







# Brockarchaeota, a novel archaeal phylum with unique and versatile carbon cycling pathways

Valerie De Anda <sup>1</sup>, Lin-Xing Chen<sup>2</sup>, Nina Dombrowski <sup>1,3</sup>, Zheng-Shuang Hua <sup>4</sup>, Hong-Chen Jiang<sup>5</sup>, Jillian F. Banfield <sup>2,6</sup>, Wen-Jun Li <sup>7,8</sup>✉ & Brett J. Baker <sup>1</sup>✉

Geothermal environments, such as hot springs and hydrothermal vents, are hotspots for carbon cycling and contain many poorly described microbial taxa. Here, we reconstructed 15 archaeal metagenome-assembled genomes (MAGs) from terrestrial hot spring sediments in China and deep-sea hydrothermal vent sediments in Guaymas Basin, Gulf of California. Phylogenetic analyses of these MAGs indicate that they form a distinct group within the TACK superphylum, and thus we propose their classification as a new phylum, 'Brockarchaeota', named after Thomas Brock for his seminal research in hot springs. Based on the MAG sequence information, we infer that some Brockarchaeota are uniquely capable of mediating non-methanogenic anaerobic methylotrophy, via the tetrahydrofolate methyl branch of the Wood-Ljungdahl pathway and reductive glycine pathway. The hydrothermal vent genotypes appear to be obligate fermenters of plant-derived polysaccharides that rely mostly on substrate-level phosphorylation, as they seem to lack most respiratory complexes. In contrast, hot spring lineages have alternate pathways to increase their ATP yield, including anaerobic methylotrophy of methanol and trimethylamine, and potentially use geothermally derived mercury, arsenic, or hydrogen. Their broad distribution and their apparent anaerobic metabolic versatility indicate that Brockarchaeota may occupy previously overlooked roles in anaerobic carbon cycling.

<sup>1</sup>Department of Marine Science, University of Texas Austin, Port Aransas, TX 78373, USA. <sup>2</sup>Department of Earth and Planetary Sciences, University of California, Berkeley, CA, USA. <sup>3</sup>NIOZ, Royal Netherlands Institute for Sea Research, Department of Marine Microbiology and Biogeochemistry, and Utrecht University, Den Burg, Netherlands. <sup>4</sup>Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei, PR China. <sup>5</sup>State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, People's Republic of China. <sup>6</sup>Department of Environmental Science, Policy and Management, University of California, Berkeley, CA, USA. <sup>7</sup>State Key Laboratory of Biocontrol, Guangdong Provincial Key Laboratory of Plant Resources and Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), School of Life Sciences, Sun Yat-Sen University, Guangzhou, People's Republic of China. <sup>8</sup>State Key Laboratory of Desert and Oasis Ecology, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, People's Republic of China. ✉email: [liwenjun3@mail.sysu.edu.cn](mailto:liwenjun3@mail.sysu.edu.cn); [acidophile@gmail.com](mailto:acidophile@gmail.com)

Advances in DNA sequencing and computational approaches have accelerated the reconstruction of metagenome assembled genomes (MAGs) from natural communities<sup>1</sup>. This approach has revealed many novel lineages on the tree of life and is advancing our understanding the ecological roles of uncultured microbes<sup>1–3</sup>. For example, many new archaeal phyla have been described from hot springs including *Geoarchaeota*<sup>4</sup>, *Marsarchaeota*<sup>5</sup>, *Aigarchaeota*<sup>6</sup>, and several Asgard phyla from deep-sea hydrothermal vents<sup>7–12</sup>. However, diversity surveys have demonstrated there are many novel taxa left to be explored<sup>13</sup>. Moreover, there are several gaps between our knowledge of active biogeochemical processes and the metabolic mechanisms and taxa mediating them. For example, the description of microbes mediating anaerobic methylo-trophy is still limited, and it is unclear which non-methanogenic heterotrophs utilize methylated compounds on the anoxic seafloor<sup>14</sup>. Little is known about the microorganisms or pathways mediating this process<sup>15</sup>.

Methylo-trophs are organisms that are capable of using simple organics including single-carbon (C1 e.g., methanol) and methylated (e.g., trimethylamine) compounds as a source of energy and carbon<sup>16,17</sup>. In nature, the most prevalent are compounds such as methanol and methylamines, which are derived from a variety of sources such as phytoplankton, plants, and the decay of organic matter<sup>15,18,19</sup>. As a result, they are ubiquitous in oceans and atmosphere and are important components of the global carbon and nitrogen cycles<sup>15</sup>. In oxic environments, methanol is converted to formaldehyde by the classical pyrrolo-quinoline quinone (PQQ)-linked methanol dehydrogenase pathway found in aerobic methylo-trophs<sup>15,18</sup>. In anoxic settings, these compounds are used as substrate for methylo-trophic-methanogenesis<sup>20–23</sup> and sulfate reduction<sup>24</sup>. Anaerobic methylo-trophs utilize the methyltransferase system (MT) to break and transfer the methyl residue to coenzyme M (in the case of methanogens) or tetrahydrofolate (H<sub>4</sub>F) (in acetogens and sulfate reducers)<sup>20–24</sup> and conserved energy via the Wood–Ljungdahl pathway (WLP). Methylo-trophic archaea include methanogenic orders (in Euryarchaeota): *Methanosarcinales*, *Methanobacteriales*, *Methanomassiliococcales*, and the recently discovered uncultured methylo-trophic phylum, *Verstraetearchaeota*<sup>20</sup>. Methylo-trophy has not been described in archaeal lineages outside of these methanogenic groups.

Here we describe a new archaeal phylum, the Brockarchaeota, whose members are metabolically versatile and can be found in geothermal environments around the world. The Brockarchaeota appear to possess diverse pathways for carbon cycling including fermentation of complex organic carbon compounds, anaerobic methylo-trophy, and chemolithotrophy.

## Results

**Genomic reconstruction.** Metagenomic sequencing, assembly, and binning of sediments from seven terrestrial hot springs in Tibet (up to 70°C) and Tengchong Yunnan, China (up to 86 °C) and deep sea hydrothermally heated Guaymas Basin (GB) sediments (10–34°C) resulted in the reconstruction of 15 draft metagenome-assembled genomes (MAGs) estimated to be 67–92% complete (Table 1). These MAGs range from 0.78 to 2.32 Mbp (average 1.47 Mbp) (Table 1). The two MAGs from GB (B48\_G17 and B27\_G9, temperature 33.62 and 10.4°C, respectively) were originally designated as “GB-API” in a prior study<sup>25</sup>. Although the GB genomes were obtained from lower temperatures these sediments experience increases in temperature due to hydrothermal circulation<sup>25</sup>. Thus, the organisms from which these genomes were derived likely prefer hot geothermal ecosystems, and anoxic conditions (Supplementary Data 1). All of these MAGs are rare members of the communities they were

recovered from, presenting 0.1 to 9% of the genomic reads (Supplementary Data 2).

**Phylogeny and distribution in nature.** Phylogenetic analyses of these MAGs based on a concatenated alignment of 37 conserved proteins (see Methods), revealed they form a distinct group within the TACK superphylum, basal to *Aigarchaeota* and *Thaumarchaeota* (Fig. 1 and Supplementary Data 3). A comparison of average amino acid identities (AAI) across 250 available TACK genomes (Supplementary Data 4), revealed that Brockarchaeota are distinct from neighboring phyla (*Geoarchaeota*, *Aigarchaeota*, and *Thaumarchaeota*) and share up to 99% genome-wide nucleotide similarity to one another (Supplementary Fig. 1 and Supplementary Data 5). The two GB MAGs (B48\_G17 and B27\_G9) are distinct from the hot springs at the AAI level (<50% similar to each other), and <45% AAI to members of *Geoarchaeota*, *Thaumarchaeota*, and *Aigarchaeota*, which is consistent with their phylogenetic placement. Phylogeny of 16S rRNA genes also indicated that Brockarchaeota do not fall within any described archaeal phyla (Fig. 2A), with <78% similarity to other TACK members. Together, these results support the classification of these MAGs as a new phylum. We propose that the phylum be named “Brockarchaeota”, after Thomas Brock, an American microbiologist known for his groundbreaking research in hot springs microbiology.

Interestingly, only three 16S rRNA gene sequences with similarity (92–96%) to Brockarchaeota sequences have been described in PCR-based surveys, highlighting the inherent bias for primer choice in diversity studies. Therefore, we searched publicly available metagenomic databases to examine the geographic distribution of this phylum. Notably, we almost exclusively found 16S rRNA gene sequences related to Brockarchaeota in sequence data generated from other hot springs from around the world (China, USA, South Africa; Fig. 2A), revealing Brockarchaeota are globally distributed in hot springs (Fig. 2B). Three sequences, which cluster together, were recovered from lake sediments in Rwanda and the Gulf of Boni in Indonesia (28° C) (see Supplementary Data 8), suggesting that some Brockarchaeota are mesophilic as well.

**Unique anaerobic methylo-trophic pathways.** To begin to understand the metabolism of the Brockarchaeota we compared the predicted proteins encoded by these genomes with a variety of functional databases (see Methods). This revealed a potential unique type of anaerobic methylo-trophic metabolism not yet described in archaea. They contain the methyltransferase system (MT), that has been shown to be essential for anaerobic methylo-trophy<sup>26</sup> and is composed of two key steps. First, specific methyltransferases break C–R bonds in a variety of substrates (MtaB for methanol, MtmB for monomethylamine, MtbB for dimethylamine, and MttB for trimethylamine) and transfer the methyl moieties to subunit MtaC. The second methyltransferase (MtaA for methanol, MtbA for methylamines), transfers the methyl-group from the corrinoïd protein to coenzyme M in methanogens, or tetrahydrofolate in acetogens<sup>24</sup>. Brockarchaeota from hot springs encode proteins predicted to be methanol-cobalamin methyltransferases (MtaB) and trimethylamine-corrinoïd protein methyltransferase (MttB) for the utilization of methanol and trimethylamine (TMA), respectively. B12-binding corrinoïd protein genes are always colocalized with *mtaB* (Supplementary Data 9), suggesting a co-transcription of both subunits of the methyltransferase complex. Due to the lack of MtaA proteins, we suggest that another undescribed protein may be involved in the transfer of the methylated compound from the corrinoïd protein to tetrahydrofolate (see details in Supplementary

**Table 1** General information on the fifteen novel Brockarchaeota MAGs.

Genome	Origin	Size (Mb)	Scaffolds (number)	Protein coding genes (number)	Predicted size(Mb)	GC (%)	Compl. (%)	Red. (%)	Largest scaffold (bp)	Strain hetero
B48_G17	Deep sea sediment (12–15 cm, 33.6 °C) Vent 2 in Dombrowski et al. <sup>25</sup>	0.78	110	911	0.92	51.06	83.1	0.93	49,468	100
B27_G9	Deep sea sediment (0–3 cm, 10.4 °C) Vent 2 in Dombrowski et al. <sup>25</sup>	0.99	223	1341	1.31	42.30	67	1.46	17,768	0
DRTY7-35_44	Hot spring sediment DiReTiYanQu-7 (collected in Jan, 2016) in Tengchong county, Yunnan, China (55.8 °C)	0.83	130	910	0.95	34.17	84	1.94	23,553	0
QC4_43	Hot spring QuCai village, Tibet, China (69.5 °C)	1.18	36	1332	1.42	43.17	80	2.43	135,049	0
QC4_48	Hot spring QuCai village, Tibet, China (69.5 °C)	1.66	86	1793	1.92	47.78	84.47	1.29	78,542	0
GD2_1	Hot spring GuDui geothermal area, Tibet, China (61.8 °C)	2.14	156	2326	2.37	47.13	89	1.94	82,330	0
QZM_A2	Hot spring QuZhuoMu village, Tibet, China (63.1 °C)	2.19	173	2447	2.59	47.44	82	1.94	96,973	0
QZM_A3	Hot spring QuZhuoMu village, Tibet, China (62.9 °C)	1.96	216	2270	2.38	47.89	78.64	3.88	70,810	0
DRTY-1.18	Hot spring sediment DiReTiYanQu-1 (collected in May, 2017) in Tengchong county, Yunnan, China (67 °C)	1.62	55	1582	1.81	34.71	88	1.94	117,455	0
DRTY-6.200	Hot spring sediment DiReTiYanQu-6 (collected in May, 2017) in Tengchong county, Yunnan, China (60 °C)	0.55	129	663	0.78	34.24	57	0	11,975	0
DRTY-6.80	Hot spring sediment DiReTiYanQu-6 (collected in May, 2017) in Tengchong county, Yunnan, China (60 °C)	2.32	55	2264	2.61	47.51	87	0	266,602	0
DRTY-7.37	Hot spring sediment DiReTiYanQu-7 (collected in Jan, 2016) in Tengchong county, Yunnan, China (55.8 °C)	1.03	189	1131	1.17	34.76	87	0.97	23,662	0
JZ-1.89	Hot spring sediment Jinze-1 (collected in May, 2017) in Tengchong county, Yunnan, China (86.5 °C)	1.48	88	1680	2.20	42.38	51	0.97	147,883	0
JZ-2.136	Hot spring sediment Jinze-2 (collected in May, 2017) in Tengchong county, Yunnan, China (63 °C)	2.04	214	2235	2.36	57.55	84	0.97	40,656	0
JZ-2.4	Hot spring sediment Jinze-2 (collected in Jan, 2016) in Tengchong county, Yunnan, China (75 °C)	1.30	61	1412	1.40	41.33	92	0	130,091	0

Genome ID, origin, number of scaffolds, number of protein-coding genes, guanine-cytosine (GC) content, estimated completeness (Compl.), estimated gene redundancy (Red.), and strain heterogeneity (Strain hetero.) are shown (Extended genomic statistics are found in Supplementary Data 1).

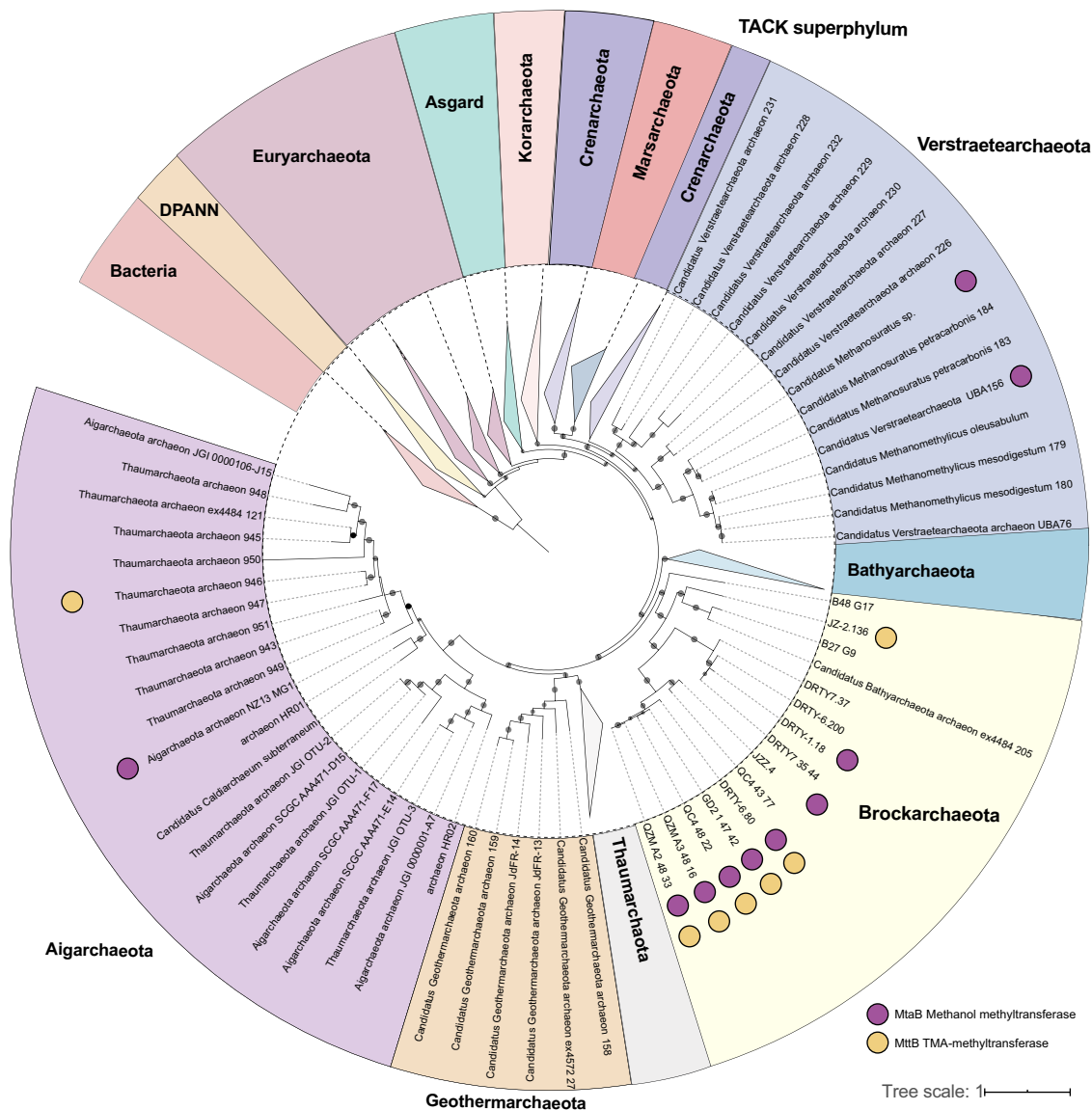
Discussion). Searches for MtaB and MttB proteins within TACK superphylum indicate methanol-MT system is a unique feature of Brockarchaeota (Fig. 1). Phylogenetic analyses of MtaB and MttB revealed Brockarchaeota form new branches within methanol and TMA methyltransferases (Fig. 3).

These results expand the distribution of anaerobic methylotrophy. Anaerobic methanol-utilization has only been described in Euryarchaeota and TACK (Verstraetaearchaeota<sup>20</sup>, Korarchaeota<sup>27</sup>) archaea, and some bacteria (Firmicutes and Deltaproteobacteria)<sup>28</sup>. However, Brockarchaeota genomes do not possess the key genes for methanogenesis including methyl-coenzyme M reductase (MCR) found in other archaea that have the MT system (Supplementary Data 10). To ensure that Brockarchaeota *mcr* genes were not overlooked, we searched metagenomic datasets from each of their communities and did not find any (See Supplementary Data 11). In addition to lacking Mcr, they lack key enzymes of the H<sub>4</sub>MPT methyl branch of the WLP, involved in the transfer and reduction of C1 moieties for methane production. The H<sub>4</sub>MPT is the natural pathway for degrading methylated compounds in methanogenic archaea<sup>20</sup> and a key module for methylotrophy, carbon

assimilation, formaldehyde detoxification in methylotrophic bacteria<sup>16,17,29,30</sup>. Furthermore, they lack the key enzyme for the carbonyl branch for the WLP, the CO-dehydrogenase–acetyl-CoA-synthase complex CODH/ACS<sup>31</sup>. Finally, Brockarchaeota do not encode the pyrroloquinoline quinone (PQQ)-linked methanol dehydrogenase pathway for aerobic methylotrophy (Supplementary Fig. 2) and recently found in deep-sea anaerobic bacteria<sup>24</sup>.

The lack of previously described methylotrophic pathways raises the question about how C1 compounds are assimilated in Brockarchaeota. Our findings suggest they metabolize formaldehyde, methanol, and trimethylamine via convergent assimilatory pathways for biosynthesis and energy conservation. A detailed comparative analysis suggests that Brockarchaeota assimilate methanol and TMA via the convergent action of the tetrahydrofolate (H<sub>4</sub>F) methyl branch of the WL pathway (H<sub>4</sub>F-WLP) and reductive glycine pathway (rGlyP).

**Formaldehyde assimilation.** Brockarchaeota MAGs (B48\_G17, DRTY7\_35\_44, DRT-1.18, JZ-2.136, DRTY7.37) encode both key enzymes of the ribulose monophosphate pathway (RuMP):



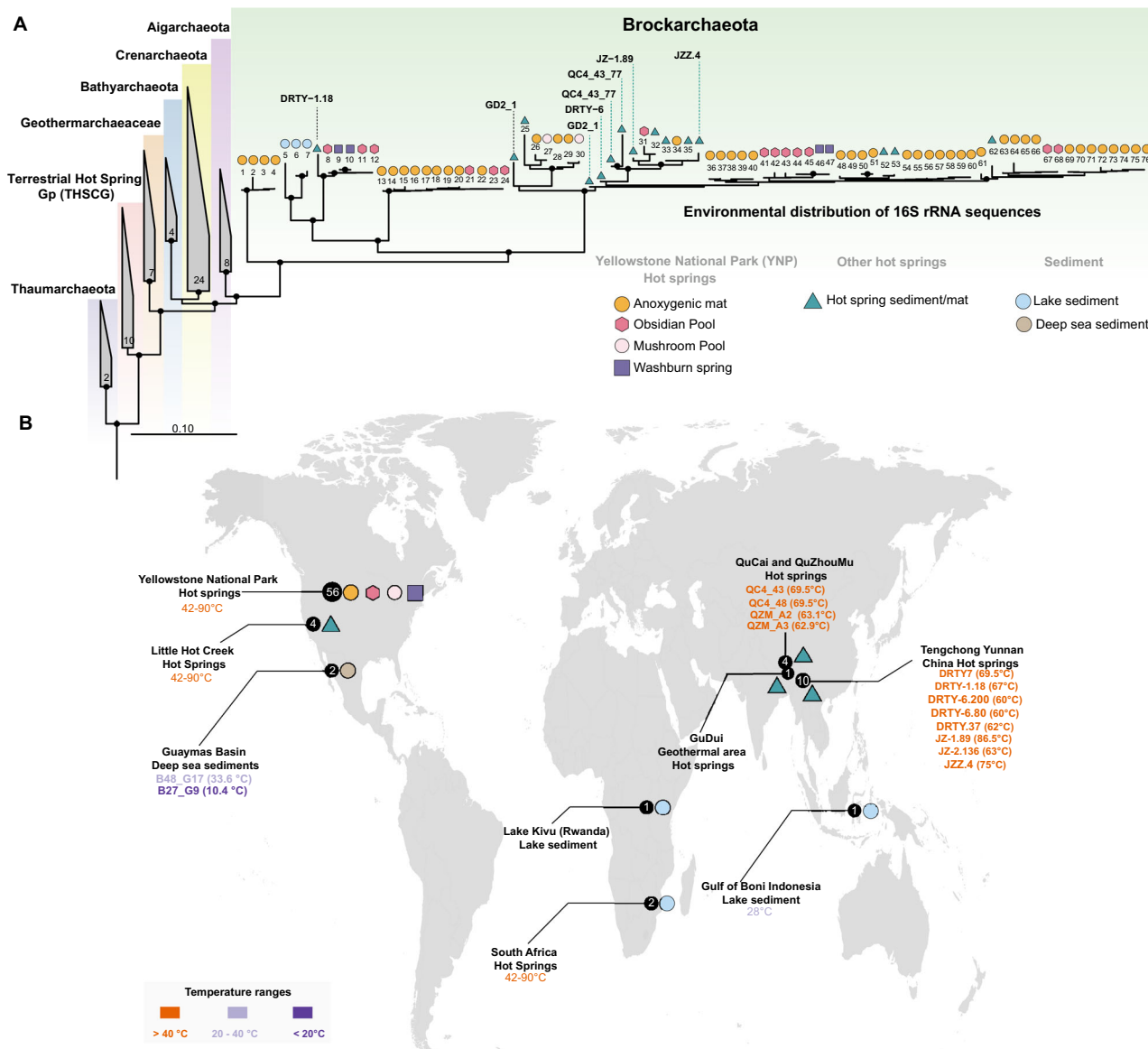
**Fig. 1 Comparison of phylogeny and distribution of methyltransferase system in Brockarchaeota and other members of TACK superphylum.** Phylogeny generated using IQtree v1.6.11 using a concatenation of 37 conserved single-copy proteins using phylsift<sup>63</sup> with a best fit LG + F + R10 model selected using the Bayesian Information Criterion (BIC), bootstrap values were calculated using non-parametric bootstrapping with 1,000 replicates (represented by gray circles, only bootstrap >70 are shown). Raw tree file is available in Supplementary Data 5; interactive version of the tree can be found at <https://itol.embl.de/tree/7618619231376741590439752#>. The presence methanol methyltransferase MtaB (PF12176) and trimethylamine methyltransferase MttB (PF06253) are shown in the outer circles. The annotation was conducted with MEBS<sup>72</sup> details can be found in Supplementary Data 6.

3-hexulose-6-phosphate synthase (HPS) and 6-phospho-3-hexuloisomerase (PHI). The RuMP pathway was originally found in methylotrophic bacteria that are able to use C1 compounds as a sole source of carbon and energy; however, it is currently recognized as a widespread prokaryotic pathway for formaldehyde fixation and detoxification<sup>32</sup>. The RuMP pathway functions as a highly efficient system for trapping free formaldehyde at relatively low concentrations. The presence of HPS and PHI in Brockarchaeota genomes suggests that formaldehyde can be fixed and detoxified via the RuMP pathway. Further oxidation of formaldehyde to formate can be carried out by the oxygen-sensitive tungsten-dependent aldehyde ferredoxin oxidoreductase (AFOR) which is present in all MAGs except those obtained from Tengchong hot springs (DRTY). In this way, Brockarchaeota can potentially play a key role in controlling formaldehyde consumption and therefore maintaining

viability of the microbial community in geothermally active environments.

**Formate, methanol, and trimethylamine assimilation.** In Brockarchaeota formate, derived by the oxidation of formaldehyde, or obtained from their geothermally surroundings<sup>33</sup>, can be further assimilated by two pathways for formate assimilation: the H<sub>4</sub>F-WLP, and the rGlyP. Brockarchaeota code for three key enzymes for the H<sub>4</sub>F-WL including formate-tetrahydrofolate ligase FTL (present in DRTY735\_44, JZ-1.89, QC4\_43\_77, QZM\_A3\_48), methylene-H<sub>4</sub>F reductase (NADPH) MTHFR (DRTY-6.200, JZ-2.136 and QZM-A3\_48), and methylenetetrahydrofolate dehydrogenase (NADP +)/methylenetetrahydrofolate cyclohydrolase FoId (QZM\_A2, JZ-1.89, GD2, JZ\_2.136, JZZ.4, B27\_69, DRTY-6-80). This suggests that formate



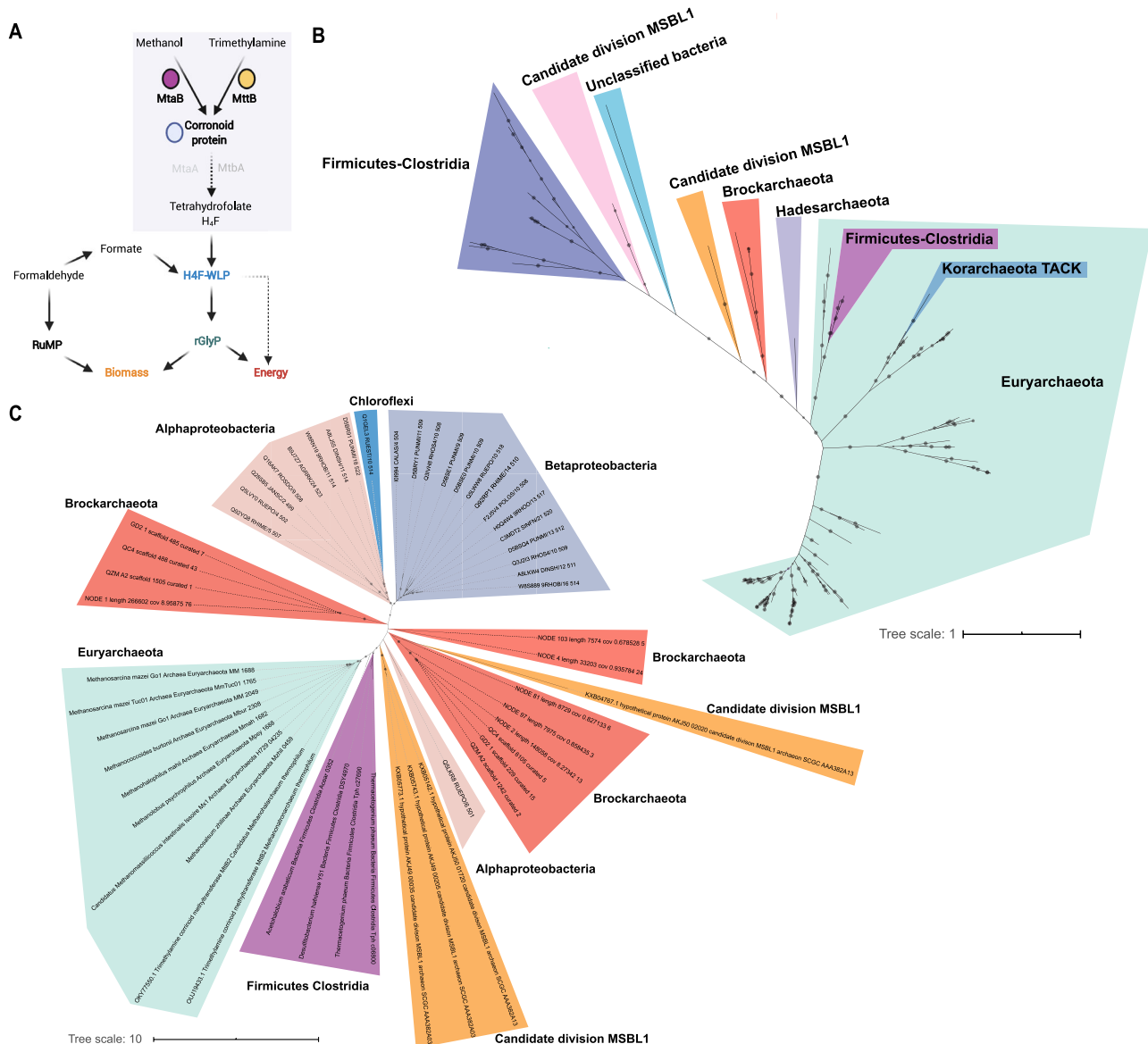


**Fig. 2** Location of samples from which *Brockarchaeota* genomes and 16S rRNA gene sequences have been recovered. **A** 16S rRNA gene tree of sequences derived from metagenomic and rRNA-based diversity surveys (NCBI accessions EU924237, KX213943, and KX213897). The eight complete 16S rRNA gene sequences of *Brockarchaeota* genomes described in this study are shown with their respective names and can be found in Supplementary data 7. Black circles in the tree represent bootstrap values using RAxML v8.2.10 with 100 replicates in the ARB software package v. 2.5b. Environmental information of each sequence were obtained from Integrated Microbial Genomes and Microbiomes database. Detailed information about the numbers displayed in the tree branches with their corresponding environmental information is described in Supplementary Data 8. **B** Geographic localization from which *Brockarchaeota* sequences were obtained. The size of the circle corresponds to the total number of *Brockarchaeota*-related sequences in each geographic location. The specific MAGs obtained in this study are shown in the map according to their temperature range. World map shape with the geographical coordinates of the *Brockarchaeota* 16S rRNA sequences shown in panel **A**, was created using ggmap<sup>80</sup>.

can be fixed to formyl-H<sub>4</sub>F, and then reduced to generate the active intermediate methylene-H<sub>4</sub>F that can be metabolized via rGlyP. The main component of the rGlyP is the glycine cleavage system (GCS), that catalyzes the reversible cleavage of glycine to CO<sub>2</sub>, CH<sub>2</sub>-H<sub>4</sub>F, and ammonia (NH<sub>3</sub>). The GCS is composed of glycine dehydrogenase (both subunits GcvPA and GcvPB present in B48\_G17 and JZ-2.136), aminomethyltransferase (GcvT, present in JZ-2.136), lipoate-binding protein (GcvH, present in B48\_G17 and JZ-2.136); Lpd, dihydrolipoyl dehydrogenase (present in DRTY735\_44 and DRTY-1.18). This suggests that in *Brockarchaeota*, like other anaerobes, the GCS operates in reductive direction by condensing the C1 moiety of CH<sub>2</sub>-H<sub>4</sub>F with CO<sub>2</sub> and ammonia to produce glycine. Glycine can react

with CH<sub>2</sub>-H<sub>4</sub>F to produce serine which can be deaminated to pyruvate by serine-dehydratase-like enzyme which the hot spring genotypes encode (including QZM\_A248\_33, QZM\_A348\_16, DRTY7.37, QC448\_22, DRTY-6.80, DRTY-1.18, DRTY-6.200, and DRTY735\_44). Then pyruvate can be further oxidized to acetate and produce ATP at substrate level phosphorylation for energy conservation and biomass production (see Fig. 4).

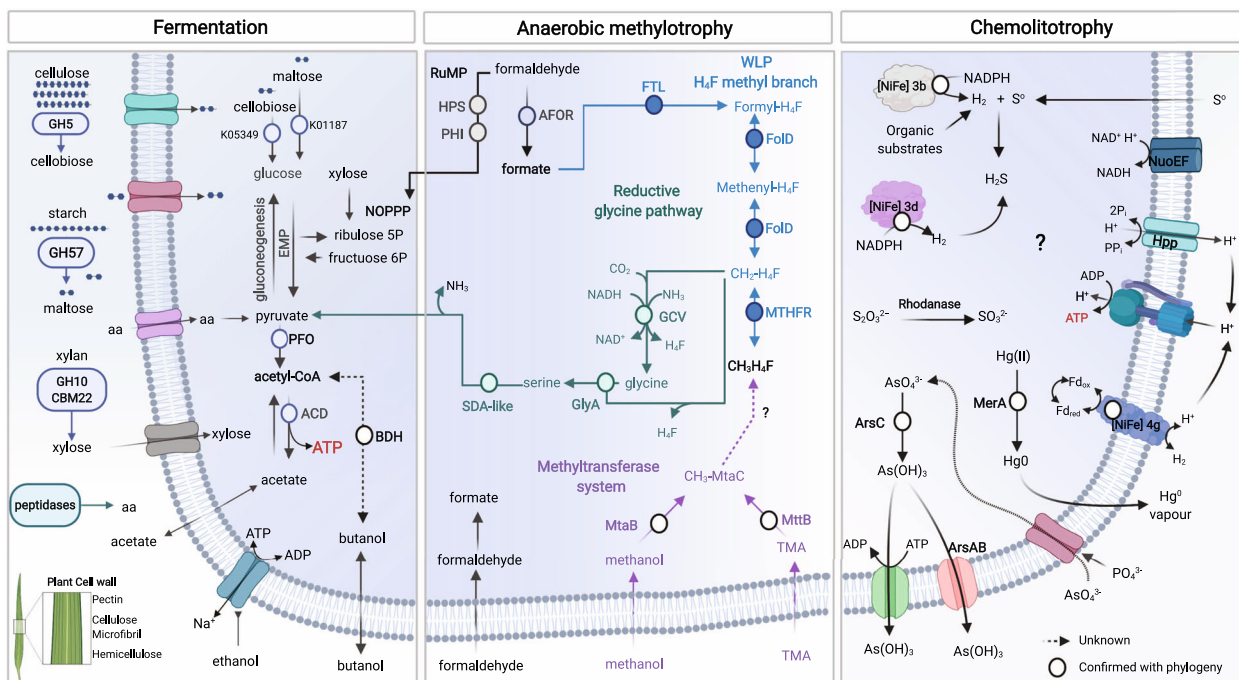
**An unknown pathway for butanol metabolism.** A phylogenetic analysis of alcohol dehydrogenases from the hot spring genomes revealed that they encode a butanol dehydrogenase BDH (Supplementary Fig. 4) that may catalyze the reversible



**Fig. 3 Proposed mechanism and phylogeny of the main enzymes of the methyltransferase system MTS.** **A** Schematic model of the hypothesized methylotrophic metabolism in Brockarchaeota. C1 units of methanol and trimethylamine TMA are transferred to tetrahydrofolate (H<sub>4</sub>F) via Methyltransferase System (MTS) to be ultimately converted to biomass or energy via convergent assimilatory routes for formaldehyde and formate reductive glycine pathway (rGlyP) and ribulose monophosphate pathway (RuMP). See more details in the text. **B** Phylogeny of methanol methyltransferases B subunit (MtaB). A total of seven methanol methyltransferase proteins were detected in Brockarchaeota and were aligned with 101 PFAM (PF12176) sequences, and 104 references obtained from Muñoz-Velasco et al.<sup>28</sup> and one Korarchaeota sequence from McKay et al.<sup>27</sup> **C** Phylogeny of trimethylamine methyltransferases A subunit (MttA). A total of 12 methanol methyltransferase proteins were detected in Brockarchaeota MAGs and were aligned with 24 PFAM (PF6253) sequences, and 18 references from Muñoz-Velasco et al.<sup>28</sup>. The alignments were done using MAFFT v7.450 (default parameters) and refined with MUSCLE v3.8.425 (default parameters), then masked (50% gaps) in Geneious Prime 2020.0.5. The trees were generated using a maximum likelihood-based approach using RAxML v8.2.10 (Gamma+LG+F model, raxmlHPC-PTHREADS-AVX -f a -m PROTGAMMAAUTO -N autoMRE -p 12345 -x 12345 -s).

conversion of butyraldehyde to butanol. Brockarchaeota BDH's are homologues to sequences from obligately anaerobic, thermophilic bacteria that can degrade complex plant saccharides such as xylan (i.e., *Caldicoprobacter oshimai*<sup>34</sup> and *Hungateiclostridium thermocellum*<sup>35</sup>) or cellulose (*Hungateiclostridium alkalicellulosi*). To investigate if Brockarchaeota can oxidize or produce butanol, we searched for genes involved in production of butanol in two model organisms; *Clostridium acetobutylicum*<sup>36,37</sup> which is one of the few organisms that produces butanol as a fermentation product, and *Saccharomyces cerevisiae*<sup>38</sup> involved in butanol and isopropanol

production. We found that Brockarchaeota genomes lack the key enzymes involved in the fermentation of pyruvate to butanol (butanal dehydrogenase, butyryl-CoA dehydrogenase, enoyl-CoA dehydratase, 3-hydroxyacyl-CoA dehydrogenase). However, most of the genomes encode a putative aldehyde dehydrogenase that could convert butyraldehyde to butyric acid. Also, we found a putative enoyl-CoA hydratase/isomerase protein that is coded by one bin (JZ-1.89), which could be involved in further converting butyric acid to acetyl-CoA. Our results suggest an alternative pathway for butanol oxidation that still remains unresolved (Supplementary Fig. 5).

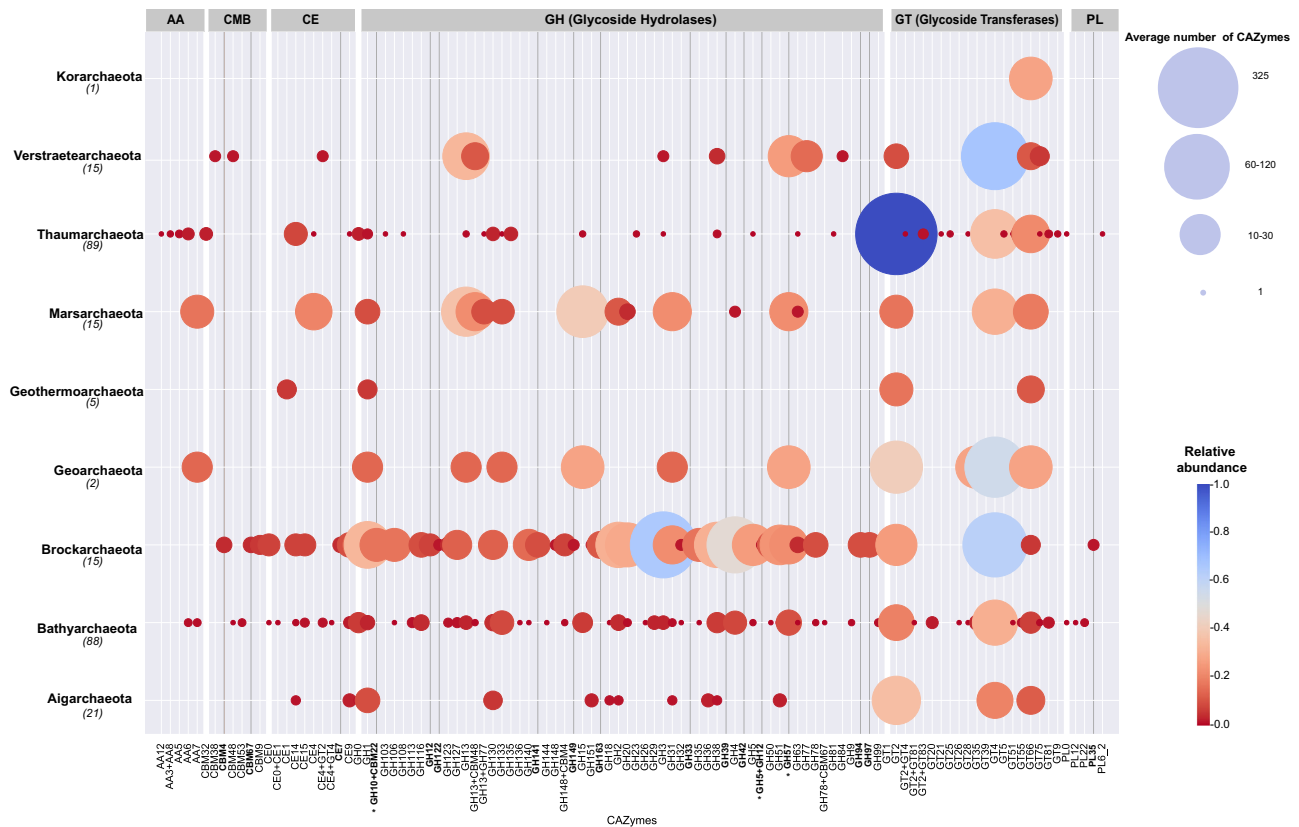


**Fig. 4 Overview of metabolic capabilities among distinct lineages within the Brockarchaeota phylum.** Fermentation panel: Brockarchaeota can breakdown and assimilate of high molecular-weight plant-derived polysaccharides via fermentation and produce ATP by substrate-level phosphorylation via concerted action of PFO (present in all except QC4\_43 and JZ-2.136), and ACD, present in hot spring genomes (GD2\_1, JZ-1.89, DRTY-6.80, JZ-1.89, QC4\_43, QZM\_A2, QZM\_A3). Acetate can also be assimilated back to acetyl-CoA by an acetyl-CoA synthetase (ACS). Anaerobic methylo-trophy panel: convergent pathways for formaldehyde and methylated compound assimilation also exist in Brockarchaeota. The RuMP (shown in black), reductive glycine pathway (rGlyP, shown in green), tetrahydrofolate methyl branch ( $H_4F$ -methyl, shown in blue), and methyltransferase system (shown in purple) that can lead to biomass production and energy conservation. An undescribed protein may be involved in the transfer of C1 moiety from the corrinoid protein to tetrahydrofolate (dashed purple arrows) and be metabolized via the  $H_4F$  methyl branch of the WLP (blue arrows). Chemolithotrophy panel: Brockarchaeota found in hot springs may increase their energy yield from the oxidation of geothermally abundant compounds such as mercury (Hg), arsenate ( $AsO_4^{3-}$ ), hydrogen ( $H_2$ ), and elemental sulfur ( $S^0$ ). Abbreviations. Pathways: EMP Embden-Meyerhof-Parnas, NOPPP Non-Oxidative Pentoses Phosphate Pathway, WLP Wood-Ljungdahl pathway. Enzymes: PFO pyruvate ferredoxin oxidoreductase, and ACD acetate-CoA ligase (ADP-forming), PFL pyruvate formate lyase, ACS acetyl-CoA synthetase, AFOR tungsten-dependent aldehyde ferredoxin oxidoreductase, BDH butanol dehydrogenase (FTL) tetrahydrofolate ( $H_4F$ ) ligase, FolD methenyl-THF cyclohydrolase/methylene-THF dehydrogenase, methylene-THF reductase MTHFR (methyltransferase system (MT), GlyA, serine hydroxymethyltransferase (GlyA), SDA-like serine deaminase-like protein, GCS glycine-cleavage system. Compounds: TMA trimethylamine. Dashed lines indicate potential novel enzymatic steps. White circles indicate enzymatic steps confirmed with phylogeny (more details can be found in Supplementary Figure 3). Created with BioRender.com.

**Utilization of extracellular organic carbon and detrital proteins.** Brockarchaeota may be able to degrade a variety of organic carbon compounds. They may utilize hexoses via Embden-Meyerhof-Parnas (EMP) pathway (Supplementary Fig. 6) and pentoses (xylose isomerase XylA and xylulose kinase XylB) via the isomerase pathway (Supplementary Fig. 7). These enzymes were previously only found in bacterial thermophiles and halophilic archaea that ferment complex compounds and degrade xylose suggesting a similar physiology in Brockarchaeota<sup>39</sup>. Once transported into the cell, carbon complex compounds could enter the central metabolism and be converted to acetate and  $H_2$  via acetogenic fermentation. The ATP conserving step for sugar or pyruvate fermentation to acetate could be catalyzed by acetate-CoA ligase in the hot spring genomes. Acetate can also be converted to acetyl-CoA by acetyl-CoA synthetase (ACS), thus acetate might be a source of carbon and energy in the absence of other substrates in hot spring Brockarchaeota. The presence of pyruvate ferredoxin oxidoreductase (PFO) that couples pyruvate oxidation to  $H_2$  production, generating acetyl-CoA, could support fermentative metabolism via degradation of either acetate, pyruvate, hexoses, or pentoses. Brockarchaeota genomes code a wide repertoire of ATPases such as the plasma-membrane proton-efflux P-type ATPase (only present in the hot spring genomes),

$Zn^{2+}/Cd^{2+}$ -exporting ATPase (present in DRTY7.37), and the V/A-type  $H^+ / Na^+$ -transporting ATPase (in most of the genomes). The existence of ATPase in Brockarchaeota suggests that members of these genotypes have the additional ability to couple acetogenic fermentation to membrane potential generation.

To complement their ability to degrade xylose, Brockarchaeota also contain a relatively high number of carbohydrate-active enzymes (average of 27 CAZymes per genome) which is 3 times what has been observed in other TACK archaea (Supplementary Data 12 and Fig. 5). Ten of the 15 Brockarchaeota genomes have genes with similarity to  $\alpha$ -L-fucosidase involved in the degradation of xyloglucan, which is the major component of hemicellulose in plant-cell walls<sup>40</sup>. All the hot spring genomes encode GH3 family proteins which may be involved in biomass degradation, but these proteins play other roles in cell wall remodeling, energy metabolism, and pathogen defense<sup>41</sup>. The hot spring genotypes contain a wider repertoire of CAZymes than the deep-sea GB genomes. Among these are four predicted to be extracellular glycoside hydrolases, which are involved in the breakdown of high molecular-weight plant-derived polysaccharides, primarily xylanes, cellulose, and starch. Comparison of the CAZymes across the TACK superphylum revealed that 17 extracellular enzymes, including those for the degradation xylanes, are unique



**Fig. 5 Carbohydrate-active enzymes (CAZymes) encoded by genomes belonging to the TACK superphylum including Brockarchaeota.** The total number of CAZymes per phylum was normalized by the total number of genomes described for each phylum shown in parenthesis. Auxiliary activities (AA), carbohydrate-binding module (CBM), carbohydrate esterases (CE), glycoside-hydrolases (GH), glycoside transferases (GT), polysaccharide lyases (PL). Unique Brockarchaeota CAZymes are shown in darker lines. Extracellular CAZymes are shown in asterisks described in Supplementary Data 12 sheet 3.

to Brockarchaeota (Supplementary Data 12). The diversity and abundance of CAZymes in members of the TACK superphylum highlights that despite the low number of sequenced Brockarchaeota genomes (15 described in this study), compared to Thaumarchaeota (89), they contain a greater variety and number than their closest relatives. Brockarchaeota also encode a wide range of peptidases. With exception of some bins (B48\_G17, DRTY-6.200, DRTY7\_35\_44, JZ-1.89, JZZ.4, and QC4\_43), they all encode potential extracellular peptidases indicating a potential role of Brockarchaeota in protein remineralization. Many of these are so unique that it was not possible to identify their specific substrates (Supplementary Data 13).

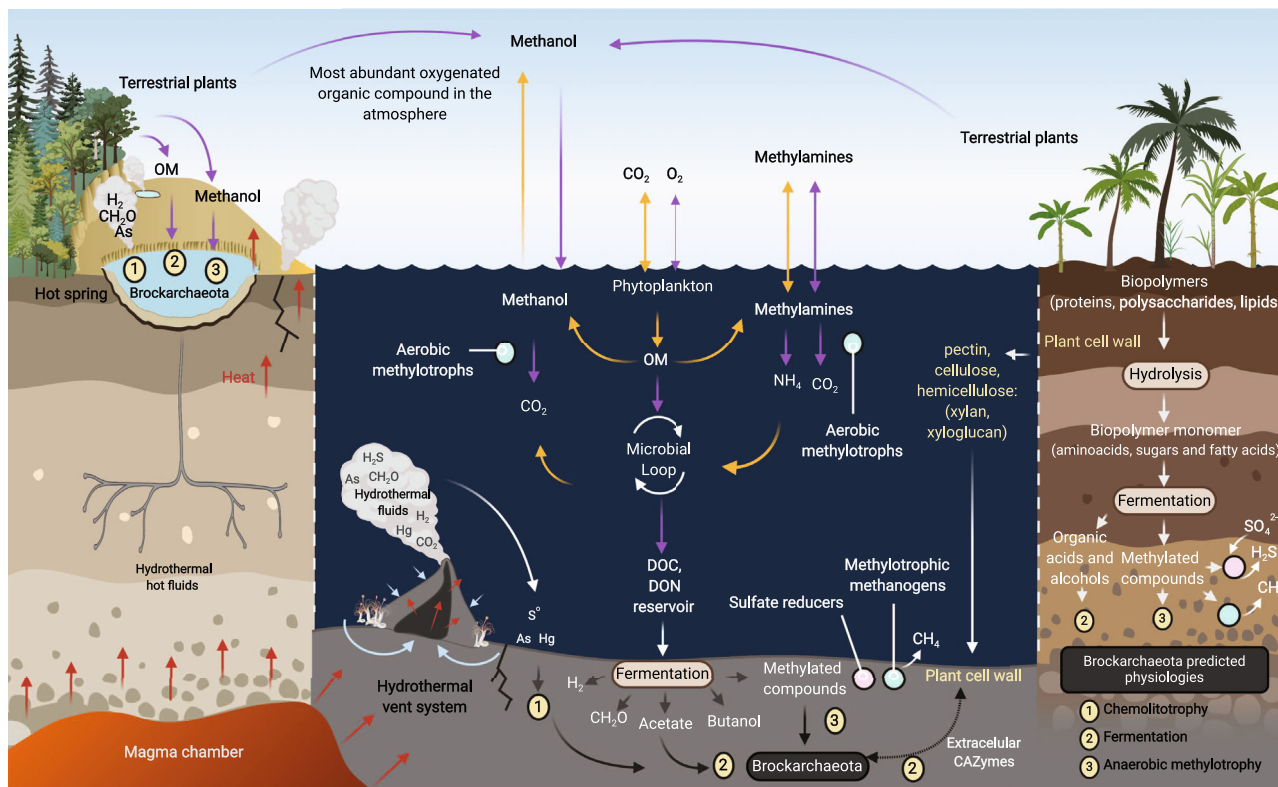
#### Mercury, sulfur, arsenate, and hydrogen chemolithotrophy.

Geothermal ecosystems such as shallow, and deep-sea vents, volcanoes, geysers, hot springs, and fumaroles are natural sources of mercury (Hg)<sup>42</sup>. Hg-resisting microorganisms are known to be enriched in deep-sea hydrothermal vents and in terrestrial geothermal springs. Three hot spring Brockarchaeota genomes (DRTY735\_44, DRTY-1.18 and DRTY7.37) encode mercuric reductase (MerA), the central enzyme in the microbial mercury detoxification system. MerA transforms the extremely toxic Hg(II) to metallic Hg(0), being potentially involved in mercury detoxification<sup>42</sup>. Brockarchaeota MerA and closely related proteins were used to generate a mercury reductase phylogeny (Supplementary Fig. 8), indicating that Brockarchaeota possess a previously uncharacterized class of MerA, which are related to other archaea (Crenarchaeota, Methanomicrobia, DPANN, and Asgard).

Interestingly, Brockarchaeota also code the arsenic detoxification system that acts by decreasing the intracellular arsenic concentration by pumping out arsenate that enters the cell, thus preventing the metals from accumulating and denaturing proteins<sup>43</sup>. The intracellular dependent arsenate reductase (ArsC, K03741) that catalyzes the reduction of arsenate  $\text{AsO}_4^{3-}$  to arsenite  $\text{As(OH)}_3$  (Fig. 4), is present in most hot spring genomes (Supplementary Data 6). Phylogeny of Brockarchaeota ArsC (Supplementary Fig. 9) indicates that they belong to a deep uncharacterized branch of Thioredoxin-coupled clade, that has been mainly described in Firmicutes<sup>44</sup>. In agreement with the geothermal origin of Brockarchaeota genomes, homologous ArsC sequences recovered from geothermally active environments belonging to uncultured Bathyarchaeota or Thaumarchaeota<sup>45,46</sup>, which could potentially be Brockarchaeota, or have a similar arsenate metabolism. The presence of ArsC and the energy-dependent related detoxification proteins, could also indicate that Brockarchaeota in hot spring genomes could use arsenate as terminal electron acceptor, as seen in other bacteria, yet the exact molecular mechanism of this process is unknown<sup>43,47</sup>.

Furthermore, similar to other heterotrophic fermentative hyperthermophilic archaea<sup>48,49</sup> Brockarchaeota might be able to reduce elemental sulfur during fermentative growth and produce  $\text{H}_2\text{S}$  due to the presence of [NiFe] Group 3b hydrogenases (Supplementary Fig. 10). During carbohydrate fermentation in the absence of sulfur, [NiFe] Group 3b hydrogenase can catalyze the production of  $\text{H}_2$  with NADPH or NAD(P)H as the electron donor. However, in the presence of sulfur, Brockarchaeota might have the ability to reduce sulfur using  $\text{H}_2$  or organic substrates as





**Fig. 6 A model of the biogeochemical roles of Brockarchaeota in the anaerobic carbon cycle.** C1 and methylated compounds, such as methanol or methylamines, are utilized biologically as carbon and energy sources in the ocean and deep-sea sediments resulting in a considerable carbon reservoir. The biodegradation of organic carbon in the water column and subsurface is a source of these compounds. The utilization of methyl compounds as precursors in methane synthesis is confined to a small group of methylotrophic methanogens (i.e., Verstraetearchaeota). The only described anaerobic methylotrophs include members of methanogenic archaea, acetogenic bacteria, and sulfate-reducing bacteria. These organisms compete for these compounds geochemically produced in anoxic settings. Brockarchaeota may recycle methanol and methylamines in anoxic environments without methane formation and may be sequestered in deep sea sediments and hot springs. Orange and purple arrows represent sources and sinks, respectively. Organic Matter (OM) includes dissolved and particulate organic matter feeding the microbial loop (adapted from Evans et al., 2019 and Zhuang et al., 2018).

electron donors, a widespread physiology in hyperthermophilic archaea living in geothermally active environments (volcanic habitats, hot springs, or marine sediments).

Hydrogen is also abundant in geothermally active systems due to volcanic processes<sup>50</sup>. Brockarchaeota might be able to use 3b [NiFe]-hydrogenases for H<sub>2</sub> oxidation with NADP<sup>+</sup> or NAD(P)<sup>+</sup> as an electron acceptor<sup>51</sup>. The hot spring genomes also encode oxygen-tolerant group 3d [NiFe]-hydrogenases, which may allow them to transfer electrons between NAD(P)H and H<sub>2</sub> depending on the availability of electron acceptors. Group 3d [NiFe]-hydrogenases are abundant in metagenomes from hot springs where microbial communities are relatively stable despite partial pressure of oxygen fluctuations<sup>52</sup>. Group [NiFe] 3b hydrogenases may also make it possible for these archaea to reduce elemental sulfur to H<sub>2</sub>S during fermentative growth. During carbohydrate fermentation in the absence of sulfur, Group 3b [NiFe]-hydrogenases might catalyze the production of H<sub>2</sub> with NADPH or NAD(P)H as the electron donor. Therefore, Brockarchaeota might have the ability to reduce sulfur, using H<sub>2</sub> or organic substrates as electron donors, which is common in hyperthermophilic archaea living in geothermally active environments<sup>53</sup>.

**Discussion**

Brockarchaeota gene content indicates they are facultative or obligate anaerobic fermentative organisms that produce acetate, CO<sub>2</sub>, and H<sub>2</sub> as byproducts (see Supplementary Information for details). Some Brockarchaeota have unique pathways for

non-methanogenic methylotrophy. This puts them a unique ecological position in nature, where they degrade abundant methylamines in anoxic environments without the production of methane (Fig. 6). Brockarchaeota are also able to degrade complex carbon compounds such as xylan. Xylans are a major structural polysaccharide in plant cells, and is the second most abundant polysaccharide in nature, accounting for approximately one-third of all renewable organic carbon on Earth after cellulose<sup>54,55</sup>. This suggests that Brockarchaeota are players in organic matter degradation in geothermally active environments. Interestingly, detrital proteins can be used as a substrate by Brockarchaeota, indicating potential role in protein remineralization in geothermally active environments.

The protein repertoire of GB and hot spring genomes have some important distinctions that reflect different anaerobic metabolisms. GB genomes appear to be obligately fermenting organisms that rely mostly on substrate-level phosphorylation since they lack all the complexes for the respiratory chain with exception of the ATPase. In contrast, hot spring genomes appear to have mechanisms to increase their ATP yield including the use of geothermally derived inorganic substrates as possible terminal electron acceptors such as mercury (Hg), arsenic (As), and hydrogen (H<sub>2</sub>). Deep-sea hydrothermal vents, hot springs, and fumaroles are natural sources of Hg<sup>42</sup>, H<sub>2</sub><sup>52</sup>, arsenic<sup>56</sup>, and sulfur<sup>57</sup>.

The discovery of Brockarchaeota genomes from sediments around the world, overlooked by conventional rRNA gene

diversity approaches, highlights the need for further exploration of subsurface microbial communities. Although they are relatively low in abundance in the communities described here, the addition of these genomes to public databases, will enhance their detection in future environmental studies, like other recently described novel archaeal lineages<sup>1,58</sup>. A lack of recognition of their existence prior to this limited our ability to fully describe sediment community structure and function. Given their broad distribution, and versatile carbon metabolism, they are likely key players in global carbon cycling. However, this first description is limited to genomic characterization, thus culturing or in activity measurements are needed to confirm their physiological activities<sup>59</sup>. Overall, the description of this new phylum enhances our understanding of biodiversity of archaea and suggests they are mediating unique roles in anoxic carbon cycling.

## Methods

**Metagenomic assembly and binning.** Two MAGs (B48\_G17 and B27\_G9) were obtained from Guaymas Basin sediments (Gulf of California; 27°N0.388, 111°W24.560) and were obtained as part of a larger study of these hydrothermal marine sediments<sup>25</sup>. Both samples were collected from the same location but G9 was sampled from 0–3 cm and G17 from 12–15 cm depth. The sediment cores from which these two MAGs were binned from were collected during Alvin dive 4571\_4 in 2009 using polycarbonate cores (45–60 cm in length, 6.25 cm interior diameter), subsampled into cm layers under N<sub>2</sub> gas in the ship's laboratory and immediately frozen at –80°C. Details on the sampling site and metagenomic sequencing effort is provided in Dombrowski et al.<sup>25</sup>. Briefly, total DNA from ≥10 g of sediment from each sample was extracted using the MoBio PowerMax soil kit using the manufacturer's instructions and adjusted to a final concentration of 10 ng/μl of each sample (using a total amount of 100 ng). Libraries for paired-end Illumina (HiSeq–2500 1TB) sequencing were prepared by the Joint Genome Institute (JGI). Sequencing was performed on an Illumina HiSeq 2500 machine using the paired-end 2 × 125 bp run-type mode. All runs combined provided a total of ~280 gigabases of sequencing data. Quality control and sequence assembly were performed by JGI. Briefly, sequences were trimmed and screened for low-quality sequences using bbtools (<https://jgi.doe.gov/data-and-tools/bbtools/>) and assembled using megahit v1.0.6 using the following options: --k-list23,43,63,83,103,123. For further binning, only scaffolds ≥2000 bps were included.

Metagenomic binning was performed on individual assemblies using the binning tools ESOM, Anvi'o (v2.2.2)<sup>60</sup> and Metabat (v1)<sup>61</sup>. ESOM bins were extracted using getClassFasta.pl and the command -loyal 51. Anvi'o was run with default parameters and metabat was run using the following settings: --minProb 75 --minContig 2000 --minContigByCorr 2000. Results from the three different binning tools were combined using DAS Tool (version 1.0) as follows: DAS\_Tool.sh -i Anvio\_contig\_list.tsv, Metabat\_contig\_list.tsv, ESOM\_contig\_list.tsv -l Anvio, Metabat, ESOM -c scaffolds.fasta --write\_bins 1.

Eight MAGs (DRTY-1.18, DRTY-6.80, DRTY-6.200, DRTY7.37, JZ-1.89, JZ-2.136, JZZ\_4, and DRITY7) were recovered from hot springs in Yunnan, China collected in January of 2016 and May of 2017 in several hot springs (Supporting Data 1). Five additional MAGs (QC4\_43, QC4\_48, GD2\_1\_47\_42, QZM\_A2, QZM\_A3) were reconstructed from hot springs in Tibet in August of 2016. Sequencing was done on an Illumina HiSeq4000 (Beijing Novogene Bioinformatics Technology Co., Ltd). These samples were assembled using metaSPADES (version 3.9.1), with a k-mer set of “21, 33, 55, 77, 99, 127”. For each sample only scaffolds larger than 2500 bp were binned using MetaBAT (v2.12.1) with default parameters, considering both tetranucleotide frequencies (TNF) and scaffold coverage information. The scaffolds from the obtained bins and the unbinned scaffolds were visualized using ESOM with a minimum length of 2500 bp and maximum length of 5000 bp as previously described<sup>62</sup> and the bins were modified by removing any out-of-range scaffolds (indicated by sequence points) or adding any unbinned scaffolds using ESOM related scripts<sup>37</sup>. MAGs from Tibet hot springs with scaffolds ≥1000 bp were uploaded to ggkbase (<http://ggkbase.berkeley.edu/>), and the bins from ESOM analyses were evaluated and modified manually at ggkbase based on GC content, coverage, and taxonomic information of scaffolds. MAGs from Tengchong hot springs were reassembled using SPAdes (version 3.9.1) under the “careful” mode with the same k-mers. During this step, the reads used for the assemblies were recruited by mapping clean reads to the curated genome bins using BBmap (v35.85; <http://sourceforge.net/projects/bbmap/>). The accuracy of all the MAGs was evaluated by calculating the percentage of completeness and gene duplications using CheckM lineage\_wf (v1.0.5).

**Phylogenetic analyses.** A phylogenetic tree was generated as recently described in ref. 1. Briefly, 37 conserved marker proteins were extracted using phylosift<sup>63</sup>, in a genomic dataset containing 3549 archaeal genomes including Brockarchaeota, and 40 bacterial genomes. An alignment of the proteins extracted from a total of 3599 genomes was generated using MAFFT v7.450<sup>64</sup> (algorithm autoselection) with a

BLOSUM62 scoring and contains 4962 characters after masking gaps present in at least 50% of the taxa. A tree was constructed with IQtree<sup>65</sup> v1.6.11 with a best fit LG + F + R10 model selected using the Bayesian Information Criterion (BIC) and bootstraps are based on 1,000 replicated trees. Command-line options -bb 1000 -bnni The bacterial genomes were used as an outgroup. The 16S rRNA sequences were extracted from Brockarchaeota genomes using Barnmap v.0.9 (<https://github.com/Victorian-Bioinformatics-Consortium/barnmap>) with the following parameters: --kingdom arc --lencutoff 0.2 --reject 0.3 --evaluate. The obtained sequences (Supplementary Data 7). were used for a 16S rRNA gene phylogeny that included sequences derived from metagenomic surveys (NCBI accession EU924237, KX213943, and KX213897) and the IMG database. We used the 16S rRNA genes from the MAGs to search these databases to identify additional Brockarchaeota genes. The rRNA phylogeny was generated using RAxML within the ARB software package (v. 2.5b). using default parameters.

**Metabolic predictions.** Gene predictions for individual genomes were performed using Prodigal<sup>66</sup> (V2.6.2, default settings). Predicted genes of individual genomes were further characterized using a combination of several databases: KofamKOALA<sup>67</sup>, Interproscan v5.31.70<sup>68</sup>, HydDB<sup>69</sup>, dbCAN2<sup>70</sup>, MEBS<sup>71</sup>, METABOLIC<sup>72</sup>, and MEROPS v12.1<sup>73</sup>. For KofamKOALA only hits above the predefined threshold for individual KOs were selected. Hydrogenases were extracted using the reference database described in Greening et al. and Søndergaard et al.<sup>52,69</sup> where there was conflict, the protein was manually reanalyzed using BLAST against non-redundant protein database, and genomic organization and annotation was confirmed using a web-based tool Operon Mapper<sup>74</sup>. The detected hydrogenases were used to generate a phylogenetic tree as previously described in Seitz et al.<sup>7</sup>. Hits for key metabolic marker genes were verified across different databases KofamKOALA, PFAMv31 and TIGR-FAMs and HydDB and were further verified using BLASTP using the NCBI web server tool. Genes encoding for carbohydrate degradation enzymes described in the carbohydrate-active enzymes (CAZymes) database<sup>75</sup> were identified by only retaining hits recovered by ≥2 tools. Protein localization of the selected CAZymes and peptidases was determined with the command line version of Psort v3.0<sup>76</sup> using the options -a and -t for archaeal genomes in tabular format files. Finally, the presence of specific protein families was obtained with MEBS. The annotation was performed in a genomic dataset of 250 publicly available TACK genomes (Supplementary Data 3) that were also used for the CAZymes annotation.

**Methyl coenzyme M reductase screening.** The *mcrA* gene was identified using GraftM v0.10.2<sup>77</sup> across metagenome assemblies where Brockarchaeota genomes from hot springs were detected<sup>78</sup>. The *mcrA*-containing scaffolds with sequence length <2.5 Kbp were discarded since scaffolds with short length were not used during the genome binning step. The taxonomic information of the corresponding bins which contain *mcrA* genes were determined using either GTDBtk v0.3.2<sup>79</sup> or phylogenetic placement (as reported in Supplementary Data 11). The *mcrABGCD* genes were identified in metagenome assemblies from deep-sea assemblies previously described<sup>25</sup> (Guay17 and Guay9; IMG genome ID 3300014887 and 3300013103, respectively).

**Relative abundance of Brockarchaeota in communities.** The relative abundance of the MAGs from deep sea samples was obtained from Supplementary Data 3 in Dombrowski et al.<sup>25</sup>. Only samples G9 and G7 are shown, and the data is sorted according to the relative abundance of those corresponding samples. GB MAGs include a post publication manual refinement in the taxonomy according to the recent archaea tree of life<sup>1</sup>. The relative abundance of MAGs from Tengchong samples was computed using the bin\_abundance.py script (<https://github.com/valdeanda/MetaGaia>). For each MAG, total length of mapped reads for individual scaffolds (mapped reads using BWA algorithm) is summed up and the total is then divided by the MAG size in bp. This number is then divided by the total number of reads to obtain the relative abundance. The final relative abundance is multiply it by 100000000 for readability purposes. For the Tibet samples, the genome bins obtained for a given sample, the sequencing coverage was determined by read mapping using Bowtie2 and coverage calculation using the jgi\_summarize\_bam\_contig\_depths script from MetaBAT<sup>61</sup>. The relative abundance of a given genome bin was calculated as its sequencing coverage divided by the total sequencing coverage of all genome bins in the corresponding sample (Tengchong samples as previously described)<sup>6</sup>.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The final assembled and annotated genomic sequences of Brockarchaeota from deep sea sediments (B27\_G9 and B48\_G17) have been deposited in NCBI under BioProject ID PRJNA362212: BioSample id SAMN09215183 and SAMN09214986, respectively. Sequence data and sample information of Brockarchaeota from hot springs are available at NCBI under Bio Project ID PRJNA544494. All the NCBI accession numbers for the MAGs described in this study are provided in Supplementary Data 1. Raw data for Figs. 1, 2 are provided as Supplementary Data 3 and 7 respectively. Any other datasets

generated and/or analyzed during the current study are available from the corresponding authors on request.

### Code availability

All custom scripts have been made available at [https://github.com/valdeanda/Useful\\_scripts](https://github.com/valdeanda/Useful_scripts). The protocol to compute MAGs relative abundance is available at MetaGaia (<https://github.com/valdeanda/MetaGaia>) bin\_abundance.py script.

Received: 2 July 2020; Accepted: 18 March 2021;

Published online: 23 April 2021

### References

- Baker, B. J. et al. Diversity, ecology and evolution of Archaea. *Nat. Microbiol.* **5**, 887–900 (2020).
- Baker, B. J., Appller, K. E. & Gong, X. New microbial biodiversity in marine sediments. *Ann. Rev. Mar. Sci.* **13**, 161–175 (2020).
- Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
- Kozubal, M. A. et al. Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* **7**, 622–634 (2013).
- Jay, Z. J. et al. Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats. *Nat. Microbiol.* **3**, 732–740 (2018).
- Hua, Z. S. et al. Genomic inference of the metabolism and evolution of the archaeal phylum Aigarchaeota. *Nat. Commun.* **9**, 2832 (2018).
- Seitz, K. W. et al. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **10**, 1822 (2019).
- Spang, A. et al. Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism. *Nat. Microbiol.* **4**, 1138–1148 (2019).
- Orsi, W. D. et al. Metabolic activity analyses demonstrate that Lokiarchaeon exhibits homoacetogenesis in sulfidic marine sediments. *Nat. Microbiol.* **5**, 248–255 (2020).
- Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
- Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
- Williams, T. A. et al. Integrative modeling of gene and genome evolution roots the archaeal tree of life. *Proc. Natl Acad. Sci. U.S.A.* **114**, E4602–E4611 (2017).
- Spang, A., Caceres, E. F. & Ettema, T. J. G. Genomic exploration of the diversity, ecology, and evolution of the archaeal domain of life. *Science* **357**, eaaf3883 (2017).
- Trembath-Reichert, E. et al. Methyl-compound use and slow growth characterize microbial life in 2-km-deep seafloor coal and shale beds. *Proc. Natl Acad. Sci. USA* **114**, E9206–E9215 (2017).
- Zhuang, G. C., Peña-Montenegro, T. D., Montgomery, A., Hunter, K. S. & Joye, S. B. Microbial metabolism of methanol and methylamine in the Gulf of Mexico: insight into marine carbon and nitrogen cycling. *Environ. Microbiol.* **20**, 4543–4554 (2018).
- Chistoserdova, L. Modularity of methylophily, revisited. *Environ. Microbiol.* **13**, 2603–2622 (2011).
- Chistoserdova, L. & Kalyuzhnaya, M. G. Current trends in methylophily. *Trends Microbiol.* **26**, 703–714 (2018).
- Sun, J., Mausz, M. A., Chen, Y. & Giovannoni, S. J. Microbial trimethylamine metabolism in marine environments. *Environ. Microbiol.* **21**, 513–520 (2018).
- Zhuang, G.-C., Montgomery, A. & Joye, S. B. Heterotrophic metabolism of C1 and C2 low molecular weight compounds in northern Gulf of Mexico sediments: controlling factors and implications for organic carbon degradation. *Geochim. Cosmochim. Acta* **247**, 243–260 (2019).
- Vanwonterghem, I. et al. Methylophily discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
- Lazar, C. S. et al. Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Environ. Microbiol.* **18**, 1200–1211 (2016).
- Zhuang, G. Methylophily methanogenesis and potential methylated substrates in marine sediment. (University of Bremen, 2014).
- Richards, M. A. et al. Exploring hydrogenotrophic methanogenesis: a genome scale metabolic reconstruction of *Methanococcus marispludis*. *J. Bacteriol.* **198**, 3379–3390 (2016).
- Sousa, D. Z. et al. The deep-subsurface sulfate reducer *Desulfotomaculum kuznetsovii* employs two methanol-degrading pathways. *Nat. Commun.* **9**, 239 (2018).
- Dombrowski, N., Teske, A. P. & Baker, B. J. Extensive metabolic versatility and redundancy in microbially diverse, dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **9**, 4999 (2018).
- Fricke, W. F. et al. The genome sequence of *Methanospaera stadmanae* reveals why this human intestinal archaeon is restricted to methanol and H<sub>2</sub> for methane formation and ATP synthesis †. *J. Bacteriol.* **188**, 642–658 (2006).
- McKay L., et al. Co-occurring genomic capacity for anaerobic methane and dissimilatory sulfur metabolisms discovered in the Korarchaeota. *Nat. Microbiol.* **4**, 614–622 (2019).
- Muñoz-Velasco, I. et al. Methanogenesis on early stages of life: ancient but not primordial. *Orig. Life Evol. Biosph.* **48**, 407–420 (2019).
- Adam, P. S., Borrel, G. & Gribaldo, S. An archaeal origin of the Wood–Ljungdahl H4MPT branch and the emergence of bacterial methylophily. *Nat. Microbiol.* **4**, 2155–2163 (2019).
- Swan, B., Reifel, K. & Valentine, D. Periodic sulfide eruptions impact microbial community structure and diversity in the water column of a hypersaline lake. *Aquat. Microb. Ecol.* **60**, 97–108 (2010).
- Adam, P. S., Borrel, G. & Gribaldo, S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *PNAS* **115**, E5837 (2018).
- Orita, I. et al. The ribulose monophosphate pathway substitutes for the missing pentose phosphate pathway in the archaeon *Thermococcus kodakaraensis*. *J. Bacteriol.* **188**, 4698–4704 (2006).
- Urschel, M. R., Kubo, M. D., Hoehler, T. M., Peters, J. W. & Boyd, E. S. Carbon source preference in chemosynthetic hot spring communities. *Appl. Environ. Microbiol.* **81**, 3834–3847 (2015).
- Yokohama, H., Wagner, I. D. & Wiegler, J. *Caldicoprobacter oshimai* gen. nov., sp. nov., an anaerobic, xylanolytic, extremely thermophilic bacterium isolated from sheep faeces, and proposal of *Caldicoprobacteraceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **60**, 67–71 (2010).
- Zhang, X. et al. *Petroclostridium xylanilyticum* gen. Nov., sp. nov., a xylan-degrading bacterium isolated from an oilfield, and reclassification of clostridial cluster iii members into four novel genera in a new hungateiclostridiaceae fam. nov. *Int. J. Syst. Evol. Microbiol.* **68**, 3197–3211 (2018).
- Girbal, L., Croux, C., Vasconcelos, I. & Soucaille, P. Regulation of metabolic shifts in *Clostridium acetobutylicum* ATCC 824. *FEMS Microbiol. Rev.* **17**, 287–297 (1995).
- Qi, F. et al. Improvement of butanol production in *Clostridium acetobutylicum* through enhancement of NAD(P)H availability. *J. Ind. Microbiol. Biotechnol.* **45**, 993–1002 (2018).
- Branduardi, P., Longo, V., Berterame, N. M., Rossi, G. & Porro, D. A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* **6**, 68 (2013).
- Johnsen, U. & Schönheit, P. Novel xylose dehydrogenase in the halophilic archaeon *Halorcula marismortui*. *J. Bacteriol.* **186**, 6198–6207 (2004).
- Ravachol, J. et al. Mechanisms involved in xyloglucan catabolism by the cellulose-producing bacterium *Ruminiclostridium cellulolyticum*. *Sci. Rep.* **6**, 22770 (2016).
- Macdonald, S. S., Blaukopf, M. & Withers, S. G. N-acetylglucosaminidases from CAZy family GH3 are really glycoside phosphorylases, thereby explaining their use of histidine as an acid/Base catalyst in place of glutamic acid. *J. Biol. Chem.* **290**, 4887–4895 (2015).
- Wang, Y. et al. Environmental conditions constrain the distribution and diversity of Archaeal merA in Yellowstone National Park, Wyoming, U.S.A. *Microb. Ecol.* **62**, 739–752 (2011).
- Nunes, C. I. P. et al. *ArsC3* from *Desulfovibrio alaskensis* G20, a cation and sulfate-independent highly efficient arsenate reductase. *J. Biol. Inorg. Chem.* **19**, 1277–1285 (2014).
- Silver, S. & Phung, L. T. Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl. Environ. Microbiol.* **71**, 599–608 (2005).
- Colman, D. R., Lindsay, M. R. & Boyd, E. S. Mixing of meteoric and geothermal fluids supports hyperdiverse chemosynthetic hydrothermal communities. *Nat. Commun.* **10**, 681 (2019).
- Zhou, Z. et al. Genome- and community-level interaction insights into carbon utilization and element cycling functions of Hydrothermarchaeota in hydrothermal sediment. *mSystems* **5**, e00795-19 (2020).
- Rabus, R., Venceslau, S. S., Lars, W., Wall, J. D. & Pereira, I. A. C. A post-genomic view of the ecophysiology, catabolism and biotechnological relevance of sulphate-reducing prokaryotes. *Adv. Micro. Physiol.* **66**, 55–321 (2015).
- Tóth, A., Takács, M., Groma, G., Rákhelyi, G. & Kovács, K. L. A novel NADPH-dependent oxidoreductase with a unique domain structure in the hyperthermophilic Archaeon, *Thermococcus litoralis*. *FEMS Microbiol. Lett.* **282**, 8–14 (2008).
- Ma, K., Weiss, R. & Adams, M. W. W. Characterization of hydrogenase II from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction. *J. Bacteriol.* **182**, 1864–1871 (2000).
- Jenney, F. E. & Adams, M. W. W. Hydrogenases of the model hyperthermophiles. *Ann. N. Y. Acad. Sci.* **1125**, 252–266 (2008).
- Van Haaster, D. J., Silva, P. J., Hagedoorn, P. L., Jongejan, J. A. & Hagen, W. R. Reinvestigation of the steady-state kinetics and physiological function of



- the soluble NiFe-hydrogenase I of *Pyrococcus furiosus*. *J. Bacteriol.* **190**, 1584–1587 (2008).
52. Greening, C. et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *ISME J.* **10**, 761–777 (2016).
  53. Stetter, K. O. Hyperthermophiles in the history of life. *Philos. Trans. R. Soc. B Biol. Sci.* **361**, 1837–1842 (2006).
  54. Collins, T., Gerday, C. & Feller, G. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol. Rev.* **29**, 3–23 (2005).
  55. Schädel, C., Richter, A., Blöchl, A. & Hoch, G. Hemicellulose concentration and composition in plant cell walls under extreme carbon source-sink imbalances. *Physiol. Plant.* **139**, 241–255 (2010).
  56. Chen, S. et al. The Great Oxidation Event expanded the genetic repertoire of arsenic metabolism and cycling. **117**, 10414–10421 (2020).
  57. Rogers, K. L. & Schulte, M. D. Organic sulfur metabolisms in hydrothermal environments. *Geobiology* **10**, 320–332 (2012).
  58. Baker, B. J. et al. Genomic inference of the metabolism of cosmopolitan subsurface Archaea, Hadesarchaea. *Nat. Microbiol.* **1**, 16002 (2016).
  59. Hatzenpichler, R., Krukenberg, V., Spietz, R. L. & Jay, Z. J. Next-generation physiology approaches to study microbiome function at single cell level. *Nat. Rev. Microbiol.* **18**, 241–256 (2020).
  60. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).
  61. Kang, D. D., Froula, J., Egan, R. & Wang, Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**, e1165 (2015).
  62. Dick, G. J. et al. Community-wide analysis of microbial genome sequence signatures. *Genome Biol.* **10**, R85 (2009).
  63. Darling, A. E. et al. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* **2**, e243 (2014).
  64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
  65. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
  66. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* **11**, 119 (2010).
  67. Aramaki, T. et al. KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**, 2251–2252 (2020).
  68. Jones, P. et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
  69. Søndergaard, D., Pedersen, C. N. S. & Greening, C. HydDB: a web tool for hydrogenase classification and analysis. *Sci. Rep.* **6**, 34212 (2016).
  70. Zhang, H. et al. DbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **46**, W95–W101 (2018).
  71. De Anda, V. et al. MEBS, a software platform to evaluate large (meta)genomic collections according to their metabolic machinery: unraveling the sulfur cycle. *Gigascience* **6**, 1–17 (2017).
  72. Zhichao, Z. et al. METABOLIC: High-throughput. profiling of microbial genomes for functional traits, biogeochemistry, and community-scale metabolic networks. Preprint at *bioRxiv* 761643 (2019).
  73. Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J. & Barrett, A. J. MEROPS: the peptidase database. *Nucleic Acids Res.* **38**, D227–D233 (2010).
  74. Taboada, B., Estrada, K., Ciria, R. & Merino, E. Operon-mapper: a web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics* **34**, 4118–4120 (2018).
  75. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
  76. Yu, N. Y. et al. PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
  77. Boyd, J. A., Woodcroft, B. J. & Tyson, G. W. GraftM: a tool for scalable, phylogenetically informed classification of genes within metagenomes. *Nucleic Acids Res.* **46**, e59 (2018).

78. Hua, Z. S. et al. Insights into the ecological roles and evolution of methyl-coenzyme M reductase-containing hot spring Archaea. *Nat. Commun.* **10**, 4574 (2019).
79. Chaumeil, P., Mussig, A. J., Parks, D. H. & Hugenholtz, P. Genome analysis GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
80. Kahle, D. & Wickham, H. ggmap: spatial visualization with ggplot2. *R. J.* **5**, 144–161 (2013).

## Acknowledgements

We thank Dr. Thomas Brock for a career of transformative microbiological research and graciously allowing us to name Brockarchaeota in his honor. This work was funded by a Simons Foundation (Award number 687165), National Science Foundation DEB: Systematics and Biodiversity Sciences (grant number 1753661), and National Science Foundation MCB – Systems and Synthetic Biology (grant number 1817354) provided to BJB. We also thank Luke Mckay and William Inskeep for the 16S rRNA Brockarchaeota sequences from Washburn Spring Yellowstone National Park. We thank Andreas for providing the sediments from Guaymas Basin. Sampling in Guaymas Basin was supported by NSF Awards OCE-0647633 to Dr. Andreas P. Teske. The sequencing was partially conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 provided to ND. This work was partially supported by National Natural Science Foundation of China (Nos. 91951205 and 91751206).

## Author contributions

J.F.B., W.-J.L., and B.J.B. conceived the study. H.C.J. and Z.S.H. performed sampling at Tibet and Yunnan hot springs. L.X.C. and N.D. reconstructed and curated the genomes. V.D.A. and B.J.B. generated the phylogenetic trees. V.D.A. and B.J.B. generated the metabolic reconstructions. V.D.A. performed the genomic comparisons and physiology inference. V.D.A. and B.J.B. wrote the manuscript with input for all the other authors.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-22736-6>.

**Correspondence** and requests for materials should be addressed to W.-J.L. or B.J.B.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021