in all the studied hypersaline environments, including the isolation area, with a relative abundance under 0.1%, the threshold for the so-called "rare-biosphere."

Description of *Aquibacillus salsiterrae* sp. nov.

Aquibacillus salsiterrae sp. nov. (sal.si.ter'rae. L. masc. adj. *salsus* salty; L. fem. n. *terra* earth, soil; N.L. gen. n. *salsiterrae* of salty soil).

Cells are Gram-stain-positive, motile rods with a size of 0.7–1.0 × 2.8–3.0 μm. Endospores are formed at terminal position. Colonies are semi-translucent and white-colored, with a size of 2.0–2.5 mm when grown in R2A medium supplemented with 7.5% (w/v) salts after 24 h of incubation at 37°C. Facultative anaerobe. The temperature range for growth is 11–45°C (optimum at 37°C). The pH values supporting growth are 6.0–8.0 (optimum at pH 6.0) and the NaCl concentration for growth is 0.5–17% (w/v) [optimum at 7% (w/v)]. Catalase and oxidase positive. Hydrolyzes aesculin but not casein, DNA, gelatin, starch, and Tween 80. Reduces nitrate and nitrite. Positive for methyl red test but negative for Voges-Proskauer test, meaning that it uses the mixed-acid pathway for glucose fermentation. Indole production, Simmons' citrate test, phenylalanine deaminase, urease, and H_2S production are negative. Acids are produced from D-arabinose, D-fructose, glycerol, D-glucose, lactose, maltose, mannitol, sucrose, D-trehalose, and D-xylose, but not from

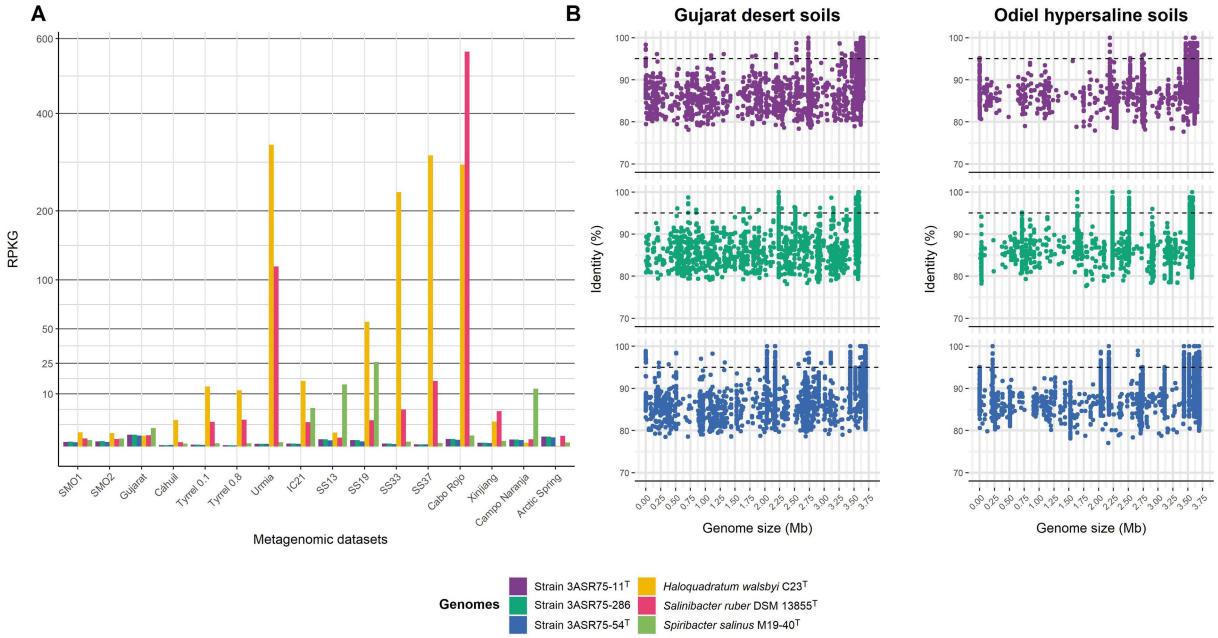


FIGURE 12
Fragment recruitment of the three new isolated strains from relevant hypersaline metagenomic datasets. Further information about these metagenomes is detailed in Supplementary Table S1. **(A)** Relative abundance represented as RPKG of strains 3ASR75-11^T, 3ASR75-286, and 3ASR75-54^T together to three reference halophilic species. Squared root transformation was performed for Y-axis in order to better visualize low values. **(B)** Plots show read recruitment across genome length of the new isolates from Gujarat desert soil (left) and Odiel Saltmarshes hypersaline soil (right) metagenomes.

D-galactose. Utilizes L-arabinose, D-cellobiose, D-maltose, D-mannose, melibiose, D-trehalose, D-xylose, butanol, dulcitol, ethanol, glycerol, mannitol, methanol, propranolol, D-sorbitol, xylitol, benzoate, formate, fumarate, hippurate, malate, and propionate as sole source of carbon and energy, but not aesculin, amygdalin, D-melezitose, ribose, starch, acetate, butyrate, glutamate, pyruvate, and valerate. Utilizes L-alanine, L-asparagine, aspartic acid, L-cysteine, glycine, L-glutamine, L-methionine, ornithine, L-phenylalanine, L-serine, L-threonine, tryptophane, and valine as sole source of carbon, nitrogen, and energy, but not arginine. Major fatty acids are anteiso-$C_{15:0}$, followed by iso-$C_{15:0}$ and anteiso-$C_{17:0}$. The genome of the type strain has a G + C content of 38.0 mol% and its approximate size is 3.70 Mb.

The type strain, 3ASR75-54^T (=CCM 9168^T = CECT 30368^T), was isolated from a hypersaline soil at the Odiel Saltmarshes Natural Area in Huelva (Southwest Spain). The accession number for the 16S rRNA gene sequence is ON652841 and that for the genome sequence is GCF_028416595.1.

Description of *Terrihalobacillus* gen. nov.

Terrihalobacillus gen. nov. (Ter.ri.ha.lo.ba.cil'lus. L. fem. n. *terra*, land; Gr. masc. n. *hals*, salt; L. masc. dim. n. *bacillus*, a small rod; N.L. masc. n. *Terrihalobacillus*, a small rod from salty land).

Cells are Gram-stain-positive, motile, and endospore-forming rods that form white-pigmented colonies. Endospores are formed at terminal position. Moderately halophilic, growing in a wide range of NaCl concentrations. Mesophile and facultative anaerobe. Catalase and oxidase positive. Genome mining reveals the biosynthetic pathway for the molybdenum cofactor and genes encoding for molybdoenzymes. Major fatty acid is anteiso-$C_{15:0}$. It belongs to the family *Bacillaceae*, order *Caryophanales*, class *Bacilli*, and phylum *Bacillota*. The DNA G + C content is 38.0–38.1 mol% (genome). The type species is *Terrihalobacillus insolitus*.

Description of *Terrihalobacillus insolitus* sp. nov.

Terrihalobacillus insolitus sp. nov. (in.so'li.tus. L. masc. adj. *insolitus*, unusual or uncommon).

Cell are Gram-stain-positive, motile rods with a size of 0.5–1.0 × 3.0–5.0 μm. Endospores are formed at terminal position. Colonies are circular, convex, opaque, and white-colored, with a size of 1 mm when growing in R2A medium supplemented with 7.5% (w/v) salts after 24 h of incubation at 37°C. Facultative anaerobe. The temperature range for growth is 10–45°C (optimum at 37°C). The pH values supporting growth are 4.0–9.0 (optimum at pH 5.0) and the NaCl concentration for growth is 0.5–20% (w/v) [optimum at 2% (w/v)]. Catalase and oxidase positive. Does not hydrolyze aesculin, casein, DNA, gelatin, starch, and Tween 80. Reduces nitrate, but not nitrite. Positive for the methyl red test but negative for the Voges-Proskauer test, meaning that it uses the mixed-acid pathway for glucose fermentation. Indole production, Simmons' citrate test, phenylalanine deaminase, urease, and H_2S production are negative. Acids are produced from D-fructose, D-galactose, D-glucose, glycerol, lactose, maltose, mannitol,

sucrose, and D-trehalose, but not from D-arabinose. Utilizes D-fructose, D-maltose, D-mannose, D-melezitose, salicin, sucrose, D-trehalose, D-xylose, mannitol, xylitol, benzoate, butyrate, fumarate, hippurate, malate, and pyruvate as sole source of carbon and energy, but not aesculin, amygdalin, D-cellobiose, D-galactose, D-glucose, D-lactose, D-raffinose, starch, butanol, acetate, glutamate, and valerate. Utilizes glycine as sole source of carbon, nitrogen, and energy source, but not L-asparagine, aspartic acid, and L-threonine. Major fatty acid is anteiso-$C_{15:0}$. The genome of the type strain has a G + C content of 38.0 mol% and its approximate size is 3.66 Mb.

The type strain, 3ASR75-11^T (=CCM 9167^T = CECT 30367^T), was isolated from a hypersaline soil at the Odiel Saltmarshes Natural Area in Huelva (Southwest Spain). The accession number for the 16S rRNA gene sequence is ON652838 and that for the genome sequence is GCF_028416575.1. Strain 3ASR75-286 is an additional strain of this species. Its DNA G + C content is 38.1 mol% (genome) and its approximate genome size is 3.59 Mb. The accession number for its 16S rRNA gene sequence is ON653021 and that for its genome sequence is GCF_028416555.1.

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

AV and CS-P conceived the study. CG, CS-P, and AV obtained the environmental samples. CG accomplished the laboratory experiments and the in-silico analysis, supported by CS-P and RRH, respectively. CG drafted the manuscript. CG, CS-P, RRH, and AV revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by grant PID2020-118136GB-I00 funded by MCIN/AEI/10.13039/501100011033 (to AV and CS-P). AV acknowledges the support from the Junta de Andalucía (grants P20_01066 and BIO-213), all with FEDER funds. CG was a recipient of a predoctoral fellowship (PRE2018-083242) from the Spanish Ministry of Science and Innovation. RRH was a recipient of a short-stay grant (PRX21/00598) from the Spanish Ministry of Universities.

Acknowledgments

We thank A. Oren from The Hebrew University of Jerusalem for his help on the nomenclature of the new species and genus.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1192059/full#supplementary-material

References

Amoozegar, M. A., Bagheri, M., Didari, M., Mehrshad, M., Schumann, P., Spröer, C., et al. (2014). *Aquibacillus halophilus* gen. nov., sp. nov., a moderately halophilic bacterium from a hypersaline lake, and reclassification of *Virgibacillus koreensis* as *Aquibacillus koreensis* comb. nov. and *Virgibacillus albus* as *Aquibacillus albus* comb. nov. *Int. J. Syst. Evol. Microbiol.* 64, 3616–3623. doi: 10.1099/ijs.0.065375-0

Anderson, L. A., McNairn, E., Leubke, T., Pau, R. N., and Boxer, D. H. (2000). ModE-dependent molybdate regulation of the molybdenum cofactor operon *moa* in *Escherichia coli*. *J. Bacteriol.* 182, 7035–7043. doi: 10.1128/JB.182.24.7035-7043.2000

Antón, J., Oren, A., Benlloch, S., Rodríguez-Valera, F., Amann, R., and Rosselló-Mora, R. (2002). *Salinibacter ruber* gen. nov., sp. nov., a novel, extremely halophilic member of the *Bacteria* from saltern crystallizer ponds. *Int. J. Syst. Evol. Microbiol.* 52, 485–491. doi: 10.1099/00207713-52-2-485

Auch, A. F., Klenk, H. P., and Göker, M. (2010). Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand. Genomic Sci.* 2, 142–148. doi: 10.4056/sigs.541628

Auguie, B. (2017). gridExtra: miscellaneous functions for grid graphics. Available at: https://cran.r-project.org/package=gridExtra

Blanco-Rivero, A., Leganés, F., Fernández-Valiente, E., Calle, P., and Fernández-Piñas, F. (2005). *mrpA*, a gene with roles in resistance to Na^+ and adaptation to alkaline pH in the cyanobacterium *Anabaena* sp. PCC7120. *Microbiology* 151, 1671–1682. doi: 10.1099/mic.0.27848-0

Boch, J., Kempf, B., and Bremer, E. (1994). Osmoregulation in *Bacillus subtilis*: synthesis of the osmoprotectant glycine betaine from exogenously provided choline. *J. Bacteriol.* 176, 5364–5371. doi: 10.1128/jb.176.17.5364-5371.1994

Booth, I. R. (2014). Bacterial mechanosensitive channels: progress towards an understanding of their roles in cell physiology. *Curr. Opin. Microbiol.* 18, 16–22. doi: 10.1016/j.mib.2014.01.005

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

Brokx, S. J., Rothery, R. A., Zhang, G., Ng, D. P., and Weiner, J. H. (2005). Characterization of an *Escherichia coli* sulfite oxidase homologue reveals the role of a conserved active site cysteine in assembly and function. *Biochemistry* 44, 10339–10348. doi: 10.1021/bi050621a

Carrasco, I. J., Márquez, M. C., Xue, Y., Ma, Y., Cowan, D. A., Jones, B. E., et al. (2008). *Sediminibacillus halophilus* gen. nov., sp. nov., a moderately halophilic, Gram-positive bacterium from a hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 58, 1961–1967. doi: 10.1099/ijs.0.65790-0

Carver, T., Thomson, N., Bleasby, A., Berriman, M., and Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119–120. doi: 10.1093/bioinformatics/btn578

Charif, D., and Lobry, J. R. (2007). "SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis"

in *Structural approaches to sequence evolution: Molecules, networks, populations*. eds. U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo (New York: Springer-Verlag), 207–232.

Chauhan, D., Srivastava, P. A., Agnihotri, V., Yennamalli, R. M., and Priyadarshini, R. (2019). Structure and function prediction of arsenate reductase from *Deinococcus indicus* DR1. *J. Mol. Model.* 25:15. doi: 10.1007/s00894-018-3885-3

Chen, S. Y., Peng, T. C., Huang, S. Z., and Chien, C. C. (2022). Isolation of an ectoine-producing *Sinobaca* sp. and identification of genes that are involved in ectoine biosynthesis. *FEMS Microbiol. Lett.* 369:fnac046. doi: 10.1093/femsle/fnac046

Chen, S. C., Sun, G. X., Yan, Y., Konstantinidis, K. T., Zhang, S. Y., Deng, Y., et al. (2020). The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. *Proc. Natl. Acad. Sci. U. S. A.* 117, 10414–10421. doi: 10.1073/pnas.2001063117

Cho, G.-Y., and Whang, K.-S. (2022). *Aquibacillus saliphilus* sp. nov., a moderately halophilic bacterium isolated from a grey saltern. *Int. J. Syst. Evol. Microbiol.* 72:5496. doi: 10.1099/ijsem.0.005496

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

Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2016). GenBank. *Nucleic Acids Res.* 44, D67–D72. doi: 10.1093/nar/gkv1276

Consejería de Medio Ambiente (1999). *Los criterios y estándares para declarar un suelo contaminado en Andalucía y la metodología y técnicas de toma de muestra y análisis para su investigación*. Sevilla: Junta de Andalucía.

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

Couto-Rodríguez, R. L., and Montalvo-Rodríguez, R. (2019). Temporal analysis of the microbial community from the crystallizer ponds in Cabo Rojo, Puerto Rico, using metagenomics. *Genes* 10:422. doi: 10.3390/genes10060422

Cowan, S. T., and Steel, K. J. (1965). *Manual for the identification of medical bacteria*. Cambridge: Cambridge University Press.

Cui, J., Sun, T., Li, S., Xie, Y., Song, X., Wang, F., et al. (2020). Improved salt tolerance and metabolomics analysis of *Synechococcus elongatus* UTEX 2973 by overexpressing Mrp antiporters. *Front. Bioeng. Biotechnol.* 8:500. doi: 10.3389/fbioe.2020.00500

Czech, L., and Bremer, E. (2018). With a pinch of extra salt—did predatory protists steal genes from their food? *PLoS Biol.* 16:e2005163. doi: 10.1371/journal.pbio.2005163

Dong, X., Quinn, P. J., and Wang, X. (2012). "Microbial metabolic engineering for L-threonine production" in *Reprogramming microbial metabolic pathways*. eds. X. Wang, J. Chen and P. Quinn (Dordrecht: Springer), 283–302.

Du, H., Zhao, Y., Wu, F., Ouyang, P., Chen, J., Jiang, X., et al. (2020). Engineering *Halomonas bluephagenesis* for L-threonine production. *Metab. Eng.* 60, 119–127. doi: 10.1016/j.ymben.2020.04.004

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

Elevi Bardavid, R., and Oren, A. (2012). Acid-shifted isoelectric point profiles of the proteins in a hypersaline microbial mat: an adaptation to life at high salt concentrations? *Extremophiles* 16, 787–792. doi: 10.1007/s00792-012-0476-6

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

Felsenstein, J. (1983). Parsimony in systematics: biological and statistical issues. *Annu. Rev. Ecol. Syst.* 14, 313–333. doi: 10.1146/annurev.es.14.110183.001525

Fernández, A. B., Ghai, R., Martin-Cuadrado, A.-B., Sánchez-Porro, C., Rodriguez-Valera, F., and Ventosa, A. (2014a). Prokaryotic taxonomic and metabolic diversity of an intermediate salinity hypersaline habitat assessed by metagenomics. *FEMS Microbiol. Ecol.* 88, 623–635. doi: 10.1111/1574-6941.12329

Fernández, A. B., Vera-Gargallo, B., Sánchez-Porro, C., Ghai, R., Papke, R. T., Rodriguez-Valera, F., et al. (2014b). Comparison of prokaryotic community structure from Mediterranean and Atlantic saltern concentrator ponds by a metagenomic approach. *Front. Microbiol.* 5:196. doi: 10.3389/fmicb.2014.00196

Galisteo, C., de la Haba, R. R., Sánchez-Porro, C., and Ventosa, A. (2023). Biotin pathway in novel *Fodinibius salsisoli* sp. nov., isolated from hypersaline soils and reclassification of the genus *Aliifodinibius* as *Fodinibius*. *Front. Microbiol.* 13:1101464. doi: 10.3389/fmicb.2022.1101464

Ghai, R., Pašić, L., Fernández, A. B., Martin-Cuadrado, A.-B., Mizuno, C. M., McMahon, K. D., et al. (2011). New abundant microbial groups in aquatic hypersaline environments. *Sci. Rep.* 1:135. doi: 10.1038/srep00135

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

Grammann, K., Volke, A., and Kunte, H. J. (2002). New type of osmoregulated solute transporter identified in halophilic members of the *Bacteria* domain: TRAP transporter TeaABC mediates uptake of ectoine and hydroxyectoine in *Halomonas elongata* DSM 2581^{T}. *J. Bacteriol.* 184, 3078–3085. doi: 10.1128/JB.184.11.3078-3085.2002

Gregory, G. J., and Boyd, E. F. (2021). Stressed out: bacterial response to high salinity using compatible solute biosynthesis and uptake systems, lessons from *Vibrionaceae*. *Comput. Struct. Biotechnol. J.* 19, 1014–1027. doi: 10.1016/j.csbj.2021.01.030

Guangchuang, Y., Shuangbin, X., and Hackl, T. (2022). Aplot: decorate a ggplot with associated information. Available at: https://cran.r-project.org/web/packages/aplot/index.html

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

Havelius, K. G. V., Reschke, S., Horn, S., Döring, A., Niks, D., Hille, R., et al. (2011). Structure of the molybdenum site in YedY, a sulfite oxidase homologue from *Escherichia coli*. *Inorg. Chem.* 50, 741–748. doi: 10.1021/ic101291j

Hille, R. (1996). The mononuclear molybdenum enzymes. *Chem. Rev.* 96, 2757–2816. doi: 10.1021/cr950061t

Hille, R., Hall, J., and Basu, P. (2014). The mononuclear molybdenum enzymes. *Chem. Rev.* 114, 3963–4038. doi: 10.1021/cr400443z

Hoffmann, T., and Bremer, E. (2016). "Management of osmotic stress by *Bacillus subtilis*: genetics and physiology," in *Stress and environmental regulation of gene expression and adaptation in Bacteria*, ed. Bruijn F. J. de (Hoboken, NJ: Wiley-Blackwell), 657–676.

Hoffmann, T., and Bremer, E. (2017). Guardians in a stressful world: the Opu family of compatible solute transporters from *Bacillus subtilis*. *Biol. Chem.* 398, 193–214. doi: 10.1515/hsz-2016-0265

Holtmann, G., Bakker, E. P., Uozumi, N., and Bremer, E. (2003). KtrAB and KtrCD: two K^+ uptake systems in *Bacillus subtilis* and their role in adaptation to hypertonicity. *J. Bacteriol.* 185, 1289–1298. doi: 10.1128/JB.185.4.1289

Hvitfeldt, E.. (2021). Paletteer: comprehensive collection of color palettes. Available at: https://github.com/EmilHvitfeldt/paletteer

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Islam, M. N., Suzauddula, M., Ahamed, Z., Rabby, M. G., Hossen, M. M., Biswas, M., et al. (2022). Phylogenetic analysis and characterization of arsenic (As) transforming bacterial marker proteins following isolation of As-tolerant indigenous bacteria. *Arch. Microbiol.* 204:660. doi: 10.1007/s00203-022-03270-5

Ito, M., Guffanti, A. A., Oudega, B., and Krulwich, T. A. (1999). Mrp, a multigene, multifunctional locus in *Bacillus subtilis* with roles in resistance to cholate and to Na^+ and in pH homeostasis. *J. Bacteriol.* 181, 2394–2402.

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

Joshi, M. S., Johnson, J. L., and Rajagopalan, K. V. (1996). Molybdenum cofactor biosynthesis in *Escherichia coli mod* and *mog* mutants. *J. Bacteriol.* 178, 4310–4312. doi: 10.1128/jb.178.14.4310-4312.1996

Jukes, T. H., and Cantor, C. R. (1969). "Evolution of protein molecules" in *Mammalian Protein Metabolism*. ed. H. Munro (New York: Academic Press), 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

Kappes, R. M., Kempf, B., and Bremer, E. (1996). Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: characterization of OpuD. *J. Bacteriol.* 178, 5071–5079. doi: 10.1128/jb.178.17.5071-5079.1996

Kassambara, A.. (2020). Ggpubr: ggplot2 based publication ready plots. Available at: https://cran.r-project.org/package=ggpubr

Kheiri, R., Mehrshad, M., Pourbabaee, A. A., Ventosa, A., and Amoozegar, M. A. (2023). Hypersaline Lake Urmia: a potential hotspot for microbial genomic variation. *Sci. Rep.* 13:374. doi: 10.1038/s41598-023-27429-2

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. (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

Koser, S. A. (1923). Utilization of the salts of organic acids by the colon-aerogenes group. *J. Bacteriol.* 8, 493–520. doi: 10.1128/jb.8.5.493-520.1923

Kosono, S., Ohashi, Y., Kawamura, F., Kitada, M., and Kudo, T. (2000). Function of a principal Na^+/H^+ antiporter, ShaA, is required for initiation of sporulation in *Bacillus subtilis*. *J. Bacteriol.* 182, 898–904. doi: 10.1128/JB.182.4.898-904.2000

Kovacs, N. (1956). Identification of *Pseudomonas pyocyanea* by the oxidase reaction. *Nature* 178:703. doi: 10.1038/178703a0

Kuhlmann, A. U., Bursy, J., Gimpel, S., Hoffmann, T., and Bremer, E. (2008a). Synthesis of the compatible solute ectoine in *Virgibacillus pantothenticus* is triggered by high salinity and low growth temperature. *Appl. Environ. Microbiol.* 74, 4560–4563. doi: 10.1128/AEM.00492-08

Kuhlmann, S. I., Terwisscha van Schelting, A. C., Bienert, R., Kunte, H. J., and Ziegler, C. M. (2008b). 1.55 Å structure of the ectoine binding protein TeaA of the osmoregulated TRAP-transporter TeaABC for *Halomonas elongata*. *Biochemistry* 47, 9475–9485. doi: 10.1021/bi8006719

Lane, D. (1991). "16S/23S rRNA sequencing" in *Nucleic acid techniques in bacterial systematics*. eds. E. Stackebrandt and M. Goodfellow (New York: Wiley), 115–175.

Lecher, J., Pittelkow, M., Zobel, S., Bursy, J., Bönig, T., Smits, S. H. J., et al. (2009). The crystal structure of UehA in complex with ectoine–a comparison with other TRAP-T binding proteins. *J. Mol. Biol.* 389, 58–73. doi: 10.1016/j.jmb.2009.03.077

Lee, J.-S., Lim, J.-M., Lee, K.-C., Lee, J.-C., Park, Y.-H., and Kim, C.-J. (2006). *Virgibacillus koreensis* sp. nov., a novel bacterium from a salt field, and transfer of *Virgibacillus picturae* to the genus *Oceanobacillus* as *Oceanobacillus picturae* comb. nov. with emended descriptions. *Int. J. Syst. Evol. Microbiol.* 56, 251–257. doi: 10.1099/ijs.0.63734-0

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

Lee, J.-C., and Whang, K.-S. (2019). *Aquibacillus sediminis* sp. nov., a moderately halophilic bacterium isolated from saltern soil. *Int. J. Syst. Evol. Microbiol.* 69, 3121–3127. doi: 10.1099/ijsem.0.003599

Legatzki, A., Grass, G., Anton, A., Rensing, C., and Nies, D. H. (2003). Interplay of the Czc system and two P-type ATPases in conferring metal resistance to *Ralstonia metallidurans*. *J. Bacteriol.* 185, 4354–4361. doi: 10.1128/JB.185.15.4354-4361.2003

Leimkühler, S. (2020). The biosynthesis of the molybdenum cofactors in *Escherichia coli*. *Environ. Microbiol.* 22, 2007–2026. doi: 10.1111/1462-2920.15003

León, M. J., Hoffmann, T., Sánchez-Porro, C., Heider, J., Ventosa, A., and Bremer, E. (2018). Compatible solute synthesis and import by the moderate halophile *Spiribacter salinus*: physiology and genomics. *Front. Microbiol.* 9:108. doi: 10.3389/fmicb.2018.00108

Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033

Li, D., Luo, R., Liu, C.-M., Leung, C.-M., Ting, H.-F., Sadakane, K., et al. (2016). MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* 102, 3–11. doi: 10.1016/j.ymeth.2016.02.020

Li, J., Zhang, B., Liu, G., Liu, Y., Yang, H., Yang, R., et al. (2020). *Radiobacillus deserti* gen. nov., sp. nov., a UV-resistant bacterium isolated from desert soil. *Int. J. Syst. Evol. Microbiol.* 70, 6338–6347. doi: 10.1099/ijsem.0.004536

Lin, W., Chai, J., Love, J., and Fu, D. (2010). Selective electrodiffusion of zinc ions in a Zrt-, Irt-like protein, ZIPB. *J. Biol. Chem.* 285, 39013–39020. doi: 10.1074/jbc.M110.180620

Logan, N. A., Berge, O., Bishop, A. H., Busse, H.-J., De Vos, P., Fritze, D., et al. (2009). Proposed minimal standards for describing new taxa of aerobic, endospore-forming bacteria. *Int. J. Syst. Evol. Microbiol.* 59, 2114–2121. doi: 10.1099/ijs.0.013649-0

Louis, P., and Galinski, E. A. (1997). Characterization of genes for the biosynthesis of the compatible solute ectoine from *Marinococcus halophilus* and osmoregulated expression in *Escherichia coli*. *Microbiology* 143, 1141–1149.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

Ma, H., Zhao, Y., Huang, W., Zhang, L., Wu, F., Ye, J., et al. (2020). Rational flux-tuning of *Halomonas bluephagenesis* for co-production of bioplastic PHB and ectoine. *Nat. Commun.* 11:3313. doi: 10.1038/s41467-020-17223-3

Magalon, A., and Mendel, R. R. (2015). Biosynthesis and insertion of the molybdenum cofactor. *EcoSal Plus* 6:2013. doi: 10.1128/ecosalplus.ESP-0006-2013

Magnuson, E., Altshuler, I., Fernández-Martínez, M. Á., Chen, Y.-J., Maggiori, C., Goordial, J., et al. (2022). Active lithoautotrophic and methane-oxidizing microbial community in an anoxic, sub-zero, and hypersaline high Arctic spring. *ISME J.* 16, 1798–1808. doi: 10.1038/s41396-022-01233-8

Mana-Capelli, S., Mandal, A. K., and Argüello, J. M. (2003). *Archaeoglobus fulgidus* CopB is a thermophilic Cu^{2+}-ATPase: functional role of its histidine-rich N-terminal metal binding domain. *J. Biol. Chem.* 278, 40534–40541. doi: 10.1074/jbc.M306907200

Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* 14:e1005944. doi: 10.1371/journal.pcbi.1005944

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

Mehrshad, M., Rodriguez-Valera, F., Amoozegar, M. A., López-García, 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

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2021). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

Meng, L., Hong, S., Liu, H., Huang, H., Sun, H., Xu, T., et al. (2014). Cloning and identification of group 1 *mrp* operon encoding a novel monovalent cation/proton antiporter system from the moderate halophile *Halomonas* zhaodongensis. *Extremophiles* 18, 963–972. doi: 10.1007/s00792-014-0666-5

MIDI (2008). *Sherlock microbial identification system operating manual, version 6.1*. Newark, DE: MIDI Inc.

Miralles-Robledillo, J. M., Torregrosa-Crespo, J., Martínez-Espinosa, R. M., and Pire, C. (2019). DMSO reductase family: phylogenetics and applications of extremophiles. *Int. J. Mol. Sci.* 20:3349. doi: 10.3390/ijms20133349

Nayfach, S., and Pollard, K. S. (2015). Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* 16:51. doi: 10.1186/s13059-015-0611-7

Neumann, M., and Leimkühler, S. (2008). Heavy metal ions inhibit molybdoenzyme activity by binding to the dithiolene moiety of molybdopterin in *Escherichia coli*. *FEBS J.* 275, 5678–5689. doi: 10.1111/j.1742-4658.2008.06694.x

Neumann, M., Seduk, F., Iobbi-Nivol, C., and Leimkühler, S. (2011). Molybdopterin dinucleotide biosynthesis in *Escherichia coli*: identification of amino acid residues of molybdopterin dinucleotide transferases that determine specificity for binding of guanine or cytosine nucleotides. *J. Biol. Chem.* 286, 1400–1408. doi: 10.1074/jbc.M110.155671

Noll, M., and Lutsenko, S. (2000). Expression of ZntA, a zinc-transporting P1-type ATPase, is specifically regulated by zinc and cadmium. *IUBMB Life* 49, 297–302. doi: 10.1080/15216540050033168

Ohyama, T., Igarashi, K., and Kobayashi, H. (1994). Physiological role of the *chaA* gene in sodium and calcium circulations at a high pH in *Escherichia coli*. *J. Bacteriol.* 176, 4311–4315. doi: 10.1128/jb.176.14.4311-4315.1994

Ono, H., Sawada, K., Khunajakr, N., Tao, T., Yamamoto, M., Hiramoto, M., et al. (1999). Characterization of biosynthetic enzymes for ectoine as a compatible solute in a moderately halophilic eubacterium, *Halomonas elongata*. *J. Bacteriol.* 181, 91–99. doi: 10.1128/jb.181.1.91-99.1999

Oren, A. (2008). Microbial life at high salt concentrations: phylogenetic and metabolic diversity. *Saline Syst.* 4:2. doi: 10.1186/1746-1448-4-2

Oren, A. (2013). Life at high salt concentrations, intracellular KCl concentrations, and acidic proteomes. *Front. Microbiol.* 4:315. doi: 10.3389/fmicb.2013.00315

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

Patel, R., Mevada, V., Prajapati, D., Dudhagara, P., Koringa, P., and Joshi, C. G. (2015). Metagenomic sequence of saline desert microbiota from wild ass sanctuary, little Rann of Kutch, Gujarat. *India. Genom. Data* 3, 137–139. doi: 10.1016/j.gdata.2015.01.003

Patiño-Ruiz, M., Ganea, C., and Călinescu, O. (2022). Prokaryotic Na^+/H^+ exchangers—transport mechanism and essential residues. *Int. J. Mol. Sci.* 23:9156. doi: 10.3390/ijms23169156

Pedrós-Alió, C. (2012). The rare bacterial biosphere. *Ann. Rev. Mar. Sci.* 4, 449–466. doi: 10.1146/annurev-marine-120710-100948

Peng, T., Xu, Y., and Zhang, Y. (2018). Comparative genomics of molybdenum utilization in prokaryotes and eukaryotes. *BMC Genomics* 19:691. doi: 10.1186/s12864-018-5068-0

Perez, M. F., Kurth, D., Farías, M. E., Soria, M. N., Castillo Villamizar, G. A., Poehlein, A., et al. (2020). First report on the plasmidome from a high-altitude lake of the Andean Puna. *Front. Microbiol.* 11:1343. doi: 10.3389/fmicb.2020.01343

Pitterle, D. M., Johnson, J. L., and Rajagopalan, K. V. (1993). *In vitro* synthesis of molybdopterin from precursor Z using purified converting factor. Role of protein-bound sulfur in formation of the dithiolene. *J. Biol. Chem.* 268, 13506–13509. doi: 10.1016/s0021-9258(19)38678-8

Plominsky, A. M., Delherbe, N., Ugalde, J. A., Allen, E. E., Blanchet, M., Ikeda, P., et al. (2014). Metagenome sequencing of the microbial community of a solar saltern crystallizer pond at Cáhuil lagoon. *Chile. Genome Announc.* 2, e01172–e01114. doi: 10.1128/genomeA.01172-14

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

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

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* 70:e102. doi: 10.1002/cpbi.102

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

Radchenko, M. V., Tanaka, K., Waditee, R., Oshimi, S., Matsuzaki, Y., Fukuhara, M., et al. (2006). Potassium/proton antiport system of *Escherichia coli*. *J. Biol. Chem.* 281, 19822–19829. doi: 10.1074/jbc.M600333200

Rensing, C., Fan, B., Sharma, R., Mitra, B., and Rosen, B. P. (2000). CopA: an *Escherichia coli* Cu(I)-translocating P-type ATPase. *Proc. Natl. Acad. Sci. U. S. A.* 97, 652–656. doi: 10.1073/pnas.97.2.652

Rensing, C., Mitra, B., and Rosen, B. P. (1997). The *zntA* gene of *Escherichia coli* encodes a Zn(II)-translocating P-type ATPase. *Proc. Natl. Acad. Sci. U. S. A.* 94, 14326–14331. doi: 10.1073/pnas.94.26.14326

Reschke, S., Sigfridsson, K. G. V., Kaufmann, P., Leidel, N., Horn, S., Gast, K., et al. (2013). Identification of a bis-molybdopterin intermediate in molybdenum cofactor biosynthesis in *Escherichia coli*. *J. Biol. Chem.* 288, 29736–29745. doi: 10.1074/jbc.M113.497453

Reshetnikov, A. S., Rozova, O. N., Trotsenko, Y. A., But, S. Y., Khmelenina, V. N., and Mustakhimov, I. I. (2020). Ectoine degradation pathway in halotolerant methylotrophs. *PLoS One* 15:e0232244. doi: 10.1371/journal.pone.0232244

Revell, L. J. (2012). Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223. doi: 10.1111/j.2041-210X.2011.00169.x

Rice, P., Longden, I., 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

Richards, L. A. (1954). "Diagnosis and improvement of saline and alkali soils" in *Agriculture handbook*. ed. L. A. Richards, vol. *60* (Washington, DC: US Department of Agriculture)

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. (2016). The Enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ. Prepr.* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1

Sainz, A., Grande, J. A., and de la Torre, M. L. (2004). Characterization of heavy metal discharge into the Ria of Huelva. *Environ. Int.* 30, 557–566. doi: 10.1016/j.envint.2003.10.013

Sainz, A., Grande, J. A., de La Torre, M. L., and Sánchez-Rodas, D. (2002). Characterisation of sequential leachate discharges of mining waste rock dumps in the Tinto and Odiel rivers. *J. Environ. Manag.* 64, 345–353. doi: 10.1006/jema.2001.0497

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

Sánchez-Porro, C., Amoozegar, M. A., Rohban, R., Hajighasemi, M., and Ventosa, A. (2009). *Thalassobacillus cyri* sp. nov., a moderately halophilic Gram-positive bacterium from a hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 59, 2565–2570. doi: 10.1099/ijs.0.010488-0

Sasser, M. (1990). Identification of bacteria by gas chromatography of cellular fatty acids. *Technical Note* #101, 1–6.

Satari, L., Guillén, A., Latorre-Pérez, A., and Porcar, M. (2021). Beyond *Archaea*: the table salt bacteriome. *Front. Microbiol.* 12:714110. doi: 10.3389/fmicb.2021.714110

Schaefer, J. K., Yagi, J., Reinfelder, J. R., Cardona, T., Ellickson, K. M., Tel-Or, S., et al. (2004). Role of the bacterial organomercury lyase (MerB) in controlling methylmercury accumulation in mercury-contaminated natural waters. *Environ. Sci. Technol.* 38, 4304–4311. doi: 10.1021/es049895w

Schoepp-Cothenet, B., van Lis, R., Philippot, P., Magalon, A., Russell, M. J., and Nitschke, W. (2012). The ineluctable requirement for the trans-iron elements molybdenum and/or tungsten in the origin of life. *Sci. Rep.* 2:263. doi: 10.1038/srep00263

Schulz, A., Stöveken, N., Binzen, I. M., Hoffmann, T., Heider, J., and Bremer, E. (2017). Feeding on compatible solutes: a substrate-induced pathway for uptake and catabolism of ectoines and its genetic control by EnuR. *Environ. Microbiol.* 19, 926–946. doi: 10.1111/1462-2920.13414

Schwibbert, K., Marin-Sanguino, A., Bagyan, I., Heidrich, G., Lentzen, G., Seitz, H., et al. (2011). A blueprint of ectoine metabolism from the genome of the industrial producer *Halomonas elongata* DSM 2581^{T}. *Environ. Microbiol.* 13, 1973–1994. doi: 10.1111/j.1462-2920.2010.02336.x

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shanmugam, K. T., Stewart, V., Gunsalus, R. P., Boxer, D. H., Cole, J. A., Chippaux, M., et al. (1992). Proposed nomenclature for the genes involved in molybdenum metabolism in *Escherichia coli* and *Salmonella typhimurium*. *Mol. Microbiol.* 6, 3452–3454. doi: 10.1111/j.1365-2958.1992.tb02215.x

Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* 16:1114. 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

Sun, H., and Shi, W. (2001). Genetic studies of *mrp*, a locus essential for cellular aggregation and sporulation of *Myxococcus xanthus*. *J. Bacteriol.* 183, 4786–4795. doi: 10.1128/JB.183.16.4786-4795.2001

Teichmann, L., Kümmel, H., Warmbold, B., and Bremer, E. (2018). OpuF, a new *Bacillus* compatible solute ABC transporter with a substrate-binding protein fused to the transmembrane domain. *Appl. Environ. Microbiol.* 84:01718. doi: 10.1128/AEM.01728-18

van Thuoc, D., Loan, T. T., Trung, T. A., van Quyen, N., Tung, Q. N., Tien, P. Q., et al. (2020). Genome mining reveals the biosynthetic pathways of polyhydroxyalkanoate and ectoines of the halophilic strain *Salinivibrio proteolyticus* M318 isolated from fermented shrimp paste. *Mar. Biotechnol.* 22, 651–660. doi: 10.1007/s10126-020-09986-z

Vandrich, J., Pfeiffer, F., Alfaro-Espinoza, G., and Kunte, H. J. (2020). Contribution of mechanosensitive channels to osmoadaptation and ectoine excretion in *Halomonas elongata*. *Extremophiles* 24, 421–432. doi: 10.1007/s00792-020-01168-y

Ventosa, A., Quesada, E., Rodriguez-Valera, F., Ruiz-Berraquero, F., and Ramos-Cormenzana, A. (1982). Numerical taxonomy of moderately halophilic Gram-negative rods. *Microbiology* 128, 1959–1968. doi: 10.1099/00221287-128-9-1959

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., Navarro-Sampedro, L., Carballo, M., and Ventosa, A. (2018). Metagenome sequencing of prokaryotic microbiota from two hypersaline soils of the Odiel Salt Marshes in Huelva, Southwestern Spain. *Genome Announc.* 6:e00140-18. doi: 10.1128/genomeA.00140-18

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

von Blohn, C., Kempf, B., Kappes, R. M., and Bremer, E. (1997). Osmostress response in *Bacillus subtilis*: characterization of a proline uptake system (OpuE) regulated by high osmolarity and the alternative transcription factor sigma B. *Mol. Microbiol.* 25, 175–187. doi: 10.1046/j.1365-2958.1997.4441809.x

Walkenhorst, H. M., Hemschemeier, S. K., and Eichenlaub, R. (1995). Molecular analysis of the molybdate uptake operon, *modABCD*, of *Escherichia coli* and *modR*, a regulatory gene. *Microbiol. Res.* 150, 347–361. doi: 10.1016/S0944-5013(11)80016-9

Wang, H. T., Xu, L., and Sun, J. Q. (2021). *Aquibacillus kalidii* sp. nov., an indole acetic acid-producing endophyte from a shoot of *Kalidium cuspidatum*, and reclassification of *Virgibacillus campisalis* Lee et al. 2012 as a later heterotypic synonym of *Virgibacillus alimentarius* Kim et al. 2011. *Int. J. Syst. Evol. Microbiol.* 71:5030. doi: 10.1099/ijsem.0.005030

Wickham, H. (2007). Reshaping data with the reshape package. *J. Stat. Softw.* 21, 1–20. Available at: http://www.jstatsoft.org/v21/i12/paper

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Wickham, H., and Seidel, D. (2020). Scales: scale functions for visualization. Available at: https://cran.r-project.org/package=scales

Wilke, C. O., and Wiernik, B. M. (2022). Ggtext: improved text rendering support for ggplot2. Available at https://cran.r-project.org/package=ggtext

Wuebbens, M. M., and Rajagopalan, K. V. (1993). Structural characterization of a molybdopterin precursor. *J. Biol. Chem.* 268, 13493–13498. doi: 10.1016/s0021-9258(19)38676-4

Xie, Y. G., Luo, Z. H., Fang, B. Z., Jiao, J. Y., Xie, Q. J., Cao, X. R., et al. (2022). Functional differentiation determines the molecular basis of the symbiotic lifestyle of Ca. *Nanohaloarchaeota*. *Microbiome* 10:172. doi: 10.1186/s40168-022-01376-y

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

Yoshinaka, T., Takasu, H., Tomizawa, R., Kosono, S., and Kudo, T. (2003). A *shaE* deletion mutant showed lower Na^{+} sensitivity compared to other deletion mutants in the *Bacillus subtilis* sodium/hydrogen antiporter (Sha) system. *J. Biosci. Bioeng.* 95, 306–309. doi: 10.1016/S1389-1723(03)80035-X

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

Yutani, H. (2021). Gghighlight: highlight lines and points in ggplot2. Available at: https://cran.r-project.org/package=gghighlight

Zhang, W.-Y., Hu, J., Zhang, X.-Q., Zhu, X.-F., and Wu, M. (2015). *Aquibacillus salifodinae* sp. nov., a novel bacterium isolated from a salt mine in Xinjiang Province, China. *Antonie van Leeuwenhoek* 107, 367–374. doi: 10.1007/s10482-014-0335-9

Zhang, Y., Rump, S., and Gladyshev, V. N. (2011). Comparative genomics and evolution of molybdenum utilization. *Coord. Chem. Rev.* 255, 1206–1217. doi: 10.1016/j.ccr.2011.02.016.Comparative

Zhang, Y. J., Zhou, Y., Ja, M., Shi, R., Chun-Yu, W. X., Yang, L. L., et al. (2012). *Virgibacillus albus* sp. nov., a novel moderately halophilic bacterium isolated from Lop Nur salt lake in Xinjiang Province, China. *Antonie van Leeuwenhoek* 102, 553–560. doi: 10.1007/s10482-012-9750-y

Zulkifli, L., Akai, M., Yoshikawa, A., Shimojima, M., Ohta, H., Guy, H. R., et al. (2010). The KtrA and KtrE subunits are required for Na^+-dependent K^+ uptake by KtrB across the plasma membrane in *Synechocystis* sp. strain PCC 6803. *J. Bacteriol.* 192, 5063–5070. doi: 10.1128/JB.00569-10

Zupok, A., Iobbi-Nivol, C., Méjean, V., and Leimkühler, S. (2019). The regulation of Moco biosynthesis and molybdoenzyme gene expression by molybdenum and iron in bacteria. *Metallomics* 11, 1602–1624. doi: 10.1039/c9mt00186g

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 July 2023
DOI 10.3389/fmicb.2023.1108245

OPEN ACCESS

EDITED BY
Mohamed Jebbar,
Université de Bretagne Occidentale, France

REVIEWED BY
Matthew Schrenk,
Michigan State University, United States
Sophie Mieszkin,
Université de Bretagne Occidentale, France

*CORRESPONDENCE
Sergey N. Gavrilov
✉ sngavrilov@gmail.com

RECEIVED 25 November 2022
ACCEPTED 26 June 2023
PUBLISHED 14 July 2023

CITATION
Zavarzina DG, Merkel AY, Klyukina AA, Elizarov IM, Pikhtereva VA, Rusakov VS, Chistyakova NI, Ziganshin RH, Maslov AA and Gavrilov SN (2023) Iron or sulfur respiration—an adaptive choice determining the fitness of a natronophilic bacterium *Dethiobacter alkaliphilus* in geochemically contrasting environments.
Front. Microbiol. 14:1108245.
doi: 10.3389/fmicb.2023.1108245

Iron or sulfur respiration—an adaptive choice determining the fitness of a natronophilic bacterium *Dethiobacter alkaliphilus* in geochemically contrasting environments

Daria G. Zavarzina[1], Alexander Yu Merkel[1], Alexandra A. Klyukina[1], Ivan M. Elizarov[1], Valeria A. Pikhtereva[1,2], Vyacheslav S. Rusakov[3], Nataliya I. Chistyakova[3], Rustam H. Ziganshin[4], Alexey A. Maslov[5] and Sergey N. Gavrilov[1]*

[1]Winogradsky Institute of Microbiology, FRC Biotechnology, Russian Academy of Sciences, Moscow, Russia, [2]Faculty of Biology, Lomonosov Moscow State University, Moscow, Russia, [3]Faculty of Physics, Lomonosov Moscow State University, Moscow, Russia, [4]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia, [5]Faculty of Geology, Lomonosov Moscow State University, Moscow, Russia

Haloalkaliphilic microorganisms are double extremophiles functioning optimally at high salinity and pH. Their typical habitats are soda lakes, geologically ancient yet widespread ecosystems supposed to harbor relict microbial communities. We compared metabolic features and their determinants in two strains of the natronophilic species *Dethiobacter alkaliphilus*, the only cultured representative of the class "*Dethiobacteria*" (*Bacillota*). The strains of *D. alkaliphilus* were previously isolated from geographically remote Mongolian and Kenyan soda lakes. The type strain AHT1^T was described as a facultative chemolithoautotrophic sulfidogen reducing or disproportionating sulfur or thiosulfate, while strain Z-1002 was isolated as a chemolithoautotrophic iron reducer. Here, we uncovered the iron reducing ability of strain AHT1^T and the ability of strain Z-1002 for thiosulfate reduction and anaerobic Fe(II) oxidation. Key catabolic processes sustaining the growth of both *D. alkaliphilus* strains appeared to fit the geochemical settings of two contrasting natural alkaline environments, sulfur-enriched soda lakes and iron-enriched serpentinites. This hypothesis was supported by a meta-analysis of *Dethiobacterial* genomes and by the enrichment of a novel phylotype from a subsurface alkaline aquifer under Fe(III)-reducing conditions. Genome analysis revealed multiheme *c*-type cytochromes to be the most probable determinants of iron and sulfur redox transformations in *D. alkaliphilus*. Phylogeny reconstruction showed that all the respiratory processes in this organism are likely provided by evolutionarily related early forms of unconventional octaheme tetrathionate and sulfite reductases and their structural analogs, OmhA/OcwA Fe(III)-reductases. Several phylogenetically related determinants of anaerobic Fe(II) oxidation were identified in the Z-1002 genome, and the oxidation process was experimentally demonstrated. Proteomic profiling revealed two distinct sets of multiheme cytochromes upregulated in iron(III)- or thiosulfate-respiring cells and the cytochromes peculiar for Fe(II) oxidizing cells. We suggest that maintaining high variation in multiheme cytochromes is an effective adaptive strategy to occupy geochemically contrasting alkaline environments. We propose that sulfur-enriched soda lakes could be secondary habitats for *D. alkaliphilus* compared to

Fe-rich serpentinites, and that the ongoing evolution of *Dethiobacterales* could retrace the evolutionary path that may have occurred in prokaryotes at a turning point in the biosphere's history, when the intensification of the sulfur cycle outweighed the global significance of the iron cycle.

KEYWORDS

***Dethiobacter alkaliphilus*, soda lakes, serpentinites, multiheme cytochromes, sulfur/thiosulfate reduction, microbial iron cycling**

Introduction

Modern soda lakes, defined by high salinity and pH, occur worldwide and are thought to harbor relict microbial communities (Zavarzin, 1993; Zhilina and Zavarzin, 1994). Despite several extreme parameters, soda lakes are characterized by high productivity and contain fully functional and diverse haloalkaliphilic microbial communities that drive the biogeochemical cycles of carbon, nitrogen, and sulfur (Jones et al., 1998; Zavarzin et al., 1999; Sorokin et al., 2011, 2014a). In modern soda lakes, the sulfur cycle is one of the most active biogeochemical processes. Redox transformations of inorganic sulfur compounds are energetically efficient enough for microorganisms to cope with costly life under polyextreme conditions. An important phenomenon observed in the sediments of Kulunda soda lakes (Altai region, Russia) is that sulfidogens using elemental sulfur and thiosulfate as electron acceptors are more active there than sulfate reducers despite the higher abundance of sulfates in the ecosystem (Sorokin et al., 2011). Inspite of the predominance of the sulfur cycle, several species and genera of iron reducers have been isolated from soda lakes and their wide distribution in these ecosystems has been further reported (Zavarzina et al., 2016). Importantly, most dissimilatory alkaliphilic iron reducers are capable of using sulfur as an alternative electron acceptor, just as many alkaliphilic sulfidogens have been shown to reduce Fe(III) (Zavarzina et al., 2018).

Dethiobacter alkaliphilus, the only cultured representative of the genus *Dethiobacter*, was isolated from mixed anaerobic sediments of northeastern Mongolian soda lakes, and was described as a sulfur- and thiosulfate-reducing, facultatively chemolithoautotrophic sulfidogen utilizing H_2 and a range of organic electron donors, strain AHT1^T (Sorokin et al., 2008). The organism is also capable of sulfur or thiosulfate disproportionation (Poser et al., 2013, 2016). The 16S rRNA gene-based phylogenetic reconstruction placed the first isolate of *D. alkaliphilus* within a deep lineage of *Firmicutes* (now *Bacillota*) related to the order *Syntrophomonadales* and syntrophic acetate-oxidizing haloalkaliphiles from soda lakes, with *Ca.* Contubernalis alkalaceticus (Zhilina et al., 2005) and *Ca.* Synthrophonatronum acetioxidans (Sorokin et al., 2014b) as the closest relatives. Further phylogenomic reconstruction based on 120 single-copy conservative markers (Parks et al., 2018) rooted *D. alkaliphilus* as a separate class "*Dethiobacteria*," order "*Dethiobacterales*," and family "*Dethiobacteraceae*" (Sorokin and Merkel, 2019). Strain Z-1002 belonging to the same species, was isolated from the sediments of a hypersaline soda lake Magadi in Kenya (Zavarzina et al., 2018) on the selective medium containing synthesized ferrihydrite (SF) as the sole electron acceptor and formate or molecular hydrogen as the energy source. Thus, the two cultured strains AHT1^T and Z-1002 of the single species of the class "*Dethiobacteria*" were isolated as sulfur/thiosulfate- and iron-reducers, respectively, from the sediments of soda lakes located on two different continents.

Interestingly, *Dethiobacter*-related phylotypes have been previously detected exclusively in alkaline environments (pH ≥ 7.5) of two different types, the sediments of soda or meromictic lakes, and subsurface ecosystems affected by serpentinization processes. Representatives of *Dethiobacter* genus were detected in the sediments of a meromictic Mahoney lake, in soda lakes Mono, Van and those of the Kulunda Steppe (Hamilton et al., 2016; Edwardson and Hollibaugh, 2017, 2018; Vavourakis et al., 2018; Ersoy Omeroglu et al., 2021), as well as in several serpentinizing ophiolites, serpentinization-based subseafloor ecosystems, and terrestrial subsurface environments (Brazelton et al., 2013; Suko et al., 2013; Suzuki et al., 2013; Tiago and Veríssimo, 2013; Crespo-Medina et al., 2014; Glaring et al., 2015; López-López et al., 2015; Postec et al., 2015; Woycheese et al., 2015; Purkamo et al., 2016, 2017; Pisapia et al., 2017; Sabuda et al., 2020, 2021). Furthermore, *Dethiobacter* was highly enriched in microcosms with serpentinizing groundwater in the absence of any external nutrients (Crespo-Medina et al., 2014; Purkamo et al., 2017). Serpentinization is a widespread geochemical phenomenon involving the aqueous alteration of ultramafic rocks, containing the Fe-rich mineral olivine [$(Mg,Fe)_2SiO_4$], to serpentine [$(Mg,Fe,)_{2-3}(Si,Fe)_2O_5(OH)_4$]. The process results in the formation of a highly alkaline (pH > 10) and strongly reducing, H_2-rich fluid, which creates thermodynamically favorable environmental conditions for abiotic organic synthesis from slab-derived inorganic carbon (McCollom and Bach, 2009). Serpentinization occurs in numerous settings on Earth, including subduction zones, mid-ocean ridges, and ophiolites and is rooted far back in the Earth's geologic history, potentially contributing to the origin and early evolution of life (Russell et al., 2010; Schrenk et al., 2013). In this view, the ecological fitness of *Dethiobacter* correlates with the putative conditions of some early Earth's ecosystems. Accordingly, deeper understanding of the phenotype and genotype of these bacteria would broaden our knowledge about the metabolic pathways that might play a crucial role in the biogeochemical cycling of elements on the early Earth, in the period of the Great Oxidation Event (2.4–1.8 Ga).

The aim of our work was to fill the knowledge gap on the metabolic pathways driving the respiration of sulfur and iron compounds in *D. alkaliphilus* strains. We also aimed to propose how the interrelations between the redox transformations of iron and sulfur compounds within a single species might affect its adaptation to geochemically contrasting extreme environments. To achieve our goals, we used a combination of classical microbiological methods with multi-omics analysis. We also

collected the data on the geology of the Lake Magadi and the Fe-depleted subsurface Yessentuki aquifer to assess the significance of geochemical factors in the evolution of *D. alkaliphilus* energy metabolism.

Materials and methods

Strains and cultivation conditions

The strain AHT1^T isolated from a mixed anaerobic sediments of northeastern Mongolian soda lakes (Sorokin et al., 2008) was kindly provided for the study by Dr. D. Yu. Sorokin. The ability of strain AHT1^T for dissimilatory iron-reduction was tested in an optimal liquid medium of the following composition (per liter distilled water): NH_4Cl 0.5 g; KH_2PO_4, 0.2 g; $MgCl_2 \times 6H_2O$ 0.1 g; $CaCl_2 \times 2H_2O$ 0.02 g; KCl 0.2 g; NaCl 3.0 g; Na_2CO_3 15.0 g; $NaHCO_3$ 20.0 g; $Na_2S \times 9H_2O$ 0.1 g; 1 mL trace element solution (Kevbrin and Zavarzin, 1992); 1 mL vitamin solution (Wolin et al., 1963); yeast extract 0.2 g. Molecular hydrogen (10% v/v in the gas phase) was used as the electron donor, and SF prepared as previously described (Zavarzina et al., 2006) was used as the electron acceptor. SF was added to the culture medium prior to sterilization to achieve a final Fe(III) content of 50 mM. The medium was prepared under pure N_2 flow. Afterwards, $NaHCO_3$, vitamins, $Na_2S \times 9H_2O$, and SF were added. The medium was dispensed in 20 mL aliquots into 50 mL flasks and the headspace was filled with pure N_2. The medium was autoclaved at 1 atm, 121°C for 20 min. The pH of the sterile medium was 9.5. After three successful transfers to the optimal medium, sodium sulfide and yeast extract were omitted for further cultivation steps in order to test the ability of strain AHT1^T to reduce SF lithoautotrophically in the absence of any reducing agents.

Strain Z-1002 was isolated in 2010 from the sediments of a hypersaline soda lake Magadi (Kenya). Anaerobic sediment samples were collected by Prof. G.A. Zavarzin in 1996 in pre-sterilized flasks, sealed on-site and stored since then at +4°C. The ability of strain Z-1002 to reduce SF during autotrophic growth with molecular hydrogen or formate in the presence of vitamin solution was previously reported (Zavarzina et al., 2018). Here we focused on determining its optimal growth conditions, the ability to reduce sulfur compounds and capacity for chemoorganotrophic growth.

The basal medium (BM) used for physiological tests of both strains contained (per liter of distilled water): NH_4Cl 0.5 g; KH_2PO_4 0.2 g; $MgCl_2 \times 6H_2O$ 0.1 g; $CaCl_2 \times 2H_2O$ 0.02 g; KCl 0.2 g; $Na_2S \cdot 9H_2O$ 0.1 g, 1 mL trace elements solution (Kevbrin and Zavarzin, 1992), 1 mL vitamins solution (Wolin et al., 1963). Sodium formate (1 g l^{-1}) and SF [final Fe(III) content of 50 mM] were added to test the Fe(III) reducing activity under optimal growth conditions. The pH growth range of strain Z-1002 was determined on BM medium supplemented with 60 g l^{-1} NaCl and 10 g L^{-1} $NaHCO_3$. pH lower than 8.0 was adjusted with 6 M HCl, in the range between 8.0 and 9.5—by titrating with 10% Na_2CO_3 solution, in the range between 9.5 and 10.3—with 12 M NaOH solution. In this experiment, three tubes with culture media were prepared for each pH value: two tubes for cultivation and one for measuring the initial pH value of a medium after autoclaving. The pH stability of the cultures was checked at the end of the incubation.

The carbonate/bicarbonate growth optimum was determined on BM medium supplemented with 20 g L^{-1} NaCl, 55 g L^{-1} $NaHCO_3$, and 95 g L^{-1} Na_2CO_3. This modified medium was gradually diluted with another one—BM containing 20 g L^{-1} NaCl but without carbonate/bicarbonate. In the latter medium, the optimal pH value for growth was maintained by the addition of 20 mM N-cyclohexyl-2-aminoethanesulfonic acid (CHES) buffer (Merck). Three subsequent transfers were performed to confirm the ability of strain Z-1002 to grow without carbonates.

The sodium chloride dependence of strain Z-1002 was determined on BM medium supplemented with 20 g L^{-1} $NaHCO_3$ and 40 g L^{-1} Na_2CO_3. NaCl was added to this medium at the concentrations of (g L^{-1}): 0, 3, 5, 10, and up to 140 in increments of 10. Three subsequent transfers were performed to confirm the ability of strain Z-1002 to grow without chloride.

In all the experiments, soluble electron donors and acceptors were added from sterile anaerobic stock solutions prior to inoculation. All the organic substrates (peptides, carbohydrates, alcohols and organic acids) were filter-sterilized using 0.2 µm pore size syringe filters (Millipore) and added to a final concentration of 3 g L^{-1}. The capability for dissimilatory reduction of sulfur compounds (S^0 (1% w/v); $S_2O_3^{2-}$ (10 mM); SO_4^{2-} (20 mM)) was tested with formate (1 g L^{-1}), H_2 (10% v/v in the gas phase), ethanol or organic acids (10 mM in each case) added as the electron donors. The ability to disproportionate sulfur or thiosulfate was tested without the addition of any of the electron donors.

The ability to oxidize a natural siderite-based mineral mixture (Fe_2CO_3 as the major component) under anaerobic conditions was tested using an optimized medium without sodium sulfide with the following composition (per liter distilled water): NH_4Cl 0.5 g; KH_2PO_4 0.2 g; $MgCl_2 \times 6H_2O$ 0.1 g; $CaCl_2 \times 2H_2O$ 0.02 g; KCl 0.2 g; NaCl 50.0 g; Na_2CO_3 40.0 g; $NaHCO_3$ 20.0 g; 1 mL trace mineral solution (Kevbrin and Zavarzin, 1992); 1 mL vitamin solution (Wolin et al., 1963). This medium was dispensed under N_2 flow in 10 mL aliquots into Hungate tubes containing 0.1 g dry mineral powder each. Uniform grains of the siderite-based mineral mixture of hydrothermal origin (Bakal deposit, the Urals, Russia) were preliminarily selected and ground to powder (<100 µm particle size) in an agate mortar. The presence of the Fe(III)-containing green rust (10.6%) and iron oxides (13%) in the mineral mixture was detected by Mössbauer spectroscopy (Table 1). The siderite-containing medium was autoclaved at 1 atm, 121°C for 20 min. Three subsequent transfers were performed to confirm the ability of strain Z-1002 to grow with siderite by Fe(II) oxidation.

All the cultivation experiments were performed in duplicate using Hungate tubes.

Sampling of Yessentuki mineral water and enrichment of a novel *Dethiobacter* phylotype

Subsurface water from the Yessentuki Mineral Water Deposit (YMWD, Stavropol Krai, Russia) was sampled in September, 2020, from Well 9 extracting Na-Ca-HCO_3-SO_4-type mineral water from the Lower Cretaceous (K_2s-m) aquifer. This well (E 42°48′10″ N 44°2′30″) is 600 m deep and has open boreholes in the interval of 485–556 m. The temperature, pH value, and salinity of the mineral water at the wellhead during the sampling were 21.9°C, 7.9, and 0.4 g L^{-1}, respectively. One hundred liter of water was filtered through a 0.2 µm pore size track membrane filters (JINR, Dubna, Russia) under the natural overpressure of the well, as described previously (Gavrilov et al., 2022). The filter was cut in half and one of its parts was used for DNA extraction and phylogenetic profiling of the natural water microbial community according to a previously described procedure (Gavrilov et al., 2022).

10.3389/fmicb.2023.1108245

TABLE 1 Hyperfine parameters of room temperature ^{57}Fe Mössbauer spectra of the siderite-based mineral mixture incubated with the growing culture of strain Z-1002 and in a sterile control medium.

Subspectrum	Phase	*I*, %	δ, mm/s	ε, mm/s	Γ, mm/s				
Siderite-based mixture from uninoculated control									
D_1	Siderite	76.3 ± 2.0	1.229 ± 0.001	0.884 ± 0.001	0.241 ± 0.002				
D_2	Green rust (Fe^{2+})	8.4 ± 2.1	1.17 ± 0.03	1.02 ± 0.05	0.64 + 0.09				
D_3	Green rust (Fe^{3+})	2.2 ± 0.5	0.37 ± 0.04	0.34 ± 0.03	0.39 ± 0.06				
		I, %	δ_{max}, mm/s	ε_{max}, mm/s	B_{max}, T	δ_{av}, mm/s	ε_{av}, mm/s	B_{av}, T	Γ, mm/s
S	Iron oxides	13.0 ± 0.4	0.379 ± 0.016	−0.092 ± 0.017	49.0 ± 0.2	0.385 ± 0.012	−0.073 ± 0.011	46.0 ± 0.3	0.21 ± 0.06
Siderite-based mixture from the grown culture of strain Z-1002									
Subspectrum	Phase	*I*, %	δ, mm/s	ε, mm/s	Γ, mm/s				
D_1	Siderite	76.1 ± 0.7	1.228 ± 0.001	0.883 ± 0.003	0.267 ± 0.001				
D_2	Green rust (Fe^{2+})	2.1 ± 0.6	1.226 ± 0.130	1.209 ± 0.120	0.43 ± 0.09				
D_3	Green rust (Fe^{3+})	4.6 ± 0.4	0.36 ± 0.08	0.34 ± 0.09	0.50 ± 0.05				
		I, %	δ_{max}, mm/s	ε_{max}, mm/s	B_{max}, T	δ_{av}, mm/s	ε_{av}, mm/s	B_{av}, T	Γ, mm/s
S	Iron oxides	17.2 ± 0.4	0.369 ± 0.012	−0.097 ± 0.012	49.3 ± 0.1	0.374 ± 0.008	−0.092 ± 0.008	46.5 ± 0.2	0.23 ± 0.03

I, relative intensity of a subspectrum; δ, Mössbauer line shift relatively α-Fe; ε, quadrupole shift; Γ, line width; δ_{max}, δ_{av}, maximum and average values of Mössbauer line shift; ε_{max}, ε_{av}, maximum and average values of a quadrupole shift; B_{max}, B_{av}, maximum and average values of hyperfine magnetic field.

Another part of the filter was placed in a pre-sterilized 100 mL flask, filled with mineral water (80 mL) and CO_2 gas headspace (20 mL), and was stored at +4°C for 1.5-years. After storage, the flask was supplemented with SF [final Fe(III) content of 10 mM], sodium sulfate, sodium acetate, and sodium formate (10 mM each), and used as a primary enrichment of sulfate and iron reducing microorganisms inhabiting the mineral water. It was incubated in the dark at 35°C for 3 weeks. Subsamples for DNA extraction and phylogenetic profiling were taken from this culture at the beginning and after the end of the incubation. Further transfers from this primary enrichment were made to the medium with the following composition (per liter distilled water): NH_4Cl 0.015 g; KH_2PO_4 0.115 g; $MgCl_2 \times 6H_2O$ 0.6 g; $CaCl_2 \times 2H_2O$ 0.1 g; $NaHCO_3$ 0.3 g; $Na_2S \cdot 9H_2O$ 0.1 g, 1 mL trace element solution (Kevbrin and Zavarzin, 1992); 1 mL vitamin solution (Wolin et al., 1963). Sodium formate (1 g L^{-1}) or sodium acetate (1 g L^{-1}) were used as the electron donors and SF (final Fe(III) content of 50 mM) as the electron acceptor. The medium for the enrichments was prepared by boiling and cooling it under pure N_2 flow, then $NaHCO_3$, vitamins, $Na_2S \cdot 9H_2O$, and SF were added, and the pH was adjusted to 8.0 with 2.5 M NaOH solution. The medium was dispensed in 10 mL aliquots into 16 mL Hungate tubes and autoclaved at 1 atm, 121°C for 20 min.

Phenotypic characterization of the strains

Growth of *D. alkaliphilus* strains was monitored by direct cell counting using an Axio Lab.A1 phase-contrast and fluorescent microscope (Zeiss, Germany). Subsamples of SF-grown cultures were pre-stained with acridine orange dye for DNA.

Molecular hydrogen consumption, the formation of gaseous metabolites, sulfide concentration, and Fe(II) production were monitored as previously described (Khomyakova et al., 2022). Acetate was analyzed using the same chromatograph with FID detector and an Optima FFAPplus 0.25 μm × 0.32 mm × 30 m capillary column (Macherey-Nagel) with argon as the carrier gas. Separation was carried out with temperature programming. Samples for gas–liquid chromatography (0.2 mL each) were pre-treated by centrifugation at 12,600 g for 2 min, followed by stepwise acidification of the clear supernatants with H_3PO_4 and 5 M formic acid to the pH of 2.0. The detection limit of the method was 0.2 mM.

Mössbauer spectroscopy

The Fe^{2+}/Fe^{3+} ratio of minerals was determined by ^{57}Fe Mössbauer spectroscopy. This method allows for the determination and quantification of different atomic environments, magnetic states, chemical states and transformations of iron-containing compounds (Kamnev and Tugarova, 2021). All Mössbauer measurements were performed at room temperature using the MS-1101 Em spectrometer, operating in the constant acceleration mode, with a ^{57}Co source in the Rh matrix. The calibration was carried out relating to α-Fe, and the spectra were fitted using SpectrRelax software. To process the spectra, a model fitting was carried out simultaneously with the extraction of the hyperfine magnetic field distribution.

DNA extraction and amplicon sequencing

The composition of the microbial community of Yessentuki mineral water and of the enrichments was determined by amplification of the hypervariable V4 region of the 16S rRNA genes, followed by sequencing and bioinformatic data processing. Total DNA was isolated with the FastDNA™ Spin Kit for Soil DNA Extraction (MP Biomedicals, United States) according to the manufacturer's instructions using the FastPrep-24™ 5G Bead Beating System (MP Bio, United States). Amplicon libraries were prepared as described by Gohl et al. (2016) using the following primers: 515F (5′-GTG BCA GCM GCC GCG GTA A-3′) and 806R (5′-GAC TAC NVG GGT MTC TAA TCC-3′) (Hugerth et al., 2014; Merkel et al., 2019) including Illumina technical sequences (Fadrosh et al., 2014). High-throughput sequencing of the libraries was performed using MiSeq

Reagent Micro Kit v2 (300-cycles) MS-103-1,002 (Illumina, United States) on a MiSeq sequencer (Illumina, United States) according to the manufacturer's instructions. Libraries were prepared and sequenced in two replicates for each sample. Amplicon sequence variants (ASV) were obtained using the Dada2 script (Callahan et al., 2016) and the SILVA 138.1 database (Quast et al., 2013). Analysis of the ASV tables was performed using Rhea (Lagkouvardos et al., 2017). All sequencing data were deposited in SRA (NCBI) under BioProject ID PRJNA945437.

Genome sequencing and analysis

A WGS library preparation and sequencing of *Dethiobacter alkaliphilus* Z-1002 was performed in BioSpark Ltd., Moscow, Russia, using KAPA HyperPlus Library Preparation Kit (KAPA Biosystems, United Kingdom) according to the manufacturer's protocol and NovaSeq 6,000 system (Illumina, San Diego, CA, United States) with a reagent kit capable of read 100 nucleotides from each end. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAPDNO000000000. Gene search, annotation, and genome-based phylogenetic reconstructions were performed as previously described (Khomyakova et al., 2022) with an additional use of PGAP service (Li et al., 2021).

Screening of the genomes of both *D. alkaliphilus* strains for multiheme cytochromes and their sequence analysis was performed as previously described (Toshchakov et al., 2018) using reported cytochromes, involved in extracellular electron transfer (EET) in *Geobacter sulfurreducens*, *Shewanella oneidensis*, "*Thermincola potens*," *Carboxydothermus ferrireducens* (Gavrilov et al., 2021 and references therein), the dataset of cytochrome query sequences was supplemented with multiheme proteins reported to be involved in Fe(III) respiration or other EET processes in the thermophilic Fe(III) reducers *Carboxydocella thermautotrophica* (Toshchakov et al., 2018) and *Melioribacter roseus* (Gavrilov et al., 2017), anaerobic Fe(II) oxidizing bacteria *Sideroxydans lithotrophicus*, *Gallionella capsiferriformans*, and *Dechloromonas aromatica* (Chakraborty et al., 2005; Emerson et al., 2007; Liu et al., 2012), Fe(III) reducing and syntrophic archaea (Mardanov et al., 2015; Smith et al., 2015; Krukenberg et al., 2018; Kashyap and Holden, 2021). Heme-binding motifs in *D. alkaliphilus* multihemes were predicted as previously described (Mardanov et al., 2015). Conservative domains, transmembrane helices and signal peptides were predicted using the hmmscan[1] web-service (Potter et al., 2018) with default parameters and all databases included.

Phylogenetic analysis of cytochromes

The cytochrome *c* protein sequences of SirA octaheme sulfite reductase from *S. oneidensis* and the MtoA Fe(II)-oxidizing decaheme of *D. aromatica* were retrieved from the non-redundant NCBI protein database on November 2022. Homologs for each of these sequences were separately screened for using the BLASTp algorithm and the obtained sets of sequences were manually curated and processed as described previously (Gavrilov et al., 2021). The resulting sets of 48 amino acid sequences for the SirA query and 342 sequences for MtoA query were amended with the sequences of SirA, or the MtoA homologs from the genomes of *D. alkaliphilus* strains (Supplementary Table S1). The immunoglobulin-like domains were omitted from the sequences of *D. alkaliphilus* to decrease non-specific alignments within the protein sets. The sets were then aligned using MAFFT 7.490 with default parameters, with 1,000 iterations of the FFT-NS-i refinement method (Katoh et al., 2019). Two final alignments (for SirA and MtoA hits) were subjected to a Bayesian inference and used to construct unrooted phylogenetic consensus trees as described previously (Gavrilov et al., 2021).

Shotgun proteomic analysis

For proteomic analysis, biomass of *D. alkaliphilus* strain AHT1^T grown with H_2 and thiosulfate or SF as the electron acceptors was harvested from 100 mL cultures by centrifugation at 16,000 g for 15 min. Both cultures were grown in triplicate. Before harvesting, the SF-grown culture was separated into liquid and mineral phases using low speed centrifugation (5 min, 1,000 g). The mineral phase was ultrasonicated inside the centrifuge bags in an ultrasonic bath (Sapphire, Russia) at 99% power, 5 min, to separate adherent cells from the minerals. The ultrasonicated mineral phase was then combined with the liquid phase, mixed thoroughly, the bulk of the minerals were held in the bags with a hand magnet and the remaining suspension was transferred to another centrifuge bag to harvest the biomass at 13,000 g for 20 min.

Cell lysis, reduction, alkylation and digestion of the proteins, as well as the processing and LC-MS/MS analysis of the obtained peptide sets were performed as previously described (Gavrilov et al., 2021). Label-free protein quantification was made by MaxQuant software version 1.5.6.5 using *D. alkaliphilus* strain AHT[1] amino acid FASTA dataset and a common contaminant database through the Andromeda search engine according to the previously described protocol (Gavrilov et al., 2021). The iBAQ algorithm (Schwanhäusser et al., 2011) implemented in MaxQuant software was used to quantify proteins in each sample. Normalization of each protein's iBAQ value to the sum of all iBAQ values generated a relative iBAQ (riBAQ) values corresponding to the molar percentage of each protein in the sample, with the entire set of proteins in the sample taken as 100% (Shin et al., 2013).

The mass spectrometry proteomics data were deposited at the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2022) partner repository with the dataset identifier PXD040929.

Analysis of environmental distribution of *Dethiobacter* genus and *Dethiobacteraceae* family

We used 16S rRNA gene sequences and related metadata to analyze the relative abundance of *Dethiobacters* in different environments. For this purpose, we took all sequences assigned to the *Dethiobacter* genus in the SILVA database 138.1 (Quast et al., 2013). Next, we analyzed the similarity of these sequences to the 16S rRNA gene sequences of *D. alkaliphilus* strains AHT1^T and Z-1002 using BLASTN (Altschul et al., 1990), discarding all sequences with less than

1 https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan

94.5% similarity to these cultured strains of the genus. Finally, to analyze the distribution of the genus *Dethiobacter* in different types of ecotopes, we obtained 108 sequences of 16S rRNA genes. Next, we analyzed the metadata related to the detection sources of these sequences.

To analyze the distribution of the representatives of the *Dethiobacteraceae* family (Sorokin and Merkel, 2019), we used the metadata on MAGs from the GTDB 207 database (Parks et al., 2022), i.e., the information on isolation sources, description of sampling sites, geological or geochemical information on them if publicly available.

Results

General phenotypic characteristics of strain Z-1002

The Fe(III)-reducing strain of *D. alkaliphilus*, Z-1002, was previously isolated in a pure culture with SF and formate under highly alkaline conditions (Zavarzina et al., 2018). Here, we describe its metabolic characteristics in detail. Strain Z-1002 appeared to be an obligate alkaliphile with the pH growth range from 7.8 to 10.1 and an optimum at pH 9.2. It is an obligate natronophile growing in the range of carbonate/bicarbonate concentration ratios from 13.5/7.5 to 82.0/45.0 g L^{-1} with an optimum at 40.0/20.0 g L^{-1}. It could not grow without carbonates in the presence of CHES buffer after the second transfer from a carbonaceous medium. It was halotolerant and grew in the range of NaCl concentrations from 0 to 120 g L^{-1} with an optimum at 50–60 g L^{-1}. Strain Z-1002 grew chemoorganotrophically with lactate, succinate, butyrate, pyruvate, propionate, or ethanol as the electron donors and SF or thiosulfate as the electron acceptors. Slow growth with a low cell yield of 5×10^5 cells mL^{-1}, accompanied by the production of 2.5 mM sulfide, was observed on thiosulfate with the addition of formate or H_2 as electron donor. The most intense thiosulfate reduction accompanied by the formation of 15 mM sulfide was observed with ethanol as the electron donor. Strain Z-1002 was unable to use elemental sulfur and sulfate as the electron acceptors, as well as it was unable to disproportionate sulfur or thiosulfate. Strain Z-1002 was unable to grow by fermentation of carbohydrates, peptides, and amino acids.

Fe(III) reducing ability of the type strain AHT1^T

Considering the metabolic features of *D. alkaliphilus* isolate from the lake Magadi, we tested the ability of the type strain AHT1^T for dissimilatory iron reduction. After three sequential transfers on the medium supplemented with SF and H_2 in the absence of yeast extract and sodium sulfide, strain AHT1^T produced 4.1 ± 0.5 mM Fe(II) from ferrihydrite with concomitant oxidation of 1.6 ± 0.2 mM molecular hydrogen (Figure 1). The mineral phase changed from brownish to dark brown color during the growth, indicating the formation of Fe(II)-containing minerals (Supplementary Figures S1). The cell yield under these conditions was $3.5 \pm 2 \times 10^7$ cell mL^{-1}. Ferrihydrite-grown cells of both *D. alkaliphilus* strains were strongly associated with iron minerals (Figure 2).

Experimental verification of the ability of strain Z-1002 for anaerobic Fe(II) oxidation

Cultures of the strain Z-1002 appeared to oxidize Fe(II) from a natural mixture of hydrothermal siderite (76%), green rust (11%), and iron oxides (13%) under anaerobic conditions in the absence of any organic compounds. The maximum cell yield under these growth conditions was 10^7 cells mL^{-1} and was observed by the 34th hour of incubation (Supplementary Figure S2). Growth was accompanied by the changes in the mineral phase as no soluble Fe(III) or Fe(II) species were detected in the culture during the incubation. Mössbauer analysis of the minerals from the grown cultures (stationary growth phase) and uninoculated controls revealed a decrease in the relative total intensity of the spectra corresponding to ferrous iron atoms from 87.7 ± 2.9% to 78.2 ± 0.9% in the grown cultures (Table 1). No such changes were observed in the controls. The relative intensity of siderite subspectrum of the microbially impacted mineral sample remained virtually unchanged, but the intensity of the doublet corresponding to Fe^{2+} atoms in the green rust structure decreased from 8.4 ± 2.1% to 2.1 ± 0.6%, while the intensity of the doublet corresponding to Fe^{3+} atoms in this phase increased from 2.2 ± 0.5% to 4.6 ± 0.4% (Table 1 and Supplementary Figure S3). This means that 75% of the bivalent iron atoms contained in the green rust were oxidized. At the end of the incubation, we also observed an increase of 4.2% in the magnetically ordered phase content of the mineral mixture. Thus, on the one hand, we have observed a rearrangement of the bivalent and trivalent iron atoms in the green rust, and on the other hand, an increase in the amount of iron oxides in the mixture.

Genome analysis of *Dethiobacter alkaliphilus* strains

Genomes statistics

The genome of the type strain AHT1^T was previously sequenced and annotated (Melton et al., 2017). Here we have sequenced and analyzed the genome of strain Z-1002 and compared its genomic

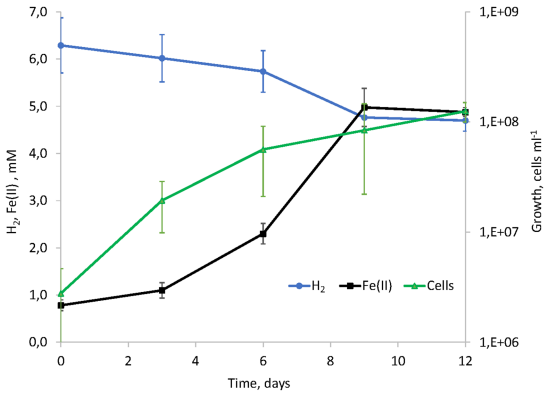


FIGURE 1
Kinetics of hydrogenotrophic growth of strain AHT1^T with ferrihydrite as the electron acceptor, hydrogen consumption and Fe(II) production in the absence of yeast extract and sodium sulfide.

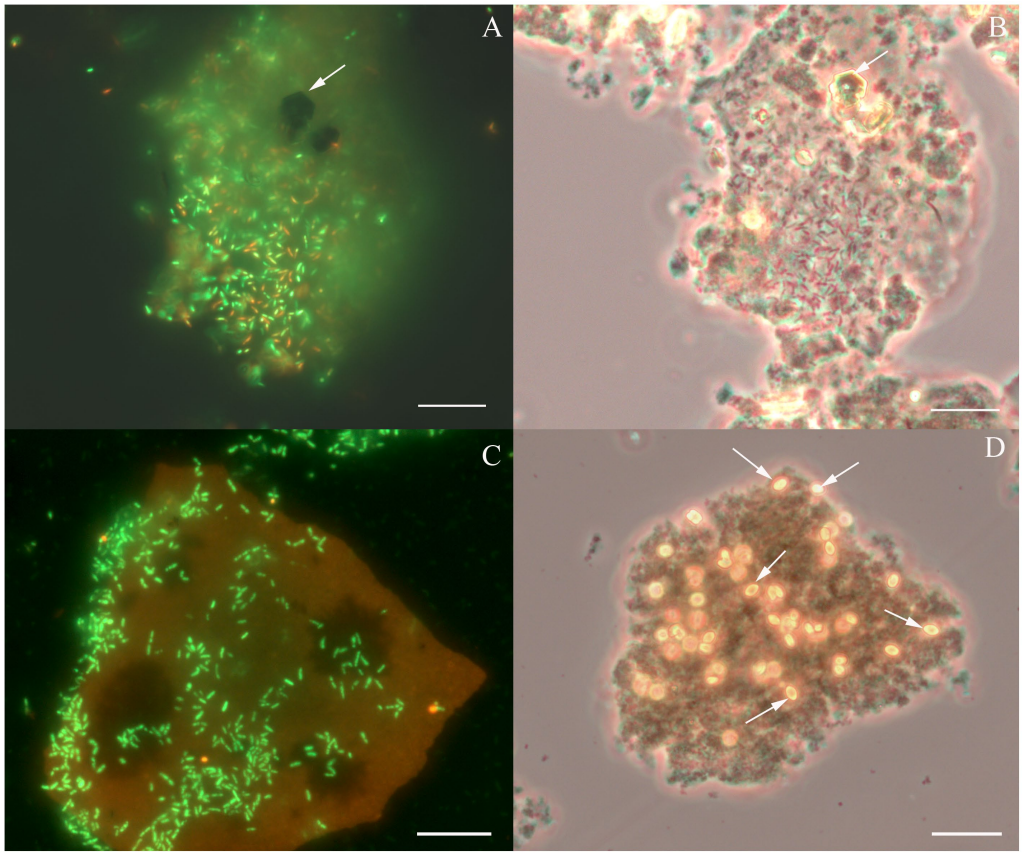


FIGURE 2
Cellular morphology of *Dethiobacter alkaliphilus* strains. **(A)** Fluorescent micrograph of acridine orange stained culture of strain Z-1002 colonizing a synthesized ferrihydrite particle during the growth with molecular hydrogen, *white arrow* indicates a newly formed magnetite crystal. **(B)** Phase contrast micrograph of the same culture (same spot), *white arrow* indicates a newly formed magnetite crystal. **(C)** Fluorescent micrograph of acridine orange stained culture of strain AHT1^T colonizing a synthesized ferrihydrite particle during the growth with molecular hydrogen. **(D)** Phase contrast micrograph of the same culture and the same spot, *white arrows* indicate newly formed siderite crystal. Each bar = 10 μm.

determinants of central carbon and energy metabolism with the type strain, as well as clarified the phylogenomic position of the genus *Dethiobacter*. Detailed statistics of both *D. alkaliphilus* genomes can be found in Supplementary Table S2.

Phylogenomic position of the genus *Dethiobacter*

Initially, the description of the phylogenetic position of the genus *Dethiobacter* was based only on the analysis of the 16S rRNA gene, which showed that this genus represents a deep phylogenetic lineage within the *Firmicutes* (*Bacillota*) phylum. Based on a modern phylogenomic reconstruction approach using 120 single-copy conserved marker genes (Parks et al., 2022), *D. alkaliphilus* has been classified as an individual class of "*Dethiobacteria*" (Sorokin and Merkel, 2022). Apart from *D. alkaliphilus*, this class includes only metagenome-assembled genomes (MAGs). According to the results of our phylogenomic analysis based on the same method, the MAGs of the *Dethiobacteraceae* family appeared to form two distinct phylogenetic clusters: one containing *D. alkaliphilus* and the MAGs from serpentinizing environments, and another containing MAGs from anaerobic digestion facilities and organic wastes (Figure 3).

Central metabolism of *Dethiobacter alkaliphilus*

As expected, both strains of *D. alkaliphilus* possess similar gene sets determining central metabolic processes. These are identical clusters of genes encoding [NiFe] uptake hydrogenase (HydABC) and its assembly proteins HypDEF, complete gene sets of V-type Na^+-ATP synthetase, and the Wood–Ljungdahl pathway for autotrophic CO_2 fixation, which includes an operon comprising a CODH gene *cooS* followed by *acsABCDE* genes of acetyl-CoA-synthetase and separately located genes encoding the methylTHF branch of the pathway and a formate dehydrogenase. Interestingly, both strains possess duplicates of the *cooS* genes, which are encoded separately by the CODH operons (DealDRAFT_0192 in strain AHT1^T and OMD50_RS09960 in Z-1002 strain). No genes of energy-converting hydrogenases have been identified in *D. alkaliphilus*. The genomes of both strains encode electrogenic membrane Na^+/H^+ Mrp antiporters, which are essential for natronophiles to expel sodium from the cell. Also, both genomes contain similar gene sets determining the processes of osmotic balancing, namely those, encoding sucrose-phosphate phosphatases and sucrose synthases, as well as ABC transporters specific for betaine and choline. Oxidation of short chain organic acids is determined in both strains by highly similar genes of tetrameric pyruvate oxidoreductases, succinate dehydrogenases/fumarate reductases

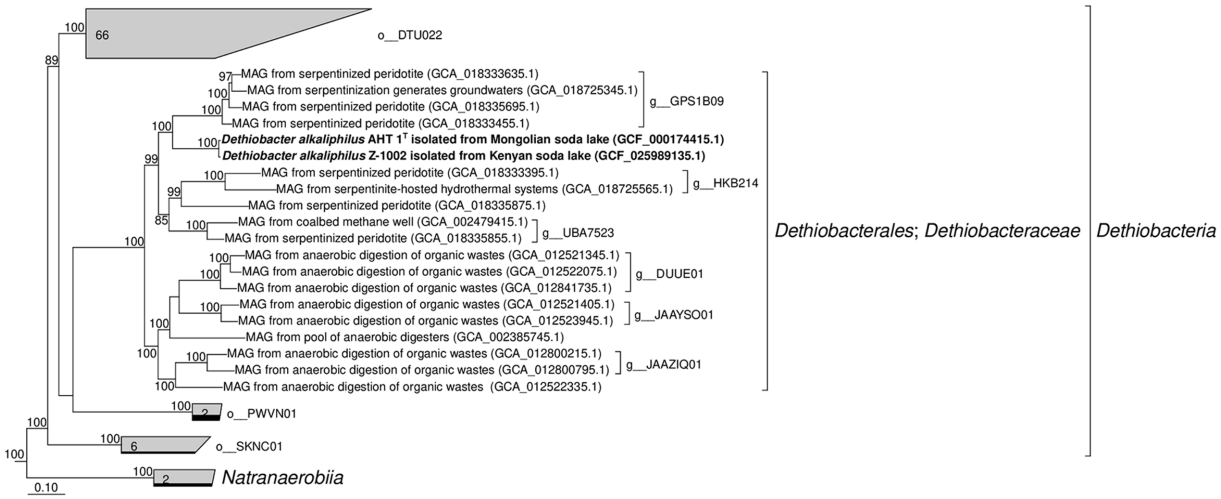


FIGURE 3
Phylogenomic placement of *Dethiobacter alkaliphilus* strains and MAGs of the *Dethiobacteraceae* family based on concatenated partial amino acid sequences of 120 bacterial single copy conserved marker genes with taxonomic designations according to the GTDB (Release 07-RS207). Bootstrap consensus tree is shown with values above 85% placed at the nodes. Bar, 0.1 changes per position.

(OMD50_RS06125-30 and DealDRAFT_0594–96 loci, respectively), and acetyl-CoA-synthetases encoded adjacently to CODH subunits.

Reduction and oxidation of iron minerals

Strain AHT1^T has been previously reported to possess a wide repertoire of multiheme *c*-type cytochromes including those thought to be involved in nitrate reduction, redox transformation of sulfur and Fe(III) compounds (Melton et al., 2017; Zavarzina et al., 2018; Sorokin and Merkel, 2019). Here we have revisited previous annotations of multiheme genes predicted in AHT1^T and compared them with cytochrome-encoding genes from strain Z-1002. A comprehensive analysis revealed 31 proteins with *c*-type multiheme conserved domains encoded in the AHT1^T genome and 27 multiheme proteins encoded in the Z-1002 genome. All the multihemes of the latter strain share homology with those of the type strain (31–100% amino acid sequence identity, Supplementary Table S1), whereas the AHT1^T genome encodes 7 multihemes that have no homology to the proteins of the strain Z-1002. The majority of multihemes from both *D. alkaliphilus* strains share similarity with the proteins previously reported to be involved in EET to Fe(III) minerals in *Geobacter*, *Shewanella*, *Thermincola*, *Carboxydothermus* species, or in Fe(III)-reducing archaea (Supplementary Table S1). In addition, 2 multihemes from AHT1^T and 3 multihemes from Z-1002 share similarities (22–27% identity) with MtoA-type outer surface cytochromes that perform the reverse EET process of Fe(II) oxidation within MtoABC porin-cytochrome complexes (Shi et al., 2012). The majority of EET-related multihemes, whether putative Fe(III) reductases or Fe(II) oxidases, are encoded by clustered genes in both strains of *D. alkaliphilus*. However, the structure of these gene clusters differs in the two strains (Figure 4). In total, 6 such clusters were identified and numbered according to their genomic coordinates in strain AHT1^T and 5—in strain Z-1002. The clusters determining the EET processes are likely to encode not only the terminal oxidoreductases, interacting with extracellular electron acceptors or donors, but also the proteins linking these oxidoreductases with the membrane-bound electron transfer chain, quinol oxidizing cytochromes and auxiliary proteins involved in the secretion and proper spatial localization of the components of the EET chain within the cell envelope and on the cell surface (Shi et al., 2012). The strain AHT1^T possesses four such clusters. The cluster 1-Fe-T (of the type strain AHT1^T) contains 3 multiheme genes homologous to those of SmhB cytochrome (38% identity) specific for soluble Fe(III) complexes in *C. ferrireducens* (Gavrilov et al., 2021), the key multiheme MtrA (22% identity) of the EET pathway in *S. oneidensis* which transfers electrons across the outer membrane to extracellular acceptors (Campbell et al., 2022), and a homolog of a quinol-oxidizing multiheme CymA (34% identity) that initiates the metal-reducing pathway in *S. oneidensis* (Shi et al., 2016). Cluster 1-Fe-T also encodes NHL- and TPR-repeat-containing proteins and includes a regulatory region downstream of the cytochrome genes. All the proteins encoded within this cluster have no homologs in the genome of strain Z-1002. Cluster 2-Fe-T in strain AHT1^T encodes two homologs (25 and 30% identity) of secreted *c*-type cytochromes from Gram-positive thermophilic Fe(III)-reducers "*T. potens*" and *C. ferrireducens* together with NHL- and TPR-repeat-containing proteins and electron transfer flavoproteins. The cluster is flanked with two regulatory genes (Figure 4). Only one protein from this cluster, DealDRAFT_1428, has a homolog in Z-1002 strain (Supplementary Table S1). The largest cluster 3-Fe-T of AHT1^T, which was previously mispredicted to include the locus DealDRAFT_1428–35 (Zavarzina et al., 2018), encodes nine multiheme cytochromes. Two of these proteins (namely, DealDRAFT_1439 and 1457) are homologous (22 and 27% identity) to MtoA-type cytochromes, while the others are similar to CymA and various auxiliary multihemes of EET pathways in *C. ferrireducens* and an ANME-2 group archaeon (Gavrilov et al., 2021; Kashyap and Holden, 2021). Interestingly, the MtoA homologs of strain AHT1^T share varying similarity (from complete to 37% identity) with six different multihemes of strain Z-1002 (OMD50_RS14355, 14365, 14375, 09855, 13505, 13520, Supplementary Table S1). In total, seven multihemes of the 3-Fe-T cluster have homologs in strain Z-1002

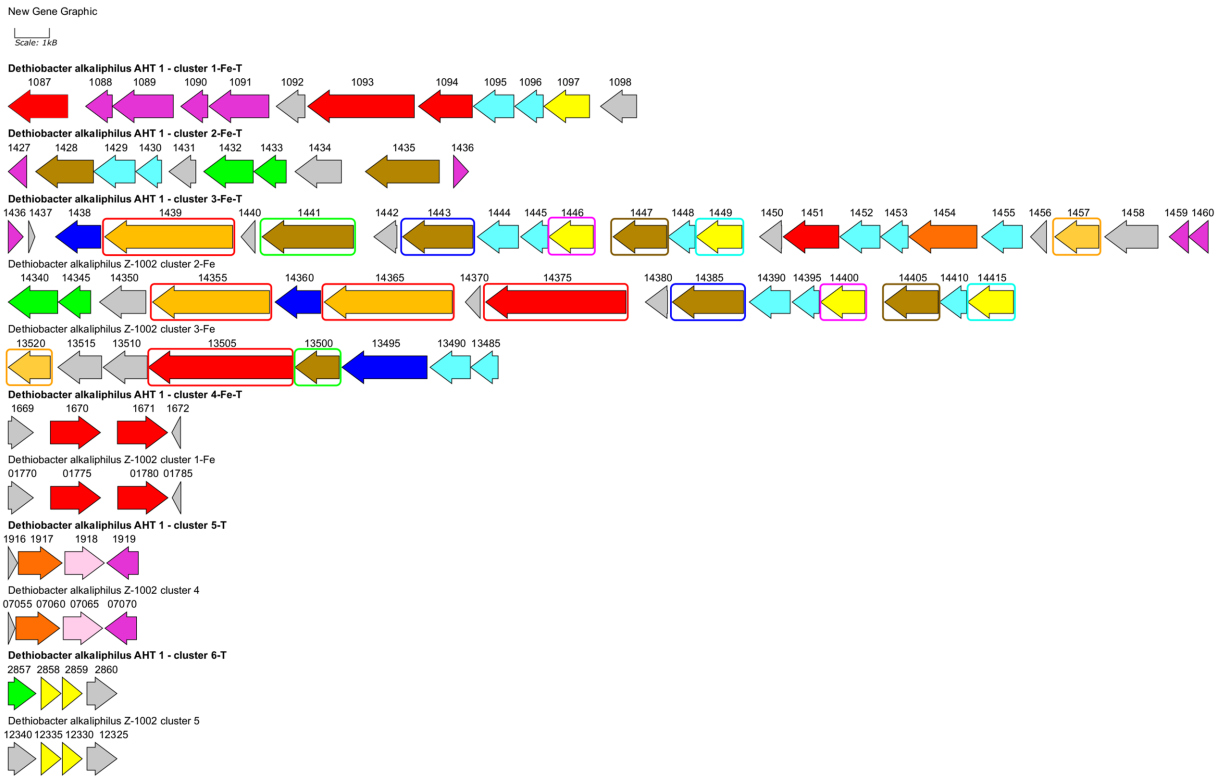


FIGURE 4
Clusters of EET-related genes and their close genomic neighborhood in *D. alkaliphilus* strains AHT1^T (marked with *bold* text) and Z-1002. Clusters are numbered according to their genomic coordinates. Gene mapping is performed with Gene Graphics web application (Harrison et al., 2018). Genes are marked with the numbers of corresponding locus tags with their prefixes omitted (DealDRAFT_ for the strain AHT1^T and OMD50_RS for the strain Z-1002). Genes are colored according to their predicted products: *green*—flavoproteins, *blue*—transport system proteins, *light blue*—NHL- and TPR-repeat-containing proteins, *pink*—rhodanese domain-containing proteins, *red*—homologs of putative terminal Fe(III) reductases, *olive*—homologs of various EET-related secreted multihemes, *light brown*—homologs of MtoA-type Fe(II) oxidases, *dark orange*—homologs of octaheme SirA sulfite reductases and Otr tetrathionate reductases, *yellow*—putative quinol oxidases, *gray*—other proteins including hypothetical ones. Similar genes of multiheme cytochromes in the clusters 3-Fe-T, 2-Fe, and 3-Fe are marked with differentially colored rectangles. Clusters 4-Fe-T and 1-Fe, 5-T and 4, as well as 6-T and 5 are almost identical to each other, that is not specially marked on the scheme. Homology values for all the mentioned similar genes are given in Supplementary Table S2.

(Figure 4). In addition to cytochromes, the 3-Fe-T cluster of strain AHT1^T contains the genes of several NHL- and TPR-repeat-containing proteins and a component of an ABC-type transport system.

In strain Z-1002, the majority of EET-related genes are organized in two clusters (Figure 4). The largest cluster 2-Fe encodes three homologs of the DealDRAFT_1,439 protein mentioned above (38–87% identity, Supplementary Table S1). Two of these homologs share similarity with the MtoA Fe(II)-oxidizing cytochrome (23 and 29% identity), and one—with phylogenetically related MtrD Fe(III)-reducing cytochrome (24% identity). Other multihemes encoded within the 2-Fe cluster of strain Z-1002 share homology with the CymA quinol-oxidizing cytochrome (35% identity) and auxiliary EET-related proteins of *C. ferrireducens* and *S. oneidensis* (29–31% identity). In addition, this cluster encodes electron transfer-related flavoproteins, NHL- and TPR-repeat-containing proteins, and an ABC transporter all of which might be involved to the secretion and proper orientation of EET-driving multihemes on the cell surface. Another large cluster of EET-related multiheme genes in strain Z-1002 is 3-Fe which also encodes an MtoA homolog (27% identity) OMD50_RS13520, completely identical to DealDRAFT_1,457, as well as the homologs (26–28% identity) of the periplasmic DmsE and outer cell surface OmcX cytochromes involved in the EET pathways of *S. oneidensis* and *G. sulfurreducens*, respectively.

Both strains of *D. alkaliphilus* possess three more EET-related gene clusters identical to each other (Figure 4). These small clusters encode the homologs of putative Fe(III) reducing multihemes (31–33% identity, clusters 4-Fe-T and 1-Fe), putative quinol-oxidizing cytochromes (29–32% identity, clusters 6-T and 5), and the cytochromes homologous to SirA/MccA-like sulfite reductases and rhodanese domain proteins (clusters 5-T and 4, see below).

Considering a rare case of the appearance of both Fe(III) reduction, and Fe(II) oxidation determinants in a single organism, we have analyzed the homologs of MtoA proteins from *D. alkaliphilus* strains in more detail. The protein DealDRAFT_1439 and all its four homologs from strain Z-1002 appeared to possess large immunoglobulin-like conserved domains at their C-terminal regions. The same domain was identified in the protein OMD50_RS09855 homologous to the MtrD Fe(III)-reducing cytochrome of *S. oneidensis* (22% identity), as well as to DealDRAFT_1457 MtoA-like protein which only possesses multiheme conserved domains (29% identity, Supplementary Table S1). Phylogenetic reconstruction of the

multiheme domains of all the MtoA-like proteins from both *D. alkaliphilus* strains (DealDRAFT_1439, 1454, 1457, and OMD50_RS14355, 14365, 13520), together with related MtrA/D homologs from this species (DealDRAFT_1094, OMD50_RS14375, 09855), revealed separate clustering and deep phylogenetic branching of all *D. alkaliphilus* proteins from previously described DmsE-family decahemes involved in Fe(II) oxidation or Fe(III) reduction (MtoA proteins or MtrA/D proteins, respectively, Supplementary Figure S4). The protein DealDRAFT_1,439 clustered with its three homologs (OMD50_RS14355, 14365, and 14375) irrespectively of their own similarity to MtoA or MtrD proteins. Comparatively deep branching of Z-1002 multihemes from those of AHT1^T strain was observed for OMD50_RS14375 and OMD50_RS09855 MtrD-like proteins.

Dissimilatory reduction of sulfur and thiosulfate

It was previously noted that the sulfur- and thiosulfate reducing type strain of *D. alkaliphilus* does not possess any canonical determinants of sulfate, sulfur, or thiosulfate respiration—neither DsrAB complexes, nor molybdopterin oxidoreductases. Instead, only *c*-type multiheme oxidoreductases have been identified in its genome (Melton et al., 2017; Sorokin and Merkel, 2019). Our analysis supported previous assumptions and revealed in both strains the genes of *c*-type cytochromes that are homologous to the Otr class of octaheme tetrathionate reductases (Simon et al., 2011) or SirA/MccA-like sulfite reductases (Kern et al., 2011; Shirodkar et al., 2011) (20 and 22–23% identity, respectively, Supplementary Table S1). In addition, we have identified the clusters encoding [NiFe] hydrogenases of the Group 3b together with beta and gamma sulfur reductase subunits of sulfhydrogenase complexes in both strains. Such enzyme complexes are proposed to harbor the activity, reducing elemental sulfur with molecular hydrogen (Ma et al., 1993). However, strain Z-1002, incapable of sulfur reduction, possesses a sulfhydrogenase cluster too, that raises doubts about the involvement of sulfhydrogenases in sulfur respiration of *D. alkaliphilus*. The difference between the two strains was observed in putative octaheme determinants of sulfur reduction. Each of the strains possesses two cytochromes. These are DealDRAFT_1917 similar to OMD50_RS07060, encoded within identical gene clusters 5-T and 4, respectively, and DealDRAFT_2033 homologous to OMD50_RS05590 as well as to the SirA sulfite reductase of *S. oneidensis*. The clusters 5-T and 4 also encode rhodanese (Cipollone et al., 2007) domain-containing proteins (Figure 4 and Supplementary Table S1). Notably, the protein DealDRAFT_1917 was previously misidentified as a thiosulfate sulfurtransferase (Melton et al., 2017). The sulfur reducing type strain AHT1^T possesses an additional homolog of Otr, DealDRAFT-1,454, which is encoded in the cluster 3-Fe-T (Figure 4) and has no homologs among the proteins of strain Z-1002. All these *D. alkaliphilus* multihemes also share homology with previously described EET-related cytochromes (Supplementary Table S1). This fact correlates with recently revealed structural similarity between SirA/MccA octaheme sulfite reductases and OcwA/OmhA terminal Fe(III) reductases (Soares et al., 2022). Phylogenetic reconstruction of all the SirA and Otr homologs from *D. alkaliphilus*, together with its OmhA and OcwA homologs taken as an outgroup, revealed a deep branching of the octahemes from a common ancestor of both Otr- and SirA-like proteins. Interestingly, the homologous proteins DealDRAFT_2033 and OMD50_RS05590 clustered together with the OcwA and OmhA Fe(III) reductases and branched off all the other analyzed cytochromes (Figure 5). The evolutionary history of the octahemes from *D. alkaliphilus* is likely to include their early branching off the common ancestor of Otr- and SirA-like proteins with rather rapid further separation of the DealDRAFT_1,454 protein and the ancestor of highly similar DealDRAFT_1917 and OMD50_RS07060 cytochromes.

Proteomic profiling of *Dethiobacter alkaliphilus* cells

To assess the relevance of the identified multihemes for EET and to detect any accessory proteins involved in electron transfer to Fe(III) or sulfur compounds, as well as from Fe(II) minerals, the results of a shotgun proteomic analysis of cells, harvested at late logarithmic growth phase, were compared across two different cultivation conditions for each strain. The strain AHT1^T was cultured with ferrihydrite or thiosulfate, and the strain Z-1002 was cultured with ferrihydrite or the mixture of Fe(II) minerals (siderite, green rust, and iron oxides). In each case, the triplicate samples were taken from three different cultures (in three biological replicas).

In the case of the AHT1^T strain, tandem MS revealed 1910 different proteins identified with a least of two unique peptides. The number of proteins identified for each sample is provided in Supplementary Table S3. Profiling of individual multiheme proteins revealed 17 EET-related multihemes that were differentially expressed under different growth conditions. A clear difference in multiheme cytochrome profiles was observed between ferrihydrite- and thiosulfate-grown cells. However, several proteins, such as DealDRAFT_1449 homologous to a CymA-like quinol oxidase (Supplementary Table S1), were expressed at the same level at both culture conditions (Figure 6A). In contrast, the protein DealDRAFT_1,457 appeared to be the only one exclusively expressed in thiosulfate reducing cells (Figure 6A). In ferrihydrite-grown cells of strain AHT1^T, the highest relative abundance was observed for DealDRAFT-1,087 which is homologous to the SmhC (23% identity) Fe(III)-specific secreted multiheme of *C. ferrireducens* (Supplementary Table S1). In thiosulfate-grown AHT1^T cells, the highest expression level was detected for the multiheme DealDRAFT_2,539 (Figure 6A) but in this case, the difference in the expression level was not statistically supported by the imputation procedure due to the absence of DealDRAFT_2,539 protein in one of the three preparations of the thiosulfate-grown biomass. Statistically relevant upregulation was observed for the proteins DealDRAFT_0324, 1087, 1093, 1435, 1439, and 1441 under ferrihydrite reduction and for the proteins DealDRAFT_1443, 1451, 1670, and 1917 (Figure 6B) under thiosulfate reduction. In addition, the cytochrome DealDRAFT_2033 was overexpressed in thiosulfate-grown cells (Figure 6A) but the absence of this protein in two of three preparations of ferrihydrite-grown biomass did not statistically support the differences in its expression level. The majority of the proteins with positive response to ferrihydrite provided as an electron acceptor are encoded in the clusters 1-Fe-T, 2-Fe-T, 3_Fe-T (Figure 4) and share homology with putative terminal Fe(III) reductases, secreted multihemes, and MtoA putative Fe(II) oxidase which is phylogenetically related to MtrA/MtrD Fe(III) oxidases (Shi et al., 2012). Interestingly, four of the six multihemes upregulated in thiosulfate-grown cells (DealDRAFT_1443, 1451, 1670, and 2539) were also homologous to various cytochromes related to Fe(III)

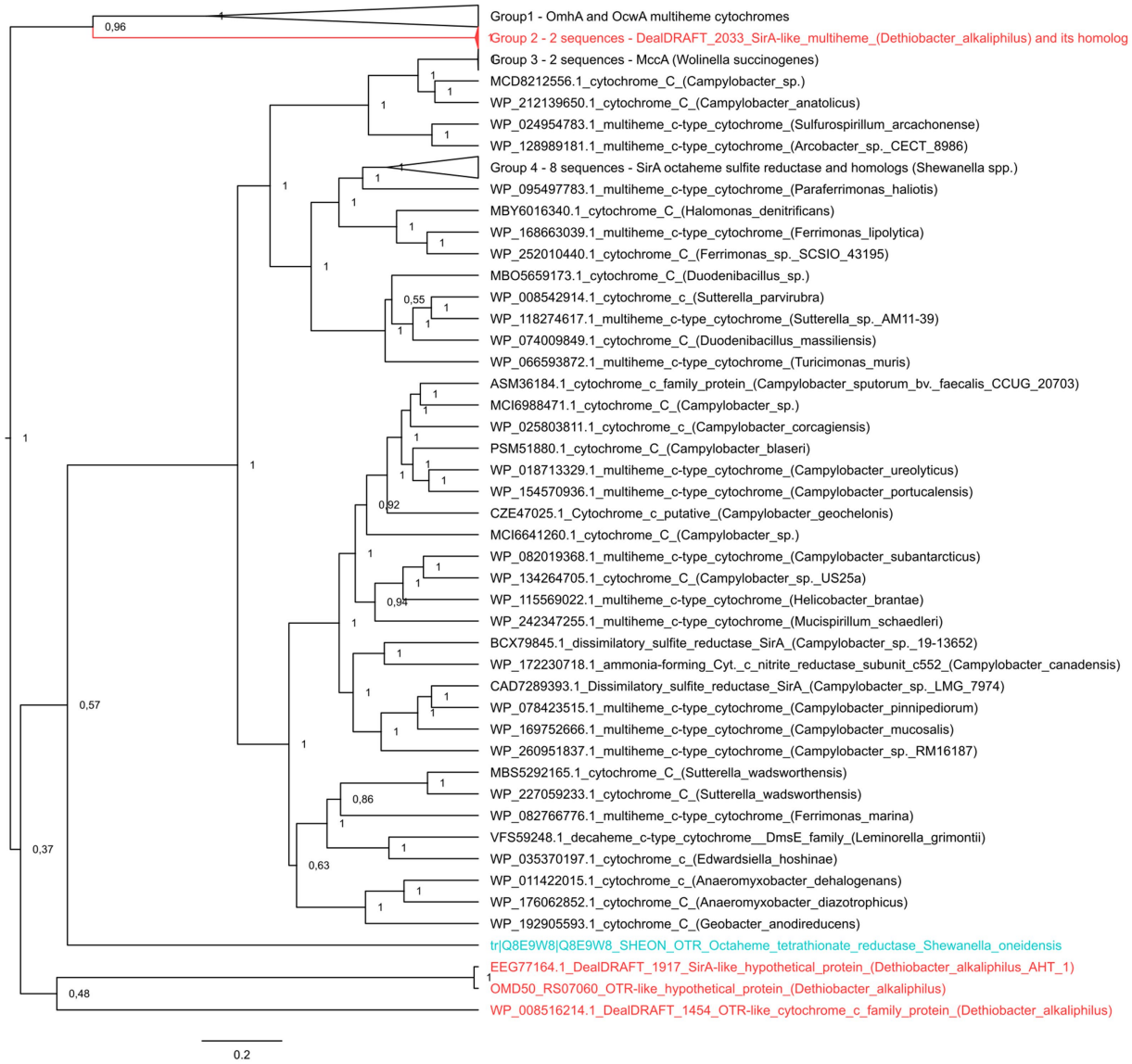


FIGURE 5
The consensus trees constructed after Bayesian inference of phylogeny from the MAFFT alignment of SirA and Otr cytochromes of *S. oneidensis*, their homologs from *D. alkaliphilus* strains, and best blast hits from public databases. Refer to "Materials and Methods" section for the detailed description of protein selection for the analysis. Homologs of SirA and Otr proteins from *D. alkaliphilus* genomes are summarized in Supplementary Table S2. The unrooted 50% majority rule consensus phylogram is displayed as a rectangular tree, for which posterior probability values are shown. 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. Labels of the proteins are colored *red* for the proteins retrieved from *D. alkaliphilus* strains, *cyan* for the manually added Otr octaheme tetrathionate reductase from *S. oneidensis* and *black* for the other proteins.

reduction (Supplementary Table S1). Only a single protein homologous to SirA sulfite reductase (22% identity), DealDRAFT_1917, was reliably upregulated during *D. alkaliphilus* growth on thiosulfate.

In the strain Z-1002, tandem MS identified 1,390 different proteins identified with at least two unique peptides. The number of proteins identified per each sample is provided in Supplementary Table S4. Profiling of individual multiheme proteins uncovered nine EET-related multihemes that were differentially expressed under ferrihydrite reduction or Fe(II) oxidation. The cytochromes OMD50_RS14355, 14365, 14415, and 14880, homologous to MtoA Fe(II) oxidases, CymA-like quinol oxidase and OmhA terminal Fe(III) reductase, respectively (Supplementary Table S1), had almost similar expression levels under ferrihydrite reduction or Fe(II) oxidation (Figure 6C). The highest expression levels under both Fe(III)-reducing and Fe(II)-oxidizing conditions were observed for the OMD50_RS14385 and OMD50_RS14880 proteins (Supplementary Table S4) which are homologous to cell surface-associated cytochromes of *C. ferrireducens* (30 and 26% identity, respectively, Supplementary Table S1). These proteins are likely to comprise the core part of the EET pathway driving the redox transformation of Fe minerals in *D. alkaliphilus*. Interestingly, the

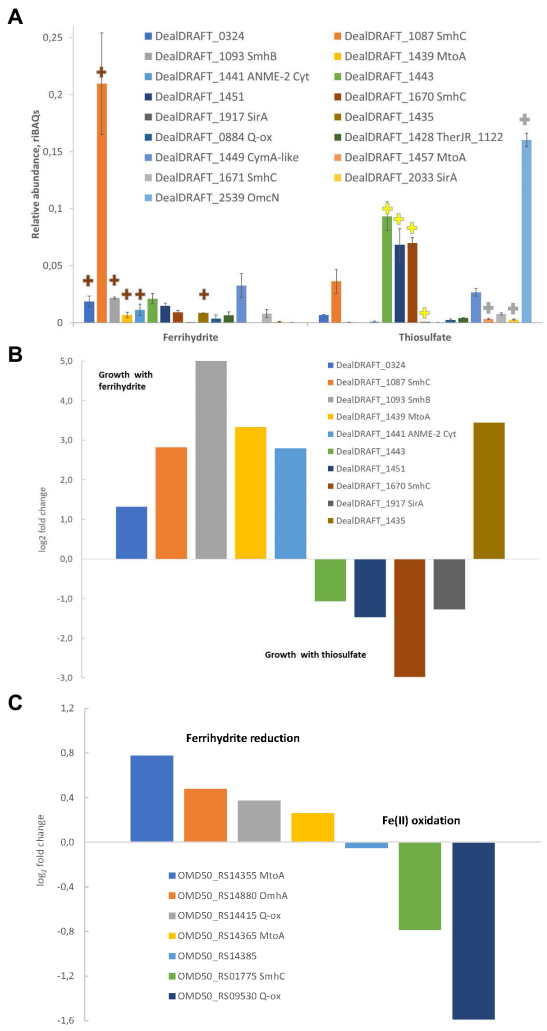


FIGURE 6
Differential expression profiles of multiheme cytochromes of *D. alkaliphilus* strains. For the strain AHT1^T, **(A,B)** ferrihydrite- and thiosulfate-grown cells are compared. For the strain Z-1002, Fe(III) reducing and Fe(II) oxidizing cells are compared. **(A)** Molar proportions (riBAQs) of individual multiheme cytochrome proteins of AHT1^T strain cultures, **(B)** statistically supported upregulation of individual multihemes in ferrihydrite-grown vs. thiosulfate-grown cells of the strain AHT1^T, **(C)** statistically supported upregulation of individual multihemes in ferrihydrite-reducing vs. Fe(II)-oxidizing cells of the strain Z-1002. **(A)** Presents the data for all the EET-related cytochromes of the strain AHT1^T, which were identified in protein profiles. *Brown plus signs* on **(A)** mark the cytochromes which upregulation in AHT1^T ferrihydrite-grown cells was statistically supported **(B)**, *yellow plus signs* mark reproducibly upregulated proteins in thiosulfate-grown cells **(B)**, *gray plus signs* indicate the proteins which differential expression was not statistically supported. On panel **(B)**, upregulation in thiosulfate grown cells is visualized by negative values of $\log_2$(fold changes). On panel **(C)**, upregulation in Fe(II)-oxidizing cells of Z-1002 strain is visualized by negative values of $\log_2$(fold changes). The cytochromes on all the panels are marked with locus tags and their close homologs among reported EET-related proteins (refer to Supplementary Table S1 for details). All the proteomic data were obtained for three biological replicates, mean values and mean deviation are presented on *A* and in the Supplementary Table S3, S4.

cytochromes OMD50_RS02905 and 7,060, homologous to OmhA Fe(III) reductase and SirA sulfite reductase (25 and 23% identity, respectively) were exclusively expressed under Fe(II) oxidation, whereas a putative quinol oxidase OMD50_RS09530 showed the strongest upregulation under the same growth conditions (Figure 6C).

Enrichment of *Dethiobacter* representatives from subsurface mineral waters of YMWD

We have detected a phylotype belonging to the genus *Dethiobacter* in a subsurface mineral water sample collected from the Lower Cretaceous aquifer of YMWD through well 9. The phylotype comprised less than 0.1% of the microbial diversity of this sample (Figure 7, 1st column). The sample was then stored at +4°C for 1.5 years and then used for the enrichment of prokaryotes inhabiting the deep subsurface aquifer using a mixture of electron donors and acceptors available in this environment. This primary enrichment containing both sulfate and ferrihydrite (as SF) and incubated at +30°C for 3 weeks was dominated by typical sulfate reducing taxa but carried out the reduction of SF to a black precipitate. No *Dethiobacter* representatives were detected in the water sample after its long-term storage (1.5 years) or in the primary enrichment at the end of its 3-weeks incubation (Figure 7, 2nd and 3rd columns). To check the presence of iron reducers in the obtained enrichment, it was transferred to selective media containing formate (10 mM) or acetate (10 mM) with SF (50 mM final Fe(III) content) as the only electron acceptor. After 2 weeks of incubation, the color of the Fe mineral changed from reddish to dark-brown, which is characteristic of ferrihydrite reduction. Fluorescence microscopy of mineral samples taken from these cultures revealed that mineral particles were densely settled with rod-shaped cells of different length. 16S rRNA gene profiling of the Fe(III) reducing enrichments revealed the predominance of *Dethiobacter* phylotypes (96% identity with *D. alkaliphilus*) in the culture with SF and formate (33.2% relative abundance) and their complete absence in the culture with SF and acetate (Figure 7, 4th and 5th columns, respectively).

Environmental distribution of *Dethiobacter*-related phylotypes

The survey of the environmental distribution of 16S rRNA gene sequences of the genus *Dethiobacter* available in public databases shows that about 25% of the sequences are retrieved from sediments of soda lakes and alkaline soils, 35% of the sequences are retrieved from Fe-enriched serpentinites, about 30%—from anaerobic bioreactors and digesters (Supplementary Figure S5). The remaining small fraction of sequences was found in microbial mats, ground and urban waters. Thus, *Dethiobacter* is the most abundant in sulfur- or iron-enriched alkaline natural habitats. However, the representatives of the genus have also been detected in anthropogenic organics-rich neutrophilic environments. Unfortunately, the lack of publicly available data on the physico-chemical conditions sustained in these anthropogenic environments, and on the sources of wastewaters treated there do not allow us to discuss the phylogenetic group of *Dethiobacteraceae* detected in bioreactors in the context of the evolution of their metabolic capacities.

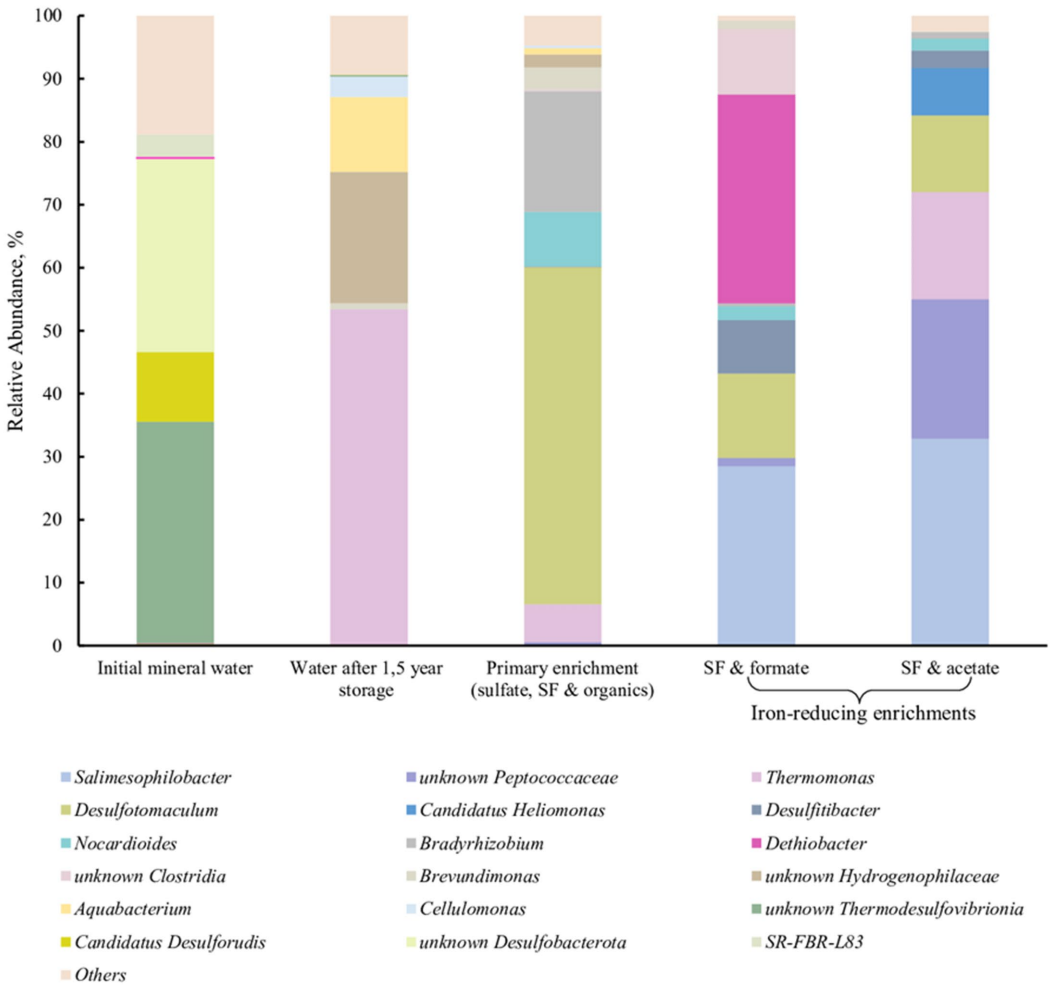


FIGURE 7
16S rRNA gene-based profiling of environmental samples and enrichment cultures from subsurface mineral water of the Lower Cretaceous aquifer of YMWD.

Discussion

Currently, *Dethiobacter alkaliphilus*, the only cultured representative of the class "*Dethiobacteria*," harbors two strains with contrasting adaptation to the conditions favoring iron or sulfur cycling. The type strain is a sulfidogen capable of elemental sulfur reduction (Sorokin et al., 2008), as well as of chemolithoautotrophic sulfur disproportionation (Poser et al., 2013, 2016), while strain Z-1002 appears to be incapable of sulfur reduction at all. However, both strains are capable of thiosulfate reduction (Table 2). Comparative genome analysis revealed the background of these physiological differences.

Non-canonical multiheme determinants of thiosulfate reduction in *Dethiobacter alkaliphilus*

Both *D. alkaliphilus* strains appeared to lack canonical genomic determinants of sulfur, sulfate or thiosulfate reduction, such as DsrAB or molybdopterin-containing complexes. Instead, each of the strains possessed two octaheme *c*-type cytochromes homologous to the SirA sulfite reductase or Otr tetrathionate reductase of *S. oneidensis* (Supplementary Table S1). The sulfur-reducing type strain possesses an additional weak homolog of Otr proteins, DealDRAFT_1454 (20% identity). Phylogenetic reconstruction of these putative determinants of sulfur compounds reduction revealed deep branching of three of them from the common ancestor of SirA and Otr proteins, and association of two *D. alkaliphilus* proteins with OcwA/OmhA Fe(III) reductases (Figure 5), which are structurally similar to, but phylogenetically distant from octaheme sulfite or tetrathionate reductases (Soares et al., 2022).

Multiheme cytochrome profiling revealed upregulation of the protein DealDRAFT_1917, homologous to SirA, in the cells of the strain AHT1^T grown with thiosulfate (Figure 6). Interestingly, the protein DealDRAFT_1457 homologous to MtoA Fe(II) oxidase was only detected in thiosulfate-grown cells. Furthermore, the most pronounced upregulation in thiosulfate-reducing cells was observed for DealDRAFT_1451 and 1670 (Figure 6) homologous to the putative Fe(III) reductases SmhB and SmhC (21 and 32% identity, respectively, Supplementary Table S1). The highest expression level in these cells

TABLE 2 Phenotypic and genotypic differences between the two strains of *Dethiobacter alkaliphilus*.

Characteristics	Strain AHT1^T	Strain Z-1002
pH range (optimum)	8.5–10.3 (9.5)[1]	7.8–10.1 (9.2)
Na^+ range (optimum), M	0.2–1.8 (0.4)[1]	0.35–2.5 (1.0)
S^0 disproportionation	+[2]	–
$S_2O_3^{2-}$ disproportionation	+[2]	–
S^0 reduction	+[1]	–
$S_2O_3^{2-}$ reduction	+[1]	+
Fe(III) reduction	+	+[3]
Fe(II) oxidation	–	+
Genome size (bp)	3,116,746	3,235,311
Protein-coding genes	3,097	3,225
Multiheme cytochrome genes	31	27
Putative gene clusters involved in Fe cycling[4]	4	3
Homologs of MtoA Fe(II) oxidases[4]	3	4
Homologs of SirA- and Otr-like octaheme cytochromes[5]	3	2

[1]Sorokin et al. (2008), [2]Poser et al. (2013), Poser et al. (2016), [3]Zavarzina et al. (2018), [4]refer to Supplementary Table S1 for details; [5]refer to Figure 5 for details.

was detected for the protein DealDRAFT_2539 homologous (24% identity) to OmcN Fe(III)-reducing outer surface cytochrome of *G. sulfurreducens* (Aklujkar et al., 2013). All these facts emphasize that the reduction of sulfur compounds in *D. alkaliphilus* is most likely driven by multiheme cytochromes homologous to Fe-cycling proteins.

Multiheme determinants of redox transformations of iron in *Dethiobacter alkaliphilus*

We have experimentally demonstrated the ability of the sulfur-reducing strain AHT1^T to reduce Fe(III) from SF and the ability of the Fe(III)-reducing strain Z-1002 to oxidize Fe(II) from green rust supplied as a mixture with siderite. The result of Fe(II) oxidation by strain Z-1002 was indicated by a 6.4% increase in the relative intensity of the Mössbauer spectra of Fe^{3+} atoms compared to that of uninoculated controls (Table 1).

Genomic analysis revealed that each of the *D. alkaliphilus* strains contained a large set of genes encoding multiheme *c*-type cytochromes, including those homologous to quinol-oxidases, membrane-associated Fe(III) reductases, such as OcwA or MtrA/D, and soluble electron shuttling cytochromes, previously identified in the model Fe(III)-respiring bacteria of the genera *Shewanella* and *Geobacter* or in Gram-positive thermophilic Fe(III) reducers (Figure 4 and Supplementary Table S1). Taken together, these cytochromes could be combined in an electron transfer chain linking the membrane-bound respiratory complexes with extracellular electron acceptors, such as Fe(III)-containing minerals, or electron donors, such as Fe(II)-bearing mineral mixtures. Proteomic profiling of strain AHT1^T for multihemes revealed strong upregulation of six different cytochromes, including the cytochrome DealDRAFT_1439 (Figure 6), encoded in three gene clusters of putative Fe(III) reductases (clusters 1-Fe-T, 2-Fe-T, 3-Fe-T, Figure 4). These cytochromes contain the homologs of all the necessary components of an extracellular electron transfer chain to Fe(III) oxides. Interestingly, the Fe(II)-oxidizing strain Z-1002 possesses four different homologs of the protein DealDRAFT_1,439 which are encoded in the 2-Fe and 3-Fe clusters. Of these, the proteins OMD50_RS14355 and OMD50_RS14365 share similarity with MtoA-type Fe(II)-oxidases (Shi et al., 2012). Phylogeny reconstruction of these and other MtoA homologs of *D. alkaliphilus* (DealDRAFT_1454, 1457, OMD50_RS14520) revealed clear separation of these proteins from DmsE-family decahemes of other organisms that perform Fe(III) reduction or Fe(II) oxidation (Supplementary Figure S4). Branching most deeply from the common ancestor of all the *D. alkaliphilus* decahemes is the phylogenetic cluster of MtoA-like cytochromes DealDRAFT_1439, OMD50_RS14355, and OMD50_RS14365 (Supplementary Figure S4), as well as the protein OMD50_RS14375 with closer homology (24% identity) to the MtrD Fe(III)-reducing cytochrome of *S. oneidensis* (Shi et al., 2012). We propose that three of the four proteins in this phylogenetic cluster might appear in strain Z-1002 by several duplication events with further minor changes occurred under the evolutionary pressure of the conditions favoring Fe-cycling, such as increased abundance of mixed Fe(III)/Fe(II)-containing minerals. Such an assumption correlates with the similarity of expression levels of two MtoA-like cytochromes in strain Z-1002 cells grown under ferrihydrite reducing or Fe(II) oxidizing conditions (Supplementary Table S4 and Figure 6C). In general, multi-omics analysis revealed that the biochemical machinery of Fe redox cycling in *D. alkaliphilus* strains is more complex and flexible than that of sulfur respiration.

Junctions of the pathways for iron and sulfur respiration in *Dethiobacter alkaliphilus*

The most likely quinol oxidizing/quinone reducing proteins of the respiratory chains in both *D. alkaliphilus* strains are the CymA-like homologs DealDRAFT_1449 and OMD50_RS14415 with their stable expression level almost independent of the electron acceptors or donors provided (Figure 6). Importantly, the majority of the cytochromes overexpressed under ferrihydrite- or thiosulfate-reducing conditions in the strain AHT1^T are encoded in three largest gene clusters (Figure 4), which also encode auxiliary flavin-containing electron transfer proteins together with NHL- and TPR-repeat-containing proteins that could be involved in protein–protein interactions mediating the assembly of multiprotein complexes (D'Andrea and Regan, 2003), such as MtrCAB Fe(III)-reducing ones (Shi et al., 2012). Interestingly, the putative Fe-cycling multiheme proteins of *D. alkaliphilus* showed greater divergence from each other than that observed among the EET-related proteins of the DmsE family from fairly different bacteria. Another important point is that the homologs of OmhA Fe(III) reductase and SirA sulfite reductase are expressed only under Fe(II) oxidation in strain Z-1002. These facts highlight the intricate weaving of EET pathways, that determine the redox transformations of sulfur and iron compounds in the organism, which could be a manifestation of the active evolution of multiheme proteins within *D. alkaliphilus* species, characteristic of the ongoing

adaptation of an evolving population to the changes in its environment (Zheng et al., 2019).

Proposal of an adaptation strategy of *Dethiobacter alkaliphilus* to geochemical settings of soda lakes

Our analysis of the distribution of *Dethiobacters* in natural environments revealed that the representatives of this genus show strong addiction to two peculiar alkaline, sulfur- or iron-enriched ecotopes, soda lakes and serpentinite-associated sediments, respectively (Supplementary Figure S5). Serpentinizing environments have been stable throughout the Earth's geological evolution, whereas soda lakes are subjected to continuous biogeochemical development, being surfaceous ecosystems (Stüeken et al., 2015; Furnes and Dilek, 2022). Their most pronounced restructuring occurred after the Great Oxidation Event followed by the accumulation of sulfates in the sediments and waters of the lakes (Montinaro and Strauss, 2016). Since then, the biogeochemical cycling of sulfur has become one of the major factors determining the physicochemical conditions in soda lakes and influencing the evolution of their inhabitants.

The phylogenetic reconstruction of the determinants of key catabolic processes, sustaining the growth of two *D. alkaliphilus* strains (Figure 5 and Supplementary Figure S4), offers insights into the evolutionary traits which led to the occupation of two different ecological niches by this bacterial species. The presence of multiheme cytochromes, sharing their phylogenetic root with OmhA/OcwA Fe(III)-reductases, as the only possible determinants of the reduction of sulfur compounds in *D. alkaliphilus* allows us to propose that soda lakes are secondary habitats for these organisms comparing to Fe-rich subsurface environments associated with serpentinites. This hypothesis correlates with the geological history and current geochemical characteristics of Magadi soda lake from which the Fe-reducing strain Z-1002 was isolated. This lake is located in a geologically young East African Rift Valley characterized by high pH, silica and carbonate concentration. The ecotope was formed *ca.* 1 million years ago due to faulting of the Rift Valley composed of alkali Pleistocene basalt, trachyte lava flows and phonolite (Jones et al., 1977; Eugster, 1980; Schagerl and Renaut, 2016). Thus, the bedrock of the modern lake appeared to be enriched with iron minerals. In this case, the Fe-reducing *D. alkaliphilus* could be an alien species for lacustrine sulfur-rich sediments, which was introduced there from the underlying Fe-rich volcanic rocks. The strain was managed to modify its multiheme cytochrome machinery under the evolutionary pressure of renewed geochemical settings with predominant sulfur cycle. A large number of different multiheme cytochrome genes and complete absence of typical genes determining the redox transformations of sulfur compounds could favor the adaptation of *Dethiobacter* to an ecotope where Fe-rich rocks are combined with sulfur-enriched sediments and waters.

Better fitness of the metabolic features of *Dethiobacters* to the life in Fe-rich alkaline environments was highlighted by the enrichment of their phylotypes from the Lower Cretaceous aquifer of the Yessentukskoye subsurface mineral water deposit (YMWD). The crystalline basement of the deposit is represented by Proterozoic-Paleozoic metamorphic and magmatic shales and granites, i.e., the rocks depleted with iron-bearing minerals. The sedimentary cover of the YMWD is represented by iron-depleted limestones, mudstones and siltstones of Meso-Cenozoic age. The Lower Cretaceous aquifer of the YMWD, penetrated by the well 9 which we have sampled, is directly connected with the recharge area and contains carbon-free sulfaceous alkaline waters (Elena et al., 2020; Filimonova et al., 2022). This environment is completely different from those formed by serpentinization processes. Not surprisingly, *Dethiobacter*-related phylotypes comprised a minor part (0.08%) of the microbial community in this aquifer and rather represented the so-called "rare biosphere," which is considered a phenotypic repository of microbial communities getting advantages upon significant changes of environmental conditions (Jousset et al., 2017). In our experiments, *Dethiobacter* phylotypes appeared to pass undetected through a long-term storage of a mineral water sample and its further cultivation under sulfate-reducing conditions. The phylotypes restored when the enrichment conditions became favorable for alkaliphilic lithotrophic Fe(III) reducers. In this case, *Dethiobacter* representatives became the dominant group of the enrichment culture (Figure 7). Thus, Fe-rich alkaline conditions seem to be optimal geochemical settings for yet uncultured *Dethiobacter* species. This fact together with the unity of origin of the sulfur, thiosulfate and Fe(III) reduction pathways in *D. alkaliphilus*, allows us to propose an adaptation strategy of the organism to the change of its environment from serpentinizing Fe-rich ecotopes to soda lakes. This strategy is based on the changes in the multiheme cytochrome repertoire aimed to get energy from the reduction of electron acceptors with lower redox potential (sulfur compounds vs. Fe(III) minerals). Such an example of intraspecific microevolution within *D. alkaliphilus* shows a possible way of a global adaptive response of prokaryotes to the activation of the sulfur cycle after the appearance of sulfates in the oceanic water and free oxygen in the atmosphere during GOE (Catling and Zahnle, 2020).

Emended description of *Dethiobacter alkaliphilus*

In addition to the characteristics given for the type strain of the species, AHT1^T (Sorokin et al., 2008; Poser et al., 2013, 2016; Sorokin and Merkel, 2019), and strain Z-1002 (Zavarzina et al., 2018), the following characteristics should be added to the formal description: both strains of the type species can grow by iron-reduction in the presence of molecular hydrogen, formate, acetate, lactate, succinate, pyruvate, butyrate, propionate or ethanol as the electron donors. Strain Z-1002 is unable to reduce elemental sulfur or disproportionate thiosulfate and elemental sulfur. Strain Z-1002 is able to reduce thiosulfate or anaerobically oxidize Fe(II)-containing minerals.

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 at: https://www.ncbi.nlm.nih.gov/genbank/, JAPDNO000000000.

Author contributions

DZ, AYM, AK, VP, VR, and NC: experimental work. DZ, AAM, and SG: field sampling. AYM and SG: genome annotation and

analysis. AYM and IE: phylogenetic analyses. VR and NC: Mössbauer spectroscopy and analysis. RZ and VP: proteomic studies. RZ: modification of routine protocols for the samples with low protein content, LC-MS/MS analysis, data processing and statistical analysis. DZ, SG, AYM, and AAM: writing the manuscript. DZ: convene the research. SG and AAM: acquire funding. All authors contributed to the article and approved the submitted version.

Funding

This research was partially funded by the Russian Science Foundation (grant no. 21-14-00333) (DZ, SG, and VP, physiological studies, enrichments, genome analysis), and by the Ministry of Science and Higher Education of the Russian Federation (AK, IE, and AYM, genome sequencing, phylogenetic analysis).

Acknowledgments

We are thankful to Dr. D. Yu. Sorokin for providing the strain AHT1^{T} and helpful discussion.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1108245/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., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Brazelton, W. J., Morrill, P. L., Szponar, N., and Schrenk, M. O. (2013). Bacterial communities associated with subsurface geochemical processes in continental serpentinite springs. *Appl. Environ. Microbiol.* 79, 3906–3916. doi: 10.1128/AEM.00330-13

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J., and Holmes, S. P. (2016). Dada2: high-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869

Campbell, I. J., Atkinson, J. T., Carpenter, M. D., Myerscough, D., Su, L. A.-F., Ajo-Franklin, C. M., et al. (2022). Determinants of multiheme cytochrome extracellular electron transfer uncovered by systematic peptide insertion. *Biochemistry* 61, 1337–1350. doi: 10.1021/ACS.BIOCHEM.2C00148

Catling, D. C., and Zahnle, K. J. (2020). The Archean atmosphere. *Sci. Adv.* 6:eaax1420. doi: 10.1126/sciadv.aax1420

Chakraborty, R., O'Connor, S. M., Chan, E., and Coates, J. D. (2005). Anaerobic degradation of benzene, toluene, ethylbenzene, and xylene compounds by Dechloromonas strain RCB. *Appl. Environ. Microbiol.* 71, 8649–8655. doi: 10.1128/AEM.71.12.8649-8655.2005

Cipollone, R., Ascenzi, P., and Visca, P. (2007). Common themes and variations in the rhodanese superfamily. *IUBMB Life* 59, 51–59. doi: 10.1080/15216540701206859

Crespo-Medina, M., Twing, K. I., Kubo, M. D. Y., Hoehler, T. M., Cardace, D., McCollom, T., et al. (2014). Insights into environmental controls on microbial communities in a continental serpentinite aquifer using a microcosm-based approach. *Front. Microbiol.* 5:604. doi: 10.3389/fmicb.2014.00604

D'Andrea, L. D., and Regan, L. (2003). TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28, 655–662. doi: 10.1016/j.tibs.2003.10.007

Edwardson, C. F., and Hollibaugh, J. T. (2017). Metatranscriptomic analysis of prokaryotic communities active in sulfur and arsenic cycling in mono Lake, California, USA. *ISME J.* 11, 2195–2208. doi: 10.1038/ismej.2017.80

Edwardson, C. F., and Hollibaugh, J. T. (2018). Composition and activity of microbial communities along the redox gradient of an alkaline, hypersaline, lake. *Front. Microbiol.* 9:14. doi: 10.3389/fmicb.2018.00014

Emerson, D., Rentz, J. A., Lilburn, T. G., Davis, R. E., Aldrich, H., Chan, C., et al. (2007). A novel lineage of proteobacteria involved in formation of marine Fe-oxidizing microbial mat communities. *PLoS One* 2:e667. doi: 10.1371/journal.pone.0000667

Ersoy Omeroglu, E., Sudagidan, M., Yurt, M. N. Z., Tasbasi, B. B., Acar, E. E., and Ozalp, V. C. (2021). Microbial community of soda Lake Van as obtained from direct and enriched water, sediment and fish samples. *Sci. Rep.* 11:18364. doi: 10.1038/s41598-021-97980-3

Eugster, H. P. (1980). Chapter 15 Lake Magadi, Kenya, and its precursors editor(s): a. Nissenbaum. *Dev. Sedimentol.* 28, 195–232. doi: 10.1016/S0070-4571(08)70239-5

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., et al. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2:6. doi: 10.1186/2049-2618-2-6

Filimonova, E., Kharitonova, N., Baranovskaya, E., Maslov, A., and Aseeva, A. (2022). Geochemistry and therapeutic properties of Caucasian mineral waters: a review. *Environ. Geochem. Health* 44, 2281–2299. doi: 10.1007/s10653-021-01160-1

Filimonova, E., Lavrushin, V., Kharitonova, N., Sartykov, A., Maximova, E., Baranovskaya, E., et al. (2020). Hydrogeology and hydrogeochemistry of mineral sparkling groundwater within Essentuki area (Caucasian mineral water region). *Environ. Earth Sci.* 79:15. doi: 10.1007/s12665-019-8721-2

Furnes, H., and Dilek, Y. (2022). Archean versus Phanerozoic oceanic crust formation and tectonics: ophiolites through time. *Geosyst. Geoenviron.* 1:100004. doi: 10.1016/j.geogeo.2021.09.004

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., Potapov, E. G., Prokof'eva, M. I., Klyukina, A. A., Merkel, A. Y., Maslov, A. A., et al. (2022). Diversity of novel uncultured prokaryotes in microbial communities of the Yessentukskoye underground mineral water deposit. *Microbiology* 91, 28–44. doi: 10.1134/S0026261722010039

Gavrilov, S. N., Zavarzina, D. G., Elizarov, I. M., Tikhonova, T. V., Dergousova, N. I., Popov, V. O., et al. (2021). Novel extracellular electron transfer channels in a gram-positive thermophilic bacterium. *Front. Microbiol.* 11:597818. doi: 10.3389/fmicb.2020.597818

Glaring, M. A., Vester, J. K., Lylloff, J. E., Abu Al-Soud, W., Sørensen, S. J., and Stougaard, P. (2015). Microbial diversity in a permanently cold and alkaline environment in Greenland. *PLoS One* 10:e0124863. doi: 10.1371/journal.pone.0124863

Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., et al. (2016). Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat. Biotechnol.* 34, 942–949. doi: 10.1038/nbt.3601

Hamilton, T. L., Bovee, R. J., Sattin, S. R., Mohr, W., Gilhooly, W. P. 3rd, Lyons, T. W., et al. (2016). Carbon and sulfur cycling below the chemocline in a meromictic Lake and

the identification of a novel taxonomic lineage in the FCB Superphylum, Candidatus Aegiribacteria. *Front. Virol.* 7:598. doi: 10.3389/fmicb.2016.00598

Harrison, K. J., 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

Hugerth, L. W., Wefer, H. A., Lundin, S., Jakobsson, H. E., Lindberg, M., Rodin, S., et al. (2014). DegePrime, a program for degenerate primer design for broad-taxonomic-range PCR in microbial ecology studies. *Appl. Environ. Microbiol.* 80, 5116–5123. doi: 10.1128/AEM.01403-14

Jones, B. F., Eugster, H. P., and Rettig, S. L. (1977). Hydrochemistry of the Lake Magadi basin, Kenya. *Geochim. Cosmochim. Acta* 41, 53–72. doi: 10.1016/0016-7037(77)90186-7

Jones, B. E., Grant, W. D., Duckworth, A. W., and Owenson, G. G. (1998). Microbial diversity of soda lakes. *Extremophiles* 2, 191–200. doi: 10.1007/s007920050060

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

Kamnev, A. A., and Tugarova, A. V. (2021). Bioanalytical applications of Mössbauer spectroscopy. *Russ. Chem. Rev.* 90, 1415–1453. doi: 10.1070/RCR5006

Kashyap, S., and Holden, J. F. (2021). Microbe-mineral interaction and novel proteins for Iron oxide mineral reduction in the hyperthermophilic crenarchaeon Pyrodictium delaneyi. *Appl. Environ. Microbiol.* 87, 2320–2330. doi: 10.1128/AEM.02330-20

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

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

Kevbrin, V. V., and Zavarzin, G. A. (1992). The influence of sulfur compounds on the growth of halophilic homoacetic bacterium *Acetohalobium arabaticum*. *Microbiology* 61, 563–571.

Khomyakova, M. A., Zavarzina, D. G., Merkel, A. Y., Klyukina, A. A., Pikhtereva, V. A., Gavrilov, S. N., et al. (2022). The first cultivated representatives of the actinobacterial lineage OPB41 isolated from subsurface environments constitute a novel order Anaerosomatales. *Front. Microbiol.* 13:1047580. doi: 10.3389/fmicb.2022.1047580

Krukenberg, V., Riedel, D., Gruber-Vodicka, H. R., Buttigieg, P. L., Tegetmeyer, H. E., Boetius, A., et al. (2018). Gene expression and ultrastructure of meso- and thermophilic methanotrophic consortia. *Environ. Microbiol.* 20, 1651–1666. doi: 10.1111/1462-2920.14077

Lagkouvardos, I., Fischer, S., Kumar, N., and Clavel, T. (2017). Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons. *PeerJ.* 5:e2836. doi: 10.7717/peerj.2836

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, J., Wang, Z., Belchik, S. M., Edwards, M. J., Liu, C., Kennedy, D. W., et al. (2012). Identification and characterization of MtoA: a decaheme c-type cytochrome of the neutrophilic Fe(II)-oxidizing bacterium Sideroxydans lithotrophicus ES-1. *Front. Microbiol.* 3:37. doi: 10.3389/fmicb.2012.00037

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

Ma, K., Schicho, R. N., Kelly, R. M., and Adams, M. W. (1993). Hydrogenase of the hyperthermophile *Pyrococcus furiosus* is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *Proc. Natl. Acad. Sci. U. S. A.* 90, 5341–5344. doi: 10.1073/pnas.90.11.5341

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

McCollom, T. M., and Bach, W. (2009). Thermodynamic constraints on hydrogen generation during serpentinization of ultramafic rocks. *Geochim. Cosmochim. Acta* 73, 856–875. doi: 10.1016/J.GCA.2008.10.032

Melton, E. D., Sorokin, D. Y., Overmars, L., Lapidus, A., Pillay, M., Ivanova, N., et al. (2017). Draft genome sequence of *Dethiobacter alkaliphilus* strain AHT1^{T}, a gram-positive sulfidogenic polyextremophile. *Stand Genom. Sci* 12:57. doi: 10.1186/s40793-017-0268-9

Merkel, A. Y., Tarnovetskii, I. Y., Podosokorskaya, O. A., and Toshchakov, S. V. (2019). Analysis of 16S rRNA primer systems for profiling of thermophilic microbial communities. *Microbiology* 88, 671–680. doi: 10.1134/S0026261719060110

Montinaro, A., and Strauss, H. (2016). Sulphur tales from the early Archean world. *Int. J. Astrobiol.* 15, 177–185. doi: 10.1017/S1473550415000531

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

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., Bai, J., Bandla, C., Hewapathirana, S., García-Seisdedos, D., Kamatchinathan, S., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* 50, D543–D552. doi: 10.1093/nar/gkab1038

Pisapia, C., Gérard, E., Gérard, M., Lecourt, L., Lang, S. Q., Pelletier, B., et al. (2017). Mineralizing filamentous bacteria from the Prony Bay hydrothermal field give new insights into the functioning of serpentinization-based subseafloor ecosystems. *Front. Microbiol.* 8:57. doi: 10.3389/fmicb.2017.00057

Poser, A., Lohmayer, R., Vogt, C., Knoeller, K., Planer-Friedrich, B., Sorokin, D., et al. (2013). Disproportionation of elemental sulfur by haloalkaliphilic bacteria from soda lakes. *Extremophiles* 17, 1003–1012. doi: 10.1007/s00792-013-0582-0

Poser, A., Vogt, C., Knöller, K., Sorokin, D. Y., Finster, K. W., and Richnow, H.-H. (2016). Sulfur and oxygen isotope fractionation during bacterial sulfur disproportionation under anaerobic haloalkaline conditions. *Geomicrobiol J.* 33, 934–941. doi: 10.1080/01490451.2015.1128993

Postec, A., Quéméneur, M., Bes, M., Mei, N., Benaïssa, F., Payri, C., et al. (2015). Microbial diversity in a submarine carbonate edifice from the serpentinizing hydrothermal system of the Prony Bay (New Caledonia) over a 6-year period. *Front. Microbiol.* 6:857. doi: 10.3389/fmicb.2015.00857

Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R., and Finn, R. D. (2018). HMMER web server: 2018 update. *Nucleic Acids Res.* 46, W200–W204. doi: 10.1093/nar/gky448

Purkamo, L., Bomberg, M., Kietäväinen, R., Salavirta, H., Nyyssönen, M., Nuppunen-Puputti, M., et al. (2016). Microbial co-occurrence patterns in deep Precambrian bedrock fracture fluids. *Biogeosciences* 13, 3091–3108. doi: 10.5194/bg-13-3091-2016

Purkamo, L., Bomberg, M., Nyyssönen, M., Ahonen, L., Kukkonen, I., and Itävaara, M. (2017). Response of deep subsurface microbial community to different carbon sources and electron acceptors during ~2 months incubation in microcosms. *Front. Microbiol.* 8:232. doi: 10.3389/fmicb.2017.00232

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

Russell, M. J., Hall, A. J., and Martin, W. (2010). Serpentinization as a source of energy at the origin of life. *Geobiology* 8, 355–371. doi: 10.1111/J.1472-4669.2010.00249.X

Sabuda, M. C., Brazelton, W. J., Putman, L. I., McCollom, T. M., Hoehler, T. M., Kubo, M. D. Y., et al. (2020). A dynamic microbial sulfur cycle in a serpentinizing continental ophiolite. *Environ. Microbiol.* 22, 2329–2345. doi: 10.1111/1462-2920.15006

Sabuda, M. C., Putman, L. I., Hoehler, T. M., Kubo, M. D., Brazelton, W. J., Cardace, D., et al. (2021). Biogeochemical gradients in a Serpentinization-influenced aquifer: implications for gas exchange between the subsurface and atmosphere. *JGR Biogeosciences* 126:e2020JG006209. doi: 10.1029/2020JG006209

Schagerl, M., and Renaut, R. W. (2016). "Dipping into the soda lakes of East Africa" in *Soda Lakes of East Africa*. ed. M. Schagerl (Cham: Springer), 3–24.

Schrenk, M. O., Brazelton, W. J., and Lang, S. Q. (2013). Serpentinization, carbon, and deep life. *Rev. Mineral. Geochem.* 75, 575–606. doi: 10.2138/RMG.2013.75.18

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

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, 651–662. doi: 10.1038/nrmicro.2016.93

Shi, L., Rosso, K. M., Zachara, J. M., and Fredrickson, J. K. (2012). Mtr extracellular electron-transfer pathways in Fe(III)-reducing or Fe(II)-oxidizing bacteria: a genomic perspective. *Biochem. Soc. Trans.* 40, 1261–1267. doi: 10.1042/BST20120098

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, 365–374. doi: 10.1038/nn.3312

Shirodkar, S., Reed, S., Romine, M., and Saffarini, D. (2011). The octahaem SirA catalyses dissimilatory sulfite reduction in *Shewanella oneidensis* MR-1. *Environ. Microbiol.* 13, 108–115. doi: 10.1111/J.1462-2920.2010.02313.X

Simon, J., Kern, M., Hermann, B., Einsle, O., and Butt, J. N. (2011). Physiological function and catalytic versatility of bacterial multihaem cytochromes c involved in nitrogen and sulfur cycling. *Biochem. Soc. Trans.* 39, 1864–1870. doi: 10.1042/BST20110713

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

Soares, R., Costa, N. L., Paquete, C. M., Andreini, C., and Louro, R. O. (2022). A new paradigm of multiheme cytochrome evolution by grafting and pruning protein modules. *Mol. Biol. Evol.* 39:msac139. doi: 10.1093/MOLBEV/MSAC139

Sorokin, D. Y., Abbas, B., Tourova, T. P., Bumazhkin, B. K., Kolganova, T. V., and Muyzer, G. (2014a). Sulfate-dependent acetate oxidation at extremely natron-alkaline conditions by syntrophic associations from hypersaline soda lakes. *Microbiology* 160, 723–732. doi: 10.1099/mic.0.075093-0

Sorokin, D. Y., Berben, T., Melton, E. D., Overmars, L., Vavourakis, C. D., and Muyzer, G. (2014b). Microbial diversity and biogeochemical cycling in soda lakes. *Extremophiles* 18, 791–809. doi: 10.1007/S00792-014-0670-9

Sorokin, D. Y., Kuenen, J. G., and Muyzer, G. (2011). The microbial sulfur cycle at extremely haloalkaline conditions of soda lakes. *Front. Microbiol.* 2:44. doi: 10.3389/fmicb.2011.00044

Sorokin, D. Y., and Merkel, A. Y. (2019). *Dethiobacter. Bergey's manual of systematics of archaea and bacteria*.

Sorokin, D. Y., and Merkel, A. Y. (2022). "*Dethiobacteria* class. Nov" in *Bergey's manual of systematics of Archaea and Bacteria*. ed. W. B. Whitman (Hoboken, NJ: John Wiley and Sons), 1–3.

Sorokin, D. Y., Tourova, T. P., Mußmann, M., and Muyzer, G. (2008). *Dethiobacter alkaliphilus* gen. Nov. sp. nov., and *Desulfurivibrio alkaliphilus* gen. Nov. sp. nov.: two novel representatives of reductive sulfur cycle from soda lakes. *Extremophiles* 12, 431–439. doi: 10.1007/S00792-008-0148-8

Stüeken, E. E., Buick, R., and Schauer, A. J. (2015). Nitrogen isotope evidence for alkaline lakes on late Archean continents. *Earth Planet. Sci. Let.* 411, 1–10. doi: 10.1016/J.EPSL.2014.11.037

Suko, T., Kouduka, M., Fukuda, A., Nanba, K., Takahashi, M., Ito, K., et al. (2013). Geomicrobiological properties of tertiary sedimentary rocks from the deep terrestrial subsurface. *Phys. Chem. Earth Parts A/B/C* 58-60, 28–33. doi: 10.1016/J.PCE.2013.04.007

Suzuki, S., Ishii, S., Wu, A., Cheung, A., Tenney, A., Wanger, G., et al. (2013). Microbial diversity in the cedars, an ultrabasic, ultrareducing, and low salinity serpentinizing ecosystem. *Proc. Natl. Acad. Sci. U. S. A.* 110, 15336–15341. doi: 10.1073/pnas.1302426110

Tiago, I., and Veríssimo, A. (2013). Microbial and functional diversity of a subterrestrial high pH groundwater associated to serpentinization. *Environ. Microbiol.* 15, 1687–1706. doi: 10.1111/1462-2920.12034

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

Vavourakis, C. D., Andrei, A. S., Mehrshad, M., Ghai, R., Sorokin, D. Y., and Muyzer, G. (2018). A metagenomics roadmap to the uncultured genome diversity in hypersaline soda lake sediments. *Microbiome* 6, 168–118. doi: 10.1186/S40168-018-0548-7

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

Woycheese, K. M., Meyer-Dombard, D. R., Cardace, D., Argayosa, A. M., and Arcilla, C. A. (2015). Out of the dark: transitional subsurface-to-surface microbial diversity in a terrestrial serpentinizing seep (Manleluag, Pangasinan, the Philippines). *Front. Microbiol.* 6:44. doi: 10.3389/fmicb.2015.00044

Zavarzin, G. A. (1993). Epicontinental soda lakes are probable relict biotopes of terrestrial biota formation. *Microbiology* 62, 473–479.

Zavarzin, G. A., Zhilina, T. N., and Kevbrin, V. V. (1999). The alkaliphilic microbial community and its functional diversity. *Microbiology* 68, 503–521.

Zavarzina, D. G., Chistyakova, N. I., Shapkin, A. V., Savenko, A. V., Zhilina, T. N., Kevbrin, V. V., et al. (2016). Oxidative biotransformation of biotite and glauconite by alkaliphilic anaerobes: the effect of Fe oxidation on the weathering of phyllosilicates. *Chem. Geol.* 439, 98–109. doi: 10.1016/J.CHEMGEO.2016.06.015

Zavarzina, D. G., Gavrilov, S. N., and Zhilina, T. N. (2018). Direct Fe(III) reduction from synthetic ferrihydrite by haloalkaliphilic lithotrophic sulfidogens. *Microbiology* 87, 164–172. doi: 10.1134/S0026261718020170

Zavarzina, D. G., Kolganova, T. V., Boulygina, E. S., Kostrikina, N. A., Tourova, T. P., and Zavarzin, G. A. (2006). *Geoalkalibacter ferrihydriticus* gen. Nov. sp. nov., the first alkaliphilic representative of the family Geobacteraceae, isolated from a soda lake. *Microbiology* 75, 673–682. doi: 10.1134/S0026261706060099

Zheng, J., Payne, J. L., and Wagner, A. (2019). Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* 365, 347–353. doi: 10.1126/science.aax1837

Zhilina, T. N., and Zavarzin, G. A. (1994). Alkaliphilic anaerobic community at pH 10. *Cur. Microbiol.* 29, 109–112. doi: 10.1007/BF01575757

Zhilina, T. N., Zavarzina, D. G., Kolganova, T. V., Tourova, T. P., and Zavarzin, G. A. (2005). Candidatus "Contubernalis alkalaceticum" an obligately syntrophic alkaliphilic bacterium capable of anaerobic acetate oxidation in a co-culture with *Desulfonatronum cooperativum*. *Microbiology* 74, 695–703. doi: 10.1007/s11021-005-0126-4

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 26 July 2023
DOI 10.3389/fmicb.2023.1176500

OPEN ACCESS

EDITED BY
Brian P. Hedlund,
University of Nevada, Las Vegas, United States

REVIEWED BY
Wittaya Tawong,
Naresuan University, Thailand
Senthil Murugapiran,
Diversigen, United States

*CORRESPONDENCE
Maurycy Daroch
✉ m.daroch@pkusz.edu.cn

RECEIVED 28 February 2023
ACCEPTED 06 July 2023
PUBLISHED 26 July 2023

CITATION
Jiang Y, Tang J, Liu X and Daroch M (2023) Polyphasic characterization of a novel hot-spring cyanobacterium *Thermocoleostomius sinensis* gen et sp. nov. and genomic insights into its carbon concentration mechanism. *Front. Microbiol.* 14:1176500. doi: 10.3389/fmicb.2023.1176500

Polyphasic characterization of a novel hot-spring cyanobacterium *Thermocoleostomius sinensis* gen et sp. nov. and genomic insights into its carbon concentration mechanism

Ying Jiang[1], Jie Tang[2], Xiangjian Liu[1] and Maurycy Daroch[1]*

[1]School of Environment and Energy, Peking University Shenzhen Graduate School, Shenzhen, China, [2]School of Food and Bioengineering, Chengdu University, Chengdu, Sichuan, China

Thermophilic cyanobacteria play a crucial role as primary producers in hot spring ecosystems, yet their microbiological, taxonomic, and ecological characteristics are not extensively studied. This study aimed to characterize a novel strain of thermophilic cyanobacteria, PKUAC-SCTA174 (A174), using a combination of traditional polyphasic methods and modern genomic-based approaches. The study included 16S rRNA-based phylogeny, ITS secondary structure prediction, morphological and habitat analyses, as well as high-quality genome sequencing with corresponding phylogenomic analyses. The results of the 16S rRNA, 16S-23S ITS secondary structure, morphology, and habitat analyses supported the classification of the strain as a member of a novel genus within the family *Oculatellaceae*, closely related to *Albertania* and *Trichotorquatus*. Genomic analysis revealed the presence of a sophisticated carbon-concentrating mechanism (CCM) in the strain, involving two CO_2 uptake systems NDH-I_3, and NDH-I_4, three types of bicarbonate transporters (BCT1, bicA, sbtA,) and two distinct putative carboxysomal carbonic anhydrases (*ccaA1* and *ccaA2*). The expression of CCM genes was investigated with a CO_2 shift experiment, indicating varying transcript abundance among different carbon uptake systems. Based on the comprehensive characterization, the strain was delineated as *Thermocoleostomius sinensis*, based on the botanical code. The study of the complete genome of strain A174 contributes valuable insights into the genetic characteristics of the genus *Thermocoleostomius* and related organisms and provides a systematic understanding of thermophilic cyanobacteria. The findings presented here offer valuable data that can be utilized for future research in taxogenomics, ecogenomics, and geogenomics.

KEYWORDS

Thermocoleostomius*, 16S rRNA, 16S-23 ITS, CO_2-concentrating mechanism, *Thermophilic cyanobacterium*, *Oculatellaceae*, *Leptolyngbyaceae*, *Albertania

Introduction

Thermophilic cyanobacteria are essential primary producers of geothermal ecosystems that profoundly impact the ecology and biological productivity of these sites (Esteves-Ferreira et al., 2018). In recent years, there has been increased interest in these extremophilic organisms with a focus on their ecology, taxonomy, genomics, and biotechnology as donors of useful enzymes and metabolites (Liang et al., 2018; Patel et al., 2019) and microbial cell factories for carbon photo valorization (Liang et al., 2019). Despite an observable increase in interest in this group of photosynthetic extremophiles, many organisms remain relatively poorly described compared to their mesophilic counterparts. In recent years, low costs and increased availability of next-generation sequencing has led to increases in metagenomic studies of hot spring microbial communities (Alcorta et al., 2020). Numerous Metagenome-Assembled Genomes (MAGs) of thermophilic cyanobacteria have been recovered from these sequencing projects highlighting genomic features of new thermophilic microorganisms and potential mechanisms of adaptations to the environment (Chen et al., 2020; Kees et al., 2022). Until recently, it was difficult to conduct fundamental and applied studies on thermophilic strains due to the lack of well-described thermophilic isolates that contain complete genomic, taxonomic, physiological, and morphological data. However, as more studies characterizing thermophilic strains have been published, the new genera have been proposed (Perona et al., 2022). Simultaneously, thermophilic cyanobacteria share many of the same challenges regarding their taxonomy to their mesophilic counterparts, most notably simple morphology (Komárek and Anagnostidis, 2005). Those features, often difficult to identify and plastic, like colony and filament morphology, types and sizes of cells, and characteristic features of sheaths, aerotopes, or branches make their assignment particularly challenging. The absence of sufficient genetic, physiological, or morphological data exacerbates this problem. To effectively identify cyanobacterial isolates and resolve their precise taxonomic status, a polyphasic approach combining multiple datasets is necessary (Raabova et al., 2019; Shalygin et al., 2020).

In the past few years, there has been a significant increase in the discovery of new species of cyanobacteria (Casamatta, 2023; Strunecký et al., 2023). This can be also observed among thermophilic strains where simple unicellular organisms such as *Thermosynechococcus*, *Synechococcus,* and *Thermostichus* (Komárek et al., 2020; Tang et al., 2022c) and filamentous strains exhibiting *Leptolyngbya*-like morphology underwent a series of revisions in recent years (Sciuto and Moro, 2016; Mai et al., 2018). Meanwhile, extensive revisions have also been made in the family *Oculatellaceae*, and numerous clades and new genera have been delineated from the strains previously assigned to *Leptolyngbya*, e.g., *Drouetiella* (Mai et al., 2018), *Thermoleptolyngbya* (Sciuto and Moro, 2016), *Timaviella* (Mai et al., 2018), and *Trichotorquatus* (Pietrasiak et al., 2021).

However, despite the wide and rapid deployment of next-generation sequencing technology (NGS) in almost all areas of microbiology, cyanobacterial taxonomy often lacks sufficient genomic components. This can lead to taxonomic revisions of a well-established cyanobacterial taxon based exclusively on single-gene sequencing and microscopic observations (Nowicka-Krawczyk et al., 2019). To enhance the credibility of taxonomic revisions, it is crucial, in our opinion, to incorporate comparative genomic analyses to strain descriptions that include morphological and ecological data. Examples of revisions that have been backed by genomic data include the delineation of *Thermoleptolynbya* (Sciuto and Moro, 2016) and *Trichothermofontia* (Tang et al., 2023). While many of the delineations that originally did not include comparative genomics were already confirmed and expanded thanks to those datasets (Strunecký et al., 2023), in many other instances it was not possible due to sequence unavailability. With the renewed interest in cyanobacterial taxonomy in recent years, it is important that those consolidated community efforts based on a polyphasic approach also include genomic and physiological data, to better guide taxonomic revisions in future.

In this study, an isolate PKUAC-SCTA174 (A174 thereafter), isolated from the Erdaoqiao Hot Spring in Ganzi Prefecture, western Sichuan Province of China, previously suggested as a potential new genus (Tang et al., 2018a) is described. The strain is the third potential novel organism from the hot spring after *Thermoleptolyngbya sichuanensis* (Tang et al., 2021) and *Leptodesmis sichuanensis* (Tang et al., 2022a) described earlier. A complete dataset containing genomic, morphological, and physiological data of the isolated thermophilic filamentous strain is presented. On the basis of the data and utilizing a polyphasic approach, a new genus is proposed *Thermocoleostomius sinensis* gen. et sp. nov., based on the botanical code. Finally, the genes responsible for the strain's sophisticated carbon concentration mechanism (CCM) were studied in response to changes in inorganic carbon availability.

Materials and methods

Origins, cultivation, deposition, and basic physiological assessment

The strain A174 was initially isolated from light green biofilm deposited on the surface of calcareous sinter in the Erdaoqiao Hot Spring in the Ganzi Prefecture, western Sichuan Province, China. Original sampling was performed on 12 May 2016 under environmental conditions including an ambient temperature of 15°C, a relative humidity of 71%, and a light intensity of ~1,000 lux. The temperature and pH of the thermal spring were 40.8°C and 6.32, respectively. The total dissolved solids in the spring were 447 mmol L^{-1}. The hot spring sampling site description and preliminary taxonomic identification have been reported earlier earlier (Tang et al., 2018a,b). The uni-cyanobacterial culture was routinely cryopreserved as 10% DMSO in BG11 stocks at −80°C. Unless stated otherwise, the pre-cultures for the experiments were established as 150 ml BG-11 (Stanier et al., 1971) medium in 500 ml Erlenmeyer flasks

agitated at 100 rpm in an illuminated shaking incubator (MGC-450BP-2, Yiheng, Shanghai, China) for 14 days. Unless stated otherwise, the light parameters were set at 45 μmol m^{-2} s^{-1} provided by cool white light fluorescent tubes, temperature 45°C, and photoperiod 12L: 12D. The strain was deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB-collection) with an accession number FACHB-3572. Simultaneously, the dried inactive holotype was deposited in the Herbarium of North Minzu University with the voucher number: NMU00174 along with other strains collected in Sichuan Ganzi Prefecture, which belonged to other genera and were described earlier and deposited in the same herbarium with voucher numbers: NMU00183 (A183 strain; Tang et al., 2021), NMU00121 (A121 strain; Tang et al., 2022a), NMU00231 (B231 strain), and NMU00412 (E412 strain; Tang et al., 2022b). All the herbarium specimens were prepared by harvesting and gently drying on the filter paper their biomass before deposition. Strain capacity to utilize various nitrogen and carbon sources was tested as follows: the strain was cultured in a series of concentrations of sodium bicarbonate (0, 0.1, 0.3, 0.5, 0.7, and 1 M) and nitrogen sources, including sodium nitrate (0, 0.075, 0.5, 1.5, 5, and 7.5 g L^{-1}) and sodium nitrite (1.218 g L^{-1}). Batch experiments were performed in 250 ml flasks and 50 ml breathable capped tubes containing 100 and 25 ml culture medium, respectively. The control groups were cultured in a standard BG11 medium. Similar sizes and numbers of cell pellets containing trichoid granules were inoculated in the experimental and control groups. The temperature was maintained at 45°C, and the experiment was performed continuously for 21 days, with adjustment for the evaporative losses.

Genome sequencing, assembly, and annotation of strain A174

Genomic DNA was extracted as previously described (Tang et al., 2021) and tested for integrity using agarose gel electrophoresis. The whole-genome sequencing was conducted using a combination of long-read Oxford Nanopore Technologies (ONT) performed with the PromethION sequencer, and Illumina short-read PE150 approach using the NovaSeq 6000 sequencer. For ONT sequencing, the libraries were generated using SQK-LSK109 Kit according to the manufacturer's guidelines. ONT sequencing yielded 1,387,280,929 bp of an average read length of 10,405 bp. These long-reads were used to assemble the draft circular contig of the genome using Flye v2.7 (Kolmogorov et al., 2019) plugin to the commercially available Geneious Prime 2022.2 package (Kearse et al., 2012). Subsequently, the draft contig was error-corrected with 6,098,395 filtered PE150 reads (clean data) generated with short-read technology, using Geneious mapper on default settings. Finally, the final genome was annotated using a customized pipeline, as described before (Tang et al., 2019). In short, the gene prediction and annotation were automatically performed using the NCBI prokaryotic genome annotation pipeline (O'Leary et al., 2016) and polished utilizing the RAST annotation system (Brettin et al., 2015) to minimize poor calls. The genome annotation of strain A174 was summarized in Supplementary Table 1. The complete genome was deposited in GenBank under accession number CP113797.

Phylogenetic reconstruction of 16S rRNA

The sequences of the 16S rRNA gene were extracted from the complete genome sequence of the A174 strain. The reference sequences with high similarity to the A174 were selected based on the BLAST algorithm from the GenBank database or by identifying with the genera in the family *Oculatellaceae*. Each sequence was trimmed to a similar length of ~1,100 bp, aligned with ClustalW, refined by adjustment of poorly aligned regions, and used to generate a phylogenetic tree representing the taxonomic assignment of the A174 strain. Meanwhile, the 16S rRNA gene sequence similarities between species were calculated based on pairwise alignment using an online calculator from enveomics collection toolbox (Rodriguez-R and Konstantinidis, 2016). Independent phylogenetic relationships of cyanobacteria species within the order *Synechococcales* (including family *Oculatellaceae* and *Leptolyngbyaceae*) were simultaneously conducted using PhyML V3.0 (Guindon et al., 2010) and Bayesian analysis. The nexus file generated from the alignment was run in MrBayes 3.2.7a (Ronquist et al., 2012), available on CIPRES Science Gateway (Miller et al., 2010). The best-fit substitution evolutionary model was chosen by jModelTest based on the Bayesian Information Criterion. Bayesian analysis was run using the TVM + invariant + gamma (TVM+I+R) substitution model with the parameters: NST = 6, Rates = invgamma, and MCMC Ngen = 10,000,000. An initial 25% of samples were discarded as the burn-in fraction. The potential scale reduction factor (PSRF) for parameter values was 1.00, which realized the purpose of convergent statistics. Maximum likelihood was conducted using PhyML, employing the GTR model, with 1,000 bootstrap replications. Bayesian inference results showed the best result to classify the family and the genus and were used to map correct phylogenetic distances and evaluate the relative support of branches. All phylogenies were visualized using FigTree (Rambaut, 2007) and subsequently annotated in Adobe Illustrator.

Prediction of secondary structures

The 16S−23S intergenic spacer (ITS) region was extracted from the annotated genome of the strain and analyzed for the presence of conserved domains. The conserved D1–D1′ domain and variable domains V2 and boxB of the 16S-23S ITS region were identified as described before (Iteman et al., 2000), and their secondary structures were computed with RNAstructure and visualized with Structure Editor 1.0 (Mathews, 2014). The tRNA sequences were identified using tRNA scan-SE v.1.3.1 (Lowe and Eddy, 1997).

Genome-based analyses

A series of pairwise genomic comparisons were performed to further investigate the taxonomic relationships between A174 and related strains. The nucleotide and amino acid sequences were retrieved from GenBank databases. Pairwise average nucleotide identity (ANI) and average amino acid identity (AAI) were calculated for the A174 genome against both closely related strains and focus taxa. ANI/AAI parameters were calculated using a publicly available algorithm provided through the Environmental Microbial Genomics Laboratory at Georgia Tech (http://enve-omics.ce.gatech.edu/index) and visualized as a matrix table. Further genome similarity was compared using digital DNA–DNA hybridization (dDDH) between pairs of genomes. The estimated DDH distances were calculated using an online tool (Meier-Kolthoff et al., 2022) based on a generalized linear model (GLM) by submitting and comparing two genomes using the BLAT program to obtain HSPs/MUMs (high-scoring segment pair/maximal matches that are unique in both sequences) (Chatterjee et al., 2015) and infer distances using formula most suitable for genomes.

Phylogenomic reconstruction

Concatenated protein sequences generated with the shared single-copy genes among all focus taxa were used to ascertain the phylogenomic position of strain A174 essentially as described before (Tang et al., 2022a). In short, the homologous gene clusters were identified with OrthoMCL (Li et al., 2003) and concatenated. The resultant multisequence alignment was generated using MAFFT v7453 (Katoh and Standley, 2013) and analyzed for phylogenomic inference using IQ-TREE v2.1.3 (Minh et al., 2020). Subsequently, 546 protein models were used to create an optimal substitution model using a ModelFinder module of IQ-TREE. The final assessment of the generated tree topology was performed with UltraFast Bootstrap (Thi Hoang et al., 2017), employing 1000 bootstrap tests (replicates).

Microscopic analysis

The light micrographs of the A174 strain were obtained from ~5 μl of a healthy 2-week-old culture grown at standard conditions at 400× magnification using light microscopy (LM, DP72, OLYMPUS, Japan) equipped with an image acquisition system (U-TV0.63XC, OLYMPUS, Japan). Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) micrographs were generated using the protocols described earlier (Tang et al., 2021). In short, the cells for SEM were washed in phosphate-buffered saline (PBS), fixed for 2 h in fixation solution (Servicebio, G1102), post-fixed with 1% OsO_4, and dehydrated with ethanol and isoamyl acetate before taking micrographs with Hitachi, Tokyo, Japan, SU8100. Cells for TEM were additionally embedded in agarose and cut to 60–80 nm thin layers with 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).

Investigation of the carbon concentration mechanism in strain A174

The carbon concentration mechanism (CCM) components were identified as previously described (Tang et al., 2022c). In short, the 28 well-described protein sequences involved in the CCM of a model strain *Synechocystis* sp. PCC 6803 were retrieved as a reference. The orthologous gene set of the A174 strain was obtained using bidirectional best hit methodology employing *E*-value cutoff of 1E-6, ≥30% identity, and 70% coverage with BLASTP and subsequently manually curated for accuracy using annotated genomes and summarized in Supplementary Table 2. Nucleotide sequences of Rubisco large subunit (*rbcL*), carboxysome shell proteins *ccmK1*, *ccmK2*, and two variants of each of the four genes of the putative high-affinity bicarbonate transporter BCT1, i.e., *cmpA*, *cmpB*, *cmpC*, and *cmpD* were extracted from the genomic sequence. All nucleotide sequences were translated into protein sequences and subjected to maximum-likelihood (ML) analysis as previously described (Tang et al., 2022c).

Gene transcription analysis during CO_2 shift experiment

A CO_2 shift experiment was performed to analyze the expression profiles of 10 putative genes responsible for carbon uptake in the A174 strain. As described earlier, the cells of A174 were cultivated in a shaking incubator to induce the formation of biomass pellets of ~1 mm in diameter. The pellets were harvested by filtration, and ~6,000 pellets corresponding to the dry weight of nearly 0.2 g were used for the CO_2 shift experiment. The strain pellets have been pre-cultured in BG-11 liquid medium under a continuous supply of ambient air for 24 h to adjust to the cultivation mode. All the cell pellets have been harvested by centrifugation at 1,500 ×g, washed, and subsequently resuspended in fresh BG-11 medium and grown for 24 h in HC [4% CO_2 in air (v/v)]. At the end of this cultivation period, ~200 pellets have been collected, flash-frozen in liquid nitrogen, and used for RNA extraction (HC). The remaining cells have been shifted to lower, ambient, carbon environment by replacing the CO_2-enriched air with ambient air (0.042% CO_2, defined as low CO_2, LC). The cultivation in LC lasted for 1 h, and an equal amount of biomass was used for RNA extraction (LC1). The remaining cells were grown for another 23 h, and another batch of cells was used for RNA extraction (LC24). The total RNA from each of the three biological replicate samples was extracted using RNAiso Plus reagent (Takara, Dalian, China) and treated with DNAseI (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer's instructions at the following three time point: HC, LC1, and LC24. RNA concentration and integrity have been

assessed with Nanophotometer (Implen, Germany) and agarose gel electrophoresis, respectively.

Real-time quantitative polymerase chain reaction (RT-pPCR) was used to assess 10 genes representing predicted distinct components of the A174 CCM system. The following genes have been selected for the analysis: *ndhF3*, *ndhF4*, *cmpA1*, *cmpA2*, *bicA*, *sbtA*, *rbcL*, *ccmK1*, *ccaA1*, and *ccaA2*. The *rpoB* was used as an internal control among other genes based on the expression stability vs. target genes and relatively high Ct value. The oligonucleotides used in the qPCR reaction are summarized in Supplementary Table 3. Reverse transcription was performed with PrimeScript RT reagent Kit (Perfect Real-Time kit, Takara, China). The resultant cDNA was a template for the quantitative estimation of gene expression levels. The quantitative PCR reaction was done with TB Green Premix Ex Taq (Tli RnaseH Plus) (TaKaRa, Dalian, China) in an Applied Biosystems QuantStudio 5 Real-Time PCR System (Thermo Fisher Scientific, Waltham, America) using the following conditions: denaturation at 95°C for 30 s, annealing at 95°C for 5 s, and then 40 cycles of 60°C for 30 s for extension, followed by a melt cycle of 95°C for 15 s, 65°C for 60 s, and 95°C for 15 s. Data were analyzed by the $2^{-\Delta\Delta CT}$ method, which directly uses the threshold cycle (CT) value to calculate the relative quantification of gene expression.

Taxonomic evaluation

The taxonomic description followed the recommendations of the Botanical Code, International Code of Nomenclature for Algae, Fungi, and Plants (Shenzhen code) (Turland et al., 2018) and was based on the classification system developed by Komárek et al. (2014).

Results and discussion

General genomic characteristics of strain A174

The complete genome of the A174 strain was obtained using a combination of Oxford Nanopore and Illumina sequencing platforms, resulting in a genome coverage of 239× and 157×, respectively. No plasmids were identified in the sequencing data by the analysis of closed circular sequences assembled from long-read sequencing. The strain's final genome comprised a single circular chromosome of 5,809,202 bp and exhibited a GC content of 48.6%. The overall analysis of the A174 genome revealed the existence of three identical ribosomal RNA operons, 48 tRNA genes, 1 tmRNA, 4,864 protein-coding sequences, and 4 repeat regions. In total 2,068 protein-coding sequences have been predicted to be hypothetical proteins, consistent with previous findings on thermophilic cyanobacteria (Cheng et al., 2020; Kono et al., 2022).

Phylogeny reconstruction using 16S rRNA gene

The genome of the A174 strain has three identical copies of the ribosomal RNA operon. The sequence of 16S rRNA gene was extracted and aligned with other focus taxa. According to the results of the alignment, 91 cyanobacterial strains' sequences were selected to construct 16S rRNA phylogenetic trees (Figure 1, Supplementary Figures 1, 2). The Bayesian tree inferred from nucleotide sequences in NCBI clearly separated new strains from other well-described genera. The tree was categorized into 22 genera, belonging to the order *Synechococcales* and including families of *Oculatellaceae* and *Leptolyngbyaceae*. The branch containing members of the family *Leptolyngbyaceae* included *Stenomitos* (Shalygin et al., 2020), *Kovacikia* (Miscoe et al., 2016), *Leptothermofonsia* (Tang et al., 2022b), *Phormidesmis* (Komárek et al., 2009), *Leptodesmis* (Raabova et al., 2019), *Alkalinema* (Vaz et al., 2015), *Myxacorys* (Pietrasiak et al., 2019; Soares et al., 2019), and *Pantanalinema* (Vaz et al., 2015). Family *Oculatellaceae* included *Drouetiella* (Mai et al., 2018), *Pegethrix* (Mai et al., 2018), *Cartusia* (Mai et al., 2018), *Siamcapillus* (Tawong et al., 2022), *Elainella* (Jahodárová et al., 2018), *Albertania* (Zammit, 2018), *Trichotorquatus* (Pietrasiak et al., 2021), *Timaviella* (Sciuto et al., 2017), *Kaiparowitsia* (Mai et al., 2018), *Tildeniella* (Mai et al., 2018; Strunecky et al., 2019), *Shackletoniella* (Strunecky et al., 2019), *Thermoleptolyngbya* (Sciuto and Moro, 2016), and *Oculatella* (Zammit et al., 2012). In addition, *Gloeobacter* genus which has a distinct, independent basal position, was used as an out-group to root the resultant phylogenetic tree accurately.

The position of A174, another two sequences retrieved from the isolates of this spring A141 and A171 and clones isolated from the thermal spring of Huizhou region in China (GDTS1-11, GDTS1-24, GDTS1-25, GDTS1-27, GDTS1-28, GDTS1-29) (Zhang et al., 2019), formed a well-defined clade with high posterior probabilities and 98.96% 16S rRNA sequence identity (Supplementary Table 4) that was clearly separated from other related cyanobacteria. All strains in this clade shared 16S rRNA molecular signature, morphological, and ecological traits. Phylogenetic analysis indicated that the novel clade containing A174-like species is closely related to *Albertania* (93.86%–94.90% 16S rRNA sequence identity) and *Trichotorquatus* (91.59%–92.27% identity) genera already described (Zammit, 2018; Pietrasiak et al., 2021). The closest related clades were occupied by the strains exhibiting *Leptolyngbya*-like (*sensu lato*) morphology such as *Leptolyngbya* sp. IkmLPT16, *Leptolyngbya* sp. CY40, *Leptolyngbya* sp. INSTML01, *Leptolyngbya* sp. RV74, and *Leptolyngbya* sp. CHAB TP201703. The 16S rRNA sequence identity between A174 and the neighboring strains are as follows: *Leptolyngbya* sp. IkmLPT16 (95.36%), *Leptolyngbya* sp. CY40 (95.36%), *Leptolyngbya* sp. INSTML01 (94.9%), *Leptolyngbya* sp. RV74 (95.48%), *Albertania* sp. BACA0146 (95.36%), *Leptolyngbya* sp. CHAB TP201703 (95.36%), and Uncultured cyanobacterium clone 9B-12 (95.24%). All these strains are likely to belong to a recently proposed novel taxon "*Egbenema*" (Akagha, 2022). Meanwhile, there are insufficient data to classify the two more distant strains, a benthic isolate from Artic hot springs of Kap

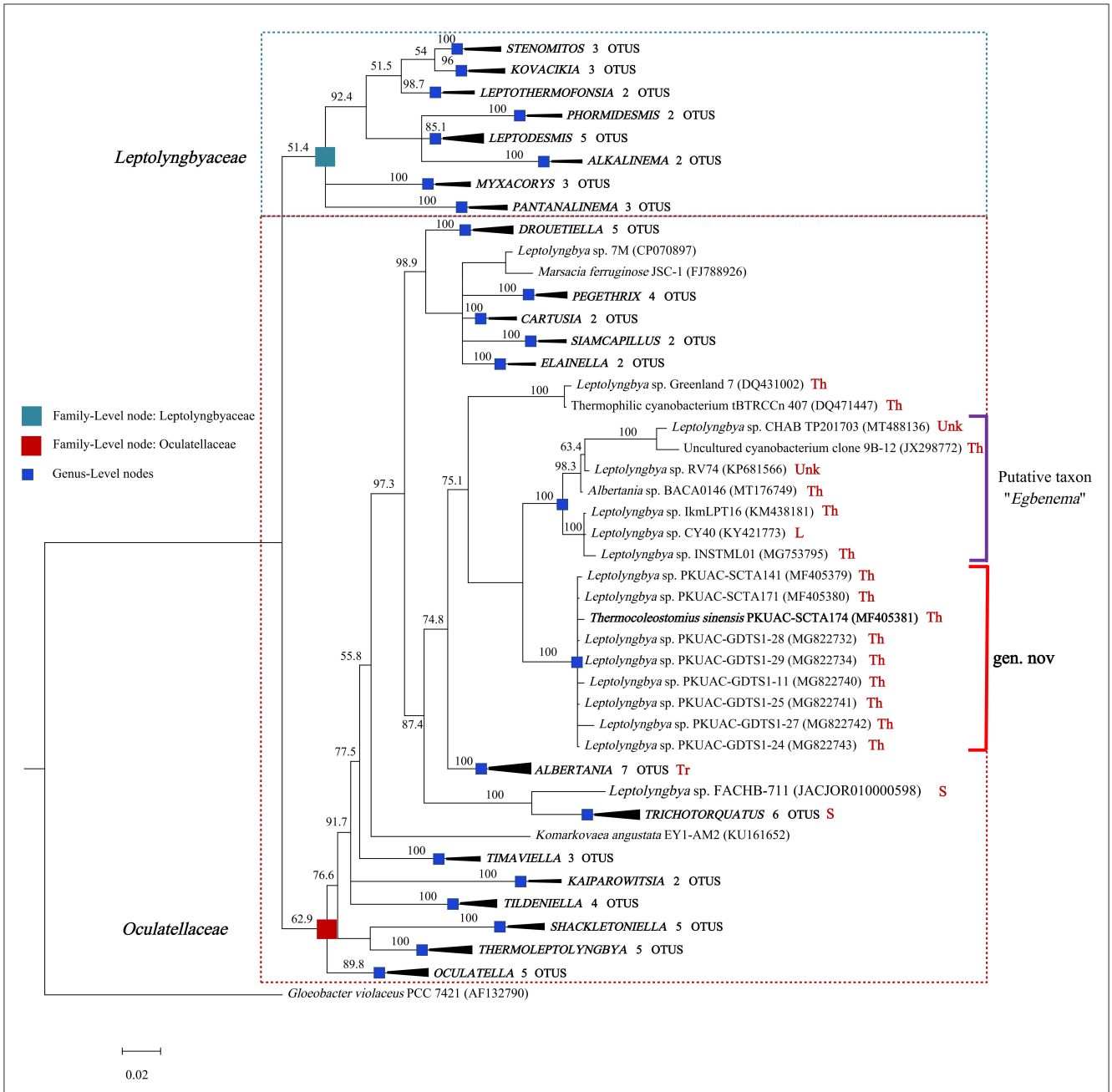


FIGURE 1
Bayesian inference of 16S rRNA gene sequences representing 91 cyanobacterial strains. Collapsed genera are indicated by black polygons, with a length corresponding to the distance from the most basal sequence to the most diverged sequence of the genus. The complete phylogram refers to Supplementary Figure 1. Posterior probabilities (%) are given above the nodes. The origin (red) of the sequences in the focus clade is provided using the following key: Th, thermal spring; L, lake; S, soil; Tr, terrestrial; Unk, unknown.

Tobin (Greenland_7) (Roeselers et al., 2007) (94.22% sequence identity) and an unidentified filamentous strain from thermal springs in Jordan (tBTRCCn407) (Bruno et al., 2009) (94.11% sequence identity). This level of sequence identity between the representatives of the two clades is at the border of differentiation of a novel genus (Rodriguez-R et al., 2018) and should be supported with additional analyses. To further classify these strains to their respective taxonomic groups, additional analyses were performed using a polyphasic approach.

Genome-based analyses

Since 16S rRNA-based taxonomy, based on a single evolutionary marker, has some limitations (Johnson et al., 2019), it is important to supplement it with other methods, especially when borderline values of genus or species delineation are observed. To make the most of the complete genome sequence of the strain and to ascertain the taxonomic position of the clade containing A174, genome-based analyses were employed. In

total, three genome-based coefficients were applied to aid the demarcation of the strain: average nucleotide identity (ANI), average amino acid identity (AAI), and digital DNA–DNA hybridization (dDDH). These parameters are presented in Table 1 and Supplementary Table 6. Methods based on ANI and AAI have proven to be useful for identifying and classifying different genera and are widely employed in taxogenomics. The ANI values (<83%) (Walter et al., 2017) and AAI values (<65%–72%) (Konstantinidis and Tiedje, 2007) are considered accepted values for genus-level discrimination. Meanwhile, two genomes are considered to belong to the same species if both ANI and AAI values between them are equal to or greater than 95% (Jain and Rodriguez, 2018).

The two coefficients can be used to verify the taxonomic position of the strains on the phylogenetic trees. The strain A174 has the closest ANI of 81.81%, to the recently sequenced *L. sichuanensis* E412. These parameters are below the level of both species and genus demarcation. The 16S rRNA taxonomy-based analysis revealed that among the strains with whole genomes sequenced, the strain most closely related to A174 was *Leptolyngbya* sp. FACHB-711. The ANI between the two strains was 75.07%, markedly lower than that of E412. Currently, there is a lack of conclusive evidence to determine the exact cause of this phenomenon. There are two non-mutually exclusive hypotheses. One possibility is the convergent evolution of A174 and E412. The two strains could independently evolve similar traits as a result of adapting to similar environments. This hypothesis is supported by the fact that the two strains were found in hot springs of similar temperatures in Kangding County of Ganzi Prefecture, and they both exhibit tolerance to elevated bicarbonate concentrations and display similar nitrogen metabolism. Another possible explanation is horizontal gene transfer. Almost a quarter of the protein-coding genes from the E412 strain was found to have been potentially acquired through this mechanism (Tang et al., 2022b). This finding suggests that the ANI of E412 may have been influenced by the transfer of genetic material from another organism resulting in an aberration of the ANI relationship resulting from the evolution. It should be noted that these explanations are not mutually exclusive, and further research is needed to fully comprehend the complex relationship between the ANIs of these two strains. Meanwhile, the AAI coefficient between the A174 and FACHB-711 was 66.56%. This parameter is at the lower boundary for genus demarcation, indicating slightly closer proximity than that of the *L. sichuanensis* E412 (63.68%).

This combined with the results of 16S rRNA phylogeny suggests that the FACHB-711 strain should be reclassified to the genus *Trichotorquatus*, closely related to the clade containing the A174 strain. ANI and AAI values between A174 and other reference genomes were in the 72–82 and 57–68% range, respectively. Those significant genome-level differences support the delineation of the clade at the genus level. The dDDH parameters further support these findings. The calculated dDDH between the A174 and other focus taxa ranged between 16.80 and 31.70%, significantly below the 70% hybridization threshold for prokaryotic species demarcation.

To further analyze the taxogenomic relationships between the strains, a phylogenomic reconstruction based on single-copy genes shared among the focus taxa with complete genome sequences was performed. A total of 642 single-copy genes were identified to be shared by all the surveyed genomes, generating concatenated alignment with a length of 199,585 aligned amino acid sites. The ML phylogram of the supergene alignment was inferred using the optimal substitution model (LG+F+R5). Each branch of the genomic phylogram is supported by strong bootstrap values (Figure 2), defining tremendous divergence among representative species from each genus. The topology of the resultant tree largely reflects that of the 16S rRNA tree (Figure 1 and Supplementary Figures 1, 2). The unique position of strain A174 confirms its classification to a novel genus. Strain FACHB-711, the closest to the A174 strain with a known genome sequence, is located in a sister clade to strain A174. The genetic distance between those two strains supports their classification to different genera, in line with the 16S rRNA phylogram and the results of genome comparisons, i.e., ANI, AAI, and dDDH. Taken together, *Leptolyngbya* sp. FACHB-711 could be reclassified to the genus *Trichotorquatus*, closely related to the novel genus containing the A174 strain if further genomic studies of well-described *Trichotorquatus* support such findings.

Analyzing all genome-based parameters and phylogenomic reconstruction based on single-copy genes, it can be concluded that the clade containing the A174 strain can be delineated from other taxa as a separate genus. Unfortunately, the scarcity of genomic sequences in the close proximity of the novel clade made its further description using genome-based parameters impossible and more traditional polyphasic methods encompassing detailed analysis of variable regions of ITS, in addition to morphological and habitat characterization need to be employed.

16S–23S ITS region and secondary structures

The ITS regions of all three ribosomal operons of strain A174 were extracted from the genome sequence and analyzed along with eight closely related strains regarding sequences and predicted structures of their ITS. Sequence alignment has shown that the three ITS regions of ribosomal operons in A174 were identical. Consequently, only one sequence was analyzed along with the related strains. The highly conserved tRNA regions were removed, and the resulting ITS sequences were processed. When selecting the corresponding operons in related organisms, the specificity of *Trichotorquatus* genus and its operon variability were also considered. Two operons (operons 2 and 3, *sensu* Pietrasiak et al., 2021) of 16S-23S ITS regions were included and compared with the target strains. The length of the resultant sequences varied from 145 to 353 bp (Supplementary Table 5). The variation could be ascribed to the significant difference in variable V2 and V3 helix structure regions. Strains in *Trichotorquatus* genus universally have V3 helixes of ~110 nucleotides. Meanwhile, in other strains, those sequences were either very short (20 nt) or absent.

The focus of the secondary structure analyses was placed on D1–D1′ and boxB regions (Supplementary Figures 3A, B). The evaluation of the predicted structures of the D1–D1′ helix revealed two structural variants that significantly differ in length. The shorter one, encompassing ~56–65 nucleotides sharing a similar

TABLE 1 Values of average nucleotide identity (ANI) and average amino acid identity (AAI) between genomes studied.

ANI \ AAI	A174	Uher 20002452	GSE-PSE-MK54-09C	CCNUW1	A121	7M	FACHB-711	JSC-1	O-77	E412	FACHB-28	LEGE06141	CLA17	A183	UHER199813D
A174	100.00	65.92	62.58	61.68	63.13	57.71	66.56	67.65	64.94	63.68	65.22	66.22	60.76	65.33	59.13
Uher 20002452	73.64	100.00	60.96	61.41	61.58	56.45	65.32	66.12	63.33	61.65	64.87	64.73	61.94	63.80	58.48
GSE-PSE-MK54-09C	74.13	73.16	100.00	58.41	60.12	52.27	60.93	61.06	64.98	60.21	61.75	62.17	58.82	65.14	58.39
CCNUW1	77.83	75.91	74.45	100.00	66.88	53.22	61.06	61.73	60.62	70.31	62.78	61.91	64.71	61.16	56.92
A121	79.07	73.49	73.74	76.89	100.00	53.89	62.45	62.91	62.77	68.77	62.86	63.07	69.17	63.07	58.03
7M	79.07	75.42	0.00	80.19	80.20	100.00	56.66	83.12	55.22	54.54	55.57	55.59	51.70	55.26	51.11
FACHB-711	75.07	76.09	72.90	76.99	75.35	77.35	100.00	66.49	63.46	62.07	64.99	65.21	61.15	63.69	58.54
JSC-1	75.81	74.34	72.95	76.50	75.01	93.96	75.81	100.00	64.18	63.49	65.14	65.67	60.58	64.54	58.88
O-77	77.57	74.67	74.69	76.14	78.58	78.43	75.70	74.88	100.00	62.74	64.28	64.29	60.41	93.57	58.99
E412	81.81	73.45	75.57	76.48	78.79	78.67	74.86	75.40	79.95	100.00	63.27	63.49	65.06	63.15	58.15
FACHB-28	73.55	76.90	73.24	77.59	74.15	77.17	77.36	75.05	73.99	74.41	100.00	68.09	62.13	64.74	59.26
LEGE06141	75.16	77.69	76.27	78.01	74.73	78.45	77.42	82.74	74.75	76.23	77.68	100.00	61.65	64.57	58.71
CLA17	77.38	79.73	74.46	78.56	75.39	0.00	77.34	73.55	74.83	75.13	77.86	79.73	100.00	60.74	57.24
A183	76.98	74.59	74.95	76.81	77.32	78.64	75.88	75.55	89.97	79.34	74.12	74.44	75.44	100.00	59.23
UHER199813D	72.26	78.06	76.78	73.67	72.76	74.35	75.37	73.54	74.50	72.54	76.51	76.63	80.58	74.53	100.00

The numbers above and below the diagonal indicate the AAI and ANI values (%), respectively.

10.3389/fmicb.2023.1176500

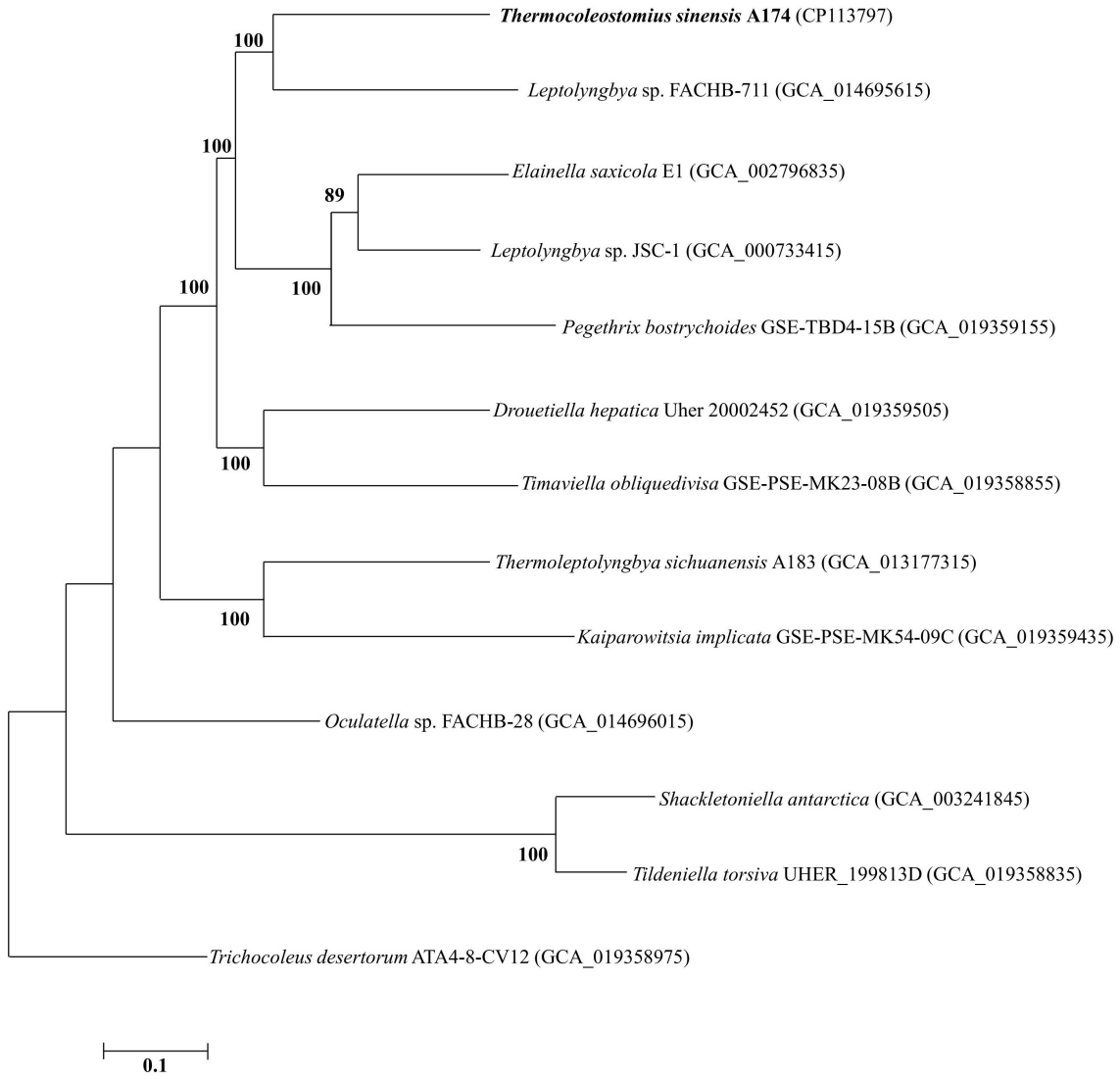


FIGURE 2
ML phylogenomic tree of a concatenated alignment of 642 single-copy genes shared by all genomes. Strain no. in bold represents the strain identified in this study. Bootstrap values (1,000 replications) are indicated at nodes. Scale bar = 5% substitutions per site.

overall fold, and a longer one of 80 nucleotides are characteristic for ITS operon 2 of the three main representatives of subclades of *Trichotorquatus*: *Trichotorquatus andrei* CMT-3FSIN-NPC37, *Trichotorquatus ladouxae* WJT66-NPBG9, and *Trichotorquatus maritimus* SMER-A. Additionally, some *Trichotorquatus* species had shorter ITS sequences present in their ITS operon 3. *Trichotorquatus maritimus* had a two nucleotide unilateral bulge when compared to other strains. Meanwhile, *Trichotorquatus* species 5 WJT32-NPBGA lacked bilateral bulge and had three unilateral bulges, which distinguishes it from other structures. It is the most similar in structure to FACHB-711, but the terminal loop of 711 had 4 nucleotides, whereas WJT32-NPBGA had seven nucleotides. One strain, *Albertania* sp. MAR67, had a unique D1–D1′ helix of 111 nucleotides. In addition, except for the similar basal clamp of five base pairs of the D1–D1′ structure, the relatively short LPT16, FACHB-711, and WJT32-NPBGA helices do not have distinct asymmetric loops of multiple nucleotides characteristic for A174 or SA373. The terminal loops of A174 and SA373 were composed of five nucleotides, while that of IkmLPT16 had 12 nucleotides. On the basis of the current, limited description of "*Egbenema*," "*Egbenema gypsiphila*" species possessed an enlarged terminal loop of 16 nucleotides, which was absent in other strains (Akagha, 2022). The length of the boxB region varied from 35 to 65 bp. Analysis of the boxB region's calculated structures shows similarity of A174 and LPT16 structures. The difference was that the end of the A174 hairpin was a 5-base asymmetric loop, while LPT16 was a 6-base symmetric loop. Meanwhile, *T. maritimus*, *T. andrei* in operon 2, and *Albertania* sp. MAR67 have shorter helices than helices in other strains, but the first two are more symmetrical. *Albertania* sp. BACA0713 has one more unilateral bulge in helices of similar length. Moreover, comparing the target strain with the putative *Egebenema* taxon, A174 has three asymmetrical additional bilateral bulge structures in BoxB helix, which all putative "*Egbenema*" strains do not have.

10.3389/fmicb.2023.1176500

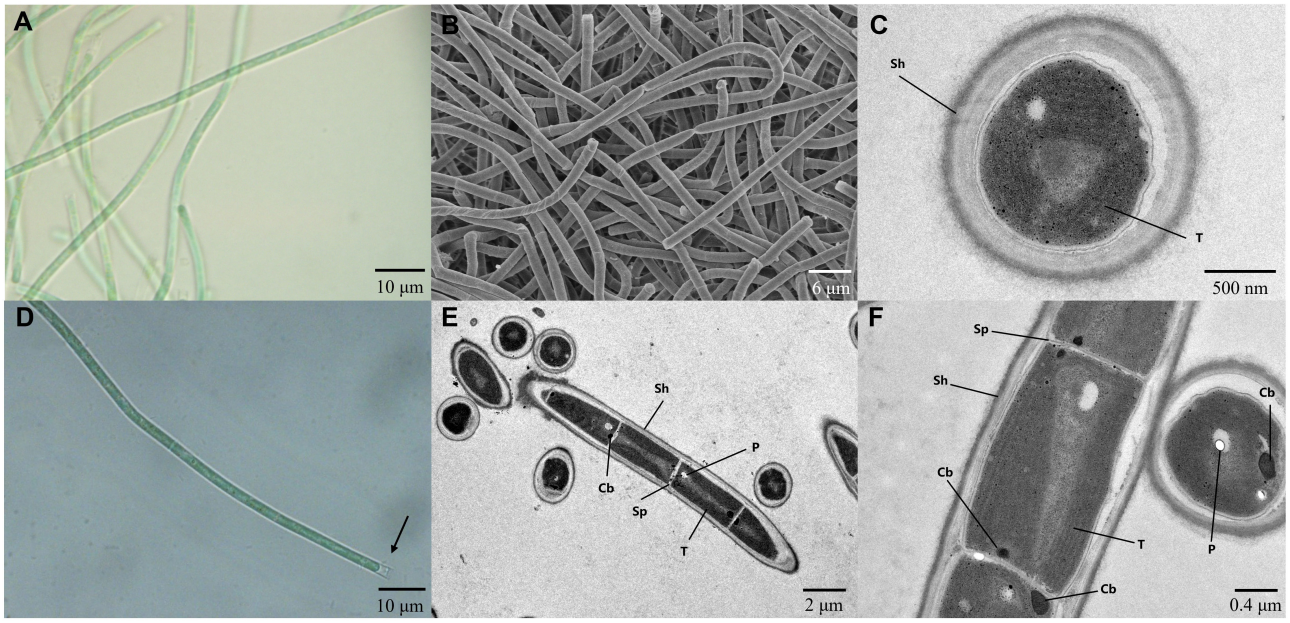


FIGURE 3
(A, D) Light micrographs of *Thermocoleostomius sinensis* sp. A174. Sheath are open at the end (arrow). Scales = 10 μm. **(B)** SEM image of the A174 strain. Scales = 6 μm. **(C, E, F)** TEM micrographs of A174. Short filament showing the round apical cell. A174 cells with carboxysome (Cb), polyphosphate body (P), sheath (Sh), septum (Sp), and thylakoid (T) in the cytoplasm. Magnifications: 400× **(A, D)**, 2,000× **(B)**, 8,000× **(C)**, 1,200× **(E)**, and 5,000× **(F)**.

The BoxB helix in *T. maritimus* in operon 3 has a unique and large unilateral bulge, which is obviously longer than in other strains. In addition, FACHB-711 is identical in structure and sequence to *Trichotorquatus* species 5 WJT32-NPBGA in BoxB helix, reinforcing that FACHB-711 most likely belongs to the *Trichotorquatus* genus.

The combined analysis of the predicted secondary structures of the D1–D1′ (Supplementary Figure 3A) and boxB (Supplementary Figure 3B) regions in the context of 16S rRNA phylogeny (Figure 1, Supplementary Figures 1, 2) and genomic analyses (Figure 2) suggests that the strain FACHB-711 should be reclassified as *Trichotorquatus*. The genus itself is composed of at least five distinctive species including *Trichotorquatus* species 5 (WJT32-NPBGA), *T. maritimus, Trichotorquatus coquimbo, T. andrei,* and *T. ladouxae*. All these strains were isolated from dryland soils in North and South America indicating that molecular evidence is consistent with their habitat. In addition, there is considerable diversity among the strains containing *Albertania,* and a currently limited amount of molecular data suggests several divergent species within this genus. Finally, the A174 strain belongs to a novel genus different from the putative taxon "*Egbenema*," *Trichotorquatus*, and *Albertania*.

Morphological and physiological characteristics of strain A174

The strain A174, when grown in the shaking flask cultures, formed very dense, isolated, small trichoid pellets of ~1 mm in diameter containing many entangled filaments. Cells aggregated when reaching a stationary phase in a liquid medium, forming bridges between granules and resulting in a dense mat. Meanwhile, the strain formed dispersed mats or thin laminae on a solid medium. The isolated cyanobacterium was investigated by light microscopy, SEM, and TEM. The image taken by light microscopy (Figures 3A, D) reveals flexuous or bent, sometimes straight trichomes. The strain was phenotypically simple and was composed of elongated cylindrical-shaped cells (Figures 3A–F) surrounded by colorless sheaths that were frequently open at the end, with the terminal cells rounded. False branching was absent. Filaments were solitary with a single trichome per sheath, blue-green in color, accompanied by some minor changes depending on the growth condition or the number of cells. TEM (Figures 3C, F) micrographs indicated that trichomes were cylindrical, composed of isodiametric cells, and the cross-walls of cells had small constrictions. The sheath was thin and colorless, 0.20–0.60 μm thick, usually longer than the trichome, and could be easily observed. Cells are circular in cross section and rectangular in the longitudinal section. Cells of the strain were isodiametric, longer than they were wide (1.5–2.5 μm wide, 2.0–5.8 μm long), with a length-to-width ratio ranging from 1.0 to 2.5 under light microscopy. The number of thylakoids varied from 4 to 6 layers. Thylakoids were arrayed in parallel lines in order at the inner edges of cells (Figures 3C, F). Sheath, septum, phycobilisome, carboxysome, and polyphosphate bodies can also be identified in the cytoplasm (Figures 3C, E, F). The morphological characteristics of other isolates from the same spring, the A141 and A171, were near-identical to the strain A174 (Supplementary Figure 13) and isolates of Huizhou region, i.e., GDTS1–24 and GDTS1–29 (Zhang et al., 2019) confirming the similar morphology of the entire clade.

TABLE 2 Comparison of morphological features and habitats of *Thermocoleostomius sinensis* and closely related strains.

Species	Morphology	Cell width (μm)	Cell length (μm)	Sheath	Thylakoids No.	Color	Habitat	References
Thermocoleostomius sinensis A174	Straight, wavy, unbranched, single trichome per sheath	1.5–2.5	2.0–5.8	Thin, colorless	4–6	Green	Thermal spring	This study
Leptolyngbya sp. PKUAC-GDTS1–24/29	Straight, wavy, mostly unbranched, few false branched	ND	ND	Thin, colorless	ND	Blue-green	Thermal spring	Zhang et al., 2019
Albertania alaskaensis	Solitary, unbranched in young culture, false branching in old culture	1.8–3.0	1.9–4	Thin, colorless	ND	Bright yellow-green	Willow roots near Meltwater Brook	Strunecky et al., 2019
Albertania sp. MAR67	Straight, wavy, unbranched, single trichome per sheath	2.0 ± 0.3/1.5–2.8	2.8 ± 0.5/1.5–4.2	Thin, colorless	ND	Green	Rocks in the Sahara Desert	Mehda et al., 2022
Albertania skiophila str. SA373 Zammit 2014	Straight, wavy, coiled, single trichome per sheath	2.0–3.0	2.0–4.0	Thin, colorless, firm, often open at the end	6–10	Green, blue-green	Humid calcareous rock walls, calcareous plasters and mortars in the cave	Zammit, 2018
Albertania egbensis N14-MA1	Straight or spiraled, entangled, rare false branching	1.6–2.6	2.0–4.6	Narrow, clear, tightly adherent to the trichomes	ND	Pinkish green	Soils or subaerial surfaces in a tropical climate area	Akagha, 2022
Albertania latericola N14-MA3	Straight, entangled, rare false branching	2.0–2.6	2.2–4.6	Narrow, clear, tightly adherent to the trichomes	ND	Pinkish green	Soils or subaerial surfaces in a tropical climate area	Akagha, 2022
Trichotorquatus maritimus SMER-A	Solitary, rarely single and double branching, single trichome per sheath	2.1–4.3	0.9–10.0	Thin, sometimes widen or absent	ND	Bright green	Dryland soils	Pietrasiak et al., 2021
Trichotorquatus coquimbo ATA2-1-KO25A	Solitary, unbranched, single trichome per sheath	1.8–3.4	1.2–5.6	Sometimes absent, thick when present	ND	Bright green	Dryland soils	Pietrasiak et al., 2021
Trichotorquatus andrei WJT9-NPBG15	Solitary, unbranched, single trichome per sheath	1.0–2.8	1.5–5.0	Sometimes absent, thick when present	ND	Bright green	Dryland soils	Pietrasiak et al., 2021
Trichotorquatus ladouxae WJT66-NPBG9	Solitary, mostly unbranched, rarely pseudo-branched, single trichome per sheath	2.0–3.6	1.2–4.0	Sometimes absent, thick when present	ND	Bright green	Dryland soils	Pietrasiak et al., 2021
Trichotorquatus sp. 5 WJT32-NPBGA	Unbranched, single trichome per sheath	2.0–2.8	1.6–6.0	Sometimes absent, thick when present	ND	Bright green	Dryland soils	Pietrasiak et al., 2021

(Continued)

10.3389/fmicb.2023.1176500

TABLE 2 (Continued)

Species	Morphology	Cell width (μm)	Cell length (μm)	Sheath	Thylakoids No.	Color	Habitat	References
"*Egbenema aeruginosa* N15-MA6"	Straight, with single trichome per sheath, infrequent false branching	2.0–2.6	1.2–2.8	Narrow, clear, tightly adherent to the trichomes	ND	Bright blue-green	Soils or subaerial surfaces in a tropical climate area	Akagha, 2022
"*Egbenema epilithica* CT225"	Straight and coiled, irregularly entangled or coiled, free or with sheaths	1.6–3.0	1.6–4.0	Narrow, clear, tightly adherent to the trichomes	ND	Blue-green	Soils or subaerial surfaces in a tropical climate area	Akagha, 2022
"*Egbenema aeruginosa*"	Straight and coiled, frequent false branching, free or with sheaths	2.0–2.6	1.8–3.6	Narrow, clear, tightly adherent to the trichomes	ND	Blue-green	Soils or subaerial surfaces in a tropical climate area	Akagha, 2022

The overall morphology of the described clade was similar to the previously described cyanobacterium *Albertania skiophila* (Zammit, 2018), with some notable exceptions. Light microscope images revealed that *A. skiophila* frequently exhibited a spiral shape, a feature, which was not observed in A174 and closely related strains. Moreover, *A. skiophila* grown at its physiological conditions exhibited cylindrical trichomes widened at their apices; meanwhile, the A174 was typically thinner than wide. Other strains belonging to *Albertania,* e.g., freshwater *A. alaskaenisis,* were characterized by apparent false branching and sometimes formed characteristic mushroom-like calyptra. These features were not observed in A174. Comparison of strain A174 to *Trichotorquatus* is problematic due to the absence of good-quality electron micrographs, but several distinguishing features can be inferred from light microscopy. *Trichotorquatus* strains described vary in their sheath description, ranging from thick to absent (Table 2). This may be attributed to the different life cycle stages and environmental conditions that these strains exhibited during taking the micrographs. Meanwhile, A174 had a defined colorless sheath under the conditions tested. Cell size, shape, and spacing were also distinguishing factors between these two genera. Moreover, the individual cells of *Trichotorquatus* were easily distinguished on light micrographs, while the cells were not well distinguishable in the A174. The cells of the representatives of a putative taxon "*Egbenema*" were often shorter than the wide based on the available light micrographs (Akagha, 2022). The average length and width were 1.9 μm and 2.2 μm in "*Egbenema aeruginosa*," which is significantly shorter than the long, cylindrical cells of A174. All "*Egbenema*" strains had visible false branches, and the width and length of each cell were very comparable. Trichomes of "*E. gypsiphila*" were mostly constricted at cross-walls.

The A174 strain was subjected to physiological tests using varying concentrations of bicarbonate and nitrate in different modifications of the BG11 medium (Table 3). It exhibited growth in most nitrogen conditions except for nitrite, and optimal growth was observed at 1.5 g/L nitrate concentration. The strain demonstrated a high tolerance for bicarbonate, with a maximum concentration of 0.7 M. When treated for 21 days with 1 M bicarbonate, the cells turned yellow-green and exhibited poor growth. The strain's tolerance to bicarbonate was attributed to its possession of bicarbonate transporters (BCT1, bicA, and sbtA). Furthermore, the strain displayed phototaxis when grown under directional light (Supplementary Figure 11).

Taxonomic position of A174 strain using polyphasic approach

In accordance with contemporary cyanobacterial taxonomy standards, Komárek et al. (2014) have defined the concept of a cyanobacterial genus based on three criteria. These criteria include the following: (i) a distinct taxonomic position with a discernible divergence (of 95% or less 16S rRNA gene sequence similarity) from the nearest sister clade, (ii) unique morphological traits or biological specificity (for example, type of cell division, formation of heterocyte or akinete, etc.) that distinguish it from other genera, and (iii) clear ecological niches that are relevant to the genus. The molecular 16S rRNA gene sequence similarities between the clade

TABLE 3 Physiological characteristic of *Thermocoleostomius sinensis* A174 grown in 45°C 150 μmol m^{-2} s^{-1}.

Dissolved inorganic carbon		Nitrogen source	
HCO_{3-} (mol/L)	Growth	NO_{3-} (g/L)	Growth
0	+++	0	++
0.1	++	0.075	++
0.3	+++	0.5	++
0.5	+++	1.5	+++
0.7	+	5	++
1.0	–	7.5	+
		1.218 (NO_{2-})	–

"–" no growth; "+" poor growth; "++" baseline growth in BG-11; "+ + +" better growth than baseline.

of interest and related clades of *Trichotorquatus* and *Albertania* were lower than 95% similarity. Delineation of A174 as a new genus was additionally supported with genomic similarity coefficients, phylogenomics, and secondary structures of D1–D1′ and boxB regions. Analysis of the habitat and morphological features reinforces the molecular evidence. Strains in *Trichotorquatus* genus were isolated from dryland soils in the desert or coastal scrub, which exhibit the sheath with a distinctive collar-like fraying and widening mid-filament. *Albertania skiophila* filaments formed biofilms on the rock surface at the crypt, and MAR67 was collected from gypsum blocks in the hyper-arid Sahara Desert. Meanwhile, Alaskan strains of *Albertania* isolated from willow roots appear unique morphologically with undulating trichomes with perpendicular cells and false branching in old cultures. Both related genera *Trichotorquatus* and *Albertania* have morphological and ecological niche differences that can support the delineation of the thermal clade containing the A174 strain as a new genus.

The exact separation between the clade and adjacent clade, putatively described as "*Egbenema*," is more problematic due to the lack of sufficient data on the strains that are not a part of this study. Analysis of the three aspects of genus-level delineation: molecular, morphological or physiological characteristics and habitat data, allows for drawing the following conclusions. Based on the 16S rRNA sequence identity between the Chinese strains (A174, A171, GDTS1–24, and GDTS1–29) and sister clade containing "*Egbenema*" (Figure 1), which ranges from 94.9 to 95.48%, indicating that based on the 16S rRNA sequence alone, it is uncertain if strains belong to the same genus. Analysis of the secondary structure of ITS helices D1–D1′ and BoxB reveals that the A174 strain has a more diverse asymmetrical additional bilateral bulge structure. The habitat data (Figure 1 and Table 2) show that strains in the focus clade originate from thermal environments supporting the ecological niche requirement for the genus delineation. Meanwhile, the strains proposed as "*Egbenema*" show diverse habitats. "*Egbenema*" strains were isolated from soil or subaerial samples in a tropical climate area and can survive in perennially wet habitat. There is also a clear morphological difference between A174 clade and putative "*Egbenema*" taxon. The members of A174 clade, contrary to strains described as "*Egbenema*," had no false branching, and their overall cell shape was different. Cells of the A174 clade are longer than wide, while those of "*Egbenema*" representatives had similar length and width. The terminal cells of the trichomes have also shown different morphologies. The characteristic pellet-forming feature of A174 was not observed in "*Egbenema*" according to the available literature.

To summarize, at least at the level of the strains with morphological data available, the focus clade of this study is consistent in morphology, habitat, and molecular data. A monophyletic cluster, comprising three strains (A171, A174, and A141) isolated from two hot springs in Erdaoqiao (Sichuan Province) and two closely related strains isolated from a single hot spring in Guangdong Province (Zhang et al., 2019), has been included in this clade. These strains exhibit similar morphology and 16S rRNA sequence identity of over 98.9% and have been found to originate from hot springs. This allows us to conclude that they belong to the same species. Consequently, the different morphology combined with very distinct environmental niches and some limited molecular evidence suggests that the clade containing A174 and the putative "*Egbenema*" clade belong to two different genera.

Genomic outlook on the carbon uptake and concentration in A174 strain

Due to the low solubility of gaseous CO_2 at elevated temperatures and higher ionic strengths, many thermophilic cyanobacteria deal with higher limitations of inorganic carbon than their mesophilic counterparts (Tang et al., 2022c). In response to this limitation, cyanobacteria frequently form microbial mats in hot springs where the microorganisms are stratified. The top layer is occupied by cyanobacteria that perform photosynthesis and release O_2, the subsequent layer contains heterotrophs that utilize oxygen for respiration and release CO_2 that is acquired by the top layer cyanobacteria (Ferris et al., 1997). In the absence of cohabitating organisms, cyanobacteria can also utilize carbon concentration mechanisms (CCM) that employ active transport to increase the amount of carbon dioxide in the carboxysome.

Like their mesophilic counterparts, thermophilic cyanobacteria perform photosynthesis in specialized protein-based structures called carboxysomes, where cytoplasm-stored bicarbonate is transported and converted into gaseous CO_2 by a carboxysomal carbonic anhydrase to generate high concentration of CO_2 around the main carboxylating enzyme, Rubisco (Figure 4). The large subunit of this protein, rbcL, has been used to discriminate between α- and β-cyanobacteria (Badger and Price, 2003). The *rbcL* gene of the A174 strain was extracted from the genome sequence and analyzed with other focus taxa using an ML phylogram (Supplementary Figure 4). The results have confirmed that the strain possesses Rubisco 1B form and thus belongs to β-cyanobacteria similarly to terrestrial, freshwater and other thermal strains in accordance with previous studies (Tang et al., 2022c). The gene itself is positioned within an operon containing genes

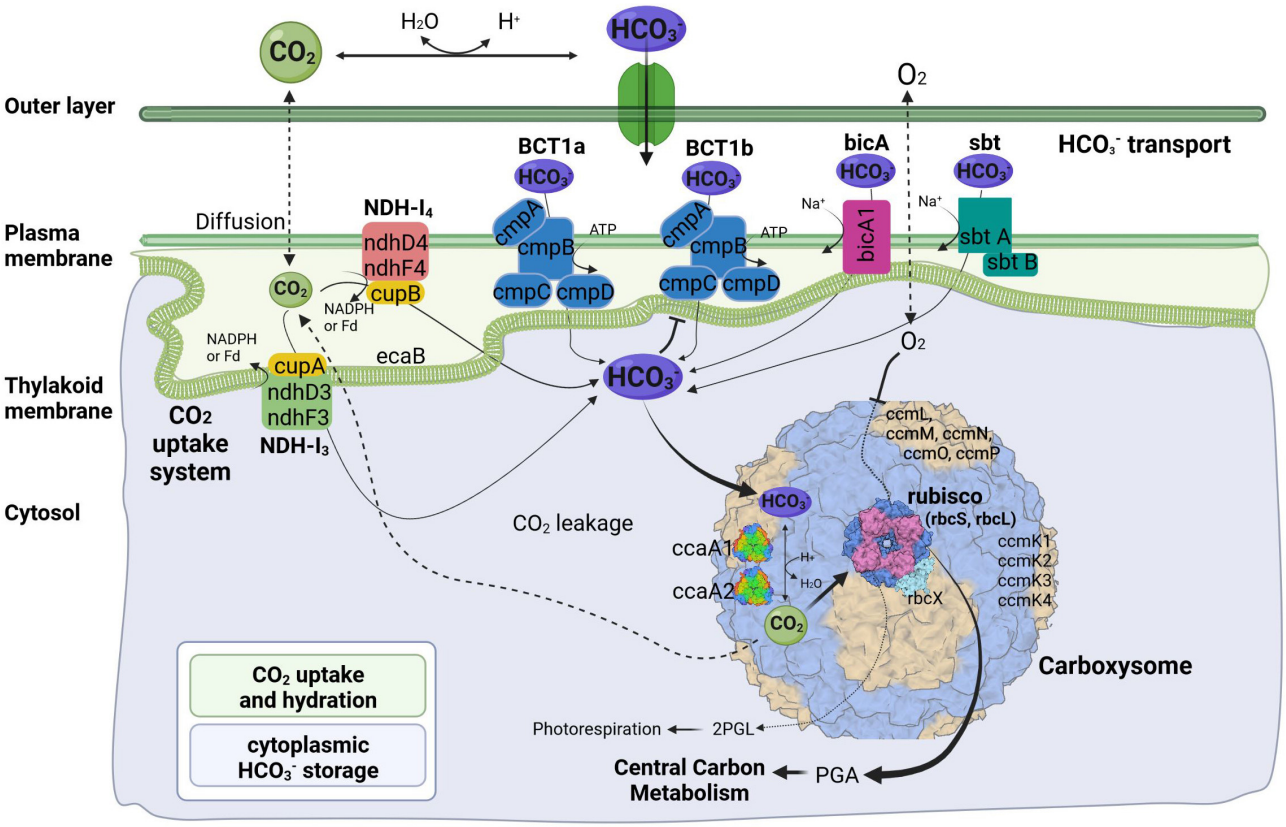


FIGURE 4
Predicted molecular components of CCM for *Thermocoleostomius sinensis* A174. 2PGL, 2-phosphoglycolate; PGA, phosphoglyceric acid. Three-dimensional structures of proteins visualized using related structures of 2YBV, 6OWF, 5SWC, 5BS1, 6KI1. Figure created with BioRender.com.

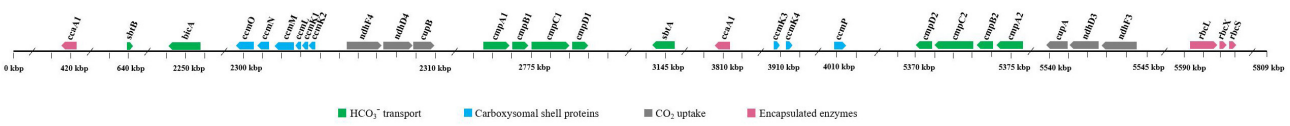


FIGURE 5
Genomic organization of CCM-related genes for *Thermocoleostomius sinensis* A174. Solid arrow boxes refer to genes and the direction of transcription.

sequentially encoding the Rubisco large subunit (*rbcL*), Rubisco chaperonin (*rbcX*), and Rubisco small subunit (*rbcS*) (Figure 5).

Cyanobacterial CCMs are primarily composed of two main components: inorganic carbon transport system responsible for carbon uptake and carboxysomes responsible for its fixation. Two complexes NDH-I_3 and NDH-I_4, responsible for CO_2 uptake, are ubiquitous in β-cyanobacteria, but there exists a significant diversity in the bicarbonate uptake systems (Price et al., 2008). Analysis of genomic components of the A174 strain reveals a similar pattern. The strain possesses a full complement of genes encoding NDH-I_3 (low-CO_2 inducible, high-affinity CO_2 uptake system) and NDH-I_4 (constitutive, low-affinity CO_2 uptake system). The amino acid sequence identity to the orthologous proteins from a mesophilic reference strain *Synechocystis* PCC 6803 of the constitutive ndhD4, ndhF4, cupB, and inducible ndhD3, and ndhF3 components varies from 55.8% (ndhF4) to 62.1% (cupB). Meanwhile, the carbonic anhydrase cupA associated with low-CO_2 inducible ndhD3 and ndhF3 is more conserved and exhibits an identity of 83.4%. These findings are in accordance with our previous findings on thermophilic strains (Tang et al., 2022c).

The genomic makeup of A174 bicarbonate transport system is diverse compared to other thermophilic strains and most similar to *Leptolyngbya* JSC-1 and *Thermoleptolyngbya* O-77 (Tang et al., 2022c). The strain shows a repertoire of known bicarbonate transporters: BCT1, bicA, and sbtA. Based on the genomic information alone, there is uncertainty about the number of ATP-dependent inducible high-affinity HCO_3^- transporters, BCT1. There are two regions of high homology to the BCT1 transporters of model strains *Synechocystis* PCC6803 and *Synechococcus* PCC7942 positioned in forms discrete operons at

2.77 and 5.37 Mbp (Figure 5). Both comprise four polypeptides cmpA-D, consistent with other thermophilic cyanobacteria (Tang et al., 2022c). The two putative bicarbonate transporters are the members of two distinct clades (Supplementary Figures 6–9), and their transcription profiles are further analyzed in the CO_2 shift experiment. In addition to BCT1, the strain possesses two genes of the sbt, the least abundant bicarbonate transporter in thermophilic cyanobacteria, *sbtA* and *sbtB*. The two copies are dispersed in distant loci of the genome. Finally, the A174 strain possesses a single bicA transporter (bicA1), similar to most thermophilic cyanobacteria except for *Thermoleptolyngbya* A183 and *Leptothermofonsia* E412, which have two gene copies of such transporter, *bicA1* and *bicA2*.

Carboxysomes are evolutionarily conserved protein structures that facilitate efficient carbon fixation in cyanobacteria. Carboxysome structures comprise a proteinaceous shell and two encapsulated enzymes, the main carboxylating protein Rubisco and carboxysomal carbonic anhydrase that transforms the soluble bicarbonate into gaseous CO_2 (Figure 4). Among β-cyanobacteria, the typical structure of shell proteins comprises ccmK-P proteins. Typically, *ccmKLMNO* genes form a discrete operon, and *ccmP* is found in a separate genomic locus. This arrangement is conserved in A174. The strain contains four *ccmK* genes, namely *ccmK1*, *K2*, *K3*, and *K4*. The genes coding for two main structural proteins *ccmK1*, *K2* (Cai et al., 2016), are conserved with other cyanobacteria but do not always cluster with their other thermophilic orthologs on the phylogenetic tree (Supplementary Figure 5). Meanwhile, the sequence of the remaining auxiliary shell proteins ccmK3 and K4 shows the highest similarity to the orthologous thermophilic proteins of *Leptolyngbya* JSC-1. When it comes to encapsulated enzymes, *rbcL* is clearly positioned within the thermophilic cluster of Rubisco but exhibits a relatively diverse sequence compared to other thermophilic cyanobacteria (Supplementary Figure 4) and 96.8% sequence identity to the enzyme of *Leptothermofonsia* E412. Interestingly, there appear to be two copies of carboxysomal carbonic anhydrase *ccaA* dispersed in two loci of the A174 genome, again highlighting higher diversity of strain's CCM than most other thermophilic cyanobacteria. Alignment of the two sequences reveals that the second copy of the gene is ~30% longer than the first (Supplementary Figure 10). More detailed studies will be required to validate the biological function of both carbonic anhydrases *in vivo*.

Gene expression profiles of the A174 strain during CO_2 shift experiment

The CO_2 shift experiment was performed to check the relative transcription levels of the A174 strain in response to carbon limitation. The cells were initially cultured in a high-carbon environment for 24 h (4% CO_2 in air) and were subsequently shifted to ambient air. The gene expressions of various bioinformatically identified genes involved in carbon uptake and conversion, including *bicA*, *ccaA1*, *ccaA2*, *cmpA1*, *cmpA2*, *ndhF3*, *ndhF4*, *rbcL*, and *sbtA*, were subsequently analyzed at two different time point: 1 h and 24 h after the shift. The data were normalized first to the relative abundance of the *rpoB* and subsequently to the transcript abundance under a carbon-replete environment (HC) and are presented in Figure 6.

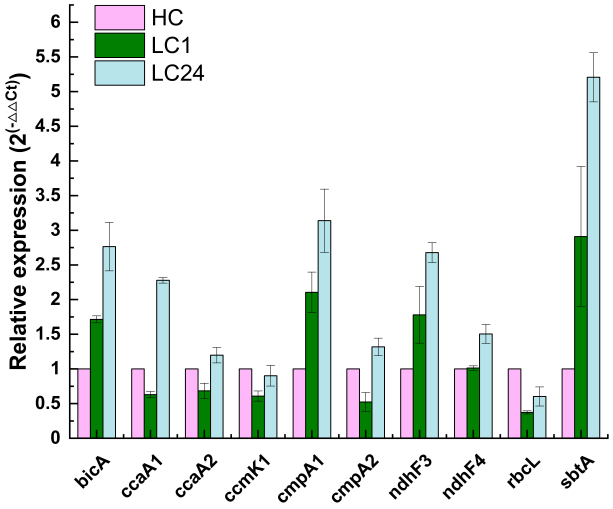


FIGURE 6
Relative changes in transcript abundance of *ndh*F3, *ndh*F4, *cmp*A1, *cmp*A2, *bic*A, *sbt*A, *rbc*L, *ccm*K1, *cca*A1, and *cca*A2 genes of *Thermocoleostomius sinensis* A174 cells during CO_2 shift experiment assessed with RT-qPCR. HC indicates high carbon for 24 h (4% CO_2), LC1 (1 h of carbon limitation), LC24 (24 h of carbon limitation). The data have been normalized to the reference gene *rpo*B and HC conditions.

The results showed that genes related to both low-affinity and high-affinity uptake systems responded positively to carbon limitation but in different magnitudes. NDH-I_3, the high-affinity system for gaseous CO_2 uptake, was found to be the dominant system in A174, with 2- to 3-fold increases in expression levels after 1 and 24 h of carbon limitation, respectively. These findings are in line with those for *Synechocystis* PCC6803 (McGinn et al., 2003), where the *ndhF3* gene is undergoing significant upregulation, and *ndhF4* is slightly downregulated during the CO_2 shift experiment. Similarly, among bicarbonate uptake systems, the transcript levels of high-affinity systems were higher than those of low-affinity ones. The gene expression level of *sbtA* was found to have increased the most compared to other transcripts assessed. Interestingly, the two putative *cmpA* genes exhibited different behaviors in response to carbon limitation, highlighting the need for further function loss studies when appropriate molecular tools become available.

Carboxysome proteins such as shell protein *ccmK1* and two encapsulated enzymes Rubisco (*rbcL*) and carbonic anhydrase *ccaA2* all show a similar pattern of transcription, with a relative drop in gene expression levels after 1 h of carbon limitation, followed by a recovery to baseline levels after 24 h, likely due to the sufficient supply of carbon by the transport components of the CCM system. Interestingly, the second of the carbonic anhydrases, *ccaA1*, exhibited markedly higher expression levels, suggesting a possible different cellular localization than suggested by bioinformatic analyses. Further functional studies are required to validate the function of both putative carbonic anhydrases on the carbon fixation of the A174 strain.

Conclusion

In this manuscript, we have characterized, using a polyphasic approach, a novel hot-spring cyanobacterium strain A174 as a representative of a novel taxon of thermophilic cyanobacteria. The subject of this study was capable of growth at 50°C and 0.7M bicarbonate. The combination of phylogenetic, taxogenomic, morphological, habitat, and physiological data suggests that the strain could belong to a novel genus of thermophilic cyanobacteria. Based on the results of 16S rRNA phylogeny, comparative genomics, secondary structures of 16S-23S ITS, and strain's distinct morphology to *Albertania*, *Trichotorquatus,* and recently proposed "*Egbenema*" support the delineation of the clade. Based on those data, we proposed a new genus *T. sinensis* as a best-described representative of this taxon and proposed its delineation within the family *Oculatellaceae*. Additionally, the inorganic carbon uptake and concentration genes have been extracted from the assembled high-quality genome sequence and analyzed. The results have shown that the strain firmly belongs to the β-cyanobacteria and has a diverse array of bicarbonate uptake proteins encompassing bicA, sbtA, and at least one variant of BCT1, and two carbonic anhydrases ccaA. The results of gene expression studies under CO_2 limitation revealed that all the bicarbonate transporters, bicA, sbtA, and BCT1 are likely to respond to carbon limitation but at different degrees, and the increase of transcript abundance in NDH-I_3 is higher than that of NDH-I_4, consistently with mesophilic counterparts.

Taxonomic treatment and description of *Thermocoleostomius sinensis* Daroch, Jiang, Tang et al. gen. et sp. nov.

The taxonomic classification methodology used for the strain proposed in this study follows the polyphasic approach as described by Komárek et al. (2014). The taxon description was consistent 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: *Oculatellaceae*

Genus: *Thermocoleostomius*, gen. nov.

Description: Filamentous cyanobacteria, forming small pellets and radiating outward, colonies intertwined, forming green and compact biofilms and new pellets. Filaments flexuous, wavy or bent, occasionally straight trichomes. Filaments with a single trichome per sheath, no false branching. Sheath frequently open at the trichome termini. Cylindrical trichomes, cross-walls of cells with minor constrictions, cells constricted toward the end. Cells variable in shape, mostly isodiametric to longer than wide, with parietal thylakoids. Apical cell rounded, shorter on average than regular vegetative cells, morphologically similar to regular cells.

Etymology: "Thermo" similar to thermophilic (high temperature tolerant), "coleo" genus epithet derived from the Greek word koleos meaning sheath or scabbard, "stomius" genus epithet derived from Greek word meaning mouth, outlet, indicating that the sheath of the strain is frequently open.

Type species: *Thermocoleostomius sinensis*.

Thermocoleostomius sinensis Daroch, Jiang, Tang et al. gen. et. sp. nov.

Diagnosis: Differing from other species of the genus based on the 16S rRNA sequence identity.

Description: Filamentous cyanobacteria (Figures 3A–F), in laboratory cultures, originating from small pellets and radiating outward in diverse directions (Supplementary Figure 14), tiny colonies intertwined and connected, forming green and compact biofilms, appearance of new pellets on solid agar plates (Supplementary Figure 11). Filaments flexuous, wavy or bent, occasionally straight trichomes (Figures 3A, D). Filaments with a single trichome per sheath, no false branching (Figures 3A–F). Sheath frequently open at the trichome termini (Figure 3D). Cylindrical trichomes, cross-walls of cells with minor constrictions, cells slightly constricted toward the end, bright green in healthy cultures, yellow in senescent cultures or in a high light environment. Cells variable in shape, mostly isodiametric to longer than wide (Figures 3A–F), length-to-width ratio 0.8:2.3. Cells size 1.5–2.5 μm long, 2.0–5.8 μm wide, parietal thylakoids arrayed in parallel lines (Figures 3C, E, F). Apical cell rounded, shorter on average than regular vegetative cells, morphologically similar to regular cells (Figure 3E). Strain capable of growth in standard BG11 medium to a maximum temperature of 50°C and medium with 0.7 M $NaHCO_3$. Successful cryopreservation of the strain in 10% DMSO for over 2 years. Phototaxis observed (Supplementary Figure 11).

Etymology: "*sinensis*" species epithet derives from the fact that all representatives of the genus have been isolated from various provinces of China.

***Type locality*:** Thermal spring. Erdaoqiao village in Ganzi Prefecture of Sichuan Province, China.

Ecology of type locality: The sample occurred as a light green biofilm deposited on the surface of calcareous sinter (Supplementary Figure 12).

***Habitat*:** Thermal spring in Ganzi Prefecture of Sichuan Province, China (30°05′14″N, 101°56′55″E).

Holotype here designated: The dried inactive holotype was deposited in the Herbarium of North Minzu University with the voucher number: NMU00174 (contact: Lei Zhang, zhangsanshi-0319@163.com).

Reference strain: The culture *of Thermocoleostomius sinensis* Daroch, Jiang, Tang et al. gen. et sp. nov. was initially denoted and deposited in Peking University Algae Collection as PKUAC-SCTA174, and it also has been deposited in the Freshwater Algae Culture Collection at the Institute of Hydrobiology (FACHB) as *Oculatellaceae* sp. FACHB-3572 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.

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

YJ: formal analysis, software, investigation, data curation, visualization, writing—original draft, and writing—reviewing and editing. JT: conceptualization, methodology, validation, formal analysis, investigation, data curation, writing—original draft, writing—reviewing and editing, visualization, supervision, project administration, and funding acquisition. MD: conceptualization, methodology, resources, data curation, writing—original draft, writing—reviewing and editing, supervision, project administration, and funding acquisition.

Funding

This research was funded by the National Natural Science Foundation of China (31970092, 32071480, and 3221101094) and Tenure-Track Fund to MD. Funding bodies had no influence over the design and execution of this research.

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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1176500/full#supplementary-material

References

Akagha, M. (2022). "*Albertania*" *and* "*Egbenema*," *Panding Biodiversity in Oculatellaceae (Cyanobacteria)*. John Carroll University.

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

Badger, M. R., and Price, G. D. (2003). CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.* 54, 609–622. doi: 10.1093/jxb/erg076

Brettin, T., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Olsen, G. J., et al. (2015). RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* 5, 8365. doi: 10.1038/srep08365

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

Cai, F., Bernstein, S. L., Wilson, S. C., and Kerfeld, C. A. (2016). Production and characterization of synthetic carboxysome shells with incorporated luminal proteins. *Plant Physiol.* 170, 1868–1877. doi: 10.1104/pp.15.01822

Casamatta, D. (2023). Giving form to the formless: an updated classification of cyanobacterial taxonomy. *J. Phycol.* 59, 9–11. doi: 10.1111/jpy.13313

Chatterjee, S., Alampalli, S. V., Nageshan, R. K., Chettiar, S. T., Joshi, S., Tatu, U. S., et al. (2015). Draft genome of a commonly misdiagnosed multidrug resistant pathogen *Candida auris*. *BMC Genomics* 16, 686. doi: 10.1186/s12864-015-1863-z

Chen, M.-Y. Y., Teng, W.-K. K., Zhao, L., Hu, C.-X. X., Zhou, Y.-K. K., Han, B.-P. P., et al. (2020). 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., Chou, L., Chiu, Y. F., Hsueh, H. T., Kuo, C. H., Chu, H. A., et al. (2020). Comparative genomic analysis of a novel strain of taiwan hot-spring cyanobacterium *Thermosynechococcus* sp. CL-1. *Front. Microbiol.* 11, 1–13. doi: 10.3389/fmicb.2020.00082

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

Ferris, M. J., Nold, S. C., Revsbech, N. P., and Ward, D. M. (1997). Population structure and physiological changes within a hot spring microbial mat community following disturbance. *Appl. Environ. Microbiol.* 63, 1367–1374. doi: 10.1128/aem.63.4.1367-1374.1997

Guindon, S., Dufayard, J.-F. F. F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., et al. (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

Iteman, I., Rippka, R., Marsac, d. e., Herdman, N. T., Tandeau, M., Herdman, D. M. N., et al. 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

Jahodárová, E., Dvorák, P., Hašler, P., Holušová, K., and Poulíčková, A. (2018). *Elainella* gen. nov.: a new tropical cyanobacterium characterized using a complex genomic approach. *Eur. J. Phycol.* 53, 39–51. doi: 10.1080/09670262.2017.1362591

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

Johnson, J. S., Spakowicz, D. J., Hong, B. Y., Petersen, L. M., Demkowicz, P., Chen, L., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat. Commun.* 10, 5029. doi: 10.1038/s41467-019-13036-1

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

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199

Kees, E. D., Murugapiran, S. K., Bennett, A. C., and Hamilton, T. L. (2022). Distribution and genomic variation of thermophilic cyanobacteria in diverse microbial mats at the upper temperature limits of photosynthesis. *mSystems* 7, e0031722. doi: 10.1128/msystems.00317-22

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

Komárek, J., and Anagnostidis, K. (2005). *Cyanoprokaryota 2. Teil/2ndPart : Oscillatoriales*. München: SpringerSpektrum.

Komárek, J., Johansen, J. R., Šmarda, J., and Strunecký, O. (2020). Phylogeny and taxonomy of Synechococcus-like cyanobacteria. *Fottea*. 20, 171–191. doi: 10.5507/fot.2020.006

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., Kaštovský, J., Ventura, S., Turicchia, S., and Šmarda, J. (2009). The cyanobacterial genus *Phormidesmis*. *Arch. Hydrobiol. Suppl. Algol. Stud.* 129, 41–59. doi: 10.1127/1864-1318/2009/0129-0041

Kono, M., Martinez, J., Sato, T., and Haruta, S. (2022). Draft genome sequence of the thermophilic unicellular cyanobacterium *Synechococcus* sp. strain C9. *Microbiol. Resour. Announc*. 11, e0029422. doi: 10.1128/mra.00294-22

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

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

Liang, Y., Kaczmarek, M. B., Kasprzak, A. K., Tang, J., Shah, M. M. R., Jin, P., et al. (2018). Thermosynechococcaceae as a source of thermostable C-phycocyanins : properties and molecular insights. *Algal Res.* 35, 223–235. doi: 10.1016/j.algal.2018.08.037

Liang, Y., Tang, J., Luo, Y., Kaczmarek, M. B., Li, X., Daroch, M., et al. (2019). Thermosynechococcus as a thermophilic photosynthetic microbial cell factory for CO 2 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 inproved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.955

Mai, T., Johansen, J. R., Pietrasiak, N., Bohunicka, M., and Martin, M. P. (2018). Revision of the *Synechococcales* (cyanobacteria) through recognition of four families including *Oculatellaceae* fam. nov., *Trichocoleaceae* fam nov., seven new genera containing 14 species. *Phytotaxa* 365, 1–59. doi: 10.11646/phytotaxa.365.1.1

Mathews, D. H. (2014). RNA secondary structure analysis using RNAstructure. *Curr. Protoc. Bioinformatics* 46, 1–25. doi: 10.1002/0471250953.bi1206s46

McGinn, P. J., Price, G. D., Maleszka, R., and Badger, M. R. (2003). Inorganic carbon limitation and light control the expression of transcripts related to the CO2-concentrating mechanism in the cyanobacterium *Synechocystis* sp. strain PCC6803. *Plant Physiol.* 132, 218–229. doi: 10.1104/pp.019349

Mehda, S., Muñoz-Martín, M. Á., Oustani, M., Hamdi-Aïssa, B., Perona, E., Mateo, P., et al. (2022). Lithic cyanobacterial communities in the polyextreme Sahara Desert: implications for the search for the limits of life. *Environ. Microbiol.* 24, 451–474. doi: 10.1111/1462-2920.15850

Meier-Kolthoff, J. P., Carbasse, J. S., Peinado-Olarte, R. L., and Göker, M. (2022). TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acids Res.* 50, D801–D807. doi: 10.1093/nar/gkab902

Miller, M. A., Pfeiffer, W., and Schwartz, T. (2010). "Creating the CIPRES science gateway for inference of large phylogenetic trees," in *2010 Gateway Computing Environments Workshop*, GCE 2010 (New Orleans, LA: IEEE). doi: 10.1109/GCE.2010.5676129

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. *Acta Botanica Hungarica*. 58, 3–4.

Nowicka-Krawczyk, P., Mühlsteinová, R., and Hauer, T. (2019). Detailed characterization of the Arthrospira type species separating commercially grown taxa into the new genus Limnospira (Cyanobacteria). *Scient. Rep*. 9, 694. doi: 10.1038/s41598-018-36831-0

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, 733–745. doi: 10.1093/nar/gkv1189

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

Perona, E., Muñoz-Martín, M. Á., and Berrendero Gómez, E. (2022). "Recent trends of polyphasic approach in taxonomy and cyanobacterial diversity," in *Expanding Horizon of Cyanobacterial Biology*, eds P. K. Singh, M. F. Fillat, and V. Sitther (Amsterdam: Elsevier), 1–49. doi: 10.1016/B978-0-323-91202-0.00008-7

Pietrasiak, N., Osorio-Santos, K., Shalygin, S., Martin, M. P., and Johansen, J. R. (2019). When is a lineage a species? A case study in *Myxacorys* gen. nov. (Synechococcales: Cyanobacteria) with the description of two new species from the Americas. *J. Phycol.* 55, 976–996. doi: 10.1111/jpy.12897

Pietrasiak, N., Reeve, S., Osorio-Santos, K., Lipson, D. A., and Johansen, J. R. (2021). Trichotorquatus gen. nov. - a new genus of soil cyanobacteria discovered from American drylands. *J. Phycol.* 57, 886–902. doi: 10.1111/jpy.13147

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

Raabova, L., Kovacik, L., Elster, J., Strunecky, O., Raabová, L., Kovacik, L., et al. (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

Rambaut, A. (2007). *FigTree, a Graphical Viewer of Phylogenetic Trees*.

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., 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

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

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029

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

Sciuto, K., Moschin, E., and Moro, I. (2017). Cryptic cyanobacterial diversity in the giant cave (Trieste, Italy): the new genus *Timaviella* (Leptolyngbyaceae). *Cryptogam. Algol.* 38, 285–323. doi: 10.7872/crya/v38.iss4.2017.285

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

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. doi: 10.11646/phytotaxa.419.1.5

Stanier, R. Y., Kunisawa, R., Mandel, M., and Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol. Rev.* 35, 171–205. doi: 10.1128/br.35.2.171-205.1971

Strunecký, O., Ivanova, A. P., and Mareš, J. (2023). An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis. *J. Phycol.* 59, 12–51. doi: 10.1111/jpy.13304

Strunecky, O., Raabova, L., Bernardova, A., Ivanova, A. P., Semanova, A., Crossley, J., et al. (2019). Diversity of cyanobacteria at the Alaska North Slope with description of two new genera: *Gibliniella* and *Shackletoniella*. *FEMS Microbiol. Ecol.* 96, iz189. doi: 10.1093/femsec/fiz189

Tang, J., Du, L.-M. L., Li, M., Yao, D., Waleron, M., Waleron, K. F., et al. (2022a). Characterization of a novel hot-spring cyanobacterium *Leptodesmis sichuanensis* sp. *Nov.,* genomic insights of molecular adaptations into its habitat. *Front. Microbiol.* 12, 739625. doi: 10.3389/fmicb.2021.739625

Tang, J., Du, L.-M. L.-M., Liang, Y.-M. 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., Li, L., Li, M., Du, L., Shah, M. R. M. M. R., Waleron, M. M. 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

Tang, J., Liang, Y., Jiang, D., Li, L., Luo, Y., Shah, M. M. R., et al. (2018b). Temperaturecontrolled thermophilic bacterial communities in hot springs of western Sichuan, China. *BMC Microbiol.* 18, 134. doi: 10.1186/s12866-018-1271-z

Tang, J., Shah, M. R., Yao, D., Du, L., Zhao, K., Li, L., et al. (2022b). 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

Tang, J., Zhou, H., Jiang, Y., Yao, D., Waleron, K. F., Du, L. M., et al. (2023). Characterization of a novel thermophilic cyanobacterium within Trichocoleusaceae, *Trichothermofontia sichuanensis* gen. et sp. nov., its CO2-concentrating mechanism. *Front. Microbiol*. 14, 1111809. doi: 10.3389/fmicb.2023.1111809

Tang, J., Zhou, H., Yao, D., Riaz, S., You, D., Klepacz-Smółka, A., et al. (2022c). Comparative genomic analysis revealed distinct molecular components and organization of CO2-concentrating mechanism in thermophilic cyanobacteria. *Front. Microbiol.* 13, 876272. doi: 10.3389/fmicb.2022.876272

Tawong, W., Pongcharoen, P., Nishimura, T., and Saijuntha, W. (2022). *Siamcapillus rubidus* gen. et sp. nov. (Oculatellaceae), a novel filamentous cyanobacterium from Thailand based on molecular and morphological analyses. *Phytotaxa* 558, 33–52. doi: 10.11646/phytotaxa.558.1.2

Thi Hoang, D., Chernomor, O., von Haeseler, A., Quang Minh, B., Sy Vinh, L., Rosenberg, M. S., et al. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol* 35, 518–522. doi: 10.1093/molbev/msx281

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*. Koeltz botanical Books.

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., *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

Walter, J. M., Coutinho, F. H., Dutilh, B. E., Swings, J., Thompson, F. L., Thompson, C. C., et al. (2017). Ecogenomics and taxonomy of cyanobacteria phylum. *Front. Microbiol.* 8, 2132. doi: 10.3389/fmicb.2017.02132

Zammit, G. (2018). Systematics and biogeography of sciophilous cyanobacteria; an ecological and molecular description of *Albertania skiophila* (Leptolyngbyaceae) gen. and sp. nov. *Phycologia* 57, 481–491. doi: 10.2216/17-125.1

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, Y., Jiang, D., Tang, J., Luo, Y., Liang, Y., Shah, M. M. R., et al. (2019). Isolation and characterization of two thermophilic Leptolyngbyaceae strains isolated from Huizhou area, China (in Chinese). *Microbiology* 46, 481–493. doi: 10.13344/j.microbiol.china.180189

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 20 September 2023
DOI 10.3389/fmicb.2023.1216591

OPEN ACCESS

EDITED BY
Brian P. Hedlund,
University of Nevada, Las Vegas, United States

REVIEWED BY
Steffen Buessecker,
Stanford University, United States
Song Zhaoqi,
Shangqiu Normal University, China
En-Min Zhou,
Yunnan University, China

*CORRESPONDENCE
Rebecca D. Prescott
✉ becks@olemiss.edu
Jimmy H. Saw
✉ jsaw@gwu.edu

†PRESENT ADDRESS
Rebecca D. Prescott,
Department of Biology, University of Mississippi, Oxford, MS, United States

‡These authors have contributed equally to this work

RECEIVED 04 May 2023
ACCEPTED 30 August 2023
PUBLISHED 20 September 2023

CITATION
Balbay MG, Shlafstein MD, Cockell C, Cady SL, Prescott RD, Lim DSS, Chain PSG, Donachie SP, Decho AW and Saw JH (2023) Metabolic versatility of *Caldarchaeales* from geothermal features of Hawai'i and Chile as revealed by five metagenome-assembled genomes. *Front. Microbiol.* 14:1216591. doi: 10.3389/fmicb.2023.1216591

Metabolic versatility of *Caldarchaeales* from geothermal features of Hawai'i and Chile as revealed by five metagenome-assembled genomes

Manolya Gul Balbay[1‡], Maximillian D. Shlafstein[1‡], Charles Cockell[2], Sherry L. Cady[3], Rebecca D. Prescott[2,4,5,6*†], Darlene S. S. Lim[7], Patrick S. G. Chain[8], Stuart P. Donachie[4], Alan W. Decho[5] and Jimmy H. Saw[1*]

[1]Department of Biological Sciences, The George Washington University, Washington, DC, United States, [2]UK Centre for Astrobiology, University of Edinburgh, Edinburgh, United Kingdom, [3]Department of Geology, Portland State University, Portland, OR, United States, [4]School of Life Sciences, University of Hawai'i at Mānoa, Honolulu, HI, United States, [5]Department of Environmental Health Sciences, University of South Carolina, Columbia, SC, United States, [6]Department of Biology, University of Mississippi, Oxford, MS, United States, [7]NASA Ames Research Center, Moffett Field, CA, United States, [8]Los Alamos National Laboratory, Los Alamos, NM, United States

Members of the archaeal order *Caldarchaeales* (previously the phylum Aigarchaeota) are poorly sampled and are represented in public databases by relatively few genomes. Additional representative genomes will help resolve their placement among all known members of *Archaea* and provide insights into their roles in the environment. In this study, we analyzed 16S rRNA gene amplicons belonging to the *Caldarchaeales* that are available in public databases, which demonstrated that archaea of the order *Caldarchaeales* are diverse, widespread, and most abundant in geothermal habitats. We also constructed five metagenome-assembled genomes (MAGs) of *Caldarchaeales* from two geothermal features to investigate their metabolic potential and phylogenomic position in the domain *Archaea*. Two of the MAGs were assembled from microbial community DNA extracted from fumarolic lava rocks from Mauna Ulu, Hawai'i, and three were assembled from DNA obtained from hot spring sinters from the El Tatio geothermal field in Chile. MAGs from Hawai'i are high quality bins with completeness >95% and contamination <1%, and one likely belongs to a novel species in a new genus recently discovered at a submarine volcano off New Zealand. MAGs from Chile have lower completeness levels ranging from 27 to 70%. Gene content of the MAGs revealed that these members of *Caldarchaeales* are likely metabolically versatile and exhibit the potential for both chemoorganotrophic and chemolithotrophic lifestyles. The wide array of metabolic capabilities exhibited by these members of *Caldarchaeales* might help them thrive under diverse harsh environmental conditions. All the MAGs except one from Chile harbor putative prophage regions encoding several auxiliary metabolic genes (AMGs) that may confer a fitness advantage on their *Caldarchaeales* hosts by increasing their metabolic potential and make them better adapted to new environmental conditions. Phylogenomic analysis of the five MAGs and over 3,000 representative archaeal genomes showed the order *Caldarchaeales* forms a monophyletic group that is sister to the clade comprising

the orders *Geothermarchaeales* (previously *Candidatus* Geothermarchaeota), *Conexivisphaerales* and *Nitrososphaerales* (formerly known as Thaumarchaeota), supporting the status of *Caldarchaeales* members as a clade distinct from the Thaumarchaeota.

KEYWORDS

Aigarchaeota, *Caldarchaeales*, extremophiles, fumaroles, hot springs, metabolism

Introduction

The archaeal candidate phylum Aigarchaeota belongs to the TACK superphylum, which initially consisted of the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota phyla (Guy and Ettema, 2011; Adam et al., 2017). The TACK superphylum now comprises several additional phyla discovered since 2011, including Bathyarchaeota (Meng et al., 2014), Verstraetearchaeota (Vanwonterghem et al., 2016), Geothermarchaeota (Jungbluth et al., 2017), and Nezhaarchaeota (Wang et al., 2019). Representatives of the Aigarchaeota, originally designated HWCG-I (Hot Water Crenarchaeotic Group I) were first discovered in a microbial mat collected from a moderately acidic geothermal stream in a subsurface gold mine (Nunoura et al., 2011). A genome fragment of HWCG-I in a fosmid clone from the metagenomic library of a microbial mat community thriving in hydrothermal fluid at the gold mine was sequenced and annotated (Nunoura et al., 2005). Intriguingly, one of the genes identified in the genome encodes a potential aerobic-type carbon monoxide dehydrogenase, which indicates that HWCG-I might be capable of chemolithotrophic growth using carbon monoxide as an electron donor and molecular oxygen as an electron acceptor, and that HWCG-I might be a facultative or obligate aerobe (Nunoura et al., 2005). Later, a composite circular genome sequence of an HWCG-I member, *Candidatus* "Caldiarchaeum subterraneum" (hereafter abbreviated as *Ca. C. subterraneum*), was assembled from the metagenomic library (Nunoura et al., 2011). Characterization of the genome of *Ca. C. subterraneum* revealed that the organism may perform hydrogenotrophy, aerobic carbon monoxide oxidation, aerobic respiration, and anaerobic respiration via nitrate or nitrite reduction (Nunoura et al., 2011). These archaea may also perform carbon fixation using the dicarboxylate/4-hydroxybutyrate pathway, though it lacks a key marker enzyme (4-hydroxybutyryl-CoA dehydratase) for this carbon fixation pathway.

Even though the presence of NiFe hydrogenases in the genome of *Ca. C. subterraneum* led to the suggestion that this organism is capable of hydrogenotrophic lifestyle (Nunoura et al., 2011), the hydrogenases in *Ca. C. subterraneum* were found to belong to Group 3B and Group 4 NiFe hydrogenases, which are involved in the regulation of redox homeostasis rather than hydrogenotrophy (Vignais, 2008; Hedlund et al., 2015). Nunoura et al. (2011) also identified genes previously found only in eukaryotes, such as a potential ubiquitin protein modification system in *Ca. C. subterraneum*. The genome of HWCG-I was proposed to belong to a novel candidate archaeal phylum, "Aigarchaeota," since the genome of *Ca. C. subterraneum* possessed unique characteristics that distinguish it from the previously described genomes of the phyla Crenarchaeota, Euryarchaeota, Thaumarchaeota, and Korarchaeota (Nunoura et al., 2011).

A single-cell genomics study, which generated 14 Aigarchaeota single-cell amplified genomes (SAGs) from sediments (~75°C–85°C) in Great Boiling Spring, Nevada (Rinke et al., 2013) revealed that they constitute five different species-level groups, and that each of those species-level groups is distinct from *Ca. C. subterraneum* (Rinke et al., 2013; Hedlund et al., 2014). Characterization of the SAGs revealed the presence of a large subunit of the ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) enzyme in a SAG, indicating that the Aigarchaeota may have the potential to fix carbon through the Calvin-Benson-Bassham cycle (CBB; Rinke et al., 2013). Some of the SAGs also have genomic potential for aerobic respiration, anaerobic respiration via dissimilatory sulfite reduction and reduction of nitrous oxide, and heterotrophic utilization of proteins and sugars (Rinke et al., 2013; Hedlund et al., 2014, 2015). In 2016, a novel member of Aigarchaeota, *Candidatus* "Calditenuis aerorheumens" was described based on both metagenome and metatranscriptome analyses of the hyperthermophilic, filamentous "pink streamer" communities in the Octopus Spring, Yellowstone National Park (Beam et al., 2016). It was reported that this archaeon could be an aerobic chemoorganotroph with autotrophic potential and is likely to able to utilize an array of organic carbon substrates, including acetate, fatty acids, amino acids and sugars, and thus may be important in cycling dissolved organic carbon (Beam et al., 2016).

Phylogenetic, phylogenomic, and comparative genomic studies have consistently unraveled a deep relationship between Thaumarchaeaota, Aigarchaeota, Crenarchaeota, and Korarchaeota in the "TACK superphylum" (Guy and Ettema, 2011; Wolf et al., 2012; Rinke et al., 2013; Hedlund et al., 2015; Petitjean et al., 2015), yet it remains controversial whether Aigarchaeota represents an independent phylum or a subclade of the phylum Thaumarchaeota (Brochier-Armanet et al., 2011; Guy and Ettema, 2011; Gribaldo and Brochier-Armanet, 2012; Spang et al., 2013; Hedlund et al., 2015). Distinct ecophysiologies of these two lineages were evident from an analysis of Aigarchaeota genome bins recovered from hot spring sediments in Tengchong, China, which revealed a strict or facultatively anaerobic lifestyle, sulfide oxidation for energy conservation, and substantial gene loss from their early ancestors (Hua et al., 2018). They also displayed diversity both in metabolic pathways and ecological roles, indicating functional partitioning and ecological divergence within a single geothermal region (Hua et al., 2018). Evolutionary genomic analyses of Aigarchaeota and its sister lineage Thaumarchaeota suggested that both phyla originated in thermal environments, sharing a large proportion of gene families with their thermophilic last common ancestor and later migration and adaptation of Thaumarchaeota to a wide range of non-thermal habitats led to the functional differentiation between these two groups of *Archaea* (Hua et al., 2018). Phylogenetic and phylogenomic analyses

of 14 Aigarchaeota and 80 Thaumarchaeota genomes in the same study provided evidence that they are different phyla (Hua et al., 2018). However, a more recent effort to unify taxonomic classification of prokaryotes led to the phylum Aigarchaeota being demoted and being reclassified and renamed as the order *Candidatus* Caldarchaeales in the Genome Taxonomy Database (GTDB; Rinke et al., 2021), and we will refer to it as the order *Caldarchaeales* without the Candidatus designation throughout the text for brevity.

Here, we report previously uncharacterized metagenome-assembled genomes (MAGs) of members of *Caldarchaeales* obtained from geothermal habitats from Hawai'i and Chile and showed that they have genomic potential for metabolic versatility. We also showed that the addition of these previously uncharacterized MAGs led to improved phylogenomic resolution of the order *Caldarchaeales* and increased the representation of poorly sampled members of this enigmatic archaeal group.

Materials and methods

Sample collection and processing

Fieldwork, sampling, and DNA extraction of lava rock samples from active fumaroles in Mauna Ulu volcanic area in Hawai'i were previously reported, and the MAGs in this study originated from sample numbered 86B (Cockell et al., 2019; Hughes et al., 2019). Biofilm samples from two hot springs in El Tatio geothermal field in Chile were collected during a field trip in August 2018. The MAGs reported here came from samples collected at two different hot spring features: Cacao East 5 (abbreviated as CE5; GPS coordinates of −22.350367, −68.008050) and Poppy (GPS coordinates of −22.333478, −68.013011) pools. A field photo of Poppy pool is shown in Supplementary Figure S1 and that of CE5 has been reported previously (Megevand et al., 2022). At the time of collection, the main CE5 pool had a temperature of 83.4°C and pH of 7.2, whereas the main Poppy pool had a temperature of 78.3°C and pH of 6.85. Sample material for DNA extraction from Poppy consisted of a brown-colored biofilm scraped from a piece of hot spring sinter that was fractured with sterile tweezers off the rim of the actively splashing hot spring. Sample material for DNA extraction from the more quiescent CE5 pool was collected with a sterile tweezer as a vertical section of a completely submerged hot spring mat located just below the water line at the pool's edge. After collection, the hot spring sinter and mat samples were placed into sterile Qorpak jars (Berlin Packaging, Clinton, PA), and transported within a few hours to a hotel where they were stored at ~4°C overnight. The sample jars were then transported on ice to the PNNL laboratory where they were stored at −80°C until they were shipped on blue ice to the University of Hawai'i laboratory, where they remained frozen until they were processed for DNA extraction. DNA was extracted from 0.5 g of each sample using the Qiagen DNeasy PowerSoil DNA kit and quantified with a Qubit instrument.

Sequencing

Illumina libraries of samples from Chile were prepared using NEBNext Ultra DNA II Library Preparation Kit (New England Biolabs, Cat. #E7645L). Sequencing libraries were generated from sample material that had input DNA amount varying between 5 and 100 ng. The DNA was fragmented with a Covaris E220, the ends made blunt, and adapters and indexes added to blunt ends for sequencing on an Illumina sequencer. Illumina libraries were eluted in DNA Elution Buffer (Zymo Research, Cat. #D3004-4-10), and concentrations of the libraries were obtained using the Qubit dsDNA HS Assay (ThermoFisher Scientific, Cat. #Q32854). The average size of the DNA in the library was determined by the Agilent High Sensitivity DNA Kit (Agilent, Cat. #5067–4,626). Libraries were quantified with the Library Quantification Kit – Illumina/Universal Kit (KAPA Biosystems, KK4824), and sequenced using a NextSeq v2.5 Reagent Cartridge (Illumina, Cat. #20024908).

Metagenome assembly, binning, and quality checks

Raw metagenomic reads were first trimmed and filtered for contamination using BBTools v38.87 (Bushnell, 2014). Quality filtering excluded adapter contaminants, read regions with scores < 20, and all reads under 50 bp. The resulting paired metagenomic reads were assembled using MetaSPAdes v3.14.0 (Nurk et al., 2017) with kmers specified as 21, 33, 55, and 77. MetaBAT2 v2.15 (Kang et al., 2019) was used to create genome bins from the assembled contigs by first using Seqtk v1.3 (Li, 2018) to remove contigs under 1 kb, then mapping the filtered contigs using BBTools v38.87 (Bushnell, 2014) with default parameters. Samtools v1.10 (Li et al., 2009) was used to sort and index the mapped contigs using default parameters, and BBTools v38.87 (Bushnell, 2014) then generated contig depths for all contigs over 1.5 kb. This contig depth file containing mean and variance of the base coverage depth was used in generating genome bins with MetaBAT2 v2.15 (Kang et al., 2019) with otherwise default parameters. The lineage workflow parameter of CheckM v1.1.3 (Parks et al., 2015) was used to assess completeness and contamination of the binning results to ensure bin quality for downstream analysis.

GTDB-Tk v1.4.0 (Chaumeil et al., 2020) was used to assign taxonomic affiliation to all genome binning results. All bins identified by GTDB-Tk as *Caldarchaeales* were selected for downstream analysis. QUAST v5.0.2 (Gurevich et al., 2013) was used to determine the number of scaffolds, genome size, N50 value, and genomic G + C content. Estimated genome size was determined from CheckM (Parks et al., 2015) completeness and the genome size result from the QUAST analysis. FastANI (Jain et al., 2018) was used to compute ANI values between the five *Caldarchaeales* MAGs and 79 *Caldarchaeales* genomes downloaded from the GTDB release 207_v2.

Genome annotation and metabolic reconstruction

Prodigal (Hyatt et al., 2010), with the "-p meta" option, was used to predict protein-coding genes in the five *Caldarchaeales* MAGs. All predicted proteins were analyzed with InterProScan (Jones et al., 2014) with default parameters to annotate protein domains and assigned to archaeal clusters of orthologous genes (arCOGs; Makarova et al., 2015b) by eggNOG-mapper v2.1.9 (Cantalapiedra et al., 2021) with default settings. Protein-coding genes were also annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database

(Kanehisa and Goto, 2000) tool GhostKOALA v2.2 (Kanehisa et al., 2016) and METABOLIC (Zhou et al., 2022); and annotations for genes of interest were confirmed by the BLASTp (Altschul et al., 1997, 2005) searches that were performed against the following databases: NCBI-nr (Sayers et al., 2021), UniRef100 (Suzek et al., 2015), and UniProtKB reference proteomes plus Swiss-Prot (UniProt Consortium, 2021; E-value cutoff $\leq 1e^{-5}$). Subcellular localization of peptidases found by METABOLIC against the MEROPS database was predicted by the PSORTb web tool v3.0.3 (Yu et al., 2010). The number of rRNA-coding sequences and presence of 16S rRNA gene in the MAGs were determined with Anvi'o (Eren et al., 2021). The Hidden Markov Model (HMM) profiling function, which utilizes Barrnap,[1] predicted the locations of rRNA genes. The tRNAscan-SE software v1.3.1 (Chan and Lowe, 2019) was used to determine the number of tRNAs in each MAG. Metabolic pathways were reconstructed based on the hits for key metabolic marker genes and the reference pathways depicted in KEGG (Kanehisa and Goto, 2000) and MetaCyc (Caspi et al., 2020) databases.

Analysis of CRISPR-Cas systems

CRISPR arrays of repeat-spacer units, and *cas* genes were identified using the combination of web tools CRISPRCasFinder (Couvin et al., 2018), CRISPRone (Zhang and Ye, 2017), CRISPRminer2 (Zhang et al., 2018), and CRISPRCasTyper (Russel et al., 2020). CRISPRone, CRISPRminer2, and CRISPRCasTyper were run with default settings. While running CRISPRCasFinder, the "General" clustering model and the "Unordered" function were used to identify *cas* genes. The CRISPRCasFinder tool has an evidence level rating that ranges from 1 to 4 for detection of CRISPR arrays. Only CRISPR arrays having evidence levels 3 and 4 were included in our analysis. While defining the *cas* locus, CRISPR locus, and CRISPR-Cas locus, and determining the type of a *cas* locus here, we followed published criteria (Zhang and Ye, 2017): (1) a *cas* locus should contain at least three *cas* genes, and at least one of those should belong to the universal *cas* genes for CRISPR adaptation (*cas1* and *cas2*) or the main components of interference module including *cas7*, *cas5*, *cas8*, *cas10*, *csf1*, *cas9*, *cpf1* (Makarova et al., 2015a); (2) CRISPR arrays that are close to each other (≤200 bps) and share very similar repeat sequences were considered to be in the same locus; and (3) the CRISPR(s), together with nearby (within 10,000 bps) *cas* genes, are defined as a CRISPR-Cas locus. The type of each *cas* locus was determined based on type signature *cas* genes listed in Makarova et al., 2015a,b The type-assignment of a *cas* locus was considered confident if it had at least three type-consistent signature *cas* genes, except for type V, which has only one signature gene, *cpf1* (Makarova et al., 2015a; Zhang and Ye, 2017). If a *cas* locus included only one or two type signature *cas* genes, we considered the type-assignment as putative.

Prediction of putative prophage regions in the MAGs

The VIBRANT v1.2.0 (Kieft et al., 2020) with default settings through the CyVerse Discovery Environment at https://de.cyverse.org/de (Merchant et al., 2016) and VirSorter v1.0.5 (Roux et al., 2015) with default settings against the Viromes database through the DOE Systems Biology KBase at http://kbase.us (Arkin et al., 2018), were used on all contigs of at least 1 kb to identify putative viral genes in the MAGs. All putative viral gene sequences detected by VIBRANT and VirSorter were queried against UniRef100 (Suzek et al., 2015), and UniProtKB reference proteomes plus Swiss-Prot (UniProt Consortium, 2021) databases using the BLAST tool (Altschul et al., 1990) with BLASTp (Altschul et al., 1997, 2005) at https://beta.uniprot.org/blast for functional annotation. Furthermore, these putative viral gene sequences were also functionally annotated by querying them against the NCBI non-redundant protein sequences (nr; Sayers et al., 2021) and IMG/VR Viral Resources v3 (Roux et al., 2021) databases using BLASTp. The BLAST hits with e-values equal to or lower than $1e^{-5}$ were accepted as statistically significant hits. HHpred, HHblits, and ProtBLAST/PSI-BLAST tools of the MPI Bioinformatics Toolkit (Zimmermann et al., 2018; Gabler et al., 2020) and NCBI CD (Conserved Domains)-Search tool (Lu et al., 2020) were also utilized for further annotation of putative viral genes in our MAGs. The search for auxiliary metabolic genes (AMGs) in the putative prophage regions of the MAGs was carried out based on the list of AMGs provided by Kieft et al. (2020).

VIBRANT uses HMM profiles from three databases: KEGG KoFam, March 2019 release (Kanehisa and Goto, 2000; Aramaki et al., 2020), Pfam (v32; Mistry et al., 2021), and Virus Orthologous Groups (VOG) release 94[2] to identify and annotate viral proteins (Kieft et al., 2020). On the other hand, VirSorter compares predicted protein sequences to Pfam (v27; Mistry et al., 2021) and to the user-selected viral reference database (Viromes database was chosen for this study) to detect viral-like genes (Roux et al., 2015).

If the protein sequence of a gene found in a putative viral region did not have any hits in KEGG, KoFam, Pfam, VOG, and Viromes databases, and this protein sequence only yielded hits to hypothetical or uncharacterized proteins when BLASTed against UniRef100, UniProtKB reference proteomes plus Swiss-Prot, NCBI nr, and IMG/VR databases, it was then annotated as a hypothetical/uncharacterized protein. We also manually inspected the KEGG, eggNOG-mapper, and InterPro annotations of the MAGs to search for viral hallmark proteins, specifically to identify any viral region not detected by VIBRANT and VirSorter tools. Viral hallmark proteins are necessary for productive infection, such as structural (e.g., capsid, tail, baseplate, portal, coat, spike), terminase, or viral holin/lysin proteins, and their presence is one of the metrics used to capture viral signatures in metagenomes (Roux et al., 2015; Kieft et al., 2020).

Anvi'o analysis

Pangenomic analyses were performed using Anvi'o v7.1 for three of the MAGs with sufficient genome completeness (>55%): 146_bin.21, 146_bin.25, and 1054_113_bin.10 and their closest relatives in the GTDB release 207_v2 (Table 1). Since two of the MAGs

1 https://github.com/tseemann/barrnap

2 vogdb.org

TABLE 1 Genomic features of the five *Caldarchaeales* MAGs.

Bin ID	1054_108_bin.3	1054_113_bin.10	1054_113_bin.2	146_bin.21	146_bin.25
Sample name/source	ET Poppy.17 (Chile)	ET 181107.CE.5 (Chile)	ET 181107.CE.5 (Chile)	100086B (Hawai'i)	100086B (Hawai'i)
No. of contigs	167	4	2	21	3
Genome size (Mbp)	0.823684	0.842732	0.632821	1.786704	1.567119
Estimated genome size (Mbp)	1.86	1.19	2.33	1.87	1.6
Largest contig size (Kbp)	12.9	628.7	575.2	371.3	857
Genomic G + C (%)	52.02	51.57	51.96	50.42	62.31
N50 value (bp)	5,339	628,726	575,203	185,085	857,045
No. of protein coding genes	956	910	702	2086	1,624
No. of rRNAs	0	2	0	1	1
Presence of 16S rRNA gene	No	Yes	No	Yes	Yes
No. of tRNAs	19	27	7	35	44
No. of genes annotated by KO	498 (52.1%)	539 (59.2%)	353 (50.3%)	1,054 (50.5%)	868 (53.4%)
Completeness (%)	44.34	70.87	27.18	95.63	98.06
Contamination (%)	0	0	0	0.97	0
Accession number	JANVYS000000000	JANVYU000000000	JANVYT000000000	JANVYQ000000000	JANVYR000000000
SeqCode name	N/A	N/A	N/A	*Pelearchaeum maunauluense*	*Calditenuis fumarioli*

generated from the Chile samples had genome completeness levels lower than 55%, they were excluded from the pangenomic analyses. All of the closest relatives of 146_bin.21, 146_bin.25, and 1054_113_bin.10 in the GTDB displaying completeness levels higher than 55% were included in the analyses. The first pangenomic analysis included 1054_113_bin.10 and three reference genomes that showed >95% ANI (Average Nucleotide Identity) with 1054_113_bin.10 and belonged to the GTDB species *Caldarchaeum subterraneum_B* (GCA_011373365.1, GCA_011364015.1, GCA_011364605.1).

The second pangenomic analysis included 146_bin.25 and two reference genomes that showed >95% ANI with 146_bin.25 and belonged to the GTDB species JGI-OTU-1 sp011364265 (GCA_011364265.1, GCA_011369755.1). Lastly, the third pangenomic analysis included 146_bin.21 and two reference genomes that showed approximately 80% ANI with 146_bin.21 and belonged to the GTDB species WAQC01 sp015520425 (GCA_015520425.1, GCA_015522085.1). Due to relative incompleteness of closely related MAGs in GTDB, we were only able to use very few genomes in the pangenomic comparisons. To ensure continuity, previous protein-coding gene annotations completed by Prodigal were first processed using "anvi-script-process-genbank" and imported into the Anvi'o contigs databases using the "—external-gene-calls" flag and "anvi-import-functions" program. After the pan-genomes were generated using the "anvi-pan-genome" program, the ANIb values between query and reference genomes were imported using PyANI from the "anvi-compute-genome-similarity" program within the Anvi'o suite.

Phylogenomic analyses

The Genome Taxonomy Database Toolkit (GTDB-Tk) version 2.1.1 was used alongside database release 207_v2 to classify the five MAGs presented here. The resulting multiple sequence alignment (MSA) containing 53 concatenated and conserved single-copy archaeal marker genes (10,153 characters) in 3492 archaeal genomes, which included the five MAGs generated in this study, was used as an input to perform a model test in the IQ-Tree program (version 2.1.4-beta; Minh et al., 2020). A maximum likelihood phylogenomic tree was constructed using the best model chosen based on the AIC and BIC criteria (LG + F + G4), with 1,000 ultra-rapid bootstrap replicates. The resulting tree was uploaded to iTOL (Interactive Tree of Life; Letunic and Bork, 2021) to collapse known clades, and edited in Inkscape for esthetics only.

Calculation of the abundance and distribution of *Caldarchaeales* from published datasets

Barrnap v0.9[3] was used to identify and extract 16S rRNA gene sequences from three of our five *Caldarchaeales* genomes

3 https://github.com/tseemann/barrnap

(1054_113_bin.10, 146_bin.21, 146_bin.25). To detect the presence of *Caldarchaeales* across diverse habitats, all three 16S rRNA gene sequences were used as queries for the Integrated Microbial NGS Platform (IMNGS; Lagkouvardos et al., 2016) to perform a Paraller similarity search (97% similarity), which analyzes all available prokaryotic 16S rRNA amplicon experimental datasets from the International Nucleotide Sequence Database Collaboration (INSDC). The INSDC includes datasets submitted from the DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), and National Center for Biotechnology Information (NCBI). The IMNGS result provided the relative abundance of reads and environmental descriptors for each hit, but data for location coordinates were searched manually in the NCBI SRA database or original literature due to inconsistencies in location reporting. The relative abundance was calculated by dividing the number of sequences identified by IMNGS to be >97% identical to the query 16S rRNA gene sequences by the total number of sequences in each 16S rRNA amplicon survey. All non-environmental 16S rRNA amplicon surveys (e.g., oral, skin, gut) were removed from downstream analysis, and amplicon sequences with identical location coordinate data were combined. The 16S rRNA gene sequences of our three *Caldarchaeales* MAGs were also searched against the Silva 16S rRNA database to obtain GenBank accession numbers for all hits. These GenBank accession numbers were downloaded and parsed using custom Python scripts to extract environmental information and coordinate location data. GenBank accessions containing location coordinate data were used in the geographic analysis, and again, identical locations were combined. Environmental features and location coordinates data for each hit were visualized in R using the ggplot2 (Wickham, 2016) and maps (Becker et al., 2022) packages. Relative abundance data were visualized for all 16S rRNA amplicon surveys found using IMNGS.

Results and discussion

Diversity, abundance, and distribution of the *Caldarchaeales*

Searches against the IMNGS and Silva 16S rRNA databases revealed that the *Caldarchaeales* are represented in a diverse array of features, including hot springs, marine geothermal features, lake and ocean sediments, soil, and other aquatic systems. They were also determined to be present in low relative abundance (<0.0001%) in plants such as *Populus deltoides*, fish gut metagenomic studies, and in surveys of human skin, oral, and gut microbial communities. Although they appear to inhabit a wide array of ecological niches, *Caldarchaeales* are especially prevalent in hot springs and other geothermal features (Figure 1; Supplementary Table S1). We cannot exclude the possibility that some of the unusual habitats the *Caldarchaeales* are found might be attributed to barcode misassignments or contaminations in the samples we obtained from these public databases. The Great Basin and Yellowstone regions of the United States host the greatest relative abundance of these archaea. These regions likely provide optimal conditions for thermophilic archaea since the abundance of affiliated sequences in all other features and areas is at least 100-fold less than in US hot springs. It should also be noted that in some soil and sediment environments, temperatures can be similar to those in many hot springs, such as a fumarolic soil sample from Antarctica in which the temperature was 65°C (Herbold et al., 2014). However, there is a possibility that primer bias could have also contributed to higher abundances of these archaea in some of the samples. Additional sampling and metagenomic studies will provide more insights into the geographical distribution of the *Caldarchaeales*.

General genomic features and phylogenomic placement

General features of the five MAGs along with genome completeness and contamination estimates are shown in Table 1. GTDB-Tk classification of the MAGs resulted in one from Hawai'i (146_bin.21) being assigned species-level novelty in the GTDB-designated genus WAQC01, in the family *Caldarchaeaceae*, order *Caldarchaeales*. In the GTDB release 207_v2, only two genomes represent WAQC01, and these are MAGs generated from hydrothermal deposits in the subsurface of a submarine volcano in the Pacific Ocean (Reysenbach et al., 2020). The other MAG from Hawai'i (146_bin.25) was assigned to the GTDB species JGI-OTU-1 sp011364265 in the GTDB genus JGI-OTU-1, family HR02, order *Caldarchaeales*. It forms a sister-level relationship with *Candidatus* "Calditenuis aerorheumensis." Three MAGs from Chile were assigned to the GTDB species *Caldarchaeum subterraneum B* in the family *Caldarchaeaceae*, order *Caldarchaeales*.

IQ-Tree phylogenomic inference of the five MAGs alongside the archaeal backbone tree from the GTDB release 207_v2, shows the three bins from Chile in a clade with *Ca. C. subterraneum* (Figure 2). 146.bin.25 groups closely with members of family HR02 and forms a sister-level relationship with a SAG from Great Boiling Spring (Rinke et al., 2013). It is interesting to note that 146_bin.21 forms a monophyletic clade with two deep-sea hydrothermal vent-associated lineages (Reysenbach et al., 2020) despite it coming from a terrestrial environment. More importantly, we obtained a monophyletic clade comprising all *Caldarchaeales* as sister to the clade comprising the orders *Geothermarchaeales* (previously *Candidatus* Geothermarchaeota), *Conexivisphaerales* and *Nitrososphaerales* (formerly classified as the phylum Thaumarchaeota).

Description of new family, genera, and species

Based on the SeqCode guidelines to assign valid names to microbial genomes with relatively high completeness and low contamination, both MAGs from Hawai'i can be given valid genus and species names as both are above 90% completeness, below 5% contamination, and have small numbers of contigs (Hedlund et al., 2022; Whitman et al., 2022). Here, we propose the names *Pelearchaeum maunauluense*^{Ts} gen. Nov., sp. nov. for 146_bin.21 after Pele, the goddess of volcanoes and fire, and *Calditenuis fumarioli* ^{Ts} sp. nov. for 146_bin.25, which belongs to the same genus as *Candidatus* "Calditenuis aerorheumensis" (ANI value of 81.07% between them; Figure 2). Due to monophyletic grouping of 146_bin.25 with 10 other members (Figure 2), we also propose to name the clade (GTDB family

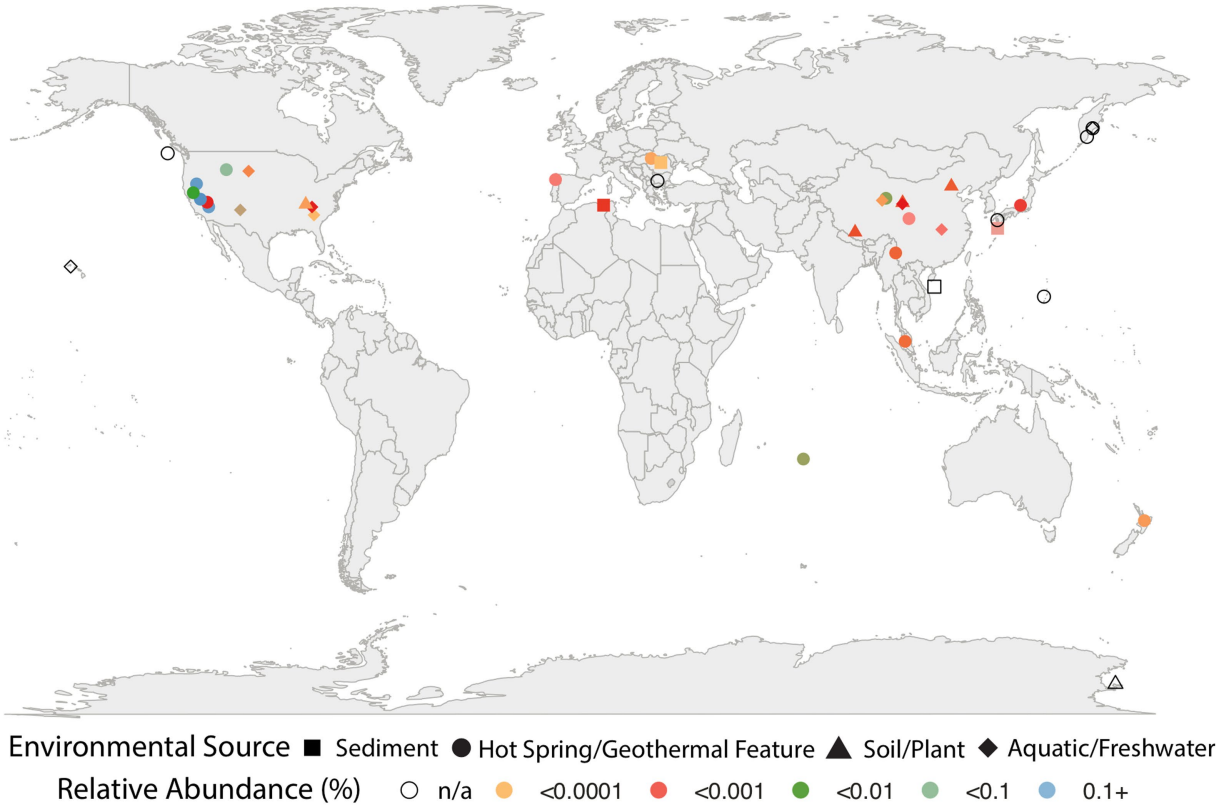


FIGURE 1
Environmental distribution of *Caldarchaeales* based on presence of 16S rRNA sequences identified in 500,048 samples from publicly available database. The symbols in various colors represent sequences from IMNGS and correspond to relative abundance (see key); symbols outlined in black represent sequences from the Silva 16S database, and do not include abundance information.

designation HR02) as *Calditenuaceae* and designate *Calditenuis fumarioli*^{Ts} as type species. A list of proposed names, etymologies, and nomenclatural types are shown in Table 2. Hereafter, we will refer to 146_bin.21 as *P. maunauluense* and 146_bin.25 as *C. fumarioli*.

Metabolic features

Metabolic features inferred from the genomes of the five MAGs are summarized in Figure 3; Supplementary Tables S2, S3, S10. GhostKOALA and eggNOG-mapper annotations of all coding sequences can be found in Supplementary Tables S4, S5, respectively. Their metabolic potentials are discussed in greater details below.

Central carbon metabolism

Only *P. maunauluense*, *C. fumarioli*, 1054_113_bin.2, and 1054_108_bin.3 include at least one marker gene of the glycolysis (Embden-Meyerhof-Parnas) pathway (Figure 3). *Pelearchaeum maunauluense*, 1054_113_bin.2, and 1054_108_bin.3 possess the metabolic potential for the Branched Entner-Doudoroff pathway to utilize glucose. These MAGs include two marker genes of the Branched ED pathway: genes encoding 2-dehydro-3-deoxygluconokinase/2-dehydro-3-deoxygalactonokinase and glycerate 2-kinase (Siebers et al., 2004; Kehrer et al., 2007; Bräsen et al., 2014; Sutter et al., 2016). The 1054_108_bin.3 and 1054_113_bin.10 lack both oxidative and non-oxidative phases of the pentose phosphate pathway (PPP), whereas *P. maunauluense*, *C. fumarioli*, and 1054_113_bin.2 possess incomplete oxidative PPP and lack the non-oxidative phase of PPP. It was reported that the Ribulose Monophosphate Pathway (RuMP) could substitute for the missing non-oxidative PPP in some archaea (Orita et al., 2006). *Pelearchaeum maunauluense* encodes genes for the bifunctional enzyme—3-hexulose-6-phosphate synthase (*hps*) and 6-phospho-3-hexuloisomerase (*phi*)—suggesting it can conduct Ru5P synthesis by RuMP, despite lacking the non-oxidative PPP.

The genes encoding fructose-1,6-bisphosphatase and phosphoenolpyruvate carboxykinase enzymes are the marker genes for the gluconeogenesis pathway (Dombrowski et al., 2018). A gene encoding fructose-1,6-bisphosphate aldolase/phosphatase (type V fructose-1,6-bisphosphatase) was identified in *P. maunauluense*, *C. fumarioli*, and 1054_113_bin.10. However, another key enzyme involved in gluconeogenesis, phosphoenolpyruvate carboxykinase was found only in the *P. maunauluense* genome. Yet, it is known that the function of this enzyme in gluconeogenesis pathway could be fulfilled by phosphoenolpyruvate synthase (pyruvate, water dikinase) or pyruvate:phosphate dikinase (Bräsen et al., 2014). Only *P. maunauluense*, *C. fumarioli*, and 1054_113_bin.10 encode pyruvate:phosphate dikinase enzyme, which could be utilized for gluconeogenesis in the absence of phosphoenolpyruvate carboxykinase. In addition to the type V fructose-1,6-bisphosphatase, we found type IV fructose-1,6-bisphosphatase

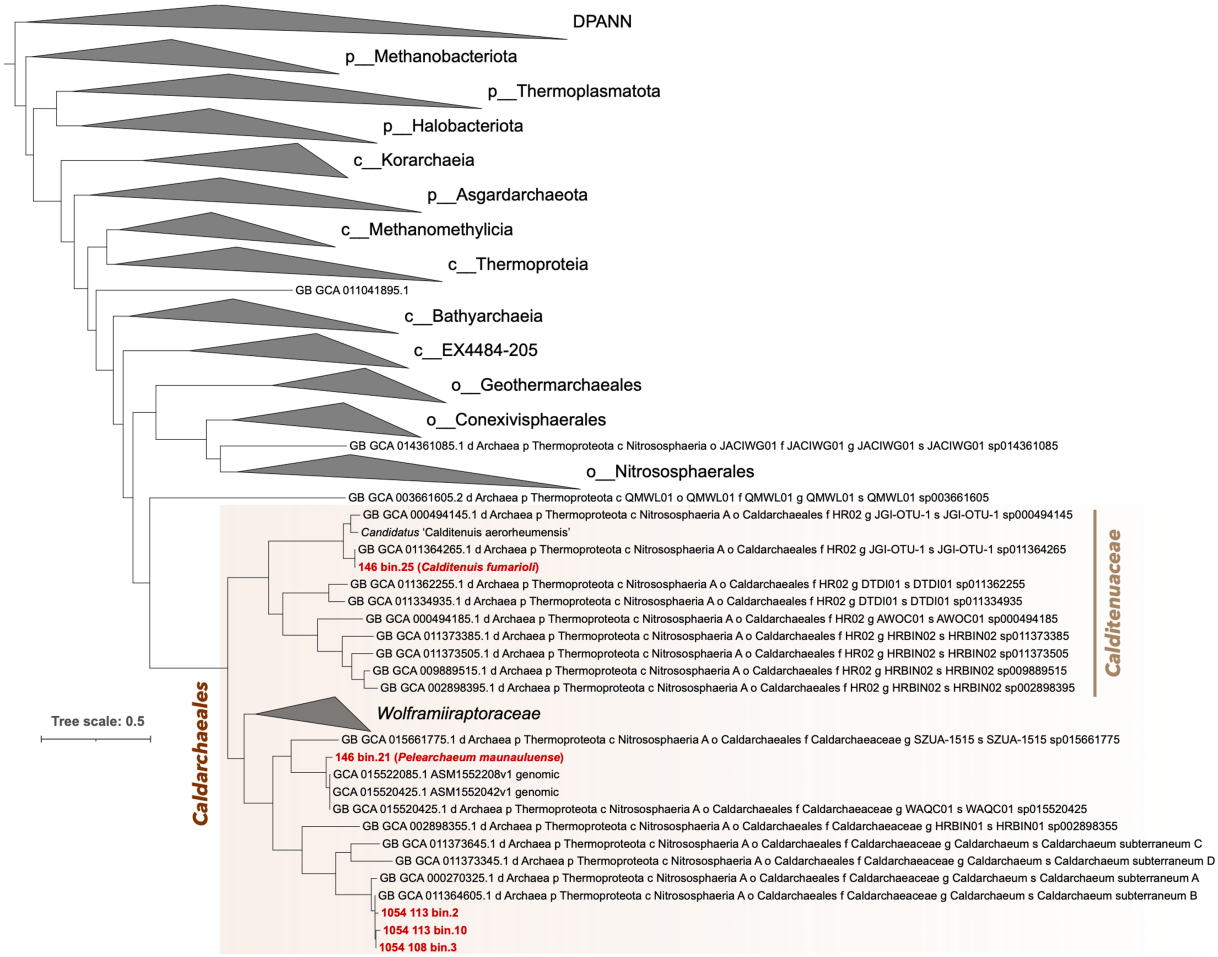


FIGURE 2
Phylogenomic placement of five *Caldarchaeales* MAGs alongside representative reference archaeal genomes obtained from the GTDB release 207_v2. Prefixes in taxa names indicate the following: p, phylum, c, class, o, order, f, family, g, genus, and s, species. The tree was rooted with DPANN archaea. Known groups of archaea were collapsed and shown as triangles. The order *Caldarchaeales* is highlighted with a light blue background. Full tree containing all taxa is available as Supplementary Data File 1.

(Stec et al., 2000; Verhees et al., 2002; Bräsen et al., 2014), which shows both fructose-1,6-bisphosphatase and inositol monophosphatase activities, in *P. maunauluense* and *C. fumarioli*.

All bins display the potential to decarboxylate pyruvate to acetyl-coenzyme A and link glycolysis with the oxidative tricarboxylic acid (oTCA) cycle (Figure 3). All except 1054_113_bin.10 encode both pyruvate ferredoxin/flavodoxin oxidoreductase (POR) and pyruvate dehydrogenase (PDH) complex. 1054_113_bin.10 only encodes PDH complex. POR catalyzes the decarboxylation of pyruvate during fermentation in anaerobes, while PDH complex catalyzes the same reaction during respiration in aerobes (Müller et al., 2012; Gould et al., 2019). Anaerobic and aerobic conditions are known to generally fluctuate in oxygen-limited environments such as deep subsurface aquifers and surface sediments. Hence, the presence of both POR and PDH complex could help microorganisms adapt to oxygen fluctuations as suggested earlier (Fang et al., 2021). While *P. maunauluense* and *C. fumarioli* encode genes for complete oTCA cycle, 1054_108_bin.3, 1054_113_bin.10, and 1054_113_bin.2 lack some of the genes. However, all five MAGs encode citrate synthase, which is one of the key enzymes of the oTCA cycle (Danson, 1993), suggesting that the cycle may be present in all the bins.

It is known that β-oxidation of fatty acids involves at least four enzymes: acyl-CoA dehydrogenase, crotonase, 3-hydroxyacyl-CoA dehydrogenase and acetyl-CoA acetyltransferase (Wang et al., 2019, 2021). Acyl-CoA dehydrogenase, fused enoyl-CoA hydratase and 3-hydroxybutyryl-CoA dehydrogenase enzyme, and acetyl-CoA C-acetyltransferase were found in *P. maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1054_113_bin.10. However, 1054_113_bin.2 encodes only acyl-CoA dehydrogenase. It was suggested that the enzymes involved in fatty acid β-oxidation could also be used for synthesizing fatty acids in *Archaea* (Dibrova et al., 2014). Hence, *P. maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1054_113_bin.10 exhibit metabolic capacity for utilization of lipids as carbon and energy sources and fatty acid biosynthesis. In addition, the presence of numerous binding/transport proteins for branched chain amino acids/dipeptides/oligopeptides and several aminotransferases and intracellular peptidases in all the MAGs indicates the capability to acquire amino acids and peptides from the surrounding environment and utilize them as a carbon and energy source. Some

TABLE 2 A table of proposed names.

Proposed taxon	Etymology	Nomenclatural type
Family *Calditenuaceae*	[Cal.di.te'nu.a.ce'ae] N.L. fem. pl. n. *Calditenuis* is the type genus of the family; –aceae ending to denote a family; N.L. fem. Pl. n. *Calditenuaceae*, the Calditenuis family	Genus *Calditenuis*
Genus *Calditenuis*	[Cal.di.te'nu.is] L. masc. Adj. *caldus*, warm; L. masc./fem. Adj. *tenuis*, thin, slender; N.L. masc. n. *Calditenuis*, a warm and slender organism.	Species *Calditenuis fumarioli*^{Ts}
Genus *Pelearchaeum*	[Pe.le.ar.chae'um] N.L. fem. n. *Pele*, name of the goddess of volcanoes and fire in Hawaiian mythology, N.L. neut. n. *archaeum*, an archaeon.	Species *Pelearchaeum maunauluense*^{Ts}
Species *Calditenuis fumarioli*^{Ts}	[fu.ma.ri.o'li] N.L. gen. n. *fumarioli*, of a fumarole, referring to the source of the nomenclatural type.	Genomic assembly: GCA_030059665.1^{Ts}
Species *Pelearchaeum maunauluense*^{Ts}	[mau.na.u.lu.en'se] N.L. neut. Adj. maunauluense, of Mauna Ulu volcano in Hawaii. In Hawaiian, ulu means to grow or sprout.	Genomic assembly: GCA_030059675.1^{Ts}

MAGs also encode a few genes for putative extracellular peptidases, implying the potential to secrete peptidases, degrade detrital proteins outside of the cell and play a role in protein remineralization (Figure 3).

Presence of xylulose kinase genes in 1054_108_bin.3 and 1054_113_bin.10 suggests the capability to degrade xylose, which is the main constituent of xylan polymers, which are the second most abundant group of polysaccharides in nature and the major component of plant cell walls (Beg et al., 2001; Knapik et al., 2019). Recently, a *Caldarchaeales* archaeon grown in enrichment culture in the lab was shown to encode xylose isomerase and xylulose kinase, use xylose as its preferred carbohydrate monomer, and was proposed to contribute to the fermentative degradation of lignocellulose at Great Boiling Spring, Nevada (Buessecker et al., 2022). The presence of similar genes in the two MAGs here from El Tatio hot springs suggests they may also be able to degrade xylose or similar molecules.

Fermentation

The detection of *acd* gene encoding ADP-forming acetyl-CoA synthetase (Acd) in *C. fumarioli* and 1054_108_bin.3 indicates the capacity for acetate fermentation even though these MAGs lack the classical phosphate acetyltransferase (Pta)—acetate kinase (Ack) pathway for acetate production (He et al., 2016). Acd is a novel enzyme that catalyzes the conversion of acetyl-CoA to acetate and couples this reaction with the synthesis of ATP via the mechanism of substrate level phosphorylation in *Archaea* and eukaryotic protists (Schönheit and Schäfer, 1995; Musfeldt and Schönheit, 2002). On the other hand, all acetate-forming *Bacteria* use classical Pta-Ack pathway for acetate formation and ATP synthesis except a few members, such as *Chloroflexus aurantiacus* (Schmidt and Schönheit, 2013) and *Propionibacterium acidipropionici* (Parizzi et al., 2012), which were suggested to use Acd for acetate formation based on the proteomic and biochemical evidence.

P. maunauluense, *C. fumarioli*, and 1054_113_bin.10 were also found to encode for AMP-forming acetyl-CoA synthetase (Acs) involved in acetate utilization by converting acetate to acetyl-CoA (Ingram-Smith et al., 2006), which suggests the possibility of both acetate production and consumption for *C. fumarioli*. However, Acs might function in the reverse direction, as well. In a recent study, acetogenic activity of both Acd and Acs is demonstrated with the enzyme assays providing strong evidence to conclude that Acs could also catalyze the conversion from acetyl-CoA to acetate in anaerobic methanotrophic archaea (Yang et al., 2020). Taken together, these findings imply that some members of the *Caldarchaeales* could have a metabolic capacity to grow as acetogens and play a role in organic carbon cycling in their ecosystems. None of the MAGs contain the gene encoding aldehyde dehydrogenase responsible for the conversion of acetate to acetaldehyde in the ethanol fermentation pathway, and the gene encoding for alcohol dehydrogenase, which generates ethanol from acetaldehyde, is only found in 146_bin25. Hence, it seems likely that none of the MAGs has the metabolic capacity to perform ethanol fermentation.

Carbon fixation

The absence of ATP citrate (pro-S)-lyase and citryl-CoA synthetase/citryl-CoA lyase was previously reported in *Caldarchaeales* genomes, indicating the lack of the reductive TCA Cycle (rTCA; Hua et al., 2018). In our study, genes encoding ATP citrate (pro-S)-lyase and citryl-CoA synthetase were not recovered from any of the MAGs. Citryl-CoA lyase was only observed in the *C. fumarioli* genome. The lack of key enzymes indicates the absence from our MAGs of the rTCA cycle.

An alternative pathway for carbon fixation, referred to as reversed oxidative TCA cycle (roTCA), may exist in *Caldarchaeales* (Hua et al., 2018). This pathway is a version of the rTCA cycle. In the roTCA cycle, citrate synthase replaces ATP-citrate lyase (or citryl-CoA synthetase / citryl-CoA lyase) of the rTCA cycle (Mall et al., 2018; Nunoura et al., 2018). The roTCA cycle was shown to operate in two obligately anaerobic bacteria, *Desulfurella acetivorans* (Mall et al., 2018) and *Thermosulfidibacter takaii* (Nunoura et al., 2018). It was also revealed that the direction of roTCA cycle was controlled by the availability of organic vs. inorganic carbon source(s), reflecting the adaptation of microorganisms to fluctuating concentrations of carbon sources (Mall et al., 2018; Nunoura et al., 2018). Furthermore, in the thermophilic sulfur-reducing Deltaproteobacterium *Hippea maritima*, the roTCA cycle was driven by high partial pressures of carbon dioxide (Steffens et al., 2021). Given that citrate synthase was identified in all MAGs analyzed in our study, and ATP-citrate lyase (or citryl-CoA synthetase/citryl-CoA lyase) is absent from all MAGs, we propose that *Caldarchaeales* lineages represented by these MAGs likely harbor a metabolic potential to carry out the roTCA cycle for CO_2 fixation as an alternative to the canonical rTCA cycle.

Previously, the dicarboxylate/4-hydroxybutyrate (DC/4HB; Nunoura et al., 2011) and the 3-hydroxypropionate/

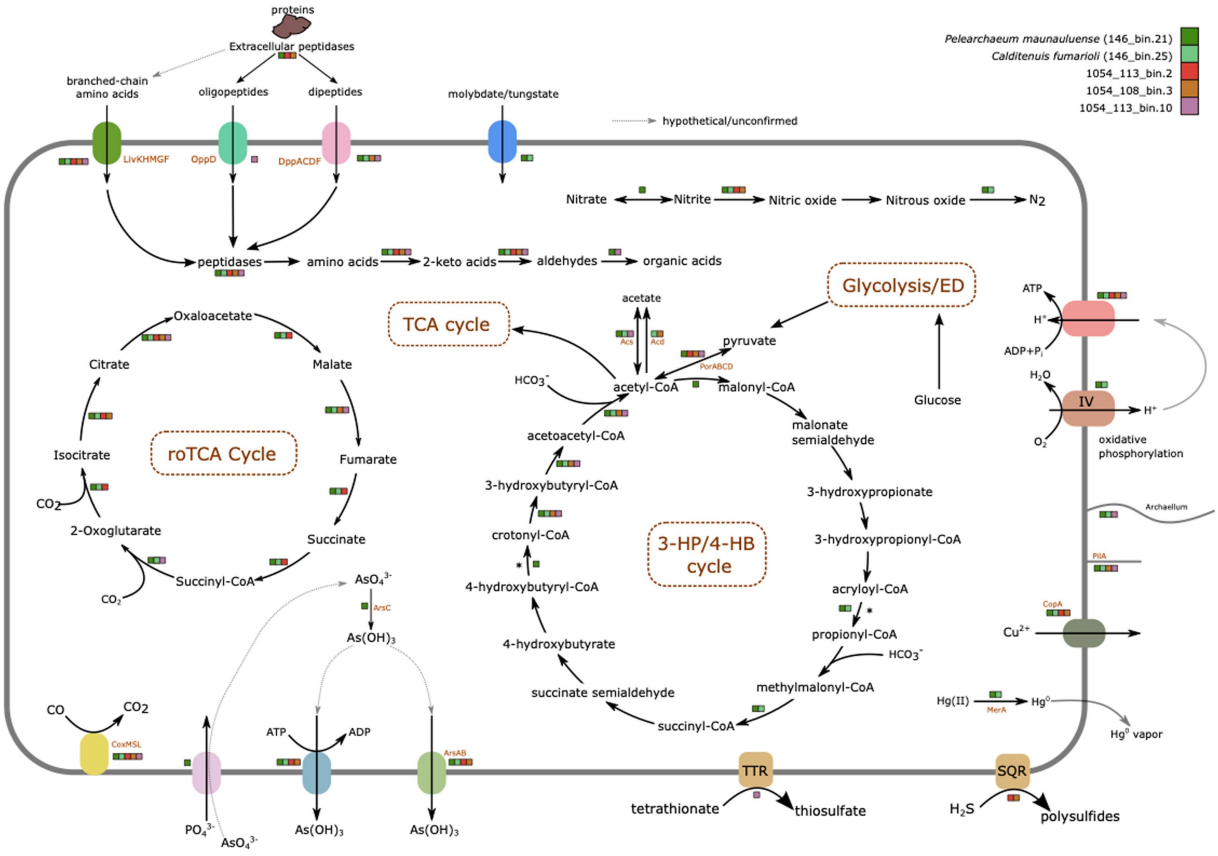


FIGURE 3
Overview of major metabolic features identified in the five MAGs from Hawai'i and Chile. MAGs are color-coded to indicate whether presence of genes encoding enzymes involved in these depictions have been identified in their genomes or not. Dotted lines surrounding the names of the pathways identified indicate partial nature of the pathways in some MAGs due to incomplete nature of the MAGs. The figure was manually drawn using Inkscape.

4-hydroxybutyrate (3HP/4HB; Beam et al., 2016) cycles were proposed to be present in *Caldarchaeales*. *Pelearchaeum maunauluense* genome contains two genes encoding key enzymes for the 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle: 4-hydroxybutyryl-CoA dehydratase and acryloyl-coenzyme A reductase (Berg et al., 2010; Hügler and Sievert, 2011; Könneke et al., 2014; Ruiz-Fernández et al., 2020), whereas *C. fumarioli* genome harbors only one key enzyme, acryloyl-coenzyme A reductase. 4-hydroxybutyryl-CoA dehydratase is also the key enzyme of the DC/4HB cycle. Yet, the gene encoding for the other key enzyme of the DC/4HB cycle, 4-hydroxybutyrate—CoA ligase was not found in any of the MAGs. Also, we did not recover any marker genes for the 3HP/4HB cycle in the MAGs from Chile. Despite the occurrence of some marker genes for the 3HP/4HB and DC/4HB cycles, we cannot assert that these cycles are functional considering that they are incomplete in *P. maunauluense* and *C. fumarioli* genomes.

C1 and oxygen metabolism

Multiple copies of aerobic carbon monoxide dehydrogenase large subunit (CoxL), aerobic carbon monoxide dehydrogenase medium subunit (CoxM), and aerobic carbon monoxide dehydrogenase small subunit (CoxS) were found in all MAGs. Also, all MAGs likely derive from organisms capable of aerobic respiration since they encode cytochrome c oxidase subunits: *P. maunauluense* has *coxA*, *coxB*, *coxC*, and *coxAC*; *C. fumarioli* has *coxA*, *coxB*, *coxD*, and *coxAC*; 1054_113_bin.2 has *coxC*; 1054_108_bin.3 has also only *coxC*; and 1054_113_bin.10 has *coxA* and *coxB*. Taken together, these findings imply the ability to grow chemolithotrophically via carbon monoxide (CO) oxidation coupled to oxygen reduction by cytochrome c oxidase complex. CO is a common gas in hot volcanic environments and hot springs since it is present in volcanic emissions and may be produced during thermal and photochemical decomposition of organic matter, or as a byproduct by some thermophilic anaerobes (Sokolova et al., 2009; Kochetkova et al., 2011; Baker et al., 2016). It was suggested that the ability to utilize CO is widespread among aerobic bacteria, and certain archaea (e.g., *Pyrobaculum* and *Sulfolobus*) likely carry out aerobic CO oxidation as well (King, 2003a). It was also proposed that CO could serve as an energy supplement when organic substrates are deficient, which is consistent with the observation that CO utilization is associated with microbial community development on Hawaiian volcanic deposits (King, 2003b). Previously, potential aerobic type carbon monoxide dehydrogenases were detected in *Caldarchaeales* (Nunoura et al., 2011; Beam et al., 2016; Hua et al., 2018), and it was speculated that *Caldarchaeales* could oxidize CO aerobically as an energy supplement in the case of limited organic substrates and the supplied energy might contribute to biomass by coupling with carbon fixation through the roTCA cycle (Hua et al., 2018).

Sulfur cycling

1054_108_bin.3 and 1054_113_bin.2 encode sulfide:quinone oxidoreductase, indicating the capacity to use hydrogen sulfide as an electron donor for chemolithotrophy. The presence of dissimilatory sulfite reductase (*dsrAB*) genes was previously reported in several *Caldarchaeales* genomes (Rinke et al., 2013; Hua et al., 2018). However, we did not recover *dsrA* and *dsrB* genes from our MAGs, which suggests they may represent organisms incapable of dissimilatory sulfite reduction. Moreover, multiple copies of the gene encoding the tetrathionate reductase subunit A were detected in 146_bin21, while all subunits of the tetrathionate reductase are present in 1054_113_bin.10, suggesting the organism from which this MAG was derived might conduct tetrathionate respiration, consistent with a previous study reporting a tetrathionate reductase complex in *Caldarchaeales* (Beam et al., 2016). In addition, Sanchez-Garcia et al., 2019 revealed that in the liquid geyser mounds at El Tatio geothermal fields, microbial metabolism is dominated by the autotrophic Calvin cycle, along with lesser sulfur and iron chemolithotrophic pathways, and iron and sulfur oxidizers and sulfate reducers are abundant in the liquid geyser mounds. Sulfate was detected in both water and sinter deposit samples of the liquid geyser mounds with the water collected from the liquid mound containing higher concentrations of sulfate ($346\,\mu g\,g^{-1}$) compared to its sinter deposits ($14\,\mu g\,g^{-1}$; Sanchez-Garcia et al., 2019). The findings of this previous study and the presence of genes involved in dissimilatory sulfur metabolism in the MAGs from El Tatio indicate that *Caldarchaeales* members might be also contributing to the sulfur cycling at El Tatio along with bacterial sulfur oxidizers and sulfate-reducing bacteria.

Nitrogen cycling

Pelearchaeum maunauluense is the only MAG encoding *narG*, *narZ*, *nxrA*; nitrate reductase/nitrite oxidoreductase, alpha subunit and *narY*, *nxrB*; nitrate reductase/nitrite oxidoreductase, beta subunit. This indicates the organism represented by *P. maunauluense* might perform the first step of denitrification, which is reduction of nitrate (NO_3^-) to nitrite (NO_2^-), and that it may also oxidize nitrite to nitrate (the second step of nitrification). Under anaerobic conditions, nitrite oxidoreductase functions as a nitrate reductase, catalyzing the reverse reaction. *Pelearchaeum maunauluense*, *C. fumarioli*, 1054_113_bin.2, and 1054_108_bin.3 encode the *nirK*; nitrite reductase (NO-forming), which suggests the organisms represented by these MAGs might perform anaerobic respiration by reducing nitrite (NO_2^-) to nitric oxide (NO; the second step of denitrification). However, nitric oxide reductase involved in the reduction of NO to nitrous oxide (N_2O; the third step of denitrification) were not identified in any of the MAGs. Since highly toxic NO is not likely to be reduced enzymatically, it could be abiotically reduced by iron as previously put forward by Kozlowski et al. (2016). This notion is supported by the studies reporting (a) the detection of iron oxides and dominance of microorganisms participating in the iron and sulfur geochemical cycles in liquid geyser mounds at El Tatio geothermal fields (Sanchez-Garcia et al., 2019) and (b) high concentrations of iron in basaltic rocks and their associated fumaroles (Cockell et al., 2019). *Pelearchaeum maunauluense* and *C. fumarioli* genomes encode *nosZ*; nitrous-oxide reductase, which suggests that the organisms represented by these MAGs might conduct anaerobic respiration by reducing N_2O to nitrogen gas (N_2; the fourth and last step of denitrification).

Versatile metabolic potentials of *Caldarchaeales* as adaptive mechanisms for life in extreme environments

The presence of some key marker genes for the 3HP/4HB and DC/4HB cycles in the MAGs from Hawai'i hints at the capacity to fix carbon. However, these cycles are not complete. All MAGs from Hawai'i and Chile also possess metabolic potential for roTCA cycle, but only further experimental validation will determine their functionality. Metabolic pathway predictions suggest the potential denitrification and tetrathionate reduction pathways might enable these lineages to respire diverse compounds other than oxygen. All MAGs except 1054_113_bin.10 encode reverse gyrase, a marker gene for hyperthermophiles, which is consistent with the environmental conditions of the habitats from which the MAGs were recovered. The occurrence of genes involved in aerobic and anaerobic respiration in the MAGs indicates a facultative anaerobic lifestyle, which would enable growth in anoxic environments. We also determined that the MAGs contain several genes implicated in oxidative stress tolerance, such as those for superoxide dismutase, thioredoxin reductase, and peroxiredoxin, which could be used to alleviate oxidative damage.

Some of our MAGs also possess genes involved in arsenic detoxification and resistance to mercury (Figure 3). Only *P. maunauluense* genome encodes genes capable of encoding arsenate reductase (*arsC*), though all other MAGs encode the gene for the ArsR family transcriptional regulator (*arsR*). Evidence for genes that encode the arsenical pump membrane protein (*arsB*) and arsenite/tail-anchored protein-transporting ATPase (*arsA*) were detected in all MAGs except 1054_113_bin.10. As for resistance to mercury, *P. maunauluense* and *C. fumarioli* genomes encode the gene for mercuric reductase (*merA*). The presence of arsenic and mercury in fumarolic rocks of Hawai'i is currently unknown. Hot springs are known to contain various levels of arsenic and at El Tatio, the concentrations can sometimes reach 40–50 mg/L of hot spring water, which is among one of the highest concentrations measured in any natural habitat (Wang et al., 2018) and therefore presence of transporters to pump out various arsenic related compounds such as arsenous acid [$As(OH)_3$] appear to indicate that *Caldarchaeales* in these habitats may be actively utilizing these detoxification systems for their survival.

Tungstoenzymes and tungstate transporters

A recent study has reported the enrichment and stable laboratory growth of *Caldarchaeales* from sediments of Great Boiling Spring (Nevada, United States), designated *Wolframiiraptor gerlachensis*, which belongs to the GTDB family NZ13-MGT (Buessecker et al., 2022). Cultivation experiments and the analysis of a MAG from *W. gerlachensis*, which includes six annotated tungsten (W)-dependent ferredoxin oxidoreductases that could play central roles in anaerobic carbohydrate degradation, have suggested that tungsten could be essential for carbohydrate metabolism, and that it may play a critical role in one or more of the annotated W-dependent ferredoxin oxidoreductases, thereby indicating that the growth of *W. gerlachensis* requires the presence of tungsten, a biologically rare trace metal (Buessecker et al., 2022). In addition, the evolutionary analysis of 78

MAGs representing four genera and 11 species of the GTDB family NZ13-MGT, designated as *Wolframiiraptoraceae*, has demonstrated that putative tungstate transporters and tungstoenzymes were ancestral in this lineage. Furthermore, homologs of 3 W-dependent ferredoxin oxidoreductases: aldehyde:ferredoxin oxidoreductases (AOR), glyceraldehyde-3-phosphate:ferredoxin oxidoreductases (GAPOR), and formaldehyde:ferredoxin oxidoreductases (FOR) are widely distributed in the family *Wolframiiraptoraceae* genomes, and sparse in other *Caldarchaeales*.

Also, the majority of the *Wolframiiraptoraceae* genomes has been detected to encode Tup ABC tungstate transporter system, whereas other ABC tungstate transporter systems: Wtp and Mod have only been identified in two species in the family *Wolframiiraptoraceae* (Buessecker et al., 2022). Given the significant role of tungsten in sugar and peptide catabolism within the *Caldarchaeales* (Buessecker et al., 2022), we searched for tungstate transporters and tungsten-associated enzymes in the five MAGs. Five AOR homologs were identified, one in 1054_113_bin.10, and four genes in *P. maunauluense*, three of which are fragmented. However, we did not recover any GAPOR or FOR homologs in the MAGs. *Pelearchaeum maunauluense* genome also encodes three subunits of the Wtp molybdate/tungstate transporter system, but no tungstate transporter system was annotated in 1054_113_bin.10, possibly due to incomplete genomes (Figure 3). Intriguingly, *C. fumarioli* includes three subunits of the Wtp molybdate/tungstate transporter system and an adjacent gene encoding a molybdate transport system regulatory protein, but no W-dependent ferredoxin oxidoreductases. The AOR homologs encoded in *P. maunauluense* and 1054_113_bin.10 may be involved in the protein catabolism by converting aldehydes, which were produced from 2-Keto acids by ferredoxin oxidoreductases to organic acids and also in the Branched Entner-Doudoroff pathway; GAPOR is only known to function in unusual Embden-Meyerhof pathways in which it converts glyceraldehyde-3-phosphate to 3-phosphoglycerate, replacing glyceraldehyde-3 phosphate dehydrogenase and phosphoglycerate kinase enzymes used in the canonical Embden-Meyerhof pathway (Mukund and Adams, 1995; Bevers et al., 2005). *Pelearchaeum maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1053_113_bin.10 encode both glyceraldehyde-3 phosphate dehydrogenase and phosphoglycerate kinase enzymes, which might explain the absence of a GAPOR homolog from these genomes.

Analysis of prophage regions

Previously, integrated mobile genetic elements (IMGEs) that include bacterial insertion sequence (IS)-like transposons, prophages, and cryptic integrated elements were reported in several *Caldarchaeales* genomes (Nunoura et al., 2011; Hua et al., 2018; Krupovic et al., 2019). In line with these studies, we also found putative prophage regions and genes encoding transposases in our MAGs, suggesting the MAGs might harbor IMGEs. Prophage-related genes and transposases identified in the MAGs are shown in Supplementary Tables S6, S7, respectively.

Viral genes were not detected in the MAG 1054_113_bin.10. However, the MAGs of *P. maunauluense*, *C. fumarioli* and 1054_113_bin.2 include prophage regions, and MAG 1054_108_bin.3 has three viral contigs, which we refer to as putative prophage regions (Figure 4). Although VIBRANT and VirSorter predicted the three contigs of MAG 1054_108_bin.3 to be entirely viral, they very likely belong to prophage regions integrated into the host genome, given MAG 1054_108_bin.3 is a very fragmented and incomplete genome of 167 contigs and 44.34% completeness.

Prophage regions in the MAGs comprise numerous viral hallmark proteins such as the DNA packaging enzyme terminase, phage integrase, a prohead protease involved in phage assembly and maturation, and structural proteins including phage capsid, tail, and portal proteins. In addition to viral hallmark proteins, prophage regions encode proteins associated with viral genome replication such as Staphylococcal nuclease homolog, which is thermostable, replicative DNA helicase Mcm (minichromosome maintenance helicase), RNA 2′,3′-cyclic 3′-phosphodiesterase, and ParB family transcriptional regulator. The replicative DNA helicase Mcm was shown to be frequently recruited from the host as the main replication protein of various crenarchaeal and euryarchaeal mobile genetic elements, including viruses and plasmids (Krupovic et al., 2010, 2019; Kazlauskas et al., 2016). RNA 2′,3′-cyclic 3′-phosphodiesterase was also shown to function as a highly heat-stable cyclic nucleotide phosphodiesterase with GTP-dependent RNA ligase activity in the hyperthermophilic archaeon *Pyrococcus furiosus* (Kanai et al., 2009). The presence of a thermostable nuclease and RNA 2′,3′-cyclic 3′-phosphodiesterase in the provirus regions suggests viruses infecting *Caldarchaeales* might have evolved strategies to replicate their genomes in high-temperature environments. The prophage regions annotated here also carry genes encoding functionally diverse proteins that might contribute to the growth of these archaea in extreme conditions. They are listed in Table 3. These findings are consistent with a previous study reporting that IMGEs found in 20 Thaumarchaeota genomes and one *Caldarchaeales* genome encode multiple AMGs, various dehydrogenases, stress response proteins, membrane transporters of cations and drugs, and chemotaxis protein receiver domains, which could contribute to the fitness, adaptation, and survival of the archaeal hosts by providing mechanisms that respond to stressful environmental conditions, and improve the host metabolic potential (Krupovic et al., 2019).

Viral gene clusters found in the MAGs might belong to Mu-like prophages, as the terminase enzymes encoded in the provirus regions of *P. maunauluense*, 1054_113_bin.2, and 1054_108_bin.3 share significant amino acid identities with Mu phage terminases, and one of the phage tail proteins encoded in the provirus region of *C. fumarioli* shares significant amino acid identities with that of the prophage Mu tail protein. The finding that some viral hallmark proteins in the provirus regions of the MAGs display homology to Mu phage proteins suggests that Mu-like phages could be infecting *Caldarchaeales* hosts. This finding also corroborates a previous study in which Mu-like prophages were found in two *Caldarchaeales* MAGs (Hua et al., 2018). Viruses identified in these MAGs might enhance the metabolic potential of their hosts through horizontal gene transfer events, facilitate niche-specialization, and may reduce competition for resources in extreme habitats such as hot springs and fumaroles found in the study sites presented here. Archaeal viruses are among the most enigmatic viruses known due to a relatively small number of them described to date compared to bacteriophages and eukaryotic viruses and due to their extraordinary genomic diversity and unique morphologies (Prangishvili et al., 2017; Munson-McGee et al., 2018; Wirth and Young, 2020). In this regard, further *in vivo* characterization of the putative *Caldarchaeales* viruses is needed to better understand

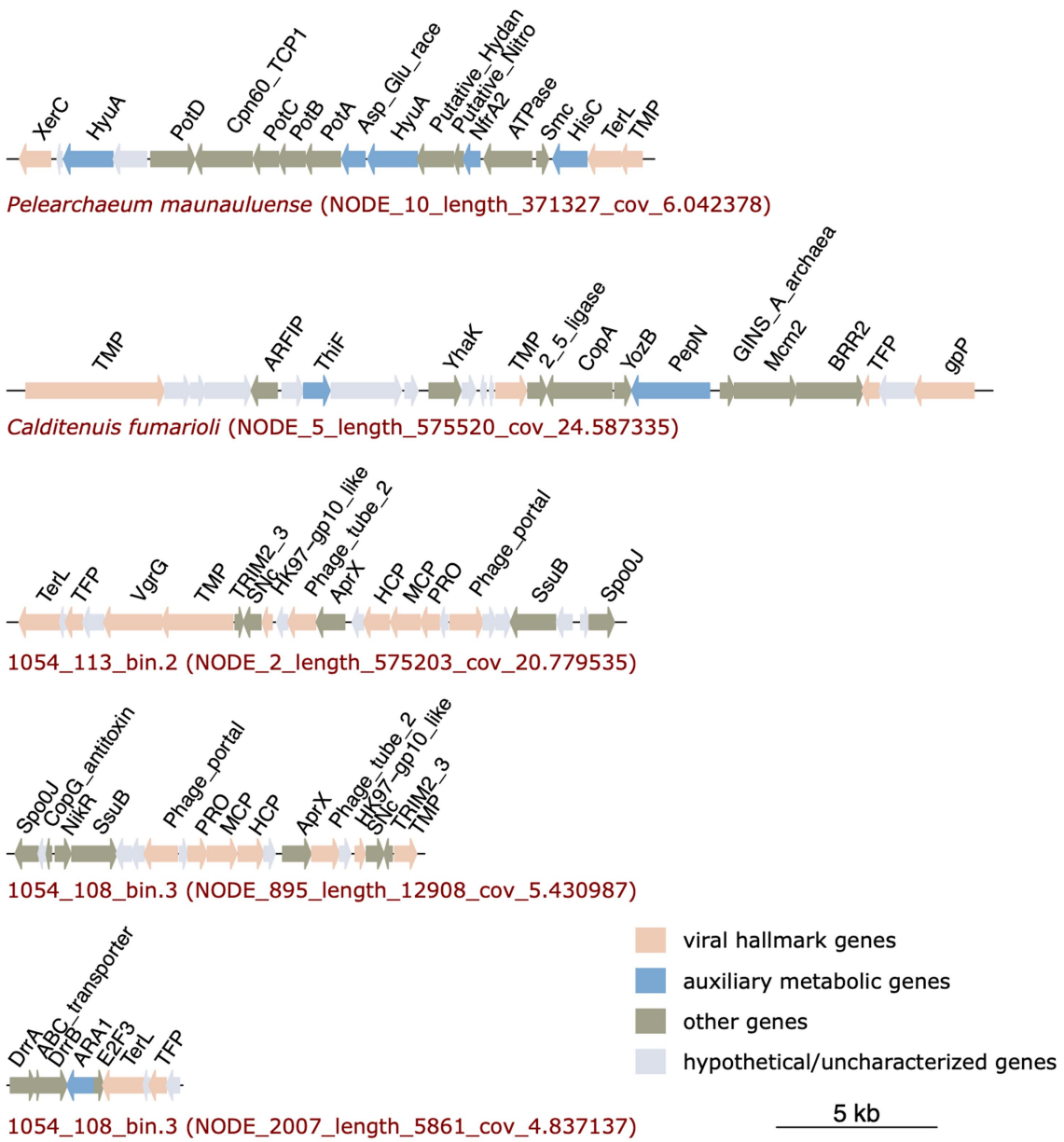


FIGURE 4
Prophage regions identified in *P. maunauluense*, *C. fumarioli*, and 1054_113_bin.2, and two viral contigs (putative prophage regions) identified in 1054_108_bin.3. The third viral contig of 1054_108_bin.3 is not shown in the figure since it does not include any genes encoding viral hallmark proteins. Full list of the genes and their annotations are shown in Supplementary Table S6.

their ecological roles and elucidate the co-evolution of them with their hosts in extreme habitats.

CRISPR-Cas systems

CRISPR-Cas systems in *Caldarchaeales* genomes have been previously reported, with MAGs harboring type-I and type-III CRISPR-Cas systems (Nunoura et al., 2011; Hua et al., 2018). Our analysis shows that type-I and type-III CRISPR-Cas systems might be also present in the MAGs reported here. We found that *P. maunauluense* has two *cas* loci classified as subtype III-D and putative subtype I-A, respectively, in two contigs (Figure 5; Supplementary Table S8). However, neither has a nearby CRISPR array (Supplementary Table S9). In the third contig of *P. maunauluense*, we detected two CRISPR loci, both of which lack *cas* genes in their vicinity. *Calditenuis fumarioli* and 1054_113_bin.2 have both subtype III-D and subtype I-A *cas* loci with nearby CRISPR arrays. 1054_108_bin.3 includes a putative subtype I-A *cas* locus, but this MAG lacks CRISPR arrays, possibly because it is highly incomplete and fragmented. Although we did not recover any *cas* locus in 1054_113_bin 10, we found that it harbors two *cas3* genes in the same contig, which does not contain a CRISPR array. Furthermore, three CRISPR arrays were found in three 1054_113_bin.10 contigs.

TABLE 3 Predicted annotations and functional categories of the provirus-encoded genes that might potentially contribute to the host fitness and survival.

Annotation	Functional category
ABC (ATP-Binding Cassette) transporters	Uptake of siderophores, heme, and vitamin B12
Putative ABC type spermidine/putrescine transport system proteins	Quorum sensing
ABC-type multidrug transport system proteins	These proteins yield significant BLAST hits to daunorubicin (an antitumor antibiotic drug) resistance ABC transporter subunits.
Type IV secretory pathway ATPase VirB11/Archaellum biosynthesis ATPase	Cell motility and secretion
Archaeal chaperonin	Heat tolerance
P-type Cu + transporter	Copper resistance
FMN reductase [NAD(P)H]	Auxiliary metabolic gene (AMG) involved in riboflavin (vitamin B2) metabolism
Histidinol-phosphate/aromatic aminotransferase	AMG involved in histidine biosynthesis
N-methylhydantoinase A/oxoprolinase/acetone carboxylase	AMG involved in arginine and proline metabolism
Asp/Glu/Hydantoin racemase	AMG involved in purine metabolism
Aldo/keto reductase	AMG involved in vitamin B6 metabolism
Molybdopterin or thiamine biosynthesis adenylyltransferase	AMG involved in sulfur relay system
Aminopeptidase N	AMG involved in glutathione metabolism
Putative nickel-responsive transcription factor	Regulation of nickel uptake
Putative CopG antitoxin of type II toxin-antitoxin system	Antibiotic resistance, biofilm formation, phage inhibition
Putative phage assembly protein	This protein shares significant sequence identities with phage tail-like nano-machines termed contractile injection systems (CISs) which mediate bacterial cell–cell interactions as either type VI secretion systems or extracellular CISs (Xu et al., 2022).
Subtilisin family serine proteases	These are extracellular peptidases that could play a role in protein remineralization and function at extreme temperatures and pH values.

Genes for cell motility and surface attachment

Pelearchaeum maunauluense, *C. fumarioli*, and 1054_113_bin.10 encode genes for the following archaeal flagellar components: FlaB, FlaG, FlaH, FlaI, FlaJ, and FlaK (Supplementary Table S3). However, 1054_108_bin.3 only has the gene encoding FlaK. No archaeal flagellar components were identified in 1054_113_bin.2, possibly due to the low level of completeness of this genome. The presence of archaeal flagellar components implies that microbial lineages represented by *P. maunauluense*, *C. fumarioli*, and 1054_113_bin.10 may be motile under certain environmental conditions. In addition, the gene encoding for an archaeal type IV pilin was identified in *P. maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1054_113_bin.10. *Archaea* type IV pili are involved in biofilm formation, including surface adhesion, microcolony formation, and regulation of flagella-dependent motility (Pohlschroder and Esquivel, 2015). In this respect, biofilm formation may protect against environmental stressors such as low or high pH and toxic chemicals and promote horizontal gene transfer and syntrophy with other microorganisms (van Wolferen et al., 2018). Given the role of type IV pili in archaeal biofilm formation, organisms from which *P. maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1054_113_bin.10 were derived may be able to form biofilms. Moreover, two chemotaxis-related genes (methyl-accepting chemotaxis protein and a heme-based aerotactic transducer) were found in *P. maunauluense* genome. Taken together, motility genes in *P. maunauluense*, *C. fumarioli*, 1054_108_bin.3, and 1054_113_bin.10, and chemotaxis genes in *P. maunauluense*, provide genomic evidence that these lineages might be able to move toward favorable environments within these geothermal habitats.

Pangenomic analysis of three distinct *Caldarchaeales* clades

The pangenomic comparison of 1054_113_bin.10 with its closest relatives from GTDB showed 1,844 genes in this pangenome, 528 of which are core, 26 which are unique, and that 19 assigned to the bin have COG categories (Figure 6A; Supplementary Table S11). Most of these gene clusters belong to "Amino acid transport and metabolism," and "Inorganic ion transport and metabolism" COG categories that comprise ABC-type nitrate/sulfonate/bicarbonate transport system components, and ABC-type dipeptide/oligopeptide/nickel transport system components. There are also two gene clusters of the COG category "Defense mechanisms" that encode toxin-antitoxin (TA) system-associated proteins. Two reference genomes from Jinze hot spring sediment environment, China (GCA_011364605.1 and GCA_011364015.1), have 265 gene clusters that were not found in 1054_113_bin.10 or the reference genome (GCA_011373365.1) from Gongxiaoshe hot spring sediment, China. Most of the 173 gene clusters assigned to COG categories belong to "Energy production and conversion," "Amino acid transport and metabolism," and "Defense mechanisms." The gene clusters of "Energy production and conversion" mainly comprise genes involved in the non-phosphorylated Entner-Doudoroff pathway, TCA cycle, and pyruvate oxidation. Gene clusters of "Defense mechanisms" mostly

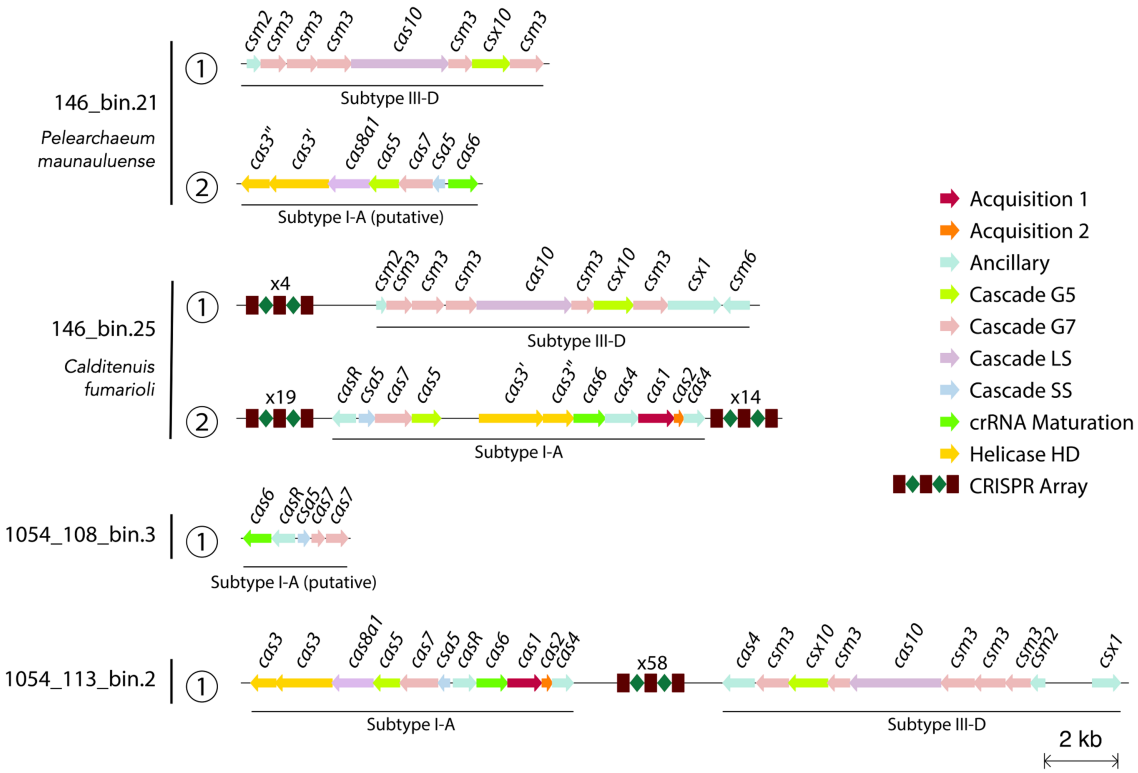


FIGURE 5
Visualization of predicted CRISPR-Cas arrays and *cas* loci in the contigs of 4 MAGs. Since the MAG 1054_113_bin.10 does not harbor any *cas* loci, CRISPR arrays and *cas* genes identified in 1054_113_bin.10 are not included in the figure. The text above the CRISPR array icon highlights the number of spacers in a CRISPR array and *cas* genes are color-coded according to their functions. The full list of the CRISPR-Cas related genes and arrays identified are shown in Supplementary Tables S8, S9, respectively.

include genes associated with toxin-antitoxin (TA) systems, CRISPR-Cas systems, and ABC-type multidrug transport systems. On the other hand, COG category "Inorganic ion transport and metabolism" contain gene clusters encoding sulfite exporters, an ABC-type Fe^{3+}-hydroxamate transport system component, ABC-type cobalamin/Fe^{3+}-siderophores transport system components, a phosphate uptake regulator, and predicted Fe^{2+}/Mn^{2+} transporter. There is also one gene cluster of the COG category "Mobilome: prophages, transposons" that encodes Mu-like prophage FluMu protein gp28, indicating the occurrence of past viral infections. The reference genome (GCA_011373365.1) from China's Gongxiaoshe hot spring sediment contains 71 unique gene clusters, 34 of which are annotated; most of the annotated gene clusters belong in the COG categories "Carbohydrate transport and metabolism," "Coenzyme transport and metabolism," and "Post-translational modification, protein turnover, chaperones" that includes gene clusters encoding chaperonin GroEL.

The pangenomic analysis of *C. fumarioli* and its closest relatives from GTDB revealed 1869 gene clusters in the pangenome, with 1,253 core gene clusters, and 146 gene clusters unique to *C. fumarioli*, 68 of which were annotated to COG categories (Figure 6B; Supplementary Table S12). The COG categories "Carbohydrate transport and metabolism" and "Defense mechanisms" contained the largest number of unique gene clusters. The "Carbohydrate transport and metabolism" category includes ABC-type sugar transport system components and a 5-Carboxyvanillate decarboxylase (LigW) that catalyzes the conversion of 5-carboxyvanillate to vanillate in the biochemical pathway for the degradation of lignin (Vladimirova et al., 2016). The "Defense mechanisms" category mainly includes CRISPR-Cas system-associated proteins, and additionally a few toxin/antitoxin (TA) system proteins, and an ATPase component of the ABC-type multidrug transport system. Considering that two reference genomes used in the pangenome analysis of *C. fumarioli* were obtained from a hot spring sediment metagenome and they belong to the same species as *C. fumarioli*, and the proteins associated with CRISPR-Cas and TA systems are among the most common unique gene clusters of *C. fumarioli*, viruses infecting *Archaea* may be more prevalent in fumaroles than in hot springs, and the unique gene clusters might reflect adaptation strategies to the fumarole environment of organisms represented by *C. fumarioli*.

For example, one of the unique gene clusters in the COG category "Energy production and conversion" encodes nitrite reductase (NO-forming), part of denitrification, and an "Intracellular trafficking, secretion, and vesicular transport" category unique gene cluster that encodes VirB4 ATPase of the Type IV secretion systems, which mediate the transfer of proteins and DNA across the bacterial cell membrane and play important roles in bacterial pathogenesis and horizontal transfer of antibiotic resistance (Walldén et al., 2012). Moreover, the presence of "Inorganic ion transport and metabolism" category unique gene clusters comprising components of ABC-type Fe^{3+} transport system, ABC-type nitrate/sulfonate/bicarbonate transport system, sulfite exporter TauE/SafE/YfcA and related

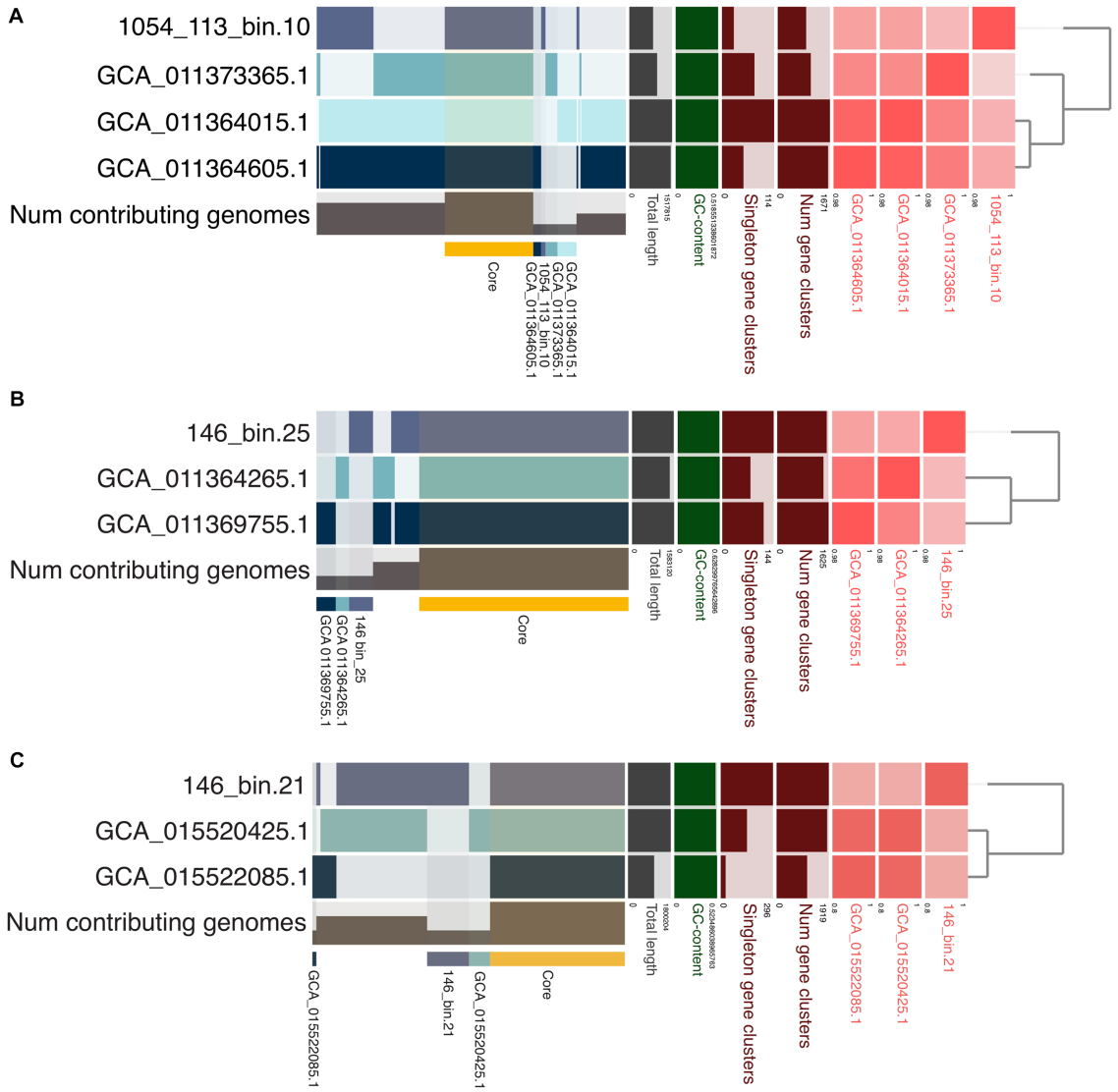


FIGURE 6
Pangenomic analysis of **(A)** 1054_113_bin.10 from Chile, **(B)** 146_bin.25 (*C. fumarioli*) from Hawai'i, and **(C)** 146_bin.21 (*P. maunauluense*) from Hawai'i. Phylogenomic trees show the relationship between the MAGs compared. The red boxes in various shades represent average nucleotide identities (ANI) between the MAGs, with the darkest shade indicating highest ANI values. Core gene clusters found in all MAGs in each comparison are highlighted with a yellow horizontal bar. Unique gene clusters present in respective MAGs are also highlighted similarly using different colors and labeled with their respective accession numbers or bin names.

permeases, and enzymes associated with assimilatory sulfate and assimilatory nitrate reductions, might provide insights into the adaptations of *Caldarchaeales* to the physicochemical characteristics of the fumaroles. Conversely, two reference genomes from the Jinze hot spring sediment share 110 gene clusters that were also not found in *C. fumarioli*. Among these, 54 were annotated, and a large number belong in the COG categories "Amino acid transport and metabolism," "Translation, ribosomal structure and biogenesis," "Nucleotide transport and metabolism," and "Replication, recombination and repair." One gene cluster in the COG category, "Inorganic ion transport and metabolism," was annotated as a predicted copper/silver-translocating P-type ATPase, which suggests potential involvement in heavy metal stress resistance. However, none of these 110 gene clusters fell into the "Defense mechanisms" category, including CRISPR-Cas and toxin/antitoxin systems.

With respect to the pangenome of *P. maunauluense* and its closest GTDB relatives, the total number of gene clusters was 2,209, with 953 core gene clusters, and 296 gene clusters unique to *P. maunauluense* is 296. Of the latter, 173 were not assigned to any COG category (Figure 6C; Supplementary Table S13). Gene clusters assigned to the COG category "Defense mechanisms" comprise the majority among 123 annotated unique gene clusters. These "Defense mechanism" gene clusters consist of proteins associated with CRISPR-Cas and toxin-antitoxin systems (TAs). TAs are small genetic modules found on bacterial mobile genetic scaffolds such as plasmids, as well as on bacterial and archaeal chromosomes (Yamaguchi et al., 2011; Yang and Walsh, 2017; Song and Wood, 2020). They typically consist of two elements: a toxin that inhibits an essential cellular process, and an antitoxin that counteracts its cognate toxin (Jurėnas et al., 2022). TAs play a critical role in the distribution and evolution of bacterial

antibiotic resistance (Yang and Walsh, 2017) and promote cell survival in their native habitat (Page and Peti, 2016; Yang and Walsh, 2017). It has also been demonstrated that they are involved in multi-resistant plasmid maintenance, stress management, and biofilm formation in *Bacteria* (Yang and Walsh, 2017).

Song and Wood (2020) reported that a primary physiological function of chromosomally encoded TA systems in bacteria is phage inhibition, that some CRISPR-Cas system elements are derived from TA systems, and some CRISPR-Cas systems mimic TA systems by reducing host metabolism to inhibit phage propagation. Given a substantial number of annotated unique gene clusters of *P. maunauluense* belong in CRISPR-Cas and TA systems, viruses of *Archaea* might be relatively more abundant in the fumarole environment from which *P. maunauluense* was recovered than in the marine hydrothermal vent from which two MAGs from the same genus were obtained. Other unique gene clusters of *P. maunauluense* comprise COG categories "Mobilome: prophages, transposons," which includes transposases, "Inorganic ion transport and metabolism," which includes a high affinity permease of iron and lead ions, and superoxide dismutases implicated in oxidative stress resistance, "Nucleotide transport and metabolism," "Carbohydrate transport and metabolism," "Amino acid transport and metabolism," "Coenzyme transport and metabolism," "Lipid transport and metabolism," "Secondary metabolites biosynthesis, transport and catabolism," "Cell motility," "Cell wall/membrane/envelope biogenesis," and "Energy production and conversion" which includes tungsten-containing aldehyde:ferredoxin oxidoreductases and tetrathionate reductase subunit A, "Post-translational modification, protein turnover, chaperones" that includes predicted dithiol-disulfide isomerases which might be involved in oxidative stress tolerance (Khairnar et al., 2013), "Replication, recombination and repair" that includes DNA repair photolyase, "Signal transduction mechanisms," "Transcription," and "Translation, ribosomal structure and biogenesis."

On the other hand, 113 gene clusters belong exclusively to the two MAGs from the marine hydrothermal vent and were not found in *P. maunauluense*. Among them, 66 were assigned to COG categories. Most of the gene clusters belong to the COG categories, "Energy production and conversion," "Amino acid transport and metabolism," "Lipid transport and metabolism," and "Defense mechanisms," with the first one containing the largest number of gene clusters. "Energy production and conversion" includes perchlorate reductase subunit alpha, which catalyzes the reduction of perchlorate to chlorite, and allows anaerobic growth on perchlorate as the electron acceptor; also included is tetrathionate reductase subunit B, a tungsten-containing aldehyde:ferredoxin oxidoreductase, and heterodisulfide reductase, subunit A (polyferredoxin)/coenzyme F420-reducing hydrogenase, delta subunit complex that may be involved in flavin-based electron bifurcation. There is also one gene cluster annotated as "Phage portal protein BeeE," which is a member of the COG category "Mobilome: prophages, transposons," implying past viral invasions. Interestingly, the "Defense mechanisms" category does not include CRISPR-Cas system-associated proteins, but it does include TA system proteins, an antibiotic ABC transporter ATP-binding protein, and an enamine deaminase RidA (reactive intermediate deaminase A) that protects metabolic enzymes against damage by reactive intermediates (Irons et al., 2020). Some other gene clusters that might contribute to the survival of *Caldarchaeales* in hydrothermal vent habitats encode manganese-containing antioxidant catalase (includes spore coat protein CotJC) in the COG category "Inorganic ion transport and metabolism," and an activator of 2-hydroxyglutaryl-CoA dehydratase (involved in the amino acid fermentation pathway) which were assigned to "Lipid transport and metabolism," plus a subtilisin family serine protease which could display hyperthermostable features, in the "Post-translational modification, protein turnover, chaperones" category.

We should also note that even though these pangenomic analyses could offer a perspective on organism- and niche-specific features and adaptations in *Caldarchaeales*, it is important to consider the incompleteness of the MAGs reported here, and of the reference genomes included in the analyses when interpreting the results. Due to the incomplete nature of many closely related MAGs in publicly available databases, we were only able to include relatively small numbers of their genomes in the pangenomic analyses. It is also important to note that the third pangenomic analysis of *P. maunauluense* MAG involved comparisons with genomes from at least two or more different species whereas the other two pangenomic included within-species level genomes. As such, a much larger number of core gene clusters may have been identified in the multi-species pangenomic analysis.

Conclusion

Metabolic reconstruction of five MAGs based on their gene content revealed that these members of the *Archaea* order *Caldarchaeales* display metabolic flexibility. They are equipped to utilize both organic substrates (i.e., glucose, amino acids/peptides, and lipids) and inorganic substrates (i.e., nitrite, hydrogen sulfide, and carbon monoxide) as electron donors, indicating a mixotrophic lifestyle.

Significantly, the fact that some MAGs encode enzymes involved in the dissimilatory tetrathionate reduction, dissimilatory reduction of nitrate and nitrite, and acetate fermentation in addition to the components of the oxidative phosphorylation complex IV attests these *Caldarchaeales* members may thrive in aerobic, microaerophilic, and anoxic environments. The presence of genes involved in oxidative stress tolerance, plus that encoding pyruvate ferredoxin/flavodoxin oxidoreductase, which enables acetyl-CoA production under anoxic conditions, and the gene encoding pyruvate dehydrogenase, which enables acetyl-CoA production under oxic conditions, suggest that these archaea have adapted to fluctuating oxygen concentrations in their habitat. Moreover, all MAGs possess the metabolic potential to fix carbon dioxide through the roTCA cycle. Also, *Caldarchaeales* members in Hawai'i might be capable of carrying out carbon dioxide fixation via the 3HP/4HB or DC/4HB cycle based on the occurrence of some key marker genes in the MAGs, although these pathways are not complete in any of the MAGs. Taken together, the wide range of metabolic capabilities displayed by these archaea might render them selectively advantageous, in case they encounter various extreme environmental conditions.

The existence of proviral regions in the MAGs indicates that these *Caldarchaeales* members might have undergone viral infections. Furthermore, the fact that all the MAGs encode Cas proteins and CRISPR arrays suggests that they developed a response mechanism against viral stress. Intriguingly, proviral regions in the MAGs were determined to encode AMGs, different membrane transporters,

ABC-type spermidine/putrescine transport system proteins implicated in quorum sensing, genes playing a role in cell secretion, and genes involved in resistance to antibiotics, heavy metals, and extreme temperature and pH values along with the viral signature genes. These findings corroborate previous studies reporting that prophage regions in microbial host genomes might comprise genes with functions which contribute to the host fitness and adaptation in myriad habitats. Lastly, increasing the number of *Caldarchaeales* representative genomes through this work has improved the phylogenomic resolution of this group of enigmatic archaea and supports their status as a clade distinct from the Thaumarchaeota. Many gaps remain in our understanding of *Caldarchaeales* biology in geothermal habitats, but the genomic information gleaned here may help us design conditions for their growth and physiological characterization in the laboratory.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found at: https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA872141.

Author contributions

CC and DL led the BASALT project to collect samples from Mauna Ulu. SC collected the samples from Chile. AD provided funding for metagenomic sequencing. RP, PC, and SD coordinated sample processing and obtained metagenome sequencing. MS performed metagenome assembly and binning, and identification of *Caldarchaeales* in public databases. MB performed metabolic pathway analysis, and analysis of viral and CRISPR regions. MS and MB performed pangenomic analyses. JS conducted taxonomic classifications and phylogenomic analyses and created the metabolic pathway overview figure. MB, MS, and JS prepared the initial draft of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by NASA Exobiology grant (80NSSC18K1064) to AD, PC, and SD; a NASA Astrobiology Institute (NAI)-CAN7 grant (13-13NAI7_2-0018) to SC; a Science and Technology Facilities Council (STFC) grant no. ST/V000586/1 to CC; CCAS startup and University Facilitating (UFF) funds from The George Washington University to JS; and the U.S. Department of Energy, Office of Science, Biological and Environmental Research Division, under award number LANLF59G to PC.

Acknowledgments

The authors would like to thank Christy Handel at the University of Hawai'i at Mānoa and Cheryl Gleasner at Los Alamos National Laboratory for assistance with sample DNA extraction and sequencing. We acknowledge that some of the samples collected and used in this study are from Hawai'i, the Native Lands of the Hawaiian people. The Hawai'i samples were collected under NASA Planetary Science and Technology Through Analog Research (PSTAR) Program (NNH14ZDA001N-PSTAR) grant (14-PSTAR14_2-0007) to DL and Hawai'i Volcanoes National Park Permit #HAVO-2016-SCI-0023. We thank the Kilauea Military Camp (KMC) staff. We also acknowledge and thank the Complejo Turìstico Tatio Mallku (Chile) for allowing access and sampling in the El Tatio geysers field as part of a NASA Astrobiology Institute CAN7 project. This is paper number BASALT-2023-OO1 in the BASALT project. This is publication #194 from the School of Life Sciences, University of Hawai'i at Mānoa.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1216591/full#supplementary-material

References

Adam, P. S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. *ISME J.* 11, 2407–2425. doi: 10.1038/ismej.2017.122

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

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schäffer, A. A., et al. (2005). Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* 272, 5101–5109. doi: 10.1111/j.1742-4658.2005.04945.x

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252. doi: 10.1093/bioinformatics/btz859

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., et al. (2018). KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat. Biotechnol.* 36, 566–569. doi: 10.1038/nbt.4163

Baker, B. J., Saw, J. H., Lind, A. E., Lazar, C. S., Hinrichs, K.-U., Teske, A. P., et al. (2016). Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* 1:16002. doi: 10.1038/nmicrobiol.2016.2

Beam, J. P., Jay, Z. J., Schmid, M. C., Rusch, D. B., Romine, M. F., Jennings, R. D. M., et al. (2016). Ecophysiology of an uncultivated lineage of Aigarchaeota from an oxic, hot

spring filamentous "streamer" community. *ISME J.* 10, 210–224. doi: 10.1038/ismej.2015.83

Becker, R., Wilks, A., and Brownrigg, R. (2022). *maps: Draw Geographical Maps*. Available at: https://cran.r-project.org/web/packages/maps/index.html (Accessed October 30, 2022).

Beg, Q. K., Kapoor, M., Mahajan, L., and Hoondal, G. S. (2001). Microbial xylanases and their industrial applications: a review. *Appl. Microbiol. Biotechnol.* 56, 326–338. doi: 10.1007/s002530100704

Berg, I. A., Kockelkorn, D., Ramos-Vera, W. H., Say, R. F., Zarzycki, J., Hügler, M., et al. (2010). Autotrophic carbon fixation in archaea. *Nat. Rev. Microbiol.* 8, 447–460. doi: 10.1038/nrmicro2365

Bevers, L. E., Bol, E., Hagedoorn, P.-L., and Hagen, W. R. (2005). WOR5, a novel tungsten-containing aldehyde oxidoreductase from *Pyrococcus furiosus* with a broad substrate specificity. *J. Bacteriol.* 187, 7056–7061. doi: 10.1128/JB.187.20.7056-7061.2005

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

Brochier-Armanet, C., Forterre, P., and Gribaldo, S. (2011). Phylogeny and evolution of the Archaea: one hundred genomes later. *Curr. Opin. Microbiol.* 14, 274–281. doi: 10.1016/j.mib.2011.04.015

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 previously uncultivated lineage of anaerobic, thermophilic Archaea. *Nat. Commun.* 13:3773. doi: 10.1038/s41467-022-31452-8

Bushnell, B. (2014). *BBTools software package*. Available at: https://jgi.doe.gov/data-and-tools/software-tools/bbtools/bb-tools-user-guide/bbmap-guide/

Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: functional annotation, Orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293

Caspi, R., Billington, R., Keseler, I. M., Kothari, A., Krummenacker, M., Midford, P. E., et al. (2020). The MetaCyc database of metabolic pathways and enzymes-a 2019 update. *Nucleic Acids Res.* 48, D445–D453. doi: 10.1093/nar/gkz862

Chan, P. P., and Lowe, T. M. (2019). tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14. doi: 10.1007/978-1-4939-9173-0_1

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. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848

Cockell, C. S., Harrison, J. P., Stevens, A. H., Payler, S. J., Hughes, S. S., Kobs Nawotniak, S. E., et al. (2019). A low-diversity microbiota inhabits extreme terrestrial basaltic terrains and their fumaroles: implications for the exploration of Mars. *Astrobiology* 19, 284–299. doi: 10.1089/ast.2018.1870

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

Danson, M. J. (1993). Chapter 1 central metabolism of the archaea. *New Comprehens Biochem.* 26, 1–24. doi: 10.1016/S0167-7306(08)60250-1

Dibrova, D. V., Galperin, M. Y., and Mulkidjanian, A. Y. (2014). Phylogenomic reconstruction of archaeal fatty acid metabolism. *Environ. Microbiol.* 16, 907–918. doi: 10.1111/1462-2920.12359

Dombrowski, N., Teske, A. P., and Baker, B. J. (2018). Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* 9:4999. doi: 10.1038/s41467-018-07418-0

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* 6, 3–6. doi: 10.1038/s41564-020-00834-3

Fang, Y., Yuan, Y., Liu, J., Wu, G., Yang, J., Hua, Z., et al. (2021). Casting light on the adaptation mechanisms and evolutionary history of the widespread Sumerlaeota. *MBio* 12, e00350–e00321. doi: 10.1128/mBio.00350-21

Gabler, F., Nam, S.-Z., Till, S., Mirdita, M., Steinegger, M., Söding, J., et al. (2020). Protein sequence analysis using the MPI bioinformatics toolkit. *Curr. Protoc. Bioinformatics* 72:e108. doi: 10.1002/cpbi.108

Gould, S. B., Garg, S. G., Handrich, M., Nelson-Sathi, S., Gruenheit, N., Tielens, A. G. M., et al. (2019). Adaptation to life on land at high O_2 via transition from ferredoxin-to NADH-dependent redox balance. *Proc. Biol. Sci.* 286:20191491. doi: 10.1098/rspb.2019.1491

Gribaldo, S., and Brochier-Armanet, C. (2012). Time for order in microbial systematics. *Trends Microbiol.* 20, 209–210. doi: 10.1016/j.tim.2012.02.006

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

Guy, L., and Ettema, T. J. G. (2011). The archaeal "TACK" superphylum and the origin of eukaryotes. *Trends Microbiol.* 19, 580–587. doi: 10.1016/j.tim.2011.09.002

He, Y., Li, M., Perumal, V., Feng, X., Fang, J., Xie, J., et al. (2016). Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat. Microbiol.* 1:16035. doi: 10.1038/nmicrobiol.2016.35

Hedlund, B. P., Chuvochina, M., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., Palmer, M., et al. (2022). SeqCode: a nomenclatural code for prokaryotes described from sequence data. *Nat. Microbiol.* 1702–1708. doi: 10.1038/s41564-022-01214-9

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.". *Extrem. Life Extreme Cond.* 18, 865–875. doi: 10.1007/s00792-014-0664-7

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

Herbold, C. W., Lee, C. K., McDonald, I. R., and Cary, S. C. (2014). Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nat. Commun.* 5:3875. doi: 10.1038/ncomms4875

Hua, Z.-S., Qu, Y.-N., Zhu, Q., Zhou, E.-M., Qi, Y.-L., Yin, Y.-R., et al. (2018). Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* 9:2832. doi: 10.1038/s41467-018-05284-4

Hughes, S. S., Haberle, C. W., Kobs Nawotniak, S. E., Sehlke, A., Garry, W. B., Elphic, R. C., et al. (2019). Basaltic terrains in Idaho and Hawai'i as planetary analogs for Mars geology and astrobiology. *Astrobiology* 19, 260–283. doi: 10.1089/ast.2018.1847

Hügler, M., and Sievert, S. M. (2011). Beyond the Calvin cycle: autotrophic carbon fixation in the ocean. *Ann. Rev. Mar. Sci.* 3, 261–289. doi: 10.1146/annurev-marine-120709-142712

Hyatt, D., Chen, G.-L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Ingram-Smith, C., Woods, B. I., and Smith, K. S. (2006). Characterization of the acyl substrate binding pocket of acetyl-CoA synthetase. *Biochemistry* 45, 11482–11490. doi: 10.1021/bi061023e

Irons, J. L., Hodge-Hanson, K., and Downs, D. M. (2020). RidA proteins protect against metabolic damage by reactive intermediates. *Microbiol. Mol. Biol. Rev.* 84, e00024–e00020. doi: 10.1128/MMBR.00024-20

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

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

Jungbluth, S. P., Amend, J. P., and Rappé, M. S. (2017). Metagenome sequencing and 98 microbial genomes from Juan de Fuca ridge flank subsurface fluids. *Sci. Data* 4:170037. doi: 10.1038/sdata.2017.37

Jurėnas, D., Fraikin, N., Goormaghtigh, F., and Van Melderen, L. (2022). Biology and evolution of bacterial toxin-antitoxin systems. *Nat. Rev. Microbiol.* 20, 335–350. doi: 10.1038/s41579-021-00661-1

Kanai, A., Sato, A., Fukuda, Y., Okada, K., Matsuda, T., Sakamoto, T., et al. (2009). Characterization of a heat-stable enzyme possessing GTP-dependent RNA ligase activity from a hyperthermophilic archaeon, *Pyrococcus furiosus*. *RNA* 15, 420–431. doi: 10.1261/rna.1122109

Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006

Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019, 1–13. doi: 10.7717/peerj.7359

Kazlauskas, D., Krupovic, M., and Venclovas, Č. (2016). The logic of DNA replication in double-stranded DNA viruses: insights from global analysis of viral genomes. *Nucleic Acids Res.* 44, 4551–4564. doi: 10.1093/nar/gkw322

Kehrer, D., Ahmed, H., Brinkmann, H., and Siebers, B. (2007). Glycerate kinase of the hyperthermophilic archaeon *Thermoproteus tenax*: new insights into the phylogenetic distribution and physiological role of members of the three different glycerate kinase classes. *BMC Genomics* 8:301. doi: 10.1186/1471-2164-8-301

Khairnar, N. P., Joe, M.-H., Misra, H. S., Lim, S.-Y., and Kim, D.-H. (2013). FrnE, a cadmium-inducible protein in *Deinococcus radiodurans*, is characterized as a disulfide isomerase chaperone in vitro and for its role in oxidative stress tolerance in vivo. *J. Bacteriol.* 195, 2880–2886. doi: 10.1128/JB.01503-12

Kieft, K., Zhou, Z., and Anantharaman, K. (2020). VIBRANT: automated recovery, annotation and curation of microbial viruses, and evaluation of viral community function from genomic sequences. *Microbiome* 8:90. doi: 10.1186/s40168-020-00867-0

King, G. (2003a). Molecular and culture-based analyses of aerobic carbon monoxide oxidizer diversity. *Appl. Environ. Microbiol.* 69, 7257–7265. doi: 10.1128/AEM.69.12.7257-7265.2003

King, G. (2003b). Contributions of atmospheric CO and hydrogen uptake to microbial dynamics on recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* 69, 4067–4075. doi: 10.1128/AEM.69.7.4067-4075.2003

Knapik, K., Becerra, M., and González-Siso, M.-I. (2019). Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios hot spring in Spain. *Sci. Rep.* 9:11195. doi: 10.1038/s41598-019-47637-z

Kochetkova, T. V., Rusanov, I. I., Pimenov, N. V., Kolganova, T. V., Lebedinsky, A. V., Bonch-Osmolovskaya, E. A., et al. (2011). Anaerobic transformation of carbon monoxide by microbial communities of Kamchatka hot springs. *Extrem. Life Extreme Cond.* 15, 319–325. doi: 10.1007/s00792-011-0362-7

Könneke, M., Schubert, D. M., Brown, P. C., Hügler, M., Standfest, S., Schwander, T., et al. (2014). Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO2 fixation. *Proc. Natl. Acad. Sci. U. S. A.* 111, 8239–8244. doi: 10.1073/pnas.1402028111

Kozlowski, J., Stieglmeier, M., Schleper, C., Klotz, M. G., and Stein, L. Y. (2016). Pathways and key intermediates required for obligate aerobic ammonia-dependent chemolithotrophy in bacteria and Thaumarchaeota. *ISME J* 10, 1836–1845. doi: 10.1038/ismej.2016.2

Krupovic, M., Gribaldo, S., Bamford, D. H., and Forterre, P. (2010). The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. *Mol. Biol. Evol.* 27, 2716–2732. doi: 10.1093/molbev/msq161

Krupovic, M., Makarova, K. S., Wolf, Y. I., Medvedeva, S., Prangishvili, D., Forterre, P., et al. (2019). Integrated mobile genetic elements in Thaumarchaeota. *Environ. Microbiol.* 21, 2056–2078. doi: 10.1111/1462-2920.14564

Lagkouvardos, I., Joseph, D., Kapfhammer, M., Giritli, S., Horn, M., Haller, D., et al. (2016). IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. *Sci. Rep.* 6:33721. doi: 10.1038/srep33721

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

Li, H. (2018). *Seqtk*. Available at: https://github.com/lh3/seqtk.

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

Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., et al. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268. doi: 10.1093/nar/gkz991

Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., et al. (2015a). An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* 13, 722–736. doi: 10.1038/nrmicro3569

Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015b). 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

Mall, A., Sobotta, J., Huber, C., Tschirner, C., Kowarschik, S., Bačnik, K., et al. (2018). Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science* 359, 563–567. doi: 10.1126/science.aao2410

Megevand, V., Carrizo, D., Lezcano, M. A., Moreno-Paz, M., Cabrol, N. A., Parro, V., et al. (2022). Lipid profiles from fresh biofilms along a temperature gradient on a hydrothermal stream at El Tatio (Chilean Andes), as a proxy for the interpretation of past and present biomarkers beyond earth. *Front. Microbiol.* 13:811904. doi: 10.3389/fmicb.2022.811904

Meng, J., Xu, J., Qin, D., He, Y., Xiao, X., and Wang, F. (2014). Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* 8, 650–659. doi: 10.1038/ismej.2013.174

Merchant, N., Lyons, E., Goff, S., Vaughn, M., Ware, D., Micklos, D., et al. (2016). The iPlant collaborative: cyberinfrastructure for enabling data to discovery for the life sciences. *PLoS Biol.* 14:e1002342. doi: 10.1371/journal.pbio.1002342

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

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., et al. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Res.* 49, D412–D419. doi: 10.1093/nar/gkaa913

Mukund, S., and Adams, M. W. (1995). Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus*. *J. Biol. Chem.* 270, 8389–8392. doi: 10.1074/jbc.270.15.8389

Müller, M., Mentel, M., van Hellemond, J. J., Henze, K., Woehle, C., Gould, S. B., et al. (2012). Biochemistry and evolution of anaerobic energy metabolism in eukaryotes. *Microbiol. Mol. Biol. Rev.* 76, 444–495. doi: 10.1128/MMBR.05024-11

Munson-McGee, J. H., Snyder, J. C., and Young, M. J. (2018). Archaeal viruses from high-temperature environments. *Genes (Basel)* 9:128. doi: 10.3390/genes9030128

Musfeldt, M., and Schönheit, P. (2002). Novel type of ADP-forming acetyl coenzyme a synthetase in hyperthermophilic archaea: heterologous expression and characterization of isoenzymes from the sulfate reducer Archaeoglobus fulgidus and the methanogen *Methanococcus jannaschii*. *J. Bacteriol.* 184, 636–644. doi: 10.1128/JB.184.3.636-644.2002

Nunoura, T., Chikaraishi, Y., Izaki, R., Suwa, T., Sato, T., Harada, T., et al. (2018). A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science* 359, 559–563. doi: 10.1126/science.aao3407

Nunoura, T., Hirayama, H., Takami, H., Oida, H., Nishi, S., Shimamura, S., et al. (2005). Genetic and functional properties of uncultivated thermophilic crenarchaeotes from a subsurface gold mine as revealed by analysis of genome fragments. *Environ. Microbiol.* 7, 1967–1984. doi: 10.1111/j.1462-2920.2005.00881.x

Nunoura, T., Takaki, Y., Kakuta, J., Nishi, S., Sugahara, J., Kazama, H., et al. (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* 39, 3204–3223. doi: 10.1093/nar/gkq1228

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

Orita, I., Sato, T., Yurimoto, H., Kato, N., Atomi, H., Imanaka, T., et al. (2006). The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* 188, 4698–4704. doi: 10.1128/JB.00492-06

Page, R., and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* 12, 208–214. doi: 10.1038/nchembio.2044

Parizzi, L. P., Grassi, M. C. B., Llerena, L. A., Carazzolle, M. F., Queiroz, V. L., Lunardi, I., et al. (2012). The genome sequence of *Propionibacterium acidipropionici* provides insights into its biotechnological and industrial potential. *BMC Genomics* 13, 1–20. doi: 10.1186/1471-2164-13-562

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

Petitjean, C., Deschamps, P., López-García, P., Moreira, D., and Brochier-Armanet, C. (2015). Extending the conserved phylogenetic core of archaea disentangles the evolution of the third domain of life. *Mol. Biol. Evol.* 32, 1242–1254. doi: 10.1093/molbev/msv015

Pohlschroder, M., and Esquivel, R. N. (2015). Archaeal type IV pili and their involvement in biofilm formation. *Front. Microbiol.* 6:190. doi: 10.3389/fmicb.2015.00190

Prangishvili, D., Bamford, D. H., Forterre, P., Iranzo, J., Koonin, E. V., and Krupovic, M. (2017). The enigmatic archaeal virosphere. *Nat. Rev. Microbiol.* 15, 724–739. doi: 10.1038/nrmicro.2017.125

Reysenbach, A.-L., St John, E., Meneghin, J., Flores, G. E., Podar, M., Dombrowski, N., et al. (2020). Complex subsurface hydrothermal fluid mixing at a submarine arc volcano supports distinct and highly diverse microbial communities. *Proc. Natl. Acad. Sci. U. S. A.* 117, 32627–32638. doi: 10.1073/pnas.2019021117

Rinke, C., Chuvochina, M., Mussig, A. J., Chaumeil, P.-A., Davín, A. A., Waite, D. W., et al. (2021). A standardized archaeal taxonomy for the genome taxonomy database. *Nat. Microbiol.* 6, 946–959. doi: 10.1038/s41564-021-00918-8

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

Roux, S., Enault, F., Hurwitz, B. L., and Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. doi: 10.7717/peerj.985

Roux, S., Páez-Espino, D., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., et al. (2021). IMG/VR v3: an integrated ecological and evolutionary framework for interrogating genomes of uncultivated viruses. *Nucleic Acids Res.* 49, D764–D775. doi: 10.1093/nar/gkaa946

Ruiz-Fernández, P., Ramírez-Flandes, S., Rodríguez-León, E., and Ulloa, O. (2020). Autotrophic carbon fixation pathways along the redox gradient in oxygen-depleted oceanic waters. *Environ. Microbiol. Rep.* 12, 334–341. doi: 10.1111/1758-2229.12837

Russel, J., Pinilla-Redondo, R., Mayo-Muñoz, D., Shah, S. A., and Sørensen, S. J. (2020). CRISPRCasTyper: automated identification, annotation, and classification of CRISPR-Cas loci. *CRISPR J.* 3, 462–469. doi: 10.1089/crispr.2020.0059

Sanchez-Garcia, L., Fernandez-Martinez, M. A., García-Villadangos, M., Blanco, Y., Cady, S. L., Hinman, N., et al. (2019). Microbial biomarker transition in high-altitude sinter mounds from El Tatio (Chile) through different stages of hydrothermal activity. *Front. Microbiol.* 9:3350. doi: 10.3389/fmicb.2018.03350

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., et al. (2021). Database resources of the National Center for biotechnology information. *Nucleic Acids Res.* 49, D10–D17. doi: 10.1093/nar/gkaa892

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

Schönheit, P., and Schäfer, T. (1995). Metabolism of hyperthermophiles. *World J. Microbiol. Biotechnol.* 11, 26–57. doi: 10.1007/BF00339135

Siebers, B., Tjaden, B., Michalke, K., Dörr, C., Ahmed, H., Zaparty, M., et al. (2004). Reconstruction of the central carbohydrate metabolism of *Thermoproteus tenax* by use of genomic and biochemical data. *J. Bacteriol.* 186, 2179–2194. doi: 10.1128/JB.186.7.2179-2194.2004

Sokolova, T. G., Henstra, A.-M., Sipma, J., Parshina, S. N., Stams, A. J. M., and Lebedinsky, A. V. (2009). Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. *FEMS Microbiol. Ecol.* 68, 131–141. doi: 10.1111/j.1574-6941.2009.00663.x

Song, S., and Wood, T. K. (2020). A primary physiological role of toxin/antitoxin systems is phage inhibition. *Front. Microbiol.* 11:1895. doi: 10.3389/fmicb.2020.01895

Spang, A., Martijn, J., Saw, J. H., Lind, A. E., Guy, L., and Ettema, T. J. G. (2013). Close encounters of the third domain: the emerging genomic view of archaeal diversity and evolution. *Archaea* 2013:202358. doi: 10.1155/2013/202358

Stec, B., Yang, H., Johnson, K. A., Chen, L., and Roberts, M. F. (2000). MJ0109 is an enzyme that is both an inositol monophosphatase and the "missing" archaeal fructose-1,6-bisphosphatase. *Nat. Struct. Biol.* 7, 1046–1050. doi: 10.1038/80968

Steffens, L., Pettinato, E., Steiner, T. M., Mall, A., König, S., Eisenreich, W., et al. (2021). High CO2 levels drive the TCA cycle backwards towards autotrophy. *Nature* 592, 784–788. doi: 10.1038/s41586-021-03456-9

Sutter, J.-M., Tästensen, J.-B., Johnsen, U., Soppa, J., and Schönheit, P. (2016). Key enzymes of the Semiphosphorylative Entner-Doudoroff pathway in the Haloarchaeon *Haloferax volcanii*: characterization of glucose dehydrogenase, gluconate dehydratase, and 2-Keto-3-Deoxy-6-Phosphogluconate aldolase. *J. Bacteriol.* 198, 2251–2262. doi: 10.1128/JB.00286-16

Suzek, B. E., Wang, Y., Huang, H., McGarvey, P. B., and Wu, C. H.UniProt Consortium (2015). UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31, 926–932. doi: 10.1093/bioinformatics/btu739

UniProt Consortium (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489. doi: 10.1093/nar/gkaa1100

van Wolferen, M., Orell, A., and Albers, S.-V. (2018). Archaeal biofilm formation. *Nat. Rev. Microbiol.* 16, 699–713. doi: 10.1038/s41579-018-0058-4

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

Verhees, C. H., Akerboom, J., Schiltz, E., de Vos, W. M., and van der Oost, J. (2002). Molecular and biochemical characterization of a distinct type of fructose-1,6-bisphosphatase from *Pyrococcus furiosus*. *J. Bacteriol.* 184, 3401–3405. doi: 10.1128/JB.184.12.3401-3405.2002

Vignais, P. M. (2008). Hydrogenases and H(+)-reduction in primary energy conservation. *Results Probl. Cell Differ.* 45, 223–252. doi: 10.1007/400_2006_027

Vladimirova, A., Patskovsky, Y., Fedorov, A. A., Bonanno, J. B., Fedorov, E. V., Toro, R., et al. (2016). Substrate distortion and the catalytic reaction mechanism of 5-Carboxyvanillate decarboxylase. *J. Am. Chem. Soc.* 138, 826–836. doi: 10.1021/jacs.5b08251

Walldén, K., Williams, R., Yan, J., Lian, P. W., Wang, L., Thalassinos, K., et al. (2012). Structure of the VirB4 ATPase, alone and bound to the core complex of a type IV secretion system. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11348–11353. doi: 10.1073/pnas.1201428109

Wang, Y., Feng, X., Natarajan, V. P., Xiao, X., and Wang, F. (2019). Diverse anaerobic methane- and multi-carbon alkane-metabolizing archaea coexist and show activity in Guaymas Basin hydrothermal sediment. *Environ. Microbiol.* 21, 1344–1355. doi: 10.1111/1462-2920.14568

Wang, Y., Li, P., Guo, Q., Jiang, Z., and Liu, M. (2018). Environmental biogeochemistry of high arsenic geothermal fluids. *Appl. Geochem.* 97, 81–92. doi: 10.1016/j.apgeochem.2018.07.015

Wang, Y., Wegener, G., Ruff, S. E., and Wang, F. (2021). Methyl/alkyl-coenzyme M reductase-based anaerobic alkane oxidation in archaea. *Environ. Microbiol.* 23, 530–541. doi: 10.1111/1462-2920.15057

Whitman, W. B., Chuvochina, M., Hedlund, B. P., Hugenholtz, P., Konstantinidis, K. T., Murray, A. E., et al. (2022). Development of the SeqCode: a proposed nomenclatural code for uncultivated prokaryotes with DNA sequences as type. *Syst. Appl. Microbiol.* 45:126305. doi: 10.1016/j.syapm.2022.126305

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. *2nd edn*. Switzerland: Springer.

Wirth, J., and Young, M. (2020). The intriguing world of archaeal viruses. *PLoS Pathog.* 16:e1008574. doi: 10.1371/journal.ppat.1008574

Wolf, Y. I., Makarova, K. S., Yutin, N., and Koonin, E. V. (2012). Updated clusters of orthologous genes for Archaea: a complex ancestor of the Archaea and the byways of horizontal gene transfer. *Biol. Direct* 7, 1–15. doi: 10.1186/1745-6150-7-46

Xu, J., Ericson, C. F., Lien, Y.-W., Rutaganira, F. U. N., Eisenstein, F., Feldmüller, M., et al. (2022). Identification and structure of an extracellular contractile injection system from the marine bacterium *Algoriphagus machipongonensis*. *Nat. Microbiol.* 7, 397–410. doi: 10.1038/s41564-022-01059-2

Yamaguchi, Y., Park, J.-H., and Inouye, M. (2011). Toxin-antitoxin systems in bacteria and archaea. *Annu. Rev. Genet.* 45, 61–79. doi: 10.1146/annurev-genet-110410-132412

Yang, S., Lv, Y., Liu, X., Wang, Y., Fan, Q., Yang, Z., et al. (2020). Genomic and enzymatic evidence of acetogenesis by anaerobic methanotrophic archaea. *Nat. Commun.* 11:3941. doi: 10.1038/s41467-020-17860-8

Yang, Q. E., and Walsh, T. R. (2017). Toxin-antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbiol. Rev.* 41, 343–353. doi: 10.1093/femsre/fux006

Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249

Zhang, Q., and Ye, Y. (2017). Not all predicted CRISPR-Cas systems are equal: isolated cas genes and classes of CRISPR like elements. *BMC Bioinformatics.* 18:92. doi: 10.1186/s12859-017-1512-4

Zhang, F., Zhao, S., Ren, C., Zhu, Y., Zhou, H., Lai, Y., et al. (2018). CRISPRminer is a knowledge base for exploring CRISPR-Cas systems in microbe and phage interactions. *Commun. Biol.* 1:180. doi: 10.1038/s42003-018-0184-6

Zhou, Z., Tran, P. Q., Breister, A. M., Liu, Y., Kieft, K., Cowley, E. S., et al. (2022). METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. *Microbiome.* 10:33. doi: 10.1186/s40168-021-01213-8

Zimmermann, L., Stephens, A., Nam, S.-Z., Rau, D., Kübler, J., Lozajic, M., et al. (2018). A completely Reimplemented MPI bioinformatics toolkit with a new HHpred server at its Core. *J. Mol. Biol.* 430, 2237–2243. doi: 10.1016/j.jmb.2017.12.007

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 14 November 2023
DOI 10.3389/fmicb.2023.1211915

OPEN ACCESS

EDITED BY
Rafael R. de la Haba,
University of Sevilla, Spain

REVIEWED BY
Gaosen Zhang,
Chinese Academy of Sciences (CAS), China
Huang Gang,
Fujian Normal University, China

*CORRESPONDENCE
Jianwei Chen
✉ chenjianwei@genomics.cn
Yuxian Wang
✉ yxwang@njtech.edu.cn
Ling Jiang
✉ jiangling@njtech.edu.cn

RECEIVED 25 April 2023
ACCEPTED 11 October 2023
PUBLISHED 14 November 2023

CITATION
Zhang Z, Zhu J, Ghenijan O, Chen J, Wang Y and Jiang L (2023) Prokaryotic taxonomy and functional diversity assessment of different sequencing platform in a hyper-arid Gobi soil in Xinjiang Turpan Basin, China. *Front. Microbiol.* 14:1211915. doi: 10.3389/fmicb.2023.1211915

Prokaryotic taxonomy and functional diversity assessment of different sequencing platform in a hyper-arid Gobi soil in Xinjiang Turpan Basin, China

Zhidong Zhang[1], Jing Zhu[1], Osman Ghenijan[1], Jianwei Chen[2]*, Yuxian Wang[3]* and Ling Jiang[3,4]*

[1]Xinjiang Key Laboratory of Special Environmental Microbiology, Institute of Applied Microbiology, Xinjiang Academy of Agricultural Sciences, Urumqi, China, [2]BGI Research, Qingdao, China, [3]College of Food Science and Light Industry, Nanjing Tech University, Nanjing, China, [4]State Key Laboratory of Materials-Oriented Chemical Engineering, Nanjing Tech University, Nanjing, China

Turpan Basin located in the eastern Xinjiang is a typical arid inland basin with extremely scarce water resources and a fragile ecosystem. Prokaryotic communities with unique genetic and physiological modifications can survive and function in such harsh environments, offering diverse microbial resources. However, numerous microbes can enter the viable but non-culturable state because of drought stress in the desert soil. In this work, next generation sequencing (NGS) technology based on DNA nanoball sequencing platform (DNBSEQ-G400) and sequencing-by-synthesis platform (NovaSeq 6000) were applied to analyze the prokaryotic diversity in three hyper-arid Gobi soils from Flaming Mountain, Toksun, and Kumtag. The comparison between two platforms indicated that DNBSEQ-G400 had better repeatability and could better reflect the prokaryotic community of this hyper-arid region. The diversity analysis based on DNBSEQ-G400 identified a total of 36 bacterial phyla, including *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, *Acidobacteriota*, *Nitrososphaerota*, and *Planctomycetota*. The environmental factors, including soluble salt, available potassium, total nitrogen, and organic matter, were positively correlated with the abundance of most prokaryote. In addition, the prokaryotic community assembly in hyper-arid soil was well described by neutral-based models, indicating that the community assembly was mainly controlled by stochastic processes. Finally, the phylogenetic analysis of *Actinomycetota* proved that such extremophiles played an important role in the ecosystems they colonize. Overall, our result provides a reference for choosing the appropriate sequencing platform and a perspective for the utilization of soil microbial resources from hyper-arid regions.

KEYWORDS

Turpan Basin, next generation sequencing, hyper-arid soil, prokaryotic diversity, community assembly

Introduction

Climate extremes have a significant impact on the global ecology, which will be further aggravated continuously (Zhang Q. et al., 2021; Zhang Y. et al., 2021). Among them, dryland zones are one of the typical ecologically fragile regions in the global ecosystem. It has been reported that dryland zones, including semi-arid, arid and hyper-arid areas, account for 40% of Earth's terrestrial surface (Yao et al., 2020). Within dryland conditions, hyper-arid areas have the driest environment characterized by extremely low mean annual precipitation, extremely high ultraviolet (UV) irradiation and an atmospheric relative humidity that usually drops to zero in the afternoon, resulting in barren land without vegetation (Mao et al., 2018; Belov et al., 2019). Although the living conditions in hyper-arid areas are extremely harsh, they have rich and diversified microbial resources, which are an important part of desert ecosystems (Zhang et al., 2020). Therefore, it is of great significance to comprehensively and deeply understand the impact of high temperature and drought on microbial behavior and ecosystem functions.

Xinjiang Uygur Autonomous Region constitutes more than 80% of the arid zone in northwestern China (He et al., 2023). The Turpan Basin, located in eastern Xinjiang, is a typical arid inland basin with extremely scarce water resources and a fragile ecosystem (Huang et al., 2021). The extreme evaporation combined with low precipitation rates results in the hyper-arid continental climate of this zone. In this extreme environment, only highly drought-tolerant oligotrophic microbial species can survive, and even these are discontinuously distributed. Nevertheless, a series of recent studies demonstrated that the hyper-arid soils found in these zones can support complex microbial assemblages (Zhang et al., 2013; Li et al., 2020). Li et al. isolated a total of 13 cultivable bacterial strains from volcanic soil samples taken in the Turpan Basin (Li et al., 2020). Zhang et al. constructed a metagenomic library from soil samples of the Turpan Basin, and screened a novel β-galactosidase from an unculturable microorganism, which exhibited high thermostability and tolerance to reaction products (Zhang et al., 2013). Therefore, the investigation of prokaryotic community diversity in the hyper-arid environment is important not only for predicting the responses of ecosystems to environmental changes and enhancing adaptability across various ecological systems, but also for facilitating the biotechnological application of soil bacteria thriving in challenging and life-limiting conditions.

Although there is currently a significant amount of research focused on arid regions, studies on the biodiversity of hyper-arid areas in Xinjiang remain relatively scarce because the abundance of microbes in such environments can be very low, resulting in a narrow niche and low resource competitiveness (Feng et al., 2020). Thus, it is difficult to isolate microbes from extreme environments using traditional pure culture methods in the laboratory (Li et al., 2020). With the continuous development of sequencing technology in recent years, the research on extreme soil microbial ecology has made significant advances (Shu and Huang, 2022), among which the microbial community structure and functional regulation have become research hotspots (Shu and Huang, 2022; Viruel et al., 2022). Modern high-throughput sequencing technologies classified as next generation sequencing (NGS) can not only detect very small amounts of microbial DNA, but also deliver results that are closer to the real community structure of microorganisms due to the large sequencing volume and high-throughput of these methods. Hwang et al. investigated viral genomes from the Atacama Desert using Illumina HiSeq 2500 to reveal the diversity and ecological impact of viruses inhabiting hyper-arid soils (Hwang et al., 2021). Using the same sequencing platform, Le et al. compared the metagenomes of soils from extreme hyper-arid deserts to understand the relation between prokaryotic communities and stress responses in soil systems (Le et al., 2016). Although NGS machines based on Illumina sequencing have dominated the sequencing market, it has been reported that Illumina index hopping can introduce false-positive contamination, causing interference in studies on community composition, diversity, and community assembly mechanisms, particularly when studying microorganisms with low abundance. Recently, MGI Tech has introduced a series of new sequencers, including the DNA nanoball (DNB) sequencing platform (DNBSEQ-G400), which promises to provide high-quality sequencing data faster and at lower prices than Illumina sequencers (Jia et al., 2022).

Here, we decided to compare the performance of Illumina sequencing (NovaSeq 6000) with DNBSEQ-G400 for deep sequencing of prokaryotic 16S rRNA genes from hyper-arid Gobi soils in the Turpan Basin. Based on the results, we evaluated the impact of the sequencing platform on community analysis in extremely low biomass environments, and revealed the predominant prokaryotic community composition, diversity, and functions. We hope this work can provide guidance for the choice of sequencing platform, so as to better inform the sustainable development and reconstruction of fragile ecosystems in hyper-arid areas.

Materials and methods

Sampling sites

Field sampling was undertaken in June 2020 in the Turpan Basin, which is located at the middle east of Xinjiang Uyghur Autonomous Region of China (42°30′-43°20' N, 87°50′-91°10′E). The Gobi Desert soil samples were collected at three sites (Figure 1), Toksun Desert (TKS), Flaming Mountain (FM), and Kumtag Desert (KMTG). The topography of this region is characterized by interlaced hills and plains with elevations ranging from −155 to 3,600 m (Pei et al., 2015). As a consequence of its extremely continental warm temperate location, Turpan has a climate with particularly high temperature, large temperature differences between day and night, large amounts of sunshine, strong solar irradiation, scarce rainfall, as well as frequent and strong winds (Pei et al., 2015). Therefore, this region experiences a hyper-arid climate with a mean annual precipitation of 6.9–25 mm, an average evaporation capacity of 2,727–3837.8 mm, and a mean annual temperature of 13°C, while the maximum temperature is over 49.6°C (Pei et al., 2015; Eminniyaz et al., 2017). The soil matrix in this region is mainly classified as sandy and gravelly, leading to the severe scarcity of vegetation and microorganisms in Gobi soil (Liu C. et al., 2021; Liu J. et al., 2021).

Collection of soil samples

Eight samples were collected approximately 1 km apart in a circle with a diameter of 3 km at each sampling location (TKS, FM, and

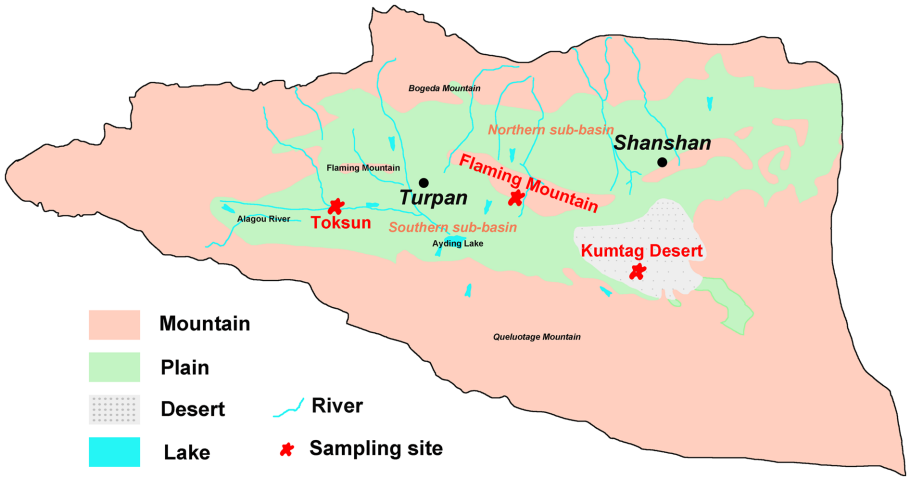


FIGURE 1
Hydrogeological map and sampling sites of Turpan Basin.

KMTG). At each sampling site, three samples were taken randomly from a 5 × 5 m homogeneous area and thoroughly mixed in sterile polypropylene bags after removing impurities. Finally, a total of 24 samples from three sampling locations were collected. All samples were obtained from a depth of 10–15 cm beneath the surface of the ground using a shovel. Finally, 2 kg of soil samples were obtained and stored at 4°C for soil physicochemical analyses and microbial quantification.

Soil physicochemical analysis

Soil samples from each site were taken to the laboratory in Xinjiang Key Laboratory of Special Environmental Microbiology to conduct physicochemical and microbiological analyses. The soil water content was measured via the oven drying method. The pH of the soil was measured using a conventional pH meter by suspending the soil in distilled water at a ratio of 1:2.5 (w/v). The number of microbial colony forming units (CFU/g) was obtained using the serial dilution method and the spread plate counting method (He et al., 2023). Other indexes of soil samples were determined according to standard agrochemical analysis methods, including organic matter, total nitrogen, available nitrogen, available potassium, and soluble salt (He et al., 2023).

DNA extraction and 16S rDNA amplicon sequencing

The genomic DNA was extracted from soil samples using a MP FastDNA 50 mL Spin Kit for Soil (MP Biomedicals, Santa Ana, CA, United States) according to the manufacturer's instructions. In order to obtain as much DNA as possible, 200 g soil samples were partitioned into 20 portions of 10 g each and DNA was extracted in parallel. All extracted DNA were combined and washed with 1 mL of elution buffer. Finally, the purity and concentration of the extracted DNA were determined by agarose gel electrophoresis. An aliquot of extracted DNA was used as a template for amplification. The V4 region of the prokaryotic 16S rRNA gene was amplified using the universal primer pair 515F (5′-GTGCCAGCMGCCGCGGTAA-3′) and 806R (5′-GGACTACHVGGGTWTCTAAT-3′) (Jia et al., 2022).

For DNBSEQ-G400 amplicon sequencing library construction, a two-step PCR procedure was used as described previously (Jia et al., 2022). Briefly, the first-step PCR was conducted by inserting zero to three random nucleotides before each primer pair to ensure balanced nucleotide proportion at each position, thereby improving the accuracy of base-calling. The PCR temperature program encompassed an initial denaturation step at 95°C for 10 min, followed by 20 cycles of denaturation at 98°C for 20 s, annealing at 58°C for 30 s, and elongation at 72°C for 30 s, with a final elongation step at 72°C for 10 min. For the second PCR amplification, the primer with a sample barcode and the DNBSEQ sequencer adapter were used with a temperature program encompassing initial denaturation at 95°C for 5 min, followed by 15 cycles of denaturation at 98°C for 20 s, annealing at 58°C for 30 s, and elongation at 72°C for 30 s, with a final elongation step at 72°C for 10 min. After that, the PCR products were verified and purified by agarose gel electrophoresis. Finally, DNA nanoballs were constructed for sequencing on the paired-end 200-bp DNBSEQ-G400 platform (BGI-Qingdao). To construct libraries for Illumina NovaSeq 6000 amplicon sequencing, only a one-step PCR procedure was carried out using an initial denaturation step at 98°C for 1 min, followed by 30 cycles of denaturation at 98°C for 10 s, annealing at 50°C for 30 s, and elongation at 72°C for 30 s, with a final elongation step at 72°C for 5 min. After the purification of PCR products, the Illumina TruSeq DNA PCR-Free Library Preparation Kit (Illumina, United States) was used to construct sequencing libraries according to the manufacture's protocol. Finally, the qualified libraries were sequenced on the paired-end 250-bp Illumina NovaSeq 6000 platform at Novogene Co., Ltd. (Beijing, China).

Bioinformatic analysis

The paired-end (PE) reads generated by the high-throughput sequencing platforms were assigned to samples based on their unique

barcodes and truncated by cutting off the barcode and primer sequence. The raw reads were filtered using SOAPnuke (v1.5.6) to remove adapter sequences and low-quality reads. Then, these high-quality clean reads were assembled into clean tags using FLASH (v1.2.11). Reads generated by each of the two sequencing platforms were combined and the denoising clustering algorithm unoise3 was applied to produce zero-radius operational taxonomic units (ZOTUs) in USEARCH (v10.0.240). The ZOTU taxonomic assignment at different taxonomic levels (from phylum to genus) was analyzed using the RDP training set (v18) with a 0.8 confidence cutoff value. QIIME (v1.9.1) was used to evaluate α-diversity (Observed_species, Simpson and Chao 1 indices) as well as the weighted and unweighted UniFrac and Bray-Curtis β-diversity distances (Liu et al., 2023), and rarefaction curves of observed species were drawn by the function "plot" of R.

Statistical analysis

All statistical analyses were performed using R software (v3.4.1). The significance of differences in α-diversity or phylogenetic diversity was assessed using the Wilcoxon-test. Differences of weighted and unweighted UniFrac distances were assessed using PERMANOVA in the "vegan" R package. The differences in community structure of the different samples and groups were analyzed using principal coordinate analysis (PCoA) based on weighted and unweighted calculations. The Sloan neutral community model prediction was carried out using the "MicEco" package in R. The beta Nearest Taxon Index (βNTI) values were calculated using the "picante" package in R. The co-occurrence networks of the prokaryotes were constructed using the SparCC algorithm and visualized by "igraph" package in R. PICRUST (v1.1.1) was used for predicting the functions from the 16S rRNA gene sequences according to the taxonomy affiliations.

Results

Soil physicochemical characteristics

The soil physicochemical characteristics were measured at three different hyper-arid Gobi sampling sites in Turpan Basin (TKS, FM and KMTG). The geochemical parameters were recorded for 8 soil samples at each site, including organic matter (OM), total nitrogen (N), available nitrogen (IonN), available potassium (IonK), soluble salt (Ss), soil water (Sw), and pH. As shown in Table 1, all samples were slightly alkaline, with pH values ranging from 8.42 ± 0.13 to 8.90 ± 0.14, and had extremely low moisture content (0.189 ± 0.133 ~ 0.251 ± 0.044%), which was consistent with other dry deserts (Liu L. et al., 2022; Liu S. et al., 2022). The concentration of OM was higher at FM (7.875 ± 3.029 g/kg) than at KMTG (3.188 ± 1.208 g/kg) and TKS (3.625 ± 1.370 g/kg), resulting in a high microbial count at FM (198 ± 37 CFU), while a low microbial counts at KMTG (59 ± 24 CFU) and TKS (67 ± 9 CFU). Furthermore, the content of N, IonK, and Ss were also significantly higher at FM than at KMTG and TKS. However, the soil characteristics did not vary significantly between KMTG and TKS, with the exception of IonN and Ss.

Overview of sequencing data

Two typical NGS technologies were used to sequence the 16S rRNA genes of microorganisms isolated from this hyper-arid environment. The sequencing statistics were summarized in Supplementary Table S1. A total of 2,176,485 and 3,088,056 PE reads were generated from 24 examined samples (8 samples per site) using Illumina NovaSeq 6000 and DNBSEQ-G400, respectively. After PE assembly and filtering from raw tags, 2,046,090 (NovaSeq 6000) and 2,275,748 (DNBSEQ-G400) clean tags with high-quality were obtained, from which 1,560,103 (NovaSeq 6000) and 2,146,520 (DNBSEQ-G400) effective tags were further produced by removing the chimeric sequences from clean tags, accounting for 71.7% (NovaSeq 6000) and 69.5% (DNBSEQ-G400) of the total quantified sequences. The average effective tags length across different samples was 253 bp, and Q20 was more than 96.61%. After the evaluation of sequencing data quality based on GC content and Q30, we concluded that all of the parameters met the demands for further analysis.

Comparison between the MGI and Illumina sequencing platforms

In order to more accurately reflect the composition of prokaryotic communities in the low-biomass hyper-arid environment, the prokaryotic species diversity was assessed using both Illumina NovaSeq 6000 and DNBSEQ-G400. According to the analysis of the high-throughput sequencing results, the samples contained a total of 36 phyla (Figure 2A). The dominant phyla were *Pseudomonadota*, *Bacillota*, *Bacteroidota*, and *Actinomycetota*, with an average relative

TABLE 1 Different physicochemical parameters and microbial count (CFU/g) of hyper-arid Gobi soils in FM, KMTG, and TKS.

Physicochemical parameters	FM	KMTG	TKS
Organic matter (OM, g/kg)	7.875 ± 3.029	3.188 ± 1.208	3.625 ± 1.370
Total nitrogen (N, g/kg)	0.18 ± 0.067	0.091 ± 0.075	0.068 ± 0.073
Available nitrogen (IonN, mg/kg)	8.738 ± 5.619	0.638 ± 0.819	17.05 ± 16.892
Available potassium (IonK, mg/kg)	275.825 ± 112.216	151.963 ± 50.153	155.362 ± 50.146
soluble salt (Ss, g/kg)	420.3 ± 181.056	125.313 ± 107.243	33.65 ± 30.065
pH	8.90 ± 0.14	8.54 ± 0.21	8.42 ± 0.13
soil water (Sw, %)	0.251 ± 0.044	0.189 ± 0.113	0.229 ± 0.102
Microbial count (CFU/g)	198 ± 37	59 ± 24	67 ± 9

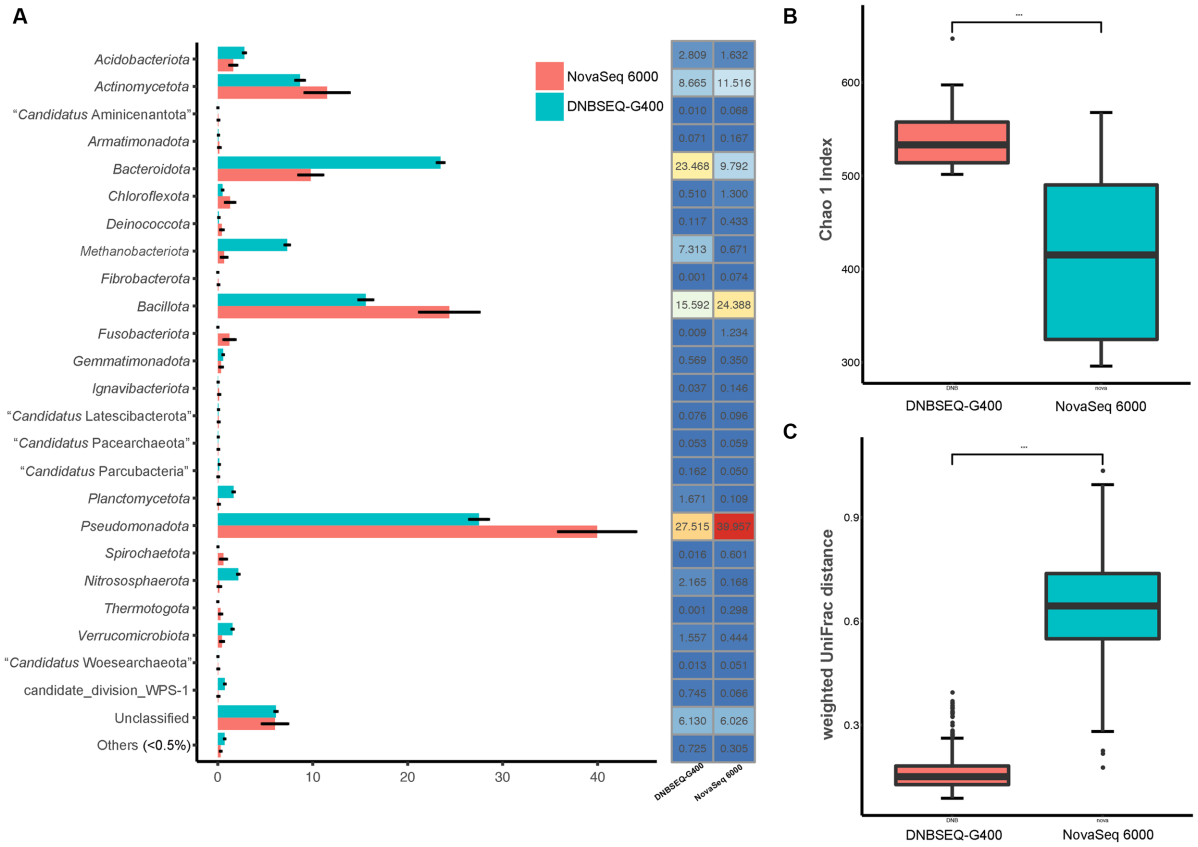


FIGURE 2
(A) Relative abundance at phyla level, **(B)** α-diversity (Chao 1 index) and **(C)** β-diversity index (weighted UniFrac distance) of prokaryotic community in Turpan Basin identified by NovaSeq 6000 and DNBSEQ-G400 platforms (***$p < 0.001$).

abundance of 39.96, 25.39, 9.79, and 11.52% according to NovaSeq 6000, as well as 27.52, 15.59, 23.47, and 8.67% according to DNBSEQ-G400. However, 4 phyla (*Elusimicrobiota*, *Lentisphaerota*, "*Candidatus* Microgenomatota," "*Candidatus* Poribacteriota") were only identified by DNBSEQ-G400, and 3 phyla ("*Candidatus* Acetothermia," "*Candidatus* Cloacimonadota," *Deferribacterota*) were only identified by NovaSeq 6000, with low abundance (<0.5%). Although the identification at the phylum level was relatively consistent between NovaSeq 6000 and DNBSEQ-G400, the former platform exhibited marginally larger error bars, indicating the poor repeatability. In addition, the Chao1 index, one of the indices used to evaluate microbial α-diversity in ecology, was much higher according to the DNBSEQ-G400 platform than according to the NovaSeq 6000 (Figure 2B), indicating that a larger proportion of prokaryotic diversity was identified by DNBSEQ-G400. Similarly, β-diversity (weighted UniFrac distance) result showed that the DNBSEQ-G400 platform presented good repeatability with relatively close distances between samples (Figure 2C).

Next, a potential niche-neutrality balancing model was used to compare the internal assembly mechanism of the prokaryotic community and its abundance distributions (Wang L. et al., 2021; Wang X. et al., 2021). The Sloan neutral model was used to investigate the microbial community assembly process. There are two important and complementary types of processes controlling the assembly of microbial communities, deterministic and stochastic process (Wang L. et al., 2021; Wang X. et al., 2021; Jia et al., 2022). As shown in Figure 3A, the Sloan neutral model-based analysis revealed that community assemblages of the DNBSEQ-G400 platform were well described by neutral-based models, with a relatively high coefficient fit ($R^2 = 0.876$), which indicated that the prokaryotic community assembly process was primarily mediated by stochastic process (Jia et al., 2022). By contrast, the fitting result of NovaSeq 6000 platform showed a relatively weak coefficient ($R^2 = 0.338$). Jiao et al. reported that the deterministic assembly was dominant in microbial communities in agricultural, forest, and grassland soils, whereas stochastic assembly contributed a larger fraction to the assembly of microbial communities in desert soils (Jiao et al., 2022). In addition, we also found that more prokaryotic phyla identified by DNBSEQ-G400 were significantly correlated with environmental factors (Figure 3B). In particular, *Actinomycetota* and *Fusobacteriota* presented negative correlation with all environmental factors, while *Bacteroidota* and *Spirochaetota* were positively correlated with environmental factors. It has been reported that *Actinomycetota* and *Bacillota* are the dominant phyla in arid soils worldwide, as they are well adapted and can survive in such barren soil with drought and high salt content (Gao et al., 2019).

In conclusion, the comparison between the NovaSeq 6000 and DNBSEQ-G400 sequencing platforms finally verified that the latter offered high accuracy for the identification of community diversity, whereby its results better matched the actual environmental conditions

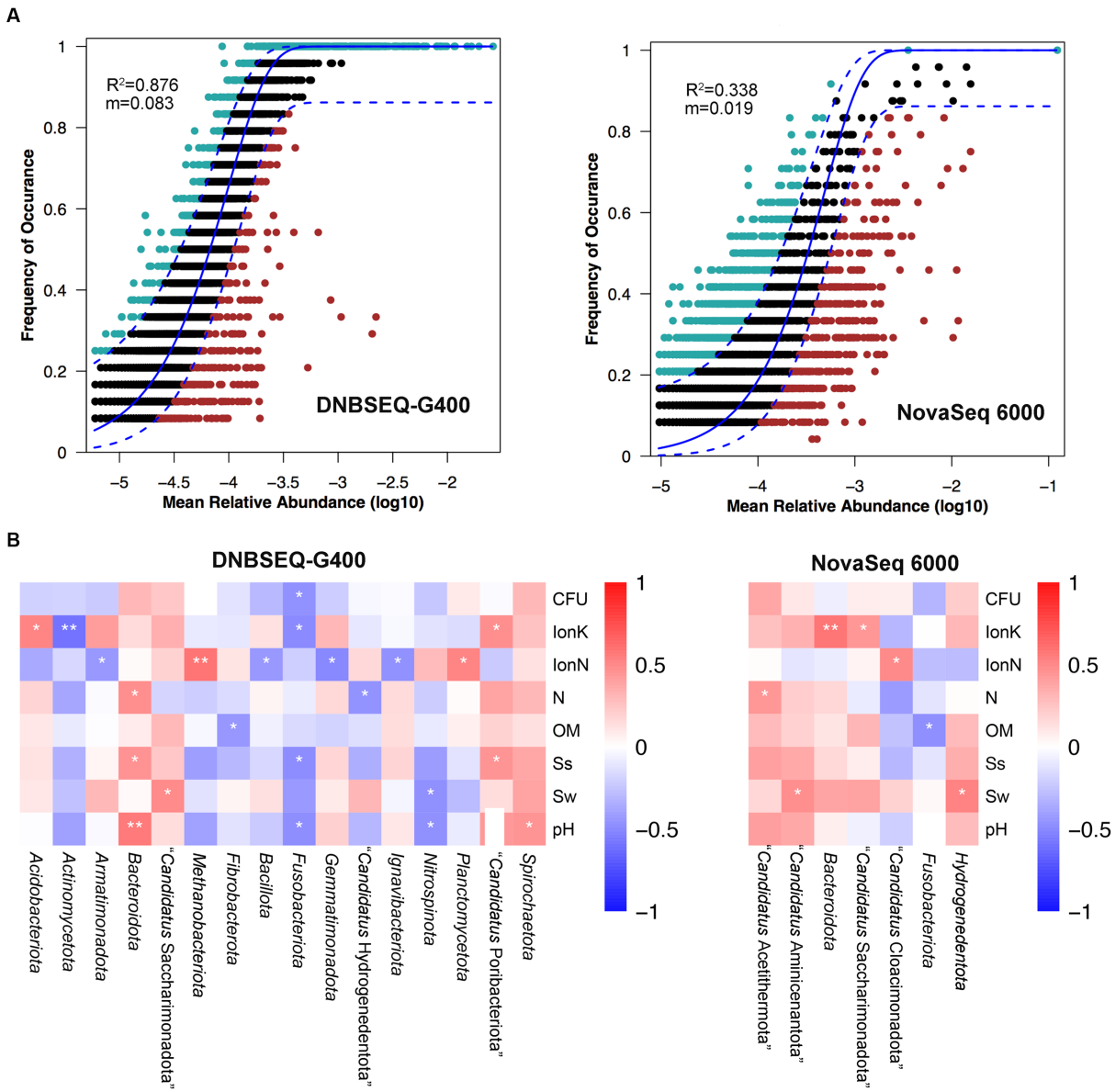


FIGURE 3
(A) Fit of Sloan's neutral model for analysis of community assembly processes. **(B)** The relationship between top phylum and environment factors of Spearman heatmap correlation analysis (**$p \leq 0.01$, *$0.01 < p \leq 0.05$).

for the prokaryotic assembly mechanism. Therefore, we further investigated the predominant prokaryotic community composition, diversity, and prokaryotic functions at three typical hyper-arid sites using only the DNBSEQ-G400 sequencing results in the following study.

Diversity analysis using the DNBSEQ-G400 sequencing platform

After clustered into ZOTUs at 100% identity thresholds, the clean tags sequenced by DNBSEQ-G400 platform were clustered into 13,161 ZOTUs. Among them, 447 ZOTUs were unique to TKS, 668 to FM, and 445 to KMTG (Supplementary Figure S1), indicating a negligible difference of prokaryotic ZOTU numbers across the soil samples. Moreover, 51.8% of prokaryotic ZOTUs were shared by all samples. The rarefaction curves of all samples tended to approach saturation plateau (Supplementary Figure S2), indicating that the sequencing depth could reasonably explain the diversity of prokaryotic communities (Liang et al., 2020). As mentioned above, a total of 36 prokaryotic phyla were detected across all soil samples. Dominant phyla were *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, *Acidobacteriota*, *Nitrososphaerota*, and *Planctomycetota* (Supplementary Figure S3). In addition, the unclassified prokaryotic communities accounted for an average of 8% of the total relative abundance at the phylum level.

The calculated α-diversity indices were used to evaluate the diversity of the prokaryotic communities. The Invsimpson indices of all samples displayed a high value (>0.974) across the three regions, indicating a great diversity of prokaryotic communities (Figure 4A). No statistically significantly differences in α-diversity were found. However, the

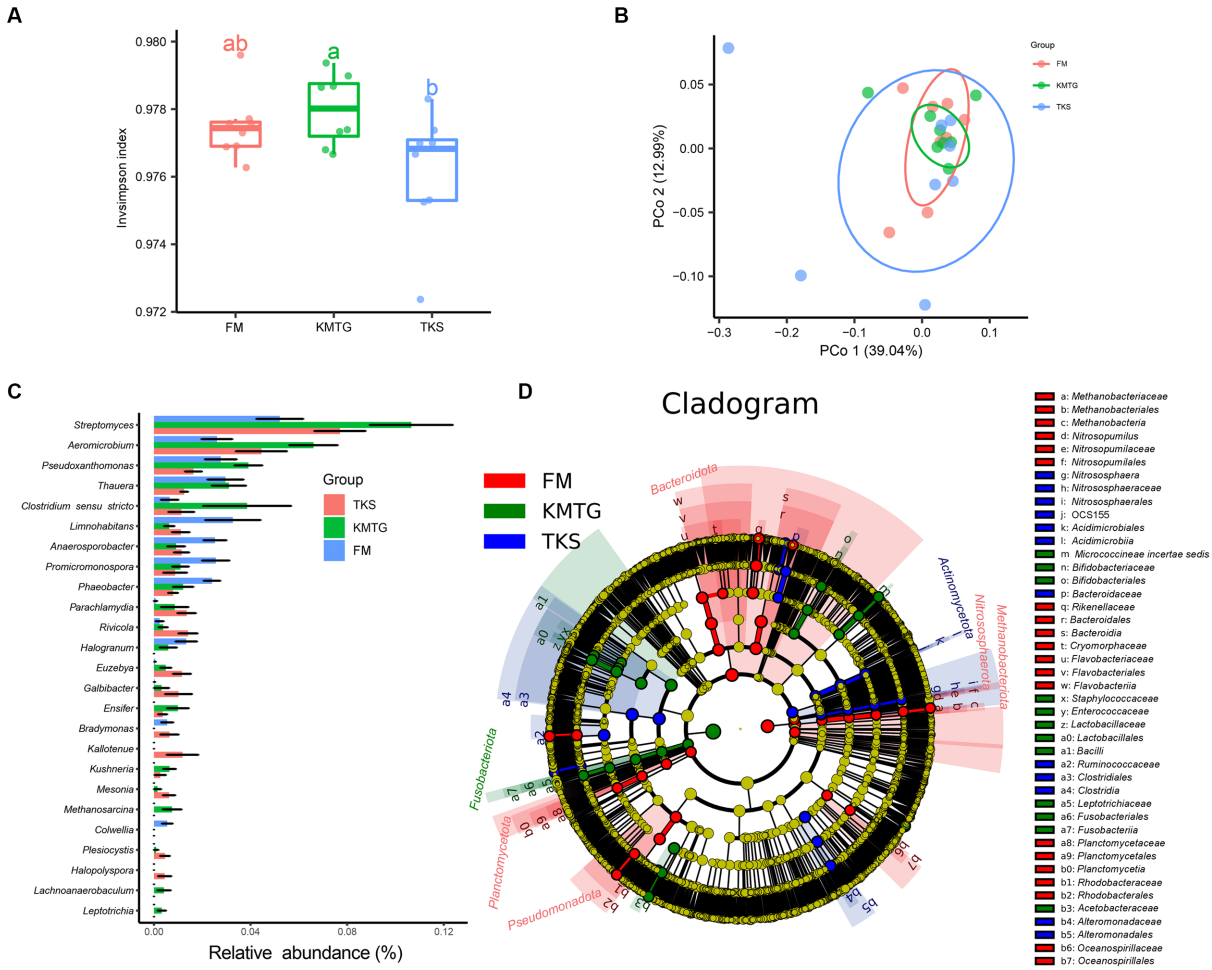


FIGURE 4
α- and β-diversity of prokaryotic communities at FM, KMTG, and TKS reflected by **(A)** Invsimpson indices and **(B)** Principal co-ordinate analysis (PCoA) (PERMANOVA test: 999 permutations, $p = 0.375$). **(C)** The 20 genera with significant difference by Kruskal-Wallis test. **(D)** LEfSe analysis of soil prokaryotic abundance at six-level cladogram (from kingdom to genus).

diversity of samples from FM was relatively concentrated, while it was more discrete for KMTG and TKS. Principal coordinates analysis (PCoA) was performed to evaluate the prokaryotic community composition based on Bray–Curtis dissimilarity index. In the hyper-arid soil, PCoA1 (39.04%) and PCoA2 (12.99%) explained 52.03% of the total microbial variation (Figure 4B). Prokaryotic community from three sites grouped together with close Bray-Curtis distance, indicating that there were some similarities between all sampling sites. The results also showed that the repeatability of samples from FM was better than that for KMTG and TKS, which was relatively discrete.

To further study the prokaryotic community structure, the differences of prokaryotic community composition in the three regions was analyzed at the taxonomy level. The Kruskal-Wallis test was first used to compare the relative abundance at the genus level among the three sites. As shown in Figure 4C, 20 genera with significant differences between FM, KMTG and TKS were identified. The top 5 prokaryotic genera from KMTG, i.e., *Streptomyces*, *Aeromicrobium*, *Pseudoxanthomonas*, *Thauera*, and *Clostridium sensu stricto* had a significantly higher abundance than at TKS and FM. However, the relative abundance of *Limnohabitans*, *Anaerosporobacter*, *Promicromonospora*, and *Phaeobacter* was higher at FM than at KMTG and TKS. Then, LEfSe analysis (linear discriminant analysis effect size; LDA cutoff ≥3) was used to illustrate the taxonomic differences from phylum to genus. In Figure 4D, each circle at a different classification level in the evolution map represents a classification at that level. The results showed that different groups at different taxonomic levels could be distinguished across sites. A total of 44 biomarkers were detected in the three regions, whereby more prokaryotic taxa were detected at FM (20 biomarkers) than at KMTG (12 biomarkers) and TKS (12 biomarkers), namely *Bacteroidota* (7 biomarkers from class to family), *Planctomycetota* (3 biomarkers from order to family), *Nitrosophaerota* (3 biomarkers from order to genus), *Methanobacteriota* (3 biomarkers from class to order), and *Pseudomonadota* (4 biomarkers from order to family). *Fusobacteriota* and *Actinomycetota* acted as a leading discriminant clade at KMTG and TKS, respectively.

Co-occurrence network patterns of prokaryotic communities affected by environmental factors

The co-occurrence network of soil prokaryotic communities was constructed to explore the correlation of core microbial taxa and

environmental factors. The correlation was statistically significant, and the soil prokaryotic network consist of 37 nodes (phylum) and 87 edges. As shown in Figure 5A, generally, the soil microbes in this co-occurrence group showed different correlations with different environmental factors. For example, *Nitrospinota* and *Fusobacteriota* exhibited negative correlations with pH, while *Spirochaetota* showed a positive correlation, which was consistent with the results shown in Figure 3A. The abundance of the majority of microorganisms was positively correlated with Ss, IonK, N, OM, and CFU, resulting in a more complex network. Such strong correlations with these environmental factors indicated that they have an important role in the composition of the prokaryotic community in hyper-arid soil. We then calculated the β-nearest taxon index (βNTI) to evaluate the changes in the relative influences of deterministic and stochastic assembly processes in hyper-arid soil. As shown in Figure 5B, the fractions of community assembly process explained by homogeneous selection ($\beta NTI \leq -2$), stochastic process ($|\beta NTI| < 2$), and variable selection ($\beta NTI \geq 2$) were determined (Jia et al., 2022). The distribution of βNTI between −2 and 2 accounted for 46.01%, indicating that prokaryotic community assembly in hyper-arid area was driven by stochastic process, which was consistent with the results of Sloan's neutral model. In addition, variable selection (50.72%) also greatly influenced the prokaryotic community assembly, demonstrating that community composition was susceptible to environmental microbial migration in hyper-arid soil.

Prokaryotic functional predictive analysis

PICRUSt software was employed to predict gene functions of soil microbiota identified in hyper-arid soil based on the annotations in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Wu et al., 2021). According to the prediction results (Figure 6A), 24 samples collected from three regions were mainly enriched in 37 metabolic pathways. The heatmap of metabolic pathways showed that the relative abundance of functions was generally similar among the samples. Among them, the majority of predicted sequences were associated with prokaryotic functions involved in carbohydrate metabolism, membrane transport, amino acid metabolism, replication and repair, as well as energy metabolism, indicating that the soil microorganisms in the Turban Basin had a particularly high utilization of carbon sources (Liu L. et al., 2022; Liu S. et al., 2022). The metabolic pathways with the most significant differences were biotin metabolism, CAM ligands, cytosolic DNA-sensing pathway, ECM-receptor interaction, focal adhesion, and primary bile acid biosynthesis (Supplementary Figure S4). In addition, a Principal Component Analysis (PCA) based on KEGG database was used to compare the global metabolic changes of the microbiota. The results showed that the three groups were not significantly different in the functional pathways (Figure 6B), illustrating that the soil microbial communities from different hyper-arid regions of the Turpan Basin were functionally convergent.

The special prokaryotic populations analysis

To identify the prokaryotic populations in this hyper-arid region, the relative abundance of taxa at the phylum, class, order, family, and genus levels of prokaryotic communities was analyzed. As elaborated in Supplementary Figure S5, the top 6 phyla were *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, and

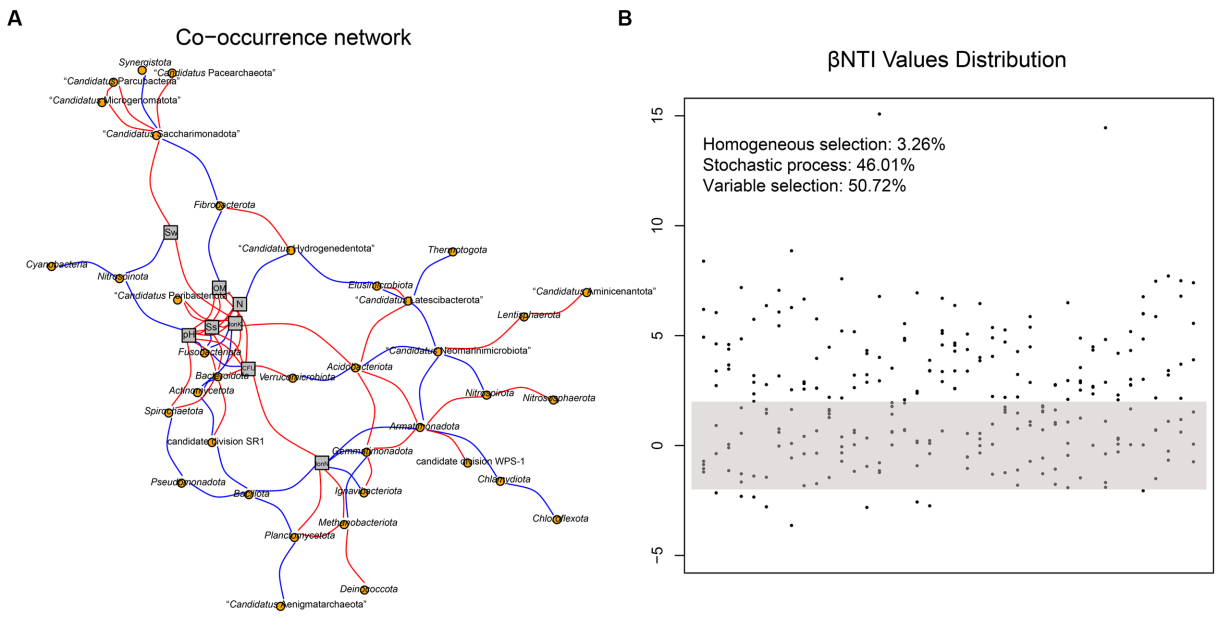


FIGURE 5
(A) Co-occurrence network based on correlation analysis between species at phylum level and environmental factors. Sw, Soil water content; OM, Organic Matter; N, Total Nitrogen; Ss, Soil salt; IonK, available potassium; IonN, available Nitrogen. **(B)** The distribution of βNTI values. Each point represents a βNTI value. A |βNTI| value of less than 2 (gray shaded region) indicates stochastic assembly processes; a βNTI value of less than −2 indicates a homogeneous selection; and a βNTI value of greater than 2 indicates a variable selection.

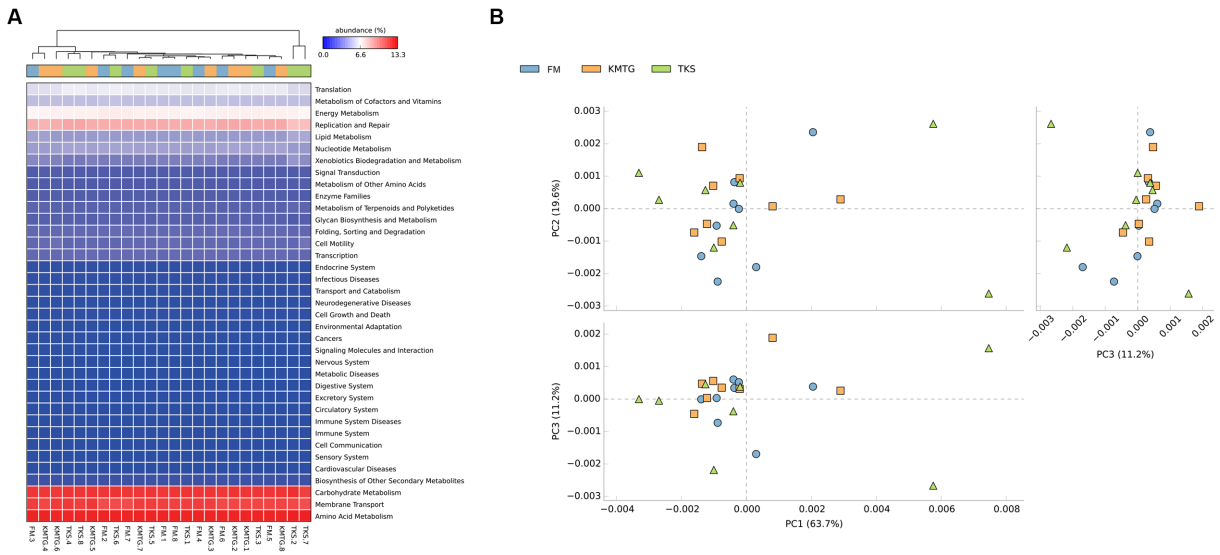


FIGURE 6
(A) Heatmap of the relative abundances of PICRUSt predicted genes for different samples. **(B)** Principal Component Analysis (PCA) of metabolic pathways (ANOSIM test: 999 permutations, $R = 0.045$, $p = 0.375$).

Acidobacteriota. However, more than 50% of the 16S rRNA sequences could not be classified into genera and species, indicating that large numbers of rare or unknown genera and species were existed in this hyper-arid region. In our previous work, we found that *Actinomycetota* were the major phyla at FM (He et al., 2023), therefore the diversity of *Actinomycetota* were further discussed here. The phylogenetic analysis showed that the microorganisms of the Turpan Basin represented a considerable level of taxonomic diversity (Figure 7). There were 247 *Actinomycetota* belonging to 12 orders, including 0319-7 L14, *Acidimicrobiales*, *Actinomycetales*, *Bifidobacteriales*, *Coriobacteriales*, *Euzebyales*, *Gaiellales*, *Nitriliruptorales*, *Rubrobacterales*, *Solirubrobacterales*, *Thermoleophilales*, and WCHB1-81. At the lower taxonomic levels, there were 86 families, 214 genera, and 247 species, demonstrating the diversity of *Actinomycetota* in this hyper-arid environment.

Discussion

The climate of the Turpan Basin is extremely dry and hot, especially in Toksun County, where the highest recorded temperature is 49.6°C, and the average annual precipitation is only 6.9 mm, making it the most arid place in China (Eminniyaz et al., 2017). It is precisely due to the extreme climate that the unique ecological environment in Turpan area has been formed, also including FM and KMTG, which constitutes a good model system to investigate the impact of hyper-arid conditions on microbial assemblages in the soil. All regolith soils in this area are mainly composed of sandy with saline-alkaline soils (Liu C. et al., 2021; Liu J. et al., 2021). The physicochemical parameters indicated that all samples in this study had very low contents of moisture, organic matter and total nitrogen, which was similar to previously described samples from the Qaidam Basin (Liu L. et al., 2022; Liu S. et al., 2022). By comparing the three regions, we found that the organic matter, total nitrogen, available potassium, and soluble salt in FM area had a relative high content than at KMTG and TKS, indicating that more nutrients were available for the growth of soil microorganisms. Consistently, the average microbial count in the FM region was more than three times that of the KMTG and TKS samples. In summary, the soil characteristics of different regions of the Turpan Basin exhibited obvious differences, leading to the spatial heterogeneity of this area.

The extreme habitat of the Turpan Basin is inhospitable to most forms of life, but it provides a haven for the thriving diversity of thermophiles (Parihar et al., 2022). However, information on the prevalence and complexity of these soil microbiota cannot be obtained by traditional cultivation methods because most microorganisms from the soil cannot be cultured in the laboratory (Li et al., 2020). In recently years, 16S rRNA amplicon sequencing technology has provided an effective strategy to analyze the microbial diversity of extreme environments (Parihar et al., 2022). Two major DNA sequencing platforms from two companies (Illumina Inc., NovaSeq 6000; MGI Tech Co., DNBSEQ-G400) were used to compare which was better able to reflect microbial diversity in hyper-arid soils with extremely low biomass abundance. The SOAPnuke results showed that both platforms provided high-quality sequencing data with sufficient coverage that were suitable for ZOTU analysis. A comparison of the two platforms revealed that the performance of DNBSEQ-G400 was mostly concordant with NovaSeq 6000, and might break the domination of Illumina in the sequencing market by offering lower prices (Anslan et al., 2021). Moreover, according to the analysis of the high-throughput sequencing results, more prokaryotic phyla could be identified based on DNBSEQ-G400 results. Some of these had notable features, such as *Elusimicrobiota*, "*Candidatus* Poribacteriota," and *Lentisphaerota*, isolated from semi-arid savanna soil and hyper-arid intermontane basin in the Qaidam Basin (Cheng et al., 2016; Xing et al., 2019), which were capable of reducing nitrate or sulfur compounds. The relative abundance and diversity index obtained from the DNBSEQ-G400 platform resulted in a relatively low

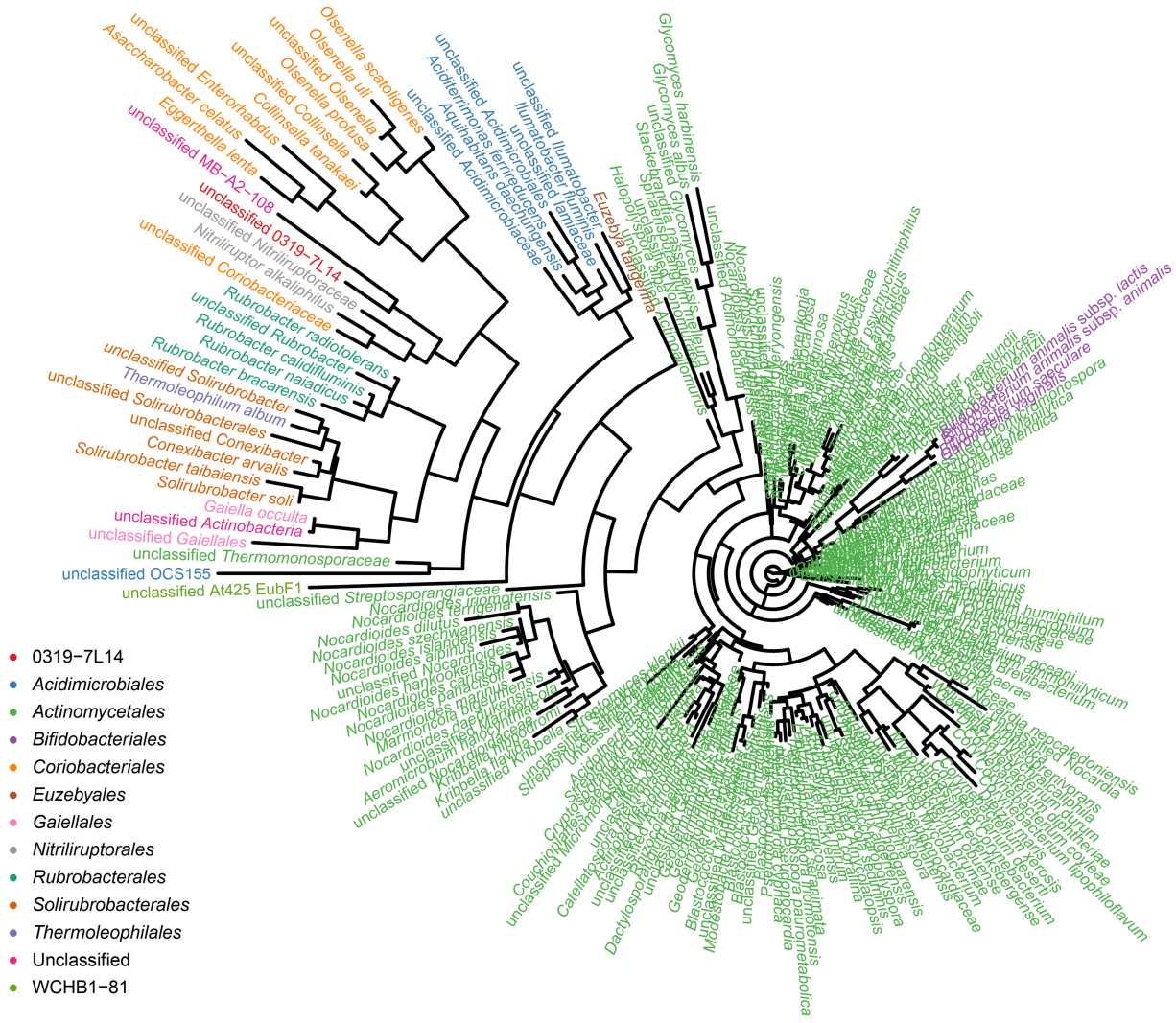


FIGURE 7
Phylogenetic tree for *Actinomycetota* species.

error range and a high value, indicating the better repeatability and high prokaryotic diversity. Notably, it was reported that 0.2 ~ 6% of index misassignment rate could occur on the Illumina sequencing platform, leading to the potential misinterpretation of sequencing results (Jia et al., 2022). Our previous work also observed a significantly lower fraction of potential false positive reads for DNBSEQ-G400 compared to NovaSeq 6000, indicating the superiority of the DNBSEQ-G400 sequencing platform (Jia et al., 2022).

The importance of disentangling community assembly mechanisms is widely recognized in microbial ecology, which is beneficial for better understanding the maintenance and generation of terrestrial microbial diversity (Zhang et al., 2019). The assembly of microbial communities is controlled by stochastic and deterministic processes, with each process governing differential fractions of microbial community compositions across diverse ecosystems. The Sloan's neutral model analyzed for DNBSEQ-G400 revealed a prominent role of stochastic process in forming the prokaryotic community. It had been reported that soil biota tended to enter a dormant state to cope with hyper-arid environment, which contributes to the resistance to environmental stressors and consequently weakens deterministic processes (Kang et al., 2022). In addition, null model analyses based on phylogenetic turnover revealed that soil prokaryotic community from hyper-arid regions were subject to the combined effects of stochastic process ($|\beta NTI| < 2$) and variable selection ($\beta NTI > 2$). Our results revealed the prominent role of stochastic process in forming the prokaryotic community of the hyper-arid Turpan Basin. Meanwhile, the contribution of variable selection in regulating prokaryotic community structure should not be neglected, indicating that community composition is susceptible to the migration of environmental microbes.

Further, Spearman correlation analysis showed that more prokaryotic phyla identified by DNBSEQ-G400 were related to soil physicochemical parameters. It had been reported that pH was the key environmental factor affecting the distribution of prokaryotic community structure, which is directly or indirectly related to available nitrogen, phosphorus, organic carbon, and metal ions (Li et al., 2021). A previous study found that the relative abundance of *Bacteroidota* and *Acidobacteriota* in the desert soil of the Ebinur Lake

Basin had a significantly correlation with the pH (Li et al., 2021). *Bacteroidota* and *Acidobacteriota* had a significant positive correlation and an extremely significantly negative correlation with pH, respectively. In our study, a similar result was also obtained for DNBSEQ-G400 sequencing platform. Moreover, the Sloan neutral model results indicated the general dominance of stochastic processes, which were reported to be more pronounced in drier soils, so that the phylogenetic structure of the community was more randomly assembled (Lee et al., 2018). All these results depicted that the prokaryotic community structure and assembly processes gleaned from the DNBSEQ-G400 data were more consistent with the characteristics of the real environment.

Considering the superiority of the DNBSEQ-G400 sequencing platform, the prokaryotic diversity was analyzed according to its sequencing results. In this study, we found that the three sampling areas had no significant differences in Simpson indices and PCoA, indicating that the prokaryotic diversity is generally similar in this area. The composition of soil microbes in the Turpan Basin arid desert was homogeneous, and dominated by xerotolerant, halotolerant, and radioresistant *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, *Acidobacteriota*, *Nitrososphaerota*, and *Planctomycetota*. Among them, *Pseudomonadota* was the predominant prokaryotic phylum, which are abundant free-living bacteria in many oligotrophic habitats, such as the hyper-arid Atacama Desert of northern Chile, or the sandy subsurface soils of Virginia and Delaware (Cao et al., 2017). In addition, environmental variables showed a significant correlation with the prokaryotic community composition. A large number of studied showed that soil prokaryotic communities are highly sensitive to changes of pH, soil organic matter and the availability of soil mineral nutrients (Daniel et al., 2018; Shao et al., 2019; Liu C. et al., 2021; Liu J. et al., 2021). In this work, we found that several dominant environmental factors, including pH, soluble salt, organic matter, total nitrogen, and available potassium, exhibited a complex network with greater connectivity, which was considered more robust to environmental stresses than simple networks with less connectivity (Wang L. et al., 2021; Wang X. et al., 2021). In addition, these environmental factors could also impact the distribution of dominant prokaryotic phyla. Overall, the prokaryotic community was mainly influenced by positive correlations, indicating that available nutrients and minerals in the soil were the main limiting factors for prokaryotic colonization.

The microbiota of hyper-arid soil are important participants in biogeochemical cycles, and PICRUSt functional predictive analysis based on high-throughput sequencing has begun to be applied in the assessment of metabolic functions, which is beneficial for exploring the adaptation of microbial communities to different environmental conditions (Zhang Q. et al., 2021; Zhang Y. et al., 2021). This study found that there was a high diversity of prokaryotic functions at the three sampling sites, and PCA results indicated that there were no significant differences in the functional pathways, showing the functional convergence of prokaryotes in the Turpan Basin. Among them, amino acid metabolism, membrane transport, and carbohydrate metabolism were the main metabolic pathways at three regions, which was in line with the previous literature on arid land (Sun et al., 2020). Amino acid metabolism enables microbes to convert ammonium salts, nitrates, and other inorganic nitrogen absorbed from the environment into proteins. Carbohydrate metabolism is a vital biochemical process that regulates the formation, decomposition, and mutual transformation of carbohydrates in microbes. Finally, membrane transport can regulate the osmotic potential of microbial cell to adapt to drought stress (Xiao et al., 2021). The enrichment of the above three metabolic pathways affected the surrounding soil nutrients in hyper-arid areas, promoting the growth of vegetation, and thus creating biodiversity in arid desert environments.

Microbial resources in hyper-arid areas have attracted significant attention from microbiologists, as water availability is a major limiting factor for all forms of life. Although low moisture and water, extreme temperature, and poor nutrients limit the growth of microorganisms, some extremophiles could still adapt to live under these harsh conditions. *Actinomycetota* are among the most frequent groups, generally accounting for over 35% of all microorganisms in hyper-arid areas (He et al., 2023). It has been reported that the *Actinomycetota* communities plays an important role in the ecosystems they colonize, as they are capable of fixing CO_2 via the Calvin-Benson-Bassham cycle to supply organic carbon to other species in oligotrophic desert ecosystems (Liu L. et al., 2022; Liu S. et al., 2022). This feature in particular illustrates their importance in nutrient cycling, which contributes to their dominance in the communities. These extremophiles as well as their metabolites have high thermal stability and bioavailability, which offers obvious application advantages and lays a foundation for their utilization as bioaugmentation agents in arid areas.

Conclusion

In conclusion, we demonstrated that MGI-Tech DNBSEQ-G400 and Illumina NovaSeq 6000 platforms both reflected the composition of prokaryotic communities in the hyper-arid soil of the Turpan Basin. However, Spearman correlation analysis showed that more prokaryotic phyla identified by DNBSEQ-G400 were negatively correlated with soil physicochemical parameters, such as pH, available nitrogen, soluble salt, and soil moisture. In addition, Sloan neutral model revealed that the prokaryotic community assemblages revealed by the DNBSEQ-G400 platform were well described by neutral-based models, with relatively high coefficient fit. These results showed that the DNBSEQ-G400 could even better illustrate the diversity of prokaryotic communities in hyper-arid environments because of the good repeatability and reasonable features of the resulting assembly. Therefore, the DNBSEQ-G400 sequencing platform was further used to analyze the prokaryotic diversity at three typical hyper-arid sites, Flaming Mountain, Toksun, and Kumtag. A total of 36 prokaryotic phyla were identified across all samples, among which the eight dominant phyla were *Pseudomonadota*, *Bacteroidota*, *Bacillota*, *Actinomycetota*, *Methanobacteriota*, *Acidobacteriota*, *Nitrososphaerota*, and *Planctomycetota*. Prokaryotic community composition showed no significant differences between the three sites, which was strongly correlated with environmental factors. In addition, Sloan neutral model and β-nearest taxon index (βNTI) indicated that prokaryotic community assembly in hyper-arid areas was driven by stochastic process. Functional annotation of the Prokaryotic community showed that carbohydrate metabolism, membrane transport, and amino acid metabolism were the main metabolic pathways at three regions. Finally, phylogenetic analysis revealed that the phylum *Actinomycetota* represented a considerable level of taxonomic diversity. Overall, this

work will lay a solid foundation for understanding the prokaryotic diversity and exploiting prokaryotic resources in hyper-arid areas.

Data availability statement

The data that support the findings of this study have been deposited in China National GeneBank Sequence Archive (CNSA) of China National GeneBank DataBase (CNGBdb) with accession number CNP0004923.

Author contributions

JC and LJ: conceptualization. ZZ and JC: methodology. ZZ and JZ: validation. JZ and YW: formal analysis and investigation. ZZ and YW: writing—original draft preparation. YW and OG: writing—review and editing. ZZ and LJ: funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (32060004, 2021YFC2102700, and U2106228), Xinjiang Academy of Agricultural Sciences Science and Technology Innovation Key Cultivation Project (xjkcpy-2021002), and the Jiangsu Synergetic Innovation Center for Advanced Bio-Manufacture (XTC2205). ZZ was supported by the Tianshan Talent Plan (2022TSYCCX0067).

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1211915/full#supplementary-material

References

Anslan, S., Mikryukov, V., Armolaitis, K., Ankuda, J., Lazdina, D., Makovskis, K., et al. (2021). Highly comparable metabarcoding results from MGI-tech and Illumina sequencing platforms. *PeerJ* 9:e12254. doi: 10.7717/peerj.12254

Belov, A. A., Cheptsov, V. S., Vorobyova, E. A., Manucharova, N. A., and Ezhelev, Z. S. (2019). Stress-tolerance and taxonomy of culturable bacterial communities isolated from a Central Mojave Desert soil sample. *Geosciences* 9:166. doi: 10.3390/geosciences9040166

Cao, C., Zhang, Y., Cui, Z., Feng, S., Wang, T., and Ren, Q. (2017). Soil bacterial community responses to revegetation of moving sand dune in semi-arid grassland. *Appl. Microbiol. Biotechnol.* 101, 6217–6228. doi: 10.1007/s00253-017-8336-z

Cheng, J., Jing, G., Wei, L., and Jing, Z. (2016). Long-term grazing exclusion effects on vegetation characteristics, soil properties and bacterial communities in the semi-arid grasslands of China. *Ecol. Eng.* 97, 170–178. doi: 10.1016/j.ecoleng.2016.09.003

Daniel, R. L., Gabriel, B., Otso, O., Leonardo, M. C., Josileis, A. Z., Masahiro, R., et al. (2018). Direct and indirect effects of a pH gradient bring insights into the mechanisms driving prokaryotic community structures. *Microbiome* 6:106. doi: 10.1186/s40168-018-0482-8

Eminniyaz, A., Qiu, J., Baskin, C. C., Baskin, J. M., and Tan, D. (2017). "Biological invasions in desert Green-Islands and grasslands" in *Biological invasions and its management in China. Invading Nature - Springer series in invasion ecology*. eds. F. Wan, M. Jiang and A. Zhan (Dordrecht: Springer)

Feng, W., Zhang, Y., Yan, R., Lai, Z., Qin, S., Sun, Y., et al. (2020). Dominant soil bacteria and their ecological attributes across the deserts in northern China. *Eur. J. Soil Sci.* 71, 524–535. doi: 10.1111/ejss.12866

Gao, J., Luo, Y., Wei, Y., Huang, Y., Zhang, H., He, W., et al. (2019). Effect of aridity and dune type on rhizosphere soil bacterial communities of *Caragana microphylla* in desert regions of northern China. *PLoS One* 14:e0224195. doi: 10.1371/journal.pone.0224195

He, Z., Wang, Y., Bai, X., Chu, M., Yi, Y., Zhu, J., et al. (2023). Bacterial community composition and isolation of Actinobacteria from the soil of Flaming Mountain in Xinjiang, China. *Microorganisms* 11:489. doi: 10.3390/microorganisms11020489

Huang, F., Ochoa, C. G., and Chen, X. (2021). Assessing environmental water requirement for groundwater-dependent vegetation in arid inland basins by combining the copula joint distribution function and the dual objective optimization: an application to the Turpan Basin, China. *Sci. Total Environ.* 799:149323. doi: 10.1016/j.scitotenv.2021.149323

Hwang, Y., Rahlff, J., Schulze-Makuch, D., Schloter, M., and Probst, A. J. (2021). Diverse viruses carrying genes for microbial extremotolerance in the Atacama Desert hyperarid soil. *mSystems* 6:3. doi: 10.1128/msystems.00385-21

Jia, Y., Zhao, S., Guo, W., Peng, L., Zhao, F., Wang, L., et al. (2022). Sequencing introduced false positive rare taxa lead to biased microbial community diversity, assembly, and interaction interpretation in amplicon studies. *Environ. Microbiome* 17:43. doi: 10.1186/s40793-022-00436-y

Jiao, S., Chu, H., Zhang, B., Wei, X., Chen, W., and Wei, G. (2022). Linking soil fungi to bacterial community assembly in arid ecosystems. *iMeta* 1:e2. doi: 10.1002/imt2.2

Kang, L., Chen, L., Zhang, D., Peng, Y., Song, Y., Kou, D., et al. (2022). Stochastic processes regulate belowground community assembly in alpine grasslands on the Tibetan plateau. *Environ. Microbiol.* 24, 179–194. doi: 10.1111/1462-2920.15827

Le, P. T., Makhalanyane, T. P., Guerrero, L. D., Vikram, S., Peer, Y. V., and Cowan, D. A. (2016). Comparative metagenomic analysis reveals mechanisms for stress response in Hypoliths from extreme hyperarid deserts. *Genome Biol. Evol.* 8, 2737–2747. doi: 10.1093/gbe/evw189

Lee, K. C., Caruso, T., Archer, S. D. J., Gillman, L. N., Lau, M. C. Y., Cary, S. C., et al. (2018). Stochastic and deterministic effects of a moisture gradient on soil microbial communities in the McMurdo dry valleys of Antarctica. *Front. Microbiol.* 9:2619. doi: 10.3389/fmicb.2018.02619

Li, Y., Chen, J., Wang, Y., Ma, D., and Rui, W. (2020). The effects of the recombinant YeaZ of *Vibrio harveyi* on the resuscitation and growth of soil bacteria in extreme soil environment. *PeerJ* 8:e10342. doi: 10.7717/peerj.10342

Li, W., Jiang, L., Zhang, Y., Teng, D., Wang, H., Wang, J., et al. (2021). Structure and driving factors of the soil microbial community associated with Alhagi sparsifolia in an arid desert. *PLoS One* 16:e0254065. doi: 10.1371/journal.pone.0254065

Liang, H., He, Z., Wang, X., Song, G., Chen, H., Lin, X., et al. (2020). Effects of salt concentration on microbial diversity and volatile compounds during suancai fermentation. *Food Microbiol.* 91:103537. doi: 10.1016/j.fm.2020.103537

Liu, J., Ding, J., Rexiding, M., Li, X., Zhang, J., Ran, S., et al. (2021). Characteristics of dust aerosols and identification of dust sources in Xinjiang, China. *Atmos. Environ.* 262:118651. doi: 10.1016/j.atmosenv.2021.118651

Liu, S., Jin, J., Yang, H., Wang, P., Liu, Q., Huang, Y., et al. (2022). Effect of feeding geese in cornfields on soil bacterial diversity and metabolic function. *Appl. Soil Ecol.* 175:104448. doi: 10.1016/j.apsoil.2022.104448

Liu, G., Li, T., Zhu, X., Zhang, X., and Wang, J. (2023). An independent evaluation in a CRC patient cohort of microbiome 16S rRNA sequence analysis methods: OUT clustering, DADA2, and Deblur. *Front. Microbiol.* 14:1178744. doi: 10.3389/fmicb.2023.1178744

Liu, L., Liu, H., Zhang, W., Chen, Y., Shen, J., Li, Y., et al. (2022). Microbial diversity and adaptive strategies in the Mars-like Qaidam Basin, north Tibetan plateau, China. *Environ. Microbiol. Rep.* 14, 873–885. doi: 10.1111/1758-2229.13111

Liu, C., Zhou, Y., Qin, H., Liang, C., Shao, S., Fuhrmann, J. J., et al. (2021). Moso bamboo invasion has contrasting effects on soil bacterial and fungal abundances, co-occurrence networks and their associations with enzyme activities in three broadleaved forests across subtropical China. *For. Ecol. Manag.* 498:119549. doi: 10.1016/j.foreco.2021.119549

Mao, B., Zhao, L., Zhao, Q., and Zeng, D. (2018). Effects of ultraviolet (UV) radiation and litter layer thickness on litter decomposition of two tree species in a semi-arid site of Northeast China. *J. Arid. Land* 10, 416–428. doi: 10.1007/s40333-018-0054-6

Parihar, J., Parihar, S. P., Suravajhala, P., and Bagaria, A. (2022). Spatial metagenomic analysis in understanding the microbial diversity of Thar Desert. *Biology* 11:461. doi: 10.3390/biology11030461

Pei, H., Fang, S., Lin, L., Qin, Z., and Wang, X. (2015). Methods and applications for ecological vulnerability evaluation in a hyper-arid oasis_ a case study of the Turpan oasis, China. *Environ. Earth Sci.* 74, 1449–1461. doi: 10.1007/s12665-015-4134-z

Shao, P., Liang, C., Kennedy, R. N., Li, X., Xie, H., and Bao, X. (2019). Secondary successional forests undergo tightly-coupled changes in soil microbial community structure and soil organic matter. *Soil Biol. Biochem.* 128, 56–65. doi: 10.1016/j.soilbio.2018.10.004

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

Sun, X., Lin, Y. L., Li, B. L., and Huang, L. F. (2020). Analysis and function prediction of soil microbial communities of *Cynomorium songaricum* in two daodi-origins. *Acta Pharm. Sin.* 55, 1334–1344. doi: 10.16438/j.0513-4870.2020-1771

Viruel, E., Fontana, C. A., Puglisi, E., Nasca, J., Banegas, N., and Cocconcelli, P. S. (2022). Land-use change affects the diversity and functionality of soil bacterial communities in semi-arid Chaco region, Argentina. *Appl. Soil Ecol.* 172:104362. doi: 10.1016/j.apsoil.2021.104362

Wang, L., Han, M., Li, X., Yu, B., Wang, H., Ginawi, A., et al. (2021). Mechanisms of niche-neutrality balancing can drive the assembling of microbial community. *Mol. Ecol.* 30, 1492–1504. doi: 10.1111/mec.15825

Wang, X., Lu, X., Li, Z., Cheng, Q., Zhou, Y., and Lei, M. (2021). Liming alters microbial community composition and its co-occurrence patterns in cd-and Pb-contaminated agricultural soil. *Appl. Soil Ecol.* 166:104064. doi: 10.1016/j.apsoil.2021.104064

Wu, S., You, F., Hall, M., and Huang, L. (2021). Native plant *Maireana brevifolia* drives prokaryotic microbial community development in alkaline Fe ore tailings under semi-arid climatic conditions. *Sci. Total Environ.* 760:144019. doi: 10.1016/j.scitotenv.2020.144019

Xiao, F., Li, Y., Li, G., He, Y., Lv, X., Zhuang, L., et al. (2021). High throughput sequencing-based analysis of the soil bacterial community structure and functions of Tamarix shrubs in the lower reaches of the Tarim River. *PeerJ* 9:e12105. doi: 10.7717/peerj.12105

Xing, R., Gao, Q., Zhang, F., Wang, J., and Chen, S. (2019). Large-scale distribution of bacterial communities in the Qaidam Basin of the Qinghai–Tibet plateau. *MicrobiologyOpen* 8:e909. doi: 10.1002/mbo3.909

Yao, J., Liu, H., Huang, J., Gao, Z., Wang, G., Li, D., et al. (2020). Accelerated dryland expansion regulates future variability in dryland gross primary production. *Nat. Commun.* 11:1665. doi: 10.1038/s41467-019-13993-7

Zhang, W., Bahadur, A., Zhang, G., Zhang, B., Wu, X., Chen, T., et al. (2020). Diverse bacterial communities from Qaidam Basin of the Qinghai–Tibet plateau: insights into variations in bacterial diversity across different regions. *Front. Microbiol.* 11:554105. doi: 10.3389/fmicb.2020.554105

Zhang, Q., Jian, S., Li, K., Wu, Z., Guan, H., Hao, J., et al. (2021). Community structure of bacterioplankton and its relationship with environmental factors in the upper reaches of the Heihe River in Qinghai plateau. *Environ. Microbiol.* 23, 1210–1221. doi: 10.1111/1462-2920.15358

Zhang, Y., Keenan, T. F., and Zhou, S. (2021). Exacerbated drought impacts on global ecosystems due to structural overshoot. *Nat. Ecol. Evol.* 5, 1490–1498. doi: 10.1038/s41559-021-01551-8

Zhang, X., Li, H., Li, C., Ma, T., Li, G., and Liu, Y. (2013). Metagenomic approach for the isolation of a thermostable β-galactosidase with high tolerance of galactose and glucose from soil samples of Turpan Basin. *BMC Microbiol.* 13, 1–10. doi: 10.1186/1471-2180-13-1

Zhang, K., Shi, Y., Cui, X., Yue, P., Li, K., Liu, X., et al. (2019). Salinity is a key determinant for soil microbial communities in a desert ecosystem. *Msystems* 4, e00225–e00218. doi: 10.1128/mSystems.00225-18

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 16 November 2023
DOI 10.3389/fmicb.2023.1293707

OPEN ACCESS

EDITED BY
Ram Karan,
University of Delhi, India

REVIEWED BY
Javier Pascual,
Darwin Bioprospecting Excellence, Spain
Elena Plotnikova,
Institute of Ecology and Genetics of Microorganisms (RAS), Russia

*CORRESPONDENCE
Rafael R. de la Haba
✉ rrh@us.es
Antonio Ventosa
✉ ventosa@us.es

RECEIVED 13 September 2023
ACCEPTED 05 October 2023
PUBLISHED 16 November 2023

CITATION
de la Haba RR, Arahal DR, Sánchez-Porro C, Chuvochina M, Wittouck S, Hugenholtz P and Ventosa A (2023) A long-awaited taxogenomic investigation of the family *Halomonadaceae*. *Front. Microbiol.* 14:1293707. doi: 10.3389/fmicb.2023.1293707

A long-awaited taxogenomic investigation of the family *Halomonadaceae*

Rafael R. de la Haba[1], David R. Arahal[2], Cristina Sánchez-Porro[1], Maria Chuvochina[3], Stijn Wittouck[4], Philip Hugenholtz[3] and Antonio Ventosa[1]*

[1]Department of Microbiology and Parasitology, Faculty of Pharmacy, University of Sevilla, Sevilla, Spain, [2]Departament of Microbiology and Ecology, University of Valencia, Valencia, Spain, [3]The University of Queensland, School of Chemistry and Molecular Biosciences, Australian Centre for Ecogenomics, St Lucia, QLD, Australia, [4]Research Group Environmental Ecology and Applied Microbiology, Department of Bioscience Engineering, University of Antwerp, Antwerp, Belgium

The family *Halomonadaceae* is the largest family composed of halophilic bacteria, with more than 160 species with validly published names as of July 2023. Several classifications to circumscribe this family are available in major resources, such as those provided by the List of Prokaryotic names with Standing in Nomenclature (LPSN), NCBI Taxonomy, Genome Taxonomy Database (GTDB), and Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB), with some degree of disagreement between them. Moreover, regardless of the classification adopted, the genus *Halomonas* is not phylogenetically consistent, likely because it has been used as a catch-all for newly described species within the family *Halomonadaceae* that could not be clearly accommodated in other *Halomonadaceae* genera. In the past decade, some taxonomic rearrangements have been conducted on the *Halomonadaceae* based on ribosomal and alternative single-copy housekeeping gene sequence analysis. High-throughput technologies have enabled access to the genome sequences of many type strains belonging to the family *Halomonadaceae*; however, genome-based studies specifically addressing its taxonomic status have not been performed to date. In this study, we accomplished the genome sequencing of 17 missing type strains of *Halomonadaceae* species that, together with other publicly available genome sequences, allowed us to re-evaluate the genetic relationship, phylogeny, and taxonomy of the species and genera within this family. The approach followed included the estimate of the Overall Genome Relatedness Indexes (OGRIs) such as the average amino acid identity (AAI), phylogenomic reconstructions using amino acid substitution matrices customized for the family *Halomonadaceae*, and the analysis of clade-specific signature genes. Based on our results, we conclude that the genus *Halovibrio* is obviously out of place within the family *Halomonadaceae,* and, on the other hand, we propose a division of the genus *Halomonas* into seven separate genera and the transfer of seven species from *Halomonas* to the genus *Modicisalibacter*, together with the emendation of the latter. Additionally, data from this study demonstrate the existence of various synonym species names in this family.

KEYWORDS

phylogenomics, signature genes, halophiles, taxonomic reclassification, genus delineation

1. Introduction

First described in 1988 to accommodate the genera *Halomonas* and *Deleya* based on 16S rRNA oligonucleotide cataloging, the family *Halomonadaceae* (Franzmann et al., 1988) is the largest bacterial family harboring halophilic microorganisms. Since then, 13 new genera with validly published names and one non-validly published name have been added to the family, as stated in the List of Prokaryotic names with Standing in Nomenclature (LPSN) (Parte et al., 2020). As of July 2023, it comprises 168 species with validly published names (according to the LPSN) that share the following features: Gram-stain-negative, non–endospore-forming rods or coccoid cells, chemo-organotrophic, aerobic or facultatively anaerobic, catalase-positive and oxidase-variable, and halophilic or halotolerant bacteria (Ventosa et al., 2021b). In accordance with the LPSN, this family belongs to the order *Oceanospirillales* (Garrity et al., 2005b), class *Gammaproteobacteria* (Garrity et al., 2005a), and other taxonomic sources indicate that it is most closely related to three genera belonging to the family *Oceanospirillaceae* (Garrity et al., 2005c): *Oceanospirillum*, *Marinospirillum*, or *Pseudospirillum*. The most widely used prokaryotic classification resources have different taxonomic opinions on this family. According to Bergey's Manual of Systematics of Archaea and Bacteria (BMSAB), the most recent description of the family *Halomonadaceae* comprises 12 genera with validly published names: *Halomonas* (type genus), *Aidingimonas*, *Carnimonas*, *Chromohalobacter*, *Cobetia*, *Halotalea*, *Kushneria*, *Larsenimonas*, *Modicisalibacter*, *Pistricoccus*, *Salinicola*, and *Zymobacter* (Ventosa et al., 2021b), as well as one genus with an effectively published name, "*Phytohalomonas*" (Liu et al., 2020). As of July 2023, LPSN and NCBI Taxonomy also assign these 13 genera to the *Halomonadaceae* and 2 additional genera: *Halovibrio* and *Terasakiispira* in LPSN (Parte et al., 2020) and *Halovibrio* and *Salicola* in NCBI Taxonomy (Schoch et al., 2020), while Genome Taxonomy Database (GTDB) (Parks et al., 2018) release 08-RS214 comprises the 13 BMSAB genera together with the genus *Terasakiispira*, the 3 *Oceanospirillaceae* genera listed above, and an unnamed genus (TA22) based on environmental sequence data. GTDB also proposes the division of the genus *Halomonas* into five distinct genera (denoted with alphabetical suffixes) for a total of 22 genera in this family. Hence, currently, there is some disagreement on the circumscription of this family, which warrants further investigation.

For the present study, we followed the taxonomy proposed by BMSAB, but irrespective of the classification adopted, most genera of the family *Halomonadaceae* are well-defined and conserved, with the notable exception of the type genus of the family, *Halomonas*. *Halomonas* is the 15th largest genus across the prokaryotes in terms of the number of species with validly published names (121 as of July 2023) according to the LPSN.[1] The species of this genus seldom form a monophyletic group and are instead scattered amongst the other genera of the family *Halomonadaceae* (de la Haba et al., 2010, 2012, 2014; Ventosa et al., 2021a,b). The polyphyly of *Halomonas* probably stems from the use of this genus as a default parent taxon for newly described members of the family that did not fall in any of the well-defined genera. Members of this genus exhibit a high degree of heterogeneity, as can be evidenced by their broad DNA G + C range (51.4–74.3 mol%), the widest across the family (Ventosa et al., 2021a). These observations point to the necessity of dividing the genus *Halomonas* and reclassifying some of its species.

After its description, three studies have proposed amendments to the family *Halomonadaceae*, accompanied by the inclusion of new genera into this family (Dobson and Franzmann, 1996; Ben Ali Gam et al., 2007; Ntougias et al., 2007). Additionally, the type genus of the family, *Halomonas*, has been amended on one occasion to accommodate nine new species in the genus (Dobson and Franzmann, 1996). Initially, a number of characteristic signatures for the 16S rRNA gene sequence were defined for members of this family (Dobson et al., 1993). Nevertheless, as new members have been described, it has been necessary to readapt the 16S rRNA gene sequence signatures (Ben Ali Gam et al., 2007; de la Haba et al., 2014). As whole-genome sequencing has become easier and cheaper, the analysis of clade-specific signature genes has been proposed as a tool to supplement the use of 16S rRNA gene sequence signatures (Zheng et al., 2020).

Several taxonomic studies have been carried out on the *Halomonadaceae*, some of them based on the 16S rRNA gene sequence analysis (Dobson et al., 1993; Mellado et al., 1995; Dobson and Franzmann, 1996), and the most recent focused on alternative phylogenetic markers, such as the 23S rRNA gene (Arahal et al., 2002; de la Haba et al., 2010) and multilocus sequence analysis based on single-copy housekeeping genes (de la Haba et al., 2012). Those approaches allowed, at that time, to establish the phylogenetic relationships among the members of this family and to propose taxonomic rearrangements accordingly. More than 10 years later, the number of species within the *Halomonadaceae* has doubled, and many of the type strains have a publicly available genome sequence. To the best of our knowledge, no studies specifically addressing the taxonomic status of the family *Halomonadaceae* in the genomic era have been conducted.

Therefore, in this study, we re-evaluate the genetic relationship, phylogeny, and taxonomy of the species and genera within the present family *Halomonadaceae*, as delineated in BMSAB, with the primary focus on its type genus. For this, we considered various overall genome relatedness indexes (OGRIs) based on average nucleotide identity (ANI), digital DNA–DNA hybridization (dDDH), average amino acid identity (AAI), core-protein average amino acid identity (cAAI), and the core-genome and GTDB phylogenies together with a signature gene analysis. To resolve the polyphyly within the genus *Halomonas*, we propose a division of the genus into seven different genera and the reclassification of several *Halomonas* species into the genus *Modicisalibacter*. Furthermore, the present study demonstrated the need to exclude the genus *Halovibrio* from the family *Halomonadaceae* and the existence of nine synonym species across this family.

2. Materials and methods

2.1. Genome retrieval and sequencing of type strains

The databases scanned for genomic data were NCBI Assembly,[2] NCBI Sequence Read Archive (SRA),[3] JGI Genome Portal,[4] ATCC

1 https://lpsn.dsmz.de/text/largest-genera

2 https://www.ncbi.nlm.nih.gov/assembly

3 https://www.ncbi.nlm.nih.gov/sra

4 https://genome.jgi.doe.gov/portal

Genome Portal,[5] and Global Catalogue of Type Strain (Shi et al., 2021). A total of 147 genome sequences of distinct type strains of species of the family *Halomonadaceae* (as per the BMSAB resource) and the genera *Halovibrio* and *Terasakiispira* were available in those public databases at the beginning of this study. Since some databases searched contained redundant data, the final set of genomes used in this research was only retrieved from the NCBI Assembly and NCBI SRA databases. One out of those 147 available genomes was discarded due to a failed quality check according to GTDB (Parks et al., 2018), while another marked as contaminated and excluded from RefSeq was finally retained after the verification of its contamination and completeness in GTDB, giving a total number of 146 genome sequences (Supplementary Table 1). All except one of the downloaded genomic data were assembled either at contig, scaffold, or chromosome level. The remaining one was only available in the NCBI SRA database, and thus, only raw reads could be retrieved, which were processed as indicated below.

To make our study as comprehensive as possible, we sequenced the genomes of an additional 17 type strains of species of *Halomonadaceae* that were not publicly available when the analyses were designed and performed (Supplementary Table 1). Seven out of those type strains were sequenced and assembled by the Global Catalogue of Microorganisms (Shi et al., 2021) as part of the type strain sequencing project (Wu and Ma, 2019). The genomic DNA of the other 10 type strains, retrieved from culture collections, was extracted and purified as described elsewhere (Galisteo et al., 2022) and further sequenced by Novogene Europe (Cambridge, United Kingdom) on the Illumina NovaSeq PE150 platform after library preparation of genomic material using the Novogene NGS DNA Library Prep Set (Cat. No. PT004). Raw reads were quality trimmed and filtered with BBTools v.38.44 (Bushnell, 2020) and further assembled with SPAdes v3.15.2 (option "—isolate") (Prjibelski et al., 2020).

For all the retrieved or obtained sequences, assembly quality, and basic statistics were estimated with QUAST v.2.3 (Gurevich et al., 2013). Genome sequence completeness and contamination were verified by using CheckM v.1.1.5 (Parks et al., 2015). Prodigal v.2.6.3 (Hyatt et al., 2010) was employed to extract the translated coding sequences from the assembled genomes, and Prokka v.1.14.6 (Seemann, 2014) was employed for the automated annotation. When more than one genome sequence was available for the same type strain, we selected the one with a better assessment of contig number, completeness, contamination, 16S rRNA gene sequence presence and length, and the presence of 20 essential amino acids coded by distinct tRNA-coding genes, resulting in a final set of 161 genomes of type strains of the family *Halomonadaceae*.

2.2. Pangenome and core-genome inference

The complete pangenome of the selected set of genome sequences was calculated with Anvi'o suite v.7.1 (Eren et al., 2021), using BLASTP+ (Camacho et al., 2009) as the amino acid sequence similarity search program. Specifically, the "anvi-pan-genome" script was used with default values except that the MCL inflation parameter was set to 5 (in order to increase the sensitivity of the Markov Cluster Algorithm) (van Dongen, 2008), and the minimum percent identity between two amino acid sequences was set to 40%. The pangenome database containing orthologous gene clusters was employed to extract the gene family presence/absence matrix ("anvi-compute-functional-enrichment-in-pan" script) as well as the translated single-copy core genes present in at least 90% of the analyzed genomes, denoted as "core90" ("anvi-get-sequences-for-gene-clusters" script).

The "core90" proteins were individually aligned with MUSCLE v.3.8.1551 (default parameters) (Edgar, 2004). Gappy columns (>50% gaps) from each alignment were removed and further pruned using the chi2 statistic to filter out the 20% least conserved columns, as implemented in the "alignment_pruner.pl" Perl script.[6] Final alignments were concatenated whenever necessary using AMAS (Borowiec, 2016).

2.3. Estimation of amino acid substitution models

Model-based phylogenetic analyses of protein sequences strongly rely on amino acid substitution models, which are generally summarized in a 20-by-20 replacement matrix, designated as *Q* matrices. Since those matrices are computationally very expensive to estimate due to their high parametrization, they are not usually estimated during a phylogenetic analysis but selected from a pre-estimated set of *Q* matrices using model selection software (Minh et al., 2021).

With the aim of constructing very reliable phylogenies, in this study, we estimated empirical general time-reversible *Q* matrices customized for the family *Halomonadaceae* derived from multiple sequence alignments of selected gene sets. For that purpose, QMarker software (Minh et al., 2021), which follows a maximum-likelihood approach, was used. Two gene datasets were chosen to generate *Halomonadaceae*-specific *Q* matrices: the "core90" set and the "bac120" set (120 single-copy bacterial proteins used by GTDB taxonomy) (Parks et al., 2018).

2.4. Phylogenomic reconstructions and overall genome relatedness indexes

The trimmed and concatenated "core90" protein alignment was used to infer the maximum-likelihood phylogeny of the family *Halomonadaceae* with the IQ-TREE v.2.2.0 program (Minh et al., 2020), using the custom Q_core90 reversible matrix as the substitution model combined with profile mixture models C10–C60 (Quang et al., 2008). The best-fit model parameters (mixture models, amino acid frequencies, and rate heterogeneity across sites) were determined with ModelFinder (Kalyaanamoorthy et al., 2017) according to Bayesian Information Criterion (BIC) (Schwarz, 1978). Branch support was estimated via 1,000 ultrafast bootstrap approximations (Hoang et al.,

5 https://genomes.atcc.org

6 https://github.com/novigit/davinciCode/blob/master/perl

2018). Normalization of taxonomic ranks at the genus level according to relative evolutionary divergence (RED) was carried out with PhyloRank.[7]

A second phylogenomic tree including type and non-type strains as well as metagenome-assembled genomes (MAGs) was constructed to assess the cluster stability when new genome sequences are added to the tree and to analyze the relationships between members of the family *Halomonadaceae* and other closely related genera of the order *Oceanospirillales* (i.e., *Halospina*, *Marinospirillum*, *Oceanospirillum*, and *Pseudospirillum*). This second tree was based on the "bac120" protein set retrieved from GTDB Release 202 data files using GTDB-Tk v. 1.7.0 (Chaumeil et al., 2019). However, unlike the original GTDB-Tk pipeline that randomly selected 42 columns per marker to reduce computational requirements, we obtained the full-length multiple sequence alignments for each protein, which were later trimmed and concatenated as explained above for the "core90" dataset. Phylogenetic inference was carried out as stated for the first tree, but the Q_bac120 reversible matrix was used instead. Arrangements and visualization of both trees were accomplished by using ARB v.7.0 (Westram et al., 2011).

Commonly used OGRIs for taxonomic purposes were pairwise calculated among the analyzed genome set. ANI metrics were estimated with PyANI v.0.2.10 (Pritchard et al., 2016) using the ANIb method (i.e., BLASTN+ to align 1,020-bp fragments of the input sequences) (Goris et al., 2007). The genome-to-genome distance calculator (GGDC) webserver was used to infer dDDH relatedness according to formula 2 (Meier-Kolthoff et al., 2013). AAI values were determined using the "aai.rb" script from the Enveomics collection (Rodriguez-R and Konstantinidis, 2016). Since horizontal gene transfer (HGT) events can distort the molecular clock of bacterial evolution (Novichkov et al., 2004) and can, thus, impact the pairwise AAI values (Zheng et al., 2020), we calculated an additional genome relatedness index denoted as cAAI, similar to AAI but based on the protein sequences of core orthologous gene families, which are less impacted by HGT. For that purpose, the previously defined "core90" protein dataset (unaligned and untrimmed) was selected to calculate cAAI values with the "aai.rb" script.

2.5. Representation of clade-specific signature genes

The gene family presence/absence matrix inferred from the pangenome was used to detect "signature genes," which are defined as gene families exclusive to specific phylogenetic clades (phylogroups), that is, gene families present in all species of a clade and absent in all other species (Zheng et al., 2020). Several phylogroups were investigated, aimed at testing different taxonomic arrangements and splitting into the family *Halomonadaceae*. For each phylogroup, a representative species was chosen, and the remaining species under study were assigned to the phylogroup of the representative species with whom they shared a most recent common ancestor.

The signature gene UpSet plots were generated by gathering gene families with the same presence/absence pattern across the analyzed genomes, ruling out two types of trivial patterns: gene families present in a single genome (singleton genes) and gene families detected in all genomes (core genes). The patterns were then displayed in descending order of frequency (number of gene families). R packages "tidyverse" v.1.2.1 (Wickham et al., 2019) and "tidygenomes" v.0.1.3,[8] as well as other R functions and scripts,[9] were used for data processing and visualization of gene family presence/absence and signature genes.

3. Results and discussion

3.1. Core-genome phylogenomics suggest several rearrangements in the family *Halomonadaceae*

A genome-based analysis was carried out aimed at shedding light on the relationships among species and genera of the family *Halomonadaceae*. The pangenome of the 161 genome sequences, which included type strains of the species belonging to the family *Halomonadaceae*, consisted of 608,808 genes organized into 59,992 orthologous gene clusters. Only 55 single-copy core gene clusters were detected across all genomes, so a more relaxed set of 189 orthologous genes shared by at least 90% of the genomes ("core90") encompassing a total of 23,924 residues after alignment and trimming was selected to calculate a *Halomonadaceae*-specific Q reversible matrix (Q_core90; Supplementary Table 2) and to infer a phylogenomic tree (Figure 1 and Supplementary Figure 1). The best-fit model selected by ModelFinder according to the BIC criterion to construct the tree was Q_core90 + C60 + F + R10. The resulting tree showed that, apart from the genus *Halomonas*, all the remaining genera within this family formed monophyletic groups. Noteworthy, the genera *Terasakiispira* and primarily *Halovibrio* were by far the most distantly related to the other genera within *Halomonadaceae*, suggesting that they might not belong to this family (see below).

Given the polyphyletic origin of the genus *Halomonas* and the heterogeneous characteristics of the species it harbors, we evaluated several proposals to split this genus into several genera with the following constraints: (i) the new proposed genera must be monophyletic; (ii) they must be supported by a 100% bootstrap value when possible; and (iii) they must fall within the relative evolutionary divergence (RED) interval defined for the rank of genus (Parks et al., 2018). The first attempt (*proposal I*) was to divide the genus *Halomonas* into eight phylogroups, five of them denoted according to GTDB taxonomy as *Halomonas* (*sensu stricto*), *Halomonas*_A, *Halomonas*_B, *Halomonas*_E, and *Halomonas*_F, and three newly proposed phylogroups designated as *Halomonas*_G, *Halomonas*_H, and *Halomonas*_I, as well as to transfer the species *Halomonas coralii*, *Halomonas ilicicola*, *Halomonas muralis*, *Halomonas radicis*, *Halomonas xianhensis*, and *Halomonas zincidurans* to the genus *Modicisalibacter*. Subsequent proposals were similar to this but included merging some of the previous phylogroups: *proposal II* = *Halomonas* (*sensu stricto*) + *Halomonas*_G; *proposal III* = *Halomonas* (*sensu stricto*) + *Halomonas*_G + *Halomonas*_H; *proposal*

7 https://github.com/dparks1134/PhyloRank

8 https://github.com/SWittouck/tidygenomes

9 https://github.com/SWittouck/lacto_genera

10.3389/fmicb.2023.1293707

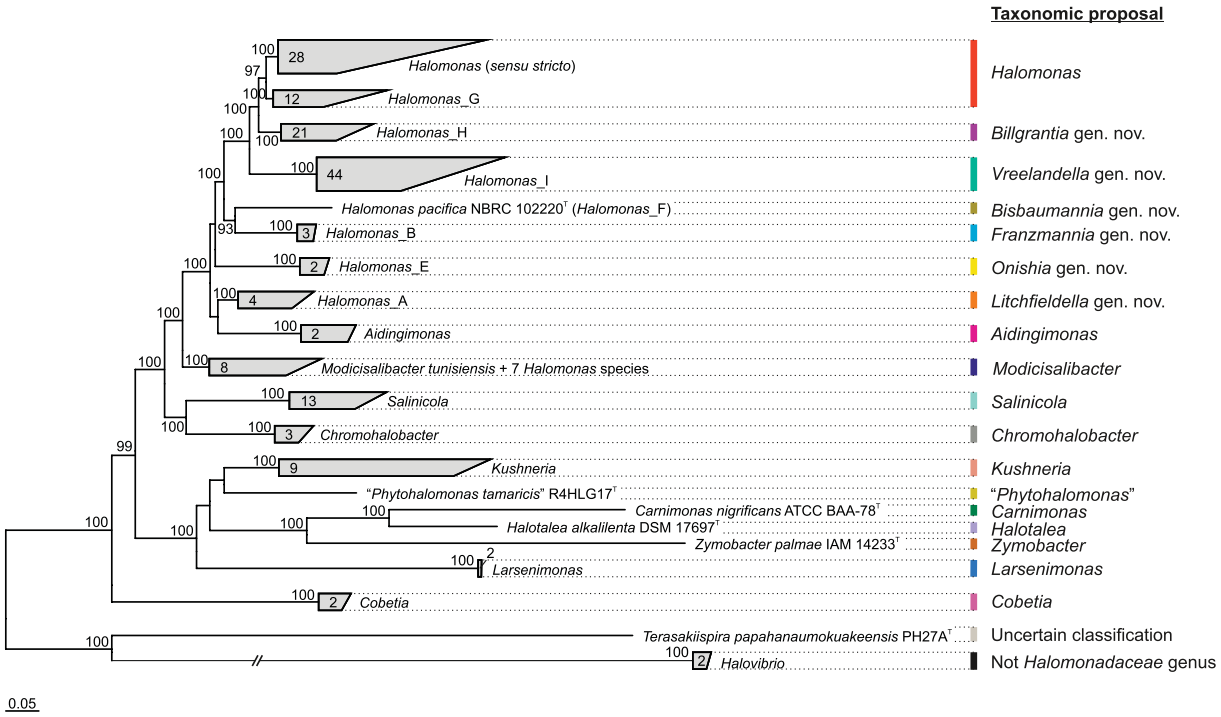


FIGURE 1
Clade collapsed maximum-likelihood phylogenomic tree based on the concatenation of the translated sequence of 189 single-copy genes shared by at least 90% of the members of the family *Halomonadaceae* under study ("core90" set). The genus *Pistricoccus* is missing because no representative genome sequences from type strains of this genus were recovered. The number of species comprised within each clade is displayed. Ultrafast bootstrap values ≥70% are shown above the branches. Bar, 0.05 changes per position.

IV = *Halomonas* (*sensu stricto*) + *Halomonas*_G + *Halomonas*_H + *Halomonas*_I. It should be noted that *proposal II* involved the formation of a phylogroup (*Halomonas* + *Halomonas*_G) supported "only" by a 97% bootstrap, which does not meet the desired constraints indicated above but is high enough to consider this option of grouping. To select the best-fitting scenario for this genus reclassification among the above proposals, we performed further comparative genomic analyses.

3.2. Assessment on the basis of AAI and cAAI values does not provide sound clues to split the genus *Halomonas*

To evaluate the proposed phylogroups under the four scenarios, we plotted the all-*vs*-all AAI and cAAI results clustered by the phylogroup while distinguishing between intra-genus and inter-genus values (Figure 2). Theoretically, a well-delimited genus or phylogroup should not display any overlap between intra-clade and inter-clade AAI/cAAI values. As shown in Figure 2, all the currently existing genera within the family *Halomonadaceae* harboring more than a single species except *Halomonas* (*sensu stricto*) (i.e., *Aidingimonas*, *Salinicola*, *Chromohalobacter*, *Kushneria*, *Larsenimonas*, and *Cobetia*), as well as the new phylogroups shared by the four proposals (*Halomonas*_B, *Halomonas*_F, *Halomonas*_E, *Halomonas*_A, and the enlarged *Modicisalibacter*) showed minimal or no overlap between the inter- and intra-cluster AAI and cAAI values, confirming their clear separation and exclusivity as different genera within this family.

By contrast, the intra-group AAI and cAAI values largely overlapped with their inter-group counterparts for the phylogroups *Halomonas*, *Halomonas*_G, *Halomonas*_H, and *Halomonas*_I (*proposal I*), indicating their heterogeneity as well as the lack of clear differentiation among them. No significant overlap reduction was observed after the consecutive merger of these phylogroups (*proposals II*, *III*, and *IV*), although *proposals II* and *III* led to slightly decreased overlap (Supplementary Figures 2–4). Based on the above observations, we believe that AAI and cAAI data may be useful for the delineation of most of the genera of the family *Halomonadaceae*, but they are not reliable for fine phylogroup demarcation within the genus *Halomonas*. The lowest intra-clade AAI and cAAI values for the current genera and *proposals I*, *II*, and *III* were 66.2 and 75.9%, respectively, while the cutoffs for *proposal IV* were 65.5 and 73.2%, respectively. These threshold values might be useful for future genus circumscription within the family *Halomonadaceae*.

3.3. Signature genes support the inclusion of *Halomonas*_G into the genus *Halomonas*

Since AAI and cAAI values did not allow us to determine the best of the four proposals to split the genus *Halomonas*, we analyzed whether the currently existing genera and suggested phylogroups can be defined by clade-specific signature genes. For that purpose, we examined the gene family presence/absence matrix extracted from the pangenome of the family *Halomonadaceae* to detect gene families

10.3389/fmicb.2023.1293707

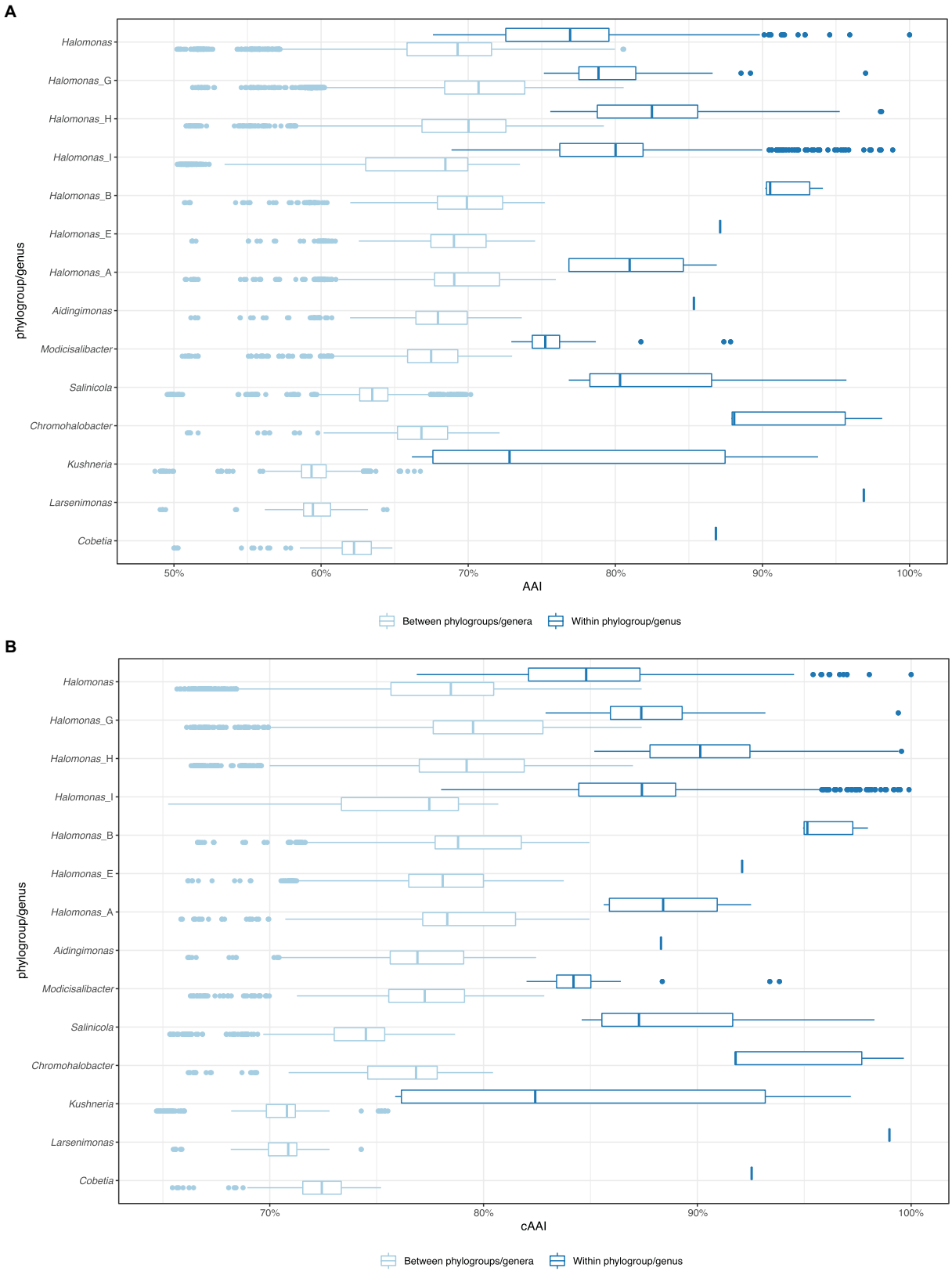


FIGURE 2
Box plot of the pairwise intra-clade and inter-clade AAI **(A)** and cAAI **(B)** values for the potential phylogroups/genera in the family *Halomonadaceae*, according to proposal I.

present in all genomes of a certain group but lacking in all other genomes. Signature genes may be the result of a shared evolutionary history within a phylogenetic clade, or they could be acquired by horizontal gene transfer reflecting common lifestyles, ecologies, and physiological properties (Zheng et al., 2020). Therefore, signature genes enable estimates of the evolutionary forces that shaped the cluster, and, thus, they might be of help to opt for one of the proposals over the others.

Except for *Halomonas*, all the current genera of the family *Halomonadaceae* according to the BMSAB classification containing more than a single species, together with *Halomonas*_B, *Halomonas*_E, *Halomonas*_A, the newly proposed phylogroups *Halomonas*_H and *Halomonas*_I, and the enlarged *Modicisalibacter*, were well supported by signature genes ranging from 5 to 548 (Figure 3 and Supplementary Figure 5). As can be expected, phylogroups/genera with a smaller number of species showed, generally, a larger number of signature genes; nevertheless, the impact of the clade size is limited, as can be evidenced from larger genera, such as *Salinicola* and *Kushneria*, displaying a relatively large number of signature genes. Considering that singleton genes (those present in a single genome) were omitted from the analysis, it was impossible to evaluate the signature genes harbored by phylogroups/genera containing only one species (i.e., *Halomonas*_F, "*Phytohalomonas*," *Carnimonas*, *Halotalea*, and *Zymobacter*). Interestingly, no signature genes were identified for phylogroups *Halomonas* (*sensu stricto*) and *Halomonas*_G, which *a priori* dismisses *proposal I*. However, when merging these phylogroups (*proposal II*), a signature gene was detected (Figure 3). Although it is a single signature gene, it is conserved across a large number of genomes (40), comprising the phylogroups *Halomonas* and *Halomonas*_G and, therefore, it likely reflects a common evolutionary history of both phylogroups providing an important hint into the delineation of the genus *Halomonas*. *Proposal III*, which considers the merging of phylogroups *Halomonas*, *Halomonas*_G, and *Halomonas*_H, was also endorsed by a single signature gene (Supplementary Figure 5); however, *proposal II* is favored over *proposal III* due to the fact that phylogroup *Halomonas*_H is very well supported by 21 signature genes and, thus, it is preferable to keep it as a separate cluster. The last proposal, *proposal IV*, consisting of the fusion of phylogroups *Halomonas*, *Halomonas*_G, *Halomonas*_H, and *Halomonas*_I, can also be rejected because the combined clade did not show any signature genes (Supplementary Figure 5) and, in addition, phylogroup *Halomonas*_I was supported by five signature genes.

3.4. GTDB-based phylogeny is consistent with *proposal II* and delineates the family *Halomonadaceae*

Although signature gene analyses endorsed *proposal II* as the most reliable one to split the genus *Halomonas*, this proposal involves the creation of a cluster with "only" a 97% branch bootstrap, as stated above. While this support value cannot be regarded as too low, we attempted to establish a long-term and trustworthy classification of the family for which only 100% supported phylogroups are preferred. This concern raises questions about the suitability of *proposal II*, making it necessary to provide some additional evidence to accept or reject it.

Accordingly, an alternative family-specific *Q* time-reversible matrix (*Q*_bac120; Supplementary Table 3) and phylogenomic tree (Figure 4) were calculated on the basis of the "bac120" marker protein dataset used to infer the bacterial GTDB taxonomy (Parks et al., 2018), comprising a total of 32,354 aligned columns after trimming. These ubiquitous single-copy proteins have been identified as being suitable for phylogenetic inference (Parks et al., 2017) and may yield better resolved trees than those obtained by using the almost entire set of core proteins ("core90" set). In fact, some of the "core90" non-curated protein-coding genes might have been laterally transferred or undergone homologous recombination (Stott and Bobay, 2020) or might lack congruent phylogenetic signals or sufficient homology to make comparisons valid and conclusive (Wu et al., 2013; Capella-Gutierrez et al., 2014). In addition, genomes from non-type strains and MAGs were also included in this complementary reconstruction to evaluate tree topology preservation.

The topology of the "bac120"-based tree (Figure 4) recovered the same phylogroups observed in the "core90"-based phylogeny (Figure 1). Although it is well-known that the selection of different phylogenetic markers can result in different topologies (Capella-Gutierrez et al., 2014; Ankenbrand and Keller, 2016; Tsai et al., 2019), in our analysis, the bias associated with marker choice did not affect the delineation of genera and phylogroups, demonstrating the stability of the obtained genome-based trees with independence of the protein set used for their inference. A good agreement between core-genome and "bac120" phylogenies has been previously demonstrated for the taxa of the *Pseudomonadaceae* (Lalucat et al., 2020), which is another family in the class *Gammaproteobacteria*. The inclusion of 82 additional GTDB representative genomes from non-type strains and MAGs of members of the family *Halomonadaceae*, as well as genomes from the closest genera of the order *Oceanospirillales*, did not alter the retrieved clades either (Figure 4). The most remarkable difference between our "core90" and "bac120" trees is the stronger support in the latter of the branch collapsing phylogroups *Halomonas* (*sensu stricto*) and *Halomonas*_G (corresponding to *proposal II*). Hence, this 100% bootstrap value recovered in the "bac120" tree after merging both phylogroups, together with the signature gene analysis, enables *proposal II* as the most appropriate to prune the genus *Halomonas* to the species comprising phylogroups *Halomonas* (*sensu stricto*) and *Halomonas*_G, while transferring the remaining species to six new genera and to the already described genus *Modicisalibacter*.

Concerning the taxonomic status of the family *Halomonadaceae*, the "bac120"-based phylogeny evidenced that all 12 genera currently affiliated to this family according to BMSAB, together with the non-validly published genus name "*Phytohalomonas*," formed a monophyletic group of microorganisms (Figure 4), whose shared feature, which was at the same time differential to the closest related genera, was their halophily or halotolerance. The genus *Halovibrio*, also considered a member of the family *Halomonadaceae* by LPSN, was intimately related to the genus *Halospina*, belonging to the family *Hahellaceae*; [10,11] whereas the genus *Terasakiispira*, included into the *Halomonadaceae* by both LPSN and GTDB resources, was rather a taxon within the family *Oceanospirillaceae* given its closest

10 https://lpsn.dsmz.de/family/hahellaceae

11 https://www.namesforlife.com/10.1601/tx.2487

10.3389/fmicb.2023.1293707

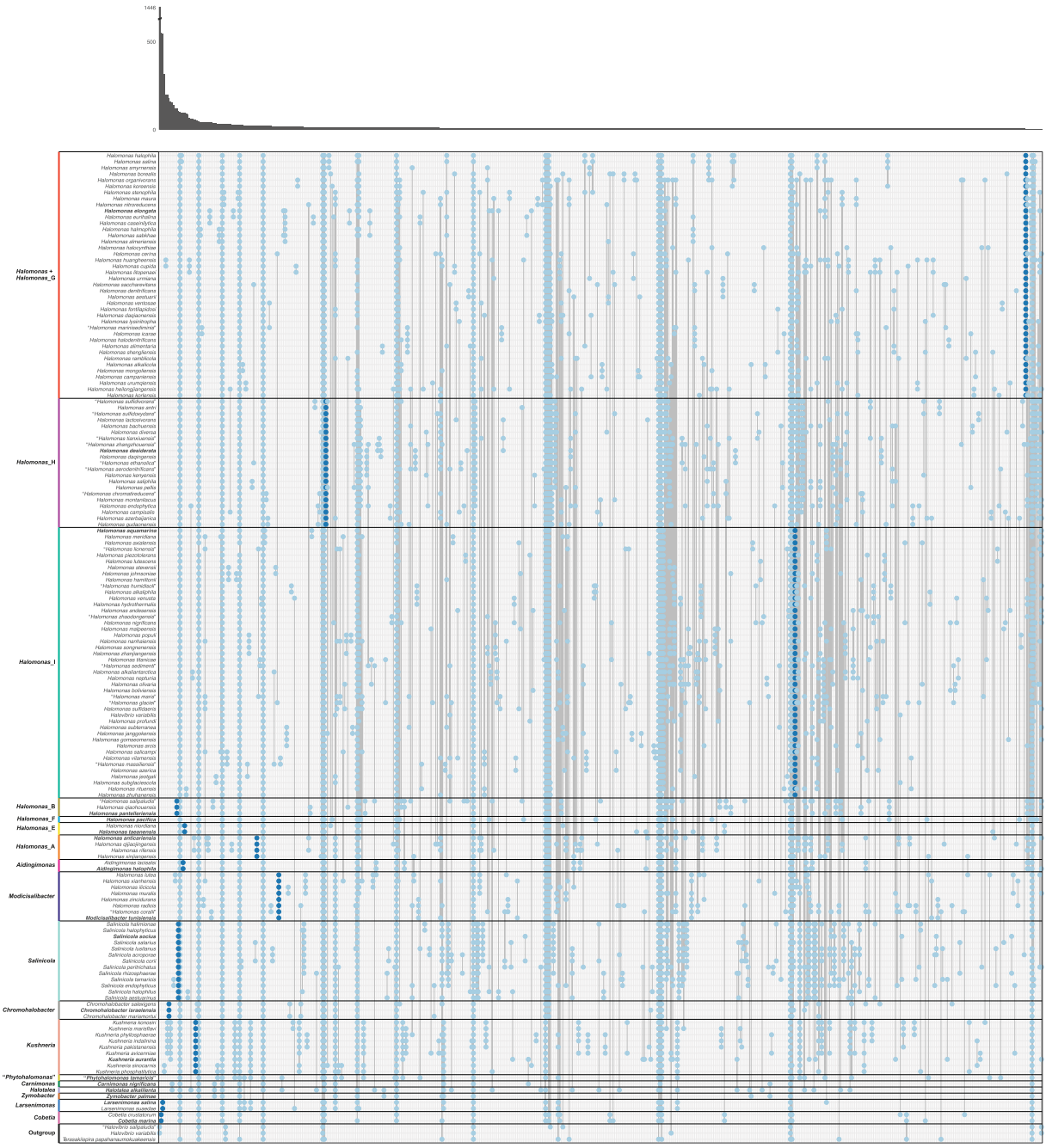


FIGURE 3
Gene family presence/absence patterns inferred from the pangenome of the family *Halomonadaceae* arranged according to *proposal II*. Each column represents a gene family pattern, where presence is indicated with a dot in the corresponding species. The absolute number of gene families that conform to each pattern is visualized in the marginal bar plot at the top. Separations between phylogroups/genera are indicated with horizontal black lines and the representative species of each phylogroup/genus is highlighted in bold. Genes that were present in all genomes of a clade and in none of the genomes outside of that clade, denoted as "signature genes," are displayed in dark blue; other genes are shown in light blue. Patterns of presence in a single species or all species are not shown. The species *Halovibrio salipaludis*, *Halovibrio variabilis*, and *Terasakiispira papahanaumokuakeensis* were not considered members of the *Halomonadaceae* and were only used as an "outgroup".

relationship to the genera *Marinospirillum*, *Oceanospirillum*, and *Pseudospirillum* (Figure 4). GTDB taxonomy (release 08-RS214) has suggested the incorporation of these three genera into the family *Halomonadaceae*, even if they harbor marine (slight) halophiles in contrast to the majority of moderately halophilic bacteria characteristic of the family *Halomonadaceae*. Therefore, it becomes clear that the genus *Halovibrio* should be kept apart from the family *Halomonadaceae*, while further investigation including genomic, phylogenomic, and phenotypic comparisons is required to delineate the taxonomic affiliation of the genera *Terasakiispira*, *Marinospirillum*, *Oceanospirillum*, and *Pseudospirillum* at the family level.

10.3389/fmicb.2023.1293707

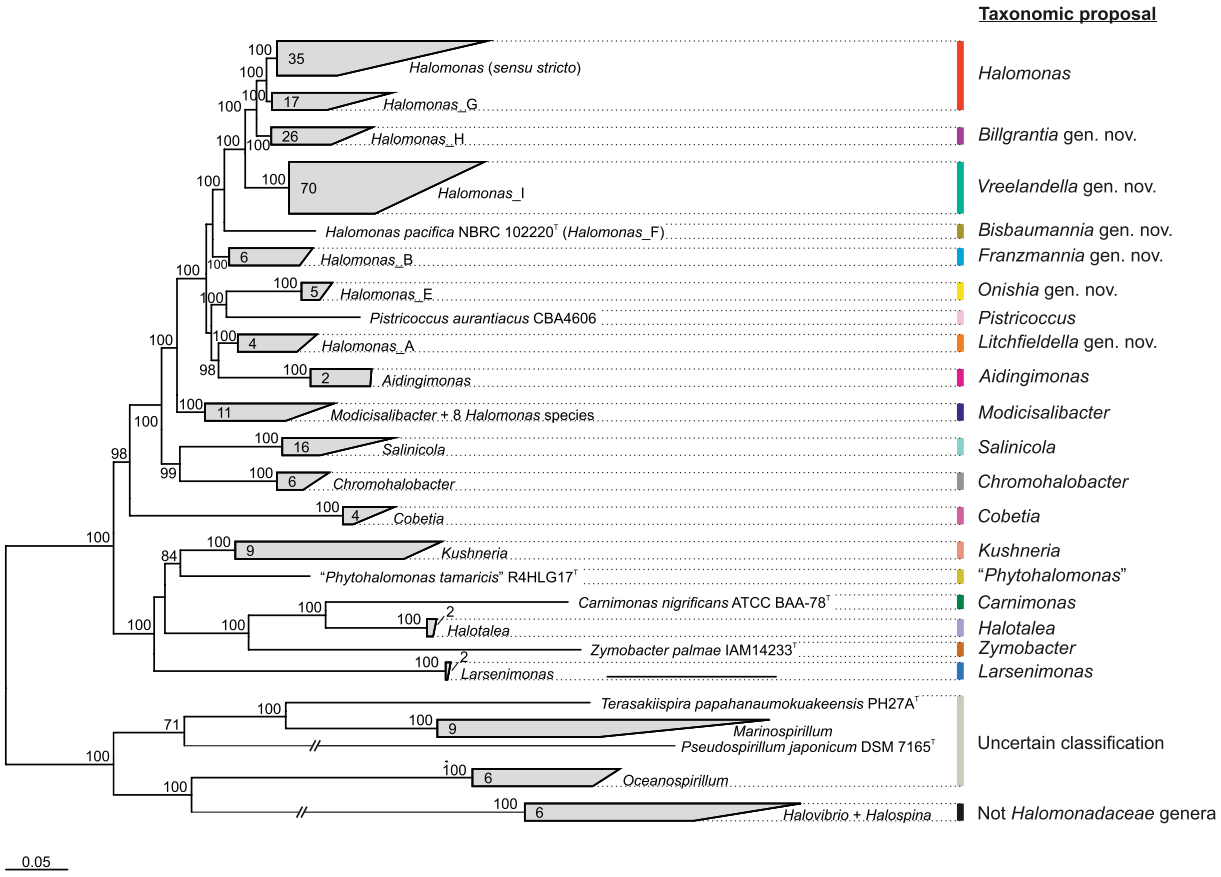


FIGURE 4
GTDB-derived maximum-likelihood phylogenomic tree based on the concatenation of 120 single-copy bacterial proteins ("bac120" set) showing the relationships among members of the family *Halomonadaceae* and related taxa. The type strain of the species *Pistricoccus aurantiacus* is missing because no genome sequences from this strain could be recovered. The number of species in each cluster is displayed inside wedges. Ultrafast bootstrap values ≥70% are shown above the branches. Bar, 0.05 changes per amino acid position.

3.5. Heterotypic synonyms revealed by ANI and dDDH values

Phylogenomic inference using both the "core90" and the "bac120" protein sets revealed the presence of some very closely related species pairs and even one set of three among the studied taxa that might be considered cases of heterotypic synonymy. To verify those hypotheses, ANI (Figure 5) and dDDH relatedness indexes were estimated between the genomes in question. It is widely accepted that cutoff values for species delineation based on these genomic indexes are 95–96% for ANI (Goris et al., 2007; Richter and Rossello-Mora, 2009; Chun and Rainey, 2014) and 70% for dDDH (Auch et al., 2010). Therefore, two or more strains can be considered to belong to the same species if both ANI values are ≥96% and dDDH values are ≥70%. According to this criterion, our results confirmed the existence of the following sets of heterotypic synonyms (ANI and dDDH values in parenthesis, respectively): *Chromohalobacter israelensis* – *Chromohalobacter salexigens* (98.1 and 83.6%); *Halomonas alkaliphila* – "*Halomonas humidisoli*" (97.7 and 80.5%); *Halomonas antri* – "*Halomonas sulfidivorans*" (97.7 and 80.2%); *Halomonas aquamarina* – *Halomonas meridiana* – *Halomonas axialensis* (96.5–97.5% and 71.2–79.1%); *Halomonas halophila* – *Halomonas salina* (100 and 100%); *Halomonas icarae* – "*Halomonas marinisediminis*" (97.4 and 79.2%); *Halomonas neptunia* – *Halomonas alkaliantarctica* (98.9 and 91.1%); *Halomonas hamiltonii* – *Halomonas johnsoniae* (97.3 and 76.8%); and *Halomonas venusta* – *Halomonas hydrothermalis* (96.8 and 73.0%).

4. Taxonomic conclusions

The comparative genomic analyses (phylogenomics, OGRI, and signature genes) conducted among the type strains of the species of the family *Halomonadaceae* allow us to propose several taxonomic rearrangements within this family:

i. The genus *Halomonas* comprises the following species: *H. elongata* (type species), *H. aestuarii*, *H. alimentaria*, *H. alkalicola*, *H. almeriensis*, *H. borealis*, *H. campaniensis*, *H. caseinilytica*, *H. cerina*, *H. cupida*, *H. daqiaonensis*, *H. denitrificans*, *H. eurihalina*, *H. fontilapidosi*, *H. halmophila*, *H. halocynthiae*, *H. halodenitrificans*, *H. halophila* (synonym *H. salina*), *H. heilongjiangensis*, *H. huangheensis*, *H. icarae* (synonym "*H. marinisediminis*"), *H. koreensis*, *H. korlensis*, *H. litopenaei*, *H. lysinitropha*, *H. maura*, *H. mongoliensis*,

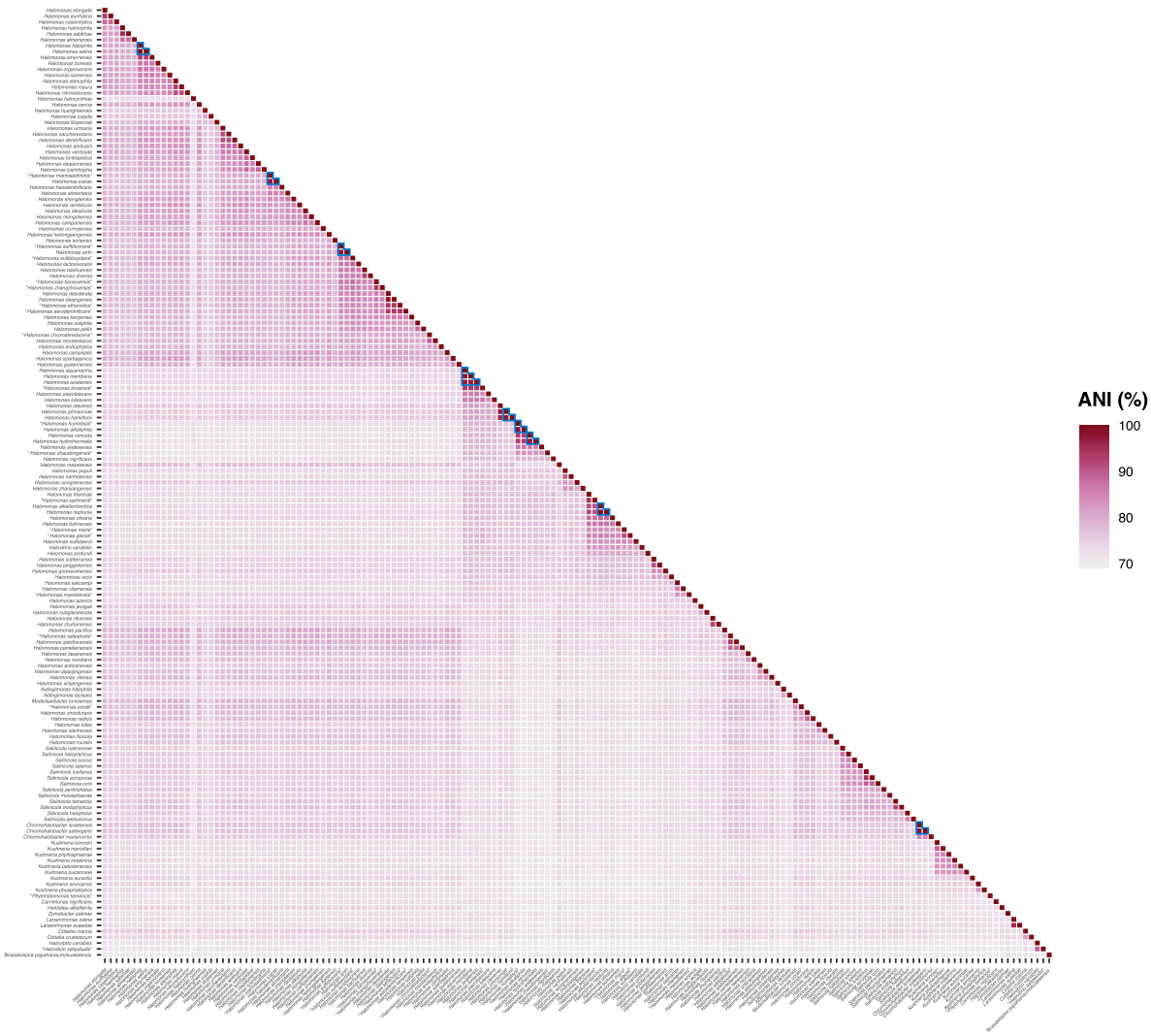


FIGURE 5
ANI values (%) from pairwise whole-genome comparisons between the type strains of members of the family *Halomonadaceae*. The ANI percentages are expressed as a heatmap in purple color. Values supporting the proposed synonym species names are surrounded by a blue triangle.

H. nitroreducens, *H. organivorans*, *H. ramblicola*, *H. sabkhae*, *H. saccharevitans*, *H. shengliensis*, *H. smyrnensis*, *H. stenophila*, *H. urmiana*, *H. urumqiensis*, and *H. ventosae*.

ii. The following species are transferred to the genus *Modicisalibacter*: "*Halomonas coralii*," *Halomonas ilicicola*, *Halomonas lutea*, *Halomonas muralis*, *Halomonas radicis*, *Halomonas xianhensis*, and *Halomonas zincidurans*, as "*Modicisalibacter coralii*," *Modicisalibacter ilicicola*, *Modicisalibacter luteus*, *Modicisalibacter muralis*, *Modicisalibacter radicis*, *Modicisalibacter xianhensis*, and *Modicisalibacter zincidurans*, respectively. Therefore, it has been necessary to extend the description of the genus *Modicisalibacter*, as indicated below.

iii. The remaining analyzed species of the genus *Halomonas* are reclassified into new genera belonging to the family *Halomonadaceae* as follows, where the type species of each genus is highlighted in bold, according to the rule of priority of publication:

- Genus *Vreelandella* (corresponding to the phylogroup *Halomonas*_I), including the species *H. alkaliphila* (synonym "*H. humidisoli*"), *H. andesensis*, ***H. aquamarina*** (synonyms *H. meridiana* and *H. axialensis*), *H. arcis*, *H. azerica*, *H. boliviensis*, "*H. glaciei*," *H. gomseomensis*, *H. hamiltonii* (synonym *H. johnsoniae*), *H. janggokensis*, *H. jeotgali*, "*H. lionensis*," *H. lutescens*, *H. malpeensis*, "*H. maris*," "*H. massiliensis*," *H. nanhaiensis*, *H. neptunia* (synonym *H. alkaliantarctica*), *H. nigrificans*, *H. olivaria*, *H. piezotolerans*, *H. populi*, *H. profundi*, *H. rituensis*, *H. salicampi*, "*H. sedimenti*," *H. songnenensis*, *H. stevensii*, *H. subglaciescola*, *H. subterranea*, *H. sulfidaeris*, *H. titanicae*, *H. utahensis*, *H. venusta* (synonym *H. hydrothermalis*), *H. vilamensis*, *H. zhanjiangensis*, "*H. zhaodongensis*," and *H. zhuhanensis*.
- Genus *Bisbaumannia* (corresponding to the phylogroup *Halomonas*_F), including the single species ***H. pacifica***.
- Genus *Billgrantia* (corresponding to the phylogroup *Halomonas*_H), including the species "*H. aerodenitrificans*,"

H. antri (synonym "*H. sulfidivorans*"), *H. azerbaijanica*, *H. bachuensis*, *H. campisalis*, "*H. chromatireducens*," ***H. desiderata*** (synonym *H. daqingensis*), *H. diversa*, *H. endophytica*, "*H. ethanolica*," *H. gudaonensis*, *H. kenyensis*, *H. lactosivorans*, "*H. montanilacus*," *H. pellis*, *H. saliphila*, "*H. sulfidoxydans*," "*H. tianxiuensis*," and "*H. zhangzhouensis*."

- Genus *Franzmannia* (corresponding to the phylogroup *Halomonas*_B), including the species ***H. pantelleriensis***, *H. qiaohouensis*, and "*H. salipaludis*."
- Genus *Litchfieldella* (corresponding to the phylogroup *Halomonas*_A), including the species ***H. anticariensis***, *H. qijiaojingensis*, *H. rifensis*, and *H. xinjiangensis*.
- Genus *Onishia* (corresponding to the phylogroup *Halomonas*_E), including the species *H. niordiana* and ***H. taeanensis***.

iv. The genus *Halovibrio* should be excluded from the family *Halomonadaceae*. According to GTBD, it should be transferred to the family *Oleiphilaceae*.

v. The following species sets are considered heterotypic synonyms (the species name holding priority in the application of the International Code of Nomenclature of Prokaryotes (Oren et al., 2023) is listed in first place): *Chromohalobacter israelensis* – *Chromohalobacter salexigens*; *Halomonas alkaliphila* – "*Halomonas humidisoli*"; *Halomonas antri* – "*Halomonas sufidivorans*"; *Halomonas aquamarina* – *Halomonas meridiana* – *Halomonas axialensis*; *Halomonas halophila* – *Halomonas salina*; *Halomonas hamiltonii* – *Halomonas johnsoniae*; *Halomonas icarae* – "*Halomonas marinisediminis*"; *Halomonas neptunia* – *Halomonas alkaliantarctica*; and *Halomonas venusta* - *Halomonas hydrothermalis*.

vi. The species whose genome sequence was not publicly available or could not be sequenced when this study was accomplished, together with the new species described after the finalization of this research, should be further studied to determine their correct placement according to the taxonomy proposed here or if they may constitute new genera within the family *Halomonadaceae*.

Moreover, we strongly recommend the use of the clade-specific amino acid substitution matrices *Q*_core90 and *Q*_bac120, empirically calculated for the family *Halomonadaceae*, for future phylogenomic studies within this family instead of the pre-computed *Q* matrices generally selected by model selection software.

Description of *Vreelandella* gen. nov.

Vreelandella (Vree.land.el'la N.L. fem. dim. n. *Vreelandella*, named after Russell Vreeland, American scientist who described the genus *Halomonas* and studied halophilic microorganisms for over 40 years).

Cells are Gram-staining-negative rods, 0.3–1.9 × 0.5–6.0 μm in size, aerobic or facultatively anaerobic, and mostly motile. Endospores are not formed. Catalase-positive and oxidase-variable. Colonies are translucent, beige, black, cream, glistening-colored, ochre, orange, pale-pink, reddish-brown, white, or yellow pigmented. Slightly to moderately halophilic, growing at 0–27% (w/v) NaCl, with optimal growth at 0–15% (w/v) NaCl. Mesophilic or psychrotolerant, thriving at −5–60°C, showing optimal growth at 20–37°C. Alkaliphilic or alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.0–10.0. Chemo-organotrophic. Nitrate reduction is variable. The major respiratory quinones are Q9, Q8, and Q6. The major fatty acids are $C_{16:0}$, $C_{16:1}\omega6c$/$C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{18:1}\omega6c$/$C_{18:1}\omega7c$, $C_{19:0}$ cyclo $\omega8c$, $C_{12:0}$ 3-OH, $C_{17:0}$ cyclo, $C_{10:0}$, iso-$C_{16:1}\omega7c$ 2-OH, and $C_{18:0}$. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content ranges between 52.1 and 63.8 mol%.

The genus *Vreelandella* belongs to the family *Halomonadaceae*. The type species is *Vreelandella aquamarina*.

Description of *Vreelandella aquamarina* comb. nov.

Vreelandella aquamarina (a.qua.ma.ri'na. L. fem. n. *aqua*, water; L. masc. adj. *marinus*, of the sea, marine; N.L. fem. adj. *aquamarina*, pertaining to seawater).

Basonym: "*Achromobacter aquamarinus*" ZoBeAll and Upham 1944.

Homotypic synonyms: *Halomonas aquamarina* (ZoBell and Upham 1944) Dobson and Franzmann, 1996; *Deleya aquamarina* (ZoBell and Upham 1944) Akagawa and Yamasato 1989; *Alcaligenes aquamarinus* (ZoBell and Upham 1944) Hendrie *et al.* 1974 (Approved Lists 1980).

Cells are Gram-stain-negative rods, 0.6–1.9 × 1.0–4.5 μm in size, with rounded ends, occurring singly or in doublets, facultatively anaerobic, and motile by means of peritrichous or lateral flagella. Endospores are not formed. Catalase and oxidase are positive. Colonies are convex, smooth, circular, entire, whitish or cream-colored, and 2 mm in diameter after 2 days of incubation on Marine Agar 2,216 at 20°C. Slightly to moderately halophilic, growing at 0–30% (w/v) NaCl, with optimal growth at 0.5–10% (w/v) NaCl. Psychrotolerant, thriving at −1–47°C, showing optimal growth at 20–40°C. Alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.0–8.0. Chemo-organotrophic. Negative for Simmons' citrate utilization, indole production, methyl red and Voges–Proskauer tests, lysine decarboxylase, and β-galactosidase. Aesculin, casein, gelatin, and DNA are not hydrolyzed. Variable for nitrate reduction to nitrite, H_2S production, urease, hydrolysis of starch and Tween 80, ornithine decarboxylase, phosphatase, and phenylalanine deaminase. Produces acid but no gas from glucose. Does not ferment glycerol or xylose. Utilizes L-arabinose, D-fructose, D-glucose, maltose, sucrose, and fumarate as sole carbon and energy sources. The major respiratory quinone is Q9. The major fatty acids are $C_{16:0}$, $C_{18:1}\omega7c$, $C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content is 56.7–57.0 mol%.

The type strain is 558^T = ATCC 14400^T = BCRC 12878^T = CCUG 16157^T = CECT 5000^T = CGMCC 1.2324^T = CIP 105454^T = DSM 30161^T = IAM 12550^T = KCTC 22193^T = LMD 73.17^T = LMG 2853^T = NCIMB 557^T. The genome size of the type strain is 3.50 Mbp, and its DNA G + C content is 56.7 mol%. Isolated from seawater (Pacific Ocean).

Type strain genome sequence accession number: GCA_900110265.1.

Type strain 16S rRNA gene sequence accession number: AJ306888.

Halomonas meridiana and *Halomonas axialensis* should be regarded as heterotypic synonyms of *Vreelandella aquamarina*.

Description of *Vreelandella venusta* comb. nov.

Vreelandella venusta (ve.nus'ta. L. fem. adj. *venusta*, lovely, beautiful).

Basonym: *Alcaligenes venustus* Baumann *et al.* 1972 (Approved Lists 1980).

Homotypic synonyms: *Halomonas venusta* (Baumann *et al.* 1972) Dobson and Franzmann 1996; *Deleya venusta* (Baumann *et al.* 1972) Baumann *et al.* 1983.

Cells are Gram-stain-negative rods, 1.5 × 2.0–3.0 μm in size, occurring singly or in doublets, aerobic or facultatively anaerobic, and motile by means of peritrichous flagella. Accumulate β-polyhydroxybutyrate as an intracellular reserve product. Endospores are not formed. Catalase and oxidase are positive. Colonies are round, smooth, and yellow or cream-colored. Slightly to moderately halophilic, growing at 0–22% (w/v) NaCl, with optimal growth at 0.5–7% (w/v) NaCl. Psychrotolerant, thriving at 2–40°C, showing optimal growth at 30°C. Alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.0–8.0. Chemo-organotrophic. Positive for nitrate reduction to nitrite. Negative for indole production and methyl red and Voges–Proskauer tests. DNA is hydrolyzed, but alginate, casein, chitin, gelatin, starch, and Tween 80 are not. Variable for H_2S production, lysine decarboxylase, ornithine decarboxylase, and phenylalanine deaminase. Acid production from D-glucose is variable, but negative from L-arabinose, D-fructose, D-galactose, lactose, maltose, D-mannose, sucrose, D-trehalose, D-xylose, or glycerol. Utilizes D-glucose, glycerol, acetate, butyrate, citrate, fumarate, DL-malate, propionate, L-glutamate, L-lysine, and L-proline as sole carbon (nitrogen) and energy sources. The major respiratory quinone is Q9. The major fatty acids are $C_{16:0}$, $C_{18:1}\omega7c$, $C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content is 52.6–53.2 mol%.

The type strain is 86^T = ATCC 27125^T = CCUG 16063^T = CIP 103201^T = DSM 4743^T = JCM 20634^T = LMG 3445^T = NBRC 102221^T. The genome size of the type strain is 4.27 Mbp, and its DNA G + C content is 52.6 mol%. Isolated from seawater (off the coast of Oahu, Hawaii, United States).

Type strain genome sequence accession number: GCA_007989605.1.

Type strain 16S rRNA gene sequence accession number: AJ306894.

Halomonas hydrothermalis should be regarded as a heterotypic synonym of *Vreelandella venusta*.

Description of *Vreelandella subglaciescola* comb. nov.

Vreelandella subglaciescola (sub.gla.ci.es'co.la. L. prep. *sub*-, under, below; L. fem. n. *glacies*, ice; L. masc. n. suff. *-cola*, dweller; N.L. fem. n. *subglaciescola*, dwelling below the ice).

Basonym: *Halomonas subglaciescola* Franzmann *et al.* 1987.

The description is as given in the proposal of the basonym (Franzmann et al., 1987), with the following addition. The genome size of the type strain is 3.11 Mbp. The DNA G + C content is 60.8 mol%.

Isolated from saline water from Organic Lake, Antarctica.

The type strain is ACAM 12^T = ATCC 43668^T = CIP 104042^T = DSM 4683^T = IAM 14167^T = JCM 21045^T = LMG 8824^T = NBRC 14766^T = UQM 2926^T.

Type strain genome sequence accession number: GCA_900142895.1.

Type strain 16S rRNA gene sequence accession number: AJ306892.

Description of *Vreelandella neptunia* comb. nov.

Vreelandella neptunia (nep.tu'ni.a. L. fem. adj. *neptunia*, Neptunian, pertaining to Neptunus, the Roman god of the sea).

Basonym: *Halomonas neptunia* Kaye *et al.* 2004.

Cells are Gram-stain-negative rods, 0.5–1.0 × 1.0–3.0 μm in size, occurring singly or in doublets, aerobic or facultatively anaerobic, and motile by means of peritrichous flagella. Produces exopolysaccharides and accumulates glycine-betaine but not β-polyhydroxybutyrate. Endospores are not formed. Oxidase-positive, but catalase-variable. Colonies are round, smooth, and cream-colored. Slightly to moderately halophilic, growing at 0.5–27% (w/v) NaCl, with optimal growth at 2–10% (w/v) NaCl. Psychrotolerant, thriving at −1–37°C, showing optimal growth at 30°C. Alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.0–9.0. Chemo-organotrophic. Positive for nitrate reduction to nitrite, but nitrite is not reduced. Negative for Simmons' citrate utilization, H_2S production, methyl red and Voges–Proskauer tests, lysine decarboxylase, ornithine decarboxylase, and phenylalanine deaminase. Able to synthesize α-glucosidase. Casein, DNA, gelatin, starch, and Tween 80 are not hydrolyzed. Variable for indole production. Acid production from D-galactose and D-glucose is positive, but negative from L-arabinose, D-fructose, lactose, maltose, D-mannose, sucrose, D-trehalose, D-xylose, or glycerol. Utilizes D-cellobiose, D-fructose, D-galactose, D-glucose, maltose, D-ribose, sucrose, D-trehalose, glycerol, and acetate as sole carbon and energy sources. The only respiratory quinone is Q9. The major fatty acids are $C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega7c$, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. The major polar lipids are phosphatidylglycerol, diphosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content is 54.8–55.0 mol%.

The type strain is Eplume1^T = ATCC BAA-805^T = CCM 7107^T = CECT 5815^T = DSM 15720^T. The genome size of the type strain is 4.93 Mbp, and its DNA G + C content is 55.0 mol%. Isolated from a deep-sea hydrothermal plume (NE Pacific Ocean, Juan de Fuca Ridge).

Type strain genome sequence accession numbers: GCA_030409295.1 and GCA_019903445.1.

Type strain 16S rRNA gene sequence accession number: AF212202.

Halomonas alkaliantarctica should be regarded as a heterotypic synonym of *Vreelandella neptunia*.

Description of *Vreelandella sulfidaeris* comb. nov.

Vreelandella sulfidaeris (sul.fid.ae'ris. N.L. neut. n. *sulfidum*, sulfide; L. neut. n. *aes*, ore; N.L. gen. n. *sulfidaeris*, from sulfide ore).

Basonym: *Halomonas sulfidaeris* Kaye *et al.* 2004.

The description is as given in the proposal of the basonym (Kaye et al., 2004), with the following addition. The genome size of the type strain is 4.48 Mbp. The DNA G + C content is 53.7 mol%.

Isolated from a deep-sea sulfide rock (NE Pacific Ocean, Juan de Fuca Ridge).

The type strain is Esulfide1^T = ATCC BAA-803^T = CCM 7108^T = CECT 5817^T = DSM 15722^T.

Type strain genome sequence accession number: GCA_007182875.1.

Type strain 16S rRNA gene sequence accession number: AF212204.

Description of *Vreelandella boliviensis* comb. nov.

Vreelandella boliviensis (bo.li.vi.en'sis. N.L. fem. adj. *boliviensis*, from Bolivia, relating to the country where the bacteria were isolated).

Basonym: *Halomonas boliviensis* Quillaguamán *et al.* 2004.

The description is as given in the proposal of the basonym (Quillaguamán et al., 2004), with the following addition. The genome size of the type strain is 4.21 Mbp. The DNA G + C content is 54.7 mol%.

Isolated from the soil around a Bolivian hypersaline lake.

The type strain is LC1^T = ATCC BAA-759^T = DSM 15516^T.

Type strain genome sequence accession numbers: GCA_000236035.1 and GCA_002265845.1.

Type strain 16S rRNA gene sequence accession number: AY245449.

Description of *Vreelandella utahensis* comb. nov.

Vreelandella utahensis (u.ta.hen'sis. N.L. fem. adj. *utahensis*, referring to Utah).

Basonym: *Halomonas utahensis* Sorokin and Tindall, 2006.

The description is as given in the proposal of the basonym (Sorokin and Tindall, 2006), with the following addition. The genome size of the type strain is 3.73 Mbp. The DNA G + C content is 55.8 mol%.

Isolated from surface water from the North Arm of Great Salt Lake (United States).

The type strain is isolate III^T = ATCC 49240^T = CECT 5286^T = CIP 105504^T = DSM 3051^T = IAM 14440^T = JCM 21223^T = NBRC 102410^T.

Type strain genome sequence accession number: GCA_007991175.1.

Type strain 16S rRNA gene sequence accession number: AJ306893.

Description of *Vreelandella gomseomensis* comb. nov.

Vreelandella gomseomensis (gom.se.om.en'sis. N.L. fem. adj. *gomseomensis*, referring to Gomseom in Anmyeondo, from where the first strains were isolated).

Basonym: *Halomonas gomseomensis* Kim *et al.* 2007.

The description is as given in the proposal of the basonym (Kim et al., 2007), with the following addition. The genome size of the type strain is 3.73 Mbp. The DNA G + C content is 59.8 mol%.

Isolated from saline water of the Gomseom solar saltern in Anmyeondo (Korea).

The type strain is M12^T = CIP 109897^T = DSM 18042^T = KCTC 12662^T.

Type strain genome sequence accession number: GCA_031451645.1.

Type strain 16S rRNA gene sequence accession number: AM229314.

Description of *Vreelandella janggokensis* comb. nov.

Vreelandella janggokensis (jang.gok.en'sis. N.L. fem. adj. *janggokensis*, referring to Janggok in Anmyeondo, from where the first strains were isolated).

Basonym: *Halomonas janggokensis* Kim *et al.* 2007.

The description is as given in the proposal of the basonym (Kim et al., 2007), with the following addition. The genome size of the type strain is 3.87 Mbp. The DNA G + C content is 57.3 mol%.

Isolated from saline water of the Janggok solar saltern in Anmyeondo (Korea).

The type strain is M24^T = CIP 109896^T = DSM 18043^T = KCTC 12663^T.

Type strain genome sequence accession number: GCA_031451615.1.

Type strain 16S rRNA gene sequence accession number: AM229315.

Description of *Vreelandella arcis* comb. nov.

Vreelandella arcis (ar'cis. L. gen. n. *arcis*, of a height, summit, or peak, referring to the isolation of the organism from a salt lake on the Qinghai-Tibet Plateau).

Basonym: *Halomonas arcis* Xu *et al.* 2007.

The description is as given in the proposal of the basonym (Xu et al., 2007), with the following addition. The genome size of the type strain is 4.14 Mbp. The DNA G + C content is 55.9 mol%.

Isolated from the water of a salt lake located in Altun Mountain on the Qinghai-Tibet Plateau (China).

The type strain is AJ282^T = CGMCC 1.6494^T = DSM 23549^T = JCM 14607^T = LMG 23978^T.

Type strain genome sequence accession number: GCA_900103865.1.

Type strain 16S rRNA gene sequence accession number: EF144147.

Description of *Vreelandella subterranea* comb. nov.

Vreelandella subterranea (sub.ter.ra'ne.a. L. fem. adj. *subterranea*, underground, subterranean, referring to the isolation of the organism from the subterranean brines).

Basonym: *Halomonas subterranea* Xu *et al.* 2007.

The description is as given in the proposal of the basonym (Xu et al., 2007), with the following addition. The genome size of the type strain is 3.73 Mbp. The DNA G + C content is 58.0 mol%.

Isolated from water of a subterranean saline well at Si-Chuan Basin (China).

The type strain is ZG16^T = CIP 109673^T = CGMCC 1.6495^T = DSM 23550^T = JCM 14608^T = LMG 23977^T.

Type strain genome sequence accession number: GCA_900111305.1.

Type strain 16S rRNA gene sequence accession number: EF144148.

Description of *Vreelandella alkaliphila* comb. nov.

Vreelandella alkaliphila (al.ka.li'phi.la. N.L. neut. n. *alkali*, the ashes of saltwort [al-qaliy]; N.L. masc. adj. *philus* [from Gr. masc. adj. *philos*], friend, loving; N.L. fem. adj. *alkaliphila*, loving alkaline media).

Basonym: *Halomonas alkaliphila* Romano *et al.* 2007.

Cells are Gram-stain-negative rods, 0.3–0.7 × 0.5–2.6 μm in size, aerobic, and motile or non-motile. Produces exopolysaccharides and accumulates β-polyhydroxybutyrate, ectoine, and glycine-betaine. Catalase and oxidase are positive. Colonies are circular, wet, convex, and beige or cream-colored. Moderately halophilic, growing at 0–25% (w/v) NaCl, with optimal growth at 4–10% (w/v) NaCl. Mesophilic, thriving at 5–50°C, showing optimal growth at 30–37°C. Alkaliphilic or alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.5–9.0. Chemo-organotrophic. Positive for nitrate reduction to nitrite and tyrosine decomposition. Negative for H_2S production. Hippurate and pectin are hydrolyzed, but casein, gelatin, phenylalanine, starch, Tween 80, and tyrosine are not. Variable for urease production. Utilizes D-glucose, D-galactose, D-fructose, sucrose, D-maltose, D-cellobiose, D-trehalose, glycerol, and acetate as sole carbon and energy sources. The major respiratory quinones are Q9, Q8, and Q6. The major fatty acids are $C_{18:1}\omega6c/C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega6c/C_{16:1}\omega7c$, and $C_{18:0}$. The major polar lipids are phosphatidylethanolamine, diphosphatidylglycerol, and phosphatidylglycerol.

The DNA G + C content is 52.5–52.6 mol%.

The type strain is 18bAG^T = ATCC BAA-953^T = DSM 16354^T. The genome size of the type strain is 4.10 Mbp, and its DNA G + C content is 52.6 mol%. Isolated from a salt pool in Campania (Italy).

Type strain genome sequence accession number: GCA_016107625.1.

Type strain 16S rRNA gene sequence accession number: AJ640133.

Description of *Vreelandella zhanjiangensis* comb. nov.

Vreelandella zhanjiangensis (zhan.ji.ang.en'sis. N.L. fem. adj. *zhanjiangensis*, pertaining to Zhanjiang, a city in China near where the sample was collected).

Basonym: *Halomonas zhanjiangensis* Chen *et al.*, 2009.

The description is as given in the proposal of the basonym (Chen et al., 2009), with the following addition. The genome size of the type strain is 4.06 Mbp. The DNA G + C content is 54.5 mol%.

Isolated from a sea urchin, *Hemicentrotus pulcherrimus*, South China Sea, tidal flat of Naozhou Island near Zhanjiang (China).

The type strain is JSM 078169^T = CCTCC AB 208031^T = DSM 21076^T = KCTC 22279^T.

Type strain genome sequence accession number: GCA_000377665.1.

Type strain 16S rRNA gene sequence accession number: FJ429198.

Description of *Vreelandella stevensii* comb. nov.

Vreelandella stevensii (ste.ven'si.i. N.L. gen. n. *stevensii*, of Stevens, named after Dr. David A. Stevens, a physician/epidemiologist who isolated and characterized the first strains).

Basonym: *Halomonas stevensii* Kim *et al.* 2010.

The description is as given in the proposal of the basonym (Kim K. K. et al., 2010), with the following addition. The genome size of the type strain is 3.69 Mbp. The DNA G + C content is 60.3 mol%.

Isolated from the blood of a renal care patient at California, San Jose, Santa Clara Valley Medical Center (United States).

The type strain is S18214^T = DSM 21198^T = KCTC 22148^T.

Type strain genome sequence accession number: GCA_000275725.1.

Type strain 16S rRNA gene sequence accession number: AM941388.

Description of *Vreelandella hamiltonii* comb. nov.

Vreelandella hamiltonii (ha.mil.to'ni.i. N.L. gen. n. *hamiltonii*, of Hamilton, named after Dr. John R. Hamilton, a microbiologist who isolated and characterized the first strains).

Basonym: *Halomonas hamiltonii* Kim *et al.* 2010.

The description is as given in the proposal of the basonym (Kim K. K. et al., 2010), with the following addition. Cells are 0.7–1.0 × 1.5–4.0 μm in size and motile with lateral or lateral/polar flagella. Optimal growth occurs at sea-salt concentrations of 0–7.5%, w/v. Grown on cetrimide agar. Nitrate reduction to nitrite, the Voges–Proskauer test, and urease are variable. Utilization of adipate, D-galactose, malonate, L-lysine, DL-isoleucine, and L-valine as sole sources of carbon (nitrogen) and energy is variable. Acid production from L-arabinose, D-fucose, D-galactose, methyl α-D-glucoside, glycerol, D-mannitol, and melezitose is variable. Susceptibility to neomycin, penicillin G, and chloramphenicol is strain-dependent. The DNA G + C content is 60.1 mol%.

The type strain is W1025^T = DSM 21196^T = KCTC 22154^T. The genome size of the type strain is 3.93 Mbp, and its DNA G + C content is 60.1 mol%. Isolated from the blood of a dialysis machine drain at California, San Jose, Santa Clara Valley Medical Center (USA).

Type strain genome sequence accession number: GCA_014651775.1.

Type strain 16S rRNA gene sequence accession number: AM941396.

Halomonas johnsoniae should be regarded as a heterotypic synonym of *Vreelandella hamiltonii*.

Description of *Vreelandella andesensis* comb. nov.

Vreelandella andesensis (an.de.sen'sis. N.L. fem. adj. *andesensis*, pertaining to the Andes).

Basonym: *Halomonas andesensis* Guzmán *et al.* 2010.

The description is as given in the proposal of the basonym (Guzmán et al., 2010), with the following addition. The genome size of the type strain is 3.91 Mbp. The DNA G + C content is 52.1 mol%.

Isolated from water from saline lake Laguna Colorada (22° 12' S 67° 49' W), 4,300 m above sea level (Bolivia).

The type strain is LC6^T = CCUG 54844^T = DSM 19434^T = LMG 24243^T.

Type strain genome sequence accession number: GCA_003989795.1.

Type strain 16S rRNA gene sequence accession number: EF622233.

Description of *Vreelandella titanicae* comb. nov.

Vreelandella titanicae (ti.tan'ic.ae. N.L. fem. n. *titanica*, the ship Titanic; N.L. gen. fem. n. *titanicae*, of or from the ship Titanic).

Basonym: *Halomonas titanicae* Mann *et al.* 2010.

The description is as given in the proposal of the basonym (Sánchez-Porro et al., 2010), with the following addition. The genome size of the type strain is 5.34 Mbp. The DNA G + C content is 54.6 mol%.

Isolated from the rusticles of the RMS Titanic wreck.

The type strain is BH1^T = ATCC BAA-1257^T = CECT 7585^T = DSM 22872^T = JCM 16411^T = LMG 25388^T.

Type strain genome sequence accession number: GCA_000336575.1.

Type strain 16S rRNA gene sequence accession number: FN433898.

Description of *Vreelandella vilamensis* comb. nov.

Vreelandella vilamensis (vi.la.men'sis. N.L. fem. adj. *vilamensis*, pertaining to Laguna Vilama, Jujuy, Argentina).

Basonym: *Halomonas vilamensis* Menes *et al.* 2011.

The description is as given in the proposal of the basonym (Menes et al., 2011), with the following addition. The genome size of the type strain is 3.47 Mbp. The DNA G + C content is 55.2 mol%.

Isolated from the sediment of hypersaline lake Laguna Vilama (22° 35' S 66° 55' W, 4,600 m above sea level) at Andean Puna desert, Jujuy (Argentina).

The type strain is SV325^T = DSM 21020^T = LMG 24332^T.

Type strain genome sequence accession number: GCA_031451755.1.

Type strain 16S rRNA gene sequence accession number: EU557315.

Description of *Vreelandella jeotgali* comb. nov.

Vreelandella jeotgali (je.ot.ga'li. N.L. gen. n. *jeotgali*, of jeotgal, a traditional Korean fermented seafood).

Basonym: *Halomonas jeotgali* Kim *et al.* 2011.

The description is as given in the proposal of the basonym (Kim M.-S. et al., 2010), with the following addition. The genome size of the type strain is 2.85 Mbp. The DNA G + C content is 62.9 mol%.

Isolated from jeotgal, a traditional Korean fermented seafood.

The type strain is Hwa^T = JCM 15645^T = KCTC 22487^T.

Type strain genome sequence accession number: GCA_000334215.1.

Type strain 16S rRNA gene sequence accession number: EU909458.

Description of *Vreelandella nanhaiensis* comb. nov.

Vreelandella nanhaiensis (nan.hai.en'sis. N.L. fem. adj. *nanhaiensis*, pertaining to Nanhai, a sea in South China where the sample was collected).

Basonym: *Halomonas nanhaiensis* Long *et al.* 2013.

The description is as given in the proposal of the basonym (Long et al., 2013), with the following addition. The genome size of the type strain is 4.03 Mbp. The DNA G + C content is 54.4 mol%.

Isolated from a sample of marine sediment at a depth of 310 m (74°52′35" S 163°53′03" E), South China Sea.

The type strain is YIM M 13059^T = CCTCC AB 2012911^T = DSM 25561^T = JCM 18142^T.

Type strain genome sequence accession number: GCA_003990185.1.

Type strain 16S rRNA gene sequence accession number: JX870002.

Description of *Vreelandella olivaria* comb. nov.

Vreelandella olivaria (o.li.va'ri.a. L. fem. adj. *olivaria*, of or belonging to olives, related to olive-processing effluents from where the type strain was isolated).

Basonym: *Halomonas olivaria* Amouric *et al.* 2014.

The description is as given in the proposal of the basonym (Amouric et al., 2014), with the following addition. The genome size of the type strain is 5.00 Mbp. The DNA G + C content is 55.3 mol%.

Isolated from salted olive-processing effluents from an evaporation pond (Morocco).

The type strain is TYRC17^T = CCUG 53850 B^T = DSM 19074^T.

Type strain genome sequence accession number: GCA_004295565.1.

Type strain 16S rRNA gene sequence accession number: DQ645593.

Description of *Vreelandella songnenensis* comb. nov.

Vreelandella songnenensis (song.nen.en'sis. N.L. fem. adj. *songnenensis*, pertaining to Songnen Plain, north-east China, where the type strain was isolated).

Basonym: *Halomonas songnenensis* Jiang *et al.* 2014.

The description is as given in the proposal of the basonym (Jiang et al., 2014), with the following addition. The genome size of the type strain is 3.69 Mbp. The DNA G + C content is 59.1 mol%.

Isolated from saline and alkaline soil in an oilfield (46° 36′ 05.36" N 124° 55′ 00.36″ E), Songnen Plain (China).

The type strain is NEAU-ST10-39^T = CGMCC 1.12152^T = DSM 25870^T.

Type strain genome sequence accession number: GCA_003002925.1.

Type strain 16S rRNA gene sequence accession number: JQ762289.

Description of *Vreelandella salicampi* comb. nov.

Vreelandella salicampi (sa.li.cam'pi. L. masc. n. *sal*, salt; L. masc. n. *campus*, field; N.L. gen. n. *salicampi*, of a salt field).

Basonym: *Halomonas salicampi* Lee *et al.* 2015.

The description is as given in the proposal of the basonym (Lee et al., 2015), with the following addition. The genome size of the type strain is 3.86 Mbp. The DNA G + C content is 56.2 mol%.

Isolated from a saltern soil at Gomso (Korea).

The type strain is BH103^T = KACC 17609^T = NBRC 109914^T = NCAIM B 02528^T.

Type strain genome sequence accession number: GCA_013415105.1.

Type strain 16S rRNA gene sequence accession number: KF963827.

Description of *Vreelandella lutescens* comb. nov.

Vreelandella lutescens (lu.tes'cens. L. part. adj. *lutescens*, becoming muddy, related to the muddy color of the mature colony).

Basonym: *Halomonas lutescens* Wang *et al.* 2016.

The description is as given in the proposal of the basonym (Wang et al., 2016), with the following addition. The genome size of the type strain is 3.70 Mbp. The DNA G + C content is 56.0 mol%.

Isolated from a sediment sample from Qinghai Lake (China).

The type strain is Q1U^T = CGMCC 1.15122^T = KCTC 42517^T.

Type strain genome sequence accession number: GCA_014640815.1.

Type strain 16S rRNA gene sequence accession number: KP259554.

Description of *Vreelandella nigrificans* comb. nov.

Vreelandella nigrificans (nig.rif'i.cans. L. part. adj. *nigrificans*, making black).

Basonym: *Halomonas nigrificans* Oguntoyinbo *et al.* 2018.

The description is as given in the proposal of the basonym (Oguntoyinbo et al., 2018), with the following addition. The genome size of the type strain is 4.93 Mbp. The DNA G + C content is 52.8 mol%.

Isolated from cheese (Germany).

The type strain is MBT G8648^T = DSM 105749^T = LMG 29097^T.

Type strain genome sequence accession number: GCA_002374315.1.

Type strain 16S rRNA gene sequence accession number: MG030686.

Description of *Vreelandella malpeensis* comb. nov.

Vreelandella malpeensis (mal.pe.en'sis. N.L. fem. adj. *malpeensis*, of or belonging to Malpe, a coastal town in Udupi City, Karnataka, India).

Basonym: *Halomonas malpeensis* Kämpfer *et al.* 2018.

The description is as given in the proposal of the basonym (Kämpfer et al., 2018), with the following addition. The genome size of the type strain is 3.61 Mbp. The DNA G + C content is 63.8 mol%.

Isolated from the rhizosphere of sand dune coastal plant, Coast of Malpe (India).

The type strain is YU-PRIM-29^T = CCM 8737^T = LMG 28855^T.

Type strain Genome sequence accession number: GCA_020622355.1.

Type strain 16S rRNA gene sequence accession number: JQ730736.

Description of *Vreelandella piezotolerans* comb. nov.

Vreelandella piezotolerans (pie.zo.to'le.rans. Gr. v. *piezô*, to press; L. pres. part. *tolerans*, tolerating; N.L. part. adj. *piezotolerans*, pressure-tolerating).

Basonym: *Halomonas piezotolerans* Yan *et al.* 2020.

The description is as given in the proposal of the basonym (Yan et al., 2020), with the following addition. The genome size of the type strain is 3.95 Mbp. The DNA G + C content is 57.9 mol%.

Isolated from a deep-sea sediment sample of the New Britain Trench.

The type strain is NBT06E8^T = KCTC 72680^T = MCCC 1K04228^T.

Type strain genome sequence accession numbers: GCA_012427705.1 and GCA_009660035.1.

Type strain 16S rRNA gene sequence accession number: MN435603.

Description of *Vreelandella rituensis* comb. nov.

Vreelandella rituensis (ri.tu.en'sis. N.L. fem. adj. *rituensis*, pertaining to Ritu, Tibet, where the type strain was isolated).

Basonym: *Halomonas rituensis* Gao *et al.* 2020.

The description is as given in the proposal of the basonym (Gao et al., 2020), with the following addition. The genome size of the type strain is 4.47 Mbp. The DNA G + C content is 57.2 mol%.

Isolated from a salt marsh sediment of a saline lake, Dongqian Lake, (33°31′ 51.06″N 80°14′13.64″E), Tibetan Plateau (China).

The type strain is TQ8S^T = CICC 24572^T = KCTC 62530^T.

Type strain genome sequence accession number: GCA_003336665.1.

Type strain 16S rRNA gene sequence accession number: MH071181.

Description of *Vreelandella zhuhanensis* comb. nov.

Vreelandella zhuhanensis (zhu.han.en'sis. N.L. fem. adj. *zhuhanensis*, pertaining to Zhuhan marsh on the Tibetan Plateau, where the type strain was isolated).

Basonym: *Halomonas zhuhanensis* Gao *et al.* 2020.

The description is as given in the proposal of the basonym (Gao et al., 2020), with the following addition. The genome size of the type strain is 3.25 Mbp. The DNA G + C content is 57.1 mol%.

Isolated from a saline lake, Zhuhan Lake, (33°32′50.89″N 80°09′38.51″E), Tibetan Plateau (China).

The type strain is ZH2S^T = CICC 24505^T = KCTC 62531^T.

Type strain genome sequence accession number: GCA_009793355.1.

Type strain 16S rRNA gene sequence accession number: MH071182.

Description of *Vreelandella azerica* comb. nov.

Vreelandella azerica (a.ze'ri.ca. N.L. fem. adj. *azerica*, pertaining to Azerbaijan, where the Urmia Lake is located and the type strain was isolated).

Basonym: *Halomonas azerica* Wenting *et al.* 2021.

The description is as given in the proposal of the basonym (Wenting et al., 2021), with the following addition. The genome size of the type strain is 3.42 Mbp. The DNA G + C content is 55.4 mol%.

Isolated from Urmia Lake (Iran).

The type strain is TBZ9^T = KACC 21783^T = LMG 25775^T.

Type strain genome sequence accession number: GCA_013112225.1.

Type strain 16S rRNA gene sequence accession number: MN900573.

Description of *Vreelandella profundi* comb. nov.

Vreelandella profundi (pro.fun'di. L. gen. n. *profundi*, of the depth of the sea).

Basonym: *Halomonas profundi* Wang *et al.* 2022.

The description is as given in the proposal of the basonym (Wang et al., 2022), with the following addition. The genome size of the type strain is 3.60 Mbp. The DNA G + C content is 54.0 mol%.

Isolated from the deep-sea sediment of the Mariana Trench (11.12°N, 142.32°E).

The type strain is MT13^T = KCTC 82923^T = MCCC 1K06389^T.

Type strain genome sequence accession number: GCA_019504685.1.

Type strain 16S rRNA gene sequence accession number: MZ411491.

Description of *Vreelandella populi* comb. nov.

Vreelandella populi (po'pu.li. L. gen. n. *populi*, of the genus *Populus*, referring to *Populus euphratica*).

Basonym: *Halomonas populi* Xu *et al.* 2021.

The description is as given in the proposal of the basonym (Xu et al., 2021), with the following addition. The genome size of the type strain is 3.80 Mbp. The DNA G + C content is 55.0 mol%.

Isolated from *Populus euphratica* in Ebinur Lake Wetland Nature Reserve (China).

The type strain is MC^T = MCCC 1K03942^T = JCM 33545^T.

Type strain genome sequence accession number: GCA_003989825.1.

Type strain 16S rRNA gene sequence accession number: MK045667.

Description of *Vreelandella glaciei* sp. nov.

Vreelandella glaciei (gla.ci.e'i. L. gen. n. *glaciei*, meaning of the cold).

The description is as given in the original proposal of "*Halomonas glaciei*" (Reddy et al., 2003), with the following addition. The genome size of the type strain is 4.96 Mbp. The DNA G + C content is 54.4 mol%.

Isolated from the fast ice of Adelie Land, Antarctica.

The type strain is DD 39^T = CGMCC 1.7263^T = JCM 11692^T = MTCC 4321^T.

Type strain genome sequence accession number: GCA_013415125.1.

Type strain 16S rRNA gene sequence accession number: AJ431369.

Description of *Vreelandella zhaodongensis* sp. nov.

Vreelandella zhaodongensis (zhao.dong.en'sis. N.L. fem. adj. *zhaodongensis*, pertaining to Zhaodong City, North East of China, where the strain was isolated).

The description is as given in the original proposal of "*Halomonas zhaodongensis*" (Jiang et al., 2013), with the following addition. The genome size of the type strain is 3.72 Mbp. The DNA G + C content is 53.0 mol%.

Isolated from saline-alkaline soils in Zhaodong (China).

The type strain is NEAU-ST10-25^T = CGMCC 1.12286^T = DSM 25869^T.

Type strain genome sequence accession number: GCA_013415115.1.

Type strain 16S rRNA gene sequence accession number: JQ762286.

Description of *Vreelandella lionensis* sp. nov.

Vreelandella lionensis (li.on.en'sis. N.L. fem. adj. *lionensis*, of or belonging to Golfe du Lion [Gulf of Lions], in reference to the origin of the type strain).

The description is as given in the original proposal of "*Halomonas lionensis*" (Gaboyer et al., 2014), with the following addition. The genome size of the type strain is 3.65 Mbp. The DNA G + C content is 55.9 mol%.

Isolated from the Mediterranean Sea sediment, Gulf of Lions (France).

The type strain is RHS90^T = CIP 110370^T = DSM 25632^T = UBOCC 3186^T.

Type strain genome sequence accession number: GCA_002087295.1.

Type strain 16S rRNA gene sequence accession number: HE661586.

Description of *Vreelandella massiliensis* sp. nov.

Vreelandella massiliensis (mas.si.li.en'sis. L. fem. adj. *massiliensis*, of Massilia, the old Roman name for Marseille, where the strain was isolated).

The description is as given in the original proposal of "*Halomonas massiliensis*" (Seck et al., 2016), with the following addition. The genome size of the type strain is 3.44 Mbp. The DNA G + C content is 58.4 mol%.

Isolated from the human gut (France).

The type strain is Marseille-P2426^T = CSUR P2426^T = DSM 103116^T.

Type strain genome sequence accession number: GCA_900155385.1.

Type strain 16S rRNA gene sequence accession number: LT223576.

Description of *Vreelandella maris* sp. nov.

Vreelandella maris (ma'ris. L. gen. n. *maris*, of the sea).

The description is as given in the original proposal of "*Halomonas maris*" (Qiu et al., 2021a), with the following addition. The genome size of the type strain is 4.52 Mbp. The DNA G + C content is 54.4 mol%.

Isolated from the deep-sea sediment in the Southwest Indian Ocean (China).

The type strain is QX-1^T = KCTC 82198^T = MCCC 1A17875^T = NBRC 114670^T.

Type strain genome sequence accession number: GCA_013371085.1.

Type strain 16S rRNA gene sequence accession number: MT372903.

Description of *Vreelandella sedimenti* sp. nov.

Vreelandella sedimenti (se.di.men'ti. L. gen. n. *sedimenti*, of sediment, referring to the sediment of the Southwest Indian Ocean, where the type strain was isolated).

The description is as given in the original proposal of "*Halomonas sedimenti*" (Qiu et al., 2021b), with the following addition. The genome size of the type strain is 5.06 Mbp. The DNA G + C content is 54.3 mol%.

Isolated from the deep-sea sediment in the Southwest Indian Ocean (China).

The type strain is QX-2^T = KCTC 82199^T = MCCC 1A17876^T.

Type strain genome sequence accession number: GCA_013416325.1.

Type strain 16S rRNA gene sequence accession number: MT372904.

Description of *Bisbaumannia* gen. nov.

Bisbaumannia (Bis.bau.mann'i.a. L. adv. *bis*, twice; N.L. fem. n. *Bisbaumannia*, referring to both microbiologist Linda Baumann and Paul Baumann, who first studied these microorganisms).

Cells are Gram-staining-negative straight rods, 0.8–1.1 × 1.5–3.0 μm in size, aerobic, and motile by means of peritrichous flagella. Endospores are not formed. Oxidase-positive. Colonies are convex with entire edges and cream-colored. Na^+ is required for growth. Mesophilic. Chemo-organotrophic. Nitrate reduction is negative. The major respiratory quinone is Q9. The major fatty acids are $C_{18:1}\omega7c$, $C_{19:0}$ cyclo $\omega8c$, $C_{16:0}$, and $C_{16:1}\omega7c$. The major polar lipids are phosphatidylglycerol, diphosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content is 67.2 mol%.

The genus *Bisbaumannia* belongs to the family *Halomonadaceae*. The type species is *Bisbaumannia pacifica.*

Description of *Bisbaumannia pacifica* comb. nov.

Bisbaumannia pacifica (pa.ci'fi.ca. L. fem. adj. *pacifica*, peaceful, pertaining to the Pacific Ocean).

Basonym: *Alcaligenes pacificus* corrig. Baumann *et al.* 1972 (Approved Lists 1980).

Homotypic synomyms: *Halomonas pacifica* (Baumann *et al.* 1972) Dobson and Franzmann 1996; *Deleya pacifica* (Baumann *et al.* 1972) Baumann *et al.* 1983.

The description is as given in the original proposal of the basonym (Baumann et al., 1972), with the following addition. The genome size of the type strain is 3.85 Mbp. The DNA G + C content is 67.2 mol%.

Isolated from the seawater off the coast of Oahu (Hawaii, United States), Pacific Ocean.

The type strain is 62^T = ATCC 27122^T = CIP 103200^T = DSM 4742^T = JCM 20633^T = LMG 3446^T = NBRC 102220^T = NCIMB 1977^T.

Type strain genome sequence accession number: GCA_007989625.1.

Type strain 16S rRNA gene sequence accession number: AB681734.

Description of *Billgrantia* gen. nov.

Billgrantia (Bill.grant'i.a. N.L. fem. n. *Billgrantia*, named after the microbiologist William [Bill] D. Grant for his great contribution to the study of halophilic microorganisms).

Cells are Gram-staining-negative rods, 0.3–1.1 × 0.8–6.0 μm in size, aerobic or facultatively anaerobic, and mostly motile. Endospores are not formed. Catalase and oxidase are positive for most of the strains. Colonies are brown, cream, light beige, pinkish white, white, or yellow pigmented. Slightly to moderately halophilic, growing at 0–26% (w/v) NaCl, with optimal growth at 1–15% (w/v) NaCl. Mesophilic, thriving at 4–55°C, showing optimal growth at 25–42°C. Alkaliphilic or alkalitolerant, growing at pH values in the range of 5.0–12.0, with optimal growth at pH 7.0–10.0. Chemo-organotrophic. Nitrate reduction is mostly positive. The major respiratory quinones are Q9 and Q8. The major fatty acids are $C_{18:1}\omega6c/C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega6c/C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{19:0}$ cyclo $\omega8c$, $C_{12:0}$ 3-OH, $C_{16:1}\omega9c$, and $C_{17:1}\omega9c$. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content ranges between 62.1 and 67.5 mol%.

The genus *Billgrantia* belongs to the family *Halomonadaceae*. The type species is *Billgrantia desiderata*.

Description of *Billgrantia desiderata* comb. nov.

Billgrantia desiderata (de.si.de.ra'ta. L. fem. adj. *desiderata*, wished for, the strain wished for).

Basonym: *Halomonas desiderata* Berendes *et al.* 1997.

The description is as given in the original proposal of the basonym (Berendes et al., 1996), with the following addition. The DNA G + C content is 64.7–64.9 mol%.

The type strain is FB2^T = CIP 105505^T = DSM 9502^T = LMG 19548^T. The genome size of the type strain is 4.89 Mbp, and its DNA G + C content is 64.7 mol%. Isolated from a municipal sewage treatment plant in Göttingen (Germany).

Type strain genome sequence accession number: GCA_011742915.1.

Type strain 16S rRNA gene sequence accession number: X92417.

Description of *Billgrantia campisalis* comb. nov.

Billgrantia campisalis (cam.pi.sa'lis. L. masc. n. *campus*, field, plain; L. masc. n. *sal*, salt; N.L. gen. n. *campisalis*, of the plain of salt, of the salt plain).

Basonym: *Halomonas campisalis* Mormile *et al.* 2000.

The description is as given in the original proposal of the basonym (Mormile et al., 1999), with the following addition. The genome size of the type strain is 4.27 Mbp. The DNA G + C content is 66.3 mol%.

Isolated from a soil sample collected from a salt flat south of Alkali Lake, Washington State (USA).

The type strain is 4A^T = ATCC 700597^T = CIP 106639^T = DSM 15413^T.

Type strain genome sequence accession numbers: GCA_031451595.1 and GCA_022341425.1.

Type strain 16S rRNA gene sequence accession number: AF054286.

Description of *Billgrantia gudaonensis* comb. nov.

Billgrantia gudaonensis (gu.dao.nen'sis. N.L. fem. adj. *gudaonensis*, pertaining to Gudao, in the Shengli oilfield, PR China, where the type strain was isolated).

Basonym: *Halomonas gudaonensis* Wang *et al.* 2007.

The description is as given in the original proposal of the basonym (Wang et al., 2007), with the following addition. The genome size of the type strain is 4.17 Mbp. The DNA G + C content is 64.9 mol%.

Isolated from saline soil contaminated by crude oil, Gudao (China).

The type strain is SL014B-69^T = CGMCC 1.6133^T = DSM 23417^T = LMG 23610^T.

Type strain genome sequence accession number: GCA_900100195.1.

Type strain 16S rRNA gene sequence accession number: DQ421808.

Description of *Billgrantia kenyensis* comb. nov.

Billgrantia kenyensis (ke.ny.en'sis. N.L. fem. adj. *kenyensis*, Kenyan, of Kenya, the region of isolation).

Basonym: *Halomonas kenyensis* Boltyanskaya *et al.* 2008.

The description is as given in the original proposal of the basonym (Boltyanskaya et al., 2007), with the following addition. The genome size of the type strain is 4.42 Mbp. The DNA G + C content is 63.8 mol%.

Isolated from sediments from soda lakes (Kenya).

The type strain is AIR-2^T = DSM 17331^T = VKM B-2354^T.

Type strain genome sequence accession number: GCA_013697085.1.

Type strain 16S rRNA gene sequence accession number: AY962237.

Description of *Billgrantia saliphila* comb. nov.

Billgrantia saliphila (sa.li'phi.la. L. masc. n. *sal*, salt; Gr. masc. adj. *philos*, loving; N.L. fem. adj. *saliphila*, salt-loving).

Basonym: *Halomonas saliphila* Gan *et al.* 2018.

The description is as given in the original proposal of the basonym (Gan et al., 2018), with the following addition. The genome size of the type strain is 4.34 Mbp. The DNA G + C content is 64.1 mol%.

Isolated from saline soil (China).

The type strain is LCB169^T = CGMCC 1.15818^T = KCTC 52618^T.

Type strain genome sequence accession number: GCA_002930105.1.

Type strain 16S rRNA gene sequence accession number: KX008964.

Description of *Billgrantia endophytica* comb. nov.

Billgrantia endophytica (en.do.phy'ti.ca. Gr. pref. *endo-*, within; Gr. neut. n. *phyton*, plant; L. fem. adj. suff. *-ica*, adjectival suffix used with

the sense of belonging to; N.L. fem. adj. *endophytica*, within plant, pertaining to the endophytic nature of the strain and its isolation from internal plant tissues).

Basonym: *Halomonas endophytica* Chen *et al.* 2018.

The description is as given in the original proposal of the basonym (Chen et al., 2018), with the following addition. The genome size of the type strain is 4.98 Mbp. The DNA G + C content is 62.1 mol%.

Isolated from liquid in the stems of *Populus euphratica* in Xinjiang (China).

The type strain is MC28^T = KCTC 52999^T = MCCC 1K03343^T.

Type strain genome sequence accession number: GCA_002879615.1.

Type strain 16S rRNA gene sequence accession number: MF850257.

Description of *Billgrantia montanilacus* comb. nov.

Billgrantia montanilacus (mon.ta.ni.la'cus. L. masc. adj. *montanus*, a mountain; L. masc. n. *lacus*, lake; N.L. gen. n. *montanilacus*, of a mountain lake).

Basonym: *Halomonas montanilacus* Lu *et al.* 2020.

The description is as given in the original proposal of the basonym (Lu et al., 2020), with the following addition. The genome size of the type strain is 4.79 Mbp. The DNA G + C content is 62.9 mol%.

Isolated from hypersaline Lake Pengyanco on the Tibetan Plateau (China).

The type strain is PYC7W^T = CICC 24506^T = KCTC 62529^T.

Type strain genome sequence accession number: GCA_003336675.1.

Type strain 16S rRNA gene sequence accession number: MH071180.

Description of *Billgrantia lactosivorans* comb. nov.

Billgrantia lactosivorans (lac.to.si.vo'rans. L. neut. adj. *lactosum*, lactose; L. pres. part. *vorans*, eating; N.L. part. adj. *lactosivorans*, eating lactose).

Basonym: *Halomonas lactosivorans* Ming *et al.* 2020.

The description is as given in the original proposal of the basonym (Ming et al., 2020), with the following addition. The genome size of the type strain is 4.36 Mbp. The DNA G + C content is 66.7 mol%.

Isolated from salt-lake sediment in Shanxi Province (China).

The type strain is CFH 90008^T = DSM 103220^T = KCTC 52281^T.

Type strain genome sequence accession number: GCA_003254665.1.

Type strain 16S rRNA gene sequence accession number: KY039330.

Description of *Billgrantia pellis* comb. nov.

Billgrantia pellis (pel'lis. L. gen. n. *pellis*, of a hide, indicating the source of the type strain).

Basonym: *Halomonas pellis* Li *et al.* 2020.

The description is as given in the original proposal of the basonym (Li et al., 2020), with the following addition. The genome size of the type strain is 4.35 Mbp. The DNA G + C content is 63.6 mol%.

Isolated from wet salted hides (China).

The type strain is L5^T = CGMCC 1.17335^T = KCTC 72573^T.

Type strain genome sequence accession number: GCA_008297955.1.

Type strain 16S rRNA gene sequence accession number: MN099429.

Description of *Billgrantia azerbaijanica* comb. nov.

Billgrantia azerbaijanica (a.zer.bai.ja'ni.ca. N.L. fem. adj. *azerbaijanica*, pertaining to Azerbaijan, a region in the north-west of Iran, where Urmia Lake is located and from which the type strain was isolated).

Basonym: *Halomonas azerbaijanica* Kazemi *et al.* 2021.

The description is as given in the original proposal of the basonym (Kazemi et al., 2021), with the following addition. The genome size of the type strain is 4.58 Mbp. The DNA G + C content is 67.5 mol%.

Isolated from water from Urmia Lake (Iran).

The type strain is TBZ202^T = CECT 9693^T = KCTC 62817^T.

Type strain genome sequence accession number: GCA_004551485.1.

Type strain 16S rRNA gene sequence accession number: MK138622.

Description of *Billgrantia diversa* comb. nov.

Billgrantia diversa (di.ver'sa. L. fem. part. adj. *diversa*, different, distinct).

Basonym: *Halomonas diversa* Wang *et al.* 2021.

The description is as given in the original proposal of the basonym (Wang et al., 2021), with the following addition. The genome size of the type strain is 4.49 Mbp. The DNA G + C content is 62.9 mol%.

Isolated from the deep-sea sediment of the Pacific Ocean.

The type strain is D167-6-1^T = KCTC 72441^T = MCCC 1A13316^T.

Type strain genome sequence accession number: GCA_014931605.1.

Type strain 16S rRNA gene sequence accession number: MW172430.

Description of *Billgrantia bachuensis* comb. nov.

Billgrantia bachuensis (ba.chu.en'sis. N.L. fem. adj. *bachuensis*, pertaining to Bachu, north-western China, where the strain was isolated).

Basonym: *Halomonas bachuensis* Xiao *et al.* 2021.

The description is as given in the original proposal of the basonym (Xiao et al., 2021), with the following addition. The genome size of the type strain is 4.70 Mbp. The DNA G + C content is 63.6 mol%.

Isolated from Gobi soil, Bachu (China).

The type strain is DX6^T = CCTCC AB 2020094^T = KCTC 82196^T.

Type strain genome sequence accession number: GCA_011742165.1.

Type strain 16S rRNA gene sequence accession number: MT180568.

Description of *Billgrantia antri* comb. nov.

Billgrantia antri (an'tri. L. gen. n. *antri*, of a cave, referring to the location of the isolate).

Basonym: *Halomonas antri* So *et al.* 2022.

The description is as given in the original proposal of the basonym (So et al., 2022), with the following addition. Cells are Gram-stain-negative rods, 0.6–0.9 × 1.6–2.4 μm in size, strictly aerobic or facultatively anaerobic, and motile by means of one polar flagellum or peritrichous flagella. Catalase and oxidase activities are variable. Slightly to moderately halophilic, growing at 0–20% (w/v) NaCl, with optimal growth at 1–8% (w/v) NaCl. Mesophilic, thriving at 4–55°C, showing optimal growth at 25–40°C. Alkalitolerant, growing at pH values in the range of 6.0–10.0, with optimal growth at pH 7.0–8.0. Voges–Proskauer test is variable. Starch is hydrolyzed, but Tween 40 and Tween 60 are not. Hydrolysis of DNA and urea is variable. Esterase lipase (C8), lipase (C14), cystine arylamidase, α-glucosidase, and tryptophan deaminase activities are variable. Utilization of N-acetyl-glucosamine, potassium gluconate, adipate, and citrate is variable. The major respiratory quinone is Q9. The major fatty acids are $C_{18:1}\omega6c$/$C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega6c$/$C_{16:1}\omega7c$, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content is 64.2–64.3 mol%.

The type strain is Y3S6^T = KACC 21536^T = NBRC 114315^T = TBRC 15164^T. The genome size of the type strain is 4.39 Mbp, and its DNA G + C content is 64.3 mol%. Isolated from surface seawater, Busan (Korea).

Type strain genome sequence accession number: GCA_019430905.1.

Type strain 16S rRNA gene sequence accession number: MN625868.

Description of *Billgrantia chromatireducens* sp. nov.

Billgrantia chromatireducens (chro.ma.ti.re.du'cens. N.L. masc. n. *chromas*, chromate; L. pres. part. *reducens*, converting to a different state; N.L. part. adj. *chromatireducens*, reducing chromate).

The description is as given in the original proposal of "*Halomonas chromatireducens*" (Shapovalova et al., 2009), with the following addition. The genome size of the type strain is 3.97 Mbp. The DNA G + C content is 62.8 mol%.

Isolated from soda salt marshes, Altai (Russia).

The type strain is AGD 8-3^T = NCCB 100225^T = VKM B-2497^T.

Type strain genome sequence accession number: GCA_001545155.1.

Type strain 16S rRNA gene sequence accession number: EU447163.

Description of *Billgrantia aerodenitrificans* sp. nov.

Billgrantia aerodenitrificans (a.e.ro.de.ni.tri'fi.cans. Gr. masc. n. *aêr*, air; N.L. inf. v. *denitrificare*, to denitrify; N.L. part. adj. *aerodenitrificans*, denitrifying with or in air).

The description is as given in the original proposal of "*Halomonas aerodenitrificans*" (Wang and Shao, 2021), with the following addition. The genome size of the type strain is 5.08 Mbp. The DNA G + C content is 64.0 mol%.

Isolated from coastal water, 5 m depth, from the Taiwan Strait (China).

The type strain is CYD-9^T = KCTC 72088^T = MCCC 1A11058^T.

Type strain genome sequence accession number: GCA_021404405.1.

Type strain 16S rRNA gene sequence accession number: MW205680.

Description of *Billgrantia ethanolica* sp. nov.

Billgrantia ethanolica (e.tha.no'li.ca. N.L. neut. n. *ethanol*, ethanol; L. fem. adj. suff. *-ica*, suffix used with various meanings; N.L. fem. adj. *ethanolica*, belonging to ethanol, in reference to the ability of the species to utilize ethanol as a substrate for growth).

The description is as given in the original proposal of "*Halomonas ethanolica*" (Wang and Shao, 2021), with the following addition. The genome size of the type strain is 4.57 Mbp. The DNA G + C content is 64.5 mol%.

Isolated from the sediment from shrimp culture pond, Zhangzhou (China).

The type strain is CYT3-1-1^T = KCTC 72090^T = MCCC 1A11081^T.

Type strain genome sequence accession number: GCA_021404305.1.

Type strain 16S rRNA gene sequence accession number: MW205683.

Description of *Billgrantia sulfidoxydans* sp. nov.

Billgrantia sulfidoxydans (sul.fid.o'xy.dans. N.L. neut. n. *sulfidum*, sulfide; N.L. pres. part. *oxydans*, oxidizing; N.L. part. adj. *sulfidoxydans*, oxidizing sulfides).

The description is as given in the original proposal of "*Halomonas sulfidoxydans*" (Wang and Shao, 2021), with the following addition. The genome size of the type strain is 4.49 Mbp. The DNA G + C content is 66.0 mol%.

Isolated from surface sediments in the coastal sea at Taiwan Strait (China).

The type strain is CYN-1-2^T = KCTC 72089^T = MCCC 1A11059^T.

Type strain genome sequence accession number: GCA_017868775.1.

Type strain 16S rRNA gene sequence accession number: MW205681.

Description of *Billgrantia tianxiuensis* sp. nov.

Billgrantia tianxiuensis (tian.xiu.en'sis. N.L. fem. adj. *tianxiuensis*, pertaining to the Tianxiu Hydrothermal Field, on the Northwest Indian Ridge, from where the type strain was isolated).

The description is as given in the original proposal of "*Halomonas tianxiuensis*" (Wang and Shao, 2021), with the following addition. The genome size of the type strain is 5.02 Mbp. The DNA G + C content is 63.9 mol%.

Isolated from sulfide from Tianxiu hydrothermal vents, 3,440 m depth, Northwest Indian Ocean.

The type strain is BC-M4-5^T = KCTC 72092^T = MCCC 1A14433^T.

Type strain genome sequence accession number: GCA_009834345.1.

Type strain 16S rRNA gene sequence accession number: MW205685.

Description of *Billgrantia zhangzhouensis* sp. nov.

Billgrantia zhangzhouensis (zhang.zhou.en'sis. N.L. fem. adj. *zhangzhouensis*, of or pertaining to Zhangzhou, a city in Fujian, China, where the type strain was isolated).

The description is as given in the original proposal of "*Halomonas zhangzhouensis*" (Wang and Shao, 2021), with the following addition. The genome size of the type strain is 4.41 Mbp. The DNA G + C content is 63.3 mol%.

Isolated from the sediments from a shrimp culture pond in Zhangzhou (China).

The type strain is CXT3-11^T = KCTC 72087^T = MCCC 1A11036^T.

Type strain genome sequence accession number: GCA_021404465.1.

Type strain 16S rRNA gene sequence accession number: MW205678.

Description of *Franzmannia* gen. nov.

Franzmannia (Franz.man'ni.a. N.L. fem. n. *Franzmannia*, in honor of Peter D. Franzmann, Australian microbiologist and polar researcher).

Cells are Gram-staining-negative rod-shaped or pleomorphic, 0.3–0.7 × 1.4–2.8 μm in size, aerobic, and motile or non-motile. Endospores are not formed. Catalase-variable and oxidase-positive. Colonies are cream, cream-pink, or light yellow pigmented. Moderately halophilic, growing at 0.5–20% (w/v) NaCl, with optimal growth at 2–10% (w/v) NaCl. Mesophilic, thriving at 10–45°C, showing optimal growth at 25–37°C. Alkalitolerant, growing at pH values in the range of 5.5–11.0, with optimal growth at pH 7.0–9.0. Chemo-organotrophic. Nitrate reduction is variable. The major respiratory quinone is Q9. The major fatty acids are $C_{18:1}\omega6c/C_{18:1}\omega7c$, $C_{16:0}$, $C_{19:0}$ cyclo $\omega8c$, and $C_{16:1}\omega6c/C_{16:1}\omega7c$. The major polar lipids are phosphatidylglycerol, diphosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content ranges between 63.8 and 64.5 mol%.

The genus *Franzmannia* belongs to the family *Halomonadaceae*. The type species is *Franzmannia pantelleriensis*.

Description of *Franzmannia pantelleriensis* comb. nov.

Franzmannia pantelleriensis (pan.tel.le.ri.en'sis. L. fem. adj. *pantelleriensis*, pertaining to Pantelleria Island [the place of isolation] in the south of Sicily, Italy).

Basonym: *Halomonas pantelleriensis* corrig Romano *et al.* 1997.

The description is as given in the proposal of the basonym (Romano et al., 1996), with the following addition. The genome size of the type strain is 4.40 Mbp. The DNA G + C content is 63.9 mol%.

Isolated from the hard sand of Venere Lake, Pantelleria Island (Italy).

The type strain is AAP^T = ATCC 700273^T = CIP 105506^T = DSM 9661^T = LMG 19550^T.

Type strain genome sequence accession number: GCA_900102875.1.

Type strain 16S rRNA gene sequence accession number: X93493.

Description of *Franzmannia qiaohouensis* comb. nov.

Franzmannia qiaohouensis (qiao.hou.en'sis. N.L. fem. adj. *qiaohouensis*, pertaining to Qiaohou salt mine, south-west China, where the type strain was isolated).

Basonym: *Halomonas qiaohouensis* Wang *et al.* 2015.

The description is as given in the proposal of the basonym (Wang et al., 2014), with the following addition. The genome size of the type strain is 4.65 Mbp. The DNA G + C content is 64.5 mol%.

Isolated from the salt mine soil in southwest China.

The type strain is YIM QH88^T = ACCC 60021^T = CCTCC AB 2012965^T = DSM 26770^T.

Type strain genome sequence accession number: GCA_031451695.1.

Type strain 16S rRNA gene sequence accession number: KC237714.

Description of *Franzmannia salipaludis* sp. nov.

Franzmannia salipaludis (sa.li.pa.lu'dis. L. masc. n. *sal*, salt; L. fem. n. *palus*, swamp, marsh; N.L. gen. n. *salipaludis*, of a salt marsh).

The description is as given in the original proposal of "*Halomonas salipaludis*" (Xing et al., 2021), with the following addition. The genome size of the type strain is 5.48 Mbp. The DNA G + C content is 63.8 mol%.

Isolated from the saline-alkali wetland soil (China).

The type strain is WRN001^T = ACCC 19974^T = KCTC 52853^T.

Type strain genome sequence accession number: GCA_002286975.1.

Type strain 16S rRNA gene sequence accession number: MF782428.

Description of *Litchfieldella* gen. nov.

Litchfieldella (Litch.field.el'la. N.L. fem. n. *Litchfieldella*, named after Dr. Carol D. Litchfield [1936–2012], in recognition of her many contributions to the study of halophilic microorganisms).

Cells are Gram-staining-negative rods, 0.4–1.0 × 1.2–4.4 μm in size, aerobic, and motile or non-motile. Endospores are not formed. Catalase-positive and oxidase-variable. Colonies are cream, yellow, brown-orange, or creamy-white pigmented. Moderately halophilic, growing at 0–23% (w/v) NaCl, with optimal growth at 5–13% (w/v) NaCl. Mesophilic, thriving at 15–50°C, showing optimal growth at 32–37°C. Alkalitolerant, growing at pH values in the range of 5.0–10.0, with optimal growth at pH 6.0–9.0. Chemo-organotrophic. Nitrate reduction is variable. The major respiratory quinone is Q9. The major fatty acids are $C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, and $C_{19:0}$ cyclo $\omega8c$.

The DNA G + C content ranges between 58.5 and 62.2 mol%.

The genus *Litchfieldella* belongs to the family *Halomonadaceae*. The type species is *Litchfieldella anticariensis*.

Description of *Litchfieldella anticariensis* comb. nov.

Litchfieldella anticariensis (an.ti.ca.ri.en'sis. L. fem. adj. *anticariensis*, pertaining to Antequera, originally the Roman city of Anticaria, in the province of Málaga, southern Spain, where the strains were isolated).

Basonym: *Halomonas anticariensis* Martínez-Cánovas *et al.* 2004.

The description is as given in the proposal of the basonym (Martínez-Cánovas et al., 2004), with the following addition. The genome size of the type strain is 5.07 Mbp. The DNA G + C content is 58.5 mol%.

Isolated from a soil sample from Fuente de Piedra, a saline-wetland wildfowl reserve in Málaga (Spain).

The type strain is FP35^T = CECT 5854^T = CIP 108499^T = DSM 16096^T = LMG 22089^T.

Type strain genome sequence accession numbers: GCA_000409775.1 and GCA_000428505.1.

Type strain 16S rRNA gene sequence accession number: AY489405.

Description of *Litchfieldella xinjiangensis* comb. nov.

Litchfieldella xinjiangensis (xin.ji.ang.en'sis. N.L. fem. adj. *xinjiangensis*, pertaining to Xinjiang, a region of China, from where the type strain was isolated).

Basonym: *Halomonas xinjiangensis* Guan *et al.* 2010.

The description is as given previously (Guan et al., 2010; Navarro-Torre et al., 2020), with the following addition. The genome size of the type strain is 3.79 Mbp. The DNA G + C content is 60.7 mol%.

Isolated from a soil sample from Lop Nur salt lake (4° 23′ N 9° 18′ E, 778 m altitude) in Xinjiang Province, north-west China.

The type strain is TRM 0175^T = CCTCC AB 208329^T = KCTC 22608^T.

Type strain genome sequence accession number: GCA_000759345.1.

Type strain 16S rRNA gene sequence accession number: EU822512.

Description of *Litchfieldella rifensis* comb. nov.

Litchfieldella rifensis (ri.fen'sis. N.L. fem. adj. *rifensis*, pertaining to the Rif Mountains in northern Morocco, where the strain was isolated).

Basonym: *Halomonas rifensis* Amjres *et al.* 2011.

The description is as given in the proposal of the basonym (Amjres et al., 2011), with the following addition. The genome size of the type strain is 4.83 Mbp. The DNA G + C content is 62.2 mol%.

Isolated from a solar saltern in the Rif Mountains (Morocco).

The type strain is HK31^T = CECT 7698^T = LMG 25695^T.

Type strain genome sequence accession number: GCM10020179.

Type strain 16S rRNA gene sequence accession number: HM026177.

Description of *Litchfieldella qijiaojingensis* comb. nov.

Litchfieldella qijiaojingensis (qi.jiao'jing.en'sis. N.L. fem. adj. *qijiaojingensis*, pertaining to Qijiaojing Lake, Xinjiang Province, north-west China, where the sample from which the type strain was isolated and was collected).

Basonym: *Halomonas qijiaojingensis* Chen *et al.* 2012.

The description is as given in the proposal of the basonym (Chen et al., 2011), with the following addition. The genome size of the type strain is 4.77 Mbp. The DNA G + C content is 60.8 mol%.

Isolated from the shore sediment from a salt lake, Qijiaojing Lake (43°23′01″ N 91°36′11″ E), Xinjiang province (China).

The type strain is YIM 93003^T = CCTCC AB 208133^T = DSM 22403^T = KCTC 22228^T.

Type strain genome sequence accession number: GCA_014651875.1.

Type strain 16S rRNA gene sequence accession number: HQ832735.

Description of *Onishia* gen. nov.

Onishia (O.ni'shi.a. N.L. fem. n. *Onishia*, named after Dr. Hiroshi Ōnishi [Japan], in recognition of his many contributions to the study of halophilic microorganisms).

Cells are Gram-staining-negative rods, 0.6–1.0 × 1.8–3.2 μm in size, aerobic or facultatively anaerobic, and motile. Endospores are not formed. Catalase and oxidase are positive. Colonies are cream

pigmented. Moderately halophilic, growing at 1–25% (w/v) NaCl, with optimal growth at 10–15% (w/v) NaCl. Mesophilic, thriving at 4–45°C, showing optimal growth at 22–35°C. Alkalitolerant, growing at pH values in the range of 4.0–10.0, with optimal growth at pH 7.5–8.0. Chemo-organotrophic. Nitrate reduction is positive. The major respiratory quinones are Q9 and, according to the genome sequence, also Q8. The major fatty acids are $C_{18:1}\omega6c$/$C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega6c$/$C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. According to the genome sequence, the major polar lipids are phosphatidylglycerol, phosphatidylethanolamine, and diphosphatidylglycerol.

The DNA G + C content ranges between 61.1 and 62.3 mol%.

The genus *Onishia* belongs to the family *Halomonadaceae*. The type species is *Onishia taeanensis*.

Description of *Onishia taeanensis* comb. nov.

Onishia taeanensis (tae.an.en'sis. N.L. fem. adj. *taeanensis*, belonging to Taean, from where the organism was isolated).

Basonym: *Halomonas taeanensis* Lee *et al.* 2005.

The description is as given in the proposal of the basonym (Lee et al., 2005), with the following addition. The genome size of the type strain is 3.76 Mbp. The DNA G + C content is 62.3 mol%.

Isolated from the soil from a solar saltern (Korea).

The type strain is BH539^T = CIP 109003^T = DSM 16463^T = KCTC 12284^T.

Type strain genome sequence accession number: GCA_900100755.1.

Type strain 16S rRNA gene sequence accession number: AY671975.

Description of *Onishia niordana* comb. nov.

Onishia niordana (nior.di.a'na. N.L. fem. adj. *niordana*, pertaining to Njörd [Niord], Nordic god of marine coast).

Basonym: *Halomonas niordiana* Diéguez *et al.* 2020.

The description is as given in the proposal of the basonym (Diéguez et al., 2020), with the following addition. The genome size of the type strain is 3.68 Mbp. The DNA G + C content is 61.1 mol%.

Isolated from seawater (Norway).

The type strain is ATF 5.4^T = CECT 9779^T = LMG 31227^T.

Type strain genome sequence accession number: GCA_004798965.1.

Type strain 16S rRNA gene sequence accession number: SDSD01000014.

Emended description of the genus *Modicisalibacter* Ben Ali Gam *et al.* 2007

Modicisalibacter (Mo.di.ci.sa.li.bac'ter. L. masc. adj. *modicus*, moderate, limited; L. masc. n. *sal* [gen. *salis*], salt; N.L. masc. n. *bacter*, a rod; N.L. masc. n. *Modicisalibacter*, a moderately halophilic rod).

Cells are Gram-staining-negative rods or short rods, 0.1–1.0 × 0.2–4.0 μm in size, aerobic, and mostly motile. Endospores are not formed. Catalase-positive and oxidase-variable. Colonies are cream, orange, pale/light orange, or yellow pigmented or colorless. Moderately halophilic, growing at 0–27.5% (w/v) NaCl, with optimal growth at 2.5–10% (w/v) NaCl. Mesophilic, thriving at 4–45°C, showing optimal growth at 25–37°C. Alkalitolerant, growing at pH values in the range of 5.0–10.0, with optimal growth at pH 6.5–8.5. Chemo-organotrophic. Nitrate reduction is positive. The major respiratory quinones are Q9 and Q8. The major fatty acids are $C_{18:1}\omega6c$/$C_{18:1}\omega7c$, $C_{16:0}$, $C_{16:1}\omega7c$/iso-$C_{15:0}$ 2-OH, $C_{19:0}$ cyclo $\omega8c$, and $C_{12:0}$ 3-OH. The major polar lipids are diphosphatidylglycerol, phosphatidylglycerol, and phosphatidylethanolamine.

The DNA G + C content ranges between 59.1 and 67.4 mol%.

The genus *Modicisalibacter* belongs to the family *Halomonadaceae*. The type species is *Modicisalibacter tunisiensis*.

Description of *Modicisalibacter muralis* comb. nov.

Modicisalibacter muralis (mu.ra'lis. L. masc. adj. *muralis*, pertaining or belonging to walls).

Basonym: *Halomonas muralis* Heyrman *et al.* 2002.

The description is as given in the proposal of the basonym (Heyrman et al., 2002), with the following addition. The genome size of the type strain is 4.14 Mbp. The DNA G + C content is 61.9 mol%.

Isolated from microbial biofilms colonizing the walls and murals of the Saint-Catherine chapel, Castle Herberstein (Austria).

The type strain is R-5058^T = CIP 108825^T = DSM 14789^T = LMG 20969^T.

Type strain genome sequence accession number: GCA_900102945.1.

Type strain 16S rRNA gene sequence accession number: AJ320530.

Description of *Modicisalibacter luteus* comb. nov.

Modicisalibacter luteus (lu'te.us. L. masc. adj. *luteus*, orange-colored).

Basonym: *Halomonas lutea* Wang *et al.* 2008.

The description is as given in the proposal of the basonym (Wang et al., 2008), with the following addition. The genome size of the type strain is 4.53 Mbp. The DNA G + C content is 59.1 mol%.

Isolated from a salt lake, Xinjiang province (China).

The type strain is YIM 91125^T = CCTCC AB 206093^T = DSM 23508^T = KCTC 12847^T.

Type strain genome sequence accession number: GCA_000378505.1.

Type strain 16S rRNA gene sequence accession number: EF674852.

Description of *Modicisalibacter ilicicola* comb. nov.

Modicisalibacter ilicicola (i.li.ci'co.la. L. n. *ilici*, the Roman name of Elche, the city close to the solar salterns where the type strain was

isolated; L. masc. n. suff. *-cola*, inhabitant, dweller; N.L. masc. n. *ilicicola*, inhabitant of Ilici).

Basonym: *Halomonas ilicicola* Arenas *et al.* 2009.

The description is as given in the proposal of the basonym (Arenas et al., 2009), with the following addition. The genome size of the type strain is 3.96 Mbp. The DNA G + C content is 63.2 mol%.

Isolated from saline water from a solar saltern, Santa Pola (38° 11′ 35" N 0° 35′ 45" W), Alicante (Spain).

The type strain is SP8^T = CCM 7522^T = CECT 7331^T = DSM 19980^T.

Type strain genome sequence accession number: GCA_900128925.1.

Type strain 16S rRNA gene sequence accession number: EU218533.

Description of *Modicisalibacter xianhensis* comb. nov.

Modicisalibacter xianhensis (xianh.en'sis. N.L. masc. adj. *xianhensis*, of or pertaining to Xianhe, Shandong Province, China, where the type strain was isolated).

Basonym: *Halomonas xianhensis* Zhao *et al.* 2012.

The description is as given in the proposal of the basonym (Zhao et al., 2012), with the following addition. The genome size of the type strain is 4.36 Mbp. The DNA G + C content is 61.3 mol%.

Isolated from saline soil contaminated with crude oil from Xianhe, Shandong Province (China).

The type strain is A-1^T = CGMCC 1.6848^T = JCM 14849^T.

Type strain genome sequence accession number: GCA_900113605.1.

Type strain 16S rRNA gene sequence accession number: EF421176.

Description of *Modicisalibacter zincidurans* comb. nov.

Modicisalibacter zincidurans (zin.ci.du'rans. N.L. neut. n. *zincum*, zinc; L. pres. part. *durans*, enduring, being insensible; N.L. part. adj. *zincidurans*, zinc tolerating).

Basonym: *Halomonas zincidurans* Xu *et al.* 2013.

The description is as given previously (Xu et al., 2013; Navarro-Torre et al., 2020), with the following addition. The genome size of the type strain is 3.55 Mbp. The DNA G + C content is 64.4 mol%.

Isolated from deep-sea sediment from the South Atlantic Ocean.

The type strain is B6^T = CGMCC 1.12450^T = JCM 18472^T.

Type strain genome sequence accession number: GCA_000731955.1.

Type strain 16S rRNA gene sequence accession number: JQ781698.

Description of *Modicisalibacter radicis* comb. nov.

Modicisalibacter radicis (ra'di.cis. L. gen. n. *radicis*, of a root).

Basonym: *Halomonas radicis* Navarro-Torre *et al.* 2020.

The description is as given in the proposal of the basonym (Navarro-Torre et al., 2020), with the following addition. The genome size of the type strain is 4.65 Mbp. The DNA G + C content is 64.9 mol%.

Isolated from *Arthrocnemum macrostachyum* growing in the Odiel marshes (Spain).

The type strain is EAR18^T = CECT 9077^T = LMG 29859^T.

Type strain genome sequence accession number: GCA_900961225.1.

Type strain 16S rRNA gene sequence accession number: KU320882.

Description of *Modicisalibacter coralii* sp. nov.

Modicisalibacter coralii (co.ra'li.i. L. gen. n. *coralii*, from the coral *Mussismilia braziliensis*).

The description is as given in the original proposal of "*Halomonas coralii*" (Vidal et al., 2019), with the following addition. The genome size of the type strain is 4.44 Mbp. The DNA G + C content is 66.3 mol%.

Isolated from *Mussismilia braziliensis* (Brazil).

The type strain is 362.1^T = CBAS 715^T.

Type strain genome sequence accession number: GCA_004117855.1.

Type strain 16S rRNA gene sequence accession number: QWBW01000001.

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 Supplementary material.

Author contributions

RRH: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. DRA: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. CS-P: Conceptualization, Investigation, Writing – review & editing. MC: Formal analysis, Writing – review & editing. SW: Supervision, Writing – review & editing. PH: Supervision, Writing – review & editing. AV: Conceptualization, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by grants PID2020-118136GB-I00 funded by MCIN/AEI/10.13039/501100011033, and from the Junta de Andalucía (P20_01066 and BIO-213), both with FEDER funds. RRH was a recipient of a short-stay grant (PRX21/00598) from

the Spanish Ministry of Universities. MC and PH were funded by an Australian Research Council Discovery Project (grant number DP220100900).

Acknowledgments

The authors thank A. Oren (The Hebrew University of Jerusalem) for his help on the nomenclature of the new taxa and Brian Kemish (The University of Queensland, Australia) for his generous help and advice on the computing resources used for the presented data analyses.

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 author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

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.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1293707/full#supplementary-material

References

Amjres, H., Béjar, V., Quesada, E., Abrini, J., and Llamas, I. (2011). *Halomonas rifensis* sp. nov., an exopolysaccharide-producing, halophilic bacterium isolated from a solar saltern. *Int. J. Syst. Evol. Microbiol.* 61, 2600–2605. doi: 10.1099/ijs.0.027268-0

Amouric, A., Liebgott, P.-P., Joseph, M., Brochier-Armanet, C., and Lorquin, J. (2014). *Halomonas olivaria* sp. nov., a moderately halophilic bacterium isolated from olive-processing effluents. *Int. J. Syst. Evol. Microbiol.* 64, 46–54. doi: 10.1099/ijs.0.049007-0

Ankenbrand, M. J., and Keller, A. (2016). bcgTree: automatized phylogenetic tree building from bacterial core genomes. *Genome* 59, 783–791. doi: 10.1139/gen-2015-0175

Arahal, D. R., Ludwig, W., Schleifer, K. H., and Ventosa, A. (2002). Phylogeny of the family *Halomonadaceae* based on 23S and 16S rDNA sequence analyses. *Int. J. Syst. Evol. Microbiol.* 52, 241–249. doi: 10.1099/00207713-52-1-241

Arenas, M., Bañón, P. I., Copa-Patiño, J. L., Sánchez-Porro, C., Ventosa, A., and Soliveri, J. (2009). *Halomonas ilicicola* sp. nov., a moderately halophilic bacterium isolated from a saltern. *Int. J. Syst. Evol. Microbiol.* 59, 578–582. doi: 10.1099/ijs.0.003509-0

Auch, A. F., von Jan, M., Klenk, H.-P., and Göker, M. (2010). Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genomic Sci.* 2, 117–134. doi: 10.4056/sigs.531120

Baumann, L., Baumann, P., Mandel, M., and Allen, R. D. (1972). Taxonomy of aerobic marine eubacteria. *J. Bacteriol.* 110, 402–429. doi: 10.1128/jb.110.1.402-429.1972

Ben Ali Gam, Z., Abdelkafi, S., Casalot, L., Tholozan, J. L., Oueslati, R., and Labat, M. (2007). *Modicisalibacter tunisiensis* gen. nov., sp. nov., an aerobic, moderately halophilic bacterium isolated from an oilfield-water injection sample, and emended description of the family *Halomonadaceae* Franzmann *et al.* 1989 emend Dobson and Franzmann 1996 emend. Ntougias *et al.* 2007. *Int. J. Syst. Evol. Microbiol.* 57, 2307–2313. doi: 10.1099/ijs.0.65088-0

Berendes, F., Gottschalk, G., Heine-Dobbernack, E., Moore, E. R. B., and Tindall, B. J. (1996). *Halomonas desiderata* sp. nov, a new alkaliphilic, halotolerant and denitrifying bacterium isolated from a municipal sewage works. *Syst. Appl. Microbiol.* 19, 158–167. doi: 10.1016/S0723-2020(96)80041-5

Boltyanskaya, Y. V., Kevbrin, V. V., Lysenko, A. M., Kolganova, T. V., Tourova, T. P., Osipov, G. A., et al. (2007). *Halomonas mongoliensis* sp. nov. and *Halomonas kenyensis* sp. nov., new haloalkaliphilic denitrifiers capable of N_2O reduction, isolated from soda lakes. *Microbiology (English translation of Mikrobiologiia)* 76, 739–747. doi: 10.1134/S0026261707060148

Borowiec, M. L. (2016). AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ* 4:e1660. doi: 10.7717/peerj.1660

Bushnell, B. (2020). BBMap short read aligner, and other bioinformatic tools. Available at: https://sourceforge.net/projects/bbmap. Last accessed: 2020.04.23.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 1–9. doi: 10.1186/1471-2105-10-421

Capella-Gutierrez, S., Kauff, F., and Gabaldón, T. (2014). A phylogenomics approach for selecting robust sets of phylogenetic markers. *Nucleic Acids Res.* 42:e54. doi: 10.1093/nar/gku071

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, C., Anwar, N., Wu, C., Fu, G., Wang, R., Zhang, C., et al. (2018). *Halomonas endophytica* sp. nov., isolated from liquid in the stems of *Populus euphratica*. *Int. J. Syst. Evol. Microbiol.* 68, 1633–1638. doi: 10.1099/ijsem.0.002585

Chen, C., Shi, R., Liu, B.-B., Zhang, Y.-J., Sun, H.-Z., Li, C.-T., et al. (2011). *Halomonas qijiaojingensis* sp. nov. and *Halomonas flava* sp. nov., two moderately halophilic bacteria isolated from a salt lake. *Antonie van Leeuwenhoek* 100, 365–373. doi: 10.1007/s10482-011-9591-0

Chen, Y.-G., Zhang, Y.-Q., Huang, H.-Y., Klenk, H.-P., Tang, S.-K., Huang, K., et al. (2009). *Halomonas zhanjiangensis* sp. nov., a halophilic bacterium isolated from a sea urchin. *Int. J. Syst. Evol. Microbiol.* 59, 2888–2893. doi: 10.1099/ijs.0.010173-0

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

de la Haba, R. R., Arahal, D. R., Márquez, M. C., and Ventosa, A. (2010). Phylogenetic relationships within the family *Halomonadaceae* based on comparative 23S and 16S rRNA gene sequence analysis. *Int. J. Syst. Evol. Microbiol.* 60, 737–748. doi: 10.1099/ijs.0.013979-0

de la Haba, R. R., Arahal, D. R., Sánchez-Porro, C., and Ventosa, A. (2014). "The family *Halomonadaceae*" in *The Prokaryotes – Gammaproteobacteria*, eds. E. Rosenberg, E. F. Delong, S. Lory, E. Stackebrandt, and F. Thompson (Berlin Heidelberg: Springer-Verlag), 325–360. doi: 10.1007/978-3-642-38922-1_235

de la Haba, R. R., Carmen Márquez, M., Thane Papke, R., and Ventosa, A. (2012). Multilocus sequence analysis of the family *Halomonadaceae*. *Int. J. Syst. Evol. Microbiol.* 62, 520–538. doi: 10.1099/ijs.0.032938-0

Diéguez, A. L., Balboa, S., and Romalde, J. L. (2020). *Halomonas borealis* sp. nov. and *Halomonas niordiana* sp. nov., two new species isolated from seawater. *Syst. Appl. Microbiol.* 43:126040. doi: 10.1016/j.syapm.2019.126040

Dobson, S. J., and Franzmann, P. D. (1996). Unification of the genera *Deleya* (Baumann et al. 1983), *Halomonas* (Vreeland et al. 1980), and *Halovibrio* (Fendrich 1988) and the species *Paracoccus halodenitrificans* (Robinson and Gibbons 1952) into a single genus, *Halomonas*, and placement of the genus *Zymobacter* in the family *Halomonadaceae*. *Int. J. Syst. Bacteriol.* 46, 550–558. doi: 10.1099/00207713-46-2-550

Dobson, S. J., McMeekin, T. A., and Franzmann, P. D. (1993). Phylogenetic relationships between some members of the genera *Deleya*, *Halomonas*, and *Halovibrio*. *Int. J. Syst. Bacteriol.* 43, 665–673. doi: 10.1099/00207713-43-4-665

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

Eren, A. M., Kiefl, E., Shaiber, A., Veseli, I., Miller, S. E., Schechter, M. S., et al. (2021). Community-led, integrated, reproducible multi-omics with anvi'o. *Nat. Microbiol.* 6, 3–6. doi: 10.1038/s41564-020-00834-3

Franzmann, P. D., Burton, H. R., and Mcmeekin, T. A. (1987). *Halomonas subglaciescola*, a new species of halotolerant bacteria isolated from Antarctica. *Int. J. Syst. Bacteriol.* 37, 27–34. doi: 10.1099/00207713-37-1-27

Franzmann, P. D., Wehmeyer, U., and Stackebrandt, E. (1988). *Halomonadaceae* fam. nov., a new family of the class *Proteobacteria* to accommodate the genera *Halomonas* and *Deleya*. *Syst. Appl. Microbiol.* 11, 16–19. doi: 10.1016/S0723-2020(88)80043-2

Gaboyer, F., Vandenabeele-Trambouze, O., Cao, J., Ciobanu, M.-C., Jebbar, M., Le Romancer, M., et al. (2014). Physiological features of *Halomonas lionensis* sp. nov., a novel bacterium isolated from a Mediterranean Sea sediment. *Res. Microbiol.* 165, 490–500. doi: 10.1016/j.resmic.2014.07.009

Galisteo, C., de la Haba, R. R., Sánchez-Porro, C., and Ventosa, A. (2022). Biotin pathway in novel *Fodinibius salsisoli* sp. nov., isolated from hypersaline soils and reclassification of the genus *Aliifodinibius* as *Fodinibius*. *Front. Microbiol.* 13:1101464. doi: 10.3389/fmicb.2022.1101464

Gan, L., Long, X., Zhang, H., Hou, Y., Tian, J., Zhang, Y., et al. (2018). *Halomonas saliphila* sp. nov., a moderately halophilic bacterium isolated from a saline soil. *Int. J. Syst. Evol. Microbiol.* 68, 1153–1159. doi: 10.1099/ijsem.0.002644

Gao, P., Lu, H., Xing, P., and Wu, Q. L. (2020). *Halomonas rituensis* sp. nov. and *Halomonas zhuhanensis* sp. nov., isolated from natural salt marsh sediment on the Tibetan plateau. *Int. J. Syst. Evol. Microbiol.* 70, 5217–5225. doi: 10.1099/ijsem.0.004395

Garrity, G. M., Bell, J. A., and Lilburn, T. (2005a). "Class III. *Gammaproteobacteria* class. nov." in *Bergey's Manual of Systematic Bacteriology*, eds. D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (New York: Springer), 1.

Garrity, G. M., Bell, J. A., and Lilburn, T. (2005b). "Family I. *Oceanospirillaceae* fam. nov." in *Bergey's Manual of Systematic Bacteriology*. eds. D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (New York: Springer), 271.

Garrity, G. M., Bell, J. A., and Lilburn, T. (2005c). "Order VIII. *Oceanospirillales* ord. nov." in *Bergey's Manual of Systematic Bacteriology*, eds. D. J. Brenner, N. R. Krieg, J. T. Staley, and G. M. Garrity (New York: Springer), 270. doi: 10.1007/0-387-28022-7_8

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

Guan, T.-W., Xiao, J., Zhao, K., Luo, X.-X., Zhang, X.-P., and Zhang, L.-L. (2010). *Halomonas xinjiangensis* sp. nov., a halotolerant bacterium isolated from a salt lake. *Int. J. Syst. Evol. Microbiol.* 60, 349–352. doi: 10.1099/ijs.0.011593-0

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

Guzmán, D., Quillaguamán, J., Muñoz, M., and Hatti-Kaul, R. (2010). *Halomonas andesensis* sp. nov., a moderate halophile isolated from the saline Lake Laguna Colorada in Bolivia. *Int. J. Syst. Evol. Microbiol.* 60, 749–753. doi: 10.1099/ijs.0.014522-0

Heyrman, J., Balcaen, A., De Vos, P., and Swings, J. (2002). *Halomonas muralis* sp. nov., isolated from microbial biofilms colonizing the walls and murals of the saint-Catherine chapel (castle Herberstein, Austria). *Int. J. Syst. Evol. Microbiol.* 52, 2049–2054. doi: 10.1099/00207713-52-6-2049

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

Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119

Jiang, J., Pan, Y., Hu, S., Zhang, X., Hu, B., Huang, H., et al. (2014). *Halomonas songnenensis* sp. nov., a moderately halophilic bacterium isolated from saline and alkaline soils. *Int. J. Syst. Evol. Microbiol.* 64, 1662–1669. doi: 10.1099/ijs.0.056499-0

Jiang, J., Pan, Y., Meng, L., Hu, S., Zhang, X., Hu, B., et al. (2013). *Halomonas zhaodongensis* sp. nov., a slightly halophilic bacterium isolated from saline-alkaline soils in Zhaodong, China. *Antonie van Leeuwenhoek* 104, 685–694. doi: 10.1007/s10482-013-9976-3

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

Kämpfer, P., Rekha, P. D., Busse, H.-J., Arun, A. B., Priyanka, P., and Glaeser, S. P. (2018). *Halomonas malpeensis* sp. nov., isolated from rhizosphere sand of a coastal sand dune plant. *Int. J. Syst. Evol. Microbiol.* 68, 1037–1046. doi: 10.1099/ijsem.0.002616

Kaye, J. Z., Márquez, M. C., Ventosa, A., and Baross, J. (2004). *Halomonas neptunia* sp. nov., *Halomonas sulfidaeris* sp. nov., *Halomonas axialensis* sp. nov. and *Halomonas hydrothermalis* sp. nov.: halophilic bacteria isolated from deep-sea hydrothermal-vent environments. *Int. J. Syst. Evol. Microbiol.* 54, 499–511. doi: 10.1099/ijs.0.02799-0

Kazemi, E., Tarhriz, V., Amoozegar, M. A., and Hejazi, M. S. (2021). *Halomonas azerbaijanica* sp. nov., a halophilic bacterium isolated from Urmia Lake after the 2015 drought. *Int. J. Syst. Evol. Microbiol.* 71:4578. doi: 10.1099/ijsem.0.004578

Kim, K. K., Jin, L., Yang, H. C., and Lee, S. T. (2007). *Halomonas gomseomensis* sp. nov., *Halomonas janggokensis* sp. nov., *Halomonas salaria* sp. nov. and *Halomonas denitrificans* sp. nov., moderately halophilic bacteria isolated from saline water. *Int. J. Syst. Evol. Microbiol.* 57, 675–681. doi: 10.1099/ijs.0.64767-0

Kim, K. K., Lee, K. C., Oh, H.-M., and Lee, J.-S. (2010). *Halomonas stevensii* sp. nov., *Halomonas hamiltonii* sp. nov. and *Halomonas johnsoniae* sp. nov., isolated from a renal care centre. *Int. J. Syst. Evol. Microbiol.* 60, 369–377. doi: 10.1099/ijs.0.004424-0

Kim, M.-S., Roh, S. W., and Bae, J.-W. (2010). *Halomonas jeotgali* sp. nov., a new moderate halophilic bacterium isolated from a traditional fermented seafood. *J. Microbiol.* 48, 404–410. doi: 10.1007/s12275-010-0032-y

Lalucat, J., Mulet, M., Gomila, M., and García-Valdés, E. (2020). Genomics in bacterial taxonomy: impact on the genus *Pseudomonas*. *Genes (Basel)* 11:139. doi: 10.3390/genes11020139

Lee, J.-C., Jeon, C. O., Lim, J.-M., Lee, S.-M., Lee, J.-M., Song, S.-M., et al. (2005). *Halomonas taeanensis* sp. nov., a novel moderately halophilic bacterium isolated from a solar saltern in Korea. *Int. J. Syst. Evol. Microbiol.* 55, 2027–2032. doi: 10.1099/ijs.0.63616-0

Lee, J.-C., Kim, Y.-S., Yun, B.-S., and Whang, K.-S. (2015). *Halomonas salicampi* sp. nov., a halotolerant and alkalitolerant bacterium isolated from a saltern soil. *Int. J. Syst. Evol. Microbiol.* 65, 4792–4799. doi: 10.1099/ijsem.0.000650

Li, X., Gan, L., Hu, M., Wang, S., Tian, Y., and Shi, B. (2020). *Halomonas pellis* sp. nov., a moderately halophilic bacterium isolated from wetsalted hides. *Int. J. Syst. Evol. Microbiol.* 70, 5417–5424. doi: 10.1099/ijsem.0.004426

Liu, L., Wang, S.-Y., He, C.-F., Zhang, X.-X., Chi, M., Liang, L.-X., et al. (2020). *Phytohalomonas tamaricis* gen. nov., sp. nov., an endophytic bacterium isolated from *Tamarix ramosissima* roots growing in Kumtag desert. *Arch. Microbiol.* 202, 143–151. doi: 10.1007/s00203-019-01724-x

Long, M.-R., Zhang, D.-F., Yang, X.-Y., Zhang, X.-M., Zhang, Y.-G., Zhang, Y.-M., et al. (2013). *Halomonas nanhaiensis* sp. nov., a halophilic bacterium isolated from a sediment sample from the South China Sea. *Antonie van Leeuwenhoek* 103, 997–1005. doi: 10.1007/s10482-013-9879-3

Lu, H., Xing, P., Zhai, L., Li, H., and Wu, Q. (2020). *Halomonas montanilacus* sp. nov., isolated from hypersaline Lake Pengyanco on the Tibetan plateau. *Int. J. Syst. Evol. Microbiol.* 70, 2859–2866. doi: 10.1099/ijsem.0.004109

Martínez-Cánovas, M. J., Béjar, V., Martínez-Checa, F., and Quesada, E. (2004). *Halomonas anticariensis* sp. nov., from Fuente de Piedra, a saline-wetland wildfowl reserve in Málaga, southern Spain. *Int. J. Syst. Evol. Microbiol.* 54, 1329–1332. doi: 10.1099/ijs.0.63108-0

Meier-Kolthoff, J. P., Auch, A. F., Klenk, H.-P., and Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics* 14:60. doi: 10.1186/1471-2105-14-60

Mellado, E., Moore, E. R., Nieto, J. J., and Ventosa, A. (1995). Phylogenetic inferences and taxonomic consequences of 16S ribosomal DNA sequence comparison of *Chromohalobacter marismortui*, *Volcaniella eurihalina*, and *Deleya salina* and reclassification of *V. eurihalina* as *Halomonas eurihalina* comb. nov. *Int. J. Syst. Bacteriol.* 45, 712–716. doi: 10.1099/00207713-45-4-712

Menes, R. J., Viera, C. E., Farías, M. E., and Seufferheld, M. J. (2011). *Halomonas vilamensis* sp. nov., isolated from high-altitude Andean lakes. *Int. J. Syst. Evol. Microbiol.* 61, 1211–1217. doi: 10.1099/ijs.0.023150-0

Ming, H., Ji, W.-L., Li, M., Zhao, Z.-L., Cheng, L.-J., Niu, M.-M., et al. (2020). *Halomonas lactosivorans* sp. nov., isolated from salt-lake sediment. *Int. J. Syst. Evol. Microbiol.* 70, 3504–3512. doi: 10.1099/ijsem.0.004209

Minh, B. Q., Dang, C. C., Vinh, L. S., and Lanfear, R. (2021). QMaker: fast and accurate method to estimate empirical models of protein evolution. *Syst. Biol.* 70, 1046–1060. doi: 10.1093/sysbio/syab010

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

Mormile, M. R., Romine, M. F., Garcia, M. T., Ventosa, A., Bailey, T. J., and Peyton, B. M. (1999). *Halomonas campisalis* sp. nov., a denitrifying, moderately haloalkaliphilic bacterium. *Syst. Appl. Microbiol.* 22, 551–558. doi: 10.1016/S0723-2020(99)80008-3

Navarro-Torre, S., Carro, L., Rodríguez-Llorente, I. D., Pajuelo, E., Caviedes, M. Á., Igual, J. M., et al. (2020). *Halomonas radicis* sp. nov., isolated from *Arthrocnemum macrostachyum* growing in the Odiel marshes(Spain) and emended descriptions of *Halomonas xinjiangensis* and *Halomonas zincidurans*. *Int. J. Syst. Evol. Microbiol.* 70, 220–227. doi: 10.1099/ijsem.0.003742

Novichkov, P. S., Omelchenko, M. V., Gelfand, M. S., Mironov, A. A., Wolf, Y. I., and Koonin, E. V. (2004). Genome-wide molecular clock and horizontal gene transfer in bacterial evolution. *J. Bacteriol.* 186, 6575–6585. doi: 10.1128/JB.186.19.6575-6585.2004

Ntougias, S., Zervakis, G. I., Fasseas, C., and Fasseas, C. (2007). *Halotalea alkalilenta* gen. nov., sp. nov., a novel osmotolerant and alkalitolerant bacterium from alkaline olive mill wastes, and emended description of the family *Halomonadaceae* Franzmann *et al.* 1989, emend. Dobson and Franzmann 1996. *Int. J. Syst. Evol. Microbiol.* 57, 1975–1983. doi: 10.1099/ijs.0.65078-0

Oguntoyinbo, F. A., Cnockaert, M., Cho, G.-S., Kabisch, J., Neve, H., Bockelmann, W., et al. (2018). *Halomonas nigrificans* sp. nov., isolated from cheese. *Int. J. Syst. Evol. Microbiol.* 68, 371–376. doi: 10.1099/ijsem.0.002515

Oren, A., Arahal, D. R., Göker, M., Moore, E. R. B., Rossello-Mora, R., and Sutcliffe, I. C. (2023). International Code of Nomenclature of Prokaryotes. Prokaryotic Code (2022 Revision). *Int. J. Syst. Evol. Microbiol.* 73:005585. doi: 10.1099/ijsem.0.005585

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

Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P. A., Woodcroft, B. J., Evans, P. N., et al. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat. Microbiol.* 2, 1533–1542. doi: 10.1038/s41564-017-0083-5

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

Pritchard, L., Glover, R. H., Humphris, S., Elphinstone, J. G., and Toth, I. K. (2016). Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Anal. Methods* 8, 12–24. doi: 10.1039/C5AY02550H

Prjibelski, A., Antipov, D., Meleshko, D., Lapidus, A., and Korobeynikov, A. (2020). Using SPAdes de novo assembler. *Curr. Protoc. Bioinformatics* 70:e102. doi: 10.1002/cpbi.102

Qiu, X., Cao, X., Xu, G., Wu, H., and Tang, X. (2021a). *Halomonas maris* sp. nov., a moderately halophilic bacterium isolated from sediment in the southwest Indian Ocean. *Arch. Microbiol.* 203, 3279–3285. doi: 10.1007/s00203-021-02317-3

Qiu, X., Yu, L., Cao, X., Wu, H., Xu, G., and Tang, X. (2021b). *Halomonas sedimenti* sp. nov., a halotolerant bacterium isolated from deep-sea sediment of the southwest Indian Ocean. *Curr. Microbiol.* 78, 1662–1669. doi: 10.1007/s00284-021-02425-9

Quang, L. S., Gascuel, O., and Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24, 2317–2323. doi: 10.1093/bioinformatics/btn445

Quillaguamán, J., Hatti-Kaul, R., Mattiasson, B., Alvarez, M. T., and Delgado, O. (2004). *Halomonas boliviensis* sp. nov., an alkalitolerant, moderate halophile isolated from soil around a Bolivian hypersaline lake. *Int. J. Syst. Evol. Microbiol.* 54, 721–725. doi: 10.1099/ijs.0.02800-0

Reddy, G. S. N., Raghavan, P. U. M., Sarita, N. B., Prakash, J. S. S., Nagesh, N., Delille, D., et al. (2003). *Halomonas glaciei* sp. nov. isolated from fast ice of Adelie Land Antarctica. *Extremophiles* 7, 55–61. doi: 10.1007/s00792-002-0295-2

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

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

Romano, I., Nicolaus, B., Lama, L., Manca, M. C., and Gambacorta, A. (1996). Characterization of a haloalkalophilic strictly aerobic bacterium, isolated from Pantelleria island. *Syst. Appl. Microbiol.* 19, 326–333. doi: 10.1016/S0723-2020(96)80059-2

Sánchez-Porro, C., Kaur, B., Mann, H., and Ventosa, A. (2010). *Halomonas titanicae* sp. nov., a halophilic bacterium isolated from the RMS titanic. *Int. J. Syst. Evol. Microbiol.* 60, 2768–2774. doi: 10.1099/ijs.0.020628-0

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, baaa062. doi: 10.1093/database/baaa062,

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Stat.* 6, 461–464. doi: 10.1214/aos/1176344136

Seck, E. H., Fournier, P.-E., Raoult, D., and Khelaifia, S. (2016). "*Halomonas massiliensis*" sp. nov., a new halotolerant bacterium isolated from the human gut. *New Microb. New Infect.* 14, 19–20. doi: 10.1016/j.nmni.2016.07.013

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Shapovalova, A. A., Khijniak, T. V., Tourova, T. P., and Sorokin, D. Y. (2009). *Halomonas chromatireducens* sp. nov., a new denitrifying facultatively haloalkaliphilic bacterium from solonchak soil capable of aerobic chromate reduction. *Microbiology (English translation of Mikrobiologiia)* 78, 102–111. doi: 10.1134/S0026261709010135

Shi, W., Sun, Q., Fan, G., Hideaki, S., Moriya, O., Itoh, T., et al. (2021). gcType: a high-quality type strain genome database for microbial phylogenetic and functional research. *Nucleic Acids Res.* 49, D694–D705. doi: 10.1093/nar/gkaa957

So, Y., Chhetri, G., Kim, I., Kang, M., Kim, J., Lee, B., et al. (2022). *Halomonas antri* sp. nov., a carotenoid-producing bacterium isolated from surface seawater. *Int. J. Syst. Evol. Microbiol.* 72:5272. doi: 10.1099/ijsem.0.005272

Sorokin, D. Y., and Tindall, B. J. (2006). The status of the genus name *Halovibrio* Fendrich 1989 and the identity of the strains *Pseudomonas halophila* DSM 3050 and *Halomonas variabilis* DSM 3051. Request for an opinion. *Int. J. Syst. Evol. Microbiol.* 56, 487–489. doi: 10.1099/ijs.0.63965-0

Stott, C. M., and Bobay, L.-M. (2020). Impact of homologous recombination on core genome phylogenies. *BMC Genomics* 21:829. doi: 10.1186/s12864-020-07262-x

Tsai, M.-H., Liu, Y.-Y., Soo, V.-W., and Chen, C.-C. (2019). A new genome-to-genome comparison approach for large-scale revisiting of current microbial taxonomy. *Microorganisms* 7:161. doi: 10.3390/microorganisms7060161

van Dongen, S. (2008). Graph clustering via a discrete uncoupling process. *SIAM J. Matrix Anal. Appl.* 30, 121–141. doi: 10.1137/040608635

Ventosa, A., de la Haba, R. R., Arahal, D. R., and Sánchez-Porro, C. (2021a). *"Halomonas" in Bergey's Manual of Systematics of Archaea and Bacteria*, ed. A. Oren (Hoboken, NJ: Wiley), 1–111. doi: 10.1002/9781118960608.gbm01190.pub2

Ventosa, A., de la Haba, R. R., Arahal, D. R., and Sánchez-Porro, C. (2021b). "*Halomonadaceae*" in *Bergey's Manual of Systematics of Archaea and Bacteria*, ed. A. Oren (Hoboken, NJ: Wiley), 1–10. doi: 10.1002/9781118960608.fbm00228.pub2

Vidal, L. M. R., Gonçalves, A., Venas, T. M., Campeão, M. E., Calegario, G., Walter, J. M., et al. (2019). *Halomonas coralii* sp. nov. isolated from *Mussismilia braziliensis*. *Curr. Microbiol.* 76, 678–680. doi: 10.1007/s00284-019-01674-z

Wang, Y.-N., Cai, H., Yu, S.-L., Wang, Z.-Y., Liu, J., and Wu, X.-L. (2007). *Halomonas gudaonensis* sp. nov., isolated from a saline soil contaminated by crude oil. *Int. J. Syst. Evol. Microbiol.* 57, 911–915. doi: 10.1099/ijs.0.64826-0

Wang, L., Liu, X., Lai, Q., Gu, L., and Shao, Z. (2021). *Halomonas diversa* sp. nov., isolated from deep-sea sediment of the Pacific Ocean. *Int. J. Syst. Evol. Microbiol.* 71:4790. doi: 10.1099/ijsem.0.004790

Wang, L., and Shao, Z. (2021). Aerobic denitrification and heterotrophic sulfur oxidation in the genus *Halomonas* revealed by six novel species characterizations and genome-based analysis. *Front. Microbiol.* 12:652766. doi: 10.3389/fmicb.2021.652766

Wang, Y., Tang, S.-K., Lou, K., Mao, P.-H., Jin, X., Jiang, C.-L., et al. (2008). *Halomonas lutea* sp. nov., a moderately halophilic bacterium isolated from a salt lake. *Int. J. Syst. Evol. Microbiol.* 58, 2065–2069. doi: 10.1099/ijs.0.65436-0

Wang, F., Wan, J.-J., Zhang, X.-Y., Xin, Y., Sun, M.-L., Wang, P., et al. (2022). *Halomonas profundi* sp. nov., isolated from deep-sea sediment of the Mariana trench. *Int. J. Syst. Evol. Microbiol.* 72:5210. doi: 10.1099/ijsem.0.005210

Wang, T., Wei, X., Xin, Y., Zhuang, J., Shan, S., and Zhang, J. (2016). *Halomonas lutescens* sp. nov., a halophilic bacterium isolated from a lake sediment. *Int. J. Syst. Evol. Microbiol.* 66, 4697–4704. doi: 10.1099/ijsem.0.001413

Wang, Y.-X., Xiao, W., Dong, M.-H., Zhao, Q., Li, Z.-Y., Lai, Y.-H., et al. (2014). *Halomonas qiaohouensis* sp. nov., isolated from salt mine soil in southwest China. *Antonie van Leeuwenhoek* 106, 253–260. doi: 10.1007/s10482-014-0189-1

Wenting, R., Montazersaheb, S., Khan, S. A., Kim, H. M., Tarhriz, V., Hejazi, M. A., et al. (2021). *Halomonas azerica* sp. nov., isolated from Urmia Lake in Iran. *Curr. Microbiol.* 78, 3299–3306. doi: 10.1007/s00284-021-02482-0

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

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., et al. (2019). Welcome to the Tidyverse. *J. Open Source Softw.* 4:1686. doi: 10.21105/joss.01686

Wu, D., Jospin, G., and Eisen, J. A. (2013). Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. *PLoS One* 8:e77033. doi: 10.1371/journal.pone.0077033

Wu, L., and Ma, J. (2019). The global catalogue of microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation. *Int. J. Syst. Evol. Microbiol.* 69, 895–898. doi: 10.1099/ijsem.0.003276

Xiao, Z., Shen, J., Wang, Z., Dong, F., and Zhao, J.-Y. (2021). *Halomonas bachuensis* sp. nov., isolated from Gobi soil. *Curr. Microbiol.* 78, 397–402. doi: 10.1007/s00284-020-02268-w

Xing, J., Gong, Q., Tang, L., Li, J., Fan, H., Wang, X., et al. (2021). *Halomonas salipaludis* sp. nov., isolated from a saline-alkali wetland soil. *Arch. Microbiol.* 203, 6033–6039. doi: 10.1007/s00203-021-02560-8

Xu, X.-W., Wu, Y.-H., Zhou, Z., Wang, C.-S., Zhou, Y.-G., Zhang, H.-B., et al. (2007). *Halomonas saccharevitans* sp. nov., *Halomonas arcis* sp. nov. and *Halomonas subterranea* sp. nov., halophilic bacteria isolated from hypersaline environments of China. *Int. J. Syst. Evol. Microbiol.* 57, 1619–1624. doi: 10.1099/ijs.0.65022-0

Xu, L., Xu, X.-W., Meng, F.-X., Huo, Y.-Y., Oren, A., Yang, J.-Y., et al. (2013). *Halomonas zincidurans* sp. nov., a heavy-metal-tolerant bacterium isolated from the deep-sea environment. *Int. J. Syst. Evol. Microbiol.* 63, 4230–4236. doi: 10.1099/ijs.0.051656-0

Xu, L., Ying, J.-J., Fang, Y.-C., Zhang, R., Hua, J., Wu, M., et al. (2021). *Halomonas populi* sp. nov. isolated from *Populus euphratica*. *Arch. Microbiol.* 204:86. doi: 10.1007/s00203-021-02704-w

Yan, F., Fang, J., Cao, J., Wei, Y., Liu, R., Wang, L., et al. (2020). *Halomonas piezotolerans* sp. nov., a multiple-stress-tolerant bacterium isolated from a deep-sea sediment sample of the New Britain Trench. *Int. J. Syst. Evol. Microbiol.* 70, 2560–2568. doi: 10.1099/ijsem.0.004069

Zhao, B., Wang, H., Mao, X., Li, R., Zhang, Y.-J., Tang, S., et al. (2012). *Halomonas xianhensis* sp. nov., a moderately halophilic bacterium isolated from a saline soil contaminated with crude oil. *Int. J. Syst. Evol. Microbiol.* 62, 173–178. doi: 10.1099/ijs.0.025627-0

Zheng, J., Wittouck, S., Salvetti, E., Franz, C. M. A. P., Harris, H. M. B., Mattarelli, P., et al. (2020). A taxonomic note on the genus *Lactobacillus*: description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of *Lactobacillaceae* and *Leuconostocaceae*. *Int. J. Syst. Evol. Microbiol.* 70, 2782–2858. doi: 10.1099/ijsem.0.004107

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 04 December 2023
DOI 10.3389/fmicb.2023.1230929

OPEN ACCESS

EDITED BY
Rafael R de la Haba,
University of Sevilla, Spain

REVIEWED BY
Heng-Lin Cui,
Jiangsu University, China
Satya P. Singh,
Saurashtra University, India

*CORRESPONDENCE
Savita Kerkar
✉ drsavitakerkar@gmail.com

RECEIVED 29 May 2023
ACCEPTED 17 November 2023
PUBLISHED 04 December 2023

CITATION
Gawas P and Kerkar S (2023) Bacterial diversity and community structure of salt pans from Goa, India.
Front. Microbiol. 14:1230929.
doi: 10.3389/fmicb.2023.1230929

Bacterial diversity and community structure of salt pans from Goa, India

Priti Gawas and Savita Kerkar*

School of Biological Sciences and Biotechnology, Goa University, Taleigao, Goa, India

In Goa, salt production from the local salt pans is an age-old practice. These salt pans harbor a rich diversity of halophilic microbes with immense biotechnological applications, as they tolerate extremely harsh conditions. Detecting the existence of these microbes by a metabarcoding approach could be a primary step to harness their potential. Three salt pans *viz.* Agarwado, Curca, and Nerul adjoining prominent estuaries of Goa were selected based on their unique geographical locations. The sediments of these salt pans were examined for their bacterial community and function by 16S rRNA amplicon-sequencing. These salt pans were hypersaline (400–450 PSU) and alkaline (pH 7.6–8.25), with 0.036–0.081 mg/L nitrite, 0.0031–0.016 mg/L nitrate, 6.66–15.81 mg/L sulfate, and 20.8–25.6 mg/L sulfide. The relative abundance revealed that the *Pseudomonadota* was dominant in salt pans of Nerul (13.9%), Curca (19.6%), and Agarwado (32.4%). The predominant genera in Nerul, Curca, and Agarwado salt pan sediments were *Rhodopirellula* (1.12%), *Sulfurivermis* (1.28%), and *Psychrobacter* (25.5%) respectively. The highest alpha diversity (Shannon-diversity Index) was observed in the Nerul salt pan (4.8) followed by Curca (4.3) and Agarwado (2.03). Beta diversity indicated the highest dissimilarity between Agarwado and the other two salt pans (0.73) *viz.* Nerul and Curca and the lowest dissimilarity was observed between Nerul and Curca salt pans (0.48). Additionally, in the Agarwado salt pan, 125 unique genera were detected, while in Nerul 119, and in Curca 28 distinct genera were noted. The presence of these exclusive microorganisms in a specific salt pan and its absence in the others indicate that the adjacent estuaries play a critical role in determining salt pan bacterial diversity. Further, the functional prediction of bacterial communities indicated the predominance of stress adaptation genes involved in osmotic balance, membrane modification, and DNA repair mechanisms. This is the first study to report the bacterial community structure and its functional genes in these three salt pans using Next-Generation Sequencing. The data generated could be used as a reference by other researchers across the world for bioprospecting these organisms for novel compounds having biotechnological and biomedical potential.

KEYWORDS

salt pan, halophilic, Bacteria, diversity, next generation sequencing

Introduction

Salt production is an aged-old traditional method designed by farmers called *Mittkars* for over 150 decades. Salt pans are habitats inhabiting a huge community of extremophiles, thriving in extreme salinity, temperature, pH, and minimum nutrient availability (Mani et al., 2012). Halophilic bacteria from these environments have developed strategies and

cellular systems that could cope with salt stress. Salt-in and salt-out are the two major strategies observed in halophiles to escape stress. In the salt-in strategy, the intracellular system is adapted to the salt concentration of the surrounding but the cellular energy consumed here is huge. Whereas, in salt-out strategies, organisms avoid adaptation and thereby energy loss by an accumulation of organic compatible solutes, also called a compatible solute strategy (Oren, 1999).

Halophilic bacteria from Goan salt pans are widely studied as a potent source of bioactive compounds (Tonima and Savita, 2011; Fernandes et al., 2019). The potential microorganisms in the salt pans of Goa have attracted the attention of the public as well as the researchers to conserve these habitats, *mittkars*, and their traditional knowledge of salt making. The culture-dependent approach examines a very minuscule microbial population. Modern methods such as high throughput sequencing and *in silico* analysis allow us to fill this gap (Janda and Abbott, 2007). The bacterial diversity of Agarwado, Curca, and Nerul salt pans and its comparative analysis has not been elucidated previously by next-generation sequencing. Hence, the present study is the first report accomplished to understand the diversity, distribution, and abundance with a comparative analysis of the bacterial community in these salt pans.

Materials and methods

Sample collection

Sediment samples were collected during the salt harvesting period from the crystallizer pond using a 5.5 cm diameter corer in sterile bags from the salt pans of Curca village (N15.45798, E73.88065) situated at Tiswadi taluka, Agarwado village (N15.640978, E73.76426) situated at Pernem taluka and Nerul village (N15.51979, E73.78909) situated at Bardez taluka (Figures 1A–E). Sediment sample refers to the sediments collected about 2 cm from a surface layer of salt pan. The sediment samples collected in triplicates from each salt pan site were mixed to obtain a single sediment sample. Sediment samples from Agarwado, Curca, and Nerul salt pans were indicated as AC, CC, and NC, respectively. Samples were transported in ice-cold condition to the laboratory and stored at −20°C until further processing. The temperature was recorded on-site using a thermometer of the three salt pan sediment samples. The pore water obtained after centrifugation of sediment samples was used to measure pH, salinity, nitrate, nitrite, sulfate, and sulfide. The pH was recorded by pH meter (Eutech, Germany) and the salinity was measured by refractometer (ATAGO, Japan). The nitrite and nitrate were estimated spectrophotometrically according to APHA (2012). Sulfate was estimated turbidometrically by measuring the precipitation of $BaSO_4$ (Clesceri et al., 1998). For the estimation of sulfide, the sample was fixed in 2% zinc acetate and was further analysed spectrophotometrically (Cline, 1969).

Environmental DNA extraction, library preparation, and Illumina sequencing

Environmental DNA was extracted from 1 g of each sediment sample using DNeasy PowerSoil Kit (Qiagen). The DNA obtained was checked for its concentration and quality using NanoDrop. The amplicon library was prepared using Nextera XT Index Kit (Illumina inc.) using bacterial specific forward primer 341F (GCCTACGGGNGGCWGCAG) and reverse primer 805R (ACTACHVGGGTATCTAATCC) for V3-V4 hypervariable region of 16S rRNA gene followed by adding Illumina adapters required for cluster generation. Amplicon libraries were purified by AMPureXP beads and quantified using a Qubit Fluorometer. Libraries were then loaded on MiSeq at 10–20 pM concentration for cluster generation and sequencing. The libraries were sequenced at Eurofins Genomics India Pvt. Ltd., Karnataka, India on the Illumina MiSeq platform using 2×300 bp chemistry to obtain paired-end reads.

FIGURE 1
Goan Salt pans in study **(A)** Location of Agarwado, Curca, and Nerul Salt pans in Goa. Aerial image of **(B)** Agarwado salt pan **(C)** Curca salt pan **(D)** Nerul salt pan [Source: Google earth pro, 2023] **(E)** Image of crystallizer pond under solar evaporation at Agarwado salt pan.

Sequence analysis by bioinformatics

Raw paired-end reads of AC, CC, and NC were analysed by Quantitative Insights Into Microbial Ecology 2 (QIIME 2), version 2–2023.7 (Bolyen et al., 2019). Reads obtained after sequencing were denoised via the q2-dada2 plugin (Callahan et al., 2016) which involves correction of Illumina sequenced errors, merging of reverse and forward paired-end reads, and clustering them into ASVs (amplicon sequence variants). The rarefaction curve was plotted using QIIME. ASVs obtained were further taxonomically classified into different taxonomic levels using the q2-feature-classifier plugin based on the SILVA (V.138) reference database (Quast et al., 2012). The Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUST2) method (Douglas et al., 2020) was utilized to predict the functional potential of the bacterial communities in the three salt pans mentioned above. The relative abundance of functional protein profiles in the three salt pans was represented in a heatmap constructed using PRIMER7.

Statistical analysis

The alpha (intra-sample) diversity indices [such as Simpson index, Shannon index, Menhinick's richness index, and Pielou's evenness (equitability)] were calculated at the genus level for the three sampling sites, similarly the beta diversity was measured as described by Whittaker (1960). Canonical Correspondence analysis (CCA) was done to understand the influence of environmental variables [Temperature (°C), pH, salinity (PSU), nitrate (mg/L), nitrite (mg/L), sulfate (mg/L), sulfide (mg/L)] on bacterial genera. Alpha diversity, beta diversity and CCA analysis were performed by using Paleontological statistics software (PAST), version 4.13 (Hammer et al., 2001).

Results

Sediment sample analysis

The physical parameters of sediment samples collected from Agarwado (AC), Curca (CC), and Nerul (NC) salt pans possessed temperatures of 35°C (in CC) and 36°C (AC and NC). Salinity observed was 450 PSU in AC and 400 PSU in CC and NC whereas the pH of the sediment samples was found to be 8.25 (AC), 7.6 (CC), and 8 (NC). The environmental DNA extracted from sediment samples AC, CC, and NC were measured by Nanodrop for its concentration and was found to be 28.6 ng/µl, 40.2 ng/µl, and 33.2 ng/µl, respectively.

Relative abundance of bacterial community

A total of 95,299, 104,637, and 240,563 reads were obtained in sediment samples of the Agarwado (AC), Curca (CC), and Nerul (NC) salt pan, respectively. The rarefaction curve of all three samples reached a plateau at this sequencing depth (Figure 2) indicating the information contained in the three samples captured the majority of abundant phylotypes.

In the bacterial community analysis of the AC sample, 42 phyla and 201 genera were identified. The abundant phyla such as *Pseudomonadota* (33%), *Bacillota* (22%), and "*Candidatus* Patescibacteria" (9.4%), were observed (Figure 3A). Major contribution at the class level was by *Gammaproteobacteria* (30.7%), and *Bacilli* (20.7%). At the order level, *Pseudomonadales*, and *Bacillales* were found to be abundant with 26.8 and 19.4% relative abundance, respectively. Predominant taxa observed at the family level were *Moraxellaceae* (26.5%) and *Planococcaceae* (18.9%) and the major identified genera were *Psychrobacter* (26%), *Halanaerobium* (3.4%), and *Antarcticibacterium* (1.5%) (Figure 3B).

In the CC sample, 50 phyla and 143 genera were noted. The predominant phyla that contributed to bacterial diversity include *Pseudomonadota* (20.8%), *Desulfobacterota* (11.2%), and *Bacteroidota* (9.8%), (Figure 3A). At the class level, *Gammaproteobacteria* (17%) and *Bacteroidia* (8.4%) were found dominant. At the order level, the predominance of *Desulfobulbales* (5%) and *Desulfobacterales* (2.8%) were detected. At the family level, *Desulfobulbaceae* (3.9%) was relatively abundant. Lastly, at the general level, *Sulfurivermis* (1.3%) and *Desulfatiglans* (1.1%) presented a major proportion (Figure 3B).

In the NC sediment sample, 53 phyla and 250 genera were recognized. The dominant bacterial phyla include *Pseudomonadota* (14.7%), *Planctomycetota* (13%) and *Chloroflexota* (11.7%) (Figure 3A). The abundant classes observed were *Anaerolineae* (10.7%) and *Gammaproteobacteria* (9%). At the order level, *Pirellulales* with 6.5% relative abundance was found to be the highest followed by *Desulfobacterales* (3.8%). At the family level, *Pirellulaceae* (6.5%), was the major family obtained. Further, the most abundant genus identified was *Rhodopirellula* (1.2%) (Figure 3B).

Correlation of bacterial genera with environmental parameters

The effect of environmental parameters such as temperature, pH, salinity, nitrite, nitrate, sulfate, and sulfide were assessed on the bacterial genera using CCA. The CCA tripod explained 88.67% of the variation in bacterial genera by environmental parameters (Figure 4). The result revealed the positive correlation between genera *Blastopirellula*, Sva0081 sediment group, *Rhodopirellula*, *Robiginitalea*, *Pirellula* and *Bacillus* with sulfide and sulfate. However, genera *Psychrobacter, Halanaerobium*, *Planococcus* and *Sphingobacterium* found abundant in AC were positively correlated with nitrite, nitrate, and salinity. Of all the environmental parameters sulfide and sulfate contributed significantly to bacterial genera in NC.

Alpha and beta diversity analysis at the genus level

The alpha diversity indices such as Simpson diversity index, Shannon diversity index (Shannon-Wiener index), Dominance, Menhinick's richness index, and Pielou's evenness (equitability) were calculated and depicted in Table 1. Similarly, the beta diversity and its pair-wise comparison presented in Table 2 were performed using the mathematical expression S/ā-1, where S is the total number of genera and ā is the average number of genera. The beta diversity was found highest and equal between AC and NC and AC and CC, i.e., 0.73. The lowest beta diversity was observed between CC and NC (0.48).

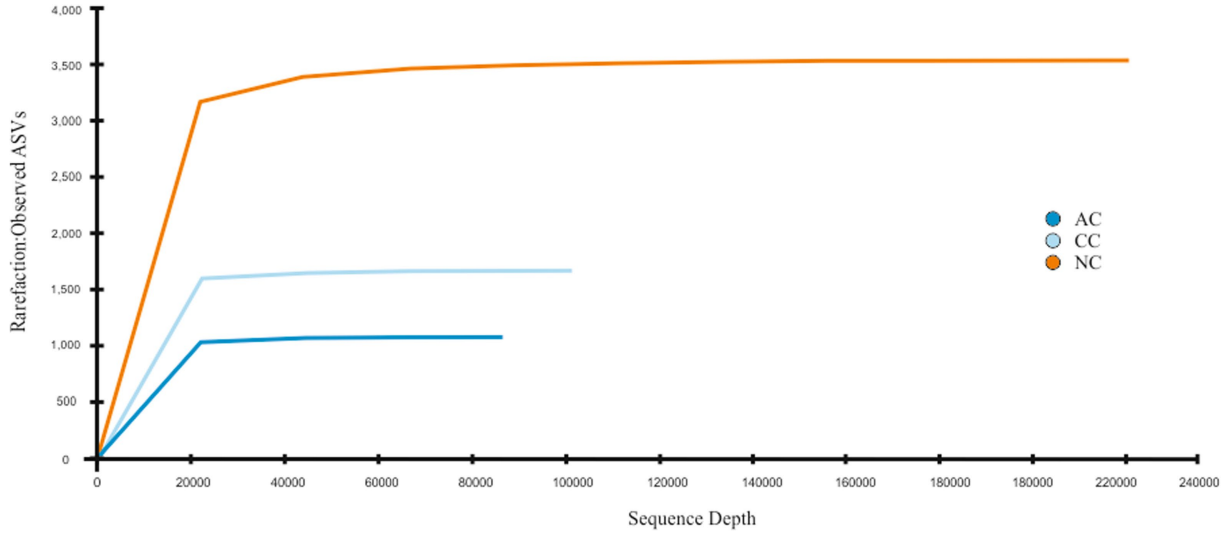


FIGURE 2
Rarefaction curve: Observed ASVs vs. Sequencing depth for three salt pan samples namely Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India.

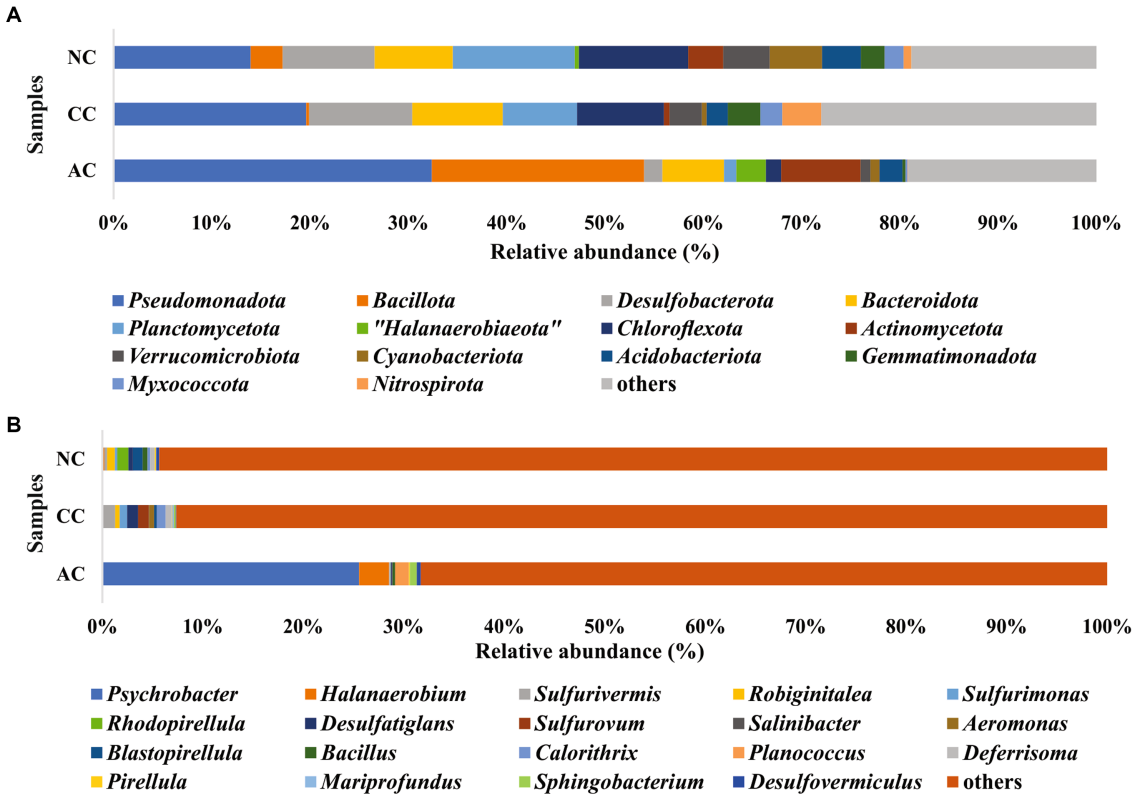


FIGURE 3
Relative abundance of bacterial communities at **(A)** phyla level and **(B)** genus level in sediment samples of Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India.

Unique and common bacterial communities among the salt pans

The Venn diagram (Figure 5A) represents the number of unique and common phyla in three salt pans. An acceptably small percentage of unique taxon was observed in AC, CC, and NC with 14.2, 6, and 5.6%, respectively. Contrastingly, the common taxa observed in all three salt pans were in a higher proportion ranging from 62 -78%. There were no common phyla found in CC and AC. The phyla found common among all three salt pans are listed in the Supplementary Table S1.

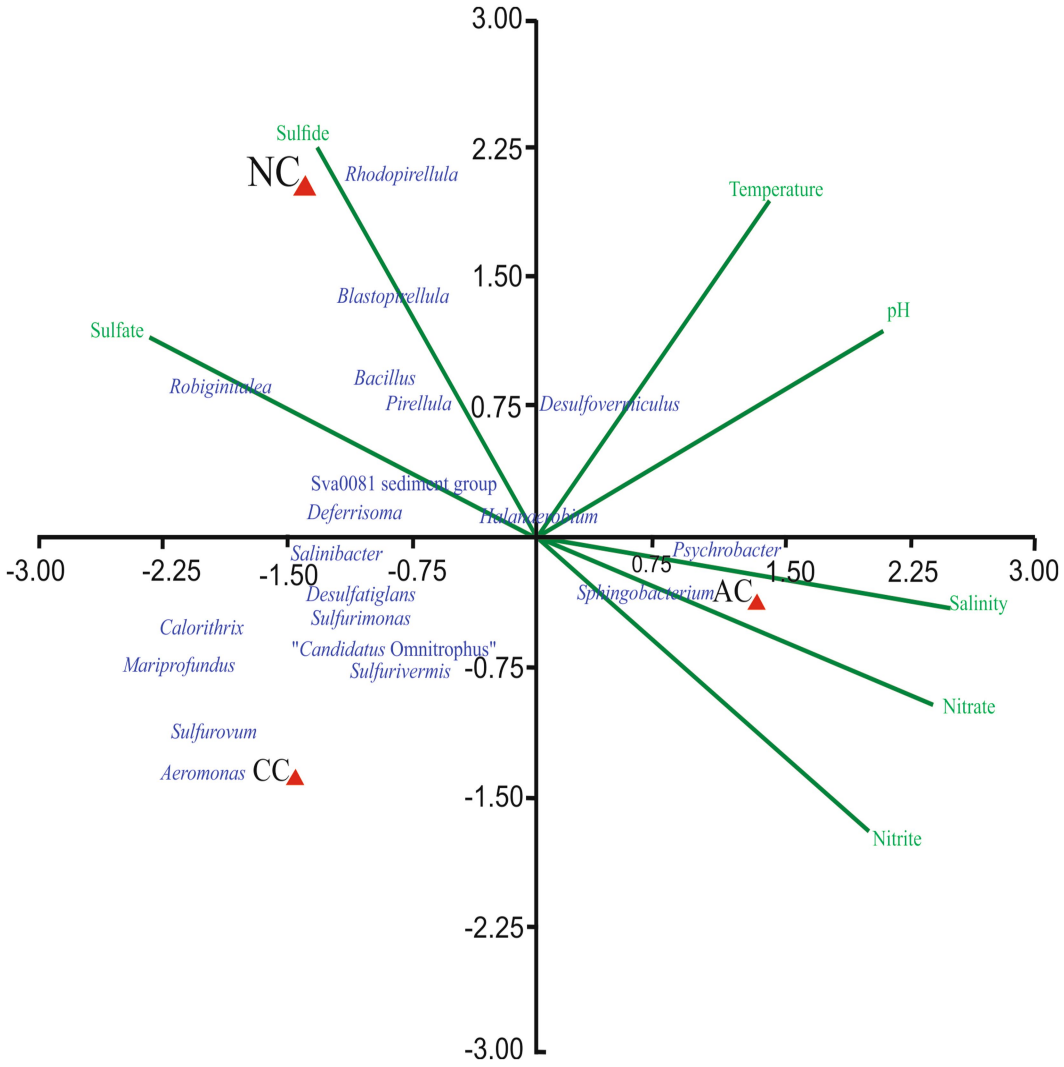


FIGURE 4
Canonical Correspondence Analysis ordination diagram of bacterial communities at generic-level at Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India associated with environmental parameters.

TABLE 1 Alpha diversity indices in Agarwado salt pan (AC), Curca salt pan (CC) and Nerul salt pan (NC) of Goa, India at genera level.

Alpha diversity indices	AC	CC	NC
Taxa (no. of genera)	201	143	250
Simpson diversity index	0.67	0.97	0.98
Pielou's evenness (equitability)	0.38	0.87	0.86
Menhinick's richness index	0.6	1	1
Dominance	0.32	0.02	0.01
Shannon diversity index	2.03	4.3	4.8

TABLE 2 Pair-wise comparison of beta diversity between Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India at the genus level measured by Whittaker using the mathematical expression S/ā-1, where S is the total number of genera and ā is the average number of genera.

	AC	CC	NC
AC	0	0.73	0.73
CC	0.73	0	0.48
NC	0.7	0.48	0

At the genera level, a total of 201, 143, 250 genera were observed in AC, CC, and NC, respectively. Although the number of genera was found highest in NC, the percentage of exclusivity was observed high in AC (61%) > NC (47%) > CC (19.4%). The common genera found in all three salt pans ranged between 13 - 23%. The Venn diagram (Figure 5B) depicts the number of unique and common genera in the three salt pans and the genera are listed in the Supplementary Table S2.

Rare taxa in the three salt pans

Rare taxa are defined as those having a relative abundance of ≤0.2% (Hugoni et al., 2013). At the phylum level, rare phyla such as

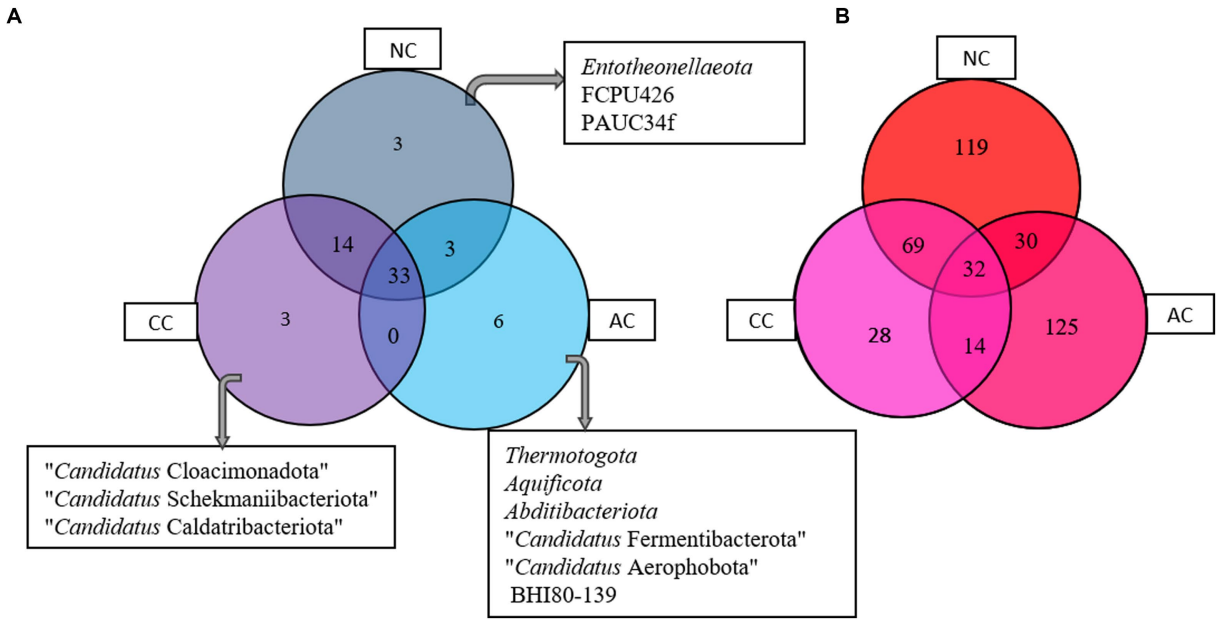


FIGURE 5
Venn diagram showing common and unique bacterial taxa at **(A)** phyla and **(B)** genus level between Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India.

Bdellovibrionota (0.06%), *Nitrospirota* (0.07%) and "*Candidatus* Sumerlaeota" (0.07%) were found in AC. In CC, "*Candidatus* Moduliflexota," *Fusobacteriota,* and "*Candidatus* Acetithermota" are rare taxa with 0.04, 0.07, and 0.06% relative abundance, respectively. Further in NC, "*Candidatus* Babelota," *Elusimicrobiota*, and *Deinococcota* are a few minor phyla with 0.11, 0.08, and 0.081% relative abundance, respectively. Secondly, "*Candidatus* Dadaibacteriota" was the only rare (≤ 0.2%) bacterial phyla found common in all three salt pans. Further, at the genus level only 11 taxa were detected common in all three salt pans (AC, CC, NC) with ≤0.2% of relative abundance (Figure 6).

Functional gene prediction in three salt pans

A large number of functional genes involved in various biological processes, including cellular metabolism and adaption to extreme environmental conditions, have been found in the indigenous bacterial communities of all the three salt pans. Figure 7 depicts the most abundant genes that were found in all the three salt pans. Further, numerous genes involved in the nitrogen, sulfur, and carbon cycles have also been found to be part of the overall functional gene profile (Supplementary Figures S1–S3).

Discussion

Traditional salt-making is one of the major and unique occupations of Goa due to its lengthy coastline of 107 Km. The three study sites *viz.* Agarwado, Curca, and Nerul salt pans were selected based on their unique geographical locations. The Agarwado salt pan is in Pernem taluka adjoining the pristine Chapora estuary in the extreme north of Goa. The Nerul and Curca salt pans are connected to the two major lifeline estuaries of Goa, i.e., Mandovi in Bardez taluka and Zuari in Tiswadi taluka, respectively. Seawater from the estuary gushes into the salt pans through a sluice gate during high tide and is evaporated to produce salt. For decades, these three distinct salterns have been conventionally used to produce local salt by traditional methods in Goa. The temperature of 35°C - 36°C, noted from these salt pans during the salt harvesting period favors the evaporation process in the crystallizer pond to achieve the salinity required for salt production. The salinity of these salt pans was found to be in the range of 400–450 PSU and pH 7.6–8.25 indicating a halophilic and alkaline nature.

Comparative analysis of the bacterial community in the salt pan

The bacterial composition of the salt pans of Goa revealed that the three sediment samples of salt pans were dominated by phylum *Pseudomonadota*. These results were parallel to the work reported by (Gorrasi et al., 2021) in other solar salterns. Within the *Pseudomonadota* phylum in the AC salt pan, only two classes such as *Gammaproteobacteria* and *Alphaproteobacteria* were observed. In addition to these two classes, *Zetaproteobacteria* was identified in CC and NC. Interestingly the only genera found in our study within the *Zetaproteobacteria* class was *Mariprofundus.* This class of bacterium is a model organism for microbial iron oxidation in marine environments (McAllister et al., 2019). The detection of genera *Mariprofundus* in CC and NC perhaps originated from the mining activities at Mandovi estuary which has deposited several tonnes of iron ore. Secondly, the Cumbarjua Canal interconnects the Mandovi

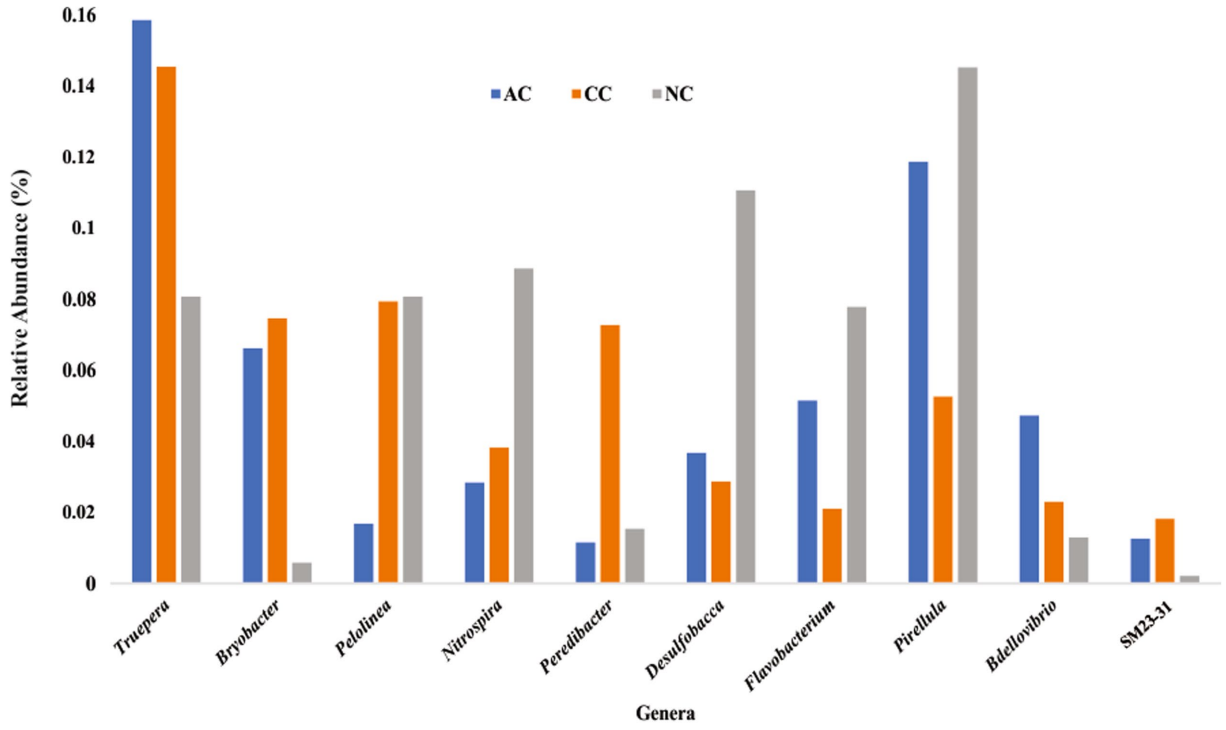


FIGURE 6
Rare bacterial genera (≤ 0.2% relative abundance) found common in all three sediment samples of Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India.

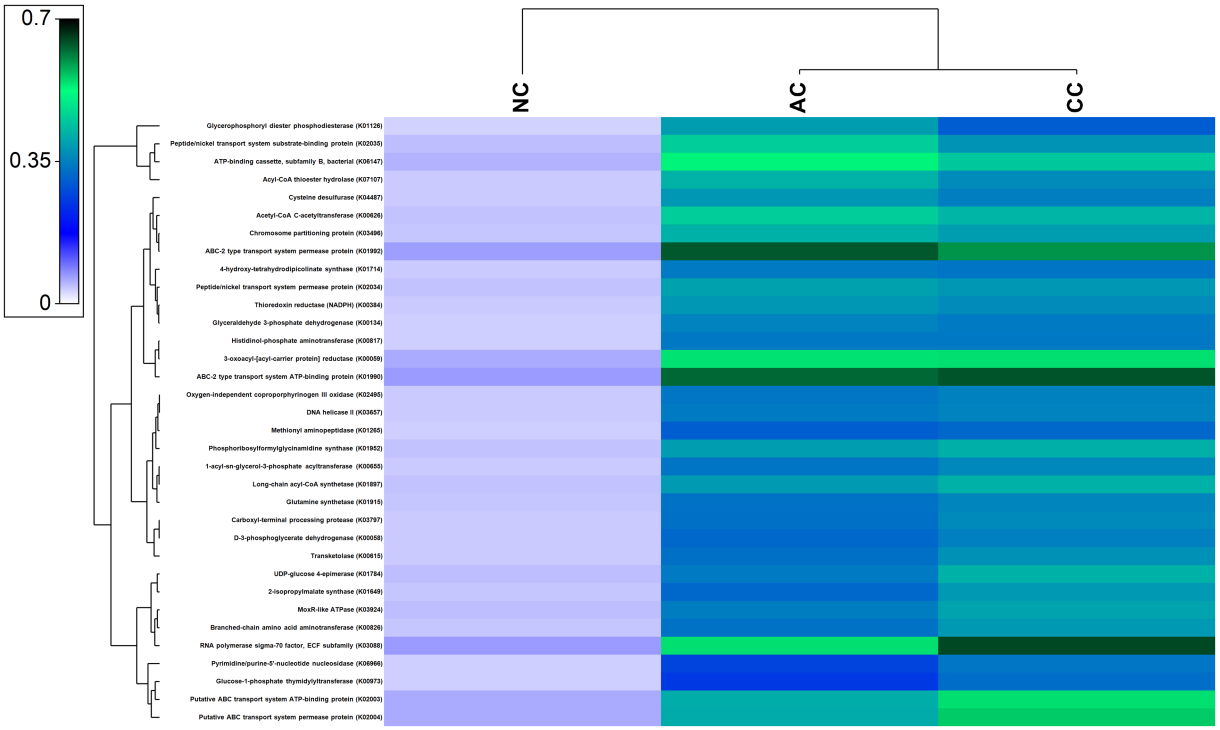


FIGURE 7
Heatmap showing the abundant functional genes predicted by using PICRUST2 algorithm in the sediment samples of Agarwado salt pan (AC), Curca salt pan (CC), and Nerul salt pan (NC) of Goa, India.

and Zuari estuaries (adjacent to NC and CC respectively) and forms a network that has served as a major route for the transportation of iron ore. Additionally, the discharge of mining rejects from the mining sites has also concentrated iron particles in these estuaries (Gauns, 2011). This demonstrates how the distribution of microorganisms in a particular area is affected by their chemical composition. Similarly, other environmental parameters like salinity, nitrite, nitrate, sulfate, and sulfide also have an influence on bacterial distribution and abundance as depicted by CCA. The dominant phyla observed in our study such as *Pseudomonadota, Bacteriodota*, *Chloroflexota,* etc. are found to have ubiquitous genetic machinery and survival weapons (Mehta et al., 2021). Hence, understanding the microbiome composition, evolving genetic machinery, and adaptive mechanisms evolved by these extremophiles make it important in biotechnology industries. The genera *Psychrobacter*, belonging to the family *Moraxellaceae*, was the most abundant genera and family, respectively, in the AC with approximately 25.5% relative abundance. Like our studies, the presence of *Psychrobacter* was also reported in the solar saltern of Bohai Bay, China, revealed by DGGE of the 16S gene (Zhang et al., 2016). The isolation of sulfidogenic bacteria *Halanaerobium,* reported from the Ribandar salt pan of Goa was observed to be the second abundant genera in AC (Das et al., 2017). The abundant genera observed in the CC and NC salt pans were *Sulfurivermis* and *Rhodopirellula, respectively.* The presence of *Sulfurivermis* shows the occurrence of active sulfur cycles in the salt pan being a sulfur oxidizing bacterium (SOB) that uses nitrate as an electron acceptor in the rTCA pathway (Kojima et al., 2017).

Alpha and Beta diversity analysis of the salt pans

The diversity of any ecosystem depends on two components "richness and evenness" (Soininen et al., 2012). In the Menhinicks richness index, an equal richness value was observed for NC and CC and was higher in comparison to AC, evident for higher community richness in Nerul and Curca salt pans. Similarly, the Pielou's evenness index for NC and CC was found to be higher as compared to AC, which indicates the distribution of an equal number of taxa in both salt pan (NC and CC) ecosystems than in the Agarwado salt pan. Further, dominance in AC was higher than NC and CC, demonstrating the presence of a more dominant taxon in the Agarwado salt pan community. Lower evenness and higher dominance observed in AC indicate lower alpha diversity. However, in Nerul and Curca salt pans high evenness and lower dominance indicate high alpha diversity. A similar observation was shown by the Shannon diversity index which considers both the richness and evenness of a community. The alpha diversity measured by the Shannon index was NC > CC > AC.

Interestingly equal dissimilarity diversity (beta diversity) was observed between AC and NC and AC and CC, which further indicated that both CC and NC are quite comparable to each other but highly divergent to AC. The similarity between NC and CC was further confirmed by the presence of high number of common genera in both (represented in Figure 5B) in addition to the beta diversity. This is due the adjoining estuaries at Nerul and Curca salt pans that are connected by the Cumbarjua canal and therefore share uniform environmental characteristic (Qasim and Gupta, 1981) influencing bacterial composition.

Bacterial taxa in ecological roles

Pseudomonadota, *Planctomycetota*, and *Chloroflexota* observed in the three salt pans are representatives of the commonly occurring phyla that are involved in a variety of ecological roles. The sulfur cycle advances in tandem with the carbon cycle and is most prevalent in hypersaline environments. Using sulfate as an electron acceptor, the sulfate-reducing bacteria break down organic molecules in an anoxic environment thereby cycling carbon anaerobically (Kerkar and Bharathi, 2007). In the current anticipated functional metagenome, genes involved in assimilatory and dissimilatory sulfate reduction and oxidation of sulfur compounds were found. The reductive nature of the salt pans sulfur cycle in the current work is indicated by the presence of the genera *Desulfovermiculus* and *Desulfatiglans* (Begmatov et al., 2021) in addition to genes such as adenylylsulfate reductase, subunits A (K00394), adenylylsulfate reductase, subunits B (K00395), dissimilatory sulfite reductase alpha subunit (K11180), dissimilatory sulfite reductase beta subunit (K11181), and sulfate adenylyltransferase (K00958). Likewise, different genes of the Sox system that are involved in sulfur oxidation: SoxA (K17222), SoxX (K17223), SoxB (K17224), SoxY (K17226), SoxZ (K17227), SoxC (K17225) and Sulfane dehydrogenase subunit were observed in the functional gene prediction (Supplementary Figure S3). *Thiohalospira* is a genus of halophilic, chemolithoautotrophic SOB that utilizes sulfur compounds as electron donors (Sorokin et al., 2020); it has only been found in the Agarwado (AC) salt pan. Secondly, the SOB genus, *Sulfuriflexus* (class: *Gammaproteobacteria*) (Kojima and Fukui, 2016; Rana et al., 2020), was exclusive to Curca (CC) sediments. *Sphingobacterium* and *Woeseia* functional in carbon-fixing (Lohmann et al., 2020) were also observed in these salt pans along with numerous genes required in carbon metabolism (Supplementary Figure S1). Unique to the Nerul (NC) salt pan, a group at genus level Cm1-21 (family: *Nitrosococcaceae*) was recently reported as ammonium oxidizers involved in the nitrification process of the nitrogen cycle (Semedo et al., 2021). Besides functional genes (Supplementary Figures S1–S3), bacterial genera like *Desulfovibrio* (observed in AC and NC sediment samples), *Pseudomonas* (AC and CC), *Nitrospina* (CC and NC), and *Nitrospira* (AC, CC, and NC) noted in our study are implicated in the carbon, nitrogen, and sulfur cycles (Megonigal et al., 2003; Meng et al., 2022; Sun et al., 2022), which illustrates that the movement of nutrients necessary to support the persistent life in these man-made salt pan ecosystems are governed by the presence of its indigenous microbes.

Significance of halophiles

Salinity is a key factor that governs the microbial community in hypersaline ecosystems such as solar saltern. High salinity imposes stress on microbial cells, challenging their survival (Fei Xi et al., 2014; Song et al., 2022). Numerous genes associated with osmoprotective mechanisms, such as salt-in and salt-out mechanisms, membrane modification, protein adaptation, DNA repair, etc. acquired by residents of hypersaline environment for normal growth and development of cells were observed to be prevalent in our study. These genes include ABC-2 type transport system permease protein (K01992), ABC-2 type transport system ATP-binding protein (K01990), ATP-binding cassette, subfamily B, bacterial (K06147),

1-acyl-sn-glycerol-3-phosphate acyltransferase (K00655), 4-hydroxy-tetrahydrodipicolinate synthase (K01714), and DNA helicase II / ATP-dependent DNA helicase PcrA (K03657) (Sleator and Hill, 2002; Yu et al., 2018; Schmitz et al., 2020).

Salt pans in Goa are influenced by three seasons: pre-monsoon, monsoon, and post-monsoon, where salinity ranges from 30–300 PSU (Mani et al., 2012). Therefore, bacteria with different salt tolerance like slightly, moderately, and extremely halophilic bacteria are expected to inhabit this ecosystem. Bacteria within the moderately halophilic group are very important as they occupy a large proportion of halophiles. As adaptive measures, the production of compatible solutes and hydrolytic enzymes to acclimatize to salinity variations is attracting industries (Baati et al., 2010; Ramos-Cormenzana, 2020). For example, the moderate halophilic genus *Micrococcus* synthesizes catalase that plays a role in maintaining colour and flavor in cured meats (Onishi and Kamekura, 1972; García-López et al., 1999). Secondly, *Rhodopirellula,* an abundant genus observed in NC sample is a moderate halophile and was reported to have a role as a food supplement to *Daphina magna* (indicator used for testing toxicity of water) (da Conceição Marinho et al., 2019). Therefore, profound knowledge and discovery of halophilic bacteria in such an ecosystem will help us to disclose numerous properties, enzymes, and proteins of halophiles that could be substantially important to different biotechnological industries.

Conclusion

Salt pans being polyextremophilic, are dynamic habitats of the resident bacteria potentially capable of tolerating extremely stressful environmental factors, i.e., high salinity (30–450 PSU), high alkalinity (pH 7.6–8.25) and fluctuating temperatures (20°C – 45°C). This is the first study profiling the bacterial communities in the Agarwado, Curca, and Nerul salt pans of Goa. The data reveals that the highest bacterial genera prevail in the Nerul salt pan sediment (250) as compared to Agarwado (201) and Curca (143). However, in the sediments of Agarwado salt pan, 125 distinct genera were found, followed by Nerul (119) and Curca (28) which were exclusively unique to these salt pans. These hypersaline natives are resistant to osmotic stress, demonstrated by the presence of genes in the predicted functional gene profile, that allows them to actively participate in the carbon, nitrogen, and sulfur cycles in these hypersaline environments. A wide microbial community and its function remains unknown, as "unculturable" as it is impractically impossible to mimic these diverse physical, chemical, biological, and mechanical factors which contribute to simulating the salt pan ecosystem in a laboratory. The high throughput sequencing has made it possible to provide an insight of the bacterial community and most of its functions in the salt pans, which could not be explored earlier. Further, a deep understanding of their functional mechanisms in this habitat will help us to assess the existence of these unique bacterial communities, and the adaptive cellular machinery/ biomolecules involved in making this hypersaline environment a conducive environment to thrive in.

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

PG: writing original draft preparation, methodology, and software analysis. SK: supervision, experimental design, review and editing. All authors contributed to the article and approved the submitted version.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by a Junior Research Fellowship for a Ph.D. scholar (award letter number: 19/06/2016(i) EU-V), funded by University Grant Commission (UGC), New Delhi, India.

Acknowledgments

The authors are thankful to the School of Biological Sciences and Biotechnology (Discipline of Biotechnology) and Goa Business School (Discipline of Computer Science and Technology), Goa University for the necessary facilities. We are grateful to Venkatachalam S., Project Scientist II at NCPOR, Goa, India for the technical assistance provided.

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

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2023.1230929/full#supplementary-material

References

APHA (2012). *Standard methods for the examination of water and waste water. 22* American Public Health Association, American Water Work Association, Water Environment Federation. Washington

Baati, H., Amdouni, R., Gharsallah, N., Sghir, A., and Ammar, E. (2010). Isolation and characterization of moderately halophilic Bacteria from Tunisian solar Saltern. *Curr. Microbiol.* 60, 157–161. doi: 10.1007/s00284-009-9516-6

Begmatov, S., Savvichev, A. S., Kadnikov, V. V., Beletsky, A. V., Rusanov, I. I., Klyuvitkin, A. A., et al. (2021). Microbial communities involved in methane, sulfur, and nitrogen cycling in the sediments of the Barents Sea. *Microorganisms* 9:2362. doi: 10.3390/microorganisms9112362

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

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

Clesceri, L. S., Greenberg, A. E., and Eaton, A. D. (1998). Standard methods for examination of water and wastewater. *American Public Health Association,* Washington.

Cline, J. D. (1969). Spectrophotometric determination of hydrogen sulfide in natural waters1. *Limnol. Oceanogr.* 14, 454–458. doi: 10.4319/lo.1969.14.3.0454

da Conceição Marinho, M., Lage, O. M., Sousa, C. D., Catita, J., and Antunes, S. C. (2019). Assessment of Rhodopirellula rubra as a supplementary and nutritional food source to the microcrustacean *Daphnia magna*. *Antonie Van Leeuwenhoek* 112, 1231–1243. doi: 10.1007/s10482-019-01255-x

Das, K. R., Kerkar, S., Meena, Y., and Mishra, S. (2017). Effects of iron nanoparticles on iron-corroding bacteria 7, 385. doi: 10.1007/s13205-017-1018-9

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. doi: 10.1038/s41587-020-0548-6

Fei Xi, X., Wang, L., Jun Hu, J., Shu Tang, Y., Hu, Y., Hua Fu, X., et al. (2014). Salinity influence on soil microbial respiration rate of wetland in the Yangtze River estuary through changing microbial community. *J. Environ. Sci.* 26, 2562–2570. doi: 10.1016/j.jes.2014.07.016

Fernandes, S., Kerkar, S., Leitao, J., and Mishra, A. (2019). Probiotic role of salt Pan Bacteria in enhancing the growth of Whiteleg shrimp, *Litopenaeus vannamei*. *Probiotics Antimicrob Proteins* 11, 1309–1323. doi: 10.1007/s12602-018-9503-y

García-López, M.-L., Santos, J.-Á., and Otero, A. (1999). MICROCOCCUS. *Encyclopedia of Food Microbiology (Elsevier)*. 20, 1344–1350. doi: 10.1006/rwfm.1999.1025

Gauns, R. (2011). *Mining havoc: Impact of mining on water resources in Goa - article from dams, Rivers, and People.* SANDRP, 10–12.

Gorrasi, S., Franzetti, A., Ambrosini, R., Pittino, F., Pasqualetti, M., and Fenice, M. (2021). Spatio-temporal variation of the bacterial communities along a salinity gradient within a Thalassohaline environment (saline di Tarquinia Salterns, Italy). *Molecules* 26:1338. doi: 10.3390/molecules26051338

Hammer, D. A. T., Ryan, P. D., Hammer, Ø., and Harper, D. A. T. (2001). Past: paleontological statistics software package for education and data analysis. *Palae. Elec.* 4, 1–9.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., et al. (2013). Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proc. Natl. Acad. Sci.* 110, 6004–6009. doi: 10.1073/pnas.1216863110

Janda, J. M., and Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: pluses, perils, and pitfalls. *J. Clin. Microbiol.* 45, 2761–2764. doi: 10.1128/JCM.01228-07

Kerkar, S., and Bharathi, P. A. L. (2007). Stimulation of sulfate-reducing activity at salt-saturation in the salterns of Ribandar, Goa, India. *Geomicrobiol J.* 24, 101–110. doi: 10.1080/01490450701266597

Kojima, H., and Fukui, M. (2016). Sulfuriflexus mobilis gen. Nov., sp. nov., a sulfur-oxidizing bacterium isolated from a brackish lake sediment. *Int. J. Syst. Evol. Microbiol.* 66, 3515–3518. doi: 10.1099/ijsem.0.001227

Kojima, H., Watanabe, M., and Fukui, M. (2017). Sulfurivermis fontis gen. Nov., sp. nov., a sulfur-oxidizing autotroph, and proposal of Thioprofundaceae fam. Nov. *Int. J. Syst. Evol. Microbiol.* 67, 3458–3461. doi: 10.1099/ijsem.0.002137

Lohmann, P., Benk, S., Gleixner, G., Potthast, K., Michalzik, B., Jehmlich, N., et al. (2020). Seasonal patterns of dominant microbes involved in central nutrient cycles in the subsurface. *Microorganisms* 8:1694. doi: 10.3390/microorganisms8111694

Mani, K., Salgaonkar, B. B., Das, D., and Bragança, J. M. (2012). Community solar salt production in Goa. *India. Aquat Biosyst* 8:30. doi: 10.1186/2046-9063-8-30

McAllister, S. M., Moore, R. M., Gartman, A., Luther, G. W., Emerson, D., and Chan, C. S. (2019). The Fe(II)-oxidizing Zetaproteobacteria: historical, ecological and genomic perspectives. *FEMS Microbiol. Ecol.* 95:fiz015. doi: 10.1093/femsec/fiz015

Megonigal, J. P., Hines, M. E., and Visscher, P. T. (2003). Anaerobic metabolism: linkages to trace gases and aerobic processes. *Treatise on Geochemistry (Elsevier).* 8, 317–424.

Mehta, P., Yadav, M., Ahmed, V., Goyal, K., Pandey, R., and Chauhan, N. S. (2021). Culture-independent exploration of the hypersaline ecosystem indicates the environment-specific microbiome evolution. *Front. Microbiol.* 12:686549. doi: 10.3389/fmicb.2021.686549

Meng, S., Peng, T., Liu, X., Wang, H., Huang, T., Gu, J.-D., et al. (2022). Ecological role of Bacteria involved in the biogeochemical cycles of mangroves based on functional genes detected through GeoChip 5.0. *mSphere* 7:21. doi: 10.1128/msphere.00936-21

Onishi, H., and Kamekura, M. (1972). *Micrococcus halobius* sp. n. *Int. J. Syst. Evol. Microbiol.* 22, 233–236. doi: 10.1099/00207713-22-4-233

Oren, A. (1999). Bioenergetic aspects of Halophilism. *Microbiol. Mol. Biol. Rev.* 63, 334–348. doi: 10.1128/MMBR.63.2.334-348.1999

Qasim, S. Z., and Gupta, R. S. (1981). Environmental characteristics of the Mandovi-Zuari estuarine system in Goa. *Estuar. Coast. Shelf Sci.* 13, 557–578. doi: 10.1016/S0302-3524(81)80058-8

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. doi: 10.1093/nar/gks1219

Ramos-Cormenzana, A. (2020). Ecology of moderately halophilic Bacteria. *The Biology of Halophilic Bacteria Edition (CRC Press)*, 55–86. doi: 10.1201/9781003069140-3

Rana, K., Rana, N., and Singh, B. (2020). Applications of sulfur oxidizing bacteria. *Physiological and Biotechnological Aspects of Extremophiles*, (Academic Press), 131–136.

Schmitz, R. A., Dietl, A., Müller, M., Berben, T., Op den Camp, H. J., and Barends, T. R. (2020). Structure of the 4-hydroxy-tetrahydrodipicolinate synthase from the thermoacidophilic methanotroph Methylacidiphilum fumariolicum SolV and the phylogeny of the aminotransferase pathway. *Acta Crystallogr, Sect F: Struct Biol Commun* 76, 199–208. doi: 10.1107/S2053230X20005294

Semedo, M., Lopes, E., Baptista, M. S., Oller-Ruiz, A., Gilabert, J., Tomasino, M. P., et al. (2021). Depth profile of nitrifying archaeal and bacterial communities in the remote oligotrophic waters of the North Pacific. *Front. Microbiol.* 12, 1–18. doi: 10.3389/fmicb.2021.624071

Sleator, R. D., and Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiol. Rev.* 26, 49–71. doi: 10.1111/j.1574-6976.2002.tb00598.x

Soininen, J., Passy, S., and Hillebrand, H. (2012). The relationship between species richness and evenness: a meta-analysis of studies across aquatic ecosystems. *Oecologia* 169, 803–809. doi: 10.1007/s00442-011-2236-1

Song, T., Liang, Q., Du, Z., Wang, X., Chen, G., Du, Z., et al. (2022). Salinity gradient controls microbial community structure and assembly in coastal solar Salterns. *Genes (Basel)* 13:385. doi: 10.3390/genes13020385

Sorokin, D. Y., Merkel, A. Y., and Muyzer, G. (2020). Thiohalospira. *Bergey's Manual of Systematics of Archaea and Bacteria.* John Wiley & Sons, Inc.

Sun, X., Kong, T., Li, F., Häggblom, M. M., Kolton, M., Lan, L., et al. (2022). Desulfurivibrio spp. mediate sulfur-oxidation coupled to Sb(V) reduction, a novel biogeochemical process. *ISME J.* 16, 1547–1556. doi: 10.1038/s41396-022-01201-2

Tonima, K., and Savita, K. (2011). Pharmaceutical potentials of bacteria from saltpans of Goa. *Intern J App Pharmaceutics* 2, 150–154.

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30, 279–338. doi: 10.2307/1943563

Yu, J., Loh, K., Song, Z. Y., Yang, H. Q., Zhang, Y., and Lin, S. (2018). Update on glycerol-3-phosphate acyltransferases: the roles in the development of insulin resistance. *Nutr. Diabetes* 8:34. doi: 10.1038/s41387-018-0045-x

Zhang, J., Ma, G., Deng, Y., Dong, J., van Stappen, G., and Sui, L. (2016). Bacterial diversity in Bohai Bay solar Saltworks, China. *Curr. Microbiol.* 72, 55–63. doi: 10.1007/s00284-015-0916-5

Frontiers in Microbiology

Explores the habitable world and the potential of microbial life

The largest and most cited microbiology journal which advances our understanding of the role microbes play in addressing global challenges such as healthcare, food security, and climate change.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact