Genome- and Community-Level Interaction Insights into Carbon Utilization and Element Cycling Functions of *Hydrothermarchaeota* in Hydrothermal Sediment

Zhichao Zhou,[a,b] Yang Liu,[a] Wei Xu,[c] Jie Pan,[a] Zhu-Hua Luo,[c] Meng Li[a]

[a]Institute for Advanced Study, Shenzhen University, Shenzhen, People's Republic of China
[b]Department of Bacteriology, University of Wisconsin—Madison, Madison, Wisconsin, USA
[c]Key Laboratory of Marine Biogenetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen, People's Republic of China

ABSTRACT Hydrothermal vents release reduced compounds and small organic carbon compounds into the surrounding seawater, providing essential substrates for microbial growth and bioenergy transformations. Despite the wide distribution of the marine benthic group E archaea (referred to as *Hydrothermarchaeota*) in the hydrothermal environment, little is known about their genomic repertoires and biogeochemical significance. Here, we studied four highly complete (>80%) metagenome-assembled genomes (MAGs) from a black smoker chimney and the surrounding sulfur-rich sediments on the South Atlantic Mid-Ocean Ridge and publicly available data sets (the Integrated Microbial Genomes system of the U.S. Department of Energy-Joint Genome Institute and NCBI SRA data sets). Genomic analysis suggested a wide carbon metabolic diversity of *Hydrothermarchaeota* members, including the utilization of proteins, lactate, and acetate; the anaerobic degradation of aromatics; the oxidation of C_1 compounds (CO, formate, and formaldehyde); the utilization of methyl compounds; CO_2 incorporation by the tetrahydromethanopterin-based Wood-Ljungdahl pathway; and participation in the type III ribulose-1,5-bisphosphate carboxylase/oxygenase-based Calvin-Benson-Bassham cycle. These microbes also potentially oxidize sulfur, arsenic, and hydrogen and engage in anaerobic respiration based on sulfate reduction and denitrification. Among the 140 MAGs reconstructed from the black smoker chimney microbial community (including *Hydrothermarchaeota* MAGs), community-level metabolic predictions suggested a redundancy of carbon utilization and element cycling functions and interactive syntrophic and sequential utilization of substrates. These processes might make various carbon and energy sources widely accessible to the microorganisms. Further, the analysis suggested that *Hydrothermarchaeota* members contained important functional components obtained from the community via lateral gene transfer, becoming a distinctive clade. This might serve as a niche-adaptive strategy for metabolizing heavy metals, C_1 compounds, and reduced sulfur compounds. Collectively, the analysis provides comprehensive metabolic insights into the *Hydrothermarchaeota*.

IMPORTANCE This study provides comprehensive metabolic insights into the *Hydrothermarchaeota* from comparative genomics, evolution, and community-level perspectives. Members of the *Hydrothermarchaeota* synergistically participate in a wide range of carbon-utilizing and element cycling processes with other microorganisms in the community. We expand the current understanding of community interactions within the hydrothermal sediment and chimney, suggesting that microbial interactions based on sequential substrate metabolism are essential to nutrient and element cycling.

KEYWORDS *Hydrothermarchaeota*, hydrothermal sediment, comparative genomics, carbon utilization, element cycling, community-level interaction, lateral gene transfer

Hydrothermal alteration and venting processes transfer reduced sulfur compounds, organic compounds (e.g., C_1 compounds, petroleum compounds, organic acids), and heavy metals to the surrounding hydrothermal sediments (1–6). Together with deposited sedimentary carbon compounds, these substrates make hydrothermal sediments a distinct ecological niche containing a variety of organic and inorganic nutrients. In hydrothermally active Guaymas Basin sediments, microorganisms syntrophically degrade hydrocarbons and lipids. Indeed, metabolic linkages among microbial groups, such as a substrate-level interdependency between fermentative members and sulfur- and nitrogen-cycling members, were proposed (6). However, the diversity and function of microorganisms that inhabit the hydrothermal environment, especially archaea, remain elusive, and the community-level microbial interactions within this environment have not yet been characterized in detail.

"*Candidatus* Hydrothermarchaeota," originally identified on the continental slope and abyssal sediments by 16S rRNA gene profiling and termed marine benthic group E (MBG-E) (7), has recently been proposed as a new archaeal phylum (8). According to a previous study, *Hydrothermarchaeota* is abundant in the deep-sea hydrothermal environment, such as the crustal fluids of the Juan de Fuca Ridge (JdFR) flank (9). A combined analysis of metagenome-assembled genomes (MAGs) and single-cell-amplified genomes of *Hydrothermarchaeota* from the crustal fluids of the Juan de Fuca Ridge flank was used to evaluate the evolutionary placement and functional potential of this new archaeal phylum (8). The results of that analysis were further consistent with those of a more recent study (10), with both studies suggesting the potential of *Hydrothermarchaeota* for carboxydotrophy and sulfate and nitrate reduction (8, 10). However, the relatively small number of genomes available to date limits the ability to understand the ecological roles and metabolism of this novel archaeal lineage.

Here, we analyzed the metagenomes from sulfur-rich hydrothermal sediments at an active deep-sea (water depth, 2,770 m) hydrothermal vent site (black smoker chimney layers and the surrounding sulfur-rich sediments) in the southern Mid-Atlantic Ridge of the South Atlantic Ocean (total S content = approximately 100 to 450 mg/g; detailed sample information is provided in Text S1 in the supplemental material). We generated two metagenomic libraries from the layer (the TVG10 library) and surrounding sediments (the TVG13 library) of an active black smoker chimney in the Mid-Atlantic Ridge (BSmoChi-MAR) of the South Atlantic Ocean (38.1 Gb for TVG10 and 30.3 Gb for TVG13). *De novo* metagenome assembly and binning resulted in 140 MAGs (>50% genome completeness) from 24 microbial groups (Table S1), including 5 archaeal MAGs and 135 bacterial MAGs. The metabolic prediction from the resolved MAGs revealed the functional redundancy and syntrophic substrate-utilizing interactions among the microorganisms. Further, based on the four *Hydrothermarchaeota* genomes of relatively high completeness (>80%) obtained in the current study, the results of a previous publication (9), and data in publicly available data sets (Table S2), we propose a metabolic scheme for this archaeal lineage. Finally, the evolutionary analysis suggested the important role of lateral gene transfer (LGT) in the niche adaptation of *Hydrothermarchaeota* to the local environment. The current study provides an in-depth insight into the genomics, community-level interactions, and evolution of *Hydrothermarchaeota*.

RESULTS AND DISCUSSION

Hydrothermarchaeota is a novel archaeal phylum. Reconstructed archaeal MAGs and scaffolds containing phylogenetically informative genes, including those for at least three ribosomal proteins (RPs) or 16S rRNA gene fragments, are summarized in Table 1 and in Table S3 in the supplemental material. Both the 16S rRNA and RP gene phylogenies placed *Hydrothermarchaeota* as a distinct lineage parallel to other *Euryarchaeota* clades, including *Thermococci*, *Methanomicrobia*, and *Hadesarchaea* (Fig. 1a and b and Fig. S1a). Within this lineage, the 16S rRNA gene sequences show median sequence identities of 80.8 to 83.9%, which supports phylum-level diversity (Fig. 1b and Text S1). The phylum name *Hydrothermarchaeota* was originally proposed for this archaeon since all the current known genomes were obtained from hydrothermal

TABLE 1 Overview of genomic statistics for archaeal MAGs constructed in the current study and reference and NCBI SRA deposits[a]

Characteristic	Value for the following *Hydrothermarchaeota* MAG:						
	SZUA-158 (TVG10)	SZUA-236 (TVG13)	SZUA-237 (TVG13)	JdFR-16	JdFR-17	JdFR-18	HyVt-292
No. of copies of individual markers							
0	9	21	49	99	87	2	30
1	139	128	136	25	59	145	141
2	1	0	3	25	29	2	17
3	0	0	0	0	12	0	0
4	0	0	0	0	1	0	0
5+	0	0	0	0	0	0	0
Completeness (%)	92.53	84.97	65.82	31.93	53.87	98.13	80.43
Contamination (%)	0.93	0.00	1.76	13.55	25.15	1.87	9.06
Strain heterogeneity (%)	0.00	0.00	33.33	84.00	71.83	0.00	35.29
Bin size (bp)	1,749,358	1,270,198	1,228,833	1,353,114	2,178,134	2,062,134	1,426,722
N_{50} (bp)	9,590	6,219	3,825	6,267	7,687	149,032	5,600
Mean scaffold length (bp)	7,706.42	5,747.50	3,510.95	5,614.58	6,331.78	93,733.36	4,988.54
No. of scaffolds	227	221	350	241	344	22	286
GC (%)	38.99	47.58	45.76	49.68	49.64	39.14	40.02
GC SD (%)	1.91	1.99	2.89	3.92	4.50	1.54	1.98
Coding density (%)	84.96	91.48	84.96	91.98	91.94	92.12	89.88

[a]HyVt-292 was reconstructed from the metagenome of deep-sea massive sulfide deposits from the Southern Mariana Trough (NCBI accession number DRR093004), which is hydrothermally inactive. JdFR MAGs were reconstructed from the data for subsurface fluids of the Juan de Fuca Ridge flank.

sediments or fluids (7–10). The 16S rRNA gene sequence data indicated that members of the *Hydrothermarchaeota* are also widely distributed in wetland, estuarine, marine, and hot spring sediments (10) (Fig. 1b). At the community level, TVG13 covered *Hydrothermarchaeota* 16S rRNA genes more diverse than those covered by TVG10 (Fig. 1c). Meanwhile, from the allele frequency analysis, a higher frequency indicates the dominance of the majority allele over minor alleles. The higher mean major allele frequency on the genome level indicates a less diverse *Hydrothermarchaeota* population in TVG10 than in TVG13 (Fig. 1d).

Mixotrophic lifestyle and versatile substrate utilization by *Hydrothermarchaeota*. We selected four representative *Hydrothermarchaeota* MAGs (SZUA-158 [TVG10], SZUA-236 [TVG13], JdFR-18, and HyVt-292) with relatively high completeness values (>80%) from the major three clades for metabolic prediction analysis (Table 1, Fig. 2a, Fig. S1b, and Tables S4 and S5). Similar to the findings of a previous study (8), almost all *Hydrothermarchaeota* MAGs contained the tetrahydromethanopterin (THMPT)-based Wood-Ljungdahl (WL) pathway (the THMPT-WL pathway) and some components of the tetrahydrofolate (THF)-based Wood-Ljungdahl pathway (the THF-WL pathway) (Fig. 2a). Since none of these MAGs contained the complete genes for the THF-WL pathway, it might well be that this pathway is not active in *Hydrothermarchaeota* (Fig. 2a). The THMPT-WL pathway in *Hydrothermarchaeota* might function in both directions, either reductively incorporating CO_2 into acetyl coenzyme A (CoA) synthesis or oxidatively converting products from central carbon metabolism (peptide and sugar carbohydrate degradation) to be channeled into energy-producing pathways. If the former direction is operational, *Hydrothermarchaeota* probably lead a mixotrophic lifestyle, using both inorganic and organic carbon sources. *Hydrothermarchaeota* lacked the methyl coenzyme M reductase (MCR) for methane metabolism. However, JdFR-18 might be able to incorporate a variety of methyl-containing compounds into the WL pathway, including mono-, di-, and trimethylamines and methanol. This feature is frequently observed in members of many other archaeal taxa, such as *Methanosarcinales*, *Methanomassiliicoccales*, *Methanofastidiosa*, *Bathyarchaeota*, and *Verstraetearchaeota* (11). In particular, JdFR-18 (clade 1) encoded HdrD (three copies) and GlcD (four copies; an FAD-containing dehydrogenase and a putative D-lactate dehydrogenase), with one pair of these two genes being colocated and responsible for heterodisulfide reduction linked to lactate utilization. This gene arrangement and
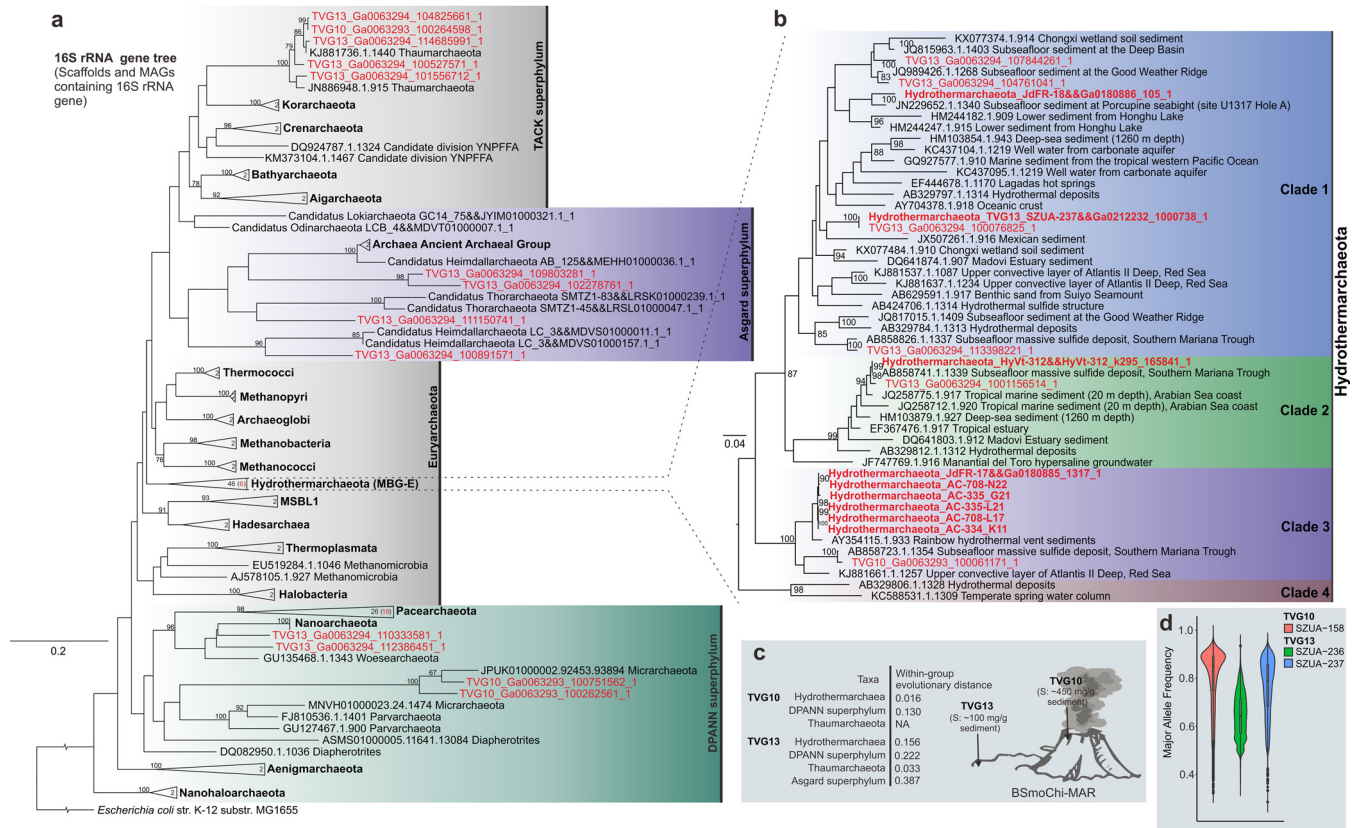
FIG 1 Phylogenetic tree of the 16S rRNA genes from TVG metagenomes. (a) RAxML maximum likelihood tree, constructed by including the metagenome 16S rRNA gene sequences (from both binned and unbinned scaffolds). The sequences obtained in the current study are colored in red; the number in red in parentheses indicates the sequence number from the current study. Bootstrap values over 75% are labeled. The tree was rooted by the *Escherichia coli* K-12 16S rRNA gene. (b) Detailed *Hydrothermarchaeota* 16S rRNA gene tree. Four clades of the representative sequences of the *Hydrothermarchaeota* 16S rRNA gene were established based on a 90% similarity cutoff. Sequences retrieved from reconstructed MAGs are highlighted in red. (c) Schematic depicting the two sampling locations and the within-group evolutionary distances of the 16S rRNA gene sequences from two samples (calculated by using the Jukes-Cantor model and pairwise comparisons based on global alignment). NA, not available because only one sequence was present within a group and a within-group evolutionary distance could not be calculated. (d) Diagram of the major allele frequency representing the diverse level of potential *Hydrothermarchaeota* in the corresponding environments where the *Hydrothermarchaeota* MAGs were identified.

lactate utilization have also been observed in the *Archaeoglobus fulgidus*, *Bathyarchaeota*, and *Verstraetearchaeota* genomes (11–13).

Hydrothermarchaeota genomes probably encode the full tricarboxylic acid (TCA) cycle but not the beta-oxidation pathway (harboring only the acetyl-CoA *C*-acetyltransferase genes; Fig. 2a). Genes encoding benzoyl-CoA reductase subunits (BcrBC) were present in HyVt-292 and JdFR-18, which indicates ATP-dependent benzoyl-CoA degradation. Benzoyl-CoA is the central intermediate of anaerobic degradation pathways of many aromatics, including benzene, phenol, 4-OH-benzoate, cresols, phenylacetate, ethyl-benzene, and others (14). Three out of four analyzed *Hydrothermarchaeota* MAGs encoded vanillate/4-hydroxybenzoate decarboxylase subunit C (BsdC) and flavin prenyltransferase (UbiX), which further supports their phenol-degrading activity. The existence of genes for both ADP-forming acetyl-CoA synthetase (Acd; EC 6.2.1.13) and acetyl-CoA synthetase (Acs; EC 6.2.1.1) suggests the possibility of both acetate fermentation and acetogenesis. In addition, further fermentation of acetate to ethanol was evident in some *Hydrothermarchaeota* MAGs, and genes for enzymes from acetaldehyde (by aldehyde ferredoxin oxidoreductase [Aor]) to ethanol (by alcohol dehydrogenase [AdhP or AdhE]) were also identified (Fig. 2a).

C₁ oxidation and other element cycling capacities. *Hydrothermarchaeota* might anaerobically oxidize CO for ferredoxin generation, considering the existence of the gene encoding the carbon monoxide dehydrogenase catalytic subunit (CooS) (clade 3)
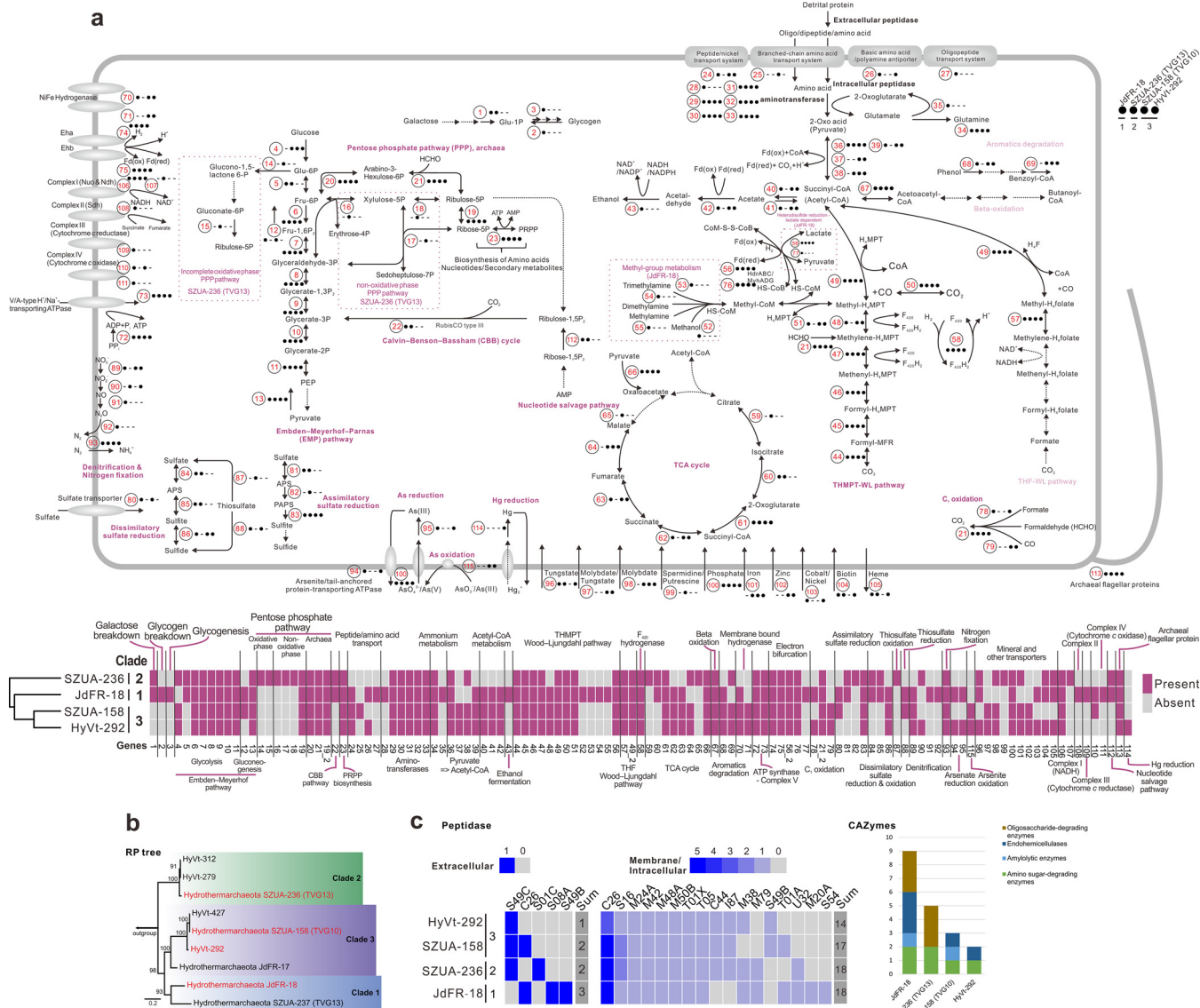
FIG 2 Metabolic pathways in *Hydrothermarchaeota*. (a) Schematic showing the metabolic pathways of the four *Hydrothermarchaeota* MAGs with highly complete genomes. red, reduction; ox, oxidation. (b) RP tree showing the phylogeny of all currently available *Hydrothermarchaeota* MAGs. (c) Heatmap and bar chart depicting the distribution of extracellular/intracellular peptidases and carbohydrate-active enzymes (CAZys) in *Hydrothermarchaeota* MAGs. Information on the glycoside hydrolase (GH) family within each functional category is summarized in the bar chart.

(8, 15). Both JdFR-18 and HyVt-292 harbored the *fdsABG* operon for NAD$^+$-dependent formate dehydrogenase, indicating that *Hydrothermarchaeota* can use formate as a bioenergy source for electron-transferring phosphorylation (16). The fused 3-hexulose-6-phosphate synthase/6-phospho-3-hexuloisomerase (Hps-Phi) (EC 4.1.2.43) and bifunctional enzyme Fae-Hps (EC 4.2.1.147) are responsible for formaldehyde fixation in the ribulose monophosphate cycle (part of the pentose phosphate pathway [PPP]) and the generation of methylene-tetrahydromethanopterin (methylene-H$_4$MPT) (the THMPT-WL pathway) (3), which can be further used for the biosynthesis of ribose and acetyl-CoA, respectively. C$_1$ compounds of various redox states are common (CO and formate) or potentially available (formaldehyde) and are generated through geochemical reactions in the hydrothermal environments (1–4). Further, some anaerobes generate CO (4). Combined with the CO$_2$ fixation ability, which is mainly the Calvin-Benson-Bassham (CBB) cycle and WL pathway, the mixotrophic lifestyle possibly makes *Hydrothermarchaeota* more adaptive within the global benthic environmental setting (17).

In addition, the process of sulfur oxidation to sulfate might take place in *Hydrothermarchaeota* because the dissimilatory sulfite reductase (DsrAB) that they encode can also facilitate sulfur oxidation (8, 18). The existence of key genes for the Sox pathway in SZUA-236 (TVG13) suggests that these organisms have the ability to oxidize thiosulfate for energy yield (Fig. 2a). The potential denitrification and sulfate reduction pathways might enable *Hydrothermarchaeota* to scavenge diverse organic matter by anaerobic respiration. Presumably, *Hydrothermarchaeota* might couple nitrate reduction with the oxidation of reduced sulfur compound (S^0, S^{2-}, and $S_2O_3^{2-}$) as an energy-generating process (19). Since the hydrothermal environment is frequently enriched in heavy metals (20), *Hydrothermarchaeota* also possess genomic components for the detoxification of As(V) (arsenate reductase [ArsC] and arsenite/tail-anchored protein-transporting ATPase [ArsA]) and Hg(II) (mercuric reductase [MerA]). They might also oxidize As(III) (cytomembrane-bound arsenite oxidase subunits [AioB]) and presumably couple the reduction of As(V) with the oxidation of reduced sulfur compounds, which suggests that As cycling could constitute one of the energy metabolism pathways in these microorganisms.

Coexistence of nucleotide salvage pathway and CBB cycle. The genomes of almost all members of the *Hydrothermarchaeota* encoded the Embden-Meyerhof-Parnas (EMP) pathway for both the glycolysis and gluconeogenesis directions (almost all members encoded fructose-1,6-bisphosphatase [FBP] and phosphoenolpyruvate [PEP] synthase/pyruvate phosphate dikinase [PPDK]) (Fig. 2a). For the former, based on the genome data, the conversion of PEP to pyruvate (catalyzed by pyruvate kinase) was lacking; however, the reverse reactions of PEP synthase/PPDK have been reported in some thermophilic archaea, including *Thermococcus* (*Euryarchaeota*) and *Thermoproteus* (*Crenarchaeota*) (21). All *Hydrothermarchaeota* clades contained an archaeon-style pentose phosphate pathway (PPP) in the genome. In addition, the SZUA-236 (TVG13) genome encoded an incomplete oxidative PPP and a nonoxidative PPP. The PPP and the phosphoribosyl pyrophosphate (PRPP) synthesis pathway are anabolic pathways for the biosynthesis of a variety of amino acids, nucleotides, and other secondary metabolites that utilize substrates from glycolysis (22). They most likely fix CO_2 by type III ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) in the CBB cycle, based on genomic predictions (Fig. 2a). The lack of phosphoribulokinase (Prk) of the CBB cycle is frequently reported in archaeal genomes (23). Meanwhile, some reports based on metabolic experiments indicate the presence of autotrophic activity of the crenarchaeotal CBB cycle, despite the lack of Prk. This suggests that although Prk-encoding genes were not detected in the *Hydrothermarchaeota* genomes in the current study, these archaea might nonetheless fix CO_2 via the CBB cycle (24).

The presence of the nucleotide salvage pathway and CBB cycle suggests that *Hydrothermarchaeota* might recover the RNA/DNA degradation products (such as AMP) for glycolysis or cycle them back into PPP for biosynthesis (Fig. 2a) (25). In addition to RNA/DNA degradation, AMP could also originate from the activities of (i) AMP-forming adenylylsulfate reductase during sulfate reduction (clades 1 and 2), (ii) PRPP synthesis (all clades), and (iii) ADP-dependent (AMP-forming) phosphofructokinase/glucokinase (clade 3) during glycolysis (26–28). The genomic components of type III RuBisCO and the nucleotide salvage function have also been identified in many other euryarchaeotal groups, including *Archaeoglobi*, *Halobacteria*, *Thermococci*, *Hadesarchaea*, and euryarchaeotal methanogens (1). The unconventional participation in the type III RuBisCO-based CBB cycle in the nucleotide salvage function of the above-mentioned archaeal taxa suggests the importance of the primary function of type III RuBisCO in the early stages of archaeal evolution (1, 25).

Limited carbohydrate assimilation in protein/peptide degraders. No genes encoding potential sugar and carbohydrate transporters were identified in the three major clades of *Hydrothermarchaeota* in the current study (Fig. 2c). The genomes of these archaea encoded genes with limited functions in carbohydrate assimilation and transformation only for galactose degradation and glycogen conversion. The annota-
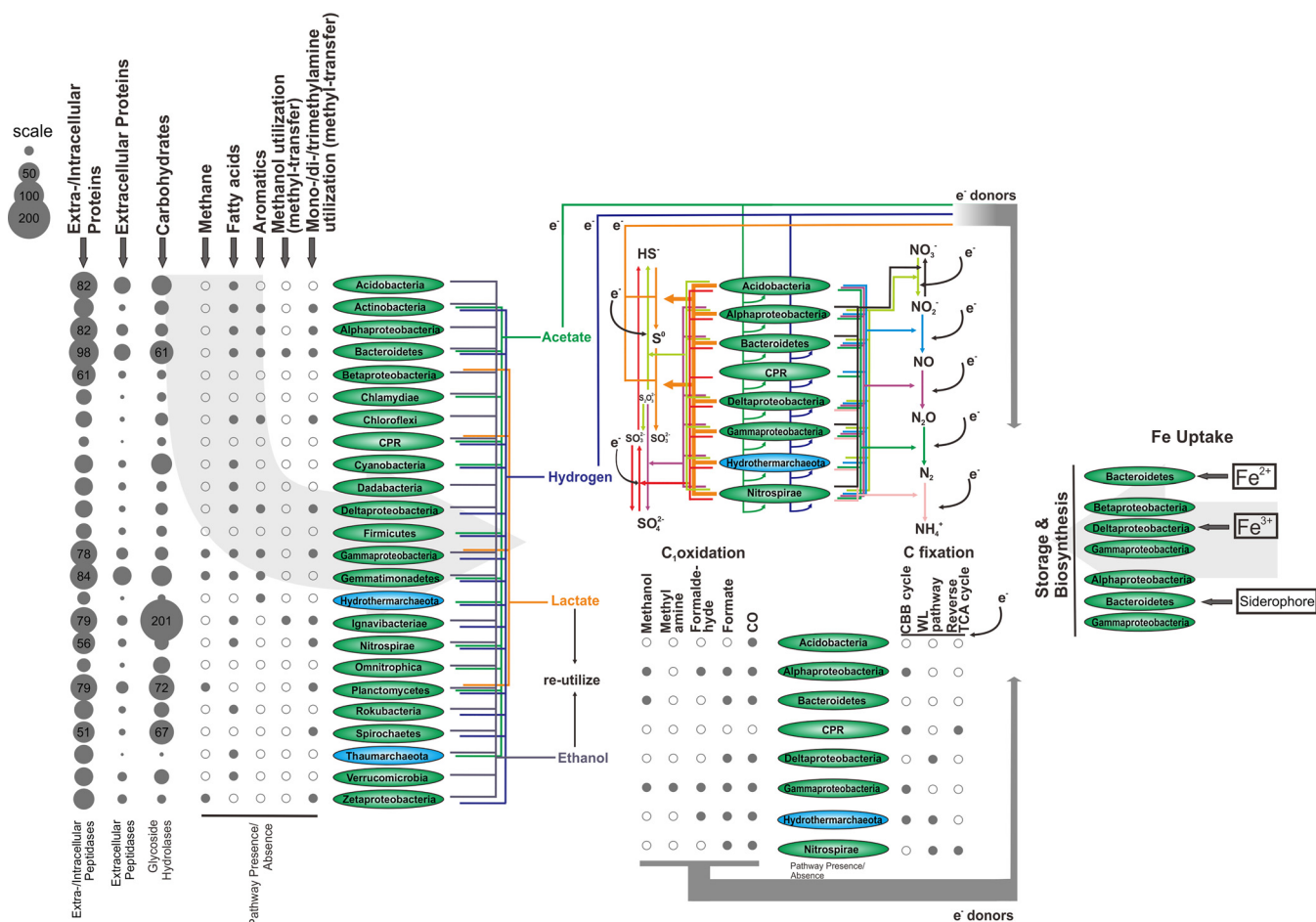
FIG 3 Metabolic prediction for the community of MAGs from TVG metagenomes. The abundance of peptidases and carbohydrate-degrading enzymes was calculated by considering the MAG completeness and taking the average for all MAGs within an individual microbial group (0 digits were used after the decimal point). The presence of a specific pathway/function within each microbial group was assigned when the pathway/function was present in any MAG within this microbial group. The Fe uptake pathway was predicted by use of the previously established database. For the metabolic prediction of N and S cycling, C_1 oxidation, and C fixation, only the major microbial groups (with at least one MAG from the group having over 15× genome coverage) are represented.

tion of carbohydrate-active enzymes (CAZys) also suggested a limited capacity for carbohydrate utilization. Among the peptidases encoded by the *Hydrothermarchaeota* genomes, serine-type endopeptidase S08A is the dominant extracellular peptidase, according to both metagenome and metatranscriptome data, in marine group I (MG-I), -II, and -III archaea from deep-sea hydrothermal plume (29); S49C is an archaeal signal peptide peptidase that destroys cleaved signal peptides; C26 is a gamma-glutamyl hydrolase that probably has glutamine amidotransferase activity (Tables S4 and S5). Further, genomic predictions in the current study indicated that while the genomes of *Hydrothermarchaeota* from clades 1 and 3 encode various peptide/amino acid transporters, the genomes of the three major clades encode the genes of six groups of aminotransferases to transfer amino residues (Tables S4 and S5) and the genes for pyruvate ferredoxin oxidoreductase (Por), indolepyruvate ferredoxin oxidoreductase (Ior), 2-oxoglutarate/2-oxoacid ferredoxin oxidoreductase (Kor), pyruvate dehydrogenase, and dihydrolipoamide dehydrogenase to assimilate 2-oxo acids (pyruvate) to succinyl-CoA (acetyl-CoA) and replenish the energy pool of reducing equivalents. The presence of the various peptide/amino acid transporters and endopeptidases/ aminotransferases encoded by the *Hydrothermarchaeota* genomes suggests that they can use detrital peptides/proteins as one of their main carbon and energy sources.

Functional redundancy and community-level interactions. We next analyzed the metabolic capacities of all the reconstructed 140 MAGs (Fig. 3 and Tables S1, S2, and S6)

to investigate microbial community interactions. Among these, *Acidobacteria*, *Bacteroidetes*, and *Gemmatimonadetes* possessed the greatest abundance of genes encoding extracellular peptidases. Presumably, they are the major players utilizing detrital proteins in marine sediments. Other microbial groups could engage in syntrophic interactions with the above-mentioned microbes to assimilate extracellular peptides/proteins using the extracellular peptidases secreted by these microbes. Furthermore, *Ignavibacteriae*, *Planctomycetes*, and *Spirochaetes* possessed the most genes encoding glycoside hydrolases, suggesting that they are the major players in carbohydrate/sugar utilization. In addition, a variety of microbial groups are predicted to be able to degrade/utilize methane, fatty acids, aromatics, methanol, and mono-, di-, and trimethylamines. The fermentation products probably include acetate, hydrogen, lactate, and ethanol. The electron pool generated by fermentation is transferred to terminal electron acceptors or CO_2 for either respiration or carbon fixation. Furthermore, the fermentation products from the first fermentation process could also be reutilized by the community members as energy and carbon sources.

The eight major microbial groups (with at least one MAG from each group with over $15\times$ genome coverage, including *Acidobacteria*, *Alphaproteobacteria*, *Bacteroidetes*, candidate phyla radiation [CPR], *Deltaproteobacteria*, *Gammaproteobacteria*, *Hydrothermarchaeota*, and *Nitrospira*) are predicted to possess multiple functions for sulfur cycling, including sulfide oxidation, sulfur oxidation, thiosulfate oxidation, sulfate reduction, and thiosulfate disproportionation. The oxidized sulfur compounds (SO_4^{2-}, SO_3^{2-}, and $S_2O_3^{2-}$), as well as nitrate/nitrite and molecular oxygen (except for CPR and *Hydrothermarchaeota*), could serve as the terminal electron acceptors for respiration involving organic or inorganic energy sources (Table S6a). This suggests that microorganisms in the chimney layers and surrounding sediments of BSmoChi-MAR have evolved various strategies to adapt to the microaerobic to anoxic environment. In addition to *Hydrothermarchaeota*, several other microbial groups are predicted to oxidize multiple C_1 compounds and have a carbon fixation capacity, as predicted by the genome contents. Further, some of these microbial groups can probably degrade carbohydrates and peptides/proteins and have sulfur cycling and denitrification abilities, such as the *Alphaproteobacteria*, *Deltaproteobacteria*, *Gammaproteobacteria*, and *Nitrospirae* (Fig. 3 and Table S6a). The inferred redundancy of carbon utilization and element cycling functions among microorganisms and the interactive processes of syntrophic and sequential utilization of substrates might make a wide range of substrates and energy sources available for the community.

Comparative genomics. We next selected representative genomes from the euryarchaeotal groups and *Hydrothermarchaeota* to compare their metabolic capacities. While a peptide degradation capacity was shared by most *Hydrothermarchaeota* and euryarchaeotal groups, other carbohydrate-degrading or -utilizing capacities, such as for starch/glycogen, aromatics, fatty acids, methanol, and mono-, di-, and trimethylamines, were distributed in a patchy manner (Fig. 4, Fig. S1c, and Table S6b). In comparison with the euryarchaeotal groups, the *Hydrothermarchaeota* genomes encoded multiple components for the cycling of N and S and the *Hydrothermarchaeota* could oxidize three important C_1 compounds (formate, formaldehyde, and carbon monoxide). At the within-phylum level, a distinct distribution of metabolic traits among the three major clades within the *Hydrothermarchaeota* was apparent (Fig. 4). Unlike other clades, clade 1 MAG (from the subsurface fluids of JdFR [SubFlu-JdFR]) might specifically utilize methanol, methanethiol, and mono-, di-, and trimethylamines. Although the chemical profile of the Juan de Fuca Ridge flank subsurface fluids is lacking (9), genomic evidence suggested the existence of methyl compounds in the surrounding environment. Clade 3 MAGs (SZUA-158 from the chimney layer of BSmoChi-MAR and HyVt-292 from sulfide deposits of the Southern Mariana Trough) were capable of sulfur oxidation (DsrAB) as the major sulfur-cycling function, which is probably attributed to the high levels of sulfide in the chimney layer and sulfide deposits. Presumably, clade 3 also mainly depended on sulfide oxidation for energy generation and exhibited
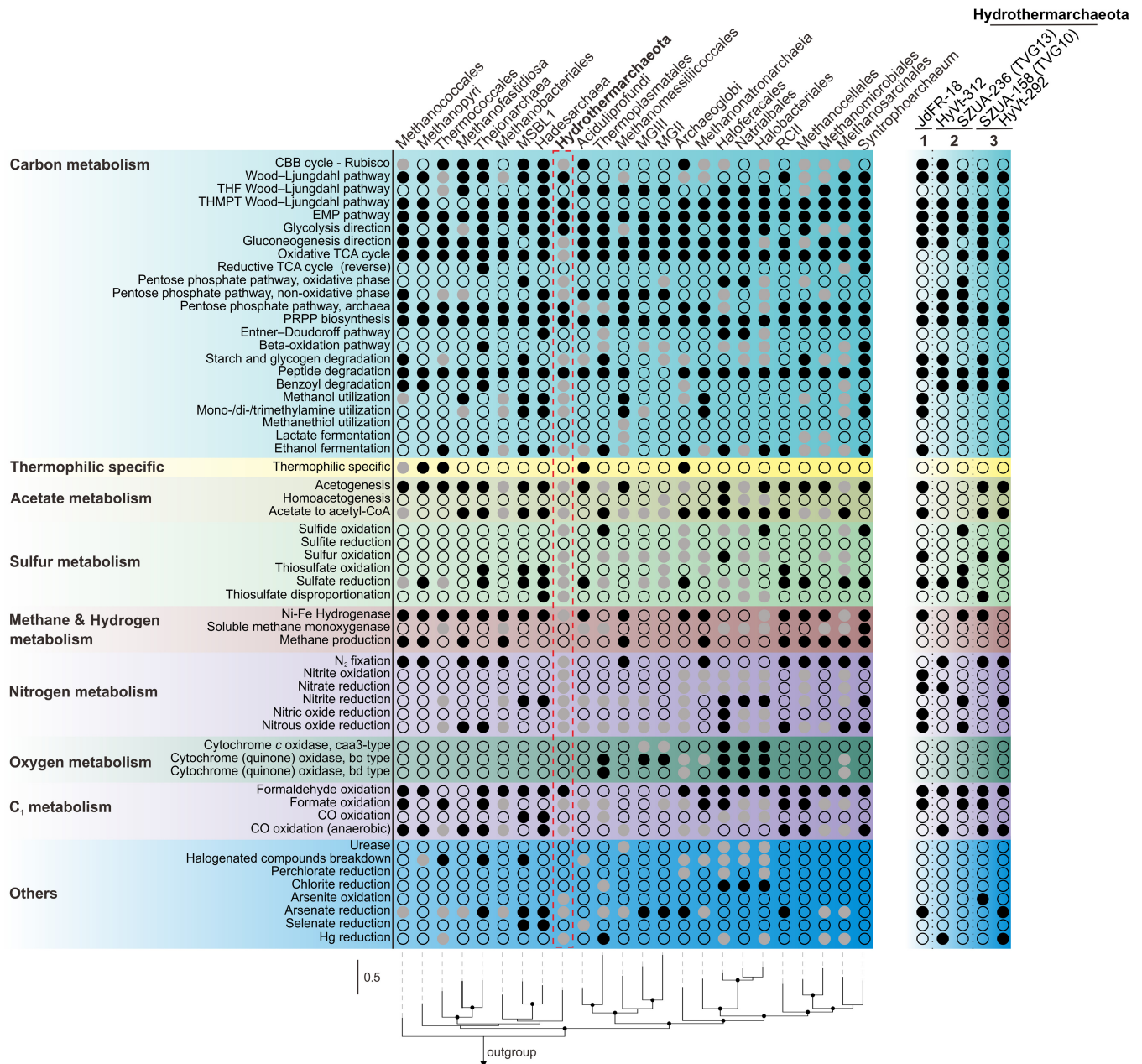
FIG 4 Comparison of the metabolic capacity of *Hydrothermarchaeota* and *Euryarchaeota*. Metabolic marker genes from the Pfam, TIGRfam, and KEGG databases were used to identify the corresponding sequences in up to five genomes within one archaeal group. Solid black dots, solid gray dots, and blank dots, present in all genomes, present in some genomes, and not present in any genome, respectively. If one marker gene was identified, it was assumed that the corresponding metabolic function exists. Because of the limited number of genomes and low genome completeness, solid black dots were used to denote gene identification in one genome for *Hadesarchaea*, *Hydrothermarchaeota*, *Theionarchaea*, *Syntrophoarchaeum*, and MSBL-1. A detailed summary of the information is provided in Text S1 in the supplemental material. The metabolic capacity of five *Hydrothermarchaeota* MAGs is also depicted. The RP phylogenetic tree was constructed by randomly selecting one genome from each group; bootstrap values over 75% are depicted as black dots on the node.

limited sugar carbohydrate degradation (Fig. 2c). Collectively, the comparative genomic analysis indicated that each clade possesses metabolic traits related to niche adaptation.

Clade-distinctive LGT. We also mapped the minimum parsimony-based prediction of gene gain and gene loss events of inferred gene orthologous groups (OGs; synonymous with gene families) to the RP-based phylogenomic tree (Fig. 5). Seven phylogenomically closely related euryarchaeotal classes or orders, including methanogens and nonmethanogens, were included. The key gene gain events at nodes 28, 26, and 24
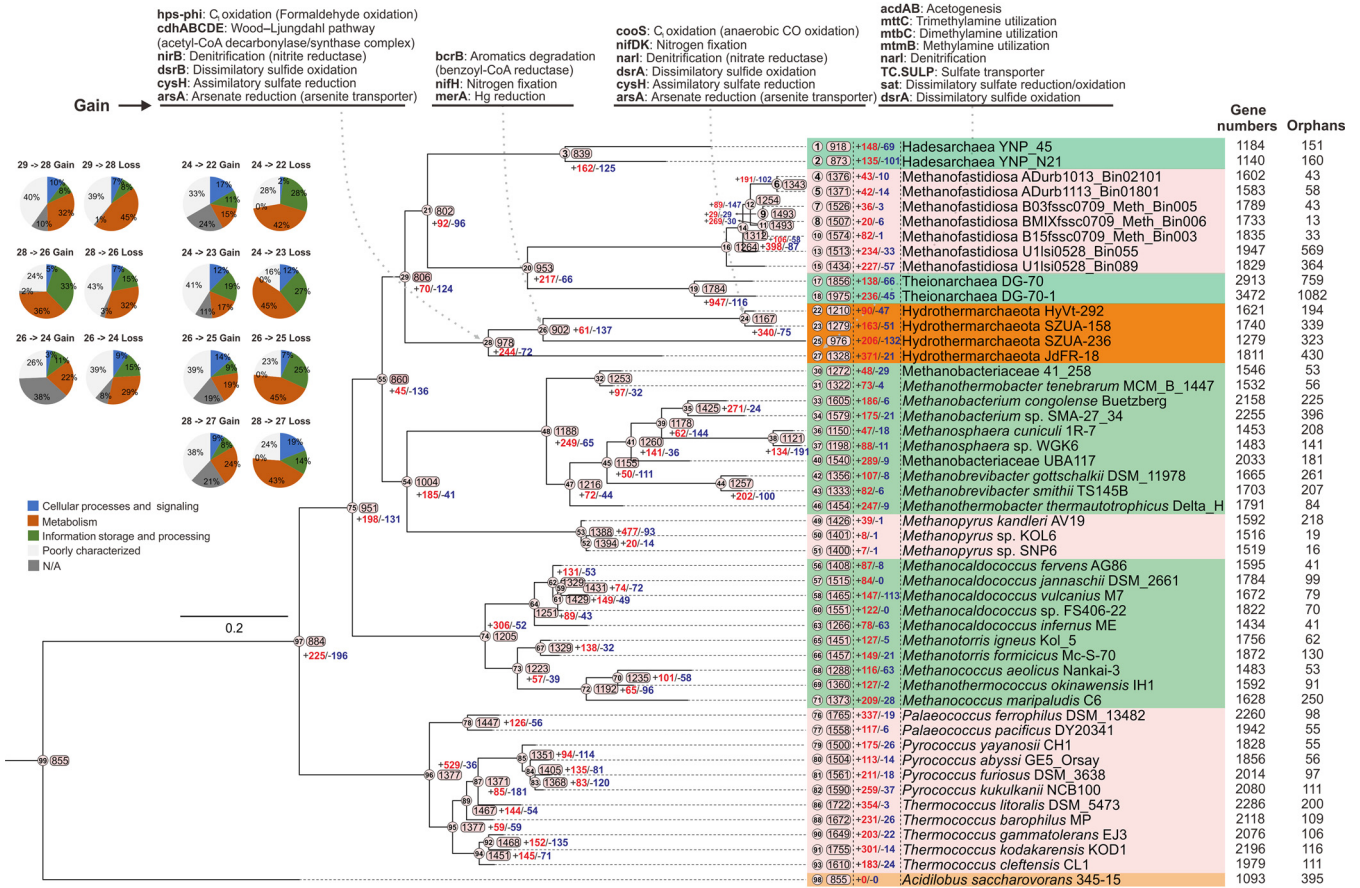
FIG 5 Estimation of orthologous group (OG) turnover events for *Hydrothermarchaeota* and related euryarchaeotal orders and classes. The numbers of OGs (numbers in rectangles) and inferred OG gain and loss numbers (red and blue numbers) are labeled accordingly on the tree nodes and tips. The numbers in circles indicate the order of nodes and tips. The COG category information of the OGs gained or lost for the *Hydrothermarchaeota* clade was parsed and is depicted. Important genes that were involved in the OG gain events in the *Hydrothermarchaeota* clade were also labeled at the corresponding nodes.

(occupying 24.9%, 6.7%, and 29.1% of the ancestral genomes, respectively) indicated that important traits of extant *Hydrothermarchaeota* have been derived from LGT. These represent many functions, such as C$_1$ oxidation of formaldehyde and carbon monoxide, CO$_2$ fixation (key components of the WL pathway), and acetyl-CoA synthesis, nitrogen and sulfur cycling, aromatic degradation, and Hg and As reduction. The gene gain events at node 27 (occupying 27.9% of the genome) probably enabled JdFR-18 to utilize tri-, di-, and monomethylamines, also allowing acetogenesis and nitrogen and sulfur cycling (Table S7a). The lost OGs at these nodes have functions mainly associated with amino acid transportation and metabolism, energy production and conversion, and transcription and translation (Fig. 6 and Table S7a). As indicated above, *Hydrothermarchaeota* might have derived functional components via LGT interactions between community members in the hydrothermal environment as an adaptive strategy in an ecosystem containing copious amounts of heavy metals, C$_1$ compounds, and reduced sulfur compounds (1–6, 19, 20, 30). The distinctive C, H, N, and S metabolism of each clade (Fig. 4 and 5 and Tables S7b and c) could also be an outcome of LGT events during adaptation in the corresponding econiches. The LGT events could result from various mechanisms, such as symbiosis (31), virus-mediated gene transfer (32), and/or recombination between microorganisms (33). Nevertheless, the LGT-derived functions probably render *Hydrothermarchaeota* a distinctive archaeal lineage able to adapt to the hydrothermal environment.

Conclusions. The phylum *Hydrothermarchaeota* is a widespread archaeal lineage in the hydrothermal sediment environment. Within the microbial community, *Hydrother-*
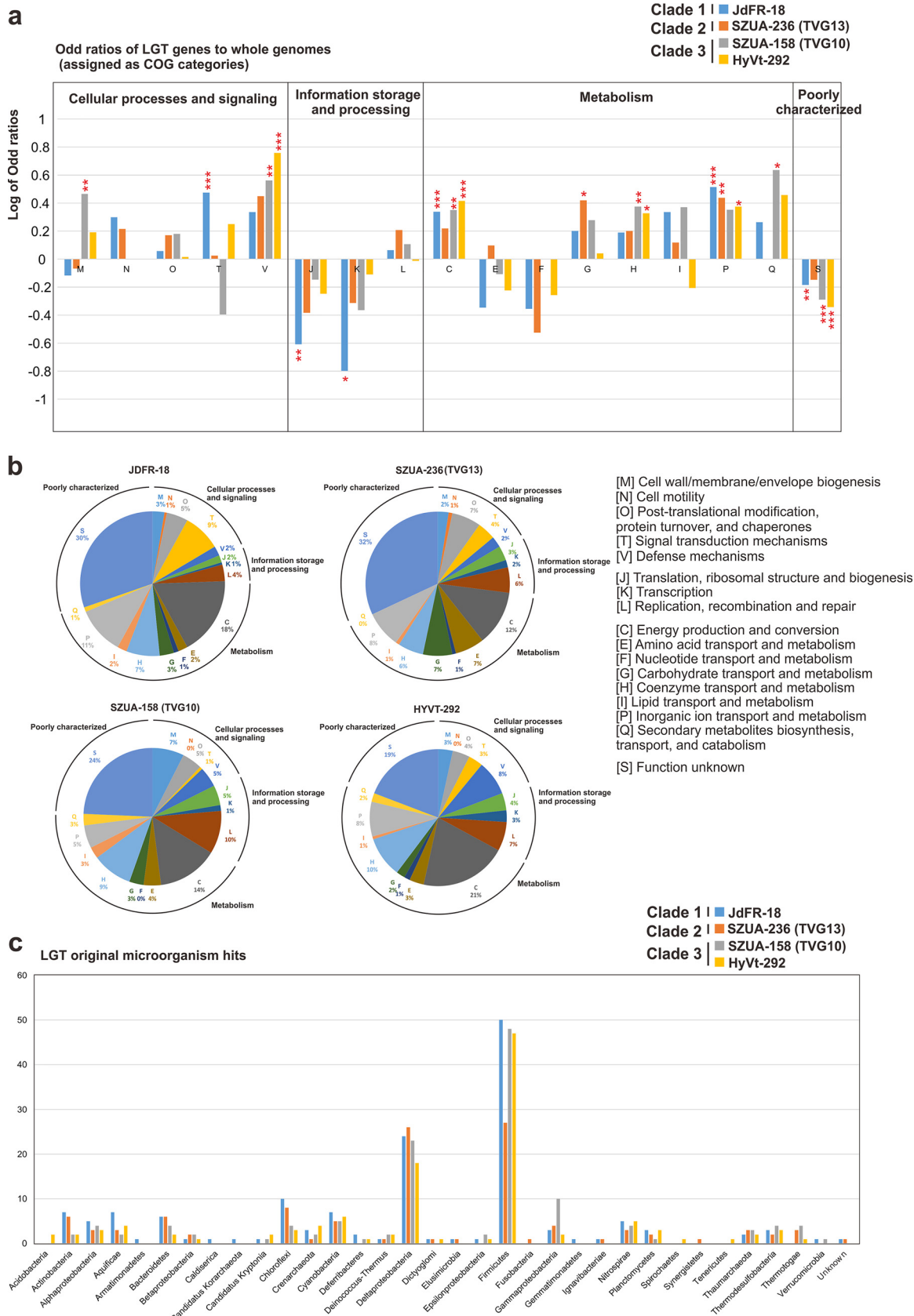
FIG 6 Lateral gene transfer analysis of *Hydrothermarchaeota*. (a) Odds ratios of potential genes horizontally transferred to whole genomes for genes assigned COG categories. The *P* value and odds ratios were calculated by Fisher's exact test. The *P* values indicate significant

(Continued on next page)

marchaeota synergistically participate in a wide range of carbon utilization and element cycling processes with other microbes. This suggests that microbial interactions are essential for nutrient and element cycling and extends the current understanding of community interactions within hydrothermal sediment and chimney environments (6, 34). Furthermore, the functional redundancy of carbon utilization and element cycling greatly increases the access of *Hydrothermarchaeota* to a wide range of carbon and energy sources.

Hydrothermarchaeota members have derived important functional components from the community via LGT. The comparative and evolutionary genomic analysis presented in the current study revealed that *Hydrothermarchaeota* members inhabiting the hydrothermal sediments form a distinctive clade in terms of genome content. LGT appears to have served as a niche-adaptive strategy of individual lineages, enabling them to metabolize heavy metals, C_1 compounds, and reduced sulfur compounds in the surrounding environment.

Originally detected in but not limited to hydrothermal environments, continental slopes, and abyssal sediments, we suspect that additional uncultured *Hydrothermarchaeota* members dwell in nonextreme environmental settings beyond the hydrothermal ecosystems, such as wetland, estuarine, coastal, and hot spring environments. The findings presented herein call for genomic studies of *Hydrothermarchaeota* from these other environments to further explore the evolutionary mechanisms and strategies used by these archaea to expand and adapt to a variety of environment types. Furthermore, the combined approaches, including genomic, transcriptomic, and enzymatic techniques, in cultivation-based experiments should facilitate research on the metabolic capacities and activities of this fascinating archaeal group in the future.

MATERIALS AND METHODS

Sample information and metagenome sequencing. Marine hydrothermal sediment samples were retrieved from an active deep-sea hydrothermal vent site (black smoker) at a depth of 2,770 m in the Mid-Atlantic Ridge of the South Atlantic Ocean during the DY125-26 cruise of R/V *Dayang Yihao* (R/V *Ocean No. 1*) in August 2012 (35). TVG10 was sampled from a layer from a black smoker chimney; TVG13 is a sulfur-rich sediment sample collected near the black smoker chimney. The samples were collected by the GTV04 deep-sea remote live-video-guided sampling device (maximum diving depth, 4,000 m). The sampler is a remotely operated mechanical clamp which is controlled by a metal shell-armed optoelectric cable to maintain a vertical attitude when sinking to the seafloor. The deep-sea camera system may be adjusted to the best position and attitude to operate the sampling work. Samples were stored at −80°C for subsequent metagenome sequencing, and physicochemical characterizations were conducted soon after collection. The latter characterizations included the total C, H, and S content, C/N ratio, and pH (35) (the chemical experimental details and results refer to those from the previous work). The metagenomic DNA was isolated from the samples using the custom procedure of the FastDNA spin kit for soil (MP Biomedicals, USA), and triple DNA extracts were used and pooled for individual samples.

Metagenome processing and genome-resolved binning. To obtain high-quality archaeal MAGs, a custom processing method involving two rounds of assemblage and binning was adopted. The metagenomes were sequenced by using an Illumina HiSeq 2000 platform. Two separate libraries for each sample were obtained and combined into one during downstream analysis. Raw reads were first dereplicated and processed by use of the Sickle tool (https://github.com/najoshi/sickle) with the default settings to trim low-quality reads. Clean reads for each sample were subjected to *de novo* metagenome assembly by use of the IDBA-UD (v1.1.1) assembler with the –mink 52 –maxk 92 –step 8 settings (36). The initial assemblies were deposited in the Integrated Microbial Genomes (IMG) system of the U.S. Department of Energy-Joint Genome Institute (DOE-JGI) database and annotated by using the DOE-JGI Microbial Genome Annotation Pipeline (MGAP; v.4) (37).

Assemblies were subjected to MetaBAT (v0.32.4) tool-based binning with 12 combinations of parameters (38). Subsequently, the Das-Tool (v1.0) program was applied to screen the MetaBAT bins, leading to high-quality and high-completeness bins (39). The CheckM (v1.0.7) program was used to assess the bin quality and phylogeny (40). The subsequently obtained archaeal MAGs were combined with (i) all available archaeal genomes from the GenBank database (updated 2 August 2017); (ii) archaeal clone, fosmid, and cosmid sequences from the NCBI nucleotide database (updated 2 August 2017); and (iii) the initial assembled scaffolds with one or more open reading frames annotated as being of archaeal

FIG 6 Legend (Continued)
deviations between horizontally transferred genes and whole genomes. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. (b) Distribution of COG categories of potential genes horizontally transferred. (c) Distribution of the original microorganisms of potential hits or horizontally transferred genes. Microorganisms were divided into the phylum level.

origin by the IMG database (only assemblies obtained in the current study) as the reference for read mapping. The BBmap program was used to obtain potential archaeal reads from raw reads with the vslow minid = 0.6 option (41). The second round of assembly by archaeal reads was assembled using the same *de novo* metagenome assembling method as that of the first round, and potential archaeon-related scaffolds were submitted to the DOE-JGI IMG database for annotation using the same method as that of the first round. The same method with the MetaBAT and Das-Tool programs was used to generate second-round MAGs; only high-quality archaeal MAGs were used for downstream analysis. Outlier scaffolds with abnormal coverage, tetranucleotide signals, and GC patterns within potentially highly contaminated MAGs (checked by use of the CheckM program) and erroneous small-subunit (SSU) sequences within MAGs were screened and decontaminated by using the RefineM (v0.0.20) program with the default settings (42). Average genome coverages were calculated by remapping raw reads to MAGs using the Bowtie2 (v2.2.8) program (43). Bacterial MAGs were obtained using binning and decontamination processes similar to those described above, but with only one round of binning. Further refinement was also conducted by manual inspection based on the VizBin program for selected MAGs (44).

Information was obtained from the SRA by using the string (((hydrothermal) AND metagenomic-[source]) AND WGS[strategy])) NOT 16S[title] NOT 454 GS[text word] AND (metagenome[organism] OR hydrothermal vent metagenome[organism] OR marine sediment metagenome[organism] OR marine metagenome[organism] OR subsurface metagenome[organism]) (updated 26 December 2017) to search hydrothermal vent sediment studies and the string (((spring sediment) AND metagenomic[source]) AND WGS[strategy]) NOT 16S[title] NOT 454 GS[text word] (updated 24 January 2018) to search for freshwater spring sediment studies deposited in the NCBI SRA. The information on the *Hydrothermarchaeota* 16S rRNA distribution and currently available *Hydrothermarchaeota* MAGs and single-cell-amplified genomes suggests that hydrothermal/spring sediments are their most frequently encountered environments. We applied metagenomic data sets from these environments to search for potential *Hydrothermarchaeota* MAGs. The SRA search results were manually inspected to confirm correctness (see Table S1 in the supplemental material). The DOE-JGI IMG deposits linked to these SRA deposits were identified, and the corresponding assemblies were used. MAGs originating from hydrothermal vent sediments (21 studies) and freshwater spring sediments (22 studies) were reconstructed from the public NCBI SRA deposits and the linked DOE-JGI IMG deposits (only one study had the IMG record but no SRA record; that study was also manually inspected to verify that it met the search criteria) (Table S1). SRA runs within one experiment and studies for one biosample underwent integrated assembly. Assembly was conducted by using the MEGAHIT (v1.1.2) program (45) with kmer iterations of k35 to k75, k45 to k95, k65 to k145, and k145 to k295 for 85-bp, 100-bp, 150-bp, and 300-bp reads, respectively, and a kmer step of 10; the preprocessing method was the same as that described above. The assembled metagenomes and reads from studies with DOE-JGI IMG deposits that passed quality control were simply used. The downstream binning methods were the same as those described above, but with one round of binning. Further refinement was also conducted by manual inspection based on the VizBin program for selective MAGs (44).

Archaeal MAG annotation. Knockout (KO) annotation was made by use of the GhostKOALA (v2.0), KAAS (v2.1), and eggNOG-mapper (v4.5.1) programs (using the first KO hit and cluster of orthologous groups [COG] hit, COG were translated to KO by use of the ko2cog.xl tool provided by the KEGG database) (46–48). For annotation by use of the NCBI nr database (updated 6 March 2017), the BLASTP result was further refined by extracting only the first meaningful hits (hits with meaningful information rather than a label as a hypothetical protein). Peptidases were called by use of the MEROPS database (using the pepunit database to avoid potential false-positive hits) via the DIAMOND BLASTP (v0.9.10.111) program with the -k 1 -e 1e-10 –subject-cover 80 –id 50 settings (49, 50). The CAZy annotation was done by using the dbCAN (v20170913) program and interpreted by using the CAZy database (self-parsed online information) (51, 52). The InterProScan (v5.26-65.0, client version) software package was used to classify protein functions, with annotations including CDD, PfamA, SMART, TIGRFAM, Phobius, and SuperFamily (53). The Phobius, PRED-SIGNAL, and PSORTb (v3.0.2) (Archaea) programs were used to predict the locations of peptidases as membrane/intracellular or extracellular (only congruent results for the extracellular location from the three methods were adopted; incongruent results were assigned a membrane/intracellular location) (54–56).

Major allele frequency analysis. For major allele frequency analysis, the anvi'o (v4.0) program was used to identify and profile single-nucleotide variants (SNVs) of *Hydrothermarchaeota* MAGs based on the mapping of the reads from the corresponding metagenomes. The strategies used to characterize the SNVs identified were performed according to the online instructions (http://merenlab.org/2015/07/20/analyzing-variability/). The major allele frequency value was the percentage of metagenomic reads mapping to a certain site with the majority SNV.

Comparative genomic analysis. The Markov cluster (MCL) algorithm implemented in the anvi'o (v4.0) program was used for protein clustering (57). The eggNOG-mapper (v4.5.1) program was used to annotate MAGs with default settings (47, 48). COG functional categories and orthologous groups parsed from the eggNOG mapping results were used to reconstruct the inner tree. The existence of specific functions or pathways was assigned according to the existence of marker genes (using the annotation results obtained as described above). Average nucleotide identity (ANI) values among *Hydrothermarchaeota* MAGs were calculated by use of the OrthoANI program with the default settings (58).

Phylogenetic reconstruction. A search of SILVA SSU128 for long marine benthic group E (*Hydrothermarchaeota*) sequences with good quality (pintail quality > 75%, sequence length > 1,000 nucleotides, sequence quality > 75%) resulted in 549 sequences (which were assigned to the *Hydrothermar-*

chaeota backbone tree HydroBBTree) (59). The alignment obtained was subjected to clustering by use of mothur software (60); 36 representative operational taxonomic unit sequences at a 90% similarity cutoff were obtained. Representative sequences in the SSURef_NR99_128_SILVA database and archaeal 16S rRNA gene sequences retrieved from metagenomic scaffolds (curated by the IMG database) and MAGs were combined (only sequences with lengths of >300 bp were considered) and subsequently aligned by using the SINA (v1.2.11) program (61). The updated 16S rRNA genes from the *Pacearchaeota* and *Asgard* superphylum genomes (deposited in the NCBI Genome database) were also used for tree construction. The SINA alignment obtained with *Escherichia coli* K-12 as the outgroup was filtered by the use of both ssuref:archaea (LTPs128_SSU) and 50% consensus filters and was subsequently used for tree construction by use of the RAxML-HPC (v8) program on the XSEDE portal implemented in the CIPRES portal with the settings of GTRCAT and 1,000 bootstrap iterations (62, 63).

The 16S rRNA gene sequences (>300 bp) that were identified by a BLAST search in the *Hydrothermarchaeota* MAGs constructed based on NCBI SRAs, previous reports, and the current study were aligned by use of the SINA (v1.2.11) program (61) and inserted into HydroBBTree by using the ARB_PARSIMONY quick-add species method in ARB software (64). Some MAGs lacked 16S rRNA gene sequences, which was not unusual, considering the low MAG completeness. The topology of the generated 16S rRNA gene tree was the same as that of HydroBBTree, and the division of clades also remained unchanged.

The masked alignment of 12 RPs (which were processed by the CheckM program and which included the L2, L3, L4, L5, L14, L16, L18, L22, S3, S8, S17, and S19 RPs) was concatenated and then subjected to tree model selection by use of the ProtTest (v3) program (40, 65). Representative archaeal genomes and reported *Hydrothermarchaeota* MAGs were included in the tree together with MAGs and scaffolds from the current study (66). A preselection was imposed on the concatenated alignment to filter sequences with less than three RPs and less than 25% aligned columns; columns with more than 50% gaps were trimmed. The RAxML-HPC (v8) program on the XSEDE portal implemented in the CIPRES portal was used to generate a phylogenetic tree with the best model as PROTGAMMAILG and 1,000 bootstrap iterations (62, 63). The *E. coli* K-12 genome was adopted as the outgroup (67).

Evolutionary analysis. The genomes of phylogenetically closely related archaeal orders/classes were retrieved from the NCBI Genome database. These included the genomes of the methanogenic organisms *Methanobacteriales*, *Methanococcales*, *Methanofastidiosa*, and *Methanopyri* and the nonmethanogenic organisms *Theionarchaea*, *Hadesarchaea*, and *Thermococcales*. One *Crenarchaeota* genome (*Acidilobus saccharovorans* strain 345-15) was used as the outgroup. The genome-picking criteria were over 80% completeness and less than 10% genome contamination; the only exceptions were two *Theionarchaea* genomes (the only two genomes available within the class) and one *Hadesarchaea* genome (77.6% completeness; one of two genomes available within the class). The analyzed genomes represented different families or genera, if possible. The phylogenomic tree of the 50 genomes retrieved was constructed by using a concatenated masked alignment of 12 RPs and the same method described above but using the IQ-TREE (v1.6.3) program (68) with the settings -m MFP -mset LG,WAG -mrate E,I,G,I+G -mfreq FU -bb 1000, which let the software precalculate the best-fit model for phylogenetic reconstruction.

The OGs of protein-coding genes shared by 50 genomes were parsed out by using the OrthoFinder (v2.2.6) program (69); orphan genes (genes present in only one genome) were not included in the OGs. The BadiRate program was used to estimate the OG turnover rate using the BDI-FR-CSP model (turnover rates-branch model-estimating procedure, being stringent on estimating turnover rates) (70), with the above-described phylogenomic tree being used as the input tree file. The output gene turnover results were parsed to OG turnover results by a custom Perl script. The OGs were annotated by the eggNOG-mapper (v4.5.1) program; each was assigned the majority annotation result. The key OG turnover events on *Hydrothermarchaeota* nodes were parsed; the related genes with function and pathway annotations were summarized.

Metabolic capacity prediction and comparison. The genomes of *Euryarchaeota* and *Hydrothermarchaeota* were retrieved from the NCBI Genome database, and five representative genomes (selected from different families, if possible) from each archaeal group were used (9). Only genomes with over 70% completeness were used. If the number of available genomes in one archaeal group was limited (less than five), all genomes were used, regardless of completeness. Metabolic marker genes were retrieved from a custom metabolic gene database and metabolic pathways annotated in the KEGG database (71, 72). The Pfam, TIGRfam, and custom metabolic gene databases were used to scan the genomes with suggested cutoff settings; KOs were assigned to genomes by using the GhostKOALA (v2.0), KAAS (v2.1), and eggNOG-mapper (v4.5.1) programs with default settings (46–48). For each metabolic marker gene, its presence/absence in the archaeal group was denoted by a solid black dot (present in all groups), solid gray dot (present in some groups), or blank dot (present in none of the groups). For each metabolic function, if one marker gene was identified, the presence of the function was assumed. Because of the limited number of genomes and low genome completeness (less than 70%) for *Hadesarchaea*, *Theionarchaea*, *Syntrophoarchaeum*, and MSBL-1 (Mediterranean Sea Brine Lakes 1), if one metabolic marker gene/metabolic function was identified in any genome within the individual archaeal group, the fact was denoted by a solid black dot.

For the community metabolic analysis of MAGs from both metagenomes, a metabolic capacity prediction method similar to the one described above was used. The abundances of peptidases and CAZys were calculated by considering the MAG completeness and taking the average values for all MAGs within an individual microbial group (0 digits were used after the decimal point). The presence of a specific pathway/function within each microbial group was assigned when the pathway/function was detected in any MAG within the microbial group. The Fe uptake pathway was predicted by the

corresponding database (73), using the DIAMOND BLASTP (v0.9.10.111) program with the settings -e 1e-20 –query-cover 80 –id 65 (50).

Data availability. Accession numbers for the initial assemblies were as follows: IMG accession number 3300003886 for TVG10 and IMG accession number 3300003885 for TVG13. Accession numbers for the second-round assemblies were as follows: IMG accession number 3300020233 for TVG10 and IMG accession number 3300020236 for TVG13. The MAGs that were resolved in the current study are deposited in NCBI under BioProject accession numbers PRJNA385762 and PRJNA480137. Detailed genomic parameters of the assembled genomes are summarized in Text S1. The NCBI and IMG genome accession numbers for the MAGs that were used in this study are NCBI accession number GCA_003230355.1 for SZUA-236, NCBI accession number GCA_003229275.1 for SZUA-237, NCBI accession number GCA_003229935.1 for SZUA-158, NCBI accession number SRR7786998 for HyVt-292, IMG accession number 2728369317 for JdFR-16, IMG accession number 2728369322 for JdFR-17, and IMG accession number 2728369320 for JdFR-18.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TEXT S1, PDF file, 0.1 MB.

FIG S1, PDF file, 0.7 MB.

TABLE S1, XLSX file, 0.04 MB.

TABLE S2, XLSX file, 0.02 MB.

TABLE S3, XLSX file, 0.01 MB.

TABLE S4, XLSX file, 0.7 MB.

TABLE S5, XLSX file, 0.02 MB.

TABLE S6, XLSX file, 0.05 MB.

TABLE S7, XLSX file, 1 MB.

DATA SET S1, XLSX file, 0.8 MB.

REFERENCES

1. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs K-U, Teske AP, Ettema T. 2016. Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. Nat Microbiol 1:16002. https://doi.org/10.1038/nmicrobiol.2016.2.

2. Cleaves HJ. 2008. The prebiotic geochemistry of formaldehyde. Precambrian Res 164:111–118. https://doi.org/10.1016/j.precamres.2008.04.002.

3. Orita I, Yurimoto H, Hirai R, Kawarabayasi Y, Sakai Y, Kato N. 2005. The archaeon *Pyrococcus horikoshii* possesses a bifunctional enzyme for formaldehyde fixation via the ribulose monophosphate pathway. J Bacteriol 187:3636–3642. https://doi.org/10.1128/JB.187.11.3636-3642.2005.

4. Sokolova TG, Henstra A-M, Sipma J, Parshina SN, Stams AJ, Lebedinsky AV. 2009. Diversity and ecophysiological features of thermophilic carboxydotrophic anaerobes. FEMS Microbiol Ecol 68:131–141. https://doi.org/10.1111/j.1574-6941.2009.00663.x.

5. Dick G, Anantharaman K, Baker B, Li M, Reed D, Sheik C. 2013. The microbiology of deep-sea hydrothermal vent plumes: ecological and biogeographic linkages to seafloor and water column habitats. Front Microbiol 4:124. https://doi.org/10.3389/fmicb.2013.00124.

6. Dombrowski N, Seitz KW, Teske AP, Baker BJ. 2017. Genomic insights into potential interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments. Microbiome 5:106. https://doi.org/10.1186/s40168-017-0322-2.

7. Vetriani C, Jannasch HW, MacGregor BJ, Stahl DA, Reysenbach AL. 1999. Population structure and phylogenetic characterization of marine benthic archaea in deep-sea sediments. Appl Environ Microbiol 65:4375–4384.

8. Carr SA, Jungbluth SP, Eloe-Fadrosh EA, Stepanauskas R, Woyke T, Rappé MS, Orcutt BN. 2019. Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. ISME J 13:1457–1468. https://doi.org/10.1038/s41396-019-0352-9.

9. Jungbluth SP, Amend JP, Rappé MS. 2017. Metagenome sequencing and 98 microbial genomes from Juan de Fuca Ridge flank subsurface fluids. Sci Data 4:170037. https://doi.org/10.1038/sdata.2017.80.

10. Kato S, Nakano S, Kouduka M, Hirai M, Suzuki K, Itoh T, Ohkuma M,

Suzuki Y. 2019. Metabolic potential of as-yet-uncultured archaeal lineages of *Candidatus* Hydrothermarchaeota thriving in deep-sea metal sulfide deposits. Microbes Environ 34:293–303. https://doi.org/10.1264/jsme2.ME19021.

11. Vanwonterghem I, Evans PN, Parks DH, Jensen PD, Woodcroft BJ, Hugenholtz P, Tyson GW. 2016. Methylotrophic methanogenesis discovered in the archaeal phylum *Verstraetearchaeota*. Nat Microbiol 1:16170. https://doi.org/10.1038/nmicrobiol.2016.170.

12. Hocking WP, Stokke R, Roalkvam I, Steen IH. 2014. Identification of key components in the energy metabolism of the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* by transcriptome analyses. Front Microbiol 5:95. https://doi.org/10.3389/fmicb.2014.00095.

13. Evans PN, Parks DH, Chadwick GL, Robbins SJ, Orphan VJ, Golding SD, Tyson GW. 2015. Methane metabolism in the archaeal phylum *Bathyarchaeota* revealed by genome-centric metagenomics. Science 350:434–438. https://doi.org/10.1126/science.aac7745.

14. Fuchs G, Boll M, Heider J. 2011. Microbial degradation of aromatic compounds—from one strategy to four. Nat Rev Microbiol 9:803–816. https://doi.org/10.1038/nrmicro2652.

15. González JM, Robb FT. 2000. Genetic analysis of *Carboxydothermus hydrogenoformans* carbon monoxide dehydrogenase genes *cooF* and *cooS*. FEMS Microbiol Lett 191:243–247. https://doi.org/10.1111/j.1574-6968.2000.tb09346.x.

16. Hartmann T, Leimkühler S. 2013. The oxygen-tolerant and NAD$^+$-dependent formate dehydrogenase from *Rhodobacter capsulatus* is able to catalyze the reduction of CO_2 to formate. FEBS J 280:6083–6096. https://doi.org/10.1111/febs.12528.

17. Zhou Z, Liu Y, Lloyd KG, Pan J, Yang Y, Gu J-D, Li M. 2018. Genomic and transcriptomic insights into the ecology and metabolism of benthic archaeal cosmopolitan, Thermoprofundales (MBG-D archaea). ISME J 13:885–901. https://doi.org/10.1038/s41396-018-0321-8.

18. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, Rappé MS, Pester M, Loy A, Thomas BC, Banfield JF. 2018. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J 12:1715–1728. https://doi.org/10.1038/s41396-018-0078-0.

19. Fortunato CS, Larson B, Butterfield DA, Huber JA. 2018. Spatially distinct, temporally stable microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal vent fluids. Environ Microbiol 20:769–784. https://doi.org/10.1111/1462-2920.14011.

20. Vetriani C, Chew YS, Miller SM, Yagi J, Coombs J, Lutz RA, Barkay T. 2005. Mercury adaptation among bacteria from a deep-sea hydrothermal vent. Appl Environ Microbiol 71:220–226. https://doi.org/10.1128/AEM.71.1.220-226.2005.

21. Bräsen C, Esser D, Rauch B, Siebers B. 2014. Carbohydrate metabolism in Archaea: current insights into unusual enzymes and pathways and their regulation. Microbiol Mol Biol Rev 78:89–175. https://doi.org/10.1128/MMBR.00041-13.

22. Madigan M, Martinko JM, Bender KS, Buckley DH, Stahl DA. 2015. Brock biology of microorganisms, 14th ed. Pearson, Boston, MA.

23. Berg IA, Kockelkorn D, Ramos-Vera WH, Say RF, Zarzycki J, Hügler M, Alber BE, Fuchs G. 2010. Autotrophic carbon fixation in archaea. Nat Rev Microbiol 8:447–460. https://doi.org/10.1038/nrmicro2365.

24. Hügler M, Huber H, Stetter KO, Fuchs G. 2003. Autotrophic CO_2 fixation pathways in archaea (Crenarchaeota). Arch Microbiol 179:160–173. https://doi.org/10.1007/s00203-002-0512-5.

25. Sato T, Atomi H, Imanaka T. 2007. Archaeal type III RuBisCOs function in a pathway for AMP metabolism. Science 315:1003–1006. https://doi.org/10.1126/science.1135999.

26. Kengen SW, Tuininga JE, de Bok FA, Stams AJ, de Vos WM. 1995. Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus*. J Biol Chem 270:30453–30457. https://doi.org/10.1074/jbc.270.51.30453.

27. Siebers B, Schönheit P. 2005. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Curr Opin Microbiol 8:695–705. https://doi.org/10.1016/j.mib.2005.10.014.

28. Schiffer A, Fritz G, Kroneck PM, Ermler U. 2006. Reaction mechanism of the iron-sulfur flavoenzyme adenosine-5′-phosphosulfate reductase based on the structural characterization of different enzymatic states. Biochemistry 45:2960–2967. https://doi.org/10.1021/bi0521689.

29. Li M, Baker BJ, Anantharaman K, Jain S, Breier JA, Dick GJ. 2015. Genomic and transcriptomic evidence for scavenging of diverse organic compounds by widespread deep-sea archaea. Nat Commun 6:8933. https://doi.org/10.1038/ncomms9933.

30. Meier DV, Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, Amann R, Meyerdierks A. 2017. Niche partitioning of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J 11:1545–1558. https://doi.org/10.1038/ismej.2017.37.

31. Hoffmeister M, Martin W. 2003. Interspecific evolution: microbial symbiosis, endosymbiosis and gene transfer. Environ Microbiol 5:641–649. https://doi.org/10.1046/j.1462-2920.2003.00454.x.

32. Koonin EV, Wolf YI. 2012. Evolution of microbes and viruses: a paradigm shift in evolutionary biology? Front Cell Infect Microbiol 2:119. https://doi.org/10.3389/fcimb.2012.00119.

33. Allen EE, Banfield JF. 2005. Community genomics in microbial ecology and evolution. Nat Rev Microbiol 3:489–498. https://doi.org/10.1038/nrmicro1157.

34. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S, Schramm A, Jorgensen BB. 2013. Predominant archaea in marine sediments degrade detrital proteins. Nature 496:215–218. https://doi.org/10.1038/nature12033.

35. Xu W, Li M, Ding J-F, Gu J-D, Luo Z-H. 2014. Bacteria dominate the ammonia-oxidizing community in a hydrothermal vent site at the Mid-Atlantic Ridge of the South Atlantic Ocean. Appl Microbiol Biotechnol 98:7993–8004. https://doi.org/10.1007/s00253-014-5833-1.

36. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.

37. Markowitz VM, Chen I-MA, Chu K, Szeto E, Palaniappan K, Grechkin Y, Ratner A, Jacob B, Pati A, Huntemann M, Liolios K, Pagani I, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. 2012. IMG/M: the integrated metagenome data management and comparative analysis system. Nucleic Acids Res 40:D123–D129. https://doi.org/10.1093/nar/gkr975.

38. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.

39. Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3:836–843. https://doi.org/10.1038/s41564-018-0171-1.

40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

41. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner, abstr 1241166. Abstr 9th Annu Genomics Energy Environ Meet, Walnut Creek, CA. Office of Science, U.S. Department of Energy, Washington, DC.

42. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542. https://doi.org/10.1038/s41564-017-0012-7.

43. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.

44. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, Van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin—an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome 3:1. https://doi.org/10.1186/s40168-014-0066-1.

45. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

46. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res 35:W182–W185. https://doi.org/10.1093/nar/gkm321.

47. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.

48. Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, Rattei T, Mende DR, Sunagawa S, Kuhn M, Jensen LJ, von Mering C, Bork P. 2016. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44:D286–D293. https://doi.org/10.1093/nar/gkv1248.

49. Rawlings ND, Barrett AJ, Finn R. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 44:D343–D350. https://doi.org/10.1093/nar/gkv1118.

50. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment

using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

51. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 40:W445–W451. https://doi.org/10.1093/nar/gks479.

52. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.

53. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

54. Bagos PG, Tsirigos KD, Plessas SK, Liakopoulos TD, Hamodrakas SJ. 2009. Prediction of signal peptides in archaea. Protein Eng Des Sel 22:27–35. https://doi.org/10.1093/protein/gzn064.

55. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, Dao P, Sahinalp SC, Ester M, Foster LJ, Brinkman F. 2010. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608–1615. https://doi.org/10.1093/bioinformatics/btq249.

56. Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res 35:W429–W432. https://doi.org/10.1093/nar/gkm256.

57. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for omics data. PeerJ 3:e1319. https://doi.org/10.7717/peerj.1319.

58. Lee I, Ouk Kim Y, Park S-C, Chun J. 2016. OrthoANI: an improved algorithm and software for calculating average nucleotide identity. Int J Syst Evol Microbiol 66:1100–1103. https://doi.org/10.1099/ijsem.0.000760.

59. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.

60. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.

61. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. https://doi.org/10.1093/bioinformatics/bts252.

62. Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees, abstr 11705685. Abstr Gateway Computing Environments Workshop (GCE), 2010, New Orleans, LA. IEEE, New York, NY.

63. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

64. Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar Buchner A, Lai T, Steppi S, Jobb G, Forster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, Konig A, Liss T, Lussmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A, Schleifer KH. 2004. ARB: a software environment for sequence data. Nucleic Acids Res 32:1363–1371. https://doi.org/10.1093/nar/gkh293.

65. Darriba D, Taboada GL, Doallo R, Posada D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27:1164–1165. https://doi.org/10.1093/bioinformatics/btr088.

66. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. Nat Microbiol 1:16048. https://doi.org/10.1038/nmicrobiol.2016.48.

67. Letunic I, Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128. https://doi.org/10.1093/bioinformatics/btl529.

68. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.

69. Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16:157. https://doi.org/10.1186/s13059-015-0721-2.

70. Librado P, Vieira FG, Rozas J. 2012. BadiRate: estimating family turnover rates by likelihood-based methods. Bioinformatics 28:279–281. https://doi.org/10.1093/bioinformatics/btr623.

71. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, Thomas BC, Singh A, Wilkins MJ, Karaoz U, Brodie EL, Williams KH, Hubbard SS, Banfield JF. 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat Commun 7:13219. https://doi.org/10.1038/ncomms13219.

72. Kanehisa M, Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30. https://doi.org/10.1093/nar/28.1.27.

73. Toulza E, Tagliabue A, Blain S, Piganeau G. 2012. Analysis of the Global Ocean Sampling (GOS) project for trends in iron uptake by surface ocean microbes. PLoS One 7:e30931. https://doi.org/10.1371/journal.pone.0030931.