Supplemental material for "Microbial life in Oceanic Crust" book chapter, *submitted to the Biological and Chemical Oceanography Database Management Office (BCO-DMO)*

Beth N. Orcutt[1], Timothy D'Angelo[1], Sean P. Jungbluth[2], Julie A. Huber[3], Jason B. Sylvan[4]

[1] Bigelow Laboratory for Ocean Sciences, 60 Bigelow Drive, East Boothbay, ME, 04544, USA; borcutt@bigelow.org

[2] Department of Energy, Joint Genome Institute, Walnut Creek, CA, USA

[3] Woods Hole Oceanographic Institution, Woods Hole, MA, USA

[4] Texas A&M University, College Station, TX, USA

**Supplemental Methods:**

*Analysis of publicly available 16S rRNA gene sequence datasets for taxonomic profiling*

To summarize crustal bacterial and archaeal taxa for this review, we synthesized publicly-available sequence datasets of the 16S rRNA gene in environmental DNA extracted from seafloor and subseafloor basalts generated using 454, Illumina and Ion Torrent amplicon platforms. These include seafloor basalts from the Dorado Outcrop [1] and the Lō'ihi Seamount [2] in the Pacific Ocean and subseafloor basalts from North Pond on the western flank of the Mid-Atlantic Ridge [3] and the Juan de Fuca Ridge flank in the northeastern Pacific Ocean [4]. Datasets from rock colonization experiments conducted in the subseafloor at the Juan de Fuca Ridge flank site [5,6] were also included, as well as microbial community surveys of the subseafloor crustal fluids from the anoxic Juan de Fuca site [7] and the oxic North Pond site [8,9]. For comparison, we included select reference datasets from oxic [10,11] and anoxic sediment [4] and the overlying bottom seawater [1] from these same study sites.

Raw sequence data from the reviewed studies were downloaded from the NCBI Short Read Archive. Sequencing reads generated using Illumina and Ion Torrent platforms were quality filtered and processed to unique Amplicon Sequence Variants (ASVs) using DADA2 [12], with taxonomy determined by the naïve Bayesian classifier in DADA2 using a training set from the SILVA v132 database [13-15]. For the 454 GS-FLX sequence datasets, operational taxonomic units (OTUs) constructed with 97% or greater sequence similarity in the original analyses were reprocessed in mothur V.1.37.6 [16] against the same SILVA database. All short read datasets were merged and summarized to the relative abundance at phylum resolution (or to class level for Proteobacteria phyla) using Phyloseq v1.24.0 [17]. Figures were produced using ggplot2 [18] in RStudio [19]. Taxonomic grouping in each sample separated taxa into common (>5% abundance in at least one sample) versus rare (never more than 5% in any sample). Supplemental Figure S1 shows the breakdown of Gammaproteobacteria families in the samples presented in Figure 4 of the main text, and Supplemental Figure S2 highlights the abundance of rare taxa (never >5% abundance in any sample). The Bray-Curtis distances between samples was calculated using the same dataset described above, summarized to relative abundance at the Family taxonomic level using Phyloseq and the Vegan package [20]. A Non-Metric Multidimensional Scaling (NMDS) ordination was produced from this distance matrix. It should be noted that common rules for beta diversity comparisons, such as common library preparation/sequencing protocols and library-size normalization, were not performed in this

analysis due to the diversity of the datasets being considered and the resulting NMDS ordination having high-stress (>20%). Therefore, the results should be viewed as broadly qualitative and not quantitative.

All data processing steps and markdown files are available via github: https://github.com/orcuttlab/ocean-crust-micro

*Analysis of publicly available 16S rRNA gene sequence datasets for inferring functional predictions*

For this review we explored the prediction of possible microbial function from phylogenetic profiles (i.e. 16S rRNA gene sequence libraries above) using the PICRUSt program (version 1.0.0; [21]). In brief, this program makes functional predictions based on the characterization of closely related microbial isolates to taxa in the phylogenetic analysis, based on the ratios of those taxa in 16S rRNA sequence libraries. A subset of the dataset described above were used in this analysis to examine for functional potential related to nitrogen, sulfur and carbon cycling as well as other physiological features. Sequences were quality-controlled to remove reads with ambiguous bases. Dereplicated or full sequence reads were divided up by sample location and average gene copy numbers were estimated using PICRUSt with default parameters and the GreenGenes/PICRUSt cog_13_5_precalculated database. Samples collected over multiple years from the same location were merged together. The following Clusters of Orthologous Groups (COGs) of proteins [22] were profiled with this approach: COG5013, nitrate reductase alpha subunit; COG1140, nitrate reductase beta subunit; COG1251, NAD(P)H-nitrite reductase; COG2146, nitrite reductase-ferredoxin; COG3256, nitric oxide reductase large subunit; COG4263, nitrous oxide reductase; COG1348, nitrogenase; COG0004, ammonia permease; COG2046, ATP sulfurylase; COG0175, PAPS reductase; COG0369, sulfite reductase alpha subunit; COG0155, sulfite reductase beta subunit; COG2221, sulfite reductase - desulfoviridin; COG4117, thiosulfate reductase cytochrome b subunit; COG1850, RuBisCO; COG4058, methyl coenzyme M reductase; COG3259, coenzyme F420 reductase alpha subunit; COG1035, coenzyme F420 reductase beta subunit; COG1941, coenzyme F420 reductase gamma subunit; COG1908, coenzyme $F_{420}$ reductase delta; COG1148, heterodisulfide reductase alpha subunit; COG2048, heterodisulfide reductase beta subunit; COG1150, heterodisulfide reductase gamma subunit; COG1785, alkaline phosphatase; COG0260, leucyl aminopeptidase; COG0006, Xaa-Pro aminopeptidase; COG0024, methionine aminopeptidase; COG0308, aminopeptidase N; COG1362, aspartyl aminopeptidase; COG1363, putative aminopeptidase ; COG1506, dipeptidyl aminopeptidase/acylaminoacyl-peptidase; COG2309, leucyl aminopeptidase (aminopeptidase.T); COG2362, D-aminopeptidase; COG3191, L-aminopeptidase/D-esterase; COG1261, flagella basal body P-ring biosynthesis; COG1815, flagellar basal-body rod; COG1558, flagellar basal-body rod; COG1843, flagellar hook capping; COG1749,flagellar hook; COG4787, flagellar basal body rod; COG4786, flagellar basal body rod; COG2063, flagellar basal body L-ring; COG1706, flagellar basal-body P-ring; COG1705, flagellum-specific muramidase; COG1256, flagellar hook-associated; COG1344, flagellin and related hook-associated proteins; COG1300, stage II sporulation; COG2359, stage V sporulation; COG2719, stage V sporulation; COG2088, stage V sporulation; COG3854, stage III sporulation; COG3874, uncharacterized spore protein; COG4326, sporulation-control. Averages are reported for the subunits of nitrate reductase, heterodisulfide reductase, and coenzyme F420 reductase and all peptidases and flagella proteins.

**Supplemental Discussion:**

*Inferring possible metabolic function from 16S rRNA gene taxonomic profiles*

As there are few functional gene surveys from the oceanic crust environment thus far, for this review we explored the prediction of possible microbial functions from phylogenetic profiles (i.e. 16S rRNA gene sequence libraries above) using the PICRUSt program (version 1.0.0; [21]). In brief, this program makes functional predictions based on the characterization of closely related microbial isolates to taxa in the phylogenetic analysis, based on the ratios of those taxa. There are limitations to this approach of inferring function from phylogeny, especially for communities with many taxa with no known cultured representatives, as is clearly the case for most environmental samples [23]. This approach is also limited by the reference databases available for comparison, as the PiCRUST reference has not been updated since 2013 although numerous new microbial dark matter groups have been described since then [24]. Finally, this analysis is limited in comparing 16S rRNA gene profiles from both clone library and amplicon sequence datasets, which have vastly different sequence amounts and may skew comparative results. However, since there are far more publicly-available 16S rRNA gene datasets, a cautiously predictive functional profiling based on inference to phylogenetically related isolates may provide some utility to constrain or predict functions in the ocean crust. A subset of the dataset described above were used in this analysis to examine for functional potential related to nitrogen, sulfur and carbon cycling as well as other physiological features.

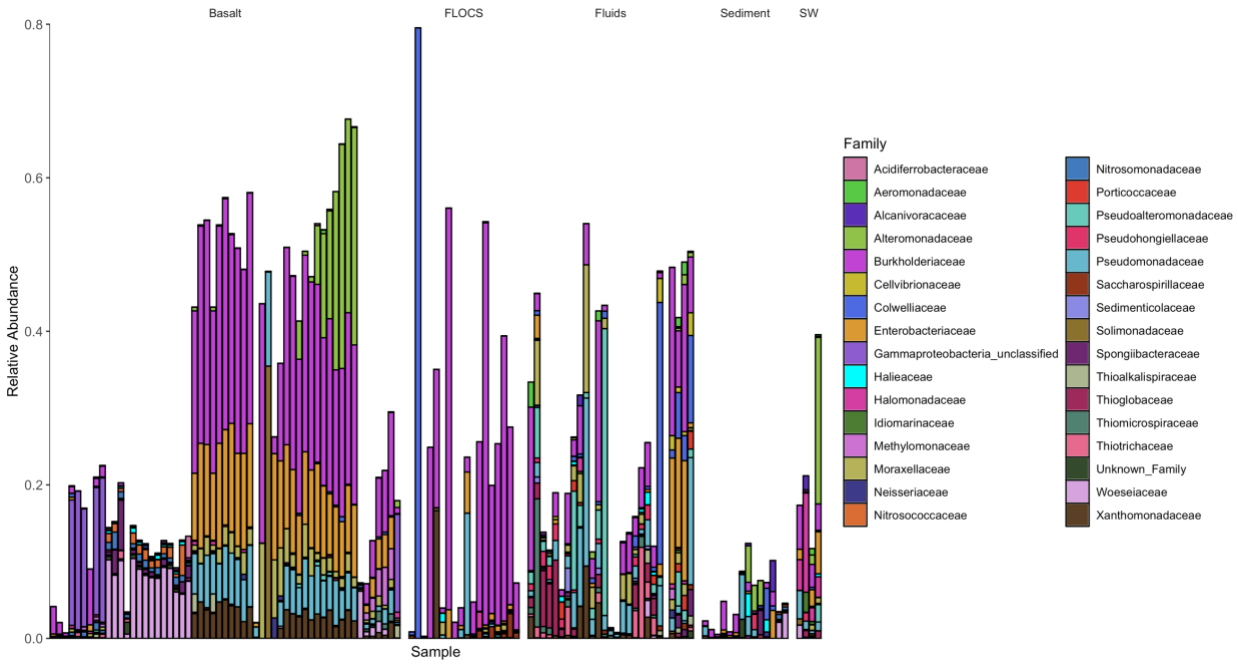
Analysis of the functional profiles generated using this approach highlight predicted cycling of nitrogen, sulfur, and carbon (to a lesser extent) in oceanic crust microbial communities (Supplemental Figure S3). The most apparent indications of nitrogen cycling were predictions of nitrite reductases and ammonia permease genes, which together suggest the oceanic crust microbial communities employ both assimilatory and dissimilatory mechanisms to access nitrogen. Genes for nitrate reduction and nitrogen fixation were sparsely predicted, which is consistent with the few genomic content descriptions from oceanic crust [25,8]. Genes predicted for nitric oxide cycling were also largely absent. Compared to nitrogen, genes predicted for sulfur cycling were more apparent across the sample types. Nearly all sample datasets predict reduction of activated sulfate to sulfite (i.e., PAPS reductase), and many also predict sulfite reductase genes (i.e., *dsrAB*). Thiosulfate cycling gene predictions were only observed in one mineral incubation sample set. While sulfate reduction potential has been confirmed in anoxic basaltic crust settings [26,27] and suggested in highly-reducing serpentinizing crust environments [28], it is somewhat counterintuitive to have sulfate reduction occurring in oxic conditions and suggest anoxic microniche in these habitats. There were few predictions for carbon fixation pathways in any of the samples, as well as few predictions for methane cycling genes. The lack of carbon fixation predictions is in contrast to demonstrated carbon fixation potential in crustal settings [29,9], and suggests that other carbon fixation pathways may be used by crustal communities. Likewise, the PiCRUST analysis does not yet take into account the recent expansion of methane cycling groups across the archaeal domain and variations in the *mcrA* gene [30]. Predictions for peptide degradation genes, as a proxy for scale of fermentation processes, were pervasive in the sample set, supporting the observations of heterotrophy as a widespread and important feature of crustal communities [31-33]. Flagellar-based motility was

also predicted to be widespread in crustal communities, though in low abundance; motility has recently been suggested to be an adaptive strategy in this habitat type [34].

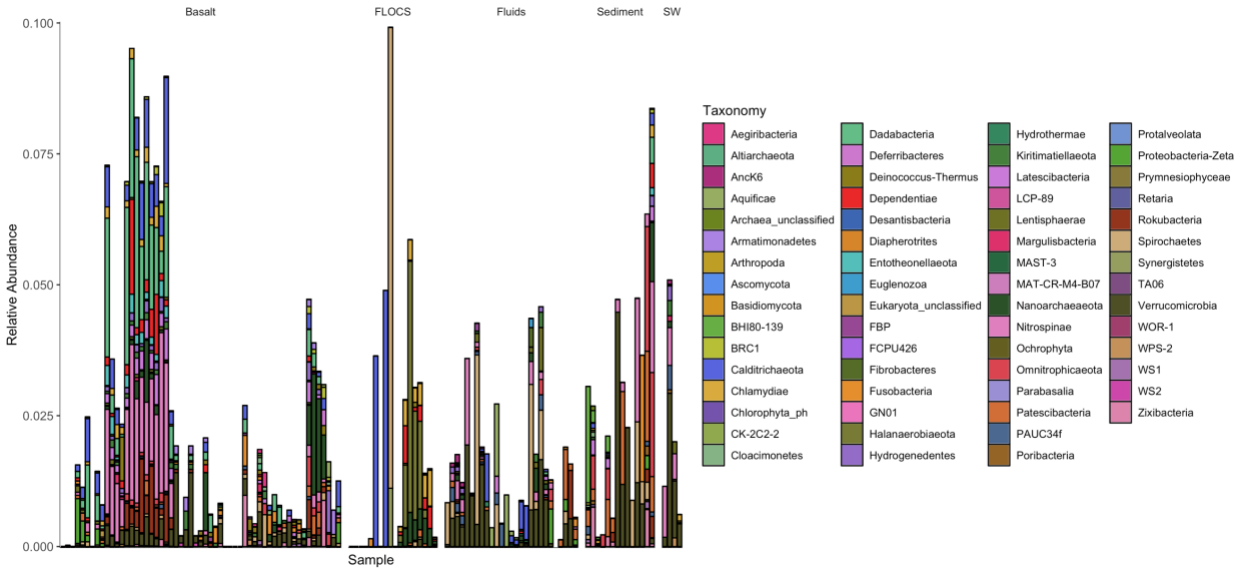
Keeping the limitations of the PiCRUST analysis in mind, these patterns (Supplemental Figure S3) suggest that common functional themes in oceanic crust microbial communities include: cycling of inorganic nitrogen (mostly as nitrite) and sulfur (mostly as sulfate), widespread amino acid fermentation, and flagellar-based motility. As the number of genomic and metagenomic datasets from oceanic crust environments increase, these functional predictions based on 16S rRNA gene surveys can be more rigorously tested.

## Supplemental Figures

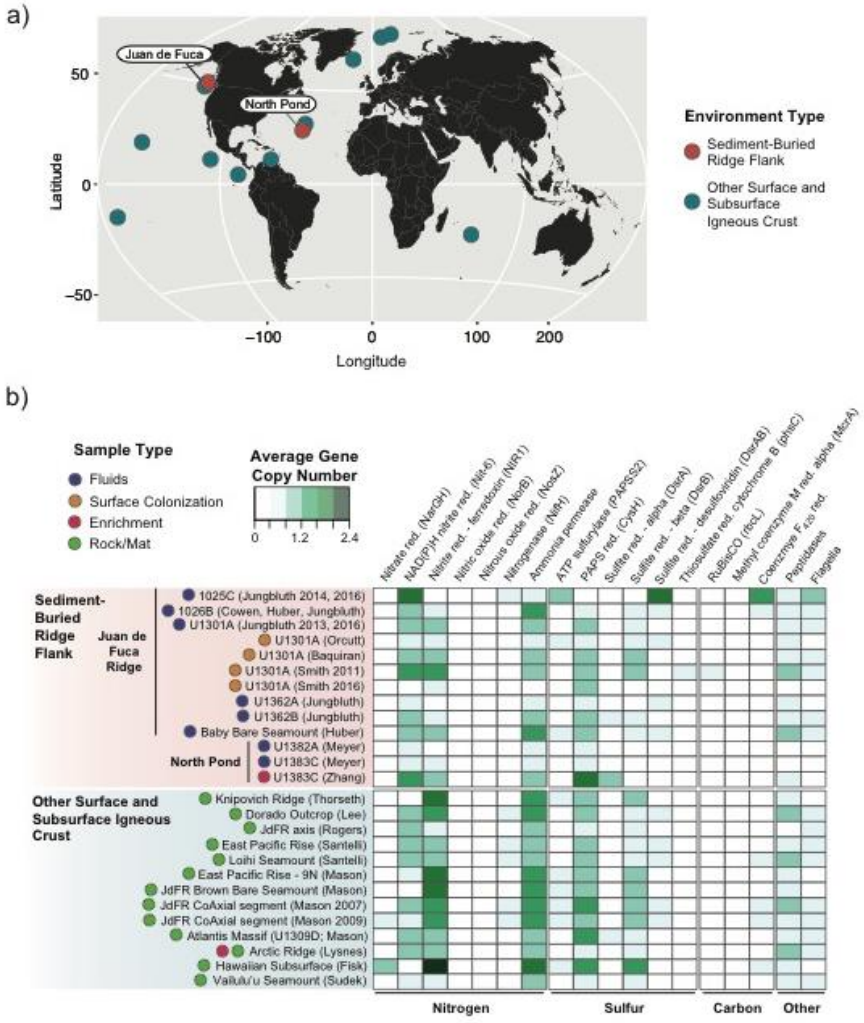
**Supplemental Fig. S1.** Relative abundance of Gammaproteobacteria families in 16S rRNA gene sequence datasets from 454, Illumina, and Ion Torrent platforms from crustal rocks, biofilms, and fluids in comparison to sediment and bottom seawater. Sample order is the same as presented in the main text and in the metadata table. Further details about sample names and original available in metadata table.



**Supplemental Figure S2.** Relative abundance of rare bacterial and archaeal taxonomic groups (phyla level for most taxa, except Proteobacteria at class level) in 16S rRNA gene sequence datasets (same as in Supplemental Figure S1).



**Supplemental Figure S3.** (a) Global map of oceanic crust samples used in the PiCRUST analysis with sediment-buried ridge flank samples (red-color) and other surface and subsurface igneous crust samples (blue-color) designated separately. Special attention is drawn to the Juan de Fuca and North Pond locations because of their relative historical importance in ocean crustal microbiology investigations. (b) Heatmap displaying PiCRUST predicted average gene copy number of selected Categories of Orthologous Groups (COGs) in ocean crust microbial communities based on analysis of 16S rRNA gene sequences. Ocean crust sample types are indicated with colored circles: fluids (blue), surface colonization (yellow), laboratory-based enrichment (pink), and rock/mat (green). Further details about sample order and origins available at in metadata table.



**References**


1. 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. Frontiers in Microbiology 6:1470. doi:10.3389/fmicb.2015.01470
2. Jacobsen Meyers ME, Sylvan JB, Edwards KJ (2014) Extracellular enzyme activity and microbial diversity measured on seafloor exposed basalts from Loihi Seamount indicate importance of basalts to global biogeochemical cycling. Applied and Environmental Microbiology. doi:10.1128/AEM.01038-14
3. Jørgensen SL, Zhao R (2016) Microbial inventory of deeply buried oceanic crust from a young ridge flank. Frontiers in Microbiology 7:820. doi:10.3389/fmicb.2016.00820
4. LaBonté JM, Lever MA, Edwards KJ, Orcutt BN (2017) Influence of igneous basement on deep sediment microbial diversity on the eastern Juan de Fuca Ridge flank. Frontiers in Microbiology 8:1434. doi:10.3389/fmicb.2017.01434
5. Smith AR, Fisk MR, Thurber AR, Flores G, Mason OU, Popa R, Colwell FS (2016) Deep crustal communities on the Juan de Fuca Ridge are governed by minearology. Geomicrobiology Journal. doi:10.1080/01490451.2016.1155001
6. Ramírez GA, Garber AI, Lecoeuvre A, D'Angelo T, Wheat CG, Orcutt BN (2019) Ecology of subseafloor crustal biofilms. Frontiers in Microbiology 10:1983. doi:10.3389/fmicb.2019.01983
7. 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. The ISME Journal 10:2033-2047. doi:10.1038/ismej.2015.248
8. Tully BJ, Wheat CG, Glazer BT, Huber JA (2017) A dynamic microbial community with high functional redundancy inhabits the cold, oxic subseafloor aquifer. The ISME Journal 12:1-16. doi:10.1038/ismej.2017.187
9. Meyer JL, Jaekel U, Tully BJ, Glazer BT, Wheat CG, Lin H-T, Hsieh C-C, Cowen JP, Hulme S, Girguis PR, Huber JA (2016) A distinct and active bacterial community in cold oxygenated fluids circulating beneath the western flank of the Mid-Atlantic ridge. Scientific Reports 6:22541. doi:10.1038/srep22541
10. Reese BK, Zinke LA, Sobol MS, LaRowe DE, Zhang X-X, Jaekel U, Wang F, Orcutt BN, Dittmar T, Defforey D, Tully BJ, Paytan A, Sylvan JB, Amend JP, Edwards KJ, Girguis PR (2018) Nitrogen cycling potential of active bacteria within oligotrophic sediment of the Mid-Atlantic Ridge Flank. Geomicrobiology Journal 35:468-483. doi:10.1080/01490451.2017.1392649
11. Zinke LA, Kiel Reese B, McManus J, Wheat CG, Orcutt BN, Amend JP (2018) Sediment microbial communities influenced by cool hydrothermal fluid migration. Frontiers in Microbiology 9:1249. doi:10/3389/fmicb.2018.01249
12. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13 (7):581-583. doi:10.1038/nmeth.3869
13. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41 (D1):D590-D596. doi:10.1093/nar/gks1219
14. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Research 42:643-648. doi:10.1093/nar/gkt1209

15. Glöckner FO, Yilmaz P, Quast C, Gerken J, Beccati A, Ciuprina A, Bruns G, Yarza P, Peplies J, Westram R, Ludwig W (2017) 25 years of serving the community with ribosomal RNA gene reference databases and tools. Journal of Biotechnology 261:169-176. doi:10.1016/j.jbiotec.2017.06.1198
16. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF (2009) Introducing *mothur*: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75:7537-7541. doi:10.1128/AEM.01541-09
17. McMurdie PJ, Holmes S (2013) Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8 (4):e61217. doi:10.1371/journal.pone.0061217
18. Wickham H, Chang W (2017) ggplot2: Create elegant data visualisations using the grammar of graphics. RStudio. R package version 2.2.1.
19. RStudio Team (2017) RStudio: Integrated Development for R. Version 1.0.153. RStudio, Inc., Boston, MA.
20. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH, Szoecs E, Wagner H (2018) vegan: Community Ecology Package. R package version 2.5-2.
21. Langille MG, Zaneveld J, Caparaso J, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C (2013) Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nature Biotechnology 31:814-821. doi:10.1038/nbt.2676
22. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Research 28 (1):33-36
23. Lloyd KG, Steen AD, Ladau J, Yin J, Crosby L (2018) Phylogenetically Novel Uncultured Microbial Cells Dominate Earth Microbiomes. mSystems 3 (5). doi:10.1128/mSystems.00055-18
24. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, Ise K (2016) A new view of the tree of life. Nature microbiology 1 (5):16048
25. Jungbluth SP, Glavina del Rio T, Tringe SG, Stepanauskas R, Rappé MS (2017) Genomic comparisons of a baterical lineage that inhabits both marine and terrestrial deep subsurface systems. PeerJ 5:e3134. doi:10.7717/peerj.3134
26. Lever MA, Rouxel OJ, Alt JC, Shimizu N, Ono S, Coggon RM, Shanks III WC, Lapham L, Elvert M, Prieto Mollar X, Hinrichs KU, Inagaki F, Teske A (2013) Evidence for microbial carbon and sulfur cycling in deeply buried ridge flank basalt. Science 339 (6125):1305-1308
27. Robador A, Jungbluth SP, LaRowe DE, Bowers RM, Rappé MS, Amend JP, Cowen JP (2015) Activity and phylogenetic diversity of sulfate-reducing microorganisms in low-temperature subsurface fluids with the upper oceanic crust. Frontiers in Microbiology 5:748
28. Lang SQ, Früh-Green G, Bernasconi SM, Brazelton WJ, Schrenk MO, McGonigle JM (2018) Deeply-sourced formate fuels sulfate reducers but not methanogens at Lost City hyfrothermal field. Scientific Reports 8:755. doi:10.1038/s41598-017-19002-5
29. Orcutt BN, Sylvan JB, Rogers DR, Delaney J, Lee RW, Girguis PR (2015) Carbon fixation by basalt-hosted microbial communities. Frontiers in Microbiology 6:904. doi:10.3389/fmicb.2015.00904

30. Boyd J, Jungbluth S, Leu A, Evans P, Woodcroft B, Chadwick G, Orphan V, Amend J, Rappe M, Tyson G (2018) Divergent methyl-coenzyme M reductase genes in a deep-subseafloor Archaeoglobi. bioRxiv. doi:10.1101/390617
31. Lang SQ, Butterfield DA, Lilley MD, Johnson HP, Hedges JI (2006) Dissolved organic carbon in ridge-axis and ridge-flank hydrothermal systems. Geochim Cosmochim Acta 70:3830-3842
32. Lin H-T, Cowen JP, Olson EJ, Amend JP, Lilley MD (2012) Inorganic chemistry, gas compositions and dissolved organic carbon in fluids from sedimented young basaltic crust on the Juan de Fuca Ridge flanks. Geochimica et Cosmochimica Acta 85:213-227
33. Shah Walker SR, Jaekel U, Osterholz H, Fisher AT, Huber JA, Pearson A, Dittmar T, Girguis PR (2018) Microbial decomposition of marine dissolved organic matter in cool oceanic crust. Nature Geoscience 11:334-339. doi:10.1038/s41561-018-0109-5
34. Carr SA, Jungbluth SP, Eloe-Fadrosh EA, Stepanauskas R, Woyke T, Rappé MS, Orcutt BN (2019) Carboxydotrophy potential of uncultivated Hydrothermarchaeota from the subseafloor crustal biosphere. The ISME Journal 13 (9):1457-1468. doi:10.1038/s41396-019-0352-9