





ARTICLE



Diverse ecophysiological adaptations of subsurface Thaumarchaeota in floodplain sediments revealed through genome-resolved metagenomics

Linta Reji ^{1,3}, Emily L. Cardarelli ^{1,4}, Kristin Boye², John R. Bargar ² and Christopher A. Francis ¹✉

© The Author(s), under exclusive licence to International Society for Microbial Ecology 2021

The terrestrial subsurface microbiome contains vastly underexplored phylogenetic diversity and metabolic novelty, with critical implications for global biogeochemical cycling. Among the key microbial inhabitants of subsurface soils and sediments are Thaumarchaeota, an archaeal phylum that encompasses ammonia-oxidizing archaea (AOA) as well as non-ammonia-oxidizing basal lineages. Thaumarchaeal ecology in terrestrial systems has been extensively characterized, particularly in the case of AOA. However, there is little knowledge on the diversity and ecophysiology of Thaumarchaeota in deeper soils, as most lineages, particularly basal groups, remain uncultivated and underexplored. Here we use genome-resolved metagenomics to examine the phylogenetic and metabolic diversity of Thaumarchaeota along a 234 cm depth profile of hydrologically variable riparian floodplain sediments in the Wind River Basin near Riverton, Wyoming. Phylogenomic analysis of the metagenome-assembled genomes (MAGs) indicates a shift in AOA population structure from the dominance of the terrestrial *Nitrososphaerales* lineage in the well-drained top ~100 cm of the profile to the typically marine *Nitrosopumilales* in deeper, moister, more energy-limited sediment layers. We also describe two deeply rooting non-AOA MAGs with numerous unexpected metabolic features, including the reductive acetyl-CoA (Wood-Ljungdahl) pathway, tetrathionate respiration, a form III RuBisCO, and the potential for extracellular electron transfer. These MAGs also harbor tungsten-containing aldehyde:ferredoxin oxidoreductase, group 4f [NiFe]-hydrogenases and a canonical heme catalase, typically not found in Thaumarchaeota. Our results suggest that hydrological variables, particularly proximity to the water table, impart a strong control on the ecophysiology of Thaumarchaeota in alluvial sediments.

The ISME Journal (2022) 16:1140–1152; <https://doi.org/10.1038/s41396-021-01167-7>

INTRODUCTION

The terrestrial subsurface is a major reservoir of phylogenetically and metabolically diverse bacteria and archaea that remain largely uncultivated. High-throughput culture-independent sequencing approaches in recent years have offered a window into the genetic diversity and metabolic novelty of subsurface microbes e.g., [1–3]. However, much remains unknown about the ecological adaptations of subsurface communities and their biogeochemical impacts. Among the most abundant members of the subsurface community are archaea of the phylum Thaumarchaeota [2–6]. Ubiquitous across terrestrial and marine systems, Thaumarchaeota are key members of the global biogeochemical nitrogen and carbon cycles. While the ubiquity of these organisms in shallow subsurface soils (<100 cm) is increasingly being documented e.g., [2, 3, 5], a systematic inquiry into their ecophysiological adaptations vertically along the soil profile is thus far lacking.

The most well-characterized members of the phylum Thaumarchaeota are the ammonia-oxidizing archaea (AOA)—a monophyletic group of predominantly mesophilic organisms that mediate the first and rate-limiting step of nitrification—the aerobic oxidation of ammonia to nitrate via nitrite. The ecology of AOA in natural

systems has been studied extensively e.g., [7–11], primarily using the molecular marker gene *amoA*, which encodes the α -subunit of the key metabolic enzyme ammonia monooxygenase. Deeply rooted lineages of Thaumarchaeota described to date, however, are not ammonia-oxidizers [12–17]. With the exception of the recently described marine heterotrophic clade [16, 17], basal lineages of Thaumarchaeota lack the ability to respire oxygen [18]. The lack of cultivated non-AOA representatives [15] limits explorations into the ecophysiological adaptations of these lineages.

Mesophilic AOA primarily constitute three major lineages: (i) *Nitrososphaerales* comprised mostly of soil AOA [19]; (ii) *Nitrosopumilales*, predominantly found in marine systems [20], and (iii) *Candidatus Nitrosotaleales*, an acidophilic lineage typically found in soils [21]. Oxygen availability and soil pH are hypothesized to have played key roles in the habitat-associated diversification of AOA over evolutionary timescales [18, 22]. In terrestrial systems, AOA population structure has been linked to soil pH [22, 23], moisture levels [24], ammonium availability [25], and temperature [26]. Notably, however, most studies assessing AOA community structure in terrestrial environments have focused on topsoil e.g., [23, 27, 28]. The few studies that examined subsurface AOA have

¹Department of Earth System Science, Stanford University, Stanford, CA, USA. ²Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ³Present address: Department of Geosciences, Princeton University, Princeton, NJ, USA. ⁴Present address: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. ✉email: caf@stanford.edu

Received: 18 June 2021 Revised: 17 November 2021 Accepted: 26 November 2021
Published online: 6 December 2021

uncovered significant changes in AOA community structure with soil depth [29, 30], which are not captured in topsoil studies.

A recent analysis of *amoA* gene abundance and diversity along a floodplain sediment profile in the Wind River Basin near Riverton, Wyoming revealed AOA as the predominant ammonia-oxidizers in these sediments, as well as regionally across four other floodplain sites spanning a 900-km north-south transect in the intermountain western United States [5]. 16S rRNA-based community profiling at a nearby site in Riverton also reported the dominance of AOA over bacterial ammonia oxidizers in the subsurface, and their community structure appeared to shift with changing soil horizons with depth [6]. Curiously, no relationship was found between AOA community structure and any of the physicochemical parameters measured, suggesting depth-associated changes in lithography was the primary determinant of shifts in phylogenetic structure [6]. Phylogenetic analysis of the *amoA* sequences from Riverton sediments also pointed to a notable shift in AOA community structure with sediment depth, which appeared to be linked to the moisture content of the sediments [5]. In order to examine the ecophysiological adaptations of these AOA, we obtained metagenome-assembled genomes (MAGs) from subsurface sediment samples collected along a ~2 m depth profile at site KB1 near Riverton, WY. The assembled genomes include diverse AOA spanning the mesophilic order-level lineages *Nitrososphaerales* and *Nitrosopumilales*, as well as a basal lineage of Thaumarchaeota that harbors unique, previously unknown metabolic adaptations.

MATERIALS AND METHODS

Sample collection, geochemical analysis, and DNA extraction

Details of field sampling and general site characteristics are presented in Cardarelli et al. [5]. The sampling site at Riverton, WY is located on an alluvial terrace within the Wind River Basin. A 234 cm deep soil pit was carved into (first ~1 m) and dug out below (bottom ~1.5 m) the terrace wall to abstract soil samples from the top soil to the groundwater aquifer at a location (KB1) close to the former river bank (Latitude: 42° 59.322804, Longitude: -108° 23.977843) in August 2015. The water table was located at 235 cm below ground surface (BGS), and the transiently flooded capillary fringe was estimated to be located between 155 and 235 cm BGS. The sediment core consisted of dry, free draining soil (i.e., at field capacity) up to ~80 cm BGS. Sediment samples for molecular analyses were collected at discrete depths at ~10–20 cm intervals (depending on soil horization) along the length of the core, flash-frozen in liquid nitrogen, and stored at -80°C until nucleic acid extraction. Geochemical characterization was performed on freeze-dried core samples, finely ground and homogenized with a mortar and pestle. For total carbon and nitrogen, 30–60 mg of sample was weighed into tin capsules and analyzed in duplicate on a Carlo Erba NA1500 elemental analyzer. Remaining samples were analyzed by using energy-dispersive X-ray fluorescence spectroscopy (XEPOS, SPECTRO Analytical, Kleve, Germany) to measure elemental composition. Key geochemical variables measured include total carbon, nitrogen and sulfur content (Fig. S1).

Approximately 0.3 g of sediment from each sample was used for DNA extraction. To lyse cells, samples were subjected to mechanical agitation in a FastPrep bead beater (MP Biomedicals, Santa Ana, CA) for 2 cycles of 30 s at setting 5.5. Following this, DNA was extracted using the PowerSoil DNA Extraction Kit (MoBio, Carlsbad, CA) following the manufacturer's instructions.

Metagenome sequencing, assembly, and genome reconstruction

Metagenome sequence libraries were constructed and sequenced 2 × 151 bp using the NovaSeq platform (Illumina) at the DOE Joint Genome Institute. Reads were quality trimmed by using BBDuk (v38.24; ref. [31]) - reads with 4 or more "N" bases were removed, and those with an average quality score >3 and minimum length ≥ 51 bp were retained. Read-correction was performed using BFC (v.r181; ref. [32]). Reads without a mate pair were removed. Quality-filtered reads from each library were assembled individually using MEGAHIT (v1.1.3; ref. [33, 34]), using a range of k-mers ($k = 21, 33, 55, 77, 99, 127$). Contigs longer than 2000 bp were binned using MetaBAT2 (v2.12.1; ref. [35]) and MaxBin2 (v2.2.6;

ref. [36, 37]). Resulting bins were refined using the bin refinement module in metaWRAP (v1.2.2; ref. [38]) and re-assembled using metaSPAdes (v3.13.0; ref. [39]). Short contigs (<2000 bp) introduced during re-assembly were removed. CheckM (v1.0.12; ref. [40]) was used to assess bin completion and redundancy. Taxonomic classifications were obtained using the Genome Taxonomy Database Toolkit (GTDB-Tk; ref. [41]) classified against GTDB Release 05-RS95 [42, 43]. Of the 27 MAGs, 13 representing species-level clusters on a ribosomal protein phylogenomic tree were selected for read recruitment in order to estimate relative abundances of the various phylogenetic clades across the soil/sediment profile. Bowtie2 [44] was used for the read recruitment analysis, with the flags "--sensitive --no-unal". Reads were recruited against each MAG individually, as well as competitively against each MAG by using a combined index built from all 13 MAGs. Mapped reads were normalized to the size of the MAG (kb) and the metagenome size (Gb), rendering values in units of reads per kilobase of genome per gigabase of metagenome.

Functional annotations

Prodigal (v2.6.3; ref. [45]) was used to predict protein-coding genes, and initial functional annotations were obtained using Prokka (v1.12; ref. [46]). KO annotations were obtained using GhostKOALA (v2.2; ref. [47]), KAAS (v2.1; ref. [48]) and eggNOG-mapper (v2; ref. [49, 50]). SEED annotations were obtained from the online Rapid Annotation using Subsystem Technology server [51]; and annotations for genes of interest were confirmed by BLASTP [52] searches against the NCBI non-redundant protein database. TransportDB (v2.0; ref. [53]) was used to predict membrane transporters. SignalP-5.0 server was used for signal peptide prediction (<http://www.cbs.dtu.dk/services/SignalP-5.0/>) (ref. [54, 55]); and transmembrane domains were identified using TMHMM-2.0 (ref. [56, 57]).

Phylogenetic analyses

Phylogenomic analysis was carried out using a concatenated alignment of ribosomal proteins, retrieved from the MAGs and thaumarchaeal reference genomes using the phylogenomics module in Anvi'o 5 [58]. The following proteins were included in the analysis: ribosomal protein L1, L13, L14, L15e, L16, L21e, L22, L23, L26, L29, L3, L31e, L32e, L37ae, L39, L4, L44, L5e, L6, S12/S23, S11, S13, S15, S17, S17e, S19, S19e, S2, S24e, S27e, S28e, S3Ae, S7, S8, S8e, and S9. A concatenated alignment of the protein sequences was generated using MUSCLE [59]. Alignment trimming was conducted by using trimAL (-gt 0.80 -resoverlap 0.55 -seqoverlap 55; ref. [60]). The trimmed alignment was used for phylogenomic tree inference using IQ-TREE [61] with 1000 bootstrap replicates [62]. The ModelFinder [63] in IQ-TREE chose LG + F + R6 as the best substitution model.

Specific functional gene sequences were identified via BLASTP searches [52], and single-protein phylogenies were computed using FastTree [64] with 100 bootstrap replicates each, based on Clustal Omega [65] alignments of protein sequences (unless otherwise specified in the figure legends). Trees were visualized in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) and edited in Adobe Illustrator to add annotations and highlight clusters.

Pangenomic analysis of basal Thaumarchaeota

Reference genomes of Thaumarchaeota lineages basal to AOA were downloaded from NCBI and the Integrated Microbial Genomes and Microbiomes (IMG/M; ref. [66]) databases. Defined family-level clusters in the GTDB release 06-RS202 were used as guides for identifying basal lineages and associated MAGs. When multiple genomes formed a species cluster, the highest quality genome was included in the reference set. Pangenome analysis was conducted in Anvi'o 6.2 [58], using a genome storage database containing HMM and NCBI COG annotations. The pangenome was summarized using 'anvi-summarize' and features of interest were manually selected to plot a presence/absence diagram alongside a phylogenomic tree of the basal MAGs computed as described earlier.

RESULTS AND DISCUSSION

Diverse Thaumarchaeota populations along the sediment depth profile

Thaumarchaeota ranked among the top 15 phyla in the KB1 metagenomes across all depths, and ranked 7th at 155 and 175 cm depths (Fig. S2). Within the top soil layers (38, 57, 67, 86 cm depths),

88–91% of the thaumarchaeal reads were classified as *Nitrososphaerales*. In contrast, *Nitrosopumilales* accounted for 76–86% of the thaumarchaeal reads in deeper sediments below 100 cm (Fig. S2). Assembly and binning of metagenomic data from the KBI metagenomes yielded 27 medium- to high-quality genomes of Thaumarchaeota (Table 1). Note that recent versions of the GTDB Toolkit [41] classify Thaumarchaeota as the monophyletic class-level lineage *Nitrososphaeria* within the phylum Thermoproteota. To minimize confusion regarding name changes, and since the conventional Thaumarchaeota taxa definitions have not yet been formally redefined to accommodate the revised taxonomic framework in the GTDB, we continue using the conventional taxa names to refer to thaumarchaeal clades in this manuscript. Where appropriate, we point out the corresponding GTDB taxa names.

On the ribosomal protein phylogenomic tree, 25 of the thaumarchaeal MAGs clustered with previously published AOA genomes (Fig. 1). The remaining two MAGs, D197_2 and D197_116, clustered as a basal non-AOA lineage, forming a sister group to the recently described marine heterotrophic Thaumarchaeota (16,17; Fig. 1). Clustering patterns in the phylogenomic tree (Fig. 1) were largely congruent with the single-gene phylogenies inferred based on the MAG-derived *amoA* and 16S rRNA genes (Fig. S3). Surface soil metagenomes mostly yielded AOA MAGs classifying within the order *Nitrososphaerales* while MAGs assembled from the moist sediment metagenomes deeper in the profile classified within the typical marine order *Nitrosopumilales* (Fig. 1). This shift in phylogenetic structure along the profile clearly reflected the depth-related shift in AOA population structure previously observed in alluvial sediments, based on the *amoA* gene diversity [5].

Metagenomes obtained from the top 4 depths of the sediment core (38, 57, 67, and 86 cm) yielded 12 MAGs classified as *Nitrososphaerales* (GTDB family *Nitrososphaeraceae*). On the ribosomal protein phylogeny (Fig. 1), these MAGs were placed within three distinct genus-level clusters: (i) the typical terrestrial genus *Nitrososphaera* ($n = 4$); (ii) a *Nitrososphaera*-sister group ($n = 4$), currently represented by genomic data alone (family NS- β as described in ref. [67]; GTDB genus UBA10452), and (iii) a sister-lineage to *Nitrosocosmicus* ($n = 4$) within the NS- ϵ lineage [67], which could not be assigned to any of the delineated AOA genera in GTDB (Table 1). The latter two groups included several MAGs from the recently published AOA MAG collection from sediments along the River Thames, UK (68; Fig. 1), suggesting a potentially wide distribution of these under-sampled AOA clusters in terrestrial subsurface and/or alluvial environments.

In contrast to the shallow soil layers, MAGs obtained from depths below 100 cm (111, 125, 155, 175, 185, 197, and 214 cm) were all classified as *Nitrosopumilales* (GTDB family *Nitrosopumilaceae*), spanning the genera *Nitrosotenuis*, *Nitrosopumilus*, *Nitrosarchaeum* and the uncultured CSP1-1 lineage (Table 1). CSP1-1, an AOA MAG assembled from alluvial sediments in Rifle, Colorado [2], was the phylogenetically closest database representative to 7 out of the 13 *Nitrosopumilales* MAGs assembled here (Table 1, Fig. 1). In the Rifle metagenomes, CSP1-1 recruited 0.7% of all sequence reads and represented the most abundant member of the sediment community [2]. We observed relatively high abundances for the Riverton MAGs clustering with CSP1-1, as up to 0.36% of metagenomic reads were mapped to the closest representative genome D185_135, whereas the remaining AOA MAGs recruited up to 0.26% of the reads across all metagenomes (Fig. 2, Table S1 in Data Set 1). Read recruitment profiles suggested a particularly high abundance of CSP1-1 AOA within the capillary fringe (i.e., >150 cm below ground surface; Fig. 2). The CSP1-1 lineage may thus represent a pervasive and dominant AOA group within subsurface alluvial sediments.

Nitrosarchaeum [68] was identified as an abundant AOA genus at depths below 100 cm at Riverton, based on *amoA* [5] and

relative 16S rRNA gene abundances [6]. While we assembled two *Nitrosarchaeum* MAGs from the 185 and 197 cm metagenomes, the genome quality estimates were relatively lower for these compared to the other AOA MAGs (Table 1). Their abundance profile suggests some degree of habitat overlap with CSP1-1 AOA, as both lineages appeared to be numerically abundant at sediment depths experiencing transient water intrusion from the water table below (Fig. 2). However, unlike CSP1-1, the *Nitrosarchaeum* MAG recruited notably fewer reads from metagenomes from 125 cm and above, suggesting different environmental controls on their distributions (Fig. 2).

The sole *Nitrosopumilus* MAG assembled in our dataset was found to be abundant only at 214 cm (Fig. 2). This MAG recruited 0.035% of the reads from 214 cm, and was the only AOA MAG recruiting >0.01% of reads from this depth. All three MAGs classified as *Nitrosarchaeum* or *Nitrosopumilus* clustered within a clade of representative AOA isolated or enriched from marine, estuarine or rhizosphere sediments (Fig. 1). The *Nitrosopumilus* MAG D214_93 shared 81.7% average nucleotide identity (ANI) with *Ca. Nitrosopumilus sediminus* AR2 enriched from Arctic sediments [69]. Similarly, the two *Nitrosarchaeum* MAGs (D185_51 and D197_10) were most closely related to *Ca. Nitrosarchaeum koreense* MY1 isolated from rhizosphere sediments [70], sharing 87.2% and 87.6% ANI with MY1, respectively.

Dominant AOA groups in mesophilic terrestrial systems typically affiliate within the genera *Nitrososphaera*, *Nitrosocosmicus*, and *Nitrosotalea* e.g., [71–73]. While the majority of *Nitrosopumilus* AOA have been identified in marine systems [74–77], they are also found in soil/sedimentary environments, particularly in aquifer-associated sediment layers [2, 3, 5, 6, 30, 78, 79]. Groundwater environments, in particular, have been found to host high abundances of *Nitrosarchaeum* and *Nitrosopumilus* AOA [29, 30]. These observations suggest that soil moisture may be an important control not only on the relative abundances of AOA in soils as described before [24], but also on their population structure. Increasing energy limitation with depth, as evidenced by decreasing amounts of total carbon and relatively higher carbon to nitrogen ratios (C:N) within the moist sediment layers (111–155 cm BGS; Fig. 2A), might also explain the AOA taxonomic shift observed in the KBI profile from the dominance of relatively generalist (i.e., greater number of metabolic adaptations as explained below) lineages in surface soils to more oligotrophic groups (i.e., with more streamlined genomes similar to marine AOA) in deeper sediments (Fig. 2B). Such a shift in phylogenetic structure with depth is likely pervasive in alluvial sediments, as similar patterns were recovered across multiple floodplain sites in the western United States, based on *amoA* gene diversity [5]. The depth-differentiation was also evident in a non-metric multi-dimensional scaling analysis of the MAG relative abundances (Fig. S4). The following taxa were identified as significantly correlating to the sample clustering in the nMDS space: *Nitrosarchaeum*, *Nitrosopumilus*, *Nitrososphaera*, unclassified genus of the NS- ϵ family, UBA0452, and CSP1-1. *Nitrosopumilales* (including CSP1-1) were particularly dominant within the capillary fringe (>155 cm), where both total carbon and nitrogen were generally lower (Fig. 2B). The elevated concentrations of sulfur below 111 cm (Fig. S1) reflect the influence of sulfate-rich groundwater plumes at the Riverton site [6], where sulfur has also been shown to be strongly correlated to salinity [5].

Notable functional features of AOA MAGs

While many of the AOA MAGs lacked relatives among the cultured reference AOA (Fig. 1), they often clustered with the MAGs described in Sheridan et al. [80]. The core genomic features of the *Nitrososphaera* MAGs appeared consistent with previously described AOA genomes. Genomic potential for ammonia oxidation was confirmed in all AOA clusters (determined by the presence of at least one of the three ammonia monooxygenase

Table 1. Metagenome-assembled genome (MAG) statistics.

MAG ID	Phylum	Class	Order	Family	Genus	Species	Conventional taxonomy - Phylum	Conventional taxonomy - Order	Completion (%)	Redundancy (%)
D38_14	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	96.1	2.9
D38_38	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	ND	Thaumarchaeota	Nitrososphaerales	79.3	4.4
D38_49	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	74.8	1.0
D57_19	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	66.0	1.8
D57_21	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	ND	Thaumarchaeota	Nitrososphaerales	94.7	1.0
D57_22	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	97.7	2.9
D67_19	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	UBA10452	ND	Thaumarchaeota	Nitrososphaerales	59.9	3.9
D67_58	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	96.1	2.9
D86_17	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	ND	Thaumarchaeota	Nitrososphaerales	83.7	1.0
D86_19	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	UBA10452	ND	Thaumarchaeota	Nitrososphaerales	75.8	1.0
D86_48	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	ND	ND	Thaumarchaeota	Nitrososphaerales	97.1	1.9
D111_44	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	CSP1-1 sp001443365	Thaumarchaeota	Nitrosopumilales	95.7	1.0
D111_58	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	ND	ND	Thaumarchaeota	Nitrosopumilales	89.8	0.0
D125_37	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotenuis	ND	Thaumarchaeota	Nitrosopumilales	93.2	5.3
D125_54	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	ND	Thaumarchaeota	Nitrosopumilales	75.0	2.9
D155_102	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotenuis	ND	Thaumarchaeota	Nitrosopumilales	97.9	1.0
D155_17	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrososphaeraceae	Nitrososphaera	ND	Thaumarchaeota	Nitrososphaerales	97.1	1.0
D155_85	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	CSP1-1 sp001443365	Thaumarchaeota	Nitrosopumilales	92.5	1.9
D175_134	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	ND	Thaumarchaeota	Nitrosopumilales	61.0	2.9
D175_87	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosotenuis	ND	Thaumarchaeota	Nitrosopumilales	77.4	1.0
D185_106	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	ND	Thaumarchaeota	Nitrosopumilales	84.5	0.5
D185_135	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	CSP1-1	CSP1-1 sp001443365	Thaumarchaeota	Nitrosopumilales	94.0	0.3
D185_51	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosarchaeum	ND	Thaumarchaeota	Nitrosopumilales	57.2	1.3
D197_10	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosarchaeum	ND	Thaumarchaeota	Nitrosopumilales	74.6	2.8
D197_116	Thermoproteota	Nitrososphaeria	Nitrososphaerales	ND	ND	ND	Thaumarchaeota	ND	98.5	1.0
D197_2	Thermoproteota	Nitrososphaeria	Nitrososphaerales	ND	ND	ND	Thaumarchaeota	ND	68.1	1.0
D214_93	Thermoproteota	Nitrososphaeria	Nitrososphaerales	Nitrosopumilaceae	Nitrosopumilus	ND	Thaumarchaeota	Nitrosopumilales	96.1	1.0

Taxonomic classifications were obtained by using the GTDB toolkit, classified against release 05-RS95. Unclassified Family/Genus/Species are indicated as "ND," not determined. Conventionally used names for the phyla and order lineages are also provided for comparison.

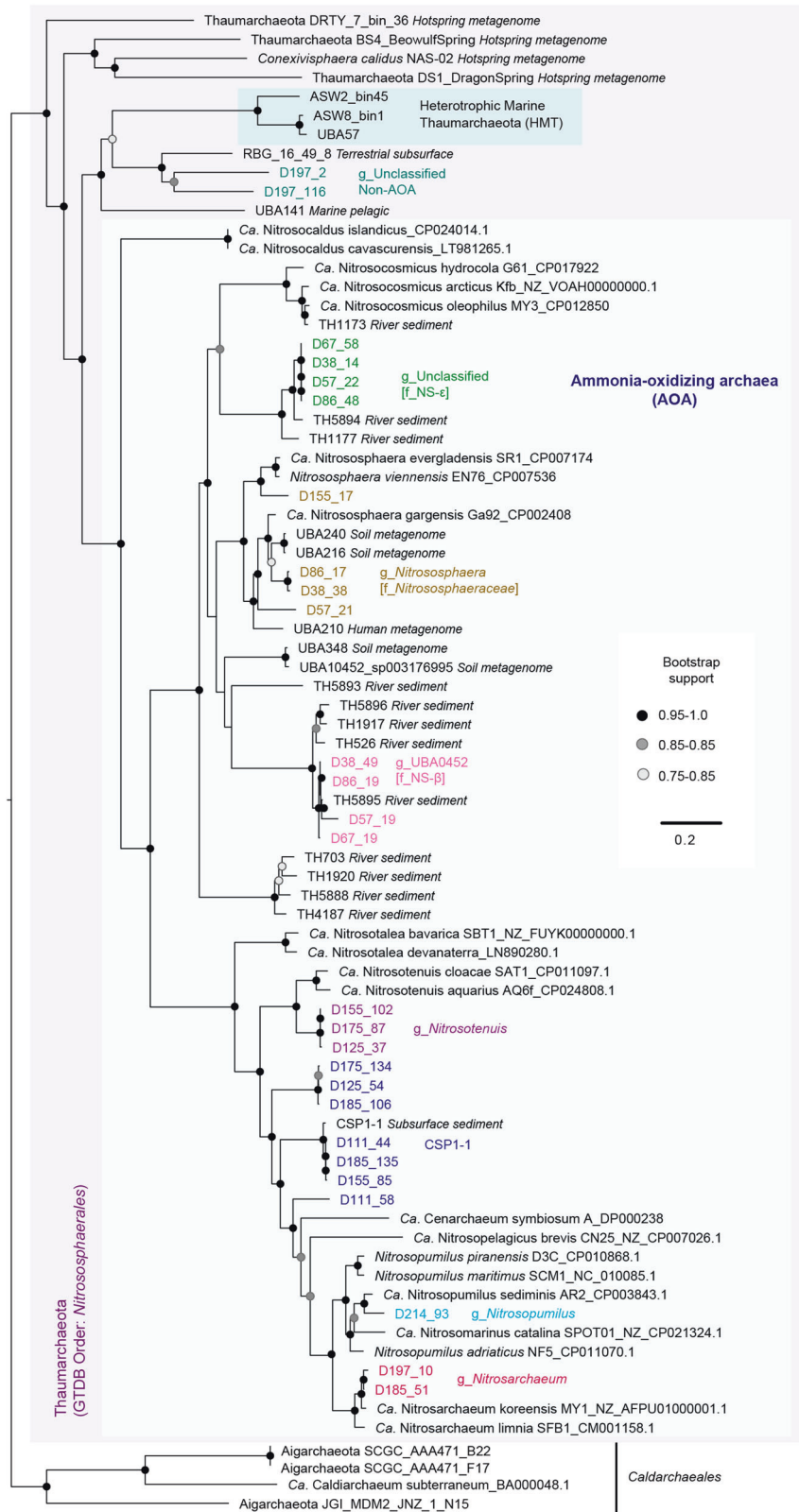


Fig. 1 Maximum-likelihood phylogenomic tree inferred using a concatenated alignment of select ribosomal proteins. See Methods for details on tree inference. Genus names listed next to each cluster (prefixed by “g_”) correspond to the classification obtained via the GTDB toolkit. In square brackets next to the genus names are the corresponding conventionally used thaumarchaeal clade names (family names prefixed by “f_”). For MAGs used as reference genomes, the habitat where the corresponding metagenomes originated have been indicated in italics. The naming convention used for naming the MAGs assembled in this study is as follows: “D[Depth cm]_[MAG #]”. For example, “D197_2” is the MAG #2 obtained from the 197 cm metagenome.

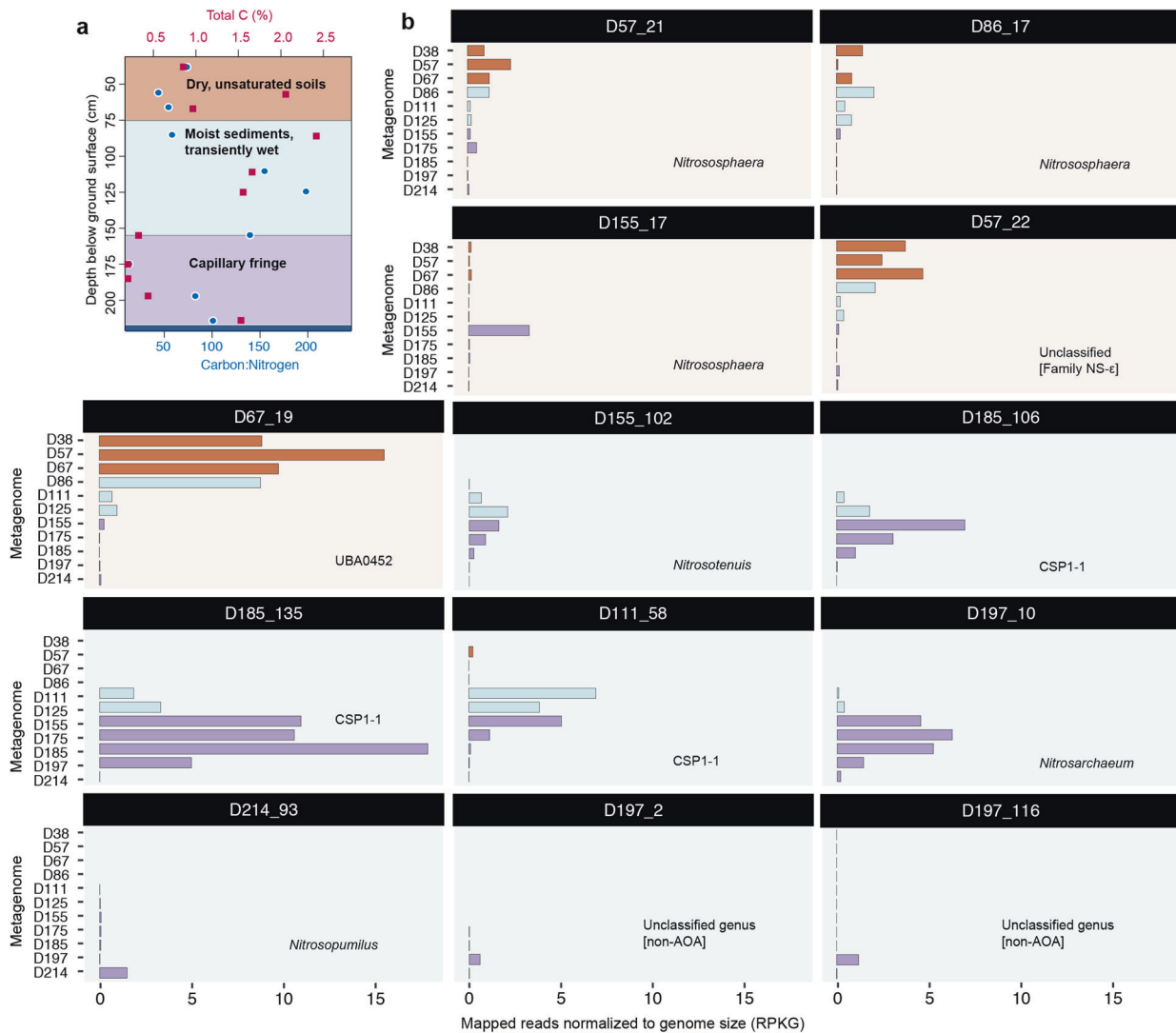


Fig. 2 Differential abundances of Thaumarchaeota lineages along the hydrological gradient. **a** Schematic representation of the sediment hydrology at KB1, along with measured values of total carbon (%) and carbon to nitrogen (C:N) ratios. The water table was located at 235 cm below ground surface, and the sediments were completely dry (at field capacity) up to ~ 80 cm. Total carbon and nitrogen were determined to be zero at 175 cm (the blue and red dots are overlapping). At 185 cm, total carbon was determined to be 0.05 % while total nitrogen was again zero, hence the missing blue dot. **b** Results of read recruitment analysis. The highest quality MAG (i.e., highest percent completion, lowest redundancy) in each major phylogenetic cluster as presented in Fig. 1 was chosen for read recruitment analysis. Panels correspond to individual MAGs to which metagenomic reads were mapped. The panel shading indicates the AOA order level lineage (orange: *Nitrososphaerales* and blue: *Nitrosopumilales*); and bar colors indicate depth layers as indicated in **a**. Abundances are expressed as the number of reads mapped per kilobase of genome per gigabase of metagenome (RPKG). Genus names obtained via the GTDB Toolkit are noted for each MAG.

gene subunits; Figs. 3, S2). 4-hydroxybutyryl-CoA dehydratase, the key enzyme of the AOA-specific CO_2 fixation pathway [81], is also ubiquitously encoded by all AOA MAGs in this study. Additionally, many of the AOA lineages found in Riverton sediments may be capable of utilizing nitrogenous organic compounds to supplement energy generation (Fig. 3). For instance, the capacity for urea hydrolysis appears to be pervasive across the subsurface lineages, as most MAGs (23 out of the 25 AOA) harbor urease subunits (Fig. 3). Many terrestrial and marine AOA are hypothesized to supplement their energy metabolism via urea hydrolysis to ammonia [69, 75, 77, 82–85]. Near-stoichiometric growth on urea has been observed for AOA strains in the mesophilic genera *Nitrosopumilus* [75, 77] and *Nitrososphaera* [86], as well as in the thermophilic genus *Nitrosocaldus* [85]. Notably, urease genes were not detected in the *Nitrosopumilus* and *Nitrosarchaeum* MAGs recovered here (Fig. 3).

Additional reduced nitrogen compounds contributing to AOA metabolism in the Riverton sediments may possibly include cyanate and nitriles. Indeed several of the *Nitrososphaera* MAGs contain homologs of cyanate and nitrile hydratases (Fig. 3). Cyanate hydratase (cyanase) catalyzes the formation of NH_3 and CO_2 from cyanate and bicarbonate [87], and can enable growth on cyanate as a nitrogen source as demonstrated for the soil thaumarchaeon *Ca. Nitrososphaera gargensis* [88]. Moreover, isotope incorporation experiments suggest that marine AOA may assimilate cyanate-derived nitrogen, despite missing the cyanase homolog in their genomes [89]. All three cyanase-encoding MAGs from the Riverton metagenomes clustered within the genus *Nitrososphaera* (Fig. 3), corroborating the limited phylogenetic distribution of this gene among AOA. Nitriles are another group of N-containing organic compounds that could potentially serve as carbon and nitrogen sources for organisms

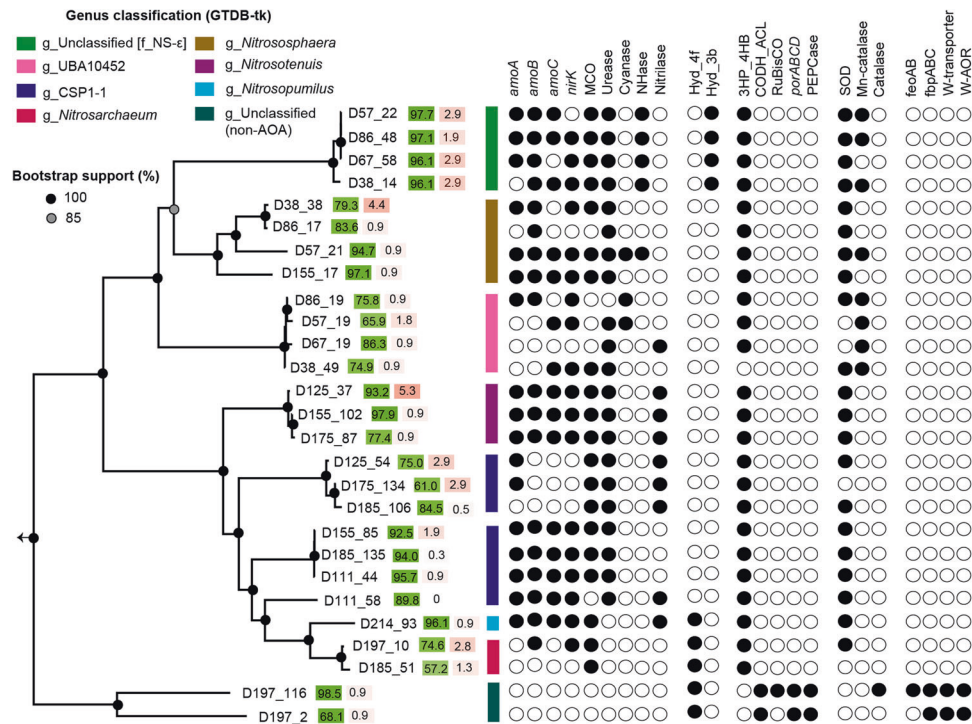


Fig. 3 Metabolic features across MAG phylogenetic clusters. The phylogeny on the left was inferred using the maximum-likelihood method in IQ-Tree, based on a concatenated alignment of ribosomal marker genes (see “Methods”). Highlighted next to each MAG name are estimates of genome completeness (green) and redundancy (red), which are also presented in Table 1. The colored bars next to each cluster indicate the genus-level classification as obtained via the GTDB toolkit. Bootstrap support values for 1000 replicates are indicated at each node. The tree was rooted with the *Ca. Caldiarchaeum subterraneum* genome. *amoABC* ammonia monooxygenase subunits A, B and C; *nirK* nitrite reductase; MCO multicopper oxidase; NHase nitrile hydratase; Hyd_4 group 4 [NiFe] hydrogenase; Hyd_3b group 3b [NiFe] hydrogenase; 3HP_HB 3-hydroxypropionate/4-hydroxybutyrate cycle (using the *hcd* gene that codes for hydroxybutyryl-CoA dehydratase as the pathway marker); CODH_ACL carbon monoxide dehydrogenase/acetyl-CoA synthase; *porABCD* pyruvate:ferredoxin oxidoreductase subunits A, B, C and D; PEPCase phosphoenolpyruvate carboxylase; SOD superoxide dismutase; *feoAB* ferrous ion transporter; *fbpABC* ferric ion transporter; W-AOR tungsten-containing aldehyde:ferredoxin oxidoreductase.

harboring the nitrile-hydrolyzing enzymes - nitrilases and nitrile hydratases. Nitrilases catalyze the formation of ammonia from nitrile compounds directly [90], whereas nitrile hydratases (NHases) catalyze the hydrolysis of nitriles to the corresponding amides, which can then be converted to ammonia and carboxylic acids by an amidase [91]. Several AOA within the *Nitrosopumilus*, *Nitrosotenuis* and *Nitrosocaldus* genera are known to carry nitrilases e.g., [92–95]; however, these appear to be phylogenetically distinct from the NHases found in the Riverton *Nitrososphaerales* MAGs (Fig. S5). The relatively patchy distribution of NHases within *Nitrososphaerales* points to potential horizontal acquisition or loss of this gene by AOA lineages (Fig. S5).

Genomic capabilities for reducing oxidative stress varied between the *Nitrosopumilales* and *Nitrososphaerales* MAGs. Superoxide dismutase, which catalyzes the conversion of highly reactive superoxide to hydrogen peroxide (H_2O_2) and oxygen, was found in 21 out of the 25 AOA MAGs (Fig. 3), as expected based on the wide distribution of this gene across AOA clades [96, 97]. H_2O_2 detoxification is most efficiently catalyzed by the gene catalase [98], which is generally absent in AOA, although there are exceptions. Among cultured AOA, manganese-containing catalases (Mn-catalases; as opposed to the typical heme-group containing catalases) have been annotated in genomes of the soil thaumarchaeon *Ca. Nitrososphaera evergladensis* [99] and in *Ca. Nitrosocosmicus exaquare* [100]. A truncated copy of Mn-catalase is also found in the *Ca. N. gargensis* genome, which appears to be horizontally acquired [97]. In accordance with this, several of the *Nitrososphaera* and *Nitrosocosmicus*-like MAGs from Riverton (but none of the *Nitrosopumilales* MAGs) contained Mn-catalases (Fig. 3). H_2O_2 , even at nanomolar levels, has been shown

to inhibit ammonia oxidation by marine AOA (i.e., *Nitrosopumilales*; ref. [101]); and culture experiments have suggested that these AOA might employ α -keto acids or co-occurring catalase-harboring bacteria as H_2O_2 scavengers [97, 102]. Whether the subsurface *Nitrosopumilales* have also adopted these strategies for H_2O_2 detoxification remains to be examined. Intriguingly, a canonical heme catalase was found in the non-AOA MAG D197_116, which clustered with catalase sequences from anaerobic archaea and bacteria, including *Methanomicrobia* and ANME-1 cluster archaea (Fig. S6). Heme-catalase sequences were found in two non-AOA thaumarchaeal MAGs assembled from a peat metagenome [103], one of which appears to be particularly distinct from the rest (Fig. S6). None of the other thaumarchaeal genomes we examined contained homologs of this gene, suggesting lateral acquisition by D197_116 and the peat Thaumarchaeota described above.

All four MAGs in the *Nitrosocosmicus* sister cluster (NS- ϵ), harbored group 3b [NiFe] hydrogenases (Fig. 3; Fig. S7), previously reported only in thermophilic AOA [95], and more recently in *Nitrososphaerales* MAGs obtained from a wastewater treatment plant [104]. The 3b-type hydrogenases are oxygen-tolerant bidirectional enzymes that couple NAD(P)H oxidation/reduction with H_2 production/consumption [105]. These may also act as sulfhydrogenases that reduce elemental sulfur or polysulfide to hydrogen sulfide [106]. Hydrogenases were not detected in any of the remaining *Nitrososphaerales* MAGs (Fig. 3). However, all thaumarchaeal MAGs assembled from 185 cm and below (except for those falling within the CSP1-1 cluster) harbored group 4f [NiFe]-hydrogenases (Fig. 3; Fig. S7). These hydrogenases potentially comprise a respiratory complex that mediates formate oxidation to CO_2 while reducing protons to generate H_2 [107], and

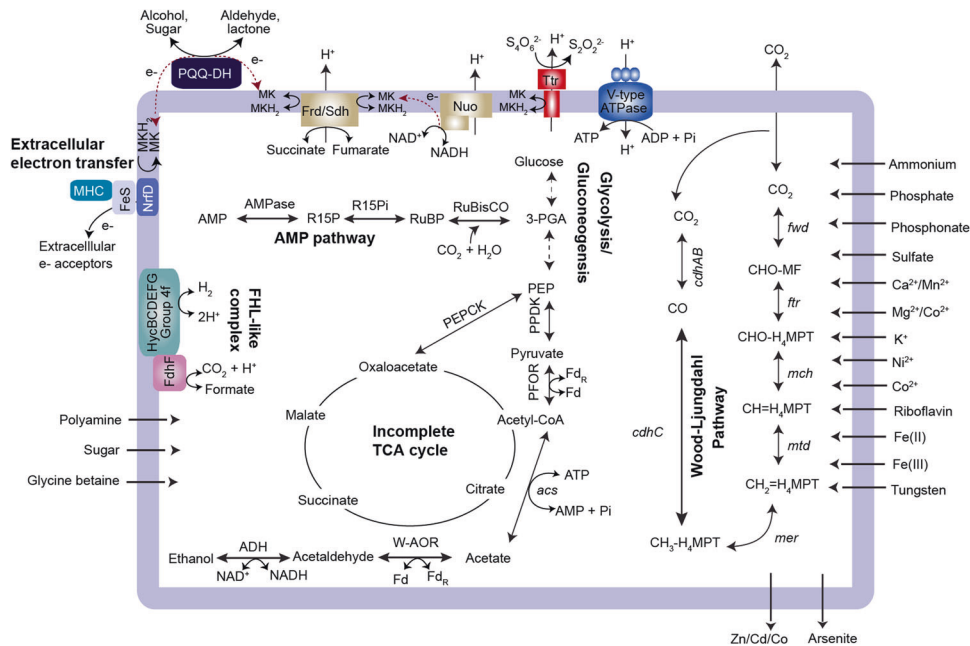


Fig. 4 Metabolic reconstruction of the non-AOA MAG D197_116. Amino acid sequences corresponding to each highlighted gene/pathway are presented in Table S3 in Data Set 1. Abbreviations: CHO-MF formyl-methanofuran, *fwd* formyl-MFR dehydrogenase, CHO-H₄MPT formyl-tetrahydromethanopterin, *ptr* formyl-MFR:H₄MPT formyltransferase, *mch* methenyl-H₄MPT cyclohydrolase, CH=H₄MPT methenyl-tetrahydromethanopterin, *mtd* F420-dependent methylene H₄MPT dehydrogenase, CH₂=H₄MPT methylene-tetrahydromethanopterin, *mer* methylene-H₄MPT reductase, CH₃-H₄MPT methyl-tetrahydromethanopterin, CO carbon monoxide, CO₂ carbon dioxide, *cdhABC* carbon monoxide dehydrogenase/acetyl-CoA synthase subunits, AMP adenosine monophosphate, AMPase AMP phosphorylase, R15P ribose 1,5-bisphosphate, R15Pi ribose 1,5-bisphosphate isomerase, RuBP ribulose 1,5-bisphosphate, RuBisCO ribulose-1,5-bisphosphate carboxylase, 3-PGA 3-phosphoglycerate, PEP phosphoenolpyruvate, PPDK pyruvate phosphate dikinase, PEPCK phosphoenolpyruvate carboxykinase, PFOR pyruvate:ferredoxin oxidoreductase, *acs* acetyl-CoA synthetase, W-AOR tungsten-dependent aldehyde:ferredoxin oxidoreductase, ADH aldehyde dehydrogenase, PQQ-DH pyrroloquinoline quinone-dependent dehydrogenase, Ttr tetrathionate reductase, MHC multiheme cytochromes.

may function in cellular redox balance. Since these genes were found across all three lineages from 185 cm and below (*Nitrosopumilus*, *Nitrosarchaeum* and the non-AOA Thaumarchaeota; Fig. 1), the group 4f hydrogenases are potentially a habitat-specific adaptation in subsurface Thaumarchaeota that aids with redox changes resulting from water table fluctuations.

Divergent non-AOA Thaumarchaeota encoding the Wood-Ljungdahl pathway, a form-III RuBisCO, and the potential for extracellular electron transfer

Two of the MAGs recovered from the 197 cm sediment sample, D197_116 and D197_2, appear to represent a non-AOA basal lineage of Thaumarchaeota (Fig. 1). On the ribosomal protein phylogeny, these MAGs form a sister group to the recently described heterotrophic marine Thaumarchaeota (HMT) lineage (Fig. 1; 16,17). While the HMT lineage was inferred to be capable of aerobic respiration, neither of the two non-AOA MAGs from Riverton harbor aerobic terminal oxidases. Despite this, these two lineages share many metabolic features, including the complete lack of ammonia-oxidation machinery. We expand upon these comparisons in later sections.

Both non-AOA MAGs encode a complete tetrahydromethanopterin (H₄MPT)-dependent Wood-Ljungdahl pathway (WLP). In addition to the key enzyme CO dehydrogenase/acetyl-CoA synthase (CODH/ACS), both MAGs encode the complete archaeal methyl-branch of the WLP (Fig. 4). Homologs of methyl-CoM reductase (McrABC) were not identified in either MAG, excluding the potential for methane metabolism in this lineage. Thus, the WLP in these Thaumarchaeota likely functions in a non-methanogenic, autotrophic CO₂ fixation pathway, as has been suggested recently for several archaeal phyla harboring the WLP without identifiable Mcr homologs [108–112]. Unlike in acetogenic

Bathyarchaeota harboring the WLP [109], the non-AOA MAGs do not contain phosphate acetyltransferase (*pta*) or acetate kinase (*ack*), which are responsible for converting acetyl-CoA to acetate with concomitant generation of ATP via substrate-level phosphorylation. Instead, acetate formation is likely catalyzed by acetyl-CoA synthetase/acetate-CoA ligase, which was identified in both MAGs (Fig. 4).

The presence of WLP within Thaumarchaeota was implied when a previous study [113] identified the CODH/ACS gene cluster in a thaumarchaeal MAG RBG_16_49_8 assembled from a subsurface aquifer sediment metagenome [3]. Following this, another study [18] also highlighted the presence of CODH/ACS subunits in RBG_16_19_8, suggesting that the loss of WLP may be a key evolutionary event marking the transition from basal anaerobic lineages of Thaumarchaeota to oxygen-respiring AOA. We were able to identify the complete H₄MPT-dependent WLP in the RBG_16_19_8 genome as well. On the phylogenomic tree, RBG_16_49_8 clustered together with the two non-AOA MAGs from Riverton (Fig. 1).

The non-AOA MAGs also harbor pyruvate:ferredoxin oxidoreductase (POR), an oxygen-sensitive enzyme catalyzing the decarboxylation of pyruvate to acetyl-CoA. The POR system was identified as a key metabolic feature of anaerobic lineages within Thaumarchaeota [18]. In the non-AOA MAGs described here, acetyl-CoA generated via the WLP or by POR activity could either enter central carbon metabolism via an incomplete TCA cycle or get reduced to ethanol (Fig. 4). The MAGs uniquely contain multiple homologs of a tungsten-containing aldehyde:ferredoxin oxidoreductase (W-AOR), which is potentially involved in the conversion of acetate to ethanol (Fig. 4; [114, 115]). W-AORs are known to have broad substrate range [116] and, therefore, might mediate various redox reactions involving organic acids and

aldehydes. While the best-characterized W-AORs are from hyperthermophilic archaea such as *Pyrococcus furiosus* and *Thermococcus litoralis* [117], homologs have been detected in various anaerobic bacteria [118] and archaea, including protein-metabolizing (cren)archaea in marine sediments [119], as well as Aigarchaeota lineages [120]. To the best of our knowledge, this is the first report of W-AORs in Thaumarchaeota. Tungsten transporters are also present in these genomes, homologs of which were detected in several non-AOA thaumarchaeal genomes analyzed (Fig. 3; Table S2 in Data Set 1).

Similar to the previously described HMT lineage [16, 17], both RVT197_2 and RVT197_116 encode several pyrroloquinoline-quinone-dependent dehydrogenases (PQQ-DHs). As membrane-bound dehydrogenases, PQQ-DHs can supply electrons directly to the membrane quinone pool, and enable growth on various alcohols and sugars, including glucose, methanol and ethanol [121]. Diverse PQQ-DHs are present in the non-AOA MAGs, all of which harbor predicted signal peptides suggesting extracellular localization (Table S3 in Data Set 1). Another similarity between D197_116 and the HMT lineage is the presence of a form III Ribulose-1,5-bisphosphate carboxylase (RuBisCO) gene (Fig. 3). Contig neighborhood comparisons between the two non-AOA MAGs suggest that D197_2 is missing the RuBisCO gene due to genome incompleteness, as the contig homologous to the RuBisCO-containing contig in D197_116 is truncated in this MAG. The D197_116 RuBisCO clusters with archaeal form III-b sequences, also found in bacteria of the Candidate Phyla Radiation (Fig. S8; [122]). RuBisCO genes previously reported in terrestrial Thaumarchaeota also affiliate within the form III-b cluster [96, 123]. The HMT RuBisCO sequences, in contrast, are homologous to the divergent form III-a sequences found in methanogens [16, 17]. The most parsimonious metabolic hypothesis for the HMT genomes suggested the involvement of RuBisCO in a potentially cyclic, anaplerotic CO₂ incorporation pathway [17]. AMP phosphorylase (AMPase), a key gene in the archaeal AMP pathway for nucleotide salvage [124], was not identified in the HMT genomes, which rendered the generation of ribose 1,5-bisphosphate (R15P) in these archaea uncertain. In contrast, D197_116 contains an AMPase in the vicinity of the RuBisCO gene, in addition to an R15P isomerase required for the generation of the RuBisCO substrate ribulose bisphosphate. However, unlike the HMT genomes, the non-oxidative pentose phosphate pathway is missing in these MAGs and, therefore, a cyclic CO₂ incorporation pathway is likely not present. Instead, the product of the AMP pathway, 3-phosphoglycerate, likely enters glycolysis to be converted to pyruvate and acetyl-CoA (Fig. 4).

Basal lineages of Thaumarchaeota described thus far, with the exception of the HMT lineage, are anaerobic heterotrophs respiring sulfate, nitrate, or iron [18, 96, 123]. The HMT lineage is genetically capable of respiring oxygen [16, 17]. Similar to other terrestrial basal groups, the non-AOA MAGs assembled here do not harbor aerobic terminal oxidases. Both MAGs, however, contained respiratory complexes I and II. The complex I (NADH:quinone oxidoreductase; Nuo) gene cluster in these genomes resemble that of the 2M-type complex I found in the HMT lineage [16], which features an extra copy of the NuoM subunit that may enhance the proton pumping efficiency of the complex [125].

D197_116 is potentially capable of extracellular electron transfer, as evidenced by the presence of multiple multiheme c-type cytochromes (MHCs) containing 10–12 heme (CXXCH) motifs each (Fig. 4). These MHCs may be involved in Fe(III) reduction as the gene neighborhood resembled that of the Fe(III)-reducing archaea in the family *Candidatus* Methanoperedenaceae [126, 127]; the MHCs in D197_116 are adjacent to an electron-transporting ferredoxin iron-sulfur protein and membrane-spanning NrfD-like proteins, as found in *Ca. Methanoperedenaceae* [126]. Feo- and Fbp-like iron transporters (ferrous and ferric iron transporters, respectively) are uniquely found in the two basal

MAGs, along with a bacterioferritin homolog that may be involved in intracellular Fe-storage (Fig. 4). Notably, many of these genes are absent in reference thaumarchaeal genomes (Table S2 in Data Set 1). Finally, D197_116 also harbors tetrathionate reductases (Fig. 4; Table S3 in Data Set 1), suggesting that these sulfur compounds may serve as external electron acceptors for these archaea. Membrane topology analysis of the tetrathionate reductase (Ttr) subunits predicted transmembrane domains for TtrC and external localization for TtrAB, potentially indicating their involvement in extracellular redox processes. The broader metabolic plasticity due to diverse pathways for energy generation and carbon assimilation might be advantageous for the survival of these archaea in the capillary fringe that experiences seasonal redox oscillations due to fluctuations in the water table [5].

Pangenomic inference into functional diversification among basal Thaumarchaeota

In order to compare the metabolic features of the non-AOA MAGs, we performed a pangenome analysis of a set of reference genomes representing basal lineages of Thaumarchaeota. The genomes were selected to represent species-level clusters on the GTDB r202 archaeal tree. Intriguingly, two of the genomes—UBA213_sp011331095 (*Nitrososphaerales* archaeon SpSt-435; ref. [128]) and UBA213_sp002713325 (Thaumarchaeota archaeon SAT137; ref. [129])—appear to be ammonia oxidizers, even though both genomes cluster basal to most non-AOA genomes on the ribosomal protein tree (Fig. 5) and harbor divergent copies of the *amoA* gene (Fig. S9). These two genomes thus represent an AOA family that falls outside the monophyletic clade of AOA families described so far.

A pangenome analysis of the basal lineages based on COG functional categories suggested independent gains or losses of functional modules among basal Thaumarchaeota (Fig. 5). For example, the patchy distribution of functional modules including urease, RuBisCO, plastocyanin, catalase, and bacterioferritin indicate that multiple basal lineages may have acquired (or lost) these genes independently along thaumarchaeal evolution (Fig. 5, Table S4). Intriguingly, the subsurface non-AOA genomes (i.e., RVT MAGs and RBG_16_49_8 from Rifle sediments) share several functional modules with the hot spring-associated thaumarchaeal families JACAEJ01 and JAAOZN01 (Fig. 5). These include CO dehydrogenase/acetyl-CoA synthase and other enzymes of the WLP, aldehyde-ferredoxin oxidoreductase, tungstate transporters, formate hydrogen lyase, and a putative DMSO reductase (Table S4; Fig. 5). Consistent with genome streamlining observed among marine AOA, many of the functional modules seem to be lost in transition to the HMT lineage (Fig. 5), with the exception of RuBisCO and PQQ-dependent dehydrogenases.

CONCLUSIONS

This study examined the phylogenetic diversity and metabolic potential of subsurface Thaumarchaeota lineages in floodplain sediments in the Wind River Basin, WY. Metagenomes obtained at discrete depths along the sediment profile yielded diverse Thaumarchaeota MAGs with distinct functional potential. Particularly notable was the shift in phylogenetic identity with sediment depth, which appeared to be linked to soil moisture as well as carbon/nitrogen content. The predominantly terrestrial *Nitrososphaerales* were dominant in the top, well-drained (dry) layers with relatively higher total C (and lower C:N), while the typically marine *Nitrosopumilales* dominated the deeper, moister layers (especially > 111 cm), including the capillary fringe where total C and N were the lowest. Non-ammonia oxidizing Thaumarchaeota MAGs were also recovered from within the capillary fringe. Thus, surface soils were dominated by relatively more generalist AOA capable of

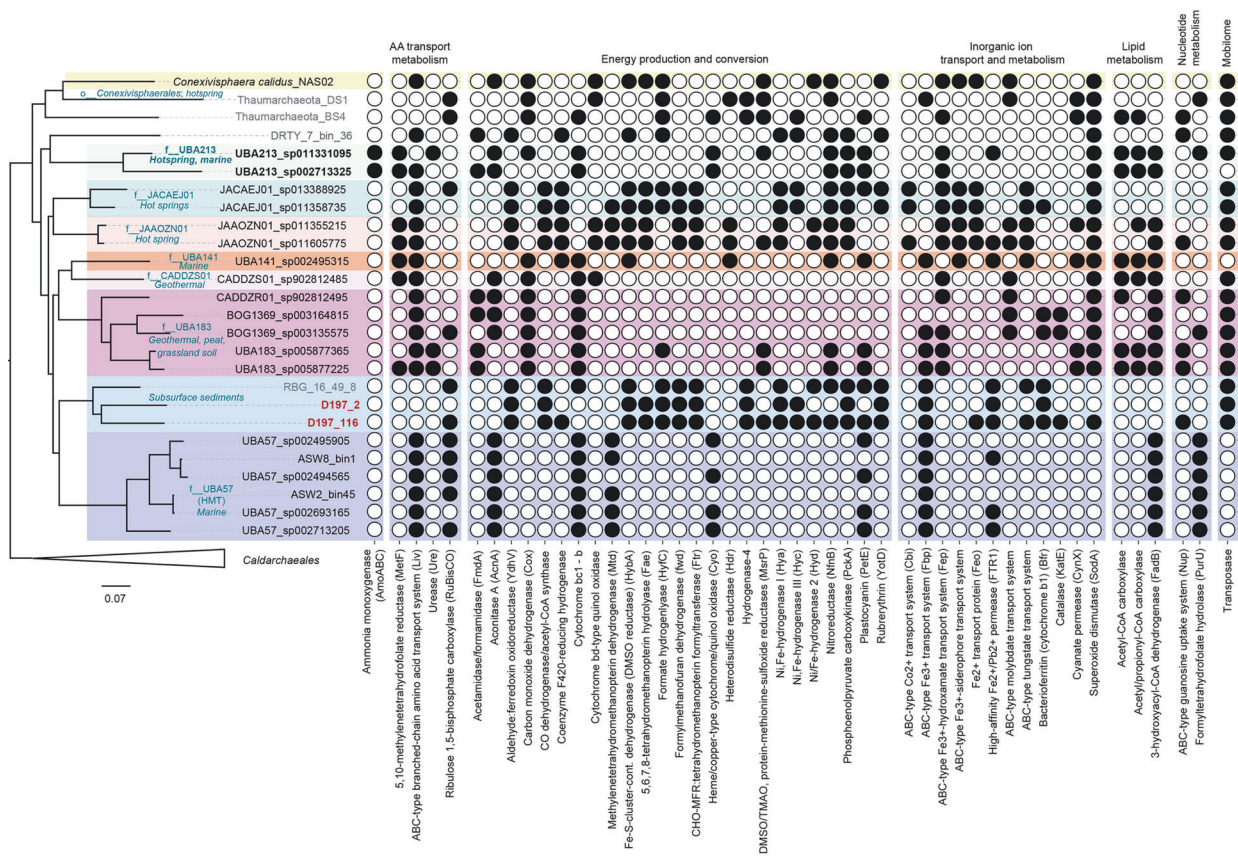


Fig. 5 Pangenomic comparison of selected functions across basal lineages of Thaumarchaeota. Included in the comparison are the two non-AOA MAGs assembled in this study (highlighted in red text), as well as genomes representing each family level clusters of *Nitrososphaerales* defined in the GTDB release 06-RS202. Previously described genomes not included in the GTDB (potentially due to failing quality checks) are indicated in gray letters. Genomes names in bold, black letters indicate the AOA family UBA213, that falls outside of the typical monophyletic AOA clade and harbors divergent copies of the *amoA* gene (Fig. S9). Filled circles indicate the presence of a function/gene. Horizontal shading of the clades signifies taxonomic lineage – either order or family (indicated by the prefixes “o_” and “f_”, respectively). For each clade, the habitats of origin for the corresponding genomes are indicated in italic typeface. Complete pangenome analysis results are presented in Table S4.

utilizing various organic compounds such as urea, cyanate and nitriles whereas typically oligotrophic AOA lineages became prominent in deeper, moister layers. This shift in phylogenetic diversity and metabolic potential from aerobic ammonia-oxidizing lineages in the surface depths to anaerobic, non-AOA lineages with diverse metabolic strategies in deeper layers likely indicates a link between thaumarchaeal population structure and subsurface hydrology, and/or a habitat partitioning pattern resulting from increasing oligotrophy in deeper sediment layers. These results, therefore, emphasize that soil moisture content should be a key variable of consideration in studies of thaumarchaeal evolution. The non-AOA MAGs representing a late-diverging basal lineage are particularly intriguing as they potentially represent an autotrophic lineage among basal Thaumarchaeota. The presence of WLP among multiple family level lineages within Thaumarchaeota suggests independent acquisition of this pathway among Thaumarchaeota. Overall, the Riverton MAGs add to the diversity of thaumarchaeal genomic data while expanding our knowledge on the functional capabilities of these ubiquitous archaea in subsurface environments.

DATA AVAILABILITY

The metagenomes analyzed in this study are available in the NCBI under the following BioProject accessions: PRJNA539634 (D38); PRJNA539635 (D57); PRJNA518423 (D67); PRJNA518424 (D86); PRJNA518421 (D111); PRJNA518422 (D125); PRJNA572113 (D155); PRJNA572114 (D175); PRJNA539171 (D185);

PRJNA539636 (D197); PRJNA572115 (D214); PRJNA518425 (D227). Genomes described in this study have been uploaded to figshare (<https://doi.org/10.6084/m9.figshare.16764199>).

REFERENCES

- Emerson JB, Thomas BC, Alvarez W, Banfield JF. Metagenomic analysis of a high carbon dioxide subsurface microbial community populated by chemolithoautotrophs and bacteria and archaea from candidate phyla. *Environ Microbiol.* 2016;18:1686–703.
- Hug LA, Thomas BC, Sharon I, Brown CT, Sharma R, Hettich RL, et al. Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ Microbiol.* 2016;18:159–73.
- Anantharaman K, Brown CT, Hug LA, Sharon I, Castelle CJ, Probst AJ, et al. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat Commun.* 2016;7:1–11.
- Lu X, Seuradge BJ, Neufeld JD. Biogeography of soil Thaumarchaeota in relation to soil depth and land usage. *FEMS Microbiol Ecol.* 2017;93:fw246.
- Cardarelli EL, Bargar JR, Francis CA. Diverse Thaumarchaeota dominate subsurface ammonia-oxidizing communities in semi-arid floodplains in the western United States. *Micro Ecol.* 2020;80:778–92.
- Tolar BB, Boye K, Bobb C, Maher K, Bargar JR, Francis CA. Stability of floodplain subsurface microbial communities through seasonal hydrological and geochemical cycles. *Front Earth Sci.* 2020;8:338.
- Francis CA, Roberts KJ, Beman JM, Santoro AE, Oakley BB. Ubiquity and diversity of ammonia-oxidizing archaea in water columns and sediments of the ocean. *PNAS.* 2005;102:14683–8.
- Treusch AH, Leininger S, Kletzin A, Schuster SC, Klenk H-P, Schleper C. Novel genes for nitrite reductase and Amo-related proteins indicate a role of

- uncultivated mesophilic crenarchaeota in nitrogen cycling. *Environ Microbiol*. 2005;7:1985–95.
9. Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW, et al. Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*. 2006;442:806–9.
 10. Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. *PNAS*. 2006;103:12317–22.
 11. Prosser JI, Nicol GW. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol*. 2012;20:523–31.
 12. Mußmann M, Brito I, Pitcher A, Damste JSS, Hatzepichler R, Richter A, et al. Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *PNAS*. 2011;108:16771–6.
 13. Weber EB, Lehtovirta-Morley LE, Prosser JI, Gubry-Rangin C, Laanbroek R. Ammonia oxidation is not required for growth of Group 1.1c soil Thaumarchaeota. *FEMS Microbiol Ecol*. 2015;91:fv001.
 14. Lin X, Handley KM, Gilbert JA, Kostka JE. Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *ISME J*. 2015;9:2740–4.
 15. Kato S, Itoh T, Yuki M, Nagamori M, Ohnishi M, Uematsu K, et al. Isolation and characterization of a thermophilic sulfur- and iron-reducing thaumarchaeote from a terrestrial acidic hot spring. *ISME J*. 2019;13:2465–74.
 16. Aylward FO, Santoro AE. Heterotrophic Thaumarchaeota with small genomes are widespread in the dark ocean. *mSystems*. 2020;5:e00415–20.
 17. Reji L, Francis CA. Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J*. 2020;14:2105–15.
 18. Ren M, Feng X, Huang Y, Wang H, Hu Z, Clingenpeel S, et al. Phylogenomics suggests oxygen availability as a driving force in Thaumarchaeota evolution. *ISME J*. 2019;13:2150–61.
 19. Kerou M, Alves RJE, Schleper C. Nitrososphaerales. In: *Bergey's manual of systematics of archaea and bacteria ed. Bergey's Manual Trust (Hoboken, NJ: John Wiley & Sons)*. 2016. <https://doi.org/10.1002/9781118960608.cbm00055>.
 20. Qin W, Martens-Habbena W, Kobelt JN, Stahl DA. Candidatus nitrosopumilales. In: *Bergey's manual of systematics of archaea and bacteria ed. Bergey's Manual Trust (Hoboken, NJ: John Wiley & Sons)*. 2016. <https://doi.org/10.1002/9781118960608.gbm01290>.
 21. Prosser JI, Nicol GW. Candidatus Nitrosotaleales. In: *Bergey's manual of systematics of archaea and bacteria ed. Bergey's Manual Trust (Hoboken, NJ: John Wiley & Sons)*. 2016. <https://doi.org/10.1002/9781118960608.obm00123>.
 22. Gubry-Rangin C, Kratsch C, Williams TA, McHardy AC, Embley TM, Prosser JI, et al. Coupling of diversification and pH adaptation during the evolution of terrestrial Thaumarchaeota. *PNAS*. 2015;112:9370–5.
 23. Nicol GW, Leininger S, Schleper C, Prosser JI. The influence of soil pH on the diversity, abundance and transcriptional activity of ammonia oxidizing archaea and bacteria. *Environ Microbiol*. 2008;10:2966–78.
 24. Szukics U, Abell GCJ, Hödl V, Mitter B, Sessitsch A, Hackl E, et al. Nitrifiers and denitrifiers respond rapidly to changed moisture and increasing temperature in a pristine forest soil. *FEMS Microbiol Ecol*. 2010;72:395–406.
 25. Höfflerle Š, Nicol GW, Pal L, Hacin J, Prosser JI, Mandić-Mulec I. Ammonium supply rate influences archaeal and bacterial ammonia oxidizers in a wetland soil vertical profile. *FEMS Microbiol Ecol*. 2010;74:302–15.
 26. Tourna M, Freitag TE, Nicol GW, Prosser JI. Growth, activity and temperature responses of ammonia-oxidizing archaea and bacteria in soil microcosms. *Environ Microbiol*. 2008;10:1357–64.
 27. He J-Z, Shen J-P, Zhang L-M, Zhu Y-G, Zheng Y-M, Xu M-G, et al. Quantitative analyses of the abundance and composition of ammonia-oxidizing bacteria and ammonia-oxidizing archaea of a Chinese upland red soil under long-term fertilization practices. *Environ Microbiol*. 2007;9:2364–74.
 28. Marusenko Y, Bates ST, Anderson I, Johnson SL, Soule T, Garcia-Pichel F. Ammonia-oxidizing archaea and bacteria are structured by geography in biological soil crusts across North American arid lands. *Ecol Process*. 2013;2:9.
 29. Opitz S, Küsel K, Spott O, Totsche KU, Herrmann M. Oxygen availability and distance to surface environments determine community composition and abundance of ammonia-oxidizing prokaryotes in two superimposed pristine limestone aquifers in the Hainich region, Germany. *FEMS Microbiol Ecol*. 2014;90:39–53.
 30. Purkamo L, Kietäväinen R, Miettinen H, Sohlberg E, Kukkonen I, Itävaara M, et al. Diversity and functionality of archaeal, bacterial and fungal communities in deep Archean bedrock groundwater. *FEMS Microbiol Ecol*. 2018;94.
 31. Bushnell B. *BBTools* software package. 2014. <http://btools.jgi.doe.gov>.
 32. Li H. BFC: correcting Illumina sequencing errors. *Bioinformatics*. 2015;31:2885–7.
 33. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node approach for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*. 2015;31:1674–6.
 34. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, et al. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11.
 35. Kang D, Li F, Kirton ES, Thomas A, Egan RS, An H, et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*. 2019;7:e7359.
 36. Wu Y-W, Tang Y-H, Tringe SG, Simmons BA, Singer SW. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*. 2014;2:26.
 37. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32:605–7.
 38. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome*. 2018;6:158.
 39. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, et al. Assembling genomes and mini-metagenomes from highly chimeric reads. In: Deng M, Jiang R, Sun F, Zhang X, editors. *Research in Computational Molecular Biology (RECOMB)*, Lecture Notes in Computer Science, Springer; Berlin, Heidelberg. 2013;7821:158–70.
 40. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2015;25:1043–55.
 41. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2020;36:1925–7.
 42. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol*. 2020;38:1079–86.
 43. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol*. 2018;36:996–1004.
 44. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
 45. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform*. 2010;11:119.
 46. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
 47. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol*. 2016;428:726–31.
 48. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAA: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res*. 2007;35:W182–5.
 49. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, Mering von C, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol*. 2017;34:2115–22.
 50. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;47:D309–14.
 51. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res*. 2013;42:D206–14.
 52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
 53. Elbourne LDH, Tetu SG, Hassan KA, Paulsen IT. TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. *Nucleic Acids Res*. 2016;45:D320–4.
 54. Nielsen H, Engelbrecht J, Brunak S, von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng*. 1997;10:1–6.
 55. Armenteros JJA, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, et al. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420–3.
 56. Sonnhammer EL, Heijne G, von, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc Int Conf Intell Syst Mol Biol*. 1998;6:175–82.
 57. Krogh A, Larsson B, Heijne G, von, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol*. 2001;305:567–80.
 58. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319.
 59. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
 60. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Nucleic Acids Res*. 2009;25:1972–3.

61. Nguyen L-T, Schmidt HA, Haeseler von A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Nucleic Acids Res.* 2015;32:268–74.
62. Hoang DT, Chernomor O, Haeseler von A, Minh BQ. Le Sy Vinh. UFBoot2: improving the ultrafast bootstrap approximation. *Nucleic Acids Res.* 2017;35:518–22.
63. Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler von A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14:587–9.
64. Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE.* 2010;5:e9490.
65. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* 2011;7:539–9.
66. Chen I-MA, Chu K, Palaniappan K, Ratner A, Huang J, Huntemann M, et al. The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 2021;49:D751–63.
67. Alves RJE, Minh BQ, Ulrich T, Haeseler A, Schleper C. Unifying the global phylogeny and environmental distribution of ammonia-oxidising archaea based on amoA genes. *Nat Commun.* 2018;9:1517.
68. Tolar BB, Mosier AC, Lund MB, Francis CA. Nitrosarchaeum. In: *Bergey's manual of systematics of archaea and bacteria ed. Bergey's Manual Trust (Hoboken, NJ: John Wiley & Sons).* 2019:1–9. <https://doi.org/10.1002/9781118960608.gbm01289>.
69. Park S-J, Kim J-G, Jung M-Y, Kim S-J, Cha I-T, Ghai R, et al. Draft genome sequence of an ammonia-oxidizing archaeon, "Candidatus Nitrosopumilus sediminis" AR2, from Svalbard in the Arctic Circle. *J Bacteriol.* 2012;194:6948–9.
70. Kim BK, Jung M-Y, Yu DS, Park S-J, Oh TK, Rhee S-K, et al. Genome sequence of an ammonia-oxidizing soil archaeon, "Candidatus Nitrosoarchaeum koreensis" MY1. *J Bacteriol.* 2011;193:5539–40.
71. Ochsenreiter T, Selezid D, Quaiser A, Bonch-Osmolovskaya L, Schleper C. Diversity and abundance of Crenarchaeota in terrestrial habitats studied by 16S RNA surveys and real time PCR. *Environ Microbiol.* 2003;5:787–97.
72. Lehtovirta-Morley LE, Stoecker K, Vilcinskas A, Prosser JL, Prosser, Nicol GW. Cultivation of an obligate acidophilic ammonia oxidizer from a nitrifying acid soil. *PNAS.* 2011;108:15892–7.
73. Lehtovirta-Morley LE, Ross J, Hink L, Weber EB, Gubry-Rangin C, Thion C, et al. Isolation of "Candidatus Nitrosocosmicus franklandus," a novel ureolytic soil archaeal ammonia oxidiser with tolerance to high ammonia concentration. *FEMS Microbiol Ecol.* 2016;92:fw057.
74. Könneke M, Bernhard AE, la Torre de JR, Walker CB, Waterbury JB, Stahl DA. Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature.* 2005;437:543–6.
75. Qin W, Amin SA, Martens-Habbena W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *PNAS.* 2014;111:12504–9.
76. Santoro AE, Dupont CL, Richter RA, Craig MT, Carini P, McIlvin MR, et al. Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": an ammonia-oxidizing archaeon from the open ocean. *PNAS.* 2015;112:1173–8.
77. Bayer B, Vojvoda J, Offre P, Alves RJE, Elisabeth NH, Garcia JA, et al. Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J.* 2015;10:1051–63.
78. Larentis M, Psenner R, Alfreider A. Prokaryotic community structure in deep bedrock aquifers of the Austrian Central Alps. *Antonie van Leeuwenhoek.* 2015;107:687–701.
79. Lazar CS, Stoll W, Lehmann R, Herrmann M, Schwab VF, Akob DM, et al. Archaeal diversity and CO₂ fixers in carbonate-/siliclastic-rock groundwater ecosystems. *Archaea.* 2017;2136287.
80. Sheridan PO, Raguideau S, Quince C, Holden J, Zhang L, Williams TA, et al. Gene duplication drives genome expansion in a major lineage of Thaumarchaeota. *Nat Commun.* 2020;11:1–12.
81. Könneke M, Schubert DM, Brown PC, Hügl M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *PNAS.* 2014;111:8239–44.
82. Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y-I, Sugahara J, et al. Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *PNAS.* 2006;103:18296–301.
83. Spang A, Pehlele A, Offre P, Zumber a gel S, Haider S, Rychlik N, et al. The genome of the ammonia-oxidizing Candidatus Nitrososphaera gargensis: insights into metabolic versatility and environmental adaptations. *Environ Microbiol.* 2012;14:3122–45.
84. Kamanda Ngugi D, Blom J, Alam I, Rashid M, Ba-Alawi W, Zhang G, et al. Comparative genomics reveals adaptations of a halotolerant thaumarchaeon in the interfaces of brine pools in the Red Sea. *ISME J.* 2015;9:396–411.
85. Abby SS, Melcher M, Kerou M, Krupovic M, Stieglmeier M, Rossel C, et al. Candidatus Nitrosocaldus cavascurensis, an ammonia oxidizing, extremely thermophilic archaeon with a highly mobile genome. *Front Microbiol.* 2018;9:28.
86. Tourna M, Stieglmeier M, Spang A, Konneke M, Schintlmeister A, Ulrich T, et al. Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *PNAS.* 2011;108:8420–5.
87. Johnson WV, Anderson PM. Bicarbonate is a recycling substrate for cyanase. *J Biol Chem.* 1987;262:9021–5.
88. Palatinszky M, Herbold C, Jehmlich N, Pogoda M, Han P, Bergen von M, et al. Cyanate as an energy source for nitrifiers. *Nature.* 2015;524:105–8.
89. Kitzinger K, Padilla CC, Marchant HK, Hach PF, Herbold CW, Kidane AT, et al. Cyanate and urea are substrates for nitrification by Thaumarchaeota in the marine environment. *Nat Microbiol.* 2019;4:234–43.
90. Pace HC, Brenner C. The nitrilase superfamily: classification, structure and function. *Genome Biol.* 2001;2:REVIEWS0001. <https://doi.org/10.1186/gb-2001-2-1-reviews0001>.
91. Ramteke PW, Maurice NG, Joseph B, Wadher BJ. Nitrile-converting enzymes: an eco-friendly tool for industrial biocatalysis. *Biotechnol Appl Biochem.* 2013;60:459–81.
92. Walker CB, la Torre de JR, Klotz MG, Urakawa H, Pinel N, Arp DJ, et al. Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *PNAS.* 2010;107:8818–23.
93. Mosier AC, Lund MB, Francis CA. Ecophysiology of an ammonia-oxidizing archaeon adapted to low-salinity habitats. *Micro Ecol.* 2012;64:955–63.
94. Lebedeva EV, Hatzenpichler R, Pelletier E, Schuster N, Hauzmayer S, Bulaev A, et al. Enrichment and genome sequence of the group 1.1a ammonia-oxidizing archaeon "Ca. Nitrosotenuis uzonensis" representing a clade globally distributed in thermal habitats. *PLoS ONE.* 2013;8:e80835.
95. Daebele A, Herbold C, Vierheilig J, Sedlacek CJ, Pjevac P, Albertsen M, et al. Cultivation and genomic analysis of "Candidatus Nitrosocaldus islandicus," an obligately thermophilic, ammonia-oxidizing thaumarchaeon from a hot spring biofilm in Graendalur valley, Iceland. *Front Microbiol.* 2018;9:193.
96. Beam JP, Jay ZJ, Kozubal MA, Inskeep WP. Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J.* 2014;8:938–51.
97. Kim J-G, Park S-J, Damste JSS, Schouten S, Rijpstra WIC, Jung M-Y, et al. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *PNAS.* 2016;113:7888–93.
98. Imlay JA. Cellular defenses against superoxide and hydrogen peroxide. *Annu Rev Biochem.* 2008;77:755–76.
99. Zhalnina KV, Dias R, Leonard MT, de Quadros PD, Camargo FAO, Drew JC, et al. Genome sequence of Candidatus Nitrososphaera evergladensis from group 1.1b enriched from everglades soil reveals novel genomic features of the ammonia-oxidizing archaea. *PLoS ONE.* 2014;9:e101648.
100. Sauder LA, Albertsen M, Engel K, Schwarz J, Nielsen PH, Wagner M, et al. Cultivation and characterization of Candidatus Nitrosocosmicus exaquare, an ammonia-oxidizing archaeon from a municipal wastewater treatment system. *ISME J.* 2017;11:1142–57.
101. Tolar BB, Powers LC, Miller WL, Wallsgrove NJ, Popp BN, Hollibaugh JT. Ammonia oxidation in the ocean can be inhibited by nanomolar concentrations of hydrogen peroxide. *Front Mar Sci.* 2016;3:237.
102. Bayer B, Pelikan C, Bittner MJ, Reinthaler T, Könneke M, Herndl GJ, et al. Proteomic response of three marine ammonia-oxidizing archaea to hydrogen peroxide and their metabolic interactions with a heterotrophic alphaproteobacterium. *mSystems.* 2019;4:e00181–19.
103. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zhayed AAF, et al. Genome-centric view of carbon processing in thawing permafrost. *Nature.* 2018;560:49–54.
104. Yang Y, Herbold CW, Jung M-Y, Qin W, Cai M, Du H, et al. Survival strategies of ammonia-oxidizing archaea (AOA) in a full-scale WWTP treating mixed landfill leachate containing copper ions and operating at low-intensity of aeration. *Water Res.* 2021;191:116798.
105. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, et al. Genomic and metagenomic surveys of hydrogenase distribution indicate H₂ is a widely utilised energy source for microbial growth and survival. *ISME J.* 2016;10:761–77.
106. Ma K, Schicho RN, Kelly RM, Adams MW. Hydrogenase of the hyperthermophile Pyrococcus furiosus is an elemental sulfur reductase or sulfhydrogenase: evidence for a sulfur-reducing hydrogenase ancestor. *PNAS.* 1993;90:5341–4.
107. Finney AJ, Sargent F. Formate hydrogenlyase: A group 4 [NiFe]-hydrogenase in tandem with a formate dehydrogenase. *Adv Micro Physiol.* 2019;74:465–86.
108. Baker BJ, Saw JH, Lind AE, Lazar CS, Hinrichs KU, Teske AP, et al. Genomic inference of the metabolism of cosmopolitan subsurface archaea, Hadesarchaea. *Nat Microbiol.* 2016;1:1–9.
109. He Y, Li M, Perumal V, Feng X, Fang J, Xie J, et al. Genomic and enzymatic evidence for acetogenesis among multiple lineages of the archaeal phylum Bathyarchaeota widespread in marine sediments. *Nat Microbiol.* 2016;1:1–9.

110. Lazar CS, Baker BJ, Seitz KW, Teske AP. Genomic reconstruction of multiple lineages of uncultured benthic archaea suggests distinct biogeochemical roles and ecological niches. *ISME J.* 2017;11:1118–29.
111. Farag IF, Biddle JF, Zhao R, Martino AJ, House CH, León-Zayas RI. Metabolic potentials of archaeal lineages resolved from metagenomes of deep Costa Rica sediments. *ISME J.* 2020;14:1345–58.
112. Orsi WD, Vuillemin A, Rodriguez P, Coskun ÖK, Gomez-Saez GV, Lavik G, et al. Metabolic activity analyses demonstrate that Lokiarchaeon exhibits homoacetogenesis in sulfidic marine sediments. *Nat Microbiol.* 2020;5:248–55.
113. Adam PS, Borrel G, Gribaldo S. Evolutionary history of carbon monoxide dehydrogenase/acetyl-CoA synthase, one of the oldest enzymatic complexes. *PNAS.* 2018;115:E1166–73.
114. Köpke M, Held C, Hujer S, Liesegang H, Wiezer A, Wollherr A, et al. Clostridium ljungdahlii represents a microbial production platform based on syngas. *PNAS.* 2010;107:13087–92.
115. Lazar CS, Baker BJ, Seitz KW, Hyde AS, Dick GJ, Hinrichs KU, et al. Genomic evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota in estuarine sediments. *Nucleic Acids Res.* 2015;18:1200–11.
116. Debnar-Daumler C, Seubert A, Schmitt G, Heider J. Simultaneous involvement of a tungsten-containing aldehyde:ferredoxin oxidoreductase and a phenylacetaldehyde dehydrogenase in anaerobic phenylalanine metabolism. *J Bacteriol.* 2014;196:483–92.
117. Kletzin A, Mukund S, Kelley-Crouse TL, Chan MK, Rees DC, Adams MW. Molecular characterization of the genes encoding the tungsten-containing aldehyde ferredoxin oxidoreductase from *Pyrococcus furiosus* and formaldehyde ferredoxin oxidoreductase from *Thermococcus litoralis*. *J Bacteriol.* 1995;177:4817–9.
118. Arndt F, Schmitt G, Winiarska A, Saft M, Seubert A, Kahnt J, et al. Characterization of an aldehyde oxidoreductase from the mesophilic bacterium *Aromatoleum aromaticum* ebn1, a member of a new subfamily of tungsten-containing enzymes. *Front Microbiol.* 2019;10. <https://doi.org/10.3389/fmicb.2019.00071>.
119. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, et al. Predominant archaea in marine sediments degrade detrital proteins. *Nature.* 2013;496:215–8.
120. Dimapilis JRR. Tungsten is essential for long-term maintenance of members of candidate archaeal genus Aigarchaeota Group 4. [dissertation on the Internet]. San Bernardino, California State University; 2019. <https://scholarworks.lib.csusb.edu/etd/927/>.
121. Anthony C. The quinoprotein dehydrogenases for methanol and glucose. *Arch Biochem Biophys.* 2004;428:2–9.
122. Jaffe AL, Castelle CJ, Dupont CL, Banfield JF. Lateral gene transfer shapes the distribution of rubisco among candidate phyla radiation bacteria and DPANN archaea. *Nucleic Acids Res.* 2019;36:435–46.
123. Herbold CW, Lehtovirta-Morley LE, Jung M-Y, Jehmlich N, Hausmann B, Han P, et al. Ammonia-oxidising archaea living at low pH: insights from comparative genomics. *Environ Microbiol.* 2017;19:4939–52.
124. Aono R, Sato T, Imanaka T, Atomi H. A pentose bisphosphate pathway for nucleoside degradation in Archaea. *Nat Chem Biol.* 2015;11:355–60.
125. Chadwick GL, Hemp J, Fischer WW, Orphan VJ. Convergent evolution of unusual complex I homologs with increased proton pumping capacity: energetic and ecological implications. *ISME J.* 2018;12:2668–80.
126. Cai C, Leu AO, Xie G-J, Guo J, Feng Y, Zhao J-X, et al. A methanotrophic archaeon couples anaerobic oxidation of methane to Fe(III) reduction. *ISME J.* 2018;12:1929–39.
127. Leu AO, McIlroy SJ, Ye J, Parks DH, Orphan VJ, Tyson GW. Lateral gene transfer drives metabolic flexibility in the anaerobic methane-oxidizing archaeal family Methanoperedenaceae. *mBio.* 2020;11:e01325–20.
128. Zhou Z, L Y, Xu W, Pan J, Luo Z-H, Li M. Genome- and community-level interaction insights into carbon utilization and element cycling functions of Hydrothermarchaeota in hydrothermal sediment. *mSystems.* 2020;5:e00795–19.
129. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. *Sci Data.* 2018; 5:170203.

ACKNOWLEDGEMENTS

This work was (partially) supported by the SLAC Floodplain Hydro-Biogeochemistry Science Focus Area (SFA), which is funded by the U.S. Department of Energy (DOE) office of Biological and Environmental Research (BER), Earth and Environmental System Science Program, under DOE contract No. DE-AC02-76SF00515 to SLAC, and by contract DE-SC0019119 to C.A.F. We gratefully acknowledge substantial logistical support from the U.S. Department of Energy, Office of Legacy Management, which provided access to the Riverton field site, obtained permits for operations, provided ES&H support, and conducted drilling operations. Sequencing was carried out as part of a Community Science Program (CSP) grant to C.A.F. from the DOE Joint Genome Institute. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Computing for this project was performed on the Sherlock 2.0 cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to the results presented here.

AUTHOR CONTRIBUTIONS

LR processed the data, performed the analyses, and drafted the manuscript in consultation with CAF. Sediment core sampling and associated geochemical analyses were performed by ELC and KB. ELC processed the samples for nucleic acid extractions and metagenome sequencing. CAF and JRB conceived and directed this part of the overall SFA project. CAF supervised the study and was in charge of overall direction and planning.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-01167-7>.

Correspondence and requests for materials should be addressed to Christopher A. Francis.

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

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