



# Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats

Authors: Zackary J. Jay, Jacob P. Beam, Mansur Dlakic, Douglas B. Rusch, Mark A. Kozubal, and William P. Inskeep

This is a postprint of an article that originally appeared in [Nature Microbiology](#) on May 14, 2018. The final version can be found at <https://dx.doi.org/10.1038/s41564-018-0163-1>.

Jay, Zackary J. , Jacob P. Beam, Mensur Dlakic, Douglas B. Rusch, Mark A. Kozubal, and William P. Inskeep. "Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats." *Nature Microbiology* 3, no. 6 (May 2018): 732-740. DOI: 10.1038/s41564-018-0163-1.

Made available through Montana State University's [ScholarWorks](#)  
[scholarworks.montana.edu](https://scholarworks.montana.edu)

# Marsarchaeota are an aerobic archaeal lineage abundant in geothermal iron oxide microbial mats

Zackary J. Jay<sup>1,4,7</sup>, Jacob P. Beam<sup>1,5,7</sup>, Mensur Dlakić<sup>2</sup>, Douglas B. Rusch<sup>3</sup>, Mark A. Kozubal<sup>1,6</sup> and William P. Inskeep<sup>1\*</sup>

**The discovery of archaeal lineages is critical to our understanding of the universal tree of life and evolutionary history of the Earth. Geochemically diverse thermal environments in Yellowstone National Park provide unprecedented opportunities for studying archaea in habitats that may represent analogues of early Earth. Here, we report the discovery and characterization of a phylum-level archaeal lineage proposed and herein referred to as the ‘Marsarchaeota’, after the red planet. The Marsarchaeota contains at least two major subgroups prevalent in acidic, microaerobic geothermal Fe(III) oxide microbial mats across a temperature range from -50–80 °C. Metagenomics, single-cell sequencing, enrichment culturing and in situ transcriptional analyses reveal their biogeochemical role as facultative aerobic chemoorganotrophs that may also mediate the reduction of Fe(III). Phylogenomic analyses of replicate assemblies corresponding to two groups of Marsarchaeota indicate that they branch between the Crenarchaeota and all other major archaeal lineages. Transcriptomic analyses of several Fe(III) oxide mat communities reveal that these organisms were actively transcribing two different terminal oxidase complexes in situ and genes comprising an  $F_{420}$ -dependent butanal catabolism. The broad distribution of Marsarchaeota in geothermal, microaerobic Fe(III) oxide mats suggests that similar habitat types probably played an important role in the evolution of archaea.**

New discoveries within Archaea have been occurring at an amazing pace since the application of random metagenome and single-cell sequencing to assemble consensus genome sequences from environmental populations<sup>1–8</sup>. Archaeal lineages, including the Korarchaeota, Nanoarchaeota, Thaumarchaeota, Aigarchaeota, Bathyarchaeota, Geoarchaeota, Verstraetearchaeota, Thorarchaeota and Lokiarchaeota (and other phyla within the Asgard group) have all been reported<sup>8–10</sup> since the initial division of archaea into two phyla—the Crenarchaeota and Euryarchaeota<sup>11</sup>. The concatenated gene sets used to construct phylogenomic analyses have varied tremendously depending on the objectives<sup>10,12–14</sup>. Consequently, there has been significant debate on the phylogeny of major groups within archaea and their relationship to eukarya<sup>6,8,10,14</sup>. Interest in unravelling the role of archaea in the evolution of eukaryotes has led to comparisons of all cellular life—eukarya, archaea and bacteria—in the same phylogenetic trees<sup>4,6,10</sup>. However, the creation of polyphyletic trees including both eukarya and archaea (or all three domains) may artificially constrain the gene sets used for concatenation, and can obfuscate phylogenetic relationships among lineages within archaea<sup>9,14</sup>.

Metagenomics has been used to characterize the consensus genome sequences of numerous different archaea within acidic (pH 2–4), geothermal Fe(III) oxide mats (Supplementary Fig. 1), including members of the Crenarchaeota, Euryarchaeota, Geoarchaeota, Nanoarchaeota and Thaumarchaeota<sup>2,3,15–18</sup>. Another undescribed lineage of archaea (referred to in previous studies<sup>3,15,19</sup> as ‘novel archaeal group 2’) has been found in high abundance in geothermal Fe(III) oxide microbial mats. Here, we report details on phylogenomic analyses, activity-based (transcriptomic) studies and direct microscopy to describe this uncharacterized lineage

of archaea. Replicate de novo assemblies, single-cell amplified genomes (SAGs) and genome sequences from an enrichment culture were used to perform phylogenomic analyses inclusive of all major archaeal groups. Our results show that extant members of the proposed ‘Marsarchaeota’ represent a separate archaeal lineage that shares metabolic attributes across several disparate archaeal phyla. The inclusion of Marsarchaeota in phylogenetic analyses will probably assist in resolving controversies regarding the archaeal root by expanding the diversity of phyla that branch between the Crenarchaeota and other lineages of Archaea.

## Results

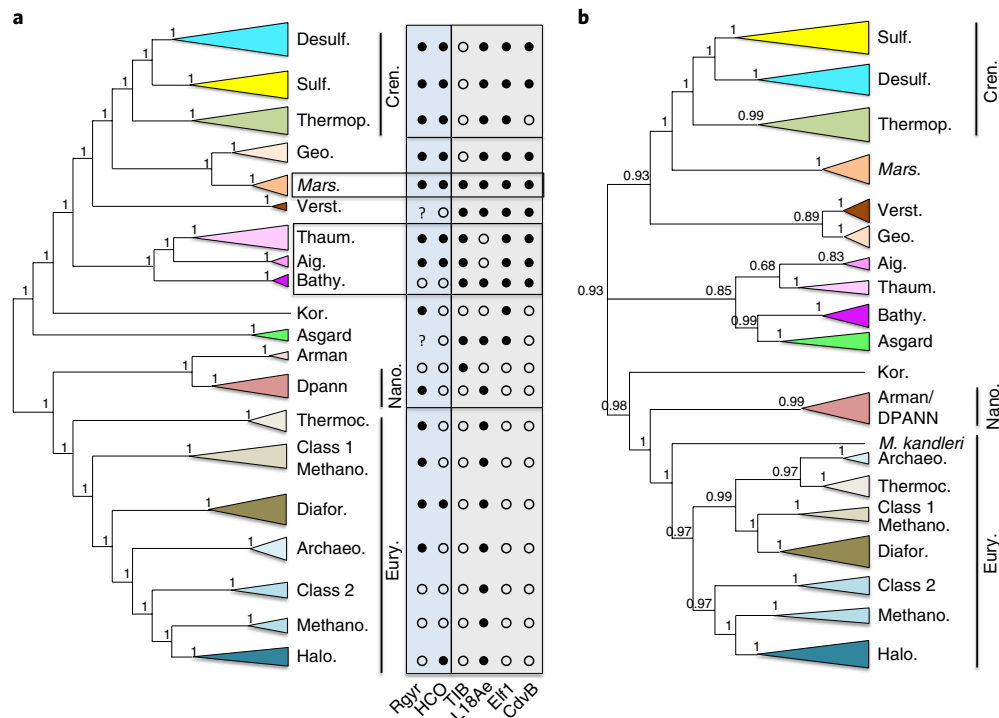
**Phylogenetic and genomic analysis of Marsarchaeota groups 1 and 2.** Several metagenomes, SAGs and genome sequences from a laboratory enrichment (Table 1 and Supplementary Fig. 2) provided replicate entries within two major subgroups of Marsarchaeota for a thorough phylogenomic analysis (concatenation of 55 protein sequences; 15,136 amino acid positions) of all major archaeal groups (Fig. 1a). The collapsed 55 protein Bayesian tree of archaea provided a coherent and consistent phylogeny with strong support at all major nodes (full version shown in Supplementary Fig. 3). This analysis shows that two separate groups of Marsarchaeota branch between the Crenarchaeota, Verstraetearchaeota and Thaumarchaeota, and are distinct from the Geoarchaeota described previously<sup>3</sup>.

The topology of the archaeal tree obtained using Bayesian methods (Fig. 1a) was consistent with maximum likelihood (ML) analysis of the same concatenated sequences, although the ML tree showed weaker support at several nodes (Supplementary Fig. 4a). We also conducted an ML phylogenomic analysis of all major archaeal lineages compared with eukarya and bacteria,

**Table 1 | Summary of metagenome and genome sequence assemblies for Marsarchaeota group 1 and 2 populations obtained from several high-temperature iron oxide microbial mats of YNP**

Mars. ID <sup>a</sup>	Source <sup>b</sup>	pH, T (°C)	Seq. (Mb) <sup>c</sup>	Scfs <sup>c</sup>	G + C (%)	Coverage	Compl. <sup>d</sup> (%)	tRNA <sup>e</sup>	Genes	Abund. <sup>f</sup> (%)	Seq. Plat. <sup>g</sup>
G1-r01	OSP_B	3.5, 73	1.08	309	44.3	4.8	72.8	21	1,255	5.2	P
G1-r02	OSP_C	3.5, 65	1.99	169	44.2	17.7	87.4	31	2,245	31.4	P
G1-r03	OSP_D	3.5, 57	1.92	192	43.9	22.7	40.8	24	2,094	17.2	P
G1-r04	BE_D	3.0, 66	2.19	195	43.8	16.8	95.1	38	2,377	16.6	P
G2-r01	OSP_D	3.5, 57	2.78	407	50.3	15.1	90.3 <sup>h</sup>	35	3,032	18.9	P
G2-r02	BE_D	3.0, 66	3.73	801	49.8	25.6	128 <sup>h</sup>	42	4,166	47.2	P
G2-cons.1	ECH_B	3.4, 66	2.55	66	49.8	425	93.2	48	2,747	9.8	I
G2-cons.2	ECH_B	3.4, 66	2.59	70	49.9	1,192	93.2	47	2,761	9.9	I
G2-cons.3	ECH_B	3.4, 66	2.47	75	50.0	1,411	93.2	47	2,648	6.7	I
G2-cons.3 <sup>i</sup>	ECH_B	3.4, 66	2.54	12	50.0	1,400	99.1	33	2,732	6.7	I

<sup>a</sup>Marsarchaeota ID: G1, group 1; G2, group 2; cons.1, cons.2 and cons.3 represent replicate archaeal consortia grown from ECH\_B inoculum. <sup>b</sup>Source: BE, Beowulf Spring East (transect positions D and E); ECH, Echinus Geyser (position B); OSP, One Hundred Spring Plain (transect positions B, C and D). <sup>c</sup>Seq., total sequence length; scfs, number of scaffolds. <sup>d</sup>Completeness estimate based on the presence of 103 marker genes specific to archaea<sup>17</sup>. <sup>e</sup>Number of transfer RNAs and total protein-coding genes determined using the Prokka pipeline<sup>46</sup>. <sup>f</sup>Abundance calculated as the amount of random sequence from Marsarchaeota relative to the total random sequence. <sup>g</sup>Sequencing platform: I, Illumina; P, 454 pyrosequencing. <sup>h</sup>Includes at least two genetically distinct populations. <sup>i</sup>Comparison using newer assembly of random sequence with SPAdes 3.11 (ref. <sup>47</sup>) and evaluated with CheckM<sup>48</sup>.



**Fig. 1 | Phylogenomic analysis of the candidate phylum Marsarchaeota (groups 1 and 2).** **a**, Bayesian tree using a concatenation of 55 deduced protein sequences (total amino acid positions aligned = 15,136) common to the Marsarchaeota and other archaeal references ( $n=108$  archaeal entries) (see Supplementary Table 1 for a list of species and aCOGs used in the analysis). The presence (filled circle) or absence (empty circle) of reverse gyrase (Rgyr), subunit I heme copper-oxidase (HCO), DNA type IB topoisomerase (TIB), 50S ribosomal protein L18Ae, elongation factor 1 (EIF1) and cell division protein CdvB (a component of the ESCRT-III sorting complex<sup>49</sup>) in each lineage is also indicated. Rgyr entries for thermophilic representatives of the Verstraetearchaeota and Asgard group have not yet been confirmed (indicated by a question mark). **b**, Bayesian tree of archaea only (same entries as in **a**) using near-full-length 16S rRNA genes. Aig., Aigarchaeota; Archaeo., Archaeoglobales; Arman., archaeal Richmond Mine acidophilic nanoorganisms; Bathy., Bathyarchaeota; Cren., Crenarchaeota; DPANN, Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaeota; Desulf., Desulfurococcales; Diafor., Diaforarchaea; Eury., Euryarchaeota; Geo., Geoarchoeota; Halo., Haloarchaea; Kor., Korarchaeota; Mars., Marsarchaeota; Methano., methanogens; Nano., Nanoarchaeota; Sulf., Sulfolobales; Thaum., Thaumarchaeota; Thermoc., Thermococcales; Thermop., Thermoproteales; Verst., Verstraetearchaeota. Numbers at nodes correspond to Bayesian posterior probabilities.

which demonstrated that the Marsarchaeota were placed reproducibly between the Crenarchaeota, Verstraetearchaeota and Thaumarchaeota (Supplementary Fig. 4b). Consequently, the

inclusion of members of the eukarya and/or bacteria did not significantly change the phylogenetic position of the Marsarchaeota. The three-domain ML tree shows the Asgard superphylum<sup>8</sup> to be

closest to the root of the eukaryal branch, although the position of the lone representative of the Korarchaeota is not strongly supported (Supplementary Fig. 4b). In fact, a three-domain Bayesian tree using 16 ribosomal proteins<sup>10</sup> shows the Korarchaeota to be closest to the eukaryal branch (Supplementary Fig. 3c). Moreover, as has been shown recently<sup>14</sup>, the removal of only one archaeal cluster of orthologous genes (arCOG) (that is, elongation factor 2) and one Asgard entry (that is, Heimdallarchaeota LC3) from the 55 arCOG tree results in the Korarchaeota being closest to the eukaryal root (Supplementary Fig. 3d). Subsequent removal of the Diapherotrites, Parvarchaeota, Aenigmarchaeota, Nanoarchaeota and Nanohaloarchaea (DPANN), as well as archaeal Richmond Mine acidophilic nanoorganism (ARMAN) representatives (that is, fast-evolving lineages<sup>14</sup>) further separates archaeal lineages from eukarya and bacteria (Supplementary Fig. 3e). In summary, our phylogenomic analysis (Fig. 1a) is consistent with recent reports<sup>6,8-10,12-14</sup> when we use a similar set of proteins. However, we also show that the position of the eukaryal root relative to different archaeal groups is dependent on the exact species and proteins used to construct the phylogeny. Neither of these alternatives results in different placement of group 1 and 2 Marsarchaeota.

A 16S ribosomal RNA (rRNA) tree (Fig. 1b) of archaea (same entries used for Fig. 1a) also shows that the Marsarchaeota branch distinctly from other phyla. Phylogenetic relationships among different archaeal groups observed using the 16S rRNA tree were similar, but not identical, to those obtained using concatenated protein sequences. For example, the 16S rRNA phylogeny (full tree; Supplementary Fig. 3f) suggests a closer relationship between the Bathyarchaeota and the Asgard group, as well as between the Geoarchaeota and Verstraetearchaeota, than was observed using 55 concatenated proteins. However, the 16S rRNA tree does not resolve archaeal phylogeny sufficiently, in part because the G+C content of 16S rRNA is influenced by adaptation to temperature (Supplementary Fig. 5). The uracil contents of 16S rRNA sequences of group 1 and 2 Marsarchaeota are intermediate compared with values observed in other archaeal phyla, which is consistent with their moderately thermal growth temperatures. The presence of reverse gyrase genes (Fig. 1a) suggests the importance of Marsarchaeota in the thermophilic ancestry of other groups of archaea (for example, Crenarchaeota and Thaumarchaeota). Before the discovery of the Marsarchaeota, reverse gyrases in archaea had been limited primarily to hyperthermophiles.

The genome sequences of group 1 and 2 Marsarchaeota contain several genes with homologues in other archaea, as well as eukarya. For example, the Marsarchaeota contain DNA type IB topoisomerase, ribosomal protein L18Ae (LXa, homologous to the eukaryotic ribosomal protein eL20 family), elongation factor 1, *cdvB* (a subunit important in cell division) and homologues to the eukaryotic Vps24 (a component of the endosomal sorting complex required for transport (ESCRT)-III sorting complex) (Fig. 1a). The presence of these genes in yet another archaeal lineage emphasizes their likely importance in a last archaeal common ancestor. Additionally, the presence of haem copper (terminal) oxidase (HCO) complexes in the Marsarchaeota suggests a potential link in aerobic catabolism between members of the Crenarchaeota, Thaumarchaeota and Euryarchaeota (for example, Thermoplasmatales), which is discussed in greater detail below.

Support for designating the Marsarchaeota as a new phylum stems from direct comparisons of the average nucleotide and amino acid identities of group 1 and 2 Marsarchaeota versus representatives of all other major archaeal groups (Fig. 2). Genome-identity values of Marsarchaeota compared with other major archaeal lineages are as low as values observed among other disparate phyla, and are lower than the threshold(s) considered for separate phyla<sup>20</sup>. For example, genome-identity values among the Bathyarchaeota, Thaumarchaeota and Aigarchaeota (or different phyla within the

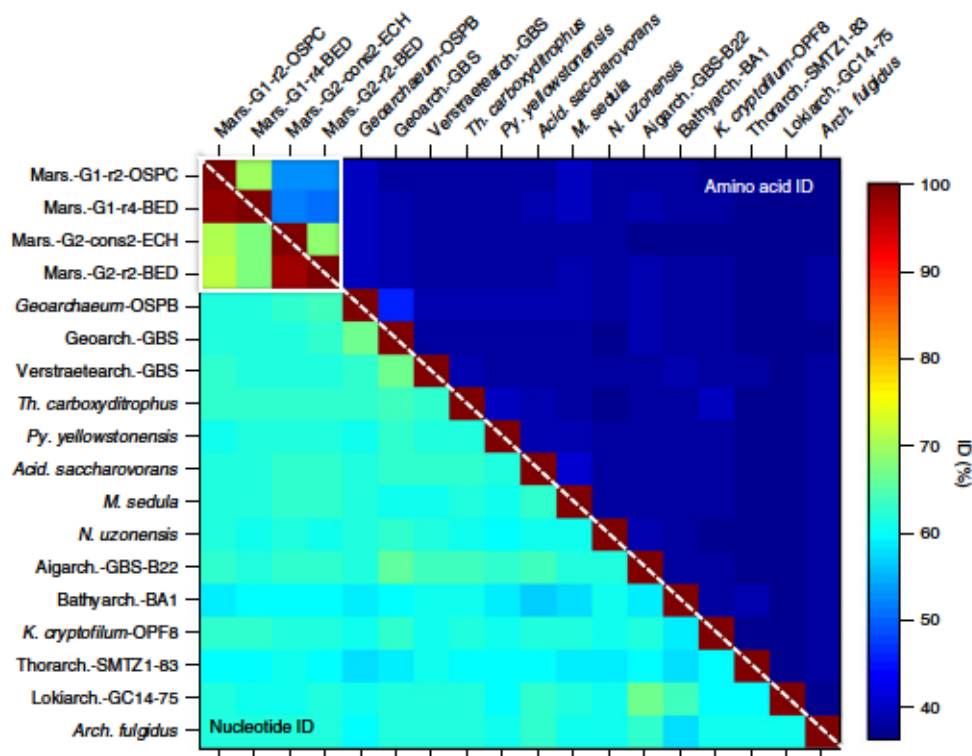
Asgard group) range from 55–65% (nucleotides) and 35–45% (amino acids), respectively (Fig. 2). Similarly, the Marsarchaeota (groups 1 and 2) are <65% (nucleotides) or <40–45% (amino acids) identical to other archaeal phyla. A comparison of genome-identity values among group 1 and 2 Marsarchaeota reveals that replicates within either group share >95% identity for many homologues, while group 1 versus group 2 entries are less than ~70% identical (nucleotides), probably comprising different ‘family-level’ lineages (Fig. 2 and Supplementary Fig. 6).

Type IB DNA topoisomerases are critical to all known eukarya, but have only been found to date in a few archaeal phyla<sup>1,8,21,22</sup>. Consequently, the discovery of another archaeal lineage with type IB DNA topoisomerases supports the possibility that these proteins were present in a common ancestor of archaea and eukarya. Phylogenetic analysis (using the Le and Gascuel (LG) model, 533 amino acids) of type IB DNA topoisomerases shows a monophyletic distribution within group 1 and 2 Marsarchaeota (Supplementary Fig. 7); however, insufficient entries for the Hadesarchaea and Bathyarchaeota preclude definitive conclusions regarding possible lateral gene transfer events. Interestingly, type IB topoisomerases of the Marsarchaeota fall between different thaumarchaeal groups, which are the most closely related to eukaryal sequences. Other genes present in the Marsarchaeota with homologues in the Thaumarchaeota, Euryarchaeota and eukarya (but absent in the Crenarchaeota) include those encoding ERCC4-type DNA repair endonucleases, heat shock proteins (Hsp70/DnaK, GrpE, Hsp40/DnaJ and Hsp90), an excinuclease (*uvrABC* endonuclease),  $F_{420}$  biosynthesis and other  $F_{420}$ -dependent proteins<sup>23</sup>.

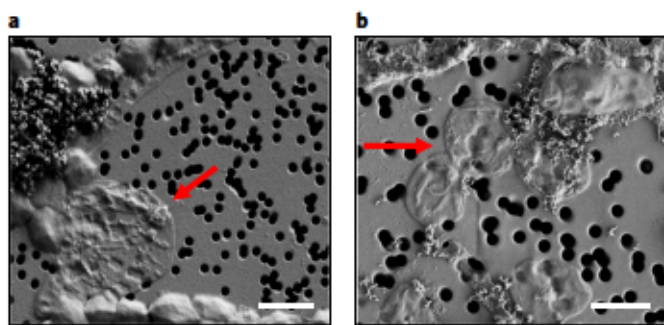
In addition to the metagenome, single-cell and genome sequences described above, numerous long-fragment (>1,200 base pair) 16S rRNA sequences have been obtained from several different Fe(III) oxide mats in Yellowstone National Park (YNP) over the past decade<sup>19</sup>. A 16S rRNA sequence identity matrix (Supplementary Fig. 8) shows numerous representatives of the two predominant groups of Marsarchaeota, and reveals strong agreement among long-fragment 16S rRNA sequences obtained from amplification and cloning, assembled metagenome sequences, SAGs and/or the enrichment culture containing a group 2 Marsarchaeota. Other 16S rRNA sequences within the Marsarchaeota have been observed in a very limited number of different locations and/or environments (see details in Supplementary Fig. 8).

**Direct identification.** Direct imaging of Marsarchaeota cells (Fig. 3) was accomplished using field-emission scanning electron microscopy (FE-SEM) of an enrichment culture from Echinus Geyser containing group 2 Marsarchaeota. Random shotgun sequencing of three replicate samples of this enrichment resulted in near-complete (~97%) genome of a group 2 Marsarchaeota (Table 1), as well as two other members of the consortium (Supplementary Fig. 2). Although the enrichment cultures also contained *Sulfolobus islandicus* and *Thermocodium* spp. (determined from genome sequencing<sup>17</sup>), the morphologies of these organisms are known. This provided a mechanism for identifying group 2 Marsarchaeota as large cocci (>1  $\mu\text{m}$  diameter) (Fig. 3). Fluorescence in situ hybridization (FISH) confirmed the presence of Marsarchaeota in Fe(III) oxide mats from Echinus Geyser (Supplementary Fig. 9). The FISH probes were designed to hybridize with members of group 1 and 2 Marsarchaeota (Cy3), all archaea (Cy5) or Aquificales (6-FAM) sequences. An overlap of red (Marsarchaeota) and blue (archaeal) probes yielded magenta cocci of ~1  $\mu\text{m}$  diameter in numerous samples from Echinus Geyser iron mats (Supplementary Fig. 9) known to contain significant amounts of group 1 and 2 Marsarchaeota.

**Abundance and activity.** The DNA abundances of group 1 and 2 Marsarchaeota were determined (based on random sequences) across numerous Fe(III) oxide mats in YNP (Supplementary



**Fig. 2 | Genome-identity correlation matrix of Marsarchaeota groups 1 and 2 compared with representatives of major archaeal lineages (top: average amino acid identity; bottom: average nucleotide identity).** *Acid.*, *Acidilobus*; Aigarch., Aigarchaeota; Arch., *Archaeoglobus*; Bathyarch., Bathyarchaeota; BED, Beowulf East Spring transect D; cons., consortium; ECH, Echinus Geyser; G1, group 1; G2, group 2; GBS, Great Boiling Spring; Geoarch., Geoarchaeota; K., *Korarchaeum*; Lokiarch., Lokiarchaeota; Mars., Marsarchaeota; M., *Metallosphaera*; N., *Nitrosopumilus*; OSPB, One Hundred Spring Plain transect B; OSPC, One Hundred Spring Plain transect C; Py., *Pyrobaculum*; r, replicate; Th., *Thermofilum*; Thorarch., Thorarchaeota; Verstraetearch., Verstraetearchaeota.

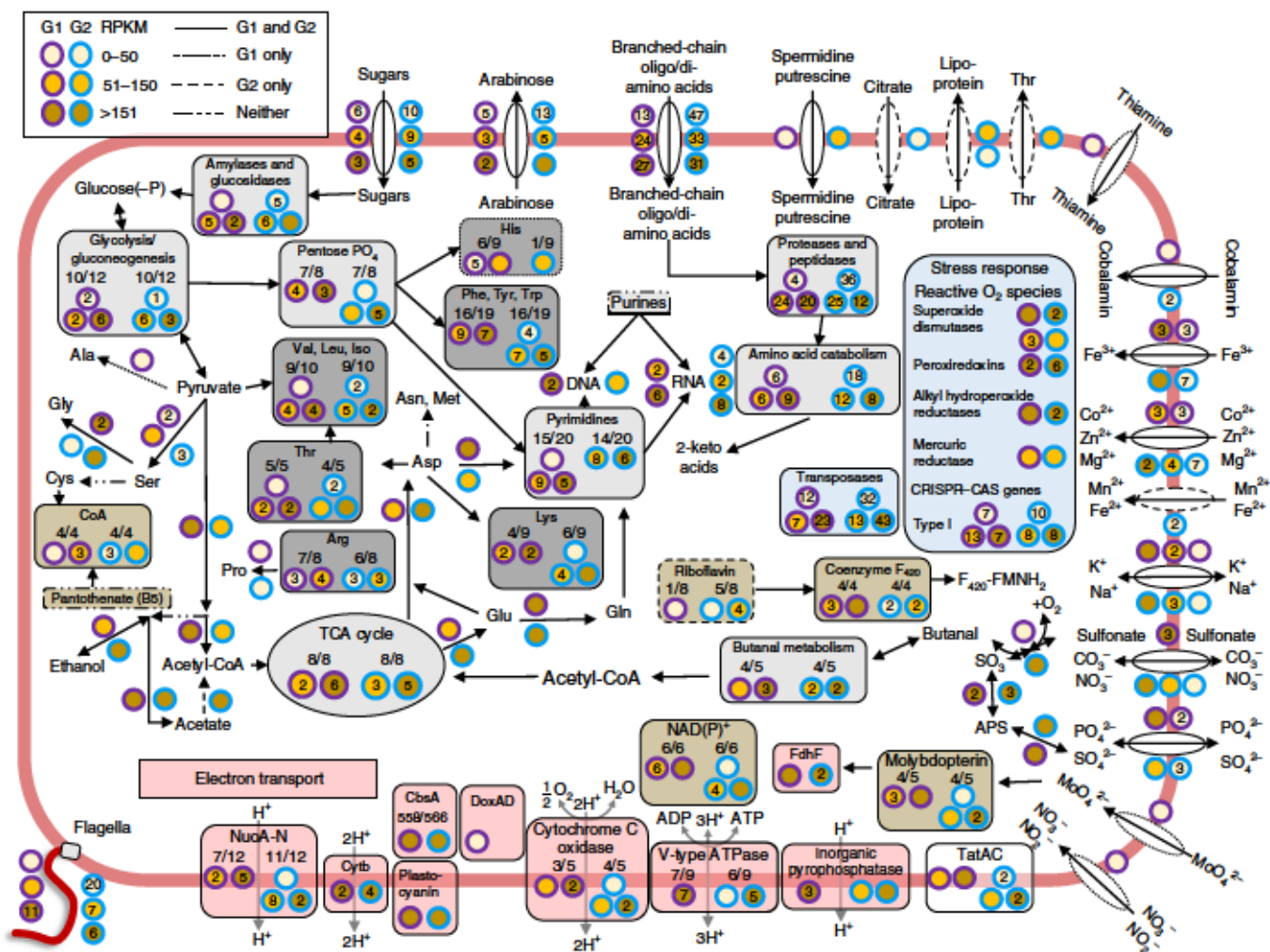


**Fig. 3 | Scanning electron micrographs (SEM) of group 2 Marsarchaeota.** **a,b**, Enrichment cultures inoculated with iron oxide sediment from Echinus Geyser were grown using 0.2% casamino acids (**a**) or 0.2% tryptone (**b**) at 75 °C and pH 3.5. The images and corresponding NWF analyses (Supplementary Fig. 2) of the assembled sequences show three major populations: group 2 Marsarchaeota (>1 μm cocci, red arrows); *Thermocodium* spp. (0.3–0.4 μm diameter filaments); and *S. islandicus*-like (<0.8 μm cocci). Field emission (FE)-SEM images were taken on 0.2 μm filters. Scale bars, 1 μm. Triplicate enrichment cultures were sequenced, with similar results (Table 1 and Supplementary Fig. 2). Images are representative of numerous others obtained from >3 subsamples from each replicate ( $n=3$ ) enrichment.

Fig. 10). Sites with temperatures ranging from 65–68 °C and pH values near 3 exhibited similar abundances to group 1 and 2 Marsarchaeota (nearly 20% each) and were the best habitats for an unbiased evaluation of possible differences in metabolic activity (transcripts) between these two lineages. Transcriptomic analysis

of several Fe(III) oxide mats showed that a significant fraction of marsarchaeotal genes were transcribed in both group 1 and group 2 populations in Beowulf Spring East (site BE\_D) and Echinus Geyser (site ECH\_B) (Supplementary Fig. 10), while the higher temperature site in One Hundred Spring Plain (site OSP\_B, 75 °C) exhibited transcripts from only the group 1 Marsarchaeota.

**Core metabolism.** Transcriptomic analyses of group 1 and 2 Marsarchaeota provided direct information on the activity of these organisms in situ (Fig. 4). A comparison of in situ activity was made primarily for Beowulf Spring East, although transcriptomic data mapped to group 1 and 2 Marsarchaeota were obtained for two additional Fe(III) oxide sites (Supplementary Table 4). Both group 1 and 2 Marsarchaeota are aerobic chemoorganotrophs that showed high expression levels of genes for carbon metabolism, oxidative phosphorylation and motility, as well as those involved in the environmental and viral stress response (Fig. 4). Highly transcribed sugar and peptide/amino acid transporters are probably responsible for the uptake of exogenous organic matter, which is then oxidized and assimilated via glycolysis (following the Embden–Meyerhof–Parnas pathway), the pentose phosphate pathway and the oxidative tricarboxylic acid cycle. At least four active pathways produce acetyl-CoA in group 1 and 2 Marsarchaeota, including synthesis from pyruvate, lipid degradation, acetate assimilation (group 2 only) and an  $F_{420}$ -dependent butanal (butyraldehyde) metabolism, which is probably a product of (butane) sulfonate dissimilation. Extracellular sulfonates, including methane sulfonate (in rain) and complex biologically synthesized sulfonates (for example, sulfolipids) are initially desulfonated via an  $F_{420}$ -dependent alkanesulfonate mono-oxygenase (Enzyme Commission number (EC) 1.14.14.5), which is highly expressed in group 2 populations and releases sulfite



**Fig. 4 | Metatranscriptomic analysis.** Key metabolic activities (reads per kilobase of transcript per million mapped reads (RPKM) values given as colour intensity) of group 1 (purple) and 2 (light blue) Marsarchaeota based on in situ metatranscriptomic analysis of an iron oxide microbial mat from Beowulf Spring East (BE\_D, 65 °C) using an assembled reference genome sequence from the same site (see Table 1). The number of genes with corresponding transcripts relative to the approximate number genes required in a pathway are shown as fractions. Numbers within circles correspond to the number of genes with the indicated RPKM shading for group 1 and 2 phylotypes. APS, adenosine phosphosulfate reductase; CRISPR, clustered regularly interspaced short palindromic repeats; Cas, CRISPR-associated; NAD(P), nicotinamide adenine dinucleotide (phosphate); NuoA-N, NADH:ubiquinone oxidoreductase subunits A-N; FMNH<sub>2</sub>, reduced flavin mononucleotide; TatAC, twin-arginine translocation subunits A and C; TCA, tricarboxylic acid.

that is subsequently oxidized to sulfate via adenosine phosphosulfate reductase (EC 1.8.99.2) and sulfurylase (ATP-dependent, EC 2.7.7.4)<sup>24</sup>.

Genes responsible for the biosynthesis of coenzyme F<sub>420</sub> and transcription of F<sub>420</sub>-dependent proteins were all highly expressed by both group 1 and 2 populations (Supplementary Table 5). These findings suggest that F<sub>420</sub>-dependent metabolism is critical to the Marsarchaeota, and recent genome sequences suggest that this cofactor is important across numerous archaeal phyla (for example, Aigarchaeota, Bathyarchaeota, Euryarchaeota, Geoarchaeota, Lokiarchaeota and Thaumarchaeota)<sup>3,23</sup>. Although the exact function of all F<sub>420</sub>-dependent enzymes is not well established<sup>23</sup>, the diverse functional role of this cofactor across numerous lineages of archaea (including methane cycling) suggests the importance of this electron carrier in the last archaeal common ancestor. Current members of the Crenarchaeota have lost the ability to utilize coenzyme F<sub>420</sub> as members of this lineage do not appear to synthesize F<sub>420</sub>, nor do they have enzymes that require this cofactor.

Genes involved in oxidative phosphorylation (including NADH dehydrogenase, HCOs, cytochrome bc complexes and V-type ATPase complexes) were highly expressed in both groups (Fig. 4 and Supplementary Table 4). Interestingly, two major types of terminal oxidase complexes are found in both group 1 and 2 populations: one contains plastocyanin(s) (blue copper proteins) and an F<sub>420</sub>-dependent oxidoreductase, while the second contains CbsA 558/566 and DoxAD cytochromes (see Supplementary Fig. 11). These findings provide evidence that two different mechanisms are responsible for the reduction of O<sub>2</sub> in Marsarchaeota. Although group 2 genomes do not contain a DoxAD cytochrome, known for its role in thiosulfate oxidation<sup>25</sup>, both populations actively transcribed *chsA* 558/566 cytochrome genes. In this regard, the Marsarchaeota show similarities to other Crenarchaeota, which contain *chsA* 558/566 in several different functional configurations<sup>26,27</sup>. The CbsA 558/566 cytochrome is found as an important component of the FoxC protein (*fox* complex) implicated in Fe(II) oxidation in the Sulfolobales<sup>26</sup> and as a separate electron transport complex in the Thermoproteales<sup>17</sup>. The Marsarchaeota represent the

most basal lineage to contain this specific cytochrome. Although it is well established that numerous O<sub>2</sub>-dependent gene families emerged during and after the Great Oxidation Event<sup>28</sup>, it is also clear that terminal oxidase complexes with high substrate affinities for O<sub>2</sub> may have supported growth on O<sub>2</sub> long before atmospheric concentrations reached current levels<sup>29</sup>. For example, even though the realized niche of Marsarchaeota contains low levels of dissolved O<sub>2</sub>, these cytochromes and associated terminal oxidase complexes are highly expressed in situ.

No genomic evidence for fermentation pathways or the use of alternate terminal electron acceptors was found in group 1 or 2 Marsarchaeota. However, laboratory enrichments containing a significant majority of group 2 Marsarchaeota (for example, >75% in two cases) released significant amounts of Fe(II) in the presence of Fe(III) oxides after the removal of O<sub>2</sub> from the headspace (Supplementary Fig. 12). Consequently, phenotypic evidence suggests that group 2 Marsarchaeota may utilize Fe(III) as a terminal electron acceptor under low O<sub>2</sub> conditions, although definitive identification of the genes responsible for the dissimilatory reduction of Fe(III) phases was not possible. The presence of a CbsA 558/566 cytochrome may suggest involvement with iron<sup>26</sup>, while the presence of a DoxAD cytochrome in group 1 populations is suggestive of possible involvement with thiosulfate<sup>25</sup>. However, DoxAD is also capable of transferring electrons to the ferricyanide couple<sup>25</sup>, which is poised similarly to iron redox couples involving Fe(III) oxides.

Formate dehydrogenases (*fdhF*) were also highly expressed in both groups 1 and 2; the oxidation of formate to CO<sub>2</sub> may represent an important source of reducing equivalents captured as NAD(P)H<sub>2</sub>. Although numerous genes were annotated as aerobic carbon monoxide dehydrogenases in both groups, only one gene in group 2 contained the putative CO active binding motif (IAYRCSFR)<sup>30</sup>, which was not highly expressed. Numerous genes implicated in various types of stress response were highly expressed, including oxidative (superoxide dismutases, peroxiredoxins and alkyl hydroperoxide reductases), mercury (mercuric reductases) and viral (CRISPR–Cas genes) stress. Numerous transposases were expressed in both groups, indicating a high frequency of transposable elements in these genomes. Finally, genes involved in motility and protein transport across the cell membrane (twin-arginine translocases) were identified and expressed at various levels. Other required metabolites, vitamins and cofactors that are probably not synthesized by group 1 and/or group 2 Marsarchaeota include purines, specific amino acids (for example, cysteine, asparagine and methionine), pantothenate (B5), cobalamin (B12) and niacin (B3).

## Discussion

Phylogenomic analysis of replicate genome sequences revealed that groups 1 and 2 of the Marsarchaeota are distinct from other established archaeal phyla (Figs. 1 and 2). Phylogenetic placement and the presence of numerous genes with homologues in other deeply rooted archaeal phyla and eukarya reveal that the Marsarchaeota represent an important link between the Crenarchaeota and other archaeal lineages. The presence and activity of two types of terminal oxidase complexes in the Marsarchaeota is consistent with the microaerobic habitat of thermal Fe(III) oxide systems, where steep O<sub>2</sub> gradients develop from high rates of microbial consumption coupled with lower rates of O<sub>2</sub> diffusion<sup>15</sup>. Members of the Marsarchaeota and Geoarchaeota have been shown to be more abundant in microaerobic mat locations, below a more oxygenated zone (~0–1 mm) containing a higher abundance of chemolithotrophic *Metallosphaera yellowstonensis* and *Hydrogenobaculum* spp.<sup>15</sup>.

The production of Fe(II) from Fe(III) oxides in enrichments containing significant fractions of group 2 Marsarchaeota (up to 82%) during O<sub>2</sub> deprivation (Supplementary Fig. 12) suggests that these organisms may utilize either O<sub>2</sub> or Fe(III) as electron acceptors, and may explain why these populations thrive in microaerobic

(~10–20 μM O<sub>2</sub>(aq)) environments with plentiful amounts of Fe(III) oxides. Alternatively, the Marsarchaeota may have benefited from other members of these consortia that were reducing Fe(III). It is highly unlikely that *M. yellowstonensis* or *S. islandicus* were reducing Fe(III) because they have not been shown to perform this metabolism in pure culture<sup>19,26</sup>; the metabolism of *Thermocodium* spp. has not been completely described<sup>17</sup>, so it is possible that this organism may be responsible for, or contribute to, the Fe(III) reduction measured in these co-cultures.

Two unique terminal oxidase complexes were noted in the Marsarchaeota (Supplementary Fig. 11), each of which is recognized in variant architectures found in other Crenarchaeota, Thaumarchaeota and Euryarchaeota (that is, Thermoplasmatales) where aerobic respiration is an important metabolism<sup>26,27</sup>. These findings provide additional evidence that the Marsarchaeota represent an extant link between members of the Crenarchaeota, Thaumarchaeota and Euryarchaeota. These gene complexes also contained the only cytochromes that might play a role in iron cycling. No definitive genomic evidence for dissimilatory Fe(III) reduction was obtained in the current study, in part because these pathways are virtually uncharacterized in archaea and it is not clear which proteins may be involved. Flavoproteins and unique cytochromes thought to be involved in Fe(III) reduction in other organisms (for example, *Shewanella*, *Ferroglobus* and *Archaeoglobus*) were not observed in these members of the Marsarchaeota. Consequently, it remains to be proven whether Fe(III), although ubiquitous in the habitats where Marsarchaeota flourish, is a critical energetic requirement.

Metabolic reconstruction and transcriptomic evidence revealed that representatives of group 1 and 2 Marsarchaeota are chemo-organotrophs and utilize a variety of reduced carbon substrates. Sulfonate may serve as a carbon source via a F<sub>420</sub>-dependent butanal metabolism, and represents an additional mechanism for generating acetyl-CoA in archaea. No evidence existed for the fixation of CO<sub>2</sub> via established pathways. Active formate dehydrogenases also suggest that Marsarchaeota can regulate carbon flux by oxidizing formate to CO<sub>2</sub>, and conserve reducing equivalents through the regeneration of NADPH and/or NADH. The presence of F<sub>420</sub>-dependent NADPH dehydrogenases ensures exchange of electrons to each cofactor.

We have provided evidence from replicate metagenome assemblies, single-cell sequences, transcriptomics, enrichment cultures and numerous 16S rRNA sequences to support the introduction of a phylum-level lineage in archaea, referred to here as the Marsarchaeota. Motivation for this designation comes from Mars (the Roman god of war and agriculture), after the ‘red’ planet, which has been shown to contain similar types of oxidized iron solid phases<sup>31,32</sup> found in acidic geothermal Fe(III) oxide mats that are central to the realized niche of extant members of this archaeal lineage. To date, members of the Marsarchaeota have been observed primarily in acidic geothermal ferric oxyhydroxide and/or jarositic microbial mats of YNP, which also contain abundant members of the Sulfolobales (for example, iron-oxidizing populations of *Metallosphaera* spp.), as well as members of the Geoarchaeota<sup>3</sup>. Geothermal iron-rich environments have been implicated as being important in early Earth and perhaps critical for the origin of life<sup>33,34,35</sup>. The discovery and characterization of several microaerobic archaeal lineages whose distribution is limited largely to acidic, geothermal Fe(III) oxide mats suggests that these habitats were important in the evolution of archaea.

## Methods

**Geothermal sites, sample collection and laboratory enrichment.** Iron oxide microbial mats from several locations in Yellowstone National Park (Supplementary Fig. 1) were sampled and subjected to DNA extraction and metagenome sequencing (Supplementary Table 2). All geothermal sites sampled

were acidic (~pH 2.5–3.5), but varied in temperature (~50–80 °C) and the composition of major ions (reported previously)<sup>19</sup>.

**Metagenomics, transcriptomics and single-cell genomes.** Metagenome sequencing of numerous iron oxide microbial mat samples was conducted using Sanger sequencing<sup>21,16</sup>, 454 pyrosequencing<sup>3,19</sup> and, more recently, random Illumina sequencing<sup>15,17</sup> (Supplementary Table 2). Sequence data were assembled and annotated using standard pipelines at the Department of Energy (DOE) Joint Genome Institute (Walnut Creek, CA, USA) as described in more detail on the Integrated Microbial Genomes and Microbiomes system<sup>35</sup>. Sanger and 454 pyrosequence assemblies from Beowulf Spring East (transect positions D and E) and One Hundred Spring Plain (transect positions B, C and D) were curated further using nucleotide word-frequency (NWF) algorithms. Briefly, NWF profiles for sequences two kilobases or more in length were reduced in dimensionality using t-distributed neighbor embedding<sup>36</sup> to present the data in three dimensions and then clustered using a semi-automated agglomerative technique as implemented in the R package dbscan<sup>37</sup>. This approach has the effect of grouping contigs sharing similar sequence characters (for example, G + C content and codon usage bias) that probably belong to the same population type (for example, genus or species level). Tests with large sets (100) of randomly selected genomes have indicated that this approach is 97–98% accurate at placing contigs ten kilobases or larger into a single cluster representing a particular bacterial or archaeal genus. Contigs belonging to specific NWF clusters were evaluated further using G + C content, coverage and BLAST scores to reference genomes and metagenomes to develop replicate genome assemblies corresponding to two major lineages within the candidate phylum Marsarchaeota. These de novo sequence assemblies (Table 1) were used as references for phylogenomic analyses, metabolic reconstruction and mapping messenger RNA from three transcriptomes of Fe(III) oxide mat samples (BE\_D, OSP\_C and ECH\_B) to curated DNA sequences. RNA was extracted from Fe(III) oxide mat samples (~2 g wet weight) by adding 1 ml of RNAPro lysis solution (MP Biomedicals) and performing 3 freeze–thaw cycles (liquid nitrogen) in a 15 ml centrifuge tube. This gentle lysis avoided bead beating and prevented the formation of single-stranded RNA–iron complexes, which are extremely difficult to reverse. After centrifugation (10,000 g for 10 min at 4 °C), the supernatant was added to 2 ml of Tri Reagent (Sigma–Aldrich) and incubated at room temperature for 15 min. Next, 400 µl of polymerase-chain-reaction-grade chloroform was added to promote nucleoprotein dissociation and incubated at room temperature for 5 min. The 15 ml tubes were then centrifuged at 4 °C for 15 min at 10,000 g, and the top supernatant containing the extracted RNA was removed (avoiding the interface) and placed into a clean 2 ml centrifuge tube on ice. The RNA was concentrated using a Zymo RNA Clean and Concentrator kit (containing DNase). The concentrated RNA was checked for DNA contamination by conventional polymerase chain reaction with universal archaeal 16S rRNA gene primers (Arch2Fb and 1406R), then sequenced (using MiSeq; Illumina) at the DOE Joint Genome Institute.

SAGs were obtained from an Fe(III) oxide microbial mat sample collected from Echinus Geyser (transect B) on 30 April 2013. The Fe(III) oxide mat was treated with 0.01 M NH<sub>4</sub> oxalate (pH 3) to assist in disaggregation and dissolution of amorphous Fe(III) oxide phases (10 g wet sample: 1 l extracting solution) and pelleted via centrifugation. Individual cells were concentrated from the mixed ‘cell-sediment’ slurry using centrifugation through a 60% Nycodenz ‘cushion’. Subsamples of the Nycodenz layer were transferred to a 10% glycerol solution before freezing at –80 °C. Samples from Nycodenz cell enrichments were diluted and subjected to cell cytometry at the DOE Joint Genome Institute. Individual cell sorts were subjected to short-fragment 16S rRNA sequencing for preliminary identification, then sequenced using random Illumina sequencing. Genome sequences of six marsarchaeotal SAGs from ECH\_B were assembled and further curated using the G + C content and NWF algorithms described above (Supplementary Table 2).

Finally, an archaeal consortium containing a representative of the group 2 lineage of Marsarchaeota was cultivated (transferred numerous times over a 1 yr period) using Fe(III) oxide microbial mat from ECH\_B as inoculum into closed-headspace incubation bottles (1 g mat, 400 ml media) containing either peptone, casamino acids or tryptic soy broth (0.1% m/v), and grown at a pH of 3.5 and a temperature of 70–75 °C. The presence of group 2 Marsarchaeota in these enrichments was confirmed using 16S rRNA sequencing before random genome sequencing of biomass sampled from three replicate consortia. Genome sequence assemblies of these enrichments were curated using NWF algorithms as described above, which resulted in three nearly complete genomes of the group 2 Marsarchaeota (Table 1). These enrichments were also used in controlled experiments mixed with a pure culture of Fe(II)-oxidizing *M. yellowstonensis* strain MK1, which was originally isolated from Beowulf Spring East Fe(III) oxide mats<sup>19</sup>. Mixed cultures in the presence of different Fe(III) oxide phases (including killed controls) were treated first with an air headspace for 0–20 h, and then with N<sub>2</sub>(g) headspace from 20–40 h. Illumina (iTag) 16S rRNA sequencing was used to verify the presence of Marsarchaeota group 2 and *M. yellowstonensis* in these cultures. Concentrations of Fe(II) were monitored over time using the FerroZine colorimetric procedure.

**Phylogenetic and genome sequence analysis.** A total of 128 taxa were included in our initial phylogenomic analyses: 108 archaea, 10 bacteria and 10 eukarya

(Supplementary Table 1). Species used for bacteria and eukarya were identical to ref. <sup>6</sup>. We started with a set of 59 arCOGs<sup>6</sup> and used HMMER3 (ref. <sup>38</sup>) to identify homologues in all genomes. Following the initial analysis, we added arCOG04223 to our list and removed arCOG00109, arCOG00782, arCOG01704, arCOG01358 and arCOG04131, as their representatives were missing in more than 10% of studied archaeal species. Sequences of 55 individual arCOGs (or 54 after removing elongation factor 2, arCOG01559) were aligned with MAFFT-L-INS-i<sup>39</sup>. After visual inspection and adjustments, trimAL<sup>40</sup> was used to trim the positions with >50% gaps, and 54–55 alignments were concatenated into a master alignment (15,136 or 14,405 amino acid positions). We used MrBayes (version 3.2.5; MPI)<sup>41</sup> for Bayesian inference analysis and tested multiple combinations of parameters (2–5 million generations, 0.1–0.25 burn-in fraction, 4 and 8 parallel chains, gamma and invgamma models for rate variation, 4 and 8 rate categories for the gamma distribution, LG model with empirical amino acid frequencies and heating factors ranging from 0.075 to 0.15). Multiple runs reached excellent convergence in terms of standard deviation of split frequencies and potential scale reduction factor diagnostic values. The same procedure was followed for the reduced datasets shown in Supplementary Fig. 3d,e. ML trees were constructed using RAXML (version 8.2.0; PROTGAMMA + I, LG model with empirical amino acid frequencies) with the slow bootstrap option. For completeness, in several instances, we used proteins from closely related species as substitutes: (1) arCOGs 01179, 04184 and 04239 of a highly related (for example, Table 1) Marsarchaeota\_G2 species were used instead of missing proteins for Marsarchaeota\_G2-r01; (2) arCOGs 00415, 01001, 01228, 04050 and 04113 of a highly related Marsarchaeota\_G1 species were used instead of missing proteins for Marsarchaeota\_G1-r04; and (3) arCOGs 00405 and 04277 of Thorarchaeota-SMTZ1-45 were used instead of missing proteins for Thorarchaeota-SMTZ1-83.

Small subunit 16S rRNA sequences were aligned using SSU-ALIGN (version 0.1.1)<sup>42</sup>. After trimming to 1,494 positions, the alignment was analysed with MrBayes for up to 10 million generations (2 parallel runs with 8 chains each, 0.25 burn-in fraction, GTR model with 4 discrete categories approximating the gamma distribution).

Genome assemblies of group 1 and 2 Marsarchaeota (Table 1) were also used for comparative analysis against one another and against reference archaeal genomes. Average 16S rRNA and amino acid identities were determined by blastn and blastp comparisons, respectively<sup>43</sup>, while average genome nucleotide identities (ANIb) were determined with JSpecies version 1.2.1 (ref. <sup>44</sup>). The transcriptome sequence from Fe(III) oxide microbial mats was mapped against curated Marsarchaeota sequences to assess genes and pathways actively transcribed in situ using BBMap (<https://sourceforge.net/projects/bbmap/>).

**Scanning electron microscopy.** FE-SEM was used to image cell preparations of the enrichment culture from Echinus Geyser (pH = 3.5, 75 °C). Subsamples of the enrichment culture were fixed in 1.5% glutaraldehyde (v/v, final concentration), filtered through 1-cm-diameter 0.2 µm polycarbonate membranes and attached with carbon tape to 1 cm aluminium stubs for imaging at a working distance of 4 mm and a voltage of 1 keV.

**Fluorescence in situ hybridization.** Approximately 1 g of iron oxide microbial mat was sampled from Echinus Geyser (pH = 3.5, 66 °C) on 4 April 2014 and fixed with 1% paraformaldehyde (final concentration) for 5 min at 4 °C. The paraformaldehyde solution was removed and the sample was washed in a 1:1 solution of 100% ethanol:1× phosphate buffered saline, and stored in 1:1 phosphate buffered saline:ethanol at –20 °C until further FISH. The paraformaldehyde-fixed iron oxide mat samples were homogenized in the storage solution, and large particles were allowed to settle for 30 s. After the large particles had settled to the bottom, 20 µl of homogenized iron oxide mat was added to 0.3 cm<sup>2</sup> Teflon printed wells on a glass microscope slide (Structure Probe; SPI Supplies). The samples were then allowed to air dry on the slides for approximately 10 min. The samples were dehydrated in an increasing ethanol series at 50, 80 and 100% for 3 min each and dried with laboratory air after the final dehydration step. They were then hybridized in 30 µl of hybridization buffer (40% formamide, 0.9 M NaCl, 20 mM Tris HCl and 0.1% sodium dodecyl sulfate) containing 1 µl of Mars654 probe (designed around group 1 and 2 Marsarchaeota 16S genes<sup>42</sup>) dual labelled<sup>45</sup> with Cy3 (5'-Cy3-CCT CTT CCG CTC CCT AGC CC-Cy3-3'), Cy5 Arch915 and 6-FAM Aqi338, which are universal archaeal and Aquificales probes, respectively (working probe solutions were 30 ng µl<sup>-1</sup> for Cy3 and Cy5 and 50 ng µl<sup>-1</sup> for 6-FAM). The samples were then hybridized for 1.5 h at 46 °C, washed for exactly 10 min in pre-warmed wash buffer (46 mM NaCl, 20 mM Tris HCl and 5 mM ethylenediaminetetraacetic acid), rinsed in room temperature, distilled water and dried with laboratory air. The samples were visualized using a 63× glycerol objective lens on a Leica SP5 inverted confocal scanning laser microscope (Leica Microsystems) at the Montana State University (MSU) Center for Biofilm Engineering Confocal Microscopy Laboratory. More details of co-hybridization are provided in Supplementary Fig. 9.

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

**Data availability.** Genome sequences (metagenomes and SAGs) are available under National Center for Biotechnology Information BioProject ID [PRJNA362583](https://ncbi.nlm.nih.gov/bioproject/PRJNA362583) (accessions: NEXA00000000–NEXO00000000). Metagenome, transcriptome and single-cell assemblies from these sites (Supplementary Table 2) are available on the Integrated Microbial Genomes and Microbiomes system under Genomes OnLine Database Study ID [GS0045212](https://doi.org/10.1093/gob/obz012) (US Department of Energy Joint Genome Institute; CSP 787081 and 701).

Received: 25 May 2017; Accepted: 16 April 2018;  
Published online: 14 May 2018

## References

- Nunoura, T. et al. Insights into the evolution of archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res.* **39**, 3204–3223 (2011).
- Inskip, W. P. et al. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front. Microbiol.* **4**, 95 (2013).
- Kozubal, M. A. et al. Geoaerchaeta: a new candidate phylum in the archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J.* **7**, 622–634 (2013).
- Rinke, C. et al. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**, 431–437 (2013).
- Castelle, C. J. et al. Genomic expansion of domain archaea highlights roles for organisms from new phyla in anaerobic carbon cycling. *Curr. Biol.* **25**, 690–701 (2015).
- Spang, A. et al. Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173–179 (2015).
- Vanwonterghem, I. et al. Methylophilic methanogenesis discovered in the archaeal phylum Verstraetearchaeota. *Nat. Microbiol.* **1**, 16170 (2016).
- Zaremba-Niedzwiedzka, K. et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).
- Forterre, P. The universal tree of life: an update. *Front. Microbiol.* **6**, 717 (2015).
- Hug, L. A. et al. A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
- Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains archaea, bacteria, and eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
- Forterre, P. The common ancestor of archaea and eucarya was not an archaeon. *Archaea* **2013**, e372396 (2013).
- Raymann, K., Forterre, P., Brochier-Armanet, C. & Gribaldo, S. Global phylogenomic analysis disentangles the complex evolutionary history of DNA replication in archaea. *Genome Biol. Evol.* **6**, 192–212 (2014).
- Da Cunha, V., Gaia, M., Gabelle, D., Nasir, A. & Forterre, P. Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* **13**, e1006810 (2017).
- Beam, J. P. et al. Assembly and succession of iron oxide microbial mat communities in acidic geothermal springs. *Front. Microbiol.* **7**, 25 (2016).
- Inskip, W. P. et al. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS ONE* **5**, e9773 (2010).
- Jay, Z. J. et al. The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. *Environ. Microbiol.* **18**, 4755–4769 (2016).
- Beam, J. P., Jay, Z. J., Kozubal, M. A. & Inskip, W. P. Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J.* **8**, 938–951 (2014).
- Kozubal, M. A. et al. Microbial iron cycling in acidic geothermal springs of Yellowstone National Park: integrating molecular surveys, geochemical processes, and isolation of novel Fe-active microorganisms. *Front. Microbiol.* **3**, 109 (2012).
- Konstantinidis, K. T. & Tiedje, J. M. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **187**, 6258–6264 (2005).
- Brochier-Armanet, C., Gribaldo, S. & Forterre, P. A DNA topoisomerase IB in Thaumarchaeota testifies for the presence of this enzyme in the last common ancestor of archaea and eucarya. *Biol. Direct* **3**, 54 (2008).
- Dahmane, N. et al. *top1b*, a phylogenetic hallmark gene of Thaumarchaeota encodes a functional eukaryote-like topoisomerase IB. *Nucleic Acids Res.* **44**, 2795–2805 (2016).
- Ney, B. et al. The methanogenic redox cofactor F420 is widely synthesized by aerobic soil bacteria. *ISME J.* **11**, 125–137 (2017).
- Cook, A. M., Smits, T. H. M. & Denger, K. in *Microbial Sulfur Metabolism* (eds Dahl, C. & Friedrich, C. G.) 170–183 (Springer, Berlin, Heidelberg, 2008).
- Müller, F. H. et al. Coupling of the pathway of sulphur oxidation to dioxygen reduction: characterization of a novel membrane-bound thiosulphate:quinone oxidoreductase. *Mol. Microbiol.* **53**, 1147–1160 (2004).
- Kozubal, M. A., Dlakić, M., Macur, R. E. & Inskip, W. P. Terminal oxidase diversity and function in “*Metallosphaera yellowstonensis*”: gene expression and protein modeling suggest mechanisms of Fe(II)-oxidation in the Sulfolobales. *Appl. Environ. Microbiol.* **77**, 1844–1853 (2011).
- Castelle, C. J. et al. The aerobic respiratory chain of the acidophilic archaeon *Ferroplasma acidiphilum*: a membrane-bound complex oxidizing ferrous iron. *Biochim. Biophys. Acta* **1847**, 717–728 (2015).
- David, L. A. & Alm, E. J. Rapid evolutionary innovation during an archaean genetic expansion. *Nature* **469**, 93–96 (2010).
- Stolper, D. A., Revsbech, N. P. & Canfield, D. E. Aerobic growth at nanomolar oxygen concentrations. *Proc. Natl Acad. Sci. USA* **107**, 18755–18760 (2010).
- Dobbek, H., Gremer, L., Kiefersauer, R., Huber, R. & Meyer, O. Catalysis at a dinuclear [CuSmO(O)OH] cluster in a CO dehydrogenase resolved at 1.1-Å resolution. *Proc. Natl Acad. Sci. USA* **99**, 15971–15976 (2002).
- Bell, J. F., McCord, T. B. & Owensby, P. D. Observational evidence of crystalline iron oxides on Mars. *J. Geophys. Res. Solid Earth* **95**, 14447–14461 (1990).
- Morris, R. V. et al. Mineralogy, composition, and alteration of Mars Pathfinder rocks and soils: evidence from multispectral, elemental, and magnetic data on terrestrial analogue, SNC meteorite, and Pathfinder samples. *J. Geophys. Res. Planets* **105**, 1757–1817 (2000).
- Vargas, M., Kashefi, K., Blunt-Harris, E. L. & Lovley, D. R. Microbiological evidence for Fe(III) reduction on early Earth. *Nature* **395**, 65–67 (1998).
- Nisbet, E. G. & Sleep, N. H. The habitat and nature of early life. *Nature* **409**, 1083–1091 (2001).
- Chen, I.-M. A. et al. IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* **45**, D507–D516 (2017).
- Van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
- Hahsler, M., Piekenbrock, M., Arya, S. & Mount, D. Package ‘*dbSCAN*’: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms (CRAN, 2016).
- Eddy, S. R. Accelerated profile HMM searches. *PLoS Comput. Biol.* **7**, e1002195 (2011).
- Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
- Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
- Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
- Nawrocki, E. *Structural RNA Homology Search and Alignment Using Covariance Models* PhD thesis, Washington Univ. (2009).
- Camacho, C. et al. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
- Richter, M. & Rosselló-Móra, R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl Acad. Sci. USA* **106**, 19126–19131 (2009).
- Stoecker, K., Dorninger, C., Daims, H. & Wagner, M. Double labeling of oligonucleotide probes for fluorescence in situ hybridization (DOPE-FISH) improves signal intensity and increases rRNA accessibility. *Appl. Environ. Microbiol.* **76**, 922–926 (2010).
- Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
- Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- Lindås, A.-C., Karlsson, E. A., Lindgren, M. T., Etema, T. J. G. & Bernander, R. A unique cell division machinery in the archaea. *Proc. Natl Acad. Sci. USA* **105**, 18942–18946 (2008).

## Acknowledgements

The authors acknowledge support from the DOE Pacific Northwest National Laboratory (subcontracts 112443 and 254840), DOE Joint Genome Institute Community Sequencing Program (CSP 787081 and CSP 701), National Science Foundation Integrative Graduate Education and Research Traineeship Program (Z.J.J. and J.P.B.) (NSF DGE 0654336) and Montana Agricultural Experiment Station (W.P.I.). Work conducted by the Pacific Northwest National Laboratory (Foundational Scientific Focus Area) and Joint Genome Institute (DOE-AC02-05CH11231) is supported by the Genomic Science Program, Office of Biological and Environmental Research, US DOE. The authors appreciate assistance from S. Tringe, T. Woyke and D. Goudeau for genome and single-cell sequencing at the DOE Joint Genome Institute, C. Carey and A. Mazurie (MSU) for Illumina metagenome and iTag data processing, R. Jennings (MSU) for genome curation, R. Carlson (MSU) for collaboration on metabolic pathway analysis, and C. Hendrix,

S. Sigler and D. Hallac (Center for Resources, YNP) for permitting this work in YNP (permits YELL-SCI-5068 and -5686). Computations were performed on the Hyalite High-Performance Computing System, operated and supported by MSU's Information Technology Center.

### Author contributions

W.P.I., Z.J.J., J.P.B. and M.A.K. designed the study. Z.J.J., J.P.B., M.A.K. and W.P.I. collected and analysed the field data. Z.J.J., M.D. and D.B.R. conducted the informatic analyses. M.D., W.P.I. and J.P.B. conducted the phylogenetic analysis. M.A.K. and J.P.B. established and analysed the enrichment cultures. Z.J.J. and J.P.B. analysed the transcriptomic data. J.P.B. and W.P.I. performed microscopy and FISH analysis. W.P.I., M.D., Z.J.J. and J.P.B. wrote the manuscript. All authors reviewed and edited the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41564-018-0163-1>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to W.P.I.

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

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### ▶ Experimental design

#### 1. Sample size

Describe how sample size was determined.

Sample size for metagenome sequencing was determined based on microbial mat quantities necessary for successful DNA extraction and sequencing at the DOE-Joint Genome Institute (Walnut Creek CA), and according to temperature ranges measured in the field corresponding to variations of < 1-2 degrees Celcius (approximately 2 gram quantities). This sample size was sufficient because it provided excess microbial cells from which to obtain a representative DNA sample from the microbial community. Moreover, replicate DNA extractions using this sample size have resulted in reproducible abundances of different community members.

#### 2. Data exclusions

Describe any data exclusions.

No data were excluded from the analysis.

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

Yes, replicate metagenomes were obtained over several years corresponding to different samples from several different geothermal springs with similar geochemical parameters (data included in manuscript). Metagenome sequencing was also performed using several different sequencing platforms. Replicate samples over repeated sampling events spanning several years were used for obtaining replicate long-fragment 16S rRNA gene sequences. Replicate assemblies were performed during manuscript revision and were consistent with prior versions. Replicate phylogenomic trees were generated and these new results have been included in the revised manuscript.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

Allocation was not applicable to the current study. Samples were collected from numerous habitat types, and experimental methods involved numerous characterization and analysis techniques.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was not applicable to the study, because results were primarily analytical in nature and were not subject to allocation.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

- |                                     |   |
|-------------------------------------|---|
| n/a                                 | Confirmed   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The <u>exact</u> sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.                             |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement indicating how many times each experiment was replicated  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as an adjustment for multiple comparisons  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> The test results (e.g. $p$ values) given as exact values whenever possible and with confidence intervals noted   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clearly defined error bars   |

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

Software described in Methods section included the following:

HMMER3 (v3.1b2)  
 MAFFT-L-INS-i (v7.310)  
 trimAL (v1.2)  
 MrBayes (v3.2.5; MPI)  
 RAxML (v8.2.0)  
 SSU-ALIGN (v0.1.1)  
 JSpecies (v1.2.1)  
 BMap (<https://sourceforge.net/projects/bbmap/>)  
 NCBI-BLAST (v2.6.0+)  
 JGI IMG-M/ER (v4)  
 SOAPdenovo2  
 Newbler (v2.7)  
 SPAdes (3.11)  
 CheckM (v1.0.7)  
 DBSCAN 0.9-8

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* [guidance for providing algorithms and software for publication](#) may be useful for any submission.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Sequence data is available to the public, and this is the primary 'material' being analyzed in the current study. Samples from subject geothermal springs in Yellowstone National Park (USA) are available to research scientists, but would require necessary permitting as dictated by the U.S. National Park Service.

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies used in the current study.

10. Eukaryotic cell lines

- a. State the source of each eukaryotic cell line used.
- b. Describe the method of cell line authentication used.
- c. Report whether the cell lines were tested for mycoplasma contamination.
- d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No eukaryotic cell lines used.

No eukaryotic cell lines used.

No eukaryotic cell lines used.

No cell lines used.

► **Animals and human research participants**

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals used.

Policy information about [studies involving human research participants](#)

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Study did not involve human research participants.