



Time-series transcriptomics from cold, oxic subseafloor crustal fluids reveals a motile, mixotrophic microbial community

Lauren M. Seyler ^{1,2,3} · Elizabeth Trembath-Reichert⁴ · Benjamin J. Tully ⁵ · Julie A. Huber ³

Received: 5 August 2020 / Revised: 27 October 2020 / Accepted: 11 November 2020
© The Author(s) 2020. This article is published with open access

Abstract

The oceanic crustal aquifer is one of the largest habitable volumes on Earth, and it harbors a reservoir of microbial life that influences global-scale biogeochemical cycles. Here, we use time series metagenomic and metatranscriptomic data from a low-temperature, ridge flank environment representative of the majority of global hydrothermal fluid circulation in the ocean to reconstruct microbial metabolic potential, transcript abundance, and community dynamics. We also present metagenome-assembled genomes from recently collected fluids that are furthest removed from drilling disturbances. Our results suggest that the microbial community in the North Pond aquifer plays an important role in the oxidation of organic carbon within the crust. This community is motile and metabolically flexible, with the ability to use both autotrophic and organotrophic pathways, as well as function under low oxygen conditions by using alternative electron acceptors such as nitrate and thiosulfate. Anaerobic processes are most abundant in subseafloor horizons deepest in the aquifer, furthest from connectivity with the deep ocean, and there was little overlap in the active microbial populations between sampling horizons. This work highlights the heterogeneity of microbial life in the subseafloor aquifer and provides new insights into biogeochemical cycling in ocean crust.

Introduction

Seawater circulating through porous, basaltic oceanic crust constitutes the largest actively flowing aquifer system on Earth, and represents ~2% of the ocean fluid volume (~27 million km³ of water; [1–3]). The convective circulation of seawater through mid-ocean ridges and ridge flanks has

profound effects on crustal and ocean chemistry [4, 5]. Studies of warm, anoxic venting fluids from mid-ocean ridges and ridge flanks indicate a diverse and active microbial community in these crustal fluids (reviewed in [6]). However, much of the remaining crustal pore volume is composed of cold, oxygenated deep ocean water that enters and exits the basaltic crust through seafloor exposures [7, 8]. These crustal fluids are chemically similar to seawater, but slightly enriched in DIC and depleted in DOC and O₂ [9–12]. Microbial communities in the crustal fluids are distinct from those in the overlying bottom water [9, 13], but the lifestyle and adaptive strategies of microbial communities and the biogeochemical processes they mediate in the oligotrophic, oxic subseafloor aquifer remain poorly constrained.

North Pond is an 8 km × 15 km sediment-filled basin located in 8 Myo volcanic crust on the western flank of the Mid-Atlantic Ridge, ~4450 m below the oligotrophic Sargasso Sea [14–16]. The overlying low-permeability sediments prevent seawater intrusion into the porous and permeable basaltic crust below. In 2011, two circulation obviation retrofit kits (CORKs [17, 18]) were installed into drill holes U1382A and U1383C which penetrated oceanic crust at North Pond during IODP Expedition 336, thus enabling sampling and monitoring of the crustal aquifer

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-00843-4>) contains supplementary material, which is available to authorized users.

✉ Lauren M. Seyler
lauren.seyler@stockton.edu

- ¹ School of Natural and Mathematical Sciences, Stockton University, Galloway, NJ, USA
- ² Blue Marble Space Institute of Science, Seattle, WA, USA
- ³ Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA
- ⁴ School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA
- ⁵ Center for Dark Energy Biosphere Investigations, University of Southern California, Los Angeles, CA, USA

[19]. The first crustal fluids were collected from the North Pond CORKs 6 months after installation in 2012, with return expeditions in 2014 and 2017. In addition, a battery-powered GeoMICROBE sled was deployed at each CORK for autonomous time series sampling from April 2012 to April 2014 [20]. 16S rRNA gene sequencing from 2012 showed that the North Pond fluid microbial community is distinct from bottom seawater and U1382A and U1383C communities are distinguishable from one another [9]. Metagenomic data from samples collected between 2012 and 2014 revealed large shifts in the dominant taxonomic groups, with a high degree of functional redundancy in the microbial community [13].

Natural abundance isotopic data from fluids sampled in 2012 and 2014 indicate early removal of dissolved organic carbon (DOC) sourced from the deep ocean, followed by the slower removal of older, more refractory components, with limited chemosynthetic production in the deepest basement fluids [10]. Experiments using fluids collected in 2012 incubated with ^{13}C -labeled bicarbonate and acetate at 5 °C and 25 °C detected both autotrophic and heterotrophic activity at a higher level than in bottom seawater [9], but low rates of metabolic activity were detected in the same fluids at 4 °C using nanocalorimetry [21]. However, these data were gathered while the North Pond aquifer was recovering from the initial drilling disturbances, and time series geochemical data suggests that by 2017, the North Pond aquifer had recovered [12]. Therefore, fluids collected from North Pond in 2017 are likely the most representative of microbial communities in the cold, oxic aquifer.

Here, we present the first metatranscriptomic data from North Pond, including samples from 2012, 2014, and 2017, analyzed to reconstruct microbial metabolic potential, transcript abundance, and community dynamics in the cool, oxic marine crustal habitat. We also generated new metagenomes from 2017 samples and compared results across all three sampling years, with an emphasis on understanding the 2017 population as it represents the community furthest from drilling disturbances. The data support the presence and activity of a motile microbial community with considerable metabolic flexibility, including the ability to carry out both autotrophy and organotrophy under oxic and anoxic conditions. We also present a conceptual model for the key microbially-mediated carbon, nitrogen, and sulfur cycling reactions occurring in the North Pond crustal fluids.

Materials and methods

Sample collection

Fluids were collected from the North Pond crustal aquifer (22°45'N and 46°05'W) in April 2012, April 2014, and

Table 1 Samples analyzed in this study with oxygen concentration data.

Sample	Date Collected	Depth (m)	Oxygen (μM) ^a
U1382A	April 25 2012	90–210	244
U1382A	April 5 2014		233
U1382A	October 12 2017		228
U1383C Shallow	April 30 2012	70–146	216
U1383C Shallow	April 2 2014		202
U1383C Shallow	October 13 2017		198
U1383C Middle	April 30 2012	146–200	n.d.
U1383C Middle	October 15 2017		205
U1383C Deep	April 20 2012	200–332	213
U1383C Deep	March 31 2014		187
U1383C Deep	October 11 2017		173
Bottom Water	October 14 2017	4397 ^b	250

^aOxygen data from [12].

^bDepth in meters below surface, remaining depths are in meters below seafloor (mbsf).

October 2017 using ROV JASON and the Mobile Pumping System (MPS [20]) as described elsewhere [9, 13, 18]. Both U1382A and U1383C CORK installations were sampled via umbilicals that are accessible at the seafloor and terminate at different depths beneath the seafloor (Table 1). CORK observatory U1382A contains a packer seal in the bottom of the casing that isolates the aquifer from the overlying sediment and one sampling depth horizon extending 90–210 meters below seafloor (mbsf). Observatory U1383C contains three sampling depth horizons below the sediment-crust interface separated by packer seals: Shallow (~70–146 mbsf), Middle (146–200 mbsf), and Deep (~200–332 mbsf). At all sampling time points, umbilical lines were flushed before sampling began and both temperature and oxygen were monitored throughout sampling [9, 12, 13]. To collect microbial biomass for -omics analyses in 2017, crustal fluid was pumped at a rate of ~0.5 liters per minute for 80 min through a 0.22 μm , 47 mm GWSP filter (Millipore). These filters were preserved on the seafloor using RNALaterTM (Ambion) as described in [13, 20]. Bottom seawater was also collected at 4397 m water depth using the MPS. Shipboard, all filters were placed in fresh RNALaterTM, incubated at 4 °C for 18 h, and then stored at –80 °C until extraction.

DNA/RNA extraction and library construction

DNA was extracted from 2017 North Pond samples (half of 47 mm flat) filters following a method adapted from [22, 23] (Supplementary Methods). Extracts were quantified using a Quant-iT PicoGreen dsDNA assay kit (Thermo Fisher Scientific) according to the manufacturer's instructions,

using a fluorescent plate reader. DNA concentrations of all samples are shown in Supplementary Table 1.

RNA was extracted from 47 mm flat filter halves from all three expeditions (2012, 2014, and 2017, Table 1) using a modified version of the mirVana RNA extraction kit (Ambion) protocol (Supplementary Methods). Extracts were quantified using a Quant-iT RiboGreen RNA assay kit (Thermo Fisher Scientific). One whole and one-half filter from each sampling horizon from 2017 were extracted in this manner and the extracts pooled for downstream analysis. For the 2012 and 2014 samples, one half filter was extracted. RNA concentrations are shown in Supplementary Table 1.

Metagenomic libraries were prepared from the 2017 DNA extractions using an Ovation Ultralow V2 DNA-Seq library preparation kit (NuGEN), according to the manufacturer's instruction as described in the Supplementary Methods.

RNA yields were too low to use an RNA sequencing library preparation kit. Instead, we used a SuperScript III First-Strand Synthesis System (Invitrogen), followed by an NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module (New England Biolabs), to produce double-stranded cDNA from our RNA extracts. The cDNA was then purified using a MinElute PCR Purification Kit (QIAGEN), and libraries were prepared as for DNA, as described in the Supplementary Methods.

Metagenomes from 2017 and metatranscriptomes from 2012, 2014, and 2017 were sequenced on an Illumina NextSeq 500 at the W.M. Keck sequencing facility at the Marine Biological Laboratory, resulting in an average read length of 151 bp.

Metagenome assembly and mapping

The 2017 metagenomes were sequenced twice due to poor clustering on the first run. Raw sequence data from both runs was visualized using fastqc [24] and quality filtered using Minoche et al. [25] quality filtering scripts [26]. Filtered R1 and R2 reads from each run were interleaved using fq2fa (<https://pypi.org/project/fq2fa/>) and the interleaved files from the two runs were concatenated for all downstream processing. Assemblies were constructed using IDBA-UD version 1.1.3 [27] and quality assessed with MetaQUAST v5.0.2 [28]. Quality filtered metagenomic reads from 2012 and 2014 were reassembled using the same pipeline. Total number of reads, quality filtering results, and assembly stats are reported in Supplementary Data.

Metagenome assemblies from all time points were uploaded to the Joint Genome Institute's Integrated Microbial Genomes and Microbiomes (IMG/M) system [29] for annotation using default IMG parameters [30].

ORF information was extracted from the resulting general feature format (.gff) and fasta files using gff2seqfeatures.py (<https://github.com/ctSkenneron/scriptShed>). The resulting fasta feature nucleotides (.ffn) files were indexed in Kallisto [31], and the R1 and R2 quality filtered read files from each metagenome were mapped to a concatenated fasta feature nucleotides (.ffn) file of all metagenome ORFs to normalize and quantify gene abundance in transcripts per million reads (TPM), which accounts for both gene length and sequencing depth, using kallisto quant (default parameters). Because the KEGG database does not differentiate between *nxrA* and *narG* or *pmoA* and *amoA* genes as they are highly orthologous, we discerned nitrite oxidation from nitrate reduction and methane oxidation from ammonia oxidation as described in Supplementary Methods.

rRNA was removed from the metagenomes by mapping the reads against the SILVA small subunit (SSU; 16S/18S) and large subunit (LSU; 23S/28S) ribosomal databases (release 132 [32]) using SortMeRNA (default parameters [33]). Taxonomy was assigned to the rRNA reads using assign_taxonomy.py in QIIME [34] with the default assignment method UCLUST [35] against the SILVA v132 SSU and LSU databases.

Metatranscriptome assembly and mapping

Quality filtering was performed on the metatranscriptomes in the same manner as for the 2017 metagenomes (Supplementary Methods). rRNA was removed from the quality-filtered metatranscriptome reads using SortMeRNA and taxonomy was assigned to the rRNA files using the SILVA SSU/LSU databases and QIIME as above. The metatranscriptomes with rRNA removed were mapped to a concatenated fasta feature nucleotides (.ffn) file of all metagenome ORFs in Kallisto.

Binning and metagenome assembled genomes (MAGs)

Assembled metagenomes from 2017 were binned with Binsanity [36] and bin quality was assessed using CheckM version 1.0.11 [37] (Supplementary Methods). High-completion bins (Supplementary Data), hereafter referred to as metagenome-assembled genomes (MAGs), were taxonomically classified using PhyloSift [38] and GTDB-Tk [39]. The MAGs were then imported into KBase [40] and a phylogenetic tree with 100 bootstrap replicates was made using SpeciesTreeBuilder v.0.1.0 with FastTree2 [41]. Tree formatting was performed using the Interactive Tree of Life [42]. Functional orthologies as defined by the KEGG database [43] were annotated using MetaSanity v1.2.0 [44] (Supplementary Methods). Genes related to iron

acquisition, storage, and oxidation/reduction were annotated using the FeGenie tool [45].

Mapping metagenomes and metatranscriptomes to MAGs

The relative abundance of each MAG in the metagenomes and metatranscriptomes was calculated using a competitive recruitment approach (Supplementary Methods). Z-scores of the normalized relative fractions were calculated using the average and sample standard deviation of the relative fraction of each MAG across all metagenomes and across all metatranscriptomes.

The metagenome and metatranscriptome MAG mapping data was also used to generate dendrograms using the pvclust package in R [46]. Relative fraction data was scaled using z-scoring, *p* values were calculated using an unweighted pair group method with arithmetic mean for hierarchical clustering, and distances between samples determined using correlation as a measure of similarity. Dendrograms were then generated with 1000 bootstrap replications.

Results

While DNA and RNA yields were generally low (<1 ng/μl, Supplementary Table 1), we were successful in generating the first metatranscriptomes from this environment. We examined over one billion high-quality paired-end Illumina sequencing reads from 11 metagenomes and 10 metatranscriptomes over three sampling periods spanning 5 years from four different subsurface fluid horizons and deep bottom water at the North Pond site (Table 1, Supplementary Data).

While similar taxa were present across all metagenomic and metatranscriptomic samples, their relative abundances fluctuated over time. Most samples were dominated by Gamma- and Alphaproteobacteria, as well as Bacteroidetes and Zetaproteobacteria (Fig. 1a; Supplementary Fig. 1A). Campylobacterota represented a large proportion of the total microbial community in 2012, but in 2017 they made up a much smaller percentage of the crustal fluid community (Fig. 1a), though they remained a significant taxonomic group in the 2017 bottom water (Supplementary Fig. 1A). In addition, Nitrospirae and Nitrospinae assigned reads, which represented ≤3% of each metagenome (Supplementary Fig. 1A), composed 31% of the metatranscriptome-mapped rRNA reads in U1383C Shallow in 2017 (Fig. 1a). Firmicutes-assigned reads (predominantly family *Clostridiaceae*) accounted for >50% of the mapped rRNA reads in the 2017 U1383C Shallow metagenome, but represented only 5% of the mapped metatranscriptome rRNA reads in U1383C Shallow in 2017 (Fig. 1a).

Metabolic, motility, biofilm, and stress transcript abundance across samples

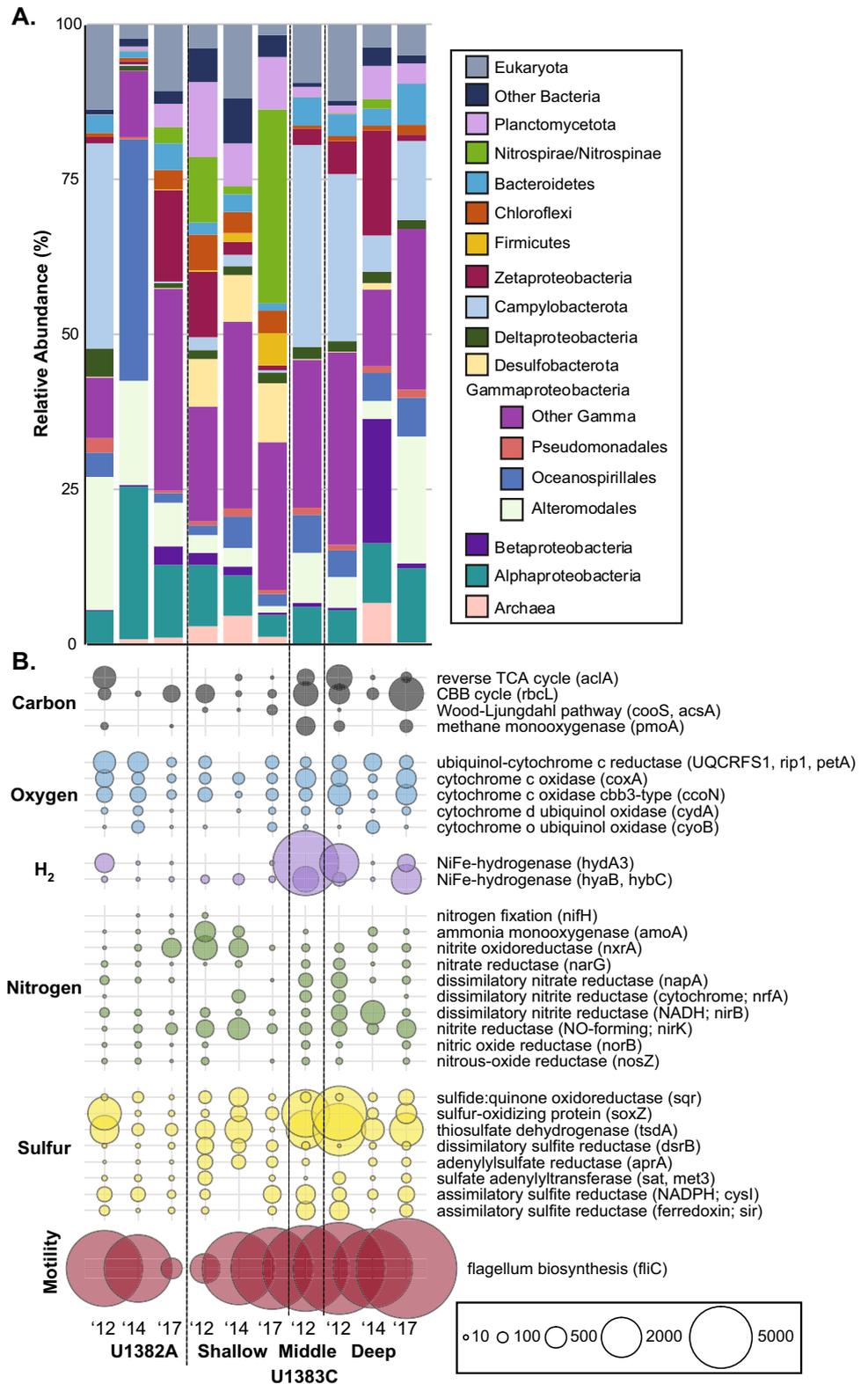
To examine functional potential at the community level across samples, we examined genes encoding the active sites of key enzymes for cycling carbon, hydrogen, oxygen, nitrogen, and sulfur. Overall, these genes were equally abundant in the metagenomes across sampling times and depth horizons (Supplementary Fig. 1B), while transcript abundance was much more variable (Fig. 1b).

In most samples, carbon fixation via the Calvin–Benson–Bassham (CBB) cycle was more abundant than the reverse tricarboxylic acid cycle (rTCA), as indicated by the number of transcripts for the large subunit of RuBisCo (*rcbL*; ~10–1400 tpm) versus the ATP citrate lyase alpha-subunit (*aclA*; ~0–300 tpm; Fig. 1b). In 2017, *rcbL* and *aclA* transcripts were higher in U1382A and U1383C Deep than in U1383C Shallow by 1–2 orders of magnitude (Fig. 1b). The catalytic subunit of carbon-monoxide dehydrogenase (*cooS*) was transcribed in U1383C Shallow in all years and U1383C Deep in 2012 (~1–80 tpm). Aerobic oxidation of methane (*pmoA*) transcripts were observed in 2012 and 2017 but not 2014 (Fig. 1b).

Genes for aerobic oxidative phosphorylation (cytochrome c oxidase (*coxA*), cytochrome o ubiquinol oxidase (*cyoB*), and ubiquinol-cytochrome c reductase (*UQCRCFS1*, *rip1*, *petA*)) and cytochromes related to microaerophily and low oxygen tolerance (*ccb₃*-type cytochrome c oxidase (*ccoN*) and cytochrome d ubiquinol oxidase (*cydA*)) were transcribed in every sample (Fig. 1b). *ccb₃*-type cytochrome c oxidases were 3–7x more abundant in 1383C Deep than in U1382A or 1383C Shallow. NiFe-hydrogenase transcripts used for anaerobic respiration and oxidative stress response (*hyaB*, *hybC*, *hydA3*) were highest in U1383C Middle in 2012 and U1383C Deep in 2012 and 2017, and were 30–1000x more abundant in U1383C Deep than in U1382A or 1383C Shallow in 2017 (Fig. 1b).

Among transcripts involved with nitrogen cycling, the most abundant were associated with ammonia oxidation, nitrate reduction, and nitrite reduction. Transcripts for *amoA* were most prevalent in U1383C Shallow in 2012 (~465 tpm) and 2014 (~110 tpm), but in 2017 were only found in U1382A and U1383C Deep, at an order of magnitude lower abundance (~10–15 tpm) (Fig. 1b). In the metagenomes, *amoA* genes displayed the highest counts in the 2017 bottom seawater sample (Supplementary Fig. 1b). The catalytic subunit of nitrite oxidoreductase (*nxrA*) was transcribed in every sample; transcripts were highest in U1383C Shallow in 2012 (~660 tpm) and 2014 (~380 tpm) and U1382A in 2017 (~390 tpm) (Fig. 1b). Nitrate reduction through either denitrification (*narG*) or the dissimilatory reduction of nitrate to ammonia (DNRA; *napA*) was also transcribed in every sample ranging from 0.5 to >250 tpm, but these

Fig. 1 Relative abundance of taxa and normalized abundance of transcripts annotated in the metatranscriptomes. a Relative abundance of taxa associated with the small subunit (16S/18S) and large subunit (23S/28S) ribosomal genes annotated in the metatranscriptomes. Ribosomal genes identified using SortMeRNA and annotated using UCLUST. **b** Normalized abundance of transcripts of key genes for carbon, oxygen, hydrogen, nitrogen, sulfur, and flagellum biosynthesis within the metatranscriptomes at each site over three sampling years, in transcripts per million reads (TPM).



processes were most abundant in 2012, in U1383C Middle and Deep (Fig. 1b). Nitrite reduction via both DNRA (*nirB*) and denitrification (*nirK*) was more transcribed than nitrate

reduction across samples and years, ranging from ~10 to 640 tpm (Fig. 1b). Nitric oxide reduction to N₂ was transcribed primarily in samples taken in 2012, but was also

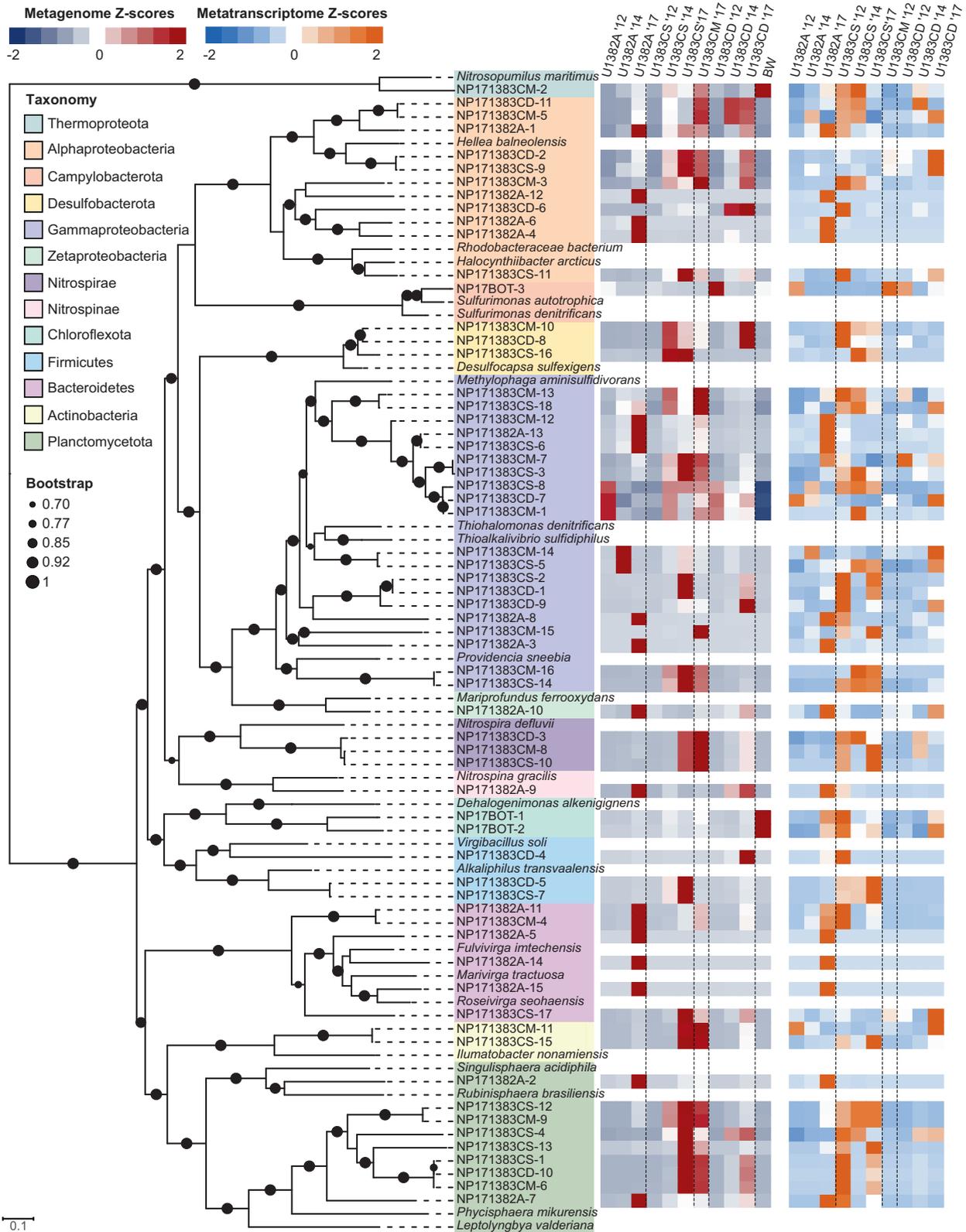


Fig. 2 Bootstrapped phylogenetic tree of “high completion” metagenome assembled genomes (MAGs). Circles at branch nodes denote bootstrapping scores (larger circles = higher scores). Includes competitive recruitment mapping of MAGs to metagenomes (red and dark blue) and metatranscriptomes (orange and blue). Relative

abundance is expressed as z-scores, calculated by subtracting the average relative abundance across all samples from the MAG relative abundance, and dividing by the standard deviation of relative MAG abundance.

observed in U1383C Deep in 2017 (Fig. 1b). Very few transcripts for nitrogen fixation (*nifH*) were detected (0–16 TPM). Nitrogen reduction genes were more abundant than nitrogen oxidation genes in the metagenomes across all samples and years (Supplementary Fig. 1b).

Genes for the oxidation or assimilation of sulfur were more transcribed (by 1–2 orders of magnitude) than for dissimilatory sulfate reduction. The most transcribed sulfur cycling genes were those involved in thiosulfate oxidation: thiosulfate oxidase (*soxZ*, up to ~3800 tpm) and thiosulfate dehydrogenase (*tsdA*, up to ~3500 tpm) (Fig. 1b). While dissimilatory sulfate reduction gene transcripts (*dsrB*, *aprA*, *sat*, *met3*) were present in every sample, they were highest in U1383C Shallow across all sampling years (~100–400 tpm) (Fig. 1b).

Transcripts associated with chemotaxis and flagellum biosynthesis were far more abundant in the metatranscriptomes than those associated with biofilm formation (Supplementary Fig. 2). Flagellin (*fliC*) was among the most abundant transcripts in every sample, and was in the top three most abundant transcripts in every 2017 sample (Fig. 1b, Supplementary Data). The top three most abundant transcripts in U1382A and U1383C Deep in 2017 also included *pilA* (type IV pilus assembly protein PilA), associated with twitching motility (Supplementary Fig. 2, Supplementary Data). Transcripts for catalases associated with oxidative stress (*katG*, *katE*, *CAT*, *catB*, *srpA*, *ahpC*), cold shock, and phage shock proteins were also high in 2017 in U1383C Deep, but displayed far lower transcription levels in other samples (Supplementary Fig. 2; Supplementary Data).

Metagenome-assembled genomes (MAGs)

A total of 447 MAGs were obtained from the 2017 metagenome using BinSanity. 63 non-redundant, high-quality, high-completion bins (hereafter referred to as metagenome-assembled genomes, or MAGs) were chosen for downstream analyses (Supplementary Data). MAGs were designated “NP” (for “North Pond”), followed by the 2-digit sampling year (“17”), followed by the sampling location (e.g. 1382A).

The taxonomic identities of these MAGs closely matched the taxonomic annotations of the 16S/23S rRNA gene mapping in the metagenomes and metatranscriptomes (Figs. 1 and 2, Supplementary Data). We compared the relative abundance of each high-completion MAG across all metagenome and metatranscriptome samples (Fig. 2). The most abundant MAGs in the metagenomes were generally observed in North Pond samples from 2017, while some MAGs were found in multiple fluid horizons and years. Transcriptome reads mapping to MAGs correlated with metagenome reads mapping to MAGs in some cases, and

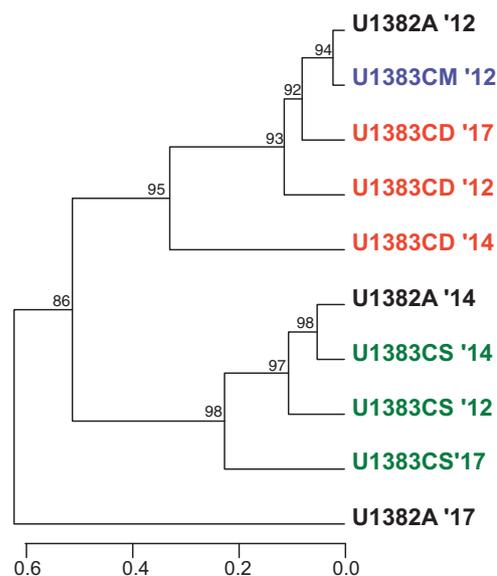


Fig. 3 Hierarchical clustering of MAG abundances in the metatranscriptomes produced using multiscale bootstrap resampling. The scale bar indicates correlation distance between samples.

hierarchical clustering of MAG transcript abundances showed that the U1383C Middle and Deep horizon samples clustered together, and the U1383C Shallow samples clustered together (Fig. 3). Clustering of MAG abundances from the metagenomes indicated all 2012 samples were most similar to one another (Supplementary Fig. 3). U1382A 2017 was distinct from other samples in both the MAG metatranscriptomes (Fig. 3) and metagenomes (Supplementary Fig. 3).

Twenty-two MAGs contained complete or mostly complete (>80%) carbon fixation pathways (Fig. 4, Supplementary Data). Most ($n = 17$) of the MAGs with the potential for carbon fixation contained pathways for the oxidation of sulfide, thiosulfate, or sulfite, making sulfur compounds a probable electron donor (Fig. 4, Supplementary Data). Three MAGs contained genes associated with the oxidation of nitrogen compounds, but only one of these also had the capability to fix carbon (Fig. 4, Supplementary Data). Nine MAGs possessed genes for the oxidation of ferrous iron and five of these, one Alphaproteobacteria and four Gammaproteobacteria, also contained the genes for RuBisCo and the CBB cycle (Fig. 4). One alphaproteobacterium and two gammaproteobacterium MAGs contained genes for NiFe and NAD-reducing hydrogenases (Fig. 4). The majority of MAGs with carbon fixation pathways also contained numerous extracellular protease and carbohydrate catabolism genes (Fig. 4, Supplementary Data). The remaining 42 MAGs lacked carbon fixation genes, but several had the ability to oxidize sulfur compounds and/or ferrous iron (Fig. 4, Supplementary Data).

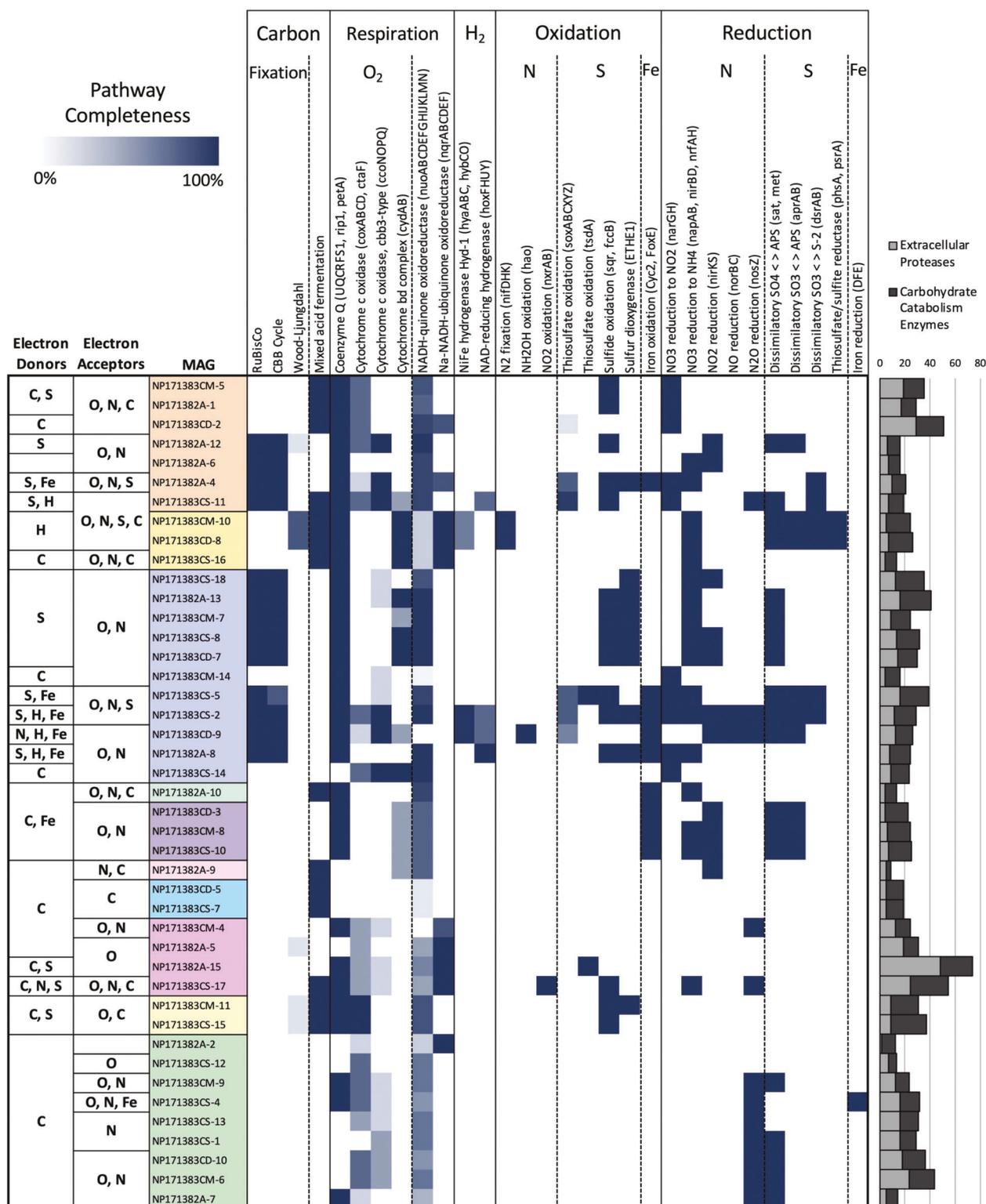


Fig. 4 Completeness of select metabolic pathways of interest in 2017 abundant high-completion MAGs determined using MetaSanity, and Fe oxidation and reduction genes annotated using FeGenie (Supplementary Data). Genes considered in calculating pathway completeness are shown in parentheses, with the exception of carbon cycling pathways, where genes considered are shown in

Supplementary Data. Completeness of each enzymatic pathway is expressed as a percentage (0–100%). Bar graph at right depicts counts of extracellular proteases and carbohydrate-catabolism enzymes in each MAG (complete dataset available in Supplementary Data). Refer to Fig. 2 for color assignments of taxa.

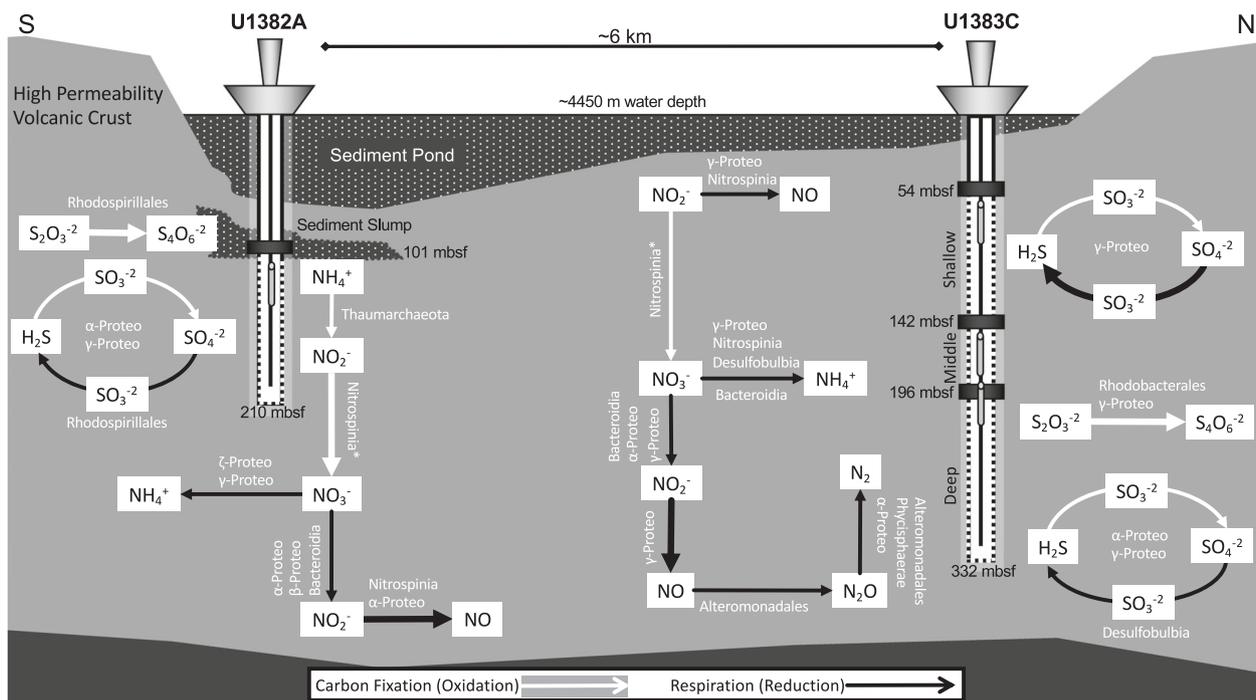


Fig. 5 Schematic of the nutrient cycling processes occurring in North Pond, inferred from metatranscriptomic reads mapped to MAGs, metagenome assemblies, and taxonomic databases. Pathways associated with carbon fixation are indicated with white arrows and respiration is indicated with black arrows. Thicker arrows

represent processes that are particularly abundant in the metatranscriptomes. In the case where a gene was not found in a MAG, we used the taxa annotations performed by IMG (these taxa are indicated with an asterisk).

While the majority (52) of the MAGs contained cytochromes for aerobic respiration (Supplementary Data), all but six aerobic MAGs also had the capability to use other electron acceptors (nitrogen or sulfur) or contained fermentative pathways (Fig. 4, Supplementary Data). All of the MAGs that contained carbon fixation and sulfur oxidation pathways had the ability to use either oxygen or oxidized nitrogen compounds as electron acceptors (Fig. 4, Supplementary Data).

Discussion

Low-temperature, off-axis environments represent the majority of global hydrothermal fluid circulation in the ocean, but the microbial communities in the subseafloor of young ridge flank oceanic crust are sparsely sampled or understood in comparison to other subseafloor habitats such as marine sediments. The goal of this study was to reconstruct microbial metabolic potential, transcript abundance, and community dynamics in the crustal fluids of the cool, oxic ridge flank North Pond using metagenomics and metatranscriptomics, with an emphasis on understanding the microbial community sampled in 2017, furthest removed from the disturbances caused by drilling. This work represents the first metatranscriptomic data recovered

from the cool, oxic subseafloor aquifer, where low biomass presents significant challenges to microbial studies.

Accessing the crustal aquifer in sedimented regions requires ocean drilling and observatory CORK installations to enable collection of crustal fluids, thus all results must be interpreted in the context of the disturbance caused by drilling and subsequent recovery from these operations. Time series sampling at North Pond from 2012–2017 has therefore been critical to resolve when the crustal aquifer returned to its pre-drilling state. Geochemical measurements taken over this 6-year period indicate the fluid chemistry rebounded by 2017 [12]. Geochemical and heat-flow data also show that U1382A has greater connectivity to bottom seawater than U1383C, although all fluids are geochemically very similar to bottom seawater [9, 47]. Heat-flow data [48, 49] and tracer experiments [12] indicate rapid lateral fluid flow from U1382A to U1383C through the porous crust, while radiocarbon data suggests potential residence times on the order of hundreds (U1382A) or thousands (U1383C) of years [10]. Within these important geochemical and hydrogeochemical contexts, we therefore focused our analyses on understanding the microbial community in 2017, the samples furthest from drilling disturbances, while also allowing us to examine how the community function has changed over time by comparing 2017 samples to those collected in 2012 and 2014 [9, 13].

Analysis of metagenomic data from North Pond samples in 2012–2014 suggested a high degree of functional redundancy despite differences in community membership across samples and years [13]. The addition of 2017 metagenomic data as well as time series metatranscriptomic data broadly supports this finding and shows that while most examined functional genes are present across many samples and taxa (Supplementary Fig. 1B), their transcript abundance varies with both time and location within the crust (Fig. 1b). Our analysis of the presence and transcription of metagenome-assembled genomes further indicates that while some members of the community are present and transcribed in multiple years and sampling horizons, many were only abundant and active in 2017 samples (Fig. 2). This is not surprising given the MAGs were constructed only from 2017 samples, but despite this, we found many MAGs at U1383C Shallow that were transcribed across all years in that horizon. Conversely, most MAGs at U1382A and U1383C Deep were only transcribed in 2017, with few shared members between these two locations, consistent with heat-flow and geochemical data [12]. Hierarchical clustering of MAG abundances in the metatranscriptomes suggests that overall, the same MAGs were active in U1383C Shallow and in U1383C Deep across all sampling years, while U1382A was more variable (Fig. 3), which may reflect a higher frequency of bottom seawater intrusions at this sampling horizon [12].

Thus, based on mapping 2017 transcripts to annotated metagenome ORFs, the taxa identified in the metatranscriptomes, and the predicted metabolisms of the most abundantly transcribed MAGs, as well as building on previous work at North Pond, we constructed a conceptual model for the key carbon, nitrogen, and sulfur cycling reactions most likely occurring in the North Pond crustal fluids in 2017 and the microbial taxa which carry out these reactions (Fig. 5). We distinguished reactions that are connected to carbon fixation (oxidation of nitrogen or sulfur) and respiration (reduction of nitrogen or sulfur). The model is a simplified representation of the most abundant microbially-mediated processes occurring at different locations and depths in the North Pond crustal aquifer.

Our analyses also indicated a number of lifestyle strategies and adaptations of microbial communities in the crustal fluids. For example, the abundance of transcripts associated with chemotaxis and flagellar motility, and the concurrent lack of transcripts for biofilm formation, suggest that the microbial community captured in these samples is motile, as biofilm formation is generally preceded by inhibition of flagellar gene transcription [50–53]. A prevalence of motility and chemotaxis genes has been described in warm, anoxic crustal fluids in Juan de Fuca Ridge flank [54]. In addition, we found evidence for several stress response mechanisms which were highly transcribed in U1383C

Deep in 2017, including oxidative stress response, survival under nutrient- and energy-limited conditions (phage shock protein A) [55], and cold shock (CspA) [56] (Supplementary Fig. 2, Supplementary Data). These results suggest that especially at greater depth in the basement at North Pond, microbial communities experience multiple stressors that may be linked to energy and temperature fluctuations.

Carbon fixation and turnover

Based on the numerous extracellular protease and carbohydrate metabolism genes detected in the majority of MAGs, combined with the presence of carbon fixation pathways in 22 of the 63 MAGs, the microbial community of the North Pond aquifer appears to be predominantly organotrophic or mixotrophic. Carbon fixation transcripts were detected in all sites and years, and in 2017 samples, carbon fixation transcripts were more abundant in U1383C Middle and Deep than at U1382A or U1383C Shallow (Fig. 1b). MAGs that possessed both carbon fixation and organic carbon degradation genes were highly expressed in both the U1383C Shallow and U1383C Deep metatranscriptomes (Figs. 2 and 4). These mixotrophs may perform heterotrophy in U1383C Shallow, and carbon fixation in the Deep horizon. These observations are consistent with radiocarbon data from North Pond, which suggests rapid turnover of semi-labile DOC sourced from the deep ocean, followed by slower removal of more refractory components of the DOC pool [10]. Chemosynthetic DOC produced in situ is likely turned over rapidly in the shallow crustal fluids, but $\Delta^{14}\text{C}$ values in the Deep horizon of U1383C suggest a net contribution of ^{14}C -enriched DOC from chemoautotrophy [10]. Two recent microbial studies of sub-seafloor rocks recovered via ocean drilling showed a dominance of heterotrophic bacteria and little evidence for autotrophic processes [57, 58].

Fixation of carbon is most likely linked to the oxidation of sulfide and thiosulfate, as evidenced by both the metatranscriptomic (Fig. 1b) and MAG data (Fig. 4, [13]). Oxidation of sulfur compounds is a well-characterized source of electrons for chemolithoautotrophy in the deep sea [59, 60]. Most of the sulfur-oxidizing chemoautotrophic MAGs were capable of using more than one type of sulfur compound, including oxidation of thiosulfate (*soxZ*, *tsdA*), as well as sulfide:quinone oxidoreductase (*sqr*) in U1383C Deep (Fig. 1b). While hydrogen sulfide has not been detected in the borehole fluids [9, 12], the oxidation of iron in sulfide complexes in the crustal rocks may allow access to both sulfide [61] and thiosulfate [62] for carbon fixation.

Previous metagenomic work at North Pond suggested the potential for some microbes to use H_2 and Fe^{2+} to drive biomass production [13], by using the redox gradient between reduced material in basalt and oxygenated aquifer

fluids [63, 64]. Oxidation of ferrous iron has been described in a low temperature deep-sea environment near the Juan de Fuca hydrothermal field [65]. Our results indicate that some members of the North Pond crustal fluid microbial community may use H_2 and/or Fe^{2+} as electron donors for carbon fixation. We identified six MAGs that contained the CBB cycle and a NiFe hydrogenase and/or the iron oxidation gene *Cyc2* (Fig. 4, Supplementary Data), an outer-membrane cytochrome that has been characterized in multiple iron-oxidizing bacterial lineages [66]. NiFe-hydrogenases catalyze the conversion of H_2 to protons and electrons and are generally inhibited by oxygen, but oxygen-tolerant and aerobic NiFe hydrogenases have been described in marine Gammaproteobacteria [67, 68]. We also identified three MAGs with high transcription in U1383C Shallow and high abundance in the U1383C Deep 2017 metagenome that were taxonomically assigned as *Desulfocapsa* (*Desulfobulbaceae*), two of which contained NiFe hydrogenases (Fig. 4; Supplementary Data). Members of this clade simultaneously oxidize and reduce sulfur compounds for energy while using electrons from H_2 to fix carbon via the Wood-Ljungdahl pathway [69]. Transcripts of the carbon-monoxide dehydrogenase catalytic subunit (*cooS/acsA*) were present in the U1383C Shallow metatranscriptome in 2017 (Fig. 1b), on contigs annotated as *Desulfobulbaceae* (Supplementary Data).

Respiration, oxygen, and anaerobiosis

MAGs possessing carbon fixation pathways all contained at least one aerobic or microaerophilic terminal oxidase (Fig. 4), suggesting that carbon fixers in the aquifer use oxygen as a terminal electron acceptor. However, if oxygen is limiting or unavailable, nitrate or nitrite may be used as an alternative (Fig. 4, [13]). The coupling of sulfur, sulfide, and thiosulfate oxidation to denitrification is well documented in multiple bacterial clades (e.g., [70, 71]). Anaerobic oxidation of sulfur [72] and sulfide [73] coupled to DNRA has also been described in marine environments.

The complete suite of transcripts for denitrification from NO_3^- to N_2 was identified in U1382A and U1383C Deep (Fig. 1b). However, transcripts for the final two steps of denitrification were not abundant in U1382A (<2 tpm; Fig. 1b), and none of the abundant MAGs in U1382A were capable of reduction of NO to N_2O (Figs. 2 and 4). Reduction of nitrogen species from NO to N_2 therefore appears to occur primarily in U1383C Deep, carried out by the Alphaproteobacteria, Gammaproteobacteria (Alteromonadales), and Planctomycetota (Phycisphaerales) (Fig. 5). The majority of MAGs that contained denitrification genes also contained aerobic terminal oxidases, except for two of the Planctomycetota (Fig. 4). While denitrification is generally a strictly anaerobic

process, aerobic denitrifying bacteria have been described [74, 75].

Sulfur reduction was a less common anaerobic respiration strategy, both in the metatranscriptomic data (Fig. 1b) and in the MAGs (Fig. 4). Six MAGs in total had the capability to reduce sulfite to sulfide, all of which contained pathways for carbon fixation: two Desulfobacterota, two Alphaproteobacteria, and two Gammaproteobacteria (Fig. 4). One MAG, a gammaproteobacterium (NP171383CS-2), contained a complete pathway for the reduction of dimethyl sulfoxide (Supplementary Data). Hydrogen sulfide has never been detected in North Pond crustal fluids [9, 12].

Anaerobic processes appear to only be relevant in U1383C Deep, based on the presence of denitrification transcripts and denitrifying MAGs, *cbb₃*-type cytochrome c oxidases, NiFe hydrogenases, and catalases associated with oxidative stress. This is consistent with the lower (173 μM [12]) oxygen concentrations present in this horizon compared to other horizons [12, 13]. In addition, samples collected in 2012, 6 months after borehole drilling ceased, included particle-laden fluids which have not been observed since [9, 12]. Organic aggregates provide microenvironments of anoxia in otherwise oxic seawater, vastly expanding the available niche space of denitrifying and sulfate-reducing bacteria in the global ocean [76]. This may explain why we observed high abundances of NiFe-hydrogenase transcripts associated with anaerobic metabolisms in 2012, and why abundances of these genes were mostly confined to the Deep horizon of U1383C in 2017, post-drilling recovery (Fig. 1b). It is possible that the crustal environment may contain slower flow paths or stagnant zones where oxygen becomes depleted, allowing for the anaerobic metabolisms we observed to take place. Furthermore, the abundance of motility and flagellar genes in the metatranscriptomes and the lack of genes for biofilm formation (Supplementary Fig. 2) also suggests that the microbial community captured in these samples is highly mobile and capable of seeking out organic particles to colonize and consume. Fractures in the basaltic basement may host more sedentary, surface-associated microbial communities with metabolisms not represented in the data presented here.

Nitrification

Cross-hole tracer experiments during the most recent cruise to North Pond detected an increase in nitrate concentrations relative to bottom seawater, most likely the result of microbial nitrification: [12] the oxidation of ammonia to nitrite and nitrite to nitrate. Ammonium concentrations in borehole fluids are usually low (<0.1 μM ammonium), and metatranscriptomic evidence for nitrogen fixation to

ammonia was lacking in the aquifer (Fig. 1b). Dissimilatory nitrate reduction to ammonia (DNRA) transcripts, however, were present in all three 2017 samples (Figs. 1b and 5) and the DNRA pathway was present in 23 of the 64 MAGs (Fig. 4). Ammonia produced by this pathway may be rapidly consumed by nitrifying archaea (Nitrososphaeria) and bacteria (Nitrospina) [77].

amoA transcripts were present in all of the 2017 samples (Fig. 1b) and were found exclusively on Nitrososphaeria-annotated contigs. We did not detect the 4-hydroxybutyrate/3-hydroxypropionate cycle used by ammonia-oxidizing archaea for carbon fixation [78] in any of our samples. However, culturing, DNA stable isotope probing, and MAG data suggest many of these archaea may be mixotrophic or heterotrophic [79–83].

Marine nitrite oxidation is generally associated with Nitrospira and Nitrospina bacteria, which may grow autotrophically or mixotrophically [84–87]. While these phyla were most abundant in the 2017 U1383C Shallow metatranscriptome (Fig. 1a), *nxrA* transcript abundance was relatively low in this sample (11–15 tpm; Fig. 1b). *nxrA* transcripts were most abundant in U1382A in 2017 (Fig. 1b) and associated with unbinned contigs that could broadly be assigned to the Nitrospina (Supplementary Data). While we obtained one Nitrospina and three Nitrospira MAGs, all of them were missing *nxrA* (Fig. 4, Supplementary Data). *nxrA* was identified on only one MAG, which belonged to phylum Bacteroidetes (genus *Thiocapsa*; Fig. 4, Supplementary Fig. 4). Based on 16S/23S data and the taxonomic annotation of the metatranscriptomic contigs, we assume that oxidation of nitrite to nitrate is being carried out primarily by Nitrospina in U1382A and U1383C Shallow (Fig. 5).

Together, our results show that the microbial community in the North Pond crustal aquifer is populated by motile mixotrophic and organotrophic bacteria that are active under both oxic and anoxic conditions. This snapshot of subsurface microbes at North Pond represents the fluids in their most pristine state, furthest removed from the disturbances caused by drilling. While low biomass presents significant challenges for such studies, this first examination of transcripts from the cool, oxic subsurface aquifer highlights the spatial heterogeneity of life in such fluids and the ability of microbes to respond and adapt to different regimes in the crustal matrix.

Data availability

Data from 2017 metagenomes and 2012, 2014, and 2017 metatranscriptomes has been deposited at SRA under the BioProject accession numbers PRJNA554681 and PRJNA522799, respectively, and raw sequence reads are available with accession numbers SRR9705319–SRR9705323, SRR8590998–SRR8591005, and SRR10499325–SRR10499326. Assembled

contigs for each 2017 metagenomic library are publicly available via IMG/MER under submission numbers 3300029216 and 3300029223–3300029226. Contigs for each individual 2017 MAG are available through FigShare at https://figshare.com/articles/dataset/North_Pond_2017_High_Completion_Bins/12389789. Raw sequence reads from Meyer et al. [9] constituting the metagenomic samples from 2012, are available under the BioProject accession no. PRJNA280201. Data from Tully et al. [13], including the 2013 and 2014 metagenomes, are available under the BioProject accession no. PRJNA391950 and raw sequence reads are available with accession no. SRX3143886–SRX3143902. Reassembled 2012–2014 metagenomes used in this study are available via IMG/MER under submission numbers 3300029924–3300029927 and 3300029949–3300029950. All accession numbers are available in Supplementary Data.

Acknowledgements We thank the captain and crew of the R/V Atlantis and Merian, the pilots and engineers of the ROV Jason II, and W. Bach, K. Becker, S. Carr, T. D'Angelo, K. Freil, P. Girguis, C-C. Hsieh, U. Jaekel, B. Kraft, H-T. Lin, M. Mullis, O. Nigro, B. Orcutt, M. Rappé, C. Sullivan, S. Shah Walter, G. Stewart, H. Villinger, C. G. Wheat, and K. Yoshimora, for their work in accomplishing the field programs. Ship time in 2012 and 2014 was provided by the German Science Foundation (DFG). Both Katrina Edwards's and James Cowen's efforts were critical to the field component and success of the North Pond project, and they are missed. The Gordon and Betty Moore Foundation sponsored most of the observatory components at North Pond through grant GBMF1609. This work was supported by NSF OCE-1062006, OCE-1745589 and OCE-1635208 to J.A.H. E.T.R. was supported by a NASA Postdoctoral Fellowship with the NASA Astrobiology Institute and a L'Oréal USA For Women in Science Fellowship. The Center for Dark Energy Biosphere Investigations (C-DEBI OCE-0939564) also supported the participation of J.A.H. and B.T. This is C-DEBI contribution number 548.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

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

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

References

1. Sclater JG, Jaupart C, Galson D. The heat flow through the oceanic and continental crust and heat loss of the Earth. *Rev Geophys*. 1980;18:269–311.
2. Stein CA, Stein S. Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *J Geophys Res*. 1994;99:3081–95.
3. Johnson HP, Pruis MJ. Fluxes of fluid and heat from the oceanic crustal reservoir. *Earth Planet Sci Lett*. 2003;216:565–74.
4. Alt JC. Seafloor processes in mid-ocean ridge hydrothermal systems. In: Humphris SE, Zierenberg RA, Mullineaux LS, Thomson RE (eds.). *Seafloor hydrothermal systems: physical, chemical, biological, and geological interactions*. American Geophysical Union; 1995. vol. 91. pp 85–114.
5. Elderfield H, Schultz A. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Ann Rev Earth Planet Sci*. 1996;24:191–224.
6. Orcutt BN, D'Angelo T, Jungbluth SP, Huber JA, Sylvan JB. Microbial life in oceanic crust. *OSF Preprints*. 2020; <https://doi.org/10.31219/osf.io/2wxe6>.
7. Edwards KJ, Fisher AT, Wheat CG. The deep subsurface biosphere in igneous ocean crust: frontier habitats for microbiological exploration. *Front Microbiol*. 2012;3:8.
8. Wheat CG, Fisher AT, McManus J, Hulme SM, Orcutt BN. Cool seafloor hydrothermal springs reveal global geochemical fluxes. *Earth Planet Sci Lett*. 2017;476:179–88.
9. Meyer JL, Jaekel U, Tully BJ, Glazer BT, Wheat CG, Lin H-T, et al. A distinct and active microbial community in cold oxygenated fluids circulating beneath the Mid-Atlantic seafloor. *Sci Rep*. 2016;6:22541.
10. Shah Walter SR, Jaekel U, Osterholtz H, Fisher AT, Huber JA, Pearson A, et al. Microbial decomposition of marine dissolved organic matter in cool oceanic crust. *Nat Geosci*. 2018;11:334–9.
11. McManus J, Wheat CG, Bach W. Carbon cycling in low temperature hydrothermal systems: the Dorado Outcrop. *Geochim Cosmochim Acta*. 2019;264:1–12.
12. Wheat CG, Becker K, Villinger H, Orcutt BN, Fournier T, Hartwell A, et al. Seafloor cross-hole tracer experiment reveals hydrologic properties, heterogeneities, and reactions in slow-spreading oceanic crust. *Geochem Geophys Geosys*. 2020;21:e2019GC008804.
13. Tully BJ, Wheat CG, Glazer BT, Huber JA. A dynamic microbial community with high functional redundancy inhabits the cold, oxic seafloor aquifer. *ISME J*. 2018;12:1–16.
14. Becker K, Bartetzko A, Davis EE. Leg 174B Synopsis: Revisiting Hole 395A for logging and long-term monitoring of off-axis hydrothermal processes in young oceanic crust. In: Becker K, Malone MJ (eds.). *Proc IODP Sci Results Vol 174B (Integrated Ocean Drilling Program Management International, 2001)*; pp 1–12.
15. Edwards KJ, Bach W, Klaus A. Mid-Atlantic Ridge flank microbiology: initiation of long-term coupled microbiological, geochemical, and hydrological experimentation within the seafloor at North Pond, western flank of the Mid-Atlantic Ridge. *IODP Sci Prosp*. 2010;336:1–62.
16. Villinger, H in Maria S Merian Cruise Report MSM37 1–42 (Senatskommission für Ozeanographie, 2014).
17. Davis EE, Becker K, Pettigrew T, Carson B, MacDonald R. CORK: a hydrologic seal and downhole observatory for deep-ocean boreholes. *Proc Ocean Drill Prog*. 1992;139:43–53.
18. Wheat CG, Jannasch HW, Kastner M, Hulme S, Cowen J, Edwards KJ, et al. Fluid sampling from oceanic borehole observatories: design and methods for CORK activities (1990–2010). Tokyo: Integrated Ocean Drilling Program; 2011.
19. Edwards KJ, Bach W, Klaus A, the Expedition 336 Scientists. Mid-Atlantic Ridge Microbiology: initiation of long-term coupled microbiological, geochemical, and hydrological experimentation within the seafloor at North Pond, western flank of the Mid-Atlantic Ridge. *Proc IODP*. 2012; 336: Tokyo (Integrated Ocean Drilling Program Management International, Inc.). <https://doi.org/10.2204/iodp.proc.336.2012>.
20. Cowen JP, Copson DA, Jolly J, Hsieh C-C, Lin H-T, Glazer BT, et al. Advanced instrument system for real-time and time-series microbial geochemical sampling of the deep (basaltic) crustal biosphere. *Deep-Sea Res*. 2012;61:43–56.
21. Robador A, La Rowe DE, Jungbluth SP, Lin H-T, Rappé MS, Neelson KH, et al. Nanocalorimetric characterization of microbial activity in deep subsurface ocean crustal fluids. *Front Microbiol*. 2016;7:454.
22. Zhou J, Bruns MA, Tidje JM. DNA recovery from soils of diverse composition. *Appl Environ Microbiol*. 1996;62:316–22.
23. Crump BC, Kling GW, Bahr M, Hobbie JE. Bacterioplankton community shifts in an arctic lake correlate with seasonal changes in organic matter source. *Appl Environ Microbiol*. 2003;69:2253–68.
24. Wingett SW, Andrews S. FastQ Screen: a tool for multi-genome mapping and quality control. Version 2 F1000Res. 2018;7:1338.
25. Minoche AE, Dohm JC, Himmelbauer H. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol*. 2011;12:R112.
26. Eren AM, Vineis JH, Morrison HG, Sogin ML. A filtering method to generate high quality short reads using Illumina paired-end technology. *PLoS ONE*. 2013;8. <https://doi.org/10.1371/annotation/afa5c40d-c604-46ae-84c4-82cb92193a5e>.
27. Peng Y, Leung HC, Yiu SM, Chin FY. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. 2012;28:1420–8.
28. Mikheenko A, Saveliev V, Gurevich A. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics*. 2016;32:1088–90.
29. Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, et al. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res*. 2019;47:D666–D677.
30. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, et al. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). *Stand Genom Sci*. 2015;10:86.
31. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34:525–7.
32. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl Acids Res*. 2013;41:D590–D596.
33. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*. 2012;28:3211–7.
34. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335–6.
35. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26:2460–1.
36. Graham ED, Heidelberg JF, Tully BJ. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. *PeerJ*. 2017;5:e3035.
37. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res*. 2014;25:1043–55.

38. Darling AE, Jospin G, Lowe E, Matsen FAIV, Bik HM, Eisen JA. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*. 2014;2:e243.
39. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. 2019;36:1925–7.
40. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol*. 2018;36:566–9.
41. Price MN, Dehal PS, Arkin AP. FastTree2 – approximately maximum-likelihood trees for large alignments. *PLoS ONE*. 2010;5:e9490.
42. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*. 2019;47:W256–W259.
43. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019;47:D590–D595.
44. Neely CJ, Graham ED, Tully BJ. MetaSanity: an integrated, customizable microbial genome evaluation and annotation pipeline. *bioRxiv* 2019. <https://doi.org/10.1101/789024>.
45. Garber AI, Nealon KH, Okamoto A, McAllister SM, Chan CS, Barco RA, et al. FeGenie: a comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol*. 2020;11:37.
46. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinform*. 2006;22:1540–2.
47. Schmidt-Schierhorn K, Kaul N, Stephan S, Villinger H. Geophysical site survey results from North Pond (Mid-Atlantic Ridge). In Edwards KJ, Bach W, Klaus A, the Expedition 336 Scientists (eds). *Proc IODP 2012*; 336: 107 (Integrated Ocean Drilling Program Management International, 2012).
48. Morin RH, Hess AE, Becker K. In situ measurements of fluid flow in DSDP Holes 395A and 534A: results from the DIANAUT Program. *Geophys Res Lett*. 1992;19:509–12.
49. Langseth MG, Becker K, Von Herzen RP, Schultheiss P. Heat and fluid flux through sediment on the western flank of the Mid-Atlantic Ridge: a hydrogeological study of North Pond. *Geophys Res Lett*. 2012;19:517–20.
50. Guttenplan SB, Kearns DB. Regulation of flagellar motility during biofilm formation. *FEMS Microbiol Rev*. 2013;37:849–71.
51. Belas R. When the swimming gets tough, the tough form a biofilm. *Mol Microbiol*. 2013;90:1–5.
52. Dressaire C, Neves Moreira R, Barahona S, Alves de Matos AP, Arraiano CM. BofA is a transcriptional switch that turns off motility and turns on biofilm development. *mBio*. 2015;6:e02352–14.
53. Wu DC, Zamorano-Sánchez D, Pagliai FA, Park JH, Floyd KA, Lee CK, et al. Reciprocal c-di-GMP signaling: Incomplete flagellum biogenesis triggers c-di-GMP signaling pathways that promote biofilm formation. *PLoS Genet*. 2020;16:e1008703.
54. Carr SA, Jungbluth SP, Eloe-Fadros EA, Stepanauskas R, Woyke T, Rappé MS, et al. Carboxydrotrophy potential of uncultivated Hydrothermarchaeota from the seafloor crustal biosphere. *ISME J*. 2019;13:1457–68.
55. Weiner L, Model P. Role of an *Escherichia coli* stress-response operon in stationary-phase survival. *Proc Natl Acad Sci*. 1994;91:2191–5.
56. Etchegaray J-P, Inoué M. CspA, CspB, and CspG, major cold shock proteins of *Escherichia coli*, are induced at low temperature under conditions that completely block protein synthesis. *J Bacteriol*. 1999;181:1827–30.
57. Suzuki Y, Yamashita S, Kouduka M, Ao Y, Mukai H, Mitsunobu S, et al. Deep microbial proliferation at the basalt interface in 33.5–104 million-year-old oceanic crust. *Commun Biol*. 2020;3:1–9.
58. Li J, Mara P, Schubotz F, Sylvan JB, Burgaud G, Klein F, et al. Recycling and metabolic flexibility dictate life in the lower oceanic crust. *Nature*. 2020;579:250–5.
59. Swan BK, Martinez-Garcia M, Preston CM, Szczyrba A, Woyke T, Lamy D, et al. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science*. 2011;333:1296–1300.
60. Anantharaman K, Breier JA, Sheik CS, Dick GJ. Evidence for hydrogen oxidation and metabolic plasticity in widespread deep-sea sulfur-oxidizing bacteria. *Proc Natl Acad Sci*. 2013;110:330–5.
61. Barco RA, Hoffman CL, Ramírez GA, Toner BM, Edwards KJ, Sylvan JB. In-situ incubation of iron-sulfur mineral reverse a diverse chemolithoautotrophic community and a new biogeochemical role for *Thiomicrospira*. *Environ Microbiol*. 2017;19:1–42.
62. Moses CO, Nordstrom DK, Herman JS. Aqueous pyrite oxidation by dissolved oxygen and by ferric iron. *Geochim Cosmochim Acta*. 1987;51:1561–71.
63. Bach W, Edwards KJ. Iron and sulfide oxidation within the basaltic ocean crust: implications for chemolithoautotrophic microbial biomass production. *Geochim Cosmochim Acta*. 2003;67:3871–87.
64. Bach W. Some compositional and kinetic controls on the biogeochemical landscapes in oceanic basement. *Front Microbiol*. 2016;7:107.
65. Edwards KJ, Rogers DR, Wirsén CO, McCollom TM. Isolation and characterization of novel psychrophilic, neutrophilic, Fe-oxidizing, chemolithoautotrophic α - and γ -Proteobacteria from the deep sea. *Appl Env Microbiol*. 2003;69:2906–13.
66. Chan CS, McAllister SM, Garber A, Hallahan BJ, Rozovsky S. Fe oxidation by a fused cytochrome-porin common to diverse Fe-oxidizing bacteria. *bioRxiv*. 2018. <https://doi.org/10.1101/228056>.
67. Maróti G, Tong Y, Yooseph S, Baden-Tillson H, Smith HO, Kovács KL, et al. Discovery of [NiFe] hydrogenase genes in metagenomic DNA: Cloning and heterologous expression in *Thiocapsa roseopersicina*. *Appl Env Microbiol*. 2009;75:5821–30.
68. Fritsh J, Scheerer P, Frielingsdorf S, Kroschinsky S, Friedrich B, Lenz O, et al. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature*. 2011;479:249–52.
69. Waldemar Finster K, Urup Kjeldsen K, Kube M, Reinhardt R, Mussman M, Amann R, et al. Complete genome sequence of *Desulfocapsa sulfexigens*, a marine deltaproteobacterium specialized in disproportionating inorganic sulfur compounds. *Stand Genom Sci*. 2013;8:58–68.
70. Beristain Cardoso R, Sierra-Alvarez R, Rowlette P, Razo Flores E, Gómez J, Field JA. Sulfide oxidation under chemolithoautotrophic denitrifying conditions. *Biotechnol Bioeng*. 2006;95:1148–57.
71. Di Capua F, Ahoranta SH, Papirio S, Lens PNL, Esposito G. Impacts of sulfur source and temperature on sulfur-driven denitrification by pure and mixed cultures of *Thiobacillus*. *Process Biochem*. 2016;51:1576–84.
72. Slobodkina GB, Mardanov AV, Ravin NV, Frolova AA, Chernyh NA, Bonch-Osmolovskaya EA, et al. Respiratory ammonification of nitrate coupled to anaerobic oxidation of elemental sulfur in deep-sea autotrophic thermophilic bacteria. *Front Microbiol*. 2017;8:87.
73. Høgslund S, Revsbech NP, Kuenen JG, Jørgensen BB, Gallardo VA, van de Vossenberg J, et al. Physiology and behaviour of marine *Thioploca*. *ISME J*. 2009;3:647–57.
74. Krul JM. Dissimilatory nitrate and nitrite reduction under aerobic conditions by an aerobically and anaerobically grown *Alcaligenes sp.* and by activated sludge. *J Appl Bacteriol*. 1976;40:245–60.

75. Wilson LP, Bouwer EJ. Biodegradation of aromatic compounds under mixed oxygen/denitrifying conditions: a review. *J Ind Microbiol Biot.* 1997;18:116–30.
76. Bianchi D, Weber TS, Kiko R, Deutsch C. Global niche of marine anaerobic metabolisms expanded by particle microenvironments. *Nat Geosci.* 2018;11:263–8.
77. Giblin AE, Tobias CR, Song B, Weston N, Banta GT, Rivera-Monroy VH. The importance of dissimilatory nitrate reduction to ammonium (DNRA) in the nitrogen cycle of coastal ecosystems. *Oceanography.* 2013;26:124–31.
78. Könneke M, Schubert DM, Brown PC, Hügler M, Standfest S, Schwander T, et al. Ammonia-oxidizing archaea use the most energy-efficient aerobic pathway for CO₂ fixation. *Proc Natl Acad Sci USA.* 2014;111:8239–44.
79. Qin W, Amin SA, Martens-Habben W, Walker CB, Urakawa H, Devol AH, et al. Marine ammonia-oxidizing archaeal isolates display obligate mixotrophy and wide ecotypic variation. *Proc Nat Acad Sci USA.* 2014;111:12504–9.
80. Seyler LM, McGuinness LM, Kerkhof LJ. Crenarchaeal heterotrophy in salt marsh sediments. *ISME J.* 2014;8:1534–43.
81. Seyler LM, McGuinness LR, Gilbert JA, Biddle JF, Gong D, Kerkhof LJ. Discerning autotrophy, mixotrophy and heterotrophy in marine TACK archaea from the North Atlantic. *FEMS Microbiol Ecol.* 2018;94:fy014.
82. Aylward FO, Santoro AE. Heterotrophic Thaumarchaea with small genomes are widespread in the dark ocean. *mSystems.* 2020;5:e00415–20.
83. Reiji L, Francis CA. Metagenome-assembled genomes reveal unique metabolic adaptations of a basal marine Thaumarchaeota lineage. *ISME J.* 2020;14:2105–15.
84. Lüscher S, Nowka B, Rattei T, Spieck E, Daims H. The genome of *Nitrospina gracilis* illuminates the metabolism and evolution of the major marine nitrite oxidizer. *Front Microbiol.* 2013;4:27.
85. Kitzinger K, Marchant HK, Bristow LA, Herbold CW, Padilla CC, Kidane AT, et al. Single cell analyses reveal contrasting life strategies of the two main nitrifiers in the ocean. *Nat Comm.* 2020;11:767.
86. Daims H, Nielsen JL, Nielsen PH, Schleifer K-H, Wagner M. In situ characterization of Nitrospira-like nitrite-oxidizing bacteria active in wastewater treatment plants. *Appl Environ Microbiol.* 2001;67:5273–84.
87. Koch H, Lüscher S, Albertsen M, Kitzinger K, Herbold C, Spieck E, et al. Expanded metabolic versatility of ubiquitous nitrite-oxidizing bacteria from the genus Nitrospira. *Proc Nat Acad Sci.* 2015;112:11371–6.