Global patterns of diversity and metabolism of microbial communities in deep-sea hydrothermal vent deposits

Zhichao Zhou[1]*, Emily St. John[2]*, Karthik Anantharaman[1+], Anna-Louise Reysenbach[2+]

[1]Department of Bacteriology, University of Wisconsin-Madison, Madison, WI 53706, USA

[2]Center for Life in Extreme Environments, Biology Department, Portland State University, Portland, OR 97201, USA

*Contributed equally to this work.

[+]Corresponding authors. Karthik Anantharaman, Email: karthik@bact.wisc.edu; Anna-Louise Reysenbach, Email: bwar@pdx.edu.

Keywords: thermophiles, deep-sea hydrothermal vents, Archaea, Bacteria, metagenomics

Running title: Diversity and metabolism of deep-sea hydrothermal vent microbes

Abstract

Background

When deep-sea hydrothermal fluids mix with cold oxygenated fluids, minerals precipitate out of solution and form hydrothermal deposits. These actively venting deep-sea hydrothermal deposits support a rich diversity of thermophilic microorganisms which are involved in a range of carbon, sulfur, nitrogen, and hydrogen metabolisms. Global patterns of thermophilic microbial diversity in deep-sea hydrothermal ecosystems have illustrated the strong connectivity between geological processes and microbial colonization, but little is known about the genomic diversity and physiological potential of these novel taxa. Here we explore this genomic diversity in 42 metagenomes from four deep-sea hydrothermal vent fields and a deep-sea volcano collected from 2004 to 2018, and document their potential implications in biogeochemical cycles.

Results

Our dataset represents 3,635 metagenome-assembled genomes encompassing 511 novel and recently identified genera from deep-sea hydrothermal settings. Some of the novel bacterial (107) and archaeal genera (30) that were recently reported from the deep-sea Brothers volcano were also detected at the deep-sea hydrothermal vent fields, while 99 bacterial and 54 archaeal genera were endemic to the deep-sea Brothers volcano deposits. We report some of the first examples of medium (≥50% complete, ≤10% contaminated) to high quality (>90% complete, <5% contaminated) MAGs from phyla and families never previously identified, or poorly sampled, from deep-sea hydrothermal environments. We greatly expand the novel diversity of Thermoproteia, Patescibacteria (Candidate Phyla Radiation, CPR), and Chloroflexota found at deep-sea hydrothermal vents and identify a small sampling of two potentially novel phyla, designated JALSQH01 and JALWCF01. Metabolic pathway analysis of metagenomes provides insights into the prevalent carbon, nitrogen, sulfur and hydrogen metabolic processes across all sites, and illustrates sulfur and nitrogen metabolic ‘handoffs’ in community interactions. We

confirm that Campylobacteria and Gammaproteobacteria occupy similar ecological guilds but their prevalence in a particular site is driven by shifts in the geochemical environment.

Conclusion

Our study of globally-distributed hydrothermal vent deposits provides a significant expansion of microbial genomic diversity associated with hydrothermal vent deposits and highlights the metabolic adaptation of taxonomic guilds. Collectively, our results illustrate the importance of comparative biodiversity studies in establishing patterns of shared phylogenetic diversity and physiological ecology, while providing many targets for enrichment and cultivation of novel and endemic taxa.

Introduction

Actively venting deep-sea hydrothermal deposits at oceanic spreading centers and arc volcanoes support a high diversity of thermophilic microorganisms. Many of these microbes acquire metabolic energy from chemical disequilibria created by the mixing of reduced high temperature endmember hydrothermal fluids with cold oxygenated seawater. Community analysis of deposits using the 16S rRNA gene has revealed a rich diversity of novel archaeal and bacterial taxa [1–4] where the community composition is strongly influenced by the abundance of redox reactive species in high temperature vent fluids [e.g. 5–7]. The variations in the composition of endmember fluids, and in turn the microbial community composition at different vent fields, reflect the temperature and pressure of fluid-rock interaction, in addition to substrate composition and entrainment of magmatic volatiles. For example, along the Mid-Atlantic Ridge, methanogens are associated with deposits from H_2-rich vents at Rainbow and are absent in H_2-poor vents at Lucky Strike [3]. At the Eastern Lau Spreading Center (ELSC), similar to other back-arc basins, the hydrothermal fluids are generally quite variable depending on differences in inputs of acidic magmatic volatiles, contributions from the subducting slab and

proximity of island arc volcanoes. Such geochemical differences are imprinted in the diversity of microbial communities [3, 4]. Similar complex community structure dynamics have also been recently reported for the communities of the submarine Brothers volcano on the Kermadec Arc [8].

While such global patterns of high temperature microbial diversity in deep-sea hydrothermal systems have demonstrated geological drivers of microbial colonization, little is known about the genomic diversity and physiological potential of the many reported novel taxa. While a few metagenomic studies of hydrothermal fluids and sediments have provided a much greater understanding of the functional potential of these communities [e.g. 7, 9–13], the metagenomic analysis of deposits has been limited to a small number of samples [e.g. 14–16]. One exception is the study of about 16 deep-sea hydrothermal deposits from Brothers volcano, which resulted in 701 medium and high quality metagenome-assembled genomes (MAGs) [8]. Further, this study demonstrated that there were functionally distinct high temperature communities associated with the volcano that could be explained through an understanding of the geological history and subsurface hydrologic regime of the volcano.

Here, we expand on the Brothers volcano study by exploring the genomic and functional diversity of hydrothermal deposits collected from deep-sea vents in the Pacific and Atlantic oceans. We greatly increase the number of novel high quality assembled genomes from deep-sea vents, many of which are endemic to vents and do not have any representatives in culture yet. We also show that known important biogeochemical cycles in hydrothermal ecosystems are accomplished by the coordination of several taxa as metabolic handoffs, where in some cases different taxa accomplish similar functions in different environments, potentially providing functional redundancy in fluctuating conditions.

Results and Discussion

Patterns of metagenomic diversity in deep-sea hydrothermal deposits

We sequenced 42 metagenomes from 40 samples (38 hydrothermal vent deposit samples and two diffuse flow fluids) collected at deep-sea hydrothermal vents and a deep-sea volcano. These represent one of the largest global collections of metagenomes from such samples (Fig. S1, S2). This study spans vent deposit collections from 2004 to 2018, from deep-sea hydrothermal vent fields in the north Atlantic (Mid-Atlantic Ridge, MAR), east and southwest Pacific (East Pacific Rise, EPR; Eastern Lau Spreading Center, ELSC), a sedimented hydrothermal system (Guaymas Basin, GB), and a deep-sea volcano (Brothers volcano, BV) (Table S1).

In this study, *de novo* assembly of sequencing data and subsequent genome binning and curation (see methods for details) resulted in 2,983 bacterial and 652 archaeal draft metagenome-assembled genomes (MAGs with ≥50% completeness, Table S2). Of these, ~21% were >90% complete, with <5% contamination, and ~36% contained a 16S rRNA gene fragment. The MAGs were initially characterized phylogenetically using the Genome Taxonomy Database Toolkit (GTDB-Tk) (Fig. 1, 2, 3, Data S1, S2, S3, S4, S5) [17]. MAGs that could not be assigned to a known genus by GTDB-Tk were assigned to new genera using AAI with the recommended cutoffs in Konstantinidis et al. [18] (Table S3A, B). Shared phyla between most of the hydrothermal deposits (excluding samples from the highly acidic Brothers volcano sites, and the diffuse flow fluids) included the Halobacteriota (e.g. Archaeoglobaceae), Methanobacteriota (e.g. Thermococcaceae), Thermoproteota (e.g. Acidilobaceae, Pyrodictiaceae), Acidobacteriota, Aquificota (e.g. Aquificaceae), Bacteroidota (e.g. Flavobacteriaceae), Campylobacterota (e.g. Sulfurimonadaceae, Nautiliaceae, Hippeaceae), Chloroflexota, Deinococcota (e.g. Marinithermaceae), Desulfobacterota (e.g. Dissulfuribacteraceae, Thermodesulfobacteriaceae),

Proteobacteria (e.g. Alphaproteobacteria, Gammaproteobacteria), and the Patescibacteria (Table S4). Many of these phyla have only a few representatives in isolated cultures and point to the importance of combining enrichment cultivation strategies with metagenomic approaches to obtain additional insights into the physiological ecology of these core lineages.

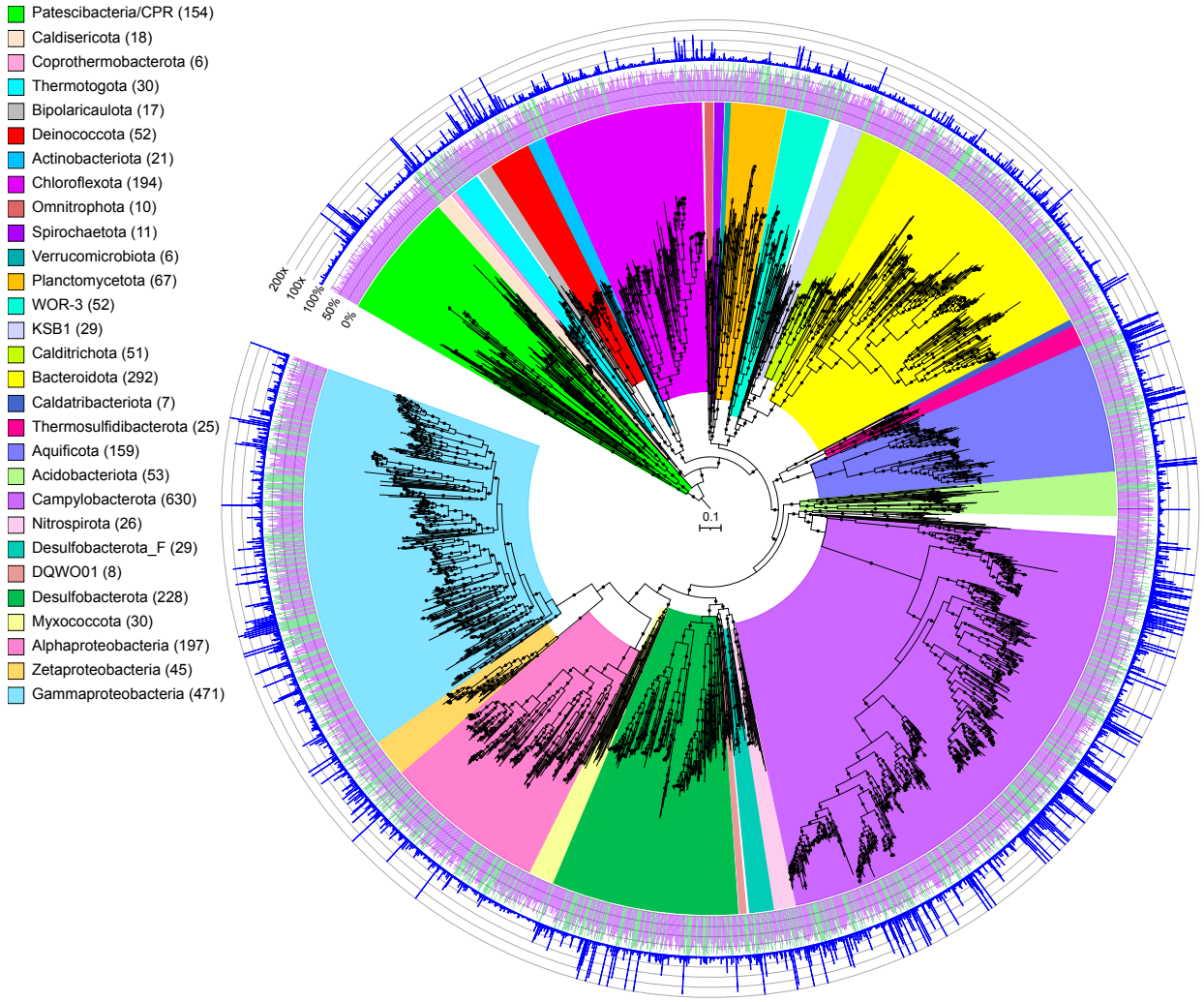


Fig. 1. Maximum-likelihood phylogenomic tree of bacterial metagenome-assembled genomes, constructed using 120 bacterial marker genes in GTDB-Tk. Major taxonomic groups are highlighted, and the number of MAGs in each taxon is shown in parentheses. See Table S2 for details. Bacterial lineages are shown at the phylum classification, except for the Proteobacteria which are split into their component classes. The inner ring displays quality (green: >90% completion, <5% contamination; purple: medium quality, ≥50% completion, ≤10% contamination), while the outer ring shows normalized read coverage up to 200x. The scale bar indicates 0.1 amino acid substitutions per site, and

filled circles are shown for SH-like support values ≥80%. The tree was artificially rooted with the Patescibacteria using iTOL. The Newick format tree used to generate this figure is available in Data S4, and the formatted tree is available online at https://itol.embl.de/shared/alrlab.

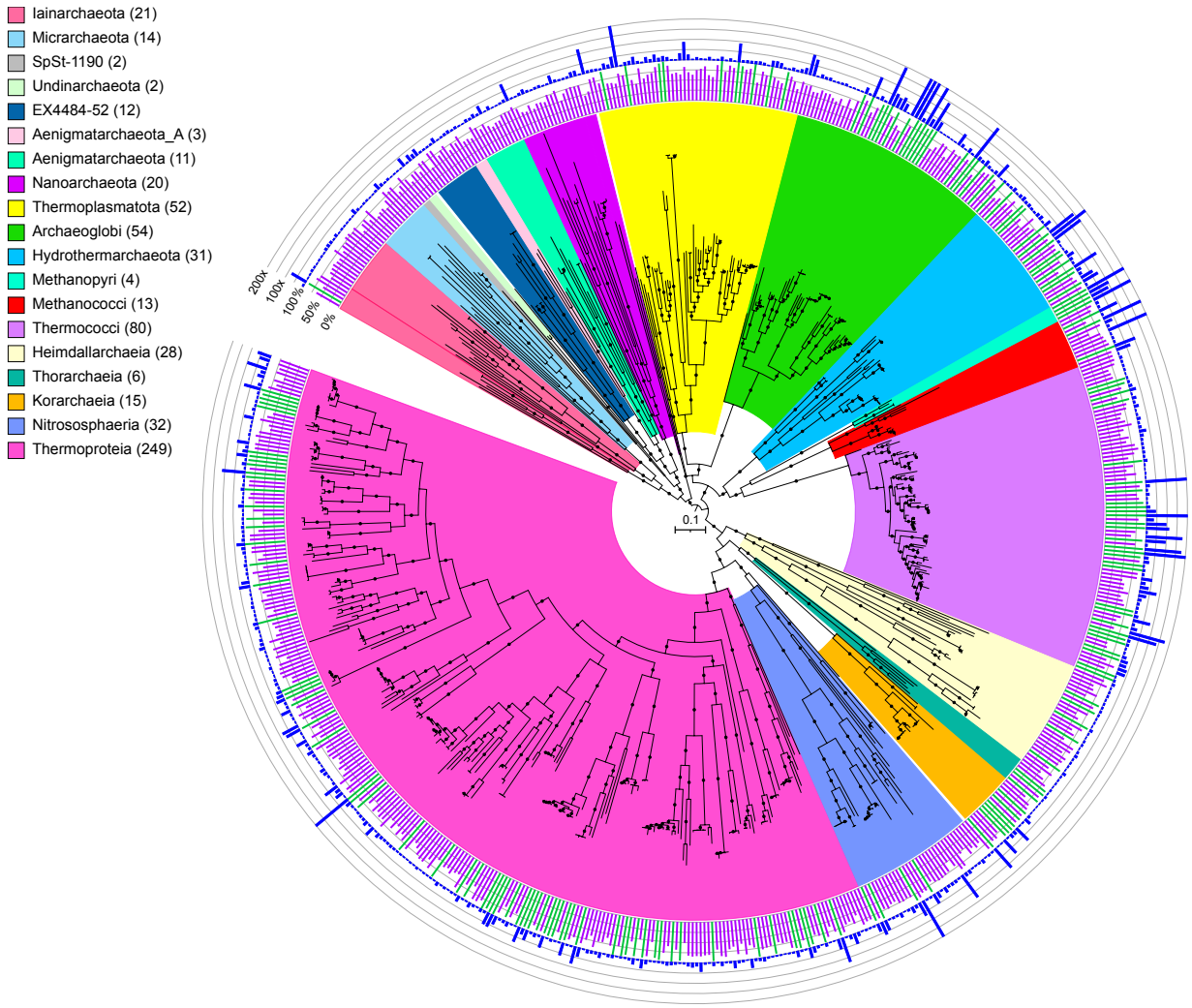


Fig. 2. Maximum-likelihood phylogenomic reconstruction of deep-sea hydrothermal vent archaeal metagenome-assembled genomes generated in GTDB-Tk. The tree was generated with 122 archaeal marker genes. Taxa are shown at the phylum level, except for the Thermoproteota, Asgardarchaeota, Halobacteriota, and Methanobacteriota, shown at the class level. The number of MAGs in each highlighted taxon is shown in parentheses. See Table S2 for details. Quality is shown on the inner ring (green: high quality, purple: medium quality, with one manually curated Nanoarchaeota MAG below the 50% completion threshold also displayed as medium quality), while the outer ring displays normalized read coverage up to 200x. SH-like support values ≥80% are indicated with filled circles, and the scale bar represents 0.1 amino acid substitutions per site. The tree was artificially rooted with the Iainarchaeota,

Micrarchaeota, SpSt-1190, Undinarchaeota, Nanohaloarchaeota, EX4484-52, Aenigmarchaeota, Aenigmarchaeota_A and Nanoarchaeota using iTOL. The tree used to create this figure is available in Newick format (Data S5), and the formatted tree is publicly available on iTOL at https://itol.embl.de/shared/alrlab.

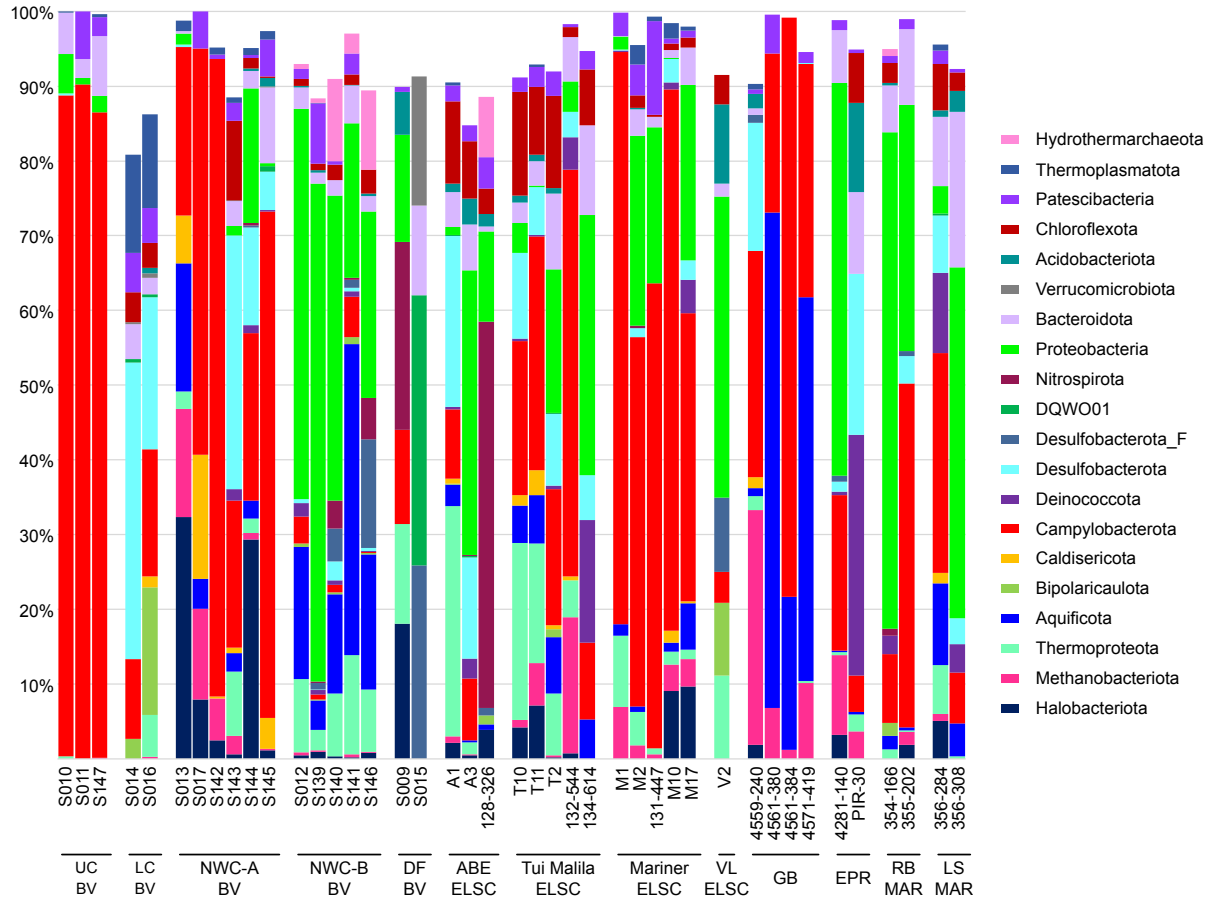


Fig. 3. Relative abundance of MAG phyla, based on normalized read coverage. The phyla shown comprise ≥10% of the MAG relative abundance in at least one metagenomic assembly. Read coverage was normalized to 100M reads per sample, and coverage values for MAGs were summed and expressed as a percent. UC, Upper Cone; LC, Lower Cone, NWC-A, Northwest Caldera Wall A; NWC-B, Northwest Caldera Wall B and Upper Caldera Wall; DF, diffuse flow; VL, Vai Lili; RB, Rainbow; LS, Lucky Strike.

While shared taxa differed in relative abundance and distribution, observable differences in community structure between vent fields were somewhat limited in this study due to small sample numbers from some of the vent fields (two samples apiece from EPR; Rainbow, MAR;

Lucky Strike, MAR), and the overall lower read depth of samples from these sites and a few other samples (Fig. S3). Therefore, obtaining statistically robust community structure patterns using MAG phylogenetic diversity for the entire dataset was not possible. However, Reysenbach et al. [8] did show that if metagenomic sequencing is deep, assembled MAG diversity tracks 16S rRNA amplicon diversity structure. Extrapolating to this study, the Brothers volcano MAG diversity patterns were retained and confirmed the amplicon observations from Reysenbach et al., 2020 [8] (Fig. S4), and in turn tracked the ELSC MAG community diversity (Fig. 4A, B). For example, sites at Brothers volcano that were hypothesized to have some magmatic inputs were predicted to be more similar in community structure to the sites along the ELSC with greater magmatic inputs, such as Mariner. Several of the samples from the more acidic Mariner vent field were more closely aligned in MAG diversity structure to those of the acidic solfataric Upper Cone sites at Brothers. The MAG data also demonstrated that the Guaymas samples were quite unique, which is not surprising, given that Guaymas Basin is a sediment-hosted system where the hydrothermal fluid geochemistry is quite different from other basalt- or andesitic-hosted hydrothermal systems (e.g., higher pH, high organics, high ammonia and methane) [19, 20].

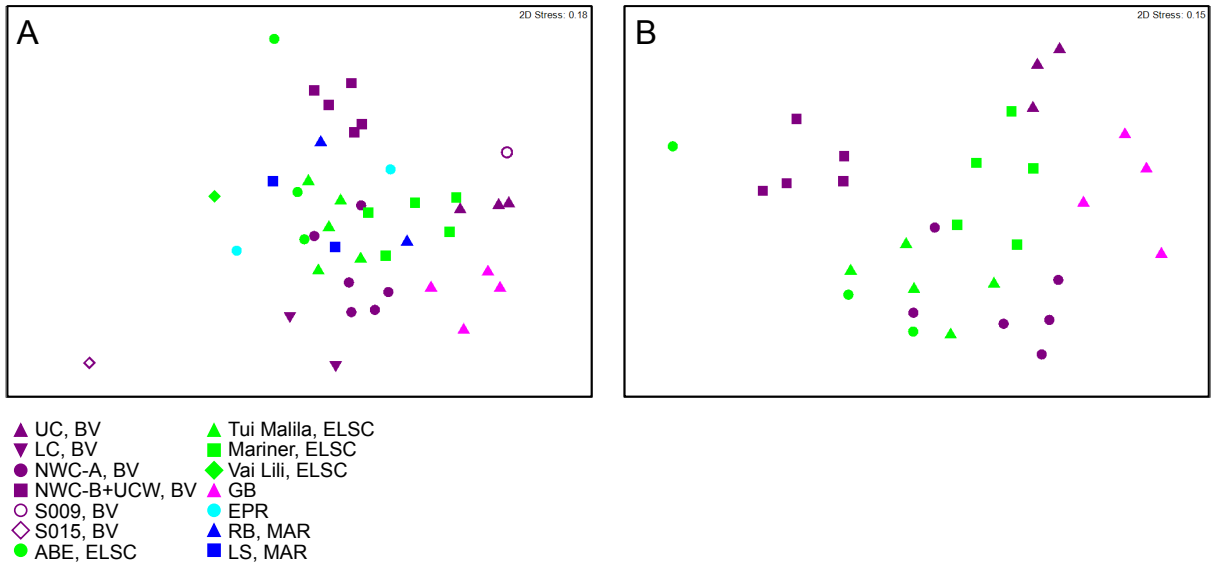


Fig. 4. Non-metric multidimensional scaling (NMDS) plots showing taxonomic diversity of MAGs. Plots depict (A) all samples in this study and (B) a subset of the data, limited to locations with three or more samples. Plots were generated using Bray-Curtis matrices of the relative abundance of GTDB taxa, based on normalized read coverage of medium- and high-quality MAGs (Table S4; set to 100M reads and expressed as a percentage of MAG read coverage per sample). Points that are closer together in the plots represent a higher degree of similarity.

Our dataset greatly broadens genomic diversity from deep-sea vents, by representing 511 novel and previously identified [8] genera, comprising 395 Bacteria and 116 Archaea. Notably, 52% (206) of these bacterial genera (Table S3A) and 72% (84) of archaeal genera (Table S3B) were found at Brothers volcano. Furthermore, 25% (99) of the recently identified bacterial genera and 47% (54) of the archaeal genera were unique to the Brothers volcano samples (Tables S3A, B), which further supports the understanding that this environment is a hotbed for novel microbial biodiversity, reflected in the volcano's complex subsurface geology [8].

While many of these novel archaeal and bacterial genera were previously reported from Brothers volcano [8], we report them again here in the context of the new data of the four deep-sea hydrothermal vent environments and the new assemblies (1000 bp contig cutoff, used for Brothers volcano samples and ELSC 2015 samples) and iterative DAS Tool binning used for all our metagenomes. Our data support that of Reysenbach et al., [8], which used MetaBAT for assemblies (2000 bp contig cutoff) of the Brothers volcano metagenomes. Namely, we recovered approximately 202 novel bacterial genera and 83 new archaeal genera from Brothers volcano communities in Reysenbach et al. [8], well within the range detected in this analysis (viz. 206 and 84, respectively). In this study, using a lower contig cutoff allowed for the recovery of a much higher number of MAGs, but many are of lower quality with higher contig counts. For example, MAGs recovered in the Reysenbach et al. [8] study had an average of 254 contigs per MAG, with ~19% (135) of MAGs comprising 100 contigs or less. In contrast, only 7% (258) of MAGs in this current study had 100 contigs or less, and the average number of contigs per

MAG was 511 (Table S2). However, using the iterative binning approach provided advantages when resolving lineages of high microdiversity, such as in the Nautiliales, with the caveat of creating some MAGs with large collections of erroneous contigs that were poorly detected by CheckM, as they had very few associated marker genes (e.g. MAGs 4571-419_metabat1_scaf2bin.008, M10_maxbin2_scaf2bin.065; Fig. S5). This points to the importance of carefully choosing assembly parameters depending on the ultimate goal of whether quality over quantity of MAGs is preferred for analyses of ecological patterns. Our data demonstrate, however, that overall patterns of MAG diversity are retained regardless of assembly techniques and parameters (Fig. S4).

Furthermore, here we document some of the first examples of medium to high quality MAGs from phyla and classes never previously identified, or poorly sampled, from deep-sea hydrothermal environments. These include Thermoproteia, Patescibacteria (formerly Candidate Phyla Radiation, CPR), Chloroflexota, and a few MAGs representing two putative new bacterial phyla, JALSQH01 (3 MAGs) and JALWCF01 (13 MAGs) (Supplementary Discussion, Fig. S6, Table S5). For example, with 249 MAGs belonging to the Thermoproteia (Table S2, Fig. S7), we have significantly expanded the known diversity and genomes from this phylum. The importance of this group at deep-sea vents was first recognized through 16S rRNA amplicon studies, where the depth of sequencing highlighted that much of this novel thermophilic diversity had been overlooked [e.g. 3, 4]. Furthermore, it is now recognized that many members of this group have several introns in the 16S rRNA gene, which explains why they were missed in original clone library assessments and may be underestimated in amplicon sequencing [21–24]. For example, 24 MAGs were related to a recently described genus of the Thermoproteia, *Zestosphaera* (GTDB family NBVN01) [24]. This genus was first isolated from a hot spring in New Zealand but is clearly a common member of many deep-sea vent sites. Further, the discovery of a 16S rRNA gene related to *Caldisphaera* at deep-sea vents [25], previously only detected in terrestrial

acidic solfataras, led to the isolation of related Thermoplasmata – *Aciduliprofundum boonei* – but the *Caldisphaera* escaped cultivation. Here we report several high-quality MAGs related to this genus (M2_metabat2_scaf2bin.319, 131-447_metabat1_scaf2bin.050, M1_metabat1_scaf2bin.025, S016_metabat2_scaf2bin.003). Additionally, we also recovered a genome from the Gearchaeales (S146_metabat1_scaf2bin.098), first discovered in iron-rich acidic mats in Yellowstone National Park [26], and members of the poorly sampled Ignicoccaceae, Ignisphaeraceae, and Thermofilaceae. While we identified several genomes from recently discovered archaeal lineages including the Micrarchaeota, Iainarchaeota and Asgardarchaeota, we also recovered 15 MAGs belonging to the Korarchaeia, 14 of which comprise two putative novel genera, and one which is closely related to a MAG previously recovered from sediment in Guaymas Basin (Genbank accession DRBY00000000.1) [27, 28]. Additionally, we recovered four MAGs from the Caldarchaeales that span two novel genera, one of which was recently proposed as *Candidatus* Benthortus lauensis [29] using a MAG generated from a previous assembly of the T2 metagenome (T2_175; Genbank accession JAHSRM000000000.1). MAGs belonging to this genus were identified at both Tui Malila, ELSC, and Brothers volcano (T2_metabat2_scaf2bin.284, S140_maxbin2_scaf2bin.281, S141_maxbin2_scaf2bin.262) with the Tui Malila MAG nearly identical (99.7% AAI similarity) to the described *Cand*. B. lauensis T2_175 MAG.

While within the Bacteria, the Gammaproteobacteria and Campylobacterota were by far the most highly represented bacterial genomes, there were other lineages for which we have very little if any data or cultures from deep-sea hydrothermal systems (Fig. 3, Fig. S7). Two such groups are the Patescibacteria and Chloroflexota, with 154 and 194 MAGs respectively.

Patescibacteria and Chloroflexota are diverse and abundant members of deep-sea hydrothermal vent deposits

The Patescibacteria/Candidate Phyla Radiation (CPR) encompasses a phylogenetically diverse branch within the bacterial tree of life that is poorly understood and rarely documented in deep-sea hydrothermal systems. Originally, the CPR was proposed to include several phylum-level lineages [30], but the entire group was later reclassified by GTDB as a single phylum, Patescibacteria [31]. Members of the Patescibacteria have been well-characterized in terrestrial soils, sediments, and groundwater [32–37], and in the mammalian oral cavity [38–40]. Several 16S rRNA gene and metagenomic studies have also identified members of the Patescibacteria from deep-sea vents, including EPR, MAR, ELSC, and Guaymas Basin [3, 4, 12, 15, 41–43], from Suiyo Seamount [44], and the Santorini submarine volcano [45], further supporting the widespread distribution of this metabolically diverse phylum.

Our study adds 56 novel genera based on AAI and GTDB classifications to the Patescibacteria phylum. These include large clades within the Gracilibacteria (10 new genera), representatives within the Microgenomatia (9 novel genera), Dojkabacteria (10 new genera), and several clades in the Paceibacteria (13 new genera) (Fig. 5A, B, Fig. S8). The Gracilibacteria and Paceibacteria were overall the most prevalent lineages of Patescibacteria in the samples but had contrasting distributions across vents (Fig. 5B). In general, when the Gracilibacteria were prevalent, the Paceibacteria appeared to be a minor component or not present, and vice versa. In particular, the Gracilibacteria MAGs were often associated with the acidic sites such as the Upper Cone at Brothers volcano (S011, S147), and the Mariner vent fields, and in the early colonization experiment from Guaymas Basin (Supplementary Discussion). This may suggest that Gracilibacteria function as early colonizers and are associated with turbulent ephemeral environments as observed previously in oil seeps [46]. Continued investigation into the ecology, evolution, and host association patterns of these groups, however, may shed more light on these distribution differences.

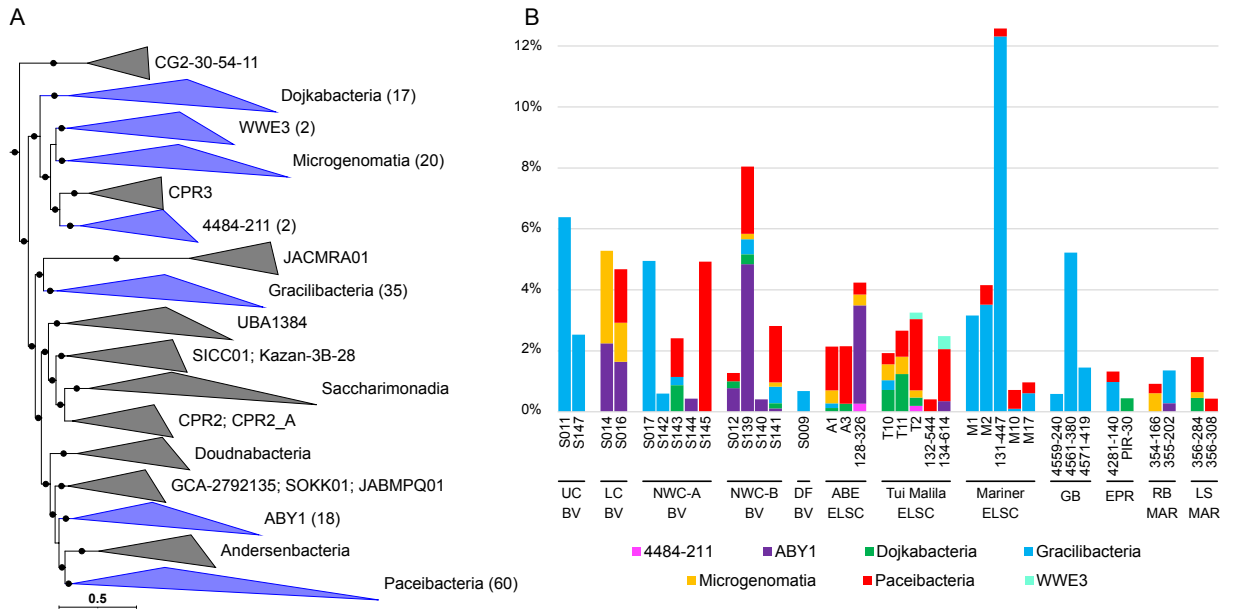


Fig. 5. Phylogenomic placement and relative abundance of Patescibacteria MAGs, displayed at the class rank. (A) Blue clades in the maximum-likelihood phylogenomic tree contain MAGs from this study, with the number of MAGs shown in parentheses. The scale bar shows 0.5 substitutions per amino acid, and filled circles indicate SH-like support (≥80%). (B) Relative abundance of Patescibacteria MAGs was calculated using normalized read coverage for MAGs in each assembly (set to 100M reads and expressed as a percentage of MAG read coverage per sample).

Consistent with previous studies [30, 34], many of the recovered Patescibacteria MAGs had very small genomes (often ~1MB or smaller; Table S2) with highly reduced metabolic potential, often lacking detectable genes for synthesis of fatty acids, nucleotides, and most amino acids (Table S6). Gene patterns also suggested that many of the organisms are obligate anaerobes, lacking aerobic respiration, and that they likely form symbiotic or parasitic associations with other microbes, as has been shown for Patescibacteria cultivated thus far from the Absconditabacterales and Saccharibacteria [39, 40, 47, 48].

We recovered several MAGs from Mariner, Guaymas Basin, and Brothers volcano that were related to the parasitic *Cand*. Vampirococcus lugosii [47] and *Cand*. Absconditicoccus praedator [48]. In order to explore if our MAGs had any hints of a parasitic lifestyle, we searched for some

of the large putative cell-surface proteins identified in the genomes of *Cand*. V. lugosii [47] and *Cand*. A. praedator [48]. Using a local BlastP of nine of the longest genes found in *Cand*. V. lugosii, we recovered high-confidence homologs (E-value=0) for alpha-2 macroglobulin genes in several MAGs from the Abscontitabacterales (based on search of *Cand*. V. lugosii protein MBS8121711.1), which may be involved in protecting parasites against host defense proteases [47]. We also recovered homologs for PKD-repeat containing proteins (MBS8122536.1; E-value=0), which are likely involved in protein-protein interactions [47]. Previous analysis of *Cand*. V. lugosii found these giant proteins are likely membrane-localized, suggesting they may potentially play a role in host/symbiont interactions. Additionally, we identified these long proteins from *Cand*. V. lugosii elsewhere in the Gracilibacteria MAGs. For example, putative homologs of the PKD repeat containing protein (MBS8122536.1), a hypothetical protein (MBS8121701.1), and the alpha-2 macroglobulin (MBS8121711.1) were identified in multiple other orders of the class Gracilibacteria (E-value ≤1E-25). The alpha-2 macroglobulin was also identified in the very distantly related Paceibacteria, and a single putative homolog of the alpha-2 macroglobulin was found in a MAG belonging to the class WWE3 (134-614_metabat1_scaf2bin.084; E-value ≤1E-24).

While the Patescibacteria likely rely on symbiotic or parasitic relationships, members of the Chloroflexota phylum are diverse and metabolically flexible organisms, capable of thriving in a wide variety of geochemical niches. Chloroflexota are abundant and widely distributed in a variety of environments, including terrestrial soils, sediments and groundwater, freshwater, pelagic oceans, and the marine subseafloor and sediments [49–55], and hydrothermal settings such as Guaymas Basin [11] and Brothers submarine volcano [8]. Genomic evidence suggests that Chloroflexota are associated with important metabolisms in the carbon cycle, including fermentation, carbon fixation, acetogenesis and the utilization of sugars, polymers, fatty acids, organic acids and other organic carbon compounds [50, 51, 54].

Here we add to the growing evidence that the Chloroflexota are diverse and metabolically versatile members of deep-sea hydrothermal vent communities. We recovered a total of 194 Chloroflexota MAGs spanning 12 orders (GTDB taxonomy), which included 22 novel genera. Of these novel genera, 14 were identified at Brothers volcano and 6 were unique to the Brothers volcano samples (Table S3A). Based on read coverage, Chloroflexota MAGs were in high relative abundance (≥7%) in several samples from the ELSC, namely, from Tui Malila and ABE, and in one NW Caldera Wall sample from Brothers volcano (Table S4). To further explore the metabolic potential of Chloroflexota in hydrothermal vent communities, we focused our analyses on ≥80%-completeness MAGs (≥80% completeness, *n*=58) distributed in 6 orders: Caldilineales, Promineofilales, Anaerolineales, Ardenticatenales, B4-G1, and SBR1031 (Fig. 6, Table S7A).

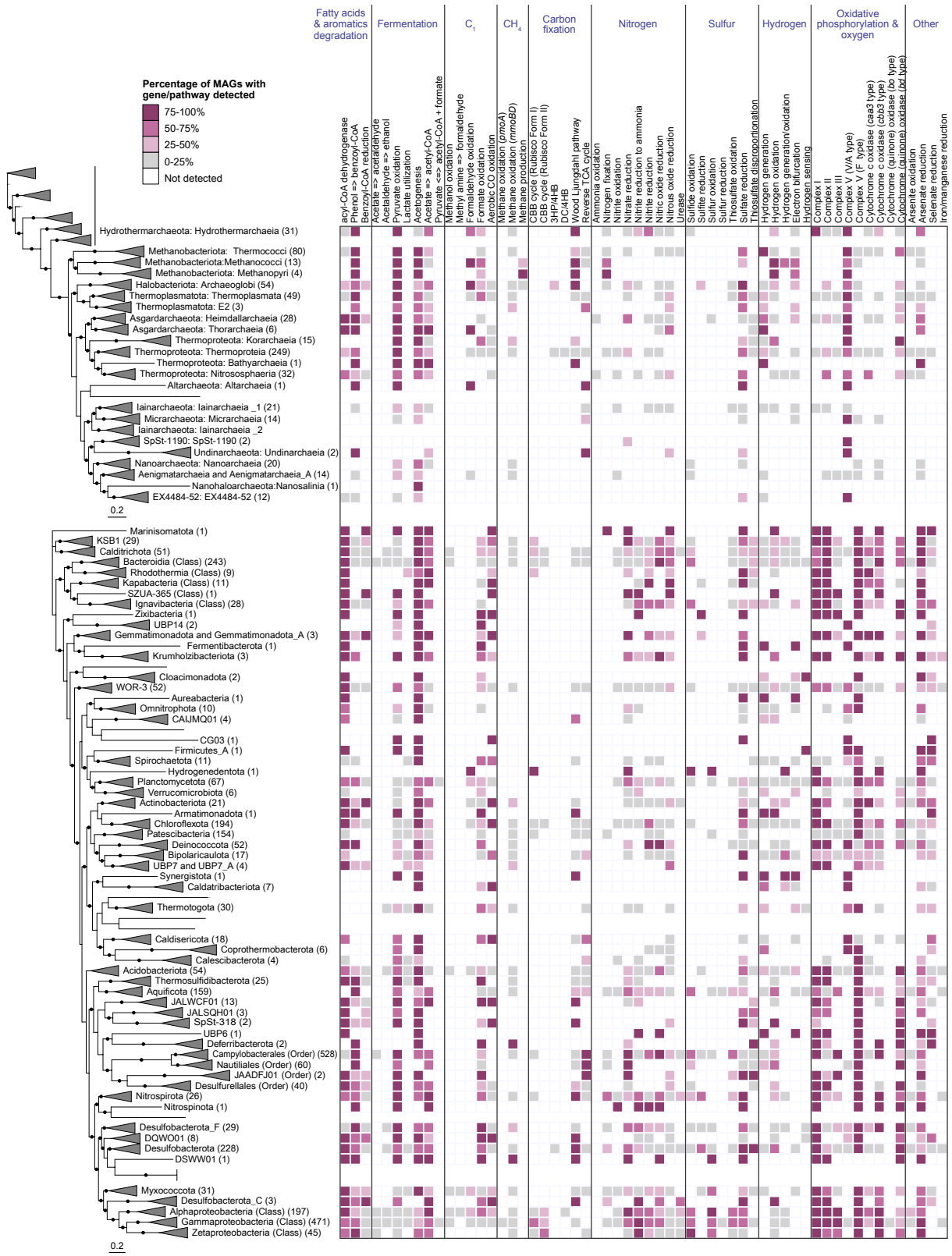
The majority (≥75%) of the ≥80%-completeness Chloroflexota MAGs encoded marker genes involved in several processes previously associated with the Chloroflexota (Table S7B), including fatty acid degradation [50, 55], formate oxidation [56], aerobic CO oxidation [57] and selenate reduction [53]. Except for the Anaerolineales, over 66% of the MAGs in the other five orders had the capacity for degradation of aromatic compounds, as previously reported for Chloroflexi from the marine subsurface [51]. While some MAGs had the potential for substrate-level phosphorylation through acetate formation, most of the MAGs contained pathways for oxidative phosphorylation and oxygen metabolism [50, 51]. The Wood–Ljungdahl pathway, the CBB cycle based on a Form I Rubisco, and the reverse TCA cycle were detected in some of the MAGs [50, 51]. Soluble methane monooxygenase genes, a metabolic potential recently also detected in a Chloroflexota MAG from the arctic [58], were identified in a total of eight of our MAGs from the orders Caldilineales, Anaerolineales, and Ardenticatenales.

Fig. 6. Phylogenetic tree of 58 ≥80%-completeness (≥80%) Chloroflexota MAGs with predicted functional capabilities. Nodes with ultrafast bootstrap support values ≥90% are shown with filled circles, and the scale bar shows 0.2 substitutions per site. One genome from the GTDB r202 database (GTDB accession GB_GCA_007123655.1) was used to re-root the tree. Hydrothermal vent fields: Brothers volcano (green), Eastern Lau Spreading Center (blue), East Pacific Rise (orange), Mid Atlantic Ridge (yellow)

Although the primary metabolic potential of the hydrothermal vent-associated Chloroflexota was in carbon cycling, we did, however, observe minor evidence for their roles in nitrogen and sulfur cycling (Fig. 6, Table S7). About 22% of the MAGs (with ≥80% completeness) encoded capacities for sulfide oxidation, as previously reported for members of this group, e.g. *Chloroflexus* spp. [59, 60]. The potential to disproportionate thiosulfate was also observed in a few MAGs. Further, thermophilic Chloroflexi grown in an enrichment culture from Yellowstone National Park were shown to oxidize nitrite. A few of our MAGs encoded genes involved in nitrite oxidation [61], while a larger proportion of the MAGs encoded genes for nitrite or nitric oxide reduction. None of the MAGs encoded complete pathways for entire sulfur oxidation or denitrification, suggesting that Chloroflexota in these environments may be associated with metabolic handoffs involving other community members (see below).

Metabolic and functional diversity in deep-sea hydrothermal vent deposits

In order to explore the metabolic and functional diversity associated with our MAGs, we utilized functional assignment results in tandem with the corresponding MAG relative abundance (Table S8). In general, genes involved in carbon, nitrogen, sulfur, and hydrogen metabolism were prevalent and shared across all hydrothermal systems in this study (Fig. 7, 8). While heterotrophy, autotrophy and mixotrophy potential were identified in all samples, 47.1% of the MAGs (by count) exhibited potential for carbon fixation. Marker genes associated with five different carbon fixation pathways were identified in the MAGs, namely, the Calvin-Benson-Bassham (CBB) cycle (form I or form II Rubisco), the 3-hydroxypropionate/4-hydroxybutyrate cycle, the dicarboxylate/4-hydroxybutyrate cycle, the reverse TCA cycle, and the Wood–Ljungdahl pathway (Fig. 7, 8). Marker gene presence also suggested the potential for widespread heterotrophic metabolism of peptides, polysaccharides, nucleotides and lipids, and fermentation via acetogenesis (Fig. 7, 8).


Percentage of MAGs with gene/pathway detected
75-100%
50-75%
25-50%
0-25%
Not detected
Fatty acids & aromatics degradation
Fermentation
C1
CH4
Carbon fixation
Nitrogen
Sulfur
Hydrogen
Oxidative phosphorylation & oxygen
Other
acyl-CoA dehydrogenase
Phenol => benzoyl-CoA
Benzoyl-CoA reduction
Acetate => acetaldehyde
Acetaldehyde => ethanol
Pyruvate oxidation
Lactate utilization
Acetogenesis
Acetate => acetyl-CoA
Pyruvate <=> acetyl-CoA + formate
Methanol oxidation
Methyl amine => formaldehyde
Formaldehyde oxidation
Formate oxidation
Aerobic CO oxidation
Methane oxidation (pmoA)
Methane oxidation (mmoBD)
Methane production
CBB cycle (Rubisco Form I)
CBB cycle (Rubisco Form II)
3HP/4HB
DC/4HB
Wood Ljungdahl pathway
Reverse TCA cycle
Ammonia oxidation
Nitrogen fixation
Nitrite oxidation
Nitrate reduction
Nitrite reduction to ammonia
Nitrite reduction
Nitric oxide reduction
Nitrous oxide reduction
Urease
Sulfide oxidation
Sulfite reduction
Sulfur oxidation
Sulfur reduction
Thiosulfate oxidation
Sulfate reduction
Thiosulfate disproportionation
Hydrogen generation
Hydrogen oxidation
Hydrogen generation/oxidation
Electron bifurcation
Hydrogen sensing
Complex I
Complex II
Complex III
Complex V (V/A type)
Complex V (F type)
Cytochrome c oxidase (caa3 type)
Cytochrome c oxidase (cbb3 type)
Cytochrome (quinone) oxidase (bo type)
Cytochrome (quinone) oxidase (bd type)
Arsenite oxidation
Arsenate reduction
Selenate reduction
Iron/manganese reduction
Hydrothermarchaeota: Hydrothermarchaeia (31)
Methanobacteriota: Thermococci (80)
Methanobacteriota:Methanococci (13)
Methanobacteriota: Methanopyri (4)
Halobacteriota: Archaeoglobi (54)
Thermoplasmatota: Thermoplasmata (49)
Thermoplasmatota: E2 (3)
Asgardarchaeota: Heimdallarchaeia (28)
Asgardarchaeota: Thorarchaeia (6)
Thermoproteota: Korarchaeia (15)
Thermoproteota: Thermoproteia (249)
Thermoproteota: Bathyarchaeia (1)
Thermoproteota: Nitrososphaeria (32)
Altarchaeota: Altarchaeia (1)
Iainarchaeota: Iainarchaeia _1 (21)
Micrarchaeota: Micrarchaeia (14)
Iainarchaeota: Iainarchaeia _2
SpSt-1190: SpSt-1190 (2)
Undinarchaeota: Undinarchaeia (2)
Nanoarchaeota: Nanoarchaeia (20)
Aenigmatarchaeia and Aenigmatarchaeia_A (14)
Nanohaloarchaeota:Nanosalinia (1)
EX4484-52: EX4484-52 (12)
0.2
Marinisomatota (1)
KSB1 (29)
Calditrichota (51)
Bacteroidia (Class) (243)
Rhodothermia (Class) (9)
Kapabacteria (Class) (11)
SZUA-365 (Class) (1)
Ignavibacteria (Class) (28)
Zixibacteria (1)
UBP14 (2)
Gemmatimonadota and Gemmatimonadota_A (3)
Fermentibacterota (1)
Krumholzibacteriota (3)
Cloacimonadota (2)
WOR-3 (52)
Aureabacteria (1)
Omnitrophota (10)
CAIJMQ01 (4)
CG03 (1)
Firmicutes_A (1)
Spirochaetota (11)
Hydrogenedentota (1)
Planctomycetota (67)
Verrucomicrobiota (6)
Actinobacteriota (21)
Armatimonadota (1)
Chloroflexota (194)
Patescibacteria (154)
Deinococcota (52)
Bipolaricaulota (17)
UBP7 and UBP7_A (4)
Synergistota (1)
Caldatribacteriota (7)
Thermotogota (30)
Caldisericota (18)
Coprothermobacterota (6)
Calescibacterota (4)
Acidobacteriota (54)
Thermosulfidibacterota (25)
Aquificota (159)
JALWCF01 (13)
JALSQH01 (3)
SpSt-318 (2)
UBP6 (1)
Deferribacterota (2)
Campylobacterales (Order) (528)
Nautiliales (Order) (60)
JAADFJ01 (Order) (2)
Desulfurellales (Order) (40)
Nitrospirota (26)
Nitrospinota (1)
Desulfobacterota_F (29)
DQWO01 (8)
Desulfobacterota (228)
DSWW01 (1)
Myxococcota (31)
Desulfobacterota_C (3)
Alphaproteobacteria (Class) (197)
Gammaproteobacteria (Class) (471)
Zetaproteobacteria (Class) (45)
0.2

393

Fig. 7. Core metabolic gene presence across phylogenetic clusters in deep-sea hydrothermal vent deposits. The number of MAGs in each clade is shown in parentheses, and MAGs belonging to unclassified lineages or falling outside their corresponding phylogenetic cluster due to unstable tree topology are shown without names. In instances where a phylum was not recovered as a monophyletic lineage within the tree (e.g. Iainarchaeia), MAG count and gene distribution for the entire phylum is only shown on one of the branches. Unless otherwise indicated, archaeal clades are shown at the class level, while bacterial clades are shown at the phylum level. Nodes with ultrafast bootstrap support ≥90% are shown with filled circles, and scale bars indicating 0.2 amino acid substitutions per site are provided for both archaeal and bacterial trees. Detailed metabolic gene presence information can be found in Table S9.

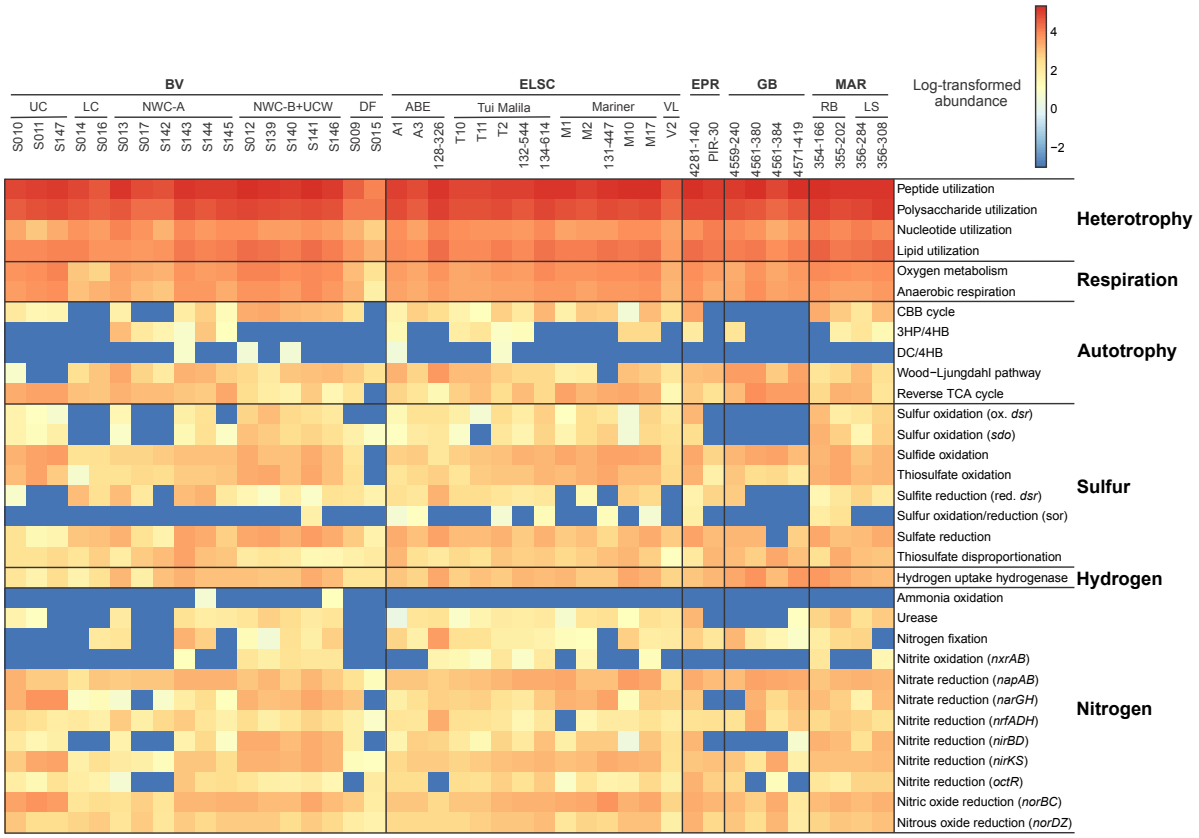


Fig. 8. Heatmap displaying the metabolic potential for each metagenome. Within each metagenomic dataset, functional abundance values were calculated as described in the methods. Functional abundances were then log-transformed, with abundance values equal to zero replaced by 10^{-3} to avoid negative infinite values.

Genes involved in nitrogen fixation, denitrification, and nitrite oxidation were identified across the different hydrothermal sites, yet the potential for anaerobic or aerobic ammonia oxidation

was rarely detected (Fig. 8). The absence of ammonia oxidation is not totally surprising, since ammonia is in very low to undetectable concentrations in deep-sea hydrothermal fluids, with the exception of sediment-hosted hydrothermal areas like at Guaymas Basin [19, 20]. In these sedimented hydrothermal systems, aerobic and anaerobic ammonia oxidation are key processes within the sediments and hydrothermal plumes [62–65], but they may not be as important in the hydrothermal deposits. Our data also expands the importance of nitrogen fixation from the first detection at deep-sea vents in *Methanocaldococcus* [66] to a greater diversity of hydrothermal Bacteria and Archaea.

Given the importance of sulfur cycling in deep-sea hydrothermal systems [67–69], it is not surprising that genes associated with elemental sulfur, sulfide, and thiosulfate oxidation, sulfate reduction, and thiosulfate disproportionation were widely distributed in MAGs from different hydrothermal samples and were associated with diverse taxonomic guilds (Fig. 7, 8). Based on metabolic gene distribution statistics (Table S9), the potential for sulfur oxidation was identified in 16% of the MAGs (577), primarily in members of the Alphaproteobacteria and Gammaproteobacteria. Genes associated with sulfide oxidation were identified in 34% of the MAGs (1216), including members of the Bacteroidia, Campylobacteria, Alphaproteobacteria, and Gammaproteobacteria. Thiosulfate oxidation genes were detected in 23% of the MAGs (836), largely comprised of the Campylobacteria, Alphaproteobacteria, and Gammaproteobacteria, while 14% of the MAGs (522) encoded genes for thiosulfate disproportionation, including the classes Bacteroidia, Campylobacteria and the phylum Desulfobacterota. The potential for dissimilatory sulfite reduction was identified in 6% of the MAGs (220) distributed across ten bacterial and archaeal phyla, namely Halobacteriota (class Archaeoglobi), Bacteroidota (class Kapabacteria), Campylobacterota (class Campylobacterales), Zixibacteria, Gemmatimonadota, Acidobacteriota, Nitrospirota, Desulfobacterota, Desulfobacterota_F and Myxococcota.

Hydrogen is highly variable in hydrothermal fluids, with some of the highest concentrations in geothermal systems hosted by ultramafic rocks, such as the Rainbow hydrothermal vent field [3], or in sediment-hosted regions like Guaymas basin [70]. In these systems, methanogens and sulfate reducers are prevalent hydrogen consumers [3, 71–74], although a wide variety of other heterotrophs and autotrophs can also derive energy from hydrogen oxidation [72]. Hydrogenase enzymes are responsible for mediating hydrogen oxidation in microbial populations but are also involved in a variety of other functions, including hydrogen evolution, electron bifurcation, and hydrogen sensing [75]. Approximately 27% of the MAGs in this study (974) encoded for at least one hydrogenase gene, and the MAGs were predominantly associated with the classes Campylobacteria, Bacteroidia, Gammaproteobacteria, and the phylum Desulfobacterota (Fig. 7, 8, Table S9). In several cases (132 MAGs), hydrogenase genes co-occurred with genes involved in the oxidation of reduced sulfur species (sulfide, elemental sulfur, sulfite or thiosulfate). This is not surprising, given that the capability to oxidize both sulfur and hydrogen has been shown in multiple isolates, including members of the Campylobacteria [76–78] and Aquificae [e.g. 79, 80].

Metabolic handoffs are a central feature of community interactions in deep-sea hydrothermal vent deposits

The microbial communities at deep-sea hydrothermal vents are shaped by a wide variety of complex interactions, including symbiosis, syntrophy, commensalism, cross-feeding, and metabolic handoffs [11, 12, 81–83]. While many of the MAGs encode genes associated with different biogeochemical cycles, as expected, the genes for a complex functional pathway often were not localized in a single MAG, but instead distributed across several MAGs. This is likened to ‘metabolic handoffs’ where the interaction between different organisms produces pathway intermediates, enabling community members to perform downstream reactions in the metabolic

pathway. For example, metagenomic analysis of a subsurface aquifer environment suggested that metabolic handoffs are commonly utilized in key biogeochemical pathways such as sulfide oxidation and denitrification [37]. Genes for sulfide oxidation were identified in all the deep-sea hydrothermal vent sites in this study, but few MAGs encoded genes for the entire three-step pathway. A much larger proportion of the MAGs, however, contained genes for a single step in sulfur oxidation (Fig. 9), consistent with a metabolic handoff scenario. Similar patterns were also

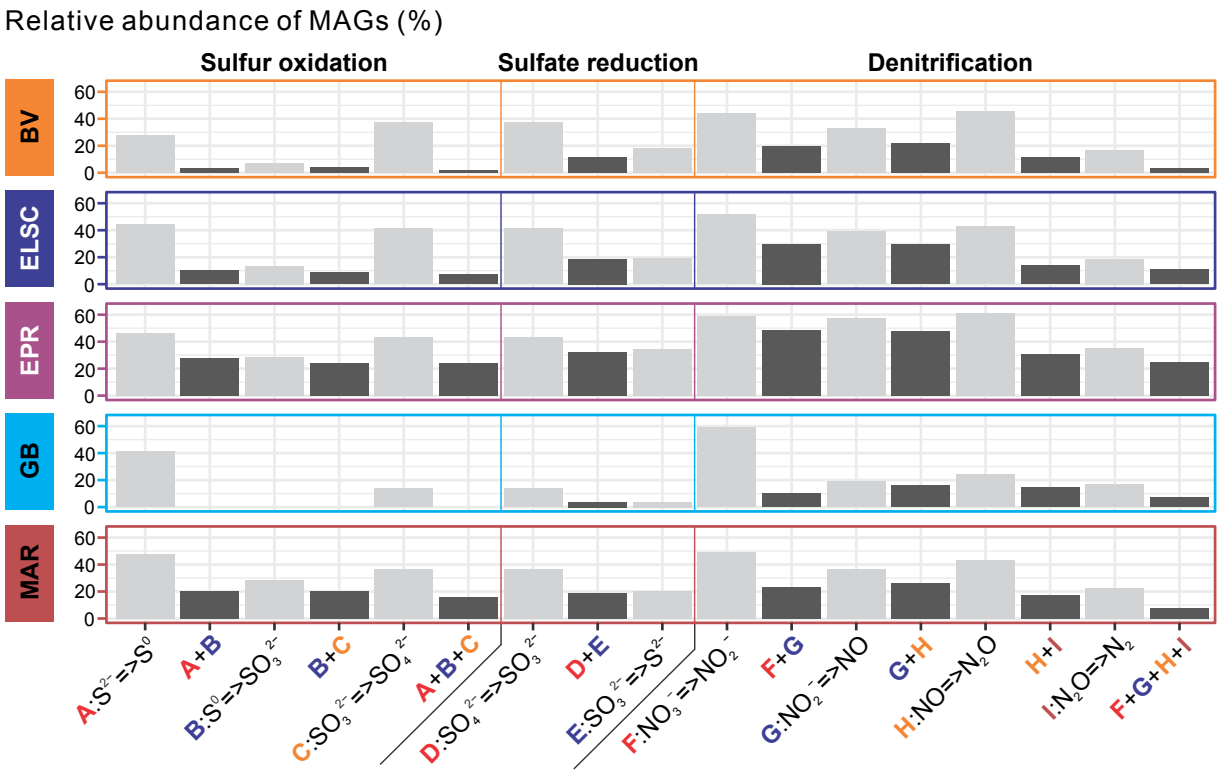


Fig. 9. Bar plots showing the sequential steps of sulfur oxidation, denitrification and sulfate reduction. Bar height indicates the percent relative abundance of MAGs in each metagenome with genes for a particular function(s), averaged across hydrothermal vent sites.

observed for sulfate reduction and denitrification (Fig. 9). Additionally, the genes for individual steps in sulfide oxidation were often found coupled with at least one gene from the denitrification pathway, which may increase the thermodynamic favorability of both pathways. Furthermore,

one or more denitrification genes co-occurred with sulfide oxidation genes in 1113 MAGs, with elemental sulfur oxidation genes in 485 MAGs and with sulfite oxidation genes in 1025 MAGs (Table S9). We recognize that some of these observations may be attributed to the incompleteness of the MAGs, however, our observations are in line with similar findings from other environments such as the terrestrial subsurface [37].

Conserved microbial functions are mediated by different taxa at different hydrothermal vent systems

Previous analyses of deep-sea hydrothermal environments and global oceans have pointed to widespread functional redundancy in microbial communities [8, 12, 84, 85], with similar metabolic potential identified across taxonomically diverse samples. For example, a study of Guaymas Basin metagenome-assembled genomes suggested that many functional genes could be identified across multiple distinct taxa [12]. In our study, members of the Campylobacteria and Gammaproteobacteria were present in almost all samples, yet showed contrasting patterns of abundance (Fig. 10). These lineages can perform several of the same functional processes including oxidation of reduced sulfur species [86], denitrification [87–89], and carbon fixation [90–93]. This can be partially explained by ecophysiological and growth differences between the groups, which are selected for by the different geochemical profiles at the various vent sites. For example, studies have suggested that Campylobacteria tend to favor higher sulfide conditions but have a broader range of oxygen tolerance than the Gammaproteobacteria, while Gammaproteobacteria tend to inhabit a narrower range of higher oxygen and lower sulfide [16, 86, 90, 94]. It is therefore not surprising that the Campylobacteria were more prevalent at several of the acidic and more turbulent sites, such as at the Upper Cone, Brothers volcano and in early colonized samples from a thermocouple array at Guaymas Basin (Table S4, Supplementary Discussion). Patwardhan et al. [95] also showed that Campylobacteria were

early colonizers of shallow marine vents followed by Gammaproteobacteria, and their differential colonization could be linked to sulfide, oxygen, and temporal differences.

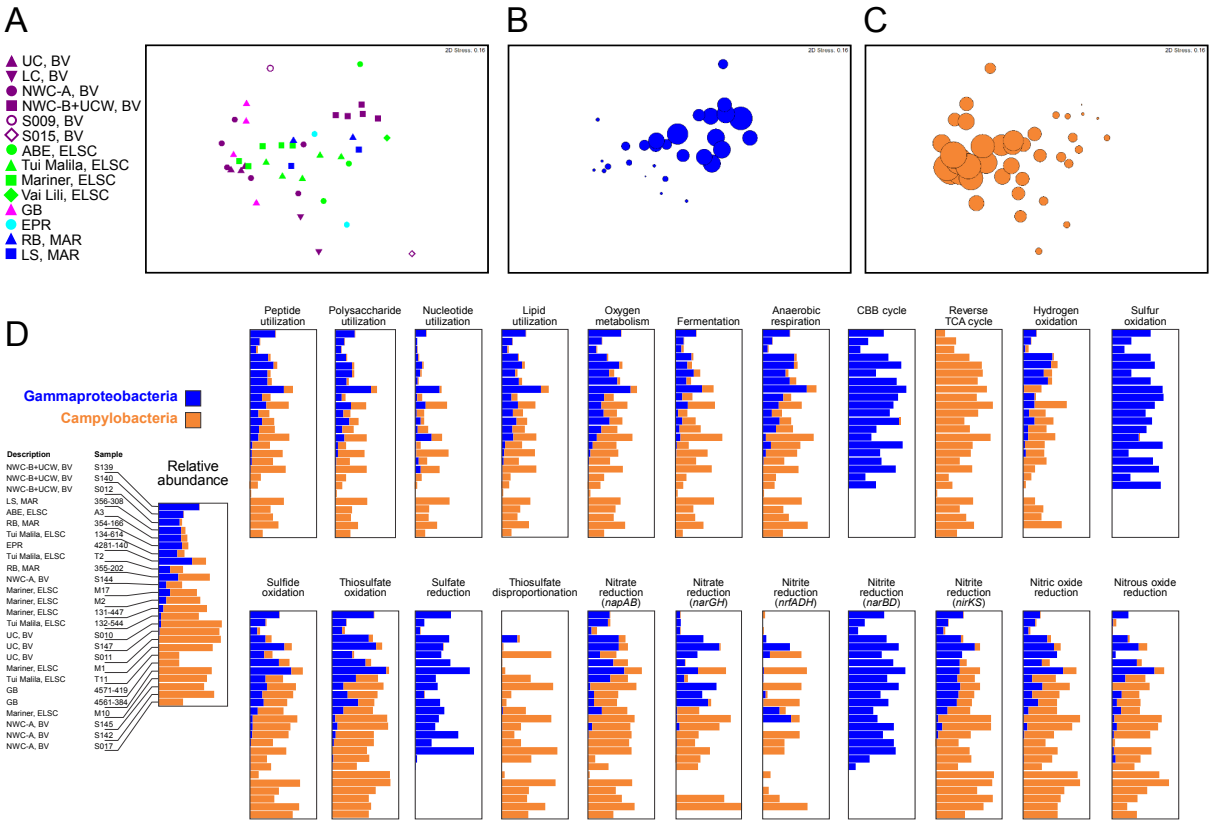


Fig. 10. Comparative taxonomic and functional gene abundance of the Campylobacteria and Gammaproteobacteria. NMDS plots were generated using a Bray-Curtis matrix of relative MAG abundance, based on GTDB-assigned taxonomy at the class level. Plots are shown for (A) all sample sites, and for all sample sites with bubbles proportional to the relative abundance of (B) Gammaproteobacteria and (C) Campylobacteria. Comparative functional distribution (D) is also shown for the Gammaproteobacteria and Campylobacteria for the 26 samples that had a summed relative abundance of both Gammaproteobacteria and Campylobacteria of ≥30%. The 22 functions depicted were selected as the Gammaproteobacteria and Campylobacteria accounted for an average of ≥20% of the total abundance for each function across the metagenomes.

516 The covariation of the Campylobacteria and Gammaproteobacteria in our data also coincided
517 with genes for key functional processes associated with these taxa (Fig. 10). Thus, the overall
518 ecological function contributed by the Campylobacteria and Gammaproteobacteria to the
519 community at all sites was similar, but carried out by either one; viz., same guild different taxa.
520 For example, relative gene abundance of individual functions tracked the relative abundance of
521 Campylobacteria and Gammaproteobacteria for 15 of 22 broadly distributed functions, including
522 heterotrophy associated with various organic carbon compounds, respiration of oxygen and
523 nitrogen compounds, and oxidation of reduced sulfur compounds. However, genes for some
524 functions were exclusively represented by either group (Fig. 10, Table S10). For example,
525 marker genes for formaldehyde oxidation, urea utilization and elemental sulfur oxidation were
526 found in the Gammaproteobacteria but were hardly detected in Campylobacteria, while genes
527 associated with thiosulfate disproportionation were attributed almost exclusively to
528 Campylobacteria (Fig. 10, Table S10). In some cases, metabolic analysis also suggested that
529 both Campylobacteria and Gammaproteobacteria had similar metabolic capabilities but
530 encoded different pathways for the same functions. For example, consistent with a previously
531 observed but non-ubiquitous trend [90–93], Campylobacteria mostly encoded genes for the
532 rTCA cycle while the Gammaproteobacteria encoded genes for the CBB cycle. Both taxa also
533 showed the potential for nitrite reduction to ammonia, with more *nrfADH* genes identified in the
534 Campylobacteria and *nirBD* only found in the Gammaproteobacteria.
535

536 **Conclusions**

537 From a comparative metagenomic analysis of 38 deep-sea hydrothermal deposits from multiple
538 globally distributed sites, we provide insights into the shared vent-specific lineages and greatly
539 expand the genomic representation of core taxa that have very few, if any, examples in
540 cultivation. Furthermore, we document many novel high quality assembled genomes that were
541 originally only identified from deep-sea vents as 16S rRNA genes. This study sheds light on the

metabolic potential and physiological ecology of such taxa. We show that overall, the different communities share similar functions, but differences in the environmental geochemistry between sites select distinct taxonomic guilds. Further, metabolic handoffs in communities provide functional interdependency between populations achieving efficient energy and substrate transformation, while functional redundancy confers higher ecosystem resiliency to perturbations and geochemical fluctuations. In summary, this study provides an integrated view of the genomic diversity and potential functional interactions within high temperature deep-sea hydrothermal deposits and has implications on their biogeochemical significance in mediating energy and substrate transformations in hydrothermal environments.

Methods

Sample collection, DNA extraction and sequencing

High-temperature, actively venting deep-sea hydrothermal deposits, a diffuse flow sample, and a water sample were collected from Brothers volcano (2018), the Eastern Lau Spreading Center (2005 and 2015), Guaymas Basin (2009), the Mid-Atlantic Ridge (2008), and the East Pacific Rise (2004 and 2006) as previously described (Flores et al., 2012a, Reysenbach et al., 2020). Expedition details, including identification numbers, research vessels and submersibles utilized for sampling are described in Table S1. Samples were processed [4] and DNA extraction was performed as previously described [4, 8, 25, 96].

Thermocouple array from Guaymas Basin

The thermocouple array experimental set up from Guaymas Basin in 2009 is described in Teske et al., 2016 [20].

Metagenomic assembling and binning

Reads from Brothers volcano and ELSC (2015) were quality filtered using FastQC v.0.11.8 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and *de novo* assembled using metaSPAdes v.3.12.0 [97] with the settings "-k 21,33,55,77,99,127 -m 400 --meta". Reads from ELSC (2005), MAR, EPR and Guaymas Basin were assembled by the Department of Energy, Joint Genome Institute (JGI) using metaSPAdes v.3.11.1 with the settings "-k 33,55,77,99,127 --only-assembler --meta". Individual assemblies were generated for each metagenomic dataset. MetaWRAP v.1.2.2 [98] was used to generate metagenome-assembled genomes (MAGs) from each assembly with the settings "--metabat2 --metabat1 --maxbin2". DAS Tool v.1.0 [99] was then applied to screen the three sets of MAGs generated by MetaWRAP, resulting in consensus MAGs with a minimum scaffold length of 1000 bp.

Metagenome-assembled genome curation and quality assessment

CheckM v.1.0.7 [100] was used to assess MAG quality and screen for the presence of 16S rRNA genes. Erroneous SSU genes were then removed using RefineM v.0.0.20 [101], which was also used to identify and remove outlier scaffolds with abnormal coverage, tetranucleotide signals, and GC patterns from highly contaminated MAGs. GTDB-Tk v.1.5.0, data release 202 [17] was used to assign taxonomy to each MAG with default settings. SSU sequences from each MAG were then re-parsed and annotated by SINA v.1.2.11 [102]. Scaffolds containing 16S rRNA gene sequences inconsistent with GTDB taxonomic classifications were deemed contaminants and were removed. Selected MAGs were then further refined and manually inspected by VizBin v.1.0.0 [103]. Final MAGs had an estimated ≥50% genome completion and ≤10% contamination, with completeness and contamination rounded to the nearest whole number.

Iterative Nanoarchaeota MAG curation

As a case study, two MAGs assigned to the Nanoarchaeota (4571-419_metabat1_scaf2bin.008, M10_maxbin2_scaf2bin.065) were iteratively curated, demonstrating that the original MAGs generated by DAS Tool contained large quantities of contaminant contigs that were not recognized by CheckM, given the low abundance of marker genes. Each MAG was visualized using the Anvi'o v.7.1 interactive interface [104], where contigs were divided into subsets based on clustering patterns in Anvi'o. Contigs in each cluster were assigned a putative taxonomy using the Contig Annotation Tool (CAT) [105]. Clusters containing most of the contigs assigned to the Nanoarchaeota were repeatedly sub-sampled and screened using the CAT pipeline until no meaningful correspondence between clustering patterns and assigned taxonomy could be identified (Fig. S5). Contigs in the final clusters were then removed if CAT definitively assigned them to a taxonomic group outside the Nanoarchaeota, while contigs assigned to the Nanoarchaeota and unclassified higher ranks were retained. A third Nanoarchaeota MAG (4281-140_maxbin2_scaf2bin.078) was also identified, but attempted curation using the above workflow revealed the presence of extensive contamination, with only a very small subset of scaffolds confidently assigned to the Nanoarchaeota. CAT analysis of a putative Nanoarchaeota MAG (JGI Bin ID 3300028417_39) separately assembled from the same read set by the JGI as part of the Genomes from Earth's Microbiomes project [106] also showed very few contigs assigned to the DPANN superphylum and extensive bacterial contamination, suggesting that this particular read set may represent a challenge for commonly utilized binning algorithms. Given the extensive contamination and difficulty identifying a valid Nanoarchaeota MAG of significant size, the 4281-140_maxbin2_scaf2bin.078 was excluded from the MAG dataset submitted to Genbank, so as to avoid contaminating the public database with erroneous information. However, the MAG was included in functional and relative abundance calculations.

MAG characterization and annotation

Open reading frames (ORFs) were predicted by Prodigal v.2.6.3 [107] with the parameter "-p meta". ORFs were then annotated by KOfam [108] and custom HMM profiles within METABOLIC v.4.0 [109] and eggNOG-emapper v.2.1.2 [110] with default settings. Transfer RNAs were predicted using tRNAscan-SE 2.0 using the general tRNA model [111]. Genomic properties, including genome coverage, genome and 16S rRNA taxonomy, tRNAs, genome completeness and scaffold parameters were parsed from results that were calculated by CheckM, tRNAscan-SE 2.0 and METABOLIC. Relative genome coverages were normalized by setting each metagenomic dataset size as 100M paired-end reads.

Prior to detailed metabolic analysis, open reading frames from the Gracilibacteria orders BD1-5 and Absconditabacterales, which are known to use genetic code 25 [e.g. 47, 48, 112, 113], were re-called using Prodigal v.2.6.3 as implemented in Prokka v.1.14.6 [114]. An additional MAG from the Gracilibacteria order GCA-2401425 (4559-240_metabat1_scaf2bin.085) was also processed using genetic code 25. Currently, the only other genome in GTDB order GCA-2401425 (Genbank accession NVTB00000000.1) [115] is publicly available in Genbank with ORFs generated using genetic code 11. However, comparative analysis of our GCA-2401425 MAG showed that ORFs called with genetic code 11 were truncated, with an average length of approximately 85 amino acids, while those called with genetic code 25 averaged 277 amino acids in length. ORFs from two additional MAGs from the Paceibacteria (A3_metabat2_scaf2bin.333 and S145_metabat2_scaf2bin.004) were also re-generated in Prokka using genetic code 11. Open reading frames were then annotated in GhostKoala [116].

Phylogenomic inference

For archaeal phylogenomic tree construction, a concatenated multiple sequence alignment (MSA) was generated in GTDB-Tk using 122 archaeal marker genes (2991 sequences, 5124 columns) [17]. IQ-TREE v.1.6.9 [117] was used to reconstruct the tree with the settings "-m MFP

644 -bb 1000 -redo -mset WAG,LG,JTT,Dayhoff -mrate E,I,G,I+G -mfreq FU -wbtl" (Data S1). The
645 bacterial phylogenomic tree was constructed in a similar manner, using a concatenated MSA of
646 120 bacterial GTDB marker genes [17]. For each GTDB bacterial phylum, no more than 15
647 reference genomes from the GTDB r202 database were used (4248 sequences, 5037 columns;
648 Data S2). Additionally, a second bacterial phylogenomic tree was inferred from the same MSA
649 using FastTree v.2.1.8 (WAG, +gamma, SH support; Data S3) [118]. Additional MSAs solely
650 using MAGs from this study were generated for the Archaea (122 marker genes) and Bacteria
651 (120 marker genes) using the GTDB-Tk identify and align commands [17]. FastTree v.2.1.10
652 (parameter: --gamma) was used to infer the phylogenomic trees, as implemented in GTDB-Tk
653 (Data S4, S5; formatted trees available online at https://itol.embl.de/shared/alrlab).
654
655 A tree was constructed in GTDB-Tk (parameter: --gamma) using MAGs assigned to the
656 Patescibacteria, along with recently described *Cand*. Vampirococcus lugosii [47] and *Cand*.
657 Absconditicoccus praedator [48], and the GTDB r202 bacterial tree-building dataset. A
658 phylogenomic tree of the Chloroflexota was also generated by extracting a concatenated MSA
659 of Chlorofexota MAGs from the entire bacterial MSA. IQ-TREE v.2.1.4 [119] was used to
660 reconstruct the tree with the settings "-m TESTMERGE -bb 1000 -bnni". An outgroup genome
661 (GCA_007123655.1) was added to reroot the phylogenomic tree. Final trees were visualized
662 using Interactive Tree of Life (iTOL) v.6 [120].
663
664 **Taxonomic assignment**
665 Initial taxonomy was assigned to each MAG using the GTDB-Tk classify pipeline. In rare
666 instances where there were discrepancies between the class-level (Archaea) or phylum-level
667 taxonomy (Bacteria) assigned by GTDB-Tk and phylogenetic tree topology, we deferred to tree
668 topology. In the Bacteria, topological taxonomic assignments were only used if confirmed by
669 both trees. MAGs that were not assigned to a known genus by GTDB-Tk were compared to

their closest relatives in this study using average amino acid identity (AAI) matrices generated in CompareM v.0.1.2 (https://github.com/dparks1134/CompareM). MAGs were assigned to novel genera using cutoffs provided by Konstantinidis et al. [18], and MAGs assigned the taxonomic status “unclassified” were automatically assigned to a novel genus.

Trophic and energy metabolism analysis

Functional genes were first characterized by METABOLIC [109]. Additional peptide utilization genes were characterized using the MEROPS database release 12.3 [121], and additional polysaccharide utilization genes were identified using dbCAN2 (2020-04-08) and the CAZy (2021-05-31) database [122, 123]. Cellular localization of peptidases/inhibitors, gene calls identified by the CAZy database, and predicted extracellular nucleases were verified using PSORTb v.3.0 [124]. Functional annotations for protein, polysaccharide, nucleic acid and lipid utilization were derived in part from previous publications [125, 126]. Iron cycling genes and hydrogenase genes were characterized based on HMMs directly obtained or indirectly parsed from FeGenie [127] and HydDB [75].

For each of these trophic and energy metabolisms, the number of functional gene calls in each genome were calculated using two different scenarios: 1) the presence of any marker gene in the complex/pathway was treated as the presence of the whole function (indicated as C), and the highest number of gene calls for an individual gene in the complex was taken to be the number of pathway ‘hits’ in the MAG. 2) Stand-alone genes that were not part of a large complex or functional pathway (indicated as A) were treated as individual accumulative gene calls for their particular function. In specific cases, marker genes were manually verified using phylogenetic trees and by inspecting operon arrangements (see below). To calculate functional abundance, all genomes were included in the analysis. Functional abundance was then calculated by multiplying normalized genome coverage (100M reads/sample) by the number of

functional gene calls for each sample. For visualization, functional abundance was then log-transformed and used to generate heatmaps with the R package pheatmap v.1.0.12 (settings: clustering_method = ward.D2). Combined functional heatmaps were also generated by summing values within larger functional groups.

To avoid potential mis-annotation by the automated methods described above, phylogenetic trees were constructed to validate predicted protein sequences for dissimilatory sulfite reductase (Dsr; Fig. S9), methyl-coenzyme M reductase subunit alpha (McrA; Fig. S10) and sulfur dioxygenase (Sdo; Fig. S11). Based on current understanding, two metabolic directions are possible for the Dsr protein: reductive Dsr, which catalyzes the reduction of sulfite to sulfide, and oxidative (or reverse) Dsr, which converts elemental sulfur oxidation to sulfite [128]. Paired DsrAB proteins were first identified in all MAGs using in-house Perl scripts. In cases where Dsr subunits were duplicated, one set of paired DsrAB proteins was manually selected. A concatenated protein alignment was then generated for DsrAB proteins from the MAGs and reference sequences using MAFFT v.7.310 [129], and the alignment was trimmed using trimAl v.1.4.rev15 [130] with the parameter "-gt 0.25". A phylogenetic tree was then constructed in IQ-TREE with settings "-m MFP -bb 1000 -redo -mset WAG,LG,JTT,Dayhoff -mrate E,I,G,I+G -mfreq FU -wbtl" (Fig. S9). Reductive and oxidative DsrAB proteins were identified based on placement in the phylogenetic tree.

Predicted proteins for McrA were first identified using the TIGR03256 HMM. Presumed false gene calls were then manually removed, including those identified in bacterial MAGs and non-methanogenic/anaerobic methanotrophic archaeal MAGs with high sequence coverage. An alignment was constructed in MAFFT v.7.310 [129] using the remaining McrA protein sequences, together with reference genes recovered from methanogens, anaerobic methanotrophs, and short-chain alkane oxidizing Archaea from the Bathyarchaeia,

Helarchaeales, *Syntrophoarchaeum* and *Polytropus* [11, 12, 131, 132]. Alignment trimming and phylogenetic tree inference were performed as described above.

Sulfur dioxygenase (Sdo) proteins were predicted using the “sulfur_dioxygenase_sdo” HMM [109]. Alignment, trimming and construction of the phylogeny were performed as described above. Positive Sdo calls were identified using two conserved amino acid residues (Asp196 and Asn244 of hETHE1, NCBI accession NP_055112) that are specific to Sdo in comparison with other metallo-β-lactamase superfamily members [133].

Statistical analysis

The relative abundance of MAGs in this study was calculated for each sample using normalized read coverage (set to 100M reads) expressed as a percentage. Bray-Curtis similarity matrices were then generated from relative abundance data at various taxonomic ranks, and nonmetric multidimensional scaling (NMDS) plots were generated from the matrices using PRIMER v.6.1.13 [134].

Acknowledgements

We thank the crew of the R/V *Roger Revelle*, R/V *Atlantis*, R/V *Thomas G. Thompson*, HOV *Alvin*, and the ROV *Jason* for assistance in collecting the samples. Many thanks to the many students who over the years helped extract the DNA, Nicole Wagner and Jennifer Meneghin for initial bioinformatic analysis assistance and to MK Tivey for thoughtful comments to the manuscript.

Declarations

Ethical approval

Not applicable.

Consent for publication

No applicable

Competing Interests

The authors declare that they have no competing interests.

Authors' contributions

A-L.R conceived of the study, collected and processed the samples and wrote the manuscript, Z.Z. and E.S.J. did the bioinformatic processing, data analysis and generated the figures and tables, K.A. assisted in project conception and data analysis, all authors read, reviewed and edited the manuscript.

Funding

This work was funded by the US-National Science Foundation grants OCE-0728391, OCE-0937404, OCE-1558795 to A-L.R, and OCE-2049478 and DBI-2047598 to K.A. We thank the Department of Energy Joint Genome Institute (Community Science Program award 339, lead Peter Girguis) for sequencing several of the samples.

Availability of data & materials

Metagenome reads are publicly available in the Sequence Read Archive (Table S1), and MAGs generated in this study are available in NCBI Genbank (BioProject PRJNA821212, Table S2).

References

773 1. Nakagawa S, Takai K, Inagaki F, Chiba H, Ishibashi JI, Kataoka S, et al. Variability in
774 microbial community and venting chemistry in a sediment-hosted backarc hydrothermal system:
775 impacts of subseafloor phase-separation. FEMS Microbiol Ecol. 2005;54:141–55.
776 2. Nunoura T, Takai K. Comparison of microbial communities associated with phase-separation-
777 induced hydrothermal fluids at the Yonaguni Knoll IV hydrothermal field, the Southern Okinawa
778 Trough. FEMS Microbiol Ecol. 2009;67:351–70.
779 3. Flores GE, Campbell JH, Kirshtein JD, Meneghin J, Podar M, Steinberg JI, et al. Microbial
780 community structure of hydrothermal deposits from geochemically different vent fields along the
781 Mid-Atlantic Ridge. Environ Microbiol. 2011;13:2158–71.
782 4. Flores GE, Shakya M, Meneghin J, Yang ZK, Seewald JS, Geoff Wheat C, et al. Inter-field
783 variability in the microbial communities of hydrothermal vent deposits from a back-arc basin.
784 Geobiology. 2012;10:333–46.
785 5. Dahle H, Økland I, Thorseth IH, Pederesen RB, Steen IH. Energy landscapes shape
786 microbial communities in hydrothermal systems on the Arctic Mid-Ocean Ridge. ISME J.
787 2015;9:1593–606.
788 6. Dahle H, Le Moine Bauer S, Baumberger T, Stokke R, Pedersen RB, Thorseth IH, et al.
789 Energy landscapes in hydrothermal chimneys shape distributions of primary producers. Front
790 Microbiol. 2018;9:1570.
791 7. Fortunato CS, Larson B, Butterfield DA, Huber JA. Spatially distinct, temporally stable
792 microbial populations mediate biogeochemical cycling at and below the seafloor in hydrothermal
793 vent fluids. Environ Microbiol. 2018;20:769–84.
794 8. Reysenbach A-L, St. John E, Meneghin J, Flores GE, Podar M, Dombrowski N, et al.
795 Complex subsurface hydrothermal fluid mixing at a submarine arc volcano supports distinct and
796 highly diverse microbial communities. Proc Natl Acad Sci U S A. 2020;117:32627–38.
797 9. Reveillaud J, Reddington E, McDermott J, Algar C, Meyer JL, Sylva S, et al. Subseafloor
798 microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-

799 Cayman Rise. Environ Microbiol. 2016;18:1970–87.
800 10. Anderson RE, Reveillaud J, Reddington E, Delmont TO, Eren AM, McDermott JM, et al.
801 Genomic variation in microbial populations inhabiting the marine subseafloor at deep-sea
802 hydrothermal vents. Nat Commun. 2017;8:1114.
803 11. Dombrowski N, Seitz KW, Teske AP, Baker BJ. Genomic insights into potential
804 interdependencies in microbial hydrocarbon and nutrient cycling in hydrothermal sediments.
805 Microbiome. 2017;5:106.
806 12. Dombrowski N, Teske AP, Baker BJ. Expansive microbial metabolic versatility and
807 biodiversity in dynamic Guaymas Basin hydrothermal sediments. Nat Commun. 2018;9:4999.
808 13. Ramírez GA, McKay LJ, Fields MW, Buckley A, Mortera C, Hensen C, et al. The Guaymas
809 Basin subseafloor sedimentary archaeome reflects complex environmental histories. iScience.
810 2020;23:101459.
811 14. Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J, et al. Comparative metagenomics of
812 microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting
813 chemistries. ISME J. 2011;5:414–26.
814 15. Hou J, Sievert SM, Wang Y, Seewald JS, Natarajan VP, Wang F, et al. Microbial succession
815 during the transition from active to inactive stages of deep-sea hydrothermal vent sulfide
816 chimneys. Microbiome. 2020;8:102.
817 16. Meier DV., Pjevac P, Bach W, Hourdez S, Girguis PR, Vidoudez C, et al. Niche partitioning
818 of diverse sulfur-oxidizing bacteria at hydrothermal vents. ISME J. 2017;11:1545–58.
819 17. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes
820 with the Genome Taxonomy Database. Bioinformatics. 2020;36:1925–7.
821 18. Konstantinidis KT, Rosselló-Móra R, Amann R. Uncultivated microbes in need of their own
822 taxonomy. ISME J. 2017;11:2399–406.
823 19. Von Damm KL, Edmond JM, Measures CI, Grant B. Chemistry of submarine hydrothermal
824 solutions at Guaymas Basin, Gulf of California. Geochim Cosmochim Acta. 1985;49:2221–37.

825 20. Teske A, de Beer D, McKay LJ, Tivey MK, Biddle JF, Hoer D, et al. The Guaymas Basin
826 hiking guide to hydrothermal mounds, chimneys, and microbial mats: complex seafloor
827 expressions of subsurface hydrothermal circulation. Front Microbiol. 2016;7:75.
828 21. Burggraf S, Larsen N, Woese CR, Stetter KO. An intron within the 16S ribosomal RNA gene
829 of the archaeon *Pyrobaculum aerophilum*. Proc Natl Acad Sci U S A. 1993;90:2547–50.
830 22. Nomura N, Morinaga Y, Kogishi T, Kim E-J, Sako Y, Uchida A. Heterogeneous yet similar
831 introns reside in identical positions of the rRNA genes in natural isolates of the archaeon
832 *Aeropyrum pernix*. Gene. 2002;295:43–50.
833 23. Jay ZJ, Inskeep WP. The distribution, diversity, and importance of 16S rRNA gene introns in
834 the order Thermoproteales. Biol Direct. 2015;10:35.
835 24. St. John E, Liu Y, Podar M, Stott MB, Meneghin J, Chen Z, et al. A new symbiotic
836 nanoarchaeote (*Candidatus* Nanoclepta minutus) and its host (*Zestosphaera tikiterensis* gen.
837 nov., sp. nov.) from a New Zealand hot spring. Syst Appl Microbiol. 2019;42:94–106.
838 25. Reysenbach A-L, Liu Y, Banta AB, Beveridge TJ, Kirshtein JD, Schouten S, et al. A
839 ubiquitous thermoacidophilic archaeon from deep-sea hydrothermal vents. Nature.
840 2006;442:444–7.
841 26. Kozubal MA, Romine M, Jennings RD, Jay ZJ, Tringe SG, Rusch DB, et al. Geoarchaeota:
842 a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone
843 National Park. ISME J. 2013;7:622–34.
844 27. McKay L, Klokman VW, Mendlovitz HP, Larowe DE, Hoer DR, Albert D, et al. Thermal and
845 geochemical influences on microbial biogeography in the hydrothermal sediments of Guaymas
846 Basin, Gulf of California. Environ Microbiol Rep. 2016;8:150–61.
847 28. Zhou Z, Liu Y, Xu W, Pan J, Luo Z-H, Li M. Genome- and community-level interaction
848 insights into carbon utilization and element cycling functions of Hydrothermarchaeota in
849 hydrothermal sediment. mSystems. 2020;5:e00795-19.
850 29. Buessecker S, Palmer M, Lai D, Dimapilis J, Mayali X, Mosier D, et al. An essential role for

851 tungsten in the ecology and evolution of a previously uncultivated lineage of anaerobic,
852 thermophilic Archaea. Nat Commun. 2022;13:3773.
853 30. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of
854 the tree of life. Nat Microbiol. 2016;1:16048.
855 31. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A
856 standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of
857 life. Nat Biotechnol. 2018;36:996–1004.
858 32. Luef B, Frischkorn KR, Wrighton KC, Holman H-YN, Birarda G, Thomas BC, et al. Diverse
859 uncultivated ultra-small bacterial cells in groundwater. Nat Commun. 2015;6:6372.
860 33. Kantor RS, Wrighton KC, Handley KM, Sharon I, Hug LA, Castelle CJ, et al. Small genomes
861 and sparse metabolisms of sediment-associated bacteria from four candidate phyla. MBio.
862 2013;4:e00708-13.
863 34. Castelle CJ, Brown CT, Anantharaman K, Probst AJ, Huang RH, Banfield JF. Biosynthetic
864 capacity, metabolic variety and unusual biology in the CPR and DPANN radiations. Nat Rev
865 Microbiol. 2018;16:629–45.
866 35. Tian R, Ning D, He Z, Zhang P, Spencer SJ, Gao S, et al. Small and mighty: adaptation of
867 superphylum Patescibacteria to groundwater environment drives their genome simplicity.
868 Microbiome. 2020;8:51.
869 36. Lemos LN, Manoharan L, Mendes LW, Venturini AM, Pylro VS, Tsai SM. Metagenome
870 assembled-genomes reveal similar functional profiles of CPR/Patescibacteria phyla in soils.
871 Environ Microbiol Rep. 2020;12:651–5.
872 37. Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of
873 microbial genomes shed light on interconnected biogeochemical processes in an aquifer
874 system. Nat Commun. 2016;7:13219.
875 38. McLean JS, Bor B, Kerns KA, Liu Q, To TT, Solden L, et al. Acquisition and adaptation of
876 ultra-small parasitic reduced genome Bacteria to mammalian hosts. Cell Rep. 2020;32:107939.

877 39. Cross KL, Campbell JH, Balachandran M, Campbell AG, Cooper SJ, Griffen A, et al.
878 Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. Nat Biotechnol.
879 2019;37:1314–21.
880 40. He X, McLean JS, Edlund A, Yooseph S, Hall AP, Liu S-Y, et al. Cultivation of a human-
881 associated TM7 phylotype reveals a reduced genome and epibiotic parasitic lifestyle. Proc Natl
882 Acad Sci U S A. 2015;112:244–9.
883 41. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, et al. Insights into
884 the phylogeny and coding potential of microbial dark matter. Nature. 2013;499:431–7.
885 42. Anantharaman K, Breier JA, Dick GJ. Metagenomic resolution of microbial functions in
886 deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J.
887 2016;10:225–39.
888 43. Beam JP, Becraft ED, Brown JM, Schulz F, Jarett JK, Bezuidt O, et al. Ancestral absence of
889 electron transport chains in Patescibacteria and DPANN. Front Microbiol. 2020;11:1848.
890 44. Kato S, Nakawake M, Kita J, Yamanaka T, Utsumi M, Okamura K, et al. Characteristics of
891 microbial communities in crustal fluids in a deep-sea hydrothermal field of the Suiyo Seamount.
892 Front Microbiol. 2013;4:85.
893 45. Oulas A, Polymenakou PN, Seshadri R, Tripp HJ, Mandalakis M, Paez-Espino AD, et al.
894 Metagenomic investigation of the geologically unique Hellenic Volcanic Arc reveals a distinctive
895 ecosystem with unexpected physiology. Environ Microbiol. 2016;18:1122–36.
896 46. Sieber CMK, Paul BG, Castelle CJ, Hu P, Tringe SG, Valentine DL, et al. Unusual
897 metabolism and hypervariation in the genome of a gracilibacterium (BD1-5) from an oil-
898 degrading community. MBio. 2019;10:e02128-19.
899 47. Moreira D, Zivanovic Y, López-Archilla AI, Iniesto M, López-García P. Reductive evolution
900 and unique predatory mode in the CPR bacterium *Vampirococcus lugosii*. Nat Commun.
901 2021;12:2454.
902 48. Yakimov MM, Merkel AY, Gaisin VA, Pilhofer M, Messina E, Hallsworth JE, et al. Cultivation

903 of a vampire: '*Candidatus* Absconditicoccus praedator.' Environ Microbiol. 2022;24:30–49.
904 49. Coutinho FH, von Meijenfeldt FAB, Walter JM, Haro-Moreno JM, Lopéz-Pérez M, van Verk
905 MC, et al. Ecogenomics and metabolic potential of the South Atlantic Ocean microbiome. Sci
906 Total Environ. 2021;765:142758.
907 50. Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR, et al. Community
908 genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum
909 and indicate roles in sediment carbon cycling. Microbiome. 2013;1:22.
910 51. Fincker M, Huber JA, Orphan VJ, Rappé MS, Teske A, Spormann AM. Metabolic strategies
911 of marine subseafloor Chloroflexi inferred from genome reconstructions. Environ Microbiol.
912 2020;22:3188–204.
913 52. Fullerton H, Moyer CL. Comparative single-cell genomics of Chloroflexi from the Okinawa
914 Trough deep-subsurface biosphere. Appl Environ Microbiol. 2016;82:3000–8.
915 53. Nuppunen-Puputti M, Kietäväinen R, Raulio M, Soro A, Purkamo L, Kukkonen I, et al.
916 Epilithic microbial community functionality in deep oligotrophic continental bedrock. Front
917 Microbiol. 2022;13:826048.
918 54. West-Roberts JA, Matheus-Carnevali PB, Schoelmerich MC, Al-Shayeb B, Thomas AD,
919 Sharrar A, et al. The Chloroflexi supergroup is metabolically diverse and representatives have
920 novel genes for non-photosynthesis based CO_2 fixation. BioRxiv.
921 https://doi.org/10.1101/2021.08.23.457424.
922 55. Liu R, Wei X, Song W, Wang L, Cao J, Wu J, et al. Novel Chloroflexi genomes from the
923 deepest ocean reveal metabolic strategies for the adaptation to deep-sea habitats. Microbiome.
924 2022;10:75.
925 56. McGonigle JM, Lang SQ, Brazelton WJ. Genomic evidence for formate metabolism by
926 Chloroflexi as the key to unlocking deep carbon in Lost City microbial ecosystems. Appl Environ
927 Microbiol. 2020;86:e02583-19.
928 57. Islam ZF, Cordero PRF, Feng J, Chen Y-J, Bay SK, Jirapanjawat T, et al. Two Chloroflexi

929 classes independently evolved the ability to persist on atmospheric hydrogen and carbon
930 monoxide. ISME J. 2019;13:1801–13.
931 58. Altshuler I, Raymond-Bouchard I, Magnuson E, Tremblay J, Greer CW, Whyte LG. Unique
932 high Arctic methane metabolizing community revealed through in situ $^{13}CH_4$-DNA-SIP
933 enrichment in concert with genome binning. Sci Rep. 2022;12:1160.
934 59. Madigan MT, Brock TD. Photosynthetic sulfide oxidation by *Chloroflexus aurantiacus*, a
935 filamentous, photosynthetic, gliding bacterium. J Bacteriol. 1975;122:782–4.
936 60. Kawai S, Martinez JN, Lichtenberg M, Trampe E, Kühl M, Tank M, et al. In-situ
937 metatranscriptomic analyses reveal the metabolic flexibility of the thermophilic anoxygenic
938 photosynthetic bacterium *Chloroflexus aggregans* in a hot spring Cyanobacteria-dominated
939 microbial mat. Microorganisms. 2021;9:652.
940 61. Spieck E, Spohn M, Wendt K, Bock E, Shively J, Frank J, et al. Extremophilic nitrite-
941 oxidizing Chloroflexi from Yellowstone hot springs. ISME J. 2020;14:364–79.
942 62. Baker BJ, Lesniewski RA, Dick GJ. Genome-enabled transcriptomics reveals archaeal
943 populations that drive nitrification in a deep-sea hydrothermal plume. ISME J. 2012;6:2269–79.
944 63. Engelen B, Nguyen T, Heyerhoff B, Kalenborn S, Sydow K, Tabai H, et al. Microbial
945 communities of hydrothermal Guaymas Basin surficial sediment profiled at 2 millimeter-scale
946 resolution. Front Microbiol. 2021;12:710881.
947 64. Beman JM, Popp BN, Francis CA. Molecular and biogeochemical evidence for ammonia
948 oxidation by marine Crenarchaeota in the Gulf of California. ISME J. 2008;2:429–41.
949 65. Speth DR, Yu FB, Connon SA, Lim S, Magyar JS, Peña-Salinas ME, et al. Microbial
950 communities of Auka hydrothermal sediments shed light on vent biogeography and the
951 evolutionary history of thermophily. ISME J. 2022;16:1750–64.
952 66. Mehta MP, Baross JA. Nitrogen fixation at 92°C by a hydrothermal vent archaeon. Science.
953 2006;314:1783–6.
954 67. Dick GJ. The microbiomes of deep-sea hydrothermal vents: distributed globally, shaped

955 locally. Nat Rev Microbiol. 2019;17:271–83.
956 68. Frank KL, Rogers DR, Olins HC, Vidoudez C, Girguis PR. Characterizing the distribution
957 and rates of microbial sulfate reduction at Middle Valley hydrothermal vents. ISME J.
958 2013;7:1391–401.
959 69. Zhou Z, Tran PQ, Adams AM, Kieft K, Breier JA, Sinha RK, et al. The sulfur cycle connects
960 microbiomes and biogeochemistry in deep-sea hydrothermal plumes. BioRxiv.
961 https://doi.org/10.1101/2022.06.02.494589.
962 70. Von Damm KL, Parker CM, Zierenberg RA, Lilley MD, Olson EJ, Clague DA, et al. The
963 Escanaba Trough, Gorda Ridge hydrothermal system: temporal stability and subseafloor
964 complexity. Geochim Cosmochim Acta. 2005;69:4971–84.
965 71. Dhillon A, Lever M, Lloyd KG, Albert DB, Sogin ML, Teske A. Methanogen diversity
966 evidenced by molecular characterization of methyl coenzyme M reductase A (*mcrA*) genes in
967 hydrothermal sediments of the Guaymas Basin. Appl Environ Microbiol. 2005;71:4592–601.
968 72. Adam N, Perner M. Microbially mediated hydrogen cycling in deep-sea hydrothermal vents.
969 Front Microbiol. 2018;9:2873.
970 73. Lever MA, Teske AP. Diversity of methane-cycling Archaea in hydrothermal sediment
971 investigated by general and group-specific PCR primers. Appl Environ Microbiol. 2015;81:1426–
972 41.
973 74. Teske A, Wegener G, Chanton JP, White D, MacGregor B, Hoer D, et al. Microbial
974 communities under distinct thermal and geochemical regimes in axial and off-axis sediments of
975 Guaymas Basin. Front Microbiol. 2021;12:633649.
976 75. Søndergaard D, Pedersen CNS, Greening C. HydDB: a web tool for hydrogenase
977 classification and analysis. Sci Rep. 2016;6:34212.
978 76. Kodama Y, Watanabe K. *Sulfuricurvum kujiense* gen. nov., sp. nov., a facultatively
979 anaerobic, chemolithoautotrophic, sulfur-oxidizing bacterium isolated from an underground
980 crude-oil storage cavity. Int J Syst Evol Microbiol. 2004;54:2297–300.

77. Takai K, Inagaki F, Nakagawa S, Hirayama H, Nunoura T, Sako Y, et al. Isolation and phylogenetic diversity of members of previously uncultivated ε-Proteobacteria in deep-sea hydrothermal fields. FEMS Microbiol Lett. 2003;218:167–74.

78. Takai K, Suzuki M, Nakagawa S, Miyazaki M, Suzuki Y, Inagaki F, et al. *Sulfurimonas paralvinellae* sp. nov., a novel mesophilic, hydrogen- and sulfur-oxidizing chemolithoautotroph within the Epsilonproteobacteria isolated from a deep-sea hydrothermal vent polychaete nest, reclassification of *Thiomicrospira denitrificans* as *Sulfurimonas denitrificans* comb. nov. and emended description of the genus *Sulfurimonas*. Int J Syst Evol Microbiol. 2006;56:1725–33.

79. Caldwell SL, Liu Y, Ferrera I, Beveridge T, Reysenbach A-L. *Thermocrinis minervae* sp. nov., a hydrogen- and sulfur-oxidizing, thermophilic member of the Aquificales from a Costa Rican terrestrial hot spring. Int J Syst Evol Microbiol. 2010;60:338–43.

80. Götz D, Banta A, Beveridge TJ, Rushdi AI, Simoneit BRT, Reysenbach A-L. *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel, thermophilic, hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. Int J Syst Evol Microbiol. 2002;52:1349–59.

81. Topçuoglu BD, Stewart LC, Morrison HG, Butterfield DA, Huber JA, Holden JF. Hydrogen limitation and syntrophic growth among natural assemblages of thermophilic methanogens at deep-sea hydrothermal vents. Front Microbiol. 2016;7:1240.

82. Webster NS. Cooperation, communication, and co-evolution: grand challenges in microbial symbiosis research. Front Microbiol. 2014;5:164.

83. Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R, et al. Hydrogen is an energy source for hydrothermal vent symbioses. Nature. 2011;476:176–80.

84. Galambos D, Anderson RE, Reveillaud J, Huber JA. Genome-resolved metagenomics and metatranscriptomics reveal niche differentiation in functionally redundant microbial communities at deep-sea hydrothermal vents. Environ Microbiol. 2019;21:4395–410.

85. Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean

1007 microbiome. Science. 2016;353:1272–7.
1008 86. Yamamoto M, Takai K. Sulfur metabolisms in Epsilon- and Gamma-proteobacteria in deep-
1009 sea hydrothermal fields. Front Microbiol. 2011;2:192.
1010 87. Giovannelli D, Chung M, Staley J, Starovoytov V, Le Bris N, Vetriani C. *Sulfurovum riftiae*
1011 sp. nov., a mesophilic, thiosulfate-oxidizing, nitrate-reducing chemolithoautotrophic
1012 epsilonproteobacterium isolated from the tube of the deep-sea hydrothermal vent polychaete
1013 *Riftia pachyptila*. Int J Syst Evol Microbiol. 2016;66:2697–701.
1014 88. Mori K, Yamaguchi K, Hanada S. *Sulfurovum denitrificans* sp. nov., an obligately
1015 chemolithoautotrophic sulfur-oxidizing epsilonproteobacterium isolated from a hydrothermal
1016 field. Int J Syst Evol Microbiol. 2018;68:2183–7.
1017 89. Nakagawa S, Takai K, Inagaki F, Horikoshi K, Sako Y. *Nitratiruptor tergarcus* gen. nov., sp.
1018 nov. and *Nitratifractor salsuginis* gen. nov., sp. nov., nitrate-reducing chemolithoautotrophs of
1019 the ε-Proteobacteria isolated from a deep-sea hydrothermal system in the Mid-Okinawa Trough.
1020 Int J Syst Evol Microbiol. 2005;55:925–33.
1021 90. Assié A, Leisch N, Meier DV, Gruber-Vodicka H, Tegetmeyer HE, Meyerdierks A, et al.
1022 Horizontal acquisition of a patchwork Calvin cycle by symbiotic and free-living
1023 Campylobacterota (formerly Epsilonproteobacteria). ISME J. 2020;14:104–22.
1024 91. Berg IA. Ecological aspects of the distribution of different autotrophic CO_2 fixation pathways.
1025 Appl Environ Microbiol. 2011;77:1925–36.
1026 92. Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hügler M, et al. Physiological
1027 proteomics of the uncultured endosymbiont of *Riftia pachyptila*. Science. 2007;315:247–50.
1028 93. Waite DW, Vanwonterghem I, Rinke C, Parks DH, Zhang Y, Takai K, et al. Comparative
1029 genomic analysis of the class Epsilonproteobacteria and proposed reclassification to
1030 Epsilonbacteraeota (phyl. nov.). Front Microbiol. 2017;8:682.
1031 94. Macalady JL, Dattagupta S, Schaperdoth I, Jones DS, Druschel GK, Eastman D. Niche
1032 differentiation among sulfur-oxidizing bacterial populations in cave waters. ISME J. 2008;2:590–

601.
95. Patwardhan S, Foustoukos DI, Giovannelli D, Yücel M, Vetriani C. Ecological succession of sulfur-oxidizing Epsilon- and Gammaproteobacteria during colonization of a shallow-water gas vent. Front Microbiol. 2018;9:2970.
96. Flores GE, Wagner ID, Liu Y, Reysenbach A-L. Distribution, abundance, and diversity patterns of the thermoacidophilic “deep-sea hydrothermal vent Euryarchaeota 2.” Front Microbiol. 2012;3:47.
97. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: a new versatile metagenomic assembler. Genome Res. 2017;27:824–34.
98. Uritskiy G V., DiRuggiero J, Taylor J. MetaWRAP -- a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018;6:158.
99. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol. 2018;3:836–43.
100. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
101. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol. 2017;2:1533–42.
102. Pruesse E, Peplies J, Glöckner FO. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics. 2012;28:1823–9.
103. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, et al. VizBin - an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome. 2015;3:1.
104. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, et al. Community-led,

integrated, reproducible multi-omics with Anvi'o. Nature Microbiology. 2021;6:3–6.
105. von Meijenfeldt FAB, Arkhipova K, Cambuy DD, Coutinho FH, Dutilh BE. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol. 2019;20:217.
106. Nayfach S, Roux S, Seshadri R, Udwary D, Varghese N, Schulz F, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39:499–509.
107. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.
108. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 2020;36:2251–2252.
109. Zhou Z, Tran PQ, Breister AM, Liu Y, Kieft K, Cowley ES, et al. METABOLIC: high-throughput profiling of microbial genomes for functional traits, metabolism, biogeochemistry, and community-scale functional networks. Microbiome. 2022;10:33.
110. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309–14.
111. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol. 2019;1962:1–14.
112. Campbell JH, O'Donoghue P, Campbell AG, Schwientek P, Sczyrba A, Woyke T, et al. UGA is an additional glycine codon in uncultured SR1 Bacteria from the human microbiota. Proc Natl Acad Sci U S A. 2013;110:5540–5.
113. Hanke A, Hamann E, Sharma R, Geelhoed JS, Hargesheimer T, Kraft B, et al. Recoding of the stop codon UGA to glycine by a BD1-5/SN-2 bacterium and niche partitioning between Alpha- and Gammaproteobacteria in a tidal sediment microbial community naturally selected in a laboratory chemostat. Front Microbiol. 2014;5:231.

114. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
115. Tully BJ, Wheat CG, Glazer BT, Huber JA. A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. ISME J. 2018;12:1–16.
116. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.
117. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
118. Price MN, Dehal PS, Arkin AP. FastTree 2 -- approximately maximum-likelihood trees for large alignments. PLoS One. 2010;5:e9490.
119. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37:1530–4.
120. Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.
121. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46:D624–32.
122. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46:W95–101.
123. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50:D571–7.
124. Yu NY, Wagner JR, Laird MR, Melli G, Rey S, Lo R, et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive

capabilities for all prokaryotes. Bioinformatics. 2010;26:1608–15.
125. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, et al. DRAM for distilling microbial metabolism to automate the curation of microbiome function. Nucleic Acids Res. 2020;48:8883–900.
126. Pérez Castro S, Borton MA, Regan K, Hrabe de Angelis I, Wrighton KC, Teske AP, et al. Degradation of biological macromolecules supports uncultured microbial populations in Guaymas Basin hydrothermal sediments. ISME J. 2021;15:3480–97.
127. Garber AI, Nealson KH, Okamoto A, McAllister SM, Chan CS, Barco RA, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. Front Microbiol. 2020;11:37.
128. Anantharaman K, Hausmann B, Jungbluth SP, Kantor RS, Lavy A, Warren LA, et al. Expanded diversity of microbial groups that shape the dissimilatory sulfur cycle. ISME J. 2018;12:1715–28.
129. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.
130. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
131. Seitz KW, Dombrowski N, Eme L, Spang A, Lombard J, Sieber JR, et al. Asgard Archaea capable of anaerobic hydrocarbon cycling. Nat Commun. 2019;10:1822.
132. Boyd JA, Jungbluth SP, Leu AO, Evans PN, Woodcroft BJ, Chadwick GL, et al. Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. ISME J. 2019;13:1269–79.
133. Liu H, Xin Y, Xun L. Distribution, diversity, and activities of sulfur dioxygenases in heterotrophic bacteria. Appl Environ Microbiol. 2014;80:1799–806.
134. Clarke KR, Gorley RN. Primer V6: user manual - tutorial. Plymouth: Plymouth Marine Laboratory; 2006.

Additional Supplementary Files

Additional File 1. Supplementary Figures (.pdf)

Fig. S1. Geographic distribution of deep-sea hydrothermal vent sampling locations. The number of samples collected in each region is shown with *n* values. **Fig. S2.** Deep-sea hydrothermal vent photographs from ELSC, EPR, MAR and Guaymas Basin. **Fig. S3.** Comparison between the number of medium- to high-quality MAGs recovered in each metagenomic assembly and the number of reads that passed quality control measures. Metagenomic assemblies are ordered by (A) increasing MAG count and (B) increasing read count. **Fig. S4.** NMDS plot showing the taxonomic diversity of Brothers volcano MAGs, based on normalized relative abundance. Clustering patterns show a high degree of similarity to NMDS plot clustering previously reported in Reysenbach et al., 2020. **Fig. S5.** Anvi'o plot showing the cluster of scaffolds (blue) predominantly corresponding to the Nanoarchaeota in M10_maxbin2_scaf2bin.065. Analysis with CAT revealed three additional contaminating scaffolds which were removed, bringing the final scaffold count to 149, with an estimated 47% completion by CheckM. Scaffold clusters that were removed (pink; 972 scaffolds) were largely assigned to taxonomic groups outside the Nanoarchaeota by CAT and had a low number of marker genes, as estimated by CheckM (6.99% completion, 0.29% contamination). **Fig. S6.** Predicted cell metabolism diagrams for the putative new phyla (A) JALSQH01 (3 MAGs) and (B) JALWCF01 (13 MAGs). Functions (F) and modules (M) were identified using METABOLIC (Table S5). Solid lines indicate the presence of a module or function, while dashed lines and a "p" in parentheses indicate that a module or function was only present sporadically (<50% of MAGs). Modules and functions not identified in any MAGs are shown with dashed lines and gray labels. **Fig. S7.** Normalized relative abundance of GTDB classes, expressed as a percentage. Classes depicted comprise ≥16% of the relative MAG abundance in at least one assembly. **Fig. S8.** Maximum-likelihood GTDB-Tk concatenated protein tree showing members of the Patescibacteria, used to generate Fig. 5A.

Lineages outside the Patescibacteria are shown as a collapsed triangle, and MAGs from this study are indicated in bold type. Filled circles represent SH-like branch support (0.8-1.0), and the scale bar shows 0.5 substitutions per amino acid. **Fig. S9.** Concatenated dissimilatory sulfite reductase (DsrAB) protein phylogenetic tree. Only the nodes with ultrafast bootstrap (UFBoot) support values over 90% were labeled with black dots. This tree included both reductive DsrAB (for reductive dissimilatory sulfite reduction to sulfide) and oxidative DsrAB (for dissimilatory sulfur oxidation to sulfite). For collapsed clades in the oxidative DsrAB clade (labeled in blue), the DsrAB call numbers and DsrAB-containing MAG numbers were labeled in square brackets. The total number for both reductive DsrAB calls and reductive DsrAB-containing MAG numbers and oxidative DsrAB calls and oxidative DsrAB-containing MAG numbers were labeled accordingly on the side of the tree. Note that one genome can have multiple paired DsrAB calls. **Fig. S10.** Phylogenetic protein tree of methyl coenzyme M reductase subunit alpha (McrA). Ultrafast bootstrap support values (>90%) are shown with filled circles. Clades comprised of predicted butane oxidation (Butane clade), X-alkane oxidation (X-alkane clade) and anaerobic methanotrophy-associated (ANME-1 and -2) McrA amino acid sequences are highlighted, and the three predicted McrA sequences from the Archaeoglobi are shown in red. **Fig. S11.** Sdo (sulfur dioxygenase) phylogenetic protein tree. Only the nodes with ultrafast bootstrap (UFBoot) support values over 90% were labeled with black dots. The positive Sdo sequences that were checked by two conservative amino acid residues were labeled yellow in the tree. Three positive Sdo clades (including ETHE1, Sdo, and Blh) were labeled yellow; the numbers of positive Sdo sequences, non-Sdo sequences, and Sdo reference sequences were labeled accordingly. Other unannotated clades and non-Sdo clades (including metallo-beta-lactamase, GloB1, and GloB2) all contained non-Sdo sequences. **Fig. S12.** Relative abundance of GTDB-assigned MAG taxa at Guaymas Basin. Abundances are shown (A) for all taxa at the genus level, and (B) for the Archaea at the order level, using read coverage normalized to 100M reads per sample and

expressed as a percentage of MAG reads per sample. Relative abundances were averaged for the two samples from the six-day thermocouple array (4561-380 and 4561-384).

Additional File 2. Supplementary Tables (.xlsx)

Table S1. Sample metadata including location, year, research vessel, number of metagenome reads and accession numbers. **Table S2.** MAG genome properties, accession numbers and taxonomic classifications. Taxonomy was assigned using GTDB-Tk, and mis-classified MAGs were taxonomically re-assigned at the phylum level (Bacteria) and class level (Archaea) using curated archaeal and bacterial phylogenetic trees. Genome quality statistics are based on completion and contamination (high quality, >90% completion, <5% contamination; medium quality, ≥50% completion, ≤10% contamination). Average contamination was 4.02%. **Table S3.** Average amino acid identity (AAI) matrices for the (A) Bacteria and (B) Archaea. Matrices are grouped by GTDB taxonomy and include MAGs that could not be assigned to a known genus by GTDB-Tk. Details are provided which recently identified MAGs were from Brothers volcano hydrothermal deposits. **Table S4.** Relative abundance of GTDB taxa by site, based on read coverage of MAGs normalized to 100M reads per sample. MAG coverage for each site was summed and expressed as a percent. **Table S5.** METABOLIC-G results for JALSQH01 (3 MAGs) and JALWCF01 (13 MAGs). In the summary rows for JALSQH01 and JALWCF01, functions and modules are listed as “present” if identified in ≥50% of all MAGs, “partially present” if found in <50% of the MAGs, and “absent” if undetected in the MAGs. **Table S6.** Selected functional genes found in Patescibacteria MAGs, based on annotation with GhostKOALA. KEGG module numbers are shown in parentheses. **Table S7.** Functional genes identified in selected > 80%-completeness MAGs from the Chloroflexota. (A) Genes are marked as present (1; green highlight) or not detected (0) in individual MAGs. (B) The proportion of > 80%-completeness MAGs in six GTDB orders that encode functional genes is also shown, with proportions ≥50% highlighted in green. **Table S8.** Identification and distribution of functional

genes in this study. (A) The HMMs, MEROPS peptidases, and CAZymes used to identify functional genes. Gene call numbers were calculated using the component (C) or accumulative (A) methods described in the methods. Genes requiring manual validation (M) are indicated. (B) Functional gene abundance, calculated as described in the methods. **Table S9.** Percentage of MAGs in phylogenetic clusters that encode core metabolic genes. Unless otherwise indicated, Archaea are shown at the class level, and Bacteria are shown at the phylum level. Genes were detected using METABOLIC, with additional validation steps for oxidative and reductive Dsr, Sdo, PmoA and McrA. **Table S10.** Comparative (A) relative abundance and (B) functional gene abundance for the Gammaproteobacteria and Campylobacteria, used to generate Fig. 10.

Additional File 3. Data S1 (.nwk)

Newick format archaeal concatenated protein phylogenetic tree, including both MAGs and GTDB reference genomes.

Additional File 4. Data S2 (.nwk)

Newick file of bacterial concatenated protein phylogenetic tree including MAGs and GTDB reference genomes, generated using IQ-TREE.

Additional File 5. Data S3 (.nwk)

Concatenated protein phylogenetic tree of bacterial MAGs and GTDB reference genomes, generated with FastTree (Newick format).

Additional File 6. Data S4 (.nwk)

MAG-only bacterial concatenated phylogenetic protein tree in Newick format.

Additional File 7. Data S5 (.nwk)

Concatenated protein phylogeny of archaeal MAGs in Newick format.

Additional File 8. Supplementary Discussion (.docx)