ARCHAEA IN THE ENVIRONMENT: VIEWS ON ARCHAEAL DISTRIBUTION, ACTIVITY, AND BIOGEOGRAPHY

EDITED BY: Andreas Teske, Ricardo Amils, Gustavo Antonio Ramírez and Anna-Louise Reysenbach

PUBLISHED IN: Frontiers in Microbiology

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-735-2
DOI 10.3389/978-2-88966-735-2

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

ARCHAEA IN THE ENVIRONMENT: VIEWS ON ARCHAEAL DISTRIBUTION, ACTIVITY, AND BIOGEOGRAPHY

Topic Editors:
Andreas Teske, University of North Carolina at Chapel Hill, United States
Ricardo Amils, Autonomous University of Madrid, Spain
Gustavo Antonio Ramírez, University of North Carolina at Chapel Hill, United States
Anna-Louise Reysenbach, Portland State University, United States

Citation: Teske, A., Amils, R., Ramírez, G. A., Reysenbach, A.-L., eds. (2021). Archaea in the Environment: Views on Archaeal Distribution, Activity, and Biogeography. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88966-735-2

Table of Contents

frontiers in Microbiology

EDITORIAL
published: 12 March 2021
doi: 10.3389/fmicb.2021.667596

Editorial: Archaea in the Environment: Views on Archaeal Distribution, Activity, and Biogeography

Andreas Teske[1]*, *Ricardo Amils*[2], *Gustavo A. Ramírez*[1,3] and *Anna-Louise Reysenbach*[4]

[1] Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2] Planetology and Habitability Department, Centro de Astrobiología, Madrid, Spain, [3] Western University of Health Sciences, College of Veterinary Medicine, Pomona, CA, United States, [4] Biology Department, Portland State University, Portland, OR, United States

Keywords: archaea, diversity, habitat, ecology, activity, environment

Editorial on the Research Topic

Editorial: Archaea in the Environment: Views on Archaeal Distribution, Activity, and Biogeography

OPEN ACCESS

Edited by:
Axel Schippers, Federal Institute for Geosciences and Natural Resources, Germany

Reviewed by:
Jerry Eichler, Ben-Gurion University of the Negev, Israel

***Correspondence:**
Andreas Teske
teske@email.unc.edu

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

Received: *13 February 2021*
Accepted: *19 February 2021*
Published: *12 March 2021*

Citation:
Teske A, Amils R, Ramírez GA and Reysenbach A-L (2021) Editorial: Archaea in the Environment: Views on Archaeal Distribution, Activity, and Biogeography. Front. Microbiol. 12:667596. doi: 10.3389/fmicb.2021.667596

On the occasion of the 10-year anniversary of Frontiers in Microbiology, this Research Topic was launched to highlight the linkages between extreme and archaeal microbiology (Teske, 2020). Archaea adapt to the physical and chemical characteristics of their habitat—such as organic matter availability, electron donor redox status, salinity, temperature, and pH—in terms of metabolic activity, community composition, gene expression patterns, and evolutionary diversification (Baker et al., 2020). Here, cultivation- and genome-based studies highlight linkages between archaea and their habitats (**Figure 1**).

An important question is the variability of archaeal genotypes and phenotypes within specific lineages at the genus- or family level. This variability defines the limits of adaptation for an archaeal group and therefore delineates its environmental range. In a study of physiological and genomic diversity within cultured isolates and high-quality environmental genomes of the extremely acidophilic, metal-and sulfur-oxidizing, facultatively autotrophic genus *Metallosphaera* (Wang P. et al.), similarity in carbon fixation, sulfur assimilation, nitrate reduction, and sulfur- and iron-oxidizing capabilities coexisted with species-specific variations in specific pathways. The evolving physiological and phylogenetic diversification within *Metallosphaera* is shaped by competing genome expansion *via* horizontal gene transfer to acquire new capabilities, and by adaptive genomic streamlining and gene loss.

As comparative genomic studies highlight ecophysiologically relevant differences among related cultured archaea, comparative field surveys of uncultured archaea infer ecophysiological traits from habitat preferences. A sequence-based archaeal survey of Thermoplasmatales-dominated metal-rich and low-pH sediments from an acid mine drainage (AMD) stream in Wales (Distaso et al.) showed that patterns of family- and genus-level uncultured lineages from this site differed from other well-studied AMD sites with distinct pH, redox status and metal and organic matter content, such as the Rio Tinto in southern Spain, Iron Mountain in California, and Carnoules Mine in France. Comparative analyses will gain ecophysiological and taxonomic specificity once more acidophilic lineages within the Thermoplasmatales are brought into culture, something that has been achieved for bacterial AMD community members to a far greater degree.

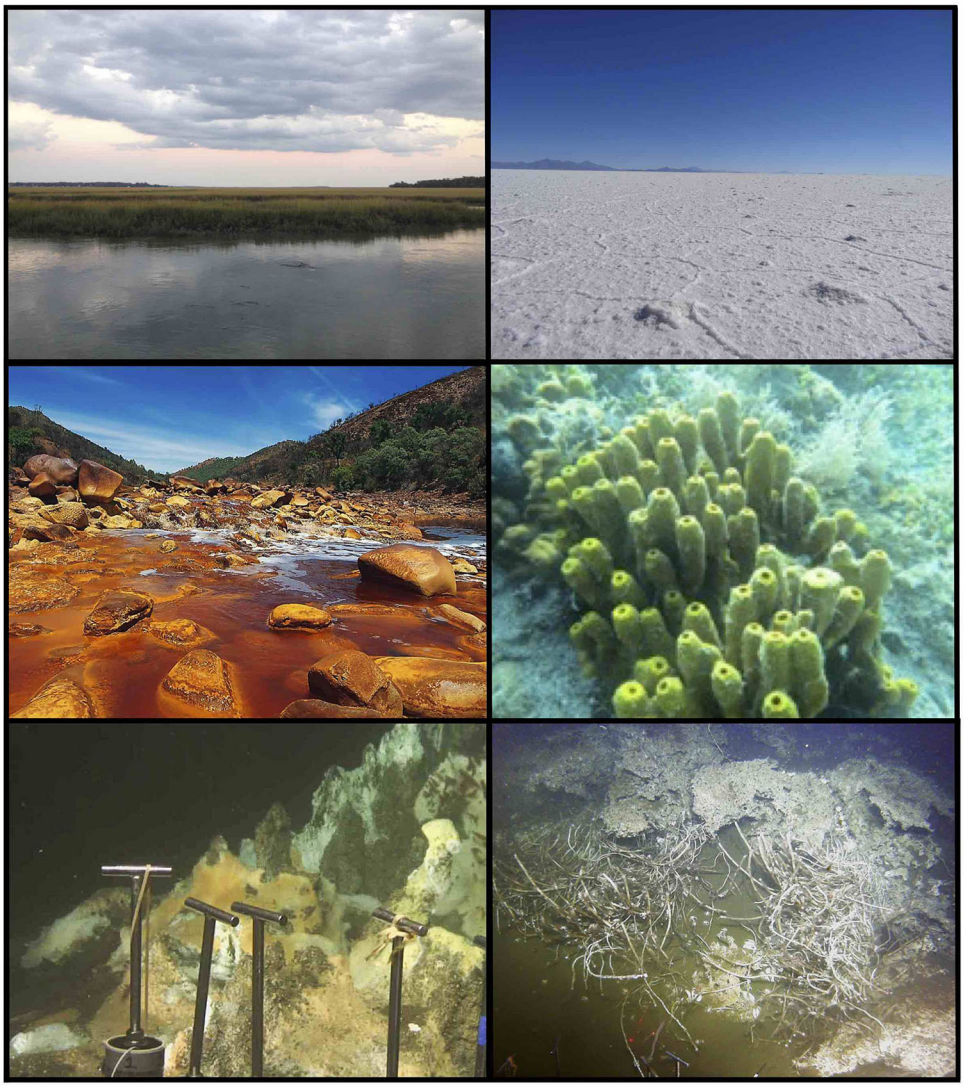

FIGURE 1 | Diverse archaeal habitats. Top row, the archaea-rich White Oak River estuary in North Carolina, USA (left), and the Salar de Uyuni, the world's largest salt flat and habitat for extremely halophilic archaea in the Andean Altiplano, Bolivia (right). Center row, the well-studied AMD model system of the acidic and metal-rich Rio Tinto, Spain (left), and the thaumarchaeotal symbiont-bearing sponge *Aplysina aerophoba*, Mediterranean Sea (right). Bottom row, hydrocarbon-rich hydrothermal sediments with microbial mat overgrowth (left), and cold seep carbonates and tubeworms in Guaymas Basin, Mexico (right). Photographs courtesy of D. Steen, N. Rodríguez, J. Segura, Steindler Lab, and WHOI Alvin group, respectively.

In part due to increasingly comprehensive global surveys (Nayfach et al., 2020), the archaeal domain is no longer regarded as a collection of extremophiles that inhabit niche environments. For example, archaeal communities thrive along the path of freshwater flow from upland soil to the sea, starting with agricultural soil, continuing with freshwater biofilms and concluding with estuarine waters (Wang H. et al.). The ammonia-oxidizing Thaumarchaeota are conspicuously present in all three environments, and their diversity gradient reflects an overall trend of greatest archaeal diversity in soil (Nitrososphaerales, Nitrosotaleales, and Nitrosopumilales), reduced diversity in freshwater biofilms (Nitrososphaerales) and lowest diversity in estuarine water (Nitrosopumilales). Within the Euryarchaota, the archaeal spectrum changes from the methanogenic Methanomasiliicoccales in soil toward other Methanomicrobia in freshwater biofilms and toward the Marine Group II (within the Thermoplasmatales) in estuarine water. These archaeal phyla have diversified and adapted successfully to colonize globally extensive, non-"extreme" habitats, and in doing so, affect global biogeochemical cycles, particularly of nitrogen and methane.

Among marine Thaumarchaeota, members of the ammonia-oxidizing, autotrophic Nitrosopumilales dominate the marine water column but also appear as endobionts in filter-feeding

sponges (Hallam et al., 2006), where they metabolize ammonium and urea produced by their host. Haber et al. compared the genomes of thaumarchaeotal sponge symbionts to those of their free-living counterparts, and found genomic modifications overlap with bacterial sponge symbionts—enriched transposes, toxin-antitoxin and restriction modification systems. The presence of branched amino-acid transporters, not found in free-living Nitrosopumilales, indicates heterotrophic amino acid metabolism and therefore a mixotrophic metabolism. This study provides an excellent example showing how archaea evolve and diversify not only as free-living microorganisms in response to environmental characteristics, but also as obligate symbionts in intimate associations with particular host species, as shown by host-specific genomic adaptions.

In some cases, the biochemistry of a particular substrate uptake and assimilation process limits the environmental competitiveness of a physiological class of archaea. Wormald et al. examine acetoclastic and hydrogenotrophic methanogenesis in alkaline soils and sediments and find that, above pH 9, hydrogenotrophic methanogenesis dominates and acetoclastic methanogenesis is no longer active; acetate continues to be utilized but only *via* syntrophic acetate fermentation with H2-utilizing methanogenic partners. Since acetate occurs as the acetate anion under alkaline conditions, the authors conclude that the energy expense for transmembrane anion uptake renders otherwise exergonic acetoclastic methanogenesis non-competitive. However, hydrogenotrophic, acetoclastic and methylotrophic methanogenesis remain detectable in serpentinization-derived alkaline fluids with pH up to 11 (Crespo-Medina et al., 2017), indicating that additional parameters are shaping the environmental persistence of methanogenic pathways.

Examples for bioprospecting of enzymes from extremophilic archaea are based on isolations of obligately halophilic, photoheterotrophic archaea from hypersaline salt lakes and salterns in India: Haloarchaeal bacteriorhodopsins and V-type ATPases from *Haloarcula*, *Halomicrobium*, and *Haloferax* isolates (Verma, Chaudhary, et al.), and an alpha-amylase from isolates related to *Halogeometricum borinquense* and *Haloferax volcanii* (Verma, Vasudeva, et al.). In contrast to pre-genomic studies, the enzymatic characterizations of these recombinant bacteriorhodopsins and alpha-amylase are supplemented by whole-genome analyses of the source organisms, to identify a fuller spectrum of candidate enzymes and their distribution across different halophilic taxa.

The final contribution of this Research Topic (Teske et al.) focuses on bacterial and archaeal communities in surficial hydrothermal sediments and recently discovered cold seeps of Guaymas Basin (**Figure 1**). Methane-cycling archaea are strongly affected by thermal regime in underlying sediments, whereas heterotrophic or mixotrophic sediment archaea (Thermoplasmata, Bathyarchaeota, and Lokiarchaeota), hyperthermophiles, and sometimes Thaumarchaeota coexist in the ecotone of temperate surficial sediments. Genomic exploration of surficial Guaymas sediments is yielding numerous novel archaeal lineages (Dombrowski et al., 2018) and is expanding the archaeal domain.

To conclude, archaeal culture- and genome-based findings are continuously providing new insights into the physiological adaptability and environmental range of archaea, and ultimately provide context to understand the evolutionary diversification of the archaeal domain throughout the microbial biosphere.

AUTHOR CONTRIBUTIONS

AT wrote the Editorial. All authors edited and commented on it.

ACKNOWLEDGMENTS

Research on hydrothermal archaea of Guaymas Basin is currently supported in the Teske lab by NSF-Molecular and Cellular Biosciences grant 1817381, NSF-Biological Oceanography grant 1829680, and the International Ocean Discovery Program. The Guaymas Basin photos in **Figure 1** were taken during cruise AT42-05 (Chief Scientist AT, UNC-CH) in HOV Alvin, Woods Hole Oceanographic Institution. RA was supported by grant PID2019-10481266GB-I00 from the Spanish Ministry of Science and Innovation.

REFERENCES

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol. Rev.* 5, 887–900. doi: 10.1038/s41564-020-0715-z

Crespo-Medina, M., Twing, K. I., Sánchez-Murillo, R., Brazelton, W. J., McCollom, T. M., and Schrenk, M. O. (2017). Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. *Front. Microbiol.* 8:916. doi: 10.3389/fmicb.2017.00916

Dombrowski, N., Teske, A. P., and Baker, B. J. (2018). Extensive microbial metabolic diversity and redundancy in Guaymas Basin hydrothermal sediments. *Nat. Comm.* 9:4999. doi: 10.1038/s41467-018-07418-0

Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., et al. (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18296–18301. doi: 10.1073/pnas.0608549103

Nayfach, S., Roux, S., Seshadri, R., Udwary, D., Varghese, N., Schulz, F., et al. (2020). A genomic catalogue of Earth's microbiomes. *Nat. Biotech.* doi: 10.1038/s41587-020-0718-6

Teske, A. (2020). 10 years of extreme microbiology: an interim reflection and future prospects. *Front. Microbiol.* 11:131. doi: 10.3389/fmicb.2020.00131

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 19 June 2020
doi: 10.3389/fmicb.2020.01192

Comparative Genomic Analysis Reveals the Metabolism and Evolution of the Thermophilic Archaeal Genus *Metallosphaera*

Pei Wang[1,2†], Liang Zhi Li[3,4†], Ya Ling Qin[1,2], Zong Lin Liang[1,2], Xiu Tong Li[3,4], Hua Qun Yin[3,4], Li Jun Liu[5], Shuang-Jiang Liu[1,2] and Cheng-Ying Jiang[1,2*]*

[1] State Key Laboratory of Microbial Resources, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China, [2] College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China, [3] School of Minerals Processing and Bioengineering, Central South University, Changsha, China, [4] Key Laboratory of Biometallurgy of Ministry of Education, Central South University, Changsha, China, [5] Department of Pathogen Biology, School of Basic Medical Science, Xi'an Medical University, Xi'an, China

OPEN ACCESS

Edited by:
Andreas Teske, The University of North Carolina at Chapel Hill, United States

Reviewed by:
Hongchen Jiang, China University of Geosciences, China
Zackary J. Jay, Montana State University, United States

***Correspondence:**
Shuang-Jiang Liu
liusj@im.ac.cn
Cheng-Ying Jiang
jiangcy@im.ac.cn

[†]These authors have contributed equally to this work

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

Received: *16 January 2020*
Accepted: *11 May 2020*
Published: *19 June 2020*

Citation:
Wang P, Li LZ, Qin YL, Liang ZL, Li XT, Yin HQ, Liu LJ, Liu S-J and Jiang C-Y (2020) Comparative Genomic Analysis Reveals the Metabolism and Evolution of the Thermophilic Archaeal Genus Metallosphaera. Front. Microbiol. 11:1192. doi: 10.3389/fmicb.2020.01192

Members of the genus *Metallosphaera* are widely found in sulfur-rich and metal-laden environments, but their physiological and ecological roles remain poorly understood. Here, we sequenced *Metallosphaera tengchongensis* Ric-A, a strain isolated from the Tengchong hot spring in Yunnan Province, China, and performed a comparative genome analysis with other *Metallosphaera* genomes. The genome of *M. tengchongensis* had an average nucleotide identity (ANI) of approximately 70% to that of *Metallosphaera cuprina*. Genes *sqr*, *tth*, *sir*, *tqo*, *hdr*, *tst*, *soe*, and *sdo* associated with sulfur oxidation, and gene clusters *fox* and *cbs* involved in iron oxidation existed in all *Metallosphaera* genomes. However, the adenosine-5′-phosphosulfate (APS) pathway was only detected in *Metallosphaera sedula* and *Metallosphaera yellowstonensis*, and several subunits of *fox* cluster were lost in *M. cuprina*. The complete 3-hydroxypropionate/4-hydroxybutyrate cycle and dicarboxylate/4-hydroxybutyrate cycle involved in carbon fixation were found in all *Metallosphaera* genomes. A large number of gene family gain events occurred in *M. yellowstonensis* and *M. sedula*, whereas gene family loss events occurred frequently in *M. cuprina.* Pervasive strong purifying selection was found acting on the gene families of *Metallosphaera*, of which transcription-related genes underwent the strongest purifying selection. In contrast, genes related to prophages, transposons, and defense mechanisms were under weaker purifying pressure. Taken together, this study expands knowledge of the genomic traits of *Metallosphaera* species and sheds light on their evolution.

Keywords: comparative genomics, *Metallosphaera*, metabolic potential, evolution, horizontal gene transfer

INTRODUCTION

Extremely acidophilic archaea of the genus *Metallosphaera* belong to the order Sulfolobales within the Crenarchaeota. Together with the genera *Acidianus* and *Sulfolobus*, physiologically versatile *Metallosphaera* spp. contribute significantly to biogeochemical element cycling and biomining processes. *Metallosphaera* spp. stand out due to their ability for facultative autotrophic growth and

tolerance to high concentrations of metal ions (Ai et al., 2016; Wheaton et al., 2016). Members of the genus *Metallosphaera* grow aerobically at low pH values (<4) and high temperatures (>60°C). They are found in sulfur-rich and metal-laden environments including solfataric areas (*Metallosphaera sedula* DSM 5348 and strains from laboratory evolution) (Huber et al., 1989; Ai et al., 2016, 2017), hot springs (such as *Metallosphaera hakonensis* HO1-1 = JCM8857, *Metallosphaera cuprina* Ar-4, *Metallosphaera tengchongensis* Ric-A, *Metallosphaera yellowstonensis* MK-1, *Metallosphaera* sp. UBA165, and *Metallosphaera* sp. Obs4) (Takayanagi et al., 1996; Kozubal et al., 2008, 2011; Liu et al., 2011a; Peng et al., 2015; Parks et al., 2017), and bioleaching-related heaps (*Metallosphaera prunae* Ron 12/II) (Fuchs et al., 1995). *Metallosphaera* species are of great potential in the extraction of base and precious metals from ores exploiting their ability to oxidize reduced inorganic sulfur compounds (RISCs) and ferrous ion, a process mediated by a set of terminal oxidases that are attached to their cell membranes (Auernik and Kelly, 2008, 2010; Orell et al., 2010). Previous studies showed that application of thermophilic archaea in chalcopyrite bioleaching could achieve faster dissolution rates and higher copper leaching yields in comparison to widely used mesophilic bioleaching bacteria because the formation of the surface passivation and diffusion layer of chalcopyrite was significantly reduced at high temperature (Rawlings, 2005; Urbieta et al., 2015; Castro and Donati, 2016).

Metallosphaera species can grow on peptides in a heterotrophic mode, fix carbon dioxide using reduced sulfur compounds as reductant in an autotrophic mode, or grow on casamino acids and FeSO4 or metal sulfides in a mixotrophic mode (Peeples and Kelly, 1995; Ulrike et al., 2005; Alber et al., 2006; Auernik and Kelly, 2008, 2010; Michael et al., 2010; Han and Kelly, 2015). Autotrophic carbon fixation through the 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle in *Metallosphaera* has been confirmed by genomics, transcriptomics, proteomics, and biochemical assays (Sebastian et al., 2011; Jiang et al., 2014). Reduced Fe and S (Fe^{2+}, HS^-, S^0, $S_4O_6{}^{2+}$, and $S_2O_3{}^{2+}$) are important electron donors for iron or sulfur-oxidizing microorganism. Unlike Fe^{2+} oxidation by the bacteria *Acidithiobacillus ferrooxidans* and *Leptospirillum ferrooxidans*, which require the blue copper protein rusticyanin and various c-type cytochromes (Rohwerder et al., 2003; Singer et al., 2008), the Fe oxidation mechanism of iron-oxidizing archaea is yet unconfirmed, although potential proteins were noticed in the genomes of *Ferroplasma*, *Sulfolobus*, and *Metallosphaera* species (Dopson et al., 2005; Bathe and Norris, 2007; Auernik et al., 2008). The genes involved in iron oxidation (Kozubal et al., 2011) such as *fox*, *cbs*, rusticyanin, and sulfocyanin have been detected in the genomes of *M. sedula* and *M. yellowstonensis*, but their pervasiveness in *Metallosphaera* is yet unknown. RISC oxidation mechanisms are complex and diverse in extremely thermoacidophilic archaea. Sulfur oxygenase reductase genes are present in the genus *Acidianus*, in the species *Sulfolobus tokadaii*, and in bacteria, but are absent from the genomes of *M. sedula* and *M. cuprina* (Auernik et al., 2008; Liu et al., 2011b), which raised the question how sulfur oxidation is initiated in *Metallosphaera*. Of the species within the genus *Metallosphaera*, genome sequences and genomic analyses have been reported for four isolates (*M. cuprina* Ar-4, *M. hakonensis* HO1-1 = JCM_8857, *M. sedula* DSM_5348, and *M. yellowstonensis* MK1) (Auernik et al., 2008; Liu et al., 2011b). *M. tengchongensis* Ric-A, the newest member of *Metallosphaera*, which was isolated from a sulfuric hot spring in Tengchong, Yunnan Province, China, showed an excellent performance in copper extraction from chalcopyrite (Peng et al., 2015). To better understand the unique metabolism, adaptation for extreme thermal and acidic conditions, roles in biogeochemical cycling, and the evolutionary history of the genus *Metallosphaera*, we performed genome sequencing of *M. tengchongensis* Ric-A and compared its genome with 18 available *Metallosphaera* genomes, of which 12 are from Genbank database of National Center for Biotechnology Information (NCBI) (*M. sedula* DSM5348, ARS50-1, ARS50-2, ARS120-1, ARS120-2, SARC-M1, and CuR1; *M. hakonensis* HO1-1 and JCM8857; *M. cuprina* Ar-4; *M. yellowstonensis* MK-1; *Metallosphaera* sp. UBA165) and six scaffold genomes are from the Integrated Microbial Genomes and Microbiomes (IMG/M) system of DOE Joint Genome Institute (JGI) (*Metallosphaera* spp. My-r02, My-r05, My-r06, YNP_08, YNP_14, and Obs4) (**Table 1**). In this work, we performed comprehensive analyses of genome-based phylogenetic relationships, metabolic pathway and gene function, heavy metal resistance, adhesion and motility, as well as mobile genetic elements (MGEs) and selective pressure. These findings will improve our understanding of the adaptive strategies of the organisms to their harsh environment and provide clues to design biomining or bioremediation processes in the future.

MATERIALS AND METHODS

Sample Collection and Sequencing

Metallosphaera tengchongensis strain Ric-A was isolated from the muddy water samples of sulfuric hot springs (24.57 N and 98.26 E, with the temperature range of 55–96°C and a pH range of 2.5–7.5, dissolved oxygen range of 0.01–1.00 mg/L) in Tengchong county of Yunnan Province, China. The hot springs are rich in S ($SO_4{}^{2-}$, 701.20–22.46 mg/L), Fe (13.89–0.05 mg/L), Ca (71.55–1.12 mg/L), K (22.64–63.01 mg/L), Al (42.43–0.13 mg/L), and other elements (Liu et al., 2011a; Peng et al., 2015; Qin et al., 2019). The sample was concentrated by tangential flow ultrafiltration through a hollow fiber membrane (Tianjin MOTIMO Membrane Technology, China). An aerobic enrichment culture in the flask with filtration membrane was established by inoculating the concentrate in basal salts medium (BSM) with elemental sulfur as energy source. The compositions of BSM were L^{-1}: $(NH_4)_2SO_4$, 3 g; $K_2HPO_4 \cdot 3H_2O$, 0.5 g; $MgSO_4 \cdot 7H_2O$, 0.5g; KCl, 0.1 g; $Ca(NO_3)_2$, 0.01 g, added with 1 ml of trace element solution ($FeCl_3 \cdot 6H_2O$, 1.1 g; $CuSO_4 \cdot 5H_2O$, 0.05 g; H_3BO_3, 0.2 g; $MnSO_4 \cdot H_2O$, 0.2 g; $Na_2MoO_4 \cdot 2H_2O$, 0.08 g; $CoCl_2 \cdot 6H_2O$, 0.06 g; $ZnSO_4 \cdot 7H_2O$, 0.09 g in 1 L of distilled water). After static culture for 5–7 days at 65°C, samples of the grown culture were spread on BSM solid plates with potassium tetrathionate ($K_2S_4O_6$, 10 mmol/L) or yeast extract (1 g/L) as energy source. The plates were incubated

TABLE 1 | The genomic statistics information and source of 19 strains used in this study.

Organism/name	Strain	Genbank or JGI accession number	Level and database	Size (Mb)	Coding density (%)	GC%	Gene	Protein	Environment	Address
M. tengchongensis	Ric-A	2821472399/CP049074	Complete NCBI	2.10	85.4	44.8	2331	2295	Acidic hot spring	China: Yunnan
M. sedula	DSM 5348	CP000682.1	Complete NCBI	2.19	89.3	46.2	2377	2298	Solfataric field	Italy
M. sedula	ARS50-1	CP012172.1	Complete NCBI	2.19	89.4	46.2	2375	2297	Lab	United States
M. sedula	ARS50-2	CP012173.1	Complete NCBI	2.19	89.4	46.2	2377	2298	Lab	United States
M. sedula	ARS120-1	CP012174.1	Complete NCBI	2.19	89.3	46.2	2376	2298	Lab	United States
M. sedula	ARS120-2	CP012175.1	Complete NCBI	2.19	89.3	46.2	2376	2298	Lab	United States
M. sedula	SARC-M1	CP012176.1	Complete NCBI	2.19	89.3	46.2	2379	2301	Lab	United States
M. sedula	CuR1	CP008822.1	Complete NCBI	2.19	89.2	46.2	2373	2289	Lab	United States
M. hakonensis	HO1-1	CP029287.1	Complete NCBI	2.54	80.3	43.7	2785	2736	Acidic hot Spring	Japan
M. hakonensis	JCM 8857	GCA_001315825.1	Draft NCBI	2.39	81.5	43.3	3312	3292	Acidic hot Spring	Japan
M. cuprina	Ar-4	CP002656.1	Complete NCBI	1.84	87.9	42.0	1968	1894	Acidic hot spring	China: Yunnan
Metallosphaera sp.	UBA165	GCA_001652185.1	Draft NCBI	1.83	88.5	45.7	3186	2821	Hot spring	Taiwan
M. yellowstonensis	MK1	GCA_000243315.1	Draft NCBI	2.82	82.4	47.7	3411	3356	Acidic hot spring	United States: Yellowstone National Park
Metallosphaera sp.	My-r02	2522125033	Draft IMG-M	1.85	82.5	47.4	2419	2366	Acidic hot spring	United States: Yellowstone National Park
Metallosphaera sp.	My-r05	2551306706	Draft IMG-M	1.85	87.9	47.4	1561	1520	Acidic hot spring	United States: Yellowstone National Park
Metallosphaera sp.	My-r06	2551306703	Draft IMG-M	1.85	89.0	48.3	1496	1455	Acidic hot spring	United States: Yellowstone National Park
Metallosphaera sp.	YNP_14	2502873002	Draft IMG-M	1.38	84.5	47.2	2093	2061	Thermal springs	United States: Yellowstone National Park
Metallosphaera sp.	YNP_08	2502894001	Draft IMG-M	1.10	78.4	43.9	2275	2252	Thermal springs	United States: Yellowstone National Park
Metallosphaera sp.	Obs4	2770939403	Draft IMG-M	2.70	83.2	50.4	3060	2968	Hot spring sediment	United States: Yellowstone National Park

for 7 days at 65°C. Colonies were picked and purified by re-plating. The purified strain Ric-A was grown at 70°C in BSM (pH 2.0) supplemented with 1 g/L yeast extract. The stationary-phase cells were harvested by centrifugation. The genomic DNA was extracted from the concentrated cells according to the instruction of "JGI Bacterial DNA isolation CTAB-2012."[1] After checking its quality, DNA was fragmented and the fragments were end-repaired and polyadenylated, and then ligated to sequencing adapter. SMRTbell DNA library was constructed by using Blue Pippin Size-Selection System; library quality was evaluated by Qubit 3.0 Fluorometer (Life Technologies, Grand Island, NY, United States) and sequenced by PacBio Biosciences (PacBio) RSII and Genome Analyzer IIx sequence platforms at Chinese National Human Genome Center at Shanghai (CHGC). After sequencing, the low-quality reads were filtered by the SMRT 2.3.0 (Chin et al., 2013), and the filtered reads were assembled to generate one contig without gaps. The hierarchical genome-assembly process (HGAP.3) pipeline implemented in SMRT Analysis 2.3.0[2] was used to correct for random errors in the long seed reads (seed length threshold 6 Kb) by aligning shorter reads from the same library against them. The resulting corrected, preassembled reads were used for *de novo* assembly. Genome data of 18 previously sequenced strains belonging to *Metallosphaera* were collected from NCBI and IMG-M database. The detailed genomic statistics information and source of 19 *Metallosphaera* strains used in this study were summarized in **Table 1**.

Average Nucleotide Identity and Whole Genome Alignments

Comparisons of average nucleotide identity (ANI) based on Blast algorithm were conducted using the pyani module[3] with default parameters. We applied the "progressive Mauve program" within Mauve v 2.3.0 (Darling et al., 2004) for constructing and visualizing multiple genome alignments of *M. tengchongensis* Ric-A with four other available complete genome sequences of *M. sedula* DSM 5348, *M. hakonensis* HO1-1, *M. cuprina* Ar-4, and *M. yellowstonensis* MK1. BlastN-based whole-genome comparison of strains *M. cuprina* Ar-4, *M. hakonensis* HO1-1, *Metallosphaera* sp. My-r06, *M. sedula* DSM 5348, *M. tengchongensis* Ric-A, and *M. yellowstonensis* MK1 were performed and represented with BRIG-0.95 (Alikhan et al., 2011), and these strains were used as references, respectively. GC content and GC skew of each genome were also indicated.

Phylogenomic Analyses

We constructed a phylogenetic tree of the 19 *Metallosphaera* spp. genomes based on whole-genome sequences with CVTree3 (Xu and Hao, 2009). The phylogenetic tree of the 19 genomes based on concatenation of the 85 core genes in a genome was constructed with the neighbor-joining (NJ), UPGMA, and maximum-likelihood (ML) method using MEGA-X (Kumar et al., 2018) with 1000 bootstrap replicates.

Genome Annotation

We applied Prokka (Seemann, 2014) and IMG Annotation Pipeline v.4.16.0 (Chen et al., 2017) for genome annotation and putative horizontally transferred gene detection. We performed Diamond BlastP v0.9.24 (Buchfink et al., 2015) with a cutoff e-value of $1e^{-5}$ together with the dbCAN database v2.0 (Le et al., 2018) for identification of genes related to carbohydrate activity enzyme (CAZymes). The BacMet v2.0 database (Pal et al., 2014) that contained genes with experimentally confirmed metal resistance function was used to identify the genes associated with metal resistance in *Metallosphaera* genomes. Gene annotations based on COG (Tatusov et al., 2001), Pfam (Finn et al., 2014), and TIGRFAM (Selengut et al., 2007) databases were performed via WebMGA (Wu et al., 2011) using Blast with a cutoff e-value of $1e^{-5}$.

Prediction of Mobile Genetic Elements

We applied the ISFinder (Siguier et al., 2006)[4] to predict and classify insertion sequences (IS) and transposases within *Metallosphaera* genomes with Blastp v0.9.24 (cutoff e-value of $1e^{-5}$). We applied the IslandViewer 4 (Bertelli et al., 2017),[5] which integrated prediction methods including IslandPath-DIMOB and SIGI-HMM that analyzed sequence composition, and another comparative genomic islands (GIs) prediction method IslandPick, to detect putative GIs distributed over *Metallosphaera* genomes. We applied PHASTER (Phage Search Tool Enhanced Release) (Arndt et al., 2016)[6] for detection and annotation of prophage and prophage remnant sequences within *Metallosphaera* genomes. We also applied CrisprCasFinder (Couvin et al., 2018)[7] for detection of CRISPRs and Cas genes within *Metallosphaera* genomes.

Comparative Genomic Analyses of Metallosphaera

The Bacterial Pan Genome Analyses tool (BPGA) pipeline (Chaudhari et al., 2016) was used to perform pan/core-genome analyses and calculation applying default parameters. The size of the *Metallosphaera* pan-genome was fitted into a power law regression function $P_s = \kappa n^{\gamma}$ with a built-in program of BPGA (Chaudhari et al., 2016), in which P_s was the total number of gene families, n stood for the number of tested genomes, and γ was free parameters. If exponent $\gamma < 0$, then the pan-genome of *Metallosphaera* was suggested to be "closed." In this case, the size of the pan-genome is relatively constant, even if new genomes were added into the analysis. On the contrary, the pan-genome was suggested to be "open" in the case of $0 < \gamma < 1$. In addition, the size of the core-genome of *Metallosphaera* was fitted into an exponential decay function $F_c = \kappa_c \exp(-n/\tau_c)$ with a built-in program of BPGA pipeline (Chaudhari et al., 2016), in which F_c stood for the number of core gene families, while κ_c and τ_c were free parameters. Gene family clustering followed by genome-wide comparisons of five

[1] https://jgi.doe.gov/user-programs/pmo-overview/protocols-sample-preparation-information/

[2] https://www.pacb.com/support/software-downloads/

[3] https://pypi.org/project/pyani/

[4] http://www-is.biotoul.fr

[5] https://www.pathogenomics.sfu.ca/islandviewer/

[6] http://phaster.ca/

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

Metallosphaera-type strains including *M. tengchongensis* Ric-A, *M. sedula* DSM 5348, *M. hakonensis* HO1-1, *M. cuprina* Ar-4, and *M. yellowstonensis* MK1 together with UniProt search, Gene Ontology (GO) Slim annotation, and GO enrichment analyses (default cutoff *p*-value is 0.05) were performed via OrthoVenn (Wang et al., 2015)[8] with default parameters.

Gene Family Evolution Analyses

Count is a software designed to analyze numerical profiles of homologous gene families on a phylogeny, which can execute ancestral reconstructions and predict family- and lineage-specific characteristics along the evolutionary tree (Miklós, 2010). We used Count software, combined BPGA v2.0 pipeline, and Wagner parsimony algorithm (Farris, 1970) for gene family clustering, ancestor genome size estimation, and detecting gene family gain and loss events, together with family expansion and contraction events with penalty ratio set to 1. We conducted the analyses only on five *Metallosphaera*-type strains, including *M. tengchongensis* Ric-A, *M. sedula* DSM 5348, *M. hakonensis* HO1-1, *M. cuprina* Ar-4, and *M. yellowstonensis* MK1 taken into consideration, due to the estimation requiring complete sets of testing gene families available only in species with high-quality genome databases.

Selective Pressure Analyses

We detected the numbers of sites under negative (purifying) or positive (diversifying) selection and estimated global dN/dS values of each gene family that contained more than three non-identical sequences (due to limitation of HyPhy) based on multiple alignments of orthologous codon sequences and a tree topology by means of HyPhy package (Pond and Muse, 2005) using the Fixed Effects Likelihood (FEL) method (Kosakovsky Pond and Frost, 2005) (applied a likelihood ratio test with default cutoff *p*-value: 0.1) via datamonkey server (Weaver et al., 2018).[9] The coding sequences of five *Metallosphaera*-type strains, including *M. tengchongensis* Ric-A, *M. sedula* DSM 5348, *M. hakonensis* HO1-1, *M. cuprina* Ar-4, and *M. yellowstonensis* MK1, were aligned with muscle codon alignment module implemented in MEGACC (Kumar et al., 2012) to obtain final codon alignments.

RESULTS AND DISCUSSION

General Genome Features of *M. tengchongensis* and Other Strains

A total of 2331 CDS, including 32 tRNA and 4 rRNA, were predicted in the genome of *M. tengchongensis* strain Ric-A using Prokka (Seemann, 2014). Whole-genome BLASTN-based ANI analyses showed only an ANI of 72.3% compared with the closest *Metallosphaera* genome (**Supplementary Table S1**). A summary of features for the 19 *Metallosphaera* genomes is listed in **Table 1**. *M. yellowstonensis* MK1 possessed the largest genome (2.82 Mb). The G + C contents of the 19 genomes ranged from 42.0 to 50.4%. These genomes varied in coding density from 78.4 to 89.4%, indicating substantial intra-genus differences. We determined that the previously unclassified strain UBA165 was a member of *M. sedula*, and strains My-r02, My-r05, My-r06, YNP_08, and YNP_14 were members of *M. yellowstonensis* based on an ANI cutoff of 96% (Richter and Rossello-Mora, 2009), as supported by further phylogenetic analyses (**Figure 1**, **Supplementary Figure S1**, and **Supplementary Table S1**). The genome alignment of *M. tengchongensis* Ric-A with four other complete genomes of *Metallosphaera* strains using Mauve (Darling et al., 2004) indicated that the chromosomal alignments of *Metallosphaera* genomes were non-conserved, as shown by the presence of hundreds of poorly organized collinear blocks and numerous inversed and rearranged regions (**Supplementary Figure S2**). Each *Metallosphaera* species harbored genomic regions that were not commonly shared; most of these harbored poorly characterized proteins as revealed by whole genome comparison of *Metallosphaera* spp. using BRIG (Alikhan et al., 2011) (**Supplementary Figures S2, S3**).

Core- and Pan-Genome and Phylogenomic Analysis of Metallosphaera

The phylogenetic trees based on the concatenated alignment of 85 core genes inferred with NJ, ML, MP, and UPGMA methods were congruent with each other, and the phylogenetic tree based on whole-genome analyses was congruent with the core-gene phylogenetic tree (**Supplementary Figures S4, S5**). Strain Obs4 is located on a clade apart from other *Metallosphaera* strains (**Supplementary Figures S1, S4, S5**). However, strain Obs4 should still be considered a member of genus *Metallosphaera* based on percentage of conserved proteins (POCP). Strain Obs4 had a POCP of 50.2% against *M. yellowstonensis* MK1, within the genus cutoff value of 50% (Qin et al., 2014). The pan-genome of 19 *Metallosphaera* strains possessed 6499 gene families, while the core-genome possessed only 85 gene families. Core- and pan-genome analyses of the 19 *Metallosphaera* genomes revealed an "open" pan-genome fitted into a power law regression function [$P_s\ (n) = 1955.63\ n^{0.372819}$] with a parameter ($\gamma$) of 0.372819 falling into the range $0 < \gamma < 1$. The core-genome was fitted into an exponential regression [$F_c\ (n) = 1886.87\ e^{-0.16335\ n}$], which had a steep slope, reaching a minimum of 85 gene families after the 19th genome was added (**Figure 2B**). The result of COG annotation revealed that the core-genome had a higher proportion of genes involved in COG categories that are associated with central biological functions translation, ribosomal structure, and biogenesis (J); posttranslational modification, protein turnover, chaperones (O); and coenzyme transport and metabolism (H) than the accessory genome and strain-unique genome. In contrast, the accessory genome had a higher proportion of genes related to COG energy production and conversion (C) and lipid transport and metabolism (I). We found that strain-specific gene families had a higher proportion of genes categorized in COG replication, recombination and repair (L), cell wall/membrane/envelope

[8]https://orthovenn2.bioinfotoolkits.net/home

[9]http://www.datamonkey.org/fel/

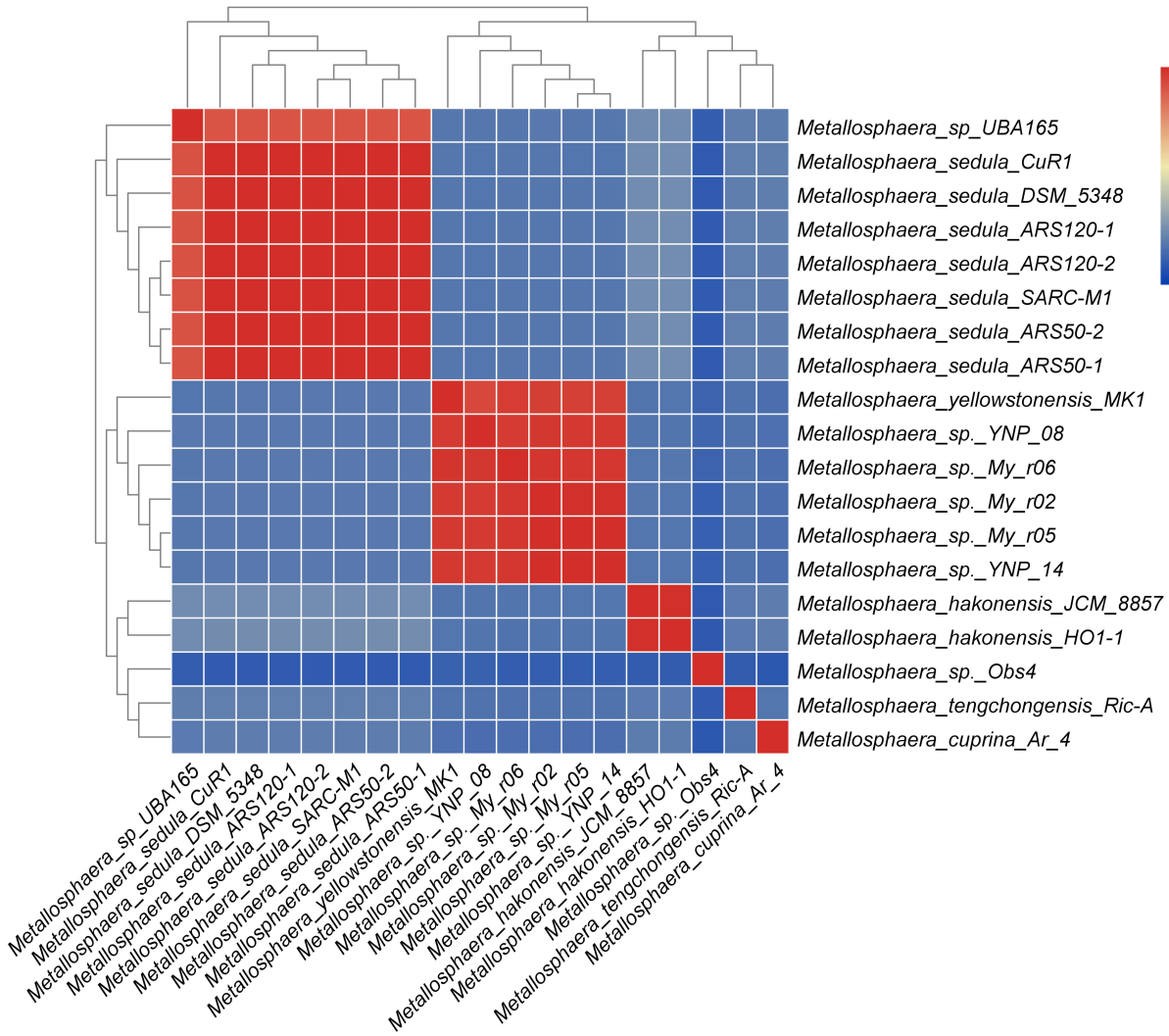

FIGURE 1 | Heat map showing species boundary of genus *Metallosphaera* based on ANI value (cutoff value = 96%), which is a universally accepted genomic measure for prokaryotic species delineation (Richter and Rossello-Mora, 2009).

biogenesis (M), inorganic ion transport and metabolism (P), and carbohydrate transport and metabolism (G) (**Figure 2**). We propose that these genes are associated with adaptive evolution within the genus *Metallosphaera*. The oligotrophic, metal-laden, and extremely acidic environments select for a highly efficient DNA injury repair system and a flexible trophic mode in *Metallosphaera*. The genomic diversity and specificity of different *Metallosphaera* strains are reflecting their distinct survival strategies in different environments.

Metabolic and Functional Potential

Sulfur Metabolism

Sulfur and hydrogen are important in energy flow in thermal environments, such as marine hydrothermal systems, continental solfataras, and hot springs, where many bacteria and archaea can grow by oxidizing hydrogen, sulfide, elemental sulfur, and thiosulfate (Amend and Shock, 2001; Zeldes et al., 2019). It was reported that *Metallosphaera* could utilize different sulfur compounds as energy sources for growth (Peng et al., 2015). The genes *sqr* encoding for sulfide:quinone oxidoreductase (SQR), which catalyzes the oxidation of hydrogen sulfide forming polysulfide, were detected in all 19 genomes of *Metallosphaera*. Unlike the genus *Acidianus* and certain members of *Sulfolobus*, the genes encoding the homologs of sulfur oxygenase/reductase (SOR), a key enzyme for archaeal sulfur oxidation (Kletzin, 1992; Chen et al., 2005; Urich et al., 2006; Dai et al., 2016), were not found in the genomes of all strains of *Metallosphaera*. However, sulfur dioxygenases (SDO, *sdo*) were encoded in all strains except for *M. cuprina* and *Metallosphaera* sp. Obs4 (**Figure 3**). In addition, sulfite-acceptor oxidoreductases (SAOR, *saor*) genes were also detected in all strains (**Figure 3**). *Metallosphaera* spp. also harbored genes encoding for tetrathionate hydrolase (TTH, *tth*), sulfite reductase (SIR, *sir*), and genomic clusters

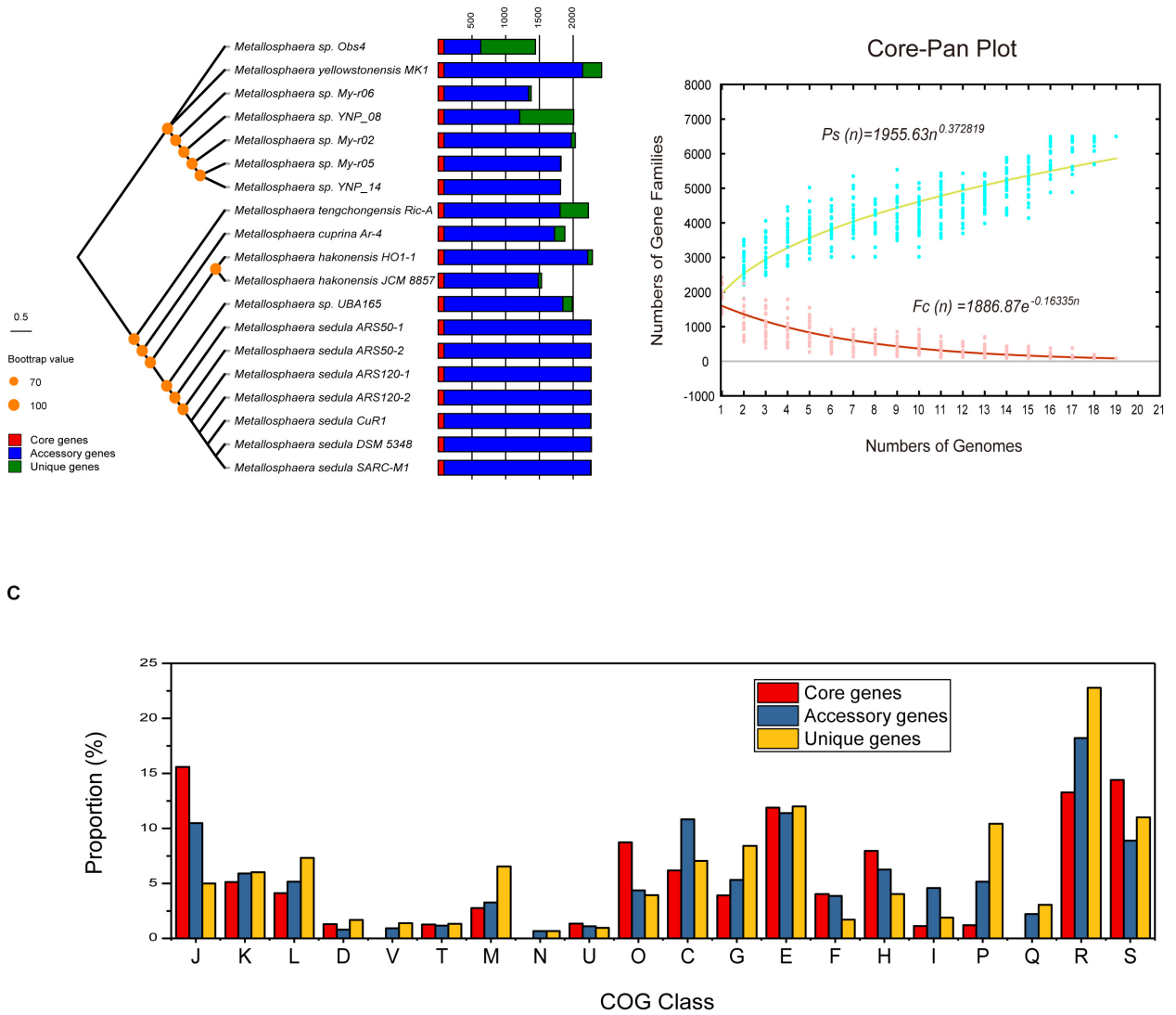

FIGURE 2 | Pan-genome analyses of strains in the genus *Metallosphaera*. **(A)** NJ phylogenetic tree based on concatenated 85 core genes (1000 bootstrap replicates) of 19 *Metallosphaera* strains (left) and stack bar diagram showing sizes of orthologous genes shared by all strains (i.e., the core-genome), number of orthologous genes shared by partial strains (i.e., the accessory genome), and number of strain-specific gene families (i.e., the unique gene) in each strain. **(B)** The curve of the fitting function and the estimation of pan-genome and core-genome sizes of *Metallosphaera*. *Ps* and F_c stand for the total number of gene families and the number of core gene families, respectively. More details for modeling approaches are presented in Section 2.6 "Comparative Genomic Analyses of *Metallosphaera*". **(C)** Bar chart showing proportions of COG categories of the different parts of *Metallosphaera* pan-genome (i.e., core, accessory, unique). COG categories description (J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; D, cell cycle control, cell division, chromosome partitioning; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, Cell motility; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown).

encoding for the thiosulfate:quinone oxidoreductase (TQO) subunits (*doxA/doxD*) and heterodisulfide reductase (HDR, *hdrC1-hdrB1A-hyp-hdrC2-hdrB2*) complex. The genes encoding for thiosulfate sulfurtransferase (TST, *tst*) were also detected in all species (**Figure 3**). According to the genome composition, we speculate that the pathway of sulfur metabolism for *Metallosphaera* might be SDO oxidized sulfur to sulfite, SAOR oxidized sulfite to sulfate, and TQO was responsible for the transformation of thiosulfate to tetrathionate, whereas TTH catalyzed the tetrathionate hydrolysis into thiosulfate, sulfur, and sulfate. However, the function of some proteins, especially for SDO and SAOR, had not been verified in *Metallosphaera* members. TST and HDR catalyzed the mutual conversion of thiol proteins (RSH) and sulfane sulfates (RSSH) in the

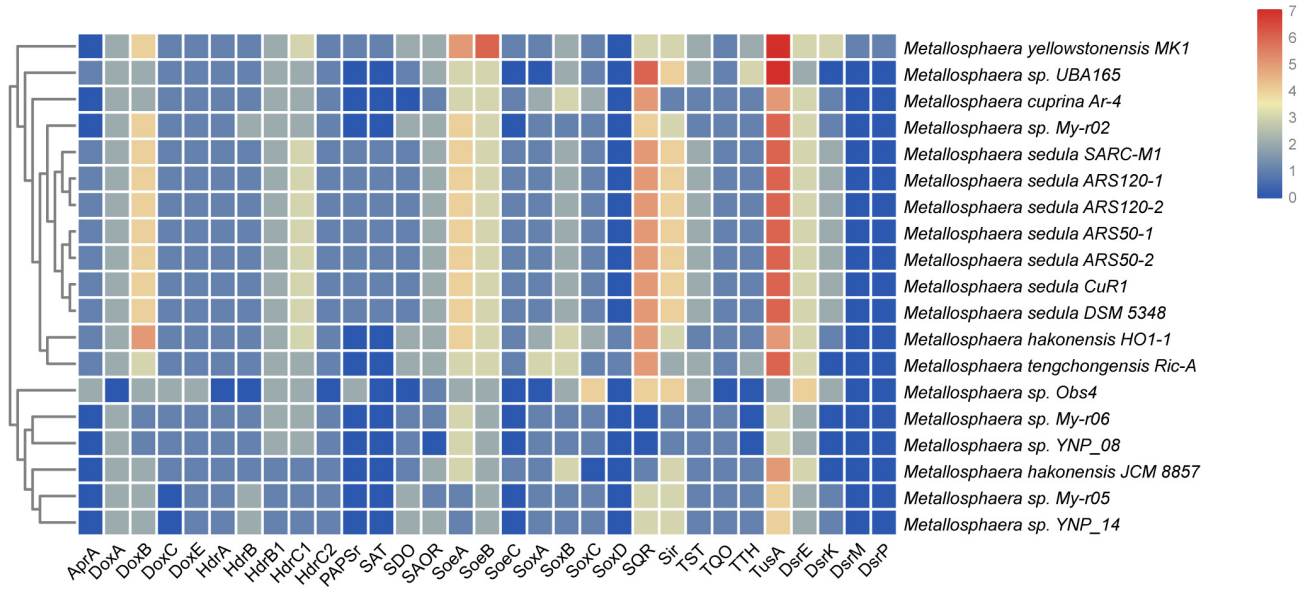

FIGURE 3 | Heat map showing the distribution and numbers of genes encoding putative proteins in sulfur metabolism in various *Metallosphaera* strains. *AprA*, APS reductase; *SAT*, ATP sulfurylase; *TST*, thiosulfate sulfurtransferase; *SDO*, sulfur dioxygenase; *DoxA/DoxD*, thiosulfate:quinone oxidoreductase subunits; *Sir*, sulfite reductase; *TTH*, tetrathionate hydrolase; *HdrA/HdrB/HdrC*, heterodisulfide reductase complex subunits; *PAPSr*, phosphoadenosine phosphosulfate reductase; *SQR*, sulfide:quinone oxidoreductase; *SoeABC*, sulfite-oxidizing enzyme; *CbsAB/SoxNL*, terminal oxidase complex; *DoxE*, sulfocyanin; *MCO*, multicopper blue protein; *doxBCE*, terminal oxidase complex; *TusA/DsrE*, cytoplasmic sulfur trafficking proteins.

cytoplasm (Chen et al., 2012), and HDR was also implicated in transferring electrons to reduce Fd_{ox} (Hua et al., 2018). Besides, *tusA-dsrE2-dsrE3A* gene clusters that functioned in cytoplasmic sulfur trafficking and dissimilatory tetrathionate oxidation were found located next to the *hdr* clusters in all *Metallosphaera* genomes, similar to other Sulfolobales members (Urbieta et al., 2017). The subunits SoeAB of the heterotrimeric membrane-bound sulfite-oxidizing enzyme complex SoeABC were detected in all *Metallosphaera* genomes, but the subunit SoeC was not identified in *Metallosphaera* sp. UBA165, *Metallosphaera* sp. Obs4, and other strains of My-r02, YNP-08, My-r06, My-r05, and YNP-14, probably resulting from the incompleteness of these genomes. Genes coding ATP sulfurylase (SAT, *sat*), adenosine-5′-phosphosulfate reductase (APR, *apr*) subunit AprA, and phosphoadenosine phosphosulfate reductase (PAPSr, *papsr*) that are involved in adenosine-5′-phosphosulfate (APS) pathway were only detected in *M. sedula* and *M. yellowstonensis* (**Figure 3**). Genes encoding for *soxABC* and *doxBCE* complex involved in sulfur oxidation and electron transfer were also annotated in *Metallosphaera* genomes. It was reported that sulfite was readily oxidized to sulfate through the direct SoeABC pathway and/or the indirect APS pathway presented in the cytoplasm in the purple sulfur bacterium *Allochromatium vinosum* (Christiane et al., 2013). Whether these pathways work in other phototrophic sulfur bacteria and sulfur-oxidized archaea remains unknown. Genes (e.g., *hyn*S, *hyn*L, *hox*M, *hyp*C, *hyp*D, and *isp* 1) encoded for the structural and auxiliary proteins of Ni/Fe hydrogenase that are potentially associated with electron transfer between hydrogen and sulfur-containing compounds were only detected in *M. sedula* and *M. yellowstonensis*; however, genes *hyp*E, *hyn*Y, and *isp* 2 were presented in *all* analysis strains. No genes of h*yp*Z, *hyn*Z, and *hyp*Y were found on the genomes of all strains (Laska et al., 2003) (**Supplementary Figure S6**).

Iron Metabolism

The *fox* genes involved in iron oxidation (Kozubal et al., 2011) were detected in all *Metallosphaera*-type species, but *M. cuprina* may have lost the subunits *foxD*, *foxE*, *foxF*, and *foxI* (**Supplementary Figure S7**). The arrangement of *fox* genes was similar in *Metallosphaera* spp. following the pattern *foxA-A′-I-B-C-D-E-F-J-G-H*, but it is different from other members of iron-oxidizing Sulfolobales (**Figure 4**), and in *M. hakonensis* HO1-1, a third copy of *foxA* was found within this cluster, probably resulting from gene duplication (**Figure 4**). The transcription initiation directions of these *fox* genes were not consistent, and their open reading frames (ORFs) were separated by spacers (**Figure 4**). Genes encoding for terminal oxidase complex *cbsAB/soxNL* were found downstream of *fox* cluster in *M. sedula* and *M. yellowstonensis*, while in strain *M. tengchongensis* Ric-A, these two clusters were separated by a genome region of about 180 kb (**Figure 4**). A gene for sulfocyanin (SoxE), a blue copper-containing protein that may function as a temporary electron storage or electron carrier in the iron-oxidizing electron transport chain (Kozubal et al., 2011), occurred in all *Metallosphaera* genomes in this study. Genes encoding for multicopper blue protein (*mco*) that contained two plastocyanin type I copper domains were only detected in *M. yellowstonensis* and *M. sedula* (Kozubal et al., 2011). These oxidases may couple the reduction of oxygen to proton translocation in cooperation with Fox complex mentioned-above (Auernik and Kelly, 2008; Kozubal et al., 2011).

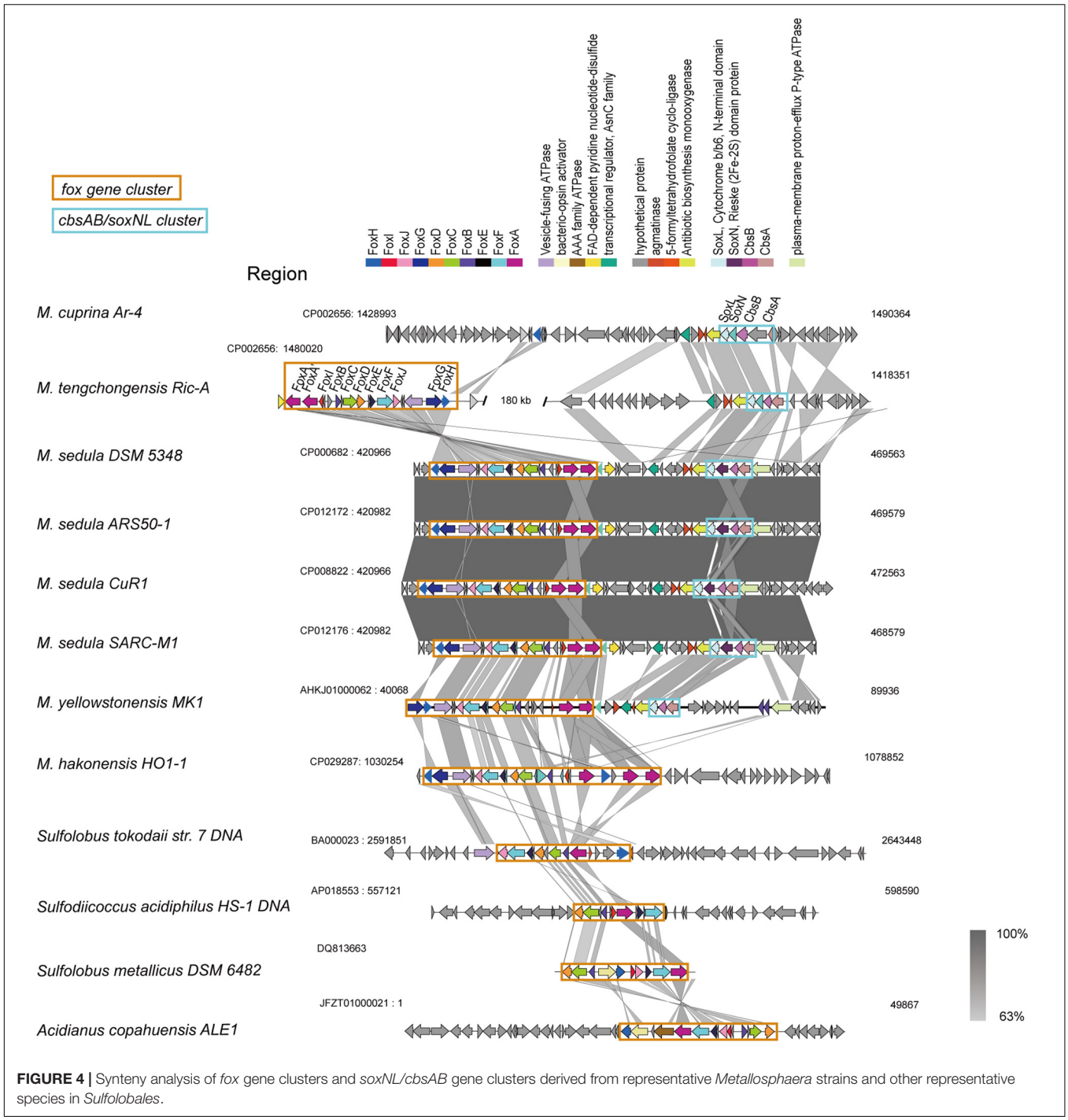

FIGURE 4 | Synteny analysis of *fox* gene clusters and *soxNL/cbsAB* gene clusters derived from representative *Metallosphaera* strains and other representative species in *Sulfolobales*.

Carbon Metabolism

Metallosphaera spp. contained an abundant repertoire of carbohydrate-active enzymes (CAZymes) including carbohydrate esterases (CEs), carbohydrate binding molecules (CBMs), glycosyltransferases (GTs), glycoside hydrolases (GHs), auxiliary activity proteins (AAs), and a small number of polysaccharide lyases (PLs), of which GTs were most abundant, and *Metallosphaera* sp. Obs4 possessed the most carbohydrate active enzymes (**Supplementary Figure S7**). The diversity of encoding genes for CAZymes possibly reflected different carbohydrate metabolism strategies in *Metallosphaera*. A complete set of genes encoding for glycolysis, gluconeogenesis, the archaeal pentose phosphate pathway, as well as an atypical TCA cycle (replacing genes encoding the alpha-ketoglutarate dehydrogenase with those encoding 2-oxoacid:ferredoxin oxidoreductase) were detected in all *Metallosphaera*

species. *Metallosphaera* spp. also contained complete non-phosphorylative and semi-phosphorylative Entner–Doudoroff (ED) pathways. Genes encoding homologous enzymes for the complete 3-hydroxypropionate/4-hydroxybutyrate (HP/HB) cycle and the dicarboxylate/4-hydroxybutyrate (DC/HB) cycle involved in autotrophic carbon fixation (Berg et al., 2007, 2010; Jiang et al., 2014) were also found in all *Metallosphaera* genomes (**Supplementary Figure S9**). It is reported that carbon monoxide was ubiquitous in hydrothermal habitats (King and Weber, 2007). All members of *Metallosphaera* in this study possessed putative type I carbon monoxide dehydrogenase (CODH) encoded by gene cluster *coxLSM*; however, the typical active site motif (VAYRCSFR) of CODH (Dobbek et al., 2002) was not observed in these proteins, which indicated that these CODH probably do not use CO; alternatively, they could possess a novel active site motif (**Supplementary Figure S9**).

Nitrogen Metabolism

All *Metallosphaera* strains possessed genes encoding for nitrate reductase and nitrite reductase involved in assimilatory nitrate reduction. Except for *nar*G genes in seven strains of *M. sedula*, the *nar*GHJI operon encoding for dissimilatory nitrate reductase was only found in *M. yellowstonensis* MK1 (**Supplementary Figure S10**). This operon was located on a GI in *M. yellowstonensis* MK1, suggesting that it might be acquired through horizontal gene transfer (HGT) events (**Supplementary Table S2**; Bertelli et al., 2017). Urease gene clusters consisting of the functional subunits (*ureAB* and *ureC*) and accessory proteins (*ureE*, *ureF*, and *ureG*) were found in all Metallosphaera species except for *M. yellowstonensis* (**Supplementary Figure S10**), indicating that urea assimilation was employed by most *Metallosphaera* strains to provide sufficient ammonia by converting one urea molecule into one carbon dioxide molecule plus two ammonia molecules. Genes encoding for nitrilase and formamidase, metabolizing organic nitrogen to ammonia, and genes encoding for ammonium transport (*amt*) and ammonia-dependent biosyntheses, such as carbamoyl-phosphate synthases (*carAB*), glutamate dehydrogenases (*gdhA*), and glutamine synthetases (*glnA*), were found in the genomes of all *Metallosphaera* species (**Supplementary Figure S10**). There is also a complete set of genes involved in arginine synthesis in all Metallosphaera species (**Supplementary Figure S10**). Polyamine derived from arginine can stabilize DNA by protecting DNA from free radical attacks and thermal denaturation (Abby et al., 2018). Polyamine biosynthesis-related genes that encoded for agmatinase (*speB*), S-adenosylmethionine decarboxylase (*speD*), and spermidine synthase (*speE*) were found in the genomes of all Metallosphaera species, while the arginine decarboxylase gene (*speA*) for the first two steps of putrescine biosynthesis was not detected; its function may have been substituted by S-adenosylmethionine decarboxylase (*speD*) (Giles and Graham, 2008).

Heavy Metal Resistance

Metallosphaera can survive natural and anthropogenic metal-rich environments, and BacMet database annotations revealed an abundant repertoire of heavy metal resistance genes in *Metallosphaera* spp., exhibiting diverse strategies to avert the deleterious effect of toxic metals on biological function (**Supplementary Figure S11**). Most of the genes related to arsenic resistance (*aioAB*: arsenite oxidase; *arsABR*: arsenical pump-driving ATPase; *arsM*: methyltransferase), divalent-cation resistance (*copARZ*, *cueA*, *cutA*, *czcD*, *corRC*, *nccN*, *nikABCDER,* and *mntRH*), mercury resistance (*merA*: mercuric reductase), and iron regulation (*fecDE*, *furA*, *fbpC*) were found in all tested *Metallosphaera* genomes. Alkylmercury lyase (encoded by gene *merB*) that cleaved mercury-alkyl bonds for mercury detoxification (Melnick and Gerard, 2007) was only detected in *M. yellowstonensis.* Genes encoding for "DNA-binding protein from starved cells" (Dps) (Blake et al., 2005), which can physically shield DNA against oxidative damage, were only detected in strains of *M. sedula* and *Metallosphaera* sp. UBA1654. Dps proteins were also able to control Fenton reaction through storing ferric oxide as a mineral core on their interior cage surface (Blake et al., 2005).

Adhesion and Motility

Etracellular polysaccharides (EPS) play a significant role in cell adhesion and biofilm formation, which is closely related to colonization, mineral solubilizing ability, and protection against adverse environmental conditions (Basak et al., 2014; Marino et al., 2018; Yu et al., 2019). Gene clusters encoding for d-TDP-glucose pyrophosphorylase (*rfbA*), d-TDP-glucose 4,6-dehydratase (*rfbB*), d-TDP-4-dehydrorhamnose 3,5-epimerase (*rfbC*), and d-TDP-4-dehydrorhamnose reductase (*rfbD*), which can convert glucose-1-phosphate to the EPS precursor d-TDP-rhamnose via a series of reactions, were found in all Metallosphaera genomes (**Supplementary Figure S9**). Motility conferred by flagella can provide a competitive advantage for microorganisms to move toward beneficial conditions. From the genomic and functional analysis, it is found that the sequenced archaea maintain a unique flagellum composition and mode of assembly, distinct from the bacteria. Archaeal flagellin commonly anchors on a *fla* locus, which was encoded by 7–13 flagella-related genes. Studies of the flagella system of the crenarchaeal model strain *Sulfolobus acidocaldarius* identified seven *fla* genes called *fla*B, *fla*X, *fla*G, *fla*F, *fla*H, *fla*I, and *fla*J, and confirmed that all seven genes were essential for assembly and motility (Thomas et al., 2001; Lassak et al., 2012; Dominik et al., 2016). However, in the genomes of 19 strains of *Metallosphaera*, genes encoding for crenarchaeal flagellin (*flaB*) and flagella accessory proteins (*flaH*, *flaI*, *flaG*, *flaF,* and *flaJ*) were only detected in strains of *M. sedula*, *M. yellowstonensis*, and *Metallosphaera* sp. UBA165. Further, a unique crenarchaeal gene *flaX*, coding for a structural part of the archaeal flagellum assembly apparatus only present in *Metallosphaera* sp. Obs4 and *Metallosphaera* sp. My-r06 (**Supplementary Figure S12**). The roles of FlaX and other accessory proteins in *Metallosphaera* members need further experimental verification. Genome comparative results indicated that the flagellum assembly apparatus did not present in every genome of these strains, which is consistent with the motility difference of *Metallosphaera* strains known by their physiological research (Peng et al., 2015). The genomes of *M. tengchongensis* Ric-A and *M. hakonensis* DSM 7519 only

contain Fla L. Both species do not possess flagella and do not have motility (Peng et al., 2015).

Mobile Genetic Elements and CRISPR-Cas Systems

Mobile genetic elements, including GIs, IS, transposons, and phages, are genome segments that display intra- and/or extracellular translocation abilities associated with HGT (Springael and Top, 2004). The prokaryotic CRISPR (clustered, regularly, interspaced, short, palindromic repeats)-Cas (CRISPR-associated genes) systems are attested to confer resistance to viral attack and mediate interactions between the host and phage. The results showed that in *Metallosphaera*, the number of transposon sequences per genome ranged from 65 (*M. cuprina* Ar-4) to 283 (*M. yellowstonensis* MK1) and the number of GI-related sequences per genome ranged from 21 (*M. cuprina* Ar-4) to 487 (*M. yellowstonensis* MK1) (**Supplementary Tables S2, S3**). GI regions of *Metallosphaera* spp. harbored genes related to the formation of extracellular polysaccharides (EPS) that account for ∼2.4% of all sequences, defense and DNA repair systems (∼2.7%), and stress resistance (∼2.0%). These additional functionalities endowed by HGT events may facilitate adaptive survival and protect *Metallosphaera* against DNA damage and protein denaturation in its acidic hot spring habitat (Yu et al., 2019). The number of prophages and prophage remnants ranged from 0 to 27 (*M. sedula* DSM5348) (**Supplementary Table S3**). Type I-A and/or type III-A/D CRISPR-Cas systems were found in all the genomes in this study except for *Metallosphaera* sp. My-r06. *M. tengchongensis* Ric-A contained the most (410) CRISPR-Cas-related genes or spacers. CRISPR-Cas-related sequences were detected in predicted GI regions of all *Metallosphaera* spp. except for *M. cuprina* Ar-4, reflecting the mobility of CRISPR-Cas systems (Koonin and Makarova, 2017; Krupovic et al., 2017; Peters et al., 2017).

Gene Ontology Enrichment and Evolutionary Analyses of Five *Metallosphaera*-Type Strains

We further applied the software OrthoVenn (Yi et al., 2015) and Count (Miklós, 2010) for gene clustering, GO enrichment analyses, and gene family evolutionary analyses of five *Metallosphaera*-type strains including *M. tengchongensis* Ric-A, *M. sedula* DSM 5348, *M. hakonensis* HO1-1, *M. cuprina* Ar-4, and *M. yellowstonensis* MK1. Results showed that 1614 (68.6%) out of 2353 identified gene families were shared by all species. *M. sedula* DSM 5348 had the most gene families (2133) in its genome, whereas *M. cuprina* Ar-4 had the fewest (1827), and *M. yellowstonensis* MK1 possessed the most strain-specific gene families (42) followed by *M. tengchongensis* Ric-A (18) (**Figure 5**). GO enrichment analyses showed that in significantly enriched (p-value < 0.05) commonly shared gene families, GO terms were associated with basic biological functions including translation (GO:0006412), rRNA binding (GO:0019843), and ribosome-related function (GO:0005840; GO:0003735). Functions significantly enriched (p-value < 0.05) in accessory genes families were related to electron transport chain (GO:0022900; GO:0008137; GO:0042773; GO:0048038; GO:0006741; GO:0008121; GO:0003951), lipid metabolism-related long-chain fatty acid-CoA ligase activity (GO:0004467), and nutrient transport and catabolic process of nitrogen sources such as urea (GO:0043419; GO:0009039) and carbon sources such as lactate (GO:0035873), mandelate (GO:0019596), glycolate (GO:0097339), benzoylformate (GO:0050695), and maleylacetate (GO:0018506) related to heterotrophic lifestyle. Proton symporter activity (GO:0015295; GO:0015538) involved in pH homeostasis, amino acid biosynthesis-related denitrification pathway (GO:0019333), and sulfur metabolism-related functions including sulfate assimilation (GO:0000103), hydrogen sulfide biosynthetic process (GO:0070814), SIR (GO:0050311) together with hydrogenase-related functions (GO:0033748; GO:0008901), arginine metabolism (GO:0016990; GO:0019547), and cysteine biosynthesis process (GO:0019344) were also enriched (p-value < 0.05) in accessory gene families (**Figure 5**). These functions and pathways probably reflected the adaptation of *Metallosphaera* spp. to acidic, sulfur-rich, heat, and metal-laden environments. *Metallosphaera* may take advantage of hydrogen sulfide for cysteine biosynthesis. Disulfide bonds in thermophilic proteins are omnipresent in thermophiles and help stabilize proteins against the harsh conditions (Beeby et al., 2005). In addition, cysteine-rich proteins such as disulfide oxidoreductase can be used by microorganisms to chelate heavy metal ions in the cytoplasm so as to reduce metal-induced reactive oxygen species (ROS), which is supported by previous results that Cu^{2+} exposure induced assimilatory sulfur metabolism for cysteine biosynthesis in *Metallosphaera* (Wheaton et al., 2016; Meslé et al., 2017). Interestingly, functions enriched (p-value < 0.05) in strain-specific gene families of *M. yellowstonensis* MK1 were related to aromatic compound metabolism such as catabolic process of benzoate (GO:0018623; GO:0043640), phthalate (GO:0046239; GO:0018796), naphthoate (GO:0018582), toluene (GO:0042203), and phenanthrene (GO:0042216) (**Figure 5**), which suggested that *M. yellowstonensis* MK1 was apt at utilizing a broader spectrum of organic carbon sources possessing great potential in bioremediation. Strain-specific gene families of *M. tengchongensis* Ric-A were enriched (p-value < 0.05) in methyltransferase-related functions (**Figure 5**); these genes are probably involved in arsenic detoxification by catalyzing the formation of volatile trimethylarsine from arsenite (Qin et al., 2006; Ai et al., 2017).

To decipher the evolutionary histories of the *Metallosphaera* species, gene family gain, loss, expansion, and contraction events were predicted by mapping the identified gene families onto the core-gene tree (**Figure 6**). A large number of gene family gain events occurred at node 1, accounting for ∼6% of gene families, and at the branches of *M. yellowstonensis* and *M. sedula*, accounting for ∼10 and 6% of gene families, respectively. By means of gene acquisition, members of *Metallosphaera* have largely expanded their genetic diversity, resulting in functional divergence, which was similar to other Archaea (Brügger et al., 2002). Of gene families undergoing gain events, about half of them were poorly characterized, and a considerable proportion of them were related to COG (X) Mobilome: prophages, transposons (∼6%), COG (K) Transcription (∼3%),

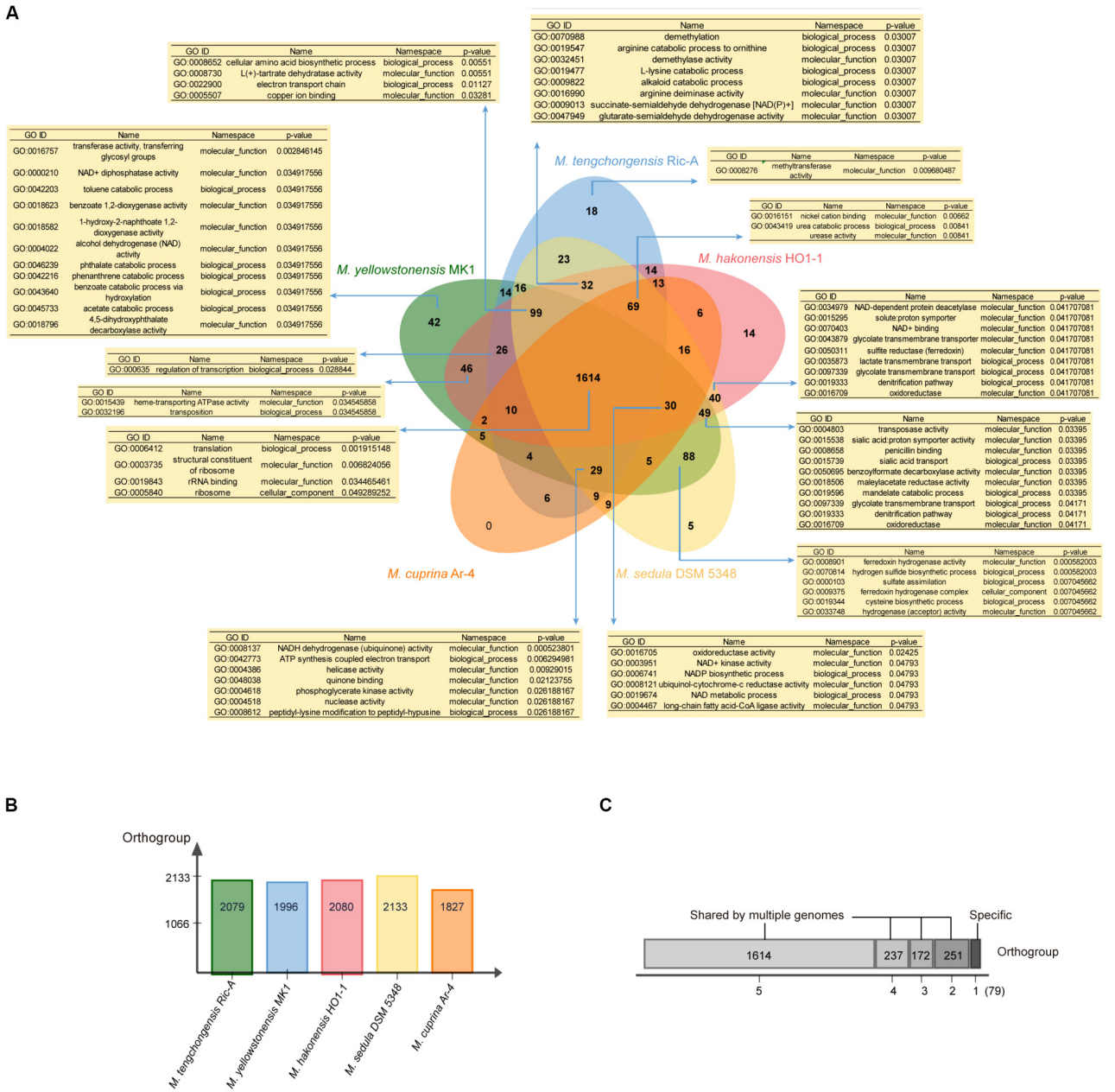

FIGURE 5 | (A) Venn diagram showing the numbers of orthogroup including the core-genome (shared by all five *Metallosphaera* type strains), the accessory genome (shared by some but not at all of these strains), and strain-specific gene families in individual genomes; enriched Gene ontology (GO) terms ($p < 0.05$) in respective orthogroups are shown in the tables. **(B)** Bar chart showing the total numbers of orthogroup identified in each genome of the five *Metallosphaera*-type strains. **(C)** The total numbers of orthogroups of strain-specific gene families (=1) or the accessory genome (shared by 2, 3, and 4 strains) and core-genome of five *Metallosphaera*-type strains.

and COG (V) Defense mechanisms (∼3%, which mostly were associated with CRISPR-Cas system) (**Figure 6**). Several sulfate assimilation and archaeal flagella biosynthesis-related genes were gained at the branches of *M. yellowstonensis* and *M. sedula*, indicating that sulfate assimilation and flagella biosynthesis were derived features of these strains (**Supplementary Table S4**). Several DNA damage repair genes (e.g., *uve* and *spl*) that helped maintain DNA fidelity were gained at the branches of *M. hakonensis*, *M. tengchongensis*, *M. sedula*, *M. yellowstonensis*, and node 1, and genes associated with EPS synthesis (e.g., *rfa* and *gal*) that enhanced colonization were gained in the branches of *M. sedula* and *M. yellowstonensis* (**Supplementary Table S4**). Several oxidoreductase encoding genes that associated with carbohydrate metabolism (e.g., *porAB* and *acoAB*) were gained at the branches of *M. sedula* and *M. tengchongensis*. A few of oxidoreductase genes related to aromatic compound

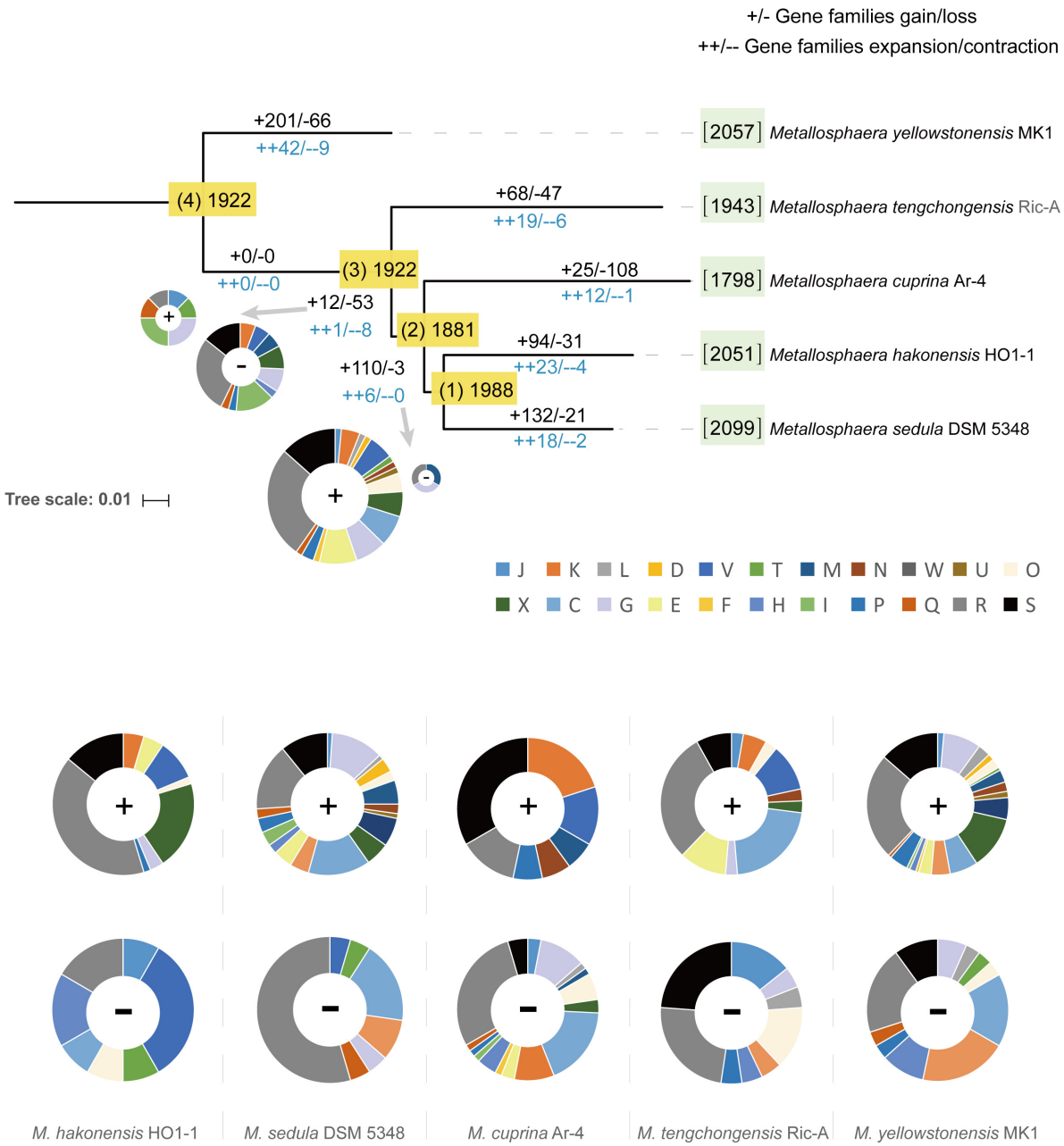

FIGURE 6 | Ancestral genome content reconstruction of *Metallosphaera* species with COUNT software. The numbers of gene families of each strain were shown in brackets before the names of strains. The numbers of gain and loss events were marked at each lineage of the tree. "+"s represent gain events and "–"s represent loss events. The pie chart shows the numbers of gained genes by COG categories. A list of gained and lost genes for the *Metallosphaera*-type strains, and nodes 1 and 2 are shown in **Supplementary Table S4**. COG categories description (J, translation, ribosomal structure and biogenesis; K, transcription; L, replication, recombination and repair; D, cell cycle control, cell division, chromosome partitioning; V, defense mechanisms; T, signal transduction mechanisms; M, cell wall/membrane/envelope biogenesis; N, cell motility; U, intracellular trafficking, secretion, and vesicular transport; O, posttranslational modification, protein turnover, chaperones; X, mobilome: prophages, transposons; C, energy production and conversion; G, carbohydrate transport and metabolism; E, amino acid transport and metabolism; F, nucleotide transport and metabolism; H, coenzyme transport and metabolism; I, lipid transport and metabolism; P, inorganic ion transport and metabolism; Q, secondary metabolites biosynthesis, transport and catabolism; R, general function prediction only; S, function unknown).

degradation (e.g., *hca* and *nfn*) were gained at the branches of *M. yellowstonensis*, *M. sedula*, and node 2. Genes encoding for putative type I CODH were expanded in *M. sedula* but contracted in *M. yellowstonensis* (**Supplementary Table S4**). Sulfocyanin encoding genes were expanded in *M. hakonensis*, *M. tengchongensis*, and Heme/copper-type cytochrome/quinol oxidase encoding genes associated with sulfur and iron oxidation (**Supplementary Table S4**) were expanded in the branches of

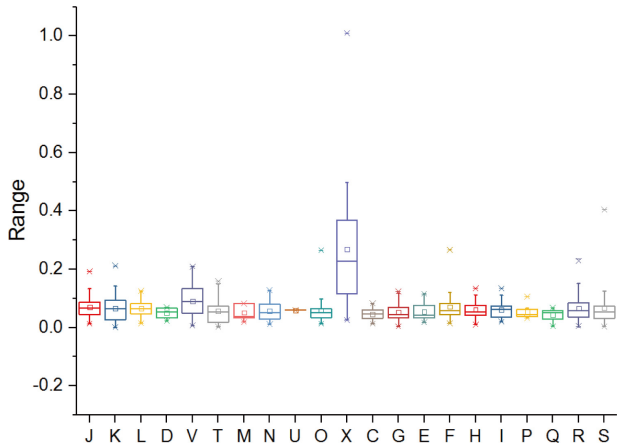

FIGURE 7 | Distribution of selection pressures (indicated by values of dN/dS) on various gene families of *Metallosphaera* spp. based on COG categories. dN, non-synonymous (amino acid-changing) substitutions; dS, synonymous (amino acid-unchanging) substitutions. dN:dS < 1 indicates the action of purifying selection (negative selection), which purges deleterious mutations to conserve the protein structure and function (Carretero-Paulet et al., 2015).

node 1 and *M. yellowstonensis*. Indicating their importance in niche adaption, several genes involved in nitrogen transport and metabolism were expanded in several species: ammonia permease encoding gene *amtB* in *M. cuprina* and *M. tengchongensis*, nitrite reductase encoding gene *nirD* in *M. hakonensis* and *M. sedula*, and adenosylmethionine decarboxylase encoding gene *speD* involved in polyamine biosynthesis in *M. yellowstonensis*. However, the urease encoding gene operon *ureABCDEF* was lost at the branch of *M. yellowstonensis* (**Supplementary Table S4**). In contrast to conspicuous gene acquisitions in other *Metallosphaera* strains, gene family loss events occurred frequently in *M. cuprina* taking up ∼6% of gene families, which mostly compromised genes belonging to COGs Energy production and conversion (C) and Carbohydrate transport and metabolism (G) such as permease of fucose, sugar, arabinose, dehydrogenase of succinate, tartrate, malate/L-lactate, aldehyde and altronate, and oxidase of sulfite (**Supplementary Table S4**). Genomic streamlining in adaptation to the acidic, thermal, and oligotrophic environment may be the main reason for gene family losses in *M. cuprina*; previous studies showed that growth temperature was negatively correlated with genome size in bacteria (Sabath et al., 2013) and that deletion of dispensable sequences from bacterial genomes led to dose-dependent growth (Kurokawa et al., 2016; Zhang et al., 2017; Ren et al., 2019).

Selective Pressure Analyses

Functions undergoing rapid evolution can be distinguished, taking into account the functional categories and selective pressure (Carretero-Paulet et al., 2015; Zhong et al., 2018). To gain insight into the conservation and evolution of various gene families in *Metallosphaera*, evolutionary pressure on each gene family that contained at least three non-identical sequences in five species of *Metallosphaera* was measured by calculating global substitution rates (dN/dS) of non-synonymous (dN) to synonymous (dS), as well as numbers of mutation sites under significant negative or positive selection in each gene family. Results indicated that there was pervasive strong purifying (negative) selection (dN/dS < 1) acting on the gene families of *Metallosphaera* with no gene families showing positive selection (dN/dS > 1), among them 86.6% gene families showing a dN/dS ratio lower than 0.1 (**Figure 7** and **Supplementary Table S5**), which emphasized purifying selection contributing largely to the long-term stability of *Metallosphaera* genomes by removing deleterious mutations. Gene families of *Metallosphaera* exhibit different degrees of purifying selection pressure as shown by COG annotation. Genes undergoing the strongest purifying selection were those related to Transcription (K), indicating that these functions were highly conserved, while genes related to mobilome: prophages, transposons (X) and Defense mechanisms (V) were under weaker purifying pressure since they showed relatively higher dN/dS ratios and more sites under diversifying (positive) selection, which indicated that these genes could gain specific adaptive mutations associated with the acquisition of new or adaptive functions (**Figure 7** and **Supplementary Table S5**).

Genome Expansion Through HGT

Horizontal gene transfer is one of the important engines that drive genomic diversity and adaptive evolution of microbes, especially in the case of those that inhabit extreme environments (Olga et al., 2009; Zhang et al., 2017; Li et al., 2019), which in this case also contributed considerably to genome contents of *Metallosphaera*. Based on COG annotation, the transferred genes in *Metallosphaera* spp. comprised mostly defensive and metabolic functions, with approximately 6.7% defense mechanisms (V), 6.4% energy production and conversion (C), 6.4% carbohydrate transport and metabolism (G), 4.9% amino acid transport and metabolism (E), 3.9% coenzyme transport and metabolism (H), 3.2% inorganic ion transport and metabolism (P), 2.1% lipid transport and metabolism (I), and 1.1% secondary metabolites biosynthesis, transport, and catabolism (Q) (**Supplementary Figure S13**). Informational proteins such as ribosomal proteins and RNA processing proteins experienced fewer HGT events in comparison with other gene families (**Supplementary Figure S13**). Most HGT events appeared to be acquired from the same domain (Archaea), which is typical among the identified HGTs (**Figure 8** and **Supplementary Table S6**). Bacteria also contributed largely to the emerging genetic diversity of *Metallosphaera* through cross-domain HGT, and acquired genes such as hydrogenase and benzoate/toluate dioxygenase also appeared to facilitate adaption to different niches, resulting in the functional divergence within the genus of *Metallosphaera* (**Figure 8** and **Supplementary Table S6**). However, it should be noted that though metagenome-assembled genomes (MAGs) included in this study have supplemented available genomic data of *Metallosphaera* and expanded sample size, these MAGs may contain contamination generated during the binning process. Thus, the results of MAGs should be interpreted carefully.

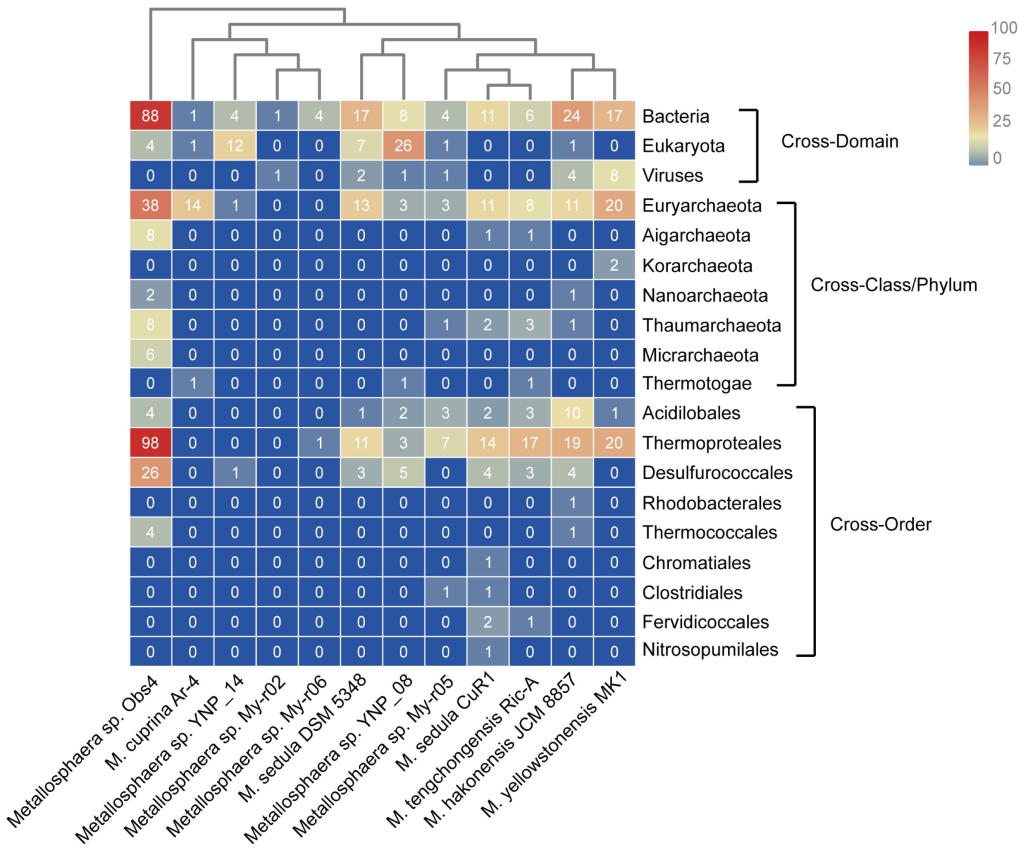

FIGURE 8 | Heat map showing the number and distribution of predicted horizontally transferred genes in different *Metallosphaera* strains and their predicted donors (right).

CONCLUSION

In this study, the genome of a *Metallosphaera* species (*M. tengchongensis*) was comparatively analyzed with 18 other genomes from *Metallosphaera* strains to enrich our understanding of the genetic traits, metabolism, and environmental adaption mechanism of the *Metallosphaera* and shed light on their evolutionary history. *Metallosphaera* microbes were widely found in sulfur-rich and metal-burdened environments, and all *Metallosphaera* strains analyzed in this study contained sulfur and iron oxidation genes. However, the APS pathway was only detected in *M. sedula* and *M. yellowstonensis*, and certain subunits of the *fox* cluster were lost in *M. cuprina*. Gene deficiency of SOR hinted that new unusual enzymes might undertake the S^0 oxidation role. Complete TCA cycles and ED pathways coexisted with HP/HB and DC/HB cycles in all genomes of these strains and supported their heterotrophic, autotrophic, and mixotrophic growth modes. The genes for assimilatory nitrate reduction were present in all genomes of 19 strains; however, a complete dissimilatory nitrate reductase gene cluster was only found in *M. yellowstonensis*, which demonstrated that *Metallosphaera* species and strains have different abilities to use inorganic nitrogen. Genes encoding for flagellin and flagella accessory proteins were only detected in strains of *M. sedula*, *M. yellowstonensis*, *Metallosphaera* sp. Obs4, *Metallosphaera* sp. UBA165, and *Metallosphaera* sp. My_r06. We also found that functions related to assimilatory sulfur metabolism and cysteine biosynthesis associated with ROS reduction were significantly enriched in accessory gene families. Evolutionary analyses showed that massive gene family gain events occurred at the branches of *M. yellowstonensis* and *M. sedula*, whereas considerable gene family loss events occurred in *M. cuprina* and pervasive strong purifying selection was found acting on the gene families of *Metallosphaera*. We also found that HGT played an important role in shaping the genetic and functional diversity of *Metallosphaera*. These findings provide a data basis for subsequent studies of metabolism and environmental adaption mechanisms in thermophilic Archaea, and for strategies to design cellular biocatalysts for the biomining process.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the genome sequence of Metallosphaera tengchongensis strain Ric-A has been deposited at JGI IMG-ER database under the IMG

Taxon OID 2821472399, and Genbank database under accession number CP049074.

AUTHOR CONTRIBUTIONS

C-YJ, S-JL, and HY designed and coordinated the study. PW, LZL, and LJL performed the bioinformatics analysis. YQ, ZL, and XL carried out the experiments and interpreted data for the work. PW and LZL wrote the manuscript. C-YJ and S-JL edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (grants 31670124 and 31600040), the Major Research Plan of National Natural Science Foundation of China (grant 91851206), Key Research Program of Chinese Academy of Sciences (ZDRW-ZS-2018-1), Basic Work of Science and Technology in the Ministry of Science and Technology (grant 2015FY110100), and the National Basic Research Program of China (2014CB846000).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.01192/full#supplementary-material

FIGURE S1 | Phylogenetic tree of *Metallosphaera* spp. with **(A)** Maximum Likelihood (ML) method; **(B)** Minimum Evolution (ME) method; **(C)** UPGMA method; **(D)** Maximum Parsimony (MP) method based on 16S rRNA sequences. Bootstrap values are indicated at each node based on a total of 1000 bootstrap replicates.

FIGURE S2 | Whole genome alignment with Mauve showing syntenic blocks between type strains *M. tengchongensis* Ric-A, *M. cuprina* Ar-4, *M. sedula* DSM 5348, and *M. hakonensis* HO1-1 (from top to bottom).

FIGURE S3 | BlastN-based whole genome comparison of strains *M. cuprina* Ar-4, *M. hakonensis* HO1-1, *Metallosphaera* sp. My-r06,*M. sedula* DSM 5348, *M. tengchongensis* Ric-A, and *M. yellowstonensis* MK1 (from inside to outside) with BRIG, and these strains were used as reference, respectively. GC content and GC skew (inner rings 1 and 2) of each genome were also indicated.

FIGURE S4 | Phylogenetic tree of *Metallosphaera* spp. rooted by *Encephalitozoon cuniculi* GB M1 and *Prochlorococcus marinus* MIT 9301 based on whole genome sequences using CVTree 3.

FIGURE S5 | Phylogenetic tree of *Metallosphaera* spp. and other members of *Sulfolobales* based on whole genome sequences using CVTree 3.

FIGURE S6 | Heat map showing the distribution and number of genes encoding putative hydrogenase related proteins in different *Metallosphaera* strains.

FIGURE S7 | Heat map showing the distribution and numbers of genes encoding putative proteins involved in iron oxidation in different *Metallosphaera* strains.

FIGURE S8 | Heat map showing the distribution and numbers of genes encoding carbohydrate active enzymes (CAZymes) in different *Metallosphaera* strains. Abbreviations in the chart: *AA*, auxiliary activities proteins; *CBM*, carbohydrate-binding molecules; *CE*, carbohydrate esterases; *GH*, glycoside hydrolases; *GT*, glycosyltransferases; *PL*, polysaccharide lyases.

FIGURE S9 | Heat map showing the distribution and numbers of genes encoding putative enzymes involved in carbon metabolism in different *Metallosphaera* strains. Abbreviation and KO term descriptions in the chart: *CoxL/M/S*, Carbon monoxide dehydrogenase large/medium/small chain; *GalE*, UDP-glucose 4-epimerase; *RfbA*, d-TDP-glucose pyrophosphorylase; *RfbB*, d-TDP-glucose 4;6-dehydratase; *RfbC*, d-TDP-4-dehydrorhamnose 3;5-epimerase; *RfbD*, d-TDP-4-dehydrorhamnose reductase; *K01895*, acetyl-CoA synthetase; *K00172*, pyruvate ferredoxin oxidoreductase gamma subunit; *K00626*, acetyl-CoA C-acetyltransferase; *K01681* aconitate hydratase; *K01007*, pyruvate; water dikinase (EC:2.7.9.2); *K01964* acetyl-CoA/propionyl-CoA carboxylase (EC:6.4.1.2); *K15016*, enoyl-CoA hydratase; *K15017*, malonyl-CoA/succinyl-CoA reductase; *K01595*, phosphoenolpyruvate carboxylase; *K15039* 3-hydroxypropionate dehydrogenase (NADP+) (EC:1.1.1.298); *K15018*, 3-hydroxypropionyl-coenzyme A synthetase; *K15019* 3-hydroxypropionyl-coenzyme A dehydratase; *K14534*, -hydroxybutyryl-CoA dehydratase; *K00031*, isocitrate dehydrogenase; *K00024*, malate dehydrogenase; *K15020*, acryloyl-coenzyme A reductase; *K14466*, 4-hydroxybutyrate-CoA ligase; *K05606*, methylmalonyl-CoA; *K00175*, 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase; *K01678*, fumarate hydratase; *K01848* methylmalonyl-CoA mutase; *K18859*, succinate dehydrogenase/fumarate reductase; *K01902*, succinyl-CoA synthetase.

FIGURE S10 | Heat map showing the distribution and numbers of genes encoding putative proteins involved in nitrogen metabolism in different *Metallosphaera* strains. Abbreviation descriptions in the chart: *GdhA*, glutamate dehydrogenase [NAD(P)+] (EC 1.4.1.3); *GlnA*, glutamine synthetase (EC 6.3.1.2); *CarA*, carbamoyl-phosphate synthase small subunit (EC 6.3.5.5); *CarB*, carbamoyl-phosphate synthase large subunit (EC 6.3.5.5); *NarG*, nitrate reductase/nitrite oxidoreductase, alpha subunit (EC 1.7.5.1; 1.7.99.4); *NarH*, nitrate reductase/nitrite oxidoreductase, beta subunit (EC 1.7.5.1; 1.7.99.4); *NarJ*, nitrate reductase delta subunit; *Nar*I, nitrate reductase gamma subunit (EC 1.7.5.1; 1.7.99.4); *Nrt*, nitrate/nitrite transporter; *Amt*, ammonium transporter; *UreAB*, *UreCD*, urease functional subunits; *UreE*, *UreF*, *UreG*, urease accessory proteins; *ArgH*, argininosuccinate lyase; *ArgI*, ornithine carbamoyltransferase; *ArcA*, arginine decarboxylase; *ArgG*, argininosuccinate synthase; *SpeB*, agmatinase; *SpeD*, S-adenosylmethionine decarboxylase; *SpeE*, spermidine synthase.

FIGURE S11 | Heat map showing patterns of the distribution and numbers of genes encoding putative proteins involved in heavy metal resistance in different *Metallosphaera* strains.

FIGURE S12 | Heat map showing patterns of the distribution and numbers of genes encoding putative proteins involved in flagella biosynthesis in different *Metallosphaera* strains.

FIGURE S13 | The COG proportions of predicted horizontal transferred genes in *Metallosphaera* spp.

TABLE S1 | Average nucleotide identity (ANI) based on whole genome alignments of *Metallosphaera* spp.

TABLE S2 | Statistics of genomic islands in the genomes of *Metallosphaera* spp.

TABLE S3 | Statistics of Mobile genetic elements (MGEs) (A), prophages and prophage remnants (B), Transposons (C), CRISPR-Cas related sequences (D) present in the genomes of *Metallosphaera* spp.

TABLE S4 | Gene families undergoing gain (A), loss (B), expansion (C), and contraction (D) events in *Metallosphaera* spp. detected with COUNT.

TABLE S5 | Selection pressure on gene families of Metallosphaera spp.

TABLE S6 | Putative horizontally transferred genes of *Metallosphaera* spp.

REFERENCES

Abby, S. S., Melcher, M., Kerou, M., Krupovic, M., Stieglmeier, M., Rossel, C., et al. (2018). Candidatus *Nitrosocaldus cavascurensis*, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front. Microbiol.* 9:28. doi: 10.3389/fmicb.2018.00028

Ai, C., McCarthy, S., Eckrich, V., Rudrappa, D., Qiu, G., and Blum, P. (2016). Increased acid resistance of the archaeon, *Metallosphaera sedula* by adaptive laboratory evolution. *J. Ind. Microbiol. Biotechnol.* 43, 1–11.

Ai, C., McCarthy, S., Liang, Y., Rudrappa, D., Qiu, G., and Blum, P. (2017). Evolution of copper arsenate resistance for enhanced enargite bioleaching using the extreme thermoacidophile *Metallosphaera sedula*. *J. Ind. Microbiol. Biotechnol.* 44, 1613–1625. doi: 10.1007/s10295-017-1973-5

Alber, B., Olinger, M., Rieder, A., Kockelkorn, D., Jobst, B., Hügler, M., et al. (2006). Malonyl-coenzyme a reductase in the modified 3-hydroxypropionate cycle for autotrophic carbon fixation in archaeal *Metallosphaera* and *Sulfolobus* spp. *J. Bacteriol.* 188, 8551–8559. doi: 10.1128/jb.00987-06

Alikhan, N. F., Petty, N. K., Zakour, N. L. B., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Amend, J. P., and Shock, E. L. (2001). Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and bacteria. *FEMS Microbiol. Rev.* 25, 175–243. doi: 10.1111/j.1574-6976.2001.tb00576.x

Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* 44, W16–W21.

Auernik, K. S., and Kelly, R. M. (2008). Identification of components of electron transport chains in the extremely thermoacidophilic crenarchaeon *Metallosphaera sedula* through iron and sulfur compound oxidation transcriptomes. *Appl. Environ. Microbiol.* 74, 7723–7732. doi: 10.1128/aem.01545-08

Auernik, K. S., and Kelly, R. M. (2010). Impact of molecular hydrogen on chalcopyrite bioleaching by the extremely thermoacidophilic archaeon *Metallosphaera sedula*. *Appl. Environ. Microbiol.* 76, 2668–2672. doi: 10.1128/aem.02016-09

Auernik, K. S., Maezato, Y., Blum, P. H., and Kelly, R. M. (2008). The genome sequence of the metal-mobilizing, extremely thermoacidophilic Archaeon *Metallosphaera sedula* provides insights into bioleaching-associated metabolism. *Appl. Environ. Microbiol.* 74, 682–692. doi: 10.1128/aem.02019-07

Basak, G., Lakshmi, V., Chandran, P., and Das, N. (2014). Removal of Zn(ii) from electroplating effluent using yeast biofilm formed on gravels: batch and column studies. *J. Environ. Health Sci. Eng.* 12:8.

Bathe, S., and Norris, P. R. (2007). Ferrous iron- and sulfur-induced genes in *Sulfolobus metallicus*. *Appl. Environ. Microbiol.* 73, 2491–2497. doi: 10.1128/aem.02589-06

Beeby, M., O'Connor, B. D., Ryttersgaard, C., Boutz, D. R., and Yeates, T. O. (2005). The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* 3:e309. doi: 10.1371/journal.pbio.0030309

Berg, I. A., Daniel, K., Wolfgang, B., and Georg, F. (2007). A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* 318, 1782–1786. doi: 10.1126/science.1149976

Berg, I. A., Ramos-Vera, W. H., Petri, A., Huber, H., and Fuchs, G. (2010). Study of the distribution of autotrophic CO_2 fixation cycles in crenarchaeota. *Microbiology* 156, 256–269. doi: 10.1099/mic.0.034298-0

Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., et al. (2017). IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res.* 45, 30–35.

Blake, W., Jesse, M., Deborah, W., Mark, Y., Dryden, K. A., Young, M., et al. (2005). An archaeal antioxidant: characterization of a Dps-like protein from *Sulfolobus solfataricus*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10551–10556. doi: 10.1073/pnas.0501497102

Brügger, K., Redder, P., She, Q. X., Confalonieri, F., Zivanovic, Y. V., and Garrett, R. A. (2002). Mobile elements in archaeal genomes. *FEMS Microbiol. Lett.* 206, 131–141. doi: 10.1016/s0378-1097(01)00504-3

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Carretero-Paulet, L., Chang, T. H., Librado, P., Ibarra-Laclette, E., Herrera-Estrella, L., Rozas, J., et al. (2015). Genome-wide analysis of adaptive molecular evolution in the carnivorous plant *Utricularia gibba*. *Genome Biol. Evol.* 7, 444–456. doi: 10.1093/gbe/evu288

Castro, C., and Donati, E. (2016). Effects of different energy sources on cell adhesion and bioleaching of a chalcopyrite concentrate by extremophilic archaeon *Acidianus copahuensis*. *Hydrometallurgy* 162, 49–56. doi: 10.1016/j.hydromet.2016.02.014

Chaudhari, N. M., Kumar, G. V., and Chitra, D. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373.

Chen, I. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Huang, J., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, D507–D573.

Chen, L. X., Ren, Y. L., Lin, J. Q., Liu, X. M., Pang, X., Lin, J. Q., et al. (2012). *Acidithiobacillus caldus* sulfur oxidation model based on transcriptome analysis between the wild type and sulfur oxygenase reductase defective mutant. *PLoS One* 7:e39470. doi: 10.1371/journal.pone.0039470

Chen, Z. W., Jiang, C. Y., She, Q., Liu, S. J., and Zhou, P. J. (2005). Key role of cysteine residues in catalysis and subcellular localization of sulfur oxygenase reductase of *Acidianus tengchongensis*. *Appl. Environ. Microbiol.* 71, 621–628. doi: 10.1128/aem.71.2.621-628.2005

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read smrt sequencing data. *Nat. Methods* 10, 563–569. doi: 10.1038/nmeth.2474

Christiane, D., Bettina, F., Daniela, H., Anne, K., and Renate, Z. (2013). Sulfite oxidation in the purple sulfur bacterium *Allochromatium vinosum*: identification of SoeABC as a major player and relevance of SoxYZ in the process. *Microbiology* 159, 2626–2638. doi: 10.1099/mic.0.071019-0

Couvin, D., Bernheim, A., Toffanonioche, 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.

Dai, X., Wang, H., Zhang, Z., Li, K., Zhang, X., Mora-López, M., et al. (2016). Genome sequencing of *Sulfolobus* sp. A20 from costa rica and comparative analyses of the putative pathways of carbon, nitrogen, and sulfur metabolism in various *Sulfolobus* strains. *Front. Microbiol* 7:1902. doi: 10.3389/fmicb.2016.01902

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R., and Meyer, O. (2002). Catalysis at a dinuclear [cusmo(=o)oh] cluster in a co dehydrogenase resolved at 1.1-a resolution. *Proc. Natl. Acad. Sci. U.S.A.* 99, 15971–15976. doi: 10.1073/pnas.212640899

Dominik, E., Lena, H., Khoa, P. T., Bräsen, C., Wen, Q., Wright, P. C., et al. (2016). Protein phosphorylation and its role in archaeal signal transduction. *FEMS Microbiol. Rev.* 40, 625–647. doi: 10.1093/femsre/fuw020

Dopson, M., Baker-Austin, C., and Bond, P. L. (2005). Analysis of differential protein expression during growth states of *Ferroplasma* strains and insights into electron transport for iron oxidation. *Microbiology* 151, 4127–4137. doi: 10.1099/mic.0.28362-0

Farris, J. S. (1970). Methods for computing Wagner trees. *Syst. Zool.* 19, 83–92. doi: 10.1093/sysbio/19.1.83

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, 222–230.

Fuchs, T., Huber, H., Teiner, K., Burggraf, S., and Stetter, K. O. (1995). *Metallosphaera prunae*, sp. nov. a novel metal-mobilizing, thermoacidophilic archaeum, isolated from a uranium mine in Germany. *Syst. Appl. Microbiol.* 8, 560–566. doi: 10.1016/s0723-2020(11)80416-9

Giles, T. N., and Graham, D. E. (2008). Crenarchaeal arginine decarboxylase evolved from an s-adenosylmethionine decarboxylase enzyme. *J. Biol. Chem.* 283, 25829–25838. doi: 10.1074/jbc.m802674200

Han, C. J., and Kelly, R. M. (2015). Biooxidation capacity of the extremely thermoacidophilic archaeon *Metallosphaera sedula* under bioenergetic challenge. *Biotechnol. Bioeng.* 58, 617–624. doi: 10.1002/(sici)1097-0290(19980620)58:6<617::aid-bit7>3.0.co;2-l

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.

Huber, G., Spinnler, C., Gambacorta, A., and Stetter, K. O. (1989). *Metallosphaera sedula* gen, and sp. nov. represents a new genus of aerobic, metal-mobilizing,

Thermoacidophilic Archaebacteria. *Syst. Appl. Microbiol.* 12, 38–47. doi: 10.1016/s0723-2020(89)80038-4

Jiang, C. Y., Liu, L. J., Guo, X., You, X. Y., Liu, S. J., and Poetsch, A. (2014). Resolution of carbon metabolism and sulfur-oxidation pathways of *Metallosphaera cuprina* Ar-4 via comparative proteomics. *J. Proteomics* 109, 276–289. doi: 10.1016/j.jprot.2014.07.004

King, G. M., and Weber, C. F. (2007). Distribution, diversity and ecology of aerobic CO-oxidizing bacteria. *Nat. Rev. Microbiol.* 5, 277–310.

Kletzin, A. (1992). Molecular characterization of the *sor* gene, which encodes the sulfur oxygenase/reductase of the thermoacidophilic Archaeum *Desulfurolobus ambivalens*. *J. Bacteriol.* 174, 5854–5859. doi: 10.1128/jb.174.18.5854-5859.1992

Koonin, E. V., and Makarova, K. S. (2017). Mobile genetic elements and evolution of CRISPR-Cas systems: all the way there and back. *Genome Biol. Evol.* 9, 2812–2825. doi: 10.1093/gbe/evx192

Kosakovsky Pond, S. L., and Frost, S. D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105

Kozubal, M., Macur, R. E., Korf, S., Taylor, W. P., Ackerman, G. G., Nagy, A., et al. (2008). Isolation and distribution of a novel iron-oxidizing crenarchaeon from acidic geothermal springs in Yellowstone national park. *Appl. Environ. Microbiol.* 74, 942–949. doi: 10.1128/aem.01200-07

Kozubal, M. A., Dlakic, M., Macur, R. E., and Inskeep, W. P. (2011). Terminal oxidase diversity and function in "*Metallosphaera yellowstonensis*": gene expression and protein modeling suggest mechanisms of Fe(II) oxidation in the *Sulfolobales*. *Appl. Environ. Microbiol.* 77, 1844–1853. doi: 10.1128/aem.01646-10

Krupovic, M., Béguin, P., and Koonin, E. V. (2017). Casposons: mobile genetic elements that gave rise to the CRISPR-Cas adaptation machinery. *Curr. Opin. Microbiol.* 38, 36–43. doi: 10.1016/j.mib.2017.04.004

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Kumar, S., Stecher, G., Peterson, D., and Tamura, K. (2012). MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis. *Bioinformatics* 28, 2685–2686. doi: 10.1093/bioinformatics/bts507

Kurokawa, M., Seno, S., Matsuda, H., and Ying, B. W. (2016). Correlation between genome reduction and bacterial growth. *DNA Res.* 23, 517–525. doi: 10.1093/dnares/dsw035

Laska, S., Lottspeich, F., and Kletzin, A. (2003). Membrane-bound hydrogenase and sulfur reductase of the hyperthermophilic and acidophilic archaeon *Acidianus ambivalens*. *Microbiology* 149, 2357–2371. doi: 10.1099/mic.0.26455-0

Lassak, K., Neiner, T., Ghosh, A., Klingl, A., Wirth, R., and Albers, S. V. (2012). Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.* 83, 110–124. doi: 10.1111/j.1365-2958.2011.07916.x

Le, H., Han, Z., Wu, P., Entwistle, S., Li, X., Tanner, Y., et al. (2018). DbCAN-seq: a database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* 46, 516–521.

Li, L., Liu, Z., Meng, D., Liu, X., Li, X., Ming, Z., et al. (2019). Comparative genomic analysis reveals the distribution, organization, and evolution of metal resistance genes in the genus *Acidithiobacillus*. *Appl. Environ. Microbiol.* 85:e02153-18. doi: 10.1128/AEM.02153-18

Liu, L. J., You, X. Y., Guo, X., Liu, S. J., and Jiang, C. Y. (2011a). *Metallosphaera cuprina* sp. nov., an acidothermophilic, metal-mobilizing archaeon. *Int. J. Syst. Evol. Microbiol.* 61, 2395–2400. doi: 10.1099/ijs.0.026591-0

Liu, L. J., You, X. Y., Zheng, H., Wang, S., Jiang, C. Y., and Liu, S. J. (2011b). Complete genome sequence of *Metallosphaera cuprina*, a metal sulfide-oxidizing archaeon from a hot spring. *J. Bacteriol.* 193, 3387–3388. doi: 10.1128/jb.05038-11

Marino, M., Maifreni, M., and Baggio, A. (2018). Innocente N1 inactivation of foodborne bacteria biofilms by aqueous and gaseous ozone. *Front. Microbiol.* 28:2024. doi: 10.3389/fmicb.2018.02024

Melnick, J. G., and Gerard, P. (2007). Cleaving mercury-alkyl bonds: a functional model for mercury detoxification by MerB. *Science* 317, 225–227. doi: 10.1126/science.1144314

Meslé, M., Beam, J. P., Jay, Z. J., Bodle, B., Bogenschutz, E. L., and Inskeep, W. P. (2017). Hydrogen peroxide cycling in high-temperature acidic geothermal springs and potential implications for oxidative stress response. *Front. Mar. Sci.* 4:130. doi: 10.3389/fmars.2017.00130

Michael, H., Krieger, R. S., Martina, J., and Georg, F. (2010). Characterization of acetyl-CoA/propionyl-CoA carboxylase in *Metallosphaera sedula*. Carboxylating enzyme in the 3-hydroxypropionate cycle for autotrophic carbon fixation. *Eur. J. Biochem.* 270, 736–744. doi: 10.1046/j.1432-1033.2003.03434.x

Miklós, C. (2010). Count: evolutionary analysis of phylogenetic profiles with parsimony and likelihood. *Bioinformatics* 26, 1910–1912. doi: 10.1093/bioinformatics/btq315

Olga, Z., Swithers, K. S., Pascal, L., Fournier, G. P., Bickhart, D. M., DeBoy, R. T., et al. (2009). On the chimeric nature, thermophilic origin, and phylogenetic placement of the Thermotogales. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5865–5870. doi: 10.1073/pnas.0901260106

Orell, A., Navarro, C. A., Arancibia, R., Mobarec, J. C., and Jerez, C. A. (2010). Life in blue: copper resistance mechanisms of bacteria and Archaea used in industrial biomining of minerals. *Biotechnol. Adv.* 28, 839–848. doi: 10.1016/j.biotechadv.2010.07.003

Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. (2014). BacMet: antibacterial biocide and metal resistance genes database. *Nucleic Acids Res.* 42, 737–743.

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/s41564-017-0012-7

Peeples, T. L., and Kelly, R. M. (1995). Bioenergetic response of the extreme thermoacidophile *Metallosphaera sedula* to thermal and nutritional stresses. *Appl. Environ. Microbiol.* 61, 2314–2321. doi: 10.1128/aem.61.6.2314-2321.1995

Peng, T. J., Liu, L. J., Liu, C., Yang, Z. F., Liu, S. J., and Jiang, C. Y. (2015). *Metallosphaera tengchongensis* sp. nov., an acidothermophilic archaeon isolated from a hot spring. *Int. J. Syst. Evol. Microbiol.* 65, 537–542. doi: 10.1099/ijs.0.070870-0

Peters, J. E., Makarova, K. S., Shmakov, S., and Koonin, E. V. (2017). Recruitment of CRISPR-Cas systems by Tn7-like transposons. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7358–7366.

Pond, S. L. K., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079

Qin, J., Rosen, B. P., Zhang, Y., Wang, G., Franke, S., and Rensing, C. (2006). Arsenic detoxification and evolution of trimethylarsine gas by a microbial arsenite S-adenosylmethionine methyltransferase. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2075–2080. doi: 10.1073/pnas.0506836103

Qin, Q. L., Xie, B. B., Zhang, X. Y., Chen, X. L., Zhou, B. C., Zhou, J., et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J. Bacteriol.* 196, 2210–2215. doi: 10.1128/jb.01688-14

Qin, Y. L., Liang, Z. L., Song, Y., Wang, B. J., Liu, S. J., and Jiang, C. Y. (2019). Amplicon-based high-throughput sequencing reveals the microbial diversity in Rehai hot springs, Tengchong, Yunnan Province, (in chinese). *Microbiol. China* 46, 2482–2493.

Rawlings, D. E. (2005). Characteristics and adaptability of iron- and sulfur-oxidizing microorganisms used for the recovery of metals from minerals and their concentrates. *Microbial. Cell Fact.* 4, 1–15.

Ren, M., Feng, X., Huang, Y., Wang, H., Hu, Z., Clingenpeel, S., et al. (2019). Phylogenomics suggests oxygen availability as a driving force in *Thaumarchaeota* evolution. *ISME J.* 13, 2150–2161. doi: 10.1038/s41396-019-0418-8

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

Rohwerder, T., Gehrke, T., Kinzler, K., and Sand, W. (2003). Bioleaching review part A: progress in bioleaching: fundamentals and mechanisms of bacterial metal sulfide oxidation. *Appl. Microbiol. Biotechnol.* 63, 239–248.

Sabath, N., Ferrada, E., Barve, A., and Wagner, A. (2013). Growth temperature and genome size in bacteria are negatively correlated, suggesting genomic streamlining during thermal adaptation. *Genome Biol. Evol.* 5, 966–977. doi: 10.1093/gbe/evt050

Sebastian, E., Michael, H., Wolfgang, E., Katharina, W., Berg, I. A., Ramos-Vera, W. H., et al. (2011). Labeling and enzyme studies of the central carbon metabolism in *Metallosphaera sedula*. *J. Bacteriol.* 193, 1191–1200. doi: 10.1128/jb.01155-10

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153

Selengut, J. D., Haft, D. H., Tanja, D., Anurhada, G., Michelle, G. G., Nelson, W. C., et al. (2007). TIGRFAMs and genome properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res.* 35, D260–D264.

Siguier, P., Perochon, J., Lestrade, L., Mahillon, J., and Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res.* 34, D32–D36.

Singer, S. W., Chan, C. S., Zemla, A., VerBerkmoes, N. C., Hwang, M., Hettich, R. L., et al. (2008). Characterization of cytochrome 579, an unusual cytochrome isolated from an iron-oxidizing microbial community. *Appl. Environ. Microbiol.* 74, 4454–4462. doi: 10.1128/aem.02799-07

Springael, D., and Top, E. M. (2004). Horizontal gene transfer and microbial adaptation to xenobiotics: new types of mobile genetic elements and lessons from ecological studies. *Trends Microbiol.* 12, 53–58. doi: 10.1016/j.tim.2003.12.010

Takayanagi, S., Kawasaki, H., Sugimori, K., Yamada, T., Sugai, A., Ito, T., et al. (1996). *Sulfolobus hakonensis* sp. nov., a novel species of acidothermophilic archaeon. *Int. J. Syst. Bacteriol.* 46, 377–382. doi: 10.1099/00207713-46-2-377

Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., et al. (2001). The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28. doi: 10.1093/nar/29.1.22

Thomas, N. A., Bardy, S. L., and Jarrell, K. F. (2001). The archaeal flagellum: a different kind of prokaryotic motility structure. *FEMS Microbiol. Rev.* 25, 147–174. doi: 10.1111/j.1574-6976.2001.tb00575.x

Ulrike, K., Sly, L. I., and Mcewan, A. G. (2005). Respiratory gene clusters of *Metallosphaera sedula* - differential expression and transcriptional organization. *Microbiology* 151, 35–43. doi: 10.1099/mic.0.27515-0

Urbieta, M. S., Donati, E. R., Chan, K. G., Shahar, S., Sin, L. L., and Goh, K. M. (2015). Thermophiles in the genomic era: biodiversity, science, and applications. *Biotechnol. Adv.* 33, 633–647. doi: 10.1016/j.biotechadv.2015.04.007

Urbieta, M. S., Rascovan, N., Vázquez, M. P., and Donati, E. (2017). Genome analysis of the thermoacidophilic archaeon *Acidianus copahuensis* focusing on the metabolisms associated to biomining activities. *BMC Genomics* 18:445. doi: 10.1186/s12864-017-3828-x

Urich, T., Gomes, C. M., Kletzin, A., and Frazao, C. (2006). X-ray structure of a self-compartmentalizing sulfur cycle metalloenzyme. *Science* 311, 996–999.

Wang, Y., Coleman-Derr, D., Chen, G., and Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, 78–84.

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., and Kosakovsky-Pond, S. L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777. doi: 10.1093/molbev/msx335

Wheaton, G. H., Mukherjee, A., and Kelly, R. M. (2016). Transcriptomes of the extremely thermoacidophilic archaeon *Metallosphaera sedula* exposed to metal "shock" reveal generic and specific metal responses. *Appl. Environ. Microbiol.* 82, 4613–4627. doi: 10.1128/aem.01176-16

Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444. doi: 10.1186/1471-2164-12-444

Xu, Z., and Hao, B. (2009). CVTree update: a newly designed phylogenetic study platform using composition vectors and whole genomes. *Nucleic Acids Res.* 37, W174–W178.

Yi, W., Colemanderr, D., Chen, G., and Gu, Y. Q. (2015). OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* 43, 78–84.

Yu, R. L., Liu, Z. H., Yu, Z. J., Wu, X. L., and Zeng, W. M. (2019). Relationship among the secretion of extracellular polymeric substances, heat resistance, and bioleaching ability of *Metallosphaera sedula*. *Int. J. Miner. Metall. Mater.* 26, 1504–1511. doi: 10.1007/s12613-019-1851-4

Zeldes, B. M., Loder, A. J., Counts, J. A., Haque, M., and Kelly, R. M. (2019). Determinants of sulfur chemolithoautotrophy in the extremely thermoacidophilic *Sulfolobales*. *Environ. Microbiol.* 21, 3696–3710. doi: 10.1111/1462-2920.14712

Zhang, X., Liu, X., Liang, Y., Guo, X., Xiao, Y., Ma, L., et al. (2017). Adaptive evolution of extreme acidophile *Sulfobacillus thermosulfidooxidans* potentially driven by horizontal gene transfer and gene loss. *Appl. Environ. Microbiol.* 83, e3016–e3098.

Zhong, C., Han, M., Yu, S., Yang, P., Li, H., and Ning, K. (2018). Pan-genome analyses of 24 *Shewanella* strains re-emphasize the diversification of their functions yet evolutionary dynamics of metal-reducing pathway. *Biotechnol. Biofuels* 11, 193–205.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 01 September 2020
doi: 10.3389/fmicb.2020.02082

Biochemical and Taxonomic Characterization of Novel Haloarchaeal Strains and Purification of the Recombinant Halotolerant α-Amylase Discovered in the Isolate

Dipesh Kumar Verma[1], Gunjan Vasudeva[2], Chandni Sidhu[2†], Anil K. Pinnaka[2], Senthil E. Prasad[3] and Krishan Gopal Thakur[1]**

[1] G. N. Ramachandran Protein Centre, Structural Biology Laboratory, Council of Scientific and Industrial Research-Institute of Microbial Technology, Chandigarh, India, [2] MTCC-Microbial Type Culture Collection and Gene Bank, CSIR-Institute of Microbial Technology, Chandigarh, India, [3] Biochemical Engineering Research and Process Development Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology, Chandigarh, India

OPEN ACCESS

Edited by:
Anna-Louise Reysenbach, Portland State University, United States

Reviewed by:
Changyi Zhang, University of Illinois at Urbana–Champaign, United States
Wenyuan Han, Huazhong Agricultural University, China

***Correspondence:**
Senthil E. Prasad
esprasad@imtech.res.in
Krishan Gopal Thakur
krishang@imtech.res.in

†Present address:
Chandni Sidhu, Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Bremen, Germany

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

Received: *12 March 2020*
Accepted: *07 August 2020*
Published: *01 September 2020*

Citation:
Verma DK, Vasudeva G, Sidhu C, Pinnaka AK, Prasad SE and Thakur KG (2020) Biochemical and Taxonomic Characterization of Novel Haloarchaeal Strains and Purification of the Recombinant Halotolerant α-Amylase Discovered in the Isolate. Front. Microbiol. 11:2082. doi: 10.3389/fmicb.2020.02082

Haloarchaea are salt-loving archaea and potential source of industrially relevant halotolerant enzymes. In the present study, three reddish-pink, extremely halophilic archaeal strains, namely wsp1 (wsp-water sample Pondicherry), wsp3, and wsp4, were isolated from the Indian Solar saltern. The phylogenetic analysis based on 16S rRNA gene sequences suggests that both wsp3 and wsp4 strains belong to *Halogeometricum borinquense* while wsp1 is closely related to *Haloferax volcanii* species. The comparative genomics revealed an open pangenome for both genera investigated here. Whole-genome sequence analysis revealed that these isolates have multiple copies of industrially/biotechnologically important unique genes and enzymes. Among these unique enzymes, for recombinant expression and purification, we selected four putative α-amylases identified in these three isolates. We successfully purified functional halotolerant recombinant Amy2, from wsp1 using pelB signal sequence-based secretion strategy using *Escherichia coli* as an expression host. This method may prove useful to produce functional haloarchaeal secretory recombinant proteins suitable for commercial or research applications. Biochemical analysis of Amy2 suggests the halotolerant nature of the enzyme having maximum enzymatic activity observed at 1 M NaCl. We also report the isolation and characterization of carotenoids purified from these isolates. This study highlights the presence of several industrially important enzymes in the haloarchaeal strains which may potentially have improved features like stability and salt tolerance suitable for industrial applications.

Keywords: haloarchaea, amylase, pangenome, carotenoids, halotolerant enzyme

INTRODUCTION

The extremely halophilic archaea have adapted to bloom even in harsh environmental conditions such as high salinity, desiccation, and intense solar radiations (Mormile et al., 2003; Gunde-Cimerman et al., 2005; Schubert et al., 2010; Stan-Lotter and Fendrihan, 2015; Winters et al., 2015). These microorganisms require at least 1.5–2.5 M NaCl concentration for their viability and typically

grow optimally in 3.5 M NaCl concentrations (Ollivier et al., 1994). Haloarchaea commonly resides in hypersaline environments such as salt lakes, salterns, heavily salted hides, meats, fish, and sauces (Radax et al., 2001; Gruber et al., 2004; Stan-Lotter and Fendrihan, 2015). Adaptation in such extreme and diverse environments makes their genome highly rich in multiple essential genes that are absent in other microorganisms (Papke et al., 2015). This essential new haloarchaeal gene pool analysis has the potential to uncover many industrially important proteins and enzymes. So, it is imperative to perform pangenome or comparative genomics analysis to understand the genetic evolution and distribution of unique and conserved genes in these microbes that help them to survive in harsh conditions (Kim et al., 2018). As aerophilic mesophiles, many haloarchaea are easy to grow in the laboratory conditions, making them one of the most extensively studied archaeal groups and thus, leading to the development of a variety of biochemical, genetic and genomic tools for better understanding of several diverse haloarchaeal species (Soppa, 2006).

Besides having the ability to thrive in high salt conditions, haloarchaea possess diverse physiologies including alkaliphiles, facultative thermophiles, thermoalkaliphiles, and psychrotolerant species (Bowers et al., 2009; Bowers and Wiegel, 2011) and diverse metabolic strategies. These interesting features make them ideal organisms for understanding archaeal biology (Falb et al., 2008). In addition, the genetic basis of these microbes to flourish in hypersaline environments may provide crucial insights to develop salt-tolerant plants for growth in currently non-arable land (Flowers and Colmer, 2015).

Halophiles produce a range of unique and stable biomolecules of commercial applications including (1) hydrolytic enzymes like gelatinases, proteases, lipases, DNAases, xylanases, and amylase. Exo-enzymes from these organisms with polymer degrading ability is of great interest in many industrial processes where high salt concentration would cause enzymatic inhibition (Oren, 2010). These unique features make haloarchaeal enzymes very useful in commercial industries such as baking industries (Oren, 2010), starch liquefaction (Chi et al., 2010), detergent industries, maltose production, etc. (Singh and Singh, 2017). (2) Membrane proteins such as bacteriorhodopsin, commercially recognized for its use in artificial retina, holograms, photoelectrical devices, optical computing, etc. (3) Biodegradable polymers such as polyhydroxyalkanoates (PHAs), produced by many haloarchaea can be used as an alternative to non-degradable plastics (Schiraldi et al., 2002; Oren, 2010). (4) Carotenoids: under certain growth conditions, microbial cells are known to accumulate different pigments, having several commercial applications (Britton, 1995; Vershinin, 1999). Carotenoids are one such class of pigments produced by microbes and plants which play a major role in protecting cells against photo-oxidative damage and hence have vital applications in the environment (Zhang et al., 2014), food and nutrition (Vilchez et al., 2011), disease control (Fiedor and Burda, 2014), and as potent antimicrobial agent (Narsing Rao et al., 2017). Haloarchaea are one of the richest sources of carotenoids compared to other microorganisms (Yatsunami et al., 2014; Montero-Lobato et al., 2018; Giani et al., 2019). The main component of the haloarchaeal carotenoid pool is bacterioruberin which reportedly has more antioxidant properties compared to plant β-carotenes (Yatsunami et al., 2014). Other than bacterioruberin, they also contain isopentenyl dehydrodopin, lycopene, and phytoene in trace amounts (Yatsunami et al., 2014).

In the present study, we isolated and characterized three haloarchaeal strains namely wsp1, wsp3, and wsp4 from the Pondicherry solar lakes, one of the high salt containing areas of the Indian solar salterns. We performed a polyphasic taxonomic classification of these isolated strains using whole-genome sequencing and biochemical assays. Using comparative genomic analysis, we identified the core, accessory, and unique gene pool of haloarchaeal proteins. We identified several industrially important enzymes encoded in the genomes of these strains. We report the expression, and purification of a halotolerant recombinant starch degrading enzyme i.e., α-amylase Amy2 isolated from *Haloferax* strain wsp1 having distinct sequence and structural features. All the strains reported here were colored and genome analysis suggested the presence of carotene biosynthesis genes in their genomes. We further report the isolation and characterization of carotenoids produced by these strains.

RESULTS

Isolation and Taxonomic Characterization of Haloarchaeal Isolates

All three strains were isolated from solar saltern samples collected from Marakkanam solar salterns, India (12°11′13.0272″N and 79°55′40.4220″E) using the dilution-plate technique on the *Halovibrio* agar medium as described previously (Verma et al., 2019). Briefly, on agar plates, the reddish-pink opaque convex colonies, 1–2 mm in diameter appeared in 7 days. The growth of wsp isolates were screened from 12 to 42°C, which suggested that they have optimum growth at 37°C. The positive growth was observed on media containing 3–5 M NaCl (**Figure 1A**), whereas no growth was observed on 0.3–2 M NaCl concentrations. Transmission electron microscopic images suggested that wsp1, wsp3, and wsp4 have pleomorphic morphology, and compared to wsp1, both wsp3, and wsp4 isolates are highly vacuolated (**Figure 1B** and **Supplementary Figure S1**).

Comparative Genome Analyses of wsp1, wsp3, and wsp4

Comparison of both 16S rRNA and whole-genome sequences suggested that all three isolates are novel strains that belong to *Haloferax* and *Halogeometricum* genera. These results were further confirmed by calculating ANI (Average nucleotide index) and digital DNA-DNA hybridization scores where the observed values were higher than the accepted cut-off values (for ANI > 95% and for DNA–DNA hybridization >91%) for novel species (**Supplementary Tables S1, S2**). The phylogenetic analysis based on 16S rRNA sequencing data suggested that wsp1 is closely related to *Haloferax volcanii*

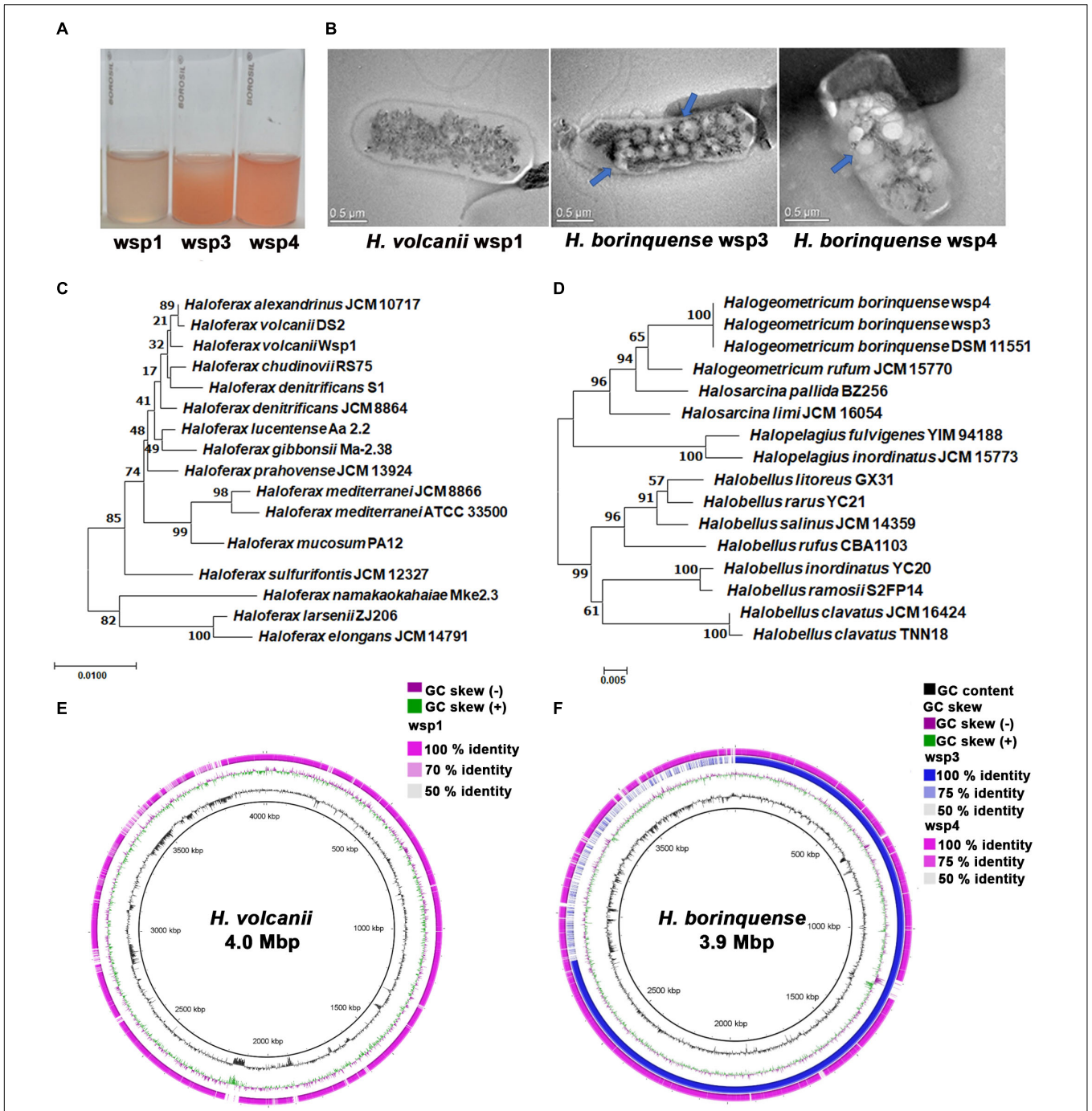

FIGURE 1 | Morphological and taxonomic characterization of the haloarchaeal isolates. **(A)** Wild type culture of haloarchaeal isolates in light reddish-pink color suggesting high carotenoid production. **(B)** TEM images suggest that wsp3 and wsp4 are highly vacuolated (shown in blue arrow) compared with wsp1 **(C,D)** The16S rRNA based phylogenetic analysis suggested that wsp1 strain is closely related to *H. volcanii* DS2 while both wsp3 and wsp4 are closely related to *H. borinquense* DSM 11551. **(E,F)** Genomic maps of the isolates and their reference species constructed using the BRIG tool. The inner black rings show the coordinates in scale and total genome size of the reference sequence, with black histogram representing GC content and purple/green histograms representing GC deviations. In the panel, **(E)** outermost ring in magenta color represents the genome map of wsp1 strain and in the panel, **(F)** blue color represents wsp3 genome map while wsp4 genome map is represented in magenta color.

DS2 and *Haloferax chudinovii* RS75 strains sharing 99% sequence similarity. Both wsp3 and wsp4 are closely related to *Halogeometricum borinquense* DSM 11551 strain sharing >99% sequence identity (**Figures 1C,D**). The genus *Haloferax* was first described by Torreblanca et al. (1986) and currently comprises twelve species, namely *H. sulfurifontis, H. mucosum,*

H. mediterranei, H. denitrificans, H. prahovense, H. larsenii, H. gibbonsii, H. elongans, H. alexandrinus, H. lucentense, H. volcanii, and *H. massiliensis.* On the other hand, genus *Halogeometricum* was first described by Montalvo-Rodriguez et al. (1998) and comprises four species, namely *H. pallidum, H. borinquense, H. limi*, and *H. rufum. H. volcanii* DS2, and *H. borenquencis* DSM 11551 are the type strains of *Haloferax* and *Halogeometricum* genus, respectively. The draft genome sequences of wsp1, wsp3, and wsp4 comprized of 37, 07,582 bp (3.7 Mb), 28, 18,554 bp (2.8 Mb) and 39, 97,080 bp (3.9 Mb) with 4047, 3040, and 4257 annotated coding sequences (CDS). The GC content of wsp1, wsp3, and wsp4 were 66.1, 61.0, and 59.7%, respectively. For further analysis, we compared wsp1 genome sequences with *H. volcanii* DS2 and both wsp3 and wsp4 genomes with the *H. borinquense* DSM 11551. Genome circular map built by using BLAST Ring Image Generator (BRIG) tool (Alikhan et al., 2011), revealed that wsp3 has high sequence similarity with both wsp4 and *H. borinquense* DSM 11551 genomes except one highly variable region of about 1 Mbp (**Figure 1F**).

Pangenome Analysis

During evolution, microbes acquired several genes that facilitate their growth and survival. The important genes in this list include genes responsible for cell signaling, metabolic regulators, antibacterial proteins, ion transporters, etc. To understand the environmental effect on gene pool variation, we performed a pangenome analysis of wsp samples with their closely related species using the Bacterial Pan Genome Analysis (BPGA) pipeline (Chaudhari et al., 2016). The original pangenome concept was developed by Tettelin et al. (2005) and it describes the total pool of genetic material comprized of all members of a species. The pangenome consists of three different groups known as the core, accessory, and unique genomes. The core genome consists of common genes that are present in all individuals, accessory or dispensable genome containing shell genes present in few individuals and a unique genome contains genes that are specifically present only in an individual member (Tettelin et al., 2005; Wolf et al., 2012; Vernikos et al., 2015). The outcomes of the pangenome analysis are discussed in the following sections.

Clusters of Orthologous Groups (COG) Distribution Plots

The whole-genome sequences of *Haloferax* and *Halogeometricum* members were retrieved from NCBI (National Center for Biotechnology Information) database. The functional annotation was carried out using Rapid Annotation using Subsystem Technology webserver (Aziz et al., 2008). The distribution of the archaeal clusters of orthologous groups (arCOG) of *Haloferax* and *Halogeometricum* genera along with our isolates with their biological functions are shown in **Figures 2A,B**. The potential functions of the unique genes from both genera appear to be widespread and linked with many different cellular functions such as cell motility, post-translational modifications, chaperones, signal transduction mechanisms, and many genes with unknown function. On the other hand, the accessory genes identified are mainly associated with carbohydrate transport, inorganic ion transport, and several genes with unknown function. The core genome consists of the genes involved in translation regulation, ribosomal structure and biogenesis, replication, recombination and repair proteins, energy production and conversion, coenzyme transport and metabolism, lipid transport and metabolism and nucleotide transport and metabolism.

Pan and Core Genome Plots

The comparative genome analysis revealed an open pangenome for both genera investigated here (**Figures 2C,D**) as the number of accessory and unique genes are increasing with the addition of new genomes. For the genus *Haloferax,* the pangenome and core genome contain 7950 and 1885 genes, respectively (**Figure 2C**). For the genus *Halogeometricum*, the pangenome and core genome contain 7197 and 1874 genes, respectively (**Figure 2D**). Similar findings were reported for the pangenome based analysis in other studies as well (Capes et al., 2012; Becker et al., 2014; Borchert et al., 2017). The pangenome analysis revealed that wsp1, wsp3, and wsp4 contain 246, 530, and 106 unique genes in their genomes. These unique genes were further annotated using online server Web Services for Metagenomic Analysis (WEBMGA) (Wu et al., 2011). We found multiple unique clusters including genes potentially encoding cation efflux proteins, integrases, and proteins potentially involved in multidrug resistance. The annotated genes also included commercially useful salt-tolerant enzymes like amylases, alcohol dehydrogenases, sulphataes, esterase's, etc.

The size of pangenome is larger than the core genome and is increasing with the addition of the new genomes suggesting both the genera investigated here have open pangenome. The current study is important for understanding the genomic variations and distribution of accessory genes providing survival advantage to the haloarchaeal strains in adverse conditions.

Biochemical Characterization of Haloarchaeal Isolates

Polyphasic taxonomic characterization were studied according to the methods in the proposed minimal standards for description of new taxa in the order *Halobacteriales* (**Table 1A**; Grant and Larsen, 1989; Oren et al., 1997). Biochemical characterization suggested that all three isolates were positive for the catalase test and negative for urease, lipase, and Voges-Proskauer tests. All three strains were also negative for acid production, casein hydrolysis, and hydrogen sulfide production. We also performed enzymatic screening which involved in amino-acid biosynthesis pathways and found that all three were negative for lysine decarboxylase, arginine decarboxylase, aesulin hydrolysis, and ornithine decarboxylase.

Wsp3 and wsp4 strains had many similar enzymatic features such as both were positive for citrate utilization, oxidase test, and methyl red test. Besides these similarities, some variations were also observed e.g., both *H. borinquense* DSM 11551 and wsp3 are positive for gelatin liquefaction and indole production while wsp4 is negative. We also found that wsp3 is unable to reduce nitrate, unlike *H. borinquense* DSM 11551 and wsp4. Genome analysis suggests that both *H. borinquense* DSM 11551

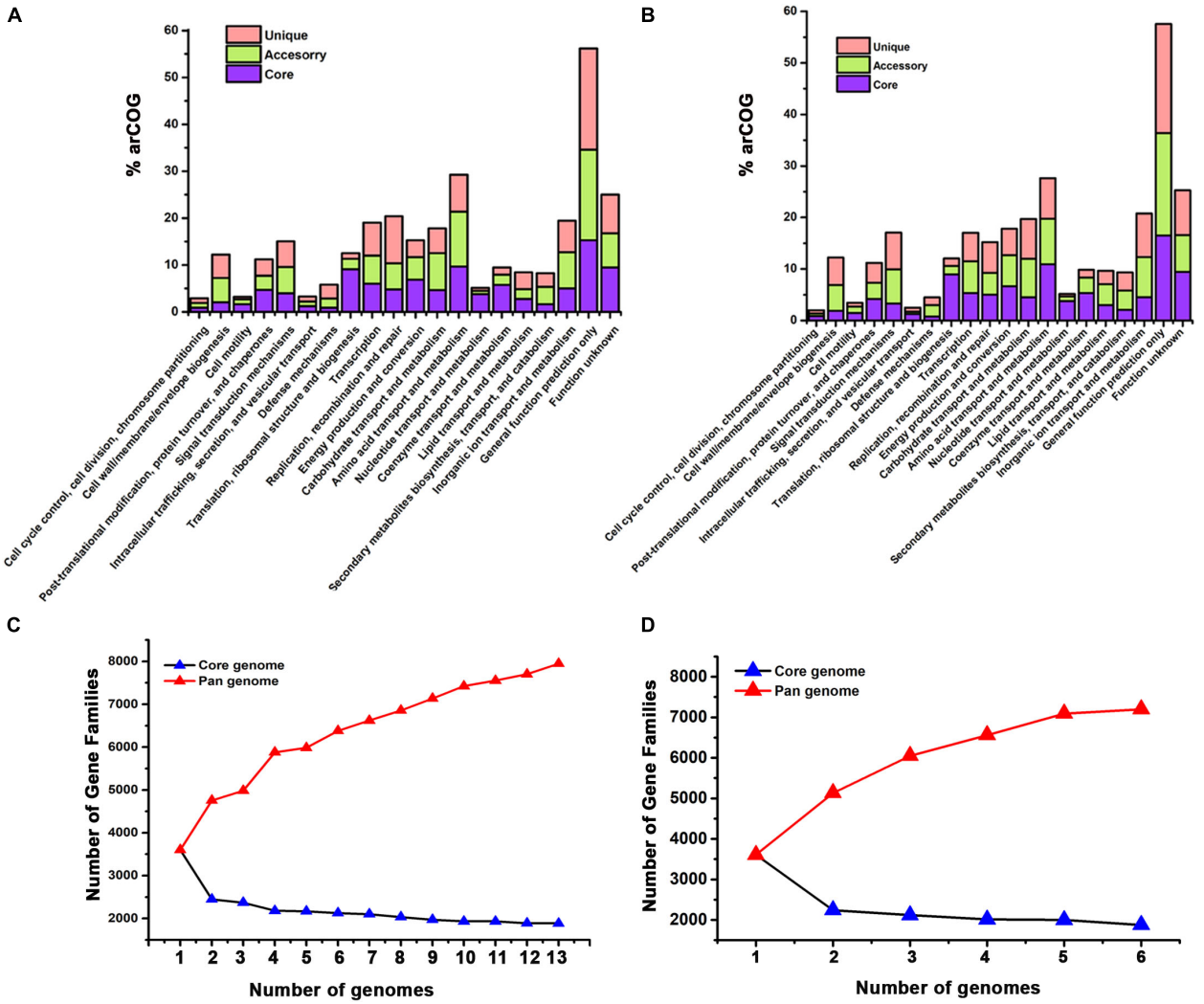

FIGURE 2 | Archaeal clusters of orthologous groups (arCOG) analysis of the genomes. Comparison of the COG distribution between the core, accessory and unique genes of the **(A)** thirteen species belonging to genus *Haloferax* including wsp1 and **(B)** four species belonging to genus *Halogeometricum* including wsp3 and wsp4 have been analyzed using default parameters of BPGA pipeline. The arCOG categories are indicated on the *X*-axis and the percentage of the genes present in each category of the arCOG classes is indicated on the *Y*-axis. **(C,D)** The Core and pangenome plots of genera *Haloferax* and *Halogeometricum*, respectively. Red and blue colored lines indicate the change in the number of pangenome and core genome gene groups added sequentially to the genome.

and wsp4 encode nitrate reductase gene which is missing in the wsp3 genome. Similarly, we also found a copy of the nitrate reductase gene in wsp1 genome, which reduces nitrate, whereas it is missing in *H. volcanii* DS2 which is negative for nitrate utilization.

Carbon Source Utilization

Carbon source preferences of haloarchaeal isolates were identified by monitoring their growth in the presence of different carbon sources. Optical density-based assay results suggest that all three strains have similar preferences for dextrose while they were unable to utilize arabinose, rhamnose, aldonitol, inositol, inulin, and melibiose (**Table 1B**). Along with enzyme-based biochemical tests, wsp3 and wsp4 had some similar and dissimilar preferences for carbon source utilization. Both wsp3 and wsp4 were positive for pyruvate and maltose utilization while negative for lactose and galactose utilization. Surprisingly, wsp3 alone was positive for three different carbon sources i.e., sucrose, cellobiose, and dulicitol while both *H. borinquense* DSM 11551 and wsp4 were negative. Similarly, wsp1 is not able to utilize sucrose and arabinose while its type strain *H. volcanii* DS2 is capable of utilizing both sugars.

Quantitative analysis of carbon source utilization shows that both wsp1 and wsp3 efficiently utilize galactose and maltose, respectively. Genome analysis suggested that both wsp1 and *H. volcanii* DS2, encode galactokinase enzyme which help them to utilize galactose as a carbon source (Anderson et al., 2011) while this enzyme is absent in both wsp3 and wsp4 strains. For maltose utilization, two alternative pathways have been proposed (Cibrario et al., 2016). The two essential enzymes

TABLE 1 | (A) Biochemical test and **(B)** carbon source utilization test as an aid to the classification of haloarchaea.

	Wsp3	Wsp4	Wsp1
(A) Biochemical test			
Catalase test	+	+	+
Oxidase Test	+	+	–
Citrate utilization	+	+	–
Production of acid	–	–	–
Production of gas	–	–	–
Lysine decarboxylase	–	–	–
Methyl red test	+	+	–
Voges-Proskauer test	–	–	–
Nitrate reduction	–	+	+
Casein hydrolysis	–	–	–
Arginine decarboxylase	–	–	–
Urease activity	–	–	–
Ornithine decarboxylase	–	–	–
Aesulin hydrolysis	+	+	+
Gelatin Hydrolysis	+	–	–
Indole production	+	–	–
Lipase production	–	–	–
(B) Carbon source			
Dextrose	W	W	W
Pyruvate	W	W	–
Lactose	–	–	+
Maltose	+++	W	–
Sucrose	W	–	–
Galactose	–	–	+++
Arabinose	–	–	–
Rhamanose	–	–	–
Cellobiose	W	–	–
Aldonitol	–	–	–
Inositol	–	–	–
Inulin	–	–	–
Dulcitol	W	–	–
Mannose	–	–	–
Melibiose	–	–	–

+++, highly positive; +, positive; –, negative, W, weakly positive.

of these pathways are *malz* (α-1, 4-glucosidase) and *malA* (α-amylase MalA). Genomic data analysis suggested that only wsp3 genome encodes for two different copies of α-1, 4-glucosidase enzyme while wsp1 and wsp4 only have an α-amylase gene in their genomes. This probably explains why wsp3 can utilize maltose more efficiently compared to both wsp1 and wsp4.

Antibiotic Susceptibility Profile

Culture isolates were also screened for antibiotic resistance (**Table 2**). The analysis suggested that all three isolates were sensitive for novobiocin and resistant to penicillin, vancomycin, chloramphenicol, cephadroxil, lincomycin, cephalexin, ceftazidine, and cephadroxil. Surprisingly wsp3 had sensitivity for cefazolin and kanamycin while wsp1 and wsp4 were resistant to both the antibiotics like their type strains. The possible reason for kanamycin resistance in wsp1 and wsp4 is the presence of aminoglycoside phosphotransferase while this enzyme is absent in wsp3 hence probably explaining the differences observed in kanamycin resistance in these isolates.

TABLE 2 | Antibiotic susceptibility test of the isolates.

Antibiotic sensitivity	wsp3	wsp4	wsp1
Novobiocin	S	S	S
Cefazolin	S	R	R
Kanamycin	S	R	R
Bacitracin	I	R	R
Penicillin	R	R	R
Vancomycin	R	R	R
Chloramphenicol	R	R	R
Cephadroxil	R	R	R
Lincomycin	R	R	R
Ciphalexin	R	R	R
Ceftazidine	R	R	R
Cephadroxil	R	R	R

S, sensitive; R, resistance; I, intermediate.

Carotenoid Isolation and Spectroscopic Characterization

In the present manuscript, we isolated carotenoids from three haloarchaeal strains (**Figures 3A,B**). The visible cell pellet color of wsp1 was different when compared to both wsp3, and wsp4 suggesting either variations in the content or molecular structure of the carotenoids produced by these strains (**Figure 3A**). For further analysis, we purified carotenoids following Yatsunami et al. (2014) protocol and characterized them using UV–visible spectroscopy. UV–visible spectra suggested that wsp1, wsp3, and wsp4 have a similar absorption spectrum having major absorption peaks at 476, 502, and 535 nm, however, there were differences in the peak heights (**Figure 3B**). These differences may explain why carotenoids isolated from wsp1 are different in color compared to wsp3 and wsp4. These three major peaks correspond to bacterioruberin (535 nm), all-Trans-lycopene (502 nm), and 13-cis-lycopene (476 nm). Bacterioruberin is a major component of haloarchaeal carotenoids which is synthesized from lycopene (Yang et al., 2015). We took the ratio of peak1/peak3 which may help us understand the bacterioruberin and lycopene contents in these strains. The lycopene has specific absorption at 502 and 476 nm while bacterioruberin absorbs at 535 nm. A slight absorption of bacterioruberin is also reported at 466 nm. The peak1/peak3 ratio was 1.0, 0.9, and 0.8 for wsp1, wsp3, and wsp4, respectively, suggesting that wsp1 has the highest bacterioruberin content among the three strains.

Recombinant Expression and Purification of Halotolerant α-Amylase From the Isolates

Haloarchaea lives in hypersaline environments such as marine salterns, saline soils, soda lakes, salted foods, etc. The cell lysis of marine planktons such as shrimps, diatoms, algae, fish, purple

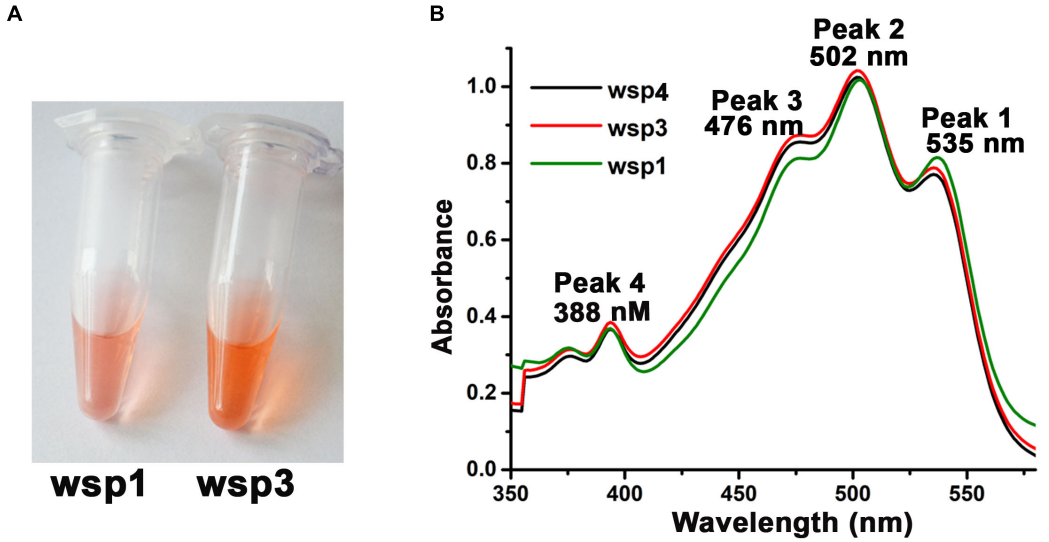

FIGURE 3 | Isolation and biophysical characterization of haloarchaeal carotenoids isolated from wsp1, wsp3, and wsp4 isolates. **(A)** The microcentrifuge tubes containing purified carotenoids from equal culture volume of the two haloarchaeal strains showing observable color variations. **(B)** UV–visible absorption spectra of the purified carotenoids showing four carotenoid standard peaks (peak I–IV) at 535, 502, 476, and 388 nm.

and green bacteria, etc. release different biomolecules including starch, cellulose, proteins, chitin, etc. To digest this biopolymer, haloarchaea needs to produce different digestive enzymes to utilize these substrates.

The whole-genome sequencing and annotation results suggested that wsp1, wsp3, wsp4 genomes code for two, one, and three putative α-amylase genes, respectively. We successfully cloned all four α-amylase genes sharing <40% sequence identity among themselves. However, only one of them, named *amy2* (locus tag G3A49-11660) isolated from wsp1, expressed well. Multiple domain analysis of Amy2 (protein ID QIB80089.1) suggested that it has a conserved α-amylase domain in the C-terminal region (residue range 227 to 635). A PSI-BLAST search using the NCBI database was performed for the additional N-terminal region (residue ranges from 1 to 226 aa). However, we did not find any significant hits, therefore, the function of this N-terminal domain is still not clear. Initially, we attempted to purify recombinant His-tagged Amy2 from the cytosolic fraction using *E. coli* Rosetta DE3 as an expression host. We successfully purified Amy2, however, the purified enzyme was inactive as suggested by the starch agar plate assay. Since most of the amylases are secretory proteins so we cloned *amy2* in pET22b vector having pelB secretion signal sequence at the N-terminus to aid secretion /expression in the periplasmic space and C-terminal 6X His-tag to aid purification of the recombinant Amy2. We used this construct for purification and further biochemical characterization of Amy2.

Biochemical Characterization of Amy2

Upon induction, we observed amylase activity even in the cell-free culture supernatant (**Figure 4A**). *E. coli* cells alone or cells carrying the pET-22b vector without amylase genes were used as negative controls in these experiments (**Figure 4A**). We successfully purified Amy2 from the periplasmic space using Ni-NTA-based affinity chromatography (**Figure 4B**). The gel-filtration profile suggested that Amy2 is predominantly monomeric in solution (**Figure 4C**) and enzymatically active on the starch agar plates. We initially tested Amy2 salt tolerance by incubating purified enzyme for 12 h with different concentrations of salt ranging from 0.25 to 4 M on the starch agar plate (**Figure 4D**). We observed amylase activity in all the samples suggesting halotolerant property of Amy2. We further performed the biochemical characterization of Amy2 to study its stability and activity at various salt concentrations, pH, and temperature.

Effect of Salt Concentration on Enzyme Activity

The effect of NaCl concentration ranging from 0–4 M on enzyme activity was examined. Data suggests that Amy2 is active in all the NaCl concentrations studied having maximum activity at 1 M NaCl concentration. These results suggest that Amy2 has wide-range salt tolerance and retains moderate activity even at 3 M NaCl concentrations (**Figure 4E**).

Effect of pH on Enzyme Activity

The effect of pH on activity was observed by performing enzyme assay under pH ranging from 3.0 to 11.0 at 1unit interval for 30 min at 37°C. The maximum enzymatic activity was observed at pH 6.0, while ∼30% activity was observed at pH 5.0 and pH 9.0 suggesting enzyme is active at wide pH range (**Figure 4F**).

Effect of Temperature on Enzyme Activity

The thermal effect on enzymatic activity was studied at the temperature range of 5–75°C. Our finding revealed that

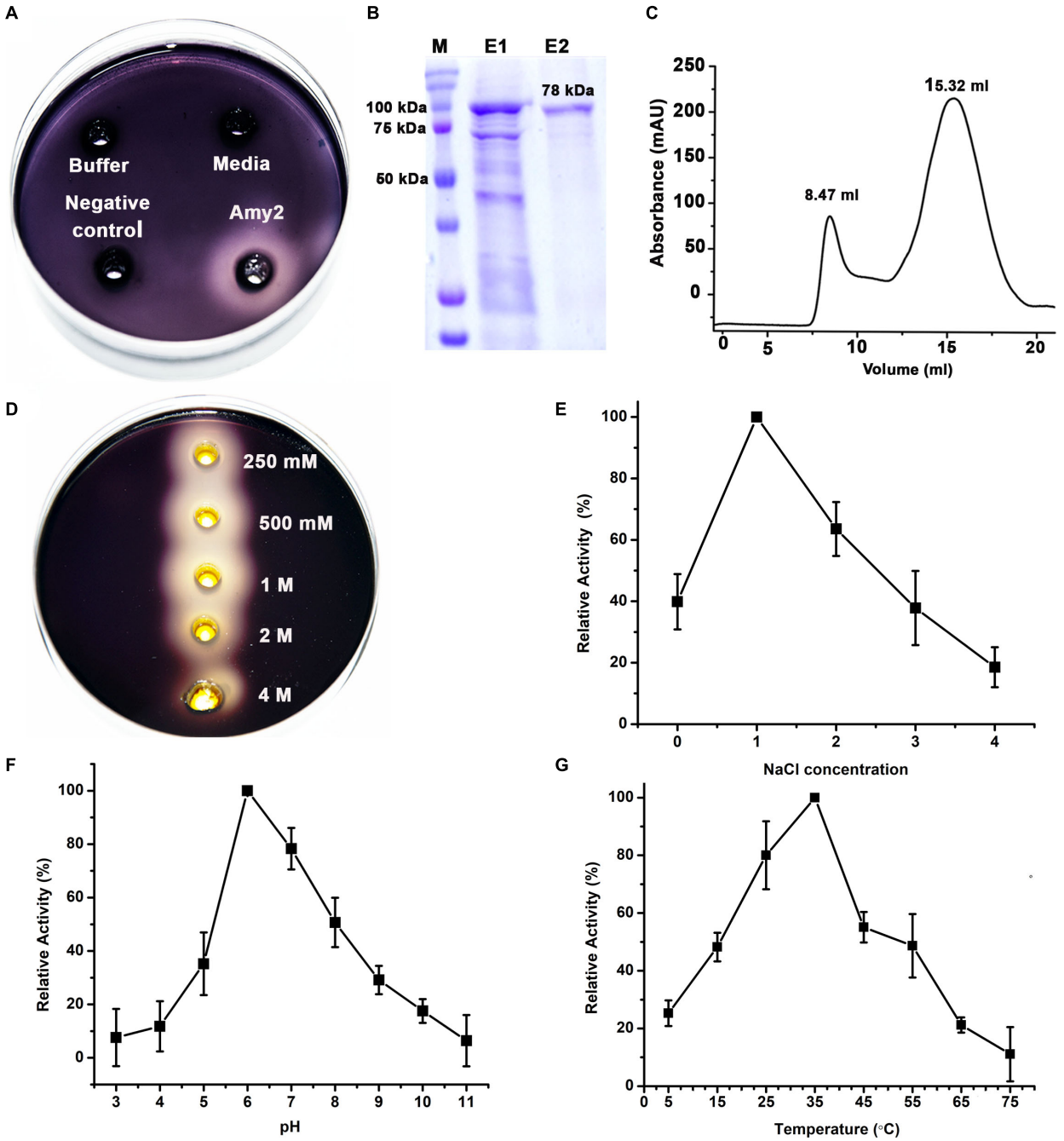

FIGURE 4 | The purification of recombinant α-amylase from *E. coli*, and carotenoids from wild type cultures. **(A)** The α-amylase activity observed from the cell-free supernatant of *E. coli* expressing recombinant Amy2. **(B)** The 15 % SDS-PAGE showing a predominant protein band at the expected molecular weight of ∼78 kDa in the gradient of imidazole elution fractions (E1, E2, E3 correspond to 20, 100, and 200 mM imidazole, respectively). **(C)** The analytical gel filtration profile suggests that Amy2 is predominantly monomeric in solution. A small fraction, corresponding to large aggregates, was also observed in the void volume (8.47 ml). **(D)** Evaluating salt tolerance of Amy2 using starch plate assay suggests that the enzyme is halotolerant. **(E)** The enzymatic activity of Amy2 against soluble starch was assayed by the iodine-starch method at the various concentrations of NaCl. Relative activity was defined as the percentage of maximum activity at 1 M NaCl. Plots showing effect of pH **(F)** and temperature **(G)** on the Activity of Amy2 and the relative activity was defined as the percentage of maximum activity observed at pH 6.0 or 35°C, respectively. All activity assays were performed three times in triplicate and mean value with standard deviation are plotted.

Amy2 showed maximum activity at 35°C, and retained 40% activity at 55°C whereas negligible activity was observed at 75°C (**Figure 4G**).

The biochemical characterization reveals attractive features of Amy2 including halotolerant, moderately thermostable, and activity at a wide pH range suitable for various industrial

applications. These biochemical characteristics reported here are comparable to α-amylases reported from other halophiles (Onodera et al., 2013; Bajpai et al., 2015).

Structure Prediction and Analysis of Amy2

Most of the archaeal α-amylases belong to the subfamily of glycosyl hydrolase GH families: GH13, GH70, and GH77 (Dong et al., 1997; Bowers et al., 2009). In recent years, several potential α-amylases from different halophilic archaea have also been added to these families (Lombard et al., 2014; Santorelli et al., 2016). The enzyme belonging to this families should have three distinct domains: a central catalytic domain harboring a $(\beta/\alpha)_8$ Tim barrel (domain A), with an irregular loop domain (domain B) usually protruding as a long loop out of the barrel connecting the third β-strand and the third α-helix and with the typical structure of an antiparallel β-sandwich (domain C) (Sarian et al., 2017).

Multiple sequence alignment using the NCBI blast portal revealed that Amy2 has a catalytic α-amylase domain encompassing 267–600 residue range (**Figure 5A**). This catalytic domain is mainly present in archaeal and bacterial species and is known to hydrolyze α-(1, 4) glycosidic linkages of glycogen, starch, related polysaccharides, and some oligosaccharides (Antranikian, 1992; Gupta et al., 2003; Sivaramakrishnan et al., 2006). The structural model of the catalytic domain of Amy2 (R269-D625) was built using PHYRE2 (Kelley et al., 2015) webserver suggested that it may have a similar structural architecture found in other members of GH family of amylases i.e., eight stranded alpha/beta-barrel that contains the active site, calcium-binding domain present between beta-strand 3rd and alpha-helix 3rd, and a carboxyl-terminal Greek key beta-barrel domain (Abe et al., 2005). Amy2 sequence blast at the RCSB database (Bernstein et al., 1977) suggested that *Halothermothrix orenii* α-amylase is the closest structural homolog sharing 28% sequence identity over 426 residues (ranges from 3 to 429 aa) (**Figure 5B**). The majority of the α-amylase enzymes are calcium-dependent metalloenzymes, where metal ion is required for both stability and enzymatic activity (Hsiu et al., 1964; Boel et al., 1990). *H. orenii* α-amylase (PDB ID-1WZA) harbors two different calcium-binding loops, loop 1: Asp44 aa to Ile52 aa and loop 2: Asp65 to Asp73 residues (Sivakumar et al., 2006). The sequence alignment and predicted structural model suggests that the calcium-binding loop 2 is missing in Amy2 (**Figures 5B,C**). Amy2 also has smaller B and C domains compared to *H. orenii* α-amylase due to deletions in the loop regions (**Figures 5B,C**).

For amylase activity, a catalytic triad consisting of Asp-Glu-Asp is required where first aspartic acid is involved in attacking the sugar anomeric center by a nucleophilic side chain and this reaction is assisted and governed by other two Glu-Asp residues (Hsiu et al., 1964; Machius et al., 1995). In *H. orenii* this triad is composed of Asp224-Glu260-Asp330. Multiple sequence alignment-based analysis suggests that Amy2 also has similar tried consisting of Asp436-Glu463-Asp529, respectively (**Figure 5D**).

Screening of CRISPR Sequences and Secondary Metabolites

The clustered regularly interspaced short palindromic repeats (CRISPR)-Cas system is used by bacteria and archaea to fend off foreign genetic elements (Horvath and Barrangou, 2010). Since its discovery, it has been developed into numerous applications like genome editing and regulation of transcription in eukaryotes and bacteria (Gilbert et al., 2013). The genomes of the strains wsp1, wsp3, and wsp4 were submitted to CRISPR Finder[1] (Grissa et al., 2007) which suggested that all three genomes have multiple CRISPR-DR (CRISPR-direct repeat) as well as spacer sequences. In wsp1, wsp3 and wsp4 genomes six, one, and four CRISPR-DR sequences were observed, respectively (**Supplementary Table S3**). Whole-genome data of wsp1, wsp3, and wsp4 were further analyzed for the presence of CRISPR associated genes (Cas). Our analysis suggested the presence of several Cas sequences in both wsp1 and wsp4 genomes although no Cas-related genes were identified in the wsp3 genome. In haloarchaea CRISPR-Cas system type I, subtype I-B, is dominantly present (Maier et al., 2017). A characteristic feature of type I systems is the presence of nuclease Cas3 protein, and subtype I-B is further characterized by the presence of Cas8b (Makarova and Koonin, 2015). The functional process of Type I-B system is mainly based on the presence of Cas proteins, namely, Cas5, Cas7, and Cas8b which combine to form a Cascade (CRISPR-associated complex for antiviral defense)-like complex in different *Haloferax* species such as *H. volcanii* (Maier et al., 2019). Genome analysis of both wsp1 and wsp4 suggested that they encode all the genes required to type I CRISPR system i.e., Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, and Cas8b genes. The presence of both Cas3 and Cas8b suggests that pws1 and pws4 harbor CRISPR-cas type IB system similar to other haloarchaeal species.

In addition to enzymes, recent studies indicate that archaea can also produce a wide range of small peptides and secondary metabolites (Pietra, 1997; Pettit, 2011; Wang et al., 2019). These small peptides and secondary metabolites could be of considerable interest in biotechnology. For the screening of secondary metabolite production, all three genome sequences were uploaded onto an online antiSMASH server (Medema et al., 2011). The output results predicted multiple copies of two important secondary metabolites i.e., siderophore and terpenes in all three strains (**Supplementary Table S4**). In the case of wsp3, no siderophore gene was predicted. Siderophores play an important role in iron transport (Neilands, 1995) while the functions of terpenes in haloarchaea are not known yet.

DISCUSSION

Haloarchaeal genera show huge variations in the diversity within different geographical locations (Oren, 2002; Naghoni et al., 2017). There are several reports which suggest that different haloarchaeal strains were routinely isolated from the different salt areas like solar salterns, brines, salt lakes, salt pans, rock

[1] https://crispr.i2bc.paris-saclay.fr/Server/

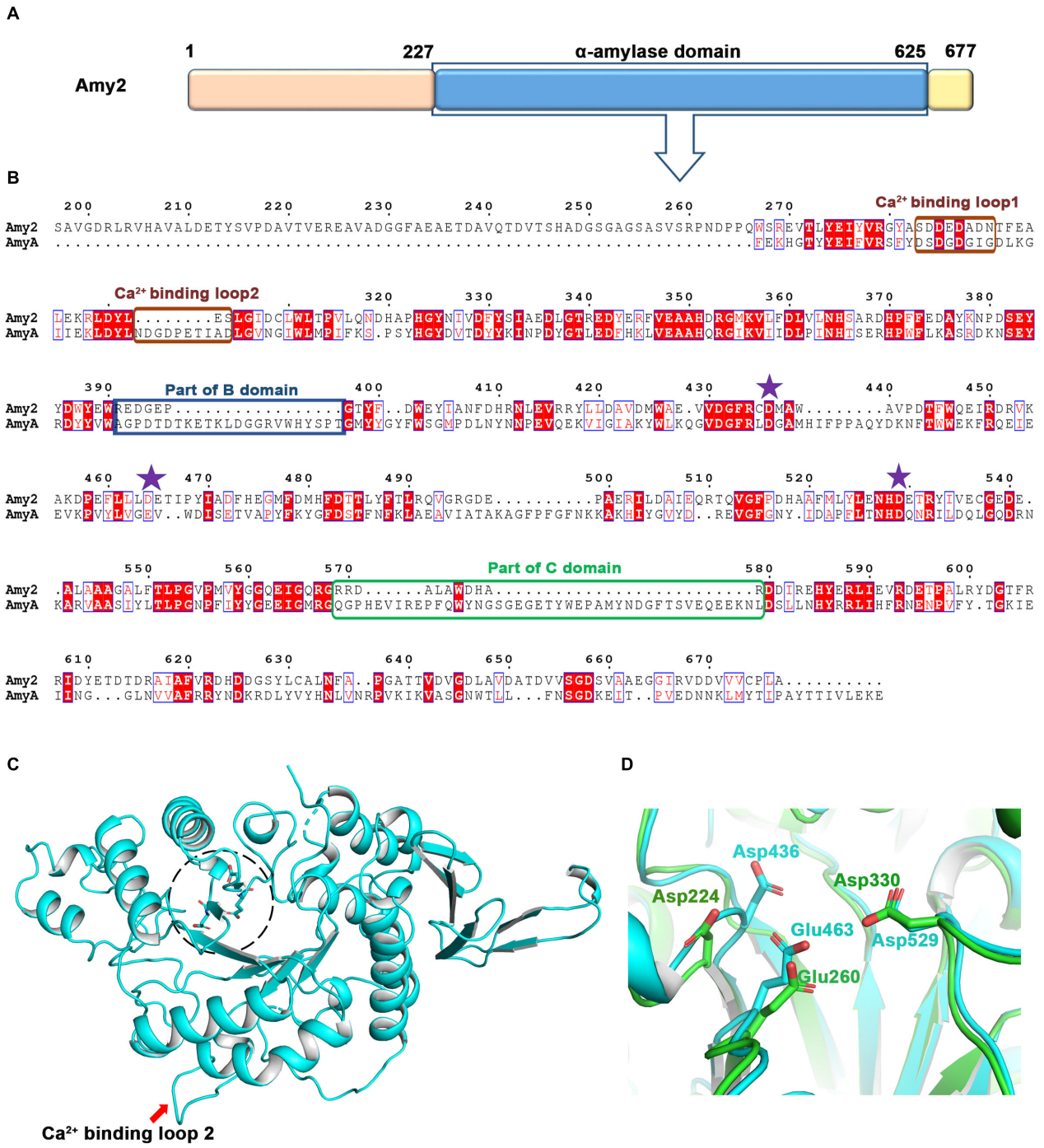

FIGURE 5 | Structural and sequence features in Amy2. **(A)** Schematic representation of predicted domain architecture in Amy2. The N-terminal (1–226) region, present in α-amylases from other haloarchaeal strains as well, did not show significant sequence similarity with proteins of known function. **(B)** Multiple sequence alignment of the Amy2 region sharing sequence conservation with *H. orenii*α α-amylase. The active site residues are marked in purple stars. The differences in the Ca^{2+} binding loops, and domains B and C are highlighted in colored boxes. **(C)** Predicted structural model of Amy2 built using the PHYRE2 online web server. The active site is encircled in black broken circle and active site residues are shown in stick representation. **(D)** The structural alignment suggested that Amy2 (cyan) and *H. orenii* α-amylase (green) have conserved active site triad consisting of structurally equivalent Asp436, Glu463, and Asp529 residues.

salts, etc. (Oren, 2002; Javor, 2012). In the present study, three haloarchaeal strains, namely wsp1, wsp3, and wsp4, were isolated from the high salinity environment and studied in detail. Haloarchaeal isolates adopt diverse morphological shapes like rods, pleomorphic rods, square, cocci, triangles, and disks (Burns et al., 2007; Minegishi et al., 2010). Few haloarchaeal strains show

rod form morphology in liquid culture while they are motile and formed nonmotile cocci on solid agar medium (Grant, 2001). All of the three isolates in this study i.e., wsp1, wsp3, and wsp4 were highly pleomorphic. Also, wsp3 and wsp4 were highly vacuolated, which may help these microbes to float on the surface of the water (Oren, 2012).

The pangenome comparison of the four *Halogeometricum* and twelve *Haloferax* haloarchaeal genomes, including three isolates reported in this study, have revealed genomic variations resulting from horizontal gene transfer, gene duplication, gene loss events, etc. which is in line with results from other studies (Vos and Didelot, 2009; DeMaere et al., 2013). This data also suggests that pangenomes of these genera are open and each genome contains about 3 to 20% unique genes. The pangenome analysis further suggested that the distribution of the unique, accessory, and core genes is widespread across different biological functions and no obvious pattern was observed.

Our data and past studies (Nagata et al., 1989; Littlechild, 2015; Karray et al., 2018; Amoozegar et al., 2019) suggest that haloarchaea codes for several enzymes suitable for industrial and research applications. Especially, their salt-tolerant and thermostable nature are attractive features suitable for specific applications like food processing, biofuel production, detergent additives, etc. However, the potential of these enzymes has not yet been commercially exploited. Amylase enzymes are typically involved in essential processes in microorganisms, such as hydrolysis of starch and glycogen, yielding glucose and maltose (Werries and Muller, 1986). There is an increased interest in exploiting amylases for the conversion of grain starch into fermentable sugars for ethanol production (Shigechi et al., 2004; Prakash and Jaiswal, 2010). However, purification of enzymes from the natural haloarchaeal source will not be commercially viable owing to the poor yield and slow growth rate (doubling time ∼4–10 h). Hence, there is a need for developing processes based on the recombinant expression of the target enzymes using fast-growing microbial hosts like *E. coli*. There are limited studies that report the recombinant haloarchaeal α-amylase production (Perez-Pomares et al., 2003; Hutcheon et al., 2005; Bajpai et al., 2015; Santorelli et al., 2016). Onodera et al., produced recombinant *Haloarcula japonica* α-amylase using *E. coli* host and found that malA is a cytosolic halotolerant α-amylase which is active even at 4 M salt concentration (Onodera et al., 2013). Our whole-genome data analysis of the wsp isolates suggested that they all code for one or more types of α-amylase genes in their genomes. We were able to successfully clone, express, and purify one of the annotated α-amylase genes named *amy2* from wsp1. Interestingly, though soluble, Amy2 purified from the cytosolic soluble fraction was not active. α-amylases are usually secretory proteins and there are five cysteine residues in Amy2. So, to direct the protein to the periplasmic space for proper folding and functional expression, we cloned Amy2 having pelB signal sequence at the N-terminus. Using this strategy, we were successful in purifying active Amy2. In future, this strategy may aid purification of other functional α-amylases as well. The purified Amy2 was studied for its activity in the presence of varying salt concentration, pH, and temperature conditions. Data suggests that Amy2 was active in a wide range of salt concentrations, temperature, and pH conditions. However, detailed studies are required to further investigate its suitability for various industrial applications.

All three isolates were characterized based on a battery of biochemical tests and their antibiotic susceptibility was also evaluated. The results for these assays for wsp1, wsp3, and wsp4 were similar to their corresponding type strains. However, some differences were also observed which we were able to explain based on the whole genome sequencing data analysis. These discrepancies were easily correlated with the presence or absence of the corresponding enzyme(s) in the genome as described in the results section. Haloarchaea is known for producing high levels of carotenoids like bacterioruberin and its C50-related pigments having high commercial potential in various industries including cosmetic, food, poultry and health (Vilchez et al., 2011; Mata-Gomez et al., 2014; Naziri et al., 2014; Rodrigo-Banos et al., 2015). Interestingly, haloarchaeal carotenoid has also been shown to have dose-dependent cytotoxicity against human liver cancer cell lines suggesting anti-cancer activity (Rodrigo-Banos et al., 2015). The isolated carotenoids from all of the three strains in this studies showed a characteristic absorption profile similar to other haloarchaeal strains (Yatsunami et al., 2014; Yang et al., 2015). The differences in the UV–Vis spectrum observed among the carotenoids purified from these strains suggest differences in the contents of bacterioruberin and other carotenoids.

To summarize, we have characterized three haloarchaeal strains isolated from Indian solar salterns. We identified several commercially important enzymes encoded in these three haloarchaeal strains and successfully purified halotolerant recombinant α-amylase using *E. coli* as an expression host. Haloarchaea is known to survive harsh environmental and nutrition depleted conditions. The detailed biochemical and biophysical characterization, in the future, may also help explore the feasibility of these enzymes for commercial applications. The future studies aimed at deciphering the function of hypothetical unique and accessory genes may also help in identifying the mechanisms of stress adaptation that aid survival of haloarchaea in harsh conditions.

MATERIALS AND METHODS

Isolation and Biochemical Characterization of Haloarchaeal Isolates

For all the biochemical tests, respective basal mediums (composition: 250.0 g NaCl, 20.0 g $MgSO_4.7H_2O$, 3.0 g trisodium citrate $2H_2O$, 2.0 g KCl, 10.0 g peptone, 25.0 g agar in 1000 ml; pH 7.2) were supplemented with a final concentration of 25.0% NaCl. The growth of the isolates was tested at 12, 15, 20, 25, 30, 37, and 42°C on Halovibrio agar (1% agar added to basal medium). Isolates were plated on Halovibrio medium (without NaCl) (HiMedia) supplemented with different concentrations of NaCl (2, 5, 7, 10, 12, 15, 18, 20, 22, 25, 27 and 30%) to test salt tolerance. Growth at different pH (5.0, 6.5, 7.0, 8.0, 9.0, and 10.0) was examined using *Halovibrio* medium (without NaCl)

supplemented with 25% NaCl as described earlier (Chen et al., 2016). The pH was adjusted using different buffers: MES (pH 5.0–6.0), PIPES (pH 6.5–7.0), HEPES (pH 7.5–8.0), and CHES (pH 9.0–10). Catalase activity, citrate utilization (using Simmons' citrate agar) and urea hydrolysis were determined as described by Cowan and Steel (Cowan and Steel, 1965). The hydrolysis of casein, gelatin, starch and indole, methyl red test, Voges–Proskauer test and oxidase activity were assessed as described by Smibert and Krieg (Smibert et al., 1994). Nitrate reduction was tested as described by Lanyi (Lányi, 1988). Acid production from various sugars was tested on minimal medium by using the method described by Smith et al. (Sasser, 1990). The utilization of different carbohydrates as sole carbon source was tested in minimal medium supplemented with 25.0 % NaCl and 2% $MgSO_47H_2O$ with 1% different carbon sources.

The Whole-Genome Sequencing and 16S rRNA Comparison

Genomic DNA extractions were performed using the Zymogen DNA isolation kit (Cat. No. D6105). The 16S rRNA gene sequences of strain wsp1, wsp3, and wsp4 were PCR-amplified using 21 forward (TCCGGTTGATCCYGCCGG) and 1453 reverse (GGGCYGCACGCGYRCTACA) standard primers. The amplified 16S rRNA gene sequences were further aligned with those of representative related taxa using the EzTaxon server (Yoon et al., 2017). The 16S rRNA gene sequences of wsp1, wsp3, and wsp4 representatives closely related species were retrieved from the EzTaxon server (Chun et al., 2007) and aligned using MEGA version X (Tamura et al., 2004; Kumar et al., 2018). Phylogenetic trees were constructed using the neighbor-joining methods (Saitou and Nei, 1987) and the percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1500 replicates) are shown next to the branches (Felsenstein, 1985). The whole-genome sequencing of all three strains were performed using Illumina NextSeq and for assembly CLC NGS Cell ver 9 (CGWB) were used. The draft genomes were annotated using the RAST online server (Aziz et al., 2008). The *H. borinquense* DSM 11551 and *H. volcanii* DS2 genomes were used as a reference for comparison and annotation. Pangenome analysis was carried out by using the BPGA pipeline (Chaudhari et al., 2016). The genomes were screened for secondary metabolite gene clusters using antiSMASH online server (Medema et al., 2011).

Genome-Sequence Submission

The genome sequence of wsp1, wsp3, and wsp4 are submitted at NCBI under accession numbers as NZ_CP048738, NZ_CP050274 and NZ_CP048739.

Transmission Electron Microscopy

Aliquots of 50 μl from samples containing wild type culture (basal medium supplemented with 25.0% NaCl) were pipetted on a carbon-coated 300-mesh copper grid (Polysciences, United States). The excess sample was blotted and followed by air drying. The grids were further imaged using a JEM 2100 electron microscope, operated at 200 keV (JEOL).

Cloning, Expression and Purification of Amy2

The full length of *amy2* was amplified and cloned in pET22b (Novagen) over-expression vector between NdeI-XhoI and NcoI-XhoI sites to yield recombinant protein having C-terminal 6X His-tag. The first set of restrictions were used for cytosolic expression (NdeI-XhoI) while the other set was used to express Amy2 fused with pelB (NcoI-XhoI) secretion signal sequence at the N-terminal. Forward primers 5′AGCTCATATGGCTGTCGGCGAGTCAGTA and 5′AGTAC CATGGCTGTCGGCGAGTCAGTA having NdeI and NcoI sites, respectively, and reverse primer 5′AGTCCTCGAGGGCGAGC GGGCAGACGAC having XhoI site were used for gene amplification. The amplified PCR products and vector were digested with NdeI-XhoI and NcoI-XhoI sites followed by an overnight ligation at 4°C. The ligated products were transformed into Top10 chemical competent cells (Thermo Scientific, United States). Positive clones were screened using colony PCR method. The positive clones were further confirmed by DNA sequencing. The protein was expressed using BL21-DE3 Rosetta cells (Thermo Scientific, United States). Transformed cells were inoculated into 10 ml media as a primary and then into 1000 ml media supplemented 100 μg/ml ampicillin. The cultures were induced at 0.6 OD_{600} by adding 0.3 mM IPTG and incubated overnight at 18°C. Amy2 was purified using Ni-NTA (Merk-Sigma, United States) based affinity chromatography method following standard manufacturer's instructions. Cell pellets were dissolved in buffer A (20 mM Tris pH 7.4, 250 mM NaCl and 10 mM $CaCl_2$) and lysed by sonication. The sample was centrifuged at 12,000 g for 30 min. After centrifugation, pellet was discarded and the cell free supernatant was mixed with Ni-NTA resin. Protein was eluted by adding increasing the concentration of imidazole (20–500 mM) to the buffer A. The eluted fractions were loaded on 15% SDS PAGE to check purity and quality of the protein samples. The elution fractions were pooled and concentrated using 30 K centrifugal amicon ultrafiltration devices (Merk Millipore, United States). The protein was further purified using gel filtration chromatography using Superdex 200 Increase column. The purified Amy2 was used for further biochemical experiments.

Analytical Gel Filtration

The 0.5 ml of the purified protein sample was injected into the Superdex 200 Increase (GE Healthcare, United States) analytical gel filtration column pre-equilibrated with 20 mM Tris pH 7.4, 250 mM NaCl and 10 mM $CaCl_2$. The flow rate was kept at 0.5 ml per min. The chromatogram was recorded at both 280 and 220 nm.

Carotenoid Isolation

The carotenoids were extracted by following protocol published by Yatsunami et al. (2014). Briefly, 20 ml of haloarchaeal cultures were centrifuged at 12,000 g for 30 min at 25°C. The supernatant was discarded and pellets were further resuspended in 5 ml of acetone and methanol solution (7:3). This solution was incubated at room temperature for

30 min in dark followed by centrifugation at 12,000 *g* for 30 min. Reddish-pink supernatant was collected and pellets were discarded. Isolated colored supernatants were centrifuged at 3000 g in the speed-vac (Eppendorf concentrator plus) to remove methanol and acetone. The Reddish-pink colored pellets were finally dissolved in 200 μl of 100% methanol. UV spectra of extracted solutions were recorded at 200–700 nm range using a spectrophotometer (CECIL CE7500).

Enzymatic Assay on a Starch Agar Plate

The α-amylase activity of the heterologously expressed Amy2 was measured using a starch agar plate assay. 1 gm of starch agar was dissolved in 100 ml of distilled water and poured into petri-plates. 50 μl of 5 μM purified Amy2 was loaded into starch agar wells and incubated overnight at 37°C. The activity was checked by staining the plates with an iodine solution [15% solution containing 5% (wt/v) I_2 and 10% (wt/v) KI].

Biochemical Characterization of Amy2

Biochemical characterization of the enzyme was performed by the iodine-starch method. The reaction solution of 50 μl contained 1.0 ISU/ml enzyme and 0.2% (w/v) starch in reaction buffer (20 mM Tris–HCl, 250 mM NaCl and 10 mM $CaCl_2$, pH 7.4). A_{660} was measured by standard enzyme assay. The effect of temperature on enzyme activity was determined by incubating the enzyme at the temperature range of 5–75°C at 10°C intervals for 30 mins. Optimum pH for enzyme activity was determined by measuring the activity at 37°C for 30 min under different pH conditions at an interval of 1.0 pH unit. Sodium acetate buffer was used for pH 3–5, phosphate buffer was used for pH 6–7, Tris buffer was used for pH 8–9, and sodium bicarbonate buffer was used for pH 10–11. Effect of salt concentration was observed by examining the activity at different NaCl concentration (0–4 M) at 37°C for 30 min. The data presented is an average of three independent experiments.

Evolutionary Relationships of Taxa

The evolutionary history was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length = 0.23099238 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1500 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. This analysis involved 16 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 1483 positions in the final dataset. Evolutionary analyses were conducted in software suite MEGA X (Kumar et al., 2018).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GCA_900112175.1, GCA_900115785.1, GCA_000172995.2, GCA_000337835.1, GCA_000337815.1, GCA_000306765.2, GCA_000337795.1, GCA_000336815.1, GCA_000336955.1, GCA_001190965.1, GCA_000336755.1, GCA_000336735.1, GCA_000336795.1, GCA_000025685.1, GCA_001368915.1, and GCA_000337095.1.

AUTHOR CONTRIBUTIONS

KT, SP, and DV conceived the study. KT, SP, and AP coordinated the study. SP provided the strains. KT and DV designed experiments, analyzed the data, and wrote the manuscript with inputs from other co-authors. DV, GV, and CS performed experiments. All authors reviewed the results and approved the final version of the manuscript.

FUNDING

This work was supported by grants to KT, SP, and AP by Council of Scientific and Industrial Research, India. DV and GV are recipient of senior research fellowship from Council of Scientific and Industrial Research and University Grants Commission, India, respectively.

ACKNOWLEDGMENTS

The authors thank Mr. Surinder Singh, Mr. Randeep Singh, and Mr. Davender Singh at CSIR-IMTECH Chandigarh for supporting Laboratory experiments. KT would like to acknowledge members of Structural Biology Laboratory for useful suggestions and discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.02082/full#supplementary-material

REFERENCES

Abe, A., Yoshida, H., Tonozuka, T., Sakano, Y., and Kamitori, S. (2005). Complexes of *Thermoactinomyces vulgaris* R-47 alpha-amylase 1 and pullulan model oligossacharides provide new insight into the mechanism for recognizing substrates with alpha-(1,6) glycosidic linkages. *FEBS J.* 272, 6145–6153. doi: 10.1111/j.1742-4658.2005.05013.x

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L., and Beatson, S. A. (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402. doi: 10.1186/1471-2164-12-402

Amoozegar, M. A., Safarpour, A., Noghabi, K. A., Bakhtiary, T., and Ventosa, A. (2019). Halophiles and their vast potential in biofuel production. *Front. Microbiol.* 10:1895. doi: 10.3389/fmicb.2019.01895

Anderson, I., Scheuner, C., Goker, M., Mavromatis, K., Hooper, S. D., Porat, I., et al. (2011). Novel insights into the diversity of catabolic metabolism from ten haloarchaeal genomes. *PLoS One* 6:e20237. doi: 10.1371/journal.pone.0020237

Antranikian, G. (1992). *Microbial Degradation of Natural Products*, ed. G. Winkelmann Weinheim: VCH.

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

Bajpai, B., Chaudhary, M., and Saxena, J. (2015). Production and Characterization of alpha-amylase from an extremely halophilic archaeon, haloferax sp. *HA*10. *Food Technol. Biotechnol.* 53, 11–17. doi: 10.17113/ftb.53.01.15.3824

Becker, E. A., Seitzer, P. M., Tritt, A., Larsen, D., Krusor, M., Yao, A. I., et al. (2014). Phylogenetically driven sequencing of extremely halophilic archaea reveals strategies for static and dynamic osmo-response. *PLoS Genet.* 10:e1004784. doi: 10.1371/journal.pgen.1004784

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr., Brice, M. D., et al. (1977). The protein data bank. A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* 80, 319–324. doi: 10.1111/j.1432-1033.1977.tb11885.x

Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., et al. (1990). Calcium binding in alpha-amylases: an X-ray diffraction study at 2.1-A resolution of two enzymes from *Aspergillus*. *Biochemistry* 29, 6244–6249. doi: 10.1021/bi00478a019

Borchert, E., Knobloch, S., Dwyer, E., Flynn, S., Jackson, S. A., Johannsson, R., et al. (2017). Biotechnological potential of cold adapted *Pseudoalteromonas* spp. Isolated from 'Deep Sea' Sponges. *Mar. Drugs* 15:184. doi: 10.3390/md15060184

Bowers, K. J., Mesbah, N. M., and Wiegel, J. (2009). Biodiversity of poly-extremophilic Bacteria: does combining the extremes of high salt, alkaline pH and elevated temperature approach a physico-chemical boundary for life? *Saline Syst.* 5:9. doi: 10.1186/1746-1448-5-9

Bowers, K. J., and Wiegel, J. (2011). Temperature and pH optima of extremely halophilic archaea: a mini-review. *Extremophiles* 15, 119–128. doi: 10.1007/s00792-010-0347-y

Britton, G. (1995). Structure and properties of carotenoids in relation to function. *FASEB J.* 9, 1551–1558. doi: 10.1096/fasebj.9.15.8529834

Burns, D. G., Janssen, P. H., Itoh, T., Kamekura, M., Li, Z., Jensen, G., et al. (2007). *Haloquadratum* walsbyi gen. nov., sp. nov., the square haloarchaeon of Walsby, isolated from saltern crystallizers in Australia and Spain. *Int. J. Syst. Evol. Microbiol.* 57, 387–392. doi: 10.1099/ijs.0.64690-0

Capes, M. D., Dassarma, P., and Dassarma, S. (2012). The core and unique proteins of haloarchaea. *BMC Genomics* 13:39. doi: 10.1186/1471-2164-13-39

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373.

Chen, S., Liu, H. C., Zhou, J., and Xiang, H. (2016). *Halorubrum pallidum* sp. nov., an extremely halophilic archaeon isolated from a subterranean rock salt. *Int. J. Syst. Evol. Microbiol.* 66, 2980–2986. doi: 10.1099/ijsem.0.001129

Chi, M. C., Chen, Y. H., Wu, T. J., Lo, H. F., and Lin, L. L. (2010). Engineering of a truncated alpha-amylase of *Bacillus* sp. strain TS-23 for the simultaneous improvement of thermal and oxidative stabilities. *J. Biosci. Bioeng.* 109, 531–538. doi: 10.1016/j.jbiosc.2009.11.012

Chun, J., Lee, J. H., Jung, Y., Kim, M., Kim, S., Kim, B. K., et al. (2007). EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2259–2261. doi: 10.1099/ijs.0.64915-0

Cibrario, A., Peanne, C., Lailheugue, M., Campbell-Sills, H., and Dols-Lafargue, M. (2016). Carbohydrate metabolism in *Oenococcus oeni*: a genomic insight. *BMC Genomics* 17:984. doi: 10.1186/s12864-016-3338-2

Cowan, S. T., and Steel, K. J. (1965). Manual for the identification of medical bacteria. *J. Clin. Pathol.* 46:975.

DeMaere, M. Z., Williams, T. J., Allen, M. A., Brown, M. V., Gibson, J. A., Rich, J., et al. (2013). High level of intergenera gene exchange shapes the evolution of haloarchaea in an isolated Antarctic lake. *Proc. Natl. Acad. Sci. U.S.A.* 110, 16939–16944. doi: 10.1073/pnas.1307090110

Dong, G., Vieille, C., Savchenko, A., and Zeikus, J. G. (1997). Cloning, sequencing, and expression of the gene encoding extracellular alpha-amylase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl. Environ. Microbiol.* 63, 3569–3576. doi: 10.1128/aem.63.9.3569-3576.1997

Falb, M., Muller, K., Konigsmaier, L., Oberwinkler, T., Horn, P., Von Gronau, S., et al. (2008). Metabolism of halophilic archaea. *Extremophiles* 12, 177–196.

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

Fiedor, J., and Burda, K. (2014). Potential role of carotenoids as antioxidants in human health and disease. *Nutrients* 6, 466–488. doi: 10.3390/nu6020466

Flowers, T. J., and Colmer, T. D. (2015). Plant salt tolerance: adaptations in halophytes. *Ann. Bot.* 115, 327–331. doi: 10.1093/aob/mcu267

Giani, M., Garbayo, I., Vilchez, C., and Martinez-Espinosa, R. M. (2019). Haloarchaeal carotenoids: healthy novel compounds from extreme environments. *Mar. Drugs* 17:524. doi: 10.3390/md17090524

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044

Grant, W. (2001). Class III. Halobacteria class nov. *Bergey's Manual Syst. Bacteriol.* 1, 294–301.

Grant, W., and Larsen, H. (1989). Extremely halophilic archaeobacteria, order Halobacteriales ord. *nov. Bergey's Manual Syst. Bacteriol.* 3, 2216–2228.

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.

Gruber, C., Legat, A., Pfaffenhuemer, M., Radax, C., Weidler, G., Busse, H. J., et al. (2004). *Halobacterium noricense* sp. nov., an archaeal isolate from a bore core of an alpine Permian salt deposit, classification of *Halobacterium* sp. NRC-1 as a strain of *H. salinarum* and emended description of *H. salinarum*. *Extremophiles* 8, 431–439. doi: 10.1007/s00792-004-0403-6

Gunde-Cimerman, N., Oren, A., and Plemenitas, A. (2005). *Adaptation to Life at High Salt Concentrations in Archaea, Bacteria and Eukarya*. New York, NY: Kluwer Academic Publishers, 9.

Gupta, R., Gigras, P., Mohapatra, H., Goswami, V. K., and Chauhan, B. (2003). Microbial α-amylases: a biotechnological perspective. *Process Biochem.* 38, 1599–1616. doi: 10.1016/s0032-9592(03)00053-0

Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167–170. doi: 10.1126/science.1179555

Hsiu, J., Fischer, E. H., and Stein, E. A. (1964). Alpha-amylases as calcium-metalloenzymes. Ii. Calcium and the catalytic activity. *Biochemistry* 3, 61–66. doi: 10.1021/bi00889a011

Hutcheon, G. W., Vasisht, N., and Bolhuis, A. (2005). Characterisation of a highly stable alpha-amylase from the halophilic archaeon Haloarcula hispanica. *Extremophiles* 9, 487–495. doi: 10.1007/s00792-005-0471-2

Javor, B. J. (2012). *Hypersaline Environments: Microbiology and Biogeochemistry*. Berlin: Springer Science & Business Media.

Karray, F., Ben Abdallah, M., Kallel, N., Hamza, M., Fakhfakh, M., and Sayadi, S. (2018). Extracellular hydrolytic enzymes produced by halophilic bacteria and archaea isolated from hypersaline lake. *Mol. Biol. Rep.* 45, 1297–1309. doi: 10.1007/s11033-018-4286-5

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053

Kim, Y. B., Kim, J. Y., Song, H. S., Lee, C., Ahn, S. W., Lee, S. H., et al. (2018). Novel haloarchaeon Natrinema thermophila having the highest growth temperature among haloarchaea with a large genome size. *Sci. Rep.* 8:7777.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

Lányi, B. (1988). 1 Classical and rapid identification methods for medically important bacteria. *Methods Microbiol.* 19, 1–67. doi: 10.1016/s0580-9517(08)70407-0

Littlechild, J. A. (2015). Archaeal enzymes and applications in industrial biocatalysts. *Archaea* 2015:147671.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495.

Machius, M., Wiegand, G., and Huber, R. (1995). Crystal structure of calcium-depleted *Bacillus licheniformis* alpha-amylase at 2.2 A resolution. *J. Mol. Biol.* 246, 545–559.

Maier, L.-K., Alkhnbashi, O. S., Backofen, R., and Marchfelder, A. (2017). "CRISPR and Salty: CRISPR-Cas Systems in Haloarchaea," in *RNA Metabolism and Gene Expression in Archaea*, *Nucleic Acids and Molecular Biology*, Vol. 32, ed. B. Clouet-d'Orval (Cham: Springer), 243–269. doi: 10.1007/978-3-319-65795-0_11

Maier, L. K., Stachler, A. E., Brendel, J., Stoll, B., Fischer, S., Haas, K. A., et al. (2019). The nuts and bolts of the Haloferax CRISPR-Cas system I-B. *RNA Biol.* 16, 469–480. doi: 10.1080/15476286.2018.1460994

Makarova, K. S., and Koonin, E. V. (2015). Annotation and classification of CRISPR-Cas systems. *Methods Mol. Biol.* 1311, 47–75. doi: 10.1007/978-1-4939-2687-9_4

Mata-Gomez, L. C., Montanez, J. C., Mendez-Zavala, A., and Aguilar, C. N. (2014). Biotechnological production of carotenoids by yeasts: an overview. *Microb. Cell Fact.* 13:12. doi: 10.1186/1475-2859-13-12

Medema, M. H., Blin, K., Cimermancic, P., De Jager, V., Zakrzewski, P., Fischbach, M. A., et al. (2011). antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res.* 39, W339–W346.

Minegishi, H., Kamekura, M., Itoh, T., Echigo, A., Usami, R., and Hashimoto, T. (2010). Further refinement of the phylogeny of the Halobacteriaceae based on the full-length RNA polymerase subunit B'(rpoB') gene. *Int. J. Syst. Evol. Microbiol.* 60, 2398–2408. doi: 10.1099/ijs.0.017160-0

Montalvo-Rodriguez, R., Vreeland, R. H., Oren, A., Kessel, M., Betancourt, C., and Lopez-Garriga, J. (1998). *Halogeometricum borinquense* gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int. J. Syst. Bacteriol.* 48(Pt 4), 1305–1312. doi: 10.1099/00207713-48-4-1305

Montero-Lobato, Z., Ramos-Merchante, A., Fuentes, J. L., Sayago, A., Fernandez-Recamales, A., Martinez-Espinosa, R. M., et al. (2018). Optimization of growth and carotenoid production by haloferax mediterranei using response surface methodology. *Mar. Drugs* 16:372. doi: 10.3390/md16100372

Mormile, M. R., Biesen, M. A., Gutierrez, M. C., Ventosa, A., Pavlovich, J. B., Onstott, T. C., et al. (2003). Isolation of *Halobacterium salinarum* retrieved directly from halite brine inclusions. *Environ. Microbiol.* 5, 1094–1102. doi: 10.1046/j.1462-2920.2003.00509.x

Nagata, S., Hyde, C. C., and Miles, E. W. (1989). The alpha subunit of tryptophan synthase. Evidence that aspartic acid 60 is a catalytic residue and that the double alteration of residues 175 and 211 in a second-site revertant restores the proper geometry of the substrate binding site. *J. Biol. Chem.* 264, 6288–6296.

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.

Narsing Rao, M. P., Xiao, M., and Li, W. J. (2017). Fungal and bacterial pigments: secondary metabolites with wide applications. *Front. Microbiol.* 8:1113. doi: 10.3389/fmicb.2017.01113

Naziri, D., Hamidi, M., Hassanzadeh, S., Tarhriz, V., Maleki Zanjani, B., Nazemyieh, H., et al. (2014). Analysis of carotenoid production by *Halorubrum* sp. TBZ126; an extremely Halophilic Archeon from Urmia Lake. *Adv. Pharm. Bull.* 4, 61–67.

Neilands, J. B. (1995). Siderophores: structure and function of microbial iron transport compounds. *J. Biol. Chem.* 270, 26723–26726. doi: 10.1074/jbc.270.45.26723

Ollivier, B., Caumette, P., Garcia, J. L., and Mah, R. A. (1994). Anaerobic bacteria from hypersaline environments. *Microbiol. Rev.* 58, 27–38. doi: 10.1128/mmbr.58.1.27-38.1994

Onodera, M., Yatsunami, R., Tsukimura, W., Fukui, T., Nakasone, K., Takashina, T., et al. (2013). Gene analysis, expression, and characterization of an intracellular alpha-amylase from the extremely halophilic archaeon *Haloarcula japonica*. *Biosci. Biotechnol. Biochem.* 77, 281–288. doi: 10.1271/bbb.120693

Oren, A. (2002). Molecular ecology of extremely halophilic Archaea and Bacteria. *FEMS Microbiol. Ecol.* 39, 1–7. doi: 10.1111/j.1574-6941.2002.tb00900.x

Oren, A. (2010). Industrial and environmental applications of halophilic microorganisms. *Environ. Technol.* 31, 825–834. doi: 10.1080/09593330903370026

Oren, A. (2012). The function of gas vesicles in halophilic archaea and bacteria: theories and experimental evidence. *Life* 3, 1–20. doi: 10.3390/life3010001

Oren, A., Ventosa, A., and Grant, W. (1997). Proposed minimal standards for description of new taxa in the order Halobacteriales. *Int. J. Syst. Evol. Microbiol.* 47, 233–238. doi: 10.1099/00207713-47-1-233

Papke, R. T., Corral, P., Ram-Mohan, N., De La Haba, R. R., Sanchez-Porro, C., Makkay, A., et al. (2015). Horizontal gene transfer, dispersal and haloarchaeal speciation. *Life* 5, 1405–1426. doi: 10.3390/life5021405

Perez-Pomares, F., Bautista, V., Ferrer, J., Pire, C., Marhuenda-Egea, F. C., and Bonete, M. J. (2003). Alpha-amylase activity from the halophilic archaeon *Haloferax mediterranei*. *Extremophiles* 7, 299–306. doi: 10.1007/s00792-003-0327-6

Pettit, R. K. (2011). Culturability and secondary metabolite diversity of extreme microbes: expanding contribution of deep sea and deep-sea vent microbes to natural product discovery. *Mar. Biotechnol.* 13, 1–11. doi: 10.1007/s10126-010-9294-y

Pietra, F. (1997). Secondary metabolites from marine microorganisms: bacteria, protozoa, algae and fungi. Achievements and prospects. *Nat. Prod. Rep.* 14, 453–464.

Prakash, O., and Jaiswal, N. (2010). alpha-Amylase: an ideal representative of thermostable enzymes. *Appl. Biochem. Biotechnol.* 160, 2401–2414. doi: 10.1007/s12010-009-8735-4

Radax, C., Gruber, C., and Stan-Lotter, H. (2001). Novel haloarchaeal 16S rRNA gene sequences from Alpine Permo-Triassic rock salt. *Extremophiles* 5, 221–228. doi: 10.1007/s007920100192

Rodrigo-Banos, M., Garbayo, I., Vilchez, C., Bonete, M. J., and Martinez-Espinosa, R. M. (2015). Carotenoids from Haloarchaea and their potential in biotechnology. *Mar. Drugs* 13, 5508–5532. doi: 10.3390/md13095508

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Santorelli, M., Maurelli, L., Pocsfalvi, G., Fiume, I., Squillaci, G., La Cara, F., et al. (2016). Isolation and characterisation of a novel alpha-amylase from the extreme haloarchaeon *Haloterrigena turkmenica*. *Int. J. Biol. Macromol.* 92, 174–184. doi: 10.1016/j.ijbiomac.2016.07.001

Sarian, F. D., Janecek, S., Pijning, T., Ihsanawati, Nurachman, Z., Radjasa, O. K., et al. (2017). A new group of glycoside hydrolase family 13 alpha-amylases with an aberrant catalytic triad. *Sci. Rep.* 7:44230.

Sasser, M. (1990). *Identification of Bacteria by Gas Chromatography of Cellular Fatty Acids*. Newark, DE: MIDI Inc.

Schiraldi, C., Giuliano, M., and De Rosa, M. (2002). Perspectives on biotechnological applications of archaea. *Archaea* 1, 75–86. doi: 10.1155/2002/436561

Schubert, B. A., Lowenstein, T. K., Timofeeff, M. N., and Parker, M. A. (2010). Halophilic Archaea cultured from ancient halite, Death Valley, California. *Environ. Microbiol.* 12, 440–454. doi: 10.1111/j.1462-2920.2009.02086.x

Shigechi, H., Koh, J., Fujita, Y., Matsumoto, T., Bito, Y., Ueda, M., et al. (2004). Direct production of ethanol from raw corn starch via fermentation by use of a novel surface-engineered yeast strain codisplaying glucoamylase and alpha-amylase. *Appl. Environ. Microbiol.* 70, 5037–5040. doi: 10.1128/aem.70.8.5037-5040.2004

Singh, A., and Singh, A. K. (2017). Haloarchaea: worth exploring for their biotechnological potential. *Biotechnol. Lett.* 39, 1793–1800. doi: 10.1007/s10529-017-2434-y

Sivakumar, N., Li, N., Tang, J. W., Patel, B. K., and Swaminathan, K. (2006). Crystal structure of AmyA lacks acidic surface and provide insights into protein stability at poly-extreme condition. *FEBS Lett.* 580, 2646–2652. doi: 10.1016/j.febslet.2006.04.017

Sivaramakrishnan, S., Gangadharan, D., Nampoothiri, K. M., Soccol, C. R., and Pandey, A. (2006). a-Amylases from microbial sources–an overview on recent developments. *Food Technol. Biotechnol.* 44, 173–184.

Smibert, R., Krieg, N., Gerhardt, P., Murray, R., and Wood, W. (1994). *Methods for General and Molecular Bacteriology*. Washington DC: American Society for Microbiology.

Soppa, J. (2006). From genomes to function: haloarchaea as model organisms. *Microbiology* 152, 585–590. doi: 10.1099/mic.0.28504-0

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

Tamura, K., Nei, M., and Kumar, S. (2004). Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11030–11035. doi: 10.1073/pnas.0404206101

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

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

Verma, D. K., Baral, I., Kumar, A., Prasad, S. E., and Thakur, K. G. (2019). Discovery of bacteriorhodopsins in Haloarchaeal species isolated from Indian solar salterns: deciphering the role of the N-terminal residues in protein folding and functional expression. *Microb. Biotechnol.* 12, 434–446. doi: 10.1111/1751-7915.13359

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016

Vershinin, A. (1999). Biological functions of carotenoids–diversity and evolution. *Biofactors* 10, 99–104. doi: 10.1002/biof.5520100203

Vilchez, C., Forjan, E., Cuaresma, M., Bedmar, F., Garbayo, I., and Vega, J. M. (2011). Marine carotenoids: biological functions and commercial applications. *Mar. Drugs* 9, 319–333. doi: 10.3390/md9030319

Vos, M., and Didelot, X. (2009). A comparison of homologous recombination rates in bacteria and archaea. *ISME J.* 3, 199–208. doi: 10.1038/ismej.2008.93

Wang, S., Zheng, Z., Zou, H., Li, N., and Wu, M. (2019). Characterization of the secondary metabolite biosynthetic gene clusters in archaea. *Comput. Biol. Chem.* 78, 165–169. doi: 10.1016/j.compbiolchem.2018.11.019

Werries, E., and Muller, F. (1986). Studies on the substrate specificity of alpha- and beta-amylase of Entamoeba histolytica. *Mol. Biochem. Parasitol.* 18, 211–221. doi: 10.1016/0166-6851(86)90039-3

Winters, Y. D., Lowenstein, T. K., and Timofeeff, M. N. (2015). Starvation-survival in Haloarchaea. *Life* 5, 1587–1609. doi: 10.3390/life5041587

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.* 7:46. doi: 10.1186/1745-6150-7-46

Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444. doi: 10.1186/1471-2164-12-444

Yang, Y., Yatsunami, R., Ando, A., Miyoko, N., Fukui, T., Takaichi, S., et al. (2015). Complete biosynthetic pathway of the C50 carotenoid bacterioruberin from lycopene in the extremely halophilic archaeon *Haloarcula japonica*. *J. Bacteriol.* 197, 1614–1623. doi: 10.1128/jb.02523-14

Yatsunami, R., Ando, A., Yang, Y., Takaichi, S., Kohno, M., Matsumura, Y., et al. (2014). Identification of carotenoids from the extremely halophilic archaeon *Haloarcula japonica*. *Front. Microbiol.* 5:100. doi: 10.3389/fmicb.2014.00100

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

Zhang, J., Sun, Z., Sun, P., Chen, T., and Chen, F. (2014). Microalgal carotenoids: beneficial effects and potential in human health. *Food Funct.* 5, 413–425.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 22 October 2020
doi: 10.3389/fmicb.2020.576661

Distinct Distribution of Archaea From Soil to Freshwater to Estuary: Implications of Archaeal Composition and Function in Different Environments

Hualong Wang[1,2], Raven Bier[3†], Laura Zgleszewski[3], Marc Peipoch[3], Emmanuel Omondi[4†], Atanu Mukherjee[4†], Feng Chen[2], Chuanlun Zhang[5,6,7] and Jinjun Kan[3,6,8]*

[1] College of Marine Life Sciences, Ocean University of China, Qingdao, China, [2] Institute of Marine and Environmental Technology, University of Maryland Center for Environmental Science, Baltimore, MD, United States, [3] Microbiology Division, Stroud Water Research Center, Avondale, PA, United States, [4] Rodale Institute, Kutztown, PA, United States, [5] Department of Ocean Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [6] Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology, Shenzhen, China, [7] Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou, China, [8] Academy for Advanced Interdisciplinary Studies, Southern University of Science and Technology, Shenzhen, China

In addition to inhabiting extreme territories, Archaea are widely distributed in common environments spanning from terrestrial to aquatic environments. This study investigated and compared archaeal community structures from three different habitats (representing distinct environments): agriculture soils (from farming system trials FST, PA, United States), freshwater biofilms (from White Clay Creek, PA, United States), and estuary water (Chesapeake Bay, United States). High-throughput sequencing of 16S rRNA genes indicated that Thaumarchaeota, Euryarchaeota, Nanoarchaeota, Crenarchaeota, and Diapherotrites were the commonly found dominant phyla across these three environments. Similar to Bacteria, distinct community structure and distribution patterns for Archaea were observed in soils vs. freshwater vs. estuary. However, the abundance, richness, evenness, and diversity of archaeal communities were significantly greater in soils than it was in freshwater and estuarine environments. Indicator species (or amplicon sequence variants, ASVs) were identified from different nitrogen and carbon cycling archaeal groups in soils (Nitrososphaerales, Nitrosotaleales, Nitrosopumilales, Methanomassiliicoccales, Lainarchaeales), freshwater biofilms (Methanobacteria, Nitrososphaerales) and Chesapeake Bay (Marine Group II, Nitrosopumilales), suggesting the habitat-specificity of their biogeochemical contributions to different environments. Distinct functional aspects of Archaea were also confirmed by functional predictions (PICRUSt2 analysis). Further, co-occurrence

Edited by:
Andreas Teske, University of North Carolina at Chapel Hill, United States

Reviewed by:
Mohea Couturier, University of Illinois at Urbana–Champaign, United States
Likui Zhang, Yangzhou University, China

***Correspondence:**
Jinjun Kan
jkan@stroudcenter.org

†Present address:
Raven Bier, Savannah River Ecology Lab University of Georgia Aiken, Aiken, SC, United States
Emmanuel Omondi, Department of Agricultural and Environmental Sciences Tennessee State University, Nashville, TN, United States
Atanu Mukherjee, Maharishi International University, Fairfield, IA, United States

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

Received: *26 June 2020*
Accepted: *30 September 2020*
Published: *22 October 2020*

Citation:
Wang H, Bier R, Zgleszewski L, Peipoch M, Omondi E, Mukherjee A, Chen F, Zhang C and Kan J (2020) Distinct Distribution of Archaea From Soil to Freshwater to Estuary: Implications of Archaeal Composition and Function in Different Environments. Front. Microbiol. 11:576661. doi: 10.3389/fmicb.2020.576661

network analysis indicated that only soil Archaea formed stable modules. Keystone species (ASVs) were identified mainly from Methanomassiliicoccales, Nitrososphaerales, Nitrosopumilales. Overall, these results indicate a strong habitat-dependent distribution of Archaea and their functional partitions within the local environments.

Keywords: Archaea, composition and distribution, soil, freshwater, estuary, 16S rRNA gene, high-throughput sequencing

INTRODUCTION

Archaea represent a diverse, abundant and widely distributed group of microorganisms in the biosphere (Karner et al., 2001; Baker et al., 2020). On the basis of cell counts and molecular studies, Archaea account for more than 20% of all prokaryotes in ocean waters (Karner et al., 2001), about 1–5% in surface soil layers (Ochsenreiter et al., 2003; Bates et al., 2011), and probably represent the dominant group of microorganisms in marine subsurface sediments (Lipp et al., 2008). Further, they are abundant in many extreme environments (Klenk et al., 1998; Takai et al., 2008; Kan et al., 2011; Baker et al., 2020). Archaea have a significant impact on biogeochemical cycling (Offre et al., 2013). These microorganisms have evolved a variety of energy metabolisms using organic and/or inorganic electron donors and acceptors (including fixing carbon from inorganic sources) and thus, play crucial roles in global geochemical cycles including influencing greenhouse gas emissions (Offre et al., 2013). For example, methanogenesis and anaerobic methane oxidation are important steps in the carbon cycle that are performed by anaerobic Archaea (Liu and Whitman, 2008; Ferry, 2010; Shen et al., 2019). Although both Archaea and Bacteria contribute to the globally important process of aerobic ammonia oxidation, the wide distribution of ammonia oxidizing Archaea in virtually all investigated aerobic habitats indicates a prominent role for these organisms (Stahl and de la Torre, 2012; Alves et al., 2018). Oxidation of ammonia to nitrite, the first step of nitrification, is performed by aerobic Thaumarchaeota, as well as by some bacterial lineages. Thaumarchaeota is abundant in oceanic plankton and also widely distributed in terrestrial environments (Leininger et al., 2006; Offre et al., 2013). It is becoming apparent that the archaeal communities have much more varied and consequential roles in biogeochemical cycles across different environments than previously thought.

As the overwhelming majority of Archaea resist cultivation in the laboratory, the availability of molecular methods, such as 16S rRNA gene cloning and high-throughput amplicon sequencing, has boosted insight into their astonishing taxonomic and metabolic diversity and omnipresence (Offre et al., 2013; Baker et al., 2020). Different types of non-extremophilic Archaea have been detected in many environments, ranging from terrestrial to marine ecosystems (Auguet et al., 2010; Flemming and Wuertz, 2019). Members of Crenarchaeota and Euryarchaeota are globally distributed, and some lineages, often uncultivated ones, are abundant in waters (Bomberg et al., 2008), soils (Pesaro and Widmer, 2002; Timonen and Bomberg, 2009), and sediments (Schleper et al., 2005; Schleper and Nicol, 2010). For example, two major groups of Euryarchaeota, MG-II and MG-III Archaea are commonly found in estuarine and oceanic waters worldwide (DeLong, 1992; Massana et al., 2000; Xie et al., 2018; Chen et al., 2020). Some MG-II organisms contain rhodopsins which are predicted to use light to boost energy yield or facilitate substrate transport, and are also capable of protein degradation (Iverson et al., 2012; Orsi et al., 2016). Even Bacteria and Archaea perform the same functions, in some habitats Archaea may exhibit greater activity. For example, Herrmann et al. (2008) found that although archaeal and bacterial ammonia monooxygenase genes (*amo*A) had similar relative abundances in freshwater sediment, the enhanced nitrification activity observed in the rhizosphere of aquatic plant (*Littorella uniflora*) was due to ammonia-oxidizing Archaea.

While numerous studies have investigated archaeal distribution and abundance, there is substantial insight to be gained from evaluating the biogeography of this domain with current technology. Most early studies focused on a single environment or with limited spatial scales (Biller et al., 2012; Yao et al., 2013), and sequencing protocols varied among these studies, including the specific primers, sequencing depth, platforms and qualities. For example, Auguet et al. (2010) investigated the global distribution of archaeal communities by using the sequences present in databases at that time which were obtained predominantly by fluorescence in site hybridization (FISH) and denaturant gradient gel electrophoresis (DGGE) band sequencing. Global distribution of specific functional archaeal groups was also researched in earlier studies, including ammonia-oxidizing Archaea (Cao et al., 2013; Alves et al., 2018) and methanogenic Archaea (Wen et al., 2017). However, systematic and detailed investigations on the composition and distribution of whole archaeal communities as well as similarities and differences across different environments (e.g., from soil to fresh water to estuary) by high-throughput sequencing are still lacking. Indeed, little is known about the archaeal composition and distribution in lotic freshwater environments (Bomberg et al., 2008; Auguet et al., 2010) compared to terrestrial (Timonen and Bomberg, 2009; Karimi et al., 2018; Flemming and Wuertz, 2019) and other aquatic environments (e.g., lakes and oceans) (Francis et al., 2005; Offre et al., 2013; Liu et al., 2018; Baker et al., 2020). Filling this knowledge gap would provide important insights into evaluating the biogeography and ecological roles of archaeal communities among distinct ecological environments.

Here we investigated and compared archaeal community structures and their distribution patterns from three different habitats representing distinct environments using identical protocols of high-throughput sequencing analysis. A total of 230 samples were collected in this study: 95 agriculture soil samples from farming system trials (FST) at Rodale Institute,

Pennsylvania, United States, 59 freshwater biofilms from White Clay Creek, Pennsylvania, United States, and 76 surface water samples from Chesapeake Bay, the biggest estuary in the North America. Deep sequencing showed more than 4,000 unique archaeal 16S rRNA gene sequences, unveiled dominant archaeal taxa, and also indicated distinct species distribution across these environments. The archaeal abundance, richness, evenness, and diversity were compared among these environments. Higher archaeal abundance and diversities were associated with soils than freshwater biofilms and estuarine environments. In addition, indicator species, co-occurrence networks, and potential ecological functions of archaeal groups from distinct environments were also explored and discussed.

MATERIALS AND METHODS

Sample Collection

In total, 230 samples were collected from three habitats representing terrestrial, lotic and estuarine environments: agriculture soils (95 samples), freshwater biofilms (59 samples), and estuary water (76 samples) (**Figure 1**). Agricultural soils were collected from Rodale Institute's Farming Systems Trial (labeled as FST in this study) in Kutztown, Pennsylvania, United States (40.5509N, −75.7297W) that spans conventional and organic agriculture (**Figure 1**). Soil samples were collected in January 2019 using a 4.5 cm soil probe and taken to 1 m depth. Each soil core was sectioned into the following depth intervals: 0–10, 10–20, 20–30, and 30–60 cm. For each sample, four different soil cores spanning one agricultural treatment field were homogenized. A subsample was stored at −80°C until DNA extraction. In freshwaters especially upstream lotic environments, biofilms are the primary and dominant life forms of microorganisms, and therefore freshwater biofilms were collected from a flume study using continuous inputs of surface water from White Clay Creek (39.8592N, −75.7837W, southeastern Pennsylvania, United States; labeled as WCC) in summer 2018 (**Figure 1**). Biofilms were grown on autoclaved rocks in the flumes and collected on days 2, 5, 9, 13, and 20 of their development during July and August. Four rocks were swabbed on 37 cm^2 of surface area and frozen at −80°C until DNA extraction. The surface water samples in Chesapeake Bay (labeled as CB) were collected at seven stations along the middle axis of the Bay in February/March, May/June, August, and October from 2003 to 2005 (**Figure 1**). Details of estuary water sample collection, sampling locations and sample preparation have been described previously (Kan, 2006). In brief, 500 ml surface water (below 2 m) were taken at each sampling station and filtered immediately through 0.2 μm Millipore polycarbonate filters (47 mm diameter; Millipore Corporation, Billerica, MA, United States). The filters were stored at −80°C. Environmental parameters were summarized to describe the three habitats (**Supplementary Table S1**), although different environmental variables were measured in each habitat and data collection was performed differently. The average measurements values of the variables for agricultural soils, freshwater biofilms, and estuary waters were grouped according to different sampling depth, time points, and sites. Although the samples from these 3 different habitats were not collected at the same time, we believe they were good representatives for each environment based on the sample collection by including different farming practices and depth for FST (Rodale Institute's Farming Systems Trial, labeled as FST in this study), the time series of WCC biofilm development, and water samples across space and time in the Chesapeake Bay. However, it is also important to note that the freshwater biofilm and soil environments represent distinct seasons while estuary samples span seasons. Thus, we interpret community differences across environments as specific to the season or seasons when collected.

DNA Extraction and High-Throughput Sequencing

Environmental DNA from soils (FST) and biofilms (WCC) were extracted using DNeasy PowerSoil kits (Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA extraction from surface water samples (CB) followed the protocol described previously (Kan et al., 2006). DNA quantity was assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc., Waltham, MA). Library preparation and sequencing followed the 16S Metagenomic Sequencing Library Preparation protocol from Illumina[1]. Hypervariable region 4 (V4) of the SSU rRNA gene primers are now widely employed for defining microbial diversity (including both Archaea and Bacteria) across many different environments by high-throughput sequencing (Caporaso et al., 2011), including the Earth Microbiome Project's exploration of the global microbiome (Gilbert et al., 2014). The V4 variable region of the 16S rRNA genes was amplified using the universal forward primer 515f (5′-GTGYCAGCMGCCGCGGTAA-3′) (Parada et al., 2016) and universal reverse primer 806r (5′-GGACTACNVGGGTWTCTAAT-3′) (Apprill et al., 2015). PCR reactions contained 25 μl 2x Premix Taq, 1 μl each primer (10 mM) and 3 μl environmental DNA (20 ng/μl) template in a volume of 50 μl, and were amplified with following thermocycling program: 5 min at 94°C for initialization; 30 cycles of 30 s denaturation at 94°C, 30 s annealing at 52°C, and 30 s extension at 72°C; followed by 10 min final elongation at 72°C. Sequencing libraries were generated by using NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs, MA, United States) following manufacturer's recommendations. Each library was quantified using Qubit 2.0 fluorometer double-stranded DNA high sensitivity DNA kit (Thermo Fisher Scientific, MA, United States). Then libraries were normalized and mixed in equidensity ratios. High-throughput sequencing of 16S rRNA genes was performed at Magigene (Magigene Biotechnology Co. Ltd., Guangzhou, China) on an Illumina Nova6000 platform (paired-end 250 bp mode), following the manufacturer's guidelines. Raw sequencing data obtained in this study are available through the GenBank database under the accession numbers: PRJNA635685 (for FST soil

[1]https://support.illumina.com/

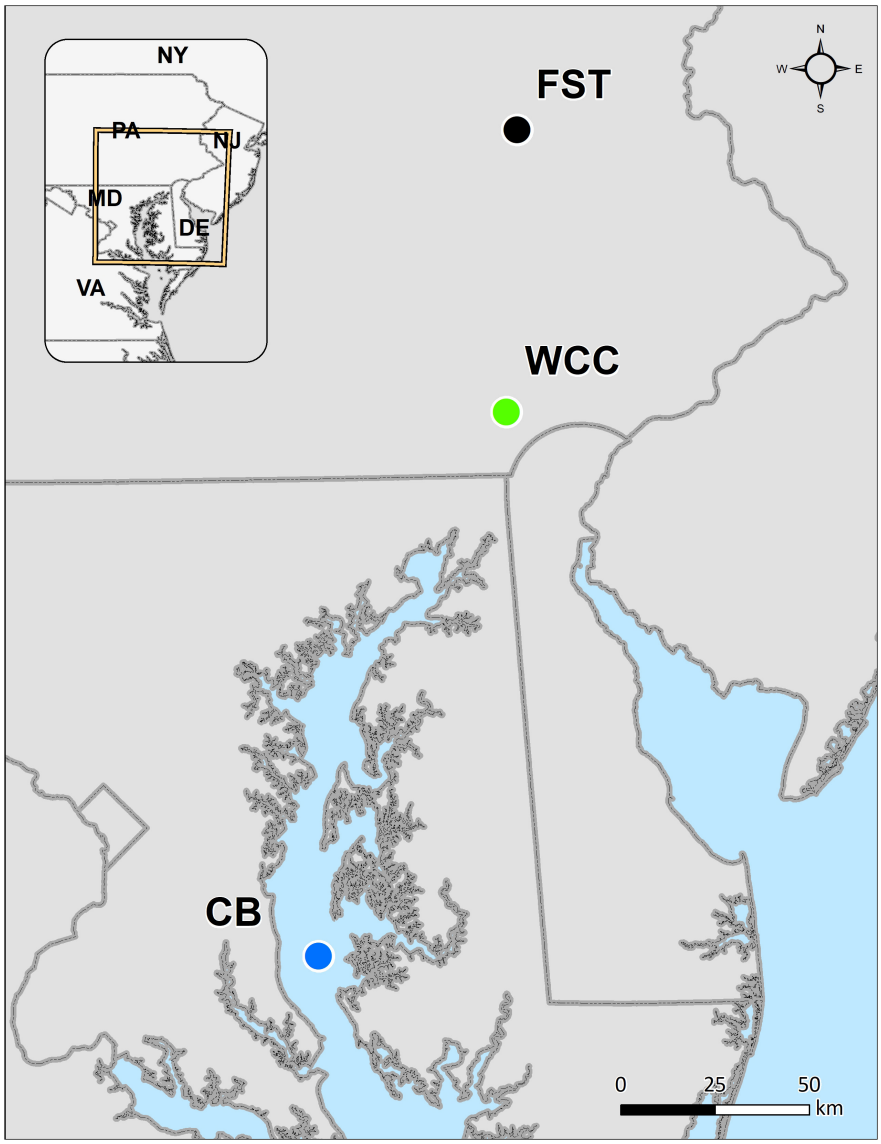

FIGURE 1 | Map of our sampling from agricultural soils (farming system trials, FST), freshwater biofilms at White Clay Creek (WCC), and the Chesapeake Bay estuary (CB).

samples), PRJNA631093 (for WCC biofilm samples), and PRJNA576689 (for CB).

Sequence Analyses

The QIIME 2 software package (version 2019.10) was used to process the raw sequence data (Bolyen et al., 2019). In brief, a total of 55,780,358 reads were obtained from these 230 samples after demultiplexing. Primers were removed with q2-DADA2 (Callahan et al., 2016), and the reads were trimmed to the same length (forward at 180 bp and reverse at 200 bp). q2-DADA2 was also used for denoising, filtering, merging, and chimera removal from these sequences and generate amplicon sequence variants (ASVs). The statistics of quality control filtering of 16S rRNA gene sequences in our samples were showed in **Supplementary Table S2**. A Naïve Bayes classifier artifact[2] was applied to assign the ASVs to taxa at 99% using the Silva classifier 132-99-515-806 dataset (Quast et al., 2012). For all ASV-based analyses, the original ASV table was rarified to a depth of 100,000 sequences per sample in order to minimize the sampling effects. An alpha rarefaction analysis at a sampling depth of 100,000 sequences was analyzed. The rarefaction curves clearly showed that our samples were sequenced to a sufficient depth in regard to prokaryotic diversity from soil to estuary (**Supplementary Figure S1**). In total, we obtained 4065 ASVs which were affiliated with the archaeal sequences and these were further evaluated. The QIIME 2 package was also used to generate Bray-Curtis distance matrices

[2]https://github.com/qiime2/q2-feature-classifier

and α-diversity metrics including evenness, observed ASVs, and Shannon-Wiener diversity.

Phylogenetic Analyses

Phylogenetic analysis was implemented in the Molecular Evolutionary Genetics Analysis (MEGA) X software (Kumar et al., 2018). Multiple sequence alignment was carried out using CLUSTAL W function (Thompson et al., 1994) with default parameters, and phylogenetic trees of 16S rRNA genes were reconstructed using the neighbor-joining method (Saitou and Nei, 1987). Bootstrap values were calculated with 1,000 re-sampling. The 16S rRNA gene sequence of *Escherichia coli* (NR 024570.1) was used as out-group for the analysis.

Co-occurrence Network Analysis

Relative abundance of archaeal ASVs were used to construct a co-occurrence network for each dataset from 3 environments. To avoid potentially erroneous sequences and improve interpretability of the dataset, we filtered out ASVs that were presented in fewer than three samples, and whose summed relative abundance was less than 0.1% in each specific network inference. All network constructions were done in R (R Core Team, 2020) (version 3.6.1) using the package "fdrtool" and "igraph" (Williams et al., 2014). We adapted the network construction code at GitHub[3]. The false discovery rate was estimated and corrected by the package "fdrtool." Co-occurrence networks for each environment were constructed using only statistically significant ($P < 0.01$) and robust (Spearman's correlation coefficient $> | 0.6|$) correlations (Barberán et al., 2012). Network visualization and topological analysis were carried out in Gephi (version 0.9.2) (Bastian et al., 2009). Other information regarding nodes (archaeal taxa), including taxonomy and relative abundances, were also imported into Gephi.

Indicator Taxa Analysis

Indicator taxa were identified for each habitat based on their specificity and fidelity to the environment. This analysis was conducted using the "multipatt" function in R-package "indicspecies" (Cáceres and Legendre, 2009) with 999 permutations and function "r.g" to account for unequal groups through correction of Pearson's phi coefficient of association. ASVs were selected as good indicators of a particular environment if the indicator value statistic was > 0.3 and $P < 0.05$ as previously recommended (Dufrêne and Legendre, 1997).

Prediction of Functional Content From Archaeal Communities

The potential functions of archaeal communities were inferred from the 16S rRNA gene high-throughput sequencing data by using PICRUSt2 (phylogenetic Investigation of Communities by Reconstruction of Unobserved States) (Douglas et al., 2019). We predicted KEGG orthology (KO) metagenomes, enzyme commission (EC) metagenomes and MetaCyc pathway abundances through a QIIME 2 module called q2-picrust2[4]. PICRUSt2 uses the 16S rRNA marker gene data to query a reference database for the closest reference genome available. Genomic-driven inference of function is then used to predict gene families, which are combined to estimate the composite metagenome. Briefly, a PICRUSt2-compatible ASV table was constructed in QIIME2. The accuracy for the predicted metagenome was tested through the Nearest Sequenced Taxon Index (NSTI), reflecting the presence of reference genomes that are closely related to the samples in the analysis.

Statistical Analyses

Statistical analyses were completed with R statistical software (version 3.6.1). Differences between major archaeal groups (phylum level) were compared using a one-way ANOVA ($P = 0.01$). Tukey's *post hoc* tests were used to test statistical significance ($P \leq 0.01$) of pair-wise comparisons. Calculation of alpha diversity (including Shannon-Wiener diversity, richness and evenness) of archaeal communities was done using the "diversity" function in the "vegan" package (Oksanen et al., 2019). Non-metric multidimensional scaling (NMDS) was used to assess differences across the three environments in community structures (Bacteria and Archaea, Bacteria only, and Archaea only). Differences of archaeal communities from three different habitats were further tested by analysis of similarities (ANOSIM). Both NMDS and ANOSIM were performed using the "metaMDS" and "anosim" functions in the "vegan" R package, respectively.

RESULTS

Uneven Diversity of Archaeal Community Across Three Habitats

Archaea comprised a broad diversity of taxa that were clearly uneven in their distribution across agriculture soils, freshwater biofilms, and estuary waters (**Figure 2**). Generally, alpha diversity of archaeal communities (Shannon-Wiener, ASV observed richness and evenness indices) were distinct among habitats (Kruskal-Wallis, $P < 0.05$) (**Figure 2**). Shannon-Wiener diversity of archaeal community was significantly greater in samples collected in agriculture soils than those collected at freshwater biofilms and estuarine surface waters ($P < 0.05$; **Figure 2A**). Similarly, evenness of archaeal communities was significantly higher in agriculture soils than in freshwater biofilms and estuarine surface water ($P < 0.05$; **Figure 2B**). Observed ASV richness in agriculture soils was also significantly higher than those in freshwater biofilms and estuary water samples ($P < 0.05$; **Figure 2C**). In addition, ASV richness in freshwater biofilms was significantly higher than that in estuarine surface waters ($P < 0.05$; **Figure 2C**). The ratio of total relative abundance of Bacteria to Archaea varied significantly for each habitat ($P < 0.05$; **Figure 2D**). The Bacteria:Archaea ratio in the agriculture soil samples was significantly lower compared to samples collected from the freshwater biofilms and estuary water, while there was no distinct difference between the freshwater biofilm samples and the estuarine surface water samples (**Figure 2D**). The low ratio of Bacteria:Archaea showed that Archaea was more abundant

[3]https://github.com/ryanjw/co-occurrence

[4]https://github.com/picrust/picrust2/wiki/q2-picrust2-Tutorial

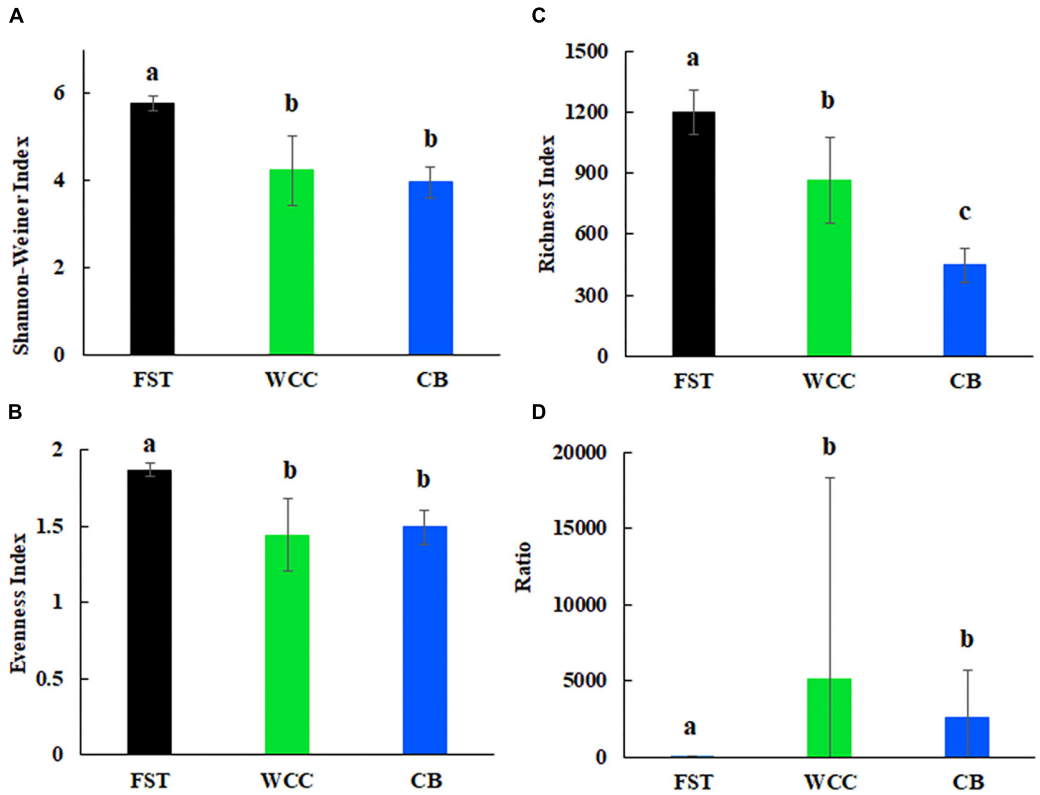

FIGURE 2 | Alpha diversity for agricultural soils (FST, n = 95), freshwater biofilms (WCC, n = 59) and estuary (CB, n = 76): **(A)** Shannon-Weiner index; **(B)** evenness; **(C)** observed richness; and **(D)** ratio of relative abundance of bacteria to archaea. Bars represent mean with standard errors. Different letters above the bars indicate significant differences between habitats ($P < 0.05$), whilst shared letters indicate no significant difference.

in the soil environment compared to the freshwater biofilms and estuary water.

Detailed Archaeal Community Structure

The percentage of total 16S rRNA gene sequences dominated by Archaea varied in the three habitats: soils (2.03–17.10%), freshwater biofilms (<0.01–0.16%), and estuary water (0.01–9.39%) (**Supplementary Table S2**). The archaeal communities were dominated by Thaumarchaeota, Euryarchaeota, Nanoarchaeota, Crenarchaeota, and Diapherotrites (**Figure 3**). In addition, the relative abundances of these major archaeal groups varied among habitats (**Figure 3**). Euryarchaeota, Thaumarchaeota and Diapherotrites were more abundant in agriculture soils than in the freshwater biofilms and estuary water ($P < 0.01$), while Crenarchaeota were more predominant in freshwater biofilms compared to estuary water ($P < 0.01$) (**Figure 3**). The proportion of Thaumarchaeota and Nanoarchaeota was significantly different across the three

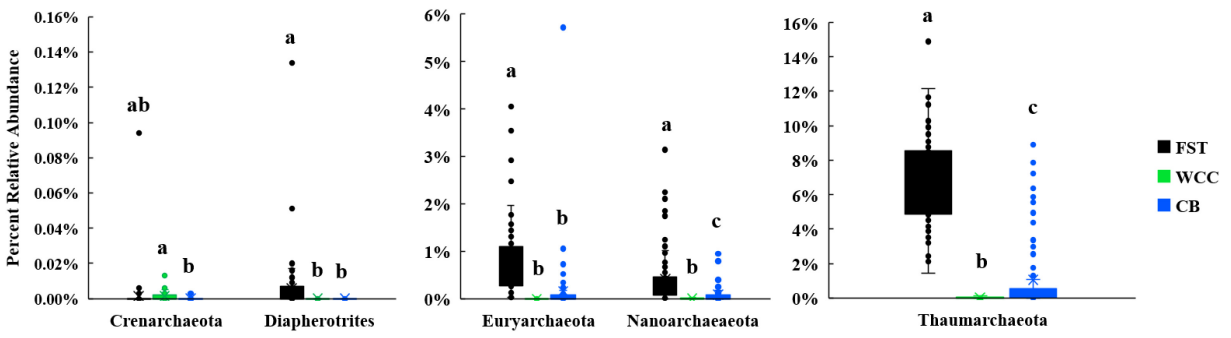

FIGURE 3 | Boxplots for major archaeal phyla from three environments [agricultural soils (FST, n = 95), freshwater biofilms (WCC, n = 59) and estuary (CB, n = 76)]. Different letters above the bars indicate significant differences between habitats ($P < 0.05$), whilst shared letters indicate no significant difference.

environments and exhibited the same distribution: they were more abundant in soils, than in estuary water, and were even less abundant in freshwater biofilms ($P < 0.01$) (**Figure 3**). In total, archaeal communities had the highest relative abundance in soils compared to the other two environments, and the major group was Thaumarchaeota (**Figure 3**).

There was a clear phylogenetic shift of major groups of archaeal communities across three habitats (**Supplementary Table S3** and **Supplementary Figures S2A–E**). For example, within the phylum Thaumarchaeota, Nitrosopumilaceae (ASV3197) dominated the archaeal communities in the Chesapeake Bay, but was essentially absent, or nearly so, in freshwater biofilms and agricultural soil environments. This implies that the organism represented by phylotype ASV3197 is probably a typical estuarine/brackish species. Two other members of Nitrosopumilaceae (ASV54 and ASV64) were dominant in the freshwater biofilms, yet they were nearly undetected in estuary waters. Furthermore, ASV66 (Nitrosotaleaceae) and ASV64 and 57 (Nitrosopumilaceae) were dominant in the soil environments (**Supplementary Table S3** and **Supplementary Figure S2E**). Similar results were also observed in the phylum Euryarchaeota

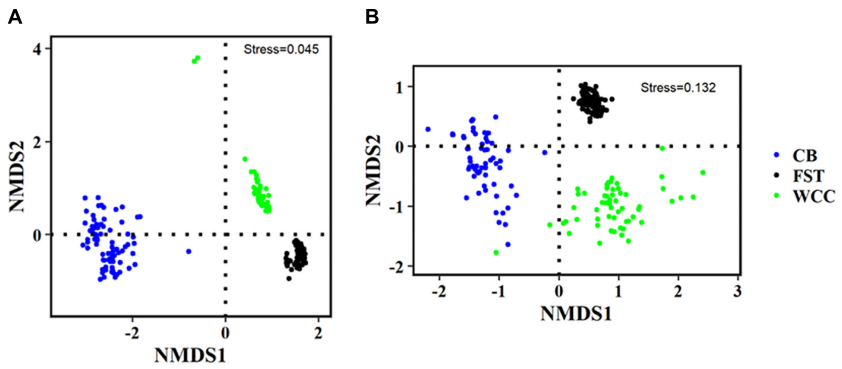

FIGURE 4 | Non-metric multidimensional scaling (NMDS) plots of different communities. **(A)** Bacteria only; and **(B)** Archaea only. FST, agricultural soils; WCC, freshwater biofilms; CB, estuary.

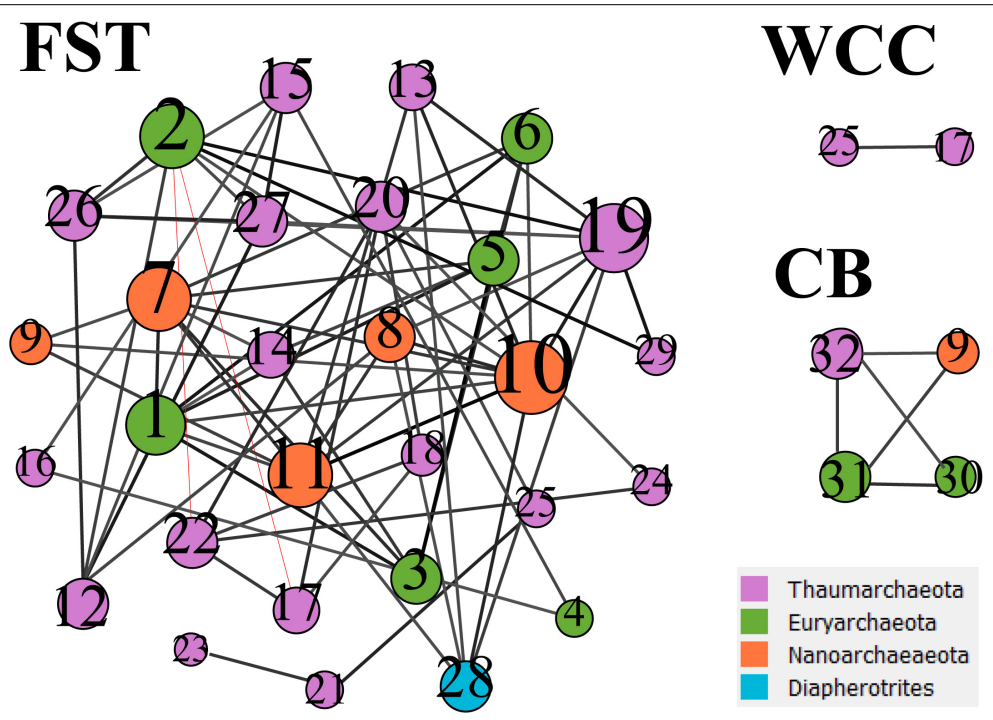

FIGURE 5 | Co-occurrence networks from three environments. FST, agricultural soils; WCC, freshwater biofilms; CB, estuary. The networks were based on archaeal ASVs which have occurred at least in three samples from each environment. Node size represents the number of associations between nodes. The number in the nodes refers to the taxa listed in **Table 1**. Black lines between nodes represent positive correlations, and red lines represent negative correlations.

(**Supplementary Figure S2C**). Two members of Marine Group II (ASV15 and ASV3190) dominated the Euryarchaeota groups in the estuary waters, while one member of Methanobacteriaceae (ASV3189) dominated in the freshwater biofilms. In soils, there were different dominant taxa within Euryarchaeota groups, including 4 members of Thermoplasmata (ASV16, ASV17, ASV19, and ASV21) (**Supplementary Table S3**).

Distribution Patterns of Archaeal Communities

Archaeal communities from agriculture soils, freshwater biofilms, and estuary water showed clear visual separation from each other in regard to the distribution patterns of the microbial community compositions (NMDS, **Figure 4** and **Supplementary Figure S3**). The three different habitats were characterized based on their physico-chemical properties (**Supplementary Table S1**), and habitat classification was a strong structuring factor of the microbial assemblages and communities clearly grouped according to their environmental types (ANOSIM test, $P < 0.01$) (**Figure 4**). The partitioned distribution patterns were observed for bacterial communities in soils vs. freshwater vs. estuary (**Figure 4A**). Similar to Bacteria, global variation of archaeal communities was strongly separated based on habitats: archaeal communities showed distinct distribution patterns across soils, freshwater, and estuary (**Figure 4B**). The abundance-weighted percentage of Archaea used for NMDS analysis accounts for a low percentage (3.75%) of combined Bacteria and Archaea. Therefore, NMDS plots of Bacteria only and combined Archaea and Bacteria were quite similar to each other with almost the same stress index (**Figure 4A**, 0.0450; **Supplementary Figure S3**, 0.0449). The ANOSIM statistic results further confirmed a clear separation of archaeal communities across 3 sampling environments ($R = 0.9134$, $P = 0.0001$).

Archaeal Networks

Co-occurrence network analysis showed that archaeal associations were distinct in three habitats. Stronger significant relationships among archaeal communities were observed in the agriculture soils (29 taxa with 68 correlations) compared to the freshwater biofilms (2 taxa with 1 correlation) and estuary surface waters (4 taxa with 5 correlations) (**Figure 5**). Those key archaeal taxa with the highest number of associations in soils were affiliated with Woesearchaeia (Nanoarchaeota; 4 members), Nitrososphaeraceae (Thaumarchaeota) and Methanomassiliicoccales (Euryarchaeota, 5 members) (**Table 1**). Four taxa that occurred in the archaeal networks from estuary waters belonged to Marine Group II (Euryarchaeota) and Nitrososphaeria and Woesearchaeia (Nanoarchaeota), while two members of freshwater networks were both affiliated to the Nitrososphaeraceae (Thaumarchaeota) (**Table 1**). Our results clearly showed that habitat differences could significantly influence the archaeal interactions and networks across different environments.

Archaeal Indicator ASVs

Archaea ASVs that were indicators for each habitat type represented five phyla and ten classes, with the majority of ASV indicators belonging to uncultured or unidentified species (**Table 2** and **Figure 6**). The number of indicator ASVs for each habitat varied widely with agricultural soils having 50 indicators while freshwater biofilms and estuaries had four and six indicator ASVs, respectively. Nitrososphaeria was the class containing the most ASVs indicative of agricultural soils (19 indicators), followed by classes Thermoplasmata and Woesearchaeia each with eight indicator ASVs. Indicator ASVs for estuaries were from classes Thermoplasmata (3 indicators), Woesearchaeia (2 indicators), and Nitrososphaeria (1 indicator). Freshwater biofilms had one archaeal indictor each from the classes Nitrososphaeria, Woesearchaeia, Bathyarchaeia, and Methanobacteria. Methanobacteria was the only class with an indicator ASV that was restricted to aquatic environments, occurring only for freshwater biofilms. Agricultural soils also contained a methanogenic ASV indicator which was of class Methanomicrobia.

Prediction of Archaeal Functional Profiles Across Habitats

The potential metabolic functions of archaeal communities and their proportions of occurrence in each step of metabolic processes were predicted by PICRUSt2. In total 352 functional pathways were found in this study. Significant differences of the contribution and quantity of the top 50 functional pathways in archaeal communities were observed across the three habitats (**Figure 7**). These pathways were classified to 13 main metabolic groups. The total percentage of these pathways ranged from 479.5 to 22.5% in soil samples and from 40.6 to 1.5% in freshwater biofilm samples, which included "Nucleoside and Nucleotide Biosynthesis," "Amino Acid Biosynthesis," "Carbohydrate Biosynthesis," "C1 Compound Utilization and Assimilation," "TCA cycle," "Fermentation" and several other pathways (**Figure 7**). The predicted pathways were most abundant in soils, followed by freshwater biofilms, and were unidentified in estuary water samples. Those most abundant pathways in soils included "incomplete reductive TCA cycle," "aerobic respiration I (cytochrome c)," "5-aminoimidazole ribonucleotide biosynthesis" and "Calvin-Benson-Bassham cycle" while the most abundant metabolism pathways in the freshwater biofilms were identified as "aerobic respiration I (cytochrome c)," "incomplete reductive TCA cycle," "L-isoleucine biosynthesis II," and "L-isoleucine biosynthesis IV" (**Figure 7**).

DISCUSSION

Heterogeneous Distribution of Archaeal Groups Across Habitats

To gain knowledge on the true ecology of a Domain, all its components should be analyzed as a whole (Auguet et al., 2010). This comparative ecological study revealed detailed archaeal composition and community distribution across three

TABLE 1 | Key species/gatekeepers for three habitats (Archaea only).

Environments	Labels in network	id	Phylum	Class	Order	Family	Genus	Number of samples it was occurred	Total reads in each habitat	Degree
FST	10	ASV41	Nanoarchaeaeota	Woesearchaeia	Uncultured euryarchaeote	Uncultured euryarchaeote	Uncultured euryarchaeote	93	19,274	10
	19	ASV57	Thaumarchaeota	Nitrososphaeria	Nitro-sosphaerales	Nitro-sosphaeraceae	Candidatus Nitrososphaera	95	90,605	9
	2	ASV16	Euryarchaeota	Thermoplasmata	Methanomassi-liicoccales	Uncultured	Uncultured archaeon	65	23,254	8
	7	ASV32	Nanoarchaeaeota	Woesearchaeia	Candidatus Staskawiczbacteria bacterium	Candidatus Staskawiczbacteria bacterium	RIFOXYA2 _FULL_32_7	93	8,360	8
	11	ASV42	Nanoarchaeaeota	Woesearchaeia	NA	NA	NA	88	7,101	8
	1	ASV12	Euryarchaeota	Thermoplasmata	Marine Group II	Uncultured archaeon	Uncultured archaeon	83	2,947	7
	3	ASV17	Euryarchaeota	Thermoplasmata	Methanomassi-liicoccales	NA	NA	95	12,145	5
	5	ASV19	Euryarchaeota	Thermoplasmata	Uncultured	Uncultured archaeon	Uncultured archaeon	95	17,243	5
	6	ASV21	Euryarchaeota	Thermoplasmata	NA	NA	NA	95	23,126	5
	8	ASV34	Nanoarchaeaeota	Woesearchaeia	Nanoarchaeota archaeon SCGC AAA011-D5	Nanoarchaeota archaeon SCGC AAA011-D5	Nanoarchaeota archaeon SCGC AAA011-D5	47	573	5
WCC	17	ASV54	Thaumarchaeota	Nitrososphaeria	Nitro-sosphaerales	Nitro-sosphaeraceae	Candidatus Nitrocosmicus	45	899	1
	25	ASV64	Thaumarchaeota	Nitrososphaeria	Nitro-sosphaerales	Nitro-sosphaeraceae	NA	30	655	1
CB	31	ASV3190	Euryarchaeota	Thermoplasmata	Marine Group II	Marine Group II_ unidentified	Marine Group II_unidentified	45	2,713	3
	32	ASV3197	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	Nitrosopumilaceae	Candidatus Nitrosopumilus	4	72	3
	30	ASV15	Euryarchaeota	Thermoplasmata	Marine Group II	Marine Group II_ unidentified	Marine Group II_unidentified	75	77,835	2
	9	ASV40	Nanoarchaeaeota	Woesearchaeia	Woesearchaeia_ unidentified	Woesearchaeia_ unidentified	Woesearchaeia_ unidentified	12	291	2

FST = agricultural soils, WCC = freshwater biofilms, CB = estuary.

TABLE 2 | Archaeal indicator ASVs for three sampling habitats.

Habitats	id	Phylum	Class	Order	Family	Genus	Species	Indicator value index	*P*
FST	ASV2	Crenarchaeota	Bathyarchaeia	NA	NA	NA	NA	0.153	0.04
FST	ASV4	Diapherotrites	Iainarchaeia	Iainarchaeales	Diapherotrites archaeon SCGC AAA011-K09	Diapherotrites archaeon SCGC AAA011-K09	Diapherotrites archaeon SCGC AAA011-K09	0.304	0.001
FST	ASV5	Diapherotrites	Iainarchaeia	Iainarchaeales	Diapherotrites archaeon SCGC AAA011-N19	Diapherotrites archaeon SCGC AAA011-N19	Diapherotrites archaeon SCGC AAA011-N19	0.379	0.001
FST	ASV7	Diapherotrites	Iainarchaeia	Iainarchaeales	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.236	0.001
FST	ASV10	Euryarchaeota	Methanomicrobia	Methanosarcinales	Methanosarcinaceae	Methanosarcina	NA	0.197	0.014
FST	ASV12	Euryarchaeota	Thermoplasmata	Marine Group II	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.616	0.001
FST	ASV13	Euryarchaeota	Thermoplasmata	Marine Group II	Uncultured euryarchaeote	Uncultured euryarchaeote	Uncultured euryarchaeote	0.279	0.001
FST	ASV14	Euryarchaeota	Thermoplasmata	Marine Group II	Uncultured haloarchaeon	Uncultured haloarchaeon	Uncultured haloarchaeon	0.314	0.001
FST	ASV17	Euryarchaeota	Thermoplasmata	Methanomassi-liicoccales	NA	NA	NA	0.673	0.001
FST	ASV16	Euryarchaeota	Thermoplasmata	Methanomassi-liicoccales	Uncultured	Uncultured archaeon	Uncultured archaeon	0.327	0.001
FST	ASV21	Euryarchaeota	Thermoplasmata	NA	NA	NA	NA	0.68	0.001
FST	ASV18	Euryarchaeota	Thermoplasmata	Uncultured	Crenarchaeote SRI-298	Crenarchaeote SRI-298	Crenarchaeote SRI-298	0.239	0.001
FST	ASV19	Euryarchaeota	Thermoplasmata	Uncultured	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.683	0.001
FST	ASV22	Nanoarchaeaeota	Nanohaloarchaeia	Aenigmarchaeales	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.36	0.001
FST	ASV25	Nanoarchaeaeota	Nanohaloarchaeia	Deep sea euryarchaeotic group (DSEG)	NA	NA	NA	0.153	0.035
FST	ASV24	Nanoarchaeaeota	Nanohaloarchaeia	Deep sea euryarchaeotic group (DSEG)	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.197	0.012
FST	ASV26	Nanoarchaeaeota	Nanohaloarchaeia	NA	NA	NA	NA	0.4	0.001
FST	ASV35	Nanoarchaeaeota	Woesearchaeia	Archaeon GW2011_AR13	Archaeon GW2011_AR13	Archaeon GW2011_AR13	Archaeon GW2011_AR13	0.292	0.001
FST	ASV28	Nanoarchaeaeota	Woesearchaeia	Candidatus Diapherotrites archaeon ADurb.Bin253	Candidatus Diapherotrites archaeon ADurb.Bin253	Candidatus Diapherotrites archaeon ADurb.Bin253	Candidatus Diapherotrites archaeon ADu	0.36	0.001
FST	ASV29	Nanoarchaeaeota	Woesearchaeia	Candidatus Nomurabacteria bacterium RIFCSPLOWO2_02_FULL_42_17	Candidatus Nomurabacteria bacterium RIFCSPLOWO2_02_FULL_42_17	Candidatus Nomurabacteria bacterium RIFCSPLOWO2_02_FULL_42_17	D	0.248	0.001

(Continued)

TABLE 2 | Continued

Habitats	id	Phylum	Class	Order	Family	Genus	Species	Indicator value index	*P*
FST	ASV32	Nanoarchaeaeota	Woesearchaeia	Candidatus Staskawiczbacteria bacterium RIFOXYA2_FULL_32_7	Candidatus Staskawiczbacteria bacterium RIFOXYA2_FULL_32_7	Candidatus Staskawiczbacteria bacterium RIFOXYA2_FULL_32_7	Candi	0.514	0.001
FST	ASV42	Nanoarchaeaeota	Woesearchaeia	NA	NA	NA	NA	0.341	0.001
FST	ASV34	Nanoarchaeaeota	Woesearchaeia	Nanoarchaeota archaeon SCGC AAA011-D5	Nanoarchaeota archaeon SCGC AAA011-D5	Nanoarchaeota archaeon SCGC AAA011-D5	Nanoarchaeota archaeon SCGC AAA011-D5	0.381	0.001
FST	ASV38	Nanoarchaeaeota	Woesearchaeia	Uncultured archaeon CLEAR-15	Uncultured archaeon CLEAR-15	Uncultured archaeon CLEAR-15	Uncultured archaeon CLEAR-15	0.226	0.002
FST	ASV41	Nanoarchaeaeota	Woesearchaeia	Uncultured euryarchaeote	Uncultured euryarchaeote	Uncultured euryarchaeote	Uncultured euryarchaeote	0.408	0.001
FST	ASV46	Thaumarchaeota	Group 1.1c	NA	NA	NA	NA	0.389	0.001
FST	ASV43	Thaumarchaeota	Group 1.1c	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.228	0.001
FST	ASV44	Thaumarchaeota	Group 1.1c	Uncultured crenarchaeote	Uncultured crenarchaeote	Uncultured crenarchaeote	Uncultured crenarchaeote	0.341	0.001
FST	ASV45	Thaumarchaeota	Group 1.1c	Uncultured thaumarchaeote	Uncultured thaumarchaeote	Uncultured thaumarchaeote	Uncultured thaumarchaeote	0.287	0.001
FST	ASV72	Thaumarchaeota	NA	NA	NA	NA	NA	0.445	0.001
FST	ASV48	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	Nitrosopumilaceae	Candidatus Nitrosoarchaeum	Uncultured archaeon	0.236	0.001
FST	ASV49	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	Nitrosopumilaceae	Candidatus Nitrosotenuis	NA	0.535	0.001
FST	ASV50	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	Nitrosopumilaceae	Uncultured archaeon	Uncultured archaeon	0.243	0.001
FST	ASV54	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Candidatus Nitrocosmicus	NA	0.602	0.001
FST	ASV53	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Candidatus Nitrocosmicus	Uncultured bacterium	0.216	0.002
FST	ASV57	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Candidatus Nitrososphaera	NA	0.559	0.001
FST	ASV55	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Candidatus Nitrososphaera	Uncultured crenarchaeote	0.659	0.001
FST	ASV56	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Candidatus Nitrososphaera	Unidentified archaeon	0.371	0.001
FST	ASV58	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Metagenome	Metagenome	0.649	0.001

(Continued)

TABLE 2 | Continued

Habitats	id	Phylum	Class	Order	Family	Genus	Species	Indicator value index	*P*
FST	ASV64	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	NA	NA	0.877	0.001
FST	ASV59	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Uncultured bacterium	Uncultured bacterium	0.809	0.001
FST	ASV60	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Uncultured compost archaeon	Uncultured compost archaeon	0.581	0.001
FST	ASV61	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Unidentified archaeon SCA1150	Unidentified archaeon SCA1150	0.574	0.001
FST	ASV62	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Unidentified archaeon SCA1151	Unidentified archaeon SCA1151	0.819	0.001
FST	ASV63	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Unidentified archaeon SCA1173	Unidentified archaeon SCA1173	0.181	0.011
FST	ASV65	Thaumarchaeota	Nitrososphaeria	Nitrosotaleales	Nitrosotaleaceae	Candidatus Nitrosotalea	Uncultured archaeon	0.451	0.001
FST	ASV69	Thaumarchaeota	Nitrososphaeria	Nitrosotaleales	Nitrosotaleaceae	NA	NA	0.241	0.001
FST	ASV66	Thaumarchaeota	Nitrososphaeria	Nitrosotaleales	Nitrosotaleaceae	Uncultured archaeon	Uncultured archaeon	0.685	0.001
FST	ASV67	Thaumarchaeota	Nitrososphaeria	Nitrosotaleales	Nitrosotaleaceae	Uncultured crenarchaeote	Uncultured crenarchaeote	0.246	0.001
FST	ASV71	Thaumarchaeota	SCGC AB-179-E04	Uncultured crenarchaeote	Uncultured crenarchaeote	Uncultured crenarchaeote	Uncultured crenarchaeote	0.154	0.035
WCC	ASV1	Crenarchaeota	Bathyarchaeia	Archaeon RBG_16_50_20	Archaeon RBG_16_50_20	Archaeon RBG_16_50_20	Archaeon RBG_16_50_20	0.332	0.001
WCC	ASV3189	Euryarchaeota	Methanobacteria	Methanobacteriales	Methanobacteriaceae	Methanobacterium	NA	0.327	0.001
WCC	ASV37	Nanoarchaeaeota	Woesearchaeia	Metagenome	Metagenome	Metagenome	Metagenome	0.549	0.001
WCC	ASV5035	Thaumarchaeota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Uncultured archaeon	Uncultured archaeon	0.269	0.001
CB	ASV3190	Euryarchaeota	Thermoplasmata	Marine Group II	Marine metagenome	Marine metagenome	Marine metagenome	0.209	0.001
CB	ASV15	Euryarchaeota	Thermoplasmata	Marine Group II	NA	NA	NA	0.172	0.001
CB	ASV3191	Euryarchaeota	Thermoplasmata	Marine Group II	Uncultured marine euryarchaeote DH148-W1	Uncultured marine euryarchaeote DH148-W1	Uncultured marine euryarchaeote DH148-W1	0.227	0.001
CB	ASV39	Nanoarchaeaeota	Woesearchaeia	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	Uncultured archaeon	0.214	0.002
CB	ASV40	Nanoarchaeaeota	Woesearchaeia	Uncultured bacterium	Uncultured bacterium	Uncultured bacterium	Uncultured bacterium	0.169	0.021
CB	ASV3197	Thaumarchaeota	Nitrososphaeria	Nitrosopumilales	Nitrosopumilaceae	Candidatus Nitrosopumilus	NA	0.377	0.001

FST, agricultural soils; WCC, freshwater biofilms; CB, estuary.

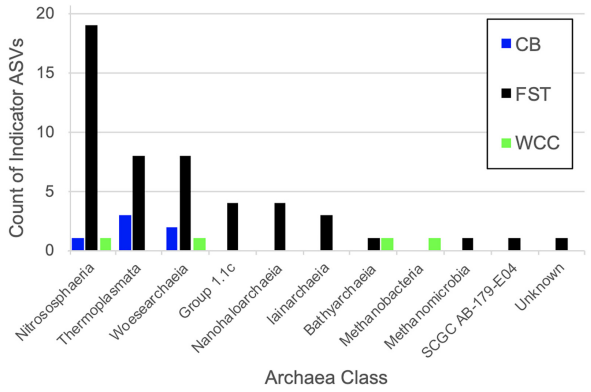

FIGURE 6 | Classes of archaeal ASVs identified as indicators for each environment type. FST, agricultural soils; WCC, freshwater biofilms; CB, estuary.

habitats by high-throughput sequencing analysis. Distinct dominant archaeal groups and phylogenetic shifts were observed in three habitat-season combinations: agriculture soils from winter, freshwater biofilms from summer, and estuary waters from all seasons. Although Thaumarchaeota, Euryarchaeota, Nanoarchaeota, Crenarchaeota, and Diapherotrites were the most commonly found dominant groups ranging from soils to waters, they had uneven abundances among habitat types. Distinct archaeal abundances across different habitats or environments were also observed in earlier studies (Auguet et al., 2010; Wen et al., 2017; Alves et al., 2018). These variances in taxa abundances were present at both phylum and ASV levels (**Figure 3** and **Supplementary Table S3**). Distinct archaeal taxa dominated different environments within each major phylum. For example, the proportion of Thaumarchaeota was significantly different across the three habitats ($P < 0.01$) (**Figure 3**), and clear phylogenetic shifts within the phylum were also observed (**Supplementary Table S3**). One member of Nitrosopumilaceae (ASV3197), probably a typical estuarine/brackish species, dominated the archaeal communities in the estuary surface water, but nearly unidentified in the other two habitats. Two other members of Nitrosopumilaceae (ASV54 and 64) were dominant in the freshwater biofilms, yet were nearly undetected in estuary waters. Furthermore, ASV66 (Nitrosotaleaceae) and ASV64 and 57 (Nitrosopumilaceae) dominated the soil environments (**Supplementary Table S3**). Thaumarchaeota are able to obtain ammonia from urea and cyanate (Baker et al., 2012; Palatinszky et al., 2015). Therefore, Thaumarchaeota have important links to climate change, as their activity has been linked to the production of the greenhouse gas nitrous oxide (N_2O) through the oxidation of ammonia to nitrite and reduction of nitrite to N_2O (Ostrom et al., 2000; Santoro et al., 2011). Thaumarchaeota are among the most abundant Archaea on the planet (Baker et al., 2020), including extreme environments such as the Yellowstone Lake and its lake floor hydrothermal vents (Kan et al., 2011). This study also demonstrates that Thaumarchaeota are relatively more abundant in soils than in the freshwater biofilms and estuary waters.

Phylogenetic shifts within a specific phylum were clearly observed among habitats (**Supplementary Figure S2**), such as those dominant taxa within Thaumarchaeota and Euryarchaeota (**Supplementary Table S3** and **Supplementary Figures S2C,E**). Similar patterns were also found in global phylogeny and environmental distribution of ammonia-oxidizing archaea (Alves et al., 2018) and methanogenic archaea (Wen et al., 2017). Phylogenetic distribution of Crenarchaeota groups was mainly found in the soil environment and less in the freshwater and estuary water (**Supplementary Figure S2A**), while the most relatively abundant taxa in the Diapherotrites groups were retrieved from soil environment (**Supplementary Figure S2B**). Further, phylogenetic affiliation and shifts of Euryarchaeota, Nanoarchaeota and Thaumarchaeota were primarily identified in the soil and estuary water and less in freshwater biofilms (**Supplementary Figures S2C–E**). This implies that phylogenetically closely related archaeal organisms have adapted to very different habitats, and also reflects the broad environmental distribution of major archaeal groups such as Thaumarchaeota and Euryarchaeota (Auguet et al., 2010; Alves et al., 2018). Euryarchaeota contain the greatest number and diversity of cultured lineages (Baker et al., 2020). They are not just involved in methane production and anaerobic methane oxidation (Orphan et al., 2002), but also participate the anaerobic oxidation of other short-chain hydrocarbons (Wang et al., 2019), suggesting that these microbes have varied roles in biogeochemical cycles. Agricultural soils and freshwater biofilms also each had an indicator ASV from the Euryarchaeota that was capable of methane production. The phylogenetic parallels of specific taxa within each archaeal phylum that dominated different habitats provide opportunities to examine interesting evolutionary tracks between soils, freshwater biofilms, and estuary water lineages.

Distinct Diversity and Distribution Patterns of Archaeal Communities

Differences in archaeal diversities and Bacteria:Archaea ratios were clearly shown among soils, freshwater biofilms, and estuary waters. The error bars to be large because the communities were sampled across time, seasons, biofilm succession, and different farming practices. Despite these factors corresponding with high variability of an Archaea community within a habitat over time and space, it is remarkable that the Archaea communities differ significantly among environments based on one-way ANOVA and ANOSIM analysis. We would be more surprised if there were very small error bars which could indicate lack of community changes over time and space (this would be highly unusual) or that the environments were not sufficiently sampled. Low Bacteria:Archaea ratios and most of the archaeal diversities were associated with soil environments, and as expected, they were different from freshwater biofilms and estuary (**Figure 2**). In fact, soil is the most diverse environment on Earth and hosts high bacterial and archaeal abundance and diversity (Griffiths et al., 2016). Almost 25% of the Bacteria and Archaea on Earth live in soils (to 8 m of depth), encompassing roughly 3×10^{29} cells, and those that live in the sea surface layer are only

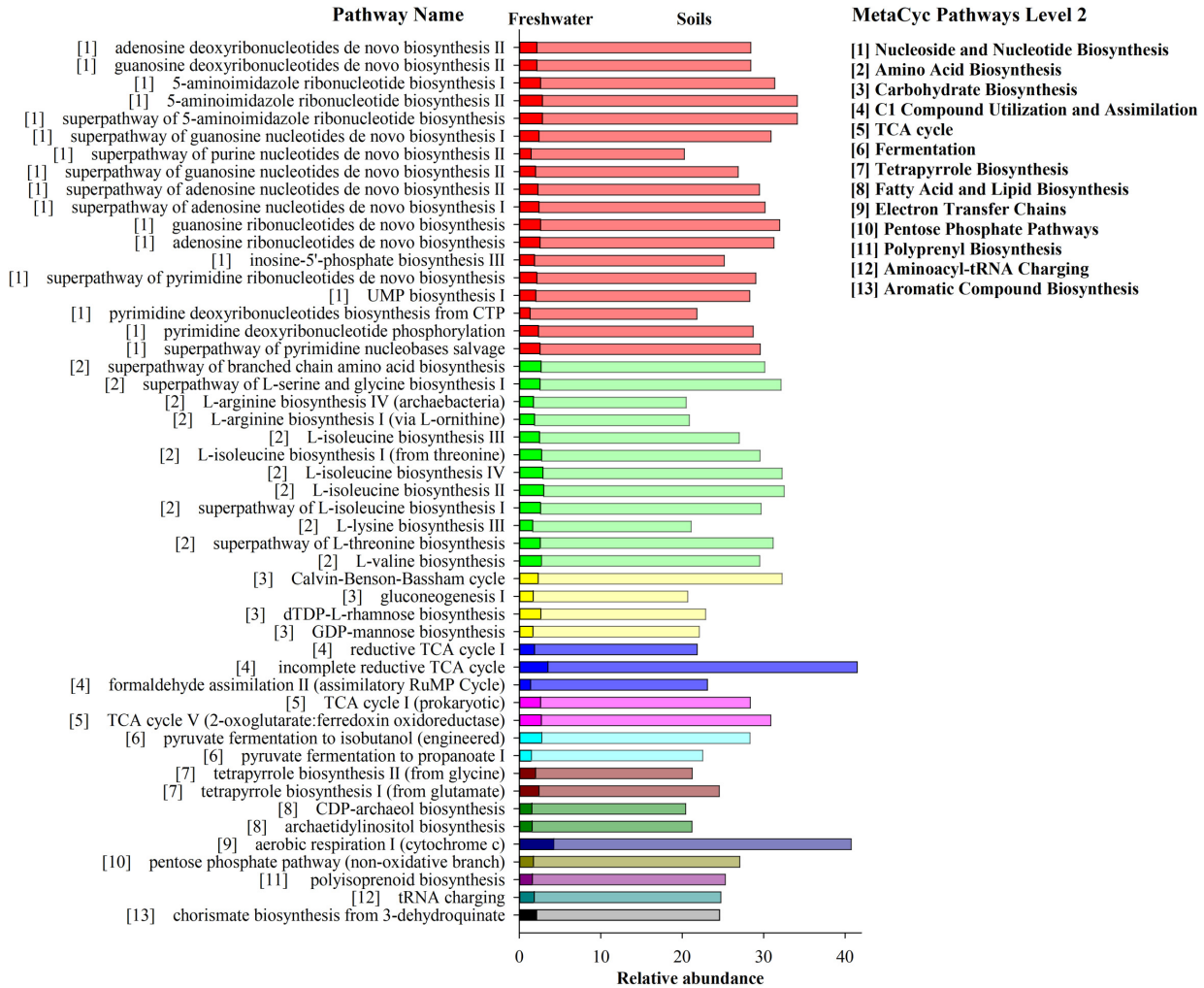

FIGURE 7 | Bar graph showing the relative abundance of archaeal genomic signatures predicted by PICRUSt2 within MetaCyc categories for agricultural soils FST (dark bar) and freshwater biofilms WCC (light bar). The estuarine samples (CB) were omitted due to unidentified functions. Pathways are displayed on the y axis, the number (1–13) in front of pathway names correspond to the number (1–13) of MetaCyc Pathways Level 2 shown on the right.

about 2×10^{23} (Flemming and Wuertz, 2019). Biofilms dominate all habitats on the surface of the Earth, except in the oceans, accounting for ∼80% of bacterial and archaeal cells. Biofilms drive the majority of biogeochemical processes and represent the main way of active bacterial and archaeal life (Flemming and Wuertz, 2019). Our study is consistent with previous studies and shows that archaeal communities are very abundant and diverse in soil compared to other environments on Earth (Baker et al., 2020), such as rivers and estuary waters.

Niche partitioning of archaeal communities clearly exists among soils, freshwater biofilms and estuarine surface waters (NMDS, **Figure 4**). Three different environments (soils, freshwater biofilms and estuary waters) differ with clear gradients in pH, oxygen and biologically relevant constituents such as CO_2, CH_4, and NH_4, which likely contributed to niche separation and differentiation, leading to the structuring and distribution of distinct archaeal physiological types in different environments (Biller et al., 2012; Reichenberger et al., 2015; Alves et al., 2018). Therefore, strong geochemical signatures (**Supplementary Table S1**) across three environments provide numerous niches capable of supporting phylogenetically and functionally diverse archaeal populations. For example, certain groups of Archaea that preferentially inhabit temperate estuarine surface waters such as ASV3197 (Nitrosopumilaceae) and two members of Marine Group II (ASV15 and ASV3190) (**Supplementary Table S3**). Similar patterns were also observed in earlier studies indicating that some microbial groups prefer to inhabit estuaries (Caporaso et al., 2011; Wen et al., 2017), even with strong geographical differentiation of archaeal communities across global estuaries (Liu et al., 2018). One important factor to recognize from this study is that our conclusions are intertwined with the season or seasons in which they were sampled. Seasonal categories can have a major influence on microbial community composition within an environment (Wang et al., 2020). For

example, Hullar et al. (2005) identified phototrophic-driven seasonal shifts in epilithic bacteria communities collected at the same White Clay Creek site (Pennsylvania Piedmont) used for this study. However, we have observed that for bacteria, season explains less of the variability in community composition than different environments within White Clay Creek (unpublished data). Because Bacteria and Archaea community samples cluster together in the NMDS, we surmise that for this study, seasonal differences within freshwater biofilms are unlikely to be more influential than those between freshwater biofilms and the other environments. While seasonal changes also occur in agricultural soils (e.g., Bossio et al., 1998 although they identified soil type as more influential than time), our DNA-based approach may have integrated soil archaeal composition across seasons. Carini et al. (2020) have shown that removal of relic DNA from a soil enhances detection of prokaryote community temporal patterns, suggesting that some composition of the community is retained from season-to-season. Therefore, in the future, there is a great need for detailed investigations and comparisons of how the archaeal community structure responds to seasonal as well as spatial variations across different environments. However, according to the nature of habitat and environment, it is necessary to carefully consider measurement of environmental parameters and the sampling strategies covering both time and space.

Our study also reveals that the relative abundances of archaeal communities from different habitats have extremely uneven phylogenetic diversities, with few clades overwhelmingly dominating overall archaeal diversity in a specific environment. For example, most members of the Thaumarchaeota (21 out of 22) were affiliated with the class Nitrososphaeria, including those taxa abundant across the three habitats, such as ASV57, ASV64, ASV66, and ASV3197 (**Supplementary Table S3**). Ammonia-oxidizing Archaea (AOA) comprise a diverse group of organisms formally defined as class Nitrososphaeria of the phylum Thaumarchaeota (Rosenberg et al., 2014; Baker et al., 2020). Plenty of amoA-based studies collectively have shown that AOA diversity and abundance in nature depend on multiple factors and are strongly partitioned by local environments, and that AOA plays a major role in nitrification, the conversion of ammonia to nitrate via nitrite (Francis et al., 2005; Biller et al., 2012; Stahl and de la Torre, 2012; Alves et al., 2018). Different ecosystems tend to harbor distinct AOA groups and niche adaptation directly or indirectly contributes to the selection of specific archaeal groups. Our study shows that considerable habitat specificity and Archaeal diversification reflects diverse niche adaptation. This potentially implies that AOA are ubiquitous and abundant from soils to freshwaters to estuaries but have uneven distribution patterns.

Differential Archaeal Networks, Indicator Species, and Functions

Co-occurrence network analysis indicated that only soil Archaea formed complex networks (29 taxa with 68 correlations) and key species (ASVs) were identified mainly from Methanomassiliicoccales, Nitrososphaerales, Nitrosopumilales (**Figure 5**). In addition, these key species were also identified as indicator species in the soil environment, suggesting unique adaptation to, or preference for, soil environments by taxa within these groups. Compared to soils, less archaeal diversity and abundance occurred in freshwater biofilms and surface water in estuaries. We speculate these archaea, either attached or free-living ones, may be more dependent on interaction with other biomes such as prokaryotes (i.e., Bacteria) or eukaryotes (e.g., microalgae). Though these biotic interactions have not been well documented and characterized, previous observations have shown that occurrence and abundances of archaeal groups coincide with diatoms, cyanobacteria, and viruses (Lima-Mendez et al., 2015; Needham and Fuhrman, 2016; Xie et al., 2018). Moreover, the potential of Archaea to shape their surroundings by a profound interaction with their biotic and abiotic environment has been researched (Valentine, 2007; Morris et al., 2013; Comolli and Banfield, 2014; Wegener et al., 2015). Moissl-Eichinger et al. (2018) summarized the basic principles of archaeal interactions, which are mainly based on the following driving factors: energetic pressure deriving from the environment, the capability for exchange of metabolites and/or electrons, genomic and structural adaptation capacity (by symbiont and host), and detoxification or facilitated horizontal gene transfer.

To investigate the potential biogeochemical implications of archaeal ASVs switching across environments, the functional capacity of the soil-, freshwater biofilm-, and estuary water-associated archaeal communities was analyzed using existing genomes within PICRUSt2 (Langille et al., 2013; Douglas et al., 2019). The predictions are sparse or lacking when PICRUSt 2 is applied using phylogenetic marker gene signatures from lesser known environments, such as the estuary Chesapeake Bay. This might be the main reason why archaeal functions were unidentified in those archaeal communities in the Bay. Although this technique is limited by the ability of 16S rRNA gene sequences to resolve ecologically important units and the phylogenetic breadth and depth of archaeal genomes, metagenomics prediction may nevertheless offer insight into the extent of both functional redundancy and differences in biogeochemical potential (but not rates) across natural environments.

Predicted pathways from the environments investigated here were most abundant in soils indicating that archaeal communities had much higher metabolic activities in the soil environments compared to the freshwater biofilms and estuary water. These predicted functional profiles are consistent with the proportion of major archaeal groups across the three environments, such as Thaumarchaeota and Nanoarchaeota which are more relatively abundant in soils than the other two environments. Earlier studies also showed distinct metabolic features of microbial communities across different environments, including water, mineral fractions, and microbial biofilms (Mesa et al., 2017). Archaeal communities in soils contain stronger abilities to perform Biosynthesis (e.g., Nucleoside and Nucleotide, Amino Acid, carbohydrate, Fatty acid, and Lipid), Generation of Precursor Metabolites and Energy (e.g., TCA cycle, Fermentation, and Electron

Transfer Chains), and Degradation/Utilization/Assimilation (C1 Compound Utilization and Assimilation). The most abundant pathway in the soil samples, "incomplete reductive TCA cycle," was performed exclusively by and widely observed in the archaeal groups of Methanobacteria, Methanococci, Methanomicrobia, including the following taxa: *Methanococcus maripaludis*, *Methanospirillum hungatei*, *Methanothermobacter thermautotrophicus*. Similarly, the pathways of "inosine-5′-phosphate biosynthesis III," "archaetidylinositol biosynthesis," "CDP-archaeol biosynthesis," and "phosphopantothenate biosynthesis III (archaebacteria)" were performed exclusively by Archaea, including Methanocaldococcus, Methanothermobacter, *Archaeoglobus fulgidus*, *Methanosarcina acetivorans*, *Thermococcus kodakarensis*. In addition, "aerobic respiration I (cytochrome c)," "incomplete reductive TCA cycle," "L-isoleucine biosynthesis II," and "L-isoleucine biosynthesis" were the most abundant metabolism pathways in freshwater biofilms, and they were mostly found in the archaeal groups of Methanobacteria, Methanococci, Methanomicrobia, and Thermoprotei. Microbiome functions were found to be responsible for interactions via nutrient exchange, but also for coping with environmental stress, to which Archaea are in general evolutionarily adapted (Valentine, 2007; Moissl-Eichinger et al., 2018). Overall, habitat differentiation from soil to freshwater to estuary could alter greatly the biogeochemical potential of archaeal communities with apparent replacement by distinct archaeal groups under different environments.

In general, archaeal networks, indicator species and their functions under each habitat further confirmed that environmental selection/adaptation has a great effect in shaping archaeal communities (Offre et al., 2013; Baker et al., 2020) and provides further evidence and knowledge on the biodiversity and complexity of archaeal communities across environmental ecosystems. Future efforts could focus on quantitative assessments of targeted archaeal groups (e.g., ammonia-oxidizing Archaea or methanogenic Archaea) and how they respond to their ambient environmental gradients in order to more precisely estimate their abundance, population dynamics and functional roles across environments.

CONCLUSION

We analyzed and compared the structure, distribution, diversity, network, indicator species, and potential functions of archaeal communities among agriculture soils, freshwater biofilms, and estuarine surface waters with 16S amplicon high-throughput sequencing. Our study highlights the heterogeneous proportions of archaeal phyla and taxa from soils to estuary, and reflects the significant influence of environment dissimilarities on archaeal abundance. Differential distribution patterns and diversity of archaeal communities in specific environments suggest potential niche-specific features of Archaea from soil, freshwater biofilms, and estuaries. Archaeal communities have complex networks, high metabolic activities and different indicator species in soil environments compared to freshwater biofilms and estuarine waters. The pressure of niche adaptation can contribute greatly to the variation of Archaea across the three habitats. This study shows the strong differentiation of archaeal communities from distinct ecosystems and provides guidance for the discovery of global diversity, distribution pattern, and ecological significance of Archaea.

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

JK, FC, and CZ designed this research project. HW, RB, and LZ analyzed the data. RB, MP, EO, AM, and JK performed the experiments and collected the samples. HW, RB, MP, EO, AM, FC, CZ, and JK contributed to manuscript draft, tables, and figures, and approved the submitted version. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the NSF LTREB program (DEB-1557063), a Visiting Professor Program at Southern University of Science and Technology, and Endowment Fund from Stroud Water Research Center. This work was also supported in part by the William Penn Foundation under Grant Award No. 188-17. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the William Penn Foundation. CZ was supported by the National Natural Science Foundation of China (91851210 and 41530105), the Shenzhen Key Laboratory of Marine Archaea Geo-Omics, Southern University of Science and Technology (ZDSYS201802081843490), and the Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou) (K19313901).

ACKNOWLEDGMENTS

We thank the research technicians, interns, and volunteers at Rodale Institute, Stroud Water Research Center, and University of Maryland for their assistance with the soil and biofilm sample collections.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.576661/full#supplementary-material

REFERENCES

Alves, R. J. E., Minh, B. Q., Urich, T., Von Haeseler, A., and Schleper, C. (2018). Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat. Commun.* 9, 1–17.

Apprill, A., Mcnally, S., Parsons, R., and Weber, L. (2015). Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat. Microb. Ecol.* 75, 129–137. doi: 10.3354/ame01753

Auguet, J.-C., Barberan, A., and Casamayor, E. O. (2010). Global ecological patterns in uncultured Archaea. *ISME J.* 4, 182–190. doi: 10.1038/ismej.2009.109

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 1–14.

Baker, B. J., Lesniewski, R. A., and Dick, G. J. (2012). Genome-enabled transcriptomics reveals archaeal populations that drive nitrification in a deep-sea hydrothermal plume. *ISME J.* 6, 2269–2279. doi: 10.1038/ismej.2012.64

Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks," in *Third International AAAI Conference on Weblogs and Social Media*, Palo Alto, CA: AAAI.

Bates, S. T., Berg-Lyons, D., Caporaso, J. G., Walters, W. A., Knight, R., and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J.* 5, 908–917. doi: 10.1038/ismej.2010.171

Biller, S. J., Mosier, A. C., Wells, G. F., and Francis, C. A. (2012). Global biodiversity of aquatic ammonia-oxidizing archaea is partitioned by habitat. *Front. Microbiol.* 3:252. doi: 10.3389/fmicb.2012.00252

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

Bomberg, M., Montonen, L., Münster, U., and Jurgens, G. (2008). Diversity and function of archaea in freshwater habitats. *Curr. Trends Microbiol.* 4, 61–89.

Bossio, D. A., Scow, K. M., Gunapala, N., and Graham, K. J. (1998). Determinants of soil microbial communities: effects of agricultural management, season, and soil type on phospholipid fatty acid profiles. *Microb. Ecol.* 36, 1–12. doi: 10.1007/s002489900087

Cáceres, M. D., and Legendre, P. (2009). Associations between species and groups of sites: indices and statistical inference. *Ecology* 90, 3566–3574. doi: 10.1890/08-1823.1

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

Cao, H., Auguet, J.-C., and Gu, J.-D. (2013). Global ecological pattern of ammonia-oxidizing archaea. *PLoS One* 8:e0052853. doi: 10.1371/journal.pone.0052853

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

Carini, P., Delgado-Baquerizo, M., Hinckley, E.-L. S., Holland-Moritz, H., Brewer, T. E., Rue, G., et al. (2020). Effects of spatial variability and relic DNA removal on the detection of temporal dynamics in soil microbial communities. *mBio* 11:e02776-9. doi: 10.1128/mBio.02776-9

Chen, S., Wang, P., Liu, H., Xie, W., Wan, X. S., Kao, S.-J., et al. (2020). Population dynamics of methanogens and methanotrophs along the salinity gradient in Pearl River Estuary: implications for methane metabolism. *Appl Microbiol Biotechnol.* 104, 1331–1346. doi: 10.1007/s00253-019-10221-6

Comolli, L. R., and Banfield, J. F. (2014). Inter-species interconnections in acid mine drainage microbial communities. *Front. Microbiol.* 5:367. doi: 10.3389/fmicb.2014.00367

DeLong, E. F. (1992). Archaea in coastal marine environments. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5685–5689.

Douglas, G. M., Maffei, V. J., Zaneveld, J., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2019). PICRUSt2: an improved and extensible approach for metagenome inference. *bioRxiv* [Preprint]. doi: 10.1101/672295

Dufrêne, M., and Legendre, P. (1997). Species assemblages and indicator species: the need for a flexible asymmetrical approach. *Ecol. Monogr.* 67, 345–366. doi: 10.2307/2963459

Ferry, J. G. (2010). How to make a living by exhaling methane. *Annu. Rev. Microbiol.* 64, 453–473. doi: 10.1146/annurev.micro.112408.134051

Flemming, H.-C., and Wuertz, S. (2019). Bacteria and archaea on Earth and their abundance in biofilms. *Nat. Rev. Microbiol.* 17, 247–260. doi: 10.1038/s41579-019-0158-9

Francis, C. A., Roberts, K. J., Beman, J. M., Santoro, A. E., and Oakley, B. B. (2005). Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14683–14688. doi: 10.1073/pnas.0506625102

Gilbert, J. A., Jansson, J. K., and Knight, R. (2014). The earth microbiome project: successes and aspirations. *BMC Biol.* 12:69. doi: 10.1186/s12915-014-0069-1

Griffiths, R. I., Thomson, B. C., Plassart, P., Gweon, H. S., Stone, D., Creamer, R. E., et al. (2016). Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. *Appl. Soil Ecol.* 97, 61–68. doi: 10.1016/j.apsoil.2015.06.018

Herrmann, M., Saunders, A. M., and Schramm, A. (2008). Archaea dominate the ammonia-oxidizing community in the rhizosphere of the freshwater macrophyte *Littorella uniflora*. *Appl. Environ. Microbiol.* 74, 3279–3283. doi: 10.1128/aem.02802-07

Hullar, M. A. J., Kaplan, L. A., and Stahl, D. A. (2005). Recurring seasonal dynamics of microbial communities in stream habitats. *Appl. Environ. Microbiol.* 72, 713–722. doi: 10.1128/aem.72.1.713-722.2006

Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., and Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335, 587–590. doi: 10.1126/science.1212665

Kan, J. (2006). *Bacterioplankton in the Chesapeake Bay: Genetic Diversity, Population Dynamics and Community Proteomics*. Dissertation, University of Maryland, College Park, MD.

Kan, J., Clingenpeel, S., Macur, R. E., Inskeep, W. P., Lovalvo, D., Varley, J., et al. (2011). Archaea in yellowstone lake. *ISME J.* 5, 1784–1795. doi: 10.1038/ismej.2011.56

Kan, J., Wang, K., and Chen, F. (2006). Temporal variation and detection limit of an estuarine bacterioplankton community analyzed by denaturing gradient gel electrophoresis (DGGE). *Aquat. Microb. Ecol.* 42, 7–18. doi: 10.3354/ame042007

Karimi, B., Terrat, S., Dequiedt, S., Saby, N. P. A., Horrigue, W., Lelièvre, M., et al. (2018). Biogeography of soil bacteria and archaea across France. *Sci. Adv.* 4, eaat1808. doi: 10.1126/sciadv.aat1808

Karner, M. B., Delong, E. F., and Karl, D. M. (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* 409, 507–510. doi: 10.1038/35054051

Klenk, H.-P., Clayton, R. A., Tomb, J.-F., White, O., Nelson, K. E., Ketchum, K. A., et al. (1998). Erratum: the complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 394:101.

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

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

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G. W., et al. (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* 442, 806–809. doi: 10.1038/nature04983

Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., et al. (2015). Determinants of community structure in the global plankton interactome. *Science* 348:1262073.

Lipp, J. S., Morono, Y., Inagaki, F., and Hinrichs, K.-U. (2008). Significant contribution of Archaea to extant biomass in marine subsurface sediments. *Nature* 454, 991–994. doi: 10.1038/nature07174

Liu, X., Pan, J., Liu, Y., Li, M., and Gu, J.-D. (2018). Diversity and distribution of Archaea in global estuarine ecosystems. *Sci. Total Environ.* 637-638, 349–358. doi: 10.1016/j.scitotenv.2018.05.016

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019

Massana, R., Delong, E. F., and Pedrós-Alió, C. (2000). A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. *Appl. Environ. Microbiol.* 66, 1777–1787. doi: 10.1128/aem.66.5.1777-1787.2000

Mesa, V., Gallego, J. L. R., González-Gil, R., Lauga, B., Sánchez, J., Méndez-García, C., et al. (2017). Bacterial, archaeal, and eukaryotic diversity across distinct microhabitats in an acid mine drainage. *Front. Microbiol.* 8:1756. doi: 10.3389/fmicb.2017.01756

Moissl-Eichinger, C., Pausan, M., Taffner, J., Berg, G., Bang, C., and Schmitz, R. A. (2018). Archaea are interactive components of complex microbiomes. *Trends Microbiol.* 26, 70–85. doi: 10.1016/j.tim.2017.07.004

Morris, B. E. L., Henneberger, R., Huber, H., and Moissl-Eichinger, C. (2013). Microbial syntrophy: interaction for the common good. *FEMS Microbiol. Rev.* 37, 384–406. doi: 10.1111/1574-6976.12019

Needham, D. M., and Fuhrman, J. A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat. Microbiol.* 1, 1–7. doi: 10.1002/9781118960608.gbm00535

Ochsenreiter, T., Selezi, D., Quaiser, A., Bonch-Osmolovskaya, L., and Schleper, C. (2003). Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ. Microbiol.* 5, 787–797. doi: 10.1046/j.1462-2920.2003.00476.x

Offre, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., et al. (2019). *vegan: Community Ecology Package. R package version 2.5-6*. Avaliable at: https://CRAN.R-project.org/package=vegan. (accessed September 01, 2019).

Orphan, V. J., House, C. H., Hinrichs, K.-U., Mckeegan, K. D., and Delong, E. F. (2002). Multiple archaeal groups mediate methane oxidation in anoxic cold seep sediments. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7663–7668. doi: 10.1073/pnas.072210299

Orsi, W. D., Smith, J. M., Liu, S., Liu, Z., Sakamoto, C. M., Wilken, S., et al. (2016). Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *ISME J.* 10, 2158–2173. doi: 10.1038/ismej.2016.20

Ostrom, N. E., Russ, M. E., Popp, B., Rust, T. M., and Karl, D. M. (2000). Mechanisms of nitrous oxide production in the subtropical North Pacific based on determinations of the isotopic abundances of nitrous oxide and di-oxygen. *Chemos. Glob. Change Sci.* 2, 281–290. doi: 10.1016/s1465-9972(00)00031-3

Palatinszky, M., Herbold, C., Jehmlich, N., Pogoda, M., Han, P., Von Bergen, M., et al. (2015). Cyanate as an energy source for nitrifiers. *Nature* 524, 105–108. doi: 10.1038/nature14856

Parada, A. E., Needham, D. M., and Fuhrman, J. A. (2016). Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ. Microbiol.* 18, 1403–1414. doi: 10.1111/1462-2920.13023

Pesaro, M., and Widmer, F. (2002). Identification of novel Crenarchaeota and Euryarchaeota clusters associated with different depth layers of a forest soil. *FEMS Microbiol. Ecol.* 42, 89–98. doi: 10.1111/j.1574-6941.2002.tb00998.x

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.

R Core Team (2020). *R: A Language and Environment for Statistical Computing (Computer software)*. Vienna: R Core Team.

Reichenberger, E. R., Rosen, G., Hershberg, U., and Hershberg, R. (2015). Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol. Evol.* 7, 1380–1389. doi: 10.1093/gbe/evv063

Rosenberg, E., Delong, E. F., Lory, S., Stackebrandt, E., and Thompson, F. (2014). *The prokaryotes: Other Major Lineages of Bacteria and the Archaea*. Berlin: Springer.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Santoro, A. E., Buchwald, C., Mcilvin, M. R., and Casciotti, K. L. (2011). Isotopic signature of N2O produced by marine ammonia-oxidizing archaea. *Science* 333, 1282–1285. doi: 10.1126/science.1208239

Schleper, C., Jurgens, G., and Jonuscheit, M. (2005). Genomic studies of uncultivated archaea. *Nat. Rev. Microbiol.* 3, 479–488. doi: 10.1038/nrmicro1159

Schleper, C., and Nicol, G. W. (2010). Ammonia-oxidising archaea - physiology, ecology and evolution. *Adv. Microb. Physiol.* 57, 1–41. doi: 10.1016/b978-0-12-381045-8.00001-1

Shen, L.-D., Ouyang, L., Zhu, Y., and Trimmer, M. (2019). Active pathways of anaerobic methane oxidation across contrasting riverbeds. *ISME J.* 13, 752–766.

Stahl, D. A., and de la Torre, J. R. (2012). Physiology and diversity of ammonia-oxidizing archaea. *Annu. Rev. Microbiol.* 66, 83–101.

Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., et al. (2008). Cell proliferation at 122 C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10949–10954.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673

Timonen, S., and Bomberg, M. (2009). Archaea in dry soil environments. *Phytochem. Rev.* 8, 505–518.

Valentine, D. L. (2007). Adaptations to energy stress dictate the ecology and evolution of the Archaea. *Nat. Rev. Microbiol.* 5, 316–323.

Wang, Y., Wegener, G., Hou, J., Wang, F., and Xiao, X. (2019). Expanding anaerobic alkane metabolism in the domain of Archaea. *Nat. Microbiol.* 4, 595–602.

Wang, H., Zhang, C., Chen, F., and Kan, J. (2020). Spatial and temporal variations of bacterioplankton in the chesapeake bay: a re-examination with high-throughput sequencing analysis. *Limnol. Oceanogr.* doi: 10.1002/lno.11572

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., and Boetius, A. (2015). Intercellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* 526, 587–590.

Wen, X., Yang, S., Horn, F., Winkel, M., Wagner, D., and Liebner, S. (2017). Global biogeographic analysis of methanogenic archaea identifies community-shaping environmental factors of natural environments. *Front. Microbiol.* 8:1339. doi: 10.3389/fmicb.2017.01339

Williams, R. J., Howe, A., and Hofmockel, K. S. (2014). Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* 5:358. doi: 10.3389/fmicb.2014.00358

Xie, W., Luo, H., Murugapiran, S. K., Dodsworth, J. A., Chen, S., Sun, Y., et al. (2018). Localized high abundance of Marine Group II archaea in the subtropical Pearl River Estuary: implications for their niche adaptation. *Environ. Microbiol.* 20, 734–754.

Yao, H., Campbell, C. D., Chapman, S. J., Freitag, T. E., Nicol, G. W., and Singh, B. K. (2013). Multi-factorial drivers of ammonia oxidizer communities: evidence from a national soil survey. *Environ. Microbiol.* 15, 2545–2556.

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 17 November 2020
doi: 10.3389/fmicb.2020.576520

High Representation of Archaea Across All Depths in Oxic and Low-pH Sediment Layers Underlying an Acidic Stream

Marco A. Distaso[1,2†], Rafael Bargiela[1†], Francesca L. Brailsford[1,2,3], Gwion B. Williams[1,2], Samuel Wright[1,2], Evgenii A. Lunev[4], Stepan V. Toshchakov[5], Michail M. Yakimov[6], David L. Jones[1,2,3], Peter N. Golyshin[1,2] and Olga V. Golyshina[1,2]*

[1] School of Natural Sciences, Bangor University, Bangor, United Kingdom, [2] Centre for Environmental Biotechnology, Bangor University, Bangor, United Kingdom, [3] School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, [4] Institute of Living Systems, Immanuel Kant Baltic Federal University, Kaliningrad, Russia, [5] National Research Centre "Kurchatov Institute", Moscow, Russia, [6] Institute for Biological Resources and Marine Biotechnology, CNR, Messina, Italy

OPEN ACCESS

Edited by:
Ricardo Amils, Autonomous University of Madrid, Spain

Reviewed by:
Ruben Michael Ceballos, University of Arkansas, United States
Felipe Gómez, Centro de Astrobiología (CSIC-INTA), Spain

***Correspondence:**
Olga V. Golyshina
o.golyshina@bangor.ac.uk

† These authors have contributed equally to this work

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

Received: *29 June 2020*
Accepted: *23 October 2020*
Published: *17 November 2020*

Citation:
Distaso MA, Bargiela R, Brailsford FL, Williams GB, Wright S, Lunev EA, Toshchakov SV, Yakimov MM, Jones DL, Golyshin PN and Golyshina OV (2020) High Representation of Archaea Across All Depths in Oxic and Low-pH Sediment Layers Underlying an Acidic Stream. Front. Microbiol. 11:576520. doi: 10.3389/fmicb.2020.576520

Parys Mountain or Mynydd Parys (Isle of Anglesey, United Kingdom) is a mine-impacted environment, which accommodates a variety of acidophilic organisms. Our previous research of water and sediments from one of the surface acidic streams showed a high proportion of archaea in the total microbial community. To understand the spatial distribution of archaea, we sampled cores (0–20 cm) of sediment and conducted chemical analyses and taxonomic profiling of microbiomes using 16S rRNA gene amplicon sequencing in different core layers. The taxonomic affiliation of sequencing reads indicated that archaea represented between 6.2 and 54% of the microbial community at all sediment depths. Majority of archaea were associated with the order Thermoplasmatales, with the most abundant group of sequences being clustered closely with the phylotype B_DKE, followed by "E-plasma," "A-plasma," other yet uncultured Thermoplasmatales with *Ferroplasma* and *Cuniculiplasma* spp. represented in minor proportions. Thermoplasmatales were found at all depths and in the whole range of chemical conditions with their abundance correlating with sediment Fe, As, Cr, and Mn contents. The bacterial microbiome component was largely composed in all layers of sediment by members of the phyla Proteobacteria, Actinobacteria, Nitrospirae, Firmicutes, uncultured Chloroflexi (AD3 group), and Acidobacteria. This study has revealed a high abundance of Thermoplasmatales in acid mine drainage-affected sediment layers and pointed at these organisms being the main contributors to carbon, and probably to iron and sulfur cycles in this ecosystem.

Keywords: acidophilic archaea and bacteria, Thermoplasmatales, "*Candidatus* Micrarchaeota", unclassified Euryarchaeota/Terrestrial Miscellaneous Euryarchaeotal Group, acid mine drainage systems, mine-impacted environments, sediment microbiome

INTRODUCTION

Parys Mountain (Parys Mt) or Mynydd Parys (Isle of Anglesey, United Kingdom) is an abandoned copper mine which contains abundant sulfidic deposits in the form of pyrite, chalcopyrite, sphalerite, and galena minerals. As with many other low pH environments associated with metal mining activity, the site is characterized by the presence of acidic streams or acid mine drainage (AMD) waters, which result from the oxidative dissolution of sulfidic minerals (Johnson, 2012). Like other AMD systems, Parys Mt streams contain large concentrations of dissolved metals and metalloids which constantly flow into the Irish Sea resulting in marine pollution (Johnson, 2012). This site attracts continuous scientific interest, as reflected in the large number of studies and the identification of many new species of acidophilic bacteria and archaea (Johnson et al., 2014; Jones and Johnson, 2015; Golyshina et al., 2016b).

Our earlier study on microbial assemblages in AMD water and sediments taken from the surface of one of acidic streams of Parys Mt revealed that archaea dominated the microbial community (Korzhenkov et al., 2019). Archaea affiliated with Euryarchaeota constituted the major group (67%) of the total shotgun reads in the community. One particular group of sequences associated with still uncultured archaea of the order Thermoplasmatales (similar to "E-plasma" metagenomic variant) was shown to represent 58% of all metagenomic reads. In the upper sediment layer, bacterial representatives (33%) were mostly related with Proteobacteria. Other bacterial reads present in low amounts (2–6%) were largely affiliated with Actinobacteria, Nitrospirae, Bacteroidetes, Acidobacteria, and Firmicutes (Korzhenkov et al., 2019).

However, in the lotic community, Proteobacteria, Nitrospirae, Acidobacteria, and Actinobacteria did collectively outnumber archaea (Korzhenkov et al., 2019).

The populations of microorganisms inhabiting sediments in AMD-affected areas have been the subject of numerous studies (Kock and Schippers, 2008; Sánchez-Andrea et al., 2011, 2012; Sun et al., 2015; Zhang et al., 2019 and others). These works established that bacteria were highly abundant in AMD sediments and thus assumed they were mainly responsible for biogeochemical cycling in these ecosystems. For example, only low numbers of archaea were reported in sediments of mine tailing dumps in Botswana, Germany, and Sweden and only in oxidized zones (Kock and Schippers, 2008). Although archaea of the order Thermoplasmatales are well-known inhabitants of AMD environments, including sediments, these organisms were found to be present in very low abundance and thus assumed to be unimportant (Kock and Schippers, 2008; Sánchez-Andrea et al., 2011, 2012; Sun et al., 2015; Zhang et al., 2019). Frequently however, the detailed information about the archaeal component is missing, or archaea were excluded from the analysis, leading to a potential underestimation of the ecological role of archaea in AMD ecosystems (Wakelin et al., 2012; Lukhele et al., 2019). To understand the patterns of archaeal distribution in sediments of an acidic stream at Parys Mt and to assess their potential role in elemental cycling, we collected shallow sediment cores (0–20 cm depth) from the AMD stream. We used a combination of chemical analysis and SSU rRNA gene amplicon sequencing to resolve, layer-by-layer, microbial composition changes with depth and across the chemical gradient in order to understand whether particular geochemical factors were associated with archaeal abundance and to assess their functional role *in situ*.

MATERIALS AND METHODS

Sampling was conducted in the acidic stream located at Parys Mt (GPS location 53.38708° −4.34968°) as described previously (**Supplementary Figure 1**; Golyshina et al., 2016a,b; Korzhenkov et al., 2019). Intact sediment cores were taken in September 2018 at three random locations each near another (within 15 cm distance) using polycarbonate tubes (50 cm-long with inner diameter of 4 cm). The tubes were gently pressed by hand into the sediment, then plugged with a butyl rubber stopper at the top. The intact cores were then carefully removed and the base of the tubes plugged with another butyl stopper and subsequently transported back to the laboratory for analysis. Upon arrival (ca. 40 min after sampling), the cores were sliced into 2–3 cm-thick disks and transferred into sterile polypropylene 50 ml Falcon tubes for consequent chemical and microbiological analyses. pH and Eh potential in the sediment surface layers were measured in the field using a SevenGo multimeter (Mettler-Toledo, Leicester, United Kingdom) and then again in the cores on return to the laboratory.

DNA Extraction and 16S rRNA Gene Amplicon Sequencing

DNA was extracted from 0.25 g of soil sample from each layer of three cores using the DNeasy PowerLyzer PowerSoil kit (QIAGEN) according to manufacturer's instructions. Two independent DNA extractions were carried out for each sample. Quality and concentration of extracted DNA were assessed by gel electrophoresis and by Qubit 4.0 Fluorometer dsDNA BR Assay Kit (Life Technologies, United States).

Libraries of 16S rRNA gene amplicons were prepared by single PCR with double-indexed fusion primers as described previously (Fadrosh et al., 2014). Hypervariable V4 16S rRNA gene fragment was amplified using modified forward primer F515 (5′-GTGBCAGCMGCCGCGGTAA-3′) and reverse R806 prokaryotic primer (5′-GGACTACHVGGGTWTCTAAT-3′), which amplify an approximately 290 bp region. Primers were designed to contain: the Illumina adapters and sequencing primers, a 12 bp barcode sequence, a heterogeneity spacer to mitigate the low sequence diversity amplicon issue, and 16S rRNA gene universal primers (Fadrosh et al., 2014). PCRs were performed using MyTaq Red DNA Polymerase (Bioline). All reactions were run with no-template negative controls. Thermocycling conditions were: initial denaturation at 95°C for 2 min, followed by 30 cycles at 95°C for 45 s, 50°C for 60 s, and 72°C for 30 s with a final elongation at 72°C for 5 min. Amplicons were visualized in a 1.5% tris-acetate agarose gels using a

GelDoc System (Bio-Rad, CA, United States). DNA bands of approximately 440 bp were gel-purified using QIAEX II Gel Extraction Kit (QIAGEN).

The purified amplicons were then quantified using Qubit 4.0 Fluorometer (Life Technologies, Carlsbad, CA, United States), pooled in equimolar amounts and the final pool was run on Illumina MiSeq platform (Illumina, San Diego, CA, United States) using 500-cycle v2 chemistry (2 × 250 bp paired-end reads) at the Centre for Environmental Biotechnology, Bangor, United Kingdom.

Bioinformatic Analysis

Raw sequencing reads were processed according to previously described protocols (Fadrosh et al., 2014; Korzhenkov et al., 2019). Briefly, the data was pre-processed in order to extract the barcodes from sequences, and then cleaned of primer sequences using tagcleaner. The barcodes and the sequences were re-matched again using in-house Python scripts. The resulting filtered reads were analyzed using QIIME v1.3.1. First, the libraries were demultiplexed based on the different barcodes. Then, the sequences were classified on operational taxonomic units (OTUs) combining *de novo* and reference-based methods (open-reference OTU generation algorithm) using the SILVA version 132 reference database.

In the case of OTUs assigned to order Thermoplasmatales, a further taxonomic assignation analysis was performed using a local Blast (Camacho et al., 2008) database based on a selection of 42 reference sequences, running a final individual blast against *nr* database for those OTU sequences with <97% of identity in their best hit against the local database.

Statistical Analysis

All statistical analysis and figures were generated using the R programming language (R Development Core Team, 2008). Principal components analysis (PCA) was undertaken using the *prcomp* function form package *stats*, included on basic R core. In the case of the non-metric multidimensional scaling (NMDS) analysis, we used the *vegan* package (Oksanen et al., 2019). For canonical correlation analysis (CCorA) internal R scripts were developed, using basic R functions.

Phylogenetic Analysis of Archaea

For phylogenetic tree construction, we selected those OTU sequences assigned to Archaea with more than 100 reads along the three cores and also 34 reference sequences belonging to different groups. Multiple alignment of sequences was developed using *Mafft* (Katoh and Standley, 2013). *UGENE* (v 1.9.8) was used for the trimming of the extremes and trimAL (Capella-Gutierrez et al., 2009) for internal trimming of the alignment, removing columns with gaps on more than the 20% of the sequences or with similarity scores lower than 0.001, producing a final multiple alignment of 293 positions. Phylogenetic tree was calculated by maximum likelihood with bootstrapping of 1,000 replicates.

CHEMICAL ANALYSIS

Background Chemical Analysis

Cores were divided by layers and subsamples removed for physicochemical analysis. Moisture content was determined for the <2 mm fraction by drying at 105°C for 24 h. The organic matter content of the sediment was measured using the loss-on-ignition method, in a muffle furnace (450°C, 16 h; Ball, 1964). Sediment C and N content was determined after oven-drying (105°C, 24 h) using a TruSpec CN analyzer (Leco Corp., St Joseph, MI, United States). Bulk elemental analysis on the dried, sieved fraction (40°C, <125 μm) was undertaken by total reflection X-ray fluorescence (TXRF) using a Bruker S2 Picofox TXRF spectrometer (Bruker Inc., MA, United States). Ion chromatography (IC) was used to determine anion concentrations (F^-, Cl^-, NO_3^-, PO_4^{3-}) in 1:10 (w/v) sediment: E-pure water (18 MΩ resistance) extracts using a 930 Compact IC Flex (Metrohm, Herisau, Switzerland).

Analysis of Black Layers (Oily Deposits) Within the Sediment

Two samples of sediment layers with an oily appearance and hydrocarbon-like odor were selected for further analysis. Samples were weighed out in aliquots of around 100 mg for extraction. The method of extraction was modified from the EPA 3550C method for extraction of non-volatile and semi-volatile organic compounds from solids such as soils, sludges, and wastes by ultrasonic extraction (USEPA, 2007). Briefly, an equal amount of anhydrous sodium sulfate was mixed with the sample to form a free-flowing powder. The sample was then spiked with an internal standard (10 mg pristine) and extracted using 0.5 ml of a 1:1 (v/v) acetone:chloroform solution. The extraction was assisted by the use of an ultrasonic bath. The sample tube was suspended in the bath at room temperature for 1 min. After extraction, the sample was separated by centrifugation, the supernatant retained, and the pellet extracted a second time as described above. The combined organic fractions were merged and evaporated to dryness at room temperature with a gentle stream of nitrogen. Once dry, the sample was resuspended in 200 μl of ethyl acetate and filtered (0.22 μm) prior to analysis.

Analysis was undertaken using a Perkin Elmer Clarus 500/580 GC-MS with a HP-5ms column (30 m, 250 μm ID and 25 μm film thickness). The carrier gas was helium, the split ratio set at 10:1, while the temperatures for the inlet, transfer line, and ionization source were 250, 180, and 200°C, respectively. The detector was set to scan between 80 and 500 μ with a 3 min solvent delay. The initial oven temperature was 60°C (10 min) followed by an 8°C/min ramp to 300°C followed by a 10 min hold. Approximate quantification of the analytes was achieved by comparing peak area to that of pristane and a response factor of 1 assumed. For pristine, a 6-point calibration curve was made between 0.5 and 50 μg/ml. Retention times of the unbranched alkanes were determined using a standard mixture of C_{10}–C_{19}.

RESULTS AND DISCUSSION

Physicochemical Data

Cores 1, 2 and 3 showed slightly different values in pH and redox potential. Cores 1 and 2 showed a similar tendency in increasing pH with depth from 1.65–1.7 (surface) to 2.4 at a depth of 8 cm in Core 1 and to 2.68 at a depth of 15 cm for Core 2. Redox was found to be positive in all layers with insignificant variations between depths and with values always >400 mV (range 413–470 mV). The three cores had visual differences in structure and exhibited mostly "oxidized colors," from mixtures of yellow/brown, to red/brown with some ochre and in some places a completely black appearance. Core 3 was distinct in comparison to other cores, being more homogeneous and with a stable pH (2.4–2.5) across the whole depth gradient (**Supplementary Table 1**).

Comparison of physical–chemical parameters between cores suggested certain variations in the content of metals and metalloids, anions, nitrogen, and organic matter (**Supplementary Table 1**). Core 1 possessed more Fe and Pb in the three upper layers (1.1, 1.2, 1.3.1) and a consistently high presence of As in all layers. Core 2 demonstrated more Rb and Ti in all layers. Both Cores 1 and 2 showed an increase in Al with depth. In contrast, Core 3 exhibited high concentrations of Cu in two layers (3.4 and 3.6, depth 9–11 and 19–21 cm), Zn (layers 3.5 and 3.6, depth 13–16 and 19–21 cm) and Rb (layers 3.3 and 3.4, depth 6–9 and 9–11 cm).

The highest amounts of organic matter were measured for Core 1 (layers 1.2 and 1.3.1) and Core 3 (layers 3.1, 3.5, and 3.6). Core 2 was found to have a low organic matter content in the sediment. The total amount of N was found to be higher in Core 2 (layers 2.2, 2.3, and 2.4) and in Core 3 (3.3, 3.4, 3.5, and 3.6). The C:N ratio was significantly higher in upper layers of Core 1 (values of 25.4 and 12.5 for layers 1.1 and 1.2, respectively) and Core 2 (26.7). In Core 3, an opposite pattern was apparent with C:N ratios of 12.4 and 17.6 seen in the deeper layers (3.6 and 3.7).

Interestingly, few fluctuations were observed in the content of fluoride, chloride, nitrate, phosphate, and sulfate. Core 1 (layer 1.3.2) possessed the highest concentrations (in mg/kg) of F^- (65.3), Cl^- (6.5), NO_3^- (653), and SO_4^{2-} (90528). Core 3 exhibited an increased content of F^-, PO_4^{3-}, and SO_4^{2-} in some layers (**Supplementary Table 1**). These observations suggest a high degree of heterogeneity in chemical composition between the cores and individual subsamples.

We analyzed 31 different chemical properties in the sediments which we divided into three categories, namely: "Carbon-Nitrogen," "Anions," and "Other elements." A preliminary PCA showed a very complex distribution of the influence of chemical variables over the different core layers. Also, some of the chemical variables overlapped and were not used in order to reduce redundancy. Measures of total C and N (mg/kg) were removed from the analysis, while sulfate (g/kg) was included. Therefore, 28 of 31 chemical properties were included in the analysis. The analysis was divided into three different parts according to each type of chemical property. Each of these analyses is composed of a PCA where the contribution percentage of each variable has been calculated and included using a color key, specific for each core (**Figure 1** and **Supplementary Figures 2–4**).

The PCA for "Carbon-Nitrogen" showed that Core 1 and 2 are quite distinct from each other while Core 3 remained in an intermediate position (**Supplementary Figure 2**).

Principal components analysis demonstrated that variance for F^-, Cl^-, NO_3^- and SO_4^{2-} are much higher on the sample 1.3.2 than for the rest of the layers, being so different those four measures overlapped on the representation. The concentration of PO_4^{3-} was the variable that contributed the most to the distribution of the samples in the PCA. Concentration of PO_4^{3-} was below the limit of detection (0.1 mg/kg) in every layer of Core 1, was detected in 2 layers out of 6 of the Core 2 in concentration <1 mg/kg (dry sediment), whereas 4 out of 7 layers of the Core 3 displayed values from 1.7 to 4.1 mg/kg, which therefore grouped together and distinctively from the rest (**Supplementary Figure 3**).

The specific PCA was conducted based on concentrations of "Other elements," primarily metals and metalloids (**Supplementary Figure 4**). Iron (Fe), arsenic (As), and manganese (Mn) were the elements which had greatest influence on the PCA; concentrations of Fe and As were much higher in Core 1, while Mn was greatest in Cores 1 and 2, but lower in Core 3. On the other side, zinc (Zn), copper (Cu), and yttrium (Y) were specifically higher in some sublayers of Core 3; however, these are the variables showing less contribution percentage to the patterns shown in the PCA.

Principal Components Analysis Using All Chemical Properties

All chemical parameters were then analyzed and included in the same PCA (**Figure 1**). Again, the sublayer 1.3.2 dropped far away from the rest of "the cloud" due to its drastic shift in values on anions concentrations, except PO_4^{3-}. For this reason, the group of the Core 1 showed a very high variance (represented by a big ellipse). However, it is also evident how the remaining variables influenced the separation of the rest of layers groups, with Fe and As concentration pushing for Core 1 group as long as Pb and Br (which was not so clear in the specific PCA for elements) (**Figure 1**).

Comparison of chemical composition of sediment from the surface and overlaying waters established previously (Korzhenkov et al., 2019) and in this study showed that Al was represented in significantly higher quantities across the gradient, exceeding its concentrations on the surface up to 5–9 times. Concentrations of K and Ti determined in sediment core layers at various depths were >2-fold higher than at the surface, whereas Cr and Mn were present at lower concentrations. Ni, Zn, Ca, As, and Sr had about the same concentrations across samples with few exceptions (e.g., more abundant in some layers). Pb was generally detected in lower quantities than on the surface, however, there were few exceptions. Fe was found in high various quantities in different layers of sediment, comparable with those at the surface (66.7 g/kg) (**Supplementary Table 1**).

Total carbon and nitrogen were less abundant in deeper layers in comparison to those at the surface (2.8 and 0.3%, respectively)

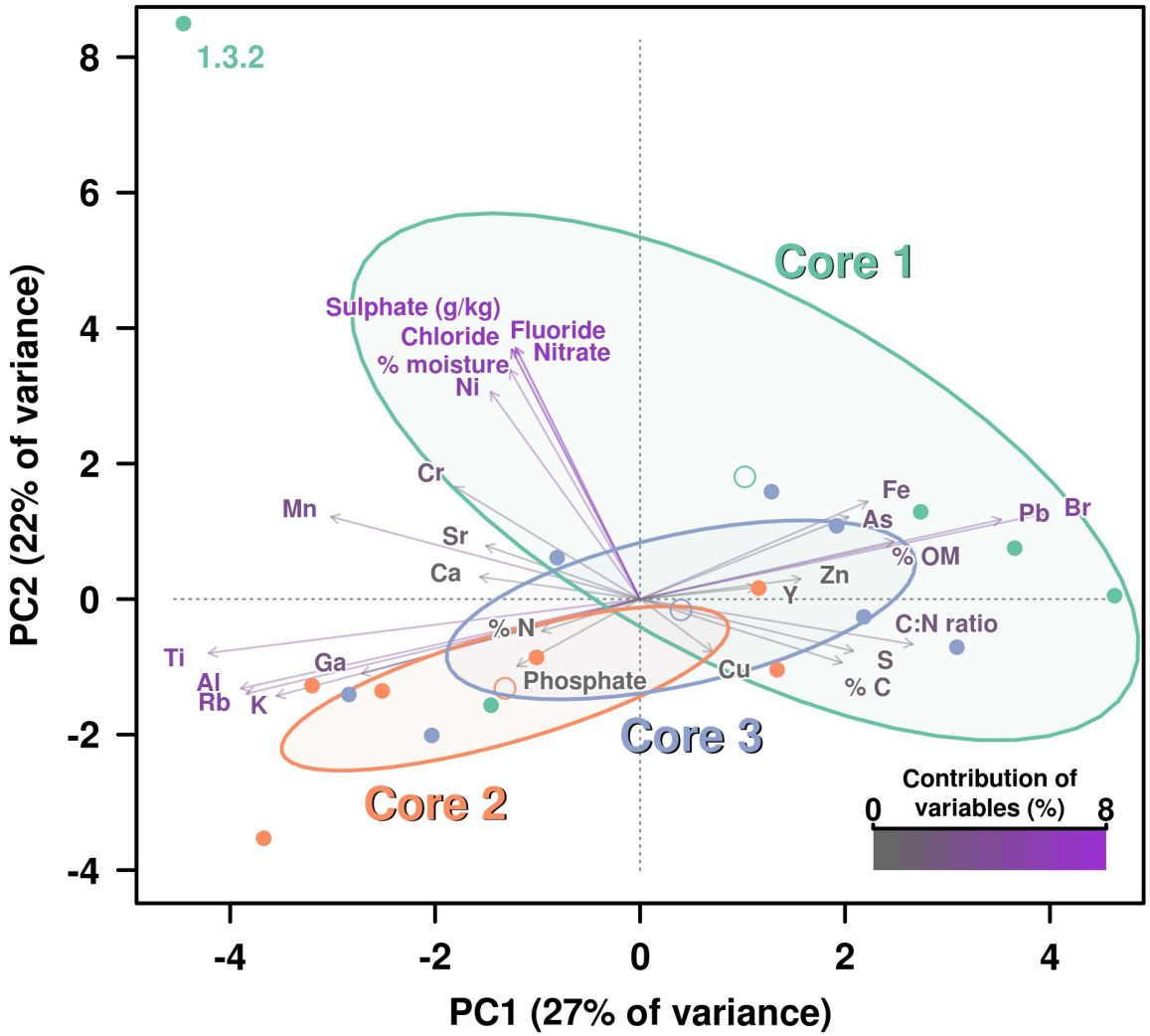

FIGURE 1 | PCA including all chemical parameters analysis by principal components analysis (PCA) of the influence of all chemical properties measured on the three cores. Contribution of each variable (chemical properties) to this graphical representation is shown by a color key from medium gray (less contribution) to violet (highest contribution). Ellipses and open dots represent the variance and mean for each core, respectively. Anion concentrations are showing the highest percentages of contribution due to the higher figures on these values for measured on layer 1.3.2, which is disrupting the variance (ellipse) corresponding to Core 1.

(**Supplementary Table 1**). C:N ratio was highly variable (0.8–26.7) across the different layers and was not dependent on depth (**Supplementary Table 1**).

GC-MS analysis of black layers (oily deposits) from Parys Mt acidic stream sediment identified hydrocarbons, specifically unbranched alkanes with C_{17} being the most abundant type.

MICROBIAL CONTENT

Taxonomic Composition of Microbial Communities in Sediment Layers

Archaea

Archaeal sequences were found in all three cores (**Figure 2**), which is in accordance with previous studies investigating surface sediments (0–3 cm) at this site (Korzhenkov et al., 2019). Across different sediment depths, archaea represented a dominant group, as judged from the total number of reads and numbers of OTUs, particularly in Cores 1 and 2. In Core 3, a very large number (ca. 30%) of archaeal reads were observed in the upper sediment layer; all deeper layers displayed a consistent decrease of archaeal reads (down to 6%) and increase in various bacterial groups.

Archaeal diversity has been mostly restricted to Euryarchaeota (or Thermoplasmatota, according to the GDTB taxonomy[1]), and among those, mainly to the members of the order Thermoplasmatales. In this study, Thermoplasmatales reads were detected in high abundance as follows: (i) in Core 1 it ranged from 49% of the total reads at the surface to 39.5% at a depth of 6–8 cm;

[1]https://gtdb.ecogenomic.org

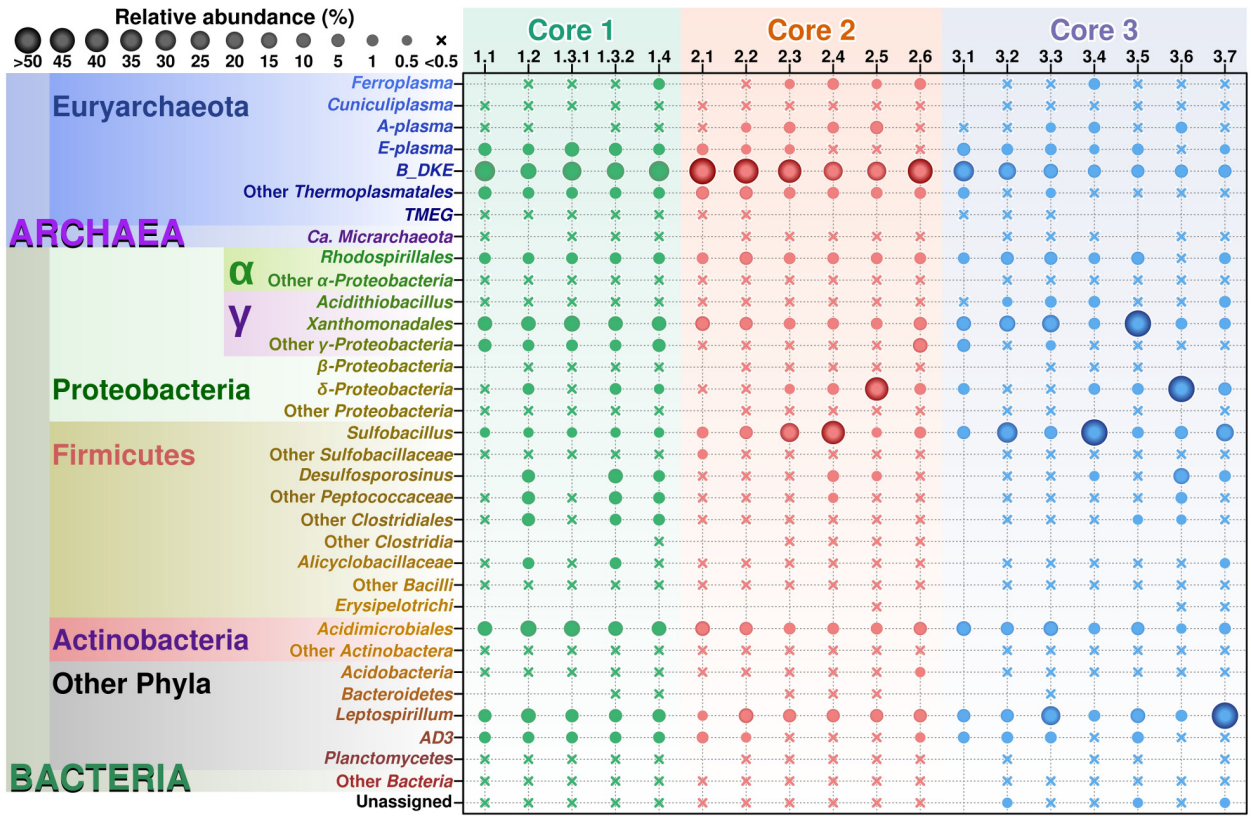

FIGURE 2 | Relative abundance of various taxonomic groups in Parys Mt sediments. OTUs found after analysis of the sequencing results were grouped by lineage on those most abundant taxons, from lowest to higher levels, with genus as the basic clustering level where possible. The final table was generated with 30 taxonomic groups. From this table, a balls diagram was produced showing the relative abundance of these taxonomic groups.

(ii) in Core 2 they represented 54.1% of total reads at the surface to 51.5% at a depth of 10–15 cm; (iii) in Core 3 they represented 39.9% at the surface to 6.2% at a depth of 20 cm.

Among Thermoplasmatales, the most abundant group of sequences across the depth gradient was affiliated to B_DKE metagenomic assembly. These sequences represented 5–45% of the total with the greatest abundance seen in Core 2. This group has also previously been reported in pyrite mine biofilm (Harz Mountains, Germany) by Krause et al. (2017). These archaea were followed by the "E-plasma" variant which was present in all three cores with varying numbers (0.5–15%) depending upon depth. In addition, within Cores 1 and 2, sequences similar to the phylotype with accession number FR683002 and to other unclassified Thermoplasmatales were detected (<0.5–10%). Reads related with "A-plasma" metagenomic assembly, *Ferroplasma acidiphilum*- (both in quantities <0.5–5%) and *Cuniculiplasma divulgatum*-related (with a relative abundance of <0.5%) organisms were also identified. These phylotypes clustered with known taxonomic clades of archaea or reference organisms, as demonstrated in **Figure 3** and **Supplementary Table 2**. No correlation of relative numbers of these taxonomic groups with sediment depth was seen. However, in the case of "A-plasma"- and *F. acidiphilum*-related organisms, their abundance gradually increased with depth down to the black-colored layer (**Figure 2**). Maximal numbers of "A-plasma" were observed at 8–10 cm (Core 2), and for *Ferroplasma*-like sequences at 6–8 cm (Core 1), 4–15 cm (Core 2), and 9–11 cm (Core 3). Interestingly, *Ferroplasma* reads were not detected in upper layers of all three cores, and their presence has not previously been reported in any other parts of the Parys Mt ecosystem (Korzhenkov et al., 2019). *Cuniculiplasma* spp. was the lowest-abundance group among other Thermoplasmatales with a relative abundance <0.5% across all layers and depths. These archaea were also earlier shown to only comprise a minor group in the upper sediment/water stream community (Korzhenkov et al., 2019).

In this study, minor quantities (0.1–0.5%) were affiliated with TMEG-related organisms (Terrestrial Miscellaneous Euryarchaeal group, or ambiguous taxa in the class Thermoplasmata, as per the SILVA database v.132). The relative abundance of this group were relatively constant with depth in Core 1, but were mostly detected in the upper sections of Cores 2 and 3. Furthermore, "*Ca.* Micrarchaeota" was present in very low abundance (<0.5%) across almost all sediment depths. Both groups were shown previously to inhabit the uppermost layer of sediments and can also be found in the overlying stream water (Korzhenkov et al., 2019).

All archaea of the order Thermoplasmatales described so far are prominent inhabitants of acidic environments and exhibit

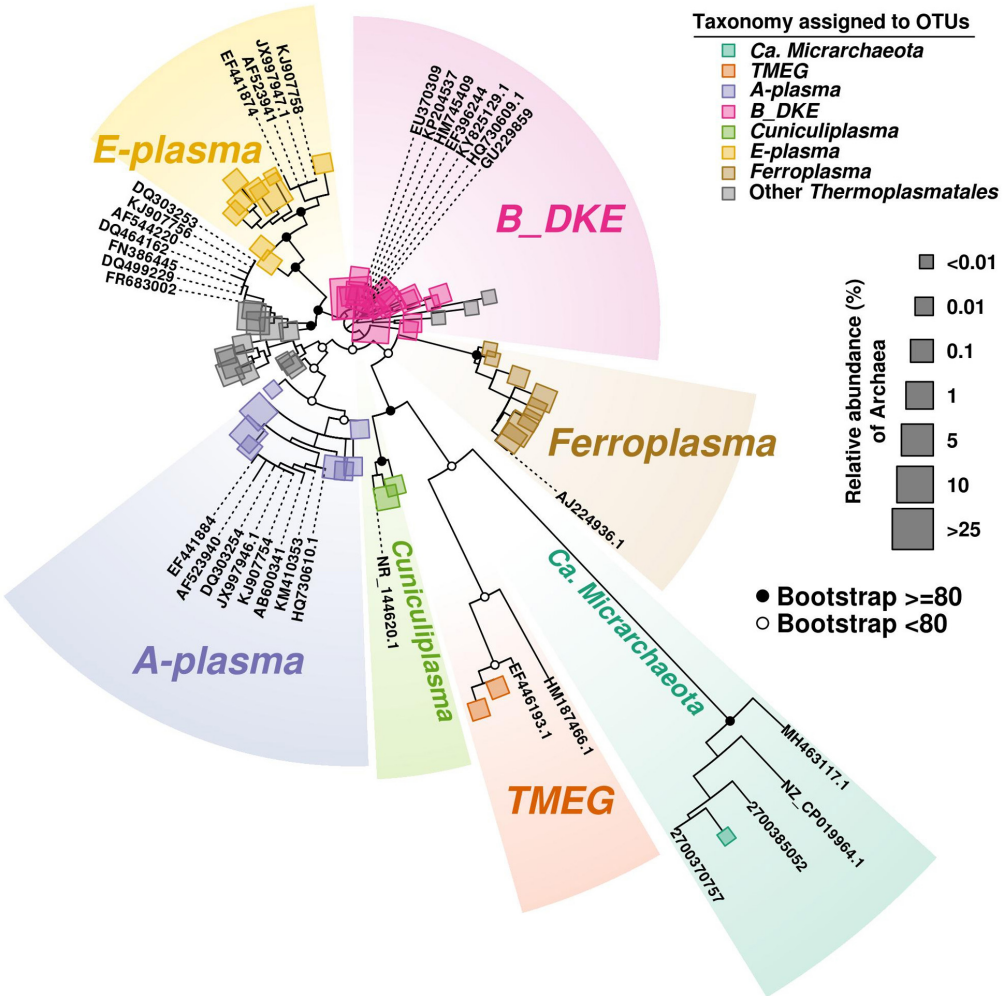

FIGURE 3 | Phylogenetic tree of Archaea. The tree was developed to include the most abundant OTU (>100 reads) sequences found along the three cores. Bootstrap values are shown on main parental nodes, where open dots represent bootstrap values under 80, while closed black dots represent values equal or higher to 80. OTU sequences are represented by colored squares corresponding to their assigned taxonomy (see section "Bioinformatics Analysis in the Materials and Methods"), while size corresponds to their relative abundance (%) relative to the amount of Archaea present. Reference sequences are represented by their accession number on GeneBank or IMG/M system.

a heterotrophic lifestyle, which is reflected in their preferential growth on complex polypeptides (Golyshina, 2011; Golyshina et al., 2016b). *Thermoplasma* was also shown to possess the potential for sulfur-driven respiration with organic carbon as an electron donor (Darland et al., 1970). Furthermore, members of the family Ferroplasmaceae are able to undertake iron oxidation/reduction (Golyshina, 2011). Heterotrophy and iron redox cycling (iron is highly available under oxidative redox conditions and low pH) together with facultatively anaerobic capability are likely present among archaeal components of these sediment communities. The occurrence of Fe (III) reduction in acidic sediments at low oxygen concentration was reported previously (Küsel et al., 2002). Sulfur respiration could potentially be another trait of these archaea. Iron redox cycling and heterotrophy were confirmed experimentally for cultured mesophilic species of *F. acidiphilum* and *C. divulgatum*, respectively (Golyshina et al., 2000, 2016b). However, since the majority of archaea populating this environment are uncultured, their metabolic properties remain to be confirmed.

It should be noted that all these archaeal phylotypes are widely found in a range of acidic environments. Archaea designated as B_DKE were identified in enrichment cultures established with biofilms obtained from a pyrite mine (Harz Mountains, Germany) (Krause et al., 2017). The organism was shown to grow in anaerobic enrichment culture when the medium was supplemented with polypeptides and ferric sulfate; furthermore, the authors suggested that these archaea could undertake ferric iron reduction (Krause et al., 2017). Similar features are highly likely for B_DKE archaea although their physiological properties still need to be confirmed in pure culture. Similar organisms are present in various low-pH environments (Krause et al., 2017). For example, almost identical SSU rRNA gene sequences

with accession numbers HQ730609, EU370309, HM745409, and EF396244 were detected in anaerobic sediments and biofilm communities from Rio Tinto (Spain), an extremely acidic, metal-rich stream (Huelva, Spain), and in La-Zarza-Perrunal acid mine effluent (Spain) (Rowe et al., 2007; Gonzalez-Torril et al., 2011; Sánchez-Andrea et al., 2011). Moreover, similar phylotypes were recovered from a low temperature (8.5°C) underground mine at Cae Coch (GU229859, Wales, United Kingdom) (Kimura et al., 2011), and in an acidic geothermal area (35°C) of Copahue (KP204537, Neuquen, Argentina) (Urbieta et al., 2015).

Other most-abundant phylotypes from Parys Mt sediments were clustered with the sequence with the accession number FR683002 from the microbial community of Pb-Zn mine, and also in acid mineral bioleaching systems of Dongxiang copper mine, Yinshan Lead-Zink mine and Yun-Fu pyrite mine (DQ464162; FN386445), all places located in China (Xiao et al., 2008; Tan et al., 2009; Huang et al., 2011). Furthermore, similar sequences were present in macroscopic filaments from Rio Tinto (Spain) (DQ303253, Garcia-Moyano et al., 2007), in cave wall biofilms from the Frasassi cave system, Italy (DQ499229; Macalady et al., 2007), in Iron Mountain AMD system, United States (AF544220; Baker and Banfield, 2003), in thermal and acidophilic biofilms, Mexico (KJ907756; unpublished) and in endolithic microbial community from Rio Tinto basin, Spain (EF441883; unpublished).

Other archaea identified in Parys Mt sediments and still awaiting their isolation are "E-plasma" and "A-plasma" (Baker and Banfield, 2003). Firstly detected in Iron Mountain (United States) metagenomic datasets, these organisms were found in the Parys Mt acidic stream surface sediment, with "E-plasma" as a dominant phylotype (Korzhenkov et al., 2019). Their metabolism was predicted as heterotrophic, which involves iron oxidation/reduction (Yelton et al., 2013). It is worth noting that both are also ubiquitous: they were found, e.g., in macroscopic filaments (DQ303254 and EF441874, correspondingly) (Garcia-Moyano et al., 2007), and in endolithic communities in the Rio Tinto basin (EF441884 and EF441874). The "A-plasma" phylotype was also detected in anaerobic sediments from Rio Tinto (HQ730610; Sánchez-Andrea et al., 2011). Furthermore, 16S rRNA gene amplicon reads of both archaea were found in forested wetland sediment samples influenced by waste coal deposits, United States (AF523940, AF523941; Brofft et al., 2002), in terrestrial subsurface cave systems, Italy (KM410353, AF523941; Hamilton et al., 2015) and in a thermal acidic biofilm, Mexico (KJ907754, KJ907758; unpublished). Additionally, sequences clustering with the "A-plasma" were present in the metagenomic data from a terrestrial acidic spring field, Japan (AB600341; Kato et al., 2011).

In relation to our results, a few points need to be highlighted. Firstly, there is still an extremely small number of archaeal taxa cultured from AMD, in comparison to bacteria. Bacterial acidophilic diversity associated with AMD sites is assigned to more than 13 genera belonging to various phyla (Acidobacteria, Actinobacteria, Firmicutes, Nitrospirae, and Proteobacteria) (Mendez-Garcia et al., 2015; Gavrilov et al., 2019). However, all cultured archaea from similar AMD environments with validly published names are affiliated with the single order, Thermoplasmatales of the phylum Euryarchaeota (genera *Ferroplasma*, *Acidiplasma*, and *Cuniculiplasma*) (Golyshina et al., 2000, 2009, 2016b; Hawkes et al., 2008). Thermophilic crenarchaeon *Metallosphaera prunae* isolated from a uranium mine is the only example of cultured representatives from another higher archaeal taxon (Fuchs et al., 1995). Thus, organisms of the order Thermoplasmatales are considered to be the most successful archaeal colonizers of mining sites, natural or anthropogenic environments with moderate temperatures, benefiting from low pH and oxygen levels. The second important point to consider when assessing sequencing data from similar environments is that the sequences submitted to the databases with the 16S rRNA sequence identity levels below 94% with the reference isolates, are often wrongly qualified as *Thermogymnomonas* spp. or *Thermoplasma* spp., which creates confusion and leads to incorrect interpretation. Importantly, *Thermogymnomonas* or *Thermoplasma* spp. were so far not detected in the low- or moderate-temperature AMD environments.

Other archaea inhabiting Parys Mt sediment belonged to "*Ca*. Micrarchaeota" detected at different depths of the three cores. These sequences showed 98–99% 16S rRNA gene identity levels to organisms from volcanic environments (GQ141757; KJ907762) and from Parys Mt surface parts (Golyshina et al., 2019). 16S rRNA sequence identity of these sediment variants to "*Ca*. Mancarchaeum acidiphilum," Mia14 was found to be 91.8%.

Bacteria

Among bacteria, members of the phylum Proteobacteria were most abundant in all cores, comprising on average 26.0 ± 3.5% of the community across all depths. Firmicutes in all layers reached moderate numbers representing 7.2 ± 3.8% of the total community (**Figure 2**). Other bacterial groups consistently present in all layers were from the phyla Nitrospirae, Actinobacteria, uncultured Chloroflexi (AD3 group), Acidobacteria and others (**Figure 2**).

No correlation of Proteobacteria distribution with sediment depth was observed. Among Proteobacteria, classes Alphaproteobacteria, Deltaproteobacteria, and Gammaproteobacteria signatures were the most prominent. Gammaproteobacteria were represented mostly by three groups of organisms: the unclassified Gammaproteobacteria, order Xanthomonadales (family Xanthomonadaceae) and the cluster RCP1-48.

Xanthomonadaceae (0.5–45%) were represented mostly by organisms closely related to *Metallibacterium scheffleri*, described as facultatively anaerobic, iron-reducing organisms (Ziegler et al., 2013). In addition, some *Stenotrophomonas* spp. and *Pseudoxanthomonas* spp. were detected. Also, *Acidithiobacillus* spp. -related OTUs, with a rather low sequence identities with type strains (<96–97%) were observed in minor amounts (<0.5–1%). Similarly, low numbers of *Acidithiobacillus* were earlier detected in the surface sediment and water, suggesting that this particular environment is not very favorable to these organisms (Korzhenkov et al., 2019). A possible reason is the extremely low pH (<2), high redox and abundance of Fe (III) in Parys Mt AMD; these factors were previously considered as less advantageous

for these organisms (Rawlings et al., 1999). Alphaproteobacteria were detected in quantities from 0.1% to a maximum of 7.3% at all sediment depths. Representatives of Rhodospirillales (family Acetobacteraceae) were seen mostly in OTUs with a very distant phylogenetic position from *Rhodophiala* spp., *Acidisoma* spp., and *Acidisphaera rubrifaciens*, making it challenging to speculate on their metabolism.

Deltaproteobacteria were associated with the order Bdellovibrionales, family Bacteriovoraceae, in which the sequences showed low homology (less than 90%) to described isolates. Patchiness was observed for the vertical distribution of these bacteria. Some increase in numbers of Bacteriovoraceae with depth was observed. Another relatively abundant bacterial phylum was Actinobacteria (<0.5–15%) with OTUs affiliated mostly with Acidimicrobiales. Among them, the sequences similar to *Aciditerrimonas* (95% identity to *Atn. ferriducens*), *Acidimicrobium* (95% identity to *Am. ferrooxidans*) and *Ferrimicrobium acidiphilum* (100%) were detected. *Aciditerrimonas* was described as facultatively anaerobic, heterotrophic and autotrophic organism, able to undertake dissimilatory reduction of ferric iron (Itoh et al., 2011). *Acidimicrobium* and *Ferrimicrobium* are known inhabitants of acidic environments, with the ability to undertake iron oxidation to undergo heterotrophic growth (Mendez-Garcia et al., 2015).

The consistent presence of Nitrospirae (1–20%) was demonstrated at various depths in all cores. Of note, at the depth of 18–20 cm in Core 3, the Nitrospirae OTUs reached 52.1%, with affiliation of all sequences to *Leptospirillum* spp. (Markosyan, 1972; Hippe, 2000; Coram and Rawlings, 2002), represented mostly by *Leptospirillum ferrooxidans*-related organisms and by new species of this genus. All validly published leptospirilli were described as aerobic and autotrophic (ferrous iron oxidizing) organisms (Markosyan, 1972; Hippe, 2000; Coram and Rawlings, 2002).

Firmicutes were found to increase their numbers with depth in Core 1 and varied in numbers in other cores, in line with the physicochemical heterogeneity of the sediments. Among them, the sequences of Sulfobacillus, YNPFFP6 group of Sulfobacillaceae-, *Alicyclobacillus*-, and *Desulfosporosinus*-related bacteria were the most representative OTUs. *Sulfobacillus* and *Alicyclobacillus* spp. are well-known inhabitants of AMD systems with facultatively anaerobic lifestyles and capable of iron oxidation and reduction, oxidation of sulfur compounds and heterotrophic or autotrophic types of carbon assimilation (Mendez-Garcia et al., 2015). Sulfate-reducing *Desulfosporosinus* members were also previously shown to inhabit AMD sediments (Alazard et al., 2010; Sánchez-Andrea et al., 2015). Firmicutes were found to be highly represented in the black-colored layers, reaching proportionally high numbers of 30–50% of total reads. Of note, at a depth of 9–11 cm in Core 3, Firmicutes represented 75.2% of the total reads. The majority of OTUs found were either *Sulfobacillus*- and *Alicyclobacillus*-related sequences, only distantly affiliated to the species with the established taxonomy. Other Firmicutes belonged to *Desulfosporosinus* and other bacteria of the family Peptococcaceae (Clostridiales). Moreover, sequences distantly related to other families of the order Clostridiales were identified in the sequencing data of "black layers." During the sampling, while inserting sampling corers into the sediments and reaching the "black horizon," we observed the development on the water surface of a thin hydrophobic film, highly likely, of hydrocarbons. We measured hydrocarbons in two selected samples of "black layers" and identified the *n*-heptadecane as a major component (19 and 43 mg/kg). This compound is known to be the most abundant product in cyanobacteria, but can also hypothetically be formed from fatty acids through reactions catalyzed by reductases and decarboxylases (Kang and Nielsen, 2017). Whatever the origin, this compound can be metabolized by acidophilic bacteria, including *Sulfobacillus* spp., as demonstrated previously (Hamamura et al., 2005; Ivanova et al., 2013).

Across the depths, other bacteria were represented by uncultured Chloroflexi (AD3 group/JG 37-AG-4) in numbers between 0.5–1% for Cores 2 and 3 and ca. 5% within Core 1. The metabolic features of these organisms previously detected in acidic ecosystems, remain unknown (Gavrilov et al., 2019).

In order to assess how the abundance profiles differs in the three cores, a non-metric multidimensional scaling (NMDS) was performed, using Bray–Curtis distances (**Figure 4**). The NMDS of the whole community suggests that the most abundant groups were in general not defining very well the differences over the three cores, hence these groups are mostly concentrated close to the center of the diagram. Additionally, NMDS results emphasized that microbial community stability decreases with depth. So, all samples from Core 1 kept a more similar taxonomic distribution profile than Cores 2 and 3 (see **Figure 4**). This can also be observed in their ellipse ranges, based on layers variance. Finally, separation among samples seems to be the result of less-abundant taxa, especially in Core 3. For instance, TMEG was detected at very low quantities in all samples, however, the layer 3.1 showed the biggest relative abundance (0.336%) which is about 12-fold higher than the average of TMEG numbers (0.027%). This was also the case with *Sulfobacillus*, which was especially abundant in layer 3.4 (>70%) (**Figure 4A**).

If we focus on the NMDS representation for Archaea, we can see a very large difference in the distribution. Samples from the Core 1 clustered very compactly showing a very similar distribution of all archaeal groups. In contrast, samples from Core 3 showed a large amount of scatter and largest variance on their ellipse (**Figure 4B**).

Correlation Analysis Between Microbial Diversity and Chemical Properties

Canonical correlation analysis was used to demonstrate the relationship between chemical properties and microbial community composition (in this case, treating microbial groups as variables). According to the CCorA, B_DKE phylotype and also other Thermoplasmatales were the groups with highest correlation with chemical variables, specifically to As, Fe, Cr, and Mn, in comparison with bacteria (**Figure 5**). All Thermoplasmatales phylotypes and '*Ca.*

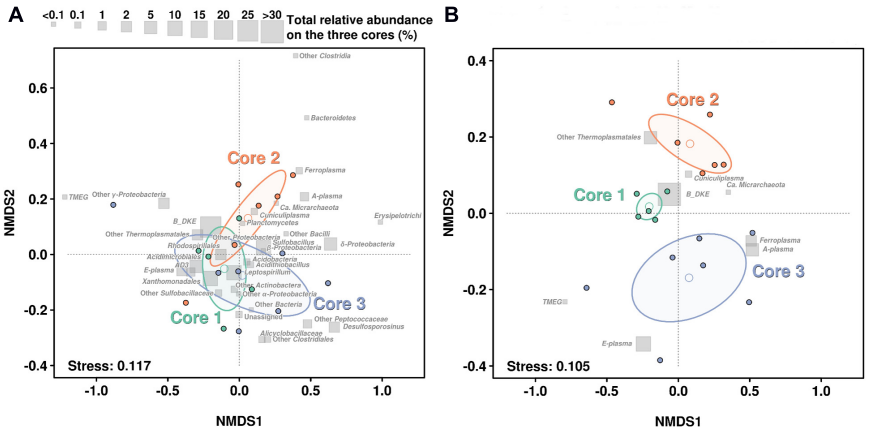

FIGURE 4 | NMDS based on the taxonomic profiles in each sediment core. **(A)** NMDS regarding the distribution of the whole community. **(B)** NMDS regarding only distribution of *Archaea*. Gray squares show the relative abundance of each taxonomic group in all layers on the three cores. Open dots show the mean of each layer while ellipse lines are based on the variance observed among each group of layers on each core. Stress level of analysis return a value of 0.118 and 0.108, which is considered a good or very good model adjustment over the 2D plane.

Mancarchaeum acidiphilum' possess a high genomic potential for metal resistance, as suggested previously in acidophiles (Dopson and Holmes, 2014). Thus, metallochaperones, heavy metal reductases, mercury (II) reductases, CopP type ATPases, arsenic efflux pump-related proteins (ArsA, ArsB, and ArsR) were found in the genomic data of reference organisms (**Supplementary Table 3**). Genes encoding these proteins were shown previously to be often located on "defense" genomic islands (Golyshina et al., 2016a, 2017).

Comparison With Other Acidic Sediments

In comparison to other AMD sediments, this particular system is characterized by positive redox potential and relatively low pH (1.7–2.5). The high abundance of archaea shown in this study seems different from earlier analyses due to a lower redox and higher pH values in the latter (Sánchez-Andrea et al., 2011, 2012; Sun et al., 2015). However, Parys Mt and Rio Tinto sediment archaeal phylotypes were found to be similar, supporting the prediction of the versatility of uncultured Thermoplasmatales in relation to the oxygen tolerance and pointing at their potential facultative anaerobic lifestyle. As in the present study, archaea of the order Thermoplasmatales were reported independently of sampling depth and spot at the Rio Tinto mine site (Sánchez-Andrea et al., 2011, 2012).

Diverse archaeal sequences were earlier revealed in the arsenic-rich creek sediment of Carnoules Mine, France (Volant et al., 2012). Archaea (Thermoplasmatales/Euryarchaeota together with Thaumarchaeota) were suggested to be important contributors to carbon and nitrogen cycles in microniches within the sediment. No overlap in archaeal phylotypes from Parys Mt and Carnoules Mine sediments could be observed, while the latter hosted archaea very distantly related with all cultured Thermoplasmatales (Volant et al., 2012). However, a relatively high similarity (about 97–98% SSU rRNA gene sequence identity) was recorded for reads from Carnoules Mine and Los Rueldos biofilm communities (Méndez-García et al., 2014).

The bacterial component in Carnoules Mine included members of genera *Gallionella*, *Thiomonas*, *Acidithiobacillus*, and *Acidiphilium*, all of which are indicative to pH values higher than in Parys Mt sediment (Bruneel et al., 2011).

Low pH favors the presence of other extremely acidophilic microorganisms, e.g., *Leptospirillum* spp. in the sediment samples. These organisms were shown to be completely absent in anoxic and higher pH sediments of Rio Tinto (Sánchez-Andrea et al., 2011, 2012). Other bacterial groups were found to be rather typical and characteristic for AMD sediments. Probably the lack of Bacteroidetes could be noted as a discrepancy in this context, because of the oxic conditions being inhibitory to the acidophilic members of this phylum. Other bacteria presented in large quantities in sedimental microniches, such as Gammaproteobacteria (*Acidibacter ferrireducens*, *M. scheffleri* and RCP1-48 group) together with Actinobacteria (*Aciditerrimonas*, *Acidimicrobium,* and *Ferrimicrobium)* point at the importance of iron metabolism in this ecosystem. Furthermore, their involvement in heterotrophic and autotrophic loops of the carbon cycle Parys Mt sediment is supported by presence of these very phylotypes. Apart from these microorganisms, Sulfobacillaceae and Alicyclobacillaceae families might take part in carbon and iron transformations in Parys Mt, which was also found in AMD sediments in other locations (Sánchez-Andrea et al., 2011, 2012).

Interestingly, the high abundance of particular archaeal taxa of the order Thermoplasmatales in Parys Mt sediments occurred across all samples, independently of variations in pH, Eh, and depth. However, once again, this group of organisms and overall the archaeal members of low-pH environments are significantly lagging behind their much better metabolically characterized bacterial counterparts. This is primarily associated with the difficulties of cultivation of archaea, for which (i.e., for the vast majority of members of

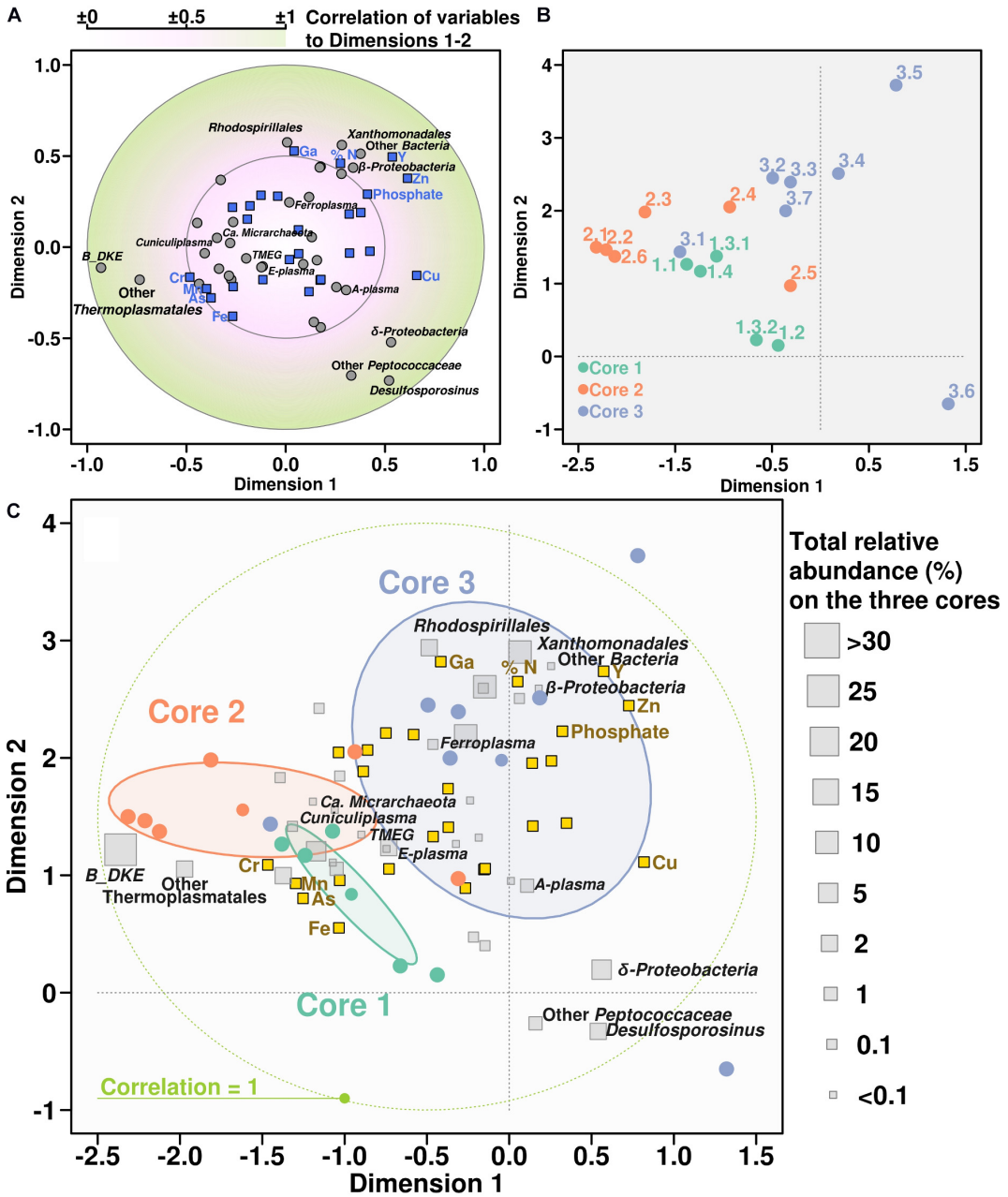

FIGURE 5 | Canonical correlation analysis between chemical variables and microbial community. Panel showing the CCorA among taxonomy distribution, chemical parameters, and the samples representation over the canonical variates. Top panels are separate representations of variables **(A)** and samples **(B)**. Below, same both representations overlapped adding the relative abundance of each taxonomic group along the whole core **(C)**.

Thermoplasmatales) only genome-informed predictions of metabolic traits are available. We suggest that the lifestyles and ecological roles of archaea in sediments of Parys Mt are based on the degradation of organic compounds from primary producers and, e.g., scavenging protein/polypeptide-rich biomass detritus and on the inorganic iron and sulfur compounds conversions. Further research is needed to understand the contribution of particular archaeal organisms inhabiting this ecosystem.

CONCLUSION

The environmental conditions in Parys Mt sediment underlying the AMD stream determined the make-up of the microbial community with a large proportion of Thermoplasmatales archaea, which were abundant at various depths and sediment layers. Bacterial community members, generally less abundant than archaea, varied in numbers more significantly across different depths, their taxonomic affiliations pointed at their

involvement in metabolism of carbon, iron, and sulfur elements. The decisive factors favoring high archaeal numbers are the low pH (1.7–2.4), the positive redox potential, availability of carbon sources (polypeptides-rich detritus/dead biomass), electron donors (ferrous iron, sulfur compounds, or carbon) and acceptors (ferric iron and oxygen). Importantly, a positive relationship was identified between Fe, As, Cr, and Mn contents and archaeal abundance, which points toward a strong tolerance of Thermoplasmatales to the high concentrations of dissolved metals and metalloids. Significant numbers of archaea in AMD sediments and the ubiquity of similar systems on our planet suggest Thermoplasmatales may have a greater impact on the global carbon, sulfur, and iron cycling than currently assumed. Further efforts are required to investigate their roles in the environment through cultivation and omics-driven analyses of their physiology and metabolism.

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 and OG conception of the work. PG, MD, GW, and FB undertook the field and laboratory work. FB, SW and DJ undertook the chemical analysis MD, EL and RB acquisition of the data for the work. RB, EL, ST, MY, PG, and OG interpretation of the data for the work. OG, RB, and PG drafted the manuscript with further contribution from all authors. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS AND FUNDING

We acknowledge the support of the Centre of Environmental Biotechnology Project funded by the European Regional Development Fund (ERDF) through the Welsh Government. RB acknowledges also the support of the Supercomputing Wales project, which is part-funded by the European Regional Development Fund (ERDF) via the Welsh Government. OG and PG acknowledge ERA Net IB2 Project MetaCat through United Kingdom Biotechnology and Biological Sciences Research Council (BBSRC) Grant BB/M029085/1. The work of ST was supported by Ministry of Science and Higher Education of Russian Federation, Grant No. 075-15-2019-1659.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.576520/full#supplementary-material

REFERENCES

Alazard, D., Joseph, M., Battaglia-Brunet, F., Cayol, J. L., and Ollivier, B. (2010). Desulfosporosinus acidiphilus sp. nov.: a moderately acidophilic sulfate-reducing bacterium isolated from acid mining drainage sediments: new taxa: firmicutes (Class *Clostridia*, Order *Clostridiales*, Family *Peptococcaceae*). *Extremophiles* 14, 305–312. doi: 10.1007/s00792-010-0309-4

Baker, B. J., and Banfield, J. F. (2003). Microbial communities in acid mine drainage. *FEMS Microbiol. Ecol.* 44, 139–152. doi: 10.1016/s0168-6496(03)00028-x

Ball, D. F. (1964). Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soil. *J. Soil Sci.* 15, 84–92. doi: 10.1111/j.1365-2389.1964.tb00247.x

Brofft, J. E., McArthur, J. V., and Shimkets, L. J. (2002). Recovery of novel bacterial diversity from a forested wetland impacted by reject coal. *Environ. Microbiol.* 4, 764–769. doi: 10.1046/j.1462-2920.2002.00337.x

Bruneel, O., Volant, A., Gallien, S., Chaumande, B., Casiot, C., Carapito, C., et al. (2011). Characterization of the active bacterial community involved in natural attenuation processes in arsenic-rich creek sediments. *Microb. Ecol.* 61, 793–810. doi: 10.1007/s00248-011-9808-9

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2008). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Coram, N. J., and Rawlings, D. E. (2002). Molecular relationship between two groups of the genus *Leptospirillum* and the finding that *Leptospirillum ferriphilum* sp. nov. dominates South African commercial biooxidation tanks that operate at 40 °C. *Appl. Environ. Microbiol.* 68, 838–845. doi: 10.1128/AEM.68.2.838-845.2002

Darland, G., Brock, T. D., Samsonoff, W., and Conti, S. F. (1970). A thermophilic, acidophilic mycoplasma isolated from a coal refuse pile. *Science* 170, 1416–1418. doi: 10.1126/science.170.3965.1416

Dopson, M., and Holmes, D. S. (2014). Metal resistance in acidophilic microorganisms and its significance for biotechnologies. *Appl. Microbiol. Biotechnol.* 98, 8133–8144. doi: 10.1007/s00253-014-5982-2

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

Fuchs, T., Huber, H., Teiner, K., Burggraf, K., and Stetter, K. O. (1995). *Metallosphaera prunae*, sp. nov., a novel metal-mobilizing, thermo-acidophilic *archaeum*, isolated from a uranium mine in Germany. *Syst. Appl. Microbiol.* 18, 560–566. doi: 10.1016/S0723-2020(11)80416-9

Garcia-Moyano, A., Gonzalez-Toril, E., Aguilera, A., and Amils, R. (2007). Prokaryotic community composition and ecology of floating macroscopic filaments from an extreme acidic environment. Rio Tinto(SW, Spain). *Syst. Appl. Microbiol.* 30, 601–614. doi: 10.1016/j.syapm.2007.08.002

Gavrilov, S. N., Korzhenkov, A. A., Kublanov, I. V., Bargiela, R., Zamana, L. V., Popova, A. A., et al. (2019). Microbial communities of polymetallic deposits"acidic ecosystems of continental climatic zone with high temperature contrasts. *Front. Microbiol.* 10:1573. doi: 10.3389/fmicb.2019.01573

Golyshina, O. V. (2011). Environmental, biogeographic, and biochemical patterns of archaea of the family *Ferroplasmaceae*. *Appl. Environ. Microbiol.* 77, 5071–5078. doi: 10.1128/AEM.00726-11

Golyshina, O. V., Bargiela, R., Toshchakov, S. V., Chernyh, N. A., Ramayah, S., and Korzhenkov, A. A. (2019). Diversity of "*Ca*. Micrarchaeota" in two distinct types of acidic environments and their associations with Thermoplasmatales. *Genes* 10:461. doi: 10.3390/genes10060461

Golyshina, O. V., Kublanov, I. V., Tran, H., Korzhenkov, A. A., Lünsdorf, H., Nechitaylo, T. Y., et al. (2016a). Biology of archaea from a novel

family *Cuniculiplasmataceae* (*Thermoplasmata*) ubiquitous in hyperacidic environments. *Sci. Rep.* 6:9034. doi: 10.1038/srep39034

Golyshina, O. V., Lünsdorf, H., Kublanov, I. V., Goldenstein, N. I., Hinrichs, K. U., and Golyshin, P. N. (2016b). The novel extremely acidophilic, cell-wall-deficient archaeon *Cuniculiplasma divulgatum* gen. nov., sp. nov. represents a new family, *Cuniculiplasmataceae* fam. nov., of the order Thermoplasmatales. *Int. J. Syst. Evol. Microbiol.* 66, 332–340. doi: 10.1099/ijsem.0.000725

Golyshina, O. V., Pivovarova, T. A., Karavaiko, G. I., Kondratéva, T. F., Moore, E. R., and Abraham, W. R. (2000). *Ferroplasma acidiphilum* gen. nov., sp. nov., an acidophilic, autotrophic, ferrous-iron-oxidizing, cell-wall-lacking, mesophilic member of the *Ferroplasmaceae* fam. nov., comprising a distinct lineage of the Archaea. *Int. J. Syst. Evol. Microbiol.* 50, 997–1006. doi: 10.1099/00207713-50-3-997

Golyshina, O. V., Tran, H., Reva, O. N., Lemak, S., Yakunin, A. F., Goesmann, A., et al. (2017). Metabolic and evolutionary patterns in the extremely acidophilic archaeon *Ferroplasma acidiphilum* YT. *Sci. Rep.* 7:3682. doi: 10.1038/s41598-017-03904-5

Golyshina, O. V., Yakimov, M. M., Lünsdorf, H., Ferrer, M., Nimtz, M., Timmis, K. N., et al. (2009). *Acidiplasma aeolicum* gen. nov., sp. nov., a euryarchaeon of the family *Ferroplasmaceae* isolated from a hydrothermal pool, and transfer of *Ferroplasma cupricumulans* to *Acidiplasma cupricumulans* comb. nov. *Int. J. Syst. Evol. Microbiol.* 59, 2815–2823. doi: 10.1099/ijs.0.009639-0

Gonzalez-Torril, E., Aguilera, A., Souza-Egipsy, V., Lopez Pamo, E., Sanchez-Espana, J., and Amils, R. (2011). Geomicrobiology of La Zarza-Perrunal acid mine effluent (Iberian Pyritic Belt, Spain). *Appl. Environ. Microbiol.* 77, 2685–2694. doi: 10.1128/AEM.02459-10

Hamamura, N., Olson, S. H., Ward, D. M., and Inskeep, W. P. (2005). Diversity and functional analysis of bacterial communities associated with natural hydrocarbon seeps in acidic soils at Rainbow Springs, Yellowstone National Park. *Appl. Environ. Microbiol.* 71, 5943–5950. doi: 10.1128/AEM.71.10.5943-5950.2005

Hamilton, T. L., Jones, D. S., Schaperdoth, I., and Macalady, J. L. (2015). Metagenomic insights into S(0) precipitation in a terrestrial subsurface lithoautotrophic ecosystem. *Front. Microbiol.* 5:756. doi: 10.3389/fmicb.2014.00756

Hawkes, R. B., Franzmann, P. D., O'Hara, G., and Plumb, J. J. (2008). Ferroplasma cupricumulans sp. nov. In List of New Names and New Combinations Previously Effectively, but not Validly, Published, Validation List no. 119. *Int. J. Syst. Evol. Microbiol.* 58, 1–2.

Hippe, H. (2000). Leptospirillium gen. nov (ex Markoysan 1972), nom. rev., including *Leptospirillium ferrooxidans* sp. nov. (ex Markoysan 1972), nom. rev. and *Leptospirillium thermoferrooxidans* sp. nov. (Golovacheva et al. 1992) *Int. J. Syst. Evol. Microbiol.* 50, 501–503. doi: 10.1099/00207713-50-2-501

Huang, L. N., Zhou, W. H., Hallberg, K. B., Wan, C. Y., Li, J., and Shu, W. S. (2011). Spatial and temporal analysis of the microbial community in the tailings of aPb-Zn mine generating acidic drainage. *Appl. Environ. Microbiol.* 77, 5540–5544. doi: 10.1128/AEM.02458-10

Itoh, T., Yamanoi, K., Kudo, T., Ohkuma, M., and Takashina, T. (2011). *Aciditerrimonas ferrireducens* gen. nov., sp. nov., an iron-reducing thermoacidophilic actinobacterium isolated from a solfataric field. *Int. J. Syst. Evol. Microbiol.* 61, 1281–1285. doi: 10.1099/ijs.0.023044-0

Ivanova, A. E., Kizilova, A. K., Kanat'eva, A. Y., Kravchenko, I. K., Kurganov, A. A., and Belyaev, S. S. (2013). A hydrocarbon-oxidizing acidophilic thermotolerant bacterial association from sulfur blocks. *Microbiology* 82, 482–489. doi: 10.1134/s0026261713040048

Johnson, D. B. (2012). Geomicrobiology of extremely acidic subsurface environments. *FEMS Microbiol. Ecol.* 81, 2–12. doi: 10.1111/j.1574-6941.2011.01293.x

Johnson, D. B., Hallberg, K. B., and Hedrich, S. (2014). Uncovering a microbial enigma: isolation and characterization of the streamer-generating, iron-oxidizing, acidophilic bacterium "*Ferrovum myxofaciens*". *Appl. Environ. Microbiol.* 80, 672–680. doi: 10.1128/AEM.03230-13

Jones, R. M., and Johnson, D. B. (2015). *Acidithrix ferrooxidans* gen. nov., sp. nov.; a filamentous and obligately heterotrophic, acidophilic member of the Actinobacteria that catalyzes dissimilatory oxido-reduction of iron. *Res. Microbiol.* 166, 111–120. doi: 10.1016/j.resmic.2015.01.003

Kang, M., and Nielsen, J. (2017). Biobased production of alkanes and alkenes through metabolic engineering of microorganisms. *J. Ind. Microbiol. Biotechnol.* 44, 613–622. doi: 10.1007/s10295-016-1814-y

Kato, S., Itoh, T., and Yamagashi, A. (2011). Archaeal diversity in a terrestrial acidic spring field revealed by a novel PCR primer targeting archaeal 16S rRNA genes. *FEMS Microbiol. Lett.* 319, 34–43. doi: 10.1111/j.1574-6968.2011.02267.x

Katoh, K., and Standley, D. M. (2013). mafft multiple sequence alignment software Version 7: improvements in performance and usability. *Mol. Biol. Evolut.* 30, 772–780. doi: 10.1093/molbev/mst010

Kimura, S., Bryan, C. G., Hallberg, K. B., and Johnson, D. B. (2011). Biodiversity and geochemistry of an extremely acidic, low-temperature subterranean environment sustained by chemolithotrophy. *Environ. Microbiol.* 13, 2092–2104. doi: 10.1111/j.1462-2920.2011.02434.x

Kock, D., and Schippers, A. (2008). Quantitative microbial community analysis of three different sulfidic mine tailing dumps generating acid mine drainage. *Appl. Environ. Microbiol.* 74, 5211–5219. doi: 10.1128/aem.00649-08

Korzhenkov, A. A., Toshchakov, S. V., Bargiela, R., Gibbard, H., Ferrer, M., and Teplyuk, A. V. (2019). Archaea dominate the microbial community in an ecosystem with low-to-moderate temperature and extreme acidity. *Microbiome* 7:11. doi: 10.1186/s40168-019-0623-8

Krause, S., Bremges, A., Munch, P. C., McHardy, A. C., and Gescher, J. (2017). Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* 7:3289.

Küsel, K., Roth, U., and Drake, H. L. (2002). Microbial reduction of Fe(III) in the presence of oxygen under low pH conditions. *Environ. Microbiol.* 4, 414–421. doi: 10.1046/j.1462-2920.2002.00314.x

Lukhele, T., Selvarajan, R., Nyoni, H., Mamba, B. B., and Msagati, T. A. M. (2019). Diversity and functional profile of bacterial communities at Lancaster acid mine drainage dam, South Africa as revealed by 16S rRNA gene high-throughput sequencing analysis. *Extremophiles* 23, 719–734. doi: 10.1007/s00792-019-01130-7

Macalady, J. L., Jones, D. S., and Lyon, E. H. (2007). Extremely acidic, pendulous cave wall biofilms from the Frasassi cave system. Italy. *Environ. Microbiol.* 9, 1402–1414. doi: 10.1111/j.1462-2920.2007.01256.x

Markosyan, G. E. (1972). A new iron-oxidizing bacterium *Leptospirillum ferrooxidans* nov. gen. nov. sp. *Biol. J. Armenia* 25, 26–29.

Méndez-García, C., Mesa, V., Sprenger, R. R., Richter, M., Diez, M. S., Solano, J., et al. (2014). Microbial stratification in low pH oxic and suboxic macroscopic growths along an acid mine drainage. *ISME J.* 8, 1259–1274. doi: 10.1038/ismej.2013.242

Mendez-Garcia, C., Pelaez, A. I., Mesa, V., Sanchez, J., Golyshina, O. V., and Ferrer, M. (2015). Microbial diversity and metabolic networks in acid mine drainage habitats. *Front. Microbiol.* 6:475. doi: 10.3389/fmicb.2015.00475

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2019). *Vegan: Community Ecology Package. R package version 2.5–6*.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rawlings, D. E., Tributsch, H., and Hansford, G. S. (1999). Reasons why 'Leptospirillum'-like species rather than *Thiobacillus ferrooxidans* are the dominant iron-oxidizing bacteria in many commercial processes for the biooxidation of pyrite and related ores. *Microbiology* 145, 5–13. doi: 10.1099/13500872-145-1-5

Rowe, O. F., Sanchez-Espana, J., Hallberg, K. B., and Johnson, D. B. (2007). Microbial communities and geochemical dynamics in an extremely acidic, metal-rich stream at an abandoned sulfide mine (Huelva, Spain) underpinned by two functional primary production systems. *Environ. Microbiol.* 9, 1761–1771. doi: 10.1111/j.1462-2920.2007.01294.x

Sánchez-Andrea, I., Knittel, K., Amann, R., Amils, R., and Sanz, J. L. (2012). Quantification of Tinto River sediment microbial communities: importance of sulfate-reducing bacteria and their role in attenuating acid mine drainage. *Appl. Environ. Microbiol.* 78, 4638–4645. doi: 10.1128/aem.00848-12

Sánchez-Andrea, I., Rodríguez, N., Amils, R., and Sanz, J. L. (2011). Microbial diversity in anaerobic sediments at Rio Tinto, a naturally acidic environment with a high heavy metal content. *Appl. Environ. Microbiol.* 77, 6085–6093. doi: 10.1128/aem.00654-11

Sánchez-Andrea, I., Stams, A. J., Hedrich, S., Ňancucheo, I., and Johnson, D. B. (2015). *Desulfosporosinus acididurans* sp. nov.: an acidophilic sulfate-reducing

bacterium isolated from acidic sediments. *Extremophiles* 19, 39–47. doi: 10.1007/s00792-014-0701-6

Sun, W., Xiao, T., Sun, M., Dong, Y., Ning, Z., Xiao, E., et al. (2015). Diversity of the sediment microbial community in the Aha Watershed (Southwest China) in response to acid mine drainage pollution gradients. *Appl. Environ. Microbiol.* 81, 4874–4884. doi: 10.1128/aem.00935-15

Tan, G. L., Shu, W. S., Zhou, W. H., Li, X. L., Lan, C. Y., and Huang, L. N. (2009). Seasonal and spatial variations in microbial community structure and diversity in the acid stream draining across an ongoing surface mining site. *FEMS Microbiol. Ecol.* 70, 121–129. doi: 10.1111/j.1574-6941.2009.00744.x

Urbieta, M. S., Gonzalez-Toril, E., Bazan, A. A., Giaveno, M. A., and Donati, E. (2015). Comparison of the microbial communities of hot springs waters and the microbial biofilms in the acidic geothermal area of Copahue (Neuquén, Argentina). *Extremophiles* 19, 437–450. doi: 10.1007/s00792-015-0729-2

USEPA (2007). *SW-846 Test Method 3550C: Ultrasonic Extraction*. Washington DC: US Environmental Protection Agency.

Volant, A., Desoeuvre, A., Casiot, C., Lauga, B., Delpoux, S., Morin, G., et al. (2012). Archaeal diversity: temporal variation in the arsenic-rich creek sediments of Carnoulès Mine, France. *Extremophiles* 16, 645–657. doi: 10.1007/s00792-012-0466-8

Wakelin, S. A., Anand, R. R., Reith, F., Gregg, A. L., Noble, R. R. P., Goldfarb, K. C., et al. (2012). Bacterial communities associated with a mineral weathering profile at a sulphidic mine tailings dump in arid Western Australia. *FEMS Microbiol. Ecol.* 79, 298–311. doi: 10.1111/j.1574-6941.2011.01215.x

Xiao, S., Xie, X., Liu, J., He, Z., and Hu, Z. (2008). Compositions and structures of archaeal communities in acid mineral bioleaching systems of Dongxiang Copper Mine and Yinshan Lead-Zinc Mine. *China. Curr. Microbiol.* 57, 239–244. doi: 10.1007/s00284-008-9183-z

Yelton, A. P., Comolli, L. R., Justice, N. B., Castelle, C., Denef, V. J., Thomas, B. C., et al. (2013). Comparative genomics in acid mine drainage biofilm communities reveals metabolic and structural differentiation of co-occurring archaea. *BMC Genomics* 14:485. doi: 10.1186/1471-2164-14-485

Zhang, X., Tang, S., Wang, M., Sun, W., Xie, Y., Peng, H., et al. (2019). Acid mine drainage affects the diversity and metal resistance gene profile of sediment bacterial community along a river. *Chemosphere* 217, 790–799. doi: 10.1016/j.chemosphere.2018.10.210

Ziegler, S., Waidner, B., Itoh, T., Schumann, P., Spring, S., and Gescher, J. (2013). *Metallibacterium scheffleri* gen. nov., sp. nov., an alkalinizing gammaproteobacterium isolated from an acidic biofilm. *Int. J. Syst. Evol. Microbiol.* 63, 1499–1504. doi: 10.1099/ijs.0.042986-0

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

frontiers in Microbiology

CORRECTION
published: 29 January 2021
doi: 10.3389/fmicb.2021.633015

Corrigendum: High Representation of Archaea Across All Depths in Oxic and Low-pH Sediment Layers Underlying an Acidic Stream

Marco A. Distaso[1,2†], Rafael Bargiela[1†], Francesca L. Brailsford[1,2,3], Gwion B. Williams[1,2], Samuel Wright[1,2], Evgenii A. Lunev[4], Stepan V. Toshchakov[5], Michail M. Yakimov[6], David L. Jones[1,2,3], Peter N. Golyshin[1,2] and Olga V. Golyshina[1,2]*

[1] School of Natural Sciences, Bangor University, Bangor, United Kingdom, [2] Centre for Environmental Biotechnology, Bangor University, Bangor, United Kingdom, [3] School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia, [4] Institute of Living Systems, Immanuel Kant Baltic Federal University, Kaliningrad, Russia, [5] National Research Centre "Kurchatov Institute", Moscow, Russia, [6] Institute for Biological Resources and Marine Biotechnology, CNR, Messina, Italy

Keywords: acidophilic archaea and bacteria, Thermoplasmatales, "*Candidatus* Micrarchaeota", unclassified Euryarchaeota/Terrestrial Miscellaneous Euryarchaeotal Group, acid mine drainage systems, mine-impacted environments, sediment microbiome

OPEN ACCESS

Edited and reviewed by:
Ricardo Amils, Autonomous University of Madrid, Spain

***Correspondence:**
Olga V. Golyshina
o.golyshina@bangor.ac.uk

†These authors have contributed equally to this work

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

Received: *24 November 2020*
Accepted: *08 January 2021*
Published: *29 January 2021*

Citation:
Distaso MA, Bargiela R, Brailsford FL, Williams GB, Wright S, Lunev EA, Toshchakov SV, Yakimov MM, Jones DL, Golyshin PN and Golyshina OV (2021) Corrigendum: High Representation of Archaea Across All Depths in Oxic and Low-pH Sediment Layers Underlying an Acidic Stream. Front. Microbiol. 12:633015. doi: 10.3389/fmicb.2021.633015

A Corrigendum on

High Representation of Archaea Across All Depths in Oxic and Low-pH Sediment Layers Underlying an Acidic Stream

by Distaso, M. A., Bargiela, R., Brailsford, F. L., Williams, G. B., Wright, S., Lunev, E. A., et al. (2020). Front. Microbiol. 11:576520. doi: 10.3389/fmicb.2020.576520

In the original article, there was a mistake. The incorrect **Figure 1** was published. The caption is correct as published. The correct **Figure 1** appears below.

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

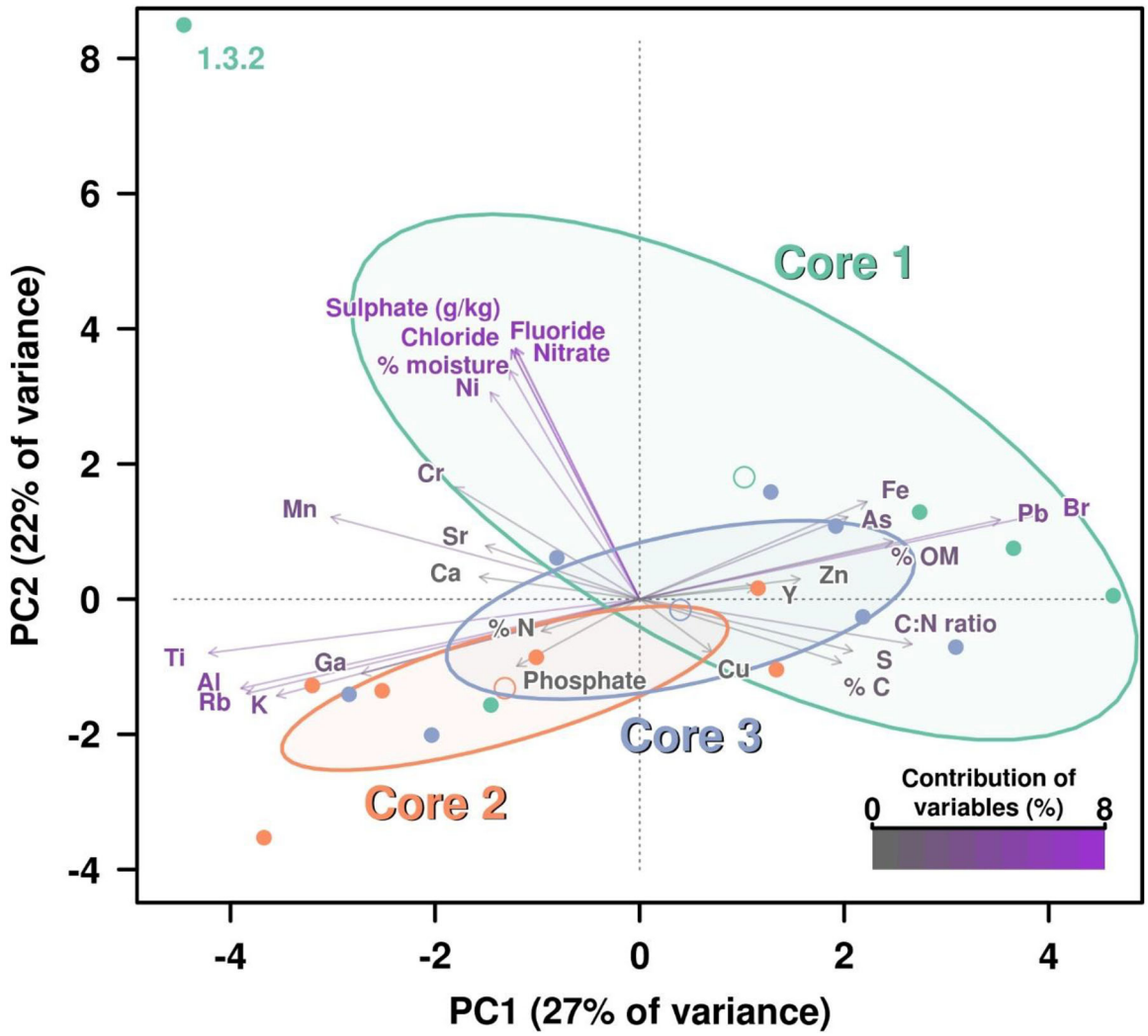

FIGURE 1 | PCA including all chemical parameters analysis by principal components analysis (PCA) of the influence of all chemical properties measured on the three cores. Contribution of each variable (chemical properties) to this graphical representation is shown by a color key from medium gray (less contribution) to violet (highest contribution). Ellipses and open dots represent the variance and mean for each core, respectively. Anion concentrations are showing the highest percentages of contribution due to the higher figures on these values for measured on layer 1.3.2, which is disrupting the variance (ellipse) corresponding to Core 1.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 03 December 2020
doi: 10.3389/fmicb.2020.614227

Hydrogenotrophic Methanogenesis Under Alkaline Conditions

*Richard M. Wormald[1], Simon P. Rout[1], William Mayes[2], Helena Gomes[2,3] and Paul N. Humphreys[1]**

[1]*Department of Biological and Geographical Sciences, University of Huddersfield, Huddersfield, United Kingdom,* [2]*Department of Geography, Geology and Environment, University of Hull, Hull, United Kingdom,* [3]*Food, Water, Waste Research Group, Faculty of Engineering, University of Nottingham, Nottingham, United Kingdom*

A cement-based geological disposal facility (GDF) is one potential option for the disposal of intermediate level radioactive wastes. The presence of both organic and metallic materials within a GDF provides the opportunity for both acetoclastic and hydrogenotrophic methanogenesis. However, for these processes to proceed, they need to adapt to the alkaline environment generated by the cementitious materials employed in backfilling and construction. Within the present study, a range of alkaline and neutral pH sediments were investigated to determine the upper pH limit and the preferred route of methane generation. In all cases, the acetoclastic route did not proceed above pH 9.0, and the hydrogenotrophic route dominated methane generation under alkaline conditions. In some alkaline sediments, acetate metabolism was coupled to hydrogenotrophic methanogenesis *via* syntrophic acetate oxidation, which was confirmed through inhibition studies employing fluoromethane. The absence of acetoclastic methanogenesis at alkaline pH values (>pH 9.0) is attributed to the dominance of the acetate anion over the uncharged, undissociated acid. Under these conditions, acetoclastic methanogens require an active transport system to access their substrate. The data indicate that hydrogenotrophic methanogenesis is the dominant methanogenic pathway under alkaline conditions (>pH 9.0).

Keywords: alkaliphiles, radioactive waste, hydrogenotrophic methanogens, acetoclastic methanogens, alkaline

OPEN ACCESS

Edited by:
Ricardo Amils, Autonomous University of Madrid, Spain

Reviewed by:
Jose Luis, Autonomous University of Madrid, Spain
Xiuzhu Dong, Chinese Academy of Sciences, China

****Correspondence:***
Paul N. Humphreys
p.n.humphreys@hud.ac.uk

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

Received: *05 October 2020*
Accepted: *11 November 2020*
Published: *03 December 2020*

Citation:
Wormald RM, Rout SP, Mayes W, Gomes H and Humphreys PN (2020) Hydrogenotrophic Methanogenesis Under Alkaline Conditions. Front. Microbiol. 11:614227. doi: 10.3389/fmicb.2020.614227

INTRODUCTION

Globally, geological disposal is largely the preferred choice for the management of intermediate level radioactive wastes (ILWs; IAEA, 2011) with a number of countries pursuing the establishment of geological disposal facilities (GDFs; Faybishenko et al., 2016). The exact design of these facilities is dependent on the host geology; however, there are often a number of common features, including the use of cementitious construction and backfill materials, the presence of metal in the construction, container and waste components, and organic, primarily cellulose, containing wastes.

The presence of metallic and organic materials means that abiotic and biotic gas generating processes (NEA, 2014) are likely to proceed within a GDF resulting in gas generation being evaluated in a number of international waste management programs (Felicione et al., 2003; NDA, 2010; Avis et al., 2011; Summerling, 2013; Poller et al., 2016). The anaerobic corrosion of metals will generate molecular hydrogen (NDA, 2010; Libert et al., 2011), which can in

turn, act as an electron donor for hydrogenotrophic methanogenesis (Libert et al., 2011; Bagnoud et al., 2016). The alkaline conditions generated by the cementitious materials will promote the alkaline hydrolysis of cellulosic wastes (Humphreys et al., 2010). This abiotic process generates a range of cellulose degradation products (CDPs; Rout et al., 2015b; Charles et al., 2019) collectively which are dominated (>70%) by isosaccharinic acids (ISAs; Almond et al., 2012).

Recent research has focused on the use of lime kiln waste sites to provide an insight into the biodegradation of CDP components under alkaline conditions (Milodowski et al., 2013; Charles et al., 2019). Despite the harsh geochemical environment (pH > 10.0), these sites support extensive and diverse bacterial and archaeal populations (Burke et al., 2012; Kyeremeh et al., 2016) capable of a wide range of metabolic and energy generating processes (Bassil et al., 2014; Rout et al., 2014, 2015a,b), including methanogenesis (Rout et al., 2015b). The information that is available for these sites suggests that the hydrogenotrophs dominate the methanogenic populations despite the availability of acetic acid and the absence of competition by other acetate consuming pathways (Charles et al., 2015; Rout et al., 2015b; Kyeremeh et al., 2016).

Given the importance of methanogenesis to gas generation within a GDF, a range of anthropogenic alkaline environments of various ages were investigated to determine the impact of environmental pH on the nature and extent of methanogenic activity.

MATERIALS AND METHODS

Sampling Site Investigations

The investigation involved sediments from three lime kiln waste sites (≈25 to ≈150 years old; New Lime sites, designated B, H, and T; Charles et al., 2019), five field kiln sites (Johnson, 2008; 200–300 years old; Old Lime sites, designated LK1, LK2, LK3, LK4, and LK5), and four steel industry waste sites (5 to ≈30 years old; Mayes et al., 2008; Steel sites, designated CW, CS, RC, and SC; see Supporting Information). In addition, a neutral pH, freshwater anaerobic sediment was employed as a control (Control). In the case of the New Lime, Steel, and Control sites, hand cored samples (dia. 30 mm) were taken at depth (~1 m). In the case of the Old Lime sites, samples were taken from a shallow depth of 10–30 cm using a stainless steel trowel. All sampling equipment was surface sterilize using a commercial hypochlorous acid spray (Redditch Medical Ltd., Stockton on Tees, UK). Sterile plastic sample containers were filled to maximum capacity with sediments and where possible with the associated pore waters in order to minimize the presence of oxygen. The pH values of the associated pore waters were determined *in situ* with a hand-held pH probe (Mettler Toledo, Columbus, OH, Unites States) and the pH of the soil/sediment determined using standard methods (B.S.I, 2005).

Microcosm Investigations

All microcosms employed in the study were incubated in the dark at 25°C, and the incubation time was dependent on the time taken for the microcosm to generate detectable methane and ranged from 2 to 8 weeks. Initial microcosms employing New Lime and Control sediments were established by mixing a soil/sediment sample (5 g) from each location with 5 ml of sterile, nitrogen purged mineral media (MM; **Supplementary Tables S1, S2**; B.S.I, 2005) adjusted to pH 10.0 with 4 M NaOH for New Lime microcosms and pH 7.0 for the Control microcosms. Slurries were then transferred to 100 ml crimp top bottles containing 100 ml of MM supplemented with 10% v/v CDP and a hydrogen:nitrogen (10:90) headspace (BOC Ltd., Guildford, UK). A CDP/hydrogen enrichment was chosen since this had previously been shown (Rout et al., 2014, 2015a; Kyeremeh et al., 2016) to support a diverse range of microbial processes and provided a source of fermentable substrates, acetate, and hydrogen. CDP was prepared as previously described (Rout et al., 2014, 2015a) All microcosm experiments were carried out in duplicate.

On the detection of methane initial enrichments were sub-cultured twice at the starting pH by inoculating sterile microcosms with 5 ml of the previous microcosm suspension (**Supplementary Figure S1**). This was carried out to reduce the impact of residual fermentable substrates, mineral components, and other electron acceptors [e.g., Iron (III) and sulphate] present within the inoculating sediment.

Once methane generation was detected at the starting pH, each sediment was sequentially exposed to a range of pH values ranging from 7.0 to 12.0 (**Supplementary Figure S1**) as described below. Every microcosm that was methane positive was sub-cultured at the next pH in the series and, hydrogen and acetate enrichments were created at the original pH (**Supplementary Figure S1**). The last pH in the series that failed to generate methane was still sub-cultured into hydrogen and acetate enrichments. These enrichments were performed by transferring 5 ml of the microcosm suspension to a 100 ml crimp top bottle containing 100 ml of MM with either a hydrogen: carbon dioxide: nitrogen (10:10:80; BOC Ltd., Guildford, United Kingdom) headspace or MM supplemented with Na-acetate (30 mM) with a 100% nitrogen headspace. A set of blank microcosms were also established as described above with MM, soil, and a 100% nitrogen atmosphere.

Chemical Analysis

Headspace hydrogen and methane was determined *via* an Agilent GC6850 equipped with HP-PLOT/Q column with particle traps ((35 m × 0.32 mm × 20 μm), Agilent Technologies, Santa Clara, CA, United States) employing thermal conductivity detection. Liquid microcosm samples (1 ml) were also removed over the course of incubation using degassed syringes and needles. Acetic acid concentrations were determined using gas chromatography (Hewlett Packard Ltd., London, United Kingdom) with a flame Ionization detector (GC-FID) as described previously (Rout et al., 2015a). To determine the presence of acetoclastic methanogenesis the metabolic inhibitor methylflouride was employed at an initial concentration of 1% as described previously (Daebeler et al., 2013). Duplicate reactors without inhibitor addition were also prepared

and served as a control. Methane produced from CO_2 (m_{CO_2}) was calculated from inhibited microcosms002C and acetate-derived methane was calculated from methane production in uninhibited microcosms ($m_{acetate} = m_{total} - m_{CO_2}$).

16S rRNA Gene Sequencing

Total genomic DNA was extracted from both soil/sediment samples and methanogenic enrichment cultures using the PowerSoil DNA Isolation Kit (Qiagen, Manchester, UK). The enrichment cultures were sampled after 2 weeks of incubation by the centrifugation (10,000 × *g*) of the microcosm fluid to pellet the cells prior to DNA extraction. The V4 region of the 16S rRNA gene was then amplified by PCR using primers 341f and 805r (Takahashi et al., 2014). Next generation sequencing of PCR products was carried out by ChunLab (South Korea) using the Illumina MiSeq platform. Sequences were then paired and clustered into OTUs using DADA2 employing R3.2.5 (Callahan et al., 2016), before being identified using the MegaBLAST search strategies (Altschul et al., 1997).

RESULTS

CDP Fed NEW Lime Microcosms

Within microcosms utilizing CDP as a carbon source, the Control enrichments demonstrated a linear reduction in methane production between pH 7.0 and 11.0 (**Figure 1**), while the New Lime enrichments demonstrated an optimum pH of 9.0 for methane generation and a linear reduction down to pH 12.0. This optimum is below the *in-situ* pH of the New Lime sites (pH 11.5–13.0, **Supplementary Table S3**), suggesting that the *in-situ* populations have alternative strategies to manage these harsh pH values (Charles et al., 2017).

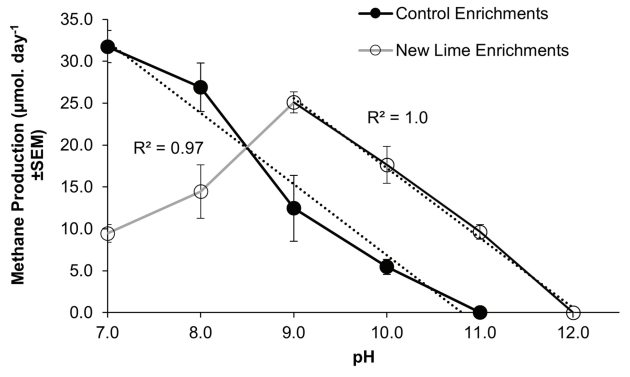

FIGURE 1 | Methane production rates in CDP-fed microcosms employing Control and New Lime sediments operating between pH 7.0 and 12.0. Control microcosms (*n* = 2 at all pH values) had a pH optimum of 7.0 (*n* = 3), in contrast to the New Lime microcosms (*n* = 2 pH 7.0, *n* = 4 pH 8.0, *n* = 6 pH 9.0, *n* = 6 pH 10.0, and *n* = 2 pH 11.0) were optimal at pH 9.0. The upper pH limit for methanogenesis was pH 10.0 for the Control sediment, and this increased to pH 11.0 in the case of the New Lime sediments. No methane generation was detected at a pH >11.0 ("*n*" denotes the number of replicates analyzed).

Mean methane production values for those microcosms producing methane are presented (**Figure 1**) in order to emphasize the overall trends observed. Across the pH range investigated the microbial populations enriched from the three New Lime sites demonstrated different responses to changing pH with the oldest site (site B; see **Supplementary Material**) generating methane across the whole range and the other two sites generating methane across narrower pH ranges (**Supplementary Figure S2**).

In the Control sediment enrichments, the archaeal community composition was composed of almost equal proportions of acetoclastic to hydrogenotrophic methanogens (46:54%; **Figure 2**) at pH 7.0 and 8.0. As the enrichment pH became more alkaline (pH ≥ 9.0), the community composition shifted toward a more hydrogenotrophic population **Figure 2** even though acetate was present in these cultures. The New Lime sediment enrichments were dominated by hydrogenotrophic methanogens (**Figure 2**), suggesting that the long-term exposure to alkaline pH had selected against acetoclastic methanogens in these sediments.

H_2:CO_2 and Acetate Fed New Lime microcosms

The trend toward hydrogenotrophic methanogenesis at alkaline pH was confirmed by the establishement of hydrogen and acetate enrichments from the CDP fed Control and New Lime enrichments at each pH. Within the H_2:CO_2 Control sediment enrichments, hydrogen consumption was at its highest at pH 7.0 at 243.3 µmol day^{-1}, and became undetectable as pH increased to pH 11.0 (**Figure 3A**). The associated methanogenic communities were most diverse at pH 7.0 with 12 different methanogenic genera detected (**Figure 3B**). *Methanobacterium* spp. were the most dominant at this pH, comprising 60.7% of the community; with *Methanoregula* spp. (14.7%) and *Methanosphaerula* spp. (10.2%) also being prevalent within the community. As pH was increased through to pH 10.0, the loss of a number of the genera observed at pH 7.0 coincided with the further dominance of *Methanobacterium* spp., rising to 97.1% of the community at pH 10.0.

These findings were reinforced by the H_2:CO_2 New Lime enrichments in which *Methanobacterium* spp. were the dominant taxa detected across all pH (7.0–11.0) enrichments (**Figure 3B**). A negative relationship between hydrogen consumption and increasing pH was seen in these enrichments as pH decreased for pH 9.0–11.0, however hydrogen consumption was still detected at pH 11.0 (77.0 µmol day^{-1}). The optimum pH for hydrogenotrophic methanogenesis was pH 9.0 with a hydrogen consumption of 167.5 µmol day^{-1}. The alkaliphilic nature of the methanogenic population in the New Lime sediments was indicated by the fact that the rates of hydrogen metabolism decreased as the pH fell below pH 9.0 (**Figure 3A**). The data here reiterate that the optimum pH for hydrogenotrophic methanogenesis is pH 9.0 for communities of this type, and that it is not the availability of substrate from fermentation (such as CDP) that is a limiting factor. As was the case with the CDP reactors, not all the New Lime enrichments were active across the whole pH range (**Supplementary Figure S3**).

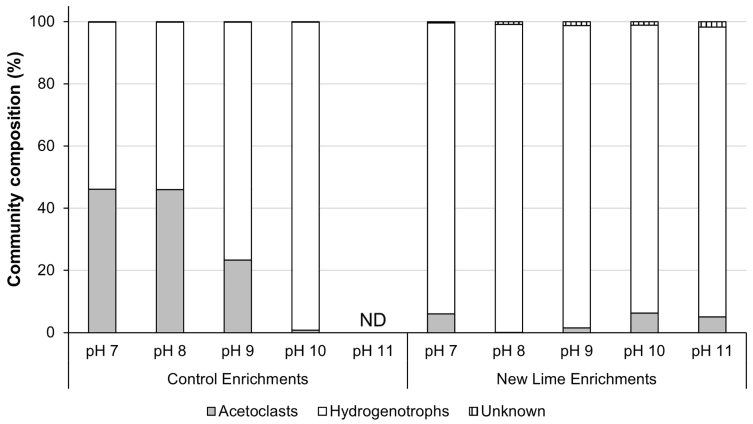

FIGURE 2 | Methanogen communities in neutral and alkaline soil enrichments. When the neutral pH Control sediment was enriched with CDP the acetoclastic and hydrogenotrophic communities were present in roughly equal proportions at pH 7.0 and 8.0, respectively. As the pH increased to pH 10.0, the acetoclastic compostion fell as pH increased, with no methanogen community detectable at pH 11.0. In contrast the alkaline New Lime enrichments were dominated by hydrogenotrophic methanogens irrespective of enrichment pH (ND, none determined).

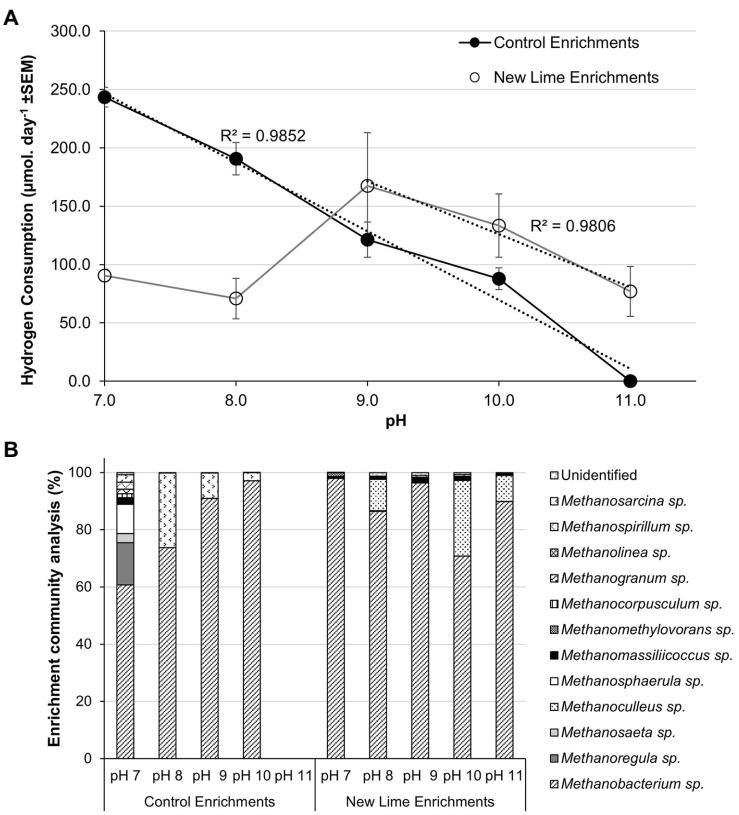

FIGURE 3 | Hydrogen consumption rates **(A)** and methanogen populations **(B)** in H_2/CO_2 enriched microcosms from Control (n = 2) and New Lime (n = 6) soils. An increasing pH negatively impacted upon hydrogen consumption in Control soil enrichments (n = 2 at all pH values), while New Lime sediments (n = 4 pH 7.0, n = 4 pH 8.0, n = 6 pH 9.0, n = 6 pH 10.0, n = 4 pH 11.0) demonstrated an optimum pH of 9.0. Across all enrichments, *Methanobacterium* sp. were the dominant genus. ND, none detected ("n" denotes the number of replicates analyzed).

Acetate fed control enrichments contained both acetoclastic and hydrogenotrophic methanogens (**Figure 4B**). Hydrogenotrophic methanogens are able to generate methane from acetate *via* syntrophic acetate oxidation (SAO). SAO is an endergonic reaction with a Gibbs free energy of +104.6 kJ mol^{-1} and is, therefore, energetically less favorable

than acetoclastic methanogenesis; however, when the oxidation of acetate is coupled to hydrogenotrophic methanogenesis, the total combined reaction is exergonic ($\Delta G0' = -31$ kJ mol^{-1}; Hattori, 2008). Few SAO bacteria have been described, but SAO is associated with the members of the phylum Firmicutes (Hattori et al., 2000; Balk et al., 2002; Westerholm et al., 2010; Manzoor et al., 2018). In the acetate fed Control enrichments (**Figure 4A**), a linear reduction ($R^2 = 0.97$) in the rate of acetate removal was observed between pH 7.0 and 9.0 from 182.7 to 33.7 μmoles day^{-1}; with a cessation of activity at pH 10.0. *Methanosarcina* spp. contributed 65.1% of the pH 7.0 and 79.3% of the pH 8.0 Control enrichment populations (**Figure 4B**). This fell to 7.1% within the pH 9.0 enrichment. The fall in acetate consumption between pH 7.0 and 8.0 coincided with the loss of both *Methanospirillium* and *Methanosaeta* spp. and a reduced proportion of *Methanosarcina* between pH 8.0 and 9.0. The loss of *Methanospirillium* spp. may suggest that SAO was active at pH 7.0 in these systems but was not able to adjust to the increase in pH.

Low levels of acetate consumption were only observed at both pH 8.0 (36.4 μmoles day^{-1}) and pH 7.0 (23.6 μmoles day^{-1}) in acetate enrichments from two of the three New Lime sites. However, the methanogenic populations in these enrichments were dominated by the hydrogenotrophic (Liu and Whitman, 2008; Dridi et al., 2012) *Methanobacterium*, *Methanoculleus*, and *Methanomassiliicossus* spp., with the potentially acetoclastic *Methanosarcina* spp. representing only 1.2% of the total archaeal populations at these pH values (**Figure 4B**). However, the members of the phylum Firmicutes which comprised 30–40% of the total populations of the pH 8.0 and 7.0 New Lime acetate enrichments, suggesting that SAO may be responsible for the low levels of acetate removal in these enrichments.

Inhibition Studies

To investigate the role of in the acetate fed enrichments, the pH 7.0 acetate fed New Lime enrichments were sub-cultured in the presence of acetate and fluoromethane, an inhibitor of acetoclastic methanogenesis (Agneessens et al., 2018). Within these enrichments (**Supplementary Figure S4**), acetate removal was observed in both the presence and absence of inhibitor (removal rates: presence: $4.1 \times 10^{-2} \pm 4 \times 10^{-3}$ mmol day^{-1}; absence: $3.5 \times 10^{-2} \pm 4 \times 10^{-3}$ mmol day^{-1}), suggesting that acetate removal in these sediments was SAO linked. The addition of fluoromethane to pH 7.0 acetate fed Control enrichment saw a cessation of acetate removal, indicating that in this case, acetoclastic methanogens was the dominant acetate removal

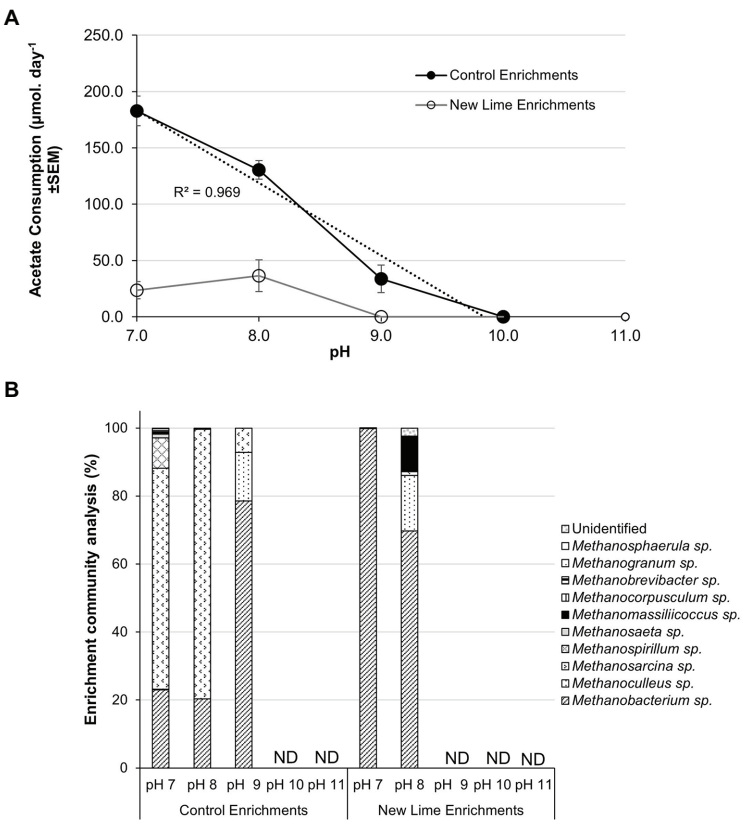

FIGURE 4 | Acetate consumption rates **(A)** and methanogen populations **(B)** in sodium acetate enriched microcosms from Control and New lime soils. Increasing pH negatively impacted upon acetate consumption with no acetate consumption observed above pH 9.0 in either the Control ($n = 2$ at all pH values) enrichments or the New Lime enrichments ($n = 2$ pH 7.0, $n = 2$ pH 8.0). *Methanosarcina* spp. and *Methanobacterium* spp. were the dominant genera in Control enrichments while *Methanobacterium* sp. were dominant in the New Lime enrichments. ND, none detected ("*n*" denotes the number of replicates analyzed).

process even though the presence of hydrogenotrophs such as *Methanospirillium* spp. may suggest some SAO activities (**Supplementary Figure S5**). These results suggest that the role of SAO in methanogen generation in alkaline environments warrants further study.

Steel and Old Lime Microcosms at pH 7 and 10

In order to determine the extent to which alkaline pH impacts upon acetoclastic methanogenesis across a wider selection of anthropogenic alkaline environments with more diverse methanogenic populations, sediments from the Steel and Old Lime sites were used to establish microcosms at pH 7.0 and 10.0. The Old Lime sediments (n = 5) had *in-situ* pH values closer to circum-neutral pH values, suggesting that the bulk environment had recovered from exposure to alkaline contamination. However, lime fragments are likely to be present within these sediments, suggesting that alkaline microsites will persist. The enrichments from these sediments were similar to the Control sediments, with both acetate and hydrogen consumptions being negatively impacted by an increase in pH from pH 7.0 to 10.0 (**Figure 5**). In particular, no acetate consumption or methane production was observed at pH 10.0. The acetate consumption rates at pH 7.0 were 2.0×10^{-2} ($\pm 7.7 \times 10^{-3}$) day^{-1} and this coincided with the detection of *Methanoculleus* sp. (70.5%), *Methanosarcina* sp. (16.6%), and *Methanomasiliicoccus* sp. (11.6%). Hydrogen consumption could still be detected within the pH 10.0 enrichments; however, the rate fell from that observed at pH 7.0 [43.1 (±6.4) μmol day^{-1}] to 17.4 (±4.6) μmol day^{-1}. At both pH 7.0 and 10.0, *Methanoculleus* sp. dominated the community composition, representing 61.6 and 89.8% of the communities, respectively.

The pH of the Steel sediments ranged from pH 9.5 to 12.9. The enrichment of these sediments with either acetate or H_2:CO_2 demonstrated similar behavior to that observed with the New Lime sediments, again no acetate consumption was detected at pH 10.0 but was observed at pH 7.0 with a rate of 36.5 (±2.0) μmol day^{-1}. At pH 7.0, acetoclastic communities from the Steel sites were dominated by *Methanosarcina* sp. (78.8%) with *Methanobacterium* sp. comprising 20.6% of the community. In contrast to the New Lime enrichments (**Figure 3A**), the Steel site sediments (**Figure 5**) demonstrated a greater rate of hydrogen consumption at pH 7.0 (90.8 ± 1.9 μmol day^{-1}) than at pH 10.0 (85.1 ± 3.2 μmol day^{-1}), although these were not significantly different (p = 0.105). The hydrogenotrophic methanogenic microcosms were again dominated by *Methanobacterium* sp.

Overall the results indicate that hydrogenotrophic methanogenesis is favored over acetoclastic methanogenesis at alkaline pH (>9.0). This observation is consistent across a wide range of calcium based anthropogenic sites of different origins and ages. Under the prevailing geochemical conditions created at these pH values the acetoclastic methanogenic pathway is unable to access its substrate due to the absence of undissociated acetic acid at these pH values.

DISCUSSION

The data presented here suggest that an alkaline pH >9.0 results in a methanogen community that is dominated by hydrogenotrophs, with populations enhanced by the activity of SAO bacteria. The methanogen populations from high pH and neutral sediments exposed to alkaline pH were both dominated by *Methanobacterium* sp., and in the case of the New Lime sediment communities, were capable of hydrogenotrophic methanogenesis at pH 11.0. Members of the Methanobacteriales have also been detected within a range of hyperalkaline environments resulting from serpentinisation processes (Brazelton et al., 2017; Crespo-Medina et al., 2017).

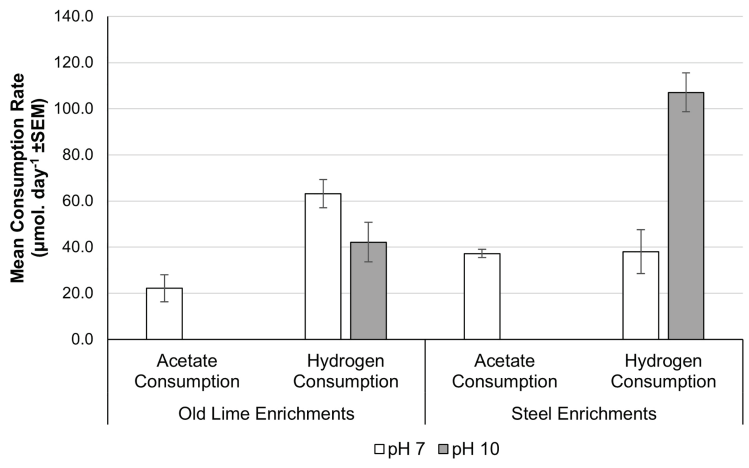

FIGURE 5 | Acetate and hydrogen consumption from microcosms prepared using Old Lime and Steel site sediments as an enrichment inoculum. The pattern observed in the Old Lime (n = 10) and Steel (n = 8) sediments mirrored that seen with the New Lime and Control sediments with acetoclastic methanogenesis absent at alkaline pH ("n" denotes the number of replicates analyzed).

A number of alkaliphilic *Methanobacterium* sp. have been isolated to date (Boone et al., 1986; Mathrani et al., 1988; Kotelnikova et al., 1998), the observations made here suggest that this genus is adaptable to aggressive pH changes as observed with Control sediment enrichments, but is also capable of surviving high pH after long-term exposure as residents of the initial New Lime sediments. This is in contrast to the Methanomicrobiales such as *Methanocalculus* sp. which are more prevalent in the hypersaline-hyperalkaline Soda lakes (Surakasi et al., 2007; Antony et al., 2013; Sorokin et al., 2015), due to the low salt tolerance of *Methanobacterium* sp. (Boone et al., 1986). In this study, *Methanosarcina* sp. were most impacted in the control sediments exposed to high pH in acetate enrichments. Pure culture studies of Lonar lake *Methanosarcina* indicated a maximum pH of 9.5 for growth upon acetate (Thakker and Ranade, 2002), which suggests that this genus may require a saline rather than a calcium dominated environment to sustain methane generation above pH 9.0.

These observations indicate that the generation of methane through the acetoclastic pathway, although energetically favorable with respect to bicarbonate (Jin and Kirk, 2018) does not proceed under alkaline conditions (>pH 9.0). High pH favors the dissociation of acetic acid to its anion ($CH3COO^-$), preventing transmembrane diffusion (Kröninger et al., 2014). Under alkaline conditions, acetate transport into the acetoclastic methanogen cell is therefore reliant on an acetate transporter, which is likely to be less energetically favorable than hydrogenotrophic methanogenesis (Kröninger et al., 2014). Much of the current understanding of methanogen ecology under alkaline conditions has been focused upon the microbiology of hypersaline, hyperalkaline Soda lakes, where methanogenesis is most commonly associated with the metabolism of C-1 compounds being released upon the biodegradation of Cyanobacterial mats (Jones et al., 1998; Sorokin et al., 2017). The current study indicates that methanogenesis in non-saline, calcium dominated alkaline environments should be considered as a different ecological niche. Within these environments, hydrogenotrophic methanogenesis is the most prevalent methanogenic pathway above pH 9.0. This observation is supported by the investigations of sediments from 13 individual sites ranging in age, origin and chemical composition. These observations suggest that under the alkaline conditions generated within a cement-based GDF, it is the hydrogenotrophic methane generation pathway that will be active rather than acetoclastic methanogenesis.

DATA AVAILABILITY STATEMENT

The data is deposited with the NCBI under Bioproject accession number PRJNA525260.

AUTHOR CONTRIBUTIONS

PH, SR, WM, and HG contributed to conception and design of the study. RW carried out the experimental work. PH, WM, HG, RW, and SR participated in the field work. SR, PH, and RW wrote the first draft. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

Richard Wormald's PhD was supported by a Radioactive Waste Management Ltd. (RWM) research bursary. RWM Ltd. is a wholly own subsidiary of the UK Nuclear Decommissioning Authority, a non-departmental public body which reports to the UK Department for Business, Energy, and Industrial Strategy. WM and HG were supported by the NERC R3AW grant (NE/L014211/1).

ACKNOWLEDGMENTS

We would like to thank the following individuals and organizations. David Johnson for his assistance regarding the field kiln sites investigated. The Canal and River Boat Trust, Hanson Aggregates, Tarmac plc, Tee Valley Wildlife Trust and British Steel.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articless/10.3389/fmicb.2020.614227/full#supplementary-material

REFERENCES

Agneessens, L. M., Ottosen, L. D. M., Andersen, M., Olesen, C. B., Feilberg, A., and Kofoed, M. V. W. (2018). Parameters affecting acetate concentrations during in-situ biological hydrogen methanation. *Bioresour. Technol.* 258, 33–40. doi: 10.1016/j.biortech.2018.02.102

Almond, M., Shaw, P. B., Humphreys, P. N., Chadha, M. J., Niemelä, K., and Laws, A. P. (2012). Behaviour of xyloisosaccharinic acid and xyloisosaccharino-1, 4-lactone in aqueous solutions at varying pHs. *Carbohydr. Res.* 363, 51–57. doi: 10.1016/j.carres.2012.09.019

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

Antony, C. P., Kumaresan, D., Hunger, S., Drake, H. L., Murrell, J. C., and Shouche, Y. S. (2013). Microbiology of Lonar Lake and other soda lakes. *ISME* 7, 468–476. doi: 10.1038/ismej.2012.137

Avis, J., Calder, N., Humphreys, P., King, F., Suckling, P., and Walsh, R. (2011). *Postclosure Safety Assessment: Gas Modelling*. Ontario, Canada: Nuclear Waste Management Organization.

B.S.I (2005). BS ISO 14853:2005 Plastics-Determination of the ultimate anaerobic biodegradation of plastic materials in an aqueous system- Method by measurement of biogas production. British Standards Institute, London, UK.

Bagnoud, A., Chourey, K., Hettich, R. L., de Bruijn, I., Andersson, A. F., Leupin, O. X., et al. (2016). Reconstructing a hydrogen-driven microbial metabolic network in Opalinus Clay rock. *Nat. Commun.* 7:12770. doi: 10.1038/ncomms12770

Balk, M., Weijma, J., and Stams, A. J. (2002). *Thermotoga lettingae* sp. nov., a novel thermophilic, methanol-degrading bacterium isolated from a thermophilic

anaerobic reactor. *Int. J. Syst. Evol. Microbiol.* 52, 1361–1368. doi: 10.1099/00207713-52-4-1361

Bassil, N. M., Bryan, N., and Lloyd, J. R. (2014). Microbial degradation of isosaccharinic acid at high pH. *ISME* 9, 310–320. doi: 10.1038/ismej.2014.125

Boone, D. R., Worakit, S., Mathrani, I. M., and Mah, R. A. (1986). Alkaliphilic methanogens from high-pH lake sediments. *Syst. Appl. Microbiol.* 7, 230–234. doi: 10.1016/S0723-2020(86)80011-X

Brazelton, W. J., Thornton, C. N., Hyer, A., Twing, K. I., Longino, A. A., Lang, S. Q., et al. (2017). Metagenomic identification of active methanogens and methanotrophs in serpentinite springs of the Voltri Massif Italy. *PeerJ* 5:e2945. doi: 10.7717/peerj.2945

Burke, I., Mortimer, R., Palani, S., Whittleston, R., Lockwood, C., Ashley, D., et al. (2012). Biogeochemical reduction processes in a hyper-alkaline affected leachate soil profile. *Geomicrobiol J.* 29, 769–779. doi: 10.1080/01490451.2011.619638

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

Charles, C. J., Rout, S. P., Garratt, E. J., Patel, K., Laws, A. P., and Humphreys, P. N. (2015). The enrichment of an alkaliphilic biofilm consortia capable of the anaerobic degradation of isosaccharinic acid from cellulosic materials incubated within an anthropogenic, hyperalkaline environment. *FEMS Microbiol. Ecol.* 91:fiv085. doi: 10.1093/femsec/fiv085

Charles, C., Rout, S., Patel, K., Akbar, S., Laws, A., Jackson, B., et al. (2017). Floc formation reduces the pH stress experienced by microorganisms living in alkaline environments. *Appl. Environ. Microbiol.* 83, e02985–e02916. doi: 10.1128/AEM.02985-16

Charles, C. J., Rout, S. P., Wormald, R., Laws, A. P., Jackson, B. R., Boxall, S. A., et al. (2019). In-Situ Biofilm Formation in Hyper Alkaline Environments. *Geomicrobiol J.* 36, 404–411. doi: 10.1080/01490451.2018.1564803

Crespo-Medina, M., Twing, K. I., Sánchez-Murillo, R., Brazelton, W. J., McCollom, T. M., and Schrenk, M. O. (2017). Methane dynamics in a tropical serpentinizing environment: the Santa Elena Ophiolite, Costa Rica. *Front. Microbiol.* 8:916. doi: 10.3389/fmicb.2017.00916

Daebeler, A., Gansen, M., and Frenzel, P. (2013). Methyl fluoride affects methanogenesis rather than community composition of methanogenic archaea in a rice field soil. *PLoS One* 8:e53656. doi: 10.1371/journal.pone.0053656

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

Faybishenko, B., Birkholzer, J., Sassani, D., and Swift, P. (2016). International Approaches for Nuclear Waste Disposal in Geological Formations: Geological Challenges in Radioactive Waste Isolation—Fifth Worldwide Review. Lawrence Berkeley National Lab, Berkeley, CA United States.

Felicione, F. S., Carney, K. P., Dwight, C. C., Cummings, D. G., and Foulkrod, L. E. (2003). "Gas-generation experiments for long-term storage of TRU wastes at WIPP" in *Proceedings of WM03*; February 23–27, 2003; Tucson, Arizona, United States.

Hattori, S. (2008). Syntrophic acetate-oxidizing microbes in methanogenic environments. *Microbes Environ.* 23, 118–127. doi: 10.1264/jsme2.23.118

Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* 50, 1601–1609. doi: 10.1099/00207713-50-4-1601

Humphreys, P. N., Laws, A., and Dawson, J. (2010). A Review of Cellulose Degradation and the Fate of Degradation Products Under Repository Conditions. Serco Contractors Report for the Nuclear Decommissioning authority, Harwell, UK.

IAEA (2011). *Geological Disopsal of Radioactive Waste*. Vienna, Austria: International Atomic Energy Agency.

Jin, Q., and Kirk, M. F. (2018). pH as a primary control in environmental microbiology: 1 thermodynamic perspective. *Front. Environ. Sci.* 6:21. doi: 10.3389/fenvs.2018.00021

Johnson, D. (2008). The archaeology and technology of early-modern lime burning in the yorkshire dales: developing a clamp kiln model. *Ind. Archaeol. Rev.* 30, 127–143. doi: 10.1179/174581908X347346

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

Kotelnikova, S., Macario, A. J., and Pedersen, K. (1998). *Methanobacterium subterraneum* sp. nov., a new alkaliphilic, eurythermic and halotolerant methanogen isolated from deep granitic groundwater. *Int. J. Syst. Bacteriol.* 48, 357–367. doi: 10.1099/00207713-48-2-357

Kröninger, L., Deppenmeier, U., and Welte, C. (2014). Experimental evidence of an acetate transporter protein and characterization of acetate activation in aceticlastic methanogenesis of *Methanosarcina mazei*. *FEMS Microbiol. Lett.* 359, 147–153. doi: 10.1111/1574-6968.12550

Kyeremeh, I. A., Charles, C. J., Rout, S. P., Laws, A. P., and Humphreys, P. N. (2016). Microbial community evolution is significantly impacted by the use of calcium isosaccharinic acid as an analogue for the products of alkaline cellulose degradation. *PLoS One* 11:e0165832. doi: 10.1371/journal.pone.0165832

Libert, M., Bildstein, O., Esnault, L., Jullien, M., and Sellier, R. (2011). Molecular hydrogen: an abundant energy source for bacterial activity in nuclear waste repositories. *Phys. Chem. Earth* 36, 1616–1623. doi: 10.1016/j.pce.2011.10.010

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019

Manzoor, S., Schnürer, A., Bongcam-Rudloff, E., and Müller, B. (2018). Genome-guided analysis of Clostridium ultunense and comparative genomics reveal different strategies for acetate oxidation and energy conservation in syntrophic acetate-oxidising bacteria. *Genes* 9:225. doi: 10.3390/genes9040225

Mathrani, I. M., Boone, D. R., Mah, R. A., Fox, G. E., and Lau, P. P. (1988). *Methanohalophilus zhilinae* sp. nov., an alkaliphilic, halophilic, methylotrophic methanogen. *Int. J. Syst. Bacteriol.* 38, 139–142. doi: 10.1099/00207713-38-2-139

Mayes, W. M., Younger, P. L., and Aumônier, J. (2008). Hydrogeochemistry of alkaline steel slag leachates in the UK. *Water Air Soil Pollut.* 195, 35–50. doi: 10.1007/s11270-008-9725-9

Milodowski, A. E., Shaw, R. P., and Stewart, D. I. (2013). Harpur Hill Site: its geology, evolutionary history and a catalogue of materials present. British Geological Survey, Keyworth, Nottingham, UK.

NDA (2010). *Gas Status Report*. Harwell, UK: Nuclear Decommissioning Authority.

NEA (2014). *Updating the NEA International FEP List: Technical Note 1: Identification and Review of Recent Project specific FEP Lists*. Paris, France: Nuclear Energy Agency.

Poller, A., Mayer, G., Darcis, M., and Smith, P. (2016). *Modelling of Gas Generation in Deep Geological Repositories after Closure*. Wettingen, Switzerland: NAGRA.

Rout, S. P., Charles, C. J., Doulgeris, C., McCarthy, A. J., Rooks, D. J., Loughnane, J. P., et al. (2015a). Anoxic biodegradation of isosaccharinic acids at alkaline pH by Natural Microbial Communities. *PLoS One* 10:e0137682. doi: 10.1371/journal.pone.0137682

Rout, S. P., Charles, C. J., Garratt, E. J., Laws, A. P., Gunn, J., and Humphreys, P. N. (2015b). Evidence of the generation of isosaccharinic acids and their subsequent degradation by local microbial consortia within hyper-alkaline contaminated soils, with relevance to intermediate level radioactive waste disposal. *PLoS One* 10:e0119164. doi: 10.1371/journal.pone.0119164

Rout, S. P., Radford, J., Laws, A. P., Sweeney, F., Elmekawy, A., Gillie, L. J., et al. (2014). Biodegradation of the alkaline cellulose degradation products generated during radioactive waste disposal. *PLoS One* 9:e107433. doi: 10.1371/journal.pone.0107433

Sorokin, D. Y., Abbas, B., Geleijnse, M., Pimenov, N. V., Sukhacheva, M. V., and van Loosdrecht, M. (2015). Methanogenesis at extremely haloalkaline conditions in the soda lakes of Kulunda Steppe (Altai, Russia). *FEMS Microbiol. Ecol.* 91:fiv016. doi: 10.1093/femsec/fiv016

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.* 2:17081. doi: 10.1038/nmicrobiol.2017.81

Summerling, T. (2013). *Assessment of Carbon-14 Bearing Gas*. Cumbria, UK: LLW Repository Ltd.

Surakasi, V. P., Wani, A. A., Shouche, Y. S., and Ranade, D. R. (2007). Phylogenetic analysis of methanogenic enrichment cultures obtained from Lonar Lake in India: isolation of *Methanocalculus* sp. and *Methanoculleus* sp. *Microb. Ecol.* 54, 697–704. doi: 10.1007/s00248-007-9228-z

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS One* 9:e105592. doi: 10.1371/journal.pone.0105592

Thakker, C. D., and Ranade, D. R. (2002). An alkalophilic *Methanosarcina* isolated from Lonar crater. *Curr. Sci.* 82, 455–458.

Westerholm, M., Roos, S., and Schnürer, A. (2010). *Syntrophaceticus schinkii* gen. nov., sp. nov., an anaerobic, syntrophic acetate-oxidizing bacterium isolated from a mesophilic anaerobic filter. *FEMS Microbiol. Lett.* 309, 100–104. doi: 10.1111/j.1574-6968.2010.02023.x

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. RWM Ltd. had no influence on the design or execution of this study or the interpretation and reporting of the data reported in this paper.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 09 December 2020
doi: 10.3389/fmicb.2020.554927

Isolation and Taxonomic Characterization of Novel Haloarchaeal Isolates From Indian Solar Saltern: A Brief Review on Distribution of Bacteriorhodopsins and V-Type ATPases in Haloarchaea

Dipesh Kumar Verma[1], *Chetna Chaudhary*[1], *Latika Singh*[1], *Chandni Sidhu*[2], *Busi Siddhardha*[3], *Senthil E. Prasad*[4]* *and Krishan Gopal Thakur*[1]*

[1] *Structural Biology Laboratory, G. N. Ramachandran Protein Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India,* [2] *MTCC-Microbial Type Culture Collection & Gene Bank, Council of Scientific and Industrial Research Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India,* [3] *Department of Microbiology, School of Life Sciences, Pondicherry University, Puducherry, India,* [4] *Biochemical Engineering Research and Process Development Centre, Council of Scientific and Industrial Research-Institute of Microbial Technology (CSIR-IMTECH), Chandigarh, India*

OPEN ACCESS

Edited by:
Anna-Louise Reysenbach, Portland State University, United States

Reviewed by:
Mohea Couturier, University of Illinois at Urbana–Champaign, United States
Likui Zhang, Yangzhou University, China

***Correspondence:**
Krishan Gopal Thakur
krishang@imtech.res.in
Senthil E. Prasad
esprasad@imtech.res.in

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

Received: *23 April 2020*
Accepted: *17 September 2020*
Published: *09 December 2020*

Citation:
Verma DK, Chaudhary C, Singh L, Sidhu C, Siddhardha B, Prasad SE and Thakur KG (2020) Isolation and Taxonomic Characterization of Novel Haloarchaeal Isolates From Indian Solar Saltern: A Brief Review on Distribution of Bacteriorhodopsins and V-Type ATPases in Haloarchaea. Front. Microbiol. 11:554927. doi: 10.3389/fmicb.2020.554927

Haloarchaea inhabit high salinity environments worldwide. They are a potentially rich source of crucial biomolecules like carotenoids and industrially useful proteins. However, diversity in haloarchaea present in Indian high salinity environments is poorly studied. In the present study, we isolated 12 haloarchaeal strains from hypersaline Kottakuppam, Tamil Nadu solar saltern in India. 16S rRNA based taxonomic characterization of these isolates suggested that nine of them are novel strains that belong to genera *Haloarcula*, *Halomicrobium,* and *Haloferax*. Transmission electron microscopy suggests the polymorphic nature of these haloarchaeal isolates. Most of the haloarchaeal species are known to be high producers of carotenoids. We were able to isolate carotenoids from all these 12 isolates. The UV-Vis spectroscopy-based analysis suggests that bacterioruberin and lycopene are the major carotenoids produced by these isolates. Based on the visual inspection of the purified carotenoids, the isolates were classified into two broad categories i.e., yellow and orange, attributed to the differences in the ratio of bacterioruberin and lycopene as confirmed by the UV-Vis spectral analysis. Using a PCR-based screening assay, we were able to detect the presence of the bacteriorhodopsin gene (*bop*) in 11 isolates. We performed whole-genome sequencing for three *bop* positive and one *bop* negative haloarchaeal isolates. Whole-genome sequencing, followed by pan-genome analysis identified multiple unique genes involved in various biological functions. We also successfully cloned, expressed, and purified functional recombinant bacteriorhodopsin (BR) from one of the isolates using *Escherichia coli* as an expression host. BR has light-driven proton pumping activity resulting in the proton gradient across the membrane, which is utilized by V-Type ATPases to produce ATP. We analyzed the distribution of *bop* and other accessory genes involved in functional BR expression and ATP synthesis in all the representative haloarchaeal

species. Our bioinformatics-based analysis of all the sequenced members of genus *Haloarcula* suggests that *bop*, if present, is usually inserted between the genes coding for B and D subunits of the V-type ATPases operon. This study provides new insights into the genomic variations in haloarchaea and reports expression of new BR variant having good expression in functional form in *E. coli*.

Keywords: haloarchaea, bacteriorhodopsin, pangenome, carotenoids, taxonomy

INTRODUCTION

A group of microbes called extremophiles can grow, adapt, and survive harsh conditions like high salinity, high or low temperature, and acidic or alkaline conditions. Haloarchaea are extremophiles that grow in the hypersaline environments such as the natural brine, Dead Sea, alkaline salt lakes, marine solar salterns, and rock salt deposits (Cayol et al., 1994; Purdy et al., 2004; Gramain et al., 2011; Stan-Lotter and Fendrihan, 2015). Besides high salinity, haloarchaea are also exposed to very stringent conditions such as high temperature, UV radiations, high ionic stresses, and alkaline pH (Bowers and Wiegel, 2011; Stan-Lotter and Fendrihan, 2015). These microbes express specialized proteins and also produce metabolites like carotenoids that aid in adaptation, survival, and growth in such harsh environmental conditions (Giani et al., 2019). These metabolites and proteins have characteristics suitable for various industrial or research applications (Littlechild, 2015; Cabrera and Blamey, 2018; Verma et al., 2020). Haloarchaea produces carotenoids in high amounts that act as antioxidants, light protection pigments, and membrane stabilizers (Rodrigo-Banos et al., 2015; Giani et al., 2019). Bacterioruberin, a haloarchaeal carotenoid, reportedly has more free radical scavenging activity compared to plant β-carotenes, being used in various food and cosmetic products (Yatsunami et al., 2014; Higa et al., 2020).

With the depleting natural resources, different non-conventional biological materials are being explored to accomplish future energy requirements. Microorganisms such as haloarchaea with their unique features have attracted the attention of researchers to investigate them for the generation of energy (Nadella and Hernandez Baltazar, 2018). These microbes code for bacteriorhodopsin protein (BR), which has a unique property of light-driven proton-pumping activity and hence finds uses in several applications, including solar cells, optical filters, hydrogen production, biosensors, optogenetics, and memory storage devices (Bogomolni et al., 1976; Saeedi et al., 2012; Li et al., 2018). BR can be potentially used to harvest abundant solar energy to produce electricity or in the photolysis of water to generate hydrogen fuel to meet future ever-increasing energy demands (Sediroglu et al., 1998). However, the production cost of BR is one of the major bottlenecks in commercializing such innovative technologies. There are few recombinant methods available that successfully enhanced the yield of BR expression using *Escherichia coli* as an expression host (Kahaki et al., 2014; Bratanov et al., 2015; Jeganathan et al., 2019; Mirfeizollahi et al., 2019). Also, many attempts have already been made to utilize recombinant BR but until now no commercial applications have been reported. Therefore, it becomes essential to explore different saltern environments for studying new BR molecules which could essentially be utilized in biophotonics and bioelectronics applications.

There are limited studies describing microbial biodiversity in the Indian solar salterns. In one study, isolation and characterization of the *Haloarcula marismortui* RR12 strain was reported from Mumbai solar saltern, India (Thombre et al., 2016). In another study, BR purification was reported from *Haloferax larsenii* RG3D.1 strain isolated from Rocky Beach of Malvan, West Coast of India (Kanekar et al., 2015). Multiple haloarchaeal strains from Goa and Mulund solar salterns have also been reported (Rajurkar and Pathak, 2014). Our group has been working on isolation of extreme haloarchaeal strains from solar salterns to study the diversity and to identify novel BR sequences, if any, from the isolates. We reported the isolation of extremely halophilic archaea *Halogeometricum borinquense* strain wsp3 and *Haloferax volcanii* strain wsp5 from Marakkanam solar salterns Pondicherry and *Haloarcul*a strain $K1^T$ from Thamaraikulam solar salterns Kanyakumari (Verma et al., 2019). In recent studies, our group has characterized several haloarchaeal strains isolated from Indian solar salterns.

In this study, we have isolated 12 haloarchaeal strains from the Kottakuppam Solar saltern, East Coast Road from Chennai to Pondicherry, Villupuram district, Tamil Nadu. The 16S rRNA gene sequencing suggested that nine of them were novel strains. We screened the strains for the presence of the bacteriorhodopsin gene (*bop*) and successfully cloned, expressed, and purified one of the recombinant BR using *E. coli* as an expression host. We also studied the genetic organization of *bop* and accessory proteins involved in the light-driven ATP synthesis from haloarchaea. We purified carotenoids from all 12 strains and performed absorption-based biophysical characterization. We further performed the comparative genomics analysis of all three *bop* positive strains, which revealed unique genes and industrially important enzymes encoded in their genomes.

RESULTS

Isolation, Taxonomic Characterization, and *bop* Screening

All pws strains were isolated from different crystallizer ponds of Kottakuppam solar saltern near Puducherry, Tamil Nadu India. The different sampling locations in the crystallizer ponds yielded different haloarchaeal isolates designated as pws 1 to 12. The single colonies of pure cultures were isolated after performing

3–4 subculturing steps. Genomic DNA was isolated from the purified cultures grown in liquid broth followed by 16S rRNA amplification. The partially amplified 16S rRNA (size ranges from 500 bp to 1,200 bp) sequences were used for multiple sequence alignment. The 16S rRNA sequence similarity scores for all twelve isolates were close to 99% with the closest reference strains for pws1, pws3, pws5, pws6, pws7, pws8, pws9, pws11, and pws12 suggesting that all these nine isolates were novel strains (**Table 1**). EzTaxon analysis suggested that strains pws1, pws3, pws7, pws10, and pws12 belong to genus *Halomicrobium*, pws2, pws4, pws5, pws6, pws8, and pws9 belong to genus *Haloarcula* and pws11 belongs to genus *Haloferax* (**Table 1**). The partial 16S rRNA sequences for pws2, pws4, and pws10 shared 100% identity score with *Haloarcula salaria* JCM 15759, *Haloarcula japonica* JCM7785 and *Halomicrobium mukohataei* JP60, respectively (**Table 1**). All the isolates were further screened for the presence of the *bop* using degenerate primers (DegF and DegR) as reported earlier (Verma et al., 2019) (**Supplementary Table 1**). The presence of an expected 450 bp PCR product on an agarose gel confirmed that all the strains except pws11 were positive for *bop* (**Supplementary Figure S1**).

Morphological Characterization of pws Strains

The Halobacteriaceae family members show extreme polymorphism, ranging from rods, pleomorphic rods, disc-shaped, cocci, square, and triangular forms (Fendrihan et al., 2006). The morphology feature of non-coccoid haloarchaea dependents on the salt concentration of the environment, and with decreasing salt concentration, different shapes like swollen, club-shaped, and bent rods appear (Mohr and Larsen, 1963). To visualize the cellular morphology of pws isolates, we performed transmission electron microscopy experiments (TEM). The TEM images of the isolates suggest that all strains are polymorphic, with size ranging from 0.5 to 3 μm (**Figure 1**). For strains pws1, pws3, and pws10, we could observe prominent rod-like morphology. TEM images of pws8 suggested that it is highly vacuolated (**Figure 1**). Several bacteria synthesize gas vesicles, and also few haloarchaea can produce these flotation devices (Walsby, 1994). These gas vacuoles are filled by diffusion with environmental gases dissolved in the water. The functional role of these gas vacuoles is to provide buoyancy, enabling cells to maintain their depth in the aqueous environment. Besides polymorphic morphology, we also observed some interesting features in TEM images shown in **Supplementary Figure S2**.

Carotenoid Isolation and Spectroscopic Analysis

Haloarchaea are known to be one of the richest sources of carotenoids (Giani et al., 2019). The major component of the haloarchaeal carotenoid pool is bacterioruberin (Yatsunami et al., 2014). We isolated carotenoids from all twelve strains as described in earlier methods (Yang et al., 2015) (**Figure 2A**). Isolated carotenoids were grossly grouped into two distinct types, i.e., yellow and orange (**Figure 2A**). UV-Vis spectra of the isolated carotenoids also show differences in their composition (**Figures 2B,C**). In the orange samples (**Figure 2B**), the ratio of peak1/peak 3 is close to 1, while in the pale-yellow samples (**Figure 2C**), the peak1/peak3 ratio is <0.8. In standard carotenoid UV-Vis spectra, peak1 (515–522 nm) corresponds to bacterioruberin absorption; peak2 represents all-trans-lycopene absorption, and peak3 represents 13-cis-lycopene absorption (Yatsunami et al., 2014). A minor absorption of bacterioruberin is also reported at 466 nm.

Both bacterioruberin and lycopene are connected through a single pathway where lycopene is converted into bacterioruberin by lycopene cyclase. Therefore, in the second profile, the low score of peak1/peak3 ratio may suggest low bacterioruberin content (peak1) and high lycopene (peak3) accumulation in pws6, pws9, pws2, and pws5 (**Figure 2C**).

The Pangenome Analysis of pws Strains

Solar salterns are not a very rich source of nutrients. Besides this, the extreme living conditions, including near saturation concentration of salts, exposure to UV light, and elevated temperature, require a set of genes to aid adaptation and survival under these harsh conditions. So, to understand the genetic diversity in haloarchaea, we performed pangenome analysis. The pangenome defines the total gene pool of a particular set of genomes. Pangenome analysis divides gene pools into three different sets: core, accessory, and unique genes based on their presence in the single or multiple organisms (Medini et al., 2005; Tettelin et al., 2005; Vernikos et al., 2015). Pangenome analysis suggests either an open or closed pangenome in given genera. Species with an open pangenome have multiple new genes added *per* sequenced genome, and hence it becomes challenging to predict the full pan-genome (Tettelin et al., 2005). On the other hand, in a closed pangenome, only limited new genes are added with the addition of a new genome. Therefore, the theoretical size of the pangenome can be calculated (Vernikos et al., 2015). The haloarchaeal isolates obtained in this study belong to three genera i.e., *Haloarcula*, *Haloferax*, and *Halomicrobium*. Therefore, we randomly picked one representative strain belonging to genera *Haloferax*, and *Halomicrobium* for whole-genome sequencing. Members of the genus *Haloarcula* are known to harbor *bop*, so we selected two isolates (pws5 and pws8) for whole-genome sequencing. The phylogenetic tree based on 16S rRNA sequences extracted from whole-genomes confirms that pws1, pws5, pws8, and pws11 share 99% sequence identity with *Halomicrobium mukohataei* DSM 12286, *Haloarcula argentinensis DSM 12282*, *Haloarcula vallismortis* ATCC 29715 and *Haloferax volcanii* DS2, respectively (**Figure 3**). These results were further confirmed by calculating average nucleotide index (ANI) and digital DNA-DNA hybridization scores where the observed values were higher than the accepted cut-off values (for ANI >95% and DNA-DNA hybridization >91%) for novel species. Hence, all four isolates are novel strains and not novel species.

The draft genome sequences of pws1, pws5, pws8, and pws11 comprised of 3.39 Mb, 4.0 Mb, 3.9 Mb and 3.67 Mb with 3,722, 4,277, 4,150 and 4,021 annotated coding sequences (CDS). The GC content of pws1, pws5, pws8 and pws11 was 65.7, 61.3, 62.2, and 66.1%, respectively. Whole-genome sequencing data confirmed the presence of *bop* in pws1, pws5, and pws8, and no

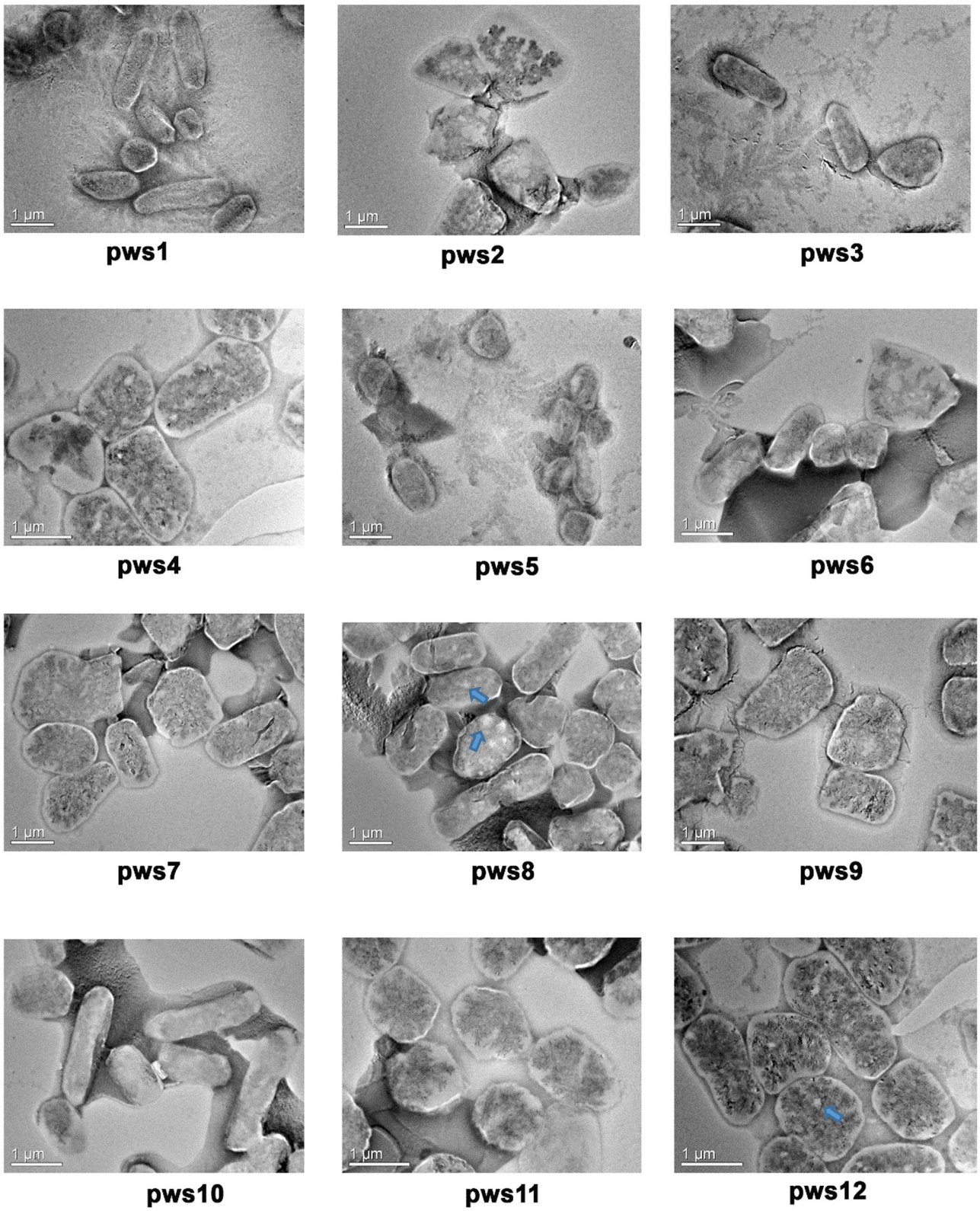

FIGURE 1 | Transmission electron microscopy images of pws isolates. TEM images reveal polymorphic morphology in the haloarchaeal isolates. Blue arrows indicate the presence of gas vacuoles.

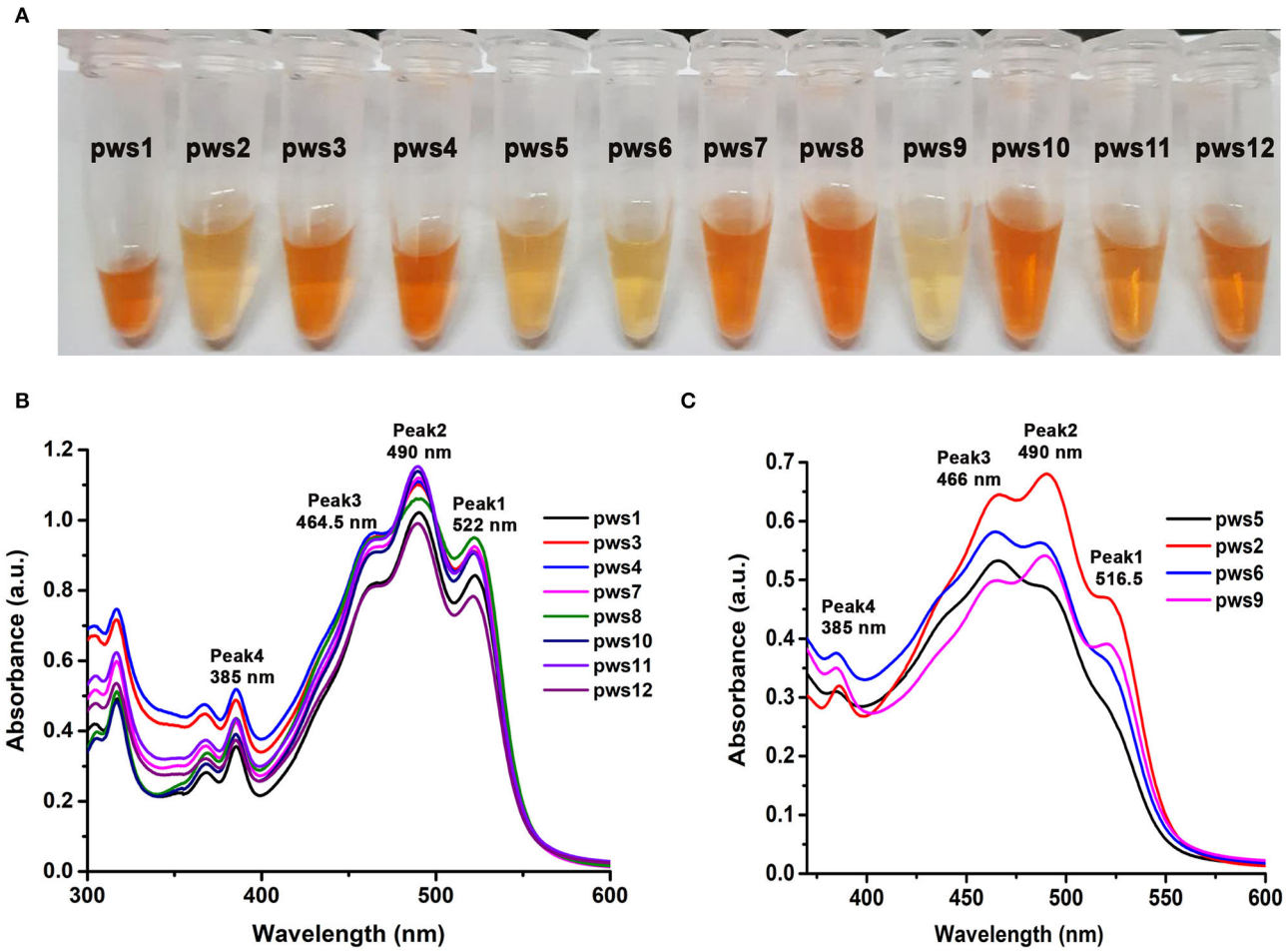

FIGURE 2 | Isolation and UV-Vis absorption spectrum based characterization of carotenoids isolated from pws strains. **(A)** Carotenoids isolated from the haloarchaeal isolates using acetone extraction method. **(B)** UV-Vis absorbance profile of yellowish colored carotenoids isolated from pws1, 3, 4, 7, 8, 10, 11, and 12 samples. **(C)** UV-Vis absorbance profile of orange colored carotenoids isolated from pws5, 2, 6, and 9. Peak 1 (522 nm), peak 2 (490 nm), peak 3 (465 nm), and peak 4 (385 nm) correspond to bacterioruberin, all-trans-lycopene, 13-cis-lycopene, and all trans-β-carotene, respectively.

bop was observed in the pws11 genome. This analysis revalidated our initial PCR-based screening results that were designed to identify isolates harboring *bop*. We selected *bop* harboring strains, pws1, pws5, and pws8, for further pangenome analysis (**Figure 4**). The results of pangenome are described under two following parts:

Pangenome Analysis of Genus *Halomicrobium*

Both 16S rRNA and whole genomic sequencing results confirmed that pws1 belongs to *Halomicrobium* genus and till now, only five different genomes are known in this genus. We used all available five genomes, including pws1 for pangenome analysis. The distribution of the archaeal clusters of orthologous groups (arCOG) plot of *Halomicrobium* genus (including pws1) suggested that pangenome and core genome contain 6,129 genes and 3,551 genes, respectively. The arCOG distribution plot also suggests that the pangenome is open and increasing with the addition of new genomes that contain several unique genes (**Figure 4A**). The frequency distribution plot suggests multiple unique and accessory genes insertion including genes coding for cell wall/membrane/envelop biogenesis, signal transduction, replication, recombination, repair, carbohydrate transport, inorganic ion transport, and metabolism-related genes (**Figure 4B**).

Pangenome Analysis of Genus *Haloarcula*

// To date, thirty different draft genomes are available under genus *Haloarcula*. We performed a pangenome analysis of pws5 and pws8 with all available thirty *Haloarcula* genomes. For the genus *Haloarcula*, the pangenome and core genome contain 13,782 genes and 690 genes, respectively. The pangenome analysis suggested that similar to *Halomicrobium*, *Haloarcula* (including pws5 and pws8) pangenome is also open and with the addition of new genomes that contain several unique genes (**Figure 4C**). They also have similarities in frequency

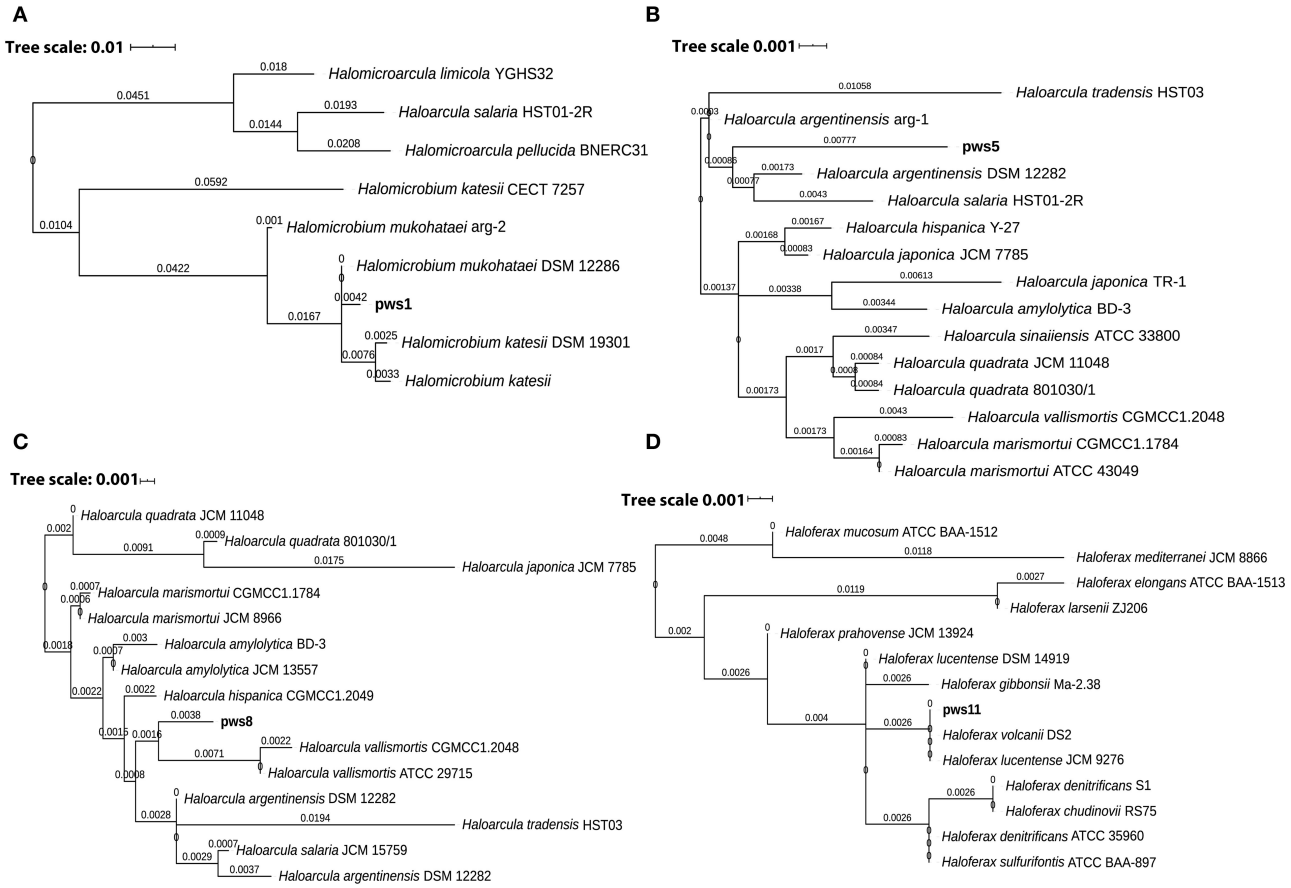

FIGURE 3 | The 16S rRNA sequence-based phylogenetic tree of the isolates and the closely related haloarchaeal species. Phylogenetic analysis suggests that isolates pws1 **(A),** pws5 **(B),** pws8 **(C)** and pws11 **(D)** are closely related to *H. mukohataei* DSM 12286, *H. argentinensis* DSM 12282, *H. vallismortis* ATCC 29715, and *H. volcanii* DS2, respectively.

distribution plots such as both genera have multiple insertions of genes coding for cell wall/membrane/envelop biogenesis, signal transduction, replication, recombination, repair, carbohydrate transport, inorganic ion transport, and metabolism-related genes. Additionally, in *Haloarcula* we found multiple other unique and accessory genes, including genes coding for post-translation modifications, chaperons, and cell motility (**Figure 4D**).

Expression and Purification of Recombinant Pws5-BR

The full-length sequence of *bop* was extracted from pws5 genome and used for designing gene-specific primers. The *bop* (750 bp) was amplified using gene-specific forward (Bop_full_F) and reverse primers (Bop_full_R). The protein sequence comparison shows that pws5-BR (locus tag -NLV14165.1) shares 92% sequence identity with *H. marismortui* BRI (HmBRI). Based on our previous study (Verma et al., 2019), we chose the pET22b expression vector to yield recombinant protein with C-terminal 6× His-tag for affinity purification. BR is a leaderless membrane protein and putting N-terminal affinity tag may interfere with the membrane localization and protein folding. Protein expression in the presence of trans-retinal yielded red-colored cell pellet suggesting proper membrane integration and functional expression of BR (**Figure 5A**). The colored cell pellet was further dissolved into DDM (n-Dodecyl-β-D-Maltoside) detergent for BR solubilization. Solubilized protein was then further passed through Ni-NTA beads. Ni-NTA based affinity purification yielded reddish colored protein corresponding to the expected size of ∼26 kDa on the 15% SDS-PAGE (**Figures 5B,C**). UV-Vis spectrum profile gave a characteristic absorbance peak at 549 nm corresponding to retinal bound-BR (**Figure 5D**). We achieved about 1 mg per liter yield of pws5-BR.

Light-Driven Proton Pumping Activity of the Recombinant Pws5-BR

To check the functional property of light-driven proton activity in the recombinant pws-BR, we used whole cell-based assay. Briefly, the pws5-BR expressing C43-Rosetta BL21 (DE3) *E. coli* cells and control *E. coli* C43-Rosetta BL21

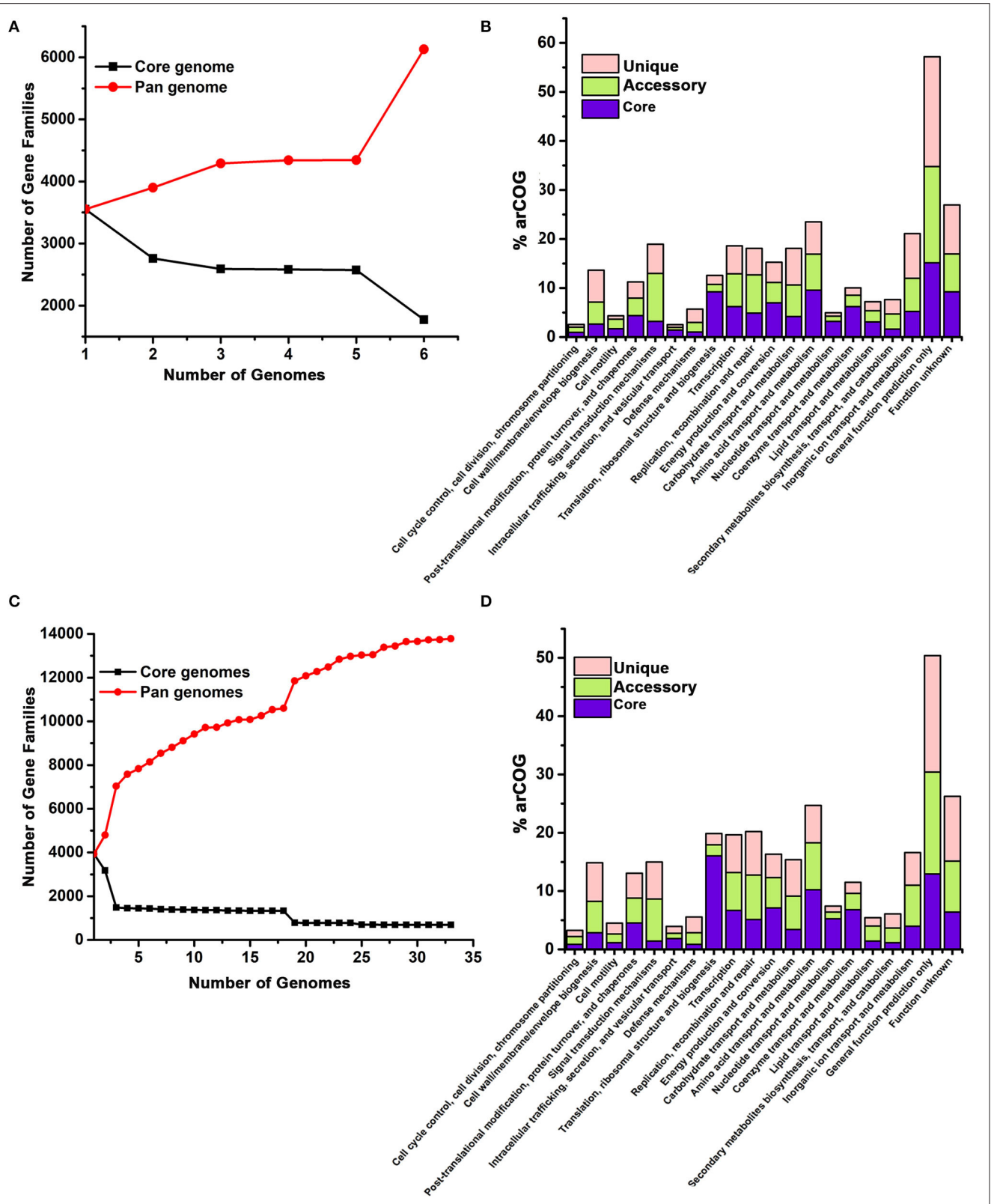

FIGURE 4 | Pangenome analysis of *Halomicrobium* genus **(A,B)** including pws1 strain and *Haloarcula* genus **(C,D)** including pws5 and pws8 strains using the BPGA pipeline. The genome sequences of *Halomicrobium* and *Haloarcula* members were retrieved from NCBI and were further analyzed by the BPGA tool. In both the panels, **(A)** and **(C)** are showing core genome (black) vs. pangenome (red) curves, while panel **(B)** and **(D)** are showing the frequency distribution of core (blue), accessory (olive) and unique genes (pink) within genomes.

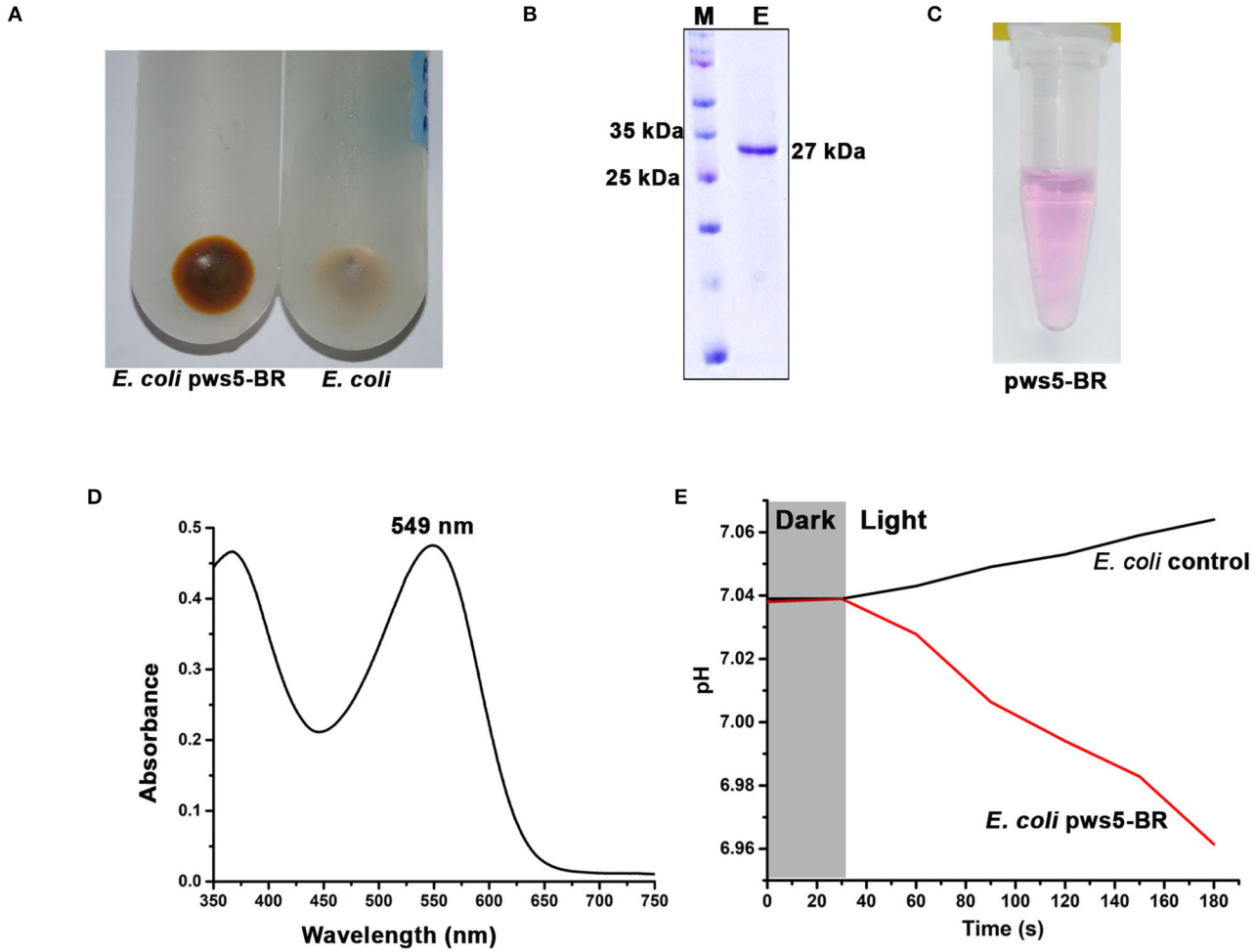

FIGURE 5 | The expression, purification and functional characterization of recombinant pws5-BR. **(A)** The presence of reddish cell pellet suggested the expression of functional pws5-BR. **(B)** The 15% SDS-PAGE showing ∼27 kDa purified protein band corresponding to the expected molecular weight of pws5-BR. **(C)** The purple color observed in the purified pws5-BR **(D)** UV-Vis spectra of the purified protein shows 549 nm characteristic BR absorption peak. **(E)** pws5-BR has light-driven proton pumping activity. In pws5-BR expressing cells, pH dropped after exposure to light while there was a marginal increase in pH in control *E. coli* cells suggesting proton pumping activity of pws5-BR.

(DE3) cells (without *pws5* positive clones) were induced with IPTG in the presence of 10 μM retinal. The cells were pelleted, washed, and resuspended in a non-buffered solution, as described in the materials and methods section. The non-buffered solution was used to detect small changes in pH generated by proton pumping activity, which will be otherwise masked by the buffer. The pH probe was dipped in a glass vial having cell suspension and incubated under dark conditions to measure changes in the pH. Under dark conditions, the pH was stable, but when white light was switched on, there was a decrease in the pH compared with the *E. coli* control cells (**Figure 5E**). This observation is in line with previously published reports (Wang et al., 2003; Kanehara et al., 2017; Verma et al., 2019). The light-induced pH shift was observed due to the proton-pumping activity of the pws5-BR.

Genetic Organization of *bop* in Haloarchaea

For the functional expression of BR and proton-driven ATP generation, several genes are required, including genes involved in retinal biosynthesis, transcription regulation, and V-type ATPases (Sharma et al., 2007). To study the genetic organization of *bop*, we performed analysis of the neighboring genes. In all three isolates, *bop* is located between B and D subunit of V-type ATPases similar to that observed in HmBRI. In the pws1 genome, we additionally found two accessory genes named bacterio-opsin-related protein (*brp*) or GAF domain-containing protein and lycopene cyclase (*crtY*) along with *bop*. In contrast, no additional genes were present in pws5 and pws8 (**Figure 6A**). The *crtY* and *brp,* along with other enzymes, help in producing retinal from lycopene (Peck et al., 2001; Tarasov et al., 2008). Surprisingly, in pws11, we found only a bacterio-opsin activator

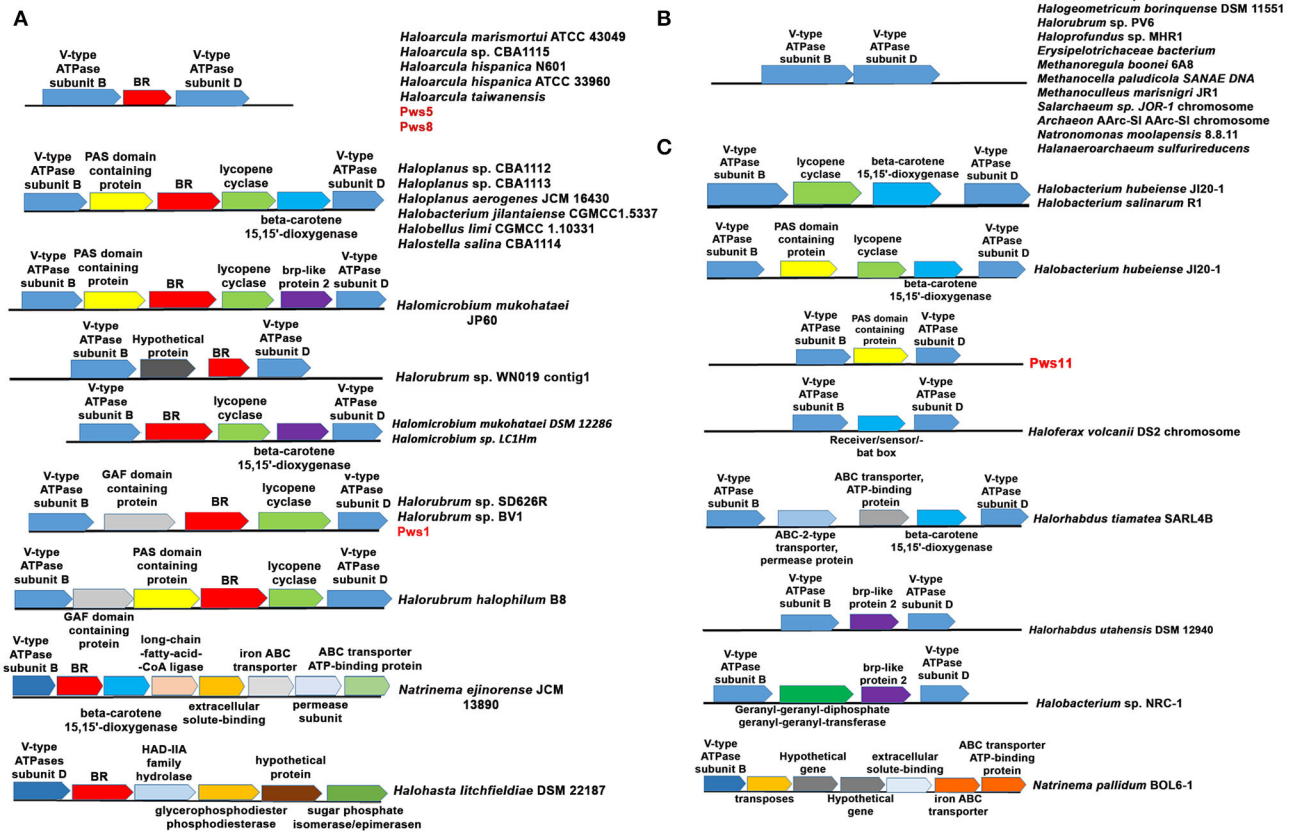

FIGURE 6 | Genetic architecture of *bop* operon in haloarchaea. **(A)** The haloarchaeal strains harboring *bop* and other genes located between the genes coding for B and D subunits of V-type ATPases operon. **(B)** Haloarchaeal members having no additional genes inserted between B and D subunits of V-type ATPases. **(C)** Haloarchaeal members having genes, excluding *bop*, inserted between B and D subunit of V-type ATPases.

TABLE 1 | 16S rRNA sequence similarity of pws isolates with the closest matches obtained using EzTaxon webserver.

S.No.	Isolate	Strain name (Accession number)	Similarity
1.	pws1	*Halomicrobium mukohataei* DSM 12286 (GCA_000023965.1)	< 99%
2.	pws2	*Haloarcula salaria* JCM 15759	100%
3.	pws3	*Halomicrobium mukohataei* DSM 12286 (GCA_000023965.1)	< 99%
4.	pws4	*Haloarcula japonica* JCM7785 (GCA_000336635.1)	100%
5.	pws5	*Haloarcula argentinensis DSM 12282* (GCA_000336895.1)	< 99%
6.	pws6	*Haloarcula salaria* JCM 15759	< 99%
7.	pws7	*Halomicrobium mukohataei* JP60 (GCA_004803735.1)	< 99%
8.	pws8	*Haloarcula vallismortis* ATCC 29715	< 99%
9.	pws9	*Haloarcula salaria* JCM 15759	< 99%
10.	pws10	*Halomicrobium mukohataei* JP60 (GCA_004803735.1)	100%
11.	pws11	*Haloferax volcanii* DS2 (GCA_007714225.1)	< 99%
12.	pws12	*Halomicrobium katesii* CECT 7257 (GCA_000379085.1)	< 99%

protein (*bat*) or Pas domain-containing protein (Montero-Lobato et al., 2018) present in between B and D subunit of V-type ATPases. However, no *bop* was observed in the genome (**Figure 6C**). The *bat* function as a transcription factor which regulates the *bop* expression (Mirfeizollahi et al., 2019). It will be interesting to study the role of accessory genes in pws1 and pws11 and the presence of *bat* in *bop* deficient strain pws11.

The differences in the genetic organization observed in the region coding for *bop* in pws strains prompted us to perform this analysis across all the sequenced haloarchaeal species. We

observed three distinct features in the genes coding for *bop* in *H. salinarum* and *H. marismortui*. (1) In *H. salinarum*, bop operon includes many additional accessory genes that are involved in retinal biosynthesis, BR over-expression, folding and membrane integration. In contrast, in *H. marismortui* HmBRI, only *bop* is present between B and D subunits of V-type ATPases. (2) *H. salinarium bop* has its specific promoter while HmBRI has no separate promoter region and may express with other neighboring genes like V-type ATPases in the predicted operon. (3) In *H. salinarum*, BR over-expression is light-dependent and regulated by the *brz* while HmBRI expresses constitutively and had no effect on expression upon light exposure (Fu et al., 2010).

All these observations suggest that during evolution, HmBRI might have randomly inserted between B and D subunit of V-type ATPases. We also found many other bops but those are not conserved and their location is also not very specific. Therefore, we have not included those set of genes in our analysis. Some of these examples are listed in **Supplementary Figure S3**. Whole-genome analysis of multiple haloarchaeal species revealed that similar to *H. marismortui* HmBRI, many other haloarchaeal species also have *bop* present in between B and D subunit of V-type ATPases (**Figures 6A,B**). Similar to pws1, multiple other *bop* related accessory genes were found between the B and D subunits of V-type ATPases in other haloarchaeal species (**Figure 6**). These *bop* related accessory genes include *crtY*, *bat,* β-carotene 15, and 15′-dioxygenases (*blh*) genes. In *H. salinarum* NRC-1 and R1 strains V-type ATPases operon, which is distinct from *bop* operon, also have *crtY* and *blh* genes in between B and D subunits (**Figure 6B**). These are an additional copy of the *crtY* and *blh* genes as a similar set of genes are also present within the *H. salinarum bop* operon. We also found many haloarchaeal strains having probably similar V-type ATPase operon organization like that observed in the pws11 strain where only BR associated (without *bop*) genes are present (**Figure 6C**). The function of BR associated genes in the absence of rhodopsin is not clear.

In *Natrinema ejinorense* JCM 13890 and *Halohasta litchfieldiae* DSM 22187, *bop* is located close to B subunit. There are many non-BR associated genes, including ABC transporter, iron transporter, extracellular solute binding protein, transposes, and permeases in the vicinity (**Figure 6A** lower panel). We also did not observe any D subunit coding gene close to B subunit in both of these V-type ATPase operons. Analysis of multiple genomes of different *Halorubrum* strains revealed that different strains of genus *Halorubrum* acquired different sets of genes between B and D subunits of V-type ATPases. For example, *Halorubrum* sp. WN019 has *bop* with an additional hypothetical gene, *Halorubrum* sp. SD616R have *bop* with GAF domain protein gene and *CrtY*, *Halorubrum halophilum* B8 has GAF domain protein, *CrtY* and *bat* (**Figure 6A**). Similarly two different type of *bop* operon found in different *Halomicrobium* strains (**Figure 6A**).

We also observed that in many haloarchaeal strains, no additional genes were present between B and D subunit of V-type ATPase (**Figure 6B**). In many cases, some genes were present however, no *bop* was observed (**Figure 6C**). These findings further suggest that *bop* and other genes might have probably randomly inserted between B and D subunits of V-type ATPases. The differences observed in the genetic organization of genes in the probable *bop* operons raises an important question about the role of different accessory genes in *bop* function and expression. Multiple sequence alignment of different haloarchaeal V-type ATPases operon sequences suggested that no insertions were observed between the B and D subunits coding regions (highlighted with an arrow in **Supplementary Figure S4**). We also searched for recombination hotspot regions and transposable element insertion sequences in the V-type ATPases, but we did not observe any such conserved sequences.

DISCUSSION

Haloarchaea have adapted to survive and grow under harsh conditions like high salt concentration, high temperature, high ionic stresses, UV light exposure, alkaline pH and nutrient limitation conditions. Several genes coding for different enzymes, proteins, and biomolecules such as BR and carotenoids facilitate haloarchaeal survival in such harsh environmental conditions. The main objective of this study was to explore the Indian solar saltern to study diversity and isolate new haloarchaeal strains harboring natural BR variants. We successfully isolated 12 haloarchaeal strains from the high salinity environment. Out of twelve, 11 were confirmed as BR harboring haloarchaeal strains using PCR-based screening assay. We also successfully cloned, expressed and purified one of the recombinant BR (pws5-BR) using *E. coli* as an expression host. It is challenging to express BR in functional form in *E. coli*. However, we were able to express pws5-BR using the strategy adopted in our previous study (Verma et al., 2019). Multiple sequencing alignment analysis suggested that pws5-BR shares 94% sequence identity with HmBRI. Considering the potential of BRs in several commercial applications, these high yielding BRs may be explored to exploit them as a substitute for *H. salinarum* BR. This will help in bringing down the cost of recombinant BR as the cost of natural BR is one of the major bottlenecks in commercializing these technologies. However, these recombinant BRs are not as stable as *H. salinarum* BR (Seyedkarimi et al., 2015; Verma et al., 2019). In the future, it will be desirable to work on improving the thermal stability of these BRs to make technologies based on these proteins a reality.

To gain insights into the genetic diversity in the isolated strains whole-genome sequencing followed by pangenome analysis were performed for pws1, pws5 and pws8 strains. The number of identified coding DNA sequences (CDS) ranged from 3,722 to 4,277. The pws5 genome having the largest genome size ∼4.0 Mb with 4,277 CDS, while pws1 has the smallest genome size ∼3.4 Mb with 3,722 CDS. Pangenome analysis of the protein-coding complements of pws1, pws5, and pws8 revealed that several functional genes were added into different categories, including metabolism and transport of amino acids, inorganic

ions, carbohydrates, and secondary metabolites as well as genes involved in signal transduction, translation and cell division. The arCOG distribution plot of pws1, pws5, and pws8 with their respected genus suggested that they have open pangenome which is increasing with the addition of new genome sequences. So, it will be interesting to sequence more isolates to study the genetic diversity in these genera, which may help us in understanding the genes involved in growth and adaptation.

Multiple BR operon analysis of different haloarchaeal genomes (including pws isolates) revealed that in *bop*, if present, is mainly present between the genes coding for B and D subunits of V-type ATPases however, in some cases, *bop* alone or along with some *bop* accessory genes is present in other locations as well. We also attempted to purify BR from the three pws isolates, but we were not successful (data not shown). In previous studies, it has been shown that BR is constitutively expressed in *H. marismortui* (Fu et al., 2010). Our data analysis also suggests, unlike *H. salinarum* the presence of light-inducible transcription regulator is absent in these isolates and many other haloarchaeal species reported in the past. This may be one of the probable reasons for poor or less expression of BR in these haloarchaeal isolates.

To summarize, this study reports four new haloarchaeal strains, isolated from the Indian solar salterns. We also report the differences observed in the *bop* operon/genetic architecture in different haloarchaeal species. We are also successful in purifying functional recombinant pws5-BR using *E. coli* as an expression host. This finding also highlights the presence of several industrially important enzymes and metabolites present in these haloarchaeal isolates.

MATERIALS AND METHODS

Isolation of Halophilic Archaea and Analysis of Samples

Solar salterns water samples were collected in bottles (500 ml each) from solar saltern situated on the ECR Highway of Kottakuppam, near to Puducherry, Tamil Nadu India (11°59′17.39″ N and 79°50′17.39″ E) and haloarchaeal species were isolating using membrane filtration technique as suggested by Montalvo-Rodriguez et al. (1998); and Elevi et al. (2004). The samples were named as pws1 to pws12, respectively. Enrichment technique was employed for the isolation of halophilic archaea. All pws samples were initially filtered using vacuum filtration techniques through 0.45-micron membrane filters, and the filtered porous membranes were transferred to halobacterium agar media (HB media) (HiMedia) and incubated for 2–3 weeks at 37°C. The filter pieces containing pink-red patches were further transferred to sterile HB liquid medium containing 250 g/l NaCl, 3 g/l, trisodium citrate, 20 g/l MgSO4. 7H20, 2 g/l KCl and 10 g/l Oxoid peptone. After 1 week of incubation at 37°C, 200 rpm red-pink color fermentation broth indicates the different carotenoid producing haloarchaeal strains. Repeated subculturing for 3–4 times in HB media yielded pure single colony red/pink pigmented colonies on the agar plates. Each colony was separately inoculated into 20 ml HB medium and incubated at 37°C with 200 rpm for genomic DNA isolation to study 16S rRNA taxonomic diversity and *bop* PCR amplification studies.

Genomic DNA Isolation and 16S rRNA Sequencing

Genomic DNA extractions of all twelve strains were performed by Zymo research genomic DNA isolation kit (Cat No. D6105). The partial 16S rRNA sequences were amplified using standard primers 21F and 1453R (**Supplementary Table 1**). The other 16S rRNA sequences used in the comparative analysis were taken from the NCBI database (https://www.ncbi.nlm.nih.gov/). The amplified partial 16S rRNA gene sequences were further used for taxonomic identification using the EzTaxon server (Chun et al., 2007).

Phylogenetic Tree Construction

The 16S rRNA sequences used for phylogenetic analysis were aligned using program MUSCLE (v3.8.31) with default settings (Edgar, 2004). Ambiguous regions i.e., poorly aligned or regions containing gaps were removed from the aligned sequences using Gblocks (v0.91b) (Castresana, 2000). The phylogenetic trees were reconstructed with PhyML program (v3.1/3.0 aLRT) using the maximum likelihood method (Guindon and Gascuel, 2003). The default substitution model was selected assuming an estimated proportion of invariant sites (of 0.748, 0.761, 0.840, and 0.940 for pws1, pws5, pws8, and pws11) and 4 gamma-distributed rate categories to account for rate heterogeneity across sites. The gamma shape parameter was estimated directly from the data. Bootstrapping method (100 bootstrap replicates) was used to assess the reliability of the internal branch (Anisimova and Gascuel, 2006). Graphical representation of the phylogenetic trees were prepared using iTOL (Letunic and Bork, 2019).

Transmission Electron Microscopy

Aliquots of 50 μl from the haloarchaeal grown cultures for 4–5 days at 37°C with constant shaking at 200 rpm were added on a carbon-coated 300-mesh copper grid (Polysciences Asia Pacific, Taiwan). The excess sample was blotted, followed by air drying. The grids were further imaged using a JEM 2100 electron microscope (JEOL), operated at 200 keV.

Carotenoid Isolation

Carotenoids were isolated following a acetone/methanol extraction method (Yatsunami et al., 2014). Briefly, 20 ml of grown culture were pelleted at 18,000 × g for 30 min. Supernatants were discarded and colored pellets were dissolved in 40 mL of acetone and methanol mixture prepared with 7:3 ratio. The samples were incubated at room temperature for 30 min in dark and again centrifuged at 18,000 × g for 30 min. Final colored supernatants were collected and speedvac at 37°C to evaporate methanol and acetone. Reddish-pink colored pellets were later dissolved in 50 μl of methanol. UV-visible spectra of extracted carotenoid solutions were recorded from 200 to 700 nm range using a UV-Vis spectrophotometer (CECIL CE7500 spectrophotometer).

Screening of *bop* From pws Isolates

The isolated strains were screened for the presence of *bop* using a PCR-based screening assay. Multiple sequence alignment was performed to identify highly conserved regions in *bop* and design a set of degenerate primers against conserved regions (**Supplementary Table 1**). Isolated whole-genome DNA samples were subjected to PCR-based amplification using these conserved primers (DegF and DrgR). The expected size of PCR product amplified using these primers is around 450 bp.

Cloning, Expression, and Purification of BRs

The full-length pws5-*bop* was amplified by gene-specific primers named Bop_full_F and Bop_full_R (**Supplementary Table 1**). The amplified fragments were further digested with NheI-XhoI restriction enzymes and cloned into pET22b vector for C-terminal 6× his-tag. The pws-5 *bop* has several rare codons. So, to aid expression we used C43-Rosetta BL21 (DE3) *E. coli* cells as reported in our earlier study (Verma et al., 2020). *E. coli* C43-Rosetta BL21 (DE3) cells are C43 (DE3) cells harboring the pRARE plasmid isolated from Rosetta (DE3) cells. The growing culture was induced at OD 0.6 with 0.5 mM IPTG and 5–10 μM retinal (Sigma-Aldrich). 100% ethanol was used to prepare trans-retinal stocks. The culture was incubated for 4 h in an incubator shaker at 37°C and kept shaking at 200 rpm. The culture was centrifuged at 9,000 × g for 15 min and resuspended in lysis buffer A (20 mM Tris, pH 7.5 and 4 M NaCl). The cells were sonicated and centrifuged at 18,000 x× g for 60 min to obtain membrane fraction in the pellet. The colored pellet was further resuspended in lysis buffer A having 0.2% DDM (Anatrace, USA) and incubated overnight. The role of DDM detergent was to extract BR from the insoluble membrane fraction. The soluble colored fraction was incubated with Ni-NTA resin for 1 h to facilitate binding, and the colored protein was eluted using elution buffer E (20 mM Tris pH 7.5, 0.02% DDM and 4 M NaCl, with 500 mM imidazole). The purified protein was concentrated and dialyzed against buffer composed of 20 mM Tris pH 7.5, 0.02% DDM and 4 M NaCl to remove imidazole.

BR Spectral Analysis and Proton Pumping Activity

The 10 μM of purified pws5-BR was used to perform UV-Vis spectral scanning (200 to 800 nm) using CECIL CE7500 spectrophotometer. The light-based proton pumping assays were performed as described earlier by Wang et al. (2003). Briefly, 20 ml culture of BR expressing C43-Rosetta BL21 (DE3) *E. coli* cells were centrifuged at 15,000 × g at 4°C, washed and resuspended in a non-buffered solution (10 mM NaCl, 10 mM MgSO4 and 100 mM $CaCl_2$). For the proton pumping experiment, OD_{600} was adjusted to 2.0 in dark. The experiment was started by illuminating the cells with the continuous high-intensity white light source for 120s. The change in pH was measured by using a JENWAY 3510 pH meter.

Genome Sequencing, Assembly, and Data Submission

Four pws isolates named pws1, pws5, pws8, and pws11 were subjected for whole-genome sequencing using Illumina NextSeq 500 and assembled using the CLC NGS Cell 9.0. The whole-genome sequencing was performed by Bionivid Technology Pvt Ltd, India. The whole-genome sequencing data of pws1, pws5, pws8, and pws11 data were submitted to the NCBI database under accession numbers WOYG00000000, WOWA00000000, WOWB00000000, and WOWC00000000, respectively. The draft genomes were annotated using the RAST pipeline online server (Aziz et al., 2008). Genome comparison and pangenome analysis were carried out by using the EzTaxon and BPGA pipeline, respectively (Chun et al., 2007; Chaudhari et al., 2016).

DATA AVAILABILITY STATEMENT

The datasets generated for 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

KT, SP, and DV conceived the study. KT and SP coordinated the study. SP and BS provided the strains. KT and DV designed experiments, analyzed data, and wrote the paper with inputs from other coauthors. DV, LS, CC, and CS performed experiments. All authors reviewed the results and approved the final version of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants to KT and SP by Council of Scientific and Industrial Research, India. DV was a recipient of a senior research fellowship, Council of Scientific and Industrial Research, India. LS was a recipient of a senior research fellowship, Department of Biotechnology, India.

ACKNOWLEDGMENTS

We thank Mr. Surinder Singh, Mr. Randeep Singh, and Mr. Davender Singh at CSIR-IMTECH Chandigarh for supporting Laboratory experiments. KT would like to acknowledge members of the Structural Biology Laboratory for useful suggestions and discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.554927/full#supplementary-material

REFERENCES

Anisimova, M., and Gascuel, O. (2006). Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* 55, 539–552. doi: 10.1080/10635150600755453

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

Bogomolni, R. A., Baker, R. A., Lozier, R. H., and Stoeckenius, W. (1976). Light-driven proton translocations in Halobacterium halobium. *Biochim. Biophys. Acta* 440, 68–88. doi: 10.1016/0005-2728(76)90114-6

Bowers, K. J., and Wiegel, J. (2011). Temperature and pH optima of extremely halophilic archaea: a mini-review. *Extremophiles* 15, 119–128. doi: 10.1007/s00792-010-0347-y

Bratanov, D., Balandin, T., Round, E., Shevchenko, V., Gushchin, I., Polovinkin, V., et al. (2015). An approach to heterologous expression of membrane proteins. The Case of Bacteriorhodopsin. *PLoS ONE* 10:e0128390. doi: 10.1371/journal.pone.0128390

Cabrera, M. A., and Blamey, J. M. (2018). Biotechnological applications of archaeal enzymes from extreme environments. *Biol. Res.* 51:37. doi: 10.1186/s40659-018-0186-3

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Cayol, J. L., Ollivier, B., Patel, B. K., Prensier, G., Guezennec, J., and Garcia, J. L. (1994). Isolation and characterization of Halothermothrix orenii gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. *Int. J. Syst. Bacteriol.* 44, 534–540. doi: 10.1099/00207713-44-3-534

Chaudhari, N. M., Gupta, V. K., and Dutta, C. (2016). BPGA- an ultra-fast pan-genome analysis pipeline. *Sci. Rep.* 6:24373. doi: 10.1038/srep24373

Chun, J., Lee, J. H., Jung, Y., Kim, M., Kim, S., Kim, B. K., et al. (2007). EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 57(Pt 10), 2259–2261. doi: 10.1099/ijs.0.64915-0

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, R., Assa, P., Birbir, M., Ogan, A., and Oren, A. (2004). Characterization of extremely halophilic Archaea isolated from the Ayvalik Saltern, Turkey. *World J. Microb. Biot.* 20, 719–725. doi: 10.1007/s11274-004-4515-z

Fendrihan, S., Legat, A., Pfaffenhuemer, M., Gruber, C., Weidler, G., Gerbl, F., et al. (2006). Extremely halophilic archaea and the issue of long-term microbial survival. *Rev. Environ. Sci. Biotechnol.* 5, 203–218. doi: 10.1007/s11157-006-0007-y

Fu, H. Y., Lin, Y. C., Chang, Y. N., Tseng, H., Huang, C. C., Liu, K. C., et al. (2010). A novel six-rhodopsin system in a single archaeon. *J. Bacteriol.* 192, 5866–5873. doi: 10.1128/JB.00642-10

Giani, M., Garbayo, I., Vilchez, C., and Martinez-Espinosa, R. M. (2019). Haloarchaeal carotenoids: healthy novel compounds from extreme environments. *Mar. Drugs* 17:527. doi: 10.3390/md17090524

Gramain, A., Diaz, G. C., Demergasso, C., Lowenstein, T. K., and McGenity, T. J. (2011). Archaeal 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

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Higa, L. H., Schilrreff, P., Briski, A. M., Jerez, H. E., de Farias, M. A., Villares Portugal, R., et al. (2020). Bacterioruberin from Haloarchaea plus dexamethasone in ultra-small macrophage-targeted nanoparticles as potential intestinal repairing agent. *Colloids Surf. B Biointerfaces* 191:110961. doi: 10.1016/j.colsurfb.2020.110961

Jeganathan, C., Thamaraiselvi, K., and Sabari Girisun, T. C. (2019). Improved production of bacteriorhodopsin from Halobacterium salinarum through direct amino acid supplement in the basal medium. *Extremophiles* 23, 133–139. doi: 10.1007/s00792-018-1067-y

Kahaki, F. A., Babaeipour, V., Memari, H. R., and Mofid, M. R. (2014). High overexpression and purification of optimized bacterio-opsin from Halobacterium Salinarum R1 in E. *coli. Appl. Biochem. Biotechnol.* 174, 1558–1571. doi: 10.1007/s12010-014-1137-2

Kanehara, K., Yoshizawa, S., Tsukamoto, T., and Sudo, Y. (2017). A phylogenetically distinctive and extremely heat stable light-driven proton pump from the eubacterium Rubrobacter xylanophilus DSM 9941(T). *Sci. Rep.* 7:44427. doi: 10.1038/srep44427

Kanekar, P. P., Kulkarni, S. O., Kanekar, S. P., Shouche, Y., Jani, K., and Sharma, A. (2015). Exploration of a haloarchaeon, Halostagnicola larsenii, isolated from rock pit sea water, West Coast of Maharashtra, India, for the production of bacteriorhodopsin. *J. Appl. Microbiol.* 118, 1345–1356. doi: 10.1111/jam.12784

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, Y. T., Tian, Y., Tian, H., Tu, T., Gou, G. Y., Wang, Q., et al. (2018). A review on bacteriorhodopsin-based bioelectronic devices. *Sensors (Basel)* 18:1368. doi: 10.3390/s18051368

Littlechild, J. A. (2015). Archaeal enzymes and applications in industrial biocatalysts. *Archaea* 2015:147671. doi: 10.1155/2015/147671

Medini, D., Donati, C., Tettelin, H., Masignani, V., and Rappuoli, R. (2005). The microbial pan-genome. *Curr. Opin. Genet. Dev.* 15, 589–594. doi: 10.1016/j.gde.2005.09.006

Mirfeizollahi, A., Yakhchali, B., Deldar, A. A., and Karkhane, A. A. (2019). In silico and experimental improvement of bacteriorhodopsin production in Halobacterium salinarum R1 by increasing DNA-binding affinity of Bat through Q661R/Q665R substitutions in HTH motif. *Extremophiles* 23, 59–67. doi: 10.1007/s00792-018-1060-5

Mohr, V., and Larsen, H. (1963). On the structural transformations and lysis of halobacterium salinarium in hypotonic and isotonic solutions. *Microbiology* 31, 267–280.

Montalvo-Rodriguez, R., Vreeland, R. H., Oren, A., Kessel, M., Betancourt, C., and Lopez-Garriga, J. (1998). Halogeometricum borinquense gen. nov., sp. nov., a novel halophilic archaeon from Puerto Rico. *Int. J. Syst. Bacteriol.* 48 (Pt 4), 1305–1312. doi: 10.1099/00207713-48-4-1305

Montero-Lobato, Z., Ramos-Merchante, A., Fuentes, J. L., Sayago, A., Fernandez-Recamales, A., Martinez-Espinosa, R. M., et al. (2018). Optimization of growth and carotenoid production by haloferax mediterranei using response surface methodology. *Mar. Drugs* 16:372. doi: 10.3390/md16100372

Nadella, R., and Hernandez Baltazar, D. (2018). Identification of new protein sources for renewable energy storage systems: A bioinformatic study. doi: 10.1063/1.5047953

Peck, R. F., Echavarri-Erasun, C., Johnson, E. A., Ng, W. V., Kennedy, S. P., Hood, L., et al. (2001). brp and blh are required for synthesis of the retinal cofactor of bacteriorhodopsin in Halobacterium salinarum. *J. Biol. Chem.* 276, 5739–5744. doi: 10.1074/jbc.M009492200

Purdy, K., Cresswell-Maynard, T., Nedwell, D., McGenity, T., Grant, W., Timmis, K., et al. (2004). Isolation of haloarchaea that grow at low salinities. *Environ. Microb.* 6, 591–595. doi: 10.1111/j.1462-2920.2004.00592.x

Rajurkar, A., and Pathak, A. (2014). Exploring the microbiota of solar saltem of Mulund, Mumbai, India. *Indian J.Geo-Mar. Sci.* 43, 634–641.

Rodrigo-Banos, M., Garbayo, I., Vilchez, C., Bonete, M. J., and Martinez-Espinosa, R. M. (2015). Carotenoids from haloarchaea and their potential in biotechnology. *Mar. Drugs* 13, 5508–5532. doi: 10.3390/md13095508

Saeedi, P., Moosaabadi, J. M., Sebtahmadi, S. S., Mehrabadi, J. F., Behmanesh, M., and Mekhilef, S. (2012). Potential applications of bacteriorhodopsin mutants. *Bioengineered* 3, 326–328. doi: 10.4161/bioe.21445

Sediroglu, V., Yucel, M., Gunduz, U., Turker, L., and Eroglu, I. (1998). "The Effect of Halobacterium halobium on Photoelectrochemical Hydrogen Production," in *BioHydrogen,* eds. O.R. Zaborsky, J.R. Benemann, T. Matsunaga, J. Miyake & A. San Pietro. (Boston, MA: Springer US), 295–304.

Seyedkarimi, M. S., Aramvash, A., and Ramezani, R. (2015). High production of bacteriorhodopsin from wild type Halobacterium salinarum. *Extremophiles* 19, 1021–1028. doi: 10.1007/s00792-015-0778-6

Sharma, A. K., Walsh, D. A., Bapteste, E., Rodriguez-Valera, F., Ford Doolittle, W., and Papke, R. T. (2007). Evolution of rhodopsin ion pumps in haloarchaea. *BMC Evol. Biol.* 7, 79. doi: 10.1186/1471-2148-7-79

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

Tarasov, V. Y., Besir, H., Schwaiger, R., Klee, K., Furtwangler, K., Pfeiffer, F., et al. (2008). A small protein from the bop-brp intergenic region of Halobacterium salinarum contains a zinc finger motif and regulates bop and crtB1 transcription. *Mol. Microbiol.* 67, 772–780. doi: 10.1111/j.1365-2958.2007.06081.x

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

Thombre, R. S., Shinde, V. D., Oke, R. S., Dhar, S. K., and Shouche, Y. S. (2016). Biology and survival of extremely halophilic archaeon Haloarcula marismortui RR12 isolated from Mumbai salterns, India in response to salinity stress. *Sci. Rep.* 6:25642. doi: 10.1038/srep25642

Verma, D. K., Baral, I., Kumar, A., Prasad, S. E., and Thakur, K. G. (2019). Discovery of bacteriorhodopsins in Haloarchaeal species isolated from Indian solar salterns: deciphering the role of the N-terminal residues in protein folding and functional expression. *Microb. Biotechnol.* 12, 434–446. doi: 10.1111/1751-7915.13359

Verma, D. K., Vasudeva, G., Sidhu, C., Pinnaka, A. K., Prasad, S. E., and Thakur, K. G. (2020). Biochemical and taxonomic characterization of novel haloarchaeal strains and purification of the recombinant halotolerant α-amylase discovered in the isolate. *Front. Microbiol.* 11:2082. doi: 10.3389/fmicb.2020.02082

Vernikos, G., Medini, D., Riley, D. R., and Tettelin, H. (2015). Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* 23, 148–154. doi: 10.1016/j.mib.2014.11.016

Walsby, A. E. (1994). Gas vesicles. *Microbiol Rev* 58, 94–144.

Wang, W. W., Sineshchekov, O. A., Spudich, E. N., and Spudich, J. L. (2003). Spectroscopic and photochemical characterization of a deep ocean proteorhodopsin. *J. Biol. Chem.* 278, 33985–33991. doi: 10.1074/jbc.M305716200

Yang, Y., Yatsunami, R., Ando, A., Miyoko, N., Fukui, T., Takaichi, S., et al. (2015). Complete biosynthetic pathway of the C50 carotenoid bacterioruberin from lycopene in the extremely halophilic archaeon Haloarcula japonica. *J. Bacteriol.* 197, 1614–1623. doi: 10.1128/JB.02523-14

Yatsunami, R., Ando, A., Yang, Y., Takaichi, S., Kohno, M., Matsumura, Y., et al. (2014). Identification of carotenoids from the extremely halophilic archaeon Haloarcula japonica. *Front. Microbiol.* 5:100. doi: 10.3389/fmicb.2014.00100

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fmicb.2020.622824

Genomic Insights Into the Lifestyles of Thaumarchaeota Inside Sponges

*Markus Haber[1,2], Ilia Burgsdorf[1], Kim M. Handley[3], Maxim Rubin-Blum[4] and Laura Steindler[1]**

[1] Department of Marine Biology, Leon H. Charney School of Marine Sciences, University of Haifa, Haifa, Israel, [2] Department of Aquatic Microbial Ecology, Institute of Hydrobiology, Biology Centre CAS, České Budějovice, Czechia, [3] School of Biological Sciences, The University of Auckland, Auckland, New Zealand, [4] Israel Oceanographic and Limnological Research Institute, Haifa, Israel

Sponges are among the oldest metazoans and their success is partly due to their abundant and diverse microbial symbionts. They are one of the few animals that have Thaumarchaeota symbionts. Here we compare genomes of 11 Thaumarchaeota sponge symbionts, including three new genomes, to free-living ones. Like their free-living counterparts, sponge-associated Thaumarchaeota can oxidize ammonia, fix carbon, and produce several vitamins. Adaptions to life inside the sponge host include enrichment in transposases, toxin-antitoxin systems and restriction modifications systems, enrichments previously reported also from bacterial sponge symbionts. Most thaumarchaeal sponge symbionts lost the ability to synthesize rhamnose, which likely alters their cell surface and allows them to evade digestion by the host. All but one archaeal sponge symbiont encoded a high-affinity, branched-chain amino acid transporter system that was absent from the analyzed free-living thaumarchaeota suggesting a mixotrophic lifestyle for the sponge symbionts. Most of the other unique features found in sponge-associated Thaumarchaeota, were limited to only a few specific symbionts. These features included the presence of exopolyphosphatases and a glycine cleavage system found in the novel genomes. Thaumarchaeota have thus likely highly specific interactions with their sponge host, which is supported by the limited number of host sponge species to which each of these symbionts is restricted.

Keywords: sponge (Porifera), archaea, thaumarchaeota, symbiosis, *Petrosia ficiformis*, *Theonella swinhoei*, *Hymedesmia (Stylopus) methanophila*

OPEN ACCESS

Edited by:
Andreas Teske, University of North Carolina at Chapel Hill, United States

Reviewed by:
Lu Fan, Southern University of Science and Technology, China
Beate M. Slaby, GEOMAR Helmholtz Center for Ocean Research Kiel, Germany

****Correspondence:***
Laura Steindler
lsteindler@univ.haifa.ac.il

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

Received: *29 October 2020*
Accepted: *14 December 2020*
Published: *11 January 2021*

Citation:
Haber M, Burgsdorf I, Handley KM, Rubin-Blum M and Steindler L (2021) Genomic Insights Into the Lifestyles of Thaumarchaeota Inside Sponges. Front. Microbiol. 11:622824. doi: 10.3389/fmicb.2020.622824

INTRODUCTION

Sponges (phylum Porifera) are sessile, soft-bodied invertebrates that inhabit marine and freshwater environments around the globe (Van Soest et al., 2012). As highly efficient filter feeders, they are a crucial link between the pelagic and benthic environments (Yahel et al., 2003; de Goeij et al., 2013). Their success has been attributed to their microbial symbionts (Taylor et al., 2007) that can make up to 38% of their tissue volume (Vacelet and Donadey, 1977). Sponge symbionts, including bacteria and archaea, can be transmitted both horizontally and vertically (Steger et al., 2008; Björk et al., 2019). The symbionts support their host in various ways. They produce secondary metabolites that are likely involved in the chemical defense of their host against predators, fouling, and pathogens (Waters et al., 2014; Mori et al., 2018); remove waste products produced by their eukaryotic host,

such as ammonia (Moeller et al., 2019); contribute nutrients (Erwin and Thacker, 2008; Mohamed et al., 2008; Rubin-Blum et al., 2019); and might even be involved in skeleton formation (Uriz et al., 2012). Given that sponges are one of the earliest-branching multi-cellular animals (Metazoa) (Feuda et al., 2017), these symbiotic interactions are considered ancient.

The microbial communities in sponges are often complex (Thomas et al., 2016) and comprise more than 60 different bacterial and archaeal phyla (Moitinho-Silva et al., 2017b). In many cases, these symbionts fall into phylogenetic clades made up entirely of sequences derived from sponges (and sometimes corals) suggesting highly specific relationships (Simister et al., 2012; Taylor et al., 2013). Comparative analyses of metagenomes from sponge-associated and seawater communities found symbiont consortia to be enriched in features such as transposable elements, defense mechanisms (e.g., CRISPR-Cas systems) and eukaryote-like proteins, potentially involved in symbiont recognition and interaction (Thomas et al., 2010; Fan et al., 2012; Horn et al., 2016; Slaby et al., 2017). Studies on individual bacterial clades revealed several additional features that differ between sponge symbionts and their free-living counterparts. The widespread sponge-associated cyanobacteria *Candidatus* Synechococcus spongiarum have more streamlined genomes than their close free-living relatives, lost genes encoding low molecular weight peptides involved in stabilizing and protecting the photosynthesis apparatus, and have fewer genes related to oxidative stress protection and signal transduction (Gao et al., 2014; Burgsdorf et al., 2015). Sponge symbionts of the Alphaproteobacteria family Rhodospirillaceae lack genes related to chemotaxis and motility compared to free-living bacteria from the same family, but have more genes related to the biosynthesis of secondary metabolites and cell detoxification (Karimi et al., 2018).

Bacteria usually dominate sponge microbial communities, but archaea are regularly found in sponge hosts and can make up substantial parts of the community (e.g., 24% of the microbial community in the Great Barrier reef sponge *Ianthella basta*) (Moeller et al., 2019), and in some cases even dominate the communities (e.g., up to 63% in the microbial community of the sponge *Axinella mexicana*) (Preston et al., 1996). A comparison of the archaeal sponge community from hosts living in shallow waters and the deep-sea indicated that sponge-associated archaea appear to be much more abundant (up to three orders of magnitude greater) in the deep-sea (Steinert et al., 2020). Most archaeal sponge symbionts are Thaumarchaeota, with some belonging to sponge-specific clades (Simister et al., 2012). Thaumarchaeota are found in sponge larvae (Steger et al., 2008) and also in adult sponges, that acquire their symbionts horizontally (Burgsdorf et al., 2014), indicating the same variability in transmission routes as found in bacterial sponge symbionts (Sipkema et al., 2015). Thaumarchaeota are well-known ammonia-oxidizers (Pester et al., 2011; Baker et al., 2020) making them a key player in sponge nitrogen metabolism and nitrogenous waste removal. While their role in the sponge nitrogen cycle is well established based on *in situ* physiology measurements (Bayer et al., 2008; Radax et al., 2012a; Moeller et al., 2019), *amoA* gene expression studies (López-Legentil et al., 2010; Zhang et al., 2014) and more recently metagenome (Moitinho-Silva et al., 2017a; Moeller et al., 2019; Zhang et al., 2019), metatranscriptome (Radax et al., 2012b; Moitinho-Silva et al., 2017a) and metaproteome analyses (Moeller et al., 2019), less is known about the genomic adaptations of Thaumarchaeota to life inside sponges.

The first published genome of a thaumarchaeal sponge symbiont is that of *Cenarchaeum symbiosum*, hosted by the demosponge *Axinella mexicana* (Hallam et al., 2006). Recently, additional genomes of archaeal symbionts have become available, including those hosted by deep-sea and glass sponges (Tian et al., 2016, 2017; Moitinho-Silva et al., 2017a; Moeller et al., 2019; Zhang et al., 2019). Only a limited number of studies compared the genomes of symbiotic and free-living Thaumarchaeota (Moeller et al., 2019; Zhang et al., 2019). Moeller et al. (2019) compared a metagenome-assembled thaumarchaeal genome (MAG) from the sponge *Ianthella basta* to those of three other sponge-associated and several free-living Thaumarchaeota. Zhang et al. (2019) compared the MAGs of three sponge-associated *Nitrosopumilus* species with those of their free-living relatives. Both studies found that the sponge symbiont genomes had elevated GC content. These symbiotic Thaumarchaeota are likely mixotrophs, due to the presence of a branched-chain amino acid transporter and are enriched in mechanisms for defense against phages, such as restriction-modification systems. Sponge-associated Thaumarchaeaota also encode eukaryotic-like proteins, which are absent from free-living Thaumarchaeota, that are assumed to be involved in interactions with the sponge hosts (Moeller et al., 2019; Zhang et al., 2019). Here, we aim to expand our knowledge of the genomic repertoire in sponge-associated Thaumarchaeota. We analyzed three new MAGs of sponge-associated Thaumarchaeota, together with the nine previously reported sponge-associated archaea genomes, and the high-quality (>95% completeness) genomes of twelve free-living Thaumarchaeota.

MATERIALS AND METHODS

Sponge Sampling and DNA Isolation

Samples of the sponges *Theonella swinhoei* and *Petrosia ficiformis* were collected by SCUBA from the Gulf of Aqaba, Red Sea (29°29′N 34°54′E) at 25 m, Israel on July 31st, 2012 and the Achziv nature marine reserve (33°00′N 35°02′E) at 20 m, Israel on May 5th, 2013, respectively, and transported on ice to the laboratory for further processing (*T. swinhoei* within 20 min to IUI-Eilat and *P. ficiformis* within 2 h to University of Haifa, respectively). Cortex tissue of *P. ficiformis* and both cortex and endosome of *T. swinhoei* were used for microbial cell enrichment, followed by DNA extraction. Microbial cell enrichment by a series of filtration and centrifugation steps and DNA extraction followed previously described methods (Thomas et al., 2010).

The asphalt-encrusting deep-sea demosponge *Hymedesmia* (*Stylopus*) *methanophila* was collected using the ROV QUEST 4,000 m (Marum, Bremen) operated from the research vessel Meteor during the M114-2 cruise to Campeche Knolls, Southern Gulf of Mexico, in March 2015 (Rubin-Blum et al., 2019). The

specimen was collected from the Chapopote knoll (21°54′ N; 93°26′ W at 2,925 m. The collection site is described in detail elsewhere (Sahling et al., 2016). Upon retrieval, the sponge tissue was removed from the underlying asphalt with a scalpel, fixed in RNAlater (Sigma®, Steiheim, Germany) according to the manufacturer's instructions and stored at −80°C. DNA was extracted with the AllPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions.

Shotgun Sequencing, Assembly, and Taxonomic Binning

T. swinhoei and *P. ficiformis* microbiomes were sequenced using an Illumina HiSeq2000 platform (2 × 100 bp, paired-end) at the Institute for Genomics and Systems Biology's Next Generation Sequencing Core (IGSB-NGS, ANL) at the University of Chicago. Libraries were generated using the TruSeq DNA standard protocol and pooled for sequencing using the Illumina HiSeq2000 platform. 69 and 129 million 100 bp sequences were generated from *T. swinhoei* and *P. ficiformis*, respectively. Sequence quality was assessed and low-quality reads (q = 3) were trimmed using the FASTX-Toolkit 0.0.13.2[1]. Sequence data sets were assembled *de novo* using IDBA-UD version 1.1.0 (Peng et al., 2012) with a kmer range of 50–70, and a step size of 5, following empirical tests. To assign contigs into genome bins, genes on contigs ≥2 kb long were predicted using Prodigal with the metagenome option (Hyatt et al., 2010, 2012). For each contig, we determined the GC content, coverage, and the taxonomic affiliation based on the best hit for each predicted protein in the Uniref90 database (accessed September 2013; Suzek et al., 2007) following UBLAST searches (usearch64; Edgar, 2010). Contigs were assigned to MAGs using these data, as well as emergent self-organizing maps (ESOM) based analysis of fragment tetranucleotide frequencies (Dick et al., 2009), as detailed by Handley et al. (2013). Among the metagenome-assembled genomes (MAGs) obtained from *T. swinhoei* was one archaea genome containing 158 contigs and 1,859,711 bp (coverage 65×). *P. ficiformis* contained two archaea genomes: one with 108 contigs and 1,791,570 bp and a second one with 161 contigs and only 521,474 bp. These two genomes (labeled bin A and B) were easily distinguishable based on their coverage (493× vs. 38×).

Genomic DNA libraries for *Hymedesmia* (*Stylopus*) *methanophila* were generated with the DNA library prep kit for Illumina (BioLABS, Frankfurt am Main, Germany) and sequenced on the Illumina HiSeq 2500 platform at the Max Planck Genome Centre (Cologne, Germany). 12.5 million 250 bp paired-end metagenomic reads were generated, while the remaining 15.5 million were generated as 150 bp paired-end reads. The metagenome was assembled with IDBA-UD (Peng et al., 2012) following decontamination, quality filtering (Q = 2) and adapter-trimming of the reads with BBDuk tool from the BBMap suite (Bushnell B)[2]. The archaeal MAG was binned based on genome coverage, GC content and taxonomic affiliation with genome-bin-tools (Seah and Gruber-Vodicka, 2015). The MAG was reassembled with Spades V3.10 (Bankevich et al., 2012) using a maximum k-mer length of 127, following re-mapping of Illumina reads to the bins using BBMap with 0.98 minimum identity. Contigs of <2 kbp were removed and the resulting bin was used in the analysis. Hereafter, we refer to this MAG as *Ca.* Nitrosopumilus sp. ESC.

[1]http://hannonlab.cshl.edu/fastx_toolkit/

[2]http://sourceforge.net/projects/bbmap/

Genome Dataset

In addition to the three sponge-associated archaeal genomes generated in this study, 20 previously published genomes were downloaded from NCBI and IMG for comparisons. These included 8 sponge symbiont genomes: *Crenarchaeum symbiosum* A (Hallam et al., 2006), *Ca.* Nitrosopumilus sp. LS AOA (Tian et al., 2016), *Ca.* Nitrosopumilus sp. Nsub (Tian et al., 2017) the most complete genome of each of *Ca.* Nitrosopumilus hexadellus, *Ca.* Nitrosopumilus detritiferus, and *Ca.* Cenporiarchaeum stylissum (Zhang et al., 2019), *Ca.* Nitrosopumilus cymbastelus (Moitinho-Silva et al., 2017a), *Ca.* Nitrosospongia ianthallae (Moeller et al., 2019). In addition, 12 free-living Thaumarchaeota genomes were retrieved that belong to the Nitrosopumilaceae family (5 from marine sediment, 5 from marine pelagic, 2 from soil) (see **Supplementary Table S1** for details). The selection of free-living genomes was based on passing an estimated genome completeness threshold of >95% based on 46 single copy COGs (see below for details of assessment).

Completeness Estimation, and Functional Annotation

All 23 genomes were uploaded to RAST (Aziz et al., 2008; Overbeek et al., 2014). Open reading frames (ORFs) were identified with the classic RAST algorithm and SEED annotations were obtained. Protein fasta files of predicted ORFs were downloaded and protein domains (PFAM) (Finn et al., 2010) and clusters of orthologous groups (COGs) (Tatusov et al., 2003) were annotated through the WebMGA annotation tool (Wu et al., 2011) with an e-value cutoff of 0.001. COGs were identified by rpsblast 2.2.15 searches against the NCBI COG database 2/2/2011. In the case of multiple COGs, only the best hit was retained. COGs were assigned to COG classes according to Galperin et al. (2015). PFAMs were assigned with hmmscan 3.0 and the Pfam database 24.0.

To select genomes for analysis, genome completeness was first determined by comparing the annotated COGs to a list of 53 essential single-copy COGs previously described to be present in all archaea (Puigbò et al., 2009). After the exclusion of COG0455 that was absent in all analyzed Thaumarchaeota genomes, and six COGs that were present in multiple copies in more than half of the analyzed genomes (present in 17–28 genomes in multiple copies), 46 COGs remained for completeness estimation. Only one of these COGs was present more than once in two or more genomes (5 times). Only sponge-associated Thaumarchaeota genomes with >90% completeness (>41 of the 46 COGs present) and free-living ones with >95% completeness (>43 COGs) were used in the analyses. Of the four newly generated sponge symbiont MAGs, three were retained. The rarer archaeal MAG

(bin B) from *P. ficiformis* contained only 19 of the 46 single-copy COGs (41.3% completeness) and thus was excluded from further analysis. For the selected genomes, we then assessed genome completeness and contamination using CheckM (Parks et al., 2015). Completeness results of both methods are reported in the **Supplementary Table S1** and details for the 46 COGs in **Supplementary Table S2**.

Comparative Genomic Analyses

COG annotations of MAGs were compared to identify functions specific to symbiotic and free-living Thaumarchaeota. To test for significant differences in COG and COG class relative abundances between free-living and sponge-associated Thaumarchaeota we used a two-tailed t-test implemented in STAMP v.2.1.3 (Parks et al., 2014). *P*-values were corrected using false discovery rate corrections according to Benjamin-Hochberg. COGs associated with multiple classes were added to each of the classes. COGs and COG classes with corrected *p*-values <0.05 were considered enriched. We determined enrichment in specific groups of COGs and Pfams with similar functions were tested in the same way using a two-tailed *t*-test and FDR correction.

Taxonomic Classification and Phylogenetic Analyses

A phylogenomic tree for the dataset of 11 symbiotic archaeal, 12 free-living archaeal and 4 outgroup genomes was based 182 single copy marker genes and constructed as follows: A concatenated alignment of protein sequences was generated using PhyloPhlAn2 (Segata et al., 2013). The alignment was then used to infer maximum-likelihood trees with RAxML (Stamatakis, 2014) with the PROTCATLG model of evolution and 1000 bootstrap replications. The best tree was visualized with iTOL (Letunic and Bork, 2019).

Near full-length 16S rRNA gene sequences were obtained from the RAST annotated genomes of the free-living and sponge-associated archaea, except bin A from *P. ficiformis*, which contained a 64 bp fragment at the end of a contig. We generated a targeted small subunit rRNA gene assembly from the *P. ficiformis* metagenome using EMIRGE (Miller et al., 2011) and a 16S rRNA gene clone library from the same DNA as used for the metagenome (see **Supplementary Material** for details). The resulting 11 sequences were used in the phylogenetic analyses.

These sequences were analyzed together with representative sequences of five previously established sponge-specific Thaumarchaeota clusters and the closest related non-host associated sequences [SC174-178 from Simister et al. (2012) as well as the closest environmental sequences for the sponge symbionts as determined by BLAST search (Altschul et al., 1990)]. The 16S rRNA gene sequences of three group I.1b thaumarchaota (*Ca.* Nitrosphaera gargensis Ga92, *Ca*. N. evergladensis SR1, and *Ca*. N. viennensis EN76) were used as outgroup. Sequences were aligned with the SINA aligner version 1.2.11 (Pruesse et al., 2012) and the resulting alignment was manually improved. A maximum likelihood tree was calculated in MEGA X (Kumar et al., 2018) using the Kimura-2-parameter substitution model with a portion of invariant sites and a gamma-shaped distribution of mutation rates, which was the best model for the data according to model test implemented in MEGA X. Sites with $>5\%$ ambiguous or missing data were omitted. The best tree was found using a heuristic search with the NNI algorithm starting from an initial Neighbor-Joining tree. The robustness of the tree was tested using 500 bootstraps replicates.

All genomes were also classified with GTDB-Tk v0.3.2 (Parks et al., 2018) with the default parameter classify_wf to determine the taxonomy of the genomes.

The Relative Abundance of Thaumarchaeal Symbionts in Different Sponge Species

To analyze the abundance of the symbiotic Thaumarchaeota in diverse sponge species from around the world, we searched the Sponge Microbiome Project (SMP) (Thomas et al., 2016; Moitinho-Silva et al., 2017b) part of the Earth Microbiome Project (EMP[3]). The project contains 16S rRNA gene amplicon data from the hypervariable V4 region. We used the 16S rRNA gene sequences from the MAGs as a query for BLASTn 2.2.30 + (Altschul et al., 1990). Sequences matching our genomes with alignment length $\geq$95 bp and a maximum of 1 mismatch (99% similarity) were used to determine the relative abundance of the symbiotic Thaumarchaeota among the SMP samples. Sequences with relative abundances below 0.1% in the samples were not used in the analysis. The binomial (presence/absence) *p*-values enrichment (Moitinho-Silva et al., 2017b) among various sponge species and environmental samples were calculated and corrected as described previously (Burgsdorf et al., 2019). *p*-values were corrected for multiple testing using the p.adjust() R function and the false-discovery (FDR) correction. Samples with corrected *p*-values of $\leq$0.05 were considered enriched. Only environmental samples and sponge species with at least three replicates were included in the analysis. Only samples from sponges collected directly from their habitat were included. Samples of unhealthy or manipulated sponges and samples with unclear taxonomy (e.g., "Porifera") were excluded. The final dataset included 2,131 tissue samples from of 180 sponge species, collected in 33 countries, as well as 305 seawater and 54 marine sediment samples. Boxplots were created using the ggplot2 package in Rstudio.

In addition, we mapped reads from 55 sponge metagenomes obtained 13 sponge species against the genomes of the sponge-associated Thaumarchaeota. All but six metagenomes were quality trimmed (quality threshold 20) with the reformat tool from the BBtools package[4]. Of the remaining six, three metagenomes (IMG accession numbers 3300003175, 3300003251, 3300003254) were quality trimmed as described in Burgsdorf et al. (2015) and the other three (SAMN11333419, SAMN11333441, SAMN12828169) according to Burgsdorf et al. (2019). The quality-filtered and -trimmed reads were mapped against the genome sequences of the sponge-associated Thaumarchaeota using bbmap tool v 37.62 from the BBtools

[3]www.earthmicrobiome.org

[4]https://jgi.doe.gov/data-and-tools/bbtools/

package with default kmer length of 13 and minimum percentage identity cutoff of 95%.

Data Deposition

Metagenomes of *T. swinhoei* and *P. ficiformis* were submitted to the Integrated Microbial Genomes (IMG) database (Markowitz et al., 2012)[5] and are accessible via the accession numbers 3300003175 and 3300003254, respectively. The draft genomes of Thaumarchaeon TS and *Ca*. Nitrosopumilus sp. Pfa have been deposited in GenBank under JAEFDA000000000 and JAEFCZ000000000 accession numbers, respectively. The metagenome of *Hymedesmia* (*Stylopus*) *methanophila* and the draft genome of its thaumarchaeal symbiont were deposited in GenBank under the project number PRJNA475438. 16S rRNA sequences from the clone library of *P. ficiformis* were deposited under accession numbers MT876199-MT876208.

RESULTS AND DISCUSSION

General Genomic Information

The estimated genome completeness based on CheckM ranged from 75.6 to 87.5% with 1.2–8.9% potential contamination for the sponge-associated and 84.6–88.9% with 3.0–9.5% potential contamination for the analyzed free-living Thaumarchaeota genomes (**Supplementary Table S1**). Estimated genome sizes and the number of predicted genes did not differ significantly between genomes from the sponge-associated and free-living archaea (two-tailed *t*-tests, $p > 0.05$). This is in-line with the general observation that genome reduction in symbiotic archaea is not seen or at least not to the same extent as in symbiotic bacteria (Kellner et al., 2018).

One of the few previously identified general patterns for sponge-associated archaeal genomes is a higher GC content compared to free-living Thaumarchaeota (Moeller et al., 2019; Zhang et al., 2019). Our analysis confirmed the overall higher %GC of the symbiotic Thaumarchaeota but also indicated that this pattern is not universal. Apart from the soil-associated *Nitrosotenius chungbukensis* MY2, which had a relatively high GC content (41.8%), the GC content of the analyzed free-living Thaumarchaeota ranged from 32.5 to 34.2%. Seven of the eleven sponge-associated had GC content between 47 and 67%, the exception being *Ca*. Nitrosopumilus cymbastelus (38.4%) and all three deep-sea sponge symbionts (31.4–33.3%). As the GC content is shaped both by phylogeny and the environment (Reichenberger et al., 2015), the Thaumarchaeota with elevated GC content are likely true symbionts and were not transient as a result of the filter feeding of the sponge host. The inclusion of the here generated genome of the thaumarchaeal symbiont from the deep-sea sponge *Hymedesmia* (*Stylopus*) *methanophila* suggests that low GC content is a general feature among deep-sea sponge-associated Thaumarchaeota. This is further supported by the recently published thaumarchaeal genomes from the deep-sea glass sponge *Vazella pourtalesii*, which had GC content between 31.8 and 40.4% (Bayer et al., 2020). The general features of the analyzed genomes are summarized in **Supplementary Table S1**.

Taxonomic Identity and Phylogeny

GTDB-tk placed the genomes generated from *P. ficiformis* and *H.* (*S.*) *methanophila* as new species within the genus *Nitrosopumilus* with an average nucleotide identity (ANI) of <88% to the closest species in GTDB (**Supplementary Table S3**). Their placement within the *Nitrosopumilus* genus was supported by the 16S rRNA gene tree (**Supplementary Figure S1**) and the phylogenomic tree (**Figure 1**). In the 16S rRNA gene tree, all *P. ficiformis* derived sequences grouped together. Hence, we refer to the two novel genomes as *Ca*. Nitrosopumilus sp. PfA and ESC, respectively. They join five previously available sponge-associated genomes in the *Nitrosopumilus* genus. The third new symbiont genome, obtained from *T. swinhoei*, had an ANI of 76.6% to the closest match in GTDB, the sponge symbiont *Cenarchaeum symbiosum*. It was assigned as new species within the *Cenarchaeum* genus by the GTDB-tk classification algorithm but a new genus by pplacer taxonomy (**Supplementary Table S3**). Given this inconsistency, the fact that the genus *Cenarchaeum* is represented by just one species, and that our MAG from *T. swinhoei* did not group in a well-supported clade with *C. symbiosum* neither in the phylogenomic nor the 16S rRNA tree, we refrained from assigning taxonomy to our MAG and refer to it as Thaumarchaeon TS.

None of the three novel genomes were part of previously identified sponge-specific clusters (Simister et al., 2012) in the 16S rRNA gene tree. Of the previously available genomes, *Ca*. Nitrosopongia ianthellae, *Ca.* Nitrosopumilus hexadellus B06 and *Ca*. Nitrosopumilus detritiferus H13 might be part of the sponge-specific cluster (SC) 174 and *Ca*. Cenporiarchaeum stylissum S13 of SC176. These results are in line with previous reports (Moeller et al., 2019; Zhang et al., 2019) and indicate a high degree of specificity of their association with sponges. A major difference between the 16S rRNA tree and the phylogenomic tree was the placement of the clade containing *Ca.* Nitrosopumilus hexadellus B06, *Ca*. Nitrosopumilus detritiferus H13, and *Ca*. Nitrosospongia ianthellae. The taxonomy algorithm of GTDB-tk placed the former two within *Nitrosopumilus* and the latter outside of it (see **Supplementary Table S3**). In the 16S rRNA tree they all grouped within the *Nitrosopumilus* genus, while in the phylogenomic tree they grouped outside of the clade containing the *Nitrosopumilus* and *Nitrosoarchaeum* genera, a placement consistent with previous analysis (see **Supplementary Figure S1B** in Zhang et al., 2019).

Distribution of Thaumarchaeal Symbionts in Sponge and Environmental Samples

To investigate how widespread the analyzed thaumarchaeal symbionts are in sponges, we used their 16S rRNA sequences as queries for BLAST searches against the Sponge Microbiome Project (SMP), an amplicon database of the V4 region of the 16S rRNA gene from >3,500 sponge specimen collected worldwide as well as seawater and sediment samples collected next to the

[5]https://img.jgi.doe.gov/cgi-bin/mer/main.cgi

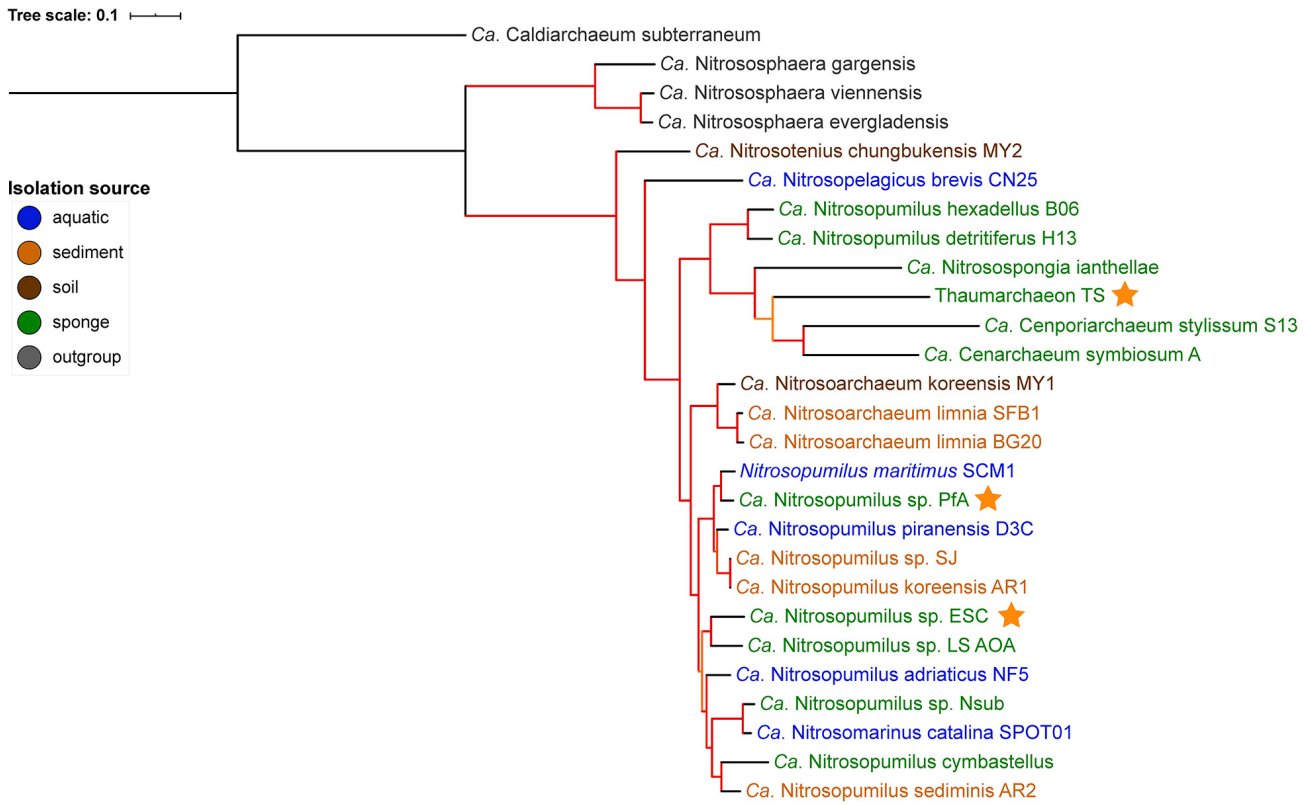

FIGURE 1 | Phylogenomic tree of free-living and sponge-associated Thaumarchaeota. Origin is indicated by font color. Genomes obtained in this study are indicated by an orange star. The tree was calculated using maximum likelihood with the PROTCATLG model of evolution and is based on 182 single copy marker genes (see "Materials and Methods" section for details). Branch colors indicate bootstrap support based on 1,000 repetitions: orange 87–89%, red > 97%. The tree was rooted with *Ca*. Caldiarchaeum subterraneaum (black branch).

sponges (Moitinho-Silva et al., 2017b). The symbionts displayed different specificity ranging from generalists (found in a wide range of sponge species and environmental samples) to specialists (which live in obligate association with a small number of sponge species).

Ca. Nitrosopumilus sp. PfA, Nsub and LS AOA (which have an identical 16S rRNA gene sequence in the V4 region used in the SMP) as well as *Ca*. Nitrosopumilus sp. ESC and *Ca*. N. hexadellus B06 are considered by us to be generalists. They were significantly associated with 23 sponge species (average relative abundance ranging from 0.2 to 20.7%) and were also found in marine sediment samples (0.5%) (Binominal test, FDR-corrected *p*-value <0.1) (**Figure 2A** and **Supplementary Table S4**). We likely underestimated their distribution due to the limited power of the statistical test (e.g., presence in 3 of 3 sponge samples was not significant). Given that the *Nitrosopumilus* species are highly related to each other, this wide-spread distribution is not surprising. To further differentiate them from one another, we performed a read recruitment analyses using 55 sponge metagenomes from 13 sponge species. *Ca*. Nitrosopumilus sp. PfA, Nsub and LS AOA clearly differed in this analysis, as most reads were recruited from sponge metagenomes of the host species (**Supplementary Table S5**). However, they were also detected in other sponge species albeit at a lower level (e.g., one to two orders of magnitude less than in their host sponge) confirming a broad distribution.

Even with the low resolution of the short sequences present in the SMP, we were able to detect an effect of biogeography on the abundance of *Ca*. Nitrosopumilus sp. PfA (**Figure 2B**). It was most abundant in samples from Israel (average relative abundance 7.1%), from where its genome was recovered (**Supplementary Table S6**). Its abundance was lower in samples from Portugal (average relative abundance 1.2%), Spain (2.1%), France (0.4%), Italy (0.3%) and Greece (3.2%). The recruitment analysis of the whole genome supports this trend as *Ca*. Nitrosopumilus sp. PfA recruited a higher proportion of reads from metagenomes of sponge specimens collected from Israel than from a Greek specimen (**Supplementary Table S6**). This is in line with the previous observation that the whole microbial community associated with its host sponge *P. ficiformis* is heavily influenced by biogeography (Burgsdorf et al., 2014).

The other six symbionts (*Ca*. Cenporiarchaeum styllisum S13, *Ca*. Nitrosospongia ianthaella, *Cenarchaeum symbiosum* A, Thaumarchaeon TS, *Ca*. Nitrosopumilus detritiferus H13, *Ca*. Nitrosopumilus cymbastellus) were significantly associated with up to 13 sponge species and not with environmental samples (**Supplementary Table S4**). They also recruited only a small proportion of reads from most sponge metagenomes apart

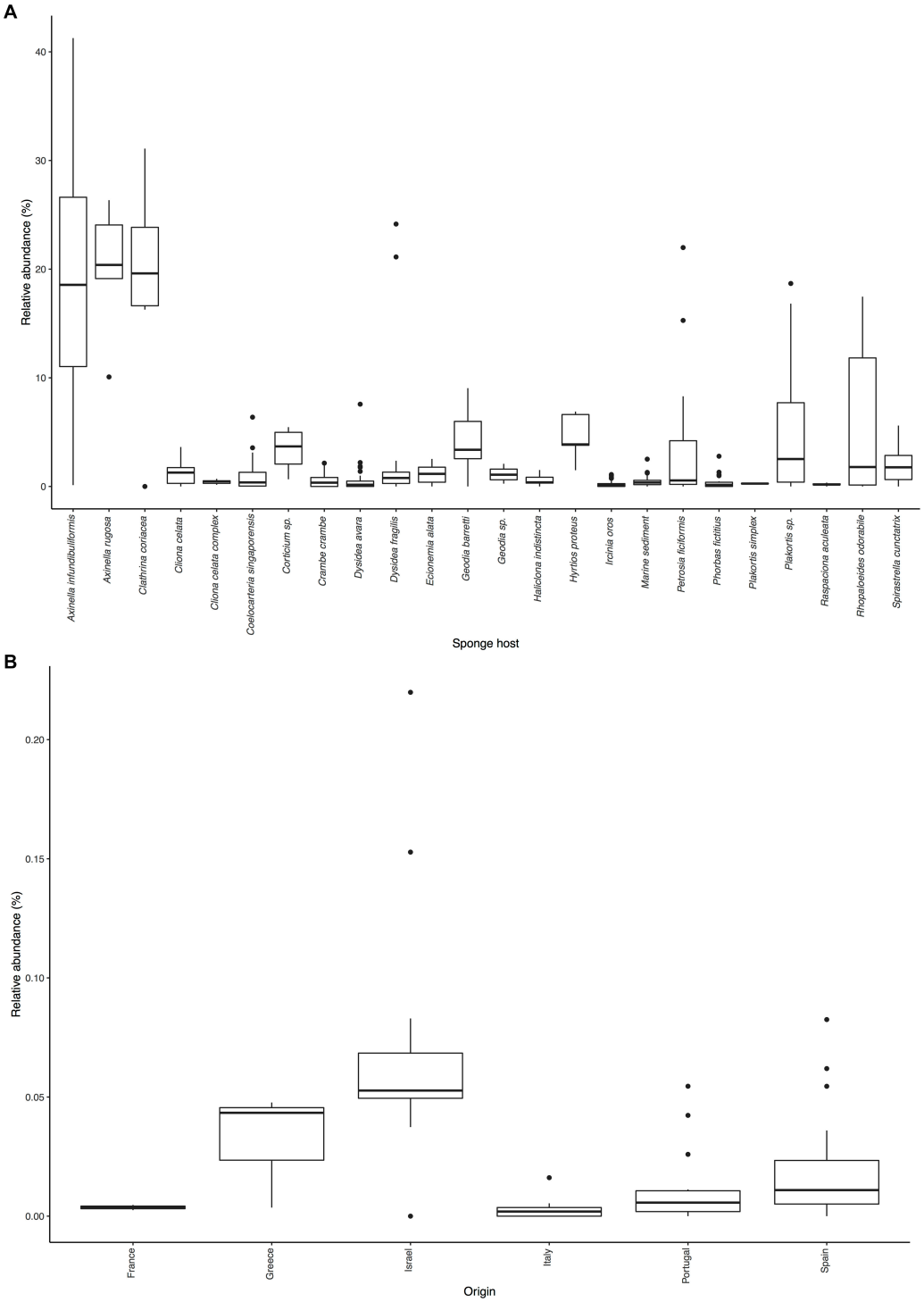

FIGURE 2 | Distribution of *Ca*. Nitrosopumilus sp. PfA.: **(A)** Sponge species and environmental samples with significant association. **(B)** Presence in samples of host sponge *Petrosia ficiformis* from different locations.

from their host sponge (**Supplementary Table S5**). Hence, we regard them as specialists. The degree of specificity varied: *Ca*. Nitrosopumilus detritiferus H13 and *Cenarchaeum symbiosum* A were not significantly associated with any sponge species indicating that they are highly specific symbionts of their host sponges, *Hexadella* cf. *dedritifera* and *Axinella mexicana*, respectively. Both sponge species were not present in the SMP dataset. Host phylogeny had a clear effect on the distribution

of *Ca*. Cenporiarchaeum styllisum S13 and *Ca*. Nitrosospongia ianthaella. The former was only significantly associated with sponges of the Agelasida C6 clade [Figure 2A in Morrow et al. (2012) shows the phylogeny of some of these host sponge species], while the latter was significantly associated only with *Ianthella* basta (22.3%), from which it was recovered, and another Verongiida sponge (16.3%) (**Supplementary Table S4**).

The differentiation between specialist and generalist did not correlate with the symbiont taxonomy as *Ca*. N. detritiferus H13 is highly specific to its host, while the closely related *Ca*. N. sp. hexadellus B06 is a generalist. However, it is possible that *Ca*. N. detritiferus H13 was not found in any sample due to the strict thresholds in our analysis and may be present in very low abundance in other sponges.

Carbon and Nitrogen Metabolism

Except for a recently discovered basal lineage (Aylward and Santoro, 2020), Thaumarchaeota can fix carbon via the highly energy-efficient 3-hydroxypropionate/4-hydroxybutyrate pathway (Könneke et al., 2014; Baker et al., 2020). The pathway was present in all Thaumarchaeota genomes analyzed here (**Supplementary Table S8**). Therefore, symbiotic Thaumarchaeota could potentially contribute to the sponge's carbon demands by transferring fixed carbon to the host.

Thaumarchaeota are also well known for their ability to perform ammonia oxidation (Baker et al., 2020). All genomes contained an ammonia transporter and all, but one had *amoA*, *amoB* and *amoC* genes (**Supplementary Table S11**). Only the newly obtained genome of Thaumarchaeon TS lacked *amoC*. The *amoA* gene and a hypothetical gene (*amoX*) of the ammonia monooxidase gene cluster were found at the end of one contig and a potential *amoB* at the end of another contig. Assuming the same gene synteny (*amoA-amoX-amoC-amoB*) as in most Nitrosopumilaceae (Park et al., 2014), this suggests that *amoC* was not assembled, and Thaumarchaeon TS is likely able to oxidize ammonia. Previous metatranscriptomic and -proteomic studies reported that ammonia oxidation and ammonia transporter genes are expressed in sponges (Liu et al., 2012; Radax et al., 2012b; Moitinho-Silva et al., 2017a; Moeller et al., 2019). Hence, Thaumarchaeota symbionts are likely involved in the removal of ammonia, a waste product of their sponge hosts (Jiménez and Ribes, 2007; Bayer et al., 2008; Hoffmann et al., 2009; Morley et al., 2016).

Some sponges excrete urea as a nitrogenous waste product (Morley et al., 2016). Eight sponge symbionts (and 3 free-living Thaumarchaeota) had a complete set of urease genes including a potential regulator and thus have the potential to produce ammonia from urea. Most of them encoded a transporter of the sodium:solute symporter family, which likely works as urea transporter in Thaumarchaeota (Offre et al., 2014; Carini et al., 2018). A cultured free-living Thaumarchaeota encoding the urease gene cluster was previously shown to grow on urea as a sole form of energy and nitrogen (Carini et al., 2018). Hence the eight sponge symbionts studied here might be able to live off sponge-derived urea as sole energy and nitrogen source.

Ammonia can also be obtained from other sources such as cyanate, creatinine, and aspartate. None of the genomes analyzed here encoded a cyanase, which catalyzes the reaction from cyanate to ammonia. However, as previously reported for some sponge-associated Thaumarchaeota (Moitinho-Silva et al., 2017a; Moeller et al., 2019), we found that all encoded a creatinine amidohydrolase (annotated either by COG1402 or PF02633). This likely enables the sponge symbionts to produce creatine from sponge-derived creatinine. In contrast to previous reports (Moitinho-Silva et al., 2017a; Moeller et al., 2019), we found a potential creatinase domain in all genomes, based on PFAM annotations (PF01321) and BLAST searches. Creatinase converts creatine into urea and sarcosine. In addition, most genomes also had a Xaa-Pro aminopeptidase (COG0006), which has been hypothesized to work as a functional analog to creatinase (Moitinho-Silva et al., 2017a). Eight sponge symbionts and three free-living Thaumarchaeota could potentially use the resulting urea to produce ammonia (**Figure 3**). The alternative pathway of creatinine degradation to sarcosine and ammonia via cytosine deaminase (COG0402), *N*-methylhydantoinase (COG0145), and *N*-carbamoylsarcosine amidase (PF00857) appeared to be incomplete in all but one genome. Cytosine deaminase (COG0402) was only missing from the sponge symbiont Thaumarchaeon TS. *N*-methylhydantoinase (COG0145) was present in three sponge-associated genomes and eight free-living ones. Two genomes (Thaumarchaeon TS, *Ca*. Nitrosopumilus sediminis AR2) had urease genes as well, giving them potentially two different routes from creatinine to ammonia. All other genomes with COG0145 lacked urease genes, while the deep-sea sponge symbiont *Ca*. Nitrosopumilus sp. LS AOA lacked both COG0145 and urease genes. *N*-carbamoylsarcosine amidase (PF00857 part of the isochorismatase family), which catalyzes the conversion to sarcosine, was only present in the sediment-dwelling *Ca*. Nitrosopumilus sediminis AR2. It is possible that other Thaumarchaeota use an alternative enzyme that has not been annotated. All but one genome encoded aspartate-ammonia lyase, which lyses L-aspartate to fumarate and ammonium (COG1027), suggesting the ability to produce ammonium from aspartate. Hence, we conclude that sponge symbionts have several ways to obtain ammonia for oxidation and energy production (**Figure 3**).

For the sponge-associated Thaumarchaeota, nitrogen availability is likely stable, based on the low copy numbers of nitrogen regulatory protein P-II-encoding genes (PF00543) in *Cenarchaeum symbiosum* and *Ca*. Nitrosospongia ianthallae (Moeller et al., 2019). Our analysis indicated that this seems to be true for most sponges and their symbionts. The nitrogen regulatory protein P-II and a negative transcriptional regulator of the NmrA-like family (PF05368) were both more common in the free-living Thaumarchaeota than the sponge-associated ones (**Supplementary Table S11**). An exception was the newly obtained genome of *Ca*. Nitrosopumilus sp. Pfa from the sponge *P. ficiformis*. This genome had 6 and 3 copies of the two regulators, respectively, which was in the range of free-living Thaumarchaeota. While *P. ficiformis* is a high microbial abundance sponge (Vacelet and Donadey, 1977; Gloeckner et al., 2014), indicating a dense associated microbial symbiont community, it obtains its symbionts by horizontal transfer (Maldonado and Riesgo, 2009). Therefore, its symbionts likely

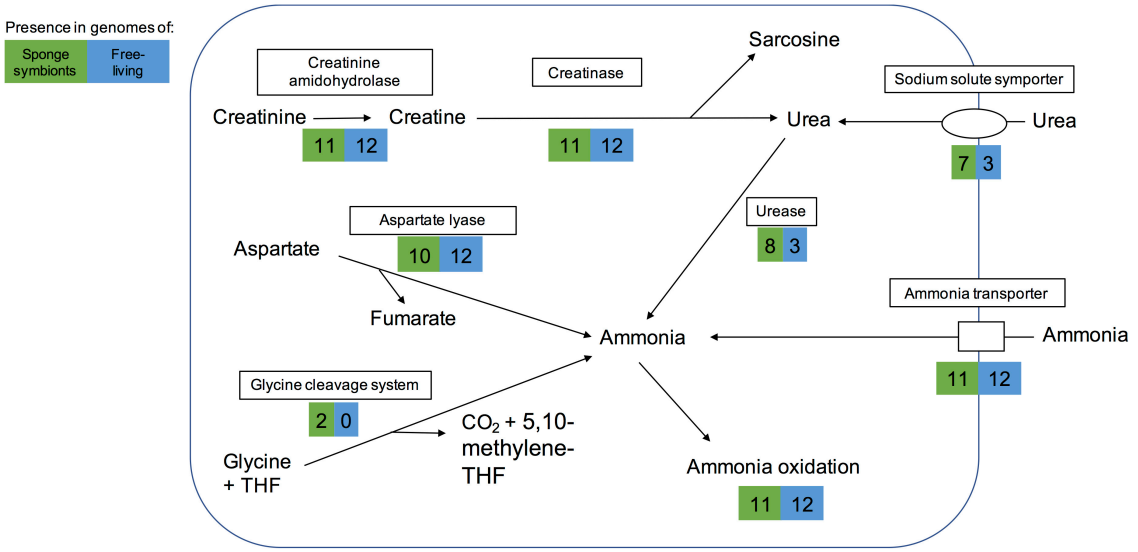

FIGURE 3 | Pathways for generating ammonia in the sponge-associated and free-living Thaumarchaeota. Boxes indicate proteins or pathways. Numbers inside colored boxes represent number of genomes found with the pathway or protein. The maximum number of sponge-associated and free-living MAGs is 11 and 12, respectively. THF, tetrahydrofolate.

need to survive outside of the sponge. This could explain why its thaumarchaeal symbiont needs more genes for nitrogen metabolism regulation than the symbionts of other sponge hosts.

Marine planktonic Thaumarchaeota are mixotrophs that can take up organic nitrogen and carbon (Ouverney and Fuhrman, 2000; Herndl et al., 2005; Kirchman et al., 2007). However, recent experimental data suggest that most assimilate only the nitrogen from amino acids rather than assimilating the complete amino acid (Dekas et al., 2019). As expected, we found ABC-type transporters for di- and oligopeptides (based on COG annotations). An amino acid transporter (COG0531) was almost exclusively limited to the free-living Thaumarchaeota. In contrast, an ABC-type branch-chained amino acid transport system was found in ten of the 11 sponge-associated Thaumarchaeota and absent from the genomes analyzed here of free-living Thaumarchaeota. Given that this transporter type can have a wide range of substrates (Hosie et al., 2002), we do not know which amino acids are taken up by the sponge symbionts. Branch-chained amino acid transporters are expressed in the sponge host based on metatranscriptome and -proteome data (Liu et al., 2012; Moitinho-Silva et al., 2017a; Moeller et al., 2019). They can also be found in the genomes of bacterial sponge symbionts (Gauthier et al., 2016). The expression of thaumarchaeal branched-chain amino acid transporters in the sponge host could indicate a mixotrophic lifestyle of its thaumarchaeal symbionts as previously suggested (Moeller et al., 2019), but experiments are necessary to determine if also the carbon is taken up and not only the amino group as shown for free-living Thaumarchaeota (Dekas et al., 2019).

Vitamin Production

Microbial symbionts of animals are important contributors to their hosts' metabolism through the production of essential nutrients such as vitamins and cofactors (Moya et al., 2008). Free-living marine Thaumarchaeota are known as vitamin B12 producers (Doxey et al., 2015; Heal et al., 2017). Analysis of the microbial community of the sponge *Ircinia ramosa* identified Thaumarchaeota symbionts as producers of riboflavin (vitamin B2), biotin (vitamin B7), and cobalamin (vitamin B12) (Engelberts et al., 2020). Using the SEED annotations, we found that all 23 Thaumarchaeota analyzed, including the sponge symbionts, have the genetic potential to synthesize pyridoxin (B6) and coenzyme F420. Most sponge-associated Thaumarchaeota could also synthesize the vitamins B1, B2, B7, and B12 (**Supplementary Table S12**). Two symbionts lacked genes for the synthesis of these vitamins. *Ca*. Cenporiarchaeum stylissum S13 lacked a single gene in the biosynthetic pathways of vitamin B1 and B7, and 5 genes for vitamin B12 synthesis. As it is the least complete genome, we cannot rule out the possibility that it can produce these vitamins. The newly obtained Thaumarchaeon TS lacked two genes for vitamin B2 synthesis, which likewise might be due to incompleteness. Sponge-associated Thaumarchaeota have, therefore, the genetic potential to provide vitamins to their host and its microbial community.

Functional Differences Between Sponge-Associated and Free-Living Thaumarchaeota

Overall, we identified 1394 COGs from 24 classes in the analyzed genomes. 156 COGs were only found in free-living and 191 only in sponge-associated genomes (**Supplementary Table S7**). Most of the COGs specific to sponge symbionts (88%) were only present in one (132 COGs) or two (36 COGs) genomes. None of these symbiont specific COGs was present in all sponge-associated Thaumarchaeota genomes. A similar trend

was found in the free-living Thaumarchaeota with 78% of the free-living specific COGs being present only in one or two genomes and none present in all (**Figure 4A**).

Sponge-associated genomes were significantly enriched in the COG classes "Replication, recombination & repair," "Nucleotide transport & metabolism," "Amino acid transport & metabolism," "Translation, ribosomal structure & biogenesis," and significantly depleted in the class "Cell wall/membrane/envelope biogenesis" (t-test, two-tailed, FDR-Benjamin-Hochberg corrected *p*-values <0.05) (**Figure 4B**). Some of these differences (e.g., the enrichment in "Replication, recombination & repair" and reduction in "Cell wall/membrane/envelope biogenesis") fit earlier reports from sponge metagenomes (Thomas et al., 2010) and bacterial symbionts (Burgsdorf et al., 2015).

10 COGs were significantly different distributed between sponge symbionts and free-living Thaumarchaeota (*t*-test, FDR corrected *p*-value <0.05) (**Figure 4C**). The four COGs (COG0410, COG0411, COG0559, COG4177) of the sponge-specific ABC branched-chain amino acid transporter were the only COGs significantly enriched in sponge symbionts. The only sponge symbiont without any part of this transporter was *Ca.* Nitrosopumilus sp. Nsub. The six COGs significantly enriched in free-living Thaumarchaeota included the above-mentioned amino acid transporter (COG0531), which was present in 11 of 12 free-living Thaumarchaeota in 1 to 3 copies and among sponge-associated Thaumarchaeota only in *Ca.* Nitrosopumilus cymbastellus in a single copy. Another transporter limited to most free-living Thaumarchaeota was COG1814, a predicted Fe^{2+}/Mn^{2+} transporter of the VIT1/CCC1 family. It was present in 9 of the 12 analyzed free-living Thaumarchaeota genomes (1–2 copies). Other COGs enriched in the free-living Thaumarchaeota included two COGs of class M Cell wall/membrane/envelope biogenesis, COG0451 (a nucleoside-diphosphate-sugar epimerase) and COG3264 (a small-conductance mechanosensitive channel). The former was present in 0–5 copies in sponge symbionts compared to 2–12 copies in free-living ones, whereas the latter had 1 copy in sponge symbionts and 2–3 copies in the free-living ones except for the two pelagic Pacific Ocean Thaumarchaeota, which also had a single copy. Finally, COG2860 and COG3272, for which no function is known (class S), were both enriched in the free-living Thaumarchaeota (**Supplementary Table S8**).

Proteins Involved in Host-Symbiont Interactions

Eukaryotic like protein (ELP) domains in the genomes of sponge symbionts have been suggested to be important for the interaction between the symbiont and their hosts (Fan et al., 2012; Hentschel et al., 2012; Reynolds and Thomas, 2016). ELP domains such as ankyrin repeats and tetratricopeptide (TPR) repeats are often enriched in sponge-associated bacteria and sponge metagenomes (e.g., Fan et al., 2012; Burgsdorf et al., 2015). Hence, we investigated if there are ELPs that are enriched in sponge-associated Thaumarchaeota and if there are ELPs that are only present in sponge-associated Thaumarchaeota. We found no common distribution pattern of ELPs in the sponge symbionts. The distribution of several known and some new ELP candidates appeared to be highly specific. ELPs that were exclusively found in sponge-associated Thaumarchaeota, were encoded in up to three sponge symbiont genomes (**Supplementary Table S11**). Ankyrin repeats, known to be enriched in various sponge-associated bacterial symbionts and sponge metagenomes (Fan et al., 2012; Burgsdorf et al., 2015) and shown to alter phagocytosis of amoeba (Nguyen et al., 2014), were only present in *Ca.* Nitrosopumilus sp. ESC. Other ELP domains previously shown to be enriched in sponge metagenomes (Fan et al., 2012) included leucine-rich repeat DUF285 and PQQ enzyme repeats. These were only present in 2 and 1 thaumarchaeal symbiont genomes, respectively, with the leucine-rich repeat found in two of the novel genomes. We also identified four ELP domains specific to sponge-associated Thaumarchaeota that have not been previously reported from bacterial sponge symbionts. These were the CUB-domain, the EPTP domain, pentatricopeptide repeats (PPR), and annexin. The deep-sea sponge symbionts *Ca.* Nitrosopumilus sp. ESC and LS AOA encoded a protein with a CUB domain. CUB-like domains are widely occurring structural motifs that in eukaryotes are found almost exclusively in extracellular and plasma membrane-associated proteins. They are involved in a wide range of biological functions, including cell signaling, axon guidance and receptor-mediated endocytosis, and are generally involved in the recognition of substrates and binding partners (Blanc et al., 2007). In the *T. swinhoei* symbiont MAG, we identified a genomic sequence that encodes an EPTP domain, which occurs in eukaryotic receptors and secreted proteins (Staub et al., 2002). PPR domains, structurally similar to ELP tetratricopeptide repeats that are well known from bacterial sponge symbionts (e.g., Fan et al., 2012; Burgsdorf et al., 2015), were encoded by *Ca.* Nitrosopumilus detritiferus H13 and *Ca.* Nitrosopumilus cymbastellus. PPRs are common in plant genomes. They play essential roles in posttranscriptional processes in mitochondria and chloroplasts (Lurin et al., 2004). Annexin, which is involved in the endocytosis and exocytosis of vesicles in eukaryotes (Gerke and Moss, 1997), was encoded by *Ca.* Nitrosopumilus cymbastellus.

Several other ELP domains were found in the genomes of some sponge-associated Thaumarchaeota, but also present in several free-living Thaumarchaeota. These included cadherins, TPRs, NHL repeats, HYR domains, Ig domains, TIR domains and WD40-like repeats (**Supplementary Table S11**). Fibronectin type III, which was previously found to be enriched in sponge-metagenomes (Fan et al., 2012) and the sponge cyanobacterial symbiont *Ca.* Synechococcus feldmannii (Burgsdorf et al., 2019), and von Willebrand factor type A domains did not show any obvious abundance difference between sponge-associated and free-living Thaumarchaeota. Laminin domains, which are enriched in Poribacteria (Siegl et al., 2011), were present in four sponge-associated Thaumarchaeota but were not enriched compared to free-living Thaumarchaeota. Overall, ELPs are likely to play a role in the recognition of sponge-associated Thaumarchaeota as symbionts by their host sponge. In contrast to their bacterial counterparts, we found high inter-species variability in ELPs, such that each ELP was only shared among a

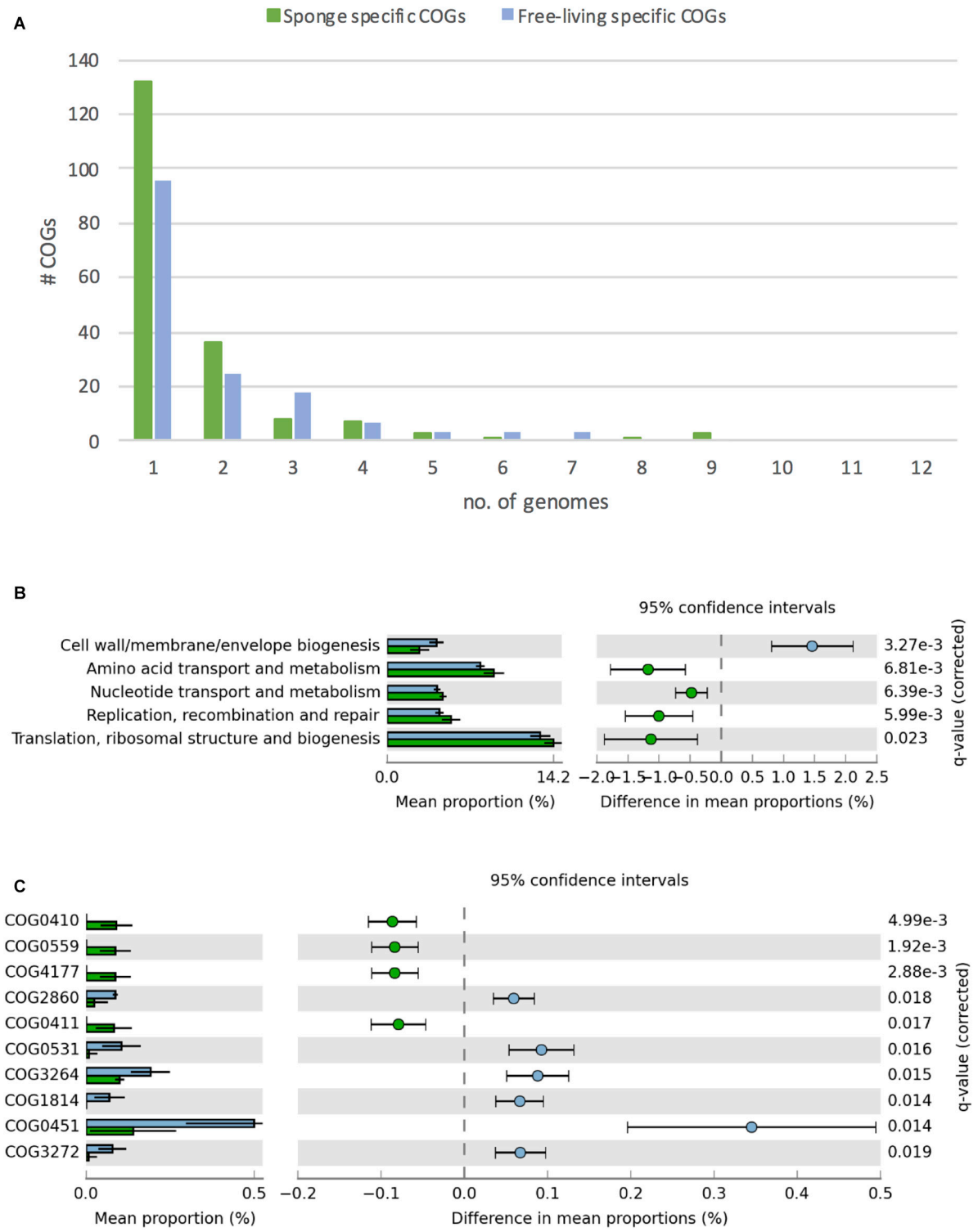

FIGURE 4 | Differences in COGs between free-living (blue) and sponge-associated (green) Thaumarchaeota. **(A)** Distribution of free-living and sponge symbiont specific COGs among genomes. The maximum for free-living Thaumarchaeota is 12 and for sponge-associated ones is 11 genomes. **(B)** COG classes with significant differences in free-living vs. sponge symbiont genomes (*t*-test, FDR Benjamin-Hochberg corrected *p*-value <0.05). **(C)** Distribution of COGs with significant differences (*t*-test, FDR Benjamin-Hochberg corrected *p*-value <0.05). COGs 0410, 0559, 4177, and 0411 are part of the branched chain amino acid transporter. Free-living enriched COGs were related to an amino acid transporter (COG0531), a Fe^{2+}/Mn^{2+} transporter (COG1814), a small-conductance mechanosensitive channel (COG2364), a nucleoside-diphosphate-sugar epimerase (COG0451), an uncharacterized membrane protein (COG2860), and a protein of unknown function (COG3272).

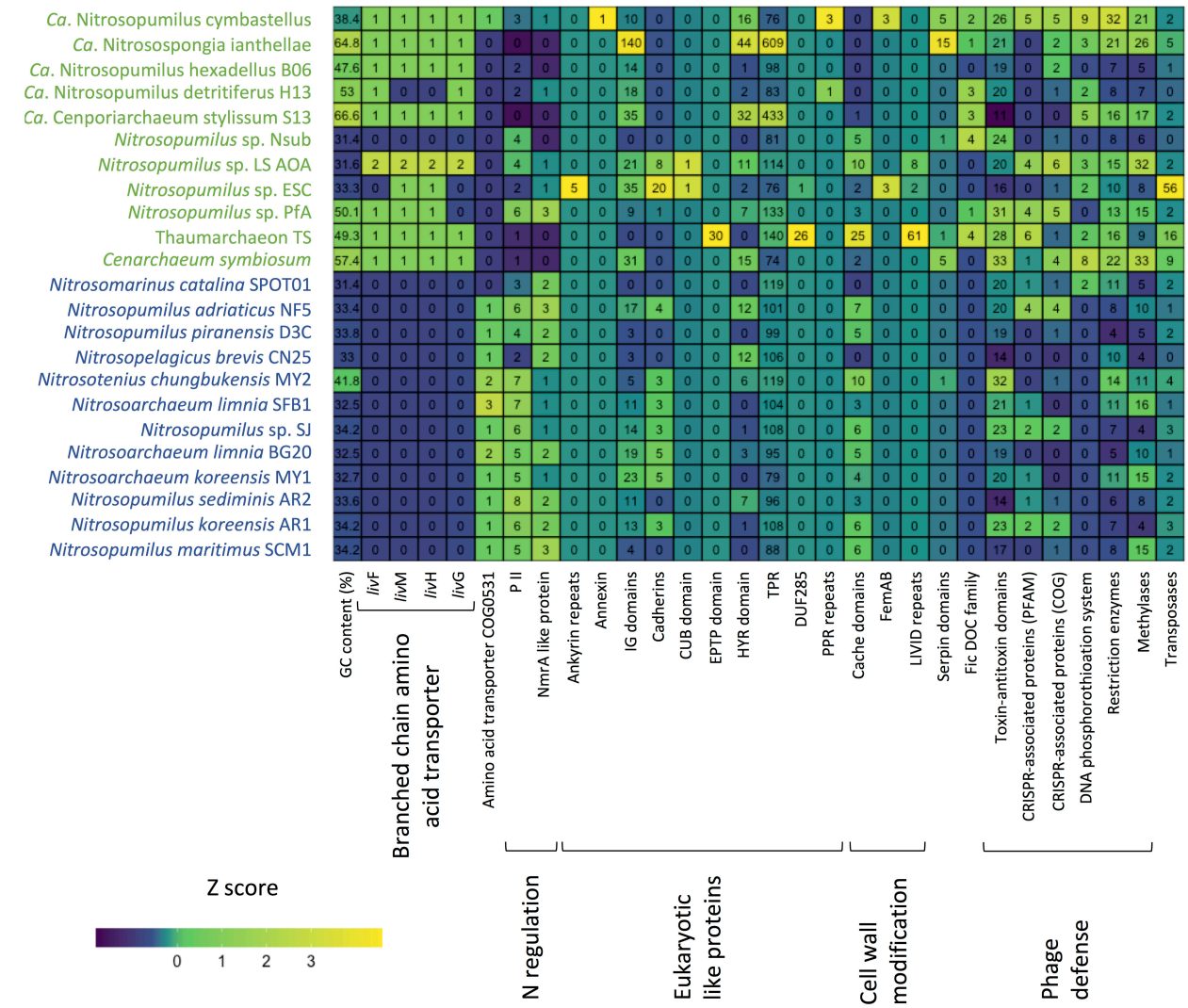

FIGURE 5 | Heatmap summarizing functional differences in the genomic repertoire between sponge-associated (green) and free-living (blue) Thaumarchaeota. Numbers indicate copies per MAG.

few genomes at most (**Figure 5**). This suggests Thaumarcheaota employ diverse strategies for interacting with their hosts.

Another proposed way for bacterial sponge symbionts to avoid digestion by their sponge host is the modification of the O-antigen on the cell wall surface. The absence of rhamnose in the sponge-associated cyanobacterium *Ca.* Synechococcus spongiarum has been suggested to aid in the evasion of sponge-phagocytosis and phage predation (Burgsdorf et al., 2015). Thaumarchaeota have surface layer glycoproteins (Rodrigues-Oliveira et al., 2017) and rhamnose can be part of these proteins in archaea (Schäffer and Messner, 2004). Of the analyzed genomes, eight sponge-associated thaumarchaeal genomes missed all four COGs necessary for rhamnose biosynthesis (COG1898 dTDP-4-dehydrorhamnose 3,5-epimerase and related enzymes, COG1091 dTDP-4-dehydrorhamnose reductase, COG1088 dTDP-D-glucose 4,6-dehydratase, COG1209 dTDP-glucose pyrophosphorylase). The other three symbiont genomes (*Ca.* Nitrosopumilus sp. LS AOA, *Ca.* Nitrosopumilus cymbastellus, *Ca.* Nitrosospongia ianthellae) encoded two out of these four COGs. Eleven of the 12 free-living Thaumarchaeota had at least two of these COGs, with only *Ca.* Nitrosomarinus catalina SPOT01 having none (**Supplementary Table S8**). Hence, the modified surface layer glycoproteins might aid thaumarchaeal symbionts to evade digestion by their host sponges as proposed for *Ca.* Synechococcus spongiarum.

Another six Pfam annotated domains involved in outer cell wall modifications were present only in one to three symbiont genomes and absent from the genomes of free-living Thaumarchaeota. The OmpA domain is known from outer membrane proteins of Gram-negative bacteria. It is necessary for pathogenesis in *E. coli* and can interact with host receptor molecules (Selvaraj et al., 2007). FemAB domains

are involved in cell wall synthesis and mutations in them decrease antibiotic resistance in *Streptococcus* strains (Rohrer and Berger-Bächi, 2003). The PEGA domain has been previously found in archaea and is structurally similar to surface-layer proteins (Adindla et al., 2004). Likewise, the LVIVD domain, which was found in three sponge symbionts (**Figure 5**), especially Thaumarchaeon TS, is a component of bacterial and archaeal surface proteins (Adindla et al., 2004). Two other sponge-specific domain annotations involved in cell wall modifications are related to bacterial O-antigen modifications and might have similar roles in archaeal surface layer proteins. The O-antigen polysaccharide polymerase Wzy links O-units via a glycosidic linage to form a long O-antigen. WbqC-like proteins might be involved in O-antigen production in bacteria. Finally, we found cache domain (3 Pfams, two sponge-specific) to be enriched in some sponge-associated archaea. Cache domains are extracellular domains likely having a role in small-molecule recognition and might be potential chemotaxis receptors (Anantharaman and Aravind, 2000). Given that the six sponge symbiont-specific domains and the enriched cache domains are on the cell surface, they are potential candidates for specific recognition between the sponge-associated Thaumarchaeota and their host sponges and thus might help the symbiont avoid digestion by its host.

Sponge-associated MAGs also encoded proteins that can potentially modulate the host metabolism and the immune system. Seven of the 11 sponge symbionts and none of the free-living Thaumarchaeota had proteins with a FIC domain (PF02661) (**Figure 5**). This domain family is widespread in bacteria and has multiple functions including the toxin component of toxin-antitoxin systems. In pathogenic bacteria, the FIC domain bearing toxins can divert host cell processes (Veyron et al., 2018). The FIC domain was overrepresented in sponge metagenomes compared to seawater metagenomes (Fan et al., 2012). Previous work identified 15 genes in *Ca*. Nitrosospongia ianthellae encoding serpin domains (Moeller et al., 2019). Serpins belong to a class of irreversible substrate inhibitors of proteases, often of the serine class. Here, we show that the domain was also present in the genomes of the sponge symbionts *Cenarchaeum symbiosum* (5 copies), *Ca*. Nitrosopumilus cymbastellus (5 copies), Thaumarchaeon TS (1 copy), *Ca*. Nitrosopumilus sp. Nsub (1 copy), while among free-living Thaumarchaeota, only the soil-inhabiting *Nitrosotenuis chungkubensis* MY2 had a single copy (**Figure 5**). Serine proteases are found as part of the innate immune system in some sponges (Riesgo et al., 2014) and are known activators of the lectin complement pathway, a highly sophisticated host defense system that detects, contains and kills pathogens (Iwaki et al., 2011). In gut bacteria, serpins have been suggested to act as protection against exogenous proteolysis (Ivanov et al., 2006). Therefore, symbiotic Thaumarchaeota might use serpins as protection against digestion by sponge amoebocytes, the "feeding cells" of sponges, with which they must coexist within the sponge mesohyl matrix.

Defense Against Phages

Due to the high filtration activity of their host, sponge symbionts are likely exposed to a vast number of phages (Hadas et al., 2006; Jahn et al., 2019). Sponge metagenomes and genomes of bacterial symbionts are often enriched in genes involved in the defense against phages such as CRISPR-Cas, restriction-modification systems, and toxin-antitoxin modules (Fan et al., 2012; Horn et al., 2016; Podell et al., 2019). This does not seem to be the case in sponge-associated Thaumarchaeota. COG class V "Defense mechanisms" did not differ significantly between the free-living and sponge-associated Thaumarchaeota analyzed. While we found components of several toxin-antitoxin modules, sponge-associated Thaumarchaeota genomes were overall not enriched in toxin-antitoxin modules, when compared to free-living Thaumarchaeota ones (irrespective of whether COG or Pfam annotations were used, both two-tailed *t*-test, $p > 0.05$) (**Supplementary Tables S8, S11**, respectively). Some modules were present only in sponge-associated Thaumarchaeota genomes. This included a putative abortive infection system (AbiE family), which has only recently been reported in archaea (Li et al., 2018) (**Supplementary Table S11**). However, another domain, an ATPase domain associated with the cellular activity (AAA) (PF13304), which in some cases has been associated with the Abi toxin, was found in all genomes in varying numbers (**Supplementary Table S11**).

While bacterial sponge symbionts are often enriched in CRISPR-associated genes compared to closely related free-living bacteria (Fan et al., 2012; Horn et al., 2016), this is not clear in sponge-associated Thaumarchaeota. Sponge-associated Thaumarchaeota did not differ significantly from free-living Thaumarchaeota in cas domains when using Pfam annotations (two-tailed *t*-test, $p = 0.31$), but were enriched in CRISPR-associated genes when using COG annotations (two-tailed *t*-test, $p = 0.047$). Irrespective of COG or Pfam annotation, all free-living Thaumarchaeota had between 0 and 2 CRISPR-associated genes except for *Nitrosopumilus adriaticus* NF5, which had 4. The sponge-associated Thaumarchaeota *Ca*. Nitrosopumilus sp. PfA, *Ca*. Nitrosopumilus sp. LS AOA, *Ca*. Nitrosopumilus cymbastellus, and, depending if COG or Pfam annotations were used either *Cenarchaeum symbiosum* or Thaumarchaeon TS had 4–6 CRISPR-associated genes surpassing most free-living Thaumarchaeota (**Figure 5** and **Supplementary Tables S8, S11**, respectively). Hence enrichment in CRISPR-associated genes is present in some Thaumarchaeota, but not as common as it appears to be for bacterial sponge symbionts [but see recent work on Verrucomicrobia for an exception (Sizikov et al., 2020)].

The DNA phosphorothioation system, which in archaea likely works as a defense system against phage (Xiong et al., 2019), appeared to be more prevalent in sponge-associated than free-living Thaumarchaeota. In addition to the previously reported presence in *Ca*. Nitrosospongia ianthellae and *Ca*. Nitrosopumilus cymbastellus (Moeller et al., 2019), we found at least one of *dndB* and *dndE* in the Thaumarchaeon TS and in *Ca*. Cenporiarchaeum stylissum S13 MAGs. Only one free-living Thaumarchaeota species, *Ca*. Nitrosomarinus catalina SPOT01, encoded this system (Ahlgren et al., 2017). When including the protein domains DUF262 (PF03235) and DUF1524 (PF07510), which have been previously found in *dndB* genes

and identified as components of restriction-modification systems (Makarova et al., 2011; Moeller et al., 2019), four more sponge symbionts were found to have at least part of the system: *Cenarchaeum symbiosum*, *Ca*. Nitrosopumilus sp. ESC, *Ca*. Nitrosopumilus sp. LS AOA and *Ca*. Nitrosopumilus detritiferus H13. Metagenome analyses showed that the DUF262 and DUF1524 domains are enriched in sponge metagenomes (Fan et al., 2012; Horn et al., 2016).

The sponge-associated Thaumarchaeota were also enriched in restriction enzymes (two-tailed *t*-test, $p = 0.006$). Nine of the 19 different Pfam domains related to restriction enzymes were found only in the sponge-associated Thaumarchaeota, whereas two domains were specific to the free-living Thaumarchaeota. The same significant difference was found when analyzing domains of methylases and methyltransferase known to be involved in restriction modifications systems (two-tailed *t*-test, $p = 0.037$). Here three of nine domains were specific to sponge symbionts. The number of domains of both elements of the restriction-modification system was significantly correlated (Pearson's r 0.7007, $p < 0.001$). The enrichment of restriction-modification systems in some sponge-associated Thaumarchaeota was noted previously (Moeller et al., 2019; Zhang et al., 2019). This enrichment is also seen in bacterial members of sponge microbial communities relative to free-living bacteria (Horn et al., 2016; Slaby et al., 2017; Burgsdorf et al., 2019), which points to the importance of phage defense mechanisms within the sponge host for both these microbial domains.

Potential for Horizontal Gene Transfer

Many bacterial sponge symbionts are enriched in transposases (Fan et al., 2012), which suggests enhanced horizontal gene transfer between members of the sponge microbiome (Pita et al., 2018). This does not seem to be the case for the majority of sponge-associated Thaumarchaeota, as the mobilome COG class X did not differ significantly between sponge-associated and free-living Thaumarchaeota. However, 11 of the 14 transposases (or inactive derivatives) annotated by COG and 13 out of 16 domains annotated by Pfam were restricted to the sponge symbiont genomes. Sponge-associated Thaumarchaeota had a larger variability in transposases ranging from 0 to 21 COGs and 0 to 56 Pfam domains per genome, compared to only 0–3 in genomes of free-living ones (with the exception of the soil-inhabiting *Nitrosotenius chungbukensis* MY2) (**Figure 5** and **Supplementary Tables S8, S11**). Based on COG annotations, the three newly analyzed genomes and *Ca*. Nitrosopumilus cymbastellus had more transposases than their marine free-living counterparts ranging from 5 COG annotated transposases in *Ca*. Nitrosopumilus cymbastellus to 21 in Thaumarchaeon TS (**Supplementary Table S8**). Pfam annotations but not COG annotations confirmed the previously reported enrichment of transposases in *Ca*. Nitrosopumilus ianthallae and *C. symbiosum* (Moeller et al., 2019) (**Supplementary Tables S8, S11**). Together this indicates a greater potential for gene transfer in some thaumarchaeal sponge symbionts, though it remains unclear why the high number of transposases is restricted to these few sponge-associated thaumarchaeal genomes.

Metabolic Functions Specific to the Novel Sponge Archaea

The two novel sponge-associated Thaumarchaeota genomes *Ca*. Nitrosopumilus sp. ESC and Thaumarchaeon TS harbored the glycine cleavage system, which was absent from all other genomes. It consisted of the genes *gcvHPT* [COG0509, COG0403/COG1003, COG0404 and COG1249, the dihydrolipoamide dehydrogenase (E3)] together with the genes necessary for lipoate production (COG0095, COG0320), which binds to GcvH. The system enables the reversible reaction glycine + tetrahydrofolate (THF) + $NAD^+ \leftrightarrow$ 5,10-methylene-THF + CO_2 + NH_3 + NADH + H^+. It has been recently found in the genomes of free-living deep-sea Thaumarchaeota (León-Zayas et al., 2015; Wang et al., 2019; Zhong et al., 2020). The methylene-THF can be used in the biosynthesis of purine and methionine (León-Zayas et al., 2015). When working in the direction of glycine cleavage, the reaction provides energy. Moreover, via the downstream oxidation of ammonia and regeneration of THF, it would enable an additional way to generate energy for the sponge symbionts (**Figure 3**).

Out of the 23 analyzed genomes, the Thaumarchaeon TS from the Red Sea sponge *T. swinhoei* and *Ca*. Nitrosopumilus sp. PfA from *P. ficiformis* collected from the Southeastern Mediterranean Sea were the only ones that encoded exopolyphosphatases (COG0248). This in addition to a high-affinity phosphate transporter consisting of PhoU, PstSABC, which was also present in six other sponge symbionts and six free-living Thaumarchaeota [based on SEED annotations (**Supplementary Table S13**)]. Phosphorus sequestration in the form of polyphosphate by microbial symbionts has been shown in Caribbean sponges (Zhang et al., 2015). In the case of the symbiont of *P. ficiformis*, the sponge was collected in the Southeastern Mediterranean Sea, which is known for phosphorus limitation (Krom et al., 1991). It is possible that *Ca*. Nitrosopumilus sp. PfA has a free-living phase (see above), during which it experiences phosphorus stress.

CONCLUSION

Free-living marine Thaumarchaeota are known as ammonium oxidizers and are key suppliers of vitamins in microbial communities. Our comparative genome analysis supports the previous notion that these traits are conserved in sponge-associated Thaumarchaeota. Some genomic features known to be enriched in bacterial sponge symbionts were also enriched in sponge-associated archaea (e.g., restriction-modification systems), suggesting that the sponge environment selects for these features. However, adaptations to life inside the sponge host appear to diverge in Thaumarchaeota symbionts, as many of these traits were specific only to a few symbiont genotypes. For example, we found a glycine cleavage system and polyphosphatases only in the three novel symbiont genomes. Thus, the analysis of additional novel Thaumarchaeota symbionts is likely to reveal more unique adaptations to the symbiosis with sponges.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

MH, IB, and LS conceived the study. IB and MR-B collected samples and isolated DNA for metagenome sequencing. MH, IB, KH, and MR-B did bioinformatics analyses. MH and LS wrote the manuscript with the contribution of all the authors. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded in part by the Gordon and Betty Moore Foundation, through Grant GBMF9352 to LS. MH was supported by an Inter-Institutional fellowship of the University of Haifa and a Helmsley fellowship. MR-B was supported by the Israeli Science Foundation (ISF grant 913/19) and thanks Nicole Dubilier, the Max Planck Institute for Marine Microbiology and the Max Planck Society for further support and funding.

ACKNOWLEDGMENTS

We acknowledge the University of Chicago Research Computing Center for support in this work and thank the staff of the Inter-University Institute in Eilat, Israel, for their help during this study. *Petrosia ficiformis* and *Theonella swinhoei* sponges were collected in compliance with the 2012/38390 and 2013/3839 permits from the Israel Nature and National Parks Protection Authority. We are grateful to the Mexican authorities for granting permission to collect *Hymedesmia* (*Stylopus*) *methanophila* samples (permission of DGOPA: 02540/14 from 5 November 2014) in the southern Gulf of Mexico and thank the MARUM and all individuals who helped during the R/V Meteor research cruise M114.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2020.622824/full#supplementary-material

REFERENCES

Adindla, S., Inampudi, K. K., Guruprasad, K., and Guruprasad, L. (2004). Identification and analysis of novel tandem repeats in the cell surface proteins of Archaeal and Bacterial genomes using computational tools. *Comp. Funct. Genomics* 5, 2–16. doi: 10.1002/cfg.358

Ahlgren, N. A., Chen, Y., Needham, D. M., Parada, A. E., Sachdeva, R., Trinh, V., et al. (2017). Genome and epigenome of a novel marine *Thaumarchaeota* strain suggest viral infection, phosphorothioation DNA modification and multiple restriction systems. *Environ. Microbiol.* 19, 2434–2452. doi: 10.1111/1462-2920.13768

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

Anantharaman, V., and Aravind, L. (2000). Cache – a signaling domain common to animal Ca2+-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends Biochem. Sci.* 25, 535–537. doi: 10.1016/S0968-0004(00)01672-8

Aylward, F. O., and Santoro, A. E. (2020). Heterotrophic Thaumarchaea with small genomes are widespread in the dark ocean. *mSystems* 5:e00415-20. doi: 10.1128/mSystems.00415-20

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

Baker, B. J., De Anda, V., Seitz, K. W., Dombrowski, N., Santoro, A. E., and Lloyd, K. G. (2020). Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* 5, 887–900. doi: 10.1038/s41564-020-0715-z

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., et al. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19, 455–477. doi: 10.1089/cmb.2012.0021

Bayer, K., Schmitt, S., and Hentschel, U. (2008). Physiology, phylogeny and *in situ* evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environ. Microbiol.* 10, 2942–2955. doi: 10.1111/j.1462-2920.2008.01582.x

Bayer, K., Busch, K., Kenchington, E., Beazley, L., Franzenburg, S., Michels, J., et al. (2020). Microbial strategies for survival in the glass sponge *Vazella pourtalesii*. *mSystems* 5:e00473-20. doi: 10.1128/mSystems.00473-20

Björk, J. R., Astudillo-García, C., Archie, E., and Montoya, J. M. (2019). Vertical transmission of sponge microbiota is weak and inconsistent. *Nat. Ecol. Evol.* 3, 1172–1183. doi: 10.1101/425009

Blanc, G., Font, B., Eichenberger, D., Moreau, C., Ricard-Blum, S., Hulmes, D. J. S., et al. (2007). Insights into How CUB Domains Can Exert Specific Functions while Sharing a Common Fold CONSERVED AND SPECIFIC FEATURES OF THE CUB1 DOMAIN CONTRIBUTE TO THE MOLECULAR BASIS OF PROCOLLAGEN C-PROTEINASE ENHANCER-1 ACTIVITY. *J. Biol. Chem.* 282, 16924–16933. doi: 10.1074/jbc.M701610200

Burgsdorf, I., Erwin, P. M., López-Legentil, S., Cerrano, C., Haber, M., Frenk, S., et al. (2014). Biogeography rather than association with cyanobacteria structures symbiotic microbial communities in the marine sponge *Petrosia ficiformis*. *Front. Microbiol.* 5:529. doi: 10.3389/fmicb.2014.00529

Burgsdorf, I., Handley, K. M., Bar-Shalom, R., Erwin, P. M., and Steindler, L. (2019). Life at home and on the roam: genomic adaptions reflect the dual lifestyle of an intracellular, facultative symbiont. *mSystems* 4:e00135-16. doi: 10.1128/mSystems.00057-19

Burgsdorf, I., Slaby, B. M., Handley, K. M., Haber, M., Blom, J., Marshall, C. W., et al. (2015). Lifestyle evolution in cyanobacterial symbionts of sponges. *mBio* 6:e00391-15. doi: 10.1128/mBio.00391-15

Carini, P., Dupont, C. L., and Santoro, A. E. (2018). Patterns of thaumarchaeal gene expression in culture and diverse marine environments. *Environ. Microbiol.* 20, 2112–2124. doi: 10.1111/1462-2920.14107

de Goeij, J. M., van Oevelen, D., Vermeij, M. J. A., Osinga, R., Middelburg, J. J., de Goeij, A. F. P. M., et al. (2013). Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* 342, 108–110. doi: 10.1126/science.1241981

Dekas, A. E., Parada, A. E., Mayali, X., Fuhrman, J. A., Wollard, J., Weber, P. K., et al. (2019). Characterizing chemoautotrophy and heterotrophy in marine Archaea and bacteria with single-cell multi-isotope NanoSIP. *Front. Microbiol.* 10:2682. doi: 10.3389/fmicb.2019.02682

Dick, G. J., Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., et al. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* 10:R85. doi: 10.1186/gb-2009-10-8-r85

Doxey, A. C., Kurtz, D. A., Lynch, M. D., 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

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26, 2460–2461. doi: 10.1093/bioinformatics/btq461

Engelberts, J. P., Robbins, S. J., de Goeij, J. M., Aranda, M., Bell, S. C., and Webster, N. S. (2020). Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME J.* 14, 1100–1110. doi: 10.1038/s41396-020-0591-9

Erwin, P. M., and Thacker, R. W. (2008). Phototrophic nutrition and symbiont diversity of two Caribbean sponge–cyanobacteria symbioses. *Mar. Ecol. Prog. Ser.* 362, 139–147. doi: 10.3354/meps07464

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N. S., et al. (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 109, E1878–E1887. doi: 10.1073/pnas.1203287109

Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., Lartillot, N., et al. (2017). Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr. Biol.* 27, 3864–3870.e4. doi: 10.1016/j.cub.2017.11.008

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., et al. (2010). The Pfam protein families database. *Nucleic Acids Res.* 38, D211–D222. doi: 10.1093/nar/gkp985

Galperin, M. Y., Makarova, K. S., Wolf, Y. I., and Koonin, E. V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* 43, D261–D269. doi: 10.1093/nar/gku1223

Gao, Z.-M., Wang, Y., Tian, R.-M., Wong, Y. H., Batang, Z. B., Al-Suwailem, A. M., et al. (2014). Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont "*Candidatus* Synechococcus spongiarum. *mBio* 5:e00079-14. doi: 10.1128/mBio.00079-14

Gauthier, M.-E. A., Watson, J. R., and Degnan, S. M. (2016). Draft genomes shed light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge *Amphimedon queenslandica*. *Front. Mar. Sci.* 3:196. doi: 10.3389/fmars.2016.00196

Gerke, V., and Moss, S. E. (1997). Annexins and membrane dynamics. *Biochim. Biophys. Acta* 1357, 129–154. doi: 10.1016/S0167-4889(97)00038-4

Gloeckner, V., Wehrl, M., Moitinho-Silva, L., Gernert, C., Schupp, P., Pawlik, J. R., et al. (2014). The HMA-LMA dichotomy revisited: an electron microscopical survey of 56 sponge species. *Biol. Bull.* 227, 78–88. doi: 10.1086/BBLv227n1p78

Hadas, E., Marie, D., Shpigel, M., and Ilan, M. (2006). Virus predation by sponges is a new nutrient-flow pathway in coral reef food webs. *Limnol. Oceanogr.* 51, 1548–1550. doi: 10.4319/lo.2006.51.3.1548

Hallam, S. J., Konstantinidis, K. T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., et al. (2006). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18296–18301. doi: 10.1073/pnas.0608549103

Handley, K. M., VerBerkmoes, N. C., Steefel, C. I., Williams, K. H., Sharon, I., Miller, C. S., et al. (2013). Biostimulation induces syntrophic interactions that impact C, S and N cycling in a sediment microbial community. *ISME J.* 7, 800–816. doi: 10.1038/ismej.2012.148

Heal, K. R., Qin, W., Ribalet, F., Bertagnolli, A. D., Coyote-Maestas, W., Hmelo, L. R., et al. (2017). Two distinct pools of B12 analogs reveal community interdependencies in the ocean. *Proc. Natl. Acad. Sci. U.S.A.* 114, 364–369. doi: 10.1073/pnas.1608462114

Hentschel, U., Piel, J., Degnan, S. M., and Taylor, M. W. (2012). Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol.* 10, 641–654. doi: 10.1038/nrmicro2839

Herndl, G. J., Reinthaler, T., Teira, E., van Aken, H., Veth, C., Pernthaler, A., et al. (2005). Contribution of Archaea to total prokaryotic production in the deep atlantic ocean. *AEM* 71, 2303–2309. doi: 10.1128/AEM.71.5.2303-2309.2005

Hoffmann, F., Radax, R., Woebken, D., Holtappels, M., Lavik, G., Rapp, H. T., et al. (2009). Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ. Microbiol.* 11, 2228–2243. doi: 10.1111/j.1462-2920.2009.01944.x

Horn, H., Slaby, B. M., Jahn, M. T., Bayer, K., Moitinho-Silva, L., Förster, F., et al. (2016). An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial Metagenomes. *Front. Microbiol.* 7:1751. doi: 10.3389/fmicb.2016.01751

Hosie, A. H. F., Allaway, D., Galloway, C. S., Dunsby, H. A., and Poole, P. S. (2002). *Rhizobium leguminosarum* has a second general amino acid permease with unusually broad substrate specificity and high similarity to branched-chain amino acid transporters (Bra/LIV) of the ABC family. *J. Bacteriol.* 184, 4071–4080. doi: 10.1128/JB.184.15.4071-4080.2002

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

Hyatt, D., LoCascio, P. F., Hauser, L. J., and Uberbacher, E. C. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28, 2223–2230. doi: 10.1093/bioinformatics/bts429

Ivanov, D., Emonet, C., Foata, F., Affolter, M., Delley, M., Fisseha, M., et al. (2006). A serpin from the gut bacterium *Bifidobacterium longum* inhibits eukaryotic elastase-like serine proteases. *J. Biol. Chem.* 281, 17246–17252. doi: 10.1074/jbc.M601678200

Iwaki, D., Kanno, K., Takahashi, M., Endo, Y., Matsushita, M., and Fujita, T. (2011). The role of mannose-binding lectin-associated serine protease-3 in activation of the alternative complement pathway. *J. Immunol.* 187, 3751–3758. doi: 10.4049/jimmunol.1100280

Jahn, M. T., Arkhipova, K., Markert, S. M., Stigloher, C., Lachnit, T., Pita, L., et al. (2019). A phage protein aids bacterial symbionts in eukaryote immune evasion. *Cell Host Microbe* 26, 542–550.e5. doi: 10.1016/j.chom.2019.08.019

Jiménez, E., and Ribes, M. (2007). Sponges as a source of dissolved inorganic nitrogen: nitrification mediated by temperate sponges. *Limnol. Oceanogr.* 52, 948–958. doi: 10.4319/lo.2007.52.3.0948

Karimi, E., Slaby, B. M., Soares, A. R., Blom, J., Hentschel, U., and Costa, R. (2018). Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS Microbiol. Ecol.* 94:fiy074. doi: 10.1093/femsec/fiy074

Kellner, S., Spang, A., Offre, P., Szöllősi, G. J., Petitjean, C., and Williams, T. A. (2018). Genome size evolution in the Archaea. *Emerg. Top. Life Sci.* 2, 595–605. doi: 10.1042/ETLS20180021

Kirchman, D. L., Elifantz, H., Dittel, A. I., Malmstrom, R. R., and Cottrell, M. T. (2007). Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol. Oceanogr.* 52, 495–507. doi: 10.4319/lo.2007.52.2.0495

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

Krom, M. D., Kress, N., Brenner, S., and Gordon, L. I. (1991). Phosphorus limitation of primary productivity in the eastern Mediterranean Sea. *Limnol. Oceanogr.* 36, 424–432. doi: 10.4319/lo.1991.36.3.0424

Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096

León-Zayas, R., Novotny, M., Podell, S., Shepard, C. M., Berkenpas, E., Nikolenko, S., et al. (2015). Single cells within the puerto rico trench suggest Hadal adaptation of microbial lineages. *Appl. Environ. Microbiol.* 81, 8265–8276. doi: 10.1128/AEM.0165915

Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Li, Z., Song, Q., Wang, Y., Xiao, X., and Xu, J. (2018). Identification of a functional toxin–antitoxin system located in the genomic island PYG1 of piezophilic hyperthermophilic archaeon *Pyrococcus yayanosii*. *Extremophiles* 22, 347–357. doi: 10.1007/s00792-018-1002-2

Liu, M., Fan, L., Zhong, L., Kjelleberg, S., and Thomas, T. (2012). Metaproteogenomic analysis of a community of sponge symbionts. *ISME J.* 6, 1515–1525. doi: 10.1038/ismej.2012.1

López-Legentil, S., Erwin, P. M., Pawlik, J. R., and Song, B. (2010). Effects of sponge bleaching on ammonia-oxidizing archaea: distribution and relative expression of ammonia monooxygenase genes associated with the barrel sponge *Xestospongia muta*. *Microb. Ecol.* 60, 561–571. doi: 10.1007/s00248-010-9662-1

Lurin, C., Andrés, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyère, C., et al. (2004). Genome-wide analysis of Arabidopsis pentatricopeptide repeat proteins

reveals their essential role in organelle biogenesis. *Plant Cell* 16, 2089–2103. doi: 10.1105/tpc.104.022236

Makarova, K. S., Wolf, Y. I., Snir, S., and Koonin, E. V. (2011). Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J. Bacteriol.* 193, 6039–6056. doi: 10.1128/JB.05535-11

Maldonado, M., and Riesgo, A. (2009). Gametogenesis, embryogenesis, and larval features of the oviparous sponge *Petrosia ficiformis* (Haplosclerida, Demospongiae). *Mar. Biol.* 156, 2181–2197. doi: 10.1007/s00227-009-1248-4

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2012). IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* 40, D115–D122. doi: 10.1093/nar/gkr1044

Miller, C. S., Baker, B. J., Thomas, B. C., Singer, S. W., and Banfield, J. F. (2011). EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. *Genome Biol.* 12:R44. doi: 10.1186/gb-2011-12-5-r44

Moeller, F. U., Webster, N. S., Herbold, C. W., Behnam, F., Domman, D., Albertsen, M., et al. (2019). Characterization of a thaumarchaeal symbiont that drives incomplete nitrification in the tropical sponge *Ianthella basta*. *Environ. Microbiol.* 21, 3831–3854. doi: 10.1111/1462-2920.14732

Mohamed, N. M., Colman, A. S., Tal, Y., and Hill, R. T. (2008). Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environ. Microbiol.* 10, 2910–2921. doi: 10.1111/j.1462-2920.2008.01704.x

Moitinho-Silva, L., Díez-Vives, C., Batani, G., Esteves, A. I., Jahn, M. T., and Thomas, T. (2017a). Integrated metabolism in sponge–microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J.* 11, 1651–1666. doi: 10.1038/ismej.2017.25

Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G. L., Cerrano, C., et al. (2017b). The sponge microbiome project. *Gigascience* 6:gix077. doi: 10.1093/gigascience/gix077

Mori, T., Cahn, J. K. B., Wilson, M. C., Meoded, R. A., Wiebach, V., Martinez, A. F. C., et al. (2018). Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 115, 1718–1723. doi: 10.1073/pnas.1715496115

Morley, S. A., Berman, J., Barnes, D. K. A., de Juan Carbonell, C., Downey, R. V., and Peck, L. S. (2016). Extreme phenotypic plasticity in metabolic physiology of Antarctic Demosponges. *Front. Ecol. Evol.* 3:157. doi: 10.3389/fevo.2015.00157

Morrow, C. C., Picton, B. E., Erpenbeck, D., Boury-Esnault, N., Maggs, C. A., and Allcock, A. L. (2012). Congruence between nuclear and mitochondrial genes in Demospongiae: a new hypothesis for relationships within the G4 clade (Porifera: Demospongiae). *Mol. Phylogenet. Evol.* 62, 174–190. doi: 10.1016/j.ympev.2011.09.016

Moya, A., Peretó, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote–animal symbioses. *Nat. Rev. Genet.* 9, 218–229. doi: 10.1038/nrg2319

Nguyen, M. T. H. D., Liu, M., and Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol. Ecol.* 23, 1635–1645. doi: 10.1111/mec.12384

Offre, P., Kerou, M., Spang, A., and Schleper, C. (2014). Variability of the transporter gene complement in ammonia-oxidizing archaea. *Trends Microbiol.* 22, 665–675. doi: 10.1016/j.tim.2014.07.007

Ouverney, C. C., and Fuhrman, J. A. (2000). Marine planktonic Archaea take up amino acids. *Appl. Environ. Microbiol.* 66, 4829–4833. doi: 10.1128/AEM.66.11.4829-4833.2000

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* 42, D206–D214. doi: 10.1093/nar/gkt1226

Park, S.-J., Ghai, R., Martín-Cuadrado, A.-B., Rodríguez-Valera, F., Chung, W.-H., Kwon, K., et al. (2014). Genomes of two new ammonia-oxidizing Archaea enriched from deep marine sediments. *PLoS One* 9:e96449. doi: 10.1371/journal.pone.0096449

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., Tyson, G. W., Hugenholtz, P., and Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123–3124. doi: 10.1093/bioinformatics/btu494

Peng, Y., Leung, H. C. M., Yiu, S. M., and Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174

Pester, M., Schleper, C., and Wagner, M. (2011). The *Thaumarchaeota*: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* 14, 300–306. doi: 10.1016/j.mib.2011.04.007

Pita, L., Rix, L., Slaby, B. M., Franke, A., and Hentschel, U. (2018). The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome* 6:46. doi: 10.1186/s40168-018-0428-1

Podell, S., Blanton, J. M., Neu, A., Agarwal, V., Biggs, J. S., Moore, B. S., et al. (2019). Pangenomic comparison of globally distributed Poribacteria associated with sponge hosts and marine particles. *ISME J.* 13, 468–481. doi: 10.1038/s41396-018-0292-9

Preston, C. M., Wu, K. Y., Molinski, T. F., and DeLong, E. F. (1996). A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. *Proc. Natl. Acad. Sci. U.S.A.* 93, 6241–6246. doi: 10.1073/pnas.93.13.6241

Pruesse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252

Puigbò, P., Wolf, Y. I., and Koonin, E. V. (2009). Search for a "Tree of Life" in the thicket of the phylogenetic forest. *J. Biol.* 8:59. doi: 10.1186/jbiol159

Radax, R., Hoffmann, F., Rapp, H. T., Leininger, S., and Schleper, C. (2012a). Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges: Archaeal nitrification in cold-water sponges. *Environ. Microbiol.* 14, 909–923. doi: 10.1111/j.1462-2920.2011.02661.x

Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H. T., Urich, T., et al. (2012b). Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environ. Microbiol.* 14, 1308–1324. doi: 10.1111/j.1462-2920.2012.02714.x

Reichenberger, E. R., Rosen, G., Hershberg, U., and Hershberg, R. (2015). Prokaryotic nucleotide composition is shaped by both phylogeny and the environment. *Genome Biol. Evol.* 7, 1380–1389. doi: 10.1093/gbe/evv063

Reynolds, D., and Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol. Ecol.* 25, 5242–5253. doi: 10.1111/mec.13812

Riesgo, A., Farrar, N., Windsor, P. J., Giribet, G., and Leys, S. P. (2014). The analysis of eight transcriptomes from all poriferan classes reveals surprising genetic complexity in sponges. *Mol. Biol. Evol.* 31, 1102–1120. doi: 10.1093/molbev/msu057

Rodrigues-Oliveira, T., Belmok, A., Vasconcellos, D., Schuster, B., and Kyaw, C. M. (2017). Archaeal S-layers: overview and current state of the art. *Front. Microbiol.* 8:2597. doi: 10.3389/fmicb.2017.02597

Rohrer, S., and Berger-Bächi, B. (2003). FemABX peptidyl transferases: a link between branched-chain cell wall peptide formation and β-Lactam resistance in gram-positive Cocci. *Antimicrob. Agents Chemother.* 47, 837–846. doi: 10.1128/AAC.47.3.837-846.2003

Rubin-Blum, M., Antony, C. P., Sayavedra, L., Martínez-Pérez, C., Birgel, D., Peckmann, J., et al. (2019). Fueled by methane: deep-sea sponges from asphalt seeps gain their nutrition from methane-oxidizing symbionts. *ISME J.* 13, 1209–1225. doi: 10.1038/s41396-019-0346-7

Sahling, H., Borowski, C., Escobar-Briones, E., Gaytán-Caballero, A., Hsu, C.-W., Loher, M., et al. (2016). Massive asphalt deposits, oil seepage, and gas venting support abundant chemosynthetic communities at the Campeche Knolls, southern Gulf of Mexico. *Biogeosciences* 13, 4491–4512. doi: 10.5194/bg-13-4491-2016

Schäffer, C., and Messner, P. (2004). Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* 14, 31R–42R. doi: 10.1093/glycob/cwh064

Seah, B. K. B., and Gruber-Vodicka, H. R. (2015). gbtools: interactive visualization of Metagenome Bins in R. *Front. Microbiol.* 6:1451. doi: 10.3389/fmicb.2015.01451

Segata, N., Börnigen, D., Morgan, X. C., and Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat. Commun.* 4:2304. doi: 10.1038/ncomms3304

Selvaraj, S. K., Periandythevar, P., and Prasadarao, N. V. (2007). Outer membrane protein A of *Escherichia coli* K1 selectively enhances the expression of intercellular adhesion molecule-1 in brain microvascular endothelial cells. *Microbes Infect.* 9, 547–557. doi: 10.1016/j.micinf.2007.01.020

Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., et al. (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* 5, 61–70. doi: 10.1038/ismej.2010.95

Simister, R. L., Deines, P., Botté, E. S., Webster, N. S., and Taylor, M. W. (2012). Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ. Microbiol.* 14, 517–524. doi: 10.1111/j.1462-2920.2011.02664.x

Sipkema, D., Caralt, S., Morillo, J. A., Al-Soud, W. A., Sorensen, S. J., Smidt, H., et al. (2015). Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission. *Environ. Microbiol.* 17, 3807–3821. doi: 10.1111/1462-2920.12827

Sizikov, S., Burgsdorf, I., Handley, K. M., Lahyani, M., Haber, M., and Steindler, L. (2020). Characterization of sponge-associated *Verrucomicrobia*: microcompartment-based sugar utilization and enhanced toxin–antitoxin modules as features of host-associated *Opitutales*. *Environ. Microbiol.* 22, 4669–4688. doi: 10.1111/1462-2920.15210

Slaby, B. M., Hackl, T., Horn, H., Bayer, K., and Hentschel, U. (2017). Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J.* 11, 2465–2478. doi: 10.1038/ismej.2017.101

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Staub, E., Pérez-Tur, J., Siebert, R., Nobile, C., Moschonas, N. K., Deloukas, P., et al. (2002). The novel EPTP repeat defines a superfamily of proteins implicated in epileptic disorders. *Trends Biochem. Sci.* 27, 441–444. doi: 10.1016/S0968-0004(02)02163-1

Steger, D., Ettinger-Epstein, P., Whalan, S., Hentschel, U., de Nys, R., Wagner, M., et al. (2008). Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environ. Microbiol.* 10, 1087–1094. doi: 10.1111/j.1462-2920.2007.01515.x

Steinert, G., Busch, K., Bayer, K., Kodami, S., Arbizu, P. M., Kelly, M., et al. (2020). Compositional and quantitative insights into bacterial and Archaeal communities of south pacific deep-sea sponges (Demospongiae and Hexactinellida). *Front. Microbiol.* 11:716. doi: 10.3389/fmicb.2020.00716

Suzek, B. E., Huang, H., McGarvey, P., Mazumder, R., and Wu, C. H. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics* 23, 1282–1288. doi: 10.1093/bioinformatics/btm098

Tatusov, R. L., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Kiryutin, B., Koonin, E. V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. doi: 10.1186/1471-2105-4-41

Taylor, M. W., Thacker, R. W., and Hentschel, U. (2007). Evolutionary insights from sponges. *Science* 316, 1854–1855. doi: 10.1126/science.1144387

Taylor, M. W., Tsai, P., Simister, R. L., Deines, P., Botte, E., Ericson, G., et al. (2013). 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME J.* 7, 438–443. doi: 10.1038/ismej.2012.111

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J. R., Easson, C., Astudillo-García, C., et al. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* 7:11870. doi: 10.1038/ncomms11870

Thomas, T., Rusch, D., DeMaere, M. Z., Yung, P. Y., Lewis, M., Halpern, A., et al. (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J.* 4, 1557–1567. doi: 10.1038/ismej.2010.74

Tian, R.-M., Sun, J., Cai, L., Zhang, W.-P., Zhou, G.-W., Qiu, J.-W., et al. (2016). The deep-sea glass sponge *Lophophysema eversa* harbours potential symbionts responsible for the nutrient conversions of carbon, nitrogen and sulfur. *Environ. Microbiol.* 18, 2481–2494. doi: 10.1111/1462-2920.13161

Tian, R.-M., Zhang, W., Cai, L., Wong, Y.-H., Ding, W., and Qian, P.-Y. (2017). Genome reduction and microbe-host interactions drive adaptation of a sulfur-oxidizing bacterium associated with a cold seep sponge. *mSystems* 2:e00184-16. doi: 10.1128/mSystems.00184-16

Uriz, M. J., Agell, G., Blanquer, A., Turon, X., and Casamayor, E. O. (2012). Endosymbiotic calcifying bacteria: A new cue to the origin of calcification in Metazoa? *Evolution* 66, 2993–2999. doi: 10.1111/j.1558-5646.2012.01676.x

Vacelet, J., and Donadey, C. (1977). Electron microscope study of the association between some sponges and bacteria. *J. Exp. Mar. Biol. Ecol.* 30, 301–314. doi: 10.1016/0022-0981(77)90038-7

Van Soest, R. W. M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., De Voogd, N. J., et al. (2012). Global diversity of sponges (Porifera). *PLoS One* 7:e0035105. doi: 10.1371/journal.pone.0035105

Veyron, S., Peyroche, G., and Cherfils, J. (2018). FIC proteins: from bacteria to humans and back again. *Pathog. Dis.* 76:fty012. doi: 10.1093/femspd/fty012

Wang, Y., Huang, J.-M., Cui, G.-J., Nunoura, T., Takaki, Y., Li, W.-L., et al. (2019). Genomics insights into ecotype formation of ammonia-oxidizing archaea in the deep ocean. *Environ. Microbiol.* 21, 716–729. doi: 10.1111/1462-2920.14518

Waters, A. L., Peraud, O., Kasanah, N., Sims, J. W., Kothalawala, N., Anderson, M. A., et al. (2014). An analysis of the sponge *Acanthostrongylophora igens*' microbiome yields an actinomycete that produces the natural product manzamine A. *Front. Mar. Sci.* 1:54. doi: 10.3389/fmars.2014.00054

Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12:444. doi: 10.1186/1471-2164-12-444

Xiong, L., Liu, S., Chen, S., Xiao, Y., Zhu, B., Gao, Y., et al. (2019). A new type of DNA phosphorothioation-based antiviral system in archaea. *Nat. Commun.* 10:1688. doi: 10.1038/s41467-019-09390-9

Yahel, G., Sharp, J. H., Marie, D., Häse, C., and Genin, A. (2003). In situ feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: bulk DOC is the major source for carbon. *Limnol. Oceanogr.* 48, 141–149. doi: 10.4319/lo.2003.48.1.0141

Zhang, F., Blasiak, L. C., Karolin, J. O., Powell, R. J., Geddes, C. D., and Hill, R. T. (2015). Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. *Proc. Natl. Acad. Sci. U.S.A.* 112, 4381–4386. doi: 10.1073/pnas.1423768112

Zhang, F., Pita, L., Erwin, P. M., Abaid, S., López-Legentil, S., and Hill, R. T. (2014). Symbiotic archaea in marine sponges show stability and host specificity in community structure and ammonia oxidation functionality. *FEMS Microbiol. Ecol.* 90, 699–707. doi: 10.1111/1574-6941.12427

Zhang, S., Song, W., Wemheuer, B., Reveillaud, J., Webster, N., and Thomas, T. (2019). Comparative genomics reveals ecological and evolutionary insights into sponge-associated *Thaumarchaeota*. *mSystems* 4:e00288-19. doi: 10.1128/mSystems.00288-19

Zhong, H., Lehtovirta-Morley, L., Liu, J., Zheng, Y., Lin, H., Song, D., et al. (2020). Novel insights into the *Thaumarchaeota* in the deepest oceans: their metabolism and potential adaptation mechanisms. *Microbiome* 8:78. doi: 10.1186/s40168-020-00849-2

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.

frontiers in Microbiology

ORIGINAL RESEARCH
published: 12 February 2021
doi: 10.3389/fmicb.2021.633649

Microbial Communities Under Distinct Thermal and Geochemical Regimes in Axial and Off-Axis Sediments of Guaymas Basin

Andreas Teske[1], Gunter Wegener[2,3], Jeffrey P. Chanton[4], Dylan White[1], Barbara MacGregor[1,5], Daniel Hoer[6,7], Dirk de Beer[2], Guangchao Zhuang[8,9,10], Matthew A. Saxton[10,11], Samantha B. Joye[10], Daniel Lizarralde[12], S. Adam Soule[12] and S. Emil Ruff[13,14]*

[1]Department of Marine Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, [2]Max-Planck-Institute for Marine Microbiology, Bremen, Germany, [3]MARUM, Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany, [4]Department of Earth, Ocean and Atmospheric Sciences, Florida State University, Tallahassee, FL, United States, [5]Department of Earth and Environmental Sciences, University of Minnesota, St. Paul, MI, United States, [6]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, United States, [7]United States Environmental Protection Agency, Research Triangle Park, NC, United States, [8]Frontiers Science Centre for Deep Ocean Multispheres and Earth System (FDOMES)/Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education, Ocean University of China, Qingdao, China, [9]Laboratory for Marine Ecology and Environmental Science, Qingdao National Laboratory for Marine Science and Technology, Ocean University of China, Qingdao, China, [10]Department of Marine Sciences, University of Georgia, Athens, GA, United States, [11]Department of Biological Sciences, Miami University, Oxford, OH, United States, [12]Geology & Geophysics Department, Woods Hole Oceanographic Institution, Woods Hole, MA, United States, [13]Marine Biological Laboratory, The Ecosystems Center, Woods Hole, MA, United States, [14]Marine Biological Laboratory, The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Woods Hole, MA, United States

OPEN ACCESS

Edited by:
Axel Schippers, Federal Institute for Geosciences and Natural Resources, Germany

Reviewed by:
Takuro Nunoura, Japan Agency for Marine-Earth Science and Technology (JAMSTEC), Japan
Gordon Webster, Cardiff University, United Kingdom

***Correspondence:**
Andreas Teske
teske@email.unc.edu

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

Received: *25 November 2020*
Accepted: *12 January 2021*
Published: *12 February 2021*

Citation:
Teske A, Wegener G, Chanton JP, White D, MacGregor B, Hoer D, de Beer D, Zhuang G, Saxton MA, Joye SB, Lizarralde D, Soule SA and Ruff SE (2021) Microbial Communities Under Distinct Thermal and Geochemical Regimes in Axial and Off-Axis Sediments of Guaymas Basin. Front. Microbiol. 12:633649. doi: 10.3389/fmicb.2021.633649

Cold seeps and hydrothermal vents are seafloor habitats fueled by subsurface energy sources. Both habitat types coexist in Guaymas Basin in the Gulf of California, providing an opportunity to compare microbial communities with distinct physiologies adapted to different thermal regimes. Hydrothermally active sites in the southern Guaymas Basin axial valley, and cold seep sites at Octopus Mound, a carbonate mound with abundant methanotrophic cold seep fauna at the Central Seep location on the northern off-axis flanking regions, show consistent geochemical and microbial differences between hot, temperate, cold seep, and background sites. The changing microbial actors include autotrophic and heterotrophic bacterial and archaeal lineages that catalyze sulfur, nitrogen, and methane cycling, organic matter degradation, and hydrocarbon oxidation. Thermal, biogeochemical, and microbiological characteristics of the sampling locations indicate that sediment thermal regime and seep-derived or hydrothermal energy sources structure the microbial communities at the sediment surface.

Keywords: cold seep, hydrothermal sediment, porewater profiles, bacteria, archaea, Guaymas Basin

INTRODUCTION

Globally, over 700 marine hydrothermal vent sites are currently known (Beaulieu and Szafranski, 2020), including sedimented hydrothermal systems at coastal and continental margin locations (Price and Giovanelli, 2017). One of the best-studied sedimented hydrothermal systems is Guaymas Basin in the Gulf of California, a young marginal rift basin characterized by active seafloor spreading and rapid deposition of organic-rich sediments from highly productive overlying waters (Lonsdale and Becker, 1985). Buried organic matter is hydrothermally transformed to methane, aliphatic and aromatic hydrocarbons, dissolved inorganic carbon, and ammonia, resulting in well-buffered and nutrient-rich vent fluids (Von Damm et al., 1985) that sustain ample microbial communities. The microbiology and biogeochemistry of hydrothermal sediments in Guaymas Basin have been studied extensively. Sulfur-oxidizing microbial mats occur in visually conspicuous hot spots where sulfide, methane-, and hydrocarbon-rich hydrothermal fluids rise to the sediment surface (McKay et al., 2012; MacGregor et al., 2013). Surficial sediments harbor complex anaerobic microbial communities adapted to these conditions, including thermophilic methane- and alkane-oxidizing archaea (Teske et al., 2002; Biddle et al., 2012; Wegener et al., 2015; Dowell et al., 2016; Laso-Pérez et al., 2016; Wang et al., 2019) and hydrocarbon-oxidizing, free-living or syntrophic sulfate-reducing bacteria (reviewed by Teske, 2019). The diversity of hydrothermal regimes in Guaymas Basin selects for microbial communities with adaptations to different thermal and geochemical niches (McKay et al., 2016; Teske et al., 2016). Here we compare thermal gradients, porewater geochemistry, and microbial community composition in the surface layer of different hydrothermal, cold seep, and background sediments (**Table 1**) to determine whether sediments with distinct porewater geochemistry and thermal regimes harbor particular bacterial and archaeal populations.

In contrast to previous small-scale surveys that focused on individual hydrothermal mounds or microbial mats (McKay et al., 2012, 2016; Dowell et al., 2016), the sampling sites are separated by distances of hundreds of meters or several miles (**Figure 1**). Within a few miles of each other, the greater Guaymas Basin geo-ecosystem includes hydrothermal areas at the southern Guaymas Basin spreading center (Teske et al., 2016), off-axis hydrothermal sites (Teske et al., 2019; Ramírez et al., 2020), hydrothermal mounds just off the northern spreading center (Berndt et al., 2016), and cold seeps on the northern flanking regions and the adjacent Sonora Margin (Geilert et al., 2018). Methane-rich off-axis cold seeps have been documented by deep-tow photography of faunal communities and carbonates, and by *in situ* measurements of methane anomalies in the deep water column (Lizarralde et al., 2011), and some cold seep sites have been sampled further by multicoring and gravity coring of sediments and dredging of seafloor minerals (Geilert et al., 2018; Núñez-Useche et al., 2018). In contrast to compression-induced seepage on massively sedimented continental margins and plate boundaries (Suess, 2010), seismic surveys have linked the Guaymas Basin cold seeps to deeply buried volcanic sills that extend from the spreading centers (Einsele et al., 1980) up to 50 km across the sedimented Guaymas flanking regions (Lizarralde et al., 2011). At specific off-axis locations, shallow hot sills drive hydrothermal fluid and gas circulation (Teske et al., 2019). Yet in most cases, off-axis sills and their seep fluids have cooled off over time, and methane-rich fluids have temperatures near ambient bottom water when they are released at the sediment surface (Lizarralde et al., 2011; Geilert et al., 2018). Guaymas Basin cold seep sites remain to be explored and sampled up close by ROV and submersible. This study concludes by surveying a cold seep site ("Central Seep") approximately equidistant from Sonora and Baja California on the northern flanking regions (Geilert et al., 2018; Núñez-Useche et al., 2018).

MATERIALS AND METHODS

Field Survey and Sampling

Guaymas Basin sites were visited and sampled with R/V *Atlantis*, HOV *Alvin*, and AUV *Sentry* during cruise AT37-06 (December 6–29, 2016). *Alvin* dives targeted previously explored sampling areas (Teske et al., 2016), or newly identified sites found by AUV *Sentry*. After *Sentry* returned from pre-programmed night dives at ca. 6 AM, dive data and bathymetries were downloaded and made available in time for the following *Alvin* dive starting at 8 AM. When *Sentry* performed seafloor photomosaic surveys running ca. 6 m above bottom, the resulting images were inspected for microbial mats and potential dive targets, for example, in the "Northern Towers" area of the southern axial trough of Guaymas Basin. Photo coverage of *Alvin* dives is available at the *Alvin* frame-grabber site.[1] The bathymetry of the hydrothermally active graben segment in southern Guaymas Basin was mapped by AUV *Sentry* during dives 407-409, 413-417; the bathymetry of Octopus Mound at the Central Seep site was mapped during *Sentry* dive 412 (**Figure 1**). Survey height was 65–70 m above the bottom. *Alvin* sampling sites were largely based on *Sentry* surveys (**Supplementary Figure S1**).

Thermal Profiles

Thermal profiles were measured in surficial sediments using *Alvin's* 50 cm or 1-m heat flow probes.[2] The 50 cm probe was used for hydrothermal sites in the southern axial valley and contains thermal sensors every 10 cm, starting 5 cm under the attached plastic disk (the "puck") that limits probe penetration and rests on the seafloor once the probe was inserted. Cold seep sediments of Octopus Mound were profiled using the 1-m probe with thermal sensors every 20 cm. After 5–10 min of temperature reading stabilization, temperature readings were recorded. Thermal profiles adjacent to sediment cores that were analyzed in this study are shown in **Supplementary Figure S3**.

[1] http://4dgeo.whoi.edu/alvin
[2] https://ndsf.whoi.edu/alvin/using-alvin/sampling-equipment/

TABLE 1 | Sampling sites and *Alvin* core numbers for 2016 sediment samples collected during Atlantis cruise AT37-06, used in parallel molecular and biogeochemical analyses.

Guaymas Basin area	Sampling location	Date	Latitude/longitude	Water depth (m)	*Alvin* core number for Sequence data	*Alvin* core number for Porewater data	Sample type	Microbial mat and/or invertebrates	T max.
Southern axial valley	Mat Mound Massif	December 12, 2016	27°00.44'/111°24.52'	2,002	4861-26	no data	Hot hydrothermal vent	White sulfur mat, snails	53°C
Southern axial valley	Mat Mound Massif	December 13, 2016	27°00.43'/111°24.56'	2,000	4862-08	4862-08	Hot hydrothermal vent	Orange *Beggiatoaceae* mat	146°C
Southern axial valley	Mat Mound Massif	December 13, 2016	27°00.43'/111°24.56'	2,000	4862-33	4862-33	Background sediment	Cold bare sediment	3.4°C
Central seep	Octopus Mound Periphery	December 18, 2016	27°28.17'/111°28.39'	1,832	4867-08	4867-32	Background sediment	Cold bare sediment	3°C
Central seep	Octopus Mound Active site	December 18, 2016	27°28.17'/111°28.39'	1,832	4867-14	4867-30	Cold seep	Tubeworms, clams, ampharetid worms	3°C
Southern axial valley	Mat Mound Massif	December 21, 2016	27°00.45'/111°24.54'	2,001	4869-25	4869-26	Hot hydrothermal vent	Orange *Beggiatoaceae* mat	85°C
Southern axial valley	Cathedral Hill	December 22, 2016	27°00.71'/111°24.22'	2,009	4870-16	4870-29	Hot hydrothermal vent	Oily sediment and fluffy white mat	80°C
Southern axial valley	Aceto Balsamico	December 22, 2016	27°00.47'/111°24.43'	2,007	4870-02	4870-32	Temperate hydrothermal vent	Yellow sulfur mat	29°C
Southern axial valley	Northern Towers	December 23, 2016	27°02.77'/111°23.09'	1,994	4871-20	4871-20	Temperate hydrothermal vent	White sulfur mat, worms	13°C
Southern axial valley	Northern Towers	December 23, 2016	27°02.75'/111°23.05'	1,990	4871-26	4871-23	Hot hydrothermal vent	White sulfur mat, no worms	92°C

Thermal gradients are tabulated in **Supplementary Tables S2** *and* **S3**, *and plotted in* **Supplementary Figure S3**. *The samples are sorted by Alvin dive number.*

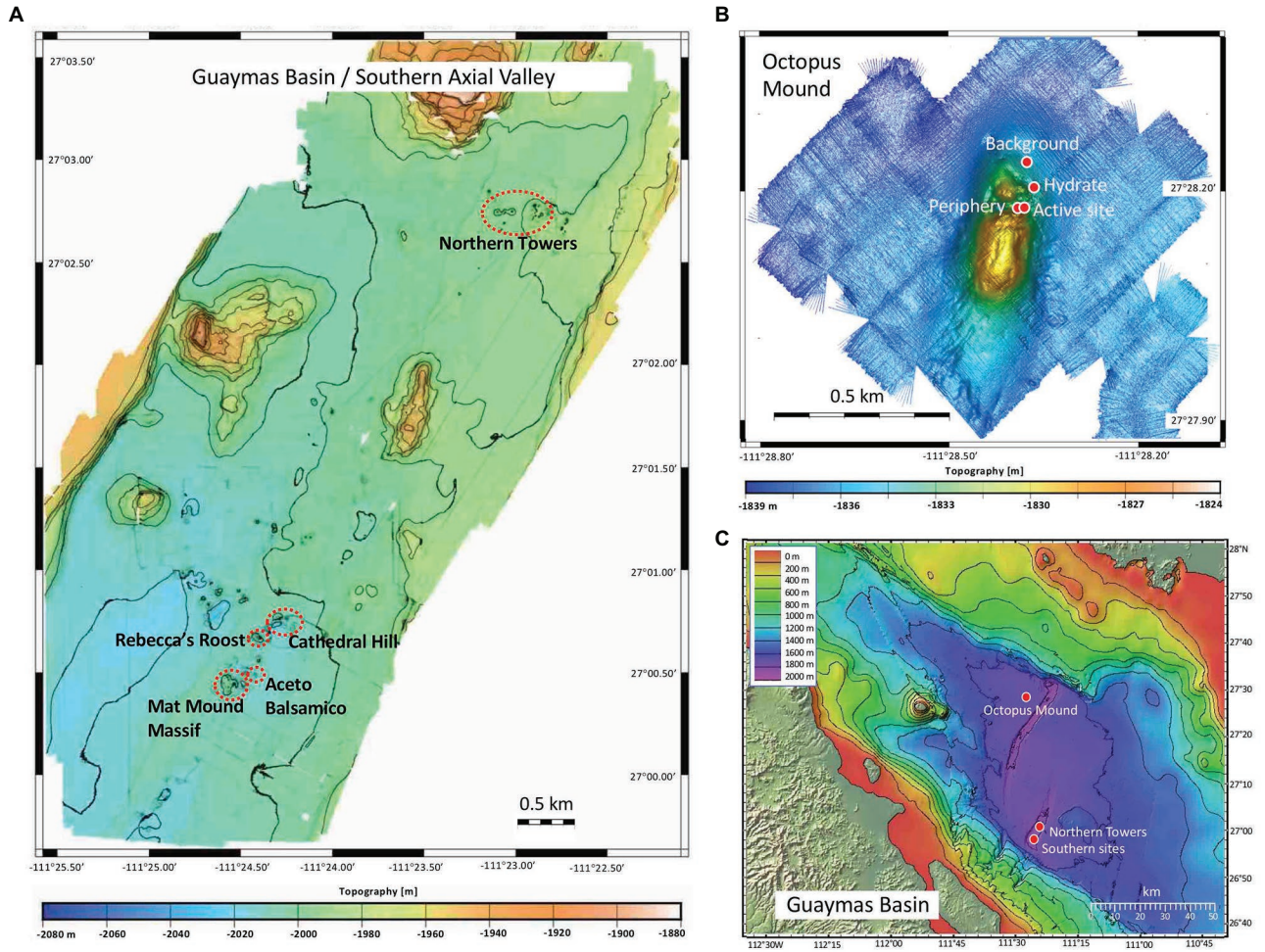

FIGURE 1 | Bathymetric maps of Guaymas Basin and Expedition AT37-06 sampling locations. **(A)** Guaymas Basin southern axial valley, mapped during *Sentry* dives 407-409 and 413-417. **(B)** Central Seep area with Octopus Mound, mapped during *Sentry* dive 412. **(C)** Guaymas Basin overview annotated with sampling sites, based on a template courtesy of C. Mortera, UNAM.

DNA Extraction, Library Preparation, and High-Throughput 16S rRNA Gene Sequencing

DNA was collected from the top 0–1 cm of sediment cores after removal of any overlying microbial mat, except for one mat-covered core (4862-08) where draining fluids had sucked the mat into the sediment before slicing. Genomic DNA was extracted from 0.5 g of sediment from each sample using the DNeasy PowerLyzer PowerSoil Kit (Cat. No. 12855-100, QIAGEN) and bead-beating at 6 m s^{-1} for 45 s using a Bead Ruptor 24 (OMNI International, Kennesaw, GA, United States). Extraction blanks were performed alongside the samples to assess laboratory contamination during the extraction process. DNA concentrations were assessed fluorometrically using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada). Bacterial 16S rRNA gene variable region v3-v4 was amplified using the "universal" primer pair 341F (5'-CCTACGGGAGGCAGCAG-3'; Klindworth et al., 2013) and Pro805R (5'-GACTACNVGGGTATCTAATCC-3'; Takahashi et al., 2014). The archaeal 16S rRNA gene v4-v5 variable region was amplified using the combined forward primers 517F (5'-GCCTAAAGCATCCGTAGC-3', 5'-GCCTAAARCGTYCGTAGC-3', 5'-GTCTAAAGGGTCYGTAGC-3', 5'-GCTTAAAGNGTYCGTAGC-3', 5'-GTCTAAARCGYYCGTAGC-3') and reverse primer 958R (5'-CCGGCGTTGANTCCAATT-3'; Topçuoglu et al., 2016). The primers were modified with Illumina MiSeq overhang adapters. Each PCR reaction consisted of 1–5 µl (~10 ng) genomic DNA template, 2.5 µl of each of the primers (final concentration 1 µM), 12.5 µl 2X Kapa HiFi HotStart ReadyMix (Kapa Biosystems, Wilmington, MA, United States) and sterile nuclease-free water to make a final volume of 25 µl gene amplification was carried out using an initial degradation at 95°C for 3 min following 25 cycles of denaturation at 95°C for 30 s, annealing at 56°C for 45 s, and extension at 72°C for 1 min, and concluded with a final extension at 72°C for 5 min. All PCR reactions were performed in triplicate, pooled, and purified using the NucleoMag NGS Clean-up and Size Select kit (Macherey-Nagel Inc., Bethlehem,

PA, United States). The purified PCR products were indexed following the instructions on Illumina's 16S amplicon library preparation guide. The concentration of dsDNA and the size of the indexed amplicons were verified using the Qubit dsDNA High Sensitivity assay kit on a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Canada) and the Agilent 2100 Bioanalyzer system (Agilent Technologies, Mississauga, ON, Canada), respectively. Indexed amplicons were pooled in equimolar amounts and sequenced using Illumina's v3 600-cycle (paired-end) reagent kit on a MiSeq benchtop sequencer (Illumina Inc., San Diego, CA, United States).

Microbial Community Analyses

Raw bacterial 16S rRNA gene amplicon sequence data were analyzed using MetaAmp (Dong et al., 2017), and archaeal raw sequence data were analyzed using DADA2 (Callaghan et al., 2016). Taxonomy of operational taxonomic units (bacterial OTUs, defined at 98% sequence identity) or amplicon sequence variants (archaeal ASVs) was assigned using the SILVA reference database v132 (Quast et al., 2013). Community analyses were performed using VisuaR[3], an R-based workflow using the packages vegan, labdsv, tidyverse (stringr, dplyr, ggplot2), UpSetR, and custom scripts (Wickham, 2016, 2018; Oksanen et al., 2012; Roberts, 2012; Conway et al., 2017; Wickham et al., 2018). The original OTU or ASV abundance tables were used to calculate richness and diversity indices, i.e., Inverse Simpson diversity (Hill et al., 2003), Shannon entropy, and Chao1 estimated richness (Chao, 1984) with a subsampling approach to ensure comparability of indices. Dissimilarities between samples were calculated using the Bray-Curtis dissimilarity coefficient (i.e., relative sequence abundance; Bray and Curtis, 1957). The resulting beta-diversity matrices were used for 2D non-metric multidimensional scaling (NMDS) ordinations with 20 random starts (Kruskal, 1964). Stress values below 0.2 indicate that the multidimensional dataset is well represented by the 2D ordination.

16S rRNA Gene Phylogeny

Phylogenetic placement of amplicon sequences was conducted using the SILVA database SSU Ref NR 132 (Quast et al., 2013) and the software ARB (Ludwig et al., 2004). ASVs were first added to the SILVA tree using the ARB "quick add" tool, neighboring near full-length sequences (>1,300 nucleotides) were selected and aligned using SINA (Prüsse et al., 2012). The alignment was manually curated based on ribosomal secondary structure and was subsequently used to calculate 100 maximum likelihood phylogenetic trees with the phyML algorithm, of which the most likely tree was automatically selected. We used a positional variability filter including only conserved positions in the alignment with a mutation rate of <3.1%. Finally, amplicon sequences were added to the consensus tree using the same positional variability filter without changing the overall tree topology.

DNA Isolation and *mcrA* Gene Sequencing

A second set of DNA extractions from sediment was performed with the MO BIO PowerSoil DNA Isolation Kit (QIAGEN, Carlsbad, CA, United States) for the purpose of *mcrA* gene amplification and sequencing. Amplification was performed with general *mcrA* primer combination mcrIRD-F (5'-GACCAGTTGTGGTTCGGAAC-3') and mcrIRD-R (5'-ATCTCGAATGGCATTCCCTC-3'; Lever and Teske, 2015). The PCR protocol contained 30 cycles of initial denaturation at 95°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min, and concluded with a final extension at 72°C for 5 min. PCR products were purified using the Wizard SV Gel and PCR Cleanup System (Promega Corporation, Madison, WI, United States) and cloned into plasmid vectors using the TOPO TA Cloning Kit (Life Technologies, Carlsbad, CA, United States). These were used to transform One Shot TOP10 *Escherichia coli* cells (Life Technologies, Carlsbad, CA, United States) which were plated on selective media. Approximately 25 colonies per sample were picked and incubated overnight in SOC medium, then plasmids were extracted using the GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific, Waltham, MA, United States). Plasmids were then sent to GeneWiz (South Plainfield, NJ, United States) for sequencing. The alignment of representative *mcrA* gene sequences was created in the MEGA software using the MUSCLE algorithm (Edgar, 2004), and the tree was created from the gene sequence alignment using the neighbor-joining method in MEGA4 (Tamura et al., 2007). The topology of the *mcrA* phylogeny was tested by 500 bootstrap runs.

Porewater Geochemistry

For porewater analysis, intact sediment cores were sampled using the Rhizons (Rhizosphere Research Products, Wageningen, NL) as described previously (Seeberg-Elverfeldt et al., 2005). The overlying water was removed from the cores. Holes were drilled into the core at designated sediment sampling depths and pretreated Rhizons (washed twice with HCl and MilliQ water) were injected and suction was applied with syringes for ~30 min. For sulfide analysis, 1 ml porewater subsamples were fixed with 0.1 ml of 0.1 M zinc acetate solution to preserve sulfide as zinc sulfide until analysis using the methylene blue method (Cline, 1969). The same fixed porewater sample was used for measuring sulfate concentrations using ion chromatography (Metrohm 930 Compact IC flex oven, Metrosep A PCC HC/4.0 preconcentration column, and Metrosep A Supp 5 Guard/4.0 chromatography column). The concentrations of ammonium, phosphate, and silicate were determined from the same porewater samples using a continuous flow nutrient analyzer (QuAAtro39; Seal Analytical) as published previously (Grasshoff et al., 2009). For combined concentration and $\delta^{13}C$ analysis of methane, 2 ml sediment subsamples were added to 30 ml serum vials containing 2 ml of 1 M sodium hydroxide solution, sealed with thick butyl rubber stoppers, crimped with aluminum seals, and stored at 4°C. Shipping problems and a resulting shortage of serum vials limited sampling capabilities and only selected sediment cores were sampled for methane. Since cores were retrieved unpressurized, outgassing may have impacted in particular the measurements of methane concentrations near and above saturation. After the cruise, the methane samples were analyzed by headspace gas chromatography-flame ionization detection (GC-FID) at Florida State University (Magen et al., 2014). Additionally, the gas samples

[3]https://github.com/EmilRuff/VisuaR

were analyzed for $\delta^{13}CH_4$ by injecting 0.1–0.5 ml of sample into a gas chromatograph interfaced to a Finnigan MAT Delta S isotope ratio Mass Spectrometer inlet system as previously described (Chanton and Liptay, 2000). Values are reported in the per mil (‰) notation relative to Vienna Pee Dee Belemnite (VPDB).

RESULTS

Guaymas Basin Survey Sites

During expedition AT37-06 to Guaymas Basin, AUV *Sentry* and submersible *Alvin* mapped and sampled hydrothermally active sediments within the southern Guaymas Basin spreading center, at an off-axis sill-driven vent site (Ringvent) with hybrid seep/vent biota on the northwestern flanking regions near Isla Tortuga (Teske et al., 2019), and at the off-axis Central Seep site on the northwestern Guaymas flanks (Geilert et al., 2018) that is approximately equidistant from the Sonora and Baja California coasts (**Figure 1**).

Different types of mat-covered sediments and thermal regimes were sampled at these locations (**Figure 2**). The Mat Mound Massif, a cluster of hydrothermal mounds and edifices, was probed from different angles on previous cruises (e.g., Dowell et al., 2016) but its full extent was only recognized during this expedition. The Aceto Balsamico area is located ca. 150 m west of Mat Mound Massif and contains moderately warm sediments covered with lime-yellow mats of sulfur-oxidizing bacteria (Teske et al., 2016). At Cathedral Hill, ca. 200 m north of Aceto Balsamico, gradually sloping sediment-covered mounds with extensive microbial mats are topped with hydrothermal edifices (Teske et al., 2016). The Northern Towers area is located ca. 5 km northeast from the other, tightly clustered sites, and was not sampled during previous expeditions in 2008/2009 (Teske et al., 2016). It has relatively few hydrothermal sediments and mats but is dominated by massive, steep hydrothermal edifices and chimneys. Here, microbial mats were located by AUV *Sentry* in photomosaic survey mode, and their coordinates were targeted during subsequent *Alvin* dives. *Alvin* push cores from all locations were collected during Expedition AT37-06 on *Alvin* dives 4861, 4862, 4867, 4869, 4870, and 4871 (**Table 1**).

Bathymetric mapping with AUV *Sentry* and reconnaissance with submersible *Alvin* at the Central Seep location revealed a mound of 200 m north-to-south and 100 m east-to-west extension that rises ca. 20 m above the seafloor. The base of the mound at its northern end, and nearby sediments harbored extensive cold seep communities and surface-breaching gas hydrates (**Figure 3**). This mound was informally named "Octopus Mound" for its abundance of cephalopods (**Supplementary Figure S1**) after it was explored and sampled during *Alvin* dives 4866 and 4867 (December 17–18, 2016). Surface sediment samples were collected by HOV *Alvin* within ~100 m of each other at the northern tip of Octopus Mound in an area with small-scale topographical diversity, diverse cold seep fauna, seafloor mineral formations, microbial mats, and hydrates (**Figure 3**). Sampling locations at Octopus Mound (**Supplementary Table S1**) were chosen based on the presence or absence of seep fauna assemblages (**Supplementary Figure S2**).

Porewater and Thermal Profiles

The Guaymas Basin cores from these diverse sampling locations can be grouped into three categories based on thermal profiles and microbial mat cover: hot hydrothermal sediments with conspicuous microbial mats and temperatures reaching >50°C within 50 cm depth, temperate hydrothermal sediments with microbial mats and a temperature range between 5 and 50°C, and background sediments without visible mats, and a temperature range of 3–5°C (**Supplementary Tables S2 and S3**, **Supplementary Figures S2, S3**). These categories were also reflected in the porewater profiles (**Figures 4, 5**).

Hydrothermal Sediments

The hot hydrothermal sediments (cores 4862-08, 4869-26, 4870-29, 4871-23) are characterized by steeply increasing temperatures into the range of >80°C by 50 cm depth (**Supplementary Table S1**, **Supplementary Figure S3**) and thick microbial mats covering the seafloor (**Supplementary Figure S2**). The extracted porewater contained high sulfide (1–10 mM) and ammonium (up to 7 mM) concentrations (**Figure 4**). Elevated concentrations of dissolved silicate (>0.2 mM) suggest hydrothermal dissolution and mobilization of solid silicate phases, and irregular sulfate profiles most likely represent seawater inmixing (**Figure 4**). Cores in the "temperate" sediment category (cores 4870-32 and 4871-20) had moderate *in situ* temperatures of 10–30°C. The ammonium concentration of core 4870-32 (5–10 mM; no ammonium data for core 4871-20) was even higher than observed at the hot hydrothermal sites. Sulfate is rapidly depleted below the sediment surface, indicating the absence of seawater sulfate inmixing; sulfide appears to be limited to surficial sediments (**Figure 5**). In the "background" category, *Alvin* cores 4862-33 and 4867-32 have temperatures near bottom seawater (3–4°C). These cores have significantly lower ammonium and silicate concentrations, in combination with near-seawater sulfate concentrations; sulfide was only found in micromolar traces (**Figure 4B**). Sediment core 4867-30 from the off-axis Octopus Mound sampling site does not fit easily into these categories; it is uniformly at bottom water temperature (2.9°C), lacks hydrothermal signatures such as high ammonium or silicate concentrations, and shows only moderate sulfate depletion; yet the core was strongly sulfidic (**Figure 5**).

Cold Seep Sites

Alvin dives 4866 and 4867 provided an opportunity to inspect different faunal assemblages and geochemical settings, and to collect sediment push cores from closely spaced targets at the edge of Octopus Mound. As determined with the *Alvin* heat flow probe, all sampling sites at Octopus Mound had *in situ* temperatures between 2.9°C at the seawater interface, increasing towards local maxima of 2.95–3.03°C at different shallow sediment depths as determined with the *Alvin* heat flow probe (**Supplementary Table S3**). The "active site," named for its conspicuous benthic invertebrate community, harbors a cold seep assemblage of tube worms, seep clams, and dense populations of ampharetid worms covering the sediment surface (**Figures 3A,B,C**). Sampling holes after core removal were deep black, indicative of

sulfidic, strongly reducing conditions. The geochemically analyzed core 4867-30 was collected at this location. The "periphery" site was selected to sample the edge of this conspicuous seep community, at a distance of ~2 m. The "background" site lacked visible seep fauna. The "gas hydrate site" (**Figures 3D,E,F**) was initially sampled for its conspicuous circular microbial mat consisting of large, filamentous sulfur-oxidizing bacteria of the family *Beggiatoaceae* (Teske and Salman, 2014), with ca. 100 µm filament diameter (**Figure 3F**). On closer inspection, the sulfur-oxidizing *Beggiatoaceae* mat was growing on top of an ampharetid worm assemblage

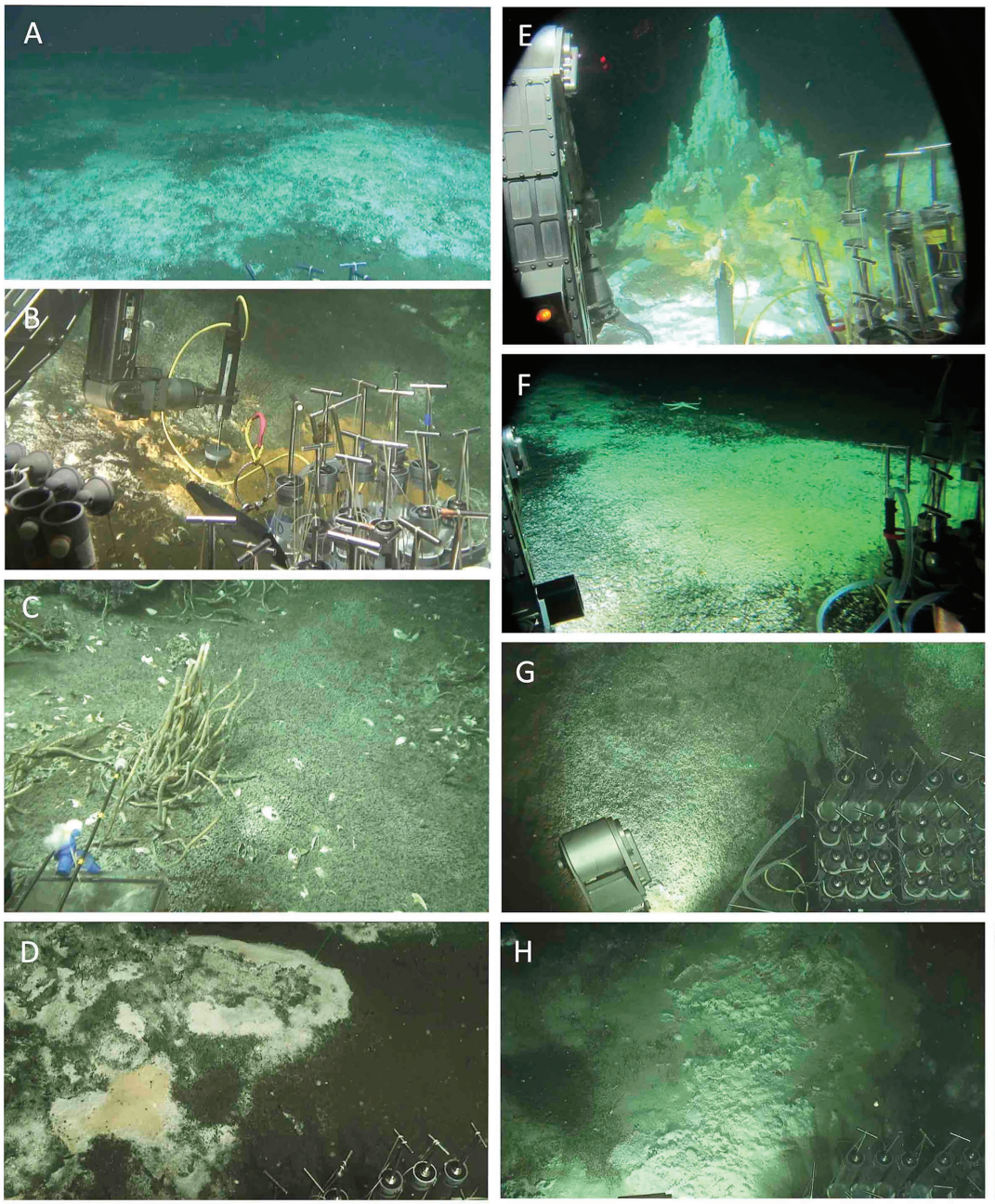

FIGURE 2 | Sampling sites for microbial and/or biogeochemical analyses. **(A)** Extensive white mat at Mat Mound Massif. Dive 4861. **(B)** Hydrothermal sediment with orange *Beggiatoaceae* mat, Mat Mound Massif. Dive 4862. **(C)** Cold seep site "active site" at Octopus Mound. Dive 4867. **(D)** Hydrothermal sediment with orange *Beggiatoaceae* mat at Mat Mound Massif. Dive 4869. **(E)** Summit and slopes at Cathedral Hill. The cores are from the white, fluffy mat area at the bottom of the photo. Dive 4870. **(F)** Temperate Aceto Balsamico mat with lime-yellow sulfur precipitates. Dive 4870. **(G)** Temperate "site 2" mat-covered sediment at Northern Towers. Dive 4871. **(H)** Hot "site 3" mat-covered hydrothermal sediment at Northern Towers. Dive 4871. The sites were photographed from inside *Alvin* (dives 4862, 4867, 4870) or documented as *Alvin* screen grabber images (dives 4861, 4869, 4871). Images courtesy of *Alvin* group, WHOI.

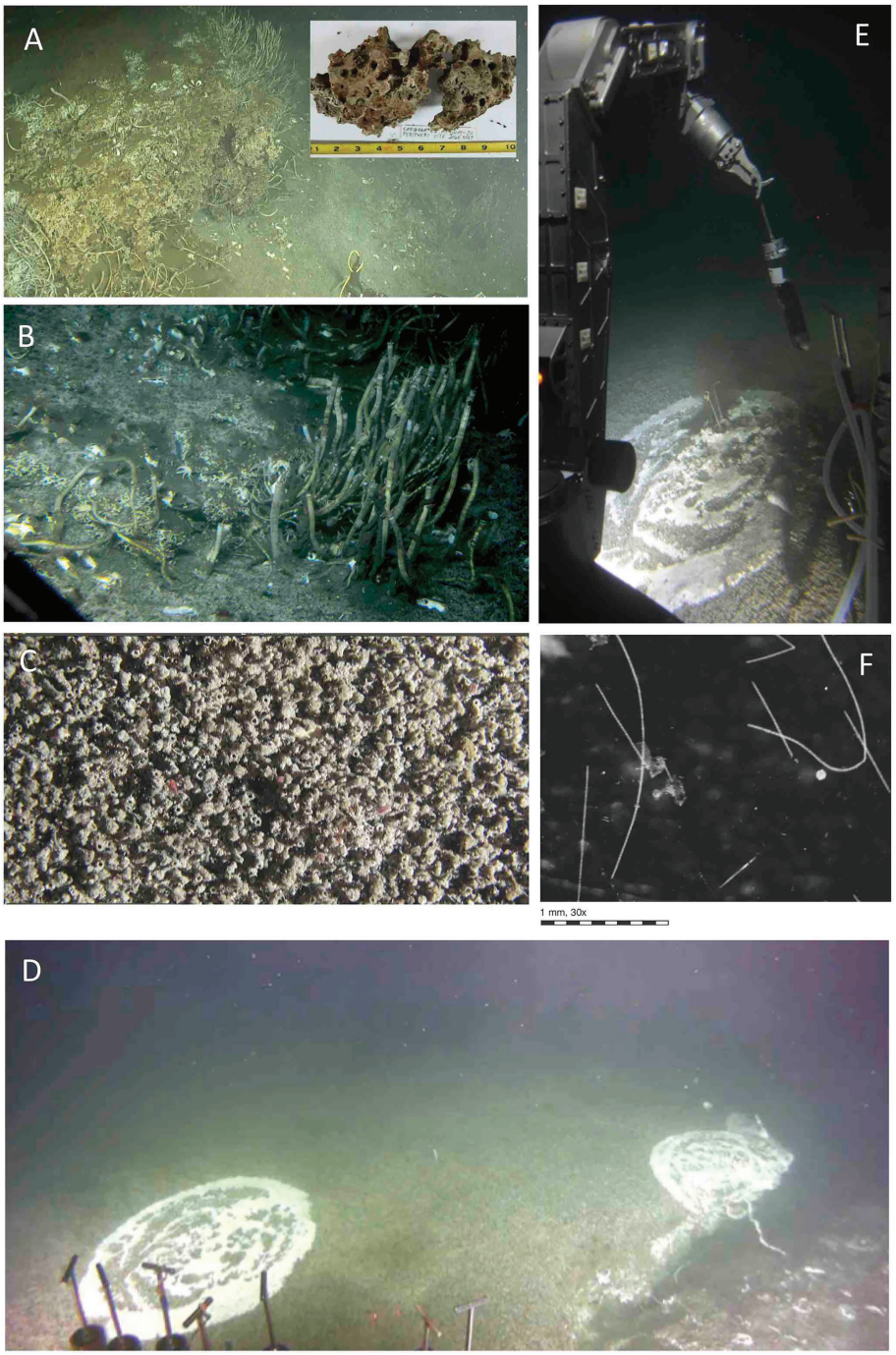

FIGURE 3 | Benthic fauna at Octopus Mound sampling sites. **(A)** Seafloor carbonate concretions with tube worms at the "active site" near the base of Octopus Mound. Insert, carbonate sample collected at this location during *Alvin* dive 4867. **(B)** Seep community with different types of tubeworms resembling *Lamellibrachia*, galatheid crabs, and clam shells. The original image was underexposed and had to be digitally manipulated, resulting in over-emphasized steel-blue hues instead of olive-green and brown tones. **(C)** Close-up view of ampharetid worm carpet, with pink worms protruding from several worm tubes. Overall view ~10 × 20 cm; video still from *Alvin's* bottom camera. **(D)** Sedimented Hydrate mound, overgrown with an extensive mat of ampharetid worms and circular spots of *Beggiatoaceae* mats. At the massive fracture to the right, the mat-covered sediment is ca. 0.5 m elevated, suggesting rising hydrate underneath. The *Beggiatoaceae* mat to the left was sampled with *Alvin* push cores. **(E)** Coring the *Beggiatoaceae* mat on top of the hydrate mound during dive 4867, the "hydrate site." The bottom of the freshly collected core contains white gas hydrate (presumably methane hydrate) that dissipates during transport to the surface. **(F)** Individual *Beggiatoaceae* filaments recovered from the hydrate mat are viewed through a dissection binocular. Filament diameters are in the range of 100–160 μm, consistent with large, colorless *Beggiatoaceae* observed previously in Guaymas Basin (McKay et al., 2012; Teske and Salman, 2014). Images A–E courtesy of the *Alvin* group, WHOI; dissection scope image of filaments by Barbara MacGregor.

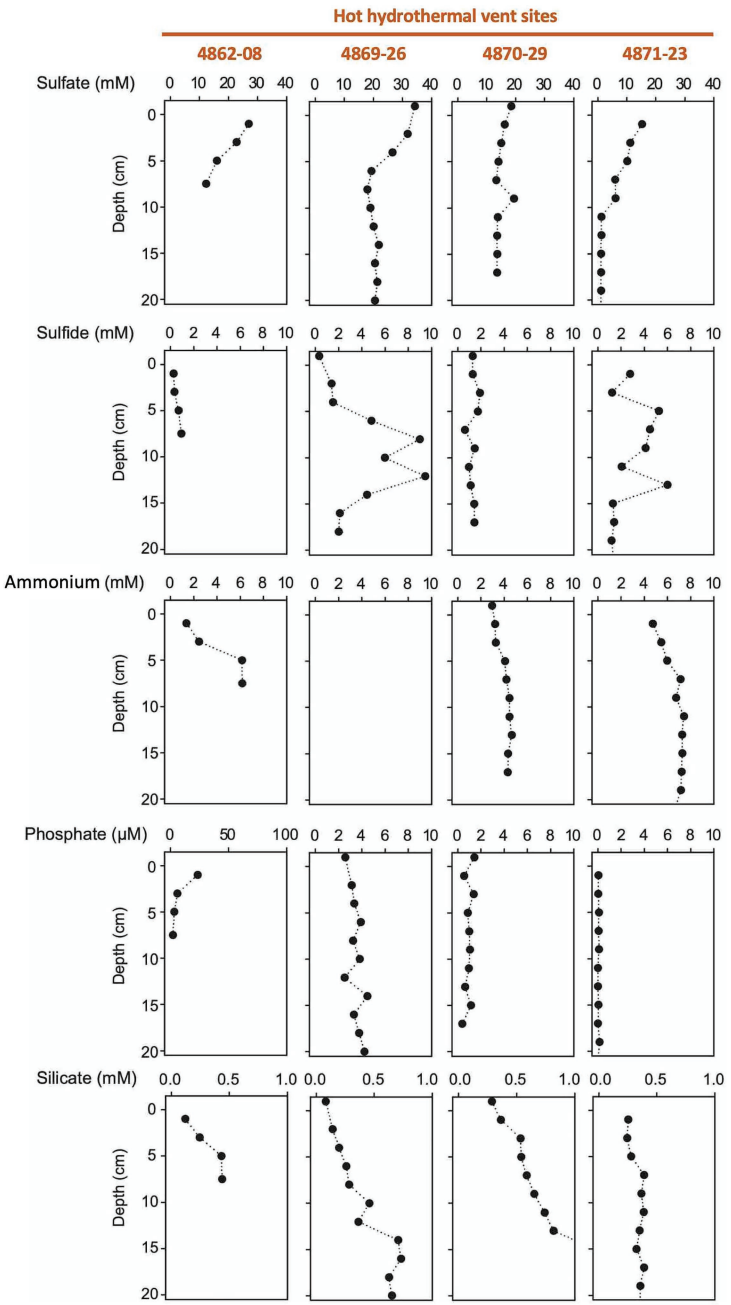

FIGURE 4 | Porewater concentration profiles of sulfate, sulfide, ammonium, phosphate, and silicate in four hydrothermal cores.

that was spreading over the entire hydrate area (**Figures 3D,E**). When push cores were removed from the sediment, the bottom end contained white hydrates, presumably methane hydrate, which dissociated during transport to the surface (**Figure 3E**). Sequence-based and geochemical analyses for these cores are briefly summarized in **Supplementary Table S1**.

Porewater nutrient profiles set the Octopus Mound sediments apart from their hydrothermal counterparts (**Figure 5**). The concentrations of ammonium were near 50 μM ("active" core 4867-30 and "background" core 4867-32), approximately two orders of magnitudes lower than measured in porewaters from the hydrothermal sites. Independently obtained ammonium profiles from nearby multicorer and gravity cores collected by the R/V *Sonne* expedition in 2015 yielded similar ammonium concentrations, around 30 μM in surficial sediments, and increasing to 300 μM at 4 m depth (Geilert et al., 2018). Phosphate concentrations remained below 2 μm in the active core and 10 μm in the background core sediments, respectively.

Porewater silicate concentrations in the *Alvin* push cores were in the range of 100–200 µm, near the bottom water concentration of ca. 175 µM in Guaymas Basin (Campbell and Gieskes, 1984), and much below the elevated dissolved silicate concentrations of 0.5–2 mM that characterize the porewater of hydrothermal Guaymas Basin cores (**Figure 5**).

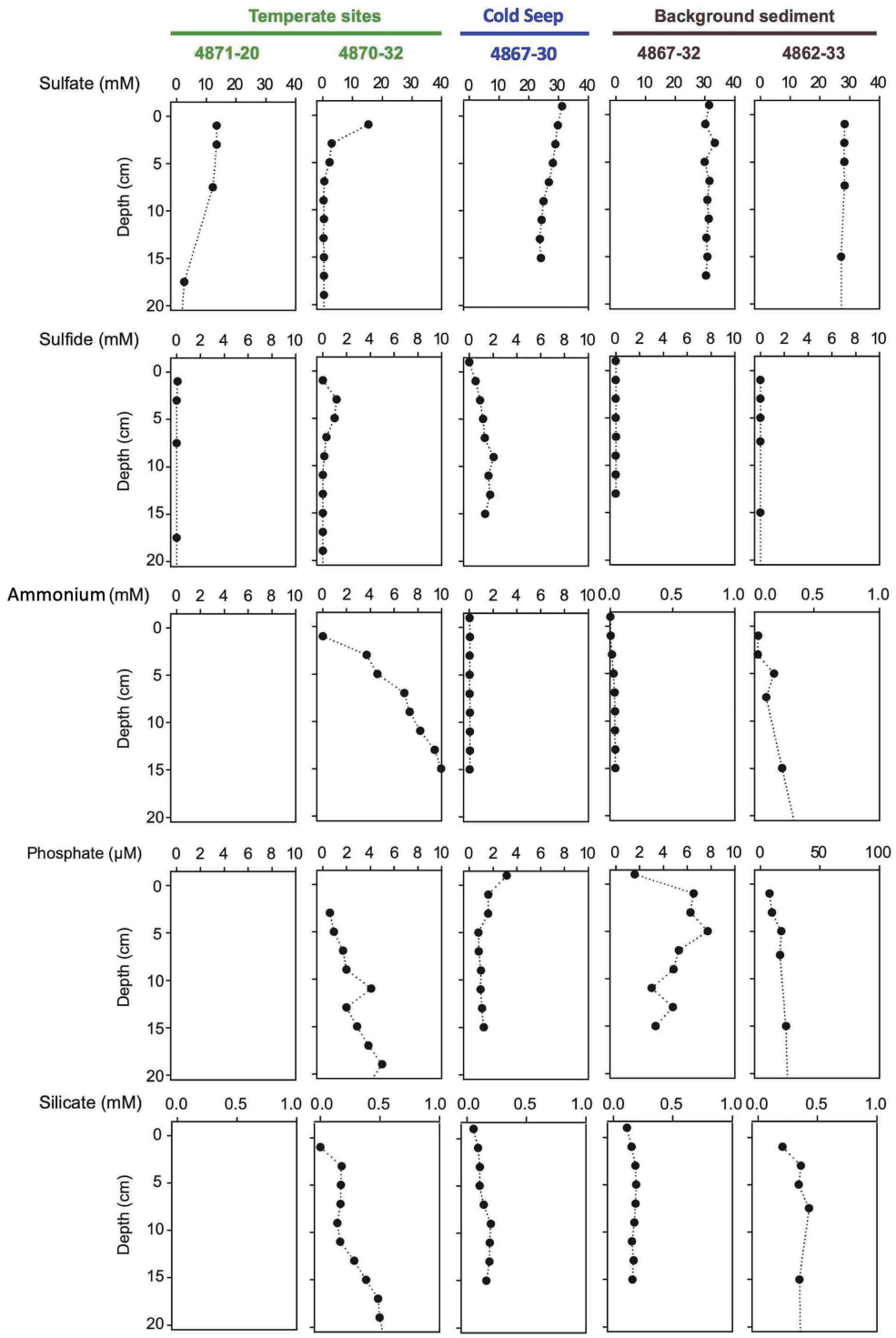

FIGURE 5 | Porewater concentration profiles of sulfate, sulfide, ammonium, phosphate, and silicate in two temperate cores, one seep core, and two background cores without visible microbial mats. Empty panels indicate data gaps.

Methane at Octopus Mound is predominantly biogenic, as indicated by Octopus Mound δ^{13}C-CH_4 values from −68 to −76‰ (**Figure 6A**), falling between the biogenic methane of Sonora Margin cold seeps (ca. −80 ‰, Vigneron et al., 2015) and the mixed biogenic/thermogenic methane at the off-axis Ringvent site (ca. −60 ‰, Teske et al., 2019). Methane concentrations in the Octopus Mound sediment cores were often supersaturated, ranging from 0.6 to 1.8 mM at the active site, and between 3.6 and 8.3 mM at the hydrate site (**Figure 6B**). Methane concentrations in sediments at the periphery and background sampling areas were below detection. A comparison to hydrothermal conditions is provided by two hydrothermal cores from the Northern Tower area, 4871-04 and 4871-27 (**Figure 6A**). Due to data gaps, these two cores were not included in broader biogeochemical and microbiological comparisons, but their microbial mat cover and their temperature maxima >100°C (**Supplementary Table S2**) identify them as hydrothermal. Their δ^{13}C-CH_4 profiles approach the values of previously studied hydrothermal cores dominated by thermogenic methane, and their multi-millimolar methane concentrations

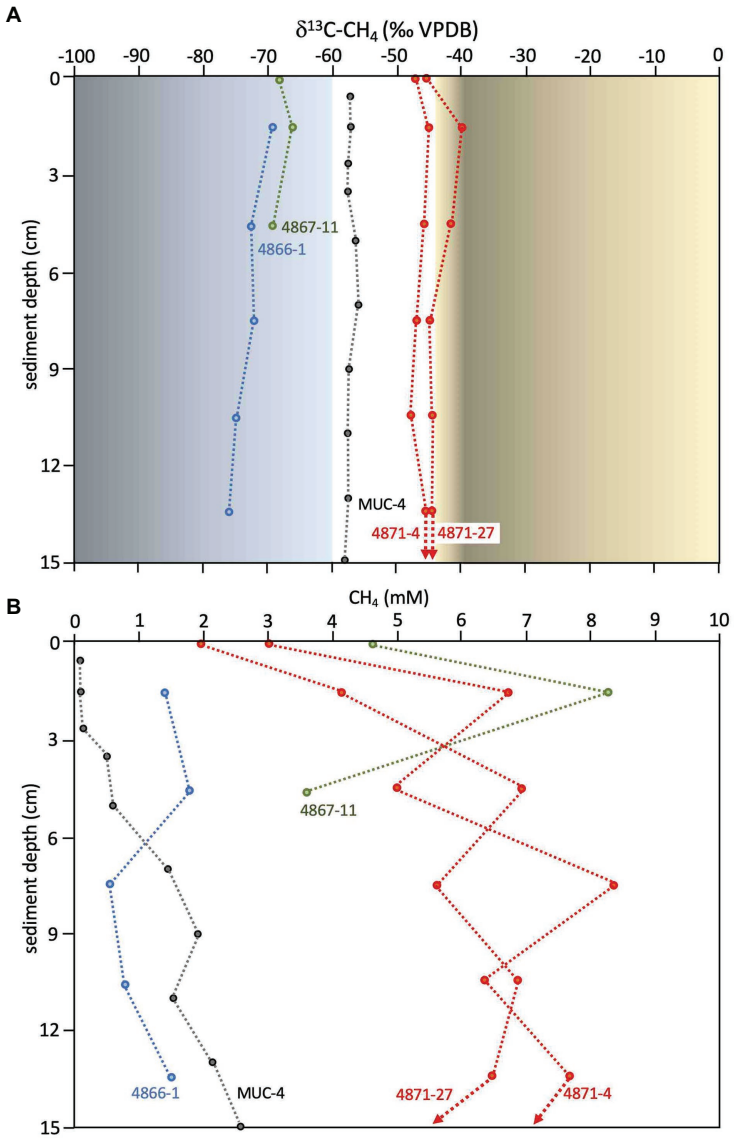

FIGURE 6 | Methane isotopic values and concentrations. **(A)** Methane $\delta^{13}CH_4$ values (VPDB) for Octopus Mound cores 4866-1 (active site, in blue) and 4867-11 (hydrate site, in green), and the nearby sediment push core MUC-4 (black), previously collected and measured independently at Central Seep (Geilert et al., 2018). Methane $\delta^{13}CH_4$ values (VPDB) for hydrothermal cores from the Northern Towers area (4871-4 and 4871-27) are plotted in red. Orange shading indicates the range of $\delta^{13}CH_4$ values for thermogenic methane in hydrothermal sediments in the southern Guaymas axial valley (McKay et al., 2016). Blue shading indicates the range of $\delta^{13}CH_4$ values for microbially produced methane in cold sediments of the Sonora Margin (Vigneron et al., 2015, Teske et al., 2019), delimited at 60‰ based on Schoell (1982) and Simoneit et al. (1986). **(B)** Methane concentrations for the same samples. Data points for supernatant samples are plotted at 0 cm depth. $\delta^{13}CH_4$ values and their standard deviations, and methane concentrations are tabulated in **Supplementary Table S4**.

are typical for hydrothermal methane-saturated sediments of Guaymas Basin (McKay et al., 2016). As illustrated by this comparison, biogenic methane dominates at Octopus Mound, whereas thermogenic methane is either entirely absent or limited to a minimal contribution.

Microbial Community Structure

The taxonomic composition of bacterial and archaeal 16S rRNA gene sequences recovered from surficial sediments is distinctly different for hot hydrothermal sediments, temperate hydrothermal sediments, and background sediments (**Figure 7**).

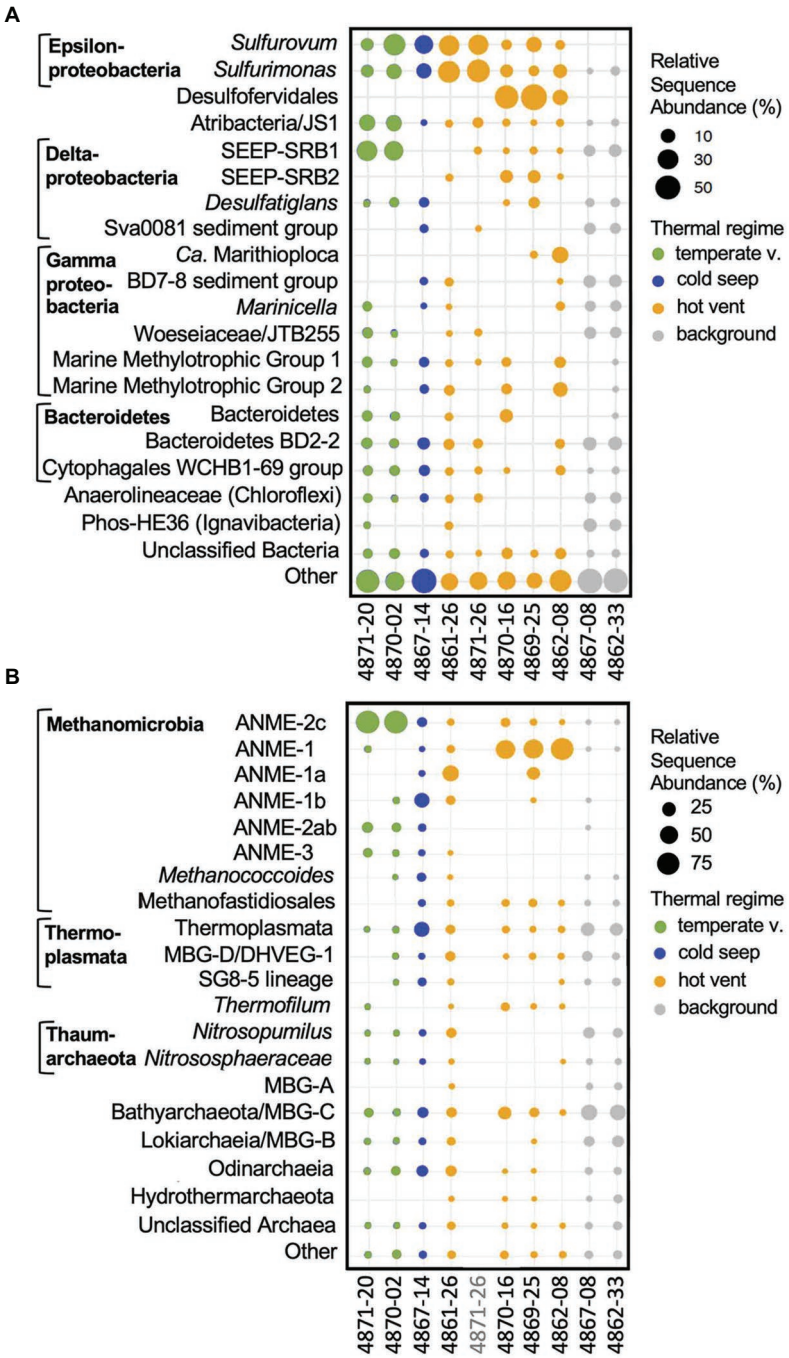

FIGURE 7 | Bacterial and archaeal community composition of Guaymas Basin sediment cores. The size of the dots indicates the relative sequence abundance of microbial clades based on 16S rRNA gene amplicon sequences. The 20 most abundant bacterial **(A)** and archaeal **(B)** family-level lineages are shown. Less abundant clades are summarized as "Other." When appropriate, taxa are annotated to supplement the automated SILVA identifications. DNA from core 4871-26 had run out after several sequencing attempts before archaeal sequencing could be finalized.

All sequence-based analyses have to be qualified by the fact that they are based on sequence frequencies, which are derived from the microbial community but do not necessarily represent it in identical proportions, and do not provide independent quantifications.

Hot Hydrothermal Sediment

Sequences of the sulfur-oxidizing chemoautotrophic genera *Sulfurovum* and *Sulfurimonas* within the Epsilonproteobacteria (Campbell et al., 2006) are broadly shared among sulfide-rich hydrothermal sediment cores of Guaymas Basin (**Figure 7A**). Mat-forming filamentous sulfur-oxidizing bacteria of the family *Beggiatoaceae* were abundant *in situ* (**Figure 2**), but the mats were removed for separate analyses before sediment slicing and DNA extraction. Therefore, *Beggiatoaceae* sequences (here assigned to *Ca*. Marithioploca, **Figure 7A**) were found in high abundance only in a single core (core 4862-08) where draining fluids had sucked the overlying mat into the surficial sediment. Sulfate-reducing bacteria are dominated by thermophiles: Sequences of the *Ca*. Desulfofervidus auxilii lineage occur in high frequency in the hot hydrothermal cores 4869-25 and 4870-16, and in reduced proportions in hydrothermal sediment core 4862-08. *Ca*. Desulfofervidus is the thermophilic bacterial syntroph of methane- and short-chain alkane-oxidizing archaea that dominates in enrichment cultures at temperatures of 37–60°C (Holler et al., 2011; Laso-Pérez et al., 2016; Krukenberg et al., 2018; Hahn et al., 2020). Interestingly, *Ca*. Desulfofervidus sequences were not recovered from core 4861-26, collected from an extensive mat area covered with white sulfur precipitates (**Figure 2A**) The intermediate temperatures there (T max. near 50°C, **Table 1**) may not suffice to sustain *Ca*. Desulfofervidus populations; the core smelled strongly sulfidic upon shipboard recovery but detailed geochemical data are not available. Sequences assigned to mesophilic sulfate-reducing bacteria of the SEEP-SRB1 cluster within the *Desulfobacteraceae* (Schreiber et al., 2010) account for smaller proportions of hydrothermal core sequences, and are absent in 4861-25. Sulfate-reducing bacteria of the SEEP-SRB2 cluster, an uncultured, presumably mesophilic and syntrophic lineage that may participate in short-chain alkane oxidation (Kleindienst et al., 2012; Krukenberg et al., 2018), occur in four out of five of the hot hydrothermal cores, except 4871-26 (**Figure 7A**). Heterotrophic, anaerobic phyla with fermentative potential (Atribacteria, Bacteroidetes, Chloroflexi) are represented by sequences from the hot hydrothermal cores, but were generally recovered from the temperature cores and background sediments (**Figure 7A**). Sequences of aerobic methylotroph MMG1 and MMG2 groups (Ruff et al., 2013) occur in hot and temperate hydrothermal cores (**Figure 7A**).

Sequences assigned to the dominant archaeal group, ANME-1, were recovered at the highest relative abundance from the fully mat-covered cores 4862-08, 4869-25, and 4870-16 (**Figure 7B**); other anaerobic methane-oxidizing archaea (ANME) types were found as well (most consistently, ANME-2c) but appear more frequently in the temperate hydrothermal cores. The most consistently found methanogenic group are methylotrophic *Methanofastidiosales* (Nobu et al., 2016b), previously detected as "Guaymas euryarchaeotal group" in Guaymas Basin hydrothermal sediments (Dhillon et al., 2005). Crenarchaeotal sequences assigned to the heterotrophic, thermophilic sulfur-reducing, and moderately acidophilic genus *Thermofilum* (Zillig et al., 1983) occur in smaller but consistently detected proportions. Uncultured sediment-associated archaea within the Thermoplasmatales, and the Bathyarchaeota and Asgardarchaeota are found in the hot and the temperate hydrothermal cores (**Figure 7B**).

Temperate Hydrothermal Sediment

These sediments share the epsilonproteobacterial sequences of the high-temperature hydrothermal cores, but they consistently lack *Ca.* Desulfofervidus (**Figure 7A**). Surface samples of cool cores 4870-2 and 4871-20 are characterized by frequent recovery of sequences attributed to the sulfate-reducing bacterial SEEP-SRB1 group (Knittel et al., 2003) which include the mesophilic syntrophic partners of ANME-1 methane-oxidizing archaea (Schreiber et al., 2010), and to the fermentative, mesophilic Atribacteria (Nobu et al., 2016a; Katayama et al., 2020) which are common inhabitants of cold seeps (Ruff et al., 2015; Chakraborty et al., 2020) and anaerobic subsurface sediments (Starnawski et al., 2017). Sequences from the aromatics-degrading sulfate-reducing *Desulfatiglans* lineage were recovered from all temperate cores, as well as from several hot hydrothermal and background samples, consistent with its cosmopolitan distribution (Jochum et al., 2018). The archaeal sequences recovered were predominantly assigned to sulfate-dependent ANME-2c methane oxidizers for the Northern Towers and Aceto Balsamico cores, and ANME-1b methane oxidizers for the Octopus Mound core (**Figure 7B**). Most other archaeal groups are shared with the background sediment, including members of the Bathyarchaeia, Thermoplasmata, Lokiarchaea, Marine Benthic Group D (renamed Thermoprofundales, Zhou et al., 2019), the uncultured SG8-5 Thermoplasmatales lineage found in estuarine sediments (Lazar et al., 2017), Thaumarchaeota of the aerobic, ammonia-oxidizing families *Nitrosopumilaceae* and *Nitrososphaeraceae* (Stieglmeier et al., 2014), methylotrophic methanogens of the uncultured *Methanofastidiosales* (Nobu et al., 2016b) and of the genus *Methanococcoides* (Liu and Whitman, 2008), and small proportions of ANME-2ab archaea. Some of these widely shared archaeal sequences show a comparatively spotty occurrence pattern in the hot hydrothermal samples, suggesting that strong hydrothermal conditions are selecting against them (**Figure 7B**). A preference for moderate over hot hydrothermal conditions characterizes the Lokiarchaeota and Odinarchaeia within the Asgardarchaeota (Zaremba-Niedzwiedzka et al., 2017) that were found recently in metagenomic surveys of Guaymas Basin hydrothermal sediments (Dombrowski et al., 2018; Seitz et al., 2019).

Background Sediment

Sequences of the sulfur-oxidizing microaerobic and nitrate-reducing chemoautotrophic genus *Sulfurimonas* occurred in reduced relative abundance in the background cores 4862-33 and 4867-08; the related sulfur-reducing genus *Sulfurovum* was no longer detected (**Figure 7A**). Sulfate-reducing bacteria were represented by SEEP-SRB-1, *Desulfatiglans*, and the Sva0081

group of uncultured Desulfobacteraceae that are hypothesized to scavenge H_2 in marine sediments (Dyksma et al., 2018). Uncultured marine sediment lineages that were frequently found include the deep-sea sediment lineage BD2-2 within the Bacteroidetes (Li et al., 1999) and the deep-sea sediment lineage BD7-8 within the Gammaproteobacteria (Li et al., 1999), the PHOS-HE36 clade (Dabert et al., 2001) within the Ignavibacteria (Iino et al., 2010), the cosmopolitan sediment-associated gammaproteobacterial JTB255 lineage (Woeseiaceae, Mussmann et al., 2017; Hoffmann et al., 2020), the WCHB1-69 lineage within the Cytophagales (Dojka et al., 1998), and diverse unclassified Anaerolineae. Based on the spotty detection of their sequences in vent and seep cores, these bacteria may tolerate some hydrothermal or seep activity, unless selected against by other extreme conditions, or they may be generally abundant in the surrounding sediment and thus occur as relic DNA in the seep cores. Sequences of aerobic methylotroph groups MMG1 and MMG2 (Ruff et al., 2013) were not detected (4867-08) or occurred in reduced proportions (4862-33; **Figure 7A**). The archaeal sequences resembled those from the temperate cores, with increased relative representation of the Bathyarchaeota and Thermoplasmatales, and the addition of Marine Benthic Group A, a sediment-dwelling sister lineage of the marine Thaumarchaeota (Lauer et al., 2016). Thus, the Guaymas Basin background sediments harbor the five globally distributed marine benthic archaeal groups of MBG-A, MBG-B (Lokiarchaeota), MBG-C (Bathyarchaeota), MBG-D (Thermoprofundales), and MBG-E (Thermoplasmata subgroup) that were originally discovered in cold North Atlantic seafloor sediments (Vetriani et al., 1999).

As previously observed for the geochemical characteristics of Octopus Mound sediments, the Octopus Mound seep core 4867-14 does not fit easily into these three categories; it differs from the background sediments by not having SEEP-SRB1 sequences among the 20 most frequently detected bacterial taxa (**Figure 7A**), from temperate sediments by its lower proportion of ANME-2c sequences, and from all sediments by its high frequency of ANME-1b sequences (**Figure 7B**).

Community Clustering

Non-metric multidimensional scaling ordination plots show distinct clustering patterns for bacterial and archaeal communities that are consistent with geochemical characteristics (**Figure 8**). The strongest separation occurs between cold background cores 4862-33 and 4867-08, which cluster together in archaeal and bacterial NMDS despite their geographical distance of ca. 50 km, and all other cores. Communities at hot and temperate sites are clearly separated from each other. The hot hydrothermal cores 4862-08, 4870-16, and 4869-25 stand apart from the temperate cores 4871-20, 4870-02, and the Octopus Mound core 4867-14, while the remaining hydrothermal cores (4871-26 and 4861-26) fall in between (**Figure 8**). Interestingly, core 4861-26 is the coolest of the hot hydrothermal cores, with a bottom temperature of 55°C (**Supplementary Figure S3**), and this intermediate thermal regime is consistent with its intermediate position in both archaeal and bacterial NMDS plots. Within the archaeal NMDS plots, the Octopus Mound core 4867-14 clusters apart from the cool hydrothermal sites (**Figures 8D,E,F**), consistent with its distinctive habitat characteristics. Except for the two background cores 4862-33 and 4867-08, no cores show overlapping NMDS clustering. These results suggest that sites with similar geochemistry can harbor microbial communities that diverge to some extent. Additional variables that were not measured could have an impact, or dynamic selection pressures due to rapidly changing hydrothermal regimes allow stochasticity to play an important role in site-specific community assembly (McKay et al., 2016). While these results contrast with a previous survey of Guaymas Basin microbial diversity that emphasized community overlap among hydrothermal sites within a constrained sampling area (Meyer et al., 2013), this survey covers a greater habitat range with distinct thermal and geochemical regimes, and thus distinct microbial communities.

Microbial Methane Oxidation

In contrast to previously sampled hydrothermal Guaymas Basin sediment communities, which were apparently dominated by ANME (Teske et al., 2002), the surface layer of the Octopus Mound site and several of the hydrothermal sites reported here have yielded 16S rRNA gene sequences of aerobic methanotrophic and methylotrophic bacteria (**Figure 7A**). Their detection could be linked to the sampling scheme; these aerobes are more likely to be detected in the top 0–1 cm layer sampled here than in surficial samples of multiple centimeters that are predominantly anoxic. The sequences are affiliated with the cultured methanotrophic genus *Methyloprofundus* (Tavormina et al., 2015), and the uncultured lineages Marine Methylotrophic Groups 2 and 3 within the Gammaproteobacteria (Ruff et al., 2013). These groups contain gene sequences from New Zealand cold seeps, the Håkon-Mosby mud volcano, methylotrophic mussel endobionts, and diverse seafloor sediments (**Figure 9**). Some Guaymas Basin OTUs are specifically related to OTUs from methane-rich seeps and mud volcanoes, the Hikurangi Margin in New Zealand that harbors abundant amphetarid worm communities (Ruff et al., 2013), and the Håkon-Mosby Mud volcano in the Norwegian Arctic Ocean (Ruff et al., 2019).

Analogous to bacterial methanotrophs, the archaeal anaerobic methane oxidizers also show considerable phylogenetic complexity (**Figure 10**). Within the broadly defined ANME-1 archaea, some Guaymas ASVs were affiliated with ANME-1a and 1b, closely related subgroups that co-occur in diverse cold seeps (Knittel et al., 2005). The ASVs assigned by the SILVA pipeline to the generic "ANME-1" category (**Figure 7B**) are phylogenetically divergent and affiliate with two different ANME-1 lineages, the uncultured and presumably thermolerant ANME-1 Guaymas lineage that occurs consistently in hydrothermal Guaymas sediments (see data compilation in Dowell et al., 2016), and a cluster termed "ANME-1a Guaymas II" that was cultured in methane-oxidizing enrichments at 50°C (Holler et al., 2011) and also appears consistently in Guaymas hydrothermal sediments (Dowell et al., 2016). Thus, the ANME-1 archaea that are specifically recovered from hydrothermal cores (4862-08, 4869-25, and 4870-16) belong to lineages that are

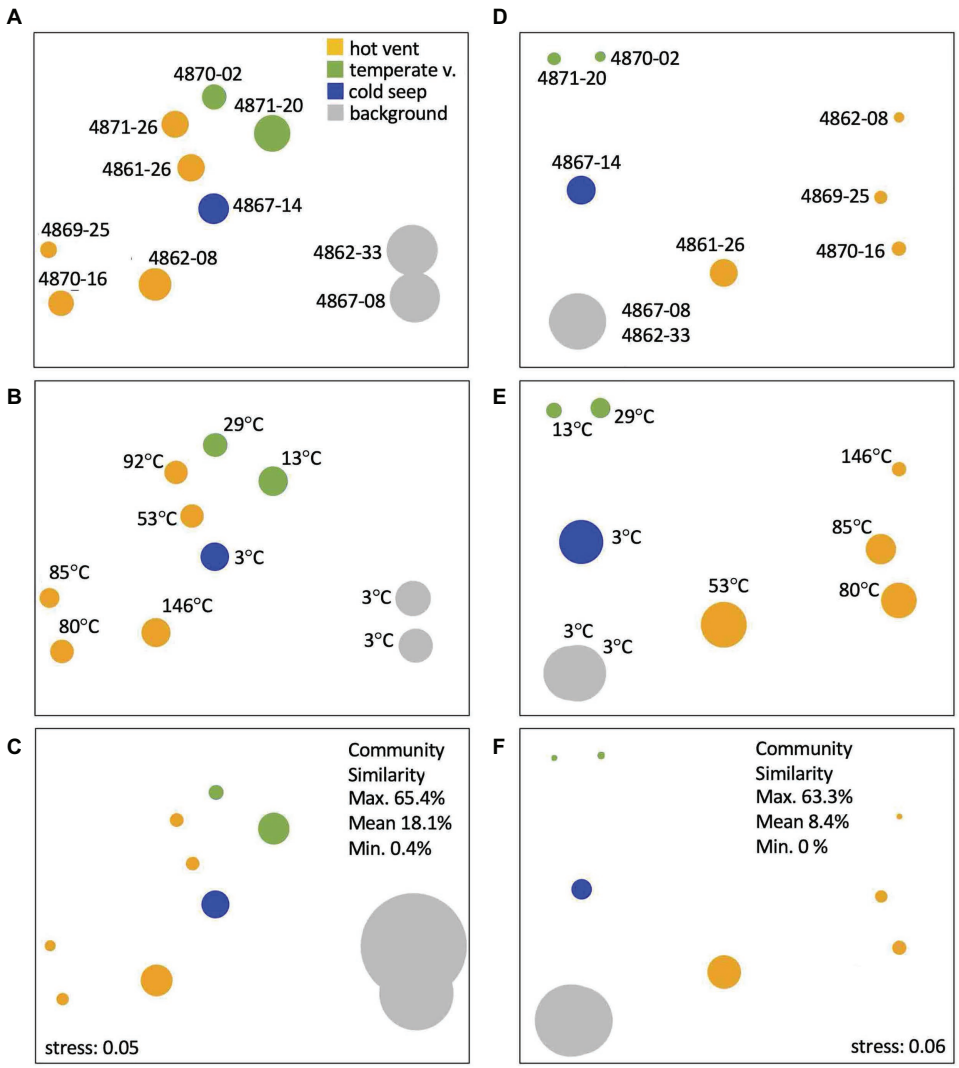

FIGURE 8 | Non-metric multidimensional scaling (NMDS) ordination plots based on 16S rRNA gene amplicons for bacterial operational taxonomic units (OTUs; **A-C**) and archaeal amplicon sequence variants (ASVs; **D-F**) from the upper cm of sediment cores, color-coded by geochemical cluster, and annotated with core number **(A,D)** and *in situ* temperature at 50 cm depth **(B,E)**. In plots **(A,D)**, circle size represents observed OTU richness. In plots **(B,E)**, circle size represents Shannon entropy. In plots **(C,F)**, circle size represents Inverse Simpson evenness.

either cultured thermophiles or show a preference for hydrothermal sediments (**Figure 10**). In contrast, sequences affiliated with the diverse ANME-2 subgroups were recovered predominantly from temperate cores (**Figure 6B**), consistent with previous observations (McKay et al., 2016). The ANME-3 subgroup, previously identified in cold seep sediments of an arctic mud volcano (Niemann et al., 2006; Lösekann et al., 2007), is also preferentially found in temperate and cold seep sediments (**Figure 7B**).

To further target anaerobic methanogenic and methane-oxidizing archaea at the Octopus Mound site, *mcrA* gene sequences that are diagnostic for these archaea were PCR-amplified and surveyed (**Figure 11**). The sediments at the ampharetid-dominated "active site," sampled by *Alvin* core 4866-1, yielded *mcrA* sequences of the ANME-2ab clades; below-surface layers of core 4867-26 from the nearby "periphery site" yielded ANME-1; and the surficial centimeters of core 4867-11 from the hydrate site yielded ANME-1 and ANME-2 sequences (**Figure 11**). No *mcrA* gene sequences of cultured methanogenic genera or families were detected.

DISCUSSION

Hydrothermalism vs. Cold Seepage

The strong contrasts between different Guaymas Basin hydrothermal sites and the Octopus Mound cold seeps in seafloor microbial community, benthic life, thermal characteristics, and porewater geochemistry highlight the diversity of geo-ecosystems in the greater Guaymas Basin region.

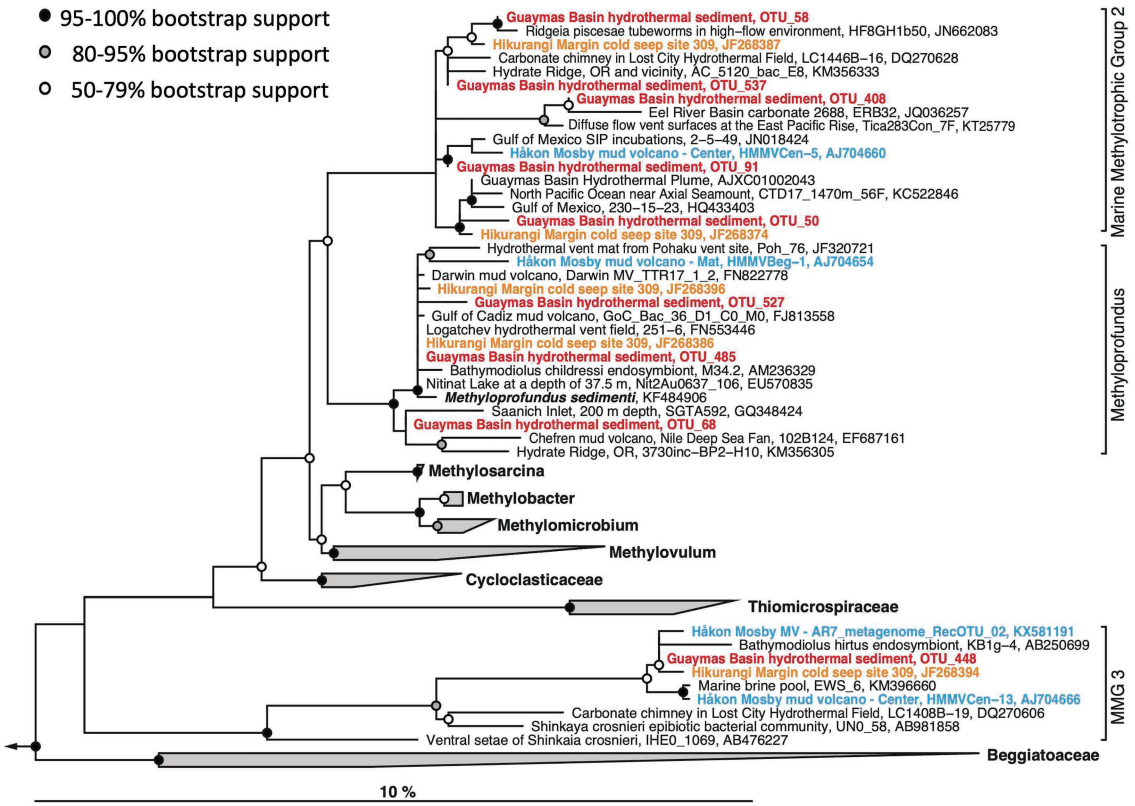

FIGURE 9 | Maximum likelihood phylogeny of gammaproteobacterial methylotrophic and methanotrophic bacterial OTUs. The tree was calculated using near full-length sequences; partial 16S rRNA gene sequences (ca. 450 nucleotide positions) obtained with primers 341F and Pro805 were added without changing the tree topology. The *Methyloprofundus* branch is synonymous with Marine Methylotroph Group 1 (MMG1), a sister lineage to MMG2 and MMG3. Sequences used for the phylogeny are listed in **Supplementary Table S5** for easy retrieval.

A closer look at the thermal and biogeochemical parameters of these different sites shows that they are not always tightly coupled, suggesting underlying transitions or hybrids between these environmental regimes.

With ammonium concentrations in the 5–10 mM range, silicate concentrations in the 0.5–2 mM range, and high sulfide concentrations of 1–3 mM, the hot hydrothermal sediment cores share key signatures of Guaymas Basin hydrothermal fluids (von Damm et al., 1985). Interestingly, these porewater indicators are not always linked to high temperatures. High ammonium and silicate concentrations are also observed in temperate hydrothermal sediments (core 4870-32 and 4871-20), suggesting that rising fluids with these hydrothermal signatures have cooled down locally before reaching the sediment surface. The lack of seawater sulfate below the upper 5 cm of sediment (**Figure 5**) indicates that the sediments are not cooled by seawater inmixing near the surface. Instead, the hydrothermal fluids must have cooled down considerably during migration through the sediments. These conditions are not indicative of a transient regime, but remain stable over many years, as observed in the Aceto Balsamico area where cool or temperate, sulfidic sediments with high acetate porewater concentrations and a conspicuous sulfur-rich lime-yellow surface layer were documented and sampled during previous cruises in 2008 and 2009 (Teske et al., 2016). These sediments were found in the same location and sampled again in 2016 and 2018. The resulting microbial niche is very different from the hydrothermal mixing regime and steep thermal gradients that characterize hot sediments overgrown with thick *Beggiatoaceae* mats (McKay et al., 2012, 2016).

Hydrothermal porewater signatures even persist in attenuated form in cores that, based on cold temperature and lack of microbial mats, have been categorized as background sediments when they were collected during *Alvin* dives. For example, the "background" core 4862-33 from the southern Guaymas Basin had ammonium and silicate concentrations of 300–400 µM, one order of magnitude above ammonium and double the silicate concentrations for the Octopus Mound background core 4867-08. Residual hydrothermal influence is also reflected in the slightly elevated sediment temperature, 3.4°C for core 4862-33 compared to 2.9°C for all Octopus Mound cores. Based on these observations, background sediments in the strict sense (no measurable hydrothermal influence at all) may have to be collected several miles away from hydrothermal features (Teske et al., 2019).

The Octopus Mound site does not fall along the spectrum between hydrothermal sediments and background endmembers.

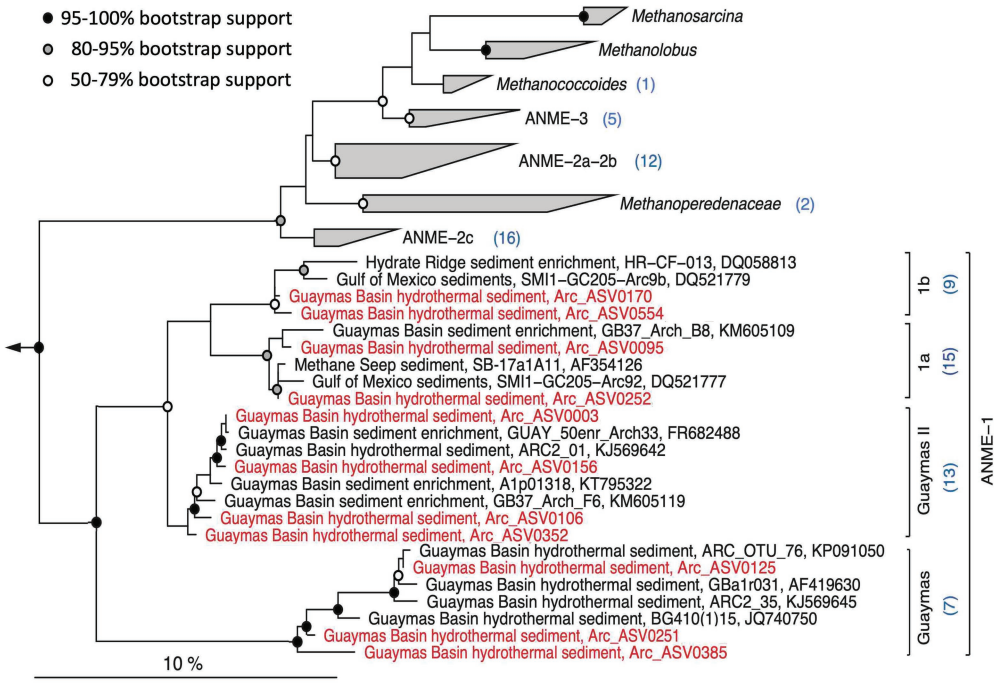

FIGURE 10 | Maximum likelihood phylogeny of ANME-1 archaea and related ANMEs and methanogens. The tree was calculated using near full-length sequences; partial 16S rRNA gene sequences (ca. 450 nucleotide positions) obtained with primers 517F and 958R were added without changing the tree topology. The annotation numbers in parentheses indicate the number of representative ASVs for different clades. Sequences used for the phylogeny are listed in **Supplementary Table S5** for easy retrieval.

Its uniformly cold temperatures profiles, extensive carbonate concretions, high sulfide and methane concentrations, white filamentous sulfur-oxidizing mats, chemosynthetic clams, and shallow methane hydrates are reminiscent of cold seeps at the base of the nearby Sonora Margin (Simoneit et al., 1990; Paull et al., 2007; Cruaud et al., 2017). Its δ^{13}C-CH_4 values are distinct from thermogenic methane in hydrothermal Guaymas Basin sediments ($\delta^{13}C \approx -38$ to 43‰; McKay et al., 2016). Instead, they occupy an intermediate position between the biogenic methane at Sonora Margin seeps ($\delta^{13}C \approx -80$‰; Vigneron et al., 2015) and mixed biogenic/thermogenic methane at the off-axis Ringvent site ($\delta^{13}C \approx -60$ ‰; Teske et al., 2019). Interestingly, Octopus Mound methane is isotopically lighter than porewater methane from cores collected elsewhere in the Central Seep area and other off-axis seep areas of Guaymas Basin ($\delta^{13}C \approx -55$ to −58‰ in MUC04; Geilert et al., 2018), and from deep subsurface sediments ($\delta^{13}C \approx -40$ to −65‰; Simoneit et al., 1986), suggesting additional biogenic contributions to the porewater methane pool.

Seafloor methane undergoes microbial oxidation, as shown by the presence of methane-derived carbonates, and ANME of different types. Carbonate samples collected at the base of Octopus Mound during *Alvin* dive 4867 showed the same morphology (**Figure 3**) as previously dredged carbonates from this location with δ^{13}C isotopic values of −46.6 and −44.7‰ (Geilert et al., 2018), and −45.2 to −47.6‰ (Núñez-Useche et al., 2018) that indicate the incorporation of methane-derived carbon. The sediments of Octopus Mound are populated with ANME archaea, as shown by 16S rRNA gene profiling (**Figure 7B**) and *mcrA* gene analysis (**Figure 11**). ANME-2 prefers surficial sediments with lower sulfide concentrations and tolerates some degree of oxygen exposure (Knittel et al., 2005; Rossel et al., 2011) that would be consistent with bio-irrigation by ampharetid worms, as shown by the identification of ANME-2ab archaea as the dominant ANME type in Ampharetid mats of the Hikurangi Margin seeps in New Zealand (Ruff et al., 2013). ANME-1 is generally found in reduced and sulfidic sediments (Rossel et al., 2011), consistent with the thick mat of sulfide-oxidizing *Beggiatoaceae* at the surface of the hydrate site, and with reducing conditions below the sediment surface in the "periphery" site, where bio-irrigating worms are absent.

Controls on Microbial Community Structure

The observation that the thermally and geochemically based core categories are generally reflected in NMDS analyses of bacterial and archaeal community composition should not be taken for granted but requires an explanation. *In situ* temperatures in the upper centimeter of hydrothermally active sediments and at the mat interface – the source of the bacterial and archaeal sequences – are almost always moderate or cool, mostly at around 5°C, or averaging ~10°C when high-temperature outliers are included (McKay et al., 2012). The integrated impact of the geochemical regime within the entire sediment core – a composite of electron acceptors and donors, nutrients, and

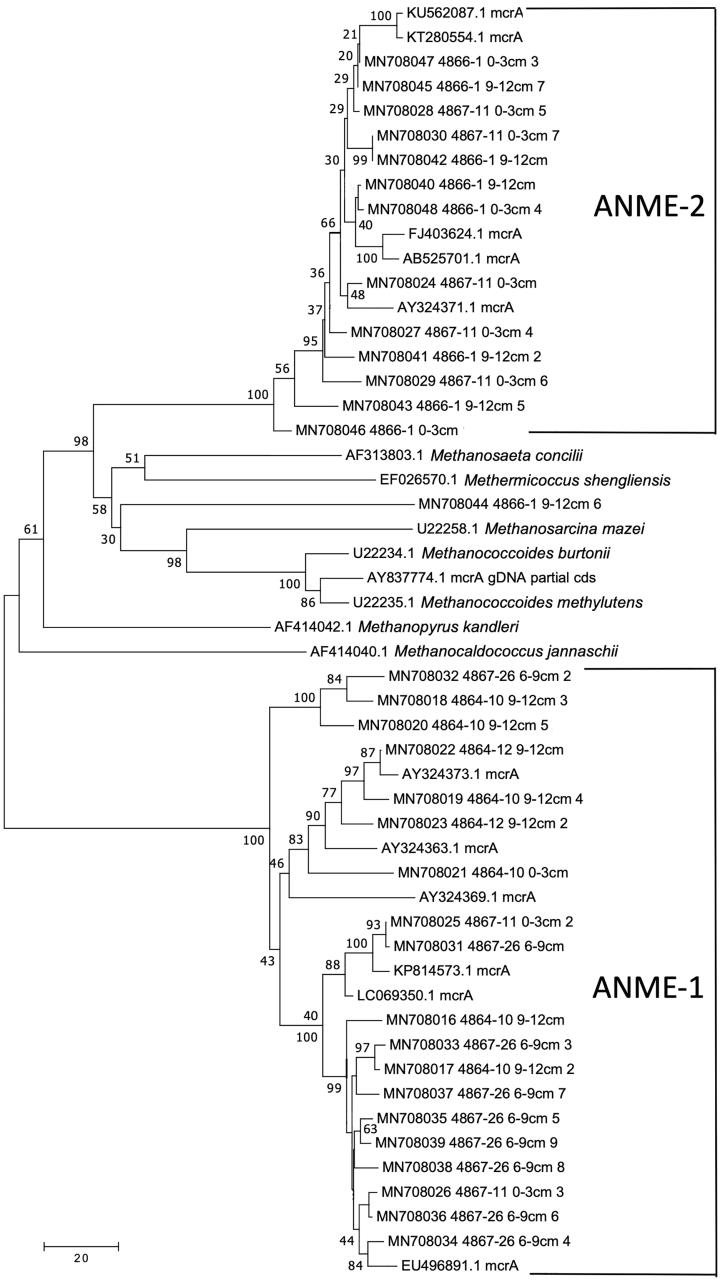

FIGURE 11 | Distance phylogeny of *mcrA* genes recovered from Octopus Mound sediments. Taxon labels begin with the Genbank number, followed by the *Alvin* Dive and core number, and (if applicable) the number of multiple clones from the same location and core that is represented by this sequence. Additional *mcrA* genes come from Ringvent sediments sampled during *Alvin* dive 4864 (Teske et al., 2019). Bootstrap numbers were obtained by 500 replicates.

carbon sources – shapes the community composition at the sediment surface. In this view, the sediment surface may function as an integrator for microbial cells that may originate deeper within the sediment and thus reflect its average geochemical regime; these cells may move upwards with rising hydrothermal and seep fluids and accumulate at the sediment surface, as recently proposed for seep microbiota (Chakraborty et al., 2020). Consistently, microbial cell numbers (Meyer et al., 2013), cell densities of cultivable thermophiles (Teske et al., 2009), microbial lipid concentrations (Guezennec et al., 1996; Teske et al., 2002; Schouten et al., 2003), and microbial activities such as sulfate reduction rates (Weber and Jørgensen, 2002; Dowell et al., 2016) and acetate oxidation rates (Zhuang et al., 2019) are strongly focused towards maximum values within the top 3–5 cm of surficial sediment. These patterns are consistent with geochemical observations that hydrothermal activity

mobilizes biomass and organic carbon from the deeper sediments towards the sediment/water interface (Lin et al., 2017).

Similar to a previous study of the hydrothermal core community in Guaymas Basin sediments (Cruaud et al., 2017), sulfate-reducing and methane-oxidizing microbial groups emerge as indicator species of hydrothermal activity: the facultatively syntrophic, hydrogen-oxidizing sulfate reducer *Ca*. Desulfofervidus (Krukenberg et al., 2016), the uncultured, presumably syntrophic sulfate-reducing and alkane-oxidizing SEEP-SRB2 lineage (Krukenberg et al., 2018), and the thermophilic ANME-1 lineages (**Figure 9**) that are distinct from the cold-seep ANME-1a and ANME-1b groups (Biddle et al., 2012) appear characteristic for hot hydrothermal sediments, consistent with a synopsis of published 16S rRNA genes from Guaymas Basin (Dowell et al., 2016). In contrast, members of the SEEP-SRB1 lineage within the Desulfobacteraceae were present in hot hydrothermal core samples, but are more frequently found in cool or background sediment cores. Interestingly, the SEEP-SRB1 bacteria were not detected in Octopus Mound core 4867-14, although the thermal conditions would have been suitable. Methane-oxidizing syntrophic SEEP-SRB1 bacteria (Schreiber et al., 2010) could be outcompeted in this location, if bio-irrigation by mat-like assemblages of ampharetid polychaetes (**Figure 3**) favors aerobic methane oxidation by widespread gammaproteobacterial methylotrophs, which occur in 4867-14 and in other cores (**Figure 5**).

The archaeal spectrum in the diverse Guaymas Basin samples overlaps with the benthic archaeal core community of ANME-1, Thermoplasmatales (Marine Benthic Group D), Lokiarchaeota, and Bathyarchaeota that was previously proposed for Guaymas Basin hydrothermal sediments (Cruaud et al., 2017). While benthic anaerobic archaea predominate, they coexist with members of the Thaumarchaeota (*Nitrosopumilaceae*) in several hydrothermal sediments. Although aerobic, ammonia-oxidizing *Nitrosopumilaceae* (Qin et al., 2017) dominate archaeal communities in the water column, ammonium provided by hydrothermal fluids or by nitrate-reducing *Beggiatoaceae* mats enriches these archaea in surficial hydrothermal sediments as well (Winkel et al., 2014). These microbial groups with different physiologies and requirements may coexist in surficial hydrothermal sediments since this habitat provides an interface with moderate temperatures where electron donors and carbon sources from the sediment overlap with electron acceptors from the water column and create highly compressed or co-existing biogeochemical niches (Schutte et al., 2018; Buckley et al., 2019). Hydrothermal circulation in surficial sediments with microbial mats (Gundersen et al., 1992; Teske et al., 2016) can transport anaerobic or thermophilic subsurface archaea to the seawater interface and mix them with aerobic Thaumarchaeota (Qin et al., 2017). The moderate temperatures just below the sediment/seawater interface would be compatible with mesophilic microorganisms, and at the same time allow the accumulation of thermophiles and hyperthermophiles that grow at higher temperatures a few centimeters downcore (Teske et al., 2009). Since surficial hydrothermal sediments harbor diverse mesophiles as well as thermophiles, they represent an attractive target to explore the highly diverse microbiota of Guaymas Basin, with outstanding potential for evolutionary and physiological discoveries (Dombrowski et al., 2018; Seitz et al., 2019).

Outlook: The Hydrothermal Landscape of Guaymas Basin

The contrasting hydrothermal regimes and microbial communities in Guaymas Basin reflect deeply sourced fluid and gas flow that is driven by underlying sill emplacement and local hydrothermal temperature gradients. Hydrothermal features in the southern Guaymas Basin spreading center are linked to underlying shallow sills (Lonsdale and Becker, 1985; Teske et al., 2016); the high local variability within hydrothermal areas indicates further differentiation of fluid and gas transport to the sediment surface (Ondréas et al., 2018). High-resolution bathymetries combined with shallow subbottom seismic profiles penetrating ca. 30–60 m below the sediment surface show that the southern vent sites, including Mat Mound Massif (synonymous with the "Orpheus" site, Ondréas et al., 2018), Rebecca's Roost, Cathedral Hill, and Aceto Balsamico, are linked to small sub-circular seafloor depressions with massive subsurface hydrothermal precipitate formation and lithification that creates convoluted flow paths skirting surface-breaching hydrothermal edifices. These complex and relatively shallow subsurface flow paths are consistent with frequent observations of hydrothermal hot spots and microbial mats at the base or the lower slope of these hydrothermal mounds (Dowell et al., 2016; Teske et al., 2016). The collapsed depressions are thought to facilitate the release of soluble light hydrocarbon (gas, oil, and condensates) that are transported by hydrothermal fluids towards shallow sediments where they accumulate (Ondréas et al., 2018). This setting has produced a vast patchwork of hydrothermal sediments and microbial mats, frequently visited by microbiological surveys (e.g., Teske et al., 2002; Cruaud et al., 2017; Dombrowski et al., 2018), including this study. In contrast, hydrothermal circulation at the northern sites, such as the Northern Towers area, would be linked to deeper faults and remobilize deeper hydrocarbons (Ondréas et al., 2018); we speculate that the resulting network of subsurface flow paths would be more channelized and less diversified compared to the southern area, and therefore generate large hydrothermal edifices, but fewer microbial mats and seafloor hot spots, as indicated by the *Sentry* survey of this area. Multichannel seismic surveys along the entire southern axial trough of Guaymas Basin would be needed to further substantiate such a scenario. Interestingly, a difference in subsurface hydrothermal circulation is suggested by the δ^{13}C-CH_4 isotopic values of hydrothermal sediments in the Northern Towers area that are slightly heavier (near −45‰, **Figure 6A**) than hydrothermal methane in the main sampling area of the southern axial trough (−39 to −43‰, McKay et al., 2016). This difference between the two sampling areas has been noticed previously (Teske et al., 2016), and could indicate the impact of isotopically heavier deep subsurface methane in the Northern Towers area (Simoneit et al., 1986).

The setting is entirely different at Octopus Mound. Multichannel seismic profiles show a gas pipe rising from a deeply buried sill underneath Octopus Mound that appears to funnel deeply sourced methane towards the sediment surface and into hydrate reservoirs (indicated by extensive shallow bottom simulating

reflectors) around the mound (Teske et al., 2018). Deep sill-driven methane seepage is the default mechanism of hydrocarbon and methane mobilization across the spreading center and flanking regions of Guaymas Basin (Lizarralde et al., 2011). Yet, significant local variability with regard to sill depth, age, and thermal stage can also characterize the off-axis sites. For example, the deep and presumably old sill intrusion, cold seepage and predominantly biogenic δ^{13}C-CH_4 isotope values at Octopus Mound contrast with the shallow, recently emplaced sill, locally high temperatures and heavier, thermogenically influenced δ^{13}C-CH_4 isotope values at Ringvent (Teske et al., 2019), and with similar thermogenically influenced δ^{13}C-CH_4 values at diverse off-axis seep locations of Guaymas Basin (Geilert et al., 2018). Another factor is the localized biological modification of isotopic signatures; the Octopus Mound data indicate that biogenic methane sources are augmenting subsurface-derived methane at active seep locations. Given its compact size, habitat diversity, and geochemical contrasts, Octopus Mound and its surrounding seep sediments and shallow hydrates in the Central Seep area provide a rewarding model system for the study of sill-driven cold seepage in Guaymas Basin.

DATA AVAILABILITY STATEMENT

The 16S rRNA miseq data are deposited under Sequence Read Archive project PRJNA626075 at Genbank. The methyl coenzyme M reductase alpha subunit gene sequences are deposited at GenBank under nucleotide accession numbers MN708024 to MN708048. The geochemical data are available from the Biological and Chemical Oceanography Data Management Office at the Woods Hole Oceanographic Institution under Project Number 474317.

AUTHOR CONTRIBUTIONS

AT headed the R/V *Atlantis* expedition in December 2016, collected thermal profiles and samples, compiled the biological seafloor observations, developed the concept for this manuscript, and wrote the manuscript with input from all authors. Porewater geochemical analyses for ammonium, phosphate, silicate, sulfate, and sulfide were supervised and compiled by GW, and by JC for methane. DW isolated DNA from Octopus Mound samples, amplified the *mcrA* genes, and constructed the *mcrA* gene phylogeny. BM kept core records and core photographs throughout the cruise, photographed filamentous *Beggiatoaceae* from the hydrate site, and selected coring sites and recorded thermal profiles during *Alvin* dive 4869. MS and DH measured the thermal sediment gradients during dives 4866 and 4867, respectively. DH also inspected *Sentry* survey images for microbial mats and suitable sampling sites, and took *Alvin* video footage of cephalopod residents at Octopus Mound. DB and GZ measured thermal profiles during *Alvin* dives 4862 and 4871 and selected coring sites, respectively. SJ co-organized the 2016 *Atlantis* expedition. SAS co-wrote the *Sentry* proposal for this cruise, compiled the *Sentry* bathymetry, and together with DL developed the interpretation of the Octopus Mound site. SER extracted DNA from surface samples, performed 16S rRNA gene sequence-based community characterizations and statistical comparisons, and inferred phylogenetic trees.

FUNDING

Research on Guaymas Basin in the Teske lab is supported by NSF Molecular and cellular Biology grant 1817381 "Collaborative Research: Next generation physiology: a systems-level understanding of microbes driving carbon cycling in marine sediments". Sampling in Guaymas Basin was supported by collaborative NSF Biological Oceanography grants 1357238 and 1357360 "Collaborative Research: Microbial carbon cycling and its interaction with sulfur and nitrogen transformations in Guaymas Basin hydrothermal sediments" to AT and SJ, respectively. SER was supported by an AITF/Eyes High Postdoctoral Fellowship and start-up funds provided by the Marine Biological Laboratory.

ACKNOWLEDGMENTS

We thank the *Alvin* and *Sentry* teams for a stellar performance during Guaymas Basin cruise AT37-06. We also thank Martina Alisch (MPI Bremen) and Claire Wilson (FSU) for generating the porewater data and Carmen Li (University of Calgary), Aleksey Morozov, Nicole Robichaud, Hilary Morrison, and Sherlynette Pérez Castro (MBL) for support with amplicon sequencing. AT highly appreciates the safe and conducive writing environment provided by the Hanse Institute for Advanced Studies (Hanse Wissenschaftskolleg) in Delmenhorst during the COVID-19 pandemic.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2021.633649/full#supplementary-material

REFERENCES

Beaulieu, S. E., and Szafranski, K. M. (2020). InterRidge Global Database of active submarine hydrothermal vent fields. Version 3.4. PANGAEA. doi: 10.1594/PANGAEA.917894

Berndt, C., Hensen, C., Mortera-Gutierrez, C., Sarkar, S., Geilert, S., Schmidt, M., et al. (2016). Rifting under steam – how magmatism triggers methane venting from sedimentary basins. *Geology* 44, 767–770. doi: 10.1130/G38049.1

Biddle, J. F., Cardman, Z., Mendlovitz, H., Albert, D. B., Lloyd, K. G., Boetius, A., et al. (2012). Anaerobic oxidation of methane at different temperature regimes in Guaymas Basin hydrothermal sediments. *ISME J.* 6, 1018–1031. doi: 10.1038/ismej.2011.164

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 326–349.

Buckley, A., MacGregor, B. J., and Teske, A. (2019). Identification, expression and activity of candidate nitrite reductases from orange *Beggiatoaceae*, Guaymas Basin. *Front. Microbiol.* 10:644. doi: 10.3389/fmicb.2019.00644

Callaghan, 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

Campbell, B., Engel, A. S., Porter, M. L., and Takai, K. (2006). The versatile epsilon-proteobacteria: key players in sulphidic habitats. *Nat. Rev. Microbiol.* 4, 458–468. doi: 10.1038/nrmicro1414

Campbell, A. T., and Gieskes, J. M. (1984). Water column anomalies associated with hydrothermal activity in the Guaymas Basin, Gulf of California. *Earth Planet. Sci. Lett.* 68, 57–72.

Chakraborty, A., Ruff, S. E., Dong, X., Ellefson, E. D., Li, C., Brooks, J. M., et al. (2020). Hydrocarbon seepage in the deep seabed links subsurface and seafloor biospheres. *Proc. Natl. Acad. Sci. U. S. A.* 117, 11029–11037. doi: 10.1073/pnas.2002289117

Chanton, J., and Liptay, K. (2000). Seasonal variation in methane oxidation in a landfill cover soil as determined by an *in situ* stable isotope technique. *Global Biogeochem. Cycles* 14, 51–60.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* 11, 265–270.

Cline, J. D. (1969). Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol. Oceanogr.* 14, 454–458.

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

Cruaud, P., Vigneron, A., Pignet, P., Caprais, J. -C., Lesongeur, F., Toffin, L., et al. (2017). Comparative study of Guaymas Basin microbiomes: cold seeps vs. hydrothermal vents sediments. *Front. Mar. Sci.* 4:417. doi: 10.3389/fmars.2017.00417

Dabert, P., Sialve, B., Delgenes, J. P., Moletta, R., and Godon, J. J. (2001). Characterization of the microbial 16S rDNA diversity of an aerobic phosphorus-removal ecosystem and monitoring of its transition to nitrate respiration. *Appl. Microbiol. Biotechnol.* 55, 500–509. doi: 10.1007/s002530000529

Dhillon, A., Lever, M., Lloyd, K. G., Albert, D. B., Sogin, M. L., and Teske, A. (2005). Methanogen diversity evidenced by molecular characterization of methyl coenzyme M reductase A (*mcrA*) genes (*mcrA*) in hydrothermal sediments of the Guaymas Basin. *Appl. Environ. Microbiol.* 71, 4592–4601.

Dojka, M. A., Hugenholtz, P., Haack, S. K., and Pace, N. R. (1998). Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* 64, 3869–3877. doi: 10.1128/AEM.64.10.3869-3877.1998

Dombrowski, N., Teske, A. P., and Baker, B. J. (2018). Extensive metabolic versatility and redundancy in microbially diverse, dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* 9:4999. doi: 10.1038/s41467-018-07418-0

Dong, X., Kleiner, M., Sharp, C. E., Thorson, E., Li, C., Liu, D., et al. (2017). Fast and simple analysis of miseq amplicon sequencing data with MetaAmp. *Front. Microbiol.* 8:1461. doi: 10.3389/fmicb.2017.01461

Dowell, F., Cardman, Z., Dasarathy, S., Kellermann, M. Y., McKay, L. J., MacGregor, B. J., et al. (2016). Microbial communities in methane and short alkane-rich hydrothermal sediments of Guaymas Basin. *Front. Microbiol.* 7:17. doi: 10.3389/fmicb.2016.00017

Dyksma, S., Pjevac, P., Ovanesov, K., and Mussmann, M. (2018). Evidence for H_2 consumption by uncultured Desulfobacterales in coastal sediments. *Environ. Microbiol.* 20, 450–461. doi: 10.1111/1462-2920.13880

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

Einsele, G., Gieskes, J. M., Curray, J., Moore, D. M., Aguago, E., Aubry, M. P., et al. (1980). Intrusion of basaltic sills into highly porous sediments and resulting hydrothermal activity. *Nature* 283, 441–445.

Geilert, S., Hensen, C., Schmidt, M., Liebetreu, V., Scholz, V., Doll, M., et al. (2018). Transition from hydrothermal vents to cold seeps records timing of carbon release in the Guaymas Basin, Gulf of California. *Biogeosciences* 15, 5715–5731. doi: 10.5194/bg-2018-12

Grasshoff, K., Kremling, K., and Ehrhardt, M. (2009). *Methods of seawater analysis*. Weinheim, Germany: John Wiley & Sons.

Guezennec, J. G., Dussauze, J., Bian, M., Rocchicioli, F., Ringelberg, D., Hedrick, D. B., et al. (1996). Bacterial community structure from Guaymas Basin, Gulf of California, as determined by analysis of phospholipid ester-linked fatty acids. *J. Mar. Biotechnol.* 4, 165–175.

Gundersen, J. K., Jørgensen, B. B., Larsen, E., and Jannasch, H. W. (1992). Mats of giant sulphur bacteria on deep-sea sediments due to fluctuating hydrothermal flow. *Nature* 360, 454–456.

Hahn, C., Laso-Pérez, R., Volcano, F., Vaziourakis, K. -M., Stokke, R., Steen, I. H., et al. (2020). *Candidatus* Ethanoperedens, a thermophilic genus of archaea mediating the anaerobic oxidation of ethane. *MBio* 11:e00600-20. doi: 10.1128/mBio.00600-20

Hill, T. C. J., Walsh, K. A., Harris, J. A., and Moffett, B. F. (2003). Using ecological diversity measures with bacterial communities. *FEMS Microbiol. Ecol.* 43, 1–11. doi: 10.1111/j.1574-6941.2003.tb01040.x

Hoffmann, K., Bienhold, C., Buttigieg, P. L., Knittel, K., Laso-Pérez, R., Rapp, J. Z., et al. (2020). Diversity and metabolism of Woeseiales bacteria, global members of marine sediment communities. *ISME J.* 14, 1042–1056. doi: 10.1038/s41396-020-0588-4

Holler, T., Widdel, F., Knittel, K., Amann, R., Kellermann, M. Y., Hinrichs, K. -U., et al. (2011). Thermophilic anaerobic oxidation of methane by marine microbial consortia. *ISME J.* 5, 1946–1956. doi: 10.1038/ismej.2011.77

Iino, T., Mori, K., Uchino, Y., Nakagawa, T., Harayama, S., and Suzuki, K. (2010). *Ignavibacterium album* gen. nov., sp. nov., a moderately thermophilic anaerobic bacterium isolated from microbial mats at a terrestrial hot spring and proposal of *Ignavibacteria* classis nov., for a novel lineage at the periphery of green sulfur bacteria. *Int. J. Syst. Evol. Microbiol.* 60, 1376–1382. doi: 10.1099/ijs.0.012484-0

Jochum, L. M., Schreiber, L., Marshall, I. P. G., Jørgensen, B. B., Schramm, A., and Kjeldsen, K. U. (2018). Single-cell genomics reveals a diverse metabolic potential of uncultivated *Desulfatiglans*-related Deltaproteobacteria widely distributed in marine sediment. *Front. Microbiol.* 9:2038. doi: 10.3389/fmicb.2018.02038

Katayama, T., Nobu, M. K., Kusada, H., Meng, X. -Y., Hosogi, N., Uematsu, K., et al. (2020). Isolation of a member of the candidate phylum Atribacteria reveals a unique cell membrane structure. *Nat. Commun.* 11:6382. doi: 10.1038/s41467-020-20149-5

Kleindienst, S., Ramette, A., Amann, R., and Knittel, K. (2012). Distribution and *in situ* abundance of sulfate-reducing bacteria in diverse marine hydrocarbon seep sediments. *Environ. Microbiol.* 14, 2689–2710. doi: 10.1111/j.1462-2920.2012.02832.x

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

Knittel, K., Boetius, A., Lemke, A., Eilers, H., Lochte, K., Pfannkuche, O., et al. (2003). Activity, distribution, and diversity of sulfate reducers and other bacteria in sediments above gas hydrate (Cascadia margin, Oregon). *Geomicrobiol J.* 20, 269–294. doi: 10.1080/01490450303896

Knittel, K., Lösekann, T., Boetius, A., Kort, R., and Amann, R. (2005). Diversity and distribution of Methanotrophic Archaea at cold seeps. *Appl. Environ. Microbiol.* 71, 467–479. doi: 10.1128/AEM.71.1.467-479.2005

Krukenberg, V., Harding, K., Richter, M., Glöckner, F. -O., Gruber-Vodicka, H. R., Adam, B., et al. (2016). *Candidatus* Desulfofervidus auxilii, a hydrogenotrophic sulfate-reducing bacterium involved in the thermophilic anaerobic oxidation of methane. *Environ. Microbiol.* 18, 3073–3091. doi: 10.1111/1462-2920.13283

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

Kruskal, J. B. (1964). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–129.

Laso-Pérez, R., Wegener, G., Knittel, K., Widdel, F., Harding, K. J., Krukenberg, V., et al. (2016). Thermophilic archaea activate butane *via* alkyl-coenzyme M formation. *Nature* 539, 396–401. doi: 10.1038/nature20152

Lauer, A., Sørensen, K. B., and Teske, A. (2016). Phylogenetic characterization of marine benthic *Archaea* in organic-poor sediments of the Eastern Equatorial Pacific Ocean (ODP site 1225). *Microorganisms* 4:32. doi: 10.3390/microorganisms4030032

Lazar, C. S., Baker, B. J., Seitz, K. W., and Teske, A. P. (2017). Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISME J.* 11, 1118–1129. doi: 10.1038/ismej.2016.189

Lever, M. A., and Teske, A. (2015). Methane-cycling archaeal diversity in hydrothermal sediment investigated by general and group-specific functional

gene and 16S rRNA gene PCR primers. *Appl. Environ. Microbiol.* 81, 1426–1441. doi: 10.1128/AEM.03588-14

Li, L., Kato, C., and Horikoshi, K. (1999). Bacterial diversity in deep-sea sediments from different depths. *Biodivers. Conserv.* 8, 659–677.

Lin, Y. -S., Koch, B. P., Feseker, T., Ziervogel, K., Goldhammer, T., Schmidt, F., et al. (2017). Near-surface heating of young rift sediment causes mass production and discharge of reactive dissolved organic matter. *Sci. Rep.* 7:44864. doi: 10.1038/srep44864

Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic and ecological diversity of the methanogenic archaea. *Ann. N. Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019

Lizarralde, D., Soule, A., Seewald, J., and Proskurowski, G. (2011). Carbon release by off-axis magmatism in a young sedimented spreading centre. *Nat. Geosci.* 4, 50–54. doi: 10.1038/ngeo1006

Lonsdale, P., and Becker, K. (1985). Hydrothermal plumes, hot springs, and conductive heat flow in the southern trough of guaymas basin. *Earth Planet. Sci. Lett.* 73, 211–225.

Lösekann, T., Knittel, K., Nadalig, T., Fuchs, B., Niemann, H., Boetius, A., et al. (2007). Diversity and abundance of aerobic and anaerobic methane oxidizers at the Haakon Mosby Mud Volcano, Barents Sea. *Appl. Environ. Microbiol.* 73, 3348–3362. doi: 10.1128/AEM.00016-07

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, A., et al. (2004). ARB: a software environment for sequence data. *Nucleic Acids Res.* 32, 1363–1371. doi: 10.1093/nar/gkh293

MacGregor, B. J., Biddle, J. F., Siebert, J. R., Staunton, E., Hegg, E., Matthysse, A. G., et al. (2013). Why orange Guaymas Basin *Beggiatoa* spp. are orange: single-filament genome-enabled identification of an abundant octaheme cytochrome with hydroxylamine oxidase, hydrazine oxidase and nitrite reductase activities. *Appl. Environ. Microbiol.* 79, 1183–1190. doi: 10.1128/AEM.02538-12

Magen, C., Lapham, L. L., Pohlman, J. W., Marshall, K., Bosman, S., Casso, M., et al. (2014). A simple headspace equilibration method for measuring dissolved methane. *Limnol. Oceanogr. Methods* 12, 637–650. doi: 10.4319/lom.2014.12.637

McKay, L., Klokman, V., Mendlovitz, H., LaRowe, D., Zabel, M., Hoer, D., et al. (2016). Thermal and geochemical influences on microbial biogeography in the hydrothermal sediments of Guaymas Basin, Gulf of California. *Environ. Microbiol. Rep.* 8, 150–161. doi: 10.1111/1758-2229.12365

McKay, L. J., MacGregor, B. J., Biddle, J. F., Mendlovitz, H. P., Hoer, D., Lipp, J. S., et al. (2012). Spatial heterogeneity and underlying geochemistry of phylogenetically diverse orange and white *Beggiatoa* mats in Guaymas Basin hydrothermal sediments. *Deep-Sea Res. I* 67, 21–31. doi: 10.1016/j.dsr.2012.04.011

Meyer, S., Wegener, G., Lloyd, K. G., Teske, A., Boetius, A., and Ramette, A. (2013). Microbial habitat connectivity across spatial scales and hydrothermal temperature gradients at Guaymas Basin. *Front. Microbiol.* 4:207. doi: 10.3389/fmic.2013.00207

Mussmann, M., Pjevac, P., Krüger, K., and Dyksma, S. (2017). Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. *ISME J.* 11, 1276–1281. doi: 10.1038/ismej.2016.185

Niemann, H., Lösekann, T., de Beer, D., Elvert, M., Nadalig, T., Knittel, K., et al. (2006). Novel microbial communities of the Haakon Mosby mud volcano and their role as a methane sink. *Nature* 443, 854–858. doi: 10.1038/nature05227

Nobu, M. K., Dodsworth, J. A., Murugapiran, S. K., Rinke, C., Gies, E. A., Webster, G., et al. (2016a). Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *ISME J.* 10, 273–286. doi: 10.1038/ismej.2015.97

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

Núñez-Useche, F., Canet, C., Liebetrau, V., Puig, T. P., Ponciano, A. C., Alfonso, P., et al. (2018). Redox conditions and authigenic mineralization related to cold seeps in central Guaymas Basin, Gulf of California. *Mar. Petrol. Geol.* 95, 1–15. doi: 10.1016/j.marpetgeo.2018.04.010

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2012). vegan: Community Ecology Package. Available at: http://cran.r-project.org/package=vegan (Accessed December 30, 2020).

Ondréas, H., Scalabrin, C., Fouquet, Y., and Godfroy, A. (2018). Recent high-resolution mapping of Guaymas hydrothermal fields (Southern Trough). *BSGF – Earth Sciences Bulletin* 189:6. doi: 10.1051/bsgf/2018005

Paull, C. K., Ussler, W. III, Peltzer, E. T., Brewer, P. G., Keaten, R., Mitts, P. J., et al. (2007). Authigenic carbon entombed in methane-soaked sediments from the northeastern transform margin of the Guaymas Basin, Gulf of California. *Deep-Sea Res. II* 54, 1240–1267. doi: 10.1016/j.dsr2.2007.04.009

Price, R. E., and Giovanelli, D. (2017). "A review on the geochemistry and microbiology of marine shallow-water hydrothermal vents" in *Reference Module in Earth Systems and Environmental Sciences*. Elsevier.

Prüsse, E., Peplies, J., and Glöckner, F. O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 28, 1823–1829. doi: 10.1093/bioinformatics/bts252

Qin, W., Heal, K. R., Ramdasi, R., Kobelt, J. N., Martens-Habbena, W., Bertagnolli, A. D., et al. (2017). *Nitrosopumilus maritimus* gen. Nov., sp. nov., *Nitrosopumilus cobalaminigenes* sp. nov., *Nitrosopumilus oxyclinae* sp. nov., and *Nitrosopumilus ureiphilus* sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. *Int. J. Syst. Evol. Microbiol.* 67, 5067–5079. doi: 10.1099/ijsem.0.002416

Quast, C., Prüsse, 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

Ramírez, G. A., McKay, L. J., Fields, M. W., Buckley, A., Mortera, C., Hensen, C., et al. (2020). The Guaymas Basin subseafloor sedimentary Archaeome reflects complex environmental histories. *iScience* 23:101459. doi: 10.1016/j.isci.2020.101459

Roberts, D. W. (2012). labdsv: Ordination and Multivariate Analysis for Ecology [Internet]. Available at: http://cran.r-project.org/package=labdsv (Accessed December 30, 2020).

Rossel, P. E., Elvert, M., Ramette, A., Boetius, A., and Hinrichs, K. -U. (2011). Factors controlling the distribution of anaerobic methanotrophic communities in marine environments: evidence from intact polar membrane lipids. *Geochim. Cosmochim. Acta* 75, 164–184. doi: 10.1016/j.gca.2010.09.031

Ruff, S. E., Arnds, J., Knittel, K., Amann, R., Wegener, G., et al. (2013). Microbial communities of deep-sea methane seeps at Hikurangi Continental Margin (New Zealand). *PLoS One* 8:e72627. doi: 10.1371/journal.pone.0072627

Ruff, S. E., Biddle, J. F., Teske, A., Knittel, K., Boetius, A., and Ramette, A. (2015). Global dispersion and local diversification of the methane seep microbiome. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4015–4020. doi: 10.1073/pnas.1421865112

Ruff, S. E., Felden, J., Gruber-Vodicka, H. R., Marcon, Y., Knittel, K., Ramette, A., et al. (2019). *In situ* development of a methanotropic microbiome in deep-sea sediments. *ISME J.* 13, 197–213. doi: 10.1038/s41396-018-0263-1

Schoell, M. (1982). "Stable isotopic analyses of interstitial gases in quaternary sediments from the Gulf of California" in *Initial reports of the Deep Sea drilling project. Vol. 64.* eds. J. Curray and D. Moore (Texas, TX: Ocean Drilling Program, College Station), 815–817.

Schouten, S., Wakeham, S. G., Hopmans, E. C., and Sinninghe Damste, J. S. (2003). Biogeochemical evidence that thermophilic archaea mediate the anaerobic oxidation of methane. *Appl. Environ. Microbiol.* 69, 1680–1686. doi: 10.1128/aem.69.3.1680-1686.2003

Schreiber, L., Holler, T., Knittel, K., Meyerdierks, A., and Amann, R. (2010). Identification of the dominant sulfate-reducing bacterial partner of anaerobic methanotrophs of the ANME-2c clade. *Environ. Microbiol.* 12, 2327–2340. doi: 10.1111/j.1462-2920.2010.02275.x

Schutte, C., Teske, A., MacGregor, B., Salman-Carvalho, V., Lavik, G., Hach, P., et al. (2018). Filamentous giant *Beggiatoaceae* from Guaymas Basin are capable of both denitrification and dissimilatory nitrate reduction to ammonium (DNRA). *Appl. Environ. Microbiol.* 84:e02860-17. doi: 10.1128/AEM.02860-17

Seeberg-Elverfeldt, J., Schlüter, M., Feseker, T., and Kölling, M. (2005). Rhizon sampling of porewaters near the sediment-water interface of aquatic systems. *Limnol. Oceanogr. Methods* 3, 361–371. doi: 10.4319/lom.2005.3.361

Seitz, K. W., Dombrowski, N., Eme, L., Spang, A., Lombard, J., Sieber, J., et al. (2019). Asgard Archaea capable of anaerobic hydrocarbon cycling. *Nature Commun.* 10:1822. doi: 10.1038/s41467-019-09364-x

Simoneit, B. R. T., Lonsdale, P., Edmond, J. M., and Shanks, W. C. III (1990). Deep-water hydrocarbon seeps in Guaymas Basin, Gulf of California. *Appl. Geochem.* 5, 41–49.

Simoneit, B. R. T., Summerhayes, C. P., and Meyers, P. A. (1986). Sources and hydrothermal alteration of organic matter in quaternary sediments: a synthesis of studies from the central Gulf of California. *Mar. Petrol. Geol.* 3, 282–297.

Starnawski, P., Bataillon, T., Ettema, T. J. G., Jochum, L. M., Schreiber, L., Chen, X., et al. (2017). Microbial community assembly and evolution in subseafloor sediment. *Proc. Natl. Acad. Sci. U. S. A.* 114, 2940–2945. doi: 10.1073/pnas.1614190114

Stieglmeier, M., Klingl, A., Alves, R. J., Rittmann, S. K., Melcher, M., Leisch, N., et al. (2014). *Nitrososphaera viennensis* gen. nov., sp. nov., an aerobic and mesophilic, ammonia-oxidizing archaeon from soil and a member of the archaeal phylum Thaumarchaeota. *Int. J. Syst. Evol. Microbiol.* 64, 2738–2752. doi: 10.1099/ijs.0.063172-0

Suess, E. (2010). "Marine cold seeps" in Handbook of hydrocarbon and lipid microbiology. ed. K. N. Timmis (Berlin, Heidelberg: Springer-Verlag), 188–302.

Takahashi, S., Tomita, J., Nishioka, K., Hisada, T., and Nishijima, M. (2014). Development of a prokaryotic universal primer for simultaneous analysis of bacteria and archaea using next-generation sequencing. *PLoS One* 9:e105592. doi: 10.1371/journal.pone.0105592

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092

Tavormina, P. L., Hatzenpichler, R., McGlynn, S., Chadwick, G., Dawson, K. S., Connon, S. A., et al. (2015). *Methyloprofundus sedimentii* gen. nov. sp. nov., an obligate methanotroph from ocean sediment belonging to the "deep sea-1" clade of marine methanotrophs. *Int. J. Syst. Evol. Microbiol.* 65, 251–259. doi: 10.1099/ijs.0.062927-0

Teske, A. (2019). "Hydrocarbon-degrading anaerobic microbial communities in natural oil seeps" in *Microbial communities utilizing hydrocarbons and lipids: Members, metagenomics and ecophysiology, handbook of hydrocarbon and lipid microbiology*. ed. T. J. McGenity (Cham: Springer).

Teske, A., de Beer, D., McKay, L., Tivey, M. K., Biddle, J. F., Hoer, D., et al. (2016). The Guaymas Basin hiking guide to hydrothermal mounds, chimneys and microbial mats: complex seafloor expressions of subsurface hydrothermal circulation. *Front. Microbiol.* 7:75. doi: 10.3389/fmicb.2016.00075

Teske, A., Edgcomb, V., Rivers, A. R., Thompson, J. R., de Vera Gomez, A., Molyneaux, S. J., et al. (2009). A molecular and physiological survey of a diverse collection of hydrothermal vent *Thermococcus* and *Pyrococcus* isolates. *Extremophiles* 13, 905–915. doi: 10.1007/s00792-009-0278-7

Teske, A., Hinrichs, K. -U., Edgcomb, V., de Vera Gomez, A., Kysela, D., Sylva, S. P., et al. (2002). Microbial diversity in hydrothermal sediments in the Guaymas Basin: evidence for anaerobic methanotrophic communities. *Appl. Environ. Microbiol.* 68, 1994–2007. doi: 10.1128/aem.68.4.1994-2007.2002

Teske, A., Lizarralde, D., and Höfig, T. W. (2018). *Expedition 385 scientific prospectus: Guaymas Basin tectonics and biosphere*. College Station, TX: International Ocean Discovery Program.

Teske, A., McKay, L. J., Ravelo, A. C., Aiello, I., Mortera, C., Núñez-Useche, F., et al. (2019). Characteristics and evolution of sill-driven off-axis hydrothermalism in Guaymas Basin – the Ringvent site. *Sci. Rep.* 9:13847. doi: 10.1038/s41598-019-50200-5

Teske, A., and Salman, V. (2014). "The family Beggiatoaceae. Chapter 6" in *The prokaryotes – Gammaproteobacteria. 4th Edn.* eds. E. Rosenberg, E. F. DeLong, F. Thompson, S. Lory and E. Stackebrandt (Berlin/Heidelberg: Springer-Verlag), 93–134.

Topçuoglu, B. D., Stewart, L. C., Morrison, H. G., Butterfield, D. A., Huber, J. A., and Holden, J. F. (2016). Hydrogen limitation and syntrophic growth among natural assemblages of thermophilic methanogens at deep-sea hydrothermal vents. *Front. Microbiol.* 7:1240. doi: 10.3389/fmicb.2016.01240

Vetriani, C., Jannasch, J. W., MacGregor, B. J., Stahl, D. A., and Reysenbach, A. -L. (1999). Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. *Appl. Environ. Microbiol.* 65, 4375–4384. doi: 10.1128/AEM.65.10.4375-4384.1999

Vigneron, A., L'Haridon, S., Godfroy, A., Roussel, E. G., Cragg, B. A., Parkes, R. J., et al. (2015). Evidence of active methanogen communities in shallow sediments of the Sonora margin cold seeps. *Appl. Environ. Microbiol.* 81, 3451–3459. doi: 10.1128/AEM.00147-15

Von Damm, K. L., Edmond, J. M., Measures, C. I., and Grant, B. (1985). Chemistry of submarine hydrothermal solutions at Guaymas Basin, Gulf of California. *Geochim. Cosmochim. Acta* 49, 2221–2237.

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

Weber, A., and Jørgensen, B. B. (2002). Bacterial sulfate reduction in hydrothermal sediments of the Guaymas Basin, Gulf of California, Mexico. *Deep Sea Res. I* 149, 827–841. doi: 10.1016/S0967-0637(01)00079-6

Wegener, G., Krukenberg, V., Riedel, D., Tegetmeyer, H. E., and Boetius, A. (2015). Intracellular wiring enables electron transfer between methanotrophic archaea and bacteria. *Nature* 526, 587–590. doi: 10.1038/nature15733

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis. 2nd Edn*. New York: Springer International Publishing.

Wickham, H. (2018). stringr: Simple, Consistent Wrappers for Common String Operations. Available at: https://www.rdocumentation.org/packages/stringr (Accessed December 30, 2020).

Wickham, H., François, R., Henry, L., and Müller, K. (2018). dplyr: A Grammar of Data Manipulation. Available at: https://dplyr.tidyverse.org (Accessed December 30, 2020).

Winkel, M., de Beer, D., Lavik, G., Peplies, J., and Mussmann, M. (2014). Close association of active nitrifyers with *Beggiatoa* mats covering deep-sea hydrothermal sediments. *Environ. Microbiol.* 16, 1612–1626. doi: 10.1111/1462-2920.12316

Zaremba-Niedzwiedzka, K., Caceres, E. F., Saw, J. H., Bäckström, D., Juzokaite, L., Vancaester, E., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358. doi: 10.1038/nature21031

Zhou, Z., Liu, Y., Lloyd, K. G., Pan, J., Yang, Y., Gu, J. -D., et al. (2019). Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan Thermoprofundales (MBG-D archaea). *ISME J.* 13, 885–901. doi: 10.1038/s41396-018-0321-8

Zhuang, G. -C., Montgomery, A., Samarkin, V. A., Song, M., Liu, J., Schubotz, F., et al. (2019). Generation and utilization of volatile fatty acids and alcohols in hydrothermally altered sediments in the Guaymas Basin, Gulf of California. *Geophys. Res. Lett.* 46, 2637–2646. doi: 10.1029/2018GL081284

Zillig, W., Gierl, A., Schreiber, G., Wunderl, S., Janekovic, D., Stetter, K. O., et al. (1983). The archaebacterium *Thermofilum pendens* represents a novel genus of the thermophilic, anaerobic sulfur-respiring Thermoproteales. *Syst. Appl. Microbiol.* 4, 79–87. doi: 10.1016/S0723-2020(83)80035-6

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership