PNAS
www.pnas.org

Supplementary Information for

Complex subsurface hydrothermal fluid mixing at a submarine arc volcano supports distinct and highly diverse microbial communities

Anna-Louise Reysenbach, Emily St. John, Jennifer Meneghin, Gilberto Flores, Mircea Podar, Nina Dombrowski, Anja Spang, Stephane L'Haridon, Susan E. Humphris, Cornel E. J. de Ronde, Fabio Caratori Tontini, Maurice Tivey, Valerie K. Stucker, Lucy C. Stewart, Alexander Diehl, Wolfgang Bach

* Anna-Louise Reysenbach
Email: bwar@pdx.edu

This PDF file includes:

- SI Text
- SI Methods
- Figures S1-S11
- List of Figshare Files
- Legends for Datasets S1 to S5
- SI References

Other supplementary materials for this manuscript include the following:

- Datasets S1 to S5

SI Text

Nomenclature.

Since the nomenclature of Bacteria and Archaea is in flux, we use the NCBI taxonomy wherever possible, and put the GTDB classification in parentheses. Members of the Epsilonbacteraeota (formerly Epsilonproteobacteria) are described by their reclassified names, as proposed by Waite et al. (1).

Brothers volcano bacterial MAG diversity.

Many MAGs recovered from Brothers volcano were related to mesophilic or mildly thermophilic Bacteria common to hydrothermal systems (Fig. 4, Dataset S3*D*). These included relatives of Epsilonbacteraeota (1) (GTDB classification Campylobacterota), related to genera such as *Hippea*, *Sulfurovum*, *Sulfurimonas* and *Nitratiruptor*. Iron oxide mats, such as those found at the diffuse venting site S009, were colonized by members of the iron-oxidizing clade Zetaproteobacteria. Thirty Chloroflexi (Chloroflexota) MAGs representing 16 new genera were recovered (Dataset S3*B*, *SI Appendix,* Fig. S7) and were related to obligately and facultatively anaerobic thermophiles within the Anaerolineae (Anaerolineales), Ardenticatenia (Ardenticatenales), Caldilineae (Caldilineales) and Thermoflexia (Thermoflexales). Significantly, the Chloroflexi MAGs recruited more than 10% of the reads that mapped back to MAGs in the Lower Cone sample S014 (Dataset S3*D*). Also prevalent in this sample were representatives of the WOR-3 lineage (~20% of the reads, Dataset S3*D*), which is related to EM3 from a Yellowstone hotspring, and members of the Candidate Phyla Radiation, or CPR (2) (Patescibacteria, ~6% reads mapped back to MAGs).

Microaerophilic Aquificae (Aquificota) were also phylogenetically diverse at Brothers volcano, representing approximately five new genera (Dataset S3*B*, *SI Appendix,* Fig. S8). In keeping with both GTDB-Tk and SILVA taxonomy, we did not consider MAGs related to the *Thermosulfidibacter* as part of the Aquificae. Aquificae MAGs were mainly derived from samples

at NWC-B+UCW. Many of the Aquificales-related MAGs encoded genes generally present in organisms adapted to either high (*coxA, coxB*) or low (*cydA, cydB*) oxygen environments (3), suggesting a range of aerobic to microaerophilic lifestyles. Several of the MAGs also encoded a sulfide:quinone oxidoreductase (*sqr*) used in sulfide oxidation. Read coverage of Aquificae was above 10% in several samples, with highest coverage in S141 (49%, Dataset S3*D*).

Like the Aquificae, the high diversity of Gammaproteobacteria MAGs was primarily restricted to samples from NWC-B+UCW (Dataset S3*D*), and the most abundant MAGs were within the GTDB order, Thiohalomonadales. Other MAGs related to thermophilic Bacteria included members of the Thermotogae (Thermotogota), Calditrichaeota (Calditrichota) and Caldiserica (Caldisericota).

Brothers volcano archaeal MAG diversity.

Furthermore, we recovered various MAGs from established archaeal phyla including from members of the proposed DPANN superphylum (2) (*SI Appendix,* Fig. S5*C* and S6) and of numerous archaeal lineages with poor genomic representation in public databases.

Diversity of DPANN superphylum MAGs.

The DPANN superphylum is a proposed group of archaeal phyla of which many, if not all, may be symbionts with reduced genome sizes (reviewed in (4)). We detected DPANN MAGs related to the Parvarchaeota, Woesearchaeota/DHVE-6 (Woesearchaeia), Pacearchaeota (Pacearchaeales), Aenigmarchaeota (Aenigmarchaeia), Nanohaloarchaeota (Nanohaloarchaea), Micrarchaeota and Diapherotrites (Iainarchaeia). While the Woesearchaeota and Aenigmarchaeota were present at several of the sites, the highest phylogenetic diversity of DPANN phyla was detected in one of the Lower Cone samples (S014), wherein the DPANN, and in particular, two thus far uncharacterized lineages, accounted for

over 60% of the archaeal reads mapped to medium to high quality MAGs (Dataset S3*E, SI Appendix,* Fig. S5*C*).

Based on our analysis, Brothers volcano DPANN MAGs represent about 26 new genera (Dataset S3*B*). Many of the MAGs share a reduced metabolic gene repertoire consistent with a symbiotic lifestyle; however, the genes that they have retained are not consistent within or among groups (Dataset S4), which may be partially due to variation in estimated genome completeness. Notably, many DPANN genomes, such as representatives of the Aenigmarchaeota and Pacearchaeota, lack the genes coding for proteins driving ATP-synthesis via the electron transport chain and membrane-bound ATP synthase. Several alternative mechanisms have been proposed to explain how these symbionts conserve energy or may obtain ATP by substrate-level phosphorylation, including a variety of fermentative pathways (reviewed in (4)) and the pentose bisphosphate pathway, which is proposed to degrade nucleosides for entry into glycolysis, followed by acetate production. Many of our Brothers volcano DPANN members likely use the latter mechanism, although even within some lineages, such as the Micrarchaeota, this ability is not consistently detectable (Dataset S4). Interestingly, like recently reported for the Nanoarchaeota (5), several of the Brothers volcano Diapherotrites may be motile as their genomes encode several archaeal flagellar proteins (*flaBD/EGHJ*) (6).

Only a limited number of DPANN members have been co-cultivated or identified with their hosts. These include representatives of the Nanoarchaeota (5), Micrarchaeota (7, 8), Nanohaloarchaeota (9) and the Huberarchaea (Huberarchaeota) (10). About 6% of the archaeal reads from Lower Cone samples S014 and S016 mapped back to Micrarchaeota MAGs, and both samples also had high read coverage (12-39% of archaeal reads) for several Thermoplasmata most closely related to *Aciduliprofundum* spp. (Dataset S3*E*). Given that other Micrarchaeota have been co-cultured with members of the Thermoplasmata (7, 8), it is possible that at least some of the Brothers Lower Cone Micrarchaeota may interact with members of the Thermoplasmata.

Enrichment of a novel woesearchaeote.

In all the above-mentioned co-cultured host-symbiont systems, the host of the DPANN is an archaeum. Here we report an enrichment culture of a Woesearchaeota from S139 (NWC-B+UCW) that appears to be associated with a bacterial host. After 7 days, at 80 ºC and at 60 ºC, we obtained a mixed culture (S21) containing a member of the Woesearchaeota, which was maintained over three successive transfers. Based on 16S rRNA gene amplicon sequencing, the enrichment cultures consisted predominantly of members of the Hydrogenothermaceae and Thermaceae (*Vulcanithermus* and *Oceanithermus*), with the woesearchaeote (6-20% of the culture) being the only archaeal amplicon detected, suggesting that the woesearchaeote may either be free-living or interacting with a bacterial partner, as these were the only other amplicons/MAGs detected in the culture. Initial amplicon data pointed to an increase in woesearchaeote abundance coinciding with a drop in *Vulcanithermus* abundance and an increased abundance of Hydrogenothermaceae. From the enrichment metagenome (11,694,970 bp assembly), we obtained seven MAGs, of which four were ≥50% complete, with ≤5% contamination (Dataset S5*A*). The only archaeal MAG was the S21 woesearchaeote (S21_7, *SI Appendix,* Fig. S9), which was similar in length (826,223 bp, 831 predicted proteins) and estimated completion to other members of the Woesearchaeota (Dataset S5*A*). No genes encoding the ATP synthase complex were identified, and substrate-level phosphorylation pathways were either partial or undetected in the MAG (Dataset S5*B*), suggesting that the S21 woesearchaeote may rely on a partner organism, as previously suggested for other members of this lineage (11). Like several other Woesearchaeota, S21_7 has a glycosyl transferase gene involved in saccharide synthesis, and it has several genes coding for glycolytic enzymes (11), but it lacks any genes for proteins of the pentose phosphate pathway and the ribulose monophosphate pathway. However, it is possible that S21_7 may encode genes for the pentose bisphosphate pathway in the un-sequenced portions of the genome, given that some genes from this pathway are found in two related MAGs, Woesearchaeota UBA119 (NCBI accession

DAWM00000000.1, 42% AAI, Dataset S5*B*) and S143_142 (45% AAI, Dataset S4). Furthermore, S21 contains a gene encoding a 4Fe-4S ferredoxin, which has been found in Woesearchaeota MAGs from anoxic environments (11) and suggests that these woesearchaeotes are anaerobes.

In order to explore what potential interactions might occur between the woesearchaeote and the bacterial co-cultures, we analyzed the enrichment culture MAGs, which revealed that the Thermaceae (*Vulcanithermus*-like S21_6 and *Oceanithermus*-related S21_1, *SI Appendix,* Fig. S10) and Hydrogenothermaceae (S21_4, *SI Appendix,* Fig. S11) MAGs encode subunits of a nitrate reductase (*narGH*) and therefore it is likely they can reduce nitrate. The Thermaceae in the enrichment can probably tolerate high levels of oxygen (3), as indicated by the presence of genes for A-family heme-copper oxygen reductases (*coxA, coxB)*, and they may be facultative anaerobic heterotrophs able to use a suite of carbohydrates and possibly amino acids (Dataset S5*C*). Conversely, the Hydrogenothermaceae representative in the enrichment likely performs microaerophilic sulfur oxidation, given the presence of the *soxYZ* gene cassette and genes encoding cytochrome *bd* oxidases (*cydA, cydB*) adapted to low oxygen environments (3) in the MAG (S21_4).

Collectively, because of the increases in the co-abundances of the woesearchaeote and Hydrogenothermaceae 16S rRNA gene amplicons and the MAG analysis showing the woesearchaeote is likely an anaerobe, we hypothesize that Woesearchaeota S21_7 forms an association with the Hydrogenothermaceae. Further studies and purification are ongoing to determine which role the other members in the enrichment play in the growth of S21. Nonetheless, the reduced metabolic repertoire of the S21 Woesearchaeota combined with the lack of any other Archaea detectable in the enrichment cultures strongly suggest that some members of the DPANN superphylum may engage in inter-domain associations with bacterial partners.

Identification of novel Archaea from poorly represented phyla.

Numerous MAGs related to other archaeal lineages were present at Brothers volcano and many of these are poorly represented in public databases. For example, we detected nine Geothermarchaeota (JdFR-13) and two novel families within the GTDB orders LC-3 and LC-2 of the Heimdallarchaeota (Heimdallarchaeia, Dataset S3*B*).

We identified four MAGs belonging to the Aigarchaeota, three of which were assigned genus-level novelty in the GTDB family Caldiarchaeaceae and one which was assigned to a new genus in the GTDB family JGI-0000106-J15. These MAGs likely represent organisms capable of aerobic growth, since they have genes coding for A-family oxygen reductases (*coxA* and/or *coxB*). Furthermore, several encode for a sulfide:quinone oxidoreductase (*sqr*), which may enable the oxidation of reduced sulfur (12). We did not recover any dissimilatory sulfite reductase (*dsrAB*) genes identified in other Aigarchaeota (12) from these MAGs.

Further, we recovered nine Hydrothermarchaeota (JdFR-18) MAGs, representing four new genera and potentially a new family. Like those from deep-sea vents elsewhere (13), several of the Hydrothermarchaeota MAGs encode genes in the Wood-Ljungdahl CO_2 fixation pathway and genes likely involved in carboxydotrophy, including the anaerobic-type carbon monoxide dehydrogenase gene *cooS*, related to proposed carboxydotrophic *cooS* from Hydrothermarchaeota JdFR-17 and clade D *cooS* from *Archaeoglobus, Geoglobus,* and *Ferroglobus*. While some Hydrothermarchaeota may be able to conserve energy by dissimilatory reduction of sulfate (13), we did not detect the genes for the first steps in this pathway (*sat, aprAB)*. However, several MAGs did encode the genes for dissimilatory sulfite reductase (*dsrAB*). As has been reported previously for some Hydrothermarchaeota (14), several MAGs also have *nifH* genes and may be capable of nitrogen fixation.

The Brothers volcano Thaumarchaeota (UBA164) MAGs most closely affiliated with thaumarchaeotes from Beowulf and Dragon Springs in Yellowstone National Park (IMG/JGI accession numbers Ga0011949, Ga0010782, Ga0010781) and Shi-Huang-Ping hot spring in

Taiwan (Genbank accession numbers DAJU00000000.1, DAXX00000000.1). Like their Yellowstone relatives, they show no evidence of ammonia oxidation capability as is common to cultivated Thaumarchaeota (ammonia monooxygenase, *amoABC*).

We also obtained a single MAG belonging to the Altiarchaeota from the Lower Cone site (S014). The Altiarchaeota were first described from cold sulfidic stream environments (15) and subsequently found to be widespread in marine and terrestrial sediments and hot springs, forming two distinct phylogenetic clades, Alti-1 and Alti-2, respectively (16). Those associated with stream environments (Alti-1) generally possess unique cellular appendages known as hami, which resemble grappling hooks (17). Although our MAG was reconstructed from a sulfidic environment, it does not appear to have the genomic capability to produce hami hooks. However, it does encode the coenzyme-A disulfide reductase predicted to be unique to the Altiarchaeota Alti-2 clade (16).

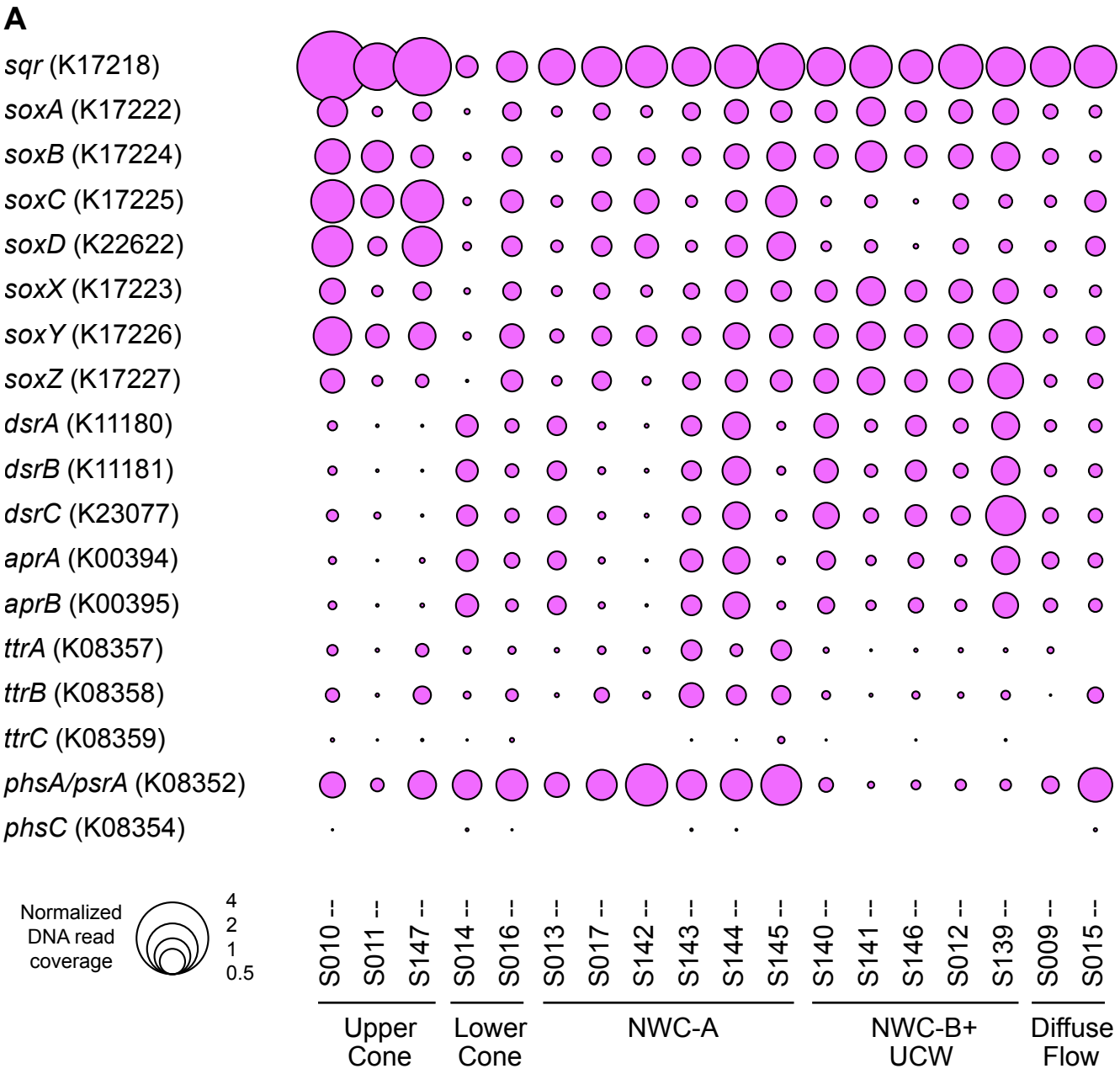
SI Methods

Annotation of MAGs of interest to this study. MAGs within the DPANN superphylum were further annotated using multiple functional databases, namely Prokka (18), the arCOG database (19), the KEGG database (20) and dbCAN-seq (21). Predicted cellular appendages were identified using previously described arCOG identifiers (6). Hydrogenases were identified using KO annotations and validated with HydDB (22) as described previously in *Materials and Methods*.

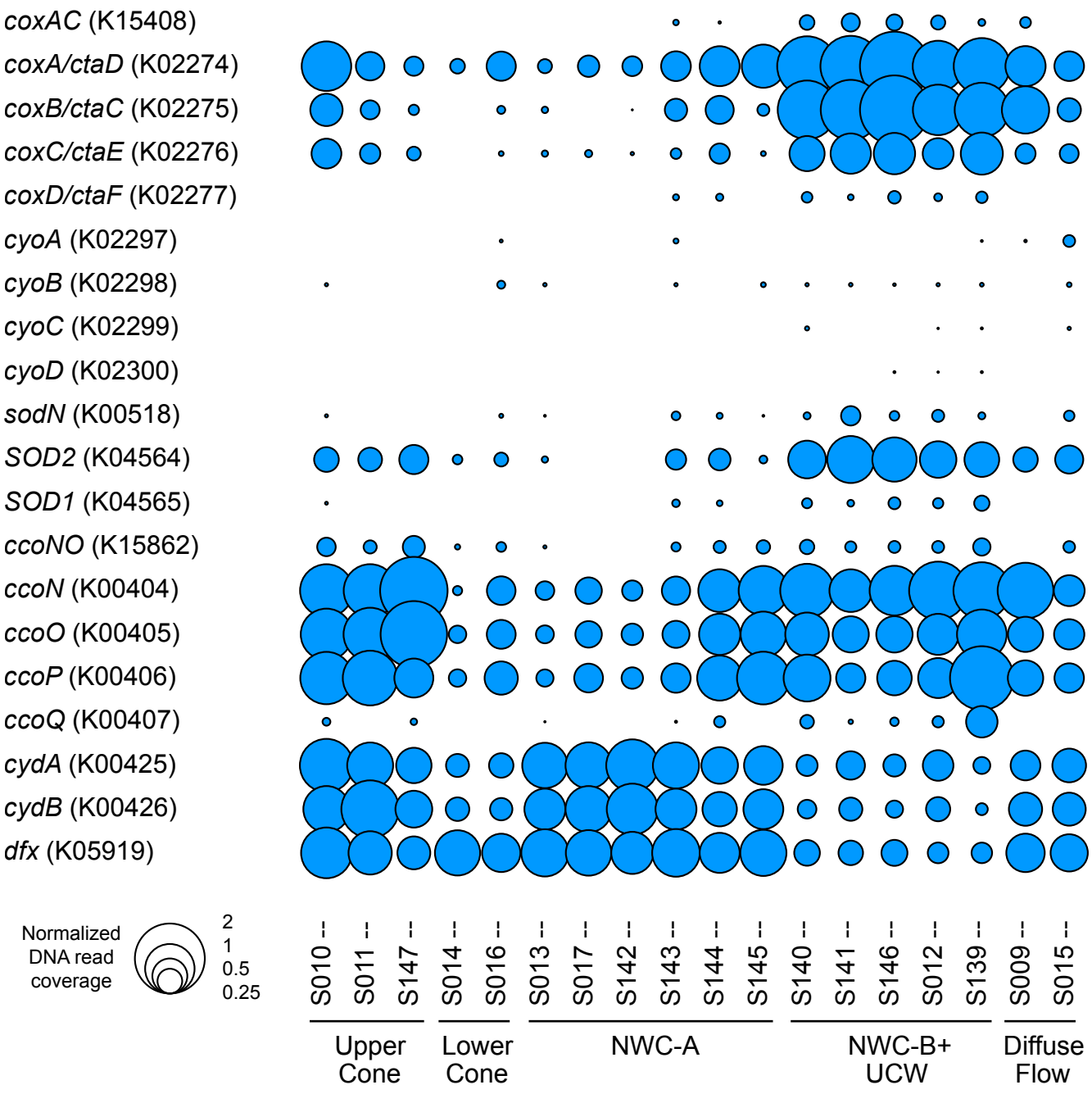
MAGs of interest other than those in the DPANN superphylum were annotated using the KEGG Automatic Annotation Server (KAAS) (23) and/or the arCOG database (19). Gene calls were confirmed using the Conserved Domain Database (24) as needed. One MAG from the Altiarchaeota was also annotated with Prokka (18) for comparison with publicly available Altiarchaeota annotations (16).

Enrichment culturing and analysis of woesearchaeote enrichment. Crushed anaerobic hydrothermal deposit slurries were inoculated (10%) into an anaerobic medium containing per liter: 20 g NaCl, 4 g $MgCl_2.H_2O$, 4 g $MgSO_4.7H_2O$, 0.3 g NH_4Cl, 0.3 g KCl, 0.3g $CaCl_2.2H_2O$, 10 ml trace element solution (Medium DSMZ 141), 2 mg $Fe(NH_4)_2SO_4$, 0.2 g Oxoid yeast extract (ThermoFisher Scientific, Waltham, MA, USA), 1 g $NaHCO_3$, 0.14 g K_2HPO_4, 0.2 mg sodium tungstate, 50 mg sodium selenate, 1 ml vitamin solution (25), 1 ml 20 amino acid solution (final concentration 1 µM), resazurin solution (1 mg/l). The medium was boiled and purged with N_2/CO_2 gas (80:20, v/v), dispensed, autoclaved and pressurized with CO (0.5 bar). The enrichment cultures were incubated at 60 ºC and 80 ºC for 4 to 7 days. DNA from cultures showing positive growth under phase microscopy was extracted using the Qiagen DNeasy Blood and Tissue kit (Hilden, Germany) and analyzed by iTAG amplicon sequencing as described in the *Material and Methods*. The metagenome of the enrichment culture was obtained using the MiSeq platform at Oak Ridge National Lab and processed as described above. Phylogenetic ribosomal protein trees were constructed as described above for the Hydrogenothermaceae (*n* = 16 taxa, 2,371 amino acid positions) and the Thermaceae (*n* = 8 taxa, 2,384 amino acid positions), and an archaeal GTDB-Tk phylogenetic tree was inferred (*n* = 1,372 taxa, 5,024 amino acid positions) to show the position of the S21 Woesearchaeota.

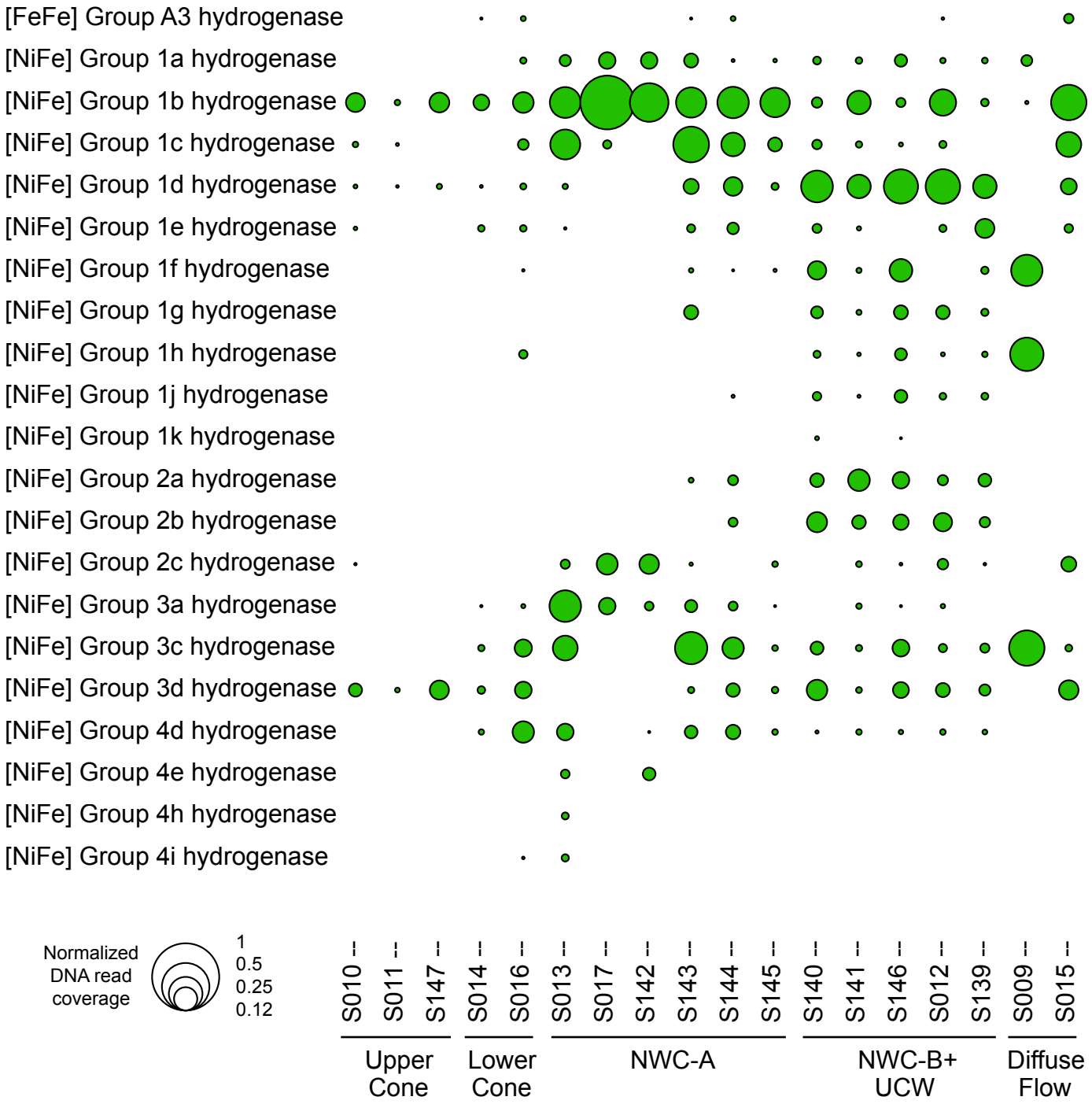
Fig. S1. Photographs of hydrothermal vent deposits collected from Brothers volcano.


A
sqr (K17218)
soxA (K17222)
soxB (K17224)
soxC (K17225)
soxD (K22622)
soxX (K17223)
soxY (K17226)
soxZ (K17227)
dsrA (K11180)
dsrB (K11181)
dsrC (K23077)
aprA (K00394)
aprB (K00395)
ttrA (K08357)
ttrB (K08358)
ttrC (K08359)
phsA/psrA (K08352)
phsC (K08354)
Normalized DNA read coverage
4
2
1
0.5
S010 --
S011 --
S147 --
S014 --
S016 --
S013 --
S017 --
S142 --
S143 --
S144 --
S145 --
S140 --
S141 --
S146 --
S012 --
S139 --
S009 --
S015 --
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow

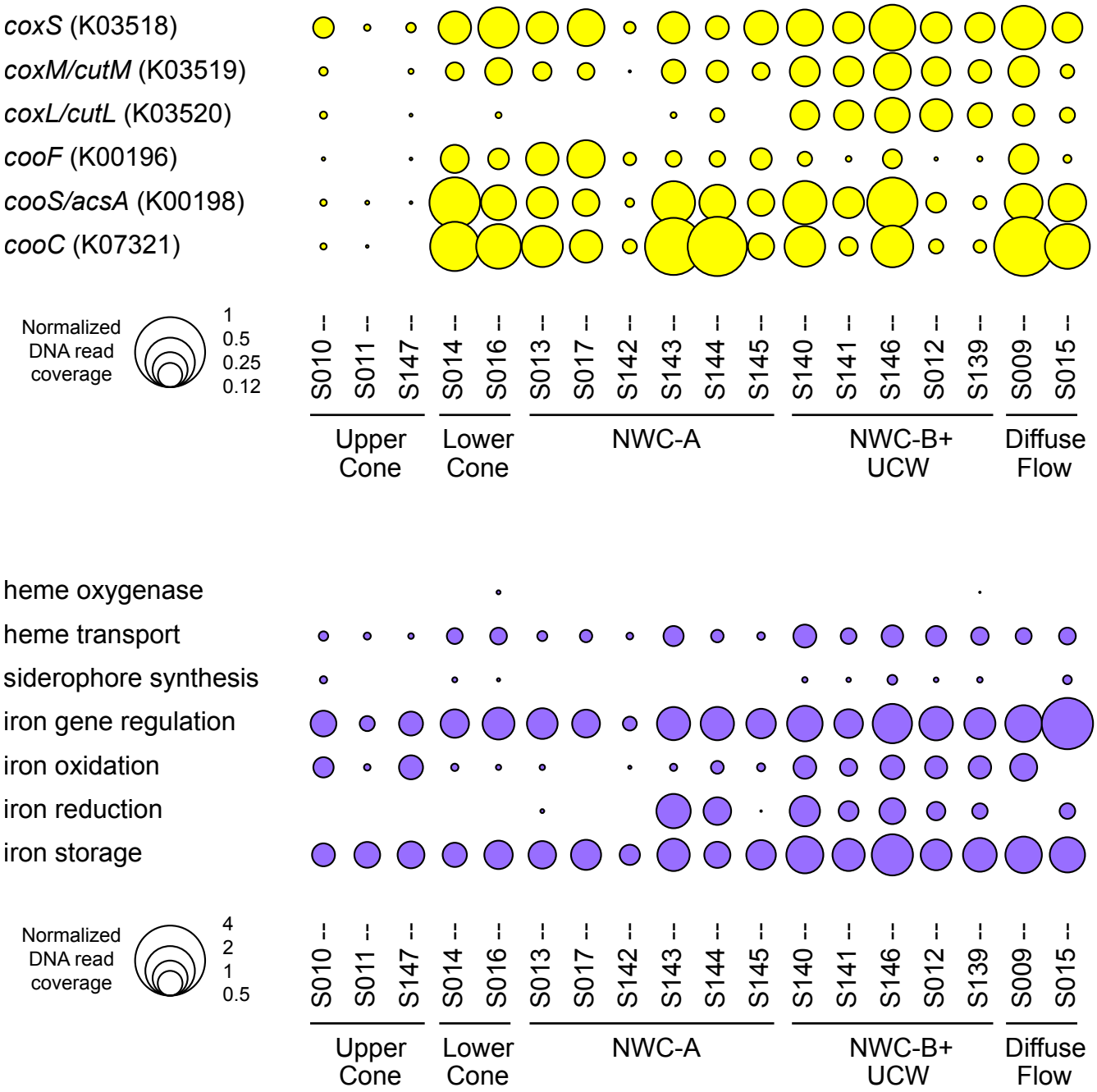
B


coxAC (K15408)
coxA/ctaD (K02274)
coxB/ctaC (K02275)
coxC/ctaE (K02276)
coxD/ctaF (K02277)
cyoA (K02297)
cyoB (K02298)
cyoC (K02299)
cyoD (K02300)
sodN (K00518)
SOD2 (K04564)
SOD1 (K04565)
ccoNO (K15862)
ccoN (K00404)
ccoO (K00405)
ccoP (K00406)
ccoQ (K00407)
cydA (K00425)
cydB (K00426)
dfx (K05919)
Normalized DNA read coverage
2
1
0.5
0.25
S010 --
S011 --
S147 --
S014 --
S016 --
S013 --
S017 --
S142 --
S143 --
S144 --
S145 --
S140 --
S141 --
S146 --
S012 --
S139 --
S009 --
S015 --
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow

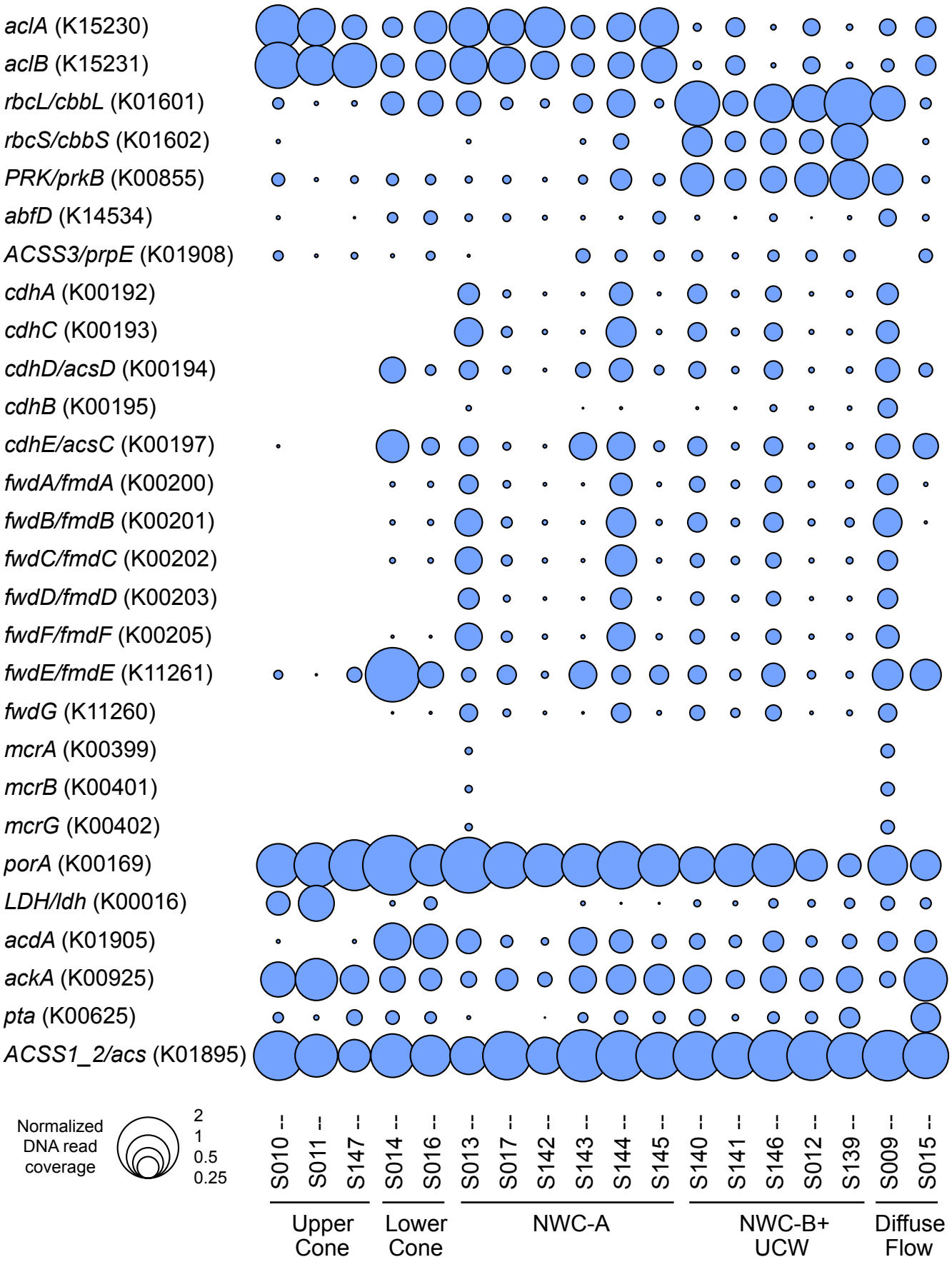
C


[FeFe] Group A3 hydrogenase
[NiFe] Group 1a hydrogenase
[NiFe] Group 1b hydrogenase
[NiFe] Group 1c hydrogenase
[NiFe] Group 1d hydrogenase
[NiFe] Group 1e hydrogenase
[NiFe] Group 1f hydrogenase
[NiFe] Group 1g hydrogenase
[NiFe] Group 1h hydrogenase
[NiFe] Group 1j hydrogenase
[NiFe] Group 1k hydrogenase
[NiFe] Group 2a hydrogenase
[NiFe] Group 2b hydrogenase
[NiFe] Group 2c hydrogenase
[NiFe] Group 3a hydrogenase
[NiFe] Group 3c hydrogenase
[NiFe] Group 3d hydrogenase
[NiFe] Group 4d hydrogenase
[NiFe] Group 4e hydrogenase
[NiFe] Group 4h hydrogenase
[NiFe] Group 4i hydrogenase
Normalized DNA read coverage
1
0.5
0.25
0.12
S010
S011
S147
S014
S016
S013
S017
S142
S143
S144
S145
S140
S141
S146
S012
S139
S009
S015
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow

D


coxS (K03518)
coxM/cutM (K03519)
coxL/cutL (K03520)
cooF (K00196)
cooS/acsA (K00198)
cooC (K07321)
Normalized DNA read coverage
1
0.5
0.25
0.12
S010
S011
S147
S014
S016
S013
S017
S142
S143
S144
S145
S140
S141
S146
S012
S139
S009
S015
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow
heme oxygenase
heme transport
siderophore synthesis
iron gene regulation
iron oxidation
iron reduction
iron storage
Normalized DNA read coverage
4
2
1
0.5
S010
S011
S147
S014
S016
S013
S017
S142
S143
S144
S145
S140
S141
S146
S012
S139
S009
S015
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow

E


aclA (K15230)
aclB (K15231)
rbcL/cbbL (K01601)
rbcS/cbbS (K01602)
PRK/prkB (K00855)
abfD (K14534)
ACSS3/prpE (K01908)
cdhA (K00192)
cdhC (K00193)
cdhD/acsD (K00194)
cdhB (K00195)
cdhE/acsC (K00197)
fwdA/fmdA (K00200)
fwdB/fmdB (K00201)
fwdC/fmdC (K00202)
fwdD/fmdD (K00203)
fwdF/fmdF (K00205)
fwdE/fmdE (K11261)
fwdG (K11260)
mcrA (K00399)
mcrB (K00401)
mcrG (K00402)
porA (K00169)
LDH/ldh (K00016)
acdA (K01905)
ackA (K00925)
pta (K00625)
ACSS1_2/acs (K01895)
Normalized DNA read coverage
2
1
0.5
0.25
S010 --
S011 --
S147 --
S014 --
S016 --
S013 --
S017 --
S142 --
S143 --
S144 --
S145 --
S140 --
S141 --
S146 --
S012 --
S139 --
S009 --
S015 --
Upper Cone
Lower Cone
NWC-A
NWC-B+ UCW
Diffuse Flow

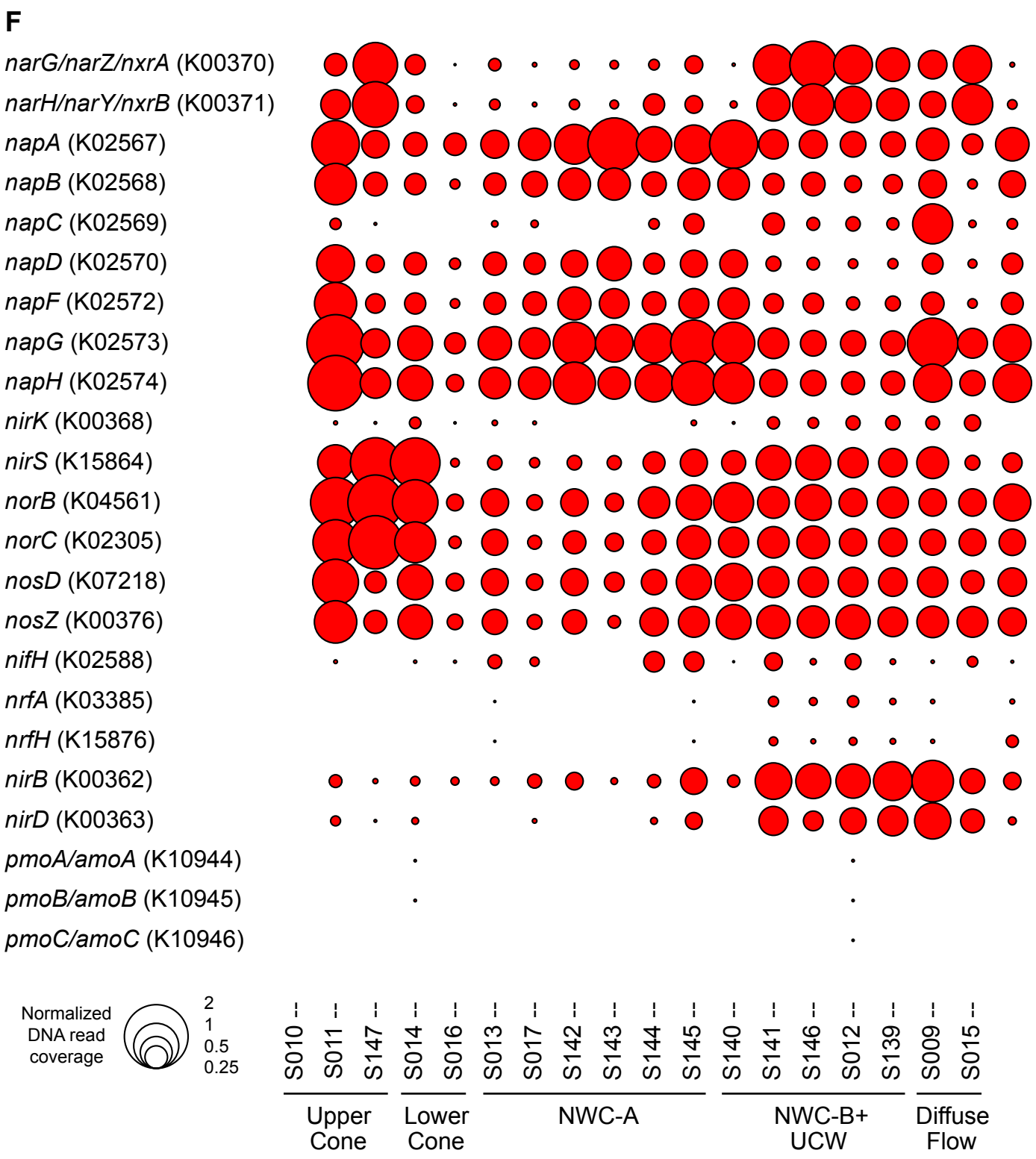


Fig. S2. Normalized abundance of select metabolic genes in Brothers volcano assemblies. Bubble size indicates the relative abundance of each gene/gene category, determined by normalizing summed read coverage of functional genes by the average summed read coverage of 14 single-copy marker genes. Genes involved in (*A*) sulfur, (*B*) oxygen, (*C*) hydrogen, (*D*) carbon monoxide and iron, (*E*) carbon, and (*F*) nitrogen metabolism are shown.

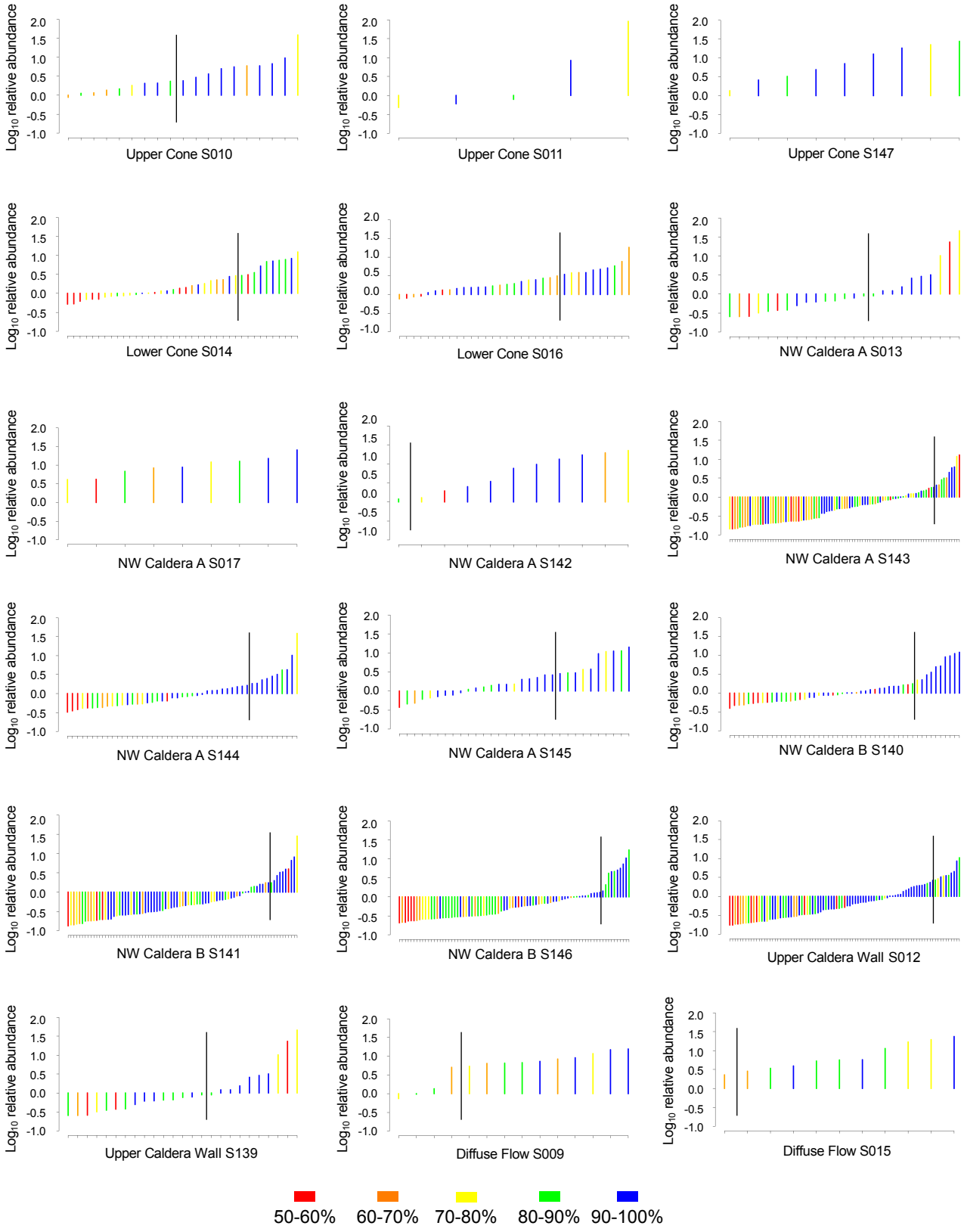
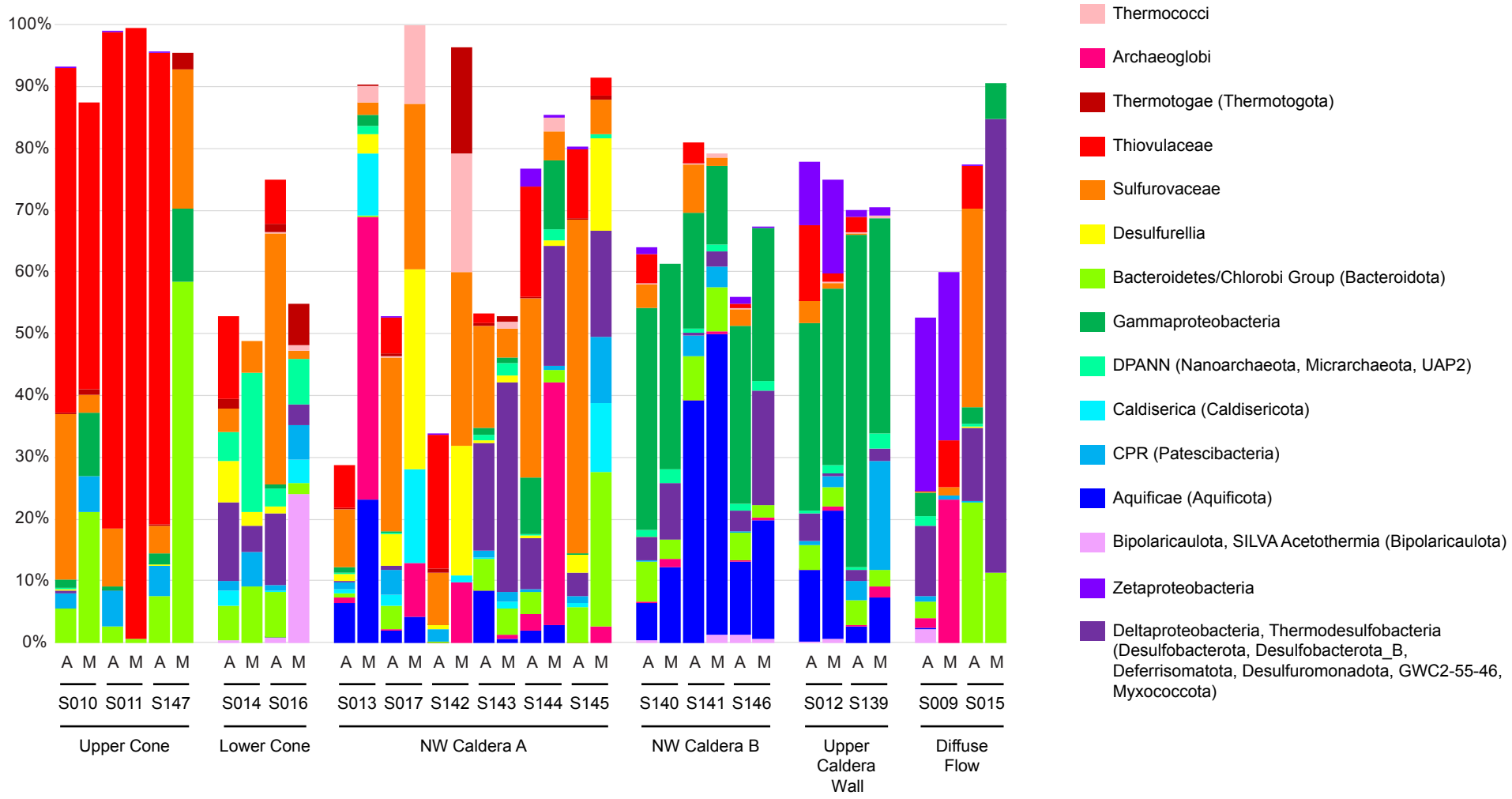


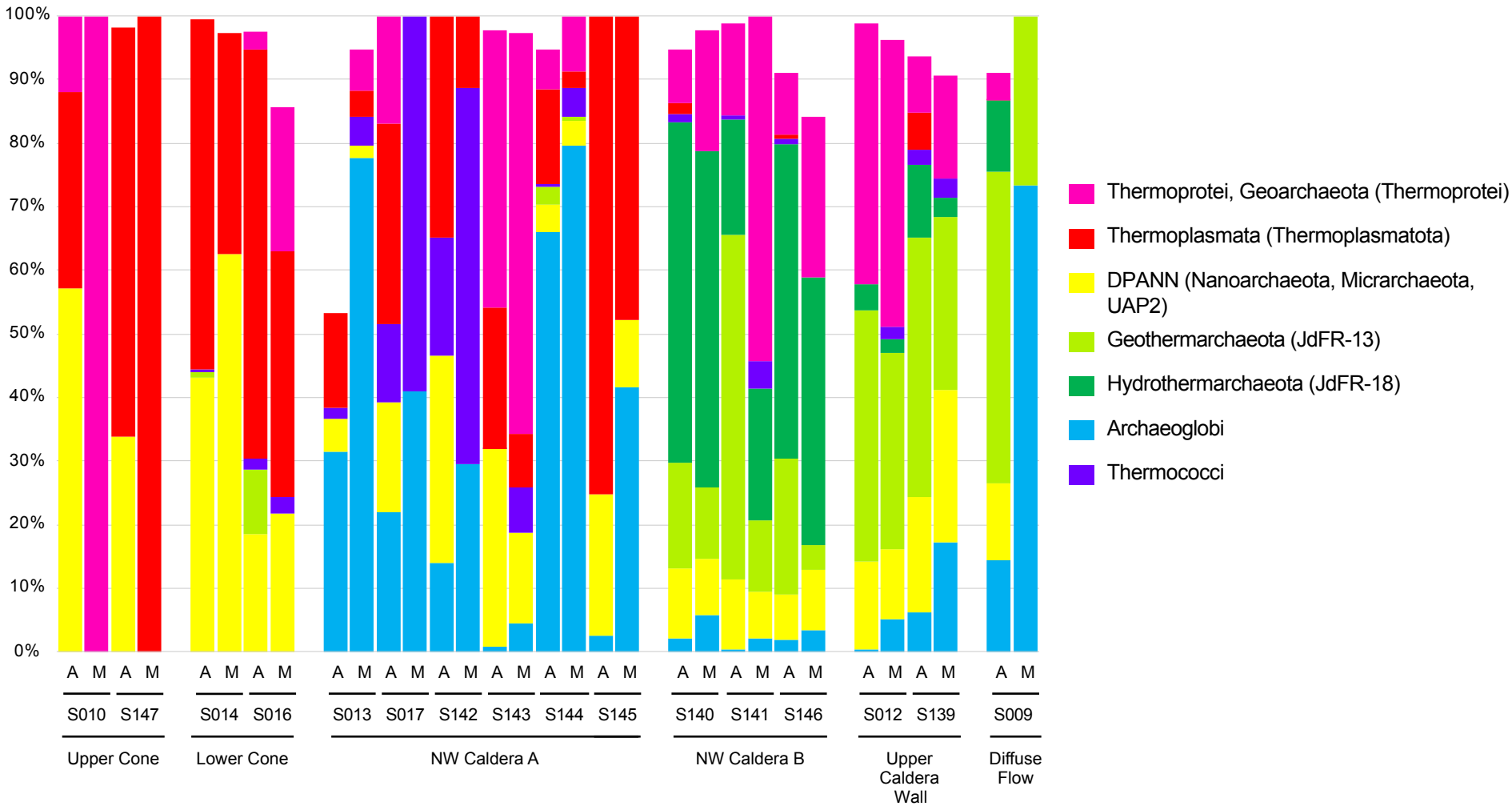
Fig. S3. Rank abundance graph of MAGs from each metagenome. Each MAG (≥50% complete, ≤10% contaminated, six or more of the 16 ribosomal proteins) is shown as a single colored line, with line color corresponding to estimated MAG completeness. MAGs are shown in order of increasing log transformed relative abundance. Black lines delineate the top ten most abundant MAGs in each sample where appropriate.

Fig. S4. Archaeal GTDB-Tk maximum-likelihood protein tree. Branch support values are shown (0.8-1.0) with black circles.
Estimated completion, contamination, and relative abundance statistics are indicated, and the scale bar shows expected amino acid substitutions.
* Curated with ESOM
**Included, although ESOM curation resulted in the loss of most ribosomal proteins

A


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M A M
S010 S011 S147 S014 S016 S013 S017 S142 S143 S144 S145 S140 S141 S146 S012 S139 S009 S015
Upper Cone
Lower Cone
NW Caldera A
NW Caldera B
Upper Caldera Wall
Diffuse Flow
Thermococci
Archaeoglobi
Thermotogae (Thermotogota)
Thiovulaceae
Sulfurovaceae
Desulfurellia
Bacteroidetes/Chlorobi Group (Bacteroidota)
Gammaproteobacteria
DPANN (Nanoarchaeota, Micrarchaeota, UAP2)
Caldiserica (Caldisericota)
CPR (Patescibacteria)
Aquificae (Aquificota)
Bipolaricaulota, SILVA Acetothermia (Bipolaricaulota)
Zetaproteobacteria
Deltaproteobacteria, Thermodesulfobacteria (Desulfobacterota, Desulfobacterota_B, Deferrisomatota, Desulfuromonadota, GWC2-55-46, Myxococcota)

B


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
A M A M
S010 S147
Upper Cone
A M A M
S014 S016
Lower Cone
A M A M A M A M A M A M
S013 S017 S142 S143 S144 S145
NW Caldera A
A M A M A M
S140 S141 S146
NW Caldera B
A M A M
S012 S139
Upper Caldera Wall
A M
S009
Diffuse Flow
Thermoprotei, Geoarchaeota (Thermoprotei)
Thermoplasmata (Thermoplasmatota)
DPANN (Nanoarchaeota, Micrarchaeota, UAP2)
Geothermarchaeota (JdFR-13)
Hydrothermarchaeota (JdFR-18)
Archaeoglobi
Thermococci

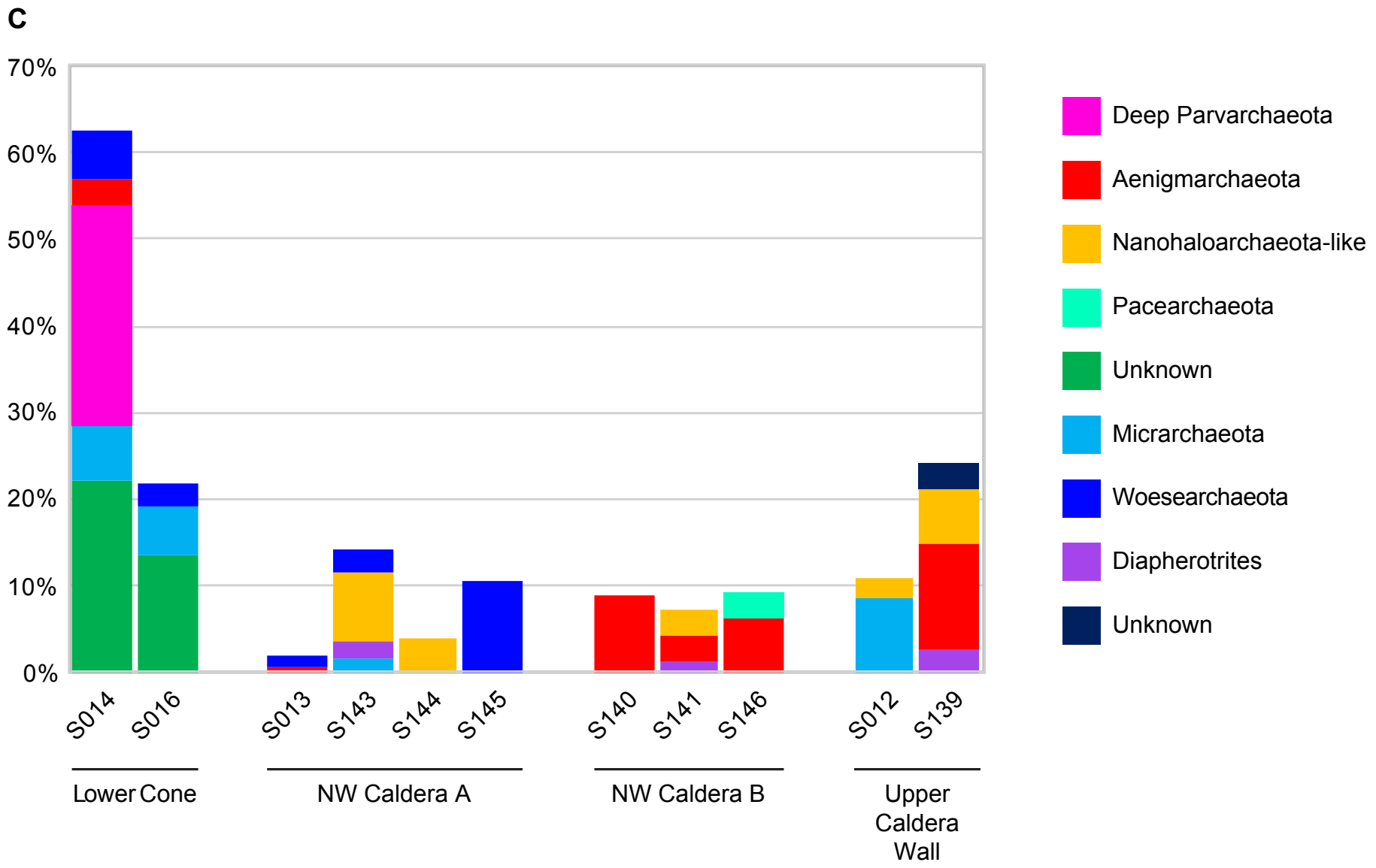


Fig. S5. Comparison of relative abundance of select MAGs (based on read coverage) and relative abundance of 16S rRNA gene amplicon OTUs. Relative abundance of clades in (*A*) the Bacteria and Archaea and (*B*) the Archaea alone are shown normalized by sample. Depicted clades were detectable in both amplicons and MAGs and represented either >15% of the total MAG read coverage or >15% of the total archaeal MAG read coverage, respectively. (*C*) Relative abundance of MAGs in the DPANN superphylum based on read coverage of archaeal MAGs. A, amplicons; M, MAGs.

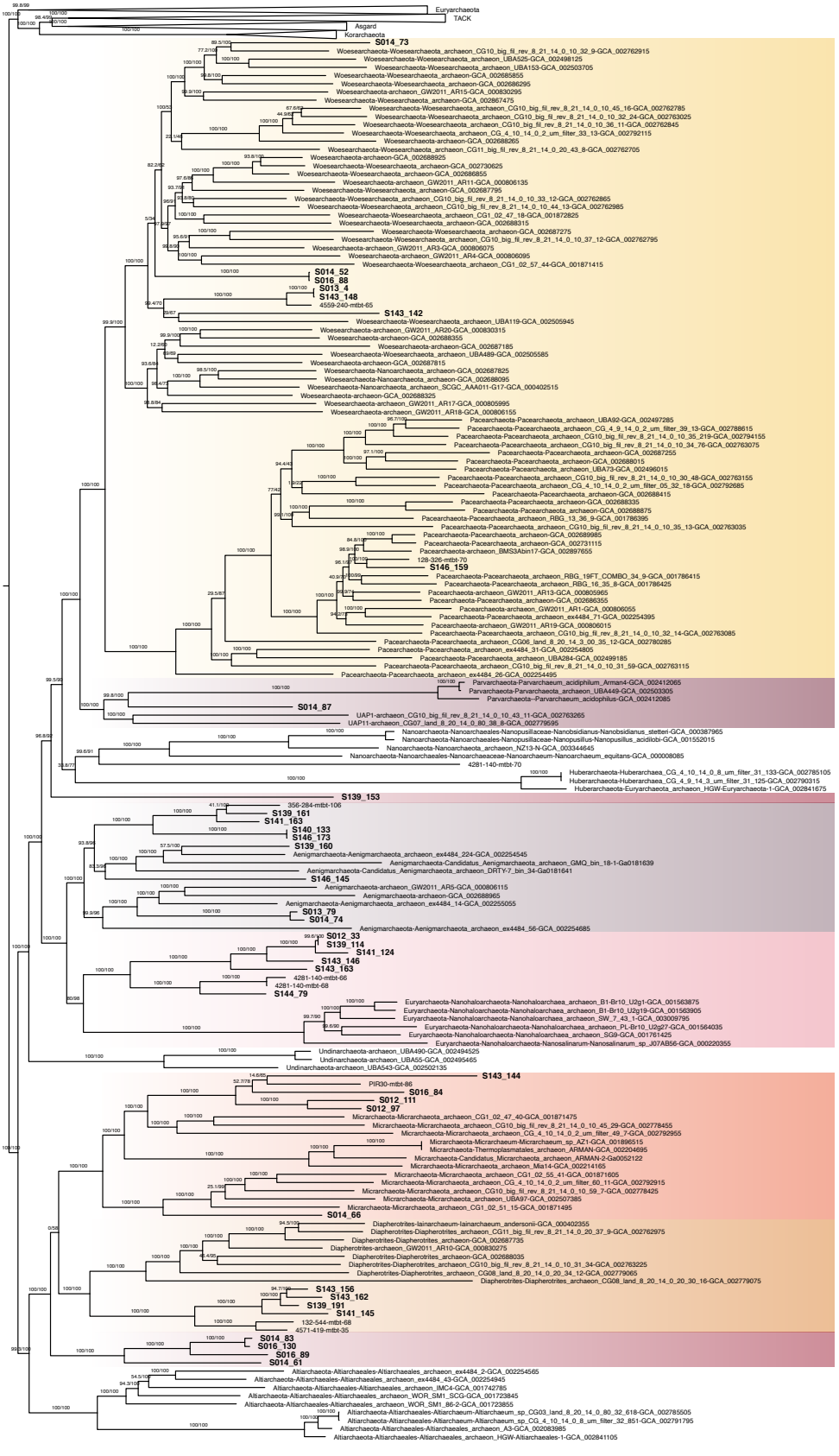


Fig. S6. Maximum-likelihood (ML) phylogenetic analysis of DPANN superphylum MAGs, based on forty-eight marker proteins and 400 archaeal species. The alignment was trimmed with BMGE (alignment length = 9,461 aa). A ML phylogenetic tree was inferred in IQ-TREE with the LG+C60+F+R model with an ultrafast bootstrap approximation (left) and SH-like approximate likelihood tests (right), each run with 1,000 replicates. The tree was artificially rooted with the DPANN archaea. Scale bar = Average number of substitutions per site.

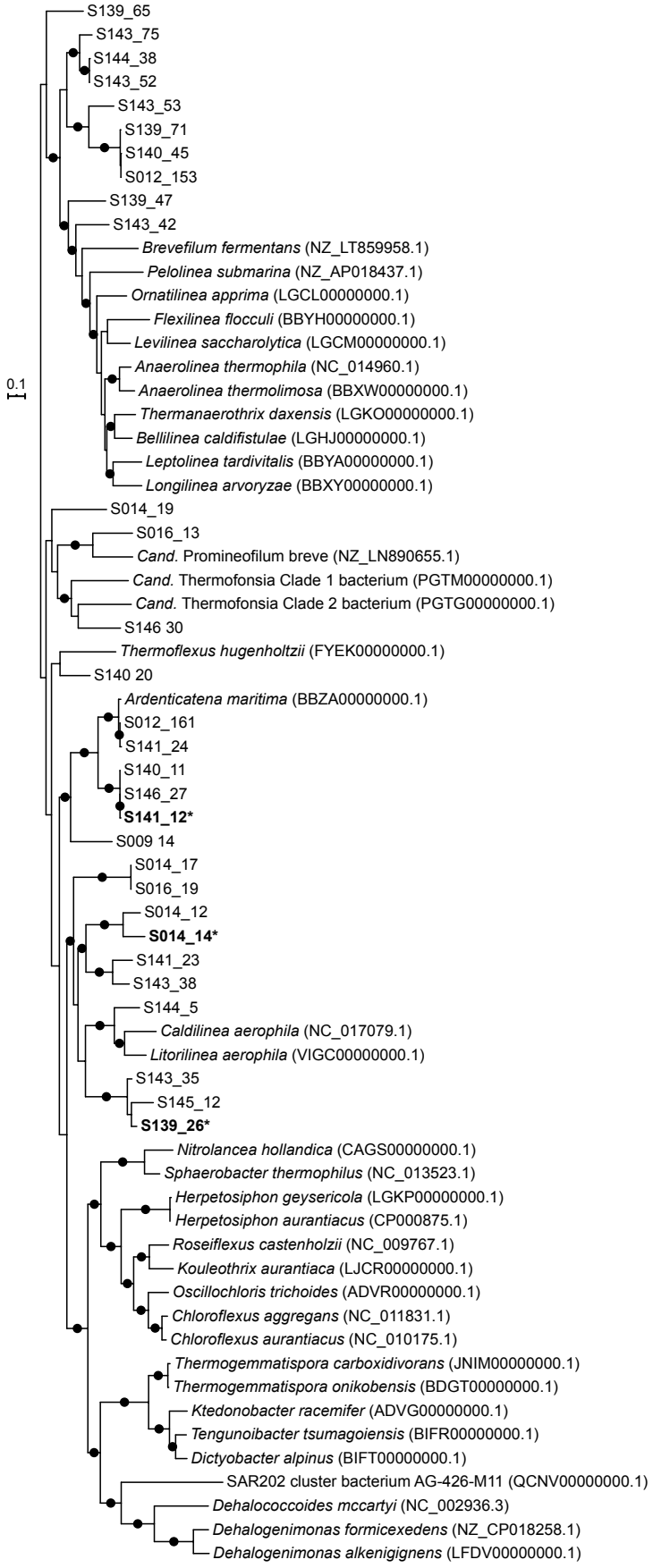


Fig. S7. Maximum-likelihood ribosomal protein tree of the Chloroflexi (Chloroflexota). The tree was constructed using 16 ribosomal proteins, inferred using RAxML with 1,000 bootstrapped replicates and rooted with *Methanocaldococcus jannaschii* (L77117.1). Bootstrap values ≥80% are indicated by black circles, and the scale bar shows expected amino acid substitutions per site.
*Curated with ESOM

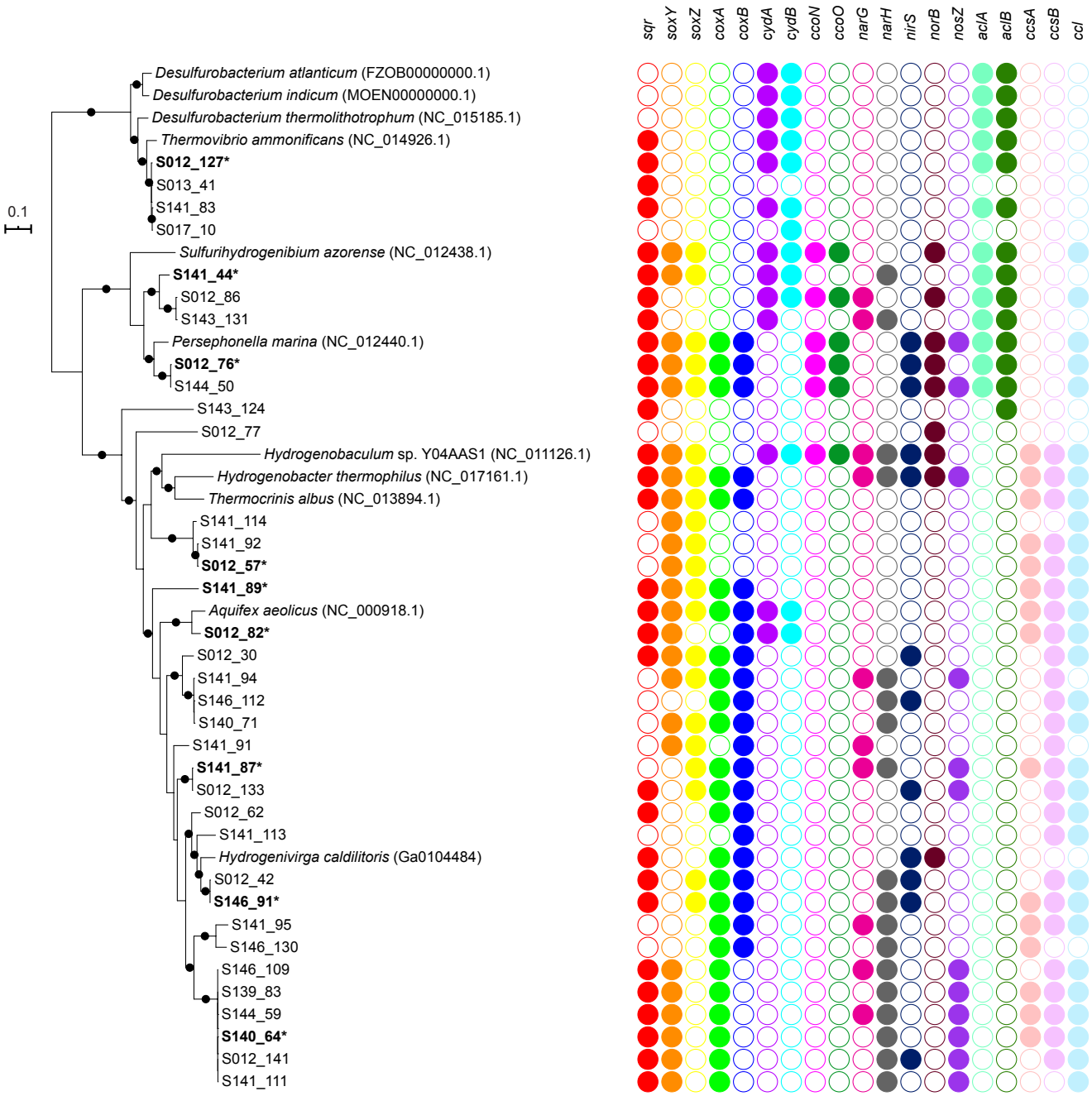


Fig. S8. Maximum-likelihood ribosomal protein tree of the Aquificae (Aquificota) constructed from 16 ribosomal proteins, using *Archaeoglobus fulgidus* (NC_000917.1) and *Methanocaldococcus jannaschii* (L77117.1) as a root. Bootstrap support based on 1,000 replicates is shown with black circles (80-100%). The presence of key genes involved in sulfur cycling, aerobic respiration, denitrification, and carbon fixation is depicted with colored circles. Full gene names and KO numbers are available on Figshare. The scale bar depicts expected amino acid substitutions per site.
* Curated with ESOM

Fig. S9. Maximum-likelihood GTDB-Tk phylogenetic tree showing the position of Woesearchaeota S21_7 within the DPANN superphylum. Branch support is shown with dark circles (0.8-1.0), and the scale bar shows expected amino acid substitutions per site.
* Curated with ESOM

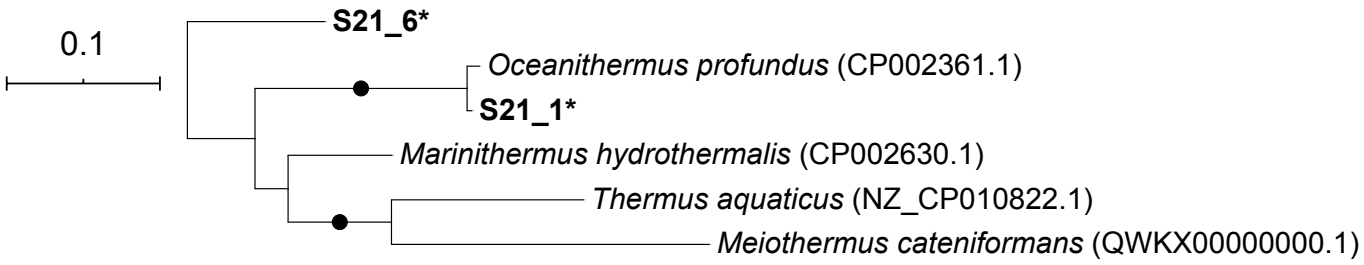


Fig. S10. Concatenated ribosomal protein tree of Thermaceae MAGs from the S21 enrichment culture metagenome, rooted with *Aquifex aeolicus* (NC_000918.1) and *Persephonella marina* (NC_012440.1). Bootstrap support (80-100%) from 1,000 bootstrapped replicates is indicated with dark circles, and the scale bar shows expected amino acid substitutions per site. Based on amplicon sequence data and a partial 16S rRNA gene sequence, the S21_6 MAG is most likely related to *Vulcanithermus medioatlanticus* (AJ507298.1; 98% similarity).
* Curated with ESOM

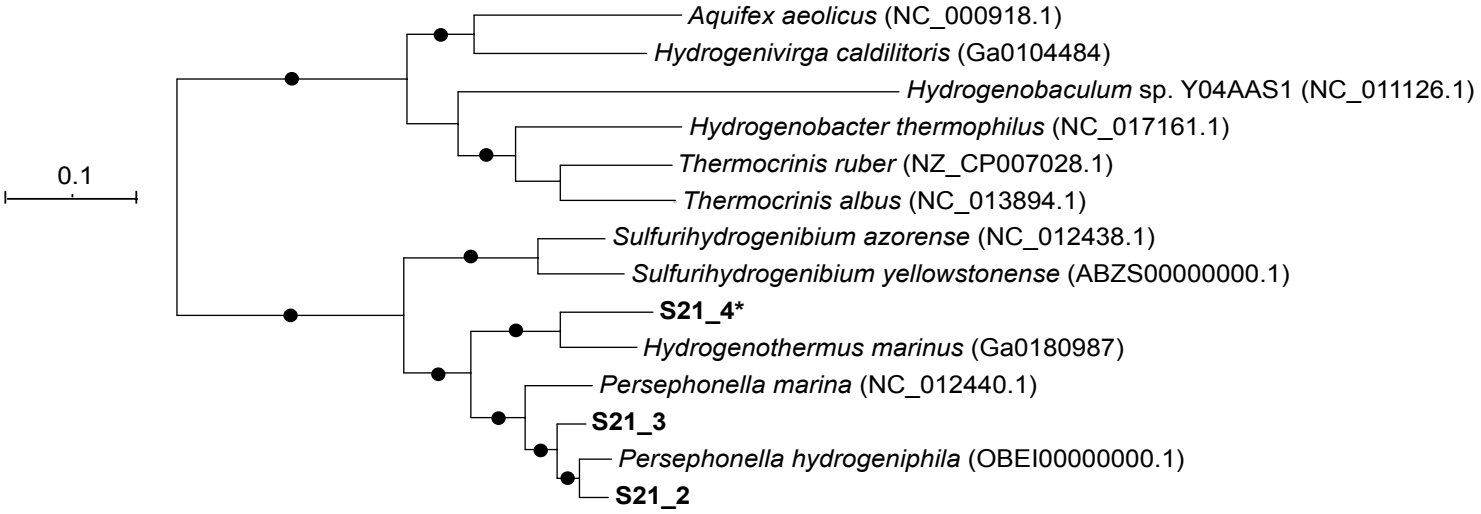


Fig. S11. Maximum-likelihood protein tree showing the S21 Hydrogenothermaceae MAGs, constructed from 16 ribosomal proteins. The tree was rooted with *Thermotoga profunda* (NZ_AP014510.1) and *Marinitoga hydrogenitolerans* (FQUI00000000.1) and bootstrap support based on 1,000 replicates is shown (80-100%) with black circles. The scale bar indicates expected amino acid substitutions per site.
* Curated with ESOM

List of Figshare files (https://doi.org/10.6084/m9.figshare.c.5099348)

Simper analyses

- SIMPER analyses of Brothers volcano taxonomic diversity.

Catabolic energy calculations

- Gibbs energy calculations for Brothers volcano vent sites.

Functional gene analyses

- Comparison of functional genes/gene categories by normalized read coverage and normalized gene counts.

Assembly translations and annotations

- Assembly translations.
- GhostKoala assembly annotation.
- FeGenie assembly annotation.
- HydDB assembly annotation.

MAG summary, translations, and annotations

- Number of MAGs assigned to Bacteria and Archaea expressed as a percent.
- MAG translations (by site).
- MAG translations individual, S009-S142.
- MAG translations individual, S143-S147.
- GhostKoala MAG annotation (by site).
- FeGenie MAG annotation.
- HydDB selected MAG annotation.
- Breakdown of selected MAG annotations from GhostKoala and HydDB.

Domain level phylogenetic tree

- GTDB-Tk Bacteria tree (phyloXML format).

Additional methods files

- QIIME2 workflow.
- arCOG genes used to create *SI Appendix,* Fig. S6.

Additional MAG annotations, translations, and summaries

- Aigarchaeota KAAS annotations.
- Aigarchaeota gene table.
- Altiarchaeota Prokka annotations.
- Aquificae KAAS annotations.
- Aquificae gene table.
- DPANN MAG annotations.
- DPANN reference annotations.
- Hydrothermarchaeota KAAS annotations.
- Hydrothermarchaeota gene table.
- Hydrothermarchaeota reference translations.
- Korarchaeota arCOG, KAAS annotations.
- Korarchaeota gene table.
- S21 MAG and reference translations.
- S21 Archaea arCOG and KAAS annotations.
- S21 Bacteria KAAS annotations.

Legends for Datasets S1 to S5

Dataset S1 (separate file). Metadata, geochemistry and 16S rRNA gene amplicon analyses. (*A*) Brothers volcano sample names, locations, BioProject and BioSample accession numbers, and Sequence Read Archive (SRA) accession numbers for 16S rRNA gene amplicon data. (*B*) Geochemistry of the hydrothermal fluids associated with the Brothers volcano deposits. (*C*) Normalized OTUs identified at Brothers volcano using 16S rRNA gene amplicon sequencing. (*D*) Subset of SIMPER results showing differentiating taxa discussed in the text, both at the class and the order level. Unabridged SIMPER outputs are available on FigShare at https://doi.org/10.6084/m9.figshare.c.5099348. (*E*) BioProject, BioSample and SRA accession numbers for 16S rRNA gene amplicon datasets from the Eastern Lau Spreading Center-Valu Fa Ridge. Datasheet abbreviations: Dataset_S1A_Metadata, Dataset_S1B_Geochemistry, Dataset_S1C_Normalized_OTUs, Dataset_S1D_SIMPER, Dataset_S1E_ELSC-VFR_accessions.

Dataset S2 (separate file). Statistics and relative normalized abundance of functional genes in metagenome assemblies. (*A*) Metagenome assembly statistics. (*B*) Relative normalized abundance of a subset of metabolic genes and iron metabolism gene categories identified in Brothers volcano contigs. The abundance was calculated by dividing the summed coverage for each gene/gene category in the assembly by the average summed coverage for 14 single copy marker genes, as described in *Materials and Methods*. Datasheet abbreviations: Dataset_S2A_Assembly_statistics, Dataset_S2B_Normalized_genes.

Dataset S3 (separate file). MAG quality, taxonomic assignments, accession numbers, AAI matrices, and relative abundance of both bacterial and archaeal MAGs and archaeal MAGs only, based on read coverage. (*A*) Number of MAGs recovered from each assembly utilizing various completion and contamination thresholds. (*B*) MAG statistics, Genbank accession

numbers, and taxonomic assignments based on GTDB-Tk and NCBI taxonomy. Clade designations for DPANN superphylum MAGs (depicted in *SI Appendix,* Fig. S6) are shown in parentheses following NCBI taxonomy where needed for clarity. MAGs curated using ESOM are indicated, and changes to MAG statistics following ESOM curation are shown. (*C*) AAI matrices of bacterial MAGs and archaeal MAGs with reference genomes. Clades are shown using NCBI taxonomy, with GTDB-Tk classifications in parentheses. (*D*-*E*) Relative abundance (using normalized read coverage) of GTDB-Tk taxa by site for (*D*) all MAG taxa and (*E*) Archaea only. Datasheet abbreviations: Dataset_S3A_MAG_quality, Dataset_S3B_Taxonomy_accessions, Dataset_S3C_AAI, Dataset_S3D_Relative_abund, Dataset_S3E_Arch_Relativ_abund.

Dataset S4 (separate file). Select metabolic genes detected in DPANN superphylum MAGs and reference genomes based on KO, arCOG and dbCAN annotations. Hydrogenases were first detected using KO annotations and confirmed using the HydDB online database.

Dataset S5 (separate file). Assembly and binning statistics, genome size and quality comparisons, and select metabolic genes detected in S21 enrichment culture MAGs and a Woesearchaeota reference genome. (*A*) Assembly and binning statistics of the S21 enrichment culture metagenome. For comparison, estimated quality and size of Woesearchaeota reference genomes are also shown. (*B*-*C*) Selected metabolic genes detected using the KEGG Automatic Annotation Server in (*B*) S21_7 and Woesearchaeota UBA119 and (*C*) bacterial S21 MAGs (≥50% completion, ≤5% contamination). Gene calls were confirmed using Conserved Domain Database (CDD) where indicated. Datasheet abbreviations: Dataset_S5A_Statistics_quality, Dataset_S5B_Archaea_genes, Dataset_S5C_Bacteria_genes.

SI References

1. D. W. Waite, *et al.*, Comparative genomic analysis of the class Epsilonproteobacteria and proposed reclassification to Epsilonbacteraeota (phyl. nov.). *Front. Microbiol.* **8**, 682 (2017).
2. L. A. Hug, *et al.*, A new view of the tree of life. *Nat. Microbiol.* **1**, 16048 (2016).
3. R. L. Morris, T. M. Schmidt, Shallow breathing: Bacterial life at low O_2. *Nat. Rev. Microbiol.* **11**, 205–212 (2013).
4. N. Dombrowski, J.-H. Lee, T. A. Williams, P. Offre, A. Spang, Genomic diversity, lifestyles and evolutionary origins of DPANN Archaea. *FEMS Microbiol. Lett.* **366**, fnz008 (2019).
5. E. St. John, *et al.*, A new symbiotic nanoarchaeote (*Candidatus* Nanoclepta minutus) and its host (*Zestosphaera tikiterensis* gen. nov., sp. nov.) from a New Zealand hot spring. *Syst. Appl. Microbiol.* **42**, 94–106 (2019).
6. K. S. Makarova, E. V. Koonin, S.-V. Albers, Diversity and evolution of type IV pili systems in Archaea. *Front. Microbiol.* **7**, 667 (2016).
7. O. V. Golyshina, *et al.*, Diversity of "*Ca*. Micrarchaeota" in two distinct types of acidic environments and their associations with Thermoplasmatales. *Genes (Basel).* **10**, 461 (2019).
8. S. Krause, A. Bremges, P. C. Münch, A. C. McHardy, J. Gescher, Characterisation of a stable laboratory co-culture of acidophilic nanoorganisms. *Sci. Rep.* **7**, 3289 (2017).
9. J. N. Hamm, *et al.*, Unexpected host dependency of Antarctic Nanohaloarchaeota. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 14661–14670 (2019).
10. K. Schwank, *et al.*, An archaeal symbiont-host association from the deep terrestrial subsurface. *ISME J.* **13**, 2135–2139 (2019).
11. X. Liu, *et al.*, Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. *Microbiome* **6**, 102 (2018).
12. Z.-S. Hua, *et al.*, Genomic inference of the metabolism and evolution of the archaeal

phylum Aigarchaeota. *Nat. Commun.* **9**, 2832 (2018).

13. S. A. Carr, *et al.*, Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. *ISME J.* **13**, 1457–1468 (2019).
14. S. Kato, *et al.*, Metabolic potential of as-yet-uncultured archaeal lineages of *Candidatus* Hydrothermarchaeota thriving in deep-sea metal sulfide deposits. *Microbes Environ.* **34**, 293–303 (2019).
15. C. Rudolph, G. Wanner, R. Huber, Natural communities of novel Archaea and Bacteria growing in cold sulfurous springs with a string-of-pearls-like morphology. *Appl. Environ. Microbiol.* **67**, 2336–2344 (2001).
16. J. T. Bird, B. J. Baker, A. J. Probst, M. Podar, K. G. Lloyd, Culture independent genomic comparisons reveal environmental adaptations for Altiarchaeales. *Front. Microbiol.* **7**, 1221 (2016).
17. C. Moissl, R. Rachel, A. Briegel, H. Engelhardt, R. Huber, The unique structure of archaeal 'hami', highly complex cell appendages with nano-grappling hooks. *Mol. Microbiol.* **56**, 361–370 (2005).
18. T. Seemann, Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
19. K. S. Makarova, Y. I. Wolf, E. V Koonin, Archaeal clusters of orthologous genes (arCOGs): An update and application for analysis of shared features between Thermococcales, Methanococcales, and Methanobacteriales. *Life (Basel)* **5**, 818–840 (2015).
20. M. Kanehisa, S. Goto, KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
21. L. Huang, *et al.*, DbCAN-seq: A database of carbohydrate-active enzyme (CAZyme) sequence and annotation. *Nucleic Acids Res.* **46**, D516–D521 (2018).
22. D. Søndergaard, C. N. S. Pedersen, C. Greening, HydDB: A web tool for hydrogenase

classification and analysis. *Sci. Rep.* **6**, 34212 (2016).

23. Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, M. Kanehisa, KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).

24. A. Marchler-Bauer, *et al.*, CDD: A conserved domain database for protein classification. *Nucleic Acids Res.* **33**, D192–D196 (2005).

25. P. Carini, L. Steindler, S. Beszteri, S. J. Giovannoni, Nutrient requirements for growth of the extreme oligotroph '*Candidatus* Pelagibacter ubique' HTCC1062 on a defined medium. *ISME J.* **7**, 592–602 (2013).