OPEN

The ISME Journal (2018) 12, 1–16
www.nature.com/ismej

ORIGINAL ARTICLE

A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer

Benjamin J Tully[1], C Geoff Wheat[2], Brain T Glazer[3] and Julie A Huber[4,5]
[1]*Center for Dark Energy Biosphere Investigations, University of Southern California, Los Angeles, CA, USA;* [2]*College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK, USA;* [3]*Department of Oceanography, University of Hawaii, Honolulu, HI, USA;* [4]*Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole, MA, USA and* [5]*Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA*

The rock-hosted subseafloor crustal aquifer harbors a reservoir of microbial life that may influence global marine biogeochemical cycles. Here we utilized metagenomic libraries of crustal fluid samples from North Pond, located on the flanks of the Mid-Atlantic Ridge, a site with cold, oxic subseafloor fluid circulation within the upper basement to query microbial diversity. Twenty-one samples were collected during a 2-year period to examine potential microbial metabolism and community dynamics. We observed minor changes in the geochemical signatures over the 2 years, yet the microbial community present in the crustal fluids underwent large shifts in the dominant taxonomic groups. An analysis of 195 metagenome-assembled genomes (MAGs) were generated from the data set and revealed a connection between litho- and autotrophic processes, linking carbon fixation to the oxidation of sulfide, sulfur, thiosulfate, hydrogen, and ferrous iron in members of the *Proteobacteria*, specifically the *Alpha*-, *Gamma*- and *Zetaproteobacteria*, the *Epsilonbacteraeota* and the *Planctomycetes*. Despite oxic conditions, analysis of the MAGs indicated that members of the microbial community were poised to exploit hypoxic or anoxic conditions through the use of microaerobic cytochromes, such as *cbb*$_3$- and *bd*-type cytochromes, and alternative electron acceptors, like nitrate and sulfate. Temporal and spatial trends from the MAGs revealed a high degree of functional redundancy that did not correlate with the shifting microbial community membership, suggesting functional stability in mediating subseafloor biogeochemical cycles. Collectively, the repeated sampling at multiple sites, together with the successful binning of hundreds of genomes, provides an unprecedented data set for investigation of microbial communities in the cold, oxic crustal aquifer.

The ISME Journal (2018) **12,** 1–16; doi:10.1038/ismej.2017.187; published online 3 November 2017

Introduction

The largest actively flowing aquifer system on Earth is circulating through oceanic crust underlying the oceans and sediments (Sclater *et al.*, 1980; Stein and Stein, 1994; Johnson and Pruis, 2003). The movement of water through the aquifer serves as a vital conduit for exchange of both microorganisms and nutrients between the ocean basins and the subseafloor and offers a route by which organisms can extract energy from the fluids and rocks beneath the seafloor (Orcutt *et al.*, 2013; Meyer *et al.*, 2016). Our understanding of life within the marine crustal aquifer has largely been shaped by studies of anaerobic and thermophilic organisms in warm ridge flank environments (Cowen *et al.*, 2003; Huber *et al.*, 2006; Jungbluth *et al.*, 2013, 2016) and crustal-source basalts exposed at the seafloor (Lysnes *et al.*, 2004; Mason *et al.*, 2009; Santelli *et al.*, 2009; Lee *et al.*, 2015). However, much of the microbial interaction with the crustal aquifer occurs within the seafloor at sites where cold, oxygenated deep ocean waters circulate through basaltic crust, entering and exiting through seafloor exposures (Fisher and Wheat, 2010; Edwards *et al.*, 2012; Wheat *et al.*, 2017). Therefore, despite advancing knowledge about microbial life in the subseafloor, our understanding is limited relative to which microorganisms live in the rocky oceanic crust, what hydrogeologic processes control subsurface fluid circulation, how these organisms harness energy in this environment, and the overall contribution to marine biogeochemical cycles is limited.

Correspondence: BJ Tully or JA Huber, Center for Dark Energy Biosphere Investigations, University of Southern California, 3616 Trousdale Parkway, Los Angeles, CA 90089, USA.
E-mail: tully.bj@gmail.com or jhuber@whoi.edu

Received 27 July 2017; revised 19 September 2017; accepted 22 September 2017; published online 3 November 2017

To more effectively study these prevalent ocean environments, several subseafloor observatories, termed circulation obviation retrofit kits (CORKs; Davis *et al.*, 1992; Wheat *et al.*, 2011), have been deployed in oceanic crust in part to allow for sampling and monitoring of the crustal aquifer (Wheat *et al.*, 2011). Two CORK observatories are installed at the well-studied site North Pond, an isolated sediment basin (8 km × 15 km, ~4484 m water depth), just west of the Mid-Atlantic ridge on 7–8 million years old crust (22°45′ N, 46°05′ W; Edwards *et al.*, 2010). At North Pond, seawater circulates between the crust and the deep ocean through the exposed ridge flanks, while sediments within the basin act as an impermeable barrier that prevents seawater exchange.

Previous studies have sought to constrain the microbial community and its activity within the basaltic aquifer at North Pond. Measurements of carbon fixation activity on basalts recovered by ocean drilling (Orcutt *et al.*, 2015) were unable to detect quantifiable rates of activity at *in situ* temperatures (4°C), while additions of nitrate and ammonia to crustal rocks stimulated microbial growth (Zhang *et al.*, 2016). Modeling of the subsurface at North Pond suggests that hydrogen and ferrous iron likely have an important role in maintaining microbial biomass, with ferrous iron estimated to support ~10% of the microbial biomass (Bach, 2016). In support of this hypothesis, a *Marinobacter* isolate capable of iron oxidation was enriched from North Pond basalts (Zhang *et al.*, 2016). PCR-based assessments of the microbial community associated with the basalts from North Pond have shown that *Gammaproteobacteria* are the dominant phylogenetic group (Jørgensen and Zhao, 2016), while the presence of genes involved in the carbon fixation through the Calvin-Benson-Bassham cycle are more common than the reverse citric acid cycle (Orcutt *et al.*, 2015).

Additional work examining the crustal fluids from the aquifer at North Pond has shown that the geochemistry of the fluids is nearly identical to the deep Atlantic bottom water (DABW), indicating a short residence time for seawater within the crustal aquifer at North Pond (Meyer *et al.*, 2016). However, basaltic formation fluids within the aquifer have concentrations of dissolved oxygen, silica and dissolved organic carbon that are different than those of the deep bottom water (Meyer *et al.*, 2016), and

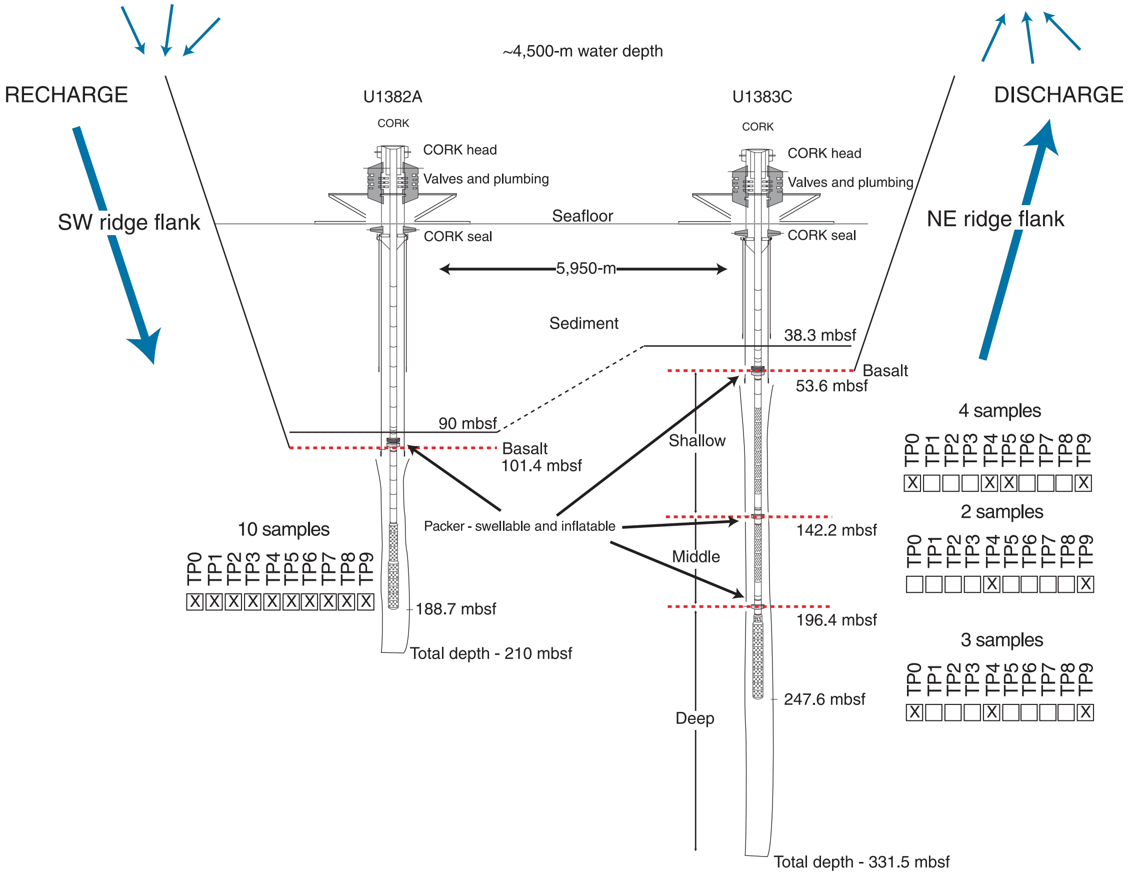

Figure 1 Idealized schematic of North Pond, CORK U1382A and CORK U1383C. Hypothesized flow of entrained seawater through the crust is represented by blue arrows. Seafloor and sediment boundaries are represented by black line. Basalt boundaries are represented by red, dashed lines. For each CORK horizon, the number of metagenomic samples are indicated, including the relative time sampled. CORK, continuous obviation retrofit kits; mbsf, meters below seafloor; TP, time point. (Modified from Edwards *et al.*, 2010).

Table 1 Collection details, cell enumeration and inorganic chemistry values for North Pond samples

CORK/bottom water	*Depth horizon*	*Date sampled*	*Time point*	*Cell counts (cells ml⁻¹ ± 95% confidence level)*	*O_2 (μmol l⁻¹)*	*NO^{3-} (μmol l⁻¹)*	*Si (μmol kg⁻¹)*
1382A	90–210	25 Apr 2012	TP0	1.4×10^4 ($\pm 6 \times 10^2$)	244 ± 1	21.1	56
1382A	90–210	6 Aug 2012	TP1	1.1×10^4 ($\pm 1.0 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	5 Oct 2012	TP2	8.1×10^3 ($\pm 1.1 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	4 Dec 2012	TP3	9.2×10^3 ($\pm 8.3 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	2 Feb 2013	TP4	N.d.	N.d	N.d	N.d
1382A	90–210	3 Apr 2013	TP5	1.5×10^4 ($\pm 6.9 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	2 Jun 2013	TP6	6.2×10^3 ($\pm 5.0 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	1 Aug 2013	TP7	1.8×10^4 ($\pm 1.1 \times 10^4$)	N.d	N.d	N.d
1382A	90–210	30 Sept 2013	TP8	5.1×10^3 ($\pm 2.5 \times 10^3$)	N.d	N.d	N.d
1382A	90–210	5 Apr 2014	TP9	2.8×10^4 ($\pm 3.8 \times 10^2$)	233 ± 1	21.9	73
1383C	70–146	30 Apr 2012	TP0	2.0×10^4 ($\pm 7 \times 10^2$)	216 ± 1	21.8	125
1383C	70–146	9 Apr 2013	TP4	N.d.	N.d	N.d	N.d
1383C	70–146	8 Jul 2013	TP5	7.7×10^3 ($\pm 9.8 \times 10^2$)	N.d	N.d	N.d
1383C	70–146	2 Apr 2014	TP9	1.1×10^4 ($\pm 1.0 \times 10^3$)	202 ± 2	22.3	146
1383C	146–200	9 Apr 2013	TP4	5.6×10^3 ($\pm 3.0 \times 10^2$)	N.d	N.d	N.d
1383C	146–200	8 Apr 2014	TP9	N.d.	185 ± 2	21.9	123
1383C	200–332	20 Apr 2012	TP0	2.1×10^4 ($\pm 8 \times 10^2$)	213 ± 1	21.8	120
1383C	200–332	9 Apr 2013	TP4	5.0×10^3 ($\pm 4.0 \times 10^2$)	N.d	N.d	N.d
1383C	200–332	31 Mar 2014	TP9	2.8×10^4 ($\pm 3.8 \times 10^2$)	187 ± 1	21.7	158
Bottom water (CTD)	~ 4400 m	26 Apr 2012	TP0	2.1×10^4 ($\pm 5 \times 10^2$)	~ 250	21.1	48
Bottom water (CTD)	~ 4400 m	10 Apr 2014	TP9	1.9×10^4 ($\pm 8.0 \times 10^2$)	~250	21.1	55

Abbreviation: N.d., not determined.

assessment of the crustal fluid microbial community through 16S rRNA gene and transcript sequencing, stable-isotope incubations, and metagenomics revealed that the aquifer community was active with a distinct community structure from bottom water. The community also had the capacity to perform both autotrophy and heterotrophy (Meyer *et al.*, 2016), with low rates of activity detected using nanocalorimetry (Robador *et al.*, 2016).

Together, these initial studies show a diverse and distinct microbial community living in the oligotrophic, oxic, basaltic crustal aquifer at North Pond with relatively low levels of metabolic activity. However, little is known about the metabolic potential and community dynamics in this understudied environment. Here, we present genomic reconstruction of North Pond crustal fluid samples collected over a span of two years, providing 21 samples for a detailed examination of potential microbial metabolism and community interactions within this subseafloor aquifer. Our high-resolution analysis of hundreds of genomes reveals a temporally and spatially dynamic microbial community and provides new insights into microbially-mediated biogeochemical cycling within the crustal aquifer.

Materials and methods

Sampling, cell quantification and chemical analysis

Crustal fluids were collected from the single horizon at U1382A and from the shallow, middle and deep horizons in U1383C (Edwards *et al.*, 2012) using a mobile pumping system designed for microbial sampling from CORK fluid delivery lines as described in Meyer *et al.* (2016) and Cowen *et al.* (2012; Figure 1). Deployed with the ROV system, mobile pumping system connectors are attached to the CORK wellhead via an umbilical to the hydrological zone of interest within the aquifer. Fluid systems were flushed and allowed to equilibrate before sampling, and dissolved oxygen concentrations were measured during pumping using an Aanderaa sensor (Meyer *et al.*, 2016). In 2012, 12 l of each fluid sample were filtered on to a 0.22 μm Sterivex-GP filter (Merck Millipore, Billerica, MA, USA) as described in Meyer *et al.* (2016). In 2014, 12 l of each sample was filtered *in situ* and immediately fixed with RNALater (Thermo Fisher Scientific, Waltham, MA, USA), as described previously (Akerman *et al.*, 2013). After sampling in 2012, a battery-powered GeoMICROBE sled was left at each CORK for time series autonomous sampling of the fluid delivery lines (Cowen *et al.*, 2012). For each filter sample, ~ 10 l of fluid were filtered *in situ* and immediately fixed with RNALater. For downstream analysis, ~ 500 ml of fluid were filtered into two Tedlar bags, one containing 54 ml of 37% formaldehyde for cell enumeration and the other with 4 ml of 10% HCl for inorganic chemistry analyses. Sleds were deployed in April 2012 and recovered in April 2014 with samples collected according to Table 1. Upon sled recovery, filters were transferred to fresh RNALater and stored at − 80 °C, while all bag samples were stored at 4 °C (Cowen *et al.*, 2012). Deep bottom water was sampled in 2012 and 2014 via a CTD at 100 m above the seafloor and filtered in the same manner as the crustal fluids onto Sterivex filters. Total microbial biomass in fluids was enumerated with DAPI (4′,6′-diamidino-2-phenylindole; Sigma-Aldrich, St Louis, MO, USA) and epifluorescent microscopy (Porter and Feig, 1980). Fluids also were analyzed for dissolved silicon and nitrate using automated colorimetric analysis and pH was measured with an electrode before a potentiometric titration for the determination of alkalinity (Wheat *et al.*, 2017).

DNA extraction and sequencing

Total genomic DNA was extracted from the filters using a phenol chloroform method, as previously described (Sogin *et al.*, 2006). DNA was sheared to 175 bp using a Covaris S-series sonicator. Metagenomics libraries were constructed using the Ovation Ultralow Library DR multiplex system (Nugen) following manufacturer's instructions. Paired-end sequencing was performed on an Illumina HiSeq 1000 at the WM Keck sequencing facility at the Marine Biological Laboratory. Raw sequence reads underwent quality control using Cutadapt (Martin, 2012; v.1.7.1; -e 0.08 --discard-trimmed --overlap = 3) to locate and remove Illumina adapter sequences from both ends of the of the read, followed by quality trimming using Trimmomatic (Bolger *et al.*, 2014; v.0.33 ; PE SLIDINGWINDOW:10:28 MINLEN:75).

Ribosomal rRNA identification and relative abundance

From the high-quality paired-end Illumina sequencing reads, 16S rRNA gene fragments were identified using Meta-RNA (Huang *et al.*, 2009; v.H3; -e 1e-10). Putative rRNA fragments and associated mate pairs from each sample were processed through EMIRGE (Miller *et al.*, 2011, 2013); emirge_amplicon.py; -l 113 -i 163 -s 33 -a 32 --phred33) to generate full-length sequences using the SILVA (Quast *et al.*, 2012) SSURef111 reference database (https://github.com/csmiller/EMIRGE). Reconstructed 16S rRNA genes were assigned taxonomy using mothur (v1.34.4) by first aligning the sequences to the SILVA SSURef123 database (align.seqs; flip = T), removing sequences that failed to align, if necessary (remove.seqs), and classifying the sequences (classify.seqs; cutoff = 80, iters = 1000).

Utilizing the high-quality sequence reads, each set of 16S rRNA sequences was used to recruit reads from the corresponding metagenomic sample, randomly subsampled using seqtk (v1.0-r82; https://github.com/lh3/seqtk) to the size of the smallest library (n = 22 142 100 reads). Reads were recruited using Bowtie2 (v.2.2.5; default parameters) and individual counts of reads per 16S rRNA were determined. Read counts were length normalized and used to calculate the relative abundance of each reconstructed 16S rRNA in the sample (Supplementary Data 15). Relative abundances were combined for sequences that shared the same mothur-ascribed Phylum (or Class for Proteobacteria) level.

Metagenomic assembly and binning

High-quality sequence reads were subjected to two rounds of assembly. A primary set of contigs was generated using IDBA-UD (Peng *et al.*, 2012; v.1.1.1; default parameters) utilizing the reads from each individual sample. A secondary set of contigs was generated in Geneious (Kearse *et al.*, 2012) v6.1.8; modified parameters used for 'High Sensitivity/Slow'; Supplementary Data 11) by combining the primary set of contigs ⩾500 bp in length from samples with the same source (that is, combining all primary contigs generated from U1382A, and so on). Secondary contigs ⩾5 kb in length from U1382A and U1383C were combined with secondary contigs ⩾3 kb generated from DABW (Supplementary Data 4).

The size-selected set of secondary contigs was used to recruit high-quality sequencing reads from each sample using Bowtie2 (as above). A coverage value, equivalent to recruited reads per bp, was determined for each contig in each sample, the coverage values were log(n + 1) transformed, and subjected to binning using affinity propagation (Frey and Dueck, 2007) and a pre-release version of BinSanity (Graham *et al.*, 2017; -p -1). CheckM (Parks *et al.*, 2015; v1.0.3; lineage_wf) was used to assess the results of the binning utilizing a ⩾50% completeness threshold to identify putative genomes. Multiple bins were identified above the completeness threshold that contained substantial estimated contamination (>55% contamination). For each suspect bin, the %G+C and coverage values for each contig were plotted against each other (data not shown) and manually assessed for putative cohesive groups (Supplementary Data 12).

Phylogeny

Each putative genome was assessed for the presence of 16 conserved ribosomal marker proteins (Hug *et al.*, 2016) based on a HMMER (Finn *et al.*, 2011) search (v.3.1b2; hmmsearch --cut_tc --notextw) of TIGRfam (Haft *et al.*, 2003) and Pfam (Bateman *et al.*, 2002) models corresponding to the proteins (Supplementary Data 13). If multiple copies of a ribosomal markers protein were detected, that protein was not included as a marker for that genome, and any genome with <8 markers was not included for further phylogenetic assessment. Ribosomal markers were collected from 1652 reference genomes representing the major Families and/or Genera from within the Bacteria. Each marker gene from the putative and reference genomes was aligned using MUSCLE (Edgar, 2004; v3.8.31; -maxiters 8), trimmed using trimAL (Capella-Gutiérrez *et al.*, 2009; v.1.2rev59; -automated1), imported in Geneious (Kearse *et al.*, 2012), and manually assessed and trimmed, if necessary. All of the individual alignments were concatenated and a phylogenetic tree was constructed using FastTree (Price *et al.*, 2010; v2.1.3; -lg -gamma).

Genomes were assessed for the presence of full-length 16S rRNA genes utilizing RNAmmer (Lagesen *et al.*, 2007; v1.2; -S bac -m ssu). The identified rRNA sequences were aligned to the SILVA SSURef123 database using the web-based SINA aligner (Pruesse *et al.*, 2012); default setting, trailing sequences removed from alignment). Aligned rRNA sequences were added to the SSURef123 NR99 ARB tree (Ludwig *et al.*, 2004) using the ARB Parsimony (Quick) tool (default parameters). Members of the ARB tree phylogenetically related to the North Pond genome rRNA sequences were extracted. ARB-based and North Pond genome 16S rRNA sequences we re-aligned, trimmed, and manually assessed (as above). The final alignment was used to construct a phylogenetic tree with FastTree (-nt -gtr -gamma). Within the genomes, 53 16S rRNAs were identified within 47 genomes. The 16S rRNA gene tree was (Supplementary Data 14) used to support or refute the assignments provided from the ribosomal marker tree (RMT) and/or CheckM (Supplementary Data 16).

Annotation and metabolic analysis

Putative CDS were predicted for the North Pond genomes using Prodigal (Hyatt *et al.*, 2012; v2.6.3; -m -p meta -q) and submitted to the GhostKoala (Kanehisa *et al.*, 2016; default parameters; genus_prokaryotes + family_eukaryotes) for annotation using the KEGG (Kanehisa *et al.*, 2016) Ontology (KO) system. Based on these KO assignments, genomes were assessed for the degree to which specific pathways and functions were complete in the individual genomes using the information on canonical pathways available as part of the KEGG Pathway Database (updated, 14 Nov 2016) and the script KEGG-decoder.py (www.github.com/bjtully/BioData/tree/master/KEGGDecoder).

Beyond specific assignments to pathways and function, for this manuscript several broad functional metabolic categories were identified for genomic bins based on the presence multiple genes. Cytochromes that participate in oxygen chemistry in aerobic organisms were defined as cytochrome *c* oxidase, *aa*$_3$-type (*coxABCD*) and cytochrome *o* ubiquinol oxidase (*cyoABCD*), while microaerobic cytochrome metabolism was defined as cytochrome *c* oxidase, *cbb*$_3$-type (*ccoPQNO*) and cytochrome *bd* complex (*cydAB*; (García-Horsman *et al.*, 1994). Similarly, sulfide oxidation was determined by the presence of sulfide:quinone oxidoreductase (*sqr*), sulfur dioxygenase (*sdo*), and/or sulfite reductase (*dsrA*), when applicable (see below). Additionally, putative thiosulfate oxidation was assessed based on components of the SOX system (*soxABCXYZ*) and/or thiosulfate dehydrogenase (*tsdA*).

Putative CDS annotated as the large subunit of ribulose-1,5-bisphosphate carboxylase (RuBisCO; K01601) were extracted, along with RuBisCO sequences representing previously described major lineages (Tabita *et al.*, 2007; Supplementary Data 17). The RuBisCO sequences were aligned, automatically trimmed, and used to construct a phylogenetic tree (as above). A similar procedure was applied to identifying molybdopterin oxidoreductases (MOBs), specifically to identify MOBs associated with Fe^{2+} oxidation. Putative MOBs were identified from the Prodigal-derived CDSs via HMMER (hmmsearch, bit score threshold $\geqslant 75$) using the molybdopterin Pfam (PF00384). Environmental MOBs were aligned with reference sequences (Supplementary Data 6), automatically trimmed, and used to construct a phylogenetic tree (as above). To differentiate between sulfite reductases present in sulfur reducing organisms and reverse sulfite reductase present in sulfur oxidizing organisms, putative CDS annotated as DsrA (K11180) were used to construct a phylogenetic tree (as above) with reference sequences (Loy *et al.*, 2009) (Supplementary Figure 6). Two HMM models designed using homologs of Cyc1$_{\text{PV-1}}$ and Cyc2$_{\text{PV-1}}$ identified in neutrophilic iron oxidizing organisms (Barco *et al.*, 2015; Tully and Heidelberg, 2016) were used to search via HMMER (hmmsearch, bit score threshold Cyc1$_{\text{PV-1}}$ $\geqslant$ 445 and Cyc2$_{\text{PV-1}}$ $\geqslant$ 55) the putative CDSs of North Pond genomes (Supplementary Data 7).

Relative abundance and community structure

Contigs composing the North Pond genomes were used to recruit high-quality sequencing reads using Bowtie2 (default parameters) from the subsampled metagenomic samples (as above). Read counts for each genome were determined using featureCounts (Liao *et al.*, 2014; v.1.5.0-p2; -F SAF) and normalized to reads per bp for the full-length of each genome. Length normalized relative abundance was determined for each sample:

$$\frac{\frac{Reads}{bp}\ per\ genome}{\sum \frac{Reads}{bp}\ all\ genome} \times \frac{\sum Recruited\ reads\ to\ genomes}{22,142,100\ reads} \times 100$$

The relative abundance values for the genomes in all 21 sample were used to cluster the genomes and samples in Past3 using the Bray-Curtis similarity measure (Supplementary Figure 9). A separate Bray-Curtis clustering step was performed only on genomes detected in the U1382A samples above 0.05% relative abundance (Supplementary Figure 8). The observed clusters of genomes were used to determine ecological units that were observed over the course of the time series sampling. Annotations for genomes in the identified U1382A ecological units were used to predict potential function. Genomes from U1382A were assigned to six functions: carbon fixation, partial denitrification (functional assignment predicts incomplete set of denitrification genes), complete denitrification, DNRA, sulfide oxidation, partial dissimilatory sulfur redox (functional assignment predicts incomplete set of sulfur redox genes), and thiosulfate oxidation. The

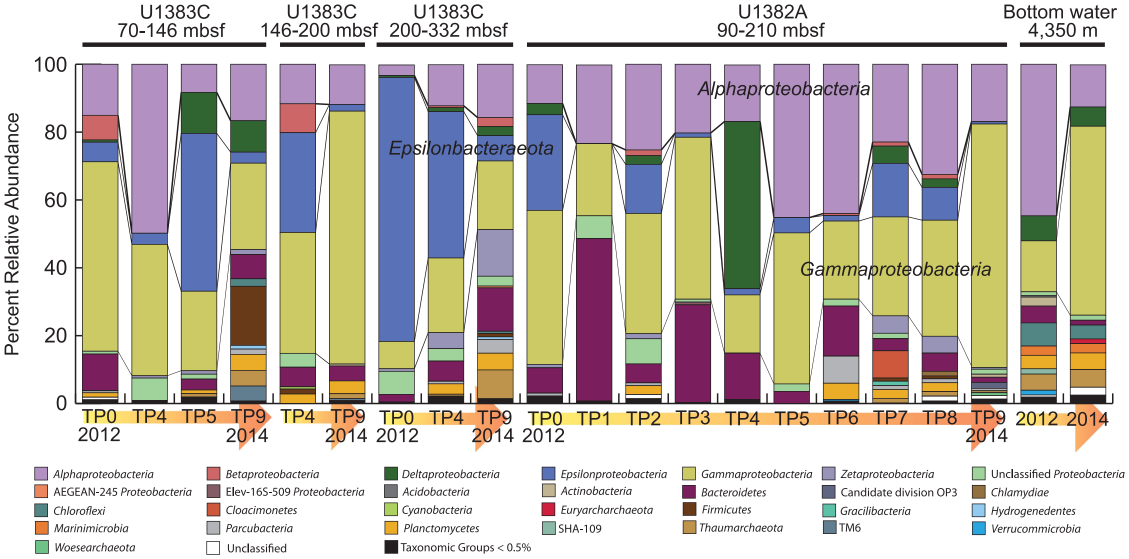

Figure 2 Microbial community structure in North Pond crustal fluids based on reconstructed full-length 16S rRNA gene sequences. Relative abundance for 16S rRNA gene sequences was combined at the Phyla (Class level for *Proteobacteria*) from holes U1383C and U13832A and the Deep Atlantic Bottom Water. Individual samples are grouped based on depth of sampling (shallow, middle, and deep in U1383C) and time (from initial sampling in 2012 to 2014 sampling time point).

fraction of the observed community with a given function was determined by:

$$\frac{\sum \textit{relative abundance of genomes with function X}}{\sum \textit{total relative abundance of genomes in time point}}$$

This process was repeated for the complete set of genomes to compare function between U1382A, U1383C and the DABW. Genomes with relative abundance >0.05% were considered for this analysis and could be assigned to multiple samples. Instead of broad functions, as for U1382A ecological units, genome counts and fractional abundance were assigned based on the presence of key genes in the nitrogen, sulfur and carbon cycles.

Results

Assessment of microorganisms in the subseafloor aquifer

The cold, oxic Mid-Atlantic subseafloor aquifer was sampled for geochemistry, cell quantification and microbial DNA from two seafloor CORK installations at the North Pond site. Water samples were collected 10 times from hole U1382A over the course of two years, and Hole U1383C was sampled 9 times from three different depth horizons that had been sealed during CORK installation. These horizons are defined by packers that seal the borehole and limit vertical mixing, defining three distinct hydrologic zones based on formation properties (Edwards *et al.*, 2012; Figure 1; Supplementary Figure 1). Two additional background seawater samples were collected from Niskin bottles that were tripped ~50 m off bottom, above the water-sediment interface (~4450-m depth) in both 2012 and 2014. These samples proved a measure of bottom water properties using the same techniques employed on those from the crustal fluids (Figure 1).

Cell counts in all 19 borehole samples ranged from 5 to 20 × 10^3 cells ml^{-1} of crustal fluid, with no discernable change during the two year period (Table 1). Geochemical data from discrete samples collected in 2012 and 2014 indicated a minor increase in silica, whereas oxygen concentrations decreased slightly at all sampling horizons. Nitrate concentrations did not change.

In total, 21 metagenomic samples were sequenced, generating 1.2 billion high-quality paired-end Illumina sequencing reads (Supplementary Data 1). 2,829 approximately full-length 16S rRNA gene sequences were reconstructed from the data set (Supplementary Data 2). The full-length 16S rRNA gene sequences in the metagenome (Figure 2) provide a snapshot of community composition in the samples, revealing a temporally and spatially dynamic community, with large shifts in the relative abundance of *Proteobacteria*, specifically in the *Alpha*-, *Gamma*- and *Deltaproteobacteria*, the *Epsilonbacteraeota*, and the *Bacteroidetes*.

After two rounds of assembly of the high-quality sequencing reads, 1.5 million contigs were produced. A subsection of contigs ⩾5 kbp in length (78 004 contigs; N50 = 25 932 bp; total bp = 1.2 Gbp) were used to reconstruct microbial genomes (Supplementary Data 3). 195 metagenome-assembled genomes (MAGs) were reconstructed and determined to be ⩾50% complete (an additional 234 genome bins were identified that were 20–50% complete, though were not analyzed further). Throughout this manuscript, the term ‘genome’ will be used to refer to the 195 binned MAGs. The genomes were given the designation NORP, for **Nor**th **P**ond genome (NORP1-195). With the exception of two genomes (NORP4 and -5), all of the genomes had ⩽10% cumulative contamination/redundancy (Supplementary Data 3). The genomes recruited between 9–61% of the sequencing reads (mean = 37.3%) from the individual samples, with the lowest recruitment rate from the 2012 bottom water sample (Supplementary Data 4).

140 MAGs had a sufficient number (⩾8) of 16 ribosomal marker proteins to be included in a phylogenetic tree with genomes from IMG (Markowitz *et al.*, 2006) that represent the major bacterial Genera and/or Families (Supplementary Data 5). The North Pond genomes were assigned to 20 Phyla, including all the lineages within the *Proteobacteria* (including *Acidithiobacillia*), the Candidate Phyla Radiation (CPR; Hug *et al.*, 2016), and the *Planctomycetes* (Supplementary Figure 2).

Based on the relative abundance of sequencing reads competitively recruited to each genome from each sample, the mean relative abundance for all genomes in all samples was 0.19% (median, 0.004%), and when examined closely, most genomes were ‘present’ in all samples at low abundance values (<0.05%; on average 151 of 195 genomes were below this threshold in each sample). NORP9 had the highest relative abundance (40.4%) in the 2012 U1383C deep sample (Figure 3; Supplementary Data 9). Genomes were subjected to Bray-Curtis clustering based on the relative abundance values (Supplementary Figure 3). Several of the genomes (for example, NORP125, -161, and -172) were cosmopolitan in the subseafloor crustal fluids, present in both holes, and at several time points and depths (Figure 3). Most of the genomes associated with the bottom water samples were not present in the crustal samples, although several genomes did have low abundances; specifically NORP160 and -164, both assigned to the *Nitrosopumilales*, and three additional genomes detected in the 2014 samples from U1382A and the middle section of U1383C. Generally, when genomes were grouped together, these groups were abundant in one or a few samples (e.g., NORP51, -54 and -55). When groups of organisms are present in multiple samples, the samples tend to be in close spatial proximity or sequential sampling events (Figure 3). In several instances, a single organism becomes highly abundant, but is only present in a single sample (for example, NORP6 or NORP73).

Metabolic potential of metagenome-assembled genomes

From the genomes, 523,212 putative coding DNA sequences (CDSs) were identified, of which, 245,902 (47%) were annotated with a KEGG ontology (KO; Kanehisa *et al.*, 2016) number/function (Supplementary Figure 4). Genomes were assessed for the presence of specific KO functions involved in numerous processes, including: carbon, nitrogen, sulfur, and hydrogen cycling, methanogenesis, motility, vitamin biosynthesis and transport, and fermentation pathways. Two of the 6 establish carbon fixation pathways (Hügler and Sievert, 2011) were identified amongst the genomes, with 32 genomes containing ribulose-1,5-bisphosphate carboxylase (RuBisCO) and elements of the Calvin-Benson-Bassham (CBB) cycle, and 7 genomes containing ATP-citrate lyase or cirtyl-CoA synthetase and citryl-CoA lyase part of the reverse citric acid cycle (rTCA; Table 2). A phylogenentic analysis of the 32 genomes possessing putative RuBisCO proteins, identified 5 genomes with Form IV RuBisCO-like-proteins, 15 genomes with Form II RuBisCO, and 12 genomes with Form I RuBisCO (Supplementary Figure 5).

Analysis of the putatively carbon fixing genomes for possible electron donors revealed at least 5 sources: sulfide (HS^-), sulfur (S^0), thiosulfate, hydrogen (H_2) and ferrous iron (Fe^{2+}; Table 2). The putative electron donors of four of the genomes could not be identified. The most prevalent electron donor as indicated by the presence within the genomes was HS^-, as 25 genomes had the potential to utilize HS^- via either sulfide:quinone reductase (*sqr*) or reverse sulfite reductase (*rdsrA*; Supplementary Figure 6). Twenty-one genomes possessed either the components for the SOX system or thiosulfate dehydrogenase (*tsdA*), suggesting a potential for thiosulfate oxidation. The presence of NAD-reducing hydrogenase, capable of the reversible H_2 redox reactions, offers an avenue for H_2 as an electron donor in 5 of the genomes capable of carbon fixation. A single genome (NORP31) possessed a sulfur dioxygenase (*sdo*) that could utilize S^0 as an electron source. Lastly, 6 genomes were identified that may mediate the oxidation of Fe^{2+} linked to carbon fixation, based on the presence of dissimilatory Fe^{2+} molybdopterin oxidoreductase (Tully and Heidelberg, 2016; Act1B; Supplementary Figure 7; Supplementary Data 6) and/or Fe^{2+} reactive cytochromes (Barco *et al.*, 2015; $Cyc1_{PV-1}$, $Cyc2_{PV-1}$; Supplementary Data 7).

In assessing the potential for aerobic respiration, 20.5% of the genomes were determined to possess low-oxygen sensitivity (aerobic; *aa_3*- and/or *bo*-type) cytochromes, 12.3% contained high-oxygen sensitivity (microaerobic; *cbb_3*- and/or *bd*-type) cytochromes, and an additional 55.9% contained cytochromes for both aerobic and mircoaerobic oxygen metabolism (Supplementary Figure 4; Supplementary Data 8). An assessment of anaerobic metabolisms showed that 7.7% of genomes possessed the potential to perform complete denitrification (*nirK*/*nirS*, *norBC*, and *nosZ*). An additional 19.5% of genomes were annotated to have the potential to perform a single step in denitrification process (nitrite reduction, nitric oxide reduction, or nitrous-oxide reduction), while 9.2% of genomes could potentially perform only two of the three steps (Supplementary Figure 4). Further, one genome (NORP6) contained sulfite reductase, necessary for complete sulfate reduction (Supplementary Figure 6).

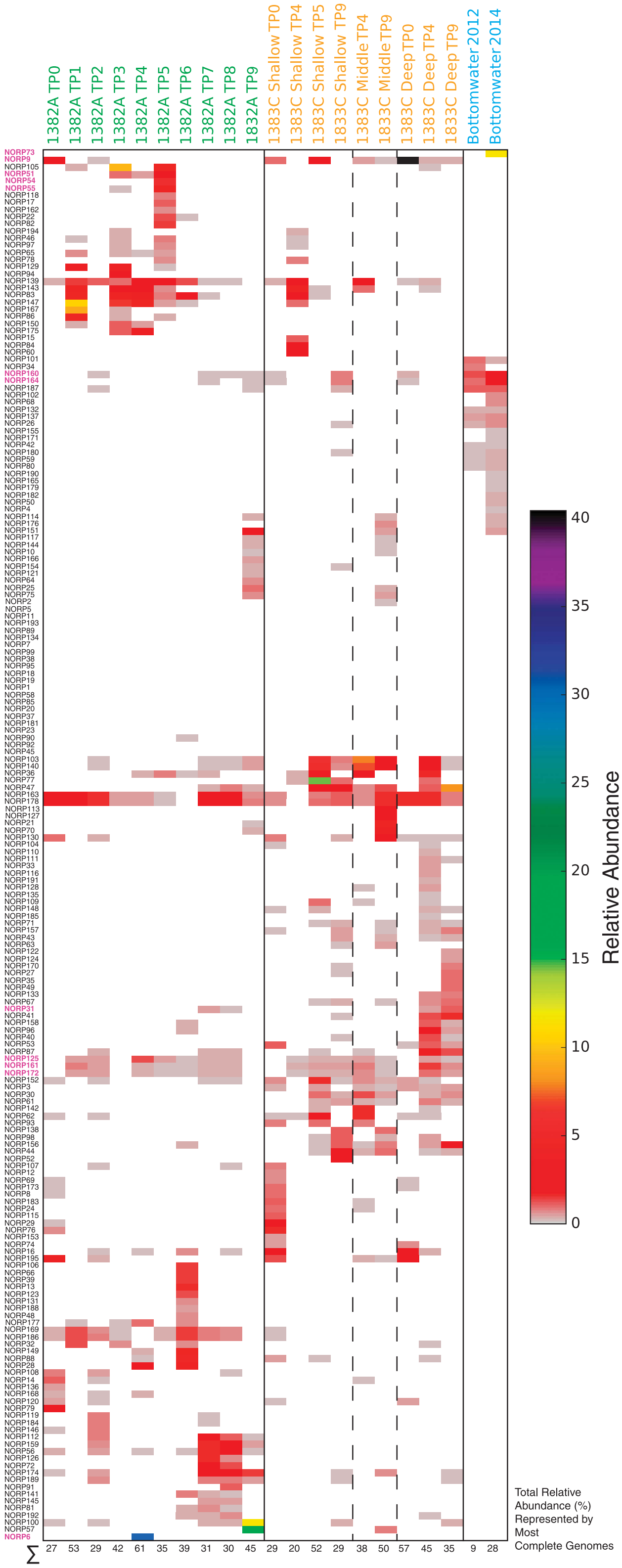

Figure 3 Heat map of the percent relative abundances for the 195 high-quality genomes. Genomes are organized and clustered based on the Bray-Curtis similarity index (Supplementary Figure 3). Genomes explicitly mentioned in text are highlighted in bold and colored purple. Values at the bottom of the column represent the total observed percent relative abundance of genomes in that sample.

Ecological units and community metabolic function

The genomes that had an appreciable presence (>0.05%, $n = 134$) in the U1382A samples were placed into 6 ecological units (Unit I–VI) representing 98 genomes that had distinguishing temporal patterns throughout the time series (Figure 4; Supplementary Figure 8). An additional ecological unit (Unit VII) consisted of eight genomes that were generally cosmopolitan in the U1382A samples. These ecological units represent a progression of community structure over the course of the time series, though in several instances members of an ecological unit re-occur in multiple time points (Figure 4). For example, Unit I consists of genomes originally sampled in TP0 that are not observed in TP1, but are observed 12 months later in the TP2 sample. This pattern can also be observed in Unit II (genomes in TP1 seen in TP3-5) and Unit V (genomes in TP2 seen in TP7-9) and supports the results of the sample clusters that indicate that TP1 and TP3-6 are more similar to one another than TP2 and TP7-8 (Supplementary Figure 9).

Each ecological unit possessed genomes capable of carbon fixation, partial and complete denitrification, DNRA, thiosulfate oxidation, and sulfur redox, with the exception of complete denitrification in Unit V and sulfur redox in Unit VII. However, analysis of the fraction of the observable community based on relative abundance with a specific functional potential changes considerably over time (Figure 5). During the time series, organisms capable of DNRA and complete denitrification are positively correlated (linear regression, $R^2 = 0.86$), which is reflected in the fact that most organisms capable of complete denitrification were also capable of DNRA, though the reciprocal was not true (Figure 5). This disparity between organisms with DNRA and complete denitrification likely explains why the fraction of DNRA capable community members was always greater than complete denitrification. Similarly, thiosulfate oxidation and sulfur redox processes were positively correlated (linear regression, $R^2 = 0.80$), though these functions generally occur in different genomes. The community fraction capable of sulfide oxidation tracks with the observed ecological units and sample clusters, as sulfide oxidation was the most prevalent sulfur pathway in TP1, 3–6, while thiosulfate oxidation was more prevalent in TP2, 7–9.

Comparison of the potential function of genomes assigned to U1382A, U1383C and DABW, the number of genomes with a predicted function, and the fraction of the observed community with that function indicates that genomes associated with the DABW do not possess the capability for carbon or nitrogen fixation, denitrification or sulfide oxidation (Figure 6). Based on relative abundance patterns, U1382A and U1383C are dominated by different microbial communities (Figure 3), but the number of genomes capable of the various steps in the nitrogen, sulfur, and carbon cycles do not vary (Figure 6). Further, there was no statistical difference (Student's

Table 2 Genomes with carbon fixation potential and putative electron sources

ID	*Phylogenetic Assignment*	*Carbon fixation pathway [CBB or rTCA] (RuBisCO Form)*	*Putative electron source (s)*	*Evidence*
NORP4	g_Methylophaga	CBB (IA+B)	H_2S	sqr
NORP17	g_Robiginitomaculum	CBB (II)	H_2S	sqr
NORP23	f_Rhodobacteraceae	CBB (II)	Thiosulfate, H_2S, H_2	soxABCXYZ, sqr, rdsrA, hoxHFUY
NORP24	f_Thiotrichaceae	CBB (IA+B)	Thiosulfate, H_2S	soxABXYZ, sqr, rdsrA
NORP31	c_Zetaproteobacteria	CBB (II)	Sulfur, H_2S, Fe^{2+}	sdo, sqr, $cyc1_{PV-1}$, $cyc2_{PV-1}$
NORP33	g_Methylophaga	CBB (IA+B)	H_2S, H_2	sqr, hoxHFUY
NORP48	g_Blastopirellula	CBB (II)	?	—
NORP54	g_Robiginitomaculum	CBB (IA+B)	H_2S	sqr
NORP55	g_Robiginitomaculum	CBB (IA+B)	H_2S	sqr
NORP56	g_Kangiella	CBB (II)	Thiosulfate, H_2S	soxABCXYZ, sqr
NORP60	—	CBB (II)	Thiosulfate, H_2S	soxCYZ, sqr
NORP65	g_Methylophaga	CBB (II)	Thiosulfate	soxCY
NORP78	f_Rhodobacteraceae	CBB (II)	Thiosulfate, H_2S	soxABXYZ, sqr, rdsrA
NORP93	g_Methylophaga	CBB (IA+B)	Thiosulfate, H_2S	soxABCXYZ, sqr
NORP100	f_Ectothiorhodospiraceae	CBB (IA+B, IC/D)	Thiosulfate, H_2S, Fe^{2+}(?)	soxABYZ, rdsrA, $cyc1_{PV-1}$
NORP103	f_Thiotrichaceae	CBB (IA+B)	H_2S, Fe^{2+}	sqr, actB1, $cyc1_{PV-1}$
NORP104	f_Methylophilaceae	CBB (IC/D)	?	soxY
NORP108	—	CBB (II)	Thiosulfate	soxABCXYZ
NORP109	g_Marinosulfonomonas	CBB (II)	Thiosulfate, H_2S, H_2	soxABCXYZ, sqr, rdsrA, hoxFUY
NORP110	g_Marinosulfonomonas	CBB (II)	Thiosulfate, H_2S	soxABCXYZ, sqr, rdsrA
NORP116	f_Rhodospirillaceae	CBB (II)	Thiosulfate, H_2S, Fe^{2+}(?)	rdsrA, actB1
NORP125	g_Robiginitomaculum	CBB (II)	H_2S	sqr
NORP128	f_Rhodospirillaceae	CBB (II)	Thiosulfate, H_2S, H_2, Fe^{2+}(?)	soxABXYZ, fccB, hoxHFUY, actB1
NORP169	o_Rhizobiales	CBB (IC/D)	?	—
NORP178	-	CBB (IA+B)	H_2S, Fe^{2+}	sqr, actB1, $cyc1_{PV-1}$
NORP181	f_Rhodobacteraceae	CBB (II)	Thiosulfate, H_2S	soxABCXYZ, sqr, rdsrA
NORP192	o_Rhizobiales	CBB (IC/D)	Thiosulfate	soxABCXYZ
NORP9	g_Sulfurimonas	rTCA	Thiosulfate, H_2S	soxABCXYZ, sqr
NORP14	g_Acrobacter	rTCA	Thiosulfate, H_2S	soxABCXYZ, tsdA, sqr
NORP62	g_Sulfurovum	rTCA	H_2	soxC, hoxHFUY
NORP87	g_Sulfurimonas	rTCA	Thiosulfate, H_2S	soxCYZ, sqr
NORP112	g_Sulfurimonas	rTCA	?	—
NORP168	g_Sulfurimonas	rTCA	Thiosulfate, H_2S	soxCY, sqr
NORP195	g_Sulfurimonas	rTCA	Thiosulfate	soxABCXYZ

Abbreviations: actB1, dissimilatory Fe^{2+} molybdopterin oxidoreductase; $cyc1_{PV-1}$, cytochrome c_4, cbb_3-type; CBB, Calvin-Benson-Bassham cycle; hoxHFUY, NAD-reducing hydrogenase; rdsrA, reverse sulfite reductase; rTCA, reverse citric acid cycle; RuBisCO, ribulose-1,5-bisphosphate carboxylase; sqr, sulfide:quinone oxidoreductase; soxABCXYZ, thiosulfate oxidation subunits/components; sdo, sulfur dioxygenase; tsdA, thiosulfate dehydrogenase.

t-test and Wilcoxon rank sum, $P < 0.5$) between U1382A and U1383C based on the fraction of the observed community capable of an ascribed metabolic reaction, with the exception of ammonia oxidation (Student's *t*-test and Wilcoxon rank sum, $P = 0.005$).

Discussion

Despite being the largest actively flowing aquifer on Earth, our understanding of microbial communities and their role in biogeochemical cycling in subseafloor crustal fluids is largely unknown. The bulk of our understanding is from studies of fluids from warm environments, including the Juan de Fuca Ridge flank in the NE Pacific Ocean and hydrothermal vents around the globe (Takai and Horikoshi, 1999; Huber *et al.*, 2002; Reveillaud *et al.*, 2016). These environments are characterized by high temperature (25–80 °C), low-oxygen fluids that are usually dominated by mesophilic and (hyper)thermophilic microorganisms with microaerobic and anaerobic metabolisms (Cowen *et al.*, 2003; Huber *et al.*, 2006; Jungbluth *et al.*, 2013; 2016). This is in contrast to North Pond, which represents a common, but understudied type of ridge flank region, where circulating fluids are cold (4–15 °C) and oxygenated (Edwards *et al.*, 2012; Meyer *et al.*, 2016). Previous work at North Pond showed that the fluids in the basaltic crust have similar chemistry to the oceanic bottom water, but that the microbial community has a distinct population structure with potential for both heterotrophic and autotrophic activity (Meyer *et al.*, 2016). Using the increased temporal and spatial sampling offered by our metagenomic time series at North Pond, we verified that the microbial community composition of the crustal fluid samples is fundamentally different from the DABW, and extended this finding to microbial communities and their genomic functional potential using MAGs (Figures 2 and 5, Supplementary Data 10). Further, we also found that the microbial communities within the crustal fluids show shifts in the dominant phyla (and proteobacterial classes) over time within a single hole and between the two holes (Figure 2). *Gammaproteobacteria* are dominant in 10 of the crustal fluid samples, but several other phylogenetic groups, *Alpha*- and *Deltaproteobacteria*, *Epsilonbacteraeota* and *Bacteroidetes*, are abundant in other samples. The initial samples were collected in 2012, approximately six months after the holes were drilled and the CORK systems were installed, therefore it is possible that the observed shifts are due to the holes returning to a natural state after the perturbation of drilling, during which surface water is pumped into the borehole to clear cuttings, inevitably pumping surface waters into the formation. Such shifts in subseafloor crustal fluid community structure have been documented in samples collected shortly after drilling and for several years afterwards on the flanks of the Juan de Fuca Ridge, a younger, warmer crustal system (Jungbluth *et al.*, 2012, 2016), highlighting the importance of time series for understanding such ecosystems and potential stresses. However, the magnitude of chemical shifts observed in discrete samples collected in 2012 and 2014 suggests only minor changes in geochemistry, including a decrease in dissolved oxygen concentrations and increase in dissolved silica concentrations at all four sampling horizons. Increases in dissolved silica may result from either diffusive exchange with sediment pore waters or water-rock reactions at low temperatures, whereas the decrease in oxygen concentrations indicates continued consumption of oxygen (Ziebis *et al.*, 2012; Meyer *et al.*, 2016), such as that inferred from a similar cool ridge flank setting at Dorado Outcrop (Wheat *et al.*, 2017).

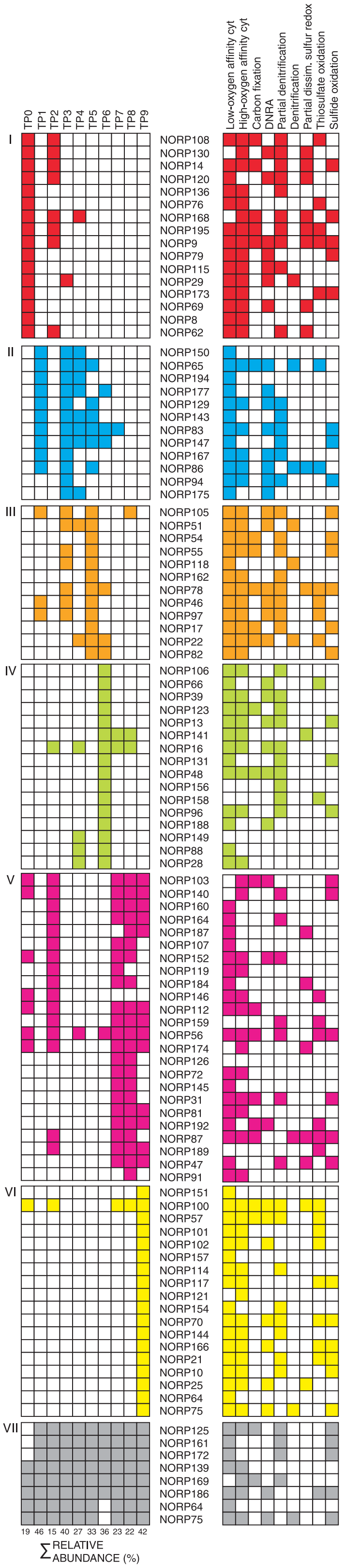

Figure 4 Ecological units and community metabolic function of U1382A. Presence (left column) and predicted function (right column) for genomes assigned to ecological units. Ecological units are ordered to illustrate the progression of community structure through time. Values at the bottom of the column represent the total relative abundance of genomes in ecological units for that time point. cyt, cytochrome; DNRA, dissimilatory nitrate reduction to ammonia; dissim., dissimilatory; TP, time point.

The high-resolution analysis, provided by the relative abundance of the reconstructed genomes, reveals that the microbial communities of U1382A, U1383C, and the DABW are composed of distinct MAGs (Figure 3). Importantly, genomes from the DABW form a cohesive group of organisms that were not present (or had a limited presence) in the crustal fluids, and conversely none of the crustal-originating genomes were detected in the DABW.

From these results, it is clear that the genomes we reconstructed represent residential subseafloor bacteria and archaea from North Pond crustal fluids, thus allowing for detailed examination of microbial metabolic functions and community dynamics and interactions within the North Pond crustal habitat. It is important to note, however, that the reconstructed genomes only represent a subset of the total micro-

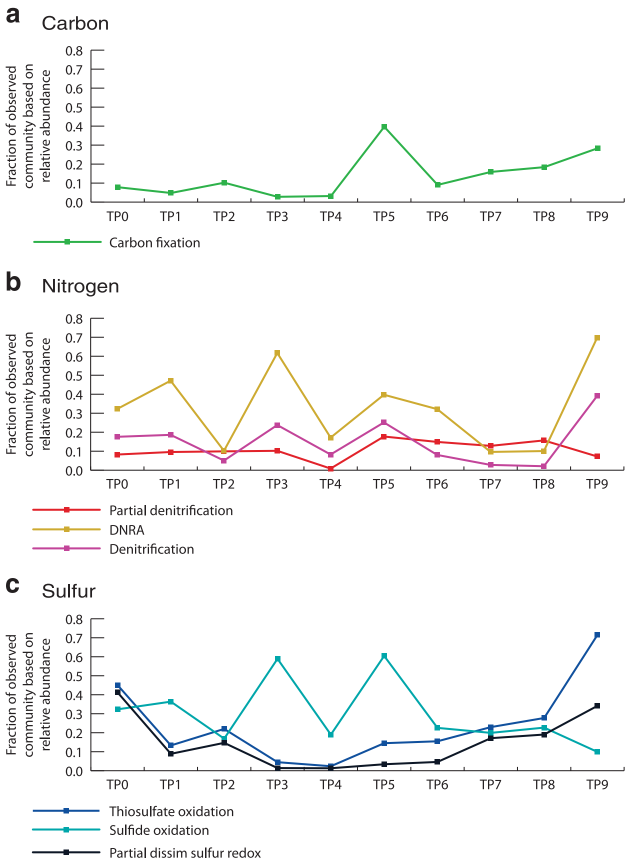

Figure 5 Fraction of the observed microbial community with potential to contribute to biogeochemically relevant processes. (**a**) Carbon fixation. (**b**) Nitrogen cycing. (**c**) Sulfur cycling. DNRA, dissimilatory nitrate reduction to ammonia; dissim., dissimilatory

bial community from any one of the metagenomic samples, thus we can only interpret results from the observed community members (Supplementary Data 4). It is likely, though, that due to the dynamics of assembly and binning that these genomes represent many of the most abundant organisms in the environment.

Carbon fixation

Previous results from North Pond samples in 2012 showed lower concentrations of dissolved organic carbon in the crustal fluids compared to seawater, as well as the potential for carbon fixation, with higher potential rates of autotrophy in the crust compared to seawater, especially at warmer temperatures (25°C) and deeper in the crust (Meyer *et al.*, 2016). In addition, limited metagenomic analysis of three samples from 2012 showed the presence of some genes associated with carbon fixation (Meyer *et al.*, 2016). Our assessment of genomes for the presence of genes representative of autotrophic carbon fixation resulted in the identification of two carbon fixation pathways: the CBB cycle and the reverse citric acid (rTCA) cycle (Table 2). All instances of the rTCA cycle were identified within the *Epsilonbacteraeota*, and the CBB cycle was identified in several different groups, including the *Alpha*-, *Gamma*-, and *Zetaproteobacteria*, as well as the *Planctomycetes*. Each of the genomes with potential for carbon fixation was also analyzed for pathways that could provide a lithotrophic source of reducing potential necessary for carbon fixation (Table 2). Results indicate that the most prevalent electron source identified amongst the putative carbon fixing genomes was sulfide, but several other electron sources were also identified, including thiosulfate, ferrous iron, sulfur, and hydrogen. These electron sources are likely coupled to the reduction of oxygen, as all but one of the genomes with predicted carbon fixation possess aerobic or microaerobic terminal oxidases. Possible additional terminal electron acceptors include nitrate and the intermediates of denitrification, as all but two of the carbon fixation genomes possess components of the denitrification or DNRA pathways (Supplementary Figure 4). While a majority of the genomes with carbon fixation potential are linked to the oxidation of sulfur compounds, a group of genomes have the potential to utilize both H_2 and Fe^{2+} to drive biomass production in support of the model proposed by Bach (2016). These putative energy couples are congruent with the hypothesis of subseafloor microbial communities that can take advantage of the redox gradient created by the presence of reduced material in volcanic-derived basalt rocks and the oxygenated aquifer fluids (Bach and Edwards, 2003; Bach, 2016). Hydrogen sulfide and iron species have not been detected in the crustal fluids at North Pond (Meyer *et al.*, 2016), but the oxidation of the iron in sulfide complexes in crustal rocks (via biotic or abiotic process) would increase access to sulfide compounds for microorganisms (Barco *et al.*, 2017) and for the abiotic oxidation of sulfide to thiosulfate (Moses *et al.*, 1987). In this manner, it would be possible to sustain carbon fixation through multiple lithotrophic pathways, which are likely important due to the oligotrophic nature of the crustal fluids. This is similar to the prevailing theory in regards to terrestrial crustal systems (Hallbeck and Pedersen, 2008), where lithoautotrophic growth in microorganisms via the CBB cycle has been found in deep terrestrial aquifers in the Fennoscandian shield (Wu *et al.*, 2015).

Genomic evidence for the prevalence of hypoxic conditions

All measurements at North Pond show that the aquifer fluids at North Pond are oxygenated, with O_2 concentrations equal to or slightly less (185–244 µM) than that of the DABW (~250 µM; Table 1; Meyer *et al.*, 2016). Therefore, it was unexpected to find that many of the North Pond genomes had genes that suggest hypoxic or potentially anoxic conditions. More than half of the genomes (56%) had terminal *c*-type cytochromes for both aerobic (*aa*$_3$- and *bo*-type) and microaerobic (*cbb*$_3$- and *bd*-

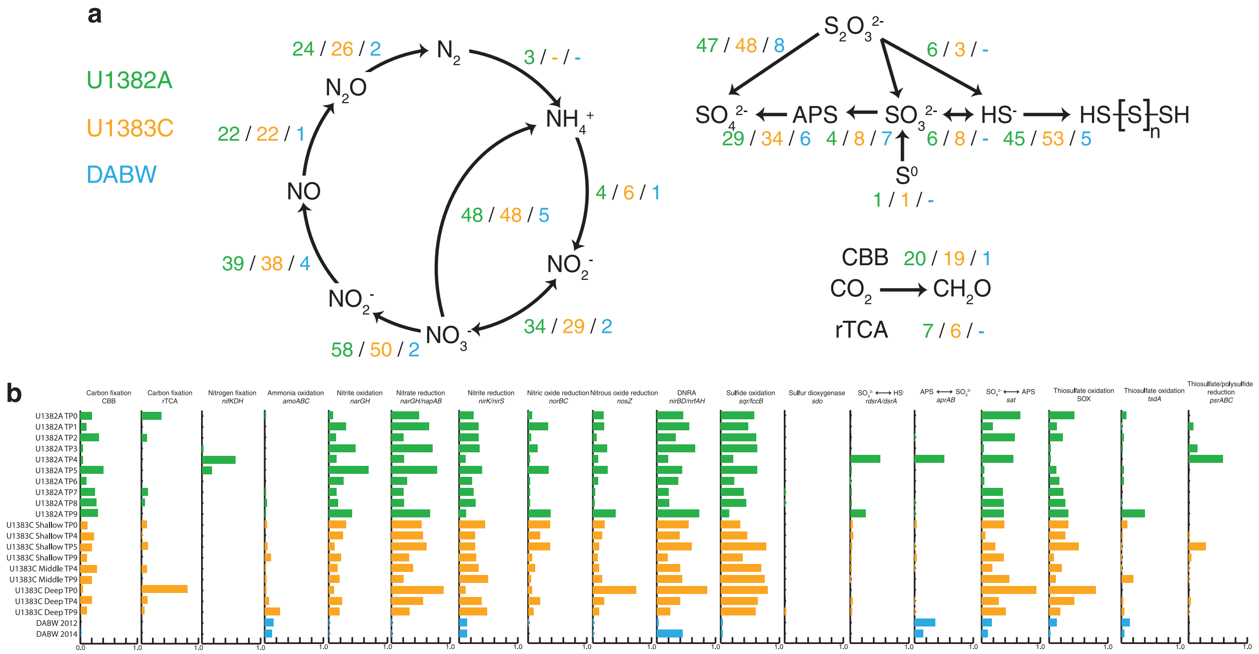

Figure 6 Comparison of putative microbial functionality between U1382A, U1383C and Deep Atlantic Bottom Water. (**a**) Metabolic pathways for nitrogen, sulfur and carbon fixation, with numbers indicating the number of genomes assigned to a particular sample type with that predicted function. (**b**) For each metabolic step represented in A, the fraction of the observed community from each sample that possesses that metabolic step. Abbreviations: DNRA, dissimilatory nitrate reduction to ammonia; TP, time point.

type) metabolisms, with an additional 13% of genomes only possessing the microaerobic cytochromes (Supplementary Figure 4). There was substantial evidence that the organisms in this environment were capable of the reduction of nitrate via both dissimilatory nitrate reduction to ammonia (DNRA; 36%) and denitrification (36%; Supplementary Figure 4). Further, NORP6 possessed the canonical sulfite reductase, necessary for the anaerobic conversion of sulfite to sulfide (Supplementary Figure 4). The role that these genes, commonly associated with anaerobic metabolisms, play in the environment is unclear. It is possible that, similar to sub-oxic microenvironments encountered in the oxic surface ocean (Ploug *et al.*, 1997), the subseafloor hosts microenvironments in which anaerobic metabolisms are ecologically viable. Like the surface ocean, one possible source of such microenvironments may be organic-rich particles, that can be readily colonized by heterotrophic microorganisms. In 2012, samples collected from North Pond crustal fluids showed a high heterogeneity of particles as detected on GFF filters (Meyer *et al.*, 2016). Another possibility may be that the complex and fractured structure of the crustal aquifer provides both oxic and sub-oxic conditions. For example, hydrogeological studies of the Juan de Fuca Ridge flank indicated that fluid flow through the crust likely only occurs through small, discrete channels, restricted to a small volume (<1%) of the crust (Fisher and Becker, 2000). Consequently fluid flow would be highly channelized through a small volume of the crustal rock. While measurements at North Pond CORKs show abundant oxygen, it is possible there are regions where fluid flow slows down and fluids could become stagnant, and anaerobic metabolisms may be more significant to the community as oxygen is consumed by heterotrophic activity or abiotic reactions. However, such stagnant fluids would likely not be indicative of the large crustal flow. Overall, the lack of an appreciable signal in the geochemical data may be the result of the extremely low biomass (~10^4 cells ml^{-1}) and relatively recent entrainment of the formation fluids, especially in U1382A.

Variable inter- and intra-borehole metabolic diversity

The microbial community observed in U1382A can be effectively assigned to seven ecological units with distinct occurrence patterns (Figure 4). These ecological units generally progress in sequential order, though several genomes within an ecological unit were detected in multiple time points, with up to 11 months between samples (TP2 vs TP7). This re-occurrence of members of the community suggest that there is mechanism for organisms to persist in the aquifer, either locally or transported from elsewhere within the subseafloor. Patterns may also be related to local geochemical conditions, where growth, and thus relative abundance, is tied to specific metabolic processes. Despite these changes in community structure over time, the genomes that are present in the ecological units are functionally redundant, with various metabolisms related to carbon fixation and nitrogen and sulfur cycling present in each of the measured time points (Figure 4). While the ecological units as a whole are functionally redundant, the fraction of the observed community capable of a specific metabolic potential shifts over the course of the time series (Figures 5a–c). Shifts in genomes capable of nitrate reduction (DNRA and complete denitrification) and sulfur oxidation (thiosulfate oxidation and sulfur redox) processes were positively correlated, suggesting that these metabolic pairs are linked to the same environmental change. Further, shifts in the fraction of the community capable of sulfide oxidation is linked to a microbial community structure that overlaps TP1 and TP3-6, while thiosulfate oxidation is linked to overlaps in TP2 and TP7-8 (Figures 5a–c; Supplementary Figure 9). This suggests that changes in availability of sulfide and thiosulfate are responsible for the changes in microbial community structure, or conversely, that microbial community metabolic potential impacts the availability of sulfide and thiosulfate.

In comparing U1382A and U1383C, several large, cohesive microbial groups were present in both boreholes (Figure 3), with organisms more abundant in U1383C clustering together, to the exclusion of organisms more abundant in U1382A. However, it was common for a group of MAGs to be more abundant in one hole and also have a reduced or minimal abundance in the other hole (Figure 3). While this result suggests there is some connectivity between the two subseafloor environments sampled by the CORKs, it is also clear that there are distinct, dominant populations within each hole, likewise there are distinct chemical signatures in both. However, the variation in community structure does not result in differences in metabolic potential, with functional redundancy in all queried processes, except for nitrogen fixation (Figure 6). This functional redundancy is further reflected in the fraction of the observed microbial community capable of participating in each metabolic step, with no statistically significant difference between the boreholes, except for ammonia oxidation (Figure 6). These results indicate that the observed differences in community structure are not related to carbon fixation or nitrogen and sulfur cycling, and are likely governed by environmental parameters that structure spatially distinct communities with a high degree of functional redundancy. A top–down control on community structure could be susceptibility to viral predation (Nigro *et al.*, 2017), while a bottom-up control may involve limits in trace nutrients or vitamin availability. Continued analysis of these data and future sampling efforts will help to elucidate the extent of these controls on the microbial community.

Concluding remarks

The microbial community in the crustal fluids of North Pond is temporally and spatially dynamic. The putative genomes extracted from our time series reveal a microbial community capable of impacting subseafloor biogeochemical cycles for carbon, nitrogen, sulfur and iron. These potential functions are redundant as community membership varies in time and space, suggesting that the communities present in both boreholes are poised to utilize the redox potential of the oceanic crust by exploiting reduced sulfur compounds and ferrous iron to drive autotrophic growth. Further research will elucidate the extent to which these organisms drive global biogeochemical processes.

Data availability

This project has been deposited at DDBJ/ENA/GenBank under the BioProject accession no. PRJNA391950, drafts of metagenome-assembled genomes are available with accession no. NVQK00000000-NVXW00000000, and raw sequence reads are available with accession no. SRX3143886-SRX3143902. Raw sequence reads from Meyer *et al.* (2016) constituting the metagenomic samples from 2012, are available under the BioProject accession no. PRJNA280201.

Additional files have been provided and are available through figshare (https://figshare.com/s/939160bb2d4156022558), such as: all primary and secondary contigs; MAGs and bins not analyzed as part of this research; and, all files described as Supplementary Data 1–17.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank the crews of the R/V Merian and ROV Jason, Wolfgang Bach, Peter Girguis Chih-Chiang Hsieh, Ulrike Jaekel, Beate Kraft, Huei-Ting Lin, Beth Orcutt, Keir Becker, Stephanie Carr, and Heiner Villinger for outstanding support in accomplishing our field programs. Ship time was generously provided by the German Science Foundation (DFG). Both Katrina Edwards and James Cowen's efforts were critical to the field component and success of this project. They are missed immensely. Leslie Murphy and Emily Reddington provided laboratory support at the WM Keck sequencing facility at the Marine Biological Laboratory. This work was supported by NSF OCE1062006 to JAH and NSF OCE1061827 to BTG. The Gordon and Betty Moore Foundation sponsored most of the observatory components at North Pond through grant GBMF1609. The Center for Dark Energy Biosphere Investigations (C-DEBI) (OCE-0939564), a National Science Foundation-funded Science and Technology Centers of Excellence also supported the participation of CGW, JAH and BJT. This is C-DEBI contribution number 395.

References

Akerman NH, Butterfield DA, Huber JA. (2013). Phylogenetic diversity and functional gene patterns of sulfur-oxidizing subseafloor Epsilonproteobacteria in diffuse hydrothermal vent fluids. *Front Microbiol* **4**: 185.

Bach W. (2016). Some compositional and kinetic controls on the bioenergetic landscapes in oceanic basement. *Front Microbiol* **7**: 9917–9918.

Bach W, Edwards KJ. (2003). Iron and sulfide oxidation within the basaltic ocean crust: implications for chemolithoautotrophic microbial biomass production. *Geochim Cosmochim Acta* **67**: 3871–3887.

Barco RA, Emerson D, Sylvan JB, Orcutt BN, Jacobson Meyers ME, Ramírez GA *et al.* (2015). New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. In: Voordouw G (ed). *Appl Environ Microbiol* **81**: 5927–5937.

Barco RA, Hoffman CL, Ramírez GA, Toner BM, Edwards KJ, Sylvan JB. (2017). *In-situ* incubation of iron-sulfur mineral reveals a diverse chemolithoautotrophic community and a new biogeochemical role for Thiomicrospira. *Environ Microbiol* **19**: 1–42.

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR *et al.* (2002). The Pfam protein families database. *Nucleic Acids Res* **30**: 276–280.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cowen JP, Copson DA, Jolly J, Hsieh C-C, Lin H-T, Glazer BT *et al.* (2012). Advanced instrument system for real-time and time-series microbial geochemical sampling of the deep (basaltic) crustal biosphere. *Deep-Sea Res* **61**: 43–56.

Cowen JP, Giovannoni SJ, Kenig F, Johnson HP, Butterfield D, Rappé MS *et al.* (2003). Fluids from aging ocean crust that support microbial life. *Science* **299**: 120–123.

Davis EE, Becker K, Pettigrew T, Carson B, MacDonald R. (1992). CORK: a hydrologic seal and downhole observatory for deep-ocean boreholes. *Proc Ocean Drill Prog* **139**: 43–53.

Edgar RC. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Edwards KJ, Bach W, Klaus A. (2010). 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* **336**: 1–62.

Edwards KJ, Fisher AT, Wheat CG. (2012). The deep subsurface biosphere in igneous ocean crust: frontier habitats for microbiological exploration. *Front Microbiol* **3**: 8.

Edwards KJ, Wheat CG, Orcutt BN, Hulme S, Becker K, Jannasch HW *et al.* (2012) Design and deployment of borehole observatories and experiments during IODP Expedition 336, Mid-Atlantic Ridge flank at North Pond. In: Edwards KJ, Bach W, Klaus A (eds). *Proceedings of the Integrated Ocean Drilling Program, Expedition 336* **Vol. 336**. Integrated Ocean Drilling Program Management International, Inc: Tokyo, pp, 1–43.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Fisher AT, Becker K. (2000). Channelized fluid flow in oceanic crust reconciles heat-flow and permeability data. *Nature* **403**: 71–74.

Fisher AT, Wheat CG. (2010). Seamounts as conduits for massive fluid, heat, and solute fluxes on ridge flanks. *Oceanography* **23**: 74–87.

Frey BJ, Dueck D. (2007). Clustering by passing messages between data points. *Science* **315**: 972–976.

García-Horsman JA, Barquera B, Rumbley J, Ma J, Gennis RB. (1994). The superfamily of heme-copper respiratory oxidases. *J Bacteriol* **176**: 5587–5600.

Graham ED, Heidelberg JF, Tully BJ. (2017). BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. *PeerJ* **5**: e3035–19.

Haft DH, Selengut JD, White O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res* **31**: 371–373.

Hallbeck L, Pedersen K. (2008). Characterization of microbial processes in deep aquifers of the Fennoscandian Shield. *Appl Geochem* **23**: 1796–1819.

Huang Y, Gilna P, Li W. (2009). Identification of ribosomal RNA genes in metagenomic fragments. *Bioinformatics* **25**: 1338–1340.

Huber JA, Butterfield DA, Baross JA. (2002). Temporal changes in archaeal diversity and chemistry in a mid-ocean ridge subseafloor habitat. *Appl Environ Microbiol* **68**: 1585–1594.

Huber JA, Johnson HP, Butterfield DA, Baross JA. (2006). Microbial life in ridge flank crustal fluids. *Environ Microbiol* **8**: 88–99.

Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ *et al.* (2016). A new view of the tree of life. *Nature Microbiology* **1**: 16048.

Hügler M, Sievert SM. (2011). Beyond the calvin cycle: autotrophic carbon fixation in the ocean. *Annu Rev Marine Sci* **3**: 261–289.

Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* **28**: 2223–2230.

Johnson HP, Pruis MJ. (2003). Fluxes of fluid and heat from the oceanic crustal reservoir. *Earth Planet Sci Lett* **216**: 565–574.

Jungbluth SP, Bowers RM, Lin H-T, Cowen JP, Rappé MS. (2016). Novel microbial assemblages inhabiting crustal fluids within mid-ocean ridge flank subsurface basalt. *ISME J* **10**: 2033–2047.

Jungbluth SP, Grote J, Lin H-T, Cowen JP, eacute MSR. (2012). Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME J* **7**: 161–172.

Jungbluth SP, Grote J, Lin H-T, Cowen JP, Rappé MS. (2013). Microbial diversity within basement fluids of the sediment-buried Juan de Fuca Ridge flank. *ISME J* **7**: 161–172.

Jørgensen SL, Zhao R. (2016). Microbial inventory of deeply buried oceanic crust from a young ridge flank. *Front Microbiol* **7**: 3871–14.

Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* **44**: D457–D462.

Kanehisa M, Sato Y, Morishima K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726–731.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S *et al.* (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**: 3100–3108.

Lee MD, Walworth NG, Sylvan JB, Edwards KJ, Orcutt BN. (2015). Microbial communities on seafloor basalts at dorado outcrop reflect level of alteration and highlight global lithic clades. *Front Microbiol* **6**: 403–420.

Liao Y, Smyth GK, Shi W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**: 923–930.

Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I *et al.* (2009). Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. *Environ Microbiol* **11**: 289–299.

Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al.* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.

Lysnes K, Thorseth IH, Steinsbu BRO, Ã VreÃ s L, Torsvik T, Pedersen RB. (2004). Microbial community diversity in seafloor basalt from the Arctic spreading ridges. *FEMS Microbiol Ecol* **50**: 213–230.

Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A *et al.* (2006). The integrated microbial genomes (IMG) system. *Nucleic Acids Res* **34**: D344–D348.

Martin M. (2012). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10–12.

Mason OU, Di Meo-Savoie CA, Van Nostrand JD, Zhou J, Fisk MR, Giovannoni SJ. (2009). Prokaryotic diversity, distribution, and insights into their role in biogeochemical cycling in marine basalts. *ISME J* **3**: 231–242.

Meyer JL, Jaekel U, Tully BJ, Glazer BT, Wheat CG, Lin H-T *et al.* (2016). A distinct and active bacterial community in cold oxygenated fluids circulating beneath the western flank of the Mid-Atlantic ridge. *Sci Rep* **6**: 22541.

Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. (2011). EMIRGE: reconstruction of full-length ribosomalgenes from microbial community short readsequencing data. *Genome Biol* **12**: R44.

Miller CS, Handley KM, Wrighton KC, Frischkorn KR, Thomas BC, Banfield JF. (2013). Short-read assembly of full-length 16S amplicons reveals bacterial diversity in subsurface sediments. Gilbert JA (ed). *PLoS One* **8**: e56018–11.

Moses CO, Nordstrom DK, Herman JS. (1987). Aqueous pyrite oxidation by dissolved oxygen and by ferric iron. *Geochim Cosmochim Acta* **51**: 1561–1571.

Nigro OD, Jungbluth SP, Lin H-T, Hsieh C-C, Miranda JA, Schvarcz CR *et al.* (2017). Viruses in the oceanic basement. *mBio* **8**: e02129–16–15.

Orcutt BN, Sylvan JB, Rogers DR, Delaney J, Lee RW, Girguis PR. (2015). Carbon fixation by basalt-hosted microbial communities. *Front Microbiol* **6**: 1–14.

Orcutt BN, Wheat CG, Rouxel O, Hulme S, Edwards KJ, Bach W. (2013). Oxygen consumption rates in subseafloor basaltic crust derived from a reaction transport model. *Nat Commun* **4**: 2539.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043–1055.

Peng Y, Leung HCM, Yiu SM, Chin FYL. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**: 1420–1428.

Ploug H, Kühl M, Buchholz B. (1997). Anoxic aggregates an ephemeral phenomenon in the ocean. *Aquat Microb Ecol* **13**: 285–394.

Porter KG, Feig YS. (1980). The use of dapi for identifying and counting aquatic microflora. *Limnol Oceanogr* **25**: 943–948.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2–approximately maximum-likelihood trees for large alignments. Poon, AFY (ed). *PLoS One* **5**: e9490.

Pruesse E, Peplies J, Glöckner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al.* (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Reveillaud J, Reddington E, McDermott J, Algar C, Meyer JL, Sylva S *et al.* (2016). Subseafloor microbial communities in hydrogen-rich vent fluids from hydrothermal systems along the Mid-Cayman Rise. *Environ Microbiol* **18**: 1970–1987.

Robador A, LaRowe DE, Jungbluth SP, Lin H-T, Rappé MS, Nealson KH *et al.* (2016). Nanocalorimetric characterization of microbial activity in deep subsurface oceanic crustal fluids. *Front Microbiol* **7**: 135–138.

Santelli CM, Edgcomb VP, Bach W, Edwards KJ. (2009). The diversity and abundance of bacteria inhabiting seafloor lavas positively correlate with rock alteration. *Environ Microbiol* **11**: 86–98.

Sclater JG, Jaupart C, Galson D. (1980). The heat flow through oceanic and continental crust and the heat loss of the Earth. *Rev Geophys* **18**: 269–311.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103**: 12115–12120.

Stein CA, Stein S. (1994). Constraints on hydrothermal heat flux through the oceanic lithosphere from global heat flow. *J Geophys Res* **99**: 3081–3095.

Tabita FR, Hanson TE, Li H, Satagopan S, Singh J, Chan S. (2007). Function, structure, and evolution of the RubisCO-like proteins and their RubisCO homologs. *Microbiol Mol Biol Rev* **71**: 576–599.

Takai K, Horikoshi K. (1999). Genetic diversity of archaea in deep-sea hydrothermal vent environments. *Genetics* **152**: 1285–1297.

Tully BJ, Heidelberg JF. (2016). Potential mechanisms for microbial energy acquisition in oxic deep-sea sediments. In: Drake HL (ed). *Appl Environ Microbiol* **82**: 4232–4243.

Wheat CG, Fisher AT, McManus J, Hulme SM, Orcutt BN. (2017). Cool seafloor hydrothermal springs reveal global geochemical fluxes. *Earth Planet Sci Lett* **476**: 179–188.

Wheat CG, Jannasch HW, kastner M, Hulme S, Cowen J, Edwards KJ *et al.* (2011). *Fluid sampling from oceanic borehole observatories: design and methods for CORK activities (1990-2010)*. Integrated Ocean Drilling Program: Tokyo.

Wu X, Holmfeldt K, Hubalek V, Lundin Dm MASO, Bertilsson S *et al.* (2015). Microbial metagenomes from three aquifers in the Fennoscandian shield terrestrial deep biosphere reveal metabolic partitioning among populations. *ISME J* **10**: 1–12.

Zhang X, Feng X, Wang F. (2016). Diversity and metabolic potentials of subsurface crustal microorganisms from the western flank of the Mid-Atlantic Ridge. *Front Microbiol* **7**: 363.

Ziebis W, McManus J, Ferdelman T, Schmidt-Schierhorn F, Bach W, Muratli J *et al.* (2012). Interstitial fluid chemistry of sediments underlying the North Atlantic gyre and the influence of subsurface fluid flow. *Earth Planet Sci Lett* **323-324**: 79–91.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)