12-2018

Microbial Communities and Biogeochemistry in Marine Sediments of the Baltic Sea and the High Arctic, Svalbard

Joy Buongiorno
University of Tennessee, jbuongio@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Joy Buongiorno entitled "Microbial Communities and Biogeochemistry in Marine Sediments of the Baltic Sea and the High Arctic, Svalbard." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

Karen Lloyd, Major Professor

We have read this dissertation and recommend its acceptance:

Alison Buchan, Terry Hazen, Jill Mickuki, Andrew Steen

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Microbial Communities and Biogeochemistry in Marine Sediments of the Baltic Sea and the High Arctic, Svalbard

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Joy Buongiorno
December 2018

Dedication

To my beloved husband

Michael Patrick Altom

my dogs

Barney and Nyla

and my cats

Muffin, Oby, Oscar, and Phantom

Acknowledgements

My deepest gratitude belongs to my major advisor, Dr. Karen Lloyd, who intentionally cultivated a work culture of patience and encouragement that allowed me to flourish as an independent researcher and find the confidence that this path requires. I am forever indebted to the people who served as both mentors and friends to me along the way: Dr. Linda Kah, Ashley R. Manning-Berg, Sarah Sheffield, Jordan Bird, and Mallory Ladd. I would like to thank my committee for challenging me and believing that I was up to that challenge. Dr. Jill Mikucki, for inspiring me to find a deeper, more impactful purpose while on this path and Dr. Heidi Goodrich-Blair for ensuring that I had the means to achieve those purpose-driven milestones. The Ladies of the Lloyd lab for being sentinels of support and encouragement. My collaborators at The Center for Geomicrobiology at Aarhus University in Denmark, The University of Vienna, and Stony Brook University for insightful discussions. The C-DEBI family for contributing to my personal growth as an academic and for providing invaluable resources that made much of this work possible. The Simons Foundation, Explorer's Club, Geological Society of America, Society for Sedimentary Geology, Sandra White, and The Graduate School for providing significant or supplemental funding for this research. And finally, my loving husband, Michael Altom, for the daily sacrifices that this long journey required of him, which he always bore with patience and humility.

Abstract

Marine sediments contain more microorganisms than all of the world's oceans, with current of estimates of 1×10^{29} microorganisms. Despite marine sediments being replete with microbial cells, the majority of these microorganisms remain uncultured in the laboratory. At present, it is estimated that over 99% of all microorganisms have evaded culture, although truer estimates likely depend upon environment. Factors responsible for the intractability of these microorganisms include very slow doubling times, predicted to be on the orders of years to centuries, as well as special physiological needs of extremophiles. Unsuccessful laboratory growth of these microorganisms requires us to rely on culture-independent tools, including molecular techniques, metagenomics, and bioinformatic tools to glean insight into their ecological structure and function.

This dissertation combines molecular and bioinformatic techniques to evaluate the biosphere within deeply buried sediments of the Baltic Sea and shallow sediments in Arctic fjords. Quantification of microbial biomass within marine sediments lays the groundwork for questions related to organic carbon and element cycling. Although essential, reliable and reproducible estimates of microbial biomass within deeply buried sediments has proved challenging. Here we present an interlaboratory comparison of quantification results from International Ocean Discovery Program Exp. 347 sediments that allowed us to define best practices that lead to meaningful quantification estimates. We then transferred these best practices to marine sediments in a Svalbard fjord (Van Keulenfjorden) to understand how glacial proximity influences microbial communities. Through 16S rRNA gene libraries, organic geochemistry, and genome reconstruction, we illustrate that cross-fjord trends in organic matter influence community structure in the sediment. In addition, we argue that biological iron and

sulfur cycling facilitates rapid recycling of electron acceptors crucial for carbon oxidation. We delved deeper into their metabolic pathways with metagenomic sequencing and contig binning. We reconstructed several genomes of the Woeseiaceae clade that can act both as a sink and a source of carbon. Ultimately, our work provides a framework for understanding how glacial proximity influences microbial community composition and metabolic function, which is important and timely with ongoing climate change and a strong threat of severe glacial retreat in this region.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Tools for Estimating Abundance of Subsurface Microorganisms

Estimates suggest that the quantity of living microorganisms in deep marine sediments is greater than that in all the world's oceans (1-3). Although the subsurface biosphere is the largest habitat for bacteria and archaea on Earth (as reviewed in (4)), scientists have only recently begun to elucidate these microorganisms' role in organic matter degradation (5-7) and have a limited understanding of the strategies that have allowed these microorganisms to subsist across geological time scales (8, 9). Accurate quantification of these buried microorganisms is important for reliable models of carbon cycling and gas flux (7), as well as cell-specific rate calculations (10, 11). Models of cell-specific energy flux require accurate abundance estimates of cells performing a particular metabolism, including metal reduction (7), sulfate reduction (12, 13), and methanogenesis (14, 15), which together make up the main terminal processes by which organic carbon (C_{org}) is degraded in marine sediments. The availability of the C_{org} that fuel these populations is controlled by sediment accumulation rate (16), the amount of primary productivity in surface waters (17), and overlying water depth. Therefore, the biogeographic distribution and abundance of subsurface microorganisms is highly variable.

The range of microbial biomass in marine sediments worldwide is extremely vast and tied closely to geography. Deep-sea sediments within the South Pacific Gyre, characterized by low surface water phytoplankton and sediment accumulation rates, have the lowest microbial cell density at only $1x10^2$ cells cm^{-3} (18). By contrast, coastal sediments in the eutrophic Baltic Sea have abundances ~$1x10^{10}$ cm^{-3} (19). Microbial abundance in marine sediments has been quantified mainly through direct count microscopy techniques using general DNA stains, such as 4 ,6-diamidino-2-phenylindole (DAPI) or acridine orange. In extremely low biomass samples and/or ones in which sediment causes non-specific binding of dyes, cell separation techniques

(20) coupled to dyes like SYBR green which are used to discriminate between biological and non-biological fluorescent particles (SYBR-spam) (21) aid in microscopic quantitation of buried cells.

Phylogenetic identification of microbes in marine sediment requires additional methods beyond direct counting with a general DNA stain. Biomass-replete sediments or sediments containing an abundance of active microorganisms with cellular ribosome content can be examined microscopically with fluorescent *in situ* hybridization (FISH). This method allows the identification of specific taxa with a rRNA-targeted oligonucleotide probe linked to a fluorophore which fluorescently labels the ribosomes of target microorganisms. However, the energy-limitation of deeply buried marine sediments (22) necessitates alternative means of quantifying specific taxa, as energy-starved cells have low ribosomal contents. The use of enzymatic signal amplification with catalyzed-reporter deposition fluorescent *in situ* hybridization (23) is a useful alternative to circumvent the problems common to FISH (Reviewed in (24)). CARD-FISH was used for the single-cell identification of bacteria within deeply buried sediments of the Peru Margin (25, 26), but the notable failure to detect any archaea in these samples caused researchers to speculate that lack of detection was an artefact of enzymatic permeabilization protocols (27). Using domain-specific enzymes for cell wall permeabilization of cells in Baltic Sea sediments, Buongiorno et al. (2017) demonstrated that bias against archaea is not a methodological artefact and determined the quantification limit of CARD-FISH is actually much higher than previously recognized. As an alternative means for taxon-specific quantification, quantitative PCR (qPCR) showed to be more reliable and relatively reproducible across laboratories (19). New advances in culture-independent means for measuring cellular activity, such as biorthogonal non-canonical amino acid tagging (BONCAT) coupled to

FISH, show promise for answering questions related to substrate preference, relative activity levels, and phylogenetic identity of active members of marine sediment communities (28-30). Although these methods allow quantification, these techniques provide limited understanding of their physiology and genetic potential.

Metagenomics and Genome Reconstruction

The majority of microorganisms remain uncultured in the laboratory (31-33). The standing estimate suggests that over 99% of all microbial diversity evades culture. However, this estimate is currently being challenged, and truer estimates may be environment-specific (34). Factors responsible for the intractability of these microorganisms include special physiological needs of extremophiles who require conditions outside of what could easily be simulated in a laboratory for growth, including very slow doubling times. These doubling times have been predicted to be on the orders of years (35) to centuries (36). Unsuccessful growth of these microorganisms under laboratory settings requires reliance on culture-independent tools, such as metagenomics and bioinformatics, which can help us glean insight into the genomes of individual populations of bacteria and archaea.

Instead of assessing the genetic information from one microorganism or synthetic community grown in the lab, metagenomic sequencing allows the assessment of genetic information within an entire natural sample. Once enough DNA is extracted from an environmental sample for sequencing, sequenced reads can be fed into a number of downstream applications depending on the desired dataset. Metagenomes, for example, have come to replace the time-consuming method of generating clone libraries for Sanger sequencing. The low costs of sequencing coupled to availability of bioinformatic tools allows the separation, classification, and clustering of microbial 16S rRNA SSU sequences (37, 38). Additionally, mapping tools such

as Bowtie2 (39) that align short sequencing reads onto long genomic contigs allow high-throughput assessment of microbial diversity, as sequencing parameters such as read coverage help to calculate the abundance of each represented taxon. In addition to microbial community structure, functional potential of a sample can be accomplished with gene annotation followed by subsequent metabolic pathway mapping with tools such as KEGG (40). As a result, novel genes may be discovered to participate in unexpected pathways, or mis-annotation of genes may generate new hypotheses about the function of novel taxa. Finally, the clustering of sequences with similar genetic signatures, such as kmer frequency (gene motifs of size k) and coverage, allow the reconstruction of individual pan-genomes, or genomes of populations, from an environmental sample called a metagenome assembled genome (MAG). This sophisticated binning method of similar sequences is a more cost-effective method than single cell genomics (SAG), whereby individual cells are physically separated, lysed, and sequenced for their genetic information. The insights provided with these new sequencing approaches allows us to interrogate standing questions of how microorganisms will respond to, and potentially participate in, climate feedbacks.

'Omics for Understanding the Roles of Microorganisms in Climate Feedbacks

In addition to understanding the genetic potential of microorganisms, metagenomics can contribute to our understanding the reciprocal feedbacks between microorganisms and their environment as it relates to climate change. Microorganisms are one of many powerful agents of atmospheric change (reviewed in (41)); however, their environmental impact in the wake of a warming climate is difficult to predict. The trajectory of greenhouse gases, such as carbon dioxide and methane, is largely dependent upon differences in local soil organic matter content (42, 43), latitude (44), and microbial community composition (45, 46). In areas that are

especially sensitive to climate change feedbacks, such as high-latitude permafrost/soils (47, 48) and the Arctic marine realm (49), metagenomics is proving to be an increasingly useful tool for the development of better predictive climate change models (42, 50), although their predictive power has yet to be tested.

Despite advances in sequencing technology and bioinformatic tools, the vast diversity and complexity of climate-affected systems presents problems for straight-forward interpretation of metagenomics data alone. Interpretation of metagenomics data is enhanced when analyzed alongside a suite of complementary 'omic datasets inside a genome-based (MAG/SAG) framework. The ongoing development of metatranscriptomics, metaproteomics, and metametabolomics along with their respective databases has allowed researchers to detect the steps and products of microbial activity, beginning with DNA encoding and ending with metabolite production. Such corresponding datasets provide the necessary bridges between genetic potential of a genome to metabolite production useful in making predictions about substrate utilization, greenhouse gas emissions, and fluxes (50).

References

1. Parkes RJ, Cragg B, Roussel E, Webster G, Weightman A, Sass H. 2014. A review of prokaryotic populations and processes in sub-seafloor sediments, including biosphere: geosphere interactions. Marine Geology 352:409-425.
2. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. PNAS 109:16213-16216.
3. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: The unseen majority. PNAS 95:6578-6583.
4. Jørgensen BB, Marshall IP. 2016. Slow microbial life in the seabed. Annual review of marine science 8:311-332.
5. D'Hondt S, Inagaki F, Zarikian CA, Abrams LJ, Dubois N, Engelhardt T, Evans H, Ferdelman T, Gribsholt B, Harris RN, Hoppie Bryce W, Hyun J-H, Kallmeyer J, Kim J, Lynch JE, McKinley Claire C, Mitsunobu S, Morono Y, Murray RW, Pockalny R, Sauvage J, Shimono T, Shiraishi F, Smith DC, Smith-Duque Christopher E, Spivack AJ, Steinsbu BO, Suzuki Y, Szpak M, Toffin L, Uramoto G, Yamaguchi YT, Zhang G-l, Zhang X-H, Ziebis W. 2015. Presence of oxygen and aerobic communities from sea floor to basement in deep-sea sediments. Nature Geoscience 8:299.
6. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S. 2013. Predominant archaea in marine sediments degrade detrital proteins. Nature 496:215-218.
7. D'Hondt S, Jørgensen BB, Miller DJ, Batzke A, Blake R, Cragg BA, Cypionka H, Dickens GR, Ferdelman T, Hinrichs K-U, Holm NG, Mitterer R, Spivack A, Wang G, Bekins B, Engelen B, Ford K, Gettemy G, Rutherford SD, Sass H, Skilbeck CG, Aiello IW, Guèrin G, House CH, Inagaki F, Meister P, Naehr T, Niitsuma S, Parkes RJ, Schippers A, Smith DC, Teske A, Wiegel

J, Padilla CN, Acosta JLS. 2004. Distributions of Microbial Activities in Deep Subseafloor Sediments. Science 306:2216-2221.

8. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, Sano Y, Inagaki F. 2011. Carbon and nitrogen assimilation in deep subseafloor microbial cells. Proceedings of the National Academy of Sciences:201107763.

9. Røy H, Kallmeyer J, Adhikari RR, Pockalny R, Jørgensen BB, D'Hondt S. 2012. Aerobic microbial respiration in 86-million-year-old deep-sea red clay. Science 336:922-925.

10. Parkes RJ, Cragg BA, Fry JC, Herbert R, Wimpenny J, Allen J, Whitfield M. 1990. Bacterial Biomass and Activity in Deep Sediment Layers from the Peru Margin [and Discussion]. Philos Trans R Soc London, Ser A 331:139-153.

11. Altmann D, Stief P, Amann R, De Beer D, Schramm A. 2003. *In situ* distribution and activity of nitrifying bacteria in freshwater sediment. Environ Microbiol 5:798-803.

12. Sahm K, MacGregor BJ, Jørgensen BB, Stahl DA. 1999. Sulphate reduction and vertical distribution of sulphate-reducing bacteria quantified by rRNA slot-blot hybridization in a coastal marine sediment. Environ Microbiol 1:65-74.

13. Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C, Jørgensen BB. 2009. Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. Environ Microbiol 11:1278-1291.

14. Biddle JF, Lipp JS, Lever MA, Lloyd KG, Sørensen KB, Anderson R, Fredricks HF, Elvert M, Kelly TJ, Schrag DP, Sogin ML, Brenchley JE, Teske A, House CH, Hinrichs K-U. 2006. Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. PNAS 103:3846-3851.

15. D'Hondt S, Rutherford S, Spivack AJ. 2002. Metabolic activity of subsurface life in deep-

sea sediments. Science 295:2067-2070.

16. Henrichs SM, Reeburgh WS. 1987. Anaerobic mineralization of marine sediment organic matter: Rates and the role of anaerobic processes in the oceanic carbon economy. Geomicrobiology Journal 5:191-237.

17. Davies JM, Payne R. 1984. Supply of organic matter to the sediment in the northern North Sea during a spring phytoplankton bloom. Marine Biology 78:315-324.

18. D'Hondt S, Spivack AJ, Pockalny R, Ferdelman TG, Fischer JP, Kallmeyer J, Abrams LJ, Smith DC, Graham D, Hasiuk F. 2009. Subseafloor sedimentary life in the South Pacific Gyre. PNAS 106:11651-11656.

19. Buongiorno J, Turner S, Webster G, Asai M, Shumaker AK, Roy T, Weightman A, Schippers A, Lloyd KG. 2017. Interlaboratory quantification of Bacteria and Archaea in deeply buried sediments of the Baltic Sea (IODP Expedition 347). FEMS microbiology ecology 93.

20. Morono Y, Terada T, Kallmeyer J, Inagaki F. 2013. An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. Environ Microbiol 15:2841-2849.

21. Morono Y, Terada T, Masui N, Inagaki F. 2009. Discriminative detection and enumeration of microbial life in marine subsurface sediments. The ISME Journal 3:503.

22. Jørgensen BB. 2011. Deep subseafloor microbial cells on physiological standby. PNAS 108:18193-18194.

23. Pernthaler A, Pernthaler J, Amann R. 2002. Fluorescence *in situ* hybridization and catalyzed reporter deposition for the identification of marine bacteria. Appl Environ Microbiol 68:3094-3101.

24. Amann R, Fuchs BM. 2008. Single-cell identification in microbial communities by

improved fluorescence *in situ* hybridization techniques. Nature Reviews Microbiology 6:339.

25. Schippers A, Neretin LN, Kallmeyer J, Ferdelman TG, Cragg BA, Parkes RJ, Jørgensen BB. 2005. Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. Nature 433:861-864.

26. Teske AP. 2005. The deep subsurface biosphere is alive and well. Trends in Microbiology 13:402-404.

27. Lloyd KG, May MK, Kevorkian RT, Steen AD. 2013. Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. Appl Environ Microbiol 79:7790-7799.

28. Hatzenpichler R, Orphan VJ. 2015. Detection of protein-synthesizing microorganisms in the environment via bioorthogonal non-canonical amino acid tagging (BONCAT). Hydrocarbon and Lipid Microbiology Protocols:145.

29. Hatzenpichler R, Scheller S, Tavormina PL, Babin BM, Tirrell DA, Orphan VJ. 2014. *In situ* visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environmental microbiology 16:2568-2590.

30. Hatzenpichler R, Connon SA, Goudeau D, Malmstrom RR, Woyke T, Orphan VJ. 2016. Visualizing *in situ* translational activity for identifying and sorting slow-growing archaeal–bacterial consortia. Proceedings of the National Academy of Sciences 113:E4069-E4078.

31. Rappé MS, Giovannoni SJ. 2003. The uncultured microbial majority. Annual Reviews in Microbiology 57:369-394.

32. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature 499:431-437.

33. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nature microbiology 2:1533.

34. Lloyd KG, Ladau J, Steen AD, Yin J, Crosby L. 2018. Phylogenetically novel uncultured microbial cells dominate Earth microbiomes. bioRxiv:303602.

35. Kevorkian R, Bird JT, Shumaker A, Lloyd KG. 2018. Estimating Population Turnover Rates by Relative Quantification Methods Reveals Microbial Dynamics in Marine Sediment. Applied and environmental microbiology 84:e01443-17.

36. Hoehler TM, Jørgensen BB. 2013. Microbial life under extreme energy limitation. Nat Rev Micro 11:83-94.

37. Miller CS, Baker BJ, Thomas BC, Singer SW, Banfield JF. 2011. EMIRGE: reconstruction of full-length ribosomal genes from microbial community short read sequencing data. Genome biology 12:R44.

38. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211-3217.

39. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359.

40. Kanehisa M, Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28:27-30.

41. Singh BK, Bardgett RD, Smith P, Reay DS. 2010. Microorganisms and climate change: terrestrial feedbacks and mitigation options. Nature Reviews Microbiology 8:779.

42. Lau MCY, Stackhouse BT, Layton AC, Chauhan A, Vishnivetskaya TA, Chourey K, Ronholm J, Mykytczuk NCS, Bennett PC, Lamarche-Gagnon G, Burton N, Pollard WH, Omelon

CR, Medvigy DM, Hettich RL, Pfiffner SM, Whyte LG, Onstott TC. 2015. An active atmospheric methane sink in high Arctic mineral cryosols. The Isme Journal 9:1880.

43. Schuur EA, Vogel JG, Crummer KG, Lee H, Sickman JO, Osterkamp T. 2009. The effect of permafrost thaw on old carbon release and net carbon exchange from tundra. Nature 459:556.

44. Marotta H, Pinho L, Gudasz C, Bastviken D, Tranvik LJ, Enrich-Prast A. 2014. Greenhouse gas production in low-latitude lake sediments responds strongly to warming. Nature Climate Change 4:467.

45. Allison SD, Treseder KK. 2008. Warming and drying suppress microbial activity and carbon cycling in boreal forest soils. Global change biology 14:2898-2909.

46. Steidinger B, Bhatnagar J, Vilgalys R, Taylor J, Bruns T, Peay KG. 2018. Global climate changes will lead to regionally divergent trajectories for ectomycorrhizal communities in North American Pinaceae forests. bioRxiv:393009.

47. Zhuang Q, Melillo JM, Kicklighter DW, Prinn RG, McGuire AD, Steudler PA, Felzer BS, Hu S. 2004. Methane fluxes between terrestrial ecosystems and the atmosphere at northern high latitudes during the past century: A retrospective analysis with a process-based biogeochemistry model. Global Biogeochemical Cycles 18:3.

48. Bradley JA, Singarayer JS, Anesio AM. 2014. Microbial community dynamics in the forefield of glaciers. Proc. R. Soc. B 281:1795.

49. Wassmann P, Duarte CM, Agusti S, Sejr MK. 2011. Footprints of climate change in the Arctic marine ecosystem. Global change biology 17:1235-1249.

50. Woodcroft BJ, Singleton CM, Boyd JA, Evans PN, Emerson JB, Zayed AAF, Hoelzle RD, Lamberton TO, McCalley CK, Hodgkins SB, Wilson RM, Purvine SO, Nicora CD, Li C, Frolking S, Chanton JP, Crill PM, Saleska SR, Rich VI, Tyson GW. 2018. Genome-centric view of carbon

processing in thawing permafrost. Nature 560:49-54.

Chapter 2: Inter-laboratory quantification of Bacteria and Archaea in deeply buried sediments of the Baltic Sea (IODP Exp. 347)

This chapter is a revised version of published work:

Buongiorno, Joy, Stephanie Turner, Gordon Webster, Masanori Asai, Alexander K. Shumaker, Taylor Roy, Andrew Weightman, Axel Schippers, and Karen G. Lloyd. "Interlaboratory quantification of Bacteria and Archaea in deeply buried sediments of the Baltic Sea (IODP Expedition 347)."FEMS Microbiology Ecology 93, no. 3 (2017).

My primary contributions to this paper include: (i) microscopy and CARD-FISH data collection, (ii) data synthesis and statistical analysis, (iii) figure and table generation, (iv) and writing the manuscript.

Abstract

Two common quantification methods for sub-seafloor microorganisms are catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) and quantitative PCR (qPCR). Using these methods, we quantified bacteria and archaea in Baltic Sea basin (IODP Exp. 347) sediments down to 90 meters below sea floor (mbsf), testing the following in an inter-laboratory comparison: 1) proteinase K permeabilization of archaea increases CARD-FISH accuracy, and 2) qPCR varies by more than an order of magnitude between laboratories using similar protocols. CARD-FISH counts did not differ between permeabilization treatments. Thus, proteinase K did not increase accuracy of CARD-FISH counts, however, 91% of these counts were below the quantification limit of 1.3×10^7 cells cm^{-3}. qPCR data varied between laboratories but were largely within the same order of magnitude if the same primers were used, with 88% of samples being above the quantification limit. Yields were elevated by preparing a sediment slurry before DNA extraction: 3.88×10^6 to 2.34×10^9 copies cm^{-3} vs. 1.39×10^7 to 1.87×10^9 total cells cm^{-3}. By qPCR, bacteria were more abundant than archaea, although they

usually were within the same order of magnitude. Overall, qPCR is more sensitive than CARD-FISH, but both require optimization to consistently achieve both precision and accuracy.

Introduction

Estimates of the global distribution and abundance of microorganisms suggest that there is nearly as much living microbial matter in deep marine sediments as there is in all of the world's oceans (1). Energetic considerations of the subseafloor biosphere are giving great insights into the range of possibilities for life on Earth (2). Although presumably very energy limited, this expansive subsurface microbiome has been suggested to play key roles in global biogeochemical cycles (3, 4). The deeply buried bacteria and archaea within marine sediments remain elusive, as they are dominated by clades with no cultured representatives (5, 6). Because of this, our current understanding of the subsurface biosphere is based on information obtained from culture-independent molecular techniques, which have revealed that the microbial community is composed of phylogenetically and physiologically diverse members. Differences in physiological characteristics of bacteria and archaea may define their relative abundance within marine sediments, however, a consistent method for their quantification has yet to be established (7, 8).

Three methods are commonly used to quantify specific bacteria and archaea in marine environments; fluorescence *in situ* hybridization (FISH) (9), catalyzed reporter deposition FISH (CARD-FISH) (10), and quantitative PCR (qPCR). In FISH and CARD-FISH, individual cells that fluoresce with a DNA probe matching the target ribosome primary sequence can be enumerated using a microscope. This allows the recognition of taxonomically identifiable targets. In CARD-FISH, the fluorescence is amplified when a large horseradish peroxidase bound to the DNA probe affixes fluorescent tyramides to cellular proteins. CARD-FISH appears to be necessary to visualize cells from low activity marine sediments and requires that cell walls be permeabilized with an enzyme to allow the large enzyme to enter (10). Biomass can also be

estimated from qPCR when physiologically or taxonomically diagnostic genes are amplified from DNA extracts and quantified with automated fluorescence measurements.

Because CARD-FISH and qPCR can be used to quantify phylogenetically or physiologically distinct microorganisms, these data are crucial for precise estimates of cell-specific maintenance requirements and energy flux in the deep subsurface (11). Calculations of cell turnover and element cycling within complex subseafloor communities have demonstrated that there is a spectrum of cell-specific metabolic capabilities coinciding with sedimentary organic matter content (12), sediment age (13, 14), and temperature. Greater precision in cellular rates is especially important for cells that are presumed to have turnover rates on the scale of hundreds to thousands of years in sediment habitats (11, 15-17). Energetic considerations of microbial life in the deep subsurface have shed light onto the maintenance requirements for life on Earth (see (2, 18) for review). However, few studies undertake cell-specific rate calculations, which require reliable quantification of cells performing a particular metabolism (19-21) as well as reliable measurement of the metabolism in question. Some specific energy flux calculations, such as those related to sulfate reduction, are based on assumptions about the relative proportion of sulfate reducers within the sedimentary microbial community (4). While this is a valid assumption when making first order approximations of energetic limitation, greater precision in cell-specific energy flux is achieved when geochemical speciation data is coupled to cell quantification data (20, 22-24).

Each of the cell quantification techniques useful in energetic models provides their own set of limitations that possibly lead to over- or under-estimating the numbers of living cells. For example, inadequate permeabilization of archaea during CARD-FISH (8), variable extraction efficiencies of DNA (25, 26), and biased primers used in qPCR (6) could potentially lead to the

over- or under-representation of community members. Because of this, we quantified bacteria and archaea in sediment samples collected from IODP Expedition 347, Baltic Sea Paleoenvironment in three separate laboratories to assess the degree of replicability in independently-working groups. The Baltic Sea contains a sedimentological record spanning the transition from the last glacial maximum to the current interglacial period, allowing the investigation of any shifts in microbial abundance in accord with climate shifts.

We examined microbial abundance across four sites (five bore holes) within the Baltic Sea basin through qPCR and CARD-FISH, as compared to total cell counts. We compared results obtained from overlapping sediment samples examined by three independent laboratories employing similar analytical procedures. Specifically, we tested the following hypotheses that were proposed from a recent meta-analysis of published CARD-FISH and qPCR data (27): i) proteinase K permeabilization of archaea increases CARD-FISH accuracy over the more commonly used permeabilization with lysozyme, and ii) qPCR varies by more than 10-fold between laboratories using similar DNA extraction protocols due to the random variability of DNA extraction efficiencies. In addition, we applied a new non-enzymatic whole-cell reporter method called fluorescence *in situ* hybridization chain reaction (DNA-HCR) to assess its utility for cell detection in marine sediment samples (28, 29).

Methods

Sample collection

Samples were collected during IODP Expedition 347 at Baltic Sea sites M0059 (Little Belt; holes C and E), M0060 (Anholt Loch; hole B), M0063 (Landsort Deep; hole E), and M0065 (Bornholm Basin; hole C) (Figure 2.1; 30) in 2013. Latitude and longitude data for each site can be found on the Biological and Chemical Oceanography Data Management Office

(BCO-DMO) website (http://www.bco-dmo.org/dataset/641342/data). Site M0060 had very low organic matter content (<1% TOC), whereas the other sites were organic-rich upper marine deposits (3-8 % TOC) overlying organic-poor (<0.5% TOC) lacustrine deposits (30). Onboard ship, perfluorocarbon tracer (PFC) testing was conducted in order to assess contamination. Because all core exteriors contain known contamination (30), the samples in this study were collected only from the interior of cores where contamination was minimal. Whole round cores were frozen at -80°C shipboard for later DNA extraction and qPCR in Hannover (Germany), Cardiff (UK), and Knoxville (TN, US) or shipped at 4°C and stored under nitrogen in sealed aluminum bag within one week to Hannover (Germany), where they were fixed for CARD-FISH with 4% formaldehyde solution following previously published protocols (20). A subset of the samples fixed in Hannover were shipped at 4°C to Knoxville (TN) for comparison analyses.

Total cell counts and CARD-FISH

Total cell numbers were determined for fixed sonicated samples through direct cell counts with SYBR Gold, SYBR Green I, and 4', 6-diamidino-2-phenylindole (DAPI) DNA staining. Epifluorescence microscopy was conducted on a Zeiss Axio Imager M2 with Axiocam MRM (Lloyd lab) and an Olympus BX60 (Schippers Lab). Probes ARCH915 and EUB338 (I – III mix) were used for archaea and bacteria, respectively (Table 2.1). Permeabilization of cell membrane to allow the entrance of the horseradish peroxidase (HRP) enzyme was carried out at either 37°C for 30 min (10 mg ml^{-1} proteinase K, (31)) or 1 h (10 mg ml^{-1}, lysozyme, Pernthaler *et al.* 2002) (Schippers Lab) or at room temperature for 20 min (Lloyd lab). Filter sections were treated with either lysozyme, proteinase K, or both enzymes together. For each sample and treatment (lysozyme, proteinase K), three separate filters were analyzed (Schippers Lab). The Lloyd lab did not count replicate filters. Cell disruption by permeabilization treatment was tested

on triplicate DAPI-stained filters. Counts were performed blind for the permeabilization treatment experiment. NON338, which is antisense to EUB338, was used as negative control (32).

Cell counts below the quantification limit for CARD-FISH were not considered in any comparison analyses, although they are plotted for completeness. The quantification limit for CARD-FISH is defined as falling within the 95% confidence interval around 30% of the mean (33). We therefore defined our quantification limit as 30 cells counted per 30 fields of view for one sample, or $1.3x10^7$ cells/cm^3. Depths for which either bacteria or archaea was below quantification limit were not evaluated for relative archaeal abundance. Raw counts were transformed to milliliters of wet sediment by dividing raw counts by depth-specific density to account for porosity of clay-rich sediment ((30); Lloyd lab).

To test the agreement of total cell counts obtained in the lab with those acquired shipboard, SYBR Green I and SYBR Gold datasets (reported from the Schippers and Lloyd Labs, respectively) were tested against AODC shipboard counts using a paired Wilcoxon singed rank test in the R package, version 3.1.1 (34). In the same manner, SYBR Green I and SYBR Gold counts were also tested. Values were considered to be significantly different when $p \leq 0.05$.

To assess the degree of CARD-FISH success, yield was determined for CARD-FISH counts relative to total cell numbers (27). A yield of 1 denotes that 100% of the microbial biomass is accounted for. Using the R package, permeabilization treatment for CARD-FISH was tested with a two sample Welch t-test on log-transformed data of yield. Yield was defined as the sum of bacteria and archaea counts divided by the total cell counts detected within the same laboratory through a non-specific DNA stain, SYBR Gold (27) for all sites except M0063E, for

which SYBR Green I counts were used (Schippers Lab). Welch t-test was used because of variability in sample size, and transformations were performed to make data normally distributed. An additional Wilcoxon signed rank test, which does not assume normality of data, was performed on raw counts above the quantification limit produced from each treatment. Values were considered to be significantly different when $p \leq 0.05$.

DNA extraction and qPCR analyses

Genomic DNA was extracted from Baltic Sea basin sediments using the FastDNA® Spin Kit for Soil (MP Biomedicals) by all three laboratories. However, a slightly modified version of the kit protocol was used by the Schippers and Weightman Labs (35). The Schippers Lab added poly-adenylic acid to the lysis mixture. Further, the Weightman and Schippers Labs tested an additional step by creating a slurry with the sediment before beginning the extraction. Therefore, 0.5 g of sediment was placed in a lysing matrix E tube (MP Biomedicals) with 200 µl of sodium phosphate buffer (MP Biomedicals) and then shaking for 10 min on a wrist action shaker at maximum speed (to break-up the 'sticky' Baltic clay sediments). The resulting sediment slurry was then further shaken for 5 min with 800 µl of sodium phosphate buffer and 120 µl MT buffer (MP Biomedicals) before lysis in a FastPrep® 24 instrument (MP Biomedicals) for 2x 30s, speed 5.5 m $*s^{-1}$. All remaining steps followed the manufacturer's protocol, except that some spin and incubation times were extended. DNA was eluted in 75 µl (Schippers Lab) or 100 µl molecular grade water (Severn Biotech Ltd.) and stored at -20°C (Schippers Lab) or -80°C until required. In an additional experiment, the Schippers Lab used the modified slurry extraction method and additionally treated samples with either hydrochloric acid, hydroiodic acid (25), or without acid addition. For all extraction protocols a non-sample control was extracted.

Total bacterial and archaeal 16S rRNA gene copy numbers were quantified with qPCR using primers listed in Table 2.1. Extracted DNA was amplified with a StepOnePlus™ Real-Time PCR System (Schippers Lab; Applied Biosystems, Carlsbad, California), BioRad iQ5 (Lloyd lab; Applied Biosystems, Foster City, California), and Mx3000P QPCR System (Weightman Lab; Agilent Technologies UK). Serial dilutions of full-length 16S rRNA gene PCR products from *Anaerolinea thermophile* DSM 14523 and *Methanococcoides methylutens* DSM 2657 (Weightman Lab), plasmids containing amplified partial 16S rRNA genes (Lloyd lab), and 16S rRNA gene PCR products of *Escherichia coli* (36), *Methanohalobium evestigatus* (37), and *Methanosarcina barkeri* (38) (Schippers Lab) were used as standards for bacteria and archaea for qPCR. Sterilized sand or water was used as a negative control. Results of qPCR were rejected if the R^2 of the standard curve was below 0.95, or if the melt curve showed evidence of primer dimers. The quantification limit was defined as having fluorescence threshold cycle numbers (Ct) well within those of the simultaneously-run standard curve and being at least 3 Ct below the non-template control Ct. TaqMan assays were used in the Schippers lab (36, 37) and SYBR green chemistry was used for all other reactions. Different master mixes were used from the companies Invitrogen (Lloyd and Schippers Labs), Quanta Biosciences (Schippers Lab, assay (37), or PCR Biosystems Ltd (Weightman Lab). Gene copy numbers were corrected for non-sample extraction control (Schippers Lab and Weightman Lab) and were converted into copies cm^{-3} wet sediment. This conversion was carried out by multiplying copy number g^{-1} dry weight by the sample's dry weight in g and the depth-specific density (g cm^{-3}; (30)), the product of which was then divided by the sample's wet weight in g. qPCR results were directly compared to total cell counts for assessment of qPCR accuracy. As copy number of 16S rRNA gene varies both phylogenetically (39) and with lifestyle (40), multiplicities of the 16S rRNA gene for

bacteria and archaea were taken into account (3.04 to 24 copies per genome; c.f. (27)). qPCR data from the Lloyd lab have been deposited in the databank managed by BCO-DMO (http://www.bco-dmo.org/dataset/641358/data).

The relative fraction of archaea within the community was determined by dividing the number of archaea reported across laboratories by the sum of bacteria and archaea counts from qPCR (27). Because the Schippers Lab used two primer sets for archaea, we used the larger of the two values for archaea in our computation. For comparison of qPCR data across laboratories, paired Wilcoxon signed rank test was performed on non-transformed qPCR copy numbers. Because qPCR was conducted on different whole round cores in the separate laboratories, copy numbers from the same 5 m intervals were compared (see Table 2.4 footnote for depths). Yield is defined as the combination of bacterial and archaeal 16S rRNA gene copy numbers divided by SYBR Gold or SYBR Green I (M0063E) cell counts.

DNA-HCR

DNA-hybridization chain reaction (HCR) was conducted in the Schippers and Lloyd Labs according to Yamaguchi *et al.* (2015) with the following modifications: both bacteria and archaea initiator probes were designed to hybridize the same amplifier C1 and C2 probes; one new mismatch probe was designed to test specificity (Table 2.1); and, when noted in the results, 10% blocking reagent was added to hybridization buffer to enhance stringency in the mismatch probe experiment. Fixed cells were embedded with 0.1% low melting point agarose on 0.2 µm polycarbonate filters (Whatman). Cells were permeabilized for 30 min at 37°C with either 1 mg ml^{-1} lysozyme (bacteria) or 1 mg ml^{-1} proteinase K (archaea). Following a 15 min rinse in Tris-NaCl-Tween buffer (TNT), filters were rinsed with MilliQ and 95% ethanol and allowed to air dry. Hybridization buffer (1 ml 1 M Tris-HCl pH 7.5, 0.9 M NaCl, 25 µl 20% SDS, and X%

formamide [see Table 2.1 for the value of X], and 0.5 µM initiator probe) was placed on parafilm-covered microscope slides. Dry filters were placed face-down on hybridization probe mix and hybridization was allowed to take place overnight at 47°C in a sealed humidification chamber. Excess hybridization probe was removed by washing filters for 30 min at 48°C in warm washing buffer (0.5 ml 0.5M EDTA pH 8.0, 1 ml 1M HCl pH 8, 25 µl 20% SDS, and either 2,150 µl (bacteria) or 460 µl (archaea) 5M NaCl). Separate tubes each containing 5 µM C1 and C2 amplifier probes were prepared with amplification buffer (0.9 M NaCl, 0.67 g $Na_2HPO_4.7H_2O$, 25 µl 20% SDS) and heated for 90 s at 95°C, then kept at 25°C for 30 min. Amplifier tubes were then mixed together for a final probe concentration of 2.5 µM each. After formamide was removed from filters with a brief wash in amplification buffer (without probe), filters were placed face-down on new parafilm-covered microscope slide containing the amplification probe mix. Hybridization occurred in a sealed humidification chamber for two hours at 46°C. Probe dissociation was prevented with 30 minute rinse in wash buffer at 4°C. Finally, filters were rinsed with MilliQ water and 95% ethanol and allowed to air dry before mounting on microscope slide with Vectashield (Vector Laboratories, California). All solutions were autoclaved and filter-sterilized. Hybridization with amplifier probes in the absence of initial initiator, with NON338 or with EUB338 with three mismatches were carried out to test specificity of signal amplification in DNA-HCR.

Results

Total cell counts

Cells were detected with DAPI and SYBR Green I or SYBR Gold at all depths that were analyzed at sites M0059, M0060, M0063, and M0065 of IODP Leg 347 (Figure 2.2). Cell morphology was dominated by cocci, although rods and *Vibrio*-shaped cells were also common in shallower (< 10 mbsf) depths (Figure 2.3a and 2.3b). Total cell counts for all sites showed

high microbial abundance, often exceeding 1 x 10^9 cells cm^{-3} in sediments shallower than 10 mbsf (Figure 2.2, Row A; Table 2.2). These cell counts exceed the global predicted regression for cell abundance with depth (41)(Parkes *et al.* 2000), as was also observed with flow cytometry and acridine orange direct counts (AODC) acquired on board during Expedition 347 (30).

Total cells produced using SYBR Green I in the Schippers Lab accounted for 74 ± 75% of shipboard AODC (Figure2. 2), with slightly diminished yields likely due to two additional washing steps introduced to samples. The Lloyd lab had similarly high yields relative to AODC of 129 ± 164% with SYBR Gold. SYBR Green I counts are not statistically different from AODC counts (paired Wilcoxon signed rank test p = 0.053; Figure 2.2, Row A). However, SYBR Gold counts were found to be statistically different from AODC values (p = 0.026). Despite this, the SYBR Gold and SYBR Green I datasets are statistically similar to each other (p = 0.098), indicating low operator bias during counting in the Lloyd and Schippers Labs.

CARD-FISH

Out of summed total of 716 samples and replicates examined between the two labs, only 67 were above the quantification limit of 30 cells per field of view (Lloyd 2014). Downcore counts are shown in Figure 2.4; however, the failure of CARD-FISH counts to reach the quantification limit in the majority of cases (as illustrated by empty symbols) prevents the interpretation of a true downcore profile. Further, no single overlapping depth was simultaneously above the quantification limit for CARD-FISH of either bacteria or archaea in both labs, precluding any interlab comparison for CARD-FISH (Figure 2.4; Table 2.3). No statistically significant difference was detected for the numbers of DAPI-stained cells treated with lysozyme or proteinase K, indicating permeabilization treatments did not differentially or detrimentally disrupt cells (p value of Welch two-sample-tests > 0.1, n = 33, data not shown). In

all samples, CARD-FISH counts with the NON338 rRNA probe were below the quantification limit. The number of samples above the quantification limit slightly increased by 4% for bacterial counts using lysozyme and about 5% for archaeal counts using proteinase K (Figure 2.5A), but these differences are not statistically significant (Welch two-sample-tests $p > 0.1$; Table 2.3). Interestingly, using lysozyme and proteinase K together decreased the proportion of samples meeting the quantification limit (Figure 2.5A); however, no significant yield differences were observed for lysozyme, proteinase K, or both together (Wilcoxon signed rank test $p > 0.05$ in all comparisons). In addition, no significant difference for percent archaea was observed across treatments (Figure 2.5B). CARD-FISH suggests that bacteria and archaea are equally distributed (Figure 2.4). However, because very few depth intervals contained counts where both domains were found to be above the limit of quantification, we cannot assess the relative abundance of bacteria and archaea.

When considering the data above the quantification limit, combined bacteria and archaea counts provide low yield relative to SYBR Green I counts, accounting only for 10% ± 13% of total cells ($n = 15$). The effect of cell loss during CARD-FISH processing was assessed by comparing DAPI cells counted after performing the CARD-FISH procedure to AODC counts. The DAPI cells were statistically lower than AODC counts (Figure 2.2, Row C, paired t-test $p = 0.01$), indicating that cell loss during CARD-FISH processing is a factor that decreases yield. However, when CARD-FISH bacteria and archaea counts above the quantification limit are combined and compared to DAPI counts, CARD-FISH yield increased to 67% ± 75%, indicating that cell loss during processing alone does not account for low yield.

The choice of permeabilization solution played no role in yield loss, with post-CARD-FISH DAPI counts being significantly lower than AODC for both lysozyme ($p = 0.0128$) and

proteinase K ($p = 0.0122$) treatments. Cell numbers decreased during CARD-FISH processing at all sites, and in both the Lloyd and Schippers Labs (Figure 2.2, gray and black squares). However, it should be noted that the Lloyd lab did not count triplicate filters, and so error bars are not reported in Figure 2.2. To test whether transport and shipment results in similar yield loss, we compared the yields before ($n = 45$) and after CARD-FISH processing with different permeabilization solutions (lysozyme, $n = 22$; proteinase K, $n = 22$). Results of the Welch two-sample t-tests show that yields acquired before and after CARD-FISH are statistically different from each other. This is true when considering both permeabilization treatments separately (lysozyme, $p < 0.001$; proteinase K, $p = <.001$) or together ($p < 0.001$, $n = 44$).

qPCR

The majority of qPCR counts of bacterial and archaeal 16S rRNA gene copies were above the quantification limit for all runs combined (87% of archaea and 100% of bacteria for the Lloyd lab, 77% of archaea and 67% of bacteria for the Schippers Lab, and 100% of both archaea and bacteria for the Weightman Lab; Figure 2.6). Only qPCR results for the Lloyd and Schippers Labs using the same primer sets (Bac340/Bac806r with Bac probe and Arch915f/Arch1059r) and the basic extraction protocol (Fast DNA Spin kit) were included in the inter-laboratory comparison for qPCR. Since the Weightman Lab used a slightly different extraction protocol, these results will be discussed below. There was decent agreement between the Schippers and Lloyd lab qPCR measurements for both bacterial and archaeal 16S rRNA gene copy numbers (Figure 2.6, purple triangles and orange squares). When samples within 5 meters of vertical depth from each other are binned, the Schippers and Lloyd Lab qPCR copy numbers are not statistically different from each other for bacteria in 1/3 cases and in 2/3 cases for archaea ($p > 0.05$; Table 2.4). However, bacterial and archaeal qPCR measurements from the Schippers

lab were generally lower than those of the Lloyd lab (14/18 and 17/20 measurements, respectively), and were 30 and 38% of Lloyd values, respectively. Lower values recorded by the Schippers Lab may be due to the Schippers Lab's use of a more specific TaqMan assay instead of SYBR green chemistry. These results show that while it is possible to reproduce copy numbers within the same order of magnitude in independent laboratories, results are not consistently reproducible, and therefore, precision is lost. Smaller variability in archaeal copy numbers was also observed between primer sets. Depths for which archaeal 16S rRNA genes were amplified with both Arch915f/Arch1059r and Arch349f/Arch806r primer sets by the Schippers Lab demonstrated that the Arch915f/Arch1059r primer set produced higher copy numbers (29/32 depths); however, the difference was too small to be significant (paired t-test for all sites, $p > 0.1$).

The methodological changes introduced by the Weightman Lab, which included an additional slurry preparation step, increased copy numbers of bacteria and archaea at most sites relative to the Schippers and Lloyd datasets (Figure 2.6). In some cases, this improvement was exceptionally great (e.g. M0060B), resulting in statistical differences between Weightman values and those of the Lloyd and Schippers Lab (Table 2.4). Although some sites had no improvement, with copy numbers that were statistically the same as those measured in the Lloyd and Schippers Labs (pluses in Table 2.4), at none of the sites was the Weightman data significantly lower than those of the Schippers and Lloyd Labs for bacteria or archaea. In fact, when compared to the combined quantification efforts of the Lloyd and Schippers Labs, the Weightman Lab had the highest abundances of bacteria and archaea in 100% and 74% of the cases, respectively. Although relatively greater copy numbers were produced by the Weightman Lab overall, qPCR results for archaea were statistically the same as those measured in the Lloyd Lab (Table 2.4).

The main differences in the Weightman data compared to the Lloyd and Schippers datasets lie with bacterial copy numbers, which were statistically different in the majority of cases (3/5 comparisons; Table 2.4).

The yield of qPCR counts was highly variable with depth, as well as across labs and cores. For example, the Weightman data had relatively high yields for all qPCR measurements made for site M0063E (between 0.01 and 1.7). However, qPCR measurements of site M0059E within the same lab demonstrate that bacterial and archaeal 16S rRNA gene copy numbers converge on total cell counts only in the upper sediment layers (Figure 2.6). Despite diminishing accuracy of measurements with depth, the Weightman Lab yield average for site M0059E is 0.94, considerably higher than that of the Schippers (0.03) or Lloyd Labs (0.14). Consistent with yield values, the Weightman Lab demonstrates the most accurate quantification across laboratories. Nine out of 22 measurements fall either on the black 1:1 line (Figure 2.6) or within the known range of copy numbers of 16S rRNA genes in a genome (average 3.04, dark blue, maximum 24, light blue; cf. (27)).

Efforts to raise the yield of qPCR measurements were performed by incorporating additional wash and slurry preparation steps in the Schippers Lab on site M0063E samples. Archaeal 16S rRNA gene copy numbers increased with the addition of a slurry step in 12 out of the 14 depths examined with both archaeal primers (Arch915f/Arch1059r and Arch349f/Arch806r) (Figure 2.5A). Notably, however, yields were systematically diminished with the addition of a preceding hydrochloric acid or hydroiodic acid wash step (red and green symbols in Figure 2.7, respectively). In contrast, bacterial 16S rRNA gene copy numbers were not improved from original values (black diamonds, Figure 2.7B) with any protocol modification.

Across all sites, bacteria are usually higher in qPCR abundance relative to archaea and their dominance did not follow a pattern downcore (black line, Figure 2.6). However, archaea regularly comprise up to 50% or higher of total qPCR counts, suggesting that archaea, although in lower abundance than bacteria, are nevertheless a numerically significant portion of the subsurface population.

DNA-HCR

The three types of negative controls for DNA-HCR (NON338 and two EUB338 probes each with 3 mismatches) yielded statistically indistinguishable cell counts from EUB338 on sample 1H-2 site 59E, 2.05 mbsf (Table 2.5). Furthermore, for cell-like particles that were positive for EUB338 with DNA-HCR, the corresponding DAPI counterstain was not visible, although other DAPI-stained cells that were not DNA-HCR positive could be visualized. Also, bright signals were not visible with *E. coli* cultures. These problems were replicated in the Schippers and Lloyd Llabs. Adding blocking reagent did not prevent the non-specific binding (Table S1). For these reasons, we conducted no further experiments with DNA-HCR.

Discussion

The heightened sensitivity of CARD-FISH relative to FISH makes it an attractive option for uncovering cellular abundance and community structure in the marine subsurface, where energetic limitations contribute to low microbial activity. CARD-FISH studies of the subsurface have typically revealed that archaea represent a quantitatively negligible fraction of the biosphere, outnumbered by the more dominant bacteria (7, 42, 43). In fact, a collection of all published marine sediment CARD-FISH counts suggested that the log-log decrease in cells observed with depth (44) does not apply to bacteria counted by CARD-FISH, in which bacterial cell numbers do not decrease below ~10 mbsf (27). This meta-analysis also investigated methodological artefacts that could result in low yields of sediment archaea. Archaeal

permeabilization procedure was identified as a possible cause of low archaeal cell counts of deep subsurface (>10 mbsf) marine sediments (27). This is because lysozyme, which is commonly used for cell permeabilization in archaea, breaks down peptidoglycan, which has only been observed in bacteria. Because some archaeal cell walls have a large protein component in their S-layer, Lloyd and coauthors hypothesized that, as described earlier in Teira *et al.* (2004), proteinase K as a permeabilization procedure would increase archaeal CARD-FISH counts and therefore total CARD-FISH yields. On the contrary, we found that no matter which permeabilization method was used (lysozyme, proteinase K, or both enzymes together), nearly all CARD-FISH counts from the Baltic Sea sediments were below the quantification limit, independent of the site or the sample organic matter content. We therefore reject the hypothesis that use of proteinase K alone is sufficient to overcome low yields of bacteria and archaea counted by CARD-FISH in deep subsurface sediments in the Baltic Sea. Furthermore, the values of the bacterial counts included in the meta-analysis (27) were often close to the quantification limit identified in our current study (10^7 cells per cm^3 of sediment) suggesting that the observed lack of a downcore trend in Lloyd *et al.* (2013) was an artefact due to non-robust cell counts. Therefore, the quantification limit for CARD-FISH appears to be much higher than that of total cell counts. Separating cells from their sediment matrix before quantification has been useful in decreasing the detection limit of total cells (45). Perhaps similar techniques would be useful for bringing CARD-FISH above quantification limits even in relatively high biomass samples, such as the Baltic Sea basin.

Much of the low CARD-FISH yields in this study overall appear to be due to cell loss associated with washing steps during CARD-FISH processing. However, loss of cells during washing steps alone does not entirely account for the low yields since nearly a third of cells

remaining after processing are unaccounted for by combined CARD-FISH counts of archaea and bacteria. The low CARD-FISH cell counts may indicate that the ribosomal contents of these bacteria and archaea were too low to allow sufficient probe binding, ribosomal sequences mismatched the probes, or that detrital grains were mistaken for cells during total cell counting but not CARD-FISH counting, thereby inflating total cell counts.

It has been suggested previously that deep subsurface cells likely have very low ribosomal contents due to their low *in situ* metabolic activity (4). In agreement with this, starvation experiments in pure culture have shown that FISH hybridization of species-specific probes declines strikingly with starvation time relative to a DNA stain (46), highlighting the impact that cellular physiological condition has on hybridization efficacy. This could indicate that the ribosomal contents of energy-starved cells in the deep subsurface may drop below the limit of quantification, despite CARD-FISH being able to amplify signals even from very low ribosomal contents in theory (10). CARD-FISH yield may also be affected by relative differences in ribosomal contents between archaea and bacteria; however, our understanding of the protein content of members of subseafloor sediment communities is limited. Insufficient probe coverage of target cells should not be responsible for low CARD-FISH yield, as our *in silico* analysis showed that the probes used in this study match >90% and 84% of bacterial and archaeal clades in the Silva database, respectively. Furthermore, the probes used in this study match bacterial and archaeal groups that typically compose deep sediment communities (15). The third option of overcounting is also an implausibility because the high consistency in total cell counts between labs makes it unlikely that large, variable numbers of false positives were included. For these reasons, coupled with the fact that all samples were above the detection (but not quantification) limit for both archaea and bacteria, suggests that a failure of the CARD-FISH

method for these samples can be attributed to significant cell loss during processing of samples that contain an already low number of ribosomes.

qPCR measurements have been used previously to calculate metabolic rates, including nitrification in estuaries (47), and sulfate reduction in sediments of the Baltic Sea (23, 24), the Black Sea (48, 49), and the forearc basins off Sumatra (50). In the upper 20-30 meters of the Baltic Sea sediments, the percentage of archaea as revealed through qPCR increased at sites that contained enough data to show a trend (M0060B, M0059C, and M0063E). This agrees with the global average, which shows a slight increase in percent archaea in the upper ~10 meters of marine sediments (27). Below these depths, however, the trends of archaea with depth are different by site. The percentage of archaea relative to total cells increases with depth at M0060B, and decreases with depth at M0059C. No robust depth trend can be observed below ~30 meters at sites M0059E or M0063E, although M0063E has a consistently lower average percentage of archaea below 30 mbsf. The data from M0059C suggest that bacteria are more numerous, and more stable over time in the lacustrine samples relative to the overlying marine sequence. M0060B, on the other hand, is a fully marine sequence, although the %TOC is as low as that of the lacustrine portion of M0059C. Since the percentage of archaea increase slightly with depth at M0060B, but not M0059C, this suggests that the source of the material (marine vs. lacustrine) is more important than total amount of organic matter, with marine sediments favoring the archaea. On average, archaea comprise 20 ± 16% of total 16S rRNA gene copies, agreeing with the suggestion that bacteria and archaea can be in the same order of magnitude in anoxic marine sediments (7).

Our qPCR data suggest that, while it is possible to reproduce quantification results within one order of magnitude in laboratories working independently, reproducibility depends heavily

upon uniformity in extraction protocols and primers for amplification. The copy number reproducibility was better than what has been shown in other studies that demonstrate that even within a single laboratory, wide variability in DNA extraction efficiencies (Mumy and Findlay 2004) and hampering effects of qPCR inhibitors (15) contribute to poor reproducibility in results. The Lloyd and Schippers Labs reproduced each others' qPCR values well because they used nearly identical methods and primer sets. The Weightman Lab's values were systematically higher, both for bacteria and archaea. This was likely due to the use of different primer sets and addition of a slurry step before DNA extraction. When the Schippers Lab replicated the slurry step on a subset of samples, they also observed an increase in qPCR values. Our results suggest that, across independent laboratories, DNA extraction yields from deep subsurface marine sediments may actually be precise, if methods are standardized.

The relatively high degree of precision for these qPCR results, however, was not matched by high accuracy relative to SYBR Green I or SYBR Gold cell counts. This indicates that qPCR values are best interpreted as a relative measure of abundance rather than absolute abundance. This interpretation is in agreement with meta-analysis of qPCR data produced from a wide variety of marine sediments (27). The qPCR values across all three laboratories were closer to each other than they were to total cell counts at most sites, even when potential multiplicities of 16S rRNA gene operons were considered (39, 40). The qPCR values were systematically lower than total cell counts, suggesting that extraction inefficiencies, or coextraction of PCR inhibitors, decreased values. Further, at sites M0060B, M0059C, and M0065C, the qPCR values were not only lower than total cell counts, they also did not correlate with them. This lack of correlation was consistent across laboratories, suggesting that it was intrinsic to the samples, not to the laboratories or their methods. Yield of qPCR values relative to total cell counts appeared to be

higher in organic rich samples of 5-8% TOC: site M0063E, and the upper ~50 meters of sites M0059C, M0059E, and M0065C. Site M0060B was highly re-worked before long-term burial, resulting in < 1% TOC content throughout the core, and sites M0059C and M0059E experience a transition to < 1% TOC content at the lacustrine interface just below 50 mbsf. In these low organic matter samples, the qPCR values were much lower than total cell counts, the values of which were not much lower than cell counts at higher % TOC intervals. Since this effect was true for all laboratories, it is likely that our qPCR methods simply worked better in the organic rich sediments than in the organic poor ones. This agrees with previous studies showing variable DNA extraction efficiencies with sample type (51), although elevated organic matter content is typically associated with increased qPCR inhibition (26). Interestingly, it appears that the percentages of archaea relative to total cells are higher in marine sequences characterized by low amounts of organic matter, in agreement with previous findings (25).

High yield loss associated with CARD-FISH observed in this study drove us to explore the utility of a new fluorescent, non-enzymatic quantification technique, DNA-HCR (29). This technique employs chain reaction binding of fluorescently-tagged oligonucleotides to linearly amplify signals for whole cell detection with microscopy. It has been successfully applied to environmental samples collected from anaerobic sludge and seawater and with the use of domain specific probes, it has been shown to produce high yield quantifications of both bacteria and archaea (29). DNA-HCR was attractive for use in marine sediments because it was less dependent on cell permeabilization and had less washing steps relative to CARD-FISH. Unfortunately, despite strong fluorescent signals of cell-like particles, counts with DNA-HCR did not have good DAPI counterstaining. Additionally, negative control experiments with mismatched probes and antisense probes suggest that unspecific binding of probe to non-targets

occurred in these sediments. It is possible that the lack of blocking reagents in the DNA-HCR protocol allowed non-target binding of probes to silica mineral grains; however, addition of blocking reagents did not alleviate the problems. Although DNA-HCR appears to be useful on many environmental samples (29), our results suggest that major modifications are required before it can be successfully applied to marine sediments.

Conclusion

In conclusion, although promising in theory, CARD-FISH appears to suffer from large yield losses that are not overcome by changing the cell permeabilization procedure, and at the moment DNA-HCR does not appear to be a viable alternative. Therefore, CARD-FISH is mainly useful for non-quantitative visualization of cells in their natural sediment environment and is not a reliable means to acquire abundance estimates of specific microbial taxa. By contrast, the values of qPCR were commonly above the limit of quantification and comparison across laboratories shows this method provided relatively precise biomass estimates of specific microbial taxa. However, comparison of qPCR results with direct cell counts indicates that qPCR estimates of cell density commonly underestimated total cellular abundance in this and other studies. Thus, qPCR is the most reliable and precise quantification technique for deep marine sediments, although it is not very accurate and is therefore most useful for relative comparisons of microbial taxa.

Dependable alternative means for assessing absolute *in situ* cellular abundance in marine sediments are not yet available; however, new advances in culture-independent measures of cellular activity, such as biorthogonal non-canonical amino acid tagging (BONCAT), show promise in being able to address questions regarding active members of marine sediment communities (52). An attractive alternative to CARD-FISH, BONCAT is a newly developed

microscopy technique which requires few steps for the fluorescent tagging of proteins. BONCAT can be coupled to FISH for the phylogenetic identification of active cells, in turn providing information about the most abundant active members of a marine sediment community. However, although BONCAT may serve as a semiquantitative means for comparing cells across target clades, the absolute abundance of those cells is likely not yet achievable due to the non-uniform nature of fluorescent incorporation across proteins (53, 54).

Our qPCR data suggested that the samples from IODP Expedition 347 contain both bacteria and archaea at all depths measured (down to 90 mbsf). Measurements compared across laboratories show varying degrees of inter-lab precision that were not matched by accuracy relative to total cell counts, which remains low at all sites analyzed. Accuracy increased with amendment of DNA extraction protocols, highlighting the importance of sample-specific modifications to maximize the yield of DNA extracted from deep marine sediments. Models of cellular functions and respiration in the marine subsurface often use qPCR to determine cell-specific rates from reactive-transport models, so it is important that qPCR values accurately quantify the target subsurface community members. Ineffective or incomplete extractions may lead to specific rate calculations that either under- or overestimate cell-specific reactions, and in this way, hinder our understanding of cellular maintenance states in energy-limited marine sediments.

References

1. Kallmeyer J, Pockalny R, Adhikari RR, Smith DC, D'Hondt S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. PNAS 109:16213-16216.
2. Hoehler TM, Jørgensen BB. 2013. Microbial life under extreme energy limitation. Nat Rev Micro 11:83-94.
3. Jørgensen BB. 1982. Mineralization of organic matter in the sea bed: the role of sulphate reduction. Nature 296:643 - 645.
4. D'Hondt S, Rutherford S, Spivack AJ. 2002. Metabolic activity of subsurface life in deep-sea sediments. Science 295:2067-2070.
5. Parkes RJ, Webster G, Cragg BA, Weightman AJ, Newberry CJ, Ferdelman TG, Kallmeyer J, Jørgensen BB, Aiello IW, Fry JC. 2005. Deep sub-seafloor prokaryotes stimulated at interfaces over geological time. Nature 436:390-394.
6. Teske A, Sørensen KB. 2008. Uncultured archaea in deep marine subsurface sediments: have we caught them all? ISME J 2:3-18.
7. Schippers A, Kock D, Höft C, Köweker G, Siegert M. 2012. Quantification of microbial communities in subsurface marine sediments of the Black Sea and off Namibia. Front Microbiol 3.
8. Lloyd KG, Schreiber L, Petersen DG, Kjeldsen KU, Lever MA, Steen AD, Stepanauskas R, Richter M, Kleindienst S, Lenk S. 2013. Predominant archaea in marine sediments degrade detrital proteins. Nature 496:215-218.
9. Amann RI, Krumholz L, Stahl DA. 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J Bacteriol 172:762-770.
10. Pernthaler A, Pernthaler J, Amann R. 2002. Fluorescence *in situ* hybridization and

catalyzed reporter deposition for the identification of marine bacteria. Appl Environ Microbiol 68:3094-3101.

11. Jørgensen BB. 2011. Deep subseafloor microbial cells on physiological standby. PNAS 108:18193-18194.

12. D'Hondt S, Spivack AJ, Pockalny R, Ferdelman TG, Fischer JP, Kallmeyer J, Abrams LJ, Smith DC, Graham D, Hasiuk F. 2009. Subseafloor sedimentary life in the South Pacific Gyre. PNAS 106:11651-11656.

13. Røy H, Kallmeyer J, Adhikari RR, Pockalny R, Jørgensen BB, D'Hondt S. 2012. Aerobic microbial respiration in 86-million-year-old deep-sea red clay. Science 336:922-925.

14. Knoblauch C, Jørgensen BB, Harder J. 1999. Community Size and Metabolic Rates of Psychrophilic Sulfate-Reducing Bacteria in Arctic Marine Sediments. Appl Environ Microbiol 65:4230-4233.

15. Biddle JF, Lipp JS, Lever MA, Lloyd KG, Sørensen KB, Anderson R, Fredricks HF, Elvert M, Kelly TJ, Schrag DP, Sogin ML, Brenchley JE, Teske A, House CH, Hinrichs K-U. 2006. Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. PNAS 103:3846-3851.

16. Jørgensen BB, Boetius A. 2007. Feast and famine - microbial life in the deep-sea bed. Nat Rev Micro 5:770-781.

17. Lomstein BA, Langerhuus AT, D'Hondt S, Jørgensen BB, Spivack AJ. 2012. Endospore abundance, microbial growth and necromass turnover in deep sub-seafloor sediment. Nature 484:101-104.

18. Price PB, Sowers T. 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. PNAS 101:4631-4636.

19. Parkes RJ, Cragg BA, Fry JC, Herbert R, Wimpenny J, Allen J, Whitfield M. 1990. Bacterial Biomass and Activity in Deep Sediment Layers from the Peru Margin [and Discussion]. Philos Trans R Soc London, Ser A 331:139-153.

20. Ravenschlag K, Sahm K, Amann R. 2001. Quantitative Molecular Analysis of the Microbial Community in Marine Arctic Sediments (Svalbard). Appl Environ Microbiol 67:387-395.

21. Altmann D, Stief P, Amann R, De Beer D, Schramm A. 2003. *In situ* distribution and activity of nitrifying bacteria in freshwater sediment. Environ Microbiol 5:798-803.

22. Sahm K, MacGregor BJ, Jørgensen BB, Stahl DA. 1999. Sulphate reduction and vertical distribution of sulphate-reducing bacteria quantified by rRNA slot-blot hybridization in a coastal marine sediment. Environ Microbiol 1:65-74.

23. Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C, Jørgensen BB. 2009. Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. Environ Microbiol 11:1278-1291.

24. Holmkvist L, Ferdelman TG, Jørgensen BB. 2011. A cryptic sulfur cycle driven by iron in the methane zone of marine sediment (Aarhus Bay, Denmark). Geochim Cosmochim Acta 75:3581-3599.

25. Breuker A, Schippers A. 2013. Data Report: Total cell counts and qPCR abundance of Archaea and Bacteria in shallow subsurface marine sediments of North Pond: Gravity cores collected on site survey cruise prior to IODP Expedition 336. In Proc. IODP Volume, vol. 336, p. 2. 2013.

26. Lloyd KG, MacGregor BJ, Teske A. 2010. Quantitative PCR methods for RNA and DNA in marine sediments: maximizing yield while overcoming inhibition. FEMS Microbiol Ecol

72:143-151.

27. Lloyd KG, May MK, Kevorkian RT, Steen AD. 2013. Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. Appl Environ Microbiol 79:7790-7799.

28. Noguera DR. 2015. Could *in situ* DNA-hybridization chain reaction enable simple and effective detection of identity and function in whole cell hybridizations? Environ Microbiol 17:2559-2561.

29. Yamaguchi T, Kawakami S, Hatamoto M, Imachi H, Takahashi M, Araki N, Yamaguchi T, Kubota K. 2015. *In situ* DNA-hybridization chain reaction (HCR): a facilitated *in situ* HCR system for the detection of environmental microorganisms. Environ Microbiol 17:2532-2541.

30. Andrén T, Jørgensen, B.B., and Cotterill, C., and the Expedition 347 Scientists. 2015. Proc. IODP, 347: College Station, TX (Integrated Ocean Drilling Program).

31. Teira E, Reinthaler T, Pernthaler A, Pernthaler J, Herndl GJ. 2004. Combining catalyzed reporter deposition-fluorescence *in situ* hybridization and microautoradiography to detect substrate utilization by bacteria and archaea in the deep ocean. Appl Environ Microbiol 70:4411-4414.

32. Wallner G, Amann R, Beisker W. 1993. Optimizing fluorescent *in situ* hybridization with rRNA-targeted oligonucleotide probes for flow cytometric identification of microorganisms. Cytometry 14:136-143.

33. Fry JC. 1990. Direct methods and biomass estimation. Method Microbiol 22:1-85.

34. RCoreTeam. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

35. Webster G, Newberry CJ, Fry JC, Weightman AJ. 2003. Assessment of bacterial

community structure in the deep sub-seafloor biosphere by 16S rDNA-based techniques: a cautionary tale. Journal of Microbiological Methods 55:155-164.

36. Nadkarni MA, Martin FE, Jacques NA, Hunter N. 2002. Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology 148:257-266.

37. Takai K, Horikoshi K. 2000. Rapid detection and quantification of members of the archaeal community by quantitative PCR using fluorogenic probes. Appl Environ Microbiol 66:5066-5072.

38. Kubo K, Lloyd KG, Biddle JF, Amann R, Teske A, Knittel K. 2012. Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. ISME J 6:1949-1965.

39. Kembel SW, Wu M, Eisen JA, Green JL. 2012. Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. PLoS Comput Biol 8:e1002743.

40. Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S, DeMaere MZ, Ting L, Ertan H, Johnson J. 2009. The genomic basis of trophic strategy in marine bacteria. Proceedings of the National Academy of Sciences 106:15527-15533.

41. Parkes RJ, Cragg BA, Wellsbury P. 2000. Recent studies on bacterial populations and processes in subseafloor sediments: a review. Hydrogeol J 8:11-28.

42. Schippers A, Neretin LN, Kallmeyer J, Ferdelman TG, Cragg BA, Parkes RJ, Jørgensen BB. 2005. Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. Nature 433:861-864.

43. Webster G, Blazejak A, Cragg BA, Schippers A, Sass H, Rinna J, Tang X, Mathes F,

Ferdelman TG, Fry JC. 2009. Subsurface microbiology and biogeochemistry of a deep, cold-water carbonate mound from the Porcupine Seabight (IODP Expedition 307). Environ Microbiol 11:239-257.

44. Parkes RJ, Cragg BA, Bale S, Getlifff J, Goodman K, Rochelle PA, Fry JC, Weightman AJ, Harvey S. 1994. Deep bacterial biosphere in Pacific Ocean sediments. Nature 371:410-413.

45. Morono Y, Terada T, Kallmeyer J, Inagaki F. 2013. An improved cell separation technique for marine subsurface sediments: applications for high-throughput analysis using flow cytometry and cell sorting. Environ Microbiol 15:2841-2849.

46. Oda Y, Slagman S-J, Meijer WG, Forney LJ, Gottschal JC. 2000. Influence of growth rate and starvation on fluorescent *in situ* hybridization of *Rhodopseudomonas palustris*. FEMS Microbiol Ecol 32:205-213.

47. Caffrey JM, Bano N, Kalanetra K, Hollibaugh JT. 2007. Ammonia oxidation and ammonia-oxidizing bacteria and archaea from estuaries with differing histories of hypoxia. ISME J 1:660-662.

48. Holmkvist L, Kamyshny A, Bruchert V, Ferdelman TG, Jørgensen BB. 2014. Sulfidization of lacustrine glacial clay upon Holocene marine transgression (Arkona Basin, Baltic Sea). Geochim Cosmochim Acta 142:75-94.

49. Leloup J, Loy A, Knab NJ, Borowski C, Wagner M, Jørgensen BB. 2007. Diversity and abundance of sulfate-reducing microorganisms in the sulfate and methane zones of a marine sediment, Black Sea. Environ Microbiol 9:131-142.

50. Schippers A, Köweker G, Höft C, Teichert BM. 2010. Quantification of microbial communities in forearc sediment basins off Sumatra. Geomicrobiol J 27:170-182.

51. Zhou J, Bruns MA, Tiedje JM. 1996. DNA recovery from soils of diverse composition.

Appl Environ Microbiol 62:316-322.

52. Hatzenpichler R, Scheller S, Tavormina PL, Babin BM, Tirrell DA, Orphan VJ. 2014. *In situ* visualization of newly synthesized proteins in environmental microbes using amino acid tagging and click chemistry. Environmental microbiology 16:2568-2590.

53. Hatzenpichler, Roland, and Victoria J. Orphan. "Detection of protein-synthesizing microorganisms in the environment via bioorthogonal noncanonical amino acid tagging (BONCAT)." In Hydrocarbon and Lipid Microbiology Protocols, pp. 145-157. Springer, Berlin, Heidelberg, 2015.

54. Hatzenpichler R, Connon SA, Goudeau D, Malmstrom RR, Woyke T, Orphan VJ. 2016. Visualizing *in situ* translational activity for identifying and sorting slow-growing archaeal−bacterial consortia. Proceedings of the National Academy of Sciences 113:E4069-E4078.

Appendix I: Tables and Figures

Table 2.1. Primers and probes used in this study.

Primer/probe	Sequence (5'-3'[|])	Target group	[FA]	Coverage[a] target, non-target	Lab	Source
Bac340f	TCCTACGGGAGGCAGCAGT	Bacteria		89%, 0%	1, 2	Nadkarni *et al*. (2002)
Bac806r	GGACTACCAGGGTATCTAATCCTGTT	Bacteria		79%, 0.1%	1, 2	Nadkarni *et al*. (2002)
Bac (TaqMan probe)	CGTATTACCGCGGCTGCTGGCAC	Bacteria		79%, 0%	2	Nadkarni *et al*. (2002)
Bac534f	GCCAGCAGCCGCGGTAAT	Bacteria		87%, 0.3%	3	Muyzer *et al*. (1993)
Bac907r	CCGTCAATTCCTTTGAGTTT	Bacteria		75%, 0%	3	Muyzer and Smalla (1998)
Arch915f	AGGAATTGGCGGGGGAGCAC	Archaea		84%, 0%	1,2	Stahl and Amann (1991)
Arch1059r	GCCATGCACCWCCTCT	Archaea		83%, 0%	1,2	Kubo *et al*. (2012)
S-D-Arch-0025-a-S-17f	CTGGTTGATCCTGCCAG	Archaea		3%, 0%	3	Vetriani *et al*. (1999)
S-D-Arch-0344-a-S-20r	ACGGCTACCTTGTTACGACTT	Archaea		38%, 0%	3	Vetriani *et al*. (1999)
Arch349f	GYGCASCAGKCGMGAAW	Archaea		81%, 0%	2	Takai and Horikoshi (2000)
Arch516f (TaqMan probe)	TGYCAGCCGCCGCGGTAAHACCVGC	Archaea		83%, 0%	2	Takai and Horikoshi (2000)
Arch806r	GGACTACYVGGGTATCTAAT	Archaea		88%, 89%	2	Takai and Horikoshi (2000)
Arch915	GTGCTCCCCCGCCAATTCCT	Archaea	55%	84%, 0%	1, 2	Stahl and Amann (1991)
EUB338	GCTGCCTCCCGTAGGAGT	Bacteria	45%	90%, 0%	1, 2	Amann, Krumholz and Stahl (1990)
EUB338 II	GCAGCCACCCGTAGGTGT	Bacteria	45%	0.8%, 0%	1, 2	Daims *et al*. (1999)
EUB338III	GCTGCCACCCGTAGGTGT	Bacteria	45%	1.3%, 0%	1, 2	Daims *et al*. (1999)
EUB338-initiatorC[o]	CCAGTTATCAGTAGTCC-GTCCTTCATTTTTTGCTGC-CTCCCGTAGGAGT	Bacteria	20%		1, 2	Yamaguchi *et al*. (2015)
Arch915-initiatorC[o]	CCAGTTATCAGTAGTCC-GTCCTTCATTTTTTGTGCT-CCCCCGCCAATTCCT	Archaea	40%		1, 2	Yamaguchi *et al*. (2015)

Table 2.1 Continued.

Primer/probe	**Sequence (5'-3'1)**	Target group	[FA]	Coverage[a] target, non-target	Lab	Source
Amplifier C1[o]	**ATGAAGGACGgactactgataactgg-** GACTTCCATAccagttatcagtagtc*		0%		1, 2	Yamaguchi *et al.* (2015)
Amplifier C2[o]	*ccagttatcagtagtcCGTCCTTCAT-gactactgataactggTATGGAAGTC		0%		1, 2	Yamaguchi *et al.* (2015)
NON338	CCGAATACAAAGCATCAAGC-ACTAGAAAAAAACTCCTACGG-GAGGCAGC		45%		1, 2	Wallner, Amann and Beisker (1993)
MMA	CCAGTTATCAGTAGTCCGTCCT-TCATTTTTTGCAGCCACCCGTAG-GTGT		45%	, 0%	1	Yamaguchi *et al.* (2015)
MMB	CCAGTTATCAGTAGTCCGTCCTT-CATTTTTTGCTCCCTGCCGTAGCAGT		45%	, 0%	1	Yamaguchi *et al.* (2015)

[a]Results of Silva TestProbe 3.0 analysis at the time of publication. No mismatches allowed.

*indicates the side on which the fluorophore is attached.

[o]indicates probe exclusively used for DNA-HCR. [FA]—formamide concentration for DNA-HCR or CARD-FISH, v/v. 1—Lloyd Lab; 2—Schippers Lab; 3—Weightman Lab.

Single underlined letters indicate DNA-HCR probe for rRNA.

Double underlined letters indicate DNA-HCR initiator sequence for chain reaction. Lowercase letters indicate the stem structure of DNA hairpin.

Bold letters are complimentary to the initiator sequence on initiator probe. Highlighted letters are locations of mismatches to EUB338 probe.

Table 2.2. Total cell counts from all sites in this study. Results are presented as cells per cubic cm (cm^3), with values adjusted for depth-specific density.

Core (M00-, sample)	Depth (mbsf)	AODC	SYBR Gold*[1]	SYBR Green I*[2]	Avg. DAPI lysozyme	Avg. DAPI proteinase K^0
59C, 1-1	1.4		4.91E+08	1.14E+09	4.01E+08	4.64E+08
59C, 1-1	1.53	1.37E+09				
59C, 1–2	2.96	1.40E+09				
59C, 2-2	4.53	1.44E+09				
59C, 2-2	5.96	1.17E+09				
59C, 3-1	7.7		3.67E+08	1.26E+09	3.76E+08	3.86E+08
59C, 3-2	7.83	1.15E+09				
59C, 3-2	8.2		4.98E+08	1.49E+09	2.86E+08	4.49E+08
59C, 3-2	9.26	5.52E+08				
59C, 4-1	11					
59C,4-2	11.13	6.41E+08				
59C, 5-1	14.05		5.04E+08	4.57E+08	2.17E+08	1.64E+08
59C, 5-2	14.43	6.87E+08				
59C, 6-1	17.6		3.22E+08	5.70E+08	1.32E+08	1.20E+08
59C, 6-2	17.73	5.66E+08				
59C, 9-1	27.5		3.87E+08	3.12E+08	5.48E+07	5.79E+07
59C, 9-2	27.63	3.00E+08				
59C, 12-2	37.5		6.04E+08	3.93E+08	2.57E+07	2.82E+07
59C, 12-2	37.53	9.03E+08				
59C, 15-1	44.6		1.20E+08	1.24E+08	2.81E+07	1.49E+07
59C, 15-2	45.02			2.08E+08	3.04E+07	1.20E+07
59C, 15-2	45.05	1.81E+08				
59C, 18-2	54.95	1.15E+08	7.59E+08	1.49E+07	3.37E+06	2.13E+06
59C, 20–1	61.57					4.64E+08
59C, 21–2	64.85	8.15E+07				
59C, 21–2	65		3.42E+08			
59C, 22-1	68.02					3.86E+08
59C, 24-1	74.62		6.61E+07			4.49E+08

Table 2.2 continued.

Core (M00-, sample)	Depth (mbsf)	AODC	SYBR Gold*[1]	SYBR Green I*[2]	Avg. DAPI lysozyme	Avg. DAPI proteinase K[0]
59C, 24-1	74.75	1.14E+08				1.64E+08
59C, 25-1	77.92					
59C, 25-1	77.93	1.90E+08				
59E, 1-1	1.41		5.79E+08	1.27E+09	1.08E+09	6.97E+08
59E, 1-1	1.52	2.52E+09				
59E, 1–2	2.05		7.39E+08	8.11E+08	5.61E+08	5.12E+08
59E, 2-2	2.96	1.61E+09				
59E, 2-2	4.82	1.47E+09				
59E, 2-2	6.26	8.45E+08				
59E, 4-2	11.42	9.52E+08	4.52E+08	6.61E+08	2.91E+08	2.39E+08
59E, 6-2	18.02	1.04E+09				
59E, 7-1	21.2		3.96E+08	3.54E+08	1.34E+08	1.23E+08
59E, 7-2	22.4		4.25E+08	3.954E+08	1.15E+08	6.60E+07
59E, 10–2	31.62		3.07E+08	3.31E+08	1.21E+08	1.12E+08
59E, 13-2	42.3		6.23E+08	2.72E+08	5.08E+07	4.63E+07
59E, 13-2	41.1		9.12E+08	1.07E+08	6.92E+07	6.19E+07
59E, 16-2	51		4.08E+08	2.83E+07	2.30E+07	1.78E+07
59E, 16-2	52.47	3.43E+08				
59E, 17-2	54.33	5.85E+07				
59E, 17-2	55.76	1.02E+08				
59E, 18-2	59.05	1.04E+08				
59E, 19-2	60.9		2.32E+08			
59E, 19-2	60.93	1.44E+08				
59E, 19-2	62.15		2.72E+08			
59E, 19-2	62.36	2.94E+08				
59E, 20–2	65.66	9.27E+07				
59E, 22-2	70.81		1.64E+08			
59E, 22-2	71.64		1.57E+08			
59E, 24-2	77.42	3.52E+07				
59E, 25-2	80.72		1.97E+08			

Table 2.2 continued.

Core (M00-, sample)	Depth (mbsf)	AODC	SYBR Gold*[1]	SYBR Green I*[2]	Avg. DAPI lysozyme	Avg. DAPI proteinase K[0]
59E, 25-2	81.7		6.55E+07			
60B, 1–2	2.58	1.57E+07				
60B, 3-2	4.33	2.80E+07				
60B, 5-2	10.93	8.26E+08				
60B, 6-2	14.23	3.00E+08				
60B, 6-2	14.3		9.29E+08			
60B, 9-1	24		5.40E+08			
60B, 9-2	24.13	6.26E+08				
60B, 12-2	34		6.93E+08			
60B, 12-2	34.03	5.85E+08				
60B, 13-2	37.35		3.00E+08			
60B, 14-2	40.6		4.19E+08			
60B, 15-2	43.9		3.73E+08			
60B, 15-2	43.93	2.17E+08				
60B, 18-2	53.8		2.44E+08			
60B, 18-2	53.83	1.04E+09				
60B, 21–2	63.7		3.13E+08			
60B, 21–2	63.73	5.58E+08				
60B, 23-2	69.55		2.97E+08			
60B, 24-2	72		4.33E+08			
60B, 24-2	72.03	4.24E+08				
60B, 27-2	81.1		5.12E+08			
60B, 27-2	81.13	3.14E+08				
60B, 27-2	81.5		4.34E+08			
60B, 28-1	84.42		2.91E+08			
60B, 28-2	84.43	2.07E+08				
63E, 1-1	1.1			1.40E+09	7.02E+07	2.04E+08
63E, 1–2	1.12			4.17E+08	3.66E+08	5.09E+08
63E, 1–2	1.22	8.26E+09				
63E, 1–2	1.65			1.63E+09		
63E, 1–2	2.65	1.09E+10				

Table 2.2 continued.

Core (M00-, sample)	Depth (mbsf)	AODC	SYBR Gold*[1]	SYBR Green I*[2]	Avg. DAPI lysozyme	Avg. DAPI proteinase K[0]
63E, 2-2	3.52	5.98E+09				
63E, 2-2	4.95	9.06E+09				
63E, 3-1	5.02			1.87E+09	3.34E+08	1.35E+08
63E, 6-1	10.83			9.33E+08	3.87E+07	4.62E+07
63E, 6-2	11.8		1.75E+08	9.27E+08	1.68E+07	4.19E+07
63E, 8-2	15.44	5.85E+09				
63E, 9-1	17.04			6.47E+08	6.34E+06	4.18E+07
63E, 10–2	19.03	2.86E+09				
63E, 10–2	19.86	3.14E+09				
63E, 12-2	23.5		4.35E+08	6.36E+08	1.16E+08	7.40E+07
63E, 12-2	24.29	2.13E+09				
63E, 13-2	25.03	1.60E+09				
63E, 14-3	28.22	2.07E+09				
63E, 15-2	29.03	1.04E+09				
63E, 15-2	29.4		5.35E+07	9.03E+06	6.96E+06	3.24E+06
63E, 15-2	30.25	9.90E+08				
63E, 16-2	31.53	1.14E+09				
63E, 17-2	33.4	6.26E+08				
63E, 18-2	35.46		2.27E+07	5.81E+07	2.35E+06	3.68E+06
63E, 18-2	35.49	2.17E+08				
63E, 19-2	37.53	4.75E+08				
63E, 20–2	39.5	1.06E+09				
63E, 21–2	41.1	9.25E+07				
63E, 21–2	41.28	1.73E+07		2.04E+05	3.35E+06	
63E, 22-2	43.52	2.69E+08				
63E, 25-2	49.52	9.93E+07				
63E, 26-2	51.52	2.67E+08				
63E, 27-2	53.97	2.56E+08				
63E, 28-2	55.52	3.52E+08				
63E, 29-2	57.52	2.80E+08				
63E, 30–2	59.52	4.96E+08				

Table 2.2 continued.

Core (M00-, sample)	Depth (mbsf)	AODC	SYBR Gold*[1]	SYBR Green I*[2]	Avg. DAPI lysozyme	Avg. DAPI proteinase K[0]
63E, 32-2	63.53	9.48E+07				
63E, 35-2	69.53	1.85E+08				
63E, 36-2	71.52	6.01E+07				
63E, 39-2	76.82	8.26E+07				
63E, 41–2	83.42	8.65E+07				
63E, 42-2	87.02	6.12E+07				
65C, 2-2	3.5		3.81E+08	1.63E+08	8.53E+07	5.26E+07
65C, 2-2	3.53	1.54E+10				
65C, 4-2	8.06	2.93E+09				
65C, 4-1	10		5.13E+08	1.60E+08	4.97E+06	2.04E+07
65C, 4-2	10.13	2.74E+09				
65C, 4-2	10.65			8.68E+07	2.21E+07	7.49E+06
65C, 4-2	11.46	3.37E+08				
65C, 5-2	13.43	7.71E+07				
65C, 5-2	14.76	2.44E+08				
65C, 6-2	16.73	2.92E+08				
65C, 7-2	20.03	1.54E+08				
65C, 7-2	20.6		7.77E+08			
65C, 8-2	23.33	2.27E+08				
65C, 10–1	29.8		6.29E+08			
65C, 10–2	29.93	2.58E+08				
65C, 11–2	33.23	1.81E+08				
65C, 12-2	36.53	2.74E+08				

+Samples directly fixed in FA without washing.

∗Samples fixed in FA. Washed twice and stored in PBS/ethanol. Single filters counted.

oSamples fixed in FA. Washed twice and stored in PBS/ethanol, whole CARD-FISH permeabilization and hybridization procedure. Triplicate filters counted. 1—Lloyd Lab; 2—Schippers Lab.

Table 2.3. CARD-FISH counts above the quantification limit for the different permeabilization treatments (lysozyme, proteinase K and both together). The quantification limit was defined as being able to count more than 30 cells within 30 fields of view. Not all treatments yielded counts above the quantification limit within a single depth. These treatments are indicated as BQL (below quantification limit). Dashes indicate that CARD-FISH was not attempted with the permeabilization solution indicated.

Site	Lab	Sample	Depth (mbsf)	Target	Lysozyme (cells cm^{-3})	Proteinase K (cells cm^{-3})	Both (cells cm^{-3})
59C	Lloyd	1-1	1.4	Bacteria	1.39E+07	1.06E+07	9.00E+06
59C	Lloyd	9-1	27.5	Bacteria	1.43E+07	6.75E+06	
59C	Lloyd	12-2	37.5	Archaea	6.54E+06	BQL	6.14E+06
59C	Schippers	1-1	1.4	Archaea	9.54E+06	1.36E+06	BQL
59C	Schippers	12-2	37.5	Archaea	BQL	6.75E+07	–
59E	Schippers	1-1	1.41	Archaea	3.77E+07	2.91E+07	–
59E	Schippers	4-2	11.42	Archaea	2.47E+07	2.83E+07	–
59E	Schippers	7-1	21.2	Archaea	2.55E+07	4.12E+07	–
59E	Schippers	7-2	22.4	Archaea	2.47E+07	BQL	–
59E	Schippers	10-2	31.62	Archaea	3.42E+07	1.41E+07	–
59E	Schippers	13-2	42.3	Archaea	1.49E+07	1.65E+07	–
59E	Lloyd	16-2	51	Bacteria	1.21E+07	BQL	8.79E+06
63E	Lloyd	1-1	1.1	Bacteria	2.76E+07	1.39E+07	1.86E+07
63E	Lloyd	1-2	1.12	Archaea	1.27E+07	1.25E+07	2.21E+07
63E	Lloyd	3-1	5.02	Archaea	BQL	1.41E+07	BQL
63E	Lloyd	1-1	1.1	Archaea	2.68E+07	2.29E+07	
63E	Lloyd	1-2	1.12	Archaea	1.34E+07	1.36E+07	
63E	Lloyd	3-1	5.02	Archaea	BQL		
63E	Lloyd	6-2	11.8	Archaea			
65C	Schippers	2-2	3.5	Bacteria	1.22E+07	BQL	BQL
65C	Schippers	4-1	10	Archaea	BQL	1.22E+07	BQL

Table 2.4. Results of statistical comparison of binned qPCR copy numbers of Bacteria and Archaea between laboratories. Symbols represent results of paired Wilcoxon signed rank testing, with pluses (+) indicating no statistical difference and minuses (–) indicating statistical difference at the 0.05 significance level. Blank fields indicate data for comparison was either absent (Weightman values M0059, C; Lloyd values M0059E) or insufficient for analysis (n < 3; M0065C, except for Archaea in Weightman and Schippers labs).

Site, hole		Lloyd Bacteria, Archaea	Schippers Bacteria, Archaea[b]
Schippers	M0059,C	–, –	–
	M0059,E		–
	M0060,B	+, +	–
	M0063,E	–, +	–
	M0065,C		–
Weightman	M0059,C		
	M0059,E		+, –
	M0060,B	–, +	+, –
	M0063,E	–, +	–, +
	M0065,C		, +

[a]Bins are as follows:

60B: 14.3–14.9, 20.8–21.3, 24.6–27.9 (Archaea only), 27.42–29.9, 78.3–81.5, 84.42–85.4 mbsf

59C: 4.6–8.2, 11.2–17.2, 24.2–27.1, 30.5–30.8, 37.4–38.1, 43.1–44.5, 61.5–64.3, 74.6–75.3, 77.5–77.9 mbsf

59E: 2.0–5.3, 22.2–22.3, 41.9–42.2, 48.5–51.8, 61.5–62.1, 71.6–75.1 mbsf

63E: 1.1–5.8, 6.5–11.8 (6.53–10.8 for Archaea), 14.5–18.7, 19.5–23.5, 27.7–31.6, 33.8–35.4, 39.6–41.1, 46–47.9, 54.9–55.5, 69.9–71.7, 76.8–80.2, 87.6–90.1 mbsf

[b]Archaeal copy numbers with primer set Arch915f/Arch1059r was used when comparing values from the Schippers lab.

Table 2.5. Cell counts obtained through DNA-HCR after hybridization with initiator probes EUB338 (I-III) mix, NON338, or EUB338 containing 3 mismatches (MMA, MMB in Table 2.1). Sample is M0059E-1-2 (2.05 mbsf). Error ranges represent standard deviations of counts from 30 fields of view. Blocking reagent was not used in NON338 measurements.

Probe	Cells cm^{-3} (without blocking reagent)	Cells/ cm^{-3} (with blocking reagent)
EUB338 mix	5.32E+07 ± 4.32E+07	3.58E+08 ± 2.56E+08
NON338	9.38E+07 ± 4.13E+07	
MMA (3 mismatches)	9.25E+07 ± 5.55E+07	1.01E+08 ± 5.66E+07
MMB (3 mismatches)	7.59E+07 ± 5.36E+07	1.21E+08 ± 7.07E+07

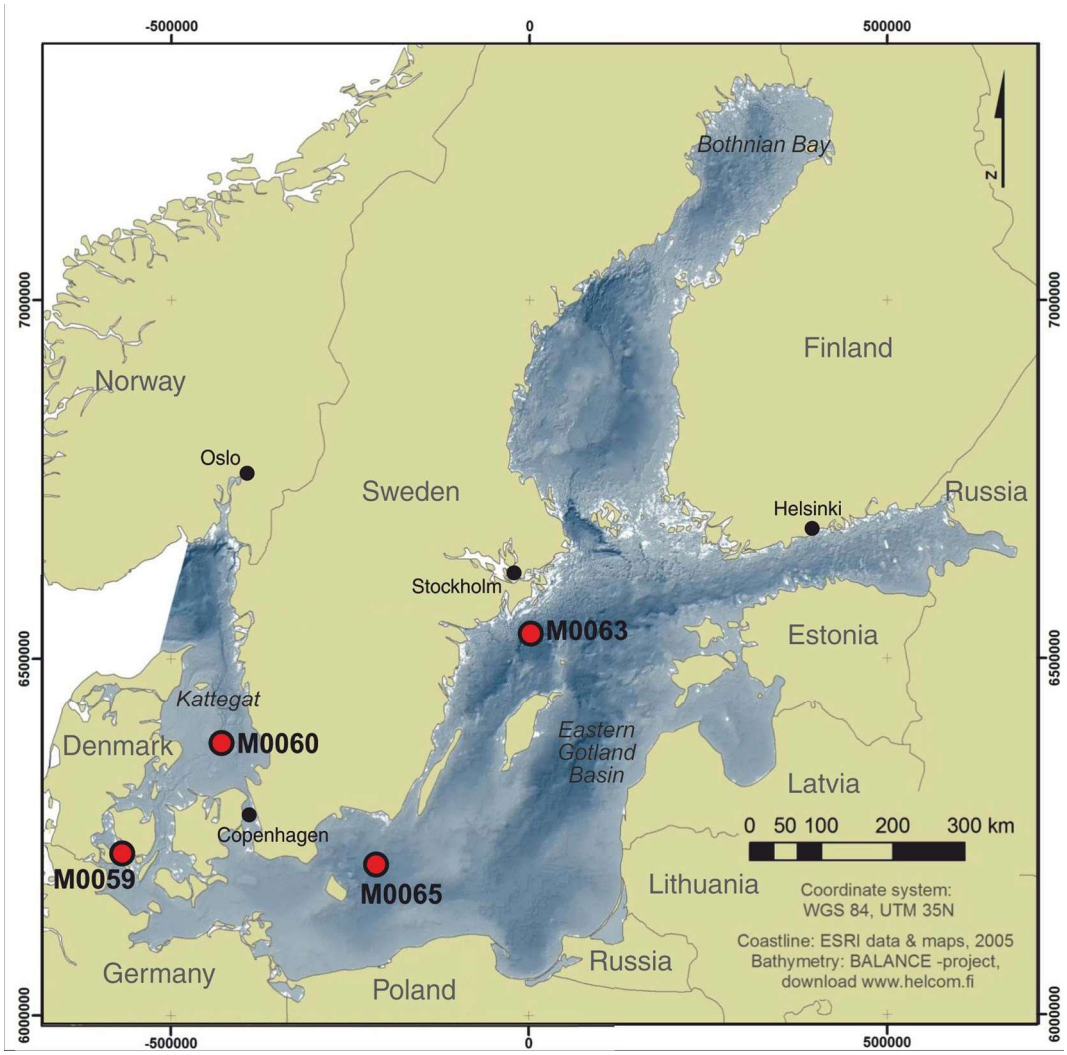

Figure 2.1. IODP Leg 347 Baltic Sea Paleoenvironment sites M0059, M0060, M0063 and M0065 are distributed around the Baltic Sea. Modified from Andren et al. (2015).

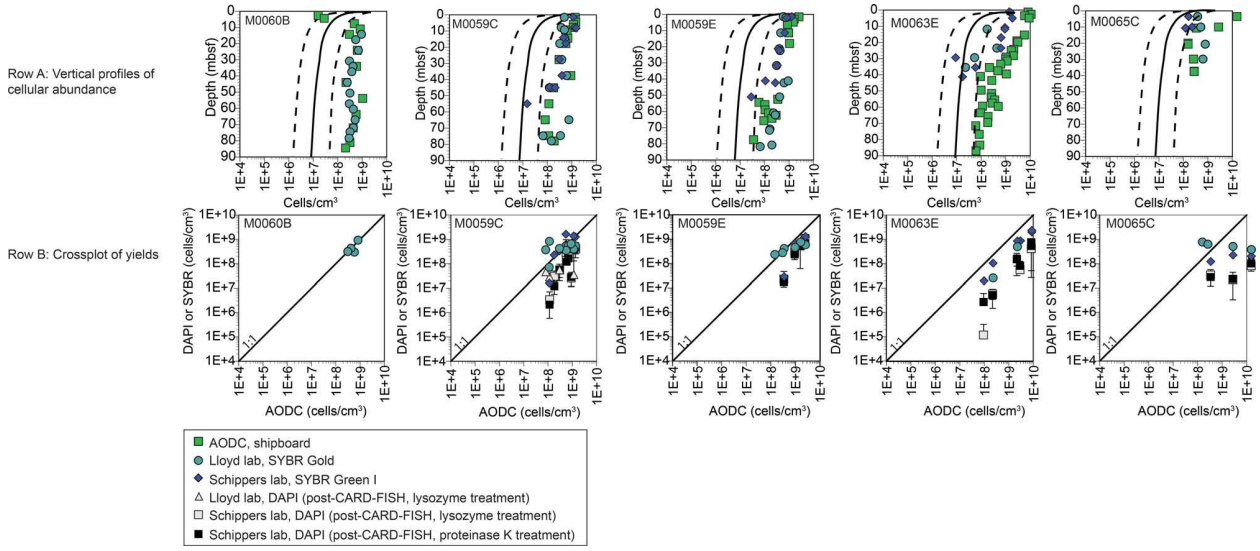

Figure 2.2. Total cell counts. Row A contains downcore profiles of SYBR Green I (blue diamonds) and SYBR Gold (teal circles) from the Schippers and Lloyd Labs, respectively. AODC shipboard counts (green squares) are published elsewhere and are shown here, with permission from the authors for comparison (30). The solid line is the global average regression with 95% confidence intervals (dashed lines; Parkes, Cragg and Wellsbury 2000). Row B illustrates the yield loss associated with CARD-FISH relative to unbinned AODC cell counts. DAPI data points represent either single counts (gray triangles; Lloyd Lab) or average counts of triplicate filters treated with either lysozyme (gray squares) or proteinase K (black squares; Schippers Lab), from which standard deviation was calculated and plotted as error bars. SYBR Green I, SYBR Gold and DAPI values have been adjusted for depth-specific density. Post-CARD-FISH counts were not conducted for Site M0060B.

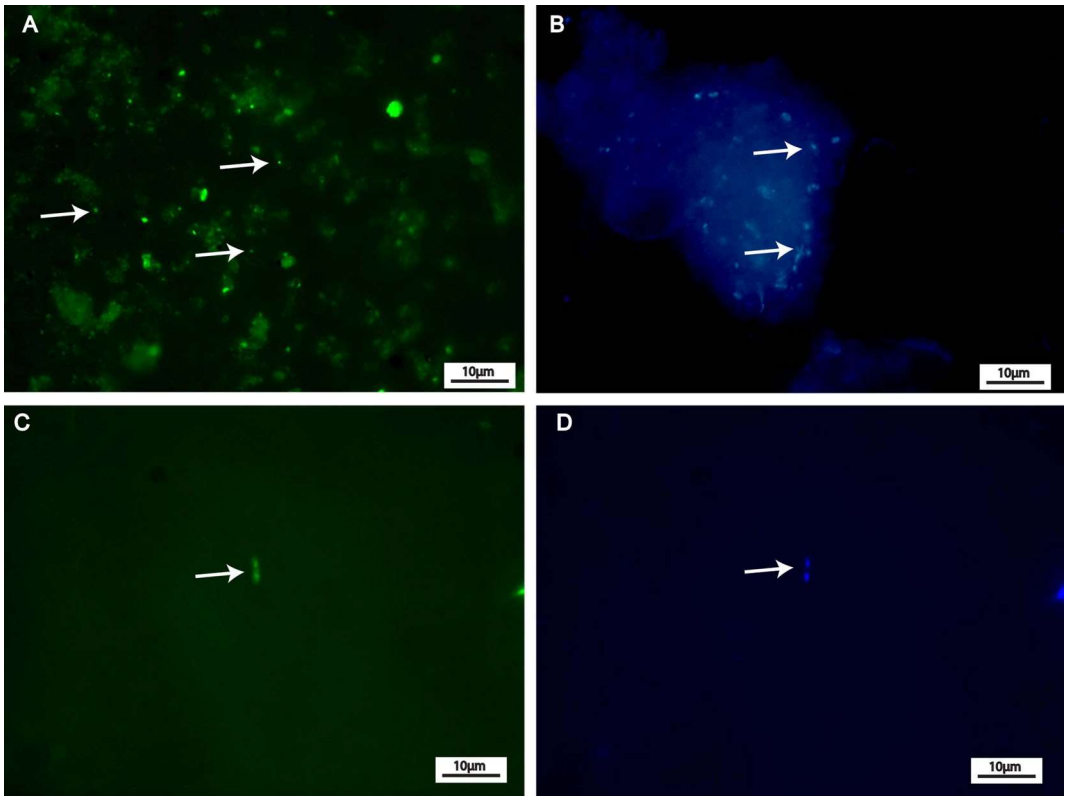

Figure 2.3. Photomicrographs of cells in sediment. (A) SYBR Gold stained cells from 23.5 mbsf in M0063E. (B) DAPI-stained cells from 1.1 mbsf in M0059C. Panels (C) and (D) show a dividing cell identified using (C) ARCH915 CARD-FISH probe and (D) DAPI in the same field of view. Arrows indicate examples of cells.

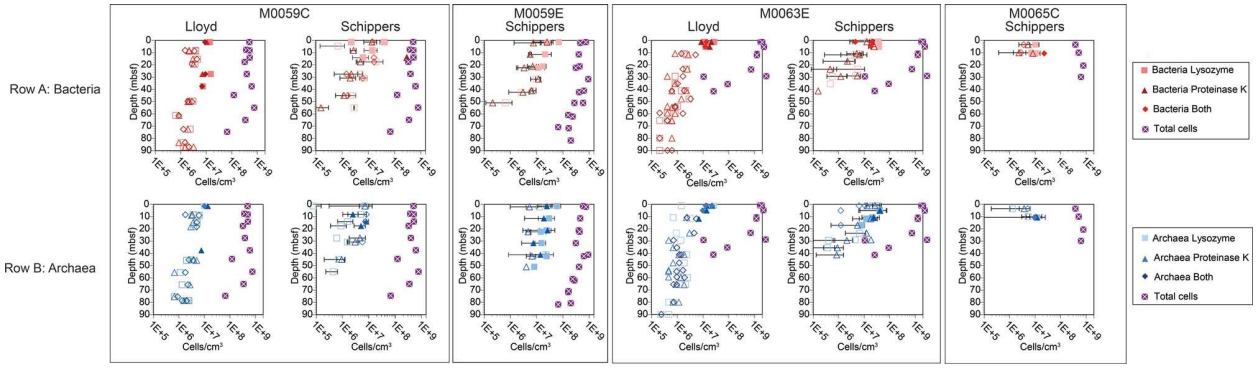

Figure 2. 4. CARD-FISH cell counts performed in the Schippers and Lloyd labs. Counts are reported as cells cm−3 and are adjusted for depth-specific density. Bacteria (shades of red) and Archaea (shades of blue) probes were applied to cells permeabilized with either lysozyme (squares), proteinase K (triangles) or both (diamonds). Total direct cell counts (purple crossed-boxes) are Lloyd Lab SYBR Gold counts for all samples except 63E, where the SYBR Green I counts are from the Schippers Lab. Error bars represent standard deviation from triplicate filter counts, which were only acquired in the Schippers Lab. Filled symbols indicate that at least one of the technical replicates was above the quantification limit of 30 cells counted per sample. Only one filter per sample was treated with both lysozyme and proteinase K treatments.

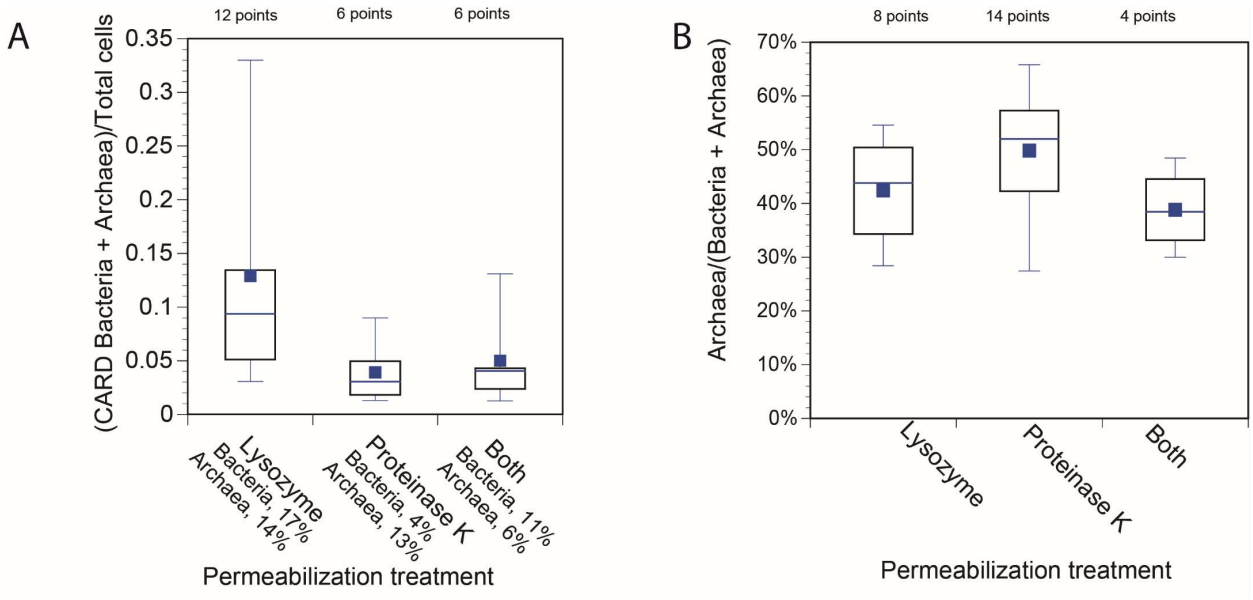

Figure 2.5. CARD-FISH yields relative to total cells as determined through DAPI direct counting for each treatments for both Bacteria and Archaea (A) and Archaea only (B). Boxes extend from the 25th to the 95th percentile and blue squares indicate data means.

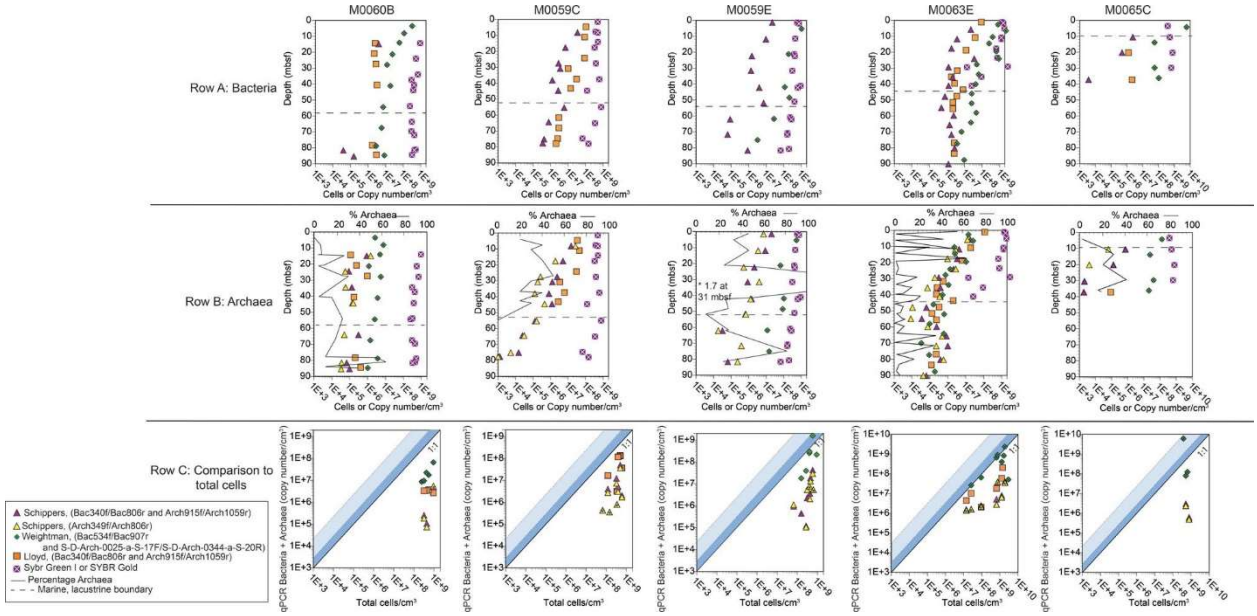

Figure 2.6. Comparison of qPCR with total cells. Downcore qPCR data for Bacteria (row A) and Archaea (row B) is reported in copy number cm−3 from all laboratories for each core. Total SYBR Gold (or Green I) cells are as described in Figure 2.2 and expressed in cells cm^{-3}. Black line in row B represents average fraction of Archaea (scale at top). Dashed lines in rows A and B indicate the depth boundary between lacustrine and overlying marine sediments. Schippers and Lloyd Bacteria copy numbers reflect quantification using Bac340f/Bac806r primers. Results of qPCR from both archaeal primer sets (Arch915f/Arch1059r and Arch349f/Arch806r) are reported for the Schippers Lab. Note x-axis for Site 65C is not shared. In row C, the solid line is the 1:1 line indicating a match between Bacteria and Archaea combined copy number cm^{-3} and total cell counts reported in cell cm^{-3} provided through SYBR Gold (or Green I for M0063E). Shaded areas indicate the known ranges of 16S rRNA gene copies per genome (3.04 copies, dark blue; 24 copies, light blue). Asterisk in row B, Site M0059E, indicates where percentage Archaea is greater than 100%. Asterisk in row C M0059C, E and M0063E indicate where Arch349f/Arch806r datapoints are hiding Arch915f/Arch1059r datapoints.

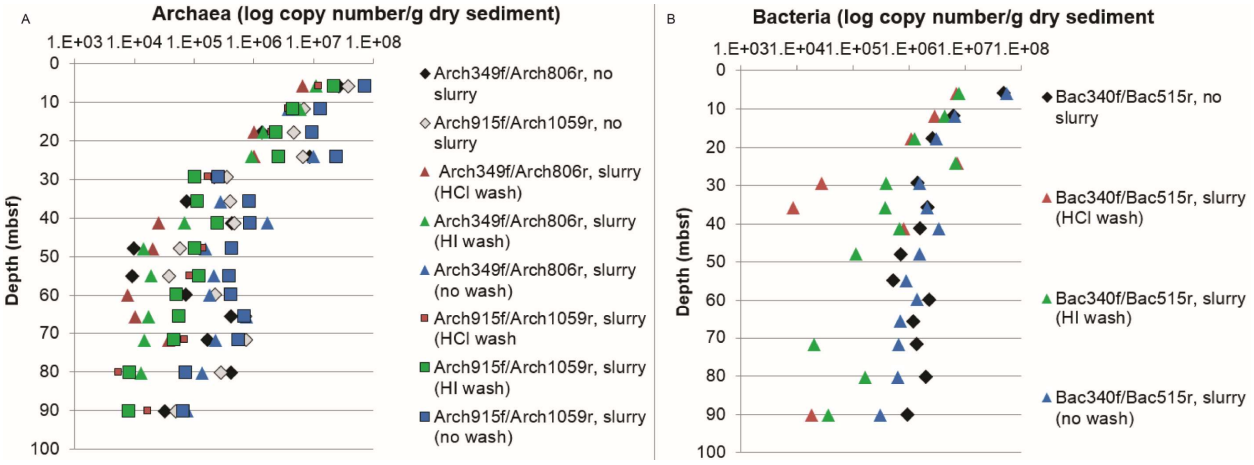

Figure 2.7. Slurry experiment results for Archaea (A) and Bacteria (B).

Chapter 3: Methanogen genome from Antarctic permafrost reveals cold adaptation and multiple pathways of methane formation

This chapter is a revised version of two published works:

Buongiorno J., Bird J.T., Lloyd K. G., Vishnivetskaya. T. (2016). Draft Genome Sequence of Antarctic Methanogen enriched from Dry Valley Permafrost. *Genome Announcements*. 4(6): e01362-16.

Vishnivetskaya, T. A., **Buongiorno, J.**, Bird, J., Krivushin, K., Spirina, E. V., Oshurkova, V., ... & Rivkina, E. M. (2018). Methanogens in the Antarctic Dry Valley Permafrost. FEMS microbiology ecology.

My primary contributions to this work include: (i) genome reconstruction, (ii) mapping and diagram of methanogenetic pathways, (iii) phylogenetic analysis, (iv) protein modeling and comparative structural analysis, and (v) writing the results of genome content, cold adaptation, and methanogenesis pathways.

Abstract

Permafrost accounts for nearly a quarter of all naturally-sourced methane produced globally. Permafrost covers over 25% of Earth's surface and is likely to produce more methane as it thaws with climate change. To understand how a warming climate may affect global methane dynamics, we first must have a greater understanding of the modern production of methane in permafrost. The biogenicity and timing of methane accumulation is not yet understood in permafrost affected soils. Here, we examined the metagenome of a methanogenic enrichment from the McMurdo Dry Valley. In order to test for the modern activity of methanogens in permafrost, incubations containing permafrost layers from Miers Valley, Antarctica, were created and monitored for methane production over 12 years. Only the enrichments containing inoculum from permafrost with *in situ* methane produced methane in the

incubation. Methane accumulation was only detected after a year of incubation. A single clade of *Methanosarcina* sp. dominated incubations characterized by methane accumulation. This shows that *in situ* methane observed in Miers Valley is likely to be biogenic in nature and not of ancient or abiotic origin. Genetic evidence suggests that this active methanogen uses cold adaptation strategies for maintaining biological function in the harsh condition of the Dry Valleys, including structural modification of key enzymes. Despite being from a different continent, the Miers Valley methanogen shares key genomic features with methanogens isolated from a Moscow fen and a Swiss lake.

Introduction

Permafrost currently contributes nearly 25% of all naturally sourced methane (1), a value that is predicted to rise significantly in coming decades (2). However, methane accumulation in permafrost environments is complex and geographically variable, making the trajectory of climate-affected methane dynamics hard to predict. Late Pleistocene permafrost from the Miers Valley (McMurdo Dry Valleys) contained methane in shallow horizons, where isotopic signatures suggested biogenic methane sources (3). In order to determine the biogenicity and timing of methane accumulation, incubation experiments were conducted. Here, we announce a nearly complete genome reconstructed from those methane-producing enrichments and describe genomic adaptations to the permanently cold permafrost environment.

Methods

Sampling, DNA extraction, and sequencing.

Anaerobic incubations of permafrost consisted of phosphate- buffered basal medium (4) and gas mixture of H_2/CO_2 (80/20) at 20°C. Methane production was first observed after one year of incubation and is ongoing today (11 years later). After seven years, samples were collected for metagenome sequencing. The total community genomic DNA from the enrichment was extracted using the PowerSoil DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA, USA), and the DNA library was prepared using the TruSeq DNA sample prep kit version 2 without whole-genome amplification. The Illumina HiSeq 2000 platform was used to acquire paired-end 2 X 100-bp metagenomic reads.

Genomic reconstruction

Adaptors and low-quality reads were trimmed with the Trimmomatic software (5) and metagenomic reads were assembled using the metaSPAdes assembler v.3.7 with k-mer size set to 21, 33, 55 and 77 (6). Contigs below 1000 bp were culled after assembly. Quality of the

assembly was assessed using the QUAST (7). The VizBin application (8) was used for visualization and subsequent binning of metagenomic sequences based on similar coverage and k-mer frequency. Gene calling and annotation of protein coding sequences was conducted with Prokka v.11 using combined curated versions of the Uniprot databases for archaea, bacteria and viruses as a reference for the BLAST alignment-based annotations (9) . The quality of the binned genome with respect to completeness and contamination was assessed using the archaeal set of single copy marker genes within CheckM (10). The 16S rRNA gene sequences were identified using RNAmmer (11).

Protein structure reconstruction

To construct ribbon depictions of the elongation factor 2 (EF2), DeepViewer (12) was used. The structural model was generated with SwissModeller and ProMod3 Version 1.0.0 (13) using automated homology modeling. Homology modeling was conducted by aligning the target EF2 sequences to the amino acid sequence of a *Sacchromyces cerevisiae* EF2 template from the SWISS-MODEL Template Library (14, 15). All target EF2 sequences had at least 30% sequence identity and 98% coverage to the template.

Phylogenetic analyses

Relationships between the 16S rRNA gene and mcrA gene within the genomic bin and other methanogens were inferred by using the Neighbor-Joining method. The sequences from *Methanospirillium hungatei* strain JF-1 and *Methanomicrobium mobile* strain BP were used as outgroups for analyses. Bootstrap percentages after 1000 replicates were calculated in MEGA 7.0 (16). Sequences were obtained by megablast with Blastn 2.4.0 (NCBI) and selected partial and complete 16S rRNA ribosomal sequences were aligned with MUSCLE (EMBL).

Results

Genome statistics and phylogeny

The final product of binning (Figure 3.1) contained 342 contigs over 1,000 bp in length, with an average coverage of 570X and 38% GC content, which we designated as *Methanosarcina sp.* strain Ant1 (17). The *Methanosarcina sp.* strain Ant1 genome contained 3,593 coding regions, 53 tRNAs, 11 predicted CRISPR regions, and several cytochromes. The 16S rRNA gene sequence found within the *Methanosarcina sp.* strain Ant1 genome has 97% nucleotide sequence identity and 100% coverage to *Methanosarcina lacustris*, a psychrotolerant methanogen isolated from a fen in Moscow (18) (Figure 3.2A). Close cultured relatives are *M. subterranea* strain HC-2 and *M. soligelidi* strain DSM 26065, isolated from a deep-subsurface diatomaceous shale formation and Siberian permafrost-affected soil, respectively (19, 20).

Methane metabolism

Methanogenesis metabolism can likely be achieved through several pathways (21) (Figure 3.3). The entire operon encoding methyl coenzyme M reductase (Mcr) and genes for hydrogenotrophic methanogenesis (*fmd*, *ftr*, *mch*, *mtd*, *mer*, *mtr*ABCDEFGH, and *hdr*ABCDE) were present. Acetoclastic genes encoding carbon monoxide dehydrogenase, acetate kinase, acetyl-coenzyme A synthetase, phosphate acetyltransferase, and the acetyl-CoA decarbonylase/synthase complex provide evidence that this organism is capable of acetoclastic methanogenesis. Methanol metabolism genes encoding the three subunits of methanol—corrinoid protein comethyltransferase— show potential for growth with methanol. *Methanosarcina sp.* strain Ant1 contains monomethylamine methyltransferase and dimethylamine corrinoid protein genes, suggesting growth with methylamines. An incomplete formate dehydrogenase operon suggests that growth with formate is not likely.

Cold adaptation strategies

Studies on elongation factor 2 (EF2) from the model archaeal psychrophile *Methanococcoides burtonii* showed the enzyme activity dependence on temperature (22, 23). The EF2 protein coding sequence was found in *Methanosarcina sp.* strain Ant1. A three-dimensional ribbon model of EF2 enzyme (GTPase) from the genome (Figure 3.4) was compared to EF2 (GTPase) models from *Methanosarcina lacustris*, the closest relative to *Methanosarcina sp.* strain Ant1 by 16S rRNA, a mesophilic relative *Methanosarcina acetivorans*, and a psychrophilic isolate *Methanolobus psychrophilus*. The EF2 model of *Methanosarcina sp.* strain Ant1 had a higher alpha helical content compared to its mesophilic relative, *M. acetivorans*, but was similar to other psychrophilic or psychrotolerant relatives (Table 3.1).

A previous study showed that the psychrophile *M. burtonii* generates unsaturated lipids by selective saturation (24). The *Methanosarcina* sp. strain Ant1 genome contains evidence for de novo synthesis of unsaturated diether lipids through a functional mevalonate pathway (21, 25). Genes encoding the DNA DSB repair Rad50 ATPase, 15 heat shock proteins and 2 cold-shock DEAD-box proteins were detected in *Methanosarcina* sp. strain Ant1, indicating that several defense strategies against environmental stresses are available to this strain. In comparison, the mesophilic *Methanosarcina barkeri* DSM 804 contained 10 heat shock proteins, though no cold-shock proteins were detected. In addition, adaptation to low water activity can be achieved through accumulation of compatible solutes (26). The *Methanosarcina* sp. strain Ant1 genome contains several different transporters for common compatible solutes on the same contig, including glycine betaine/carnitine/choline transport ATP-binding protein opuCA, glycine

betaine/carnitine/choline transport system permease protein opuCB and choline-binding protein precursor.

Discussion

The close phylogenetic relationship of the genome reconstructed in this study to modern psychrophilic taxa pointed us toward investigating genomic adaptations that make possible the ability for a microorganism to remain viable after being locked away in -17°C permafrost for thousands of years (3). Permanently cold environments, such as permafrost, present distinct stressors to cellular functions. Challenges to maintaining membrane fluidity and substrate affinity can be overcome by structural modifications to lipids and proteins, respectively. In *Methanosarcina sp.* Ant1, genomic adaptations to the stress of the permafrost environment include *de novo* synthesis of unsaturated diether lipids through a functional mevalonate pathway, including acetyl-CoA transferase, 3-hydroxy-3-methylglutaryl-coenzyme A reductase, and mevalonate kinase (21). This is perhaps a strategy novel to psychrophilic *Methanosarcina*, as the psychrophile *Methanococcoides burtonii* performs selective saturation instead of forming unsaturated lipids *de novo* (24).

Another adaptation within the *Methanosarcina sp.* Ant1 genome includes protein structural modification. Three-dimensional modeling of the EF2 protein found within the genome illustrated that *Methanosarcina sp.* Ant1 contains increased alpha helical content relative to its mesophilic counterpart, *Methanosarcina acetivorans* (Figure 3.3), with numbers similar to other psychrophilic and psychrotolerant methanogens (Table 3.1). Modifications, such as increased alpha helical content, increased substrate affinity, stronger polar and weakened hydrophobic interactions, allow for greater flexibility in cold environments (27, 28).

Finally, other genomic adaptations include the presence of genes involved in the repair of DNA double strand breaks, which threaten microbial cell viability and are lethal if not repaired (29). The genes encoding the break repair system essential in the recognition and repair of DNA double strand breaks, endonuclease Mre11 and ATPase Rad50, were in the *Methanosarcina sp.* Ant1 genome. The conservation of this system within archaea has been noted previously (30), which expands from distantly related thermophilic archaeon, *Pyrococcus furiosus* (31), to our psychrophilic Methanosarcina genome.

Metabolic versatility regarding pathways and substrates available to *Methanosarcina sp*. Ant1 for methanogenesis was detected in the genome, suggesting the distinct possibility of methane production shortly after the onset of thawing conditions when organic material becomes bioavailable. However, because *Methanosarcina sp.* Ant1 is adapted to cold conditions, the question of whether these cold adaptations—including the structural modification of specially modified, heat-labile enzymes—will prevent this and other cold-adapted microorganisms from thriving under conditions of thaw. It remains to be seen whether warmer temperatures will induce protein denaturation and/or kinetic instability, or if microbes can quickly adapt to their warming environment. The enrichment condition under which this strain was grown (20°C) is certainly evidence that growth and energy generation are possible at higher temperatures, but without the context of a natural setting, predictions are not straightforward.

Conclusion

In conclusion, we have shown the ability of a methanogenic archaeon to remain viable after consistent burial within permafrost for thousands of years. Such long-term viability is likely attributable in part to the adaptations to cold detectable within its genome. Metabolic versatility in the way of methanogenesis suggests that with the onset of warming and permafrost melting,

there is the distinct possibility for elevated methane production and accumulation. The question remains if this production will be mitigated by oxidation before emission to the atmosphere.

References

1. Wagner D, Lipski A, Embacher A, Gattinger A. 2005. Methane fluxes in permafrost habitats of the Lena Delta: effects of microbial community structure and organic matter quality. Environmental Microbiology 7:1582-1592.
2. Anisimov OA. 2007. Potential feedback of thawing permafrost to the global climate system through methane emission. Environmental Research Letters 2:045016.
3. Gilichinsky D, Wilson G, Friedmann E, McKay C, Sletten R, Rivkina E, Vishnivetskaya T, Erokhina L, Ivanushkina N, Kochkina G. 2007. Microbial populations in Antarctic permafrost: biodiversity, state, age, and implication for astrobiology. Astrobiology 7:275-311.
4. Kenealy W, Zeikus J. 1981. Influence of corrinoid antagonists on methanogen metabolism. Journal of Bacteriology 146:133-140.
5. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics:btu170.
6. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2016. metaSPAdes: a new versatile de novo metagenomics assembler. arXiv preprint arXiv:160403071.
7. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.
8. Laczny CC, Sternal T, Plugaru V, Gawron P, Atashpendar A, Margossian HH, Coronado S, van der Maaten L, Vlassis N, Wilmes P. 2015. VizBin-an application for reference-independent visualization and human-augmented binning of metagenomic data. Microbiome 3:1.
9. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics:btu153.
10. 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 research 25:1043-1055.

11. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic acids research 35:3100-3108.

12. Kaplan W, Littlejohn TG. 2001. Swiss-PDB viewer (deep view). Briefings in bioinformatics 2:195-197.

13. Guex N, Peitsch MC. 1997. SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. electrophoresis 18:2714-2723.

14. Arnold K, Bordoli L, Kopp J, Schwede T. 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22:195-201.

15. Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. Nucleic acids research 37:D387-D392.

16. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution:msw054.

17. Buongiorno J, Bird JT, Krivushin K, Oshurkova V, Shcherbakova V, Rivkina EM, Lloyd KG, Vishnivetskaya TA. 2016. Draft genome sequence of antarctic methanogen enriched from dry valley permafrost. Genome announcements 4:e01362-16.

18. Zhilina T, Zavarzin G. 1991. Low-temperature methane production by pure culture of *Methanosarcina* sp. Doklady Akademii Nauk Sssr 317:1242-1245.

19. Shimizu S, Ueno A, Naganuma T, Kaneko K. 2015. *Methanosarcina subterranea* sp. nov., a methanogenic archaeon isolated from a deep subsurface diatomaceous shale formation. International journal of systematic and evolutionary microbiology 65:1167-1171.

20. Wagner D, Schirmack J, Ganzert L, Morozova D, Mangelsdorf K. 2013. *Methanosarcinasoligelidi* sp. nov., a desiccation-and freeze-thaw-resistant methanogenic

archaeon from a Siberian permafrost-affected soil. International journal of systematic and evolutionary microbiology 63:2986-2991.

21. Vishnivetskaya TA, Buongiorno J, Bird J, Krivushin K, Spirina EV, Oshurkova V, Shcherbakova VA, Wilson G, Lloyd KG, Rivkina EM. 2018. Methanogens in the Antarctic Dry Valley Permafrost. FEMS Microbiology Ecology doi:10.1093/femsec/fiy109:fiy109-fiy109.

22. Thomas T, Cavicchioli R. 1998. Archaeal cold-adapted proteins: structural and evolutionary analysis of the elongation factor 2 proteins from psychrophilic, mesophilic and thermophilic methanogens 1. FEBS letters 439:281-286.

23. Thomas T, Cavicchioli R. 2000. Effect of temperature on stability and activity of elongation factor 2 proteins from Antarctic and thermophilic methanogens. Journal of bacteriology 182:1328-1332.

24. Cavicchioli R. 2006. Cold-adapted archaea. Nat Rev Micro 4:331-343.

25. Smit A, Mushegian A. 2000. Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway. Genome Research 10:1468-1484.

26. Russell N, Evans R, Ter Steeg P, Hellemons J, Verheul A, Abee T. 1995. Membranes as a target for stress adaptation. International journal of food microbiology 28:255-261.

27. Thomas T, Cavicchioli R. 2002. Cold adaptation of archaeal elongation factor 2 (EF-2) proteins. Current Protein and Peptide Science 3:223-230.

28. Madigan MT, Martinko JM, Parker J. 1997. Brock biology of microorganisms, vol 514. prentice hall Upper Saddle River, NJ.

29. Wyman C, Kanaar R. 2006. DNA double-strand break repair: all's well that ends well. Annu Rev Genet 40:363-383.

30. White MF, Allers T. 2018. DNA repair in the archaea—an emerging picture. FEMS

Microbiology Reviews 42:514-526.

31. Hopkins BB, Paull TT. 2008. The P. furiosus mre11/rad50 complex promotes 5′ strand resection at a DNA double-strand break. Cell 135:250-260.

Appendix I: Tables and Figures

Table 3.1. Number of predicted EF2 alpha helical domains.

Organism	Relation to temp.	Number of alpha helices
Methanosarcina sp. strain Ant1	Antarctic permafrost at −18°C	23
Methanolobus psychrophilus	Psychrophilic	21
Methanosarcina lacustris	Psychrotolerant	22
Methanosarcina acetivorans	Mesophilic	18

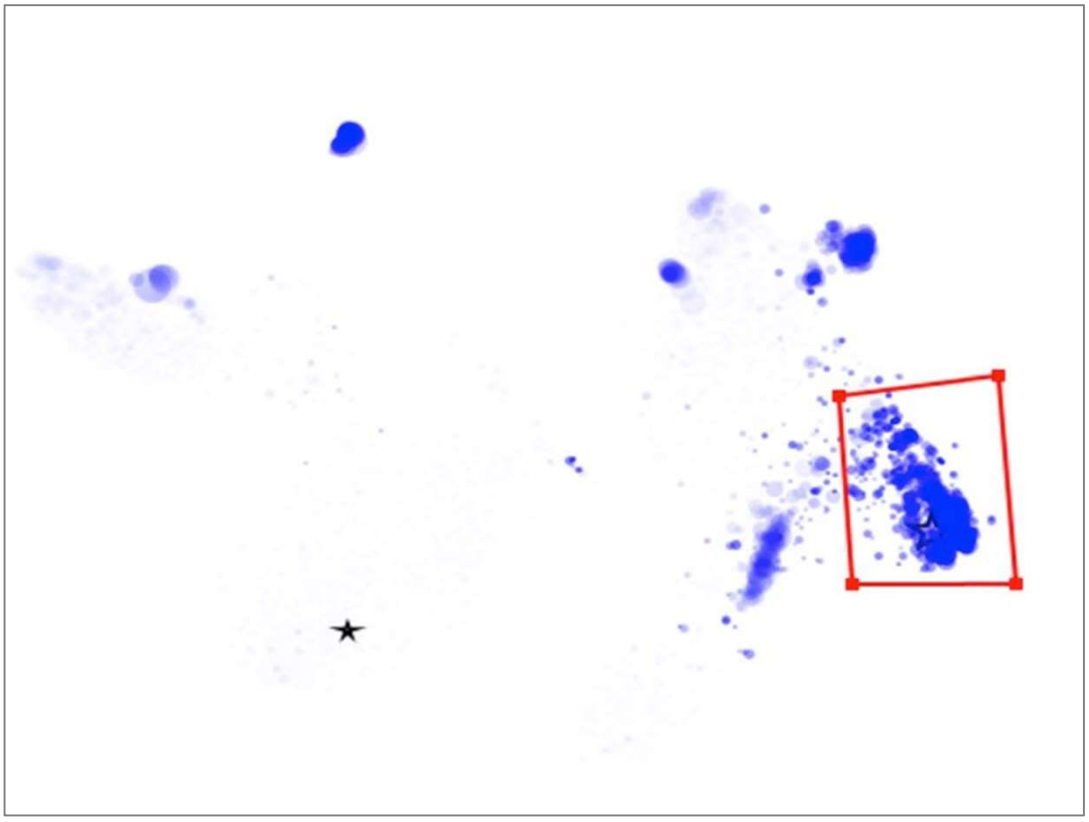

Figure 3.1. Scatter plot visualization in VizBin of the metagenomic dataset Ant1. Coverage and k-mer frequency were used to produce two-dimensional representation of bins within the metagenome Ant1. The stars highlight contigs that contain a homolog for the gene *mcrG*, which is important for the metabolism of methanogens. The red polygon was manually placed around a selection of contigs, which contained a nearly complete genome of the novel uncultured methanogen *Methanosarcina sp.* Ant1.

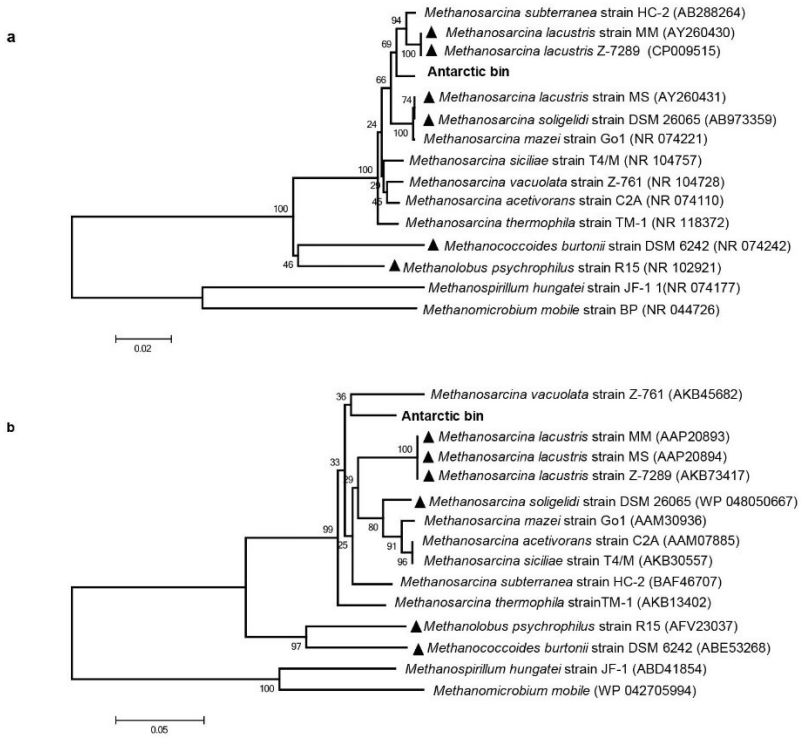

Figure 3.2. Phylogenetic tree for the 16S rRNA gene sequence (a) and mcrA amino acid sequence (b) showing the relationship between the metagenomic bin in this study and relatives of the *Methanosarcina*. Relationships were inferred by using the Neighbor-Joining method. The sequences from *Methanospirillium hungatei* strain JF-1 and *Methanomicrobium mobile* strain BP were used as outgroups for both trees. Accession numbers are in parentheses and numbers at nodes indicate the bootstrap percentages after 1000 replicates calculated in MEGA 7.0 (16). Sequences were obtained by megablast with Blastn 2.4.0 (NCBI) and selected partial and complete 16S rRNA ribosomal sequences were aligned with MUSCLE (EMBL). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. Triangles indicate psychrophilic or psychrotolerant isolates.

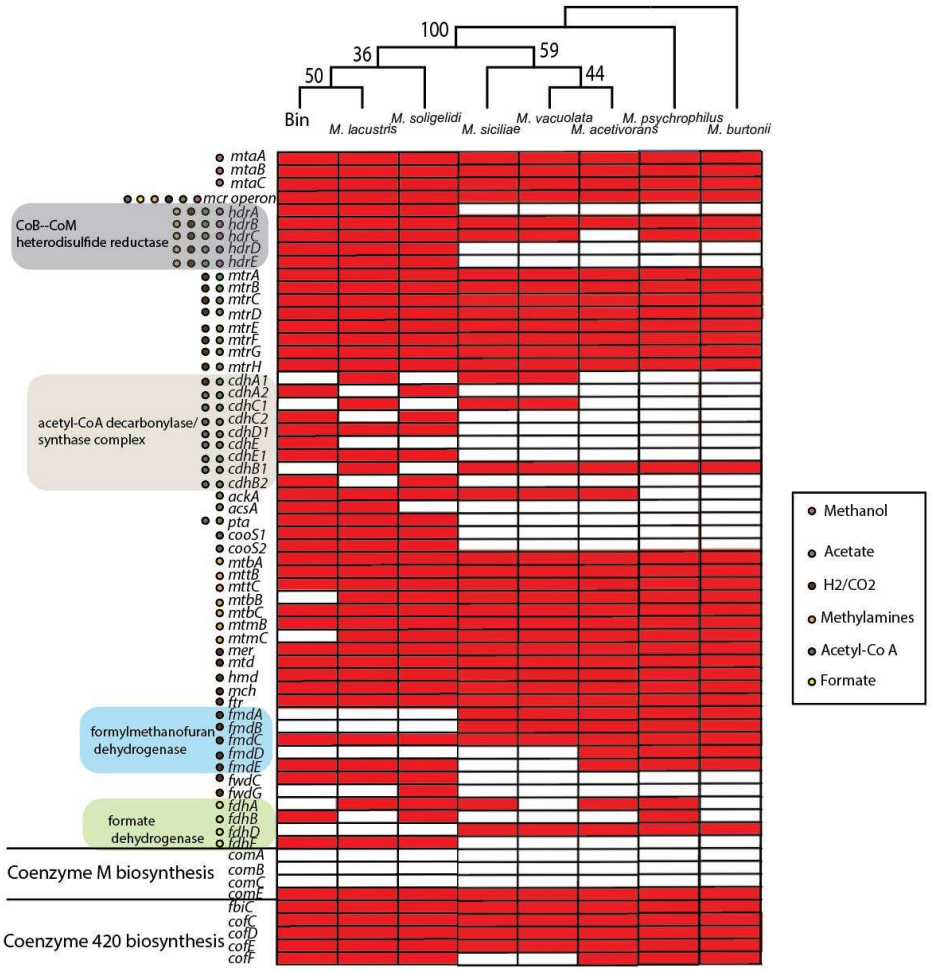

Figure 3.3. Comparative analysis of the presence (red) or absence (white) of genes involved in methane metabolism as detailed in KEGG reference pathway map 00680. Genes to the left of the columns were identified through the pathway reconstruction feature of KEGG and by manual inspection of the genomes. Different colored dots are used to indicate substrate specificity, although some genes are universal to all pathways (*mcr* operon, for example). The genome's subunit composition for the enzymes CoB—CoM heterodisulfide reductase, acetyl-CoA decarbonylase/synthase, formylmethanofuran dehydrogenase, and formate dehydrogenase demonstrate the largest differences compared to its more distant relatives in *Methanosarcina.*

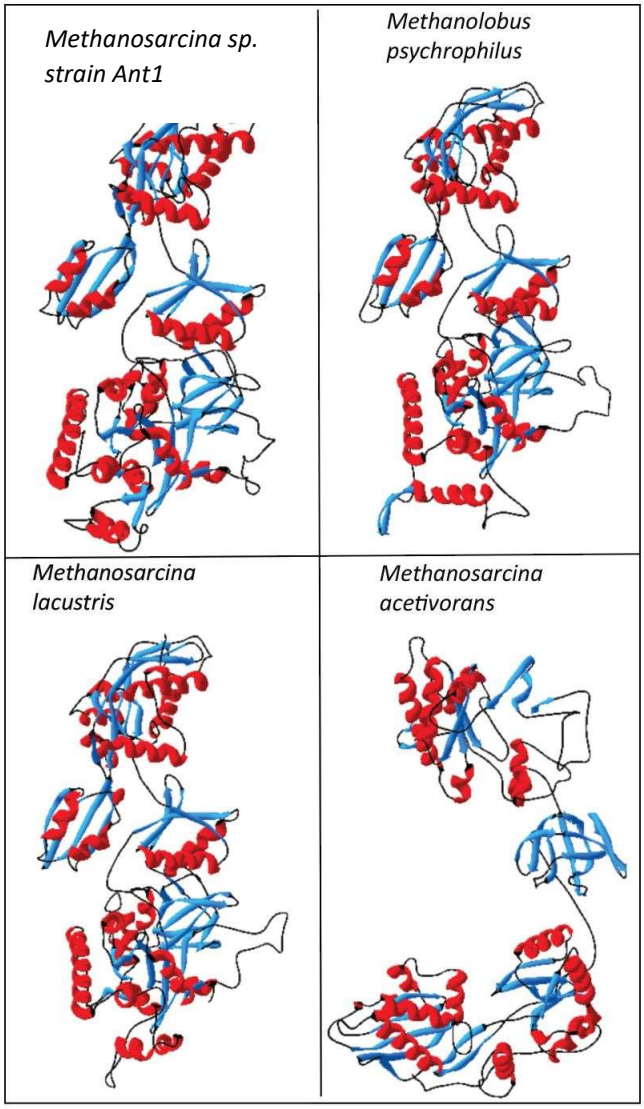

Figure 3.4. Models of Elongation Factor-2 (EF2) using the automated homology server, SwissModeller. Models were generated from the same template (2p.8y in PDBe), which had at least 30% sequence identity and 98% coverage across each target. Models had good quality scores and analysis of ramachondron plots within DeepViewer (12) showed that nearly all dihedral angels were within acceptable limits. DeepViewer was used to construct ribbon models of the figure, with alpha helices in red, beta sheets in blue, and loops in black.

Chapter 4: Cross-fjord trends of complex microbial communities control subsurface iron and sulfur cycling in Arctic sediments

This chapter is a revised version of work submitted previously for publication:

Under revision, **Buongiorno, J.**, Herbert, L., Wehrmann L., Michaud A., Laufer K., Roy H., Jorgensen B.B., Szynkiewicz A., Faiia A., Yeager K., Schindler K., Lloyd K.G., Cross-fjord trends of complex microbial communities control subsurface iron and sulfur cycling in Arctic sediments. (Applied and Environmental Microbiology).

My primary contributions to this paper include: (i) field sampling, (ii) microbiology/organic geochemistry/hydrogen sample processing and DNA extraction, (iii) organic geochemistry/hydrogen measurement collection, (iv) data analysis and network building, and (v) writing of manuscript.

Abstract

In anoxic marine sediments worldwide, complete organic matter oxidation to carbon dioxide is achieved mostly through microbial sulfate reduction, although sediments with high iron content allow substantial dissimilatory iron reduction to occur in upper sediment layers. The relative contribution of iron and sulfate reducers involved in carbon oxidation is controlled by the availability of organic carbon and oxidized iron. In many fjords in the Svalbard archipelago, iron delivery may allow iron reducers to compete successfully with sulfate reducers for common substrates. Here, we explore the biological catalysts that drive iron and sulfur cycling in sediments of Van Keulenfjord, Svalbard. We examined 16S rRNA gene libraries across sediment depth and with increasing distance from the main glacier. Near the main glacier, we found a diverse and abundant iron reducing community above 10 cm depth. Below this, the dominance of sulfate reducers increased. In contrast, at the fjord mouth, iron reducers were restricted to the

upper 5 cm and abundant sulfate reducing bacteria were supported by more labile and higher total organic carbon compared to the middle site. Microbial network analysis demonstrated that station was a strong control on co-occurrence patterns between microbial taxa and that uncultured Sva1033 was interacted with more taxa at station AC. Differences in sulfur and iron microbial communities between the studied sites point to cross-fjord trends in organic geochemistry and microbial community composition that may become increasingly important with changes in sediment loading and primary productivity brought on by glacial retreat.

Introduction

In marine sediments, the complete anaerobic remineralization of organic matter to carbon dioxide is achieved through a complex assemblage of microbial sulfate reducers and iron reducers (1-5). Electron donors for dissimilatory sulfate and iron reduction come from H_2, formate, acetate, or volatile fatty acids produced by microbial fermenters (6, 7). The resulting reduced iron and sulfur compounds interact to form pyrite minerals. When pyrite is buried without further resuspension, sulfate is depleted to free sulfide, which outcompetes microbial iron reducers in the reduction of iron oxides. In Svalbard, however, glacially-derived iron-rich plumes deliver poorly crystalline, biologically available iron oxides to fjord sediments with seasonal melting. This rejuvenates pools of bioavailable iron that could then stimulate iron respiration coupled to organic matter remineralization (8).

Despite being permanently cold (2.6 to -1.7°C (9)), Svalbard sediments demonstrate rates of sulfur cycling (9-12) and microbial activities (13) that are comparable to those of temperate sediments. The combination of the deposition of sediments with high iron to organic matter ratios and heterotrophic communities fueled by rapid iron and sulfur cycles, results in sediments with very low total organic carbon (<1%). Although the geochemical processes have been well-described in Svalbard fjords, the biological catalysts that drive them alongside organic matter remineralization have only been explored in Smeerenburgfjord, which has 16S rRNA from organisms capable of sulfate reduction, iron reduction, fermentation, aerobic heterotrophy, and sulfur oxidation (14). The sulfate reducing community of Svalbard, although diverse, is dominated by Desulfosarcina/Desulfococcus groups (15), and oxidation of free sulfide is performed by large filamentous Beggiatoa, which are absent in some of the other Svalbard fjord sediments (16). Bacterial isolates from Smeerenburgfjord include the genera *Desulfuromusa*,

Desulfuromonas, *Shewanella*, and *Desulfovibrio* which are capable of sulfate, sulfur, and iron reduction (sometimes with multiple electron acceptors used by the same isolate) (17). A high diversity of extracellular enzyme targets is paralleled by a high diversity of heterotrophs, demonstrating a robust organic matter-remineralizing community fueled by the removal of fermentative products by the rapid iron and sulfur reduction (18, 19).

Van Keulenfjorden (Figure 4.1), which is heavily influenced by hematite bedrock, contains sediments characterized by high iron accumulation which predicts enhanced iron-mediated recycling of sulfur species (20). However, the bioavailability of iron deposits, gradients of sedimentary organic carbon, and bioirrigation in sediments in the middle and outer reaches of the fjord (eg (4)) likely support a diverse subsurface community of iron and sulfur cyclers (20). It remains unknown how biological catalysts shape the geochemical environment and here we aim to understand the abundance and diversity of iron and sulfur cycling clades in Van Keulenfjorden sediment. We used qPCR to map patterns of microbial abundance across the fjord and 16S libraries to understand the depth profiles of clades involved with iron and sulfur cycling in sediments from varying distance from the main glacier. Because the main glacier has been surging (21), we sought to map the depth layer over which the suboxic zone extends with distance from the glacier. In addition, we trace changes in organic geochemistry along a spatial transect through the fjord and with depth. Finally, we predicted the members of these communities have synergistic or antagonistic relationships with each other and geochemical parameters that ultimately influence ecological structure in Van Keulenfjorden sediments. To test this, we use network analyses to unravel the connections between microbial taxa and environmental measurements. In recent years, network analysis has been shown to be an

effective tool to get at emergent properties of complex and dynamic microbial ecosystems (22, 23).

Results and Discussion

Sediment characteristics and glacial history

Three separate cores from stations AB and AC in Van Keulenfjorden (Figure 4.1) were collected only cm apart in August 2016. Sediment of Van Keulenfjorden was dark gray to black, sticky and fine-grained; sulfide smell was never detected. Gamma activity was detected for age dating, but non-steady state input of radioisotopes precluded use of ^{210}Pb for age dating (Figure 4.2A). A distinct ^{137}Cs peak at 16 – 17 cm below seafloor (cmbsf), however, indicated the year 1963 (Figure 4.2B)(24), giving a mean sediment accumulation rate of 0.31 ± 0.02 cm y^{-1} over the last ~50 years. Previous measurements in the area have shown a similar sediment accumulation rate of 0.06 cm y^{-1} (21, 25). The near absence of ^{137}Cs coincided with a layer of coarse material, and this layer could have been deposited near instantaneously.

Organic and isotope geochemistry

Total organic carbon (TOC) values averaged 1.4 ± 0.08 wt % at inner station HA, 1.4 ± 0.07 wt % at middle station AC, and 1.5 ± 0.09 wt % at the outer station AB (Figure 4.3A; Table 4.1). In the upper 8 cm, TOC concentrations increased with increasing distance from the glacier. After statistical outliers are removed (Figure 4.4) and when all data from each core are pooled, TOC was statistically higher at outer station AB than at middle station AC (P value of Welsh two-sample t-test = 0.004) and inner station HA (P value of Welsh two-sample t-test = 0.0002). Low TOC content is typical of Svalbard fjords (20, 26, 27) and the pattern of increased TOC with increasing distance from glacial outflow reflects physical processes that hinder the production of autochthonous marine organic material at fjord heads. First, surface water turbidity associated with the suspended load of glacial outflow during summer months limits light

penetration and, in turn, primary production in surface waters closest to this zone of runoff (28). In addition, these low salinity glacial meltwaters create freshwater conditions for which marine zooplankton are not suited (29-31). The limitations brought about by turbidity and low salinity conditions are alleviated further toward the fjord mouth, where phytoplankton readily bloom (28). Because of tight pelagic-benthic coupling (28, 32, 33), cross-fjord signatures of increased water column productivity toward the fjord mouth are captured in the sediment geochemistry (eg. (20)).

The average isotope compositions of carbon ($\delta^{13}C_{org}$) within organic matter of Van Keulenfjord sediment were -26.1 ± 0.24 ‰ at HA, -26.0 ± 0.34 ‰ at AC, and -25.3 ± 0.76 ‰ at AB (Figure 4.2B; Table 4.1). Carbon to nitrogen ratios (C/N) averages were 13.4 ± 0.47 at HA, 13.4 ± 0.52 at AC, and 12.9 ± 0.47 at AB (Figure 4.2C; Table 4.1), with an overall average value of ~13. When C/N was plotted against $\delta^{13}C_{org}$, an overall seaward trend of higher isotope values was observed. At the inner (HA) and middle (AC) sites, organic matter signatures related to terrestrially-derived coal (average -26‰)(34), soil (average -25 ‰)(34) and C_3 land plants (-25 to -35 ‰) (35-37) were detected, while signals of relatively labile, marine phytoplankton, from -22 to -25 ‰ (32, 38), were exclusive to outer site AB (Figure 4.2D). Like $\delta^{13}C_{org}$, C/N ratios can be used to identify the relative contribution of marine versus terrestrial sources to organic carbon pools, with allochthonous, terrestrially-derived organic matter typically ~20 and marine-derived organics ~6 (26, 38). There is general agreement with respect to organic matter source between isotope signals and C/N signals; however, at AB, C/N ratios are greater than average phytoplankton values (32), which we interpret as the preferential removal of nitrogen from bulk organic matter during early diagenesis in the seabed (39).

The seaward gradient of increased carbon amount and lability along the long axis of the fjord has been suggested previously for this and other nearby fjords (20, 26, 28). Previous geochemical measurements demonstrated that increased primary production coupled with decreased sedimentation rates toward the mouth of Van Keulenfjorden results in elevated subsurface aerobic respiration and a shallower zone of metal reduction coupled to organic matter remineralization (20). However, the composition and structure of the microbial communities potentially participating in these metabolic processes in the sediment have not been examined. This drove us to explore how the observed spatial variability in organic matter amount and quality may influence microbial abundance and community structure.

Quantitative PCR

Low DNA extraction yields from station HA sediments prevented us from obtaining quantitative PCR (qPCR) values for this station, despite having used the same methods successfully at stations AB and AC. Either the microorganisms at station HA were in lower abundance, or these sediments had higher concentrations of coextracted inhibitors than sediments at stations AC and AB. At station AB, average bacterial 16S rRNA gene copy numbers ranged from 1.05×10^{8} at 18-19 cmbsf to 1.33×10^{11} 16S rRNA gene copies g fresh sediment^{-1} at 0-1 cmbsf (Figure 4.5; Table 4.2). At AB, most 16S rRNA gene copy numbers are above the range captured in our standard curve (1×10^{9} copies, black dashed line), and extrapolated values are high, even compared to temperate, eutrophic sediments (40). High copy numbers could be due, in part, to differences in 16S rRNA gene copies per cell, which has been shown to average 3.04 copies per cell (41), or difficulties in absolute versus relative quantification with qPCR (42). The high copies of the 16S rRNA gene observed here is supported by previous high rRNA recovery from sediments Hornsund, Svalbard (13), although significant correlation between the number of

prokaryotic cells and rRNA contents was only observed below 6 cm depth. Archaeal 16S rRNA gene copy numbers were lower than bacteria, ranging from 7.42×10^4 16S rRNA gene copies g fresh sediment-1 at 18-19 cmbsf in core to 3.92×10^8 16S rRNA gene copies g fresh sediment^{-1} at 4-5 cmbsf (Figure 4.5). These values were congruent with qPCR measurements of archaea within Smeerenburgfjord, Svalbard, which averaged 1.9×10^8 16S rRNA gene copies g sediment^{-1} in the first 7 cm (43). Despite differences in values, bacteria and archaea had similar 16S rRNA gene copy patterns with depth within each core, and both exhibited only gradual changes with depth. However, these trends differed for the two AB cores which had opposite depth trends over the upper 5 cm, and then values diverged for the deeper depths.

In contrast to site AB, bacterial and archaeal 16S rRNA gene copies were much more dynamic with depth at AC, differing by as much as two to four orders of magnitude in adjacent depth layers. Maximum values were similar at the two sites, but minimum values were much lower at AC than AB. These large oscillations with depth were likely not due to experimental error, since replicate measurements were not statistically significantly different. The patterns of bacterial gene copy numbers were largely synced between the two AC cores in sediments above 15 cm, while the AB cores exhibited low variability between depths using the same measurement techniques. Given the high variability of bacteria and archaea at station AC, it is not clear whether the overall values decrease with depth, as is observed for station AB.

The seaward increase in 16S rRNA copy numbers observed here may reflect the increase in the quality and quantity in organic matter we detected along the long axis of the fjord. Because of tight benthic-pelagic coupling, higher primary productivity at the fjord mouth may support a higher sediment community that receives seasonal input of labile organics relatively undiluted by terrigenous sediment. Further, differences in downcore profiles between the two sites may result

from physical processes that disrupt sediment communities and prokaryote abundance in the middle and inner fjord, such as highly episodic deposition of sediments with meltwater plumes (20), bioturbation (4, 44, 45), and glacial surge events (21).

Community composition

After normalization, we generated a total of 52 libraries across the two stations that produced amplifiable DNA, AB and AC. Rarefaction profiles of 16S rRNA gene sequences never plateaued (Figure 4.6), suggesting that we did not sequence enough to capture the entire breadth of diversity in these sediments. Despite this, our ability to achieve our aim of identifying the distribution and co-occurrence patters of the most abundant sequences was not diminished.

Across all libraries, the majority of reads were identified as bacteria (96 – 97% versus archaea at 3 – 4%). At the phylum level, most sequences were identified as Proteobacteria, making up ~25 to 42% of relative sequence abundance at both stations (Figure 4.7). The relative abundance of Plactomycetes sequences (~10 – 20%) remained steady downcore at both stations compared to other phyla, such as Bacteroidetes. Sequences from Bacteroidetes decreased from 16% in surface sediments to 3% relative abundance at both stations.

Ordination analysis showed that most of the variability in community composition between sites can be explained by TOC, especially within shallow sediment depths (Appendix III). With increasing sediment depth, communities converge on similar compositions. Deeper sediment communities move in ordination space close to vectors related to C/N, $\delta^{13}C_{org}$, and hydrogen . This suggests either that community composition is driven by a combination of effects related to these geochemical parameters, or perhaps that like geochemistry, differences in microbial communities between stations can only be strongly observed in shallow sediment layers.

At both stations, sequences related to anaerobic bacteria likely participating in *in situ* cycling of iron and sulfur species were present, including the Deltaproteobacterial families Desulfobacteraceae and Desulfobulbaceae. High Desulfobacteraceae relative abundance was shown previously in Smeerenburgfjord sediment, with the genera *Desulfosarcina*, *Desulfofrigus*, and *Desulfococcus* the most abundant sulfate reducers (14, 15). Unlike Smeerenburgfjord, however, where Desulfobulbaceae were not able to be detected, Desulfobulbaceae sequences were in high relative abundance across all sites in Van Keulenfjorden. This family has members that can grow through both sulfate (46) and iron reduction (47), suggesting that Van Keulenfjorden sediment, rich in both sulfate and potentially metabolizable iron (20), is perhaps better suited for supporting this clade.

Although there was overall good agreement between the two cores sequenced at each site, minor local heterogeneity in sediment communities was observed at site AB. For example, the 12 – 13 cm interval in core AB.1 had a library composition similar to more shallow depth layers (Figure 4.7). Because this feature was not station-wide (i.e., not also observed in core AB.2), it may have been related to infaunal burrowing/bioturbation (*cf* (4, 44)). Despite minor core to core variability, little difference in sequence composition and relative abundance was observed between the two sites at the family level. At the genus level, however, we observed clear differences in downcore relative sequence abundance between stations, especially in shallow depths. For example, *Desulfococcus* and *Desulfosarcina*, are more abundant in shallow depths of station AB compared to shallow depths at station AC. Specifically, while at both stations the relative sequence abundance of *Desulfococcus* reads at the 0 – 1 cm interval was 0.03 %, rapid increases at station AB allowed relative abundance to exceed 0.1 % at 4 – 5 cm depth, whereas at station AC, relative read abundance did not reach this point until 12 – 13 cm

depth (Figure 4.8A, B). Both *Desulfococcus* and *Desulfosarcina* are able to couple reduction of oxidized sulfur compounds, such as sulfate and sulfite, to the oxidation of volatile fatty acids (48, 49), aromatic compounds (50-52), and H_2 (49, 53). The rapid increases in relative read abundance for genera related to sulfate reducers at station AB was congruent with sulfate reduction rates, which were highest within the first 5 cm, exceeding 50 nmol cm $^{-3}$ d $^{-1}$ (Figure 4.9). High sulfate reduction rates at this interval corresponded with the highest TOC values (Figure 4.2A), suggesting ample electron donors were available to stimulate sulfate reduction. However, sulfate reduction rates at station AC remained low, never exceeding 5 nmol cm $^{-3}$ d $^{-1}$ on average.

The depth distribution of sequences for genera within the closely related Desulfuromonadaceae (*Desulfuromusa*) and Geobacteraceae (*Geopsychrobacter*, *Geothermobacter*, and *Geobacter*) were similarly distinct between stations. The relative abundance of these sequences was highest above ~10 cm in station AB (Figure 4.8A), whereas at station AC, sequence abundance for these genera only slightly decreased or remained steady downcore (Figure 4.8B). Specifically, at station AC, the relative read abundance for *Desulfuromusa* displayed no observable trend with depth, while sequences of Geobacteraceae genera decreased slightly. *Desulfuromusa* and its relatives use various terminal electron acceptors for growth, including Fe(III), Mn(IV), elemental sulfur, and nitrate (54-56). This highlights metabolic plasticity that may allow the use of variable electron acceptors experienced throughout the depth of the core. Similarly, Geobacteraceae contain numerous adaptations that allow them to thrive in iron-rich anoxic marine sediments, including low maintenance energy (57) and the ability to oxidize common fermentation products while reducing Fe(III) or Mn(IV) (56, 58, 59). Differences in depth gradients in the same core between these closely related clades

suggests the potential for metabolic niche differentiation. Further, our genetic data suggests that reducible iron (ferrihydrite, hematite, goethite, and iron carbonates) is present deeper at station AC sediment than at station AB. One potential cause for a deepening of the iron reduction zone is a high sediment deposition rate at AC compared to outer station AB. Spatial differences between middle and outer fjord sediments related to iron accumulation have been suggested to play an important role in biogeochemical cycling of iron and sulfur within nearby Van Mijenfjorden (6) and the same drivers may be at work in Van Keulenfjorden.

At both stations, the vertical zonation between sequences related to iron reducers and those related to sulfur reducers agrees with what is predicted through thermodynamic sorting and energy yield of reduction with Fe(III) and sulfur species (60, 61). However, recent studies have shown that the distribution of iron-cycling bacteria is decoupled from traditional geochemical zonation in sediments and may be driven instead by microniche distribution and metabolic flexibility (62). Increased abundance of sequences related to sulfur reducers below ~5 cm depth at station AB and ~12 cm at station AC is likely an artifact of an absolute decrease in iron-cycling bacteria, as can be interpreted from decreases in bacterial 16S rRNA gene copies past these depths (Figure 4.3A, C). Decreased absolute abundance of clades that reduce iron may occur because of the decreasing availability of high-energy, bioavailable iron (oxyhydr)oxides or manganese oxides with depth (20). In particular, oxidized iron depletion could occur through dissimilatory iron reduction or through abiotic interactions with sulfide generated from microbial sulfate reduction (63, 64). The rapid decline at station AB in *Desulfuromonas* and Geobacteraceae sequences with depth suggests that oxidized iron is quickly exhausted in shallow depths, perhaps from a combination of high rates of dissimilatory iron reduction and sufficient sulfide from sulfate reduction for the chemical scavenging of oxidized iron.

Clades with cultured representatives that oxidize reduced forms of sulfur were also present at both sites, but more abundant at station AC. Sequences for *Arcobacter*, *Sulfurimonas*, *Sulfurovum* (Epsilonproteobacteria), *Cocleimonas* (Gammaproteobacteria), and *Thiobacillus* (Betaproteobacteria) all maintained relatively high sequence abundance with depth at AC (Figure 4.8). These populations could be supported though the presence of sulfur intermediates that are likely generated through abiotic interactions with reducible iron. Specifically, if reducible iron penetrates deeper in station AC sediment, redox conditions remain suboxic in a cryptic iron-sulfur cycle wherein reduced sulfur intermediates, such as elemental sulfur and thiosulfate, become replenished (65-67). This cryptic iron-sulfur cycle then could provide a consistent source of sulfur intermediates that are useful in biological sulfur oxidation for groups like *Cocleimonas* and *Sulfurovum* (68-71). The biological reoxidation of reduced sulfur species and abiotic reoxidation with reducible iron and manganese together may explain the conservation of pore water sulfate with depth previously noted within Van Keulenfjorden sediments (20).

Inconsistent depth trends in the relative abundance of *Mariprofundus* sequences may be related to the distribution of suboxic microniches. Like sequences related to clades known to oxidize sulfur compounds, *Mariprofundus* sequences were more abundant and penetrated deeper in station AC. The two isolates from this group, *Mariprofundus ferroxydans* and *Mariprofundus micogutta*, oxidize Fe(II) with molecular oxygen under microaerophilic conditions (72-74). *Gallionella* sequences were more abundant at station AC and were present at nearly every depth where *Mariprofundus* was found. However, while *Mariprofundus* sequences extended to 15 cmbsf at AB, *Gallionella* sequences were mostly restricted to the first 2 cm at this station (Figure 4.8A). Because station AB is situated near the source of marnie waters to the fjord, these observations agree with environmental studies suggesting that *Mariprofundus* is a strict marine

iron oxidizer, while *Gallionella* is restricted to freshwater systems or maintains low abundance in marine systems (75-77).

Desulfobulbus sequences were highly abundant at both stations and generally increased with depth (Figure 5). *Desulfobulbus* contains members with diverse metabolism, including *Desulfobulbus propionicus*, which can grow while performing dissimilatory iron reduction (47) and *Desulfobulbus alkaliphilus*, which can grow using sulfate and sulfite in the complete oxidation of organic matter (78). Because of this metabolic flexibility, we consider this genus to reduce either iron or sulfur species. Sequences of the uncultured Sva1033 unclassified group displayed depth gradients very similar to *Desulfobulbus* sequences at both stations. Sva1033 was first identified through gene clone libraries of Smeerenburgfjord sediment, where it phylogenetically grouped within the order Desulfuromondales (14). This study found that its closest relative by 16S rRNA gene identity (93.7%) is *Desulfuromonas palmitatis*, a dissimilatory iron reducer capable of oxidizing long-chain fatty acids (79). Sva0081 sediment group sequences increased with depth at both sites. 16S rRNA gene clones of Sva0081 sediment group from Smeerenbergfjord sediment were identified as members of the Desulfobacteraceae (14) and has since been identified in diverse sediment habitats, including from the North Pacific (80), the North Sea (81), and in the Wadden Sea (81). Metagenomic and single cell genome analysis suggests that *Sva0081 sediment group* is an important scavenger of H_2 in marine sediments (81).

Microbial networks

Networks were built to understand how the most abundant (top 30) operational taxonomic units (OTUs) and those with cultured representatives that cycle iron and/or sulfur occur together and with geochemistry (cf (82)). Individual microbial co-occurrence networks

were generated for each core (Figure 4.11) and then merged to find replicated patterns (Figure 4.12). Neither geochemistry (TOC, $\delta^{13}C_{org}$, C/N, H_2) nor SRR were found to have statistically significant relationships, and instead connections were limited to interactions between microbial taxa. Most nodes in both networks represent phylogenetically diverse members of the Deltaproteobacteria which mainly have positive co-occurrence relationships between them. This indicates similar abundance patterns among the Deltaproteobacteria at the two stations. This agrees with observations of relative abundance for *Desulfococcus* and SEEP-SRB1 which showed increased relative abundance with sediment depth, likely related to favorable anoxic conditions (Figure 4.8).

Within the AB network, 53% of nodes were among the most abundant OTUs, including *Desulfobulbus*, *Desulfococcus*, and *Geopsychrobacter*. By contrast, within the larger AC network, most nodes were relatively rare abundance OTUs. We tested if relatively rare taxa are important members of the community by calculating betweenness centrality, or average number of shortest paths. The betweenness centrality metric can be used to identify key members of a microbial community and help generate hypotheses about the functional role of these microorganisms *in situ* (83, 84). At station AB, a relatively low-abundance *Nitrosomonas* OTU had the highest betweenness centrality (Figure 4.12A). Members of the *Nitrosomonas* are chemolithoautotrophs that gain energy through the oxidation of ammonia to nitrate (85, 86) and are crucial nitrogen cyclers in marine sediments (87-89). Nitrate generated by *Nitrosomonas* could perhaps benefit members of the community that rely on nitrate for their metabolism, allowing this relatively rare OTU to impart control on how other members of its community occur together. At station AC, a *Desulfobulbus* OTU had the highest betweenness centrality and the highest relative sequence abundance (Figure 4.12). Further, this OTU had the most

connections with other taxa, suggesting that this OTU represents a "hub" that connects many nodes that aren't directly connected (90). Future work should explore the *in situ* metabolic activity of *Desulfobulbus* and its potential interactions with microbial counterparts in these sediments using targeted genomics and/or incubation approaches.

Epsilonproteobacteria were more represented in the AC network, agreeing with relative sequence abundance for genera like *Arcobacter*, *Sulfurimonas*, and Sulfurovum, which were more abundant at station AC (Figure 4.8). However, sequence abundance did not always predict network results. For example, although no clear distinction in sequence abundance between stations was observed for Sva1033, their OTUs were exclusive to the AC network, in which they only positively correlated with each other and members of Desulfobulbaceae and Desulfobacteraceae, Some genera that were more prevalent at station AC, like *Geothermobacter* and *Geopsychrobacter*, were present in both networks, while others, such as *Gallionella* and *Mariprofundus*, did not appear in our networks at all. Further, station-specific co-occurrence patterns observed for the same OTUs, such as *Desulfococcus* and *Geopsychrobacter*, suggests that distance from the glacier was a strong control on interactions between microbial taxa.

Conclusions

In conclusion, the sediments of Van Keulenfjorden contain a highly abundant and diverse consortium of bacteria and archaea that is supported by a cross-fjord gradient of increasing amount and bioavailability of organic matter moving toward the mouth of the fjord. Our work supports previous hypotheses that glacial proximity predicts sediment microbial community composition and structure. Sequence analysis suggested that the upper sediment carbon oxidizing niches between the outer and middle stations were occupied by sulfate reducers and iron reducers, respectively. Interactions between iron and sulfur chemistry in the sediment

encouraged high rates of sulfate reduction in shallow depths at station AB, which was matched by our sequence analysis. At station AC, low rates of sulfate reduction and sulfide production allowed a deeper zone of reducible iron available for dissimilatory iron reduction, which agreed with sequence analysis that showed that iron cycling genera penetrated deeper in station AC sediments. Interaction networks suggested that co-occurrence patterns between microbial taxa are not strongly influenced by glacial proximity for most taxa, with the notable exception of the uncultured Sva1033 clade, which network connections allow us to hypothesize occupies a similar ecological niche as members of Desulfobulbaceae and Desulfobacteraceae. More work is necessary to uncover the biological and environmental conditions that favor members of Sva1033.

Our results support the hypothesized alteration in reduced iron delivery to the open ocean along Western Svalbard predicted by Wehrmann et al. (2014). Enhanced sulfate reduction occurring farthest from the glaciers chemically binds up any free reduced iron that is generated through iron reduction with the generation of iron sulfide minerals. As glaciers continue to recede, we predict that conditions that prevail at AB will also characterize AC, and thus the delivery of reduced iron to the open ocean will further decline. Limited export of reduced iron may impact primary production along the shelf, where removal of this key micronutrient will decrease phytoplankton populations that represent a large sink for carbon dioxide in the atmosphere.

Material and methods

Sample collection

Cores from stations AB and AC in Van Keulenfjorden were collected in August 2016. Poly-carbonate core liners were used to subsample HAPs corers (91) at each site, with each core

(AB.1, AB.2 at site AB and AC.1, AC.2 at AC) taken centimeters apart, down to a depth of ~20 cm below sea floor (cmbsf). Cores were stored at 4°C until they were ready for processing within 8 hours. A metal plate and collar were used to section at 1 cm intervals. Cores destined for molecular work were processed sterilely outside, where air temperatures remained near *in situ* temperatures (~4°C). Cores for geochemical analyses were processed inside the Kings Bay Marine Lab at room temperature. Sediment samples for organic geochemistry were stored at -80°C until processed.

Sedimentation accumulation rate

Frozen sediment was shipped on dry ice to University of Kentucky for analysis of natural and anthropogenic ɣ-emitters via low-level ɣ-spectroscopy. Sediment accumulation was then calculated from the depth where the maximum activity of ^{137}Cs was found, divided by the time since 1963. This model is based on the assumption of limited vertical mobility of cesium in sediments (92-94).

Organic and isotope geochemistry

Sediment for analysis of organic matter was freeze-dried after thawing from -80°C and subjected to acid fumigation overnight before analysis (95). Total organic carbon as well as isotope composition of carbon and nitrogen from bulk organic matter was measured using a Thermo-Finnigan Delta XL mass spectrometer coupled to an elemental analyzer at The University of Tennessee, Knoxville. Carbon to nitrogen (C/N) ratios were calculated by dividing percent C by percent N. Isotopic values were calibrated against the USGS40 and USGS41 international standards. In-house standard sets were run every 12 samples. Outliers were determined using Cook's distance (96) in R (97). Across multiple runs, one standard deviation was 0.1-0.2 ‰ for $\delta^{13}C_{org}$, 1.1-1.8 % for mgN, and 1.0-2.2 % for mgC.

Quantitative PCR

Genomic DNA was extracted from approximately 2 g of Svalbard sediment per depth using the RNeasy Power Soil kit for RNA extraction with the DNA accessory kit (QIAGEN, Valencia, CA). DNA extracts were stored at -80 °C until required. We tested 1:1 dilutions and 1:40 dilutions to identify the most suitable concentrations of DNA for qPCR, but found that undiluted DNA extracts provided the lowest Ct values. Total 16S rRNA gene copy numbers of bacteria and archaea were quantified with qPCR using domain-specific primers. The sequence for the bacterial primer pair Bac340f/Bac515r was, 5'- TCCTACGGGAGGCAGCAGT-3' for the forward primer, and 5'GGACTACCAGGGTATCTAATCCTGTT-3' for the reverse primer (74). The sequence for the archaeal primer pair Arch806f/Arch915r was 5'-ATT AGA TAC CCS BGT AGT CC-3' for the forward primer and 5'- GTG CTC CCC CGC CAA TTC CT-3' for the reverse primer (75, 76). Extracted DNA was amplified with a BioRad DNA Engine Option 2 system (Applied Biosystems, Foster City, CA) using SYBR Green chemistry (Invitrogen master mix). Serial dilutions of extracted plasmids containing amplified partial 16S rRNA genes were used as standards for bacteria and archaea, ranging from 1. Nuclease free water was used as a negative control and undiluted DNA extracts were used as templates. Results of qPCR were rejected if the R^2 of the standard curve was below 0.95, or if there was evidence of primer dimers within the melt curve. The quantification limit of qPCR was defined as having fluorescence threshold cycle numbers (Ct) well within those of the simultaneously run standard curve and being at least 3 Ct below the non-template control Ct. Across multiple runs, the standard curve ranged in copy numbers from 1×10^2 and 1×10^9. Gene copy numbers were converted into gene copies g^{-1} fresh sediment by accounting for how much sediment was used for each extraction. For most depths within each core, two technical replicates were performed.

16S rRNA gene libraries

Taxonomic diversity of Svalbard sediments was evaluated using 16S rRNA gene library sequencing. Genomic DNA extracts from AB.1, AB.2, AC.1, and AC.2 were used to generate 16S rRNA amplicon libraries. The Phusion Master Mix (Thermo Fisher) was used with the primer set 515F/806R (98) at the Center for Environmental Biology at The University of Tennessee, Knoxville for amplification. Reads were sequenced with Illumina MiSeq and trimmed for quality with Trimmomatic using a window 10 base pairs wide and a minimum phred score of 28 (99). Trimmed reads were then processed in mothur 1.35.1 (100) using the computational cluster at the Bioinformatics Resource Facility at The University of Tennessee, Knoxville. OTUs were clustered *de novo* at the 97% similarity level with the SILVA release 123 (101). Rarefaction analysis was calculated in mothur with "rarefaction.single" and reads were normalized with "normalize.shared" (norm = 60000).

Hydrogen

Samples for hydrogen analysis consisted of 1 mL of sediment placed into a dark glass serum vial which was then crimp sealed, hand shaken, and gassed with N_2 for 15 min prior to storage at 4°C. Headspace was measured with glass syringes on a Peak Performer GC at The University of Tennessee, Knoxville after 4 days.

Microbial network analysis

To evaluate the co-correlation of target OTUs, we generated microbial networks using relative abundance at the OTU level from all four cores with the Pearson correlation coefficient calculated in the extended local similarity analysis (eLSA) program (81, 82). While abundance measures with 16S rRNA genes are likely not true measures of total abundance, as primer bias can underrepresent or overrepresent specific sequences (102), relative sequence abundance may still be related to actual abundance *in situ*. Networks excluded OTUs whose sum did not reach

0.1% of reads across all libraries from a core. PercentileZ normalization was used in network construction and a strict P-value cutoff of <0.001 was used to determine statistically significant co-occurrence patterns, which ranged in Pearson's r values from -0.95 to 1. At this P-value, the false discovery rate, or q-estimation, was 0.

Networks were visualized with the organic layout in Cytoscape 3.5.1 (103). Betweenness was calculated with the Analyze Network module in Cytoscape by treating edges as undirected (85). The randomness of the generated networks was tested through examination of the degree distribution. Degree is a node attribute that is simply the sum of all direct connections involving that node. As random networks are characterized by a degree distribution fitting a Poisson distribution (104), we used a Chi Square (χ^2) test to determine the goodness of fit between observed and expected degree distributions if originating from a Poisson distribution and found that our networks were not random (105).

Sulfate reduction rates

In situ sulfate reduction rates (SRR) were determined via the whole-core injection method (106) in 2.5 cm wide and ca. 20 long sub-cores that were taken from a HAPs core. Per 1-cm depth interval, 50 kBq of $^{35}S\text{-}SO_4^{2-}$ was injected through pre-drilled holes in the coring tube that were sealed with polyurethane-based elastic glue. Whole cores were incubated for 14 to16 hours at 2°C. The incubation was stopped by splicing the core in 1cm sections and mixing each section with 10 ml of 10% zinc acetate. Samples were stored at -20°C before radiolabeled total reduced inorganic sulfur (TRIS) was recovered and separated from $^{35}S\text{-}SO_4^{2-}$ using the cold chromium distillation method (107). Radioactivities of the distillate and of sulfate in the sample were analyzed using scintillation counting and sulfate reduction rates were calculated according to Jørgensen (1978).

References

1. Thamdrup B. 2000. Bacterial manganese and iron reduction in aquatic sediments, p 41-84, Advances in microbial ecology. Springer.
2. Canfield DE, Thamdrup B, Hansen JW. 1993. The anaerobic degradation of organic matter in Danish coastal sediments: iron reduction, manganese reduction, and sulfate reduction. Geochimica et Cosmochimica Acta 57:3867-3883.
3. Jørgensen BB. 1982. Mineralization of organic matter in the sea bed: the role of sulphate reduction. Nature 296:643 - 645.
4. Kostka JE, Thamdrup B, Glud RN, Canfield DE. 1999. Rates and pathways of carbon oxidation in permanently cold Arctic sediments. Marine Ecology Progress Series:7-21.
5. Schippers A, Jørgensen BB. 2002. Biogeochemistry of pyrite and iron sulfide oxidation in marine sediments. Geochimica et Cosmochimica Acta 66:85-92.
6. Lovley DR, Phillips EJ. 1988. Novel mode of microbial energy metabolism: organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. Applied and environmental microbiology 54:1472-1480.
7. Capone DG, Kiene RP. 1988. Comparison of microbial dynamics in marine and freshwater sediments: contrasts in anaerobic carbon catabolism. Limnology and Oceanography 33:725-749.
8. van de Velde S, Meysman FJ. 2016. The influence of bioturbation on iron and sulphur cycling in marine sediments: a model analysis. Aquatic Geochemistry 22:469-504.
9. Arnosti C, Jørgensen B, Sagemann J, Thamdrup B. 1998. Temperature dependence of microbial degradation of organic matter in marine sediments: polysaccharide hydrolysis, oxygen consumption, and sulfate reduction. Marine Ecology Progress Series:59-70.
10. Kostka JE, Gribsholt B, Petrie E, Dalton D, Skelton H, Kristensen E. 2002. The rates and pathways of carbon oxidation in bioturbated saltmarsh sediments. Limnology and Oceanography

47:230-240.

11. Knoblauch C, Jørgensen BB, Harder J. 1999. Community Size and Metabolic Rates of Psychrophilic Sulfate-Reducing Bacteria in Arctic Marine Sediments. Appl Environ Microbiol 65:4230-4233.

12. Sagemann J, Jørgensen B, Greeff O. 1998. Temperature dependence and rates of sulfate reduction in cold sediments of Svalbard, Arctic Ocean. Geomicrobiology Journal 15:85-100.

13. Sahm K, Berninger U-G. 1998. Abundance, vertical distribution, and community structure of benthic prokaryotes from permanently cold marine sediments (Svalbard, Arctic Ocean). Marine Ecology Progress Series:71-80.

14. Ravenschlag K, Sahm K, Pernthaler J, Amann R. 1999. High bacterial diversity in permanently cold marine sediments. Applied and environmental microbiology 65:3982-3989.

15. Ravenschlag K, Sahm K, Knoblauch C, Jørgensen BB, Amann R. 2000. Community structure, cellular rRNA content, and activity of sulfate-reducing bacteria in marine arctic sediments. Applied and Environmental Microbiology 66:3592-3602.

16. Jørgensen BB, Dunker R, Grünke S, Røy H. 2010. Filamentous sulfur bacteria, Beggiatoa spp., in arctic marine sediments (Svalbard, 79 N). FEMS microbiology ecology 73:500-513.

17. Vandieken V, Finke N, Jørgensen BB. 2006. Pathways of carbon oxidation in an Arctic fjord sediment (Svalbard) and isolation of psychrophilic and psychrotolerant Fe (III)-reducing bacteria. Marine Ecology Progress Series 322:20-41.

18. Teske A, Durbin A, Ziervogel K, Cox C, Arnosti C. 2011. Microbial community composition and function in permanently cold seawater and sediments from an Arctic fjord of Svalbard. Applied and Environmental Microbiology 77:2008-2018.

19. Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen A, Teske A. 2014.

Verrucomicrobia are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. Applied and environmental microbiology 80:3749-3756.

20. Wehrmann LM, Formolo MJ, Owens JD, Raiswell R, Ferdelman TG, Riedinger N, Lyons TW. 2014. Iron and manganese speciation and cycling in glacially influenced high-latitude fjord sediments (West Spitsbergen, Svalbard): Evidence for a benthic recycling-transport mechanism. Geochim Cosmochim Acta 141:628-655.

21. Sund M, Lauknes T, Eiken T. 2014. Surge dynamics in the Nathorstbreen glacier system, Svalbard. The Cryosphere 8:623-638.

22. Ruan Q, Dutta D, Schwalbach MS, Steele JA, Fuhrman JA, Sun F. 2006. Local similarity analysis reveals unique associations among marine bacterioplankton species and environmental factors. Bioinformatics 22:2532-2538.

23. Fuhrman JA, Steele JA. 2008. Community structure of marine bacterioplankton: patterns, networks, and relationships to function. Aquatic Microbial Ecology 53:69-81.

24. Ritchie JC, McHenry JR. 1990. Application of radioactive fallout cesium-137 for measuring soil erosion and sediment accumulation rates and patterns: a review. Journal of environmental quality 19:215-233.

25. Kempf P, Forwick M, Laberg JS, Vorren TO. 2013. Late Weichselian and Holocene sedimentary palaeoenvironment and glacial activity in the high-arctic Van Keulenfjorden, Spitsbergen. The Holocene 23:1607-1618.

26. Winkelmann D, Knies J. 2005. Recent distribution and accumulation of organic carbon on the continental margin west off Spitsbergen. Geochemistry, Geophysics, Geosystems 6.

27. Buongiorno J, Kenneth W, Loy A, Lloyd KG. In preparation. Genomic and transcriptional evidence for physiological responses to burial of the dominant carbon-fixing clade Woeseiaceae

in Arctic fjord sediment. ISME J.

28. Hop H, Pearson T, Hegseth EN, Kovacs KM, Wiencke C, Kwasniewski S, Eiane K, Mehlum F, Gulliksen B, Wlodarska-Kowalczuk M. 2002. The marine ecosystem of Kongsfjorden, Svalbard. Polar Research 21:167-208.

29. Zajaczkowski M, Legezynska J. 2001. Estimation of zooplankton mortality caused by an Arctic glacier outflow. Oceanologia 43:341-351.

30. Buchholz F, Buchholz C, Weslawski JM. 2010. Ten years after: krill as indicator of changes in the macro-zooplankton communities of two Arctic fjords. Polar Biology 33:101-113.

31. Wçslawski JM, Legezytńska J. 1998. Glaciers caused zooplankton mortality? Journal of Plankton Research 20:1233-1240.

32. McMahon KW, Ambrose Jr WG, Johnson BJ, Sun M-Y, Lopez GR, Clough LM, Carroll ML. 2006. Benthic community response to ice algae and phytoplankton in Ny Ålesund, Svalbard. Marine Ecology Progress Series 310:1-14.

33. Søreide JE, Carroll ML, Hop H, Ambrose WG, Hegseth EN, Falk-Petersen S. 2013. Sympagic-pelagic-benthic coupling in Arctic and Atlantic waters around Svalbard revealed by stable isotopic and fatty acid tracers. Marine Biology Research 9:831-850.

34. Kim J-H, Peterse F, Willmott V, Kristensen DK, Baas M, Schouten S, Sinninghe Damsté J. 2011. Large ancient organic matter contributions to Arctic marine sediments (Svalbard). Limnology and Oceanography 56:1463-1474.

35. Fry B, Sherr EB. δ13C Measurements as Indicators of Carbon Flow in Marine and Freshwater Ecosystems, p 196-229. In (ed), Springer New York,

36. O'Leary MH. 1981. Carbon isotope fractionation in plants. Phytochemistry 20:553-567.

37. O'Leary MH. 1988. Carbon isotopes in photosynthesis. Bioscience 38:328-336.

38. Schubert CJ, Calvert SE. 2001. Nitrogen and carbon isotopic composition of marine and terrestrial organic matter in Arctic Ocean sediments: implications for nutrient utilization and organic matter composition. Deep Sea Research Part I: Oceanographic Research Papers 48:789-810.

39. Lehmann MF, Bernasconi SM, Barbieri A, McKenzie JA. 2002. Preservation of organic matter and alteration of its carbon and nitrogen isotope composition during simulated and in situ early sedimentary diagenesis. Geochimica et Cosmochimica Acta 66:3573-3584.

40. Blazejak A, Schippers A. 2010. High abundance of JS-1- and Chloroflexi-related Bacteria in deeply buried marine sediments revealed by quantitative, real-time PCR. FEMS Microbiology Ecology 72:198-207.

41. Lloyd KG, May MK, Kevorkian RT, Steen AD. 2013. Meta-analysis of quantification methods shows that archaea and bacteria have similar abundances in the subseafloor. Appl Environ Microbiol 79:7790-7799.

42. Buongiorno J, Turner S, Webster G, Asai M, Shumaker AK, Roy T, Weightman A, Schippers A, Lloyd KG. 2017. Interlaboratory quantification of Bacteria and Archaea in deeply buried sediments of the Baltic Sea (IODP Expedition 347). FEMS microbiology ecology 93.

43. Kubo K, Lloyd KG, Biddle JF, Amann R, Teske A, Knittel K. 2012. Archaea of the Miscellaneous Crenarchaeotal Group are abundant, diverse and widespread in marine sediments. ISME J 6:1949-1965.

44. Glud RN, Holby O, Hoffmann F, Canfield DE. 1998. Benthic mineralization and exchange in Arctic sediments (Svalbard, Norway). Marine Ecology Progress Series:237-251.

45. Jørgensen BB, Glud RN, Holby O. 2005. Oxygen distribution and bioirrigation in Arctic fjord sediments (Svalbard, Barents Sea). Marine Ecology Progress Series 292:85-95.

46. Widdel F, Pfennig N. 1981. Studies on dissimilatory sulfate-reducing bacteria that decompose fatty acids. Archives of microbiology 129:395-400.

47. Holmes DE, Bond DR, Lovley DR. 2004. Electron transfer by Desulfobulbus propionicus to Fe (III) and graphite electrodes. Applied and Environmental Microbiology 70:1234-1237.

48. Platen H, Temmes A, Schink B. 1990. Anaerobic degradation of acetone by Desulfococcus biacutus spec. nov. Archives of microbiology 154:355-361.

49. Harms G, Zengler K, Rabus R, Aeckersberg F, Minz D, Rosselló-Mora R, Widdel F. 1999. Anaerobic Oxidation of o -Xylene, m -Xylene, and Homologous Alkylbenzenes by New Types of Sulfate-Reducing Bacteria. Applied and Environmental Microbiology 65:999-1004.

50. Evans WC, Fuchs G. 1988. Anaerobic degradation of aromatic compounds. Annual Reviews in Microbiology 42:289-317.

51. Dagley S. 1971. Catabolism of aromatic compounds by micro-organisms, p 1-46, Advances in microbial physiology, vol 6. Elsevier.

52. Rabus R, Nordhaus R, Ludwig W, Widdel F. 1993. Complete oxidation of toluene under strictly anoxic conditions by a new sulfate-reducing bacterium. Applied and Environmental Microbiology 59:1444-1451.

53. Imhoff-Stuckle D, Pfennig N. 1983. Isolation and characterization of a nicotinic acid-degrading sulfate-reducing bacterium, *Desulfococcus niacini* sp. nov. Archives of microbiology 136:194-198.

54. Greene AC, Patel BK, Yacob S. 2009. *Geoalkalibacter subterraneus* sp. nov., an anaerobic Fe (III)-and Mn (IV)-reducing bacterium from a petroleum reservoir, and emended descriptions of the family Desulfuromonadaceae and the genus Geoalkalibacter. International journal of systematic and evolutionary microbiology 59:781-785.

55. Liesack W, Finster K. 1994. Phylogenetic analysis of five strains of gram-negative, obligately anaerobic, sulfur-reducing bacteria and description of *Desulfuromusa* gen. nov., including *Desulfuromusa kysingii* sp. nov., *Desulfuromusa bakii* sp. nov., and *Desulfuromusa succinoxidans* sp. nov. International Journal of Systematic and Evolutionary Microbiology 44:753-758.

56. Holmes DE, Nicoll JS, Bond DR, Lovley DR. 2004. Potential role of a novel psychrotolerant member of the family Geobacteraceae, *Geopsychrobacter electrodiphilus* gen. nov., sp. nov., in electricity production by a marine sediment fuel cell. Applied and Environmental Microbiology 70:6023-6030.

57. Lin B, Westerhoff HV, Röling WF. 2009. How Geobacteraceae may dominate subsurface biodegradation: physiology of *Geobacter metallireducens* in slow-growth habitat-simulating retentostats. Environmental microbiology 11:2425-2433.

58. Prakash O, Gihring TM, Dalton DD, Chin K-J, Green SJ, Akob DM, Wanger G, Kostka JE. 2010. *Geobacter daltonii* sp. nov., an Fe (III)-and uranium (VI)-reducing bacterium isolated from a shallow subsurface exposed to mixed heavy metal and hydrocarbon contamination. International journal of systematic and evolutionary microbiology 60:546-553.

59. Nevin KP, Holmes DE, Woodard TL, Hinlein ES, Ostendorf DW, Lovley DR. 2005. *Geobacter bemidjiensis* sp. nov. and *Geobacter psychrophilus* sp. nov., two novel Fe (III)-reducing subsurface isolates. International Journal of Systematic and Evolutionary Microbiology 55:1667-1674.

60. Froelich PN, Klinkhammer GP, Bender ML, Luedtke NA, Heath GR, Cullen D, Dauphin P, Hammond D, Hartman B, Maynard V. 1979. Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. Geochim Cosmochim Acta

43:1075-1090.

61. Orcutt BN, Sylvan JB, Knab NJ, Edwards KJ. 2011. Microbial Ecology of the Dark Ocean above, at, and below the Seafloor. Microbiology and Molecular Biology Reviews 75:361-422.

62. Otte JM, Harter J, Laufer K, Blackwell N, Straub D, Kappler A, Kleindienst S. 2018. The distribution of active iron-cycling bacteria in marine and freshwater sediments is decoupled from geochemical gradients. Environmental microbiology 20:2483–2499.

63. Enning D, Garrelfs J. 2014. Corrosion of iron by sulfate-reducing bacteria: new views of an old problem. Appl Environ Microbiol 80:1226-1236.

64. Enning D, Venzlaff H, Garrelfs J, Dinh HT, Meyer V, Mayrhofer K, Hassel AW, Stratmann M, Widdel F. 2012. Marine sulfate-reducing bacteria cause serious corrosion of iron under electroconductive biogenic mineral crust. Environmental microbiology 14:1772-1787.

65. Elsgaard L, Jørgensen BB. 1992. Anoxic transformations of radiolabeled hydrogen sulfide in marine and freshwater sediments. Geochimica et Cosmochimica Acta 56:2425-2435.

66. Berg P, Rysgaard S, Thamdrup B. 2003. Dynamic modeling of early diagenesis and nutrient cycling. A case study in an artic marine sediment. American journal of science 303:905-955.

67. Holmkvist L, Ferdelman TG, Jørgensen BB. 2011. A cryptic sulfur cycle driven by iron in the methane zone of marine sediment (Aarhus Bay, Denmark). Geochim Cosmochim Acta 75:3581-3599.

68. Klatt JM, Polerecky L. 2015. Assessment of the stoichiometry and efficiency of CO(2) fixation coupled to reduced sulfur oxidation. Frontiers in Microbiology 6:484.

69. Grimm F, Franz B, Dahl C. 2008. Thiosulfate and sulfur oxidation in purple sulfur bacteria, p 101-116, Microbial Sulfur Metabolism (Dahl C & Friedrich CG, eds), pp. 101–116. Springer

Verlag, Berlin

70. Tanaka N, Romanenko LA, Iino T, Frolova GM, Mikhailov VV. 2011. *Cocleimonas flava* gen. nov., sp. nov., a gammaproteobacterium isolated from sand snail (Umbonium costatum). International journal of systematic and evolutionary microbiology 61:412-416.

71. Inagaki F, Takai K, Nealson KH, Horikoshi K. 2004. *Sulfurovum lithotrophicum* gen. nov., sp. nov., a novel sulfur-oxidizing chemolithoautotroph within the ε-Proteobacteria isolated from Okinawa Trough hydrothermal sediments. International Journal of Systematic and Evolutionary Microbiology 54:1477-1482.

72. Singer E, Emerson D, Webb EA, Barco RA, Kuenen JG, Nelson WC, Chan CS, Comolli LR, Ferriera S, Johnson J. 2011. Mariprofundus ferrooxydans PV-1 the first genome of a marine Fe (II) oxidizing Zetaproteobacterium. PloS one 6:e25386.

73. Makita H, Tanaka E, Mitsunobu S, Miyazaki M, Nunoura T, Uematsu K, Takaki Y, Nishi S, Shimamura S, Takai K. 2017. *Mariprofundus micogutta* sp. nov., a novel iron-oxidizing zetaproteobacterium isolated from a deep-sea hydrothermal field at the Bayonnaise knoll of the Izu-Ogasawara arc, and a description of Mariprofundales ord. nov. and Zetaproteobacteria classis nov. Archives of Microbiology 199:335-346.

74. Scott JJ, Breier JA, Luther III GW, Emerson D. 2015. Microbial iron mats at the Mid-Atlantic Ridge and evidence that Zetaproteobacteria may be restricted to iron-oxidizing marine systems. PLoS One 10:e0119284.

75. Emerson D, Fleming EJ, McBeth JM. 2010. Iron-oxidizing bacteria: an environmental and genomic perspective. Annual review of microbiology 64:561-583.

76. McBeth JM, Little BJ, Ray RI, Farrar KM, Emerson D. 2011. Neutrophilic iron-oxidizing "Zetaproteobacteria" and mild steel corrosion in nearshore marine environments. Applied and

Environmental Microbiology 77:1405-1412.

77. McBeth JM, Fleming EJ, Emerson D. 2013. The transition from freshwater to marine iron-oxidizing bacterial lineages along a salinity gradient on the S heepscot R iver, M aine, USA. Environmental Microbiology Reports 5:453-463.

78. Sorokin DY, Tourova TP, Panteleeva AN, Muyzer G. 2012. *Desulfonatronobacter acidivorans* gen. nov., sp. nov. and *Desulfobulbus alkaliphilus* sp. nov., haloalkaliphilic heterotrophic sulfate-reducing bacteria from soda lakes. International Journal of Systematic and Evolutionary Microbiology 62:2107-2113.

79. Coates JD, Lonergan DJ, Philips EJP, Jenter H, Lovley DR. 1995. *Desulfuromonas palmitatis* sp. nov., a marine dissimilatory Fe(III) reducer that can oxidize long-chain fatty acids. Archives of Microbiology 164:406-413.

80. Li H, Yu Y, Luo W, Zeng Y, Chen B. 2009. Bacterial diversity in surface sediments from the Pacific Arctic Ocean. Extremophiles 13:233-246.

81. Dyksma S, Pjevac P, Ovanesov K, Mussmann M. 2017. Evidence for H2 consumption by uncultured Desulfobacterales in coastal sediments. Environmental microbiology.

82. Fuhrman JA. 2009. Microbial community structure and its functional implications. Nature 459:193-199.

83. Lupatini M, Suleiman AKA, Jacques RJS, Antoniolli ZI, de Siqueira Ferreira A, Kuramae EE, Roesch LFW. 2014. Network topology reveals high connectance levels and few key microbial genera within soils. Frontiers in Environmental Science 2.

84. Berry D, Widder S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. Frontiers in microbiology 5:219.

85. Chain P, Lamerdin J, Larimer F, Regala W, Lao V, Land M, Hauser L, Hooper A, Klotz

M, Norton J. 2003. Complete genome sequence of the ammonia-oxidizing bacterium and obligate chemolithoautotroph Nitrosomonas europaea. Journal of bacteriology 185:2759-2773.

86. Loveless J, Painter H. 1968. The influence of metal ion concentrations and pH value on the growth of a Nitrosomonas strain isolated from activated sludge. Microbiology 52:1-14.

87. Li J, Nedwell DB, Beddow J, Dumbrell AJ, McKew BA, Thorpe EL, Whitby C. 2015. amoA gene abundances and nitrification potential rates suggest that benthic ammonia-oxidizing bacteria and not archaea dominate N cycling in the Colne Estuary, United Kingdom. Appl Environ Microbiol 81:159-165.

88. Hansen JI, Henriksen K, Blackburn TH. 1981. Seasonal distribution of nitrifying bacteria and rates of nitrification in coastal marine sediments. Microbial Ecology 7:297-304.

89. Wankel SD, Mosier AC, Hansel CM, Paytan A, Francis CA. 2011. Spatial Variability in Nitrification Rates and Ammonia-Oxidizing Microbial Communities in the Agriculturally Impacted Elkhorn Slough Estuary, California. Applied and Environmental Microbiology 77:269-280.

90. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. 2004. The architecture of complex weighted networks. Proceedings of the national academy of sciences 101:3747-3752.

91. Kanneworff E, Nicolaisen W. 1972. The “Haps” a frame-supported bottom corer. Ophelia 10:119-128.

92. Huntley SL, Wenning RJ, Su SH, Bonnevie NL, Paustenbach DJ. 1995. Geochronology and sedimentology of the lower Passaic River, New Jersey. Estuaries and Coasts 18:351-361.

93. Winkels H, Kroonenberg S, Lychagin MY, Marin G, Rusakov G, Kasimov N. 1998. Geochronology of priority pollutants in sedimentation zones of the Volga and Danube delta in comparison with the Rhine delta. Applied Geochemistry 13:581-591.

94. Valero-Garcés BL, Navas A, Machín J, Walling D. 1999. Sediment sources and siltation in mountain reservoirs: a case study from the Central Spanish Pyrenees. Geomorphology 28:23-41.

95. Harris D, Horwáth WR, Van Kessel C. 2001. Acid fumigation of soils to remove carbonates prior to total organic carbon or carbon-13 isotopic analysis. Soil Science Society of America Journal 65:1853-1856.

96. Cook RD. 1979. Influential observations in linear regression. Journal of the American Statistical Association 74:169-174.

97. RCoreTeam. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

98. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108:4516-4522.

99. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114-2120.

100. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537-7541.

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

102. Parada AE, Fuhrman JA. 2017. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. The ISME journal.

103. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research 13:2498-2504.

104. Assenov Y, Ramírez F, Schelhorn S-E, Lengauer T, Albrecht M. 2007. Computing topological parameters of biological networks. Bioinformatics 24:282-284.

105. Wasserman S, Faust K. 1994. Social network analysis: Methods and applications, vol 8. Cambridge university press.

106. Jørgensen BB. 1978. A comparison of methods for the quantification of bacterial sulfate reduction in coastal marine sediments. Geomicrobiology Journal 1:49-64.

107. Røy H, Weber HS, Tarpgaard IH, Ferdelman TG, Jørgensen BB. 2014. Determination of dissimilatory sulfate reduction rates in marine sediment via radioactive 35S tracer. Limnology and Oceanography: Methods 12:196-211.

Appendix I: Tables and Figures

Table 4.1. Geochemistry on bulk sedimentary organic matter. Data for replicate measures shown. Asterisks indicate data points that are statistical outliers.

Stn.core	Depth (cmbsf)	TOC (wt%)	$\delta^{13}C_{org}$ (‰ vs. VPDB)	C/N
AB.2	1-2	1.7	-29.8*	12.0
AB.2	3-4	1.6	-24.7	12.7
AB.2	4-5	1.6	-25.3	12.5
AB.2	5-6	1.9*	-24.2	9.2*
AB.2	6-7	1.6	-24.3	13.7
AB.2	7-8	1.5	-24.9	12.6
AB.2	8-9	1.4	-27.0	12.3
AB.2	9-10	1.6	-27.0	13.5
AB.2	11-12	1.5	-25.0	12.7
AB.2	11-12	1.5	-25.0	13.0
AB.2	12-13	1.5	-25.0	12.7
AB.2	13-14	1.5	-24.8	13.2
AB.2	14-15	1.4	-25.5	12.5
AB.2	16-17	1.5	-24.8	13.6
AB.2	17-18	1.4	-25.6	12.9
AB.2	18-19	1.4	-25.9	13.0
AB.2	19-20	1.5	-25.4	13.5
AC.2	0-1	1.4	-26.3*	13.4
AC.2	1-2	1.5	-25.6	11.5*
AC.2	2-3	1.4	-25.6	13.1
AC.2	3-4	1.5	-26.0	13.8
AC.2	4-5	1.4	-25.7	12.2
AC.2	5-6	1.4	-25.9	12.5
AC.2	6-7	1.5	-26.0	12.7
AC.2	7-8	1.6	-25.8	13.4
AC.2	8-9	1.4	-25.9	13.4
AC.2	9-10	1.5	-25.9	12.8
AC.2	10-11	1.5	-26.4	13.9
AC.2	11-12	1.4	-26.0	13.2
AC.2	12-13	1.3	-25.9	14.2
AC.2	12-13	1.4	-25.7	14.0
AC.2	13-14	1.4	-26.0	13.0
AC.2	15-16	1.3	-25.9	13.7
AC.2	16-17	1.5	-26.2	13.7
AC.2	17-18	1.4	-27.1	13.6
AC.2	19-20	1.4	-25.9	13.8
AC.2	19-20	1.3	-26.3*	13.3

Table 4.1 continued.

Stn.core	Depth (cmbsf)	TOC (wt%)	$\delta^{13}C_{org}$ (‰ vs. VPDB)	C/N
HA.2	0-1	1.4	-26.2	13.2
HA.2	1-2	1.4	-26.2	13.7
HA.2	2-3	1.3	-26.2	13.6
HA.2	3-4	1.4	-26.3	14.2
HA.2	3-4	1.3	-26.2	13.4
HA.2	4-5	1.2	-26.4	13.1
HA.2	5-6	1.3	-26.2	14.1
HA.2	7-8	1.4	-26.4	14.1
HA.2	8-9	1.4	-25.8	13.5
HA.2	8-9	1.3	-26.6	12.7
HA.2	9-10	1.3	-26.2	13.9
HA.2	10-11	1.4	-26.1	13.5
HA.2	11-12	1.5	-25.7	12.9
HA.2	12-13	1.5	-27.0*	13.3
HA.2	13-14	1.4	-25.7	13.3
HA.2	14-15	1.5	-26.0	14.5*
HA.2	15-16	1.3	-26.5	12.6
HA.2	16-17	1.4	-26.0	12.7
HA.2	17-18	1.4	-26.1	13.2
HA.2	18-19	1.4	-26.0	13.2
HA.2	22-23	1.5	-26.2	13.3

Table 4.2. Results of qPCR quantification of two cores taken at stations AB and AC. Values are reported in copy numbers g^{-1} sediment for each station (Stn) and depth (in cmbsf). Average copy numbers are taken from technical replicates Rep A and Rep B. Depths for which values are missing are attributed to either a missing sample (*) or results being below the quantification limit (+). ND = not determined.

Stn. core	Depth	Bacteria Rep A	Bacteria Rep B	Average Bacteria	Archaea Rep A	Archaea Rep B	Average Archaea
AB.1	0-1	1.66E+09	2.69E+08	9.65E+08	2.29E+07	8.05E+07	5.17E+07
AB.1	1-2	4.52E+09	4.26E+09	4.39E+09	1.31E+08	+	1.31E+08
AB.1	2-3	7.01E+09	1.18E+10	9.41E+09	2.45E+08	1.46E+08	1.96E+08
AB.1	3-4	2.21E+10	4.13E+10	3.17E+10	5.75E+07	2.55E+08	1.56E+08
AB.1	4-5	8.27E+10	5.62E+10	6.95E+10	3.62E+08	4.22E+08	3.92E+08
AB.1	5-6	+	2.90E+10	2.90E+10	7.91E+07	1.79E+08	1.29E+08
AB.1	6-7	1.55E+10	2.01E+10	1.55E+10	1.70E+08	5.56E+07	1.13E+08
AB.1	7-8	3.38E+10	+	3.38E+10	1.02E+07	7.98E+06	9.07E+06
AB.1	8-9	+	2.45E+10	2.45E+10	3.07E+06	3.82E+06	3.45E+06
AB.1	9-10	3.28E+10	2.81E+10	3.05E+10	1.07E+07	1.94E+07	1.51E+07
AB.1	10-11	1.85E+10	1.29E+10	1.57E+10	6.06E+06	8.80E+06	7.43E+06
AB.1	11-12	1.04E+10	9.84E+09	1.01E+10	8.06E+06	8.19E+06	8.13E+06
AB.1	12-13	2.12E+10	6.91E+09	1.40E+10	1.43E+07	1.29E+07	1.36E+07
AB.1	13-14	2.06E+10	1.88E+10	1.97E+10	5.06E+05	5.90E+07	2.98E+07
AB.1	14-15	1.28E+10	1.35E+10	1.32E+10	3.48E+06	2.87E+06	3.18E+06
AB.2	0-1	1.39E+11	1.28E+11	1.33E+11	2.74E+07	3.38E+07	3.06E+07
AB.2	1-2	6.64E+10	+	6.64E+10	1.77E+07	2.58E+07	2.17E+07
AB.2	2-3	+	6.50E+09	6.50E+09	2.40E+06	2.26E+06	2.33E+06
AB.2	3-4	2.49E+09	4.86E+09	3.67E+09	1.08E+06	8.34E+05	9.55E+05
AB.2	4-5	5.24E+09	3.48E+10	2.00E+10	4.51E+06	3.91E+06	4.21E+06
AB.2	5-6	1.58E+11	1.34E+10	8.56E+10	2.19E+06	ND	2.19E+06
AB.2	6-7	4.70E+09	1.67E+09	3.18E+09	1.87E+05	1.34E+05	1.61E+05
AB.2	7-8	1.11E+09	1.66E+09	1.38E+09	3.26E+06	3.10E+05	1.79E+06
AB.2	8-9	1.48E+08	4.97E+08	3.22E+08	+	+	+
AB.2	9-10	1.53E+09	1.06E+09	1.30E+09	3.80E+05	+	3.80E+05
AB.2	10-11	5.68E+08	1.68E+09	1.12E+09	4.46E+05	ND	4.46E+05
AB.2	11-12	*	*	*	*	*	*
AB.2	12-13	3.74E+09	9.29E+08	2.33E+09	6.07E+04	2.69E+05	1.65E+05
AB.2	13-14	1.46E+09	2.98E+09	2.22E+09	1.07E+06	ND	1.07E+06
AB.2	14-15	2.24E+08	3.63E+08	2.93E+08	3.94E+05	ND	3.94E+05
AB.2	15-16	*	*	*	*	*	*
AB.2	16-17	3.07E+08	9.26E+07	2.00E+08	+	+	+

Table 4.2 continued.

Stn. core	Depth	Bacteria Rep A	Bacteria Rep B	Average Bacteria	Archaea Rep A	Archaea Rep B	Average Archaea
AB.2	17-18	1.27E+09	1.04E+09	1.15E+09	4.26E+06	1.14E+06	2.70E+06
AB.2	18-19	5.48E+07	1.55E+08	1.05E+08	7.42E+04	ND	7.42E+04
AC.1	0-1	9.68E+07	1.55E+08	1.26E+08	2.21E+06	8.33E+09	4.17E+09
AC.1	1-2	7.37E+09	4.48E+07	3.71E+09	6.25E+06	3.90E+06	5.07E+06
AC.1	2-3	2.15E+09	+	2.15E+09	2.73E+07	1.10E+07	1.92E+07
AC.1	3-4	2.85E+09	2.79E+09	2.82E+09	5.06E+06	4.64E+06	4.85E+06
AC.1	4-5	+	+	+	+	+	+
AC.1	5-6	3.71E+08	3.94E+08	3.83E+08	4.10E+05	4.56E+05	4.33E+05
AC.1	6-7	7.02E+08	7.26E+08	7.14E+08	6.01E+06	1.86E+06	3.94E+06
AC.1	7-8	2.84E+09	3.45E+09	3.14E+09	1.49E+07	2.83E+07	2.16E+07
AC.1	8-9	1.17E+10	1.03E+10	1.10E+10	2.82E+07	2.51E+07	2.66E+07
AC.1	9-10	+	+	+	+	+	+
AC.1	10-11	3.42E+07	3.38E+07	3.40E+07	4.41E+05	2.45E+05	3.43E+05
AC.1	11-12	3.42E+07	5.29E+07	4.36E+07	1.23E+05	1.85E+05	1.54E+05
AC.1	12-13	1.76E+09	2.77E+09	2.26E+09	3.83E+06	3.30E+06	3.56E+06
AC.1	13-14	7.33E+07	4.78E+07	6.05E+07	7.78E+04	2.10E+05	1.44E+05
AC.1	14-15	5.64E+07	6.18E+07	5.91E+07	5.11E+05	6.86E+05	5.99E+05
AC.1	15-16	1.69E+09	+	1.69E+09	3.23E+06	5.41E+06	4.32E+06
AC.1	16-17	+	+	+	+	+	+
AC.1	17-18	7.77E+06	7.92E+06	7.85E+06	2.28E+06	2.05E+06	2.17E+06
AC.1	18-19	5.69E+06	5.72E+06	5.70E+06	3.36E+06	1.10E+06	2.23E+06
AC.1	19-20	3.73E+09	3.60E+09	3.67E+09	1.32E+06	1.24E+06	1.28E+06
AC.2	0-1	4.05E+06	3.31E+06	4.05E+06	3.67E+04	+	3.67E+04
AC.2	1-2	1.00E+11	8.09E+10	1.00E+11	3.79E+07	2.64E+07	3.21E+07
AC.2	2-3	6.46E+10	4.78E+10	6.46E+10	1.32E+07	1.23E+07	1.27E+07
AC.2	3-4	1.04E+09	1.41E+09	1.04E+09	1.60E+06	2.12E+06	1.86E+06
AC.2	4-5	5.35E+08	2.38E+09	5.35E+08	1.60E+06	2.12E+06	5.87E+06
AC.2	5-6	+	+	+	+	+	+
AC.2	6-7	1.29E+10	8.93E+09	1.09E+10	4.34E+06	1.66E+11	8.32E+10
AC.2	7-8	5.82E+09	6.40E+09	6.11E+09	2.51E+07	4.93E+07	3.72E+07
AC.2	8-9	4.76E+09	4.49E+09	4.62E+09	6.28E+06	4.12E+06	5.20E+06
AC.2	9-10	7.66E+07	1.21E+07	4.43E+07	1.51E+04	2.67E+04	2.09E+04
AC.2	10-11	2.51E+07	2.17E+07	2.34E+07	1.02E+06	1.75E+05	5.99E+05
AC.2	11-12	4.41E+09	+	4.41E+09	1.05E+07	4.71E+06	7.62E+06
AC.2	12-13	4.88E+08	2.68E+09	1.58E+09	2.67E+06	2.86E+06	2.76E+06
AC.2	13-14	7.19E+09	8.95E+09	7.19E+09	3.16E+07	5.90E+07	4.53E+07
AC.2	14-15	*	*	*	*	*	*
AC.2	15-16	4.26E+06	2.07E+06	3.17E+06	7.52E+04	+	7.52E+04
AC.2	16-17	+	+	+	+	+	+
AC.2	17-18	6.70E+09	5.08E+09	5.89E+09	7.71E+07	2.89E+07	5.30E+07

Table 4.2 continued.

Stn. core	Depth	Bacteria Rep A	Bacteria Rep B	Average Bacteria	Archaea Rep A	Archaea Rep B	Average Archaea
AC.2	18-19	*	*	*	*	*	*
AC.2	19-20	3.95E+08	3.58E+08	3.76E+08	ND	ND	ND

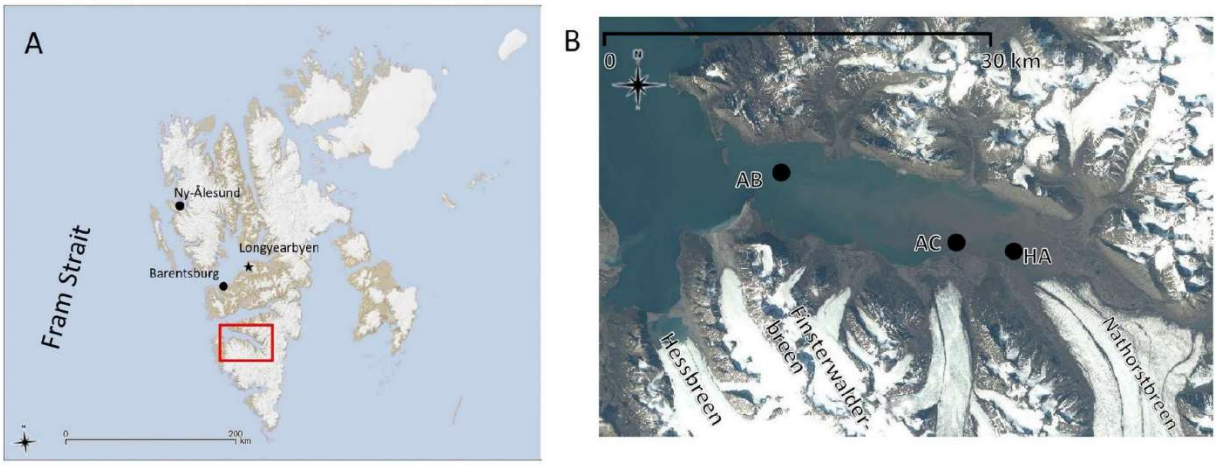

Figure 4.1. Map of field site in Svalbard. Locations of towns are noted for reference. Van Keulenfjorden is enclosed in red box. Scale bar is 200 km (A). Locations of stations within the fjord, with station AC in the middle of the fjord, and AB at the outer mouth. Scale bar is 20 km (B).

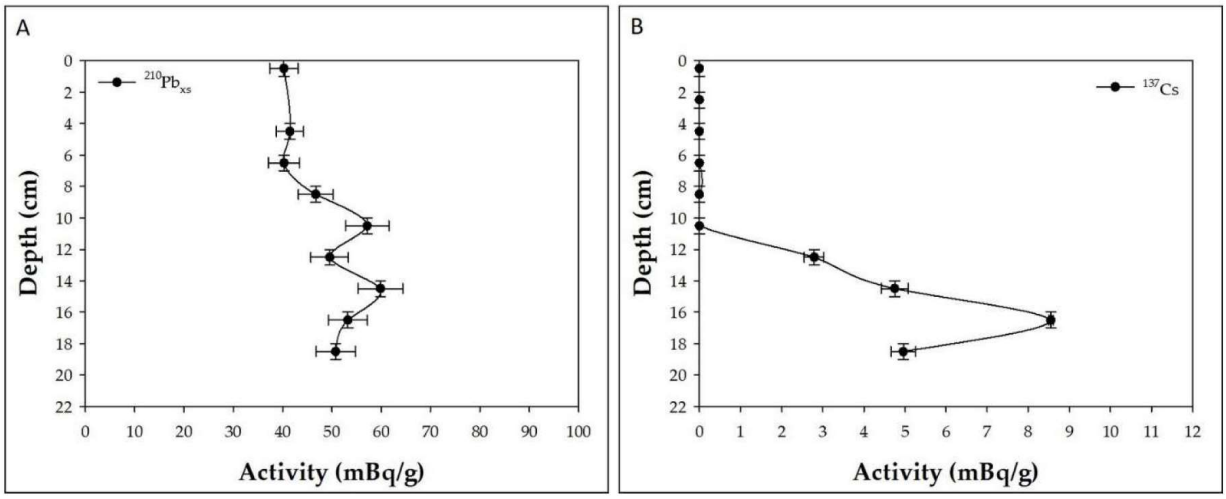

Figure 4.2. Results of age dating for site AC using 210Pb (A) and 137Cs (B). The sediment profile was too mixed for reliable 210Pb ages.

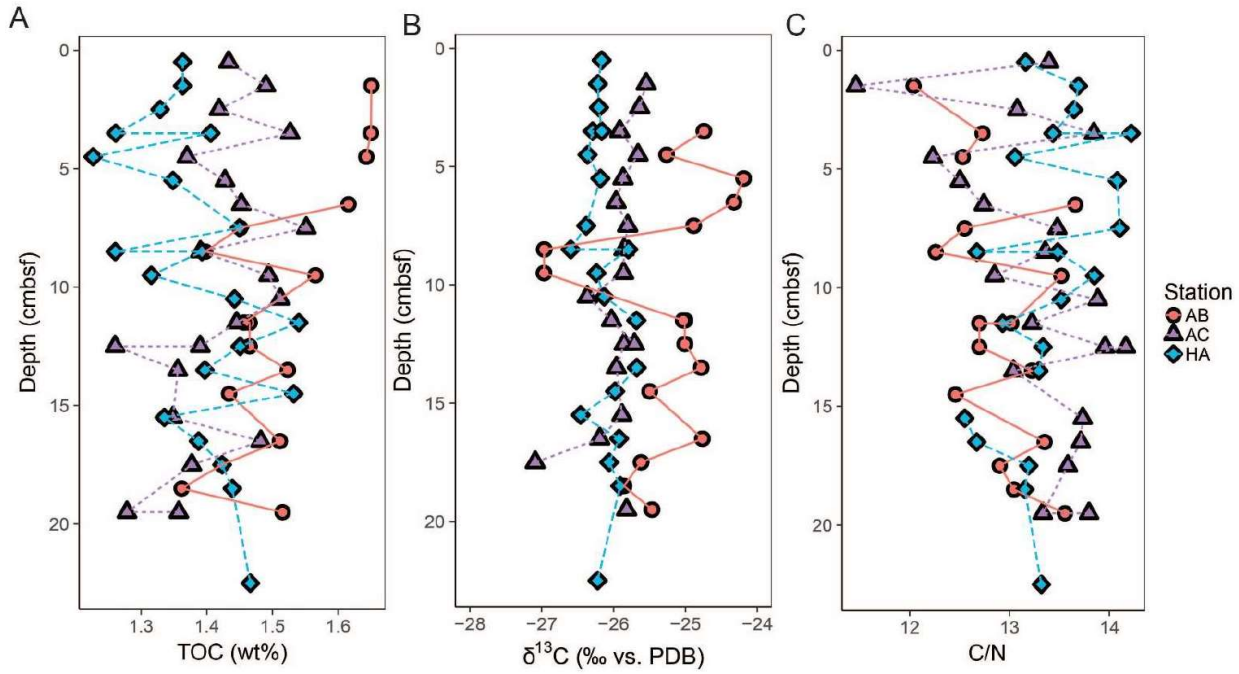

Figure 4.3. Organic geochemistry data. Downcore profiles of total organic carbon (TOC) (A), $\delta^{13}C_{org}$ (B), and carbon to nitrogen ratios (C/N) (C) for sites AB (salmon circles) and AC (purple triangles). All values are reported in Table 4.2.

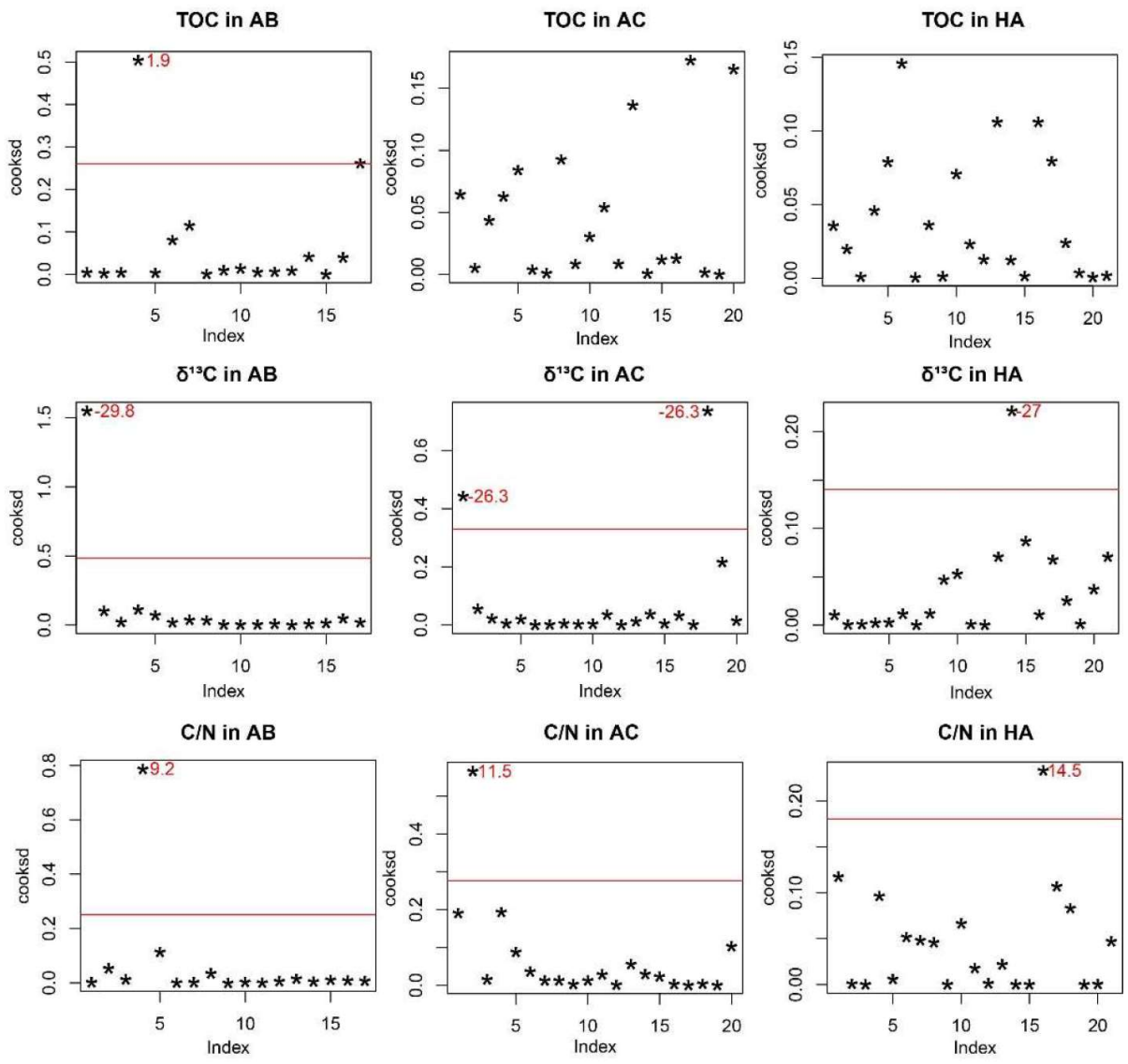

Figure 4.4. Outliers determined with Cook's distance measure for TOC (top row), carbon isotopes (middle row) and C/N ratio (bottom row) for the outer site AB (left column), middle site AC (middle column), and inner site HA (right row).

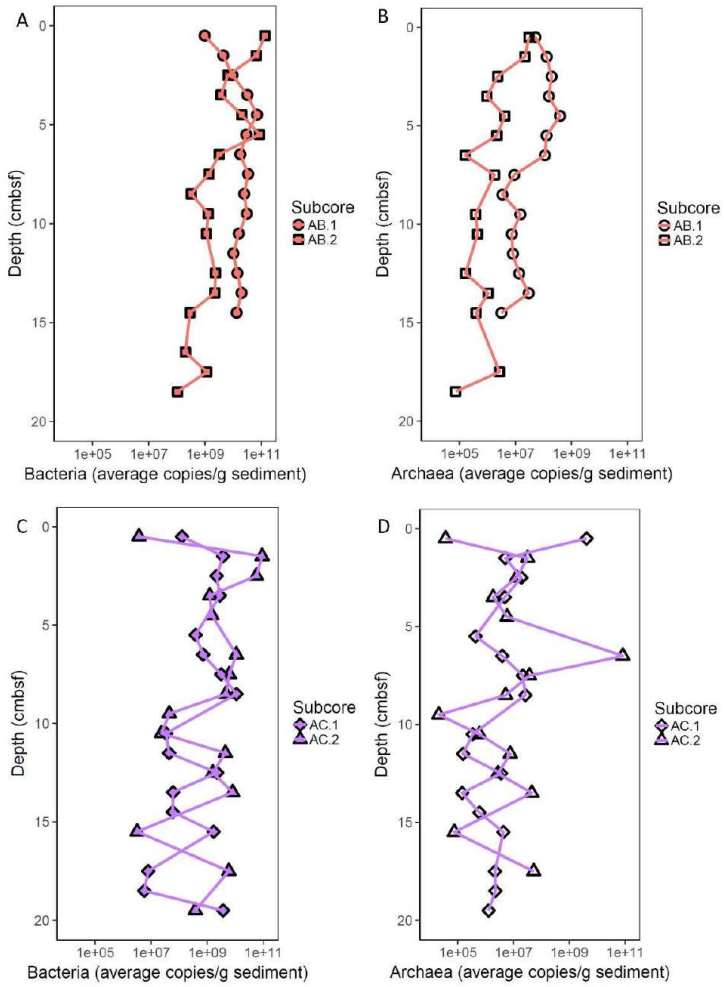

Figure 4.5. qPCR data. Downcore results of abundance of the 16S rRNA gene for bacteria (A, C) and archaea (B, D) reported in copies g fresh sediment^{-1}. Average values between technical duplicates are shown for replicate cores AB.1 and AB.2 (salmon) and AC.1 and AC.2 (purple). All values are reported in Table 4.2.

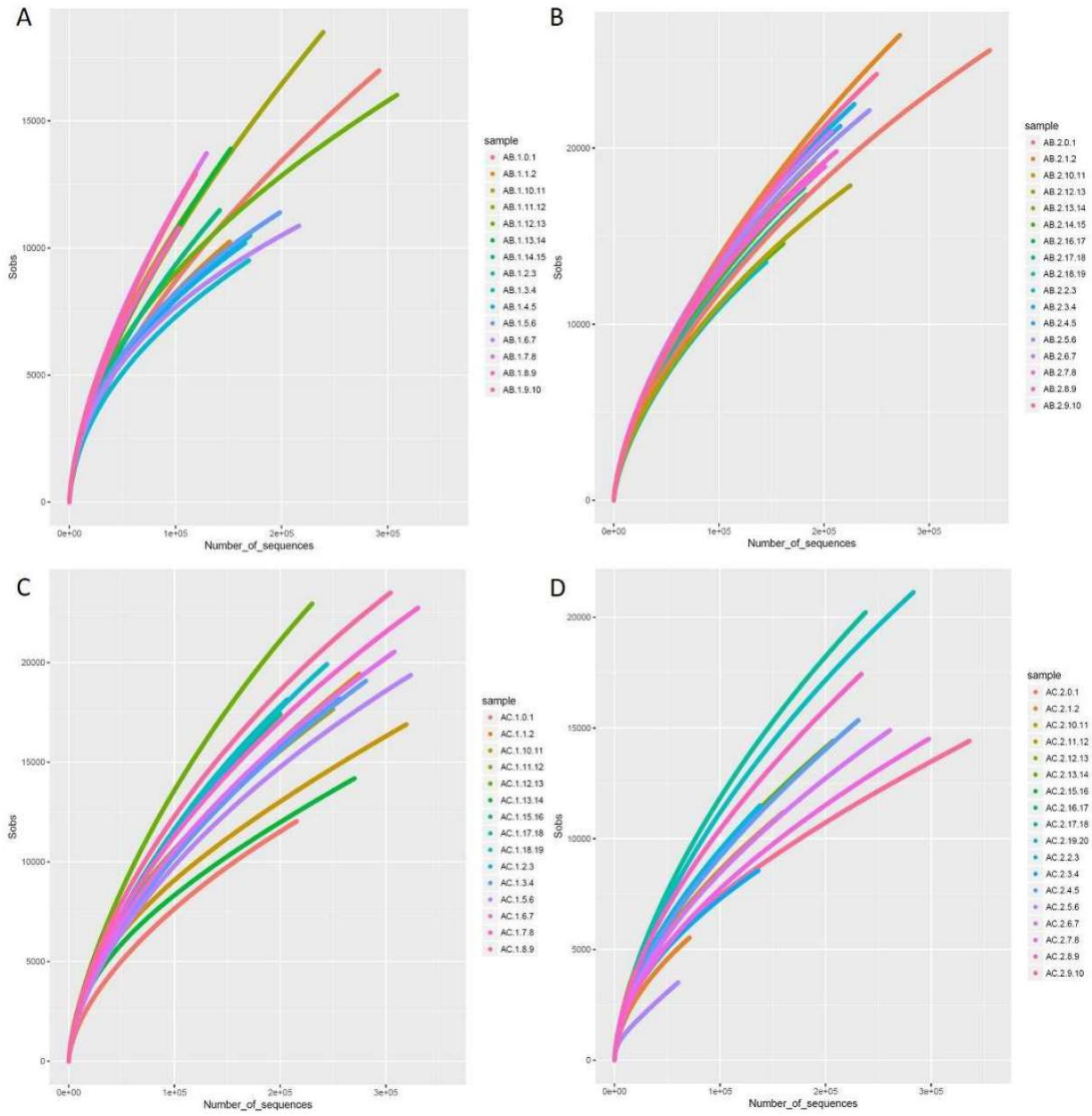

Figure 4.6. Measurement of alpha diversity of 16S rRNA gene amplicon libraries for core AB.1 (A), core AB.2 (B), core AC.1 (C), core AC.2 (D). Number of observed OTUs is plotted on the y-axis and number of sequences is on the x-axis.

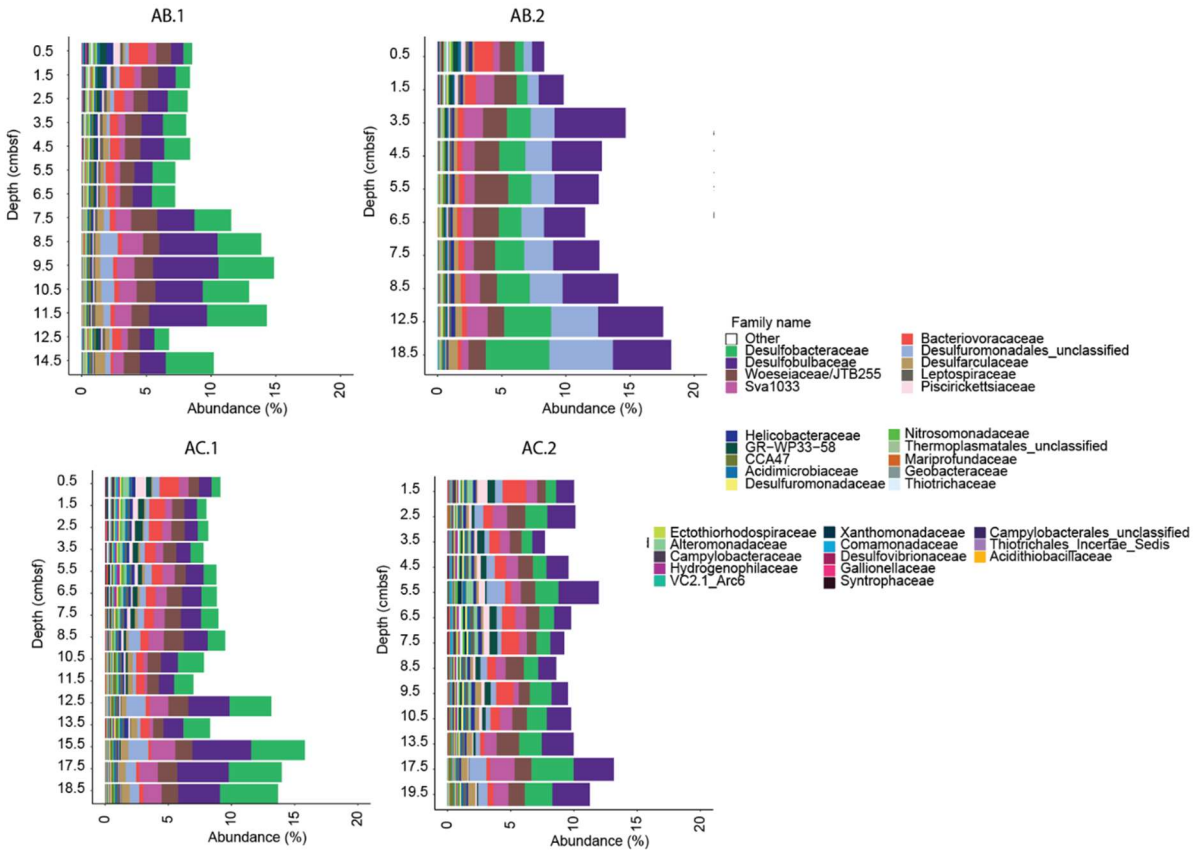

Figure 4.7. Community composition of iron and sulfur families.16S rRNA gene libraries for outer station AB (A and C) and middle station AC (C and D) are reported along a depth axis downward for two cores at each site. Only families with summed abundance greater than 1% are shown. Colors of families are shared between stations and are in order of relative abundance for each station. The x-axis scale of abundance does not reach 100% because reads are dominated by families other than those with iron and sulfur cycling representatives.

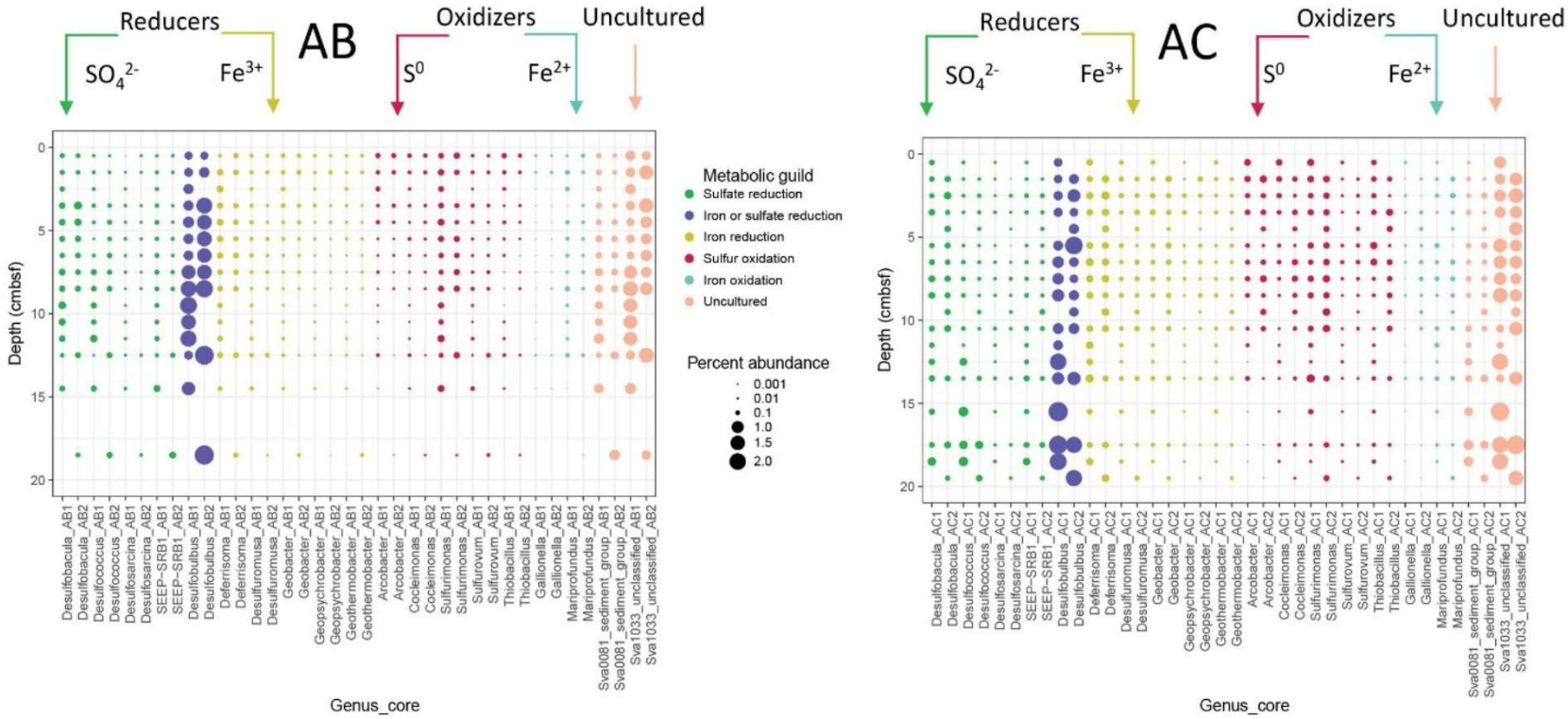

Figure 4.8. Relative abundances of sequences of iron and sulfur taxa in site AB (A) and AC (B). Sequences are sorted by metabolic guild: Sulfate reducers (green), sulfate/iron reducers (purple), iron reducers (yellow), sulfur oxidizers (pink), and iron oxidizers (teal), and uncultured (peach). The number next to the genus name on the x-axis indicates which core the sequences are from (e.g. AB.1 or AB.2, AC.1 or AC.2).

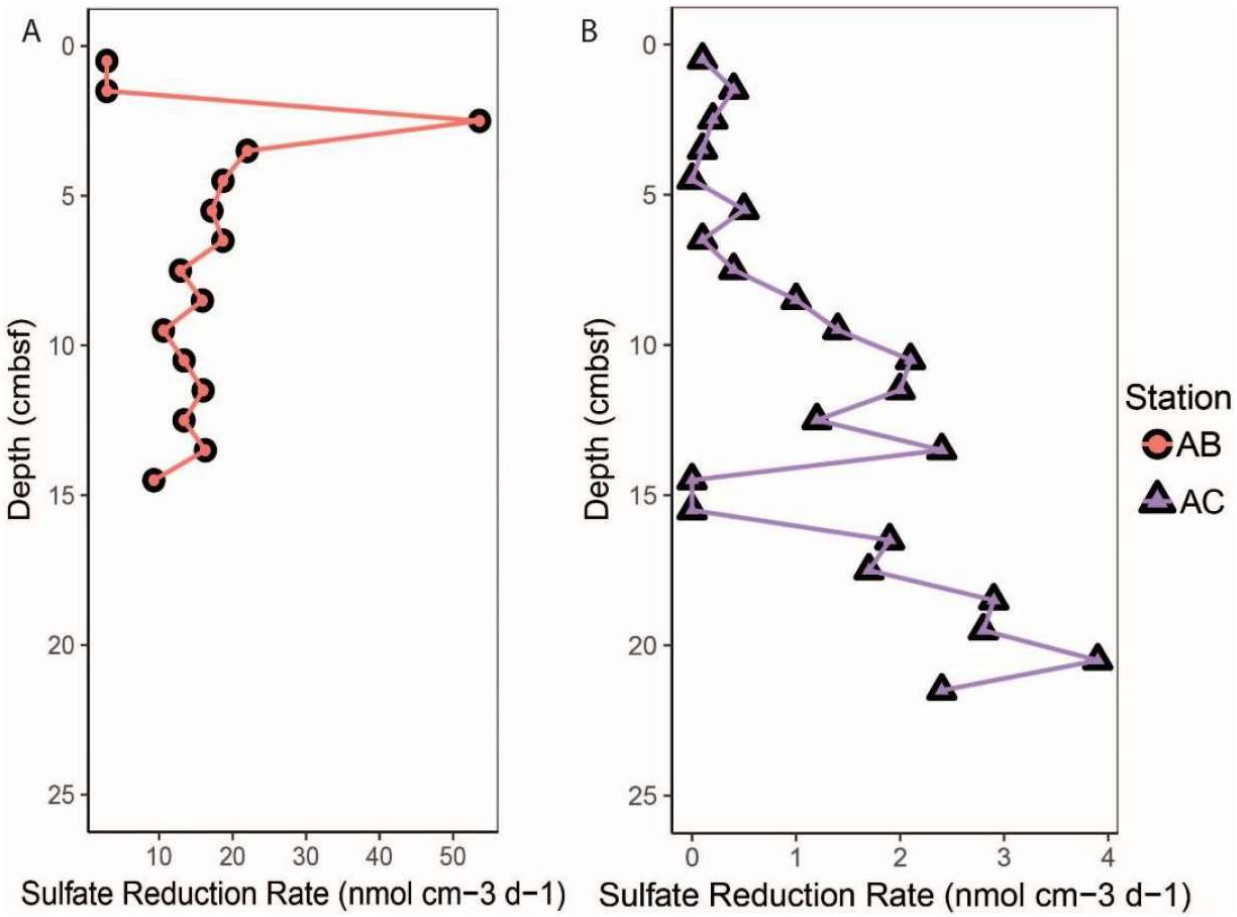

Figure 4.9. Sulfate reduction rates in sites AB (A) and AC (B). Note x- and y- axes are not shared between panels.

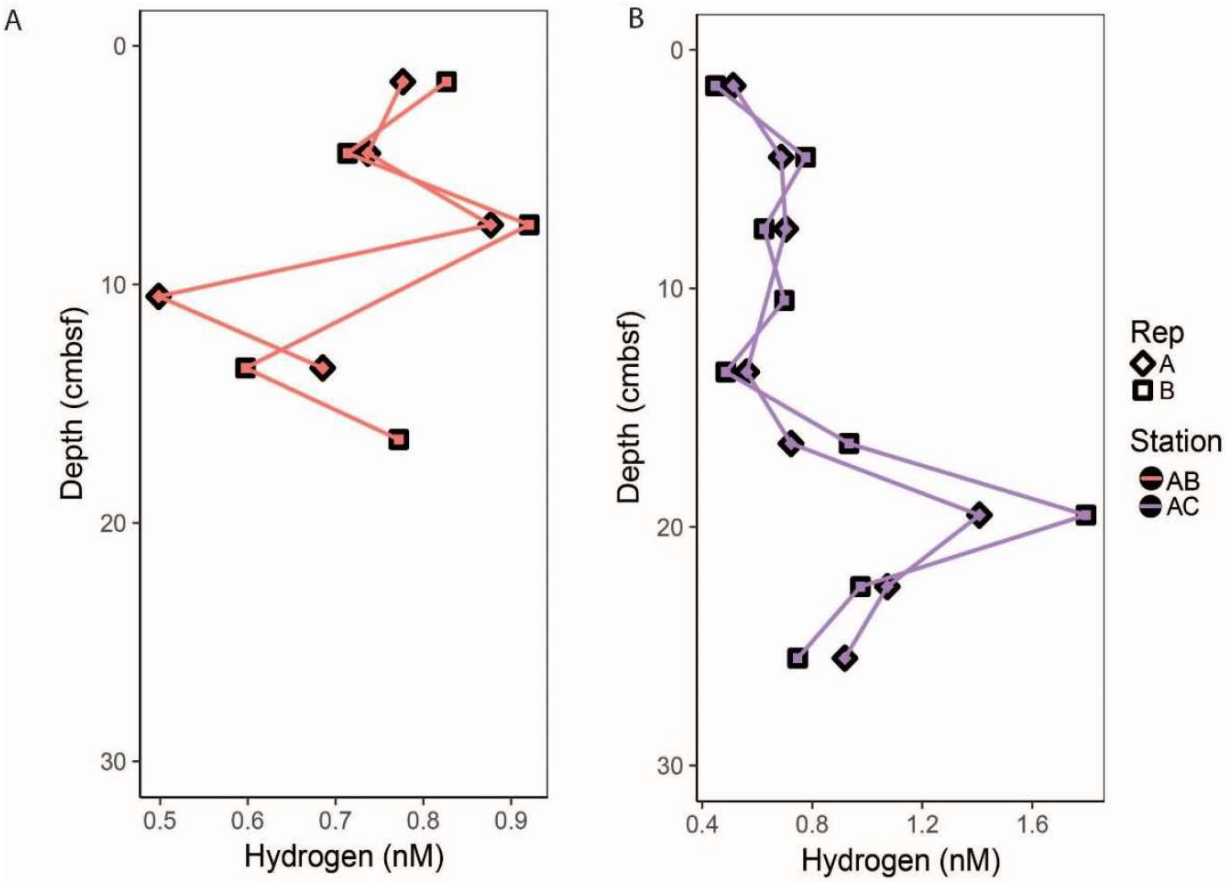

Figure 4.10. Hydrogen data for sites AB (A) and AC (B). Note x- and y- axes are not shared between panels.

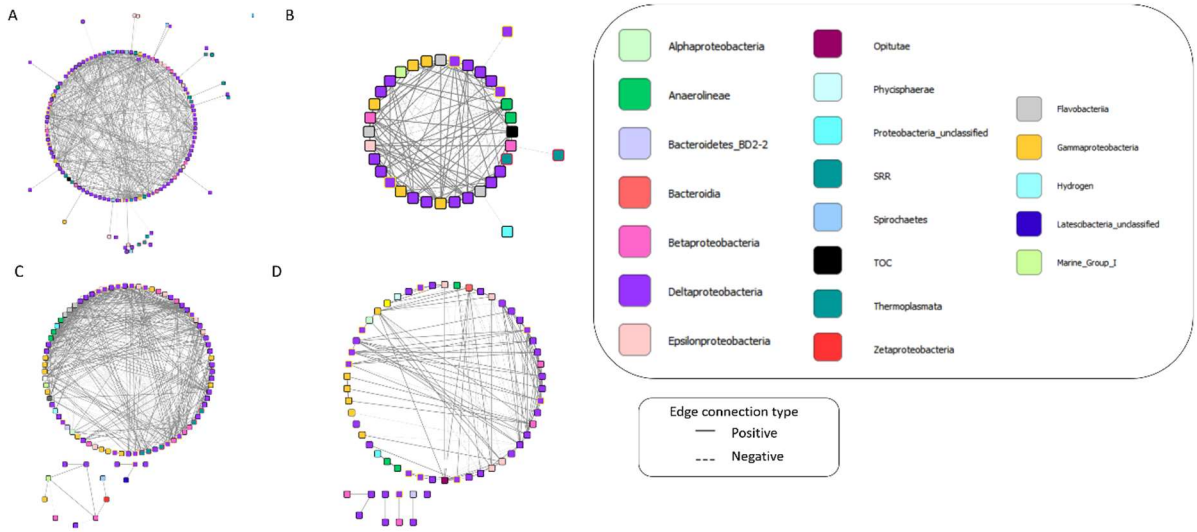

Figure 4.11. Individual microbial interaction networks for cores AB.1 (A), AB.2 (B), AC.1 (C), and AC.2 (D). Node color indicates taxonomy.

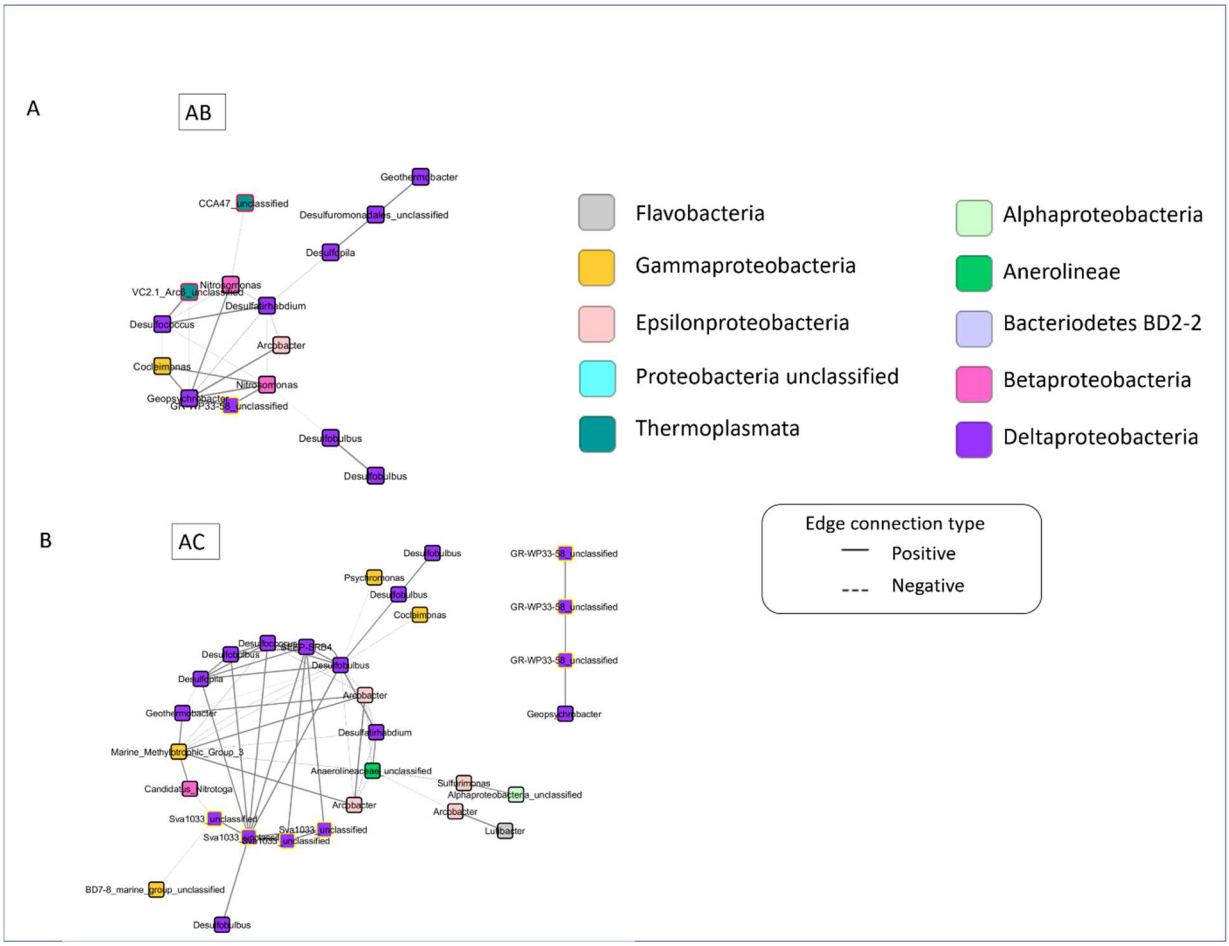

Figure 4.12. Merged microbial co-occurrence networks. Individual network characteristics have been combined to show merged networks for outer station AB (A) and middle station AC (B) to get at the core microbiome features at each site. Isolated nodes have been removed for clarity. Node color indicates taxonomy at the class level and edge relationships are indicated with solid and dashed lines for positive and negative connections, respectively.

Appendix II: R Code

R scripts for qPCR figures

```
######## downcore plots for qPCR figure ################

qPCR_melted<-read.csv("qPCR_geochem.csv", header=TRUE, stringsAsFactors = FALSE)

AB_Bac<-subset(AB_vals, Domain %in% c("Bacteria"))

AB_Arc<-subset(AB_vals, Domain %in% c("Archaea"))

VK_Bac<-subset(qPCR_melted, Domain %in% c("Bacteria"))

VK_Arc<-subset(qPCR_melted, Domain %in% c("Archaea"))

library(scales)

library(ggplot2)

library(plyr)

show_col(hue_pal()(9))

cols<-c("AB.2" = "#F8766D","AB.1" = "#F8766D", "AC.1" = "#C77CFF", "AC.2" = "#C77CFF", "AB" =
"#F8766D", "AC" = "#C77CFF", "HA" = "#00B9E3")

AB_Bacteria<-ggplot(AB_Bac[!is.na(AB_Bac$Average),], aes(x=Depth, y=Average, color=Subcore,
Shape=Subcore)) +

  geom_point(aes(fill=Subcore, shape=Subcore),colour="black", size=4, stroke=2) +

  scale_shape_manual(values = c(21,22)) +

  scale_fill_manual(values=cols) +

  scale_colour_manual(values=cols) +

  theme_bw(base_size = 20) +

  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),

        panel.background = element_blank(), axis.line = element_line(colour = "black")) +

  scale_x_reverse(limits=c(20,0)) +

  geom_line(aes(color=Subcore)) +

  scale_y_log10(limits = c(1e4,1.5e11)) +

  geom_line(size=1.5) +

  labs(x="Depth (cmbsf)", y= "Bacteria (average copies/g sediment)") +

  coord_flip()


AC_Bacteria<-ggplot(AC_Bac[!is.na(AC_Bac$Average),], aes(x=Depth, y=Average, color=Subcore)) +

  geom_point(aes(fill=Subcore, shape=Subcore), colour="black", size=4, stroke=2) +

  scale_colour_manual(values=cols) +

  scale_shape_manual(values = c(23,24)) +

  scale_fill_manual(values=cols) +

  theme_bw(base_size = 20) +

  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),

        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
```

```
  scale_x_reverse(limits=c(20,0)) +
  geom_line(aes(color=Subcore)) +
  scale_y_log10(limits = c(1e4,1.5e11)) +
  geom_line(size=1.5) +
  labs(x="Depth (cmbsf)", y= "Bacteria (average copies/g sediment)") +
  coord_flip()
AB_Archaea<-ggplot(AB_Arc[!is.na(AB_Arc$Average),], aes(x=Depth, y=Average, color=Subcore)) +
  geom_point(aes(shape=Subcore), colour="black", size=4, stroke=2) +
  scale_colour_manual(values=cols) +
  scale_shape_manual(values = c(21,22)) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_x_reverse(limits=c(20,0)) +
  geom_line(aes(color=Subcore)) +
  scale_y_log10(limits = c(1e4,1.5e11)) +
  geom_line(size=1.5) +
  labs(x="Depth (cmbsf)", y= "Archaea (average copies/g sediment)") +
  coord_flip()
AC_Archaea<-ggplot(AC_Arc[!is.na(AC_Arc$Average),], aes(x=Depth, y=Average, color=Subcore)) +
  geom_point(aes(shape=Subcore), colour="black", size=4, stroke=2) +
  scale_fill_manual(values=cols) +
  scale_shape_manual(values = c(23,24)) +
  scale_colour_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_x_reverse(limits=c(20,0)) +
  geom_line(aes(color=Subcore)) +
  scale_y_log10(limits = c(1e4,1.5e11)) +
  geom_line(size=1.5) +
  labs(x="Depth (cmbsf)", y= "Archaea (average copies/g sediment)") +
  coord_flip()
```

```
######## Cook's distance test on organic chemistry data #################

org_chem<-read.csv("TOC.csv")

# Isolate geochem from each station:

VK<-org_chem[27:84,]

VKAB<-VK[which(VK$Station == "AB"), names(VK) %in% c("Fjord", "Station", "Depth", "TOC", "CtoN",
"d13C", "d15N")]

VKAC<-org_chem[44:63,]

VKHA<-VK[which(VK$Station == "HA"), names(VK) %in% c("Fjord", "Station", "Depth", "TOC", "CtoN",
"d13C", "d15N")]

# Cook's distance to determine outliers, change data set each time to create plots

mod<-lm(CtoN ~ Depth , data=VKHA)

cooksd<-cooks.distance(mod)

plot(cooksd, pch="*", cex=2, main="Influential Obs by Cooks distance, CtoN in HA")  # plot cook's
distance

abline(h = 4*mean(cooksd, na.rm=T), col="red")  # add cutoff line

text(x=1:length(cooksd)+1, y=cooksd, labels=ifelse(cooksd>4*mean(cooksd,
na.rm=T),names(cooksd),""), col="red")  # add labels)

influential <- as.numeric(names(cooksd)[(cooksd > 4*mean(cooksd, na.rm=T))])  # influential row
numbers from original data frame

#t tests for data without outliers:

org_chem_no_out<-read.csv("TOC_outliers_removed.csv")

VK_no_out<-org_chem_no_out[27:84,]

VKAB_no_out<-VK_no_out[which(VK_no_out$Station == "AB"), names(VK_no_out) %in% c("Fjord",
"Station", "Depth", "TOC", "CtoN", "d13C", "d15N")]

VKAC_no_out<-org_chem_no_out[44:63,]

VKHA_no_out<-VK_no_out[which(VK_no_out$Station == "HA"), names(VK_no_out) %in% c("Fjord",
"Station", "Depth", "TOC", "CtoN", "d13C", "d15N")]

shapiro.test(VKAB_no_out$TOC)

shapiro.test(VKAC_no_out$TOC)

shapiro.test(VKHA_no_out$TOC)


t.test(VKAB_no_out$TOC,VKAC_no_out$TOC) #Welsh two sample t test -  0.004435

t.test(VKAB_no_out$TOC, VKHA_no_out$TOC) #Welsh two sample t test - 0.0002843

t.test(VKHA_no_out$TOC, VKAC_no_out$TOC) #Welsh two sample t test - 0.228


t.test(VKAB_no_out$d13C,VKAC_no_out$d13C) #Welsh two sample t test - 0.005297

t.test(VKAB_no_out$d13C, VKHA_no_out$d13C) #Welsh two sample t test - 0.0006274

t.test(VKHA_no_out$d13C, VKAC_no_out$d13C) #Welsh two sample t test - 0.06565
```

```
t.test(VKAB_no_out$CtoN,VKAC_no_out$CtoN) #Welsh two sample t test - 0.05487
t.test(VKAB_no_out$CtoN, VKHA_no_out$CtoN) #Welsh two sample t test - 0.005338
t.test(VKHA_no_out$CtoN, VKAC_no_out$CtoN) #Welsh two sample t test - 0.558

# downcore plots for figure, outliers removed:
library(ggplot2)
VK_cton_plot<-ggplot(VK_no_out, aes(x=Depth, y=CtoN, shape=Station, color=Station)) +
  geom_point(aes(shape=Station, fill=Station), colour="black", size=4, stroke=2) +
  scale_shape_manual(values = c(21, 24, 23)) +
  scale_fill_manual(values=cols) +
  scale_colour_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_x_reverse() +
  geom_line(aes(color=Station, linetype=Station), size=0.75) +
  labs(x="Depth (cmbsf)", y="CtoN") +
  coord_flip()

VK_d13C_plot<-ggplot(VK_no_out, aes(x=Depth, y=d13C, shape=Station, color=Station)) +
  geom_point(aes(shape=Station, fill=Station), colour="black", size=4, stroke=2) +
  scale_shape_manual(values = c(21, 24, 23)) +
  scale_fill_manual(values=cols) +
  scale_colour_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_y_continuous(limits=c(-28, -24)) +
  scale_x_reverse() +
  geom_line(aes(color=Station, linetype=Station), size=0.75) +
  labs(x="Depth (cmbsf)", y=expression(paste(delta^{13}, "C (\u2030 vs. PDB)"))) +
  coord_flip()

VK_TOC<-ggplot(VK_no_out, aes(x=Depth, y=TOC, shape=Station, color=Station)) +
```

```
geom_point(aes(shape=Station, fill=Station), colour="black", size=4, stroke=2) +
scale_shape_manual(values = c(21, 24, 23)) +
scale_x_reverse() +
scale_fill_manual(values=cols) +
scale_colour_manual(values=cols) +
theme_bw(base_size = 20) +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
      panel.background = element_blank(), axis.line = element_line(colour = "black")) +
geom_line(aes(color=Station, linetype=Station), size=0.75) +
labs(x="Depth (cmbsf)", y= "TOC (wt%)") +
coord_flip()
```

R script for relative abundance plots of 16S rRNA gene amplicon data

```
######## Relative abundance bubble plots ##########

library(tidyr)

library(dplyr)

library(plyr)

library(ggplot2)


norm_AB<-read.csv("AB_norm.csv")

head(norm_AB)

norm_AB_tidy<-gather(norm_AB, depth, abundance, X0.5:X18.5)

head(norm_AB_tidy)

write.csv(norm_AB_tidy, "norm_AB_tidy.csv")


bub_cols<-c("1"="#01b64e", "2"= "#6346f2", "3"="#cbc600", "4"="#cb0049", "5"="#3aedc7",
"6"="#ffb89c")


############### AB bubble plot #############

norm_AB_tidy_fixed<-read.csv("norm_AB_tidy_fixed.csv", stringsAsFactors = FALSE)

head(norm_AB_tidy_fixed)

norm_AB_tidy_fixed$Genus <- reorder(norm_AB_tidy_fixed$Genus,norm_AB_tidy_fixed$guild)

norm_AB_tidy_fixed$Percent_Abundance<-
as.numeric(as.character(norm_AB_tidy_fixed$Percent_Abundance))

norm_AB_tidy_fixed$guild <- as.factor(norm_AB_tidy_fixed$guild)



ggplot(norm_AB_tidy_fixed, aes(x=depth, y=Genus, color=guild, size=Percent_Abundance)) +

  geom_point(aes(fill=guild)) +

  scale_color_manual(values=bub_cols) +

  scale_size(range = c(0,10), breaks = c(0.001, 0.01, 0.1, 1, 1.5, 2)) +

  coord_flip() +

  theme_bw(base_size = 20) +

  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +

  labs(x="Depth (cmbsf)", y= "Genus") +

  scale_x_reverse(limits= c(20,0))



##############################################
```

```
norm_AC<-read.csv("AC_norm.csv")

head(norm_AC)

norm_AC_tidy<-gather(norm_AC, depth, abundance, X0.5:X19.5)

head(norm_AC_tidy)

write.csv(norm_AC_tidy, "norm_AC_tidy.csv")


####################AC Bubble plot ############

norm_AC_tidy_fixed<-read.csv("norm_AC_tidy_fixed.csv", stringsAsFactors = FALSE)

head(norm_AC_tidy_fixed)

norm_AC_tidy_fixed$Genus <- reorder(norm_AC_tidy_fixed$Genus,norm_AC_tidy_fixed$guild)

norm_AC_tidy_fixed$Percent_Abundance<-
as.numeric(as.character(norm_AC_tidy_fixed$Percent_Abundance))

norm_AC_tidy_fixed$guild <- as.factor(norm_AC_tidy_fixed$guild)


ggplot(norm_AC_tidy_fixed, aes(x=depth, y=Genus, color=guild, size=Percent_Abundance)) +

  geom_point(aes(fill=guild)) +

  scale_color_manual(values=bub_cols) +

  scale_size(range = c(0,10), breaks = c(0.001, 0.01, 0.1, 1, 1.5, 2)) +

  coord_flip() +

  theme_bw(base_size = 20) +

  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +

  labs(x="Depth (cmbsf)", y= "Genus") +

  scale_x_reverse(limits= c(20,0))
```

R script for sulfate reduction and hydrogen plots

```
######## Sulfate reduction plots ##########

# figure of SRR with depth #

library(ggplot2)

SRR<-read.csv("SRR.csv", header=TRUE)


# Both together:

ggplot(SRR, aes(x = Depth, y = SRR, color=Station, shape= Station)) +

  geom_point(aes(fill=Station), colour="black", size=4, stroke=2) +

  scale_shape_manual(values = c(21, 24)) +

  scale_fill_manual(values=cols) +

  theme_bw(base_size = 20) +

  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),

        panel.background = element_blank(), axis.line = element_line(colour = "black")) +

  scale_colour_manual(values=cols) +

  theme(text = element_text(size=20)) +

  scale_x_reverse() +

  geom_line(aes(color=Station, linetype=Station)) +

  geom_line(size=1) +

  labs(x="Depth (cmbsf)", y= "Sulfate Reduction Rate (nmol cm-3 d-1)") +

  coord_flip()


# separated:

SRR_AB<-subset(SRR, Station %in% c("AB"))

ggplot(SRR_AB, aes(x = Depth, y = SRR, color=Station, shape= Station)) +

  geom_point(aes(fill=Station), colour="black", size=4, stroke=2) +

  scale_shape_manual(values = c(21)) +

  scale_fill_manual(values=cols) +

  theme_bw(base_size = 20) +

  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),

        panel.background = element_blank(), axis.line = element_line(colour = "black")) +

  scale_colour_manual(values=cols) +

  theme(text = element_text(size=20)) +

  scale_x_reverse(limits=c(25, 0)) +

  geom_line(aes(color=Station, linetype=Station)) +

  geom_line(size=1) +
```

```
  labs(x="Depth (cmbsf)", y= "Sulfate Reduction Rate (nmol cm-3 d-1)") +
  coord_flip()

SRR_AC<-subset(SRR, Station %in% c("AC"))
ggplot(SRR_AC, aes(x = Depth, y = SRR, color=Station, shape= Station)) +
  geom_point(aes(fill=Station), colour="black", size=4, stroke=2) +
  scale_shape_manual(values = c(24)) +
  scale_fill_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_colour_manual(values=cols) +
  theme(text = element_text(size=20)) +
  scale_x_reverse(limits=c(25, 0)) +
  geom_line(aes(color=Station, linetype=Station)) +
  geom_line(size=1) +
  labs(x="Depth (cmbsf)", y= "Sulfate Reduction Rate (nmol cm-3 d-1)") +
  coord_flip()

####### Hydrogen ###########
library(ggplot2)
hyd<-read.csv("Hydrogen.csv")
H_AB<-subset(hyd, Station %in% c("AB"))

ggplot(H_AB, aes(x = Depth, y = Hydrogen, color=Station, shape= Rep)) +
  geom_point(aes(fill=Station), colour="black", size=4, stroke=2) +
  scale_shape_manual(values = c(23, 22)) +
  scale_fill_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_colour_manual(values=cols) +
  theme(text = element_text(size=20)) +
  scale_x_reverse(limits=c(30, 0)) +
  geom_line(aes(color=Station, linetype=Station)) +
  geom_line(size=1) +
```

```
  labs(x="Depth (cmbsf)", y= "Hydrogen (nM)") +
  coord_flip()



H_AC<-subset(hyd, Station %in% c("AC"))


ggplot(H_AC, aes(x = Depth, y = Hydrogen, color=Station, shape= Rep)) +
  geom_point(aes(fill=Station), colour="black", size=4, stroke=2) +
  scale_shape_manual(values = c(23, 22)) +
  scale_fill_manual(values=cols) +
  theme_bw(base_size = 20) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  scale_colour_manual(values=cols) +
  theme(text = element_text(size=20)) +
  scale_x_reverse(limits=c(30, 0)) +
  geom_line(aes(color=Station, linetype=Station)) +
  geom_line(size=1) +
  labs(x="Depth (cmbsf)", y= "Hydrogen (nM)") +
  coord_flip()
```

Code for analyses within Appendix III

```
getwd()

library(ggplot2)

library(vegan)

library(dplyr)

library(scales)

library(grid)

library(reshape2)

library(phyloseq)

theme_set(theme_bw())

sharedfile =
"Sva_All.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.unique_list.0.
03.pick.shared" #shared file for the overall sequencing depth (not just Fe/S groups) was
"Shared_all.shared"
```

```
taxfile =
"Sva.trimmed.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.opti_mcc.unique_lis
t.0.03.cons.taxonomy"

mapfile = "Metadata.csv"

mothur_data <- import_mothur(mothur_shared_file = sharedfile, mothur_constaxonomy_file = taxfile)
# Import mothur data

map <-read.csv(mapfile) # Import sample metadata (organic geochem, hydrogen, site, replciate,
fjord, Fe and Mn concentrations)

head(map)

map <- sample_data(map)

rownames(map) <- map$Sample.Id # Assign rownames to be Sample ID's

moth_merge <- merge_phyloseq(mothur_data, map) # Merge mothurdata object with sample metadata

moth_merge

colnames(tax_table(moth_merge))

colnames(tax_table(moth_merge)) <-c("Kingdom", "Phylum", "Class", "Order", "Family", "Genus")

colnames(tax_table(moth_merge))

sample_sum_df <- data.frame(sum = sample_sums(moth_merge)) # Make a data frame with a column for
the read counts of each sample

ggplot(sample_sum_df, aes(x = sum)) +

  geom_histogram(color = "black", fill = "indianred", binwidth = 2500) +

  ggtitle("Distribution of sample sequencing depth") +

  xlab("Read counts") +

  theme(axis.title.y = element_blank())

# mean, max and min of sample read counts

smin<-min(sample_sums(moth_merge))

smin #2 for library AB0-1, when this lib removed, min is 1,936

smax<-max(sample_sums(moth_merge))

smax #49166

smean<-mean(sample_sums(moth_merge))

smean #23289.94, when AB0-1 removed = 23803

sva_phylum<- moth_merge %>%

  tax_glom(taxrank = "Phylum") %>%

  transform_sample_counts(function(x) {x/sum(x)} ) %>%

  psmelt() %>%

  filter(Abundance > 0.02) %>%

  arrange(Phylum)

phylum_colors <- c(

  "#CBD588", "#5F7FC7", "orange","#DA5724", "#508578", "#CD9BCD",
```

```
  "#AD6F3B", "#673770","#D14285", "#652926", "#C84248",
  "#8569D5", "#5E738F","#D1A33D", "#8A7C64", "#599861"
)
ggplot(sva_phylum, aes(x = Sample, y = Abundance, fill = Phylum)) +
  facet_grid(Station~.) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = phylum_colors) +
  #
  guides(fill = guide_legend(reverse = TRUE, keywidth = 1, keyheight = 1)) +
  ylab("Relative Abundance (Phyla > 2%) \n") +
  ggtitle("Phylum Composition of Svalbard sediments \n Bacterial Communities by Sampling Site") +
  theme(axis.text.x = element_text(angle =90, hjust = 1))
###### ordination

# Source code files downloaded from ~/git_repos/MicrobeMiseq/R/miseqR.R
source("C:/Users/JoySpin/Documents/miseqR.R")

minlib = 15000 #minlib for all sequences (not just Fe/S) was 60000
sva_scale<-scale_reads(moth_merge, minlib) #scale reads to even depth
sample_data(sva_scale)$Depth <- factor(
  sample_data(sva_scale)$Depth,
  levels = c(0.5,
             1.5,
             2.5,
             3.5,
             4.5,
             5.5,
             6.5,
             7.5,
             8.5,
             9.5,
             10.5,
             11.5,
             12.5,
             13.5,
             14.5,
```

```
              15.5,
              16.5,
              17.5,
              18.5,
              19.5)
)
require(devtools)

install_version("vegan", version ="2.4-5", repos = "http://cran.us.r-project.org") #Phyloseq qas
written with dependency on an older Vegan package
#Restart R
library(vegan)

sva_pcoa<-ordinate(
  physeq = sva_scale,
  method = "PCoA",
  distance = "bray"
)
palette<-colfunc <- colorRampPalette(c("lightpink", "brown"))

plot_ordination(
  physeq = sva_scale,
  ordination = sva_pcoa,
  color = "Depth",
  shape = "Station",
  title = "PCoA Bray Curtis"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size=3) + scale_color_manual(values = palette(c(20)))

otu_table(sva_scale)

################ PCoA is the same if only samples with > 15000 read are examined
sva_15000<-prune_samples(sample_sums(moth_merge)>15000, moth_merge)
sva_15000
```

```
otu_table(sva_15000)
sva_15000_pcoa<-ordinate(
  physeq = sva_scale,
  method = "PCoA",
  distance = "bray"
)

plot_ordination(
  physeq = sva_15000,
  ordination = sva_15000_pcoa,
  color = "Depth",
  shape = "Station",
  title = "PCoA Bray Curtis"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size=3)



write.csv(sva_15000_pcoa$vectors, file = "sva_15000_pcoa.csv")
########################### PCoA of individual sites
sva_AB<-moth_merge %>%
  subset_samples(Station=="AB")
sva_AC<-moth_merge %>%
  subset_samples(Station=="AC")

sva_AB_pcoa<-ordinate(
  physeq = sva_AB,
  method = "PCoA",
  distance = "bray"
)

plot_ordination(
  physeq = sva_AB,
  ordination = sva_AB_pcoa,
  color = "Depth",
```

```
  title = "PCoA of VK stn AB communities Bray Curtis"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4) + scale_colour_gradient(high = "brown", low = "lightpink")


sva_AC_pcoa<-ordinate(
  physeq = sva_AC,
  method = "PCoA",
  distance = "bray"
)

plot_ordination(
  physeq = sva_AC,
  ordination = sva_AC_pcoa,
  color = "Depth",
  title = "PCoA of VK stn AC communities Bray Curtis"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4) + scale_colour_gradient(high = "brown", low = "lightpink")

############## Prune samples with small libraries

sample_data(sva_15000)
write.csv(sample_data(sva_15000), file = "sva_15000.csv")
sva_AB_15000<-prune_samples(sample_sums(sva_AB)>15000, sva_AB)
sva_AB_15000_pcoa<-ordinate(
  physeq = sva_AB_15000,
  method = "PCoA",
  distance = "bray"
)

plot_ordination(
  physeq = sva_AB_15000,
```

```
  ordination = sva_AB_15000_pcoa,
  color = "Depth",
  shape = "Station",
  title = "PCoA of VK stn AB prokaryotic communities, >15000 reads"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4, shape=19) + scale_colour_gradient(high = "brown", low = "lightpink") +
  geom_text(mapping = aes(label = Depth), size = 5, vjust = 1.5)

sva_AC_15000<-prune_samples(sample_sums(sva_AC)>15000, sva_AC)
sva_AC_15000_pcoa<-ordinate(
  physeq = sva_AC_15000,
  method = "PCoA",
  distance = "bray"
)

plot_ordination(
  physeq = sva_AC_15000,
  ordination = sva_AC_15000_pcoa,
  color = "Depth",
  shape = "Station",
  title = "PCoA of VK stn AC prokaryotic communities, >15000 reads"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4, shape=17) + scale_colour_gradient(high = "brown", low = "lightpink") +
  geom_text(mapping = aes(label = Depth), size = 5, vjust = 1.5)

set.seed(1)
#################### NMDS plot updated 11/16/17 ###################################
sva_nmds<-ordinate(
  physeq = sva_15000,
  method = "NMDS",
  distance = "bray"
)
```

```
plot_ordination(
  physeq = sva_15000,
  ordination = sva_nmds,
  color = "Depth",
  shape = "Station",
  title = "NMDS of Svalbard bacterial Communities"
) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4) + scale_colour_gradient(high = "brown", low = "lightpink") +
  geom_text(mapping = aes(label = Depth), size = 5, vjust = 1.5)


stressplot(sva_nmds)



########### Doing new distance matrix only on libraries with more than 15,000 reads (Nov. 16th,
2017) ###################
VK_15000_not_na<-sva_15000 %>%
  subset_samples(
    !is.na(Hydrogen) &
      !is.na(d13Corg) &
      !is.na("%C") &
      !is.na("C/N")
  )
colnames(sample_data(VK_15000_not_na)) <-c("Sample Id",     "Fjord",       "Station",
"Replicate", "Depth",  "Proximity_to_glacier",       "Hydrogen",    "d13Corg", "Percent_C",
       "CtoN", "Fe", "Mn")
VK_bray_15000_not_na<-phyloseq::distance(VK_15000_not_na, method = "bray")
sampledf_VK_15000<-data.frame(sample_data(sva_15000))
adonis(VK_bray_15000_not_na ~ Station, data= sampledf_VK_15000)
beta_VK_15000<-betadisper(VK_bray_15000_not_na, sampledf_VK_15000$Station)
permutest(beta_VK_15000)
sva_15000_bray<-phyloseq::distance(physeq=sva_15000, method="bray")


############ CAP ord plot ###################
cap_ord_VK <- ordinate(
```

```
  physeq = VK_15000_not_na,
  method = "CAP",
  distance = VK_bray_15000_not_na,
  formula = ~ Hydrogen + d13Corg + Percent_C + CtoN + Proximity_to_glacier + Depth + Fe + Mn)

cap_plot <- plot_ordination(
  physeq = VK_15000_not_na,
  ordination = cap_ord_VK,
  color = "Depth",
  axes = c(1,2)
) +
  aes(shape = Station) +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black")) +
  geom_point(size = 4) + scale_colour_gradient(high = "brown", low = "lightpink") +
  geom_text(mapping = aes(label = Depth), size = 5, vjust = 1.5)
arrowmat <- vegan::scores(cap_ord_VK, display = "bp")
arrowdf <- data.frame(labels = rownames(arrowmat), arrowmat)
arrow_map <- aes(xend = CAP1,
                 yend = CAP2,
                 x = 0,
                 y = 0,
                 shape = NULL,
                 color = NULL,
                 label = labels)
label_map <- aes(x = 1.3 * CAP1,
                 y = 1.3 * CAP2,
                 shape = NULL,
                 color = NULL,
                 label = labels)
arrowhead = arrow(length = unit(0.02, "npc"))

cap_plot +
  geom_segment(
    mapping = arrow_map,
    size = .75,
```

```
    data = arrowdf,
    color = "black",
    arrow = arrowhead
  ) +
  geom_text(
    mapping = label_map,
    size = 5,
    data = arrowdf,
    color = "dodgerblue3",
    show.legend = FALSE
  )

anova(cap_ord_VK, by="terms", perm.max=500)
anova(cap_ord_VK, by = "margin")
anova(cap_ord_VK)



sample_dist<-vegdist(tax_table(sva_15000), method = "bray")
mantel(sva_bray, bray_not_na)



########## diversity


pal="Set1"
moth_merge
plot_richness(sva_15000)
plot_richness(sva_15000, measures = c("Chao1", "Shannon"))
plot_richness(sva_15000, x="Depth", measures = c("Chao1", "Shannon"))
sample_data(sva_15000)$fjord<-get_variable(sva_15000, "Station") %in% c("AB", "AC")
plot_richness(sva_15000, x="Depth", color="Station", measures = c("Chao1", "Shannon"))
sample_data(sva_15000)$fjord<-get_variable(sva_AB_AC, "Station") %in% c("AB", "AC")
plot_richness(sva_15000, x="Depth", color="Station", measures = c("Chao1","Shannon"))
number_ticks<-function(n) {function(limits) pretty (limits, n)}
p<-plot_richness(sva_15000, x="Depth", color="Station", measures = c("Chao1", "Shannon",
"Simpson")) + geom_line() +
  scale_x_reverse() +
```

```
  coord_flip()
plots<-layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE),
       widths=c(3,1), heights=c(1,2))



#### plots used in Appendix
d<-plot_richness(sva_15000, x="Depth", color="Station", measures = c("Shannon")) +
  geom_line() +
  geom_point(aes(size=7)) +
  theme(text = element_text(size=20)) +
  scale_x_reverse() +
  coord_flip() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))

K<-plot_richness(sva_15000, x="Depth", color="Station", measures = c("Chao1")) +
  geom_line() +
  geom_point(aes(size=7)) +
  theme(text = element_text(size=20)) +
  scale_x_reverse() +
  coord_flip()+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))

L<-plot_richness(sva_15000, x="Depth", color="Station", measures = c("Simpson")) +
  geom_line() +
  geom_point(aes(size=7)) +
  theme(text = element_text(size=20)) +
  scale_x_reverse() +
    coord_flip() +
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "black"))

richness_AB_AC<-estimate_richness(sva_15000)
write.csv(richness_AB_AC, "Richness_AB_AC.csv")
```

Appendix III: Alpha Diversity and Ordination of 16S rRNA Gene Amplicon Libraries

Sample sequencing depth and alpha diversity

Using the Phyloseq package for R (1, 2), we calculated the distribution of sample sequencing depth (Figure 4A-III.1). The largest sample read counts was 49,166 and the average was 23289. All samples were dominated by Proteobateria (Figure 4A-III.2). Marine sediments represent one of the most diverse habitats for bacteria and archaea (3). To understand if microbial diversity within Van Keulenfjorden differed according to 16S rRNA gene amplicon analysis between sites or with depth, we calculated several different measures of alpha diversity (Shannon, Simpson, Chao1, Figure 4A-III.2).

Ordination analysis

To understand the differences across samples, unconstrained ordination analyses (Principle Coordinates Analysis or PCoA and Non-Metric Multidimensional Scaling or NMDS) on Bray-Curtis dissimilarity were applied to samples >15,000 reads. The main difference between these two measures is that PCoA solves an eigenvalue equation associated with a linear system, and NMDS can better accurately preserve high-dimensionality of complicated systems in its projections (4). NMDS ranks the distance between samples and preserves these ranks when placing them into two-dimensional ordination space (5, 6). The goodness of fit between the original NMDS plot (with *n* dimensions) into the final two-dimensional projection is captured with a stress plot.

Across both stations AB and AC in Van Keulenfjorden, partitioning between samples largely occurred along a depth gradient (Figure 4A-III.3 and Figure 4A-III.4). This result is not surprising considering typical niche partitioning according to rapid changes in redox conditions that is typical of shallow marine sediment (7, 8). A more detailed look into how environmental parameters may have influence over the community structure was achieved with constrained ordination (4). We used Canonical Analysis of Principal Coordinates (CAP) analysis to see

which environmental parameters have influence over Bray Curtis dissimilarity. The results show that shallow AC samples are influenced heavily by proximity to the main glacier (Figure 4A-III.5). Deeper samples (>12 cm depth) in AC are influenced by Fe^{2+} concentrations, which only begin to pick up in concentration (> 200 μM) after 10 cmbsf likely because of abiotic interactions with Mn. (unpublished data, Lisa Herbert). Importantly, shallow samples in AB appear to be driven by C/N ratios and TOC amounts while deeper samples (> 16 cmbsf) only appears to be strongly influenced by depth. Taken together, these results support our hypothesis that differences in community structure within shallow sediments is driven by organic matter amount and lability, both of which are environmentally determined by spatial gradients along the long axis of the fjord. Anova testing of the model wherein distance is a function of the environmental parameters shown in Figure 4A-III.5 was found to be significant (Tables 4A-III.1 through 4A-III.3).

References

1. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PloS one 8:e61217.

2. RCoreTeam. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

3. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM, Herndl GJ. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proceedings of the National Academy of Sciences 103:12115-12120.

4. Zuur A, Ieno EN, Smith GM. 2007. Analyzing ecological data. Springer Science & Business Media.

5. Shepard RN. 1966. Metric structures in ordinal data. Journal of Mathematical Psychology 3:287-315.

6. Ramette A. 2007. Multivariate analyses in microbial ecology. Fems Microbiology Ecology 62:142-160.

7. Froelich PN, Klinkhammer GP, Bender ML, Luedtke NA, Heath GR, Cullen D, Dauphin P, Hammond D, Hartman B, Maynard V. 1979. Early oxidation of organic matter in pelagic sediments of the eastern equatorial Atlantic: suboxic diagenesis. Geochim Cosmochim Acta 43:1075-1090.

8. Leloup J, Fossing H, Kohls K, Holmkvist L, Borowski C, Jørgensen BB. 2009. Sulfate-reducing bacteria in marine sediment (Aarhus Bay, Denmark): abundance and diversity related to geochemical zonation. Environ Microbiol 11:1278-1291.

Table 4A-III.1. Initial Anova testing of the model "distance ~ Hydrogen + d13Corg + Percent_C + CtoN + Proximity_to_glacier + Depth + Fe + Mn"

	Df	SumOfSqs	F	Pr(>F)	Signif. code
Model	8	3.5437	9.9614	0.001	***
Residual	45	2.0011			

F : The ratio produced by dividing the Mean Square for the Model by the Mean Square for Error.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4A-III.2. Secondary Anova with terms added sequentially for the model "distance ~ Hydrogen + d13Corg + Percent_C + CtoN + Proximity_to_glacier + Depth + Fe + Mn"

	Df	SumOfSqs	F	Pr(>F)	Signif. code
Hydrogen	1	0.10248	2.3045	0.087	.
d13Corg	1	0.19158	4.3083	0.014	*
Percent_C	1	0.12806	2.8799	0.034	*
CtoN	1	0.09975	2.2431	0.086	.
Proximity_to_glacier	1	1.92746	43.3449	0.001	***
Depth	1	0.68151	15.3259	0.001	***
Fe	1	0.08668	1.9492	0.103	
Mn	1	0.32619	7.3354	0.001	***
Residual	45	2.00106			

F : The ratio produced by dividing the Mean Square for the Model by the Mean Square for Error.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 4A-III.3. Final Anova testing with marginal effects of terms in the model "distance ~ Hydrogen + d13Corg + Percent_C + CtoN + Proximity_to_glacier + Depth + Fe + Mn"

	Df	SumOfSqs	F	Pr(>F)	Signif. code
Hydrogen	1	0.04282	0.963	0.386	
d13Corg	1	0.0231	0.5194	0.685	
Percent_C	1	0.08147	1.8321	0.128	
CtoN	1	0.10525	2.3668	0.082	.
Proximity_to_glacier	1	1.20108	27.0101	0.001	***
Depth	1	0.23357	5.2524	0.005	**
Fe	1	0.07301	1.6419	0.166	
Mn	1	0.32619	7.3354	0.002	**
Residual	45	2.00106			

F : The ratio produced by dividing the Mean Square for the Model by the Mean Square for Error.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

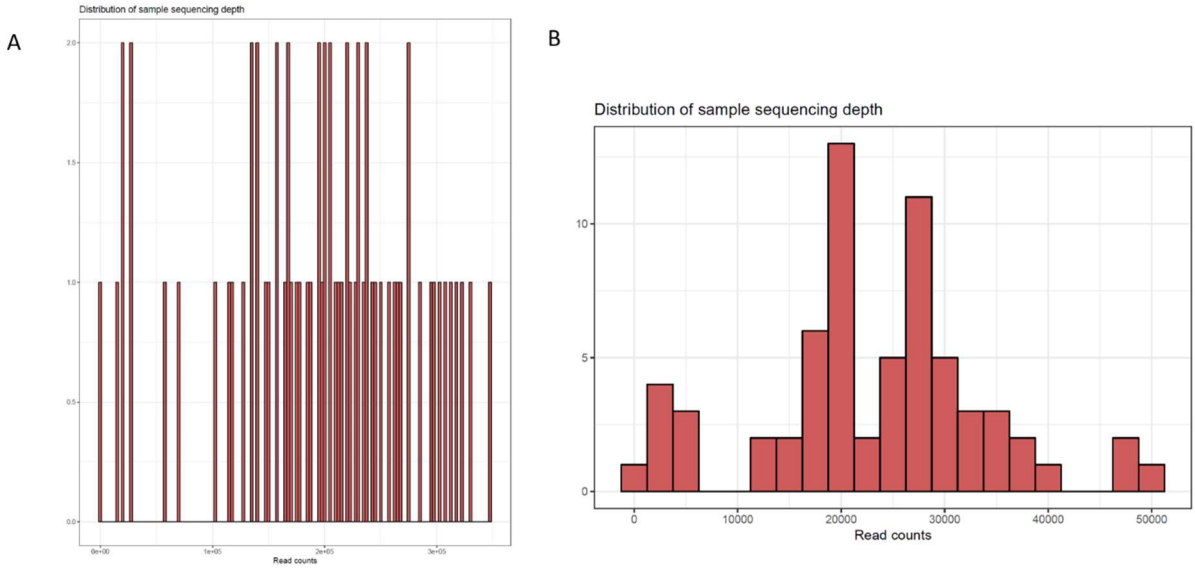

Figure 4A-III.1. Distribution of sequencing depth for all AB and AC amplicon libraries considering all sequences (A) and only those sequences identified as iron and/or sulfur groups (B).

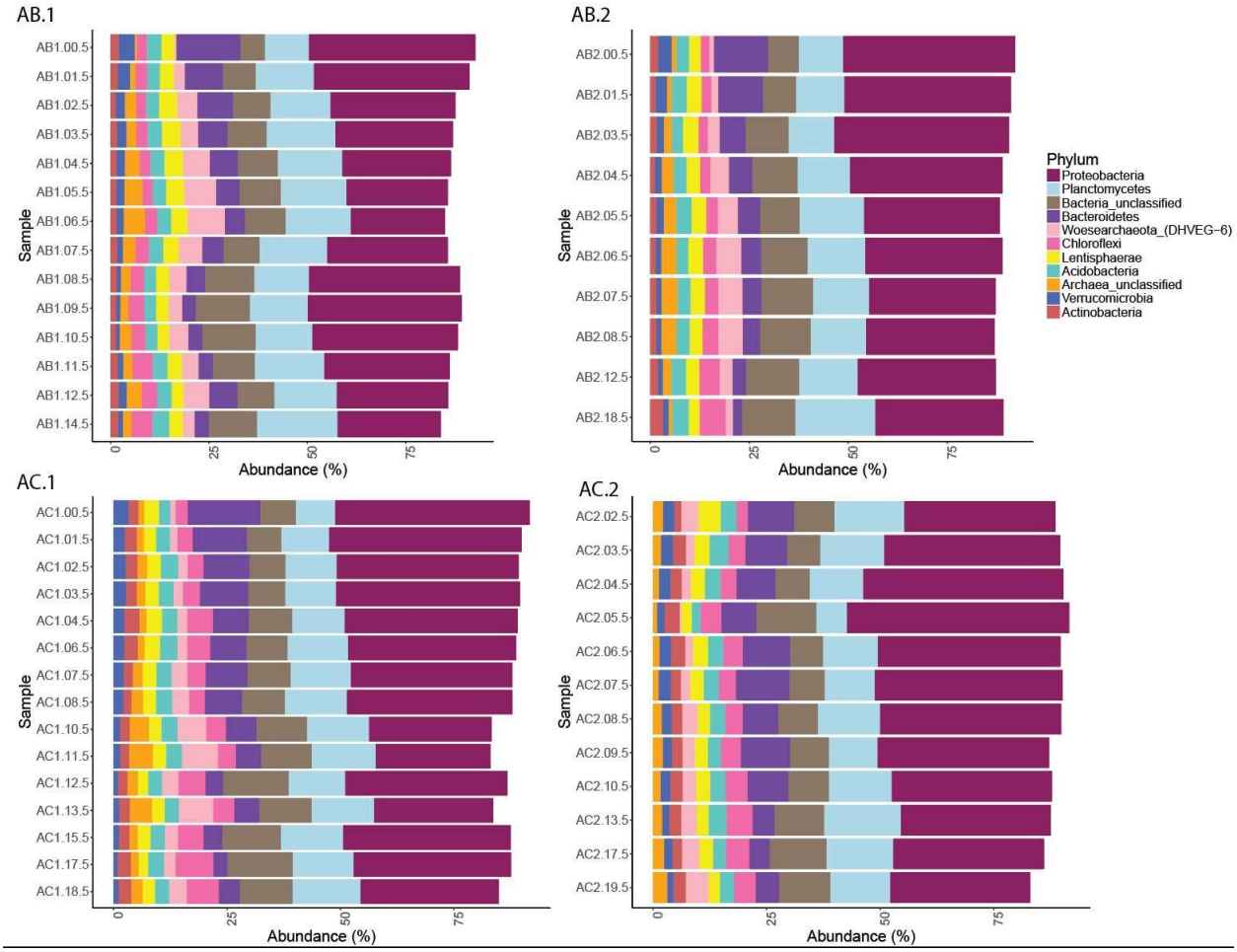

Figure 4A-III.2. Phylum distributions across all libraries.

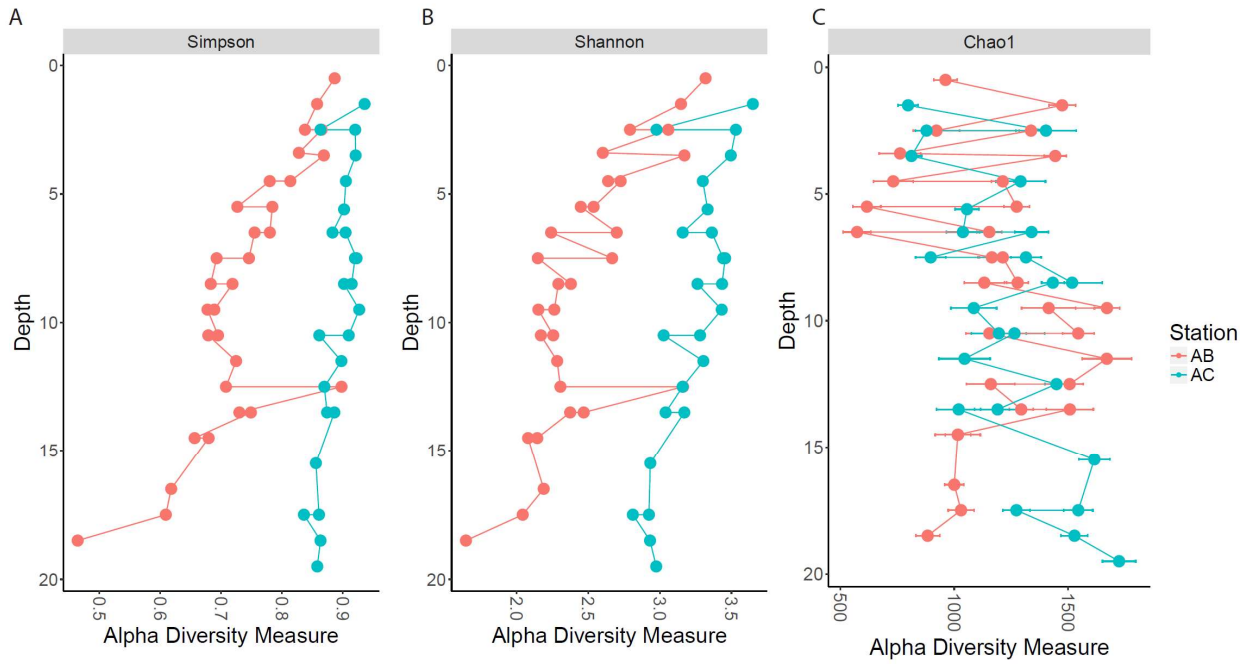

Figure 4A-III.3. Alpha diversity with depth for stations AB (salmon) and AC (teal). Results are for iron and sulfur groups only.

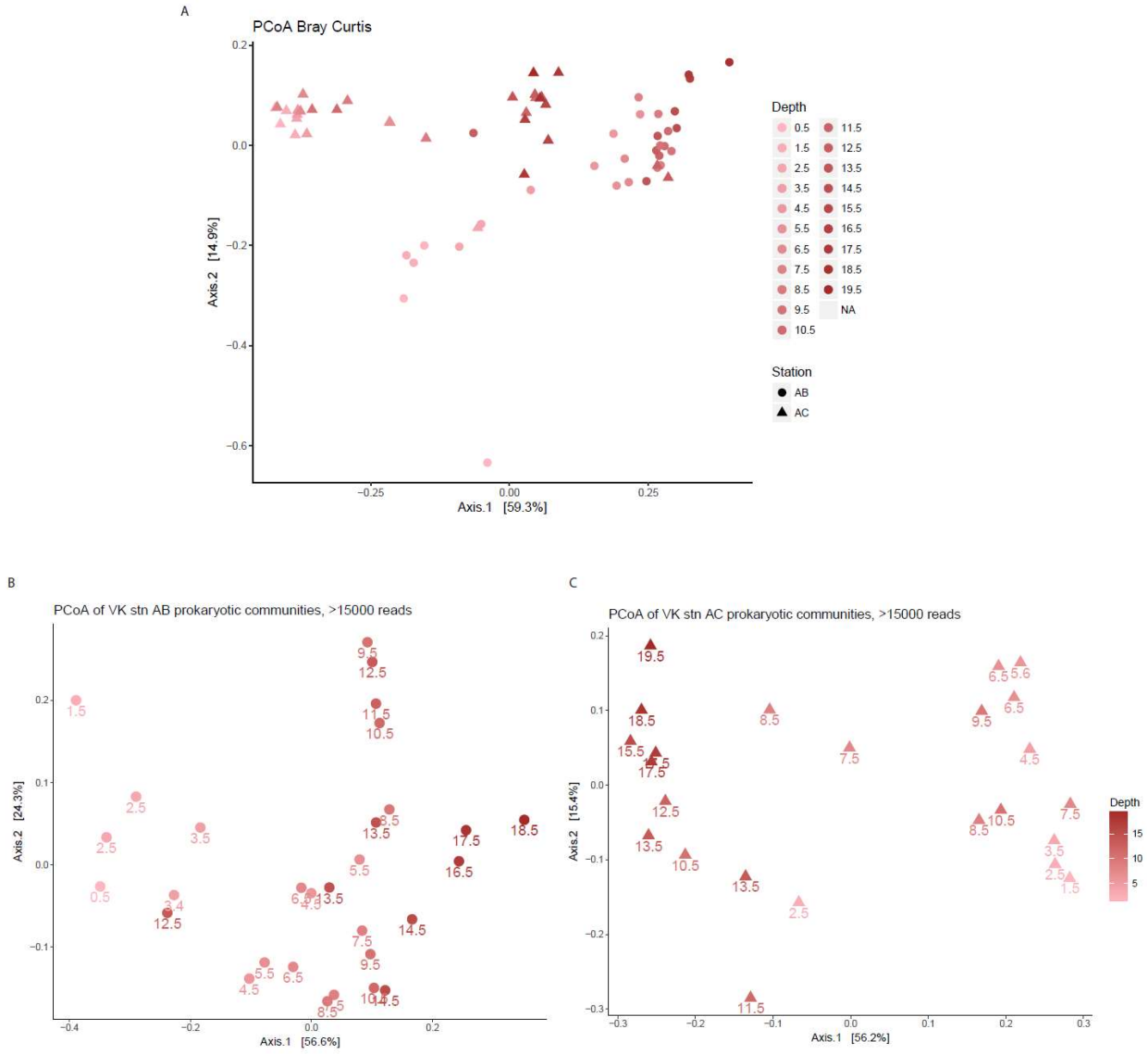

Figure 4A-III.4. Principle coordinates analysis on Bray Curtis dissimilarity for iron and sulfur taxa within sites in Van Keulenfjorden. Panel A contains both sites together, while panels B and C show individual results for stations AB and AC, respectively.

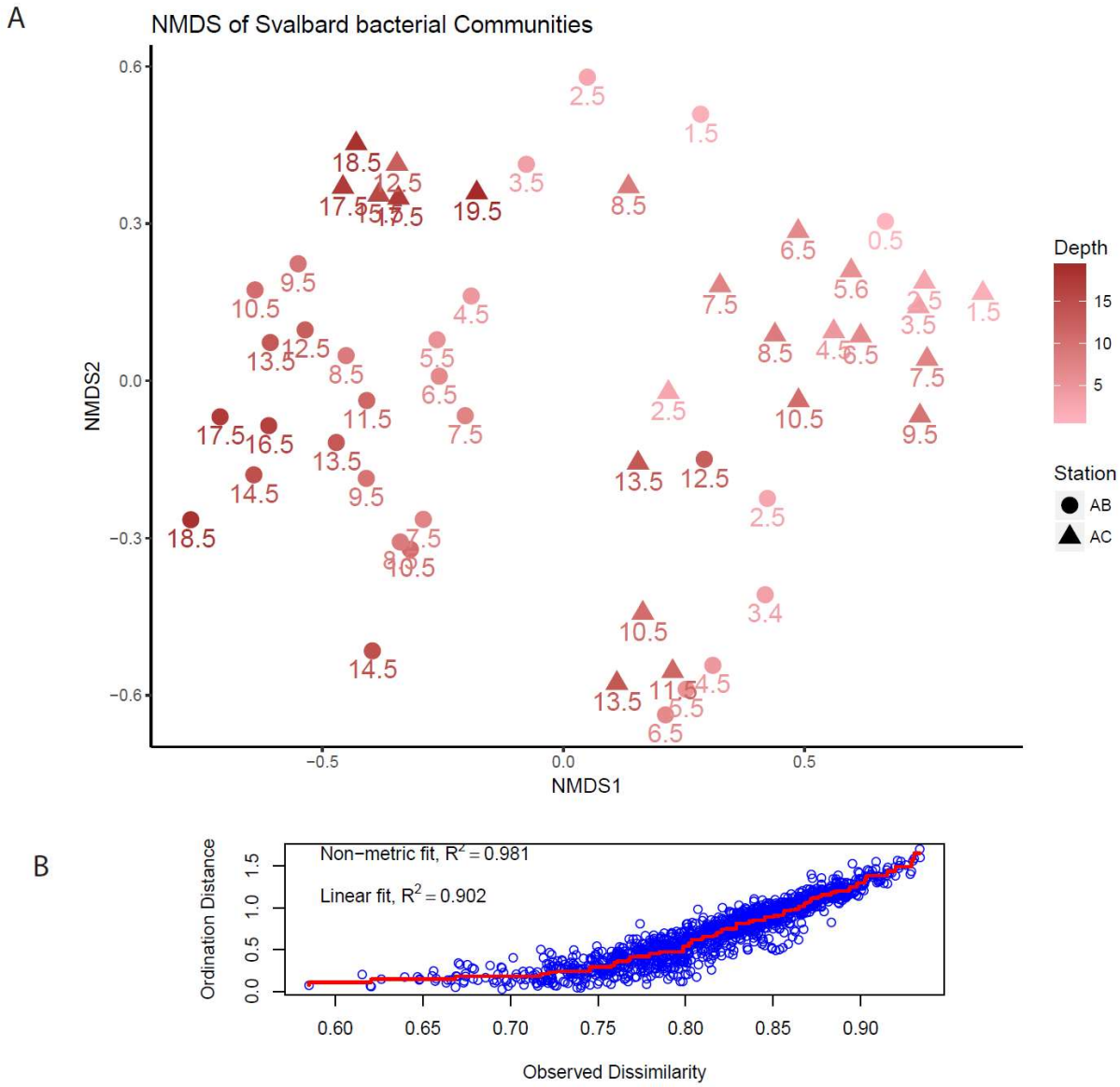

Figure 4A-III.5. NMDS for stations iron and sulfur groups in AB and AC (A) and stress plot for goodness of fit (B).

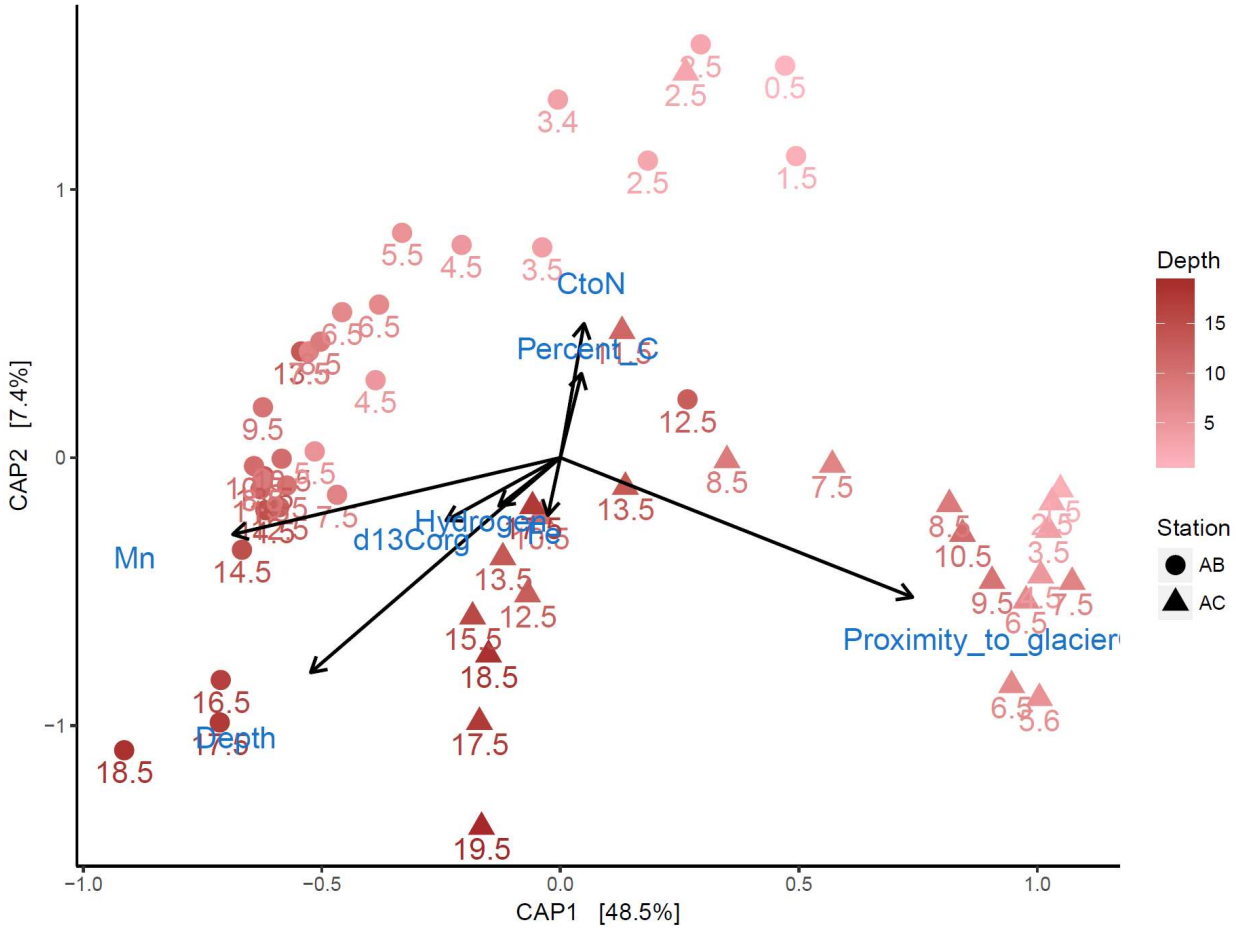

Figure 4A-III.6. Canonical Analysis of Principal Coordinates (CAP) plot with environmental variables as vectors describing compositon of iron and sulfur taxa within each depth.

Chapter 5: Genomic and transcriptional evidence for physiological responses to burial of the dominant carbon-fixing clade Woeseiaceae in Arctic fjord sediment

Abstract

Dark carbon fixation within marine sediments is performed largely by chemoautotrophic gammaproteobacteria. The most abundant and widely-distributed of these is the clade JTB255, recently identified as being a member of the Woeseiaceae. Single cell genomic sequencing and metagenomic binning of this group showed the potential for both chemolithoautotrophy and heterotrophy, highlighting the potential for Woeseiaceae to act as both a carbon source and a carbon sink in its environment. However, the only cultured representative of the Woeseiaceae was identified as a non-spore forming obligate chemoheterotroph. This suggests that uncultured Woeseiaceae clades may have fundamentally different physiologies compared to this isolate. Further, although community composition studies suggest that the Woeseiaceae are extremely abundant in marine sediments worldwide, very little is known about their transcriptional activity *in situ*, especially in Arctic marine sediments where understanding climate-affected carbon dynamics is important and timely. We used 16S rRNA gene sequencing, metagenomic binning, and transcriptomics (at 1 cm depth intervals) to uncover the *in situ* abundance, genomic content, and activity in fjord sediments of Svalbard (79°N). We reconstructed 5 Woeseiaceae genomes, whose phylogenetic placement was in the Steroidobacterales, updating previous phylogenies which placed them into Chromatiales. The genomes encoded a truncated Sox pathway for the oxidation of diverse sulfur intermediates linked to a reverse dissimilatory sulfide reductase (rDSR) pathway for the complete oxidation of thiosulfate. In addition, sulfur oxidation could generate ATP for the reduction of inorganic carbon with a complete Calvin Benson Cycle. Transcriptional recruitment is relatively high among these genomes. With increasing sediment depth, anoxic conditions appeared to stimulate the transcription of nitrite reductase (*nir*S) involved in denitrification. The cytochromes encoded in the genomes span a vast range of redox potential, suggesting that Woeseiaceae have flexible redox preferences within microaerobic to

anoxic conditions. Importantly, as Woeseiaceae continued to be buried, they increased transcription of genes related to stress-mitigation and sporulation while simultaneously decreasing transcription of genes related to growth. Sequencing at high depth resolution allowed us to capture nuanced changes that highlight the delicate interplay between redox conditions and transcriptional activity of redox-sensitive enzymes and the strategies Woeseiaceae use for subsiting after burial for future population re-seeding.

Introduction

In marine sediments worldwide, chemoautotrophic bacteria perform carbon fixation in the absence of light, which has been estimated to total at least 0.11 Pg C y^{-1} (1). Among these microorganisms, the gammaproteobacterial clade JTB255 have been shown through isotope labeling and FACS sorting studies to contribute nearly 20% of total microbial carbon fixation in marine sediments (2). This coastal group has a worldwide distribution (as reviewed by (2)) and are consistently among the most abundant groups by sequence abundance (3-5). Single cell genomic sequencing and metagenomic binning of this clade showed the potential for chemolithoautotrophy as well as heterotrophy (3). Phylogenomic analysis allowed JTB255 to be designated as a member of the recently established Woeseiaceae, which has only one cultured representative, *Woeseia oceani* XK5 (3, 6). Contrary to both laboratory and genomic evidence that suggested that Woeseiaceae fix inorganic carbon, *Woeseia oceani* XK5 was shown to be a non-spore forming, obligate heterotroph and therefore incapable of autotrophic growth (6), suggesting that it may have a fundamentally different physiology than the uncultured Woeseiaceae clades that are abundant in marine sediments.

For a clade such as this with a worldwide distribution and evidence for vast metabolic versatility, surprisingly little is known about the activity of Woeseiaceae *in situ*. Here, we aim to understand the *in situ* transcriptional activity of Woeseiaceae populations in Arctic coastal sediments where the presence of Woeseiaceae/JTB55 has been noted previously (5). We hypothesize that depth-resolved transcriptional evidence will shed light how this clade physiologically responds to changing conditions with burial. We used 16s rRNA gene libraries to determine sequence abundance with depth and metagenomically assembled genomes (MAGs) to understand if the genomic contents of these Arctic Woeseiaceae genomes are in any way tailored

to its cold, organic carbon-limited environment compared to other published genomes. Our new phylogenetic analysis showed that Woeseiaceae belong within the Steroidobacterales instead of the previously-assigned Chromatiales. The MAGs contained mechanisms for carbon oxidation through the oxidation of reduced sulfur species as well as evidence for a truncated denitrification pathway that could lead to the release of nitric oxide. Next, we used transcriptomics to uncover the transcriptional landscape in this enigmatic group related to energy metabolism, carbon fixation, and spore formation. Our work suggests that redox-sensitive regulators help Woeseiaceae perform metabolic switching from sulfur oxidation coupled to carbon fixation through the Calvin Benson Cycle to nitrite reduction with changing respiratory conditions according to sediment depth. We also present the evidence for expression of spore forming proteins that may aid in this group's ability to re-seed its populations after periods of suboptimal conditions. We place our findings within the context of total organic geochemistry measurements to understand how feedbacks associated with a warming climate may impact this key group.

Methods

Sediment collection

Sediment for sequencing analyses was collected in the summer of 2016 from different stations within Svalbard fjords (79°N). Sediments included in this study are from Stations AB (77°35.249' N, 15°05.121'E) and AC (77°32.260' N, 15°39.434' E) in Van Keuelenfjorden (outer and middle stations, respectively) and Stations F (78°55.075' N, 12°15.929' E) and P (78°57.915'N, 12°15.600'E) in Kongsfjorden (both located at areas of glacier outflow) (Figure 5.1). Smeerenbergfjoren Station J (79°42.8'N, 11°05.9'E) sediment was also used to supply additional sequencing information for the construction of higher quality draft genomes only and genomes reconstructed from Station J are not included here.

Geochemistry

Sediment was collected with a HAPS corer and subsectioned every 3 cm down to a depth of ~30 cm. Samples for hydrogen analysis consisted of 1 mL of sediment placed into a dark glass serum vial which was then crimp sealed, hand shaken, and gassed with N2 for 15 min prior to storage at 4°C. Headspace was measured with glass syringes on a Peak Performer GC at The University of Tennessee, Knoxville after 2 days. Sediment for analysis of organic matter was freeze-dried after thawing from -80°C and subjected to acid fumigation overnight before analysis (88). Total organic carbon as well as isotope composition of carbon and nitrogen from bulk organic matter was measured using a Thermo-Finnigan Delta XL mass spectrometer coupled to an elemental analyzer at The University of Tennessee, Knoxville. Carbon to nitrogen (C/N) ratios were calculated by dividing percent C by percent N. Isotopic values were calibrated against the USGS40 and USGS41 international standards. In-house standard sets were run every 12 samples. Across multiple runs, one standard deviation was 0.1-0.2 ‰ for δ^{13}Corg, 1.1-1.8 % for mgN, and 1.0-2.2 % for mgC.

DNA extraction

Cores for molecular analyses were subsectioned at 1 cm depth intervals in the Ny-Ålesund Marine Lab down to ~20 cm depth. Sediment was frozen immediately on dry ice and remained frozen during transport. Sediment was stored at -80°C until processing. Nucleic acids were extracted by both the Lloyd (Stations AB, AC, P, F) and Loy (Station AC at 18 cm depth, Station J) laboratories using the Qiagen RNeasy Powersoil® Kit for RNA with the DNA extraction accessory according to manufacturer's instructions. RNA extracts were treated with DNase in-house (Qiagen) and further DNase treatment was performed at MRDNA (Shallowater, TX), followed by sequencing of metatranscriptomic libraries with Illumina HiSeq 2500, PE 2x250 bp. Individual 1 cm-depth resolved metatranscriptomic libraries were generated with RNA

extracts from the first 5 to 6 cm of sediment from stations AB, AC, F, and P, for a total of 20 metatranscriptomes. Metagenomic libraries were generated from the combined extracts from the first 5 cm (spanning 0 to 5 cm downcore) in Stations AB and F with MRDNA with Illumina HiSeq 2500 PE 2x250 bp. The Loy Lab sequenced DNA extracts from Station AC at 18 cm depth using Illumina HiSeq 3000, PE 2x250 at The University of Vienna.

Prior to metagenomic assembly, raw reads were trimmed for quality and adapters were removed using in-house scripts in the Loy Lab (which removed the leading eight 5' bases, bases with QC < 15 and reads below 50 bp in length). The Lloyd Lab used Trimmomatic (7) for trimming both metagenomes and metatranscriptomes, with a sliding window of 10 and a Phred cut off score of 28 for all reads above 90 bp. The quality of trimmed reads was assessed with Quast 4.5 (8).

16S rRNA gene amplicon libraries

Taxonomic diversity of Van Keulenfjord sediments was evaluated using 16S rRNA gene library sequencing. Genomic DNA extracts from duplicate cores at each were used to generate 16S rRNA gene amplicon libraries. Sequencing from Stations P and F failed and are therefore not presented here. The Phusion Master Mix (Thermo Fisher) was used with the primer set 515F/806R (9) at the Center for Environmental Biology at The University of Tennessee, Knoxville for amplification. Reads were sequenced with Illumina MiSeq and trimmed for quality with Trimmomatic (7) using a window 10 base pairs wide and a minimum phred score of 28. Trimmed reads were then processed in mothur 1.35.1 (10) using the computational cluster at the Bioinformatics Resource Facility (BRC) at The University of Tennessee, Knoxville. Operational taxonomic units (OTUs) were clustered *de novo* at the 97% similarity level with the SILVA release 123 (11). Recent analysis suggests that clutering at ~100% similarity is appropriate to

identify OTUs with high taxonomic resolution (12), however, work at the BRC with controlled mixed communities identified high sequencing error rates that would preclude higher similarity cutoffs suggested by Edgar (2018). OTU counts were normalized in mothur with the normalized.shared command with a minimum library size of 60,000 sequences.

Metagenomic assembly

MetaSPAdes. Metagenomic assembly was accomplished using metaSPAdes (13) both locallyfor Linux and in KBase, browser interface with bioinformatics modules and applications (14). Station F was the only metagenome that was assembled via SPAdes version 3.11 locally. The metaSPAdes option was utilized with kmer sizes of 21, 33, 55, 77, 99, and 127. Assembled contigs were then filtered to contain only contigs with more than 5x coverage and 1000bp length using in-house scripts. All other metaSPAdes assemblies were completed on KBase with the default parameters (1000bp length and kmer sizes of 21, 33, and 55).

IDBA and Megahit. To reduce RAM utilization and wall clock time, larger sequence datasets were normalized with bbnorm in the Loy lab. Station F and Station AB metagenomes were assembled with IDBA version 1.1.3 with default settings in the Lloyd lab (on KBase) and in the Loy lab (via command line). Asemblies with Megahit were completed either locally or on KBase by the Lloyd lab. All assemblies were generated with 1000 minimum contig length, except IDBA assembly in the Loy lab, which used a 500 bp cutoff.

Taxonomic binning of contigs into metagenome assembled genome (MAGs)

Contig binning was carried out with MaxBin2 v. 2.2.3 (15), CONCOCT (16) and MetaBAT (17) in command line. Each of these binning tools utilizes genomic signatures within contigs, such as coverage and kmer frequency, to identify discrete clusters of contigs that likely represent a population's genome. MaxBin2 and CONCOCT binning were performed with default

parameters and 1000 bp minimum contig length. MetaBat binning was conducted with an interactive pipeline that applies decreasing levels of stringency within each successive iteration, collecting the best bins and their contigs each time. The remaining contigs are then placed into the next step with a lower stringency (see batch script in Appendix I).

To achieve the highest quality genomes possible, DAS Tool (18) was applied to each set of binned contigs. This bioinformatic tool takes several different MAGs as input and identifies a consensus, non-redundant genome for each MAG, leading to higher quality genomes. Quality was determined for our MAGs by categorizing CheckM (19) completeness and contamination according to Bowers et al. (2017). In brief, medium quality drafts have completeness and contamination values of ≥ 50% and <10%, respectively whereas high quality MAGs have completeness and contamination values >90% and contamination <5%. Taxonomic assignment was determined through phylogenetic analysis using a concatenated alignment of single copy marker genes included in the CheckM suite and a tree was built with FastTree (21).

Phylogenetic analysis of Woeseiaceae MAGs

Phylogenetic analysis of Woeseiacea genomes was conducted with the phylogenomic workflow in Anvi'o v. 5.1 (22) with publicly available Woeseiacea genomes downloaded from NCBI or IMG. A total of 49 ribosomal genes were identified, concatenated and aligned with Clustal in Mega v. 7 (23), wherein a maximum-likelihood tree was built with 1000 bootstraps. Members of the Rhodbacterales were used as outgroups.

RubisCO sequence analysis

Publicly available full length (> 480 aa) sequences for large and small chains of ribulose bisphosphate carboxylase (RubisCO) were downloaded from NCBI and compared with sequences annotated in our MAGs. Alignments were conducted in Mega v. 7 using Clustal.

Mega was also used to construct the neighbor-joining tree and topologies shown have >80% support after 1000 bootstraps.

Pangenomic analysis

The pangenomic workflow was implemented in Anvi'o v. 5.1 in combination with the phylogenomics workflow (22). The 5 MAGs reconstructed in this study were evaluated alongside the 5 Woeseiaceae genomes available on NCBI or IMG. Independent gene annotation was performed on these genomes alongside ours with Prokka (24) (alongside the MAGs in this study using a curated database including information for both bacteria and archaea).

Transcripts, mapping and annotation

Ribosomal sequences were identified and removed using all bacterial, archaeal, and eukaryotic databases included in SortmeRNA (25). To understand how mRNA recruited to MAG contigs, filtered transcripts were mapped to MAGs using Bowtie2 (26) with sensitive local mapping. Mapping files, Prokka gene calls, and fasta files for each MAG were then used in the metagenomics pipeline of Anvi'o v 5.1 (22) for COG identification, transcript abundance, coverage estimates, and visualization. Regression analysis on transcript coverage was performed in R (27)(R script is contained within Appendix II).

Results and Discussion

Geochemistry

Hydrogen within Kongsfjorden Stations P and F showed no clear trend with depth after an initial drop from ~1.2 nM to ~0.75 nM at both stations (Figure 5.2A). These values are too low for methanogenesis but may support sulfate reduction (28). Total organic carbon was overall very low (<1 wt %), characteristic of Kongsfjorden (Figure 5.2B) (29) and Svalbard fjords in general (30). Low surface water productivity near the fjord head, coupled to high rates of terrigenous clastic sediment results in low organic carbon contents in these sediments. In support

of this, previous organic geochemistry work showed a seaward increase in both organic matter quantity and quality in Van Keulenfjord (31). As both Stations P and F in Kongsfjord are situated at the fjord head (Figure 5.1), the organic matter that is delivered to the sediment is less labile than freshly deposited algal material. This is demonstrated with high C/N values that reach near 14 (Figure 5.2C), in line with previous C/N measurements made in Kongsfjorden (29). Isotopic signatures of organic matter ($\delta^{13}C_{org}$) range between -23 ‰ (vs. VPDB) and -26 ‰ (vs. VPDB). These values are more depleted in ^{13}C than the $\delta^{13}C_{org}$ of primary producers in Svalbard, which ranges from -15.7 to -19.7 ‰ in ice algae (32, 33) and from -22 to -24 ‰ in marine phytoplankton (34, 35). Instead, $\delta^{13}C_{org}$ signatures at Stations P and F may reflect terrestrial material, such as soil and coal, which have $\delta^{13}C_{org}$ averages of -26 and -25 ‰, respectively (34). Together, this suggests very low amounts of organic matter are delivered to the seabed to fuel subsurface communities, and the organic carbon that does reach the seabed is ancient and therefore largely not bioavailable (29).

Relative sequence abundance by 16S rRNA gene amplicon libraries

16S rRNA gene amplicon library surveys of two stations within Van Keulenfjorden showed that Woeseiaceae is consistently within the top three most abundant families, making up between 2 and 4% of reads in each of our libraries (Figure 5.3). These results are consistent with findings from recent meta-analyses, which highlighted the broad biogeographic distribution and consistently high sequence abundance of this clade (2, 3). In fact, across 65 separate studies across the globe, sequences for the Woeseiaceae were detected 92% of the time and were reliably found to be among the most abundant bacteria (2). Our sequencing data showed that the relative abundance of Woeseiaceae sequences did not have observable depth trends at site AB (Figure 5.2 A) and slightly increased with depth at site AC (Figure 5.2 B).

Despite their suspected metabolic versatility, little is known about the ability of Woeseiaceae to respond to stress associated with nutrient-depleted conditions and the question remains if these populations continue to metabolize and grow as they become buried. Microbial populations in deeply buried sediments, for example, adopt a subsistence strategy of remaining in a state of dormancy until conditions are again suitable, even in million-year-old sediments (36). These populations may remain on metabolic "standby", whereby they subsist after burial by slowing down metabolism and spending more energy on maintenance metabolism rather than growth (37, 38). To test if the relative increase in sequence abundance with depth is associated with enhanced growth, we performed metagenomic and metatranscriptomic sequencing within the first 5 cm where Woeseiaceae sequence abundance increases.

Summary of assemblies and Woeseiaceae MAGs

After binning with these assemblies and dereplication of similar contigs, 49 high and medium quality draft genomes were recovered from Stations AB, AC, and F (Table 5.1). Completeness ranged from 51% to 98% based upon single copy marker genes in the CheckM suite (19) (Table 5.2). Phylogenetic analysis with a concatenated alignment of CheckM genes in FastTree (21) allowed us to determine that binning efforts captured diverse bacterial genomes across the alpha-, delta-, and gamma-proteobacteria and as well as an archaeal genome. Importantly, we reconstructed four medium quality and one high quality genome from the Woeseiacaeae/JTB255 clade across stations AB, AC, and F (Table 5.2).

Phylogenomic analysis

A concatenated ribosomal protein alignment including *Woeseia* sp. genomes from the IMG and NCBI databases (Table 5.3) showed that these genomes grouped with the Steroidobacterales (Figure 5.4). This supports the most current Silva taxonomy and updates previous studies which classify Woeseiaceae as grouping within the Chromatiales (4, 6).

Read recruitment to Woeseiaceae genomes

Cumulative read recruitment to all reconstructed MAGs varied slightly, from 18.3% recruitment in the AC metagenome to 24.1% recruitment in the AB metagenome (metaSPAdes assemblies). Interestingly, recruitment was dominated by Woeseiaceae genomes (Figure 5.5). To determine how the Woeseiaceae genomes recruited transcripts from Stations P, F, AC, and AB, we determined abundance, or the mean coverage of MAG contigs divided by the transcriptome's overall mean coverage (22). In this way, transcript recruitment is normalized within a sample. This means that abundance patterns cannot be interpreted across metatranscriptomic samples; however, the relative abundance of transcript recruitment across MAGs within a sample can be determined. Read abundance for all MAGs in metatranscriptome samples is reported in Figure 5.6 to provide context for abundance within the Woeseiaceae genomes. The 5 Woeseiaceae genomes (marked with asterisks in Figure 5.6) had low transcript abundance within the first 2 cm at Stations F and P. At AB, the highest abundance for these genomes was observed at 2-3 cm, and there was good agreement between the two biological replicates for this depth interval.

Increased transcript abundance with depth relative to other MAGs is observed at both Kongsfjorden sites for several Woeseiaceae genomes, even those with very low transcript abundance overall, such as Woeseia_stnAC. This would seem to confirm our 16S rRNA gene sequence analysis, which showed an increase in abundance with depth; our transcript abundance data likewise suggested an increase in activity at the individual population genome level. However, it is still uncertain to what genes the transcripts are recruiting—whether they are tied to metabolism, growth, or stress responses is unclear from transcript abundance alone. To answer this question, coverage for pathway- and enzyme-specific genes were analyzed after reconstructing the main respiratory and carbon fixing pathways encoded in the genome.

Carbon fixation under low oxygen conditions

Reduction of CO_2 can occur via either the reverse TCA (rTCA) cycle or the Calvin Benson–Bassham Cycle (CBB) cycle. None of our genomes contain evidence for an rTCA cycle. Instead, our genomes encode the genes for the CBB cycle. Ribulose bisphosphate carboxylase (RubisCO, *cbb*M) was detected within all Woeseiaceae genomes and transcribed by the MAG Woeseia2_stnAC in Van Keulenfjorden and Woeseia_stnAB in both fjords (see Appendix III for heatmap of genomic and transcript content of the other non-Woeseiaceae MAGs). In support of a functional CBB cycle, we found either a complete (Woeseia_stnAB, Woeseia_stnF, Woeseia2_stnF) or nearly complete (Woeseia_stnAC, Woeseia2_stnAC) CBB pathway, with gylceraldehyde-3-phosphate dehydrogenase (GADPH, encoded with *gap*), phosphoglycerate kinase (PGK, encoded with *pgk*), and phosphoribulo kinase (PBK, encoded with *cfx*P) all detected and transcribed in some genomes in both fjords. Further, co-localization analysis of contigs showed that RubisCO genes are often nearby other CBB cycle genes.

Despite the presence of the CBB pathway, carbon fixation in these MAGs may be hampered by a few of the problems notable with the RubisCO enzyme. First, molecular oxygen competes for the active site of RubisCO, which leads to costly side-reactions associated with photorespiration. Specifically, 2-phosphoglycolate (2-PG) is a toxic side product of RubisCO during O_2 fixation that forms when the enzyme acts as an oxygenase instead of a carboxylase (39, 40). To prevent 2-PG from inhibiting CBB pathways, cells can metabolize it with phosphoglycolate phosphatase (*cbb*ZC) which converts 2-PG to glycolate (41). In our Woeseiacceae genomes, contigs containing RubisCO or PBK always contain *cbb*ZC. Participation of phosphoglycolate phosphatase in substrate generation for either the glyoxlate cycle or photorespiration is not supported in the genomes, and so cotranscription of

phosphoglycolate phosphatase with RubisCO is a proactive means of removing any 2-PG that may arise through oxygenase activity.

Another way to circumvent the oxygenase activity of RubisCO is to increase its affinity toward CO_2 compared to O_2. Type I RubiCOs, widely distributed across the alpha-, beta-, and gammaproteobacteria (42, 43), have a higher affinity for CO_2 compared to O_2 (44, 45). Sequence analysis of the RubisCO genes in our genomes indicate that all are Type II RubisCOs (Figure 5.7). Type II have a similarly wide distribution across prokaryotes, although a lower affinity for CO_2 must be overcome by alternative strategies. Higher oxygenase activity in RubisCO Type II can be mitigated by carbon dioxide-concentrating mechanism proteins (CCMs), which facilitate the active uptake of inorganic carbon species, including both bicarbonate and CO_2 (45-47). Genes encoding these proteins (*ccm*L and *ccm*K) were detected in nearly all of our Woeseiaceae genomes. Evidence for cotranscription of CCMs with CBB cycle genes, as has been observed in some *Synechococcus* species (48), was not detected in our Woeseiaceae genomes.

The detection of genes associated with strategies to circumvent molecular interactions with oxygens suggest that the microorganisms in our system are exposed to low levels of oxygen. Previous microelecrode studies of the oxygen penetration depth in and around Svalbard show that average oxygen penetrates only 6 to 8 mm within the sediment (49, 50). We searched for genes in the genomes that would provide evidence for aerobic lifestyle or microaerophily. While we were able to identify the presence of 8 genes considered to be oxygen-specific encoded in the genomes (51), none were transcribed. Likewise, we did not find evidence for the high-redox potential cytochrome c oxidase aa_3 -type (respiratory complex IV) (52), which has been suggested to be used in oxygen respiration in Woeseiaceae (3). Instead, we detected the presence of three subunits encoding the functional core and catalytic units of cytochrome c oxidase cbb_3-

type (*cco*N1, *cco*O, and *cco*P). With a high affinity for oxygen, cbb_3-type oxidases are used for respiration under low oxygen conditions and have been implicated in aerobic respiration in *Rhodobacter sphaeroides* (53), nitrogen fixation in endosymbionts (54), and sulfur oxidation by uncultured bacteria in microbial mats (55), all under microaerobic conditions. Transcription of cbb_3-type oxidases is found in Woeseia2_stnAC and Woeseia2_stnF, the latter of which encodes the membrane-associated subunit of the aerobic respiration control sensor protein complex (*arc*B) on the same contig as cbb_3-type oxidase. This complex is responsible for transcriptional regulation of enzymes involved in aerobic versus anaerobic pathways in facultative anaerobic bacteria (56, 57). ArcAB senses and responds to changing respiratory conditions, therefore preventing the unnecessary translation of enzymes involved in respiratory pathways for which the terminal electron acceptor is not present. Most of our Woeseiaceae genomes have the *arc*B gene, suggesting this clade can switch between multiple respiratory capabilities. In fact, *arc*B is co-localized together with RubisCO and nitrite reductase (*Nir*S) in Woeseia2_stnAC, which also has sulfur oxidation genes (*sox*). To understand the potentially modular metabolic capabilities of Woeseiaceae under different conditions, we reconstructed sulfur oxidation and nitrite reduction pathways in relation to carbon fixation.

Sulfur oxidation fuels the CBB cycle

In sulfur oxidizing microorganisms, carbon fixation coupled to the oxidation of reduced sulfur hinges on the production of ATP and reducing equivalents. The generation of reducing equivalents is carried out via membrane-associated reverse electron transport reactions involving the oxidation of H_2S to either zero-valent sulfur or sulfate as end products. These reactions occur through either the consumption of ATP, or through the proton motive force generated when reduced sulfur is oxidized exergonically with oxygen (58). We detected in our genomes several

mechanisms by which reduced sulfur intermediates can be completely oxidized to sulfate to generate ATP. The rarest method among our genomes is the traditional Sox pathway (*sox*BCDXAYZ) that leads directly to the complete oxidation of reduced sulfur species (thiosulfate, hydrogen sulfide, elemental sulfur, and sulfite) to sulfate (58-60) (Figure 5.8). All subunits necessary for the complete Sox pathway were found only in Woeseia_stnF (Table 5.4), although none were transcribed. In some genomes, we also found evidence for the coupling of formate dehydrogenase and polysulfide reductase in the reduction of polysulfides, generating reduced sulfur. This reduced sulfur can be fed into the branched/truncated Sox pathway involved in thiosulfate oxidation.

In a truncated system, one in which *sox*CD are missing and *sox*YZ cannot be regenerated, elemental sulfur will accumulate without being oxidized to sulfate (61, 62), leading to an inefficient source of energy. To circumvent this loss of energy, lithotrophic bacteria, such as the common marine uncultured SUP05 group, are thought to run enzymes used in dissimilatory sulfate reduction in reverse (63, 64). Reverse dissimilatory sulfite reductase (rDSR), adenylylsulfate reductase (*apr*AB), and sulfate adenylyltransferase (*sat*) are required to complete thiosulfate oxidation to sulfate and generate ATP (58, 65, 66). In our genomes, both *apr*AB and *sat* are almost always present (sometimes, but not always, encoded on the same contig) and have transcripts in some cases (Table 5.4). The only two genomes with *dsr*AB—Woeseia_stnF and Woeseia2_stnAC—share 88% and 85% amino acid sequence identity to the rDSR of an uncultured bacterium as shown through blastp, respectively. In addition, the *dsr*AB operon in Woeseia_stnF is located on the same contig as the *sox*AX gene, suggesting cotranscription and use of rDSR to complete sulfur oxidation in a truncated Sox pathway. However, identity and co-localization do not allow us to say with confidence that these *dsr*AB genes were in fact encoding

reverse Dsr enzymes, and questions of functionality can be answered in part by phylogenetic analysis with published databases (67). Our genomes also contained an additional means of bringing sulfide into the rDSR pathway, with sulfide-quinone oxidoreductase (*sqr*) and flavocytochrome c (*fcc*A). Together, these allow sulfide species (S^-, H_2S, HS^-) to be brought into the rDSR sulfur oxidation pathway through the initial oxidation to elemental sulfur. Sulfide-quinone oxidoreductase and flavocytochrome c were present either simultaneously or alone in our genomes (Table 5.4), presenting an extra way to circumvent potential loss of energy with a truncated Sox pathway.

Respiratory switching and nitrite reduction

All genomes had the gene for NO-forming nitrite reduction, *nir*S (cytochrome cd_1), which catalyzes the second step in denitrification in facultative anaerobes (68, 69). This NADH-dependent nitrite reductase had a systematic increase in transcript coverage with depth some genomes (Tables 5-9). Transcripts for NirS were found at the deepest intervals in our cores from both stations F (Kongsfjord) and AB (Van Keulenfjord), while genes for sulfur oxidation (*dsr*AB) were restricted to the top 2.5 cm on average. The increase with depth is likely a response mediated in part by the NirS-specific transcriptional response regulator encoded by *nar*L which is sensitive to changing respiratory conditions from oxygen to nitrate and nitrite (70, 71), found encoded in the Woeseia_stnF genome.

Woeseiacea may contribute to the high rates of denitrification that has been observed in Svalbard fjord sediment, which have benthic N_2 production similar to temperate sediments (72). However, we did not detect other denitrification genes (*nor*, *noz*) in the genomes, a feature common among denitrifyers with nearby microbial counterparts that would facilitate complete reduction to N_2 (73). An incomplete denitrification pathway was also noted in previous genomic

work on Woeseiaceae (3). Our genomes did however contain the genes encoding for the protein translocases necessary to move the cofactor-free NirS from the cytosol to the periplasm (*sec*Y and *sec*A), maturation proteins (*ccm*EFH), and a heme transporter (*ccm*A) (73, 74). Nitrite reductase does not oxidize quinones directly, but instead receives electrons from the cytochrome bc_1 complex by way of cytochrome c intermediates that are not membrane-bound (75). Each Woeseiaceae genome encoded one or more genes for cytochrome c-552, cytochrome c-554, and/or cytochrome c-555 (Table 5.4). Electrons are initially passed to these intermediates by respiratory complex I, NADH-quinone oxidoreductase, or by hydrogenases, all of which were encoded in our genomes (Figure 5.8).

Our genomes did not contain evidence for complex II genes encoding for succinate dehydrogenase, in contrast to what has been found previously in other Woeseiaceae genomes (3). However, in addition to *nuo* and *nqo* genes encoding for respiratory complex I, all genomes had two copies of *nqr*ABCDEF, encoding for Na^+-translocating-NADH-quinone reductase, as well as respiratory complex III, also called cytochrome bc_1 complex. The cytochrome bc_1 complex can be encoded either with the *fbc*H gene, or in pieces with the *pet*ABC operon (73, 76). Most of our genomes contained *fbc*H, which encodes the cytochrome b component at its 5' end and the cytochrome c_1 component at its 3' end, (77). In addition, *pet*ABC genes were also encoded in our genomes, which encodes the Rieske protein, cytochrome b and cytochrome c components separately. This shows that these genomes have multiple biogenesis pathways for key cytochromes needed for respiration.

Most genomes have an energy-conserving, Na^+ translocating protein complex encoded as *rnf* (*rnf*CEG). This protein complex can pump out Na^+ ions through the energy generated from reducing NAD^+ with ferredoxin and is commonly found in nitrogen fixers (78, 79). Although we

have no evidence of nitrogen fixation, our genomes contained evidence for ferredoxin generation and can likely couple the oxidation of ferredoxin with the generation of NADH (80, 81). Coupled reactions with ferredoxin and the RNF complex have been shown to be involved in ATP synthesis (82), reverse electron transfer (78), and reduction of both inorganic and organic electron donors in acetogens (83). The RNF complex also has a role in carbon reduction, whereby the NADH and hydrogen produced from Na^+ transport can be used to reduce CO_2 (81), providing an additional means by which carbon may be fixed.

Preference for microaerobic to anoxic conditions

The vast redox potential spanned by the cytochromes encoded in these genomes ranges from +420 mV with cbb_3-type cytochrome c (75) to as low as -500 mV with ferredoxin oxidation with RNF (83). Despite this broad range, evidence for respiration in fully oxygenated conditions is not supported. Instead, the evidence suggests that these Woeseiaceae genomes have flexible redox preferences within microaerobic to anoxic conditions. Interestingly, transcripts for NADH-quinone oxidoreductase subunits were more than three times higher in Kongsfjorden compared to Van Keulenfjorden, suggesting overall respiration activity is higher within Kongsfjorden sediment. This may be due to the differing conditions dominating site AB, located at the mouth of Van Keuelenfjord, compared to conditions at site AC of Van Keulenfjord and sites P and F in Kongsfjorden, all situated at the head of their respective fjords. Although sulfate does not become depleted with depth at any of these sites (29), site AB sees increased delivery of labile organic matter to surface sediments (29, 31) bringing about a shallower zone of iron reduction based on community structure analysis (31). However, transcripts for the cytochrome bc_1 complex were not found and transcripts for the high-redox potential cbb_3-type cytochrome c complex were found equally in both fjords, limiting our ability to interpret in a straight-forward

way the potential differential transcriptional activity of respiratory complexes as it relates to prevailing geochemical conditions.

Metabolic "lock down" and sporulation

Across both stations F and AB (where simultaneous metagenomes and metatranscriptomes were generated), the percentage of genes with corresponding transcript coverage (or the percentage of genes "turned on") varied across the Woeseia genomes, ranging from 13% (417 out of 3,110 genes in Woeseia_stnAB) to 22% (774 genes out of 3,406 in Woeseia2_stnF) (Table 5.10). Interestingly, out of the several hundred genes with transcripts across all genomes, very few had significant trends in transcriptional coverage with depth. This was surprising, as we anticipated to observe clear redox shifts in our transcripts. Instead, transcriptional coverage of genes implicated in transcription, growth, hydrolases, and vitamin transporters decreased in transcriptional coverage with increasing sediment depth (Tables 5.5-5.9; Figures 9-13). This included genes encoding for DNA-directed RNA polymerase, 50S ribosomal protein L10, 60kDa chaperonin, and elongation factor tu. Most genomes had decreasing transcript coverage for arylsulfatase (*ats*A), a hydrolase involved in the breakdown of phenol sulfates delivered to the seabed with phytoplankton blooms (86, 87). Lon protease, which is required for survival from DNA damage due to its role in the selective degradation of abnormal and/or mutant proteins (88, 89), also decreased with depth. Aging has been shown to decrease Lon protease expression, and so the depth signal we observed in transcript coverage likely is related to cell aging with burial (89). Stress mitigation pathways, such as DNA-binding protein HU and rubrerythrin, increased with depth, suggesting that these aging cells are primed for dealing with sub-optimal conditions deeper in the sediment.

Most notably, regression analysis showed that nearly all genomes had considerably high positive slopes for spore protein SP21 at both stations AB and F (Figures 9-13, Tables 5-9). The occurrence of highly transcribed spore proteins in our other reconstructed (non-Woeseiaceae) genomes was not observed and additional network analysis does not indicate significant correlations with other taxa (Appendix VI). The key spore protein transcribed in our genomes was SP21 (*hsp*A), which is responsible for fruiting body development in Myxobacteria (90, 91). Spores of Myxobacteria, called myxospores, develop under nutrient-limited conditions as a means for persisting until nutrients once again become plentiful (92, 93). SP21 expression has also been observed to occur after oxygen depletion, although it was not clear if oxygen depletion was the trigger for expression, or if it was more directly tied to metabolic slow-down. As our transcript evidence demonstrated the decreased expression of genes related to growth with depth, the expression of SP21 in Woseiaceae is likely tied more directly to metabolic slow-down rather than oxygen depletion. These myxospores are separate from the diverse array of sulfate reducing Firmicutes that endospore germination studies have shown to be active when induced under high temperatures in Svalbard fjord sediment (94, 95). The ability to form spores was not observed in the culture *Woeseia oceani* (6) and the potential for spore formation in this clade has not been noted previously. To determine if this feature is unique to our genomes, we performed pangenomic analysis with the other Woeseiaceae genomes in the NCBI database. SP21 genes are not enriched in our genomes compared to others (see Appendix V for detailed discussion of pangenomic and enriched functions). The formation of myxospores in these genomes is evidence that Woeseiaceae have a mechanism for persisting though conditions that are not optimal for growth until they become exposed once again to nutrients and preferable redox conditions, perhaps resulting from an episode of bioturbation. This provides to them an avenue for

continuing their populations through re-seeding and is perhaps a dominant control on their ubiquitous global distribution and high abundance.

Conclusion

Genomic content suggested that Woeseiaceae/JTB255 are fine-tuned for microaerobic to anoxic conditions and can serve primary producers in carbon-limited Arctic sediment environments. In addition to coupling sulfur oxidation to the fixation of inorganic carbon, transcriptional evidence demonstrated that these Woeseiaceae switch their respiratory metabolism to denitrification as they become buried within the sediment. Continued burial results in increased transcription of stress-mitigation proteins and spore-forming proteins. Together, these may prepare the cell for a period of dormancy while it continues to generate nitric oxide at depth. By sequencing at 1 cm depth resolution, we were able to capture nuanced changes that highlight the delicate interplay between redox conditions and transcriptional activity of redox-sensitive enzymes. Climate change in the Arctic threatens to remove the direct conduit of oxidized terrestrial material that allows the deeper penetration of highly oxidized conditions in fjord sediment. Glacial retreat therefore threatens to alter the sediment ecosystem for which these genomes are optimized, and although they have a mechanism for re-seeding their population, it is uncertain if conditions will remain optimal for re-establishment in the shallow subsurface.

References

1. Middelburg JJ. 2011. Chemoautotrophy in the ocean. Geophysical Research Letters 38.
2. Dyksma S, Bischof K, Fuchs BM, Hoffmann K, Meier D, Meyerdierks A, Pjevac P, Probandt D, Richter M, Stepanauskas R. 2016. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. The ISME journal 10:1939.
3. Mußmann M, Pjevac P, Krüger K, Dyksma S. 2017. Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. The ISME journal 11:1276.
4. Bowman JP, McCuaig RD. 2003. Biodiversity, Community Structural Shifts, and Biogeography of Prokaryotes within Antarctic Continental Shelf Sediment. Applied and Environmental Microbiology 69:2463-2483.
5. Teske A, Durbin A, Ziervogel K, Cox C, Arnosti C. 2011. Microbial community composition and function in permanently cold seawater and sediments from an Arctic fjord of Svalbard. Applied and Environmental Microbiology 77:2008-2018.
6. Du Z-J, Wang Z-J, Zhao J-X, Chen G-J. 2016. Woeseia oceani gen. nov., sp. nov., a chemoheterotrophic member of the order Chromatiales, and proposal of Woeseiaceae fam. nov. International journal of systematic and evolutionary microbiology 66:107-112.
7. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics:btu170.
8. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075.
9. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108:4516-4522.

10. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and environmental microbiology 75:7537-7541.

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

12. Edgar RC. 2018. Updating the 97% identity threshold for 16S ribosomal RNA OTUs. Bioinformatics 1:5.

13. Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2016. metaSPAdes: a new versatile de novo metagenomics assembler. arXiv preprint arXiv:160403071.

14. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: The United States Department of Energy Systems Biology Knowledgebase. Nature Biotechnology 36:566.

15. Wu Y-W, Simmons BA, Singer SW. 2015. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-607.

16. Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nature methods 11:1144.

17. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165.

18. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2017. Recovery of genomes from metagenomes via a dereplication, aggregation, and scoring strategy. bioRxiv doi:10.1101/107789.

19. 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 research 25:1043-1055.

20. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy T, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nature biotechnology 35:725.

21. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one 5:e9490.

22. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319.

23. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular biology and evolution:msw054.

24. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics:btu153.

25. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28:3211-3217.

26. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359.

27. RCoreTeam. 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2013.

28. Hoehler TM, Alperin MJ, Albert DB, Martens CS. 1994. Field and laboratory studies of methane oxidation in an anoxic marine sediment: Evidence for a methanogen-sulfate reducer consortium. Global Biogeochemical Cycles 8:451-463.

29. Wehrmann LM, Formolo MJ, Owens JD, Raiswell R, Ferdelman TG, Riedinger N, Lyons TW. 2014. Iron and manganese speciation and cycling in glacially influenced high-latitude fjord sediments (West Spitsbergen, Svalbard): Evidence for a benthic recycling-transport mechanism. Geochim Cosmochim Acta 141:628-655.

30. Wehrmann LM, Riedinger N, Brunner B, Kamyshny Jr A, Hubert CR, Herbert LC, Brüchert V, Jørgensen BB, Ferdelman TG, Formolo MJ. 2017. Iron-controlled oxidative sulfur cycling recorded in the distribution and isotopic composition of sulfur species in glacially influenced fjord sediments of west Svalbard. Chemical Geology 466:678-695.

31. Buongiorno J, Lisa Herbert, Laura Wehrmann, Alex Michaud, Katja Laufer, Hans Røy, Bo Barker Jørgensen, Anna Szynkiewicz, Anthony Faiia, Kevin Yeager, Kimberly Schindler, and Karen Lloyd. Under revision. Cross-fjord trends of complex microbial communities control subsurface iron and sulfur cycling in Arctic sediments.

32. McMahon KW, Ambrose Jr WG, Johnson BJ, Sun M-Y, Lopez GR, Clough LM, Carroll ML. 2006. Benthic community response to ice algae and phytoplankton in Ny Ålesund, Svalbard. Marine Ecology Progress Series 310:1-14.

33. Søreide JE, Carroll ML, Hop H, Ambrose WG, Hegseth EN, Falk-Petersen S. 2013. Sympagic-pelagic-benthic coupling in Arctic and Atlantic waters around Svalbard revealed by stable isotopic and fatty acid tracers. Marine Biology Research 9:831-850.

34. Schubert CJ, Calvert SE. 2001. Nitrogen and carbon isotopic composition of marine and terrestrial organic matter in Arctic Ocean sediments:: implications for nutrient utilization and organic matter composition. Deep Sea Research Part I: Oceanographic Research Papers 48:789-810.

35. Kim J-H, Peterse F, Willmott V, Kristensen DK, Baas M, Schouten S, Sinninghe Damsté J. 2011. Large ancient organic matter contributions to Arctic marine sediments (Svalbard). Limnology and Oceanography 56:1463-1474.

36. Morono Y, Terada T, Nishizawa M, Ito M, Hillion F, Takahata N, Sano Y, Inagaki F. 2011. Carbon and nitrogen assimilation in deep subseafloor microbial cells. Proceedings of the National Academy of Sciences:201107763.

37. Jørgensen BB. 2011. Deep subseafloor microbial cells on physiological standby. PNAS 108:18193-18194.

38. Price PB, Sowers T. 2004. Temperature dependence of metabolic rates for microbial growth, maintenance, and survival. PNAS 101:4631-4636.

39. Hagemann M, Eisenhut M, Hackenberg C, Bauwe H. Pathway and Importance of Photorespiratory 2-Phosphoglycolate Metabolism in Cyanobacteria, p 91-108. In Hallenbeck PC (ed), Springer New York,

40. Hagemann M, Bauwe H. 2016. Photorespiration and the potential to improve photosynthesis. Current Opinion in Chemical Biology 35:109-116.

41. Norman EG, Colman B. 1991. Purification and Characterization of Phosphoglycolate Phosphatase from the Cyanobacterium *Coccochloris peniocystis*. Plant Physiology 95:693-698.

42. Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and

structure/function relationships. Journal of experimental botany 59:1515-1524.

43. 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. Microbiology and Molecular Biology Reviews 71:576-599.

44. Hernandez JM, Baker SH, Lorbach SC, Shively JM, Tabita FR. 1996. Deduced amino acid sequence, functional expression, and unique enzymatic properties of the form I and form II ribulose bisphosphate carboxylase/oxygenase from the chemoautotrophic bacterium *Thiobacillus denitrificans*. Journal of bacteriology 178:347-356.

45. Tabita FR. 1999. Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: a different perspective. Photosynthesis Research 60:1-28.

46. Badger MR, Price GD. 1992. The CO2concentrating mechanism in cyanobactiria and microalgae. Physiologia Plantarum 84:606-615.

47. Kaplan A, Reinhold L. 1999. CO2 concentrating mechanisms in photosynthetic microorganisms. Annual review of plant biology 50:539-570.

48. Watson GM, Tabita FR. 1997. Microbial ribulose 1, 5-bisphosphate carboxylase/oxygenase: a molecule for phylogenetic and enzymological investigation. FEMS microbiology letters 146:13-22.

49. Hulth S, Blackburn TH, Hall PO. 1994. Arctic sediments (Svalbard): consumption and microdistribution of oxygen. Marine Chemistry 46:293-316.

50. Jørgensen BB, Glud RN, Holby O. 2005. Oxygen distribution and bioirrigation in Arctic fjord sediments (Svalbard, Barents Sea). Marine Ecology Progress Series 292:85-95.

51. Liu S, Du M-Z, Wen Q-F, Kang J, Dong C, Xiong L, Huang J, Guo F-B. 2018. Comprehensive exploration of the enzymes catalysing oxygen-involved reactions and COGs

relevant to bacterial oxygen utilization. Environmental Microbiology.

52. Hendler RW, Sidhu GS. 1988. A new high potential redox transition for cytochrome aa3. Biophysical journal 54:121-133.

53. Toledo-Cuevas M, Barquera B, Gennis RB, Wikström M, García-Horsman JA. 1998. The cbb3-type cytochrome c oxidase from Rhodobacter sphaeroides, a proton-pumping heme-copper oxidase. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1365:421-434.

54. Preisig O, Anthamatten D, Hennecke H. 1993. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. Proceedings of the National Academy of Sciences 90:3309-3313.

55. Tamazawa S, Takasaki K, Tamaki H, Kamagata Y, Hanada S. 2012. Metagenomic and biochemical characterizations of sulfur oxidation metabolism in uncultured large sausage-shaped bacterium in hot spring microbial mats. PloS one 7:e49793.

56. Gunsalus R, Park S-J. 1994. Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. Research in microbiology 145:437-450.

57. Alvarez AF, Georgellis D. 2010. Chapter 12 - In Vitro and In Vivo Analysis of the ArcB/A Redox Signaling Pathway, p 205-228, Methods in Enzymology, vol 471. Academic Press.

58. Klatt JM, Polerecky L. 2015. Assessment of the stoichiometry and efficiency of CO(2) fixation coupled to reduced sulfur oxidation. Frontiers in Microbiology 6:484.

59. Friedrich CG, Bardischewsky F, Rother D, Quentmeier A, Fischer J. 2005. Prokaryotic sulfur oxidation. Current Opinion in Microbiology 8:253-259.

60. Rother D, Henrich H-J, Quentmeier A, Bardischewsky F, Friedrich CG. 2001. Novel genes of the sox gene cluster, mutagenesis of the flavoprotein SoxF, and evidence for a general sulfur-oxidizing system in Paracoccus pantotrophusGB17. Journal of bacteriology 183:4499-4508.

61. Ghosh W, Dam B. 2009. Biochemistry and molecular biology of lithotrophic sulfur oxidation by taxonomically and ecologically diverse bacteria and archaea. FEMS microbiology reviews 33:999-1043.

62. Grimm F, Franz B, Dahl C. 2008. Thiosulfate and sulfur oxidation in purple sulfur bacteria, p 101-116, Microbial Sulfur Metabolism. Springer.

63. Murillo AA, Ramírez-Flandes S, DeLong EF, Ulloa O. 2014. Enhanced metabolic versatility of planktonic sulfur-oxidizing γ-proteobacteria in an oxygen-deficient coastal ecosystem. Frontiers in Marine Science 1:18.

64. Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG, Tortell PD, Hallam SJ. 2009. Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. Science 326:578-582.

65. Harada M, Yoshida T, Kuwahara H, Shimamura S, Takaki Y, Kato C, Miwa T, Miyake H, Maruyama T. 2009. Expression of genes for sulfur oxidation in the intracellular chemoautotrophic symbiont of the deep-sea bivalve Calyptogena okutanii. Extremophiles 13:895.

66. Stewart F, Dmytrenko O, DeLong E, Cavanaugh C. 2011. Metatranscriptomic analysis of sulfur oxidation genes in the endosymbiont of *Solemya velum*. Frontiers in microbiology 2:134.

67. Loy A, Duller S, Baranyi C, Mußmann M, Ott J, Sharon I, Béjà O, Le Paslier D, Dahl C, Wagner M. 2009. Reverse dissimilatory sulfite reductase as phylogenetic marker for a subgroup of sulfur-oxidizing prokaryotes. Environmental microbiology 11:289-299.

68. Nicke T, Schnitzer T, Münch K, Adamczack J, Haufschildt K, Buchmeier S, Kucklick M, Felgenträger U, Jänsch L, Riedel K. 2013. Maturation of the cytochrome cd1 nitrite reductase NirS from Pseudomonas aeruginosa requires transient interactions between the three proteins NirS, NirN and NirF. Bioscience reports 33:e00048.

69. Beller HR, Letain TE, Chakicherla A, Kane SR, Legler TC, Coleman MA. 2006. Whole-genome transcriptional analysis of chemolithoautotrophic thiosulfate oxidation by T*hiobacillus denitrificans* under aerobic versus denitrifying conditions. Journal of bacteriology 188:7005-7015.

70. Härtig E, Schiek U, Vollack K-U, Zumft WG. 1999. Nitrate and Nitrite Control of Respiratory Nitrate Reduction in Denitrifying *Pseudomonas stutzeri* by a Two-Component Regulatory System Homologous to NarXL of *Escherichia coli*. Journal of bacteriology 181:3658-3665.

71. Lin AV, Stewart V. 2010. Functional roles for the GerE-family carboxyl-terminal domains of nitrate response regulators NarL and NarP of *Escherichia coli* K-12. Microbiology 156:2933-2943.

72. Gihring TM, Lavik G, Kuypers MMM, Kostka JE. 2010. Direct determination of nitrogen cycling rates and pathways in Arctic fjord sediments (Svalbard, Norway). Limnology and Oceanography 55:740-752.

73. Graf DR, Jones CM, Hallin S. 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N2O emissions. PloS one 9:e114118.

74. Thöny-Meyer L. 1997. Biogenesis of respiratory cytochromes in bacteria. Microbiology and Molecular Biology Reviews 61:337-376.

75. Thöny-Meyer L, Künzler P. 1997. Translocation to the periplasm and signal sequence cleavage of preapocytochrome c depend on sec and lep, but not on the ccm gene products. European journal of biochemistry 246:794-799.

76. Trumpower BL. 1990. Cytochrome bc1 complexes of microorganisms. Microbiological reviews 54:101-129.

77. Chen YL, Dincturk HB, Qin H, Knaff DB. 1998. The pet operon, encoding the prosthetic group-containing subunits of the cytochrome bc1 complex, of the purple sulfur bacterium *Chromatium vinosum*. Photosynthesis research 57:139-158.

78. Thöny-Meyer L, Stax D, Hennecke H. 1989. An unusual gene cluster for the cytochrome bc1 complex in *Bradyrhizobium japonicum* and its requirement for effective root nodule symbiosis. Cell 57:683-697.

79. Kumagai H, Fujiwara T, Matsubara H, Saeki K. 1997. Membrane localization, topology, and mutual stabilization of the rnfABC gene products in *Rhodobacter capsulatus* and implications for a new family of energy-coupling NADH oxidoreductases. Biochemistry 36:5509-5521.

80. Saeki K, Kumagai H. 1998. The rnf gene products in Rhodobacter capsulatus play an essential role in nitrogen fixation during anaerobic DMSO-dependent growth in the dark. Archives of microbiology 169:464-467.

81. Biegel E, Schmidt S, González JM, Müller V. 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cellular and molecular life sciences 68:613-634.

82. Biegel E, Schmidt S, Müller V. 2009. Genetic, immunological and biochemical evidence for a Rnf complex in the acetogen *Acetobacterium woodii*. Environmental microbiology 11:1438-1443.

83. Hess V, Schuchmann K, Müller V. 2013. The ferredoxin: NAD+ oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na+ and is reversibly coupled to the membrane potential. Journal of Biological Chemistry:jbc. M113. 510255.

84. Müller V, Imkamp F, Biegel E, Schmidt S, Dilling S. 2008. Discovery of a ferredoxin: NAD+-oxidoreductase (Rnf) in Acetobacterium woodii: a novel potential coupling site in acetogens. Annals of the New York Academy of Sciences 1125:137-146.

85. Oshrain RL, Wiebe WJ. 1979. Arylsulfatase Activity in Salt Marsh Soils. Applied and Environmental Microbiology 38:337-340.

86. Meyer-Reil L-A. 1987. Seasonal and Spatial Distribution of Extracellular Enzymatic Activities and Microbial Incorporation of Dissolved Organic Substrates in Marine Sediments. Applied and Environmental Microbiology 53:1748-1755.

87. Kuroda A, Nomura K, Ohtomo R, Kato J, Ikeda T, Takiguchi N, Ohtake H, Kornberg A. 2001. Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in *E. coli*. Science 293:705-708.

88. Bota DA, Van Remmen H, Davies KJ. 2002. Modulation of Lon protease activity and aconitase turnover during aging and oxidative stress. FEBS letters 532:103-106.

89. Wireman JW, Dworkin M. 1977. Developmentally induced autolysis during fruiting body formation by *Myxococcus xanthus*. Journal of bacteriology 129:798-802.

90. Lünsdorf H, Schairer HU, Heidelbach M. 1995. Localization of the stress protein SP21 in indole-induced spores, fruiting bodies, and heat-shocked cells of *Stigmatella aurantiaca*. Journal of Bacteriology 177:7092-7099.

91. Heidelbach M, Skladny H, Schairer HU. 1993. Purification and characterization of SP21, a development-specific protein of the myxobacterium *Stigmatella aurantiaca*. Journal of bacteriology 175:905-908.

92. Heidelbach M, Skladny H, Schairer HU. 1993. Heat shock and development induce synthesis of a low-molecular-weight stress-responsive protein in the myxobacterium *Stigmatella aurantiaca*. Journal of bacteriology 175:7479-7482.

93. Hubert C, Loy A, Nickel M, Arnosti C, Baranyi C, Brüchert V, Ferdelman T, Finster K, Christensen FM, De Rezende JR. 2009. A constant flux of diverse thermophilic bacteria into the cold Arctic seabed. Science 325:1541-1544.

94. Hubert C, Arnosti C, Brüchert V, Loy A, Vandieken V, Jørgensen BB. 2010. Thermophilic anaerobes in Arctic marine sediments induced to mineralize complex organic matter at high temperature. Environmental microbiology 12:1089-1104.

Appendix I: Tables and Figures

Table 5.1. Assembly statistics, including the number of contigs, the largest contig, the N50 (the length for which the collection of all contigs of that length or longer covers at least half an assembly), and L50 (the number of contigs equal to or longer than N50, aka the minimal number of contigs that cover half the assembly).

Metagenome	Assembler	No. contigs > 1000 bp	Largest contig (bp)	N50	L50
AC_18cmbsf	metaSPAdes	47,245	161,849	3,492	13,636
AC_18cmbsf	Megahit	51,025	208,964	3,385	15,352
AC_18cmbsf	IDBA	142,046	158,899	1,408	76,848
AB_Lloydlab_0-5cmbsf	metaSPAdes	36,489	92,139	4,982	9,188
AB_Lloydlab_0-5cmbsf	Megahit	532,416	52,817	1,676	167,575
AB_Lloydlab_0-5cmbsf	IDBA	167,814	72,311	2,225	46,030
F_Lloydlab_0-5cmbsf	metaSPAdes	6,714	188,610	7,634	7,634
F_Lloydlab_0-5cmbsf	Megahit	255,297	77,432	1,535	1,199
F_Lloydlab_0-5cmbsf	IDBA	13,731	103,160	3,999	3,422

Table 5.2. Statistics for all MAGs. Woeseiaceae MAGs are in bold print. Completeness and contamination from CheckM are used to determine genome draft quality according to Bowers et al., (2017).

MAG name	Length (bp)	Number contigs	Compl.	Contam.	Strain heterogeneity	Quality
Acidimicrobiia_stnAC	1,169,838	291	64.85	1.28	0	Medium
Acidobacteria_stnAC	3,608,277	144	95.73	5.98	0	Medium
Akkermansiaceae_stnF	2,887,821	224	98.34	0.14	100	High
Akkermansiaceae2_stnF	3,011,311	738	96.42	1.36	0	High
Anaerolineae_stnAC	1,682,520	309	72.12	7.9	26.67	Medium
Anaerolineae_stnAC	1,423,899	176	54.55	3.18	14.29	Medium
Chromatiales_stnAB	2,205,383	297	85.61	8.69	4	Medium
Dadabacteria_stnAB	2,092,273	157	92.35	2.57	0	High
Desulfobacteraceae_stnAC	3,349,224	623	86.8	0.04	0	Medium
Desulfobacterales_stnF	2,907,775	192	53.82	0.07	0	Medium
Desulfosarcina_stnAB	1,156,346	321	83.74	0.56	66.67	Medium
Flavobacteria_stnAB	3,507,500	605	80.02	5.09	14.29	Medium
Gemmatimonadetes_stnAB	1,446,593	371	60.29	0	0	Medium
Gemmatimonadetes_stnAC	1,833,525	420	76.64	3.27	50	Medium
Geobacter_stnAC	2,846,301	514	70.17	2.5	25	Medium
Halieaceae_stnAB	1,382,038	391	71.35	7.12	12.12	Medium
Labilibaculum_stnF	1,309,716	349	54.54	3.66	5	Medium
Latescibacteria_stnAB	3,013,289	327	55.7	0.68	0	Medium
Lentisphaerae_stnAB	1,903,818	262	60.74	4.04	5.56	Medium
Lentisphaerae2_stnAB	1,272,728	212	92.91	5.1	0	Medium
Lentisphaerae3_stnAB	1,019,905	223	80.95	2.94	0	Medium
Lentisphaerae4_stnAB	1,996,229	354	64.78	0	0	Medium

Table 5.2 continued.

MAG name	Length (bp)	Number contigs	Compl.	Contam.	Strain heterogeneity	Quality
Lentisphaerae5_stnAB	1,678,967	398	54.01	0.98	56.25	Medium
Lentisphaerae6_stnAB	2,656,030	555	75.55	1.77	23.81	Medium
Myxococcales_stnAB	1,699,629	288	67.09	4.52	14.29	Medium
Nitrosomonadaceae_stnAB	1,907,502	429	77.51	1.54	16.67	Medium
Nitrospina_stnAB	1,344,434	295	74.01	6.64	15.38	Medium
Nitrospina_stnAC	1,920,139	407	58.38	0	0	Medium
Nitrospinae_stnAC	778,268	195	66.13	2.1	0	Medium
Nitrospiraceae_stnAC	895,756	231	57.68	1.21	50	Medium
Nitrospirae_stnAC	3,738,793	535	55.21	0.06	100	Medium
Olavius_Gamma_stnAC	2,609,566	576	80.24	4.72	42.11	Medium
Olavius_Gamma2_stnAC	1,237,209	328	90.07	7.95	26.53	Medium
Phyciphaerae_stnF	3,229,210	615	79.59	6.92	30	Medium
Phycisphaerales_stnAC	1,235,250	306	82.95	1.22	25	Medium
Planctomycetales_stnAC	2,902,147	460	53.42	4.55	0	Medium
Rhodobacterales_stnAB	1,792,237	415	63.2	1.98	50	Medium
Scalindua_stnAC	904,834	229	51.02	4.65	60	Medium
Syntrophaceae_stnAB	1,492,049	323	66.08	4.48	44.44	Medium
Syntrophaceae_stnF	602,754	166	71.56	3.58	57.14	Medium
Thiohalomonas_stnAC	1,425,011	294	68.71	4.28	53.57	Medium
Thiotrichaceae_stnAB	1,757,957	383	57.97	2.34	9.09	Medium
Thiotrichaceae_stnAC	2,156,343	305	89.45	3.62	3.7	Medium
Verrucomicrobia_stnAB	2,142,440	341	79.39	5.1	0	Medium
Woeseia_stnAB	2,732,782	388	83.46	6.31	28.57	Medium
Woeseia_stnAC	3,122,991	331	80.8	9.52	18.64	Medium
Woeseia2_stnAC	4,237,456	684	80.09	4.66	54.55	Medium
Woeseia_stnF	2,489,600	421	94.83	7.59	50	Medium
Woeseia2_stnF	3,504,067	339	90.79	2.97	22.22	High

Table 5.3. Genome statistics for publicly available Woeseiaceae genomes and the genomes in this study. Completeness and contamination determined for each genome with CheckM.

Genome name	Accession	Sampe source	Size (mbp)	No contigs	Completeness	Contamination
WOR SG8 31[a]	NCBI LJTI00000000	Estuarine sediment	5.9	266	100	97
JSS Woes1[b]	IMG 2695420981	Tidal sediment	8.1	607	91	89
20 j1[b]	IMG 2651869885	Tidal sediment	2.4	298	44	6
SAG 1868 B[b]	IMG 2651869504	Tidal sediment	2.2	358	51	0.4
Woeseia_oceani XK5[c]	NCBI NZ_CP016268	Culture from coastal sediment	4.0	1	91	2
Woeseia_stnAB[d]	IMG 2802428844	Arctic sediment	2.7	388	83	6
Woeseia_stnAC[d]	IMG 2802428845	Arctic sediment	3.1	331	80	9
Woeseia2_stnAC[d]	IMG 2802428847	Arctic sediment	4.2	684	80	4
Woeseia_stnF[d]	IMG 2802428846	Arctic sediment	2.4	421	94	7
Woeseia2_stnF[d*]	IMG 2802428848	Arctic sediment	3.5	339	90	2

Table 5.4. Genome content and transcription of genes for sulfur oxidation, carbon fixation, and denitrification in the Woseiaceae MAGs.

Gene or cytochrome name	Number of genomes	Number of genomes with transcripts in KF	Number of genomes with transcripts in VK
*sox*A	5	0	0
*sox*B	3	0	0
*sox*C	1	0	0
*sox*D	1	0	0
*sox*X	4	0	0
*sox*Y	4	0	0
*sox*Z	4	0	0
Cytochrome c-555	5	0	0
Cytochrome c-554(548)	2	0	0
Cytochrome c-552	2	0	1
*cco*N1	4	1	2
*cco*O	4	1	1
*cco*P	4	1	1
*fcc*A	3	2	1
sqr	3	0	0
Sat	5	0	0
*apr*B	4	1	2
*apr*A	5	1	2
*dsr*A	2	1	1
*dsr*B	2	1	1
*psr*A	5	0	0
*fdh*A	5	2	3
*fdh*C	5	0	0
*fdh*D	5	1	0
*fdh*F	5	1	0
*cbb*M	5	1	2
gap	3	1	1
pgk	5	0	0
*cfx*P	5	0	0
*nir*S	5	3	2
*sec*A	5	0	0
*sesc*Y	5	0	0
NADH oxidoreductase subunits of *nuo*, *nqr*, and *nqo*	5	5	5
*fbc*H	4	0	0
*pet*A	4	0	0

Table 5.4 continued.

Gene or cytochrome name	Number of genomes	Number of genomes with transcripts in KF	Number of genomes with transcripts in VK
*pet*B	5	0	0
*pet*C	5	0	0
*ccm*A	4	0	0
*ccm*E	3	0	0
*ccm*F	4	0	0
*ccm*H	4	0	0

Table 5.5.Woeseia_stnAB regression statistics for transcripts with depth in alphabetical order.

Gene_name_gene_id	Slope	Intercept	R^2	Station
60 kDa chaperonin_5902	-0.037	0.265	0.954	AB
Colicin I receptor_5875	-0.175	0.664	0.803	AB
None_5471	-0.03	0.249	0.23	AB
None_6158	-0.017	0.076	0.553	F
None_6257	0.046	0.029	0.88	AB
None_6260	-0.111	0.635	0.733	AB
None_6325	-0.043	0.544	0.656	F
None_7593	-0.1	0.391	0.995	AB
None_7607	0.096	0.334	0.822	AB
None_7630	-0.01	0.084	0.927	F
None_7796	7.684	3.28	0.919	F
None_7796	2.433	3.993	0.131	AB
Pesticin receptor_5933	-0.007	0.149	0.162	F
RNA polymerase-binding transcription factor DksA_5868	0.039	0.125	0.387	F
Spore protein SP21_5904	13.883	-5.956	0.908	F
Spore protein SP21_5904	12.054	-6.023	0.691	AB

Table 5.6.Woeseia_stnAC regression statistics for transcripts with depth in alphabetical order.

Gene_name_gene_id	Slope	Intercept	R^2	Station
60 kDa chaperonin_89881	0.023	0.157	0.328	F
None_87630	0.017	0.039	0.981	AB
None_88605	-0.091	0.982	0.971	F
None_88605	-0.051	0.509	0.232	AB
None_89268	-0.069	0.981	0.232	F
None_89268	-0.579	2.293	0.696	AB
Spore protein SP21_89878	28.051	-21.52	0.91	F
Spore protein SP21_89878	12.986	-6.037	0.544	AB

Table 5.7.Woeseia_stnF regression statistics for transcripts with depth in alphabetical order.

Gene_name_gene_id	Slope	Intercept	R^2	Station
50S ribosomal protein L10_107515	0.255	0.083	0.995	AB
60 kDa chaperonin_108179	-0.091	0.419	0.418	AB
Arylsulfatase_110079	-0.024	0.377	0.37	F
DNA topoisomerase 1_107545	-0.005	0.056	0.486	F
Elongation factor Tu_107510	0.08	0.074	0.528	AB
Elongation factor Tu_107522	-0.025	0.357	0.29	AB
Isocitrate lyase_107433	0.025	0.042	0.999	F
Lon protease_109799	-0.009	0.133	0.968	F
Methylmalonate semialdehyde dehydrogenase [acylating]_109966	0.041	-0.001	0.928	AB
Nitrite reductase_108935	0.041	0.131	0.934	F
Nitrite reductase_108935	0.038	0.209	0.255	AB
None_108374	-0.15	0.957	0.97	F
None_108905	-0.08	0.757	0.251	F
None_109691	0.338	-0.651	0.699	F
None_109809	1.991	0.571	0.713	F
None_109809	1.06	-0.181	0.357	AB
None_109908	-0.048	0.8	0.301	F
None_110192	-0.106	0.38	0.973	AB
None_111027	-0.087	0.553	0.47	AB
None_111104	-0.041	0.715	0.858	F
Rubrerythrin_108041	0.08	0.606	0.779	F
Spore protein SP21_108144	40.521	100.458	0.456	F
Spore protein SP21_108144	55.331	63.789	0.183	AB
Spore protein SP21_109810	65.103	48.737	0.594	F
Spore protein SP21_109810	35.296	10.432	0.282	AB
Spore protein SP21_110867	16.249	-17.732	0.917	F
Spore protein SP21_110867	10.773	-14.451	0.979	AB
Uptake hydrogenase large subunit_107812	-0.026	0.186	0.585	F
Vitamin B12 transporter BtuB_107397	-0.015	0.113	0.514	F
Vitamin B12-dependent ribonucleoside-diphosphate reductase_107318	-0.023	0.152	0.409	AB

Table 5.8.Woeseia2_stnAC regression statistics for transcripts with depth in alphabetical order.

Gene_name_gene_id	Slope	Intercept	R^2	Station
CoB--CoM heterodisulfide reductase iron-sulfur subunit D_64380	0.002	0.18	0.481	F
DNA-binding protein HU_64400	0.207	0.521	0.536	F
DNA-directed RNA polymerase subunit beta'_63721	-0.01	0.047	0.959	F
DNA-directed RNA polymerase subunit beta'_63721	-0.02	0.09	0.643	AB
Lon protease_64401	-0.025	0.239	0.154	F
Nitrite reductase_64141	0.031	0.118	0.982	F
None_63420	0.02	0.085	0.254	F
None_64185	43.254	-7.19	0.854	F
Spore protein SP21_65543	10.681	3.374	0.907	F
Spore protein SP21_65543	4.411	4.325	0.093	AB

Table 5.9.Woeseia2_stnF regression statistics for transcripts with depth in alphabetical order.

Gene_name_gene_id	Slope	Intercept	R^2	Station
60 kDa chaperonin_103085	-0.08	0.529	0.909	F
Arylsulfatase_103780	-0.127	0.627	0.879	AB
DNA-directed RNA polymerase subunit beta'_102956	-0.016	0.074	0.436	AB
DNA-directed RNA polymerase subunit beta_102957	0.025	0.046	0.999	AB
Fimbrial protein_102898	-0.188	1.039	0.822	AB
None_102101	-0.233	0.924	0.872	F
None_102610	-0.202	1.454	0.13	AB
None_103289	-0.231	0.804	0.971	AB
None_104127	0.173	-0.137	0.277	AB
Protein translocase subunit SecY_101612	0.028	0.084	0.691	AB
RNA polymerase-binding transcription factor DksA_101664	0.009	0.141	0.119	F

Table 5.10.Transcript statistics at stations AB and F combined for each Woeseia genome.

Genome	Number of genes[a]	Number of genes with transcript coverage[b]	Percentage of genome "turned on"
Woeseia_stnAB	3,110	417	13%
Woeseia_stnAC	2,603	417	16%
Woeseia2_stnAC	329	2,819	12%
Woeseia_stnF	863	4,370	19%
Woeseia2_stnF	774	3,406	22%

[a] Number of genes determined through counting annotations in Prokka output.

[b] Number of genes with transcripts determined through manual inspection of coverage files output by Anvio.

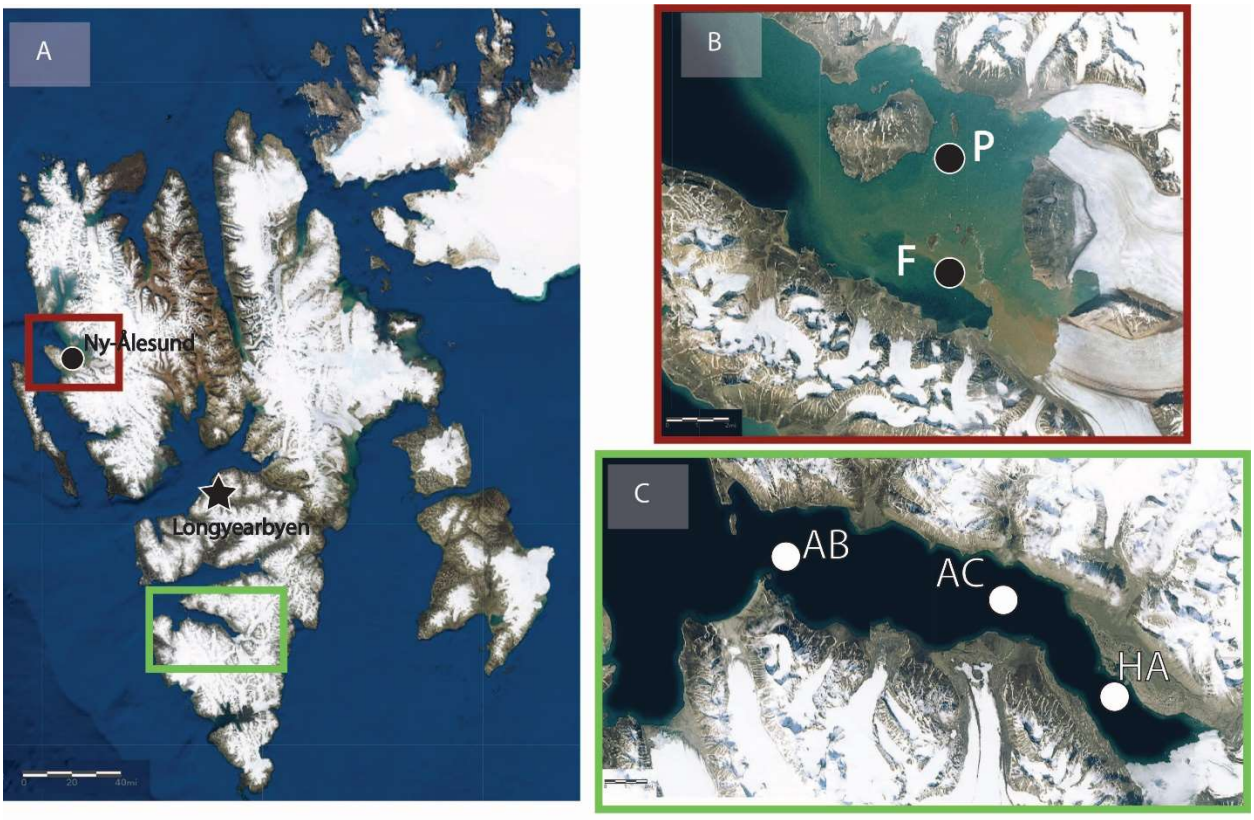

Figure 5.1. Map of field areas in Spitzbergen (A). Red box in overview map indicates Kongsfjorden (B) and green box indicates Van Keulenfjorden (C). Images taken from USGS Land Look web server (https://landlook.usgs.gov/viewer.html).

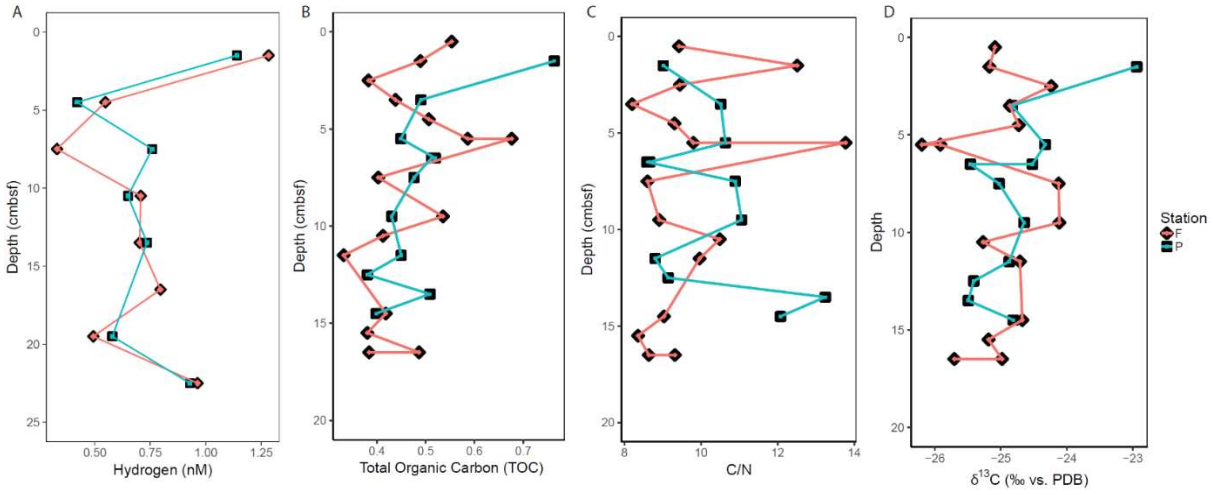

Figure 5.2. Geochemistry results. Results for hydrogen (A), total organic carbon (B), C/N (C), and carbon isotopes in bulk organic matter (D) are shown for Stations F in salmon and P in teal in Kongsfjorden. Measurements for Van Keulenfjorden are reported in the previous chapter. Note y-axis is not shared between A and B-D.

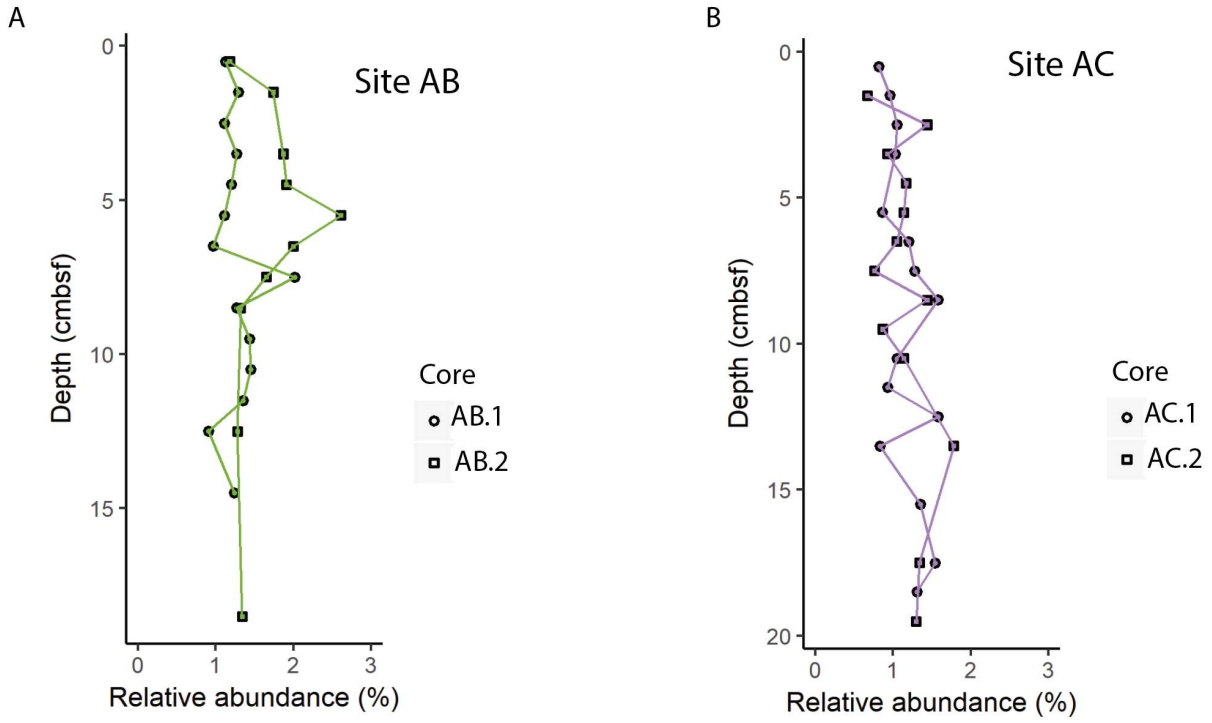

Figure 5.3. Relative 16s rRNA gene amplicon sequence abundance for Woeseiaceae in outer Station AB (A) and middle Station AC (B). Results from two different cores are shown.

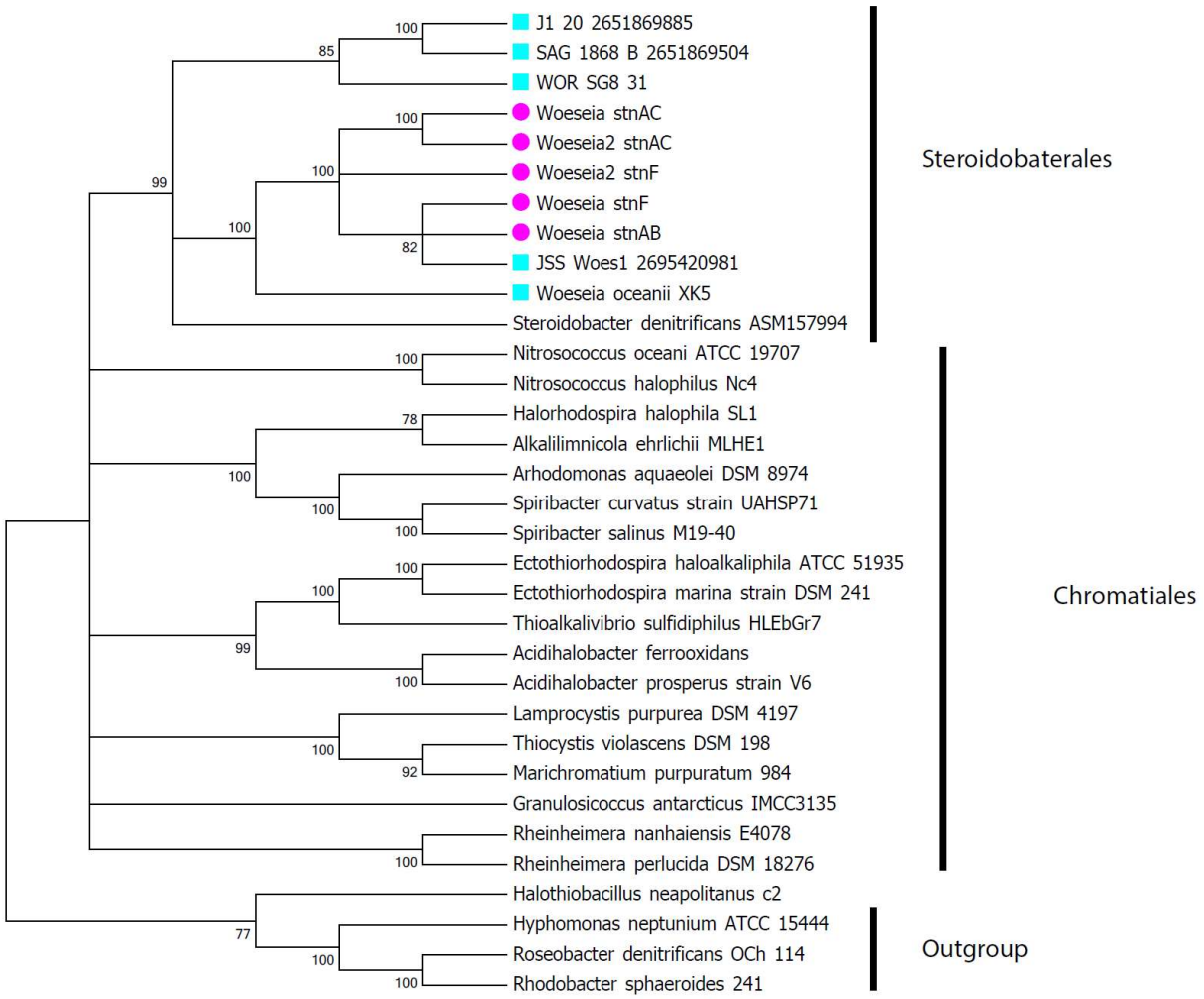

Figure 5.4. Phylogenetic tree for Woeseiaceae. Maximum likelihood was calculated in Mega v. 7 with 1000 bootstraps on a concatenated alignment of 49 ribosomal proteins. Only nodes with >75% support are shown. Genomes from the database are indicated with teal squares and MAGs from this study are indicated with magenta circles. Details about Woeseiaceae genomes included in this analysis can be found in Table 5.3.

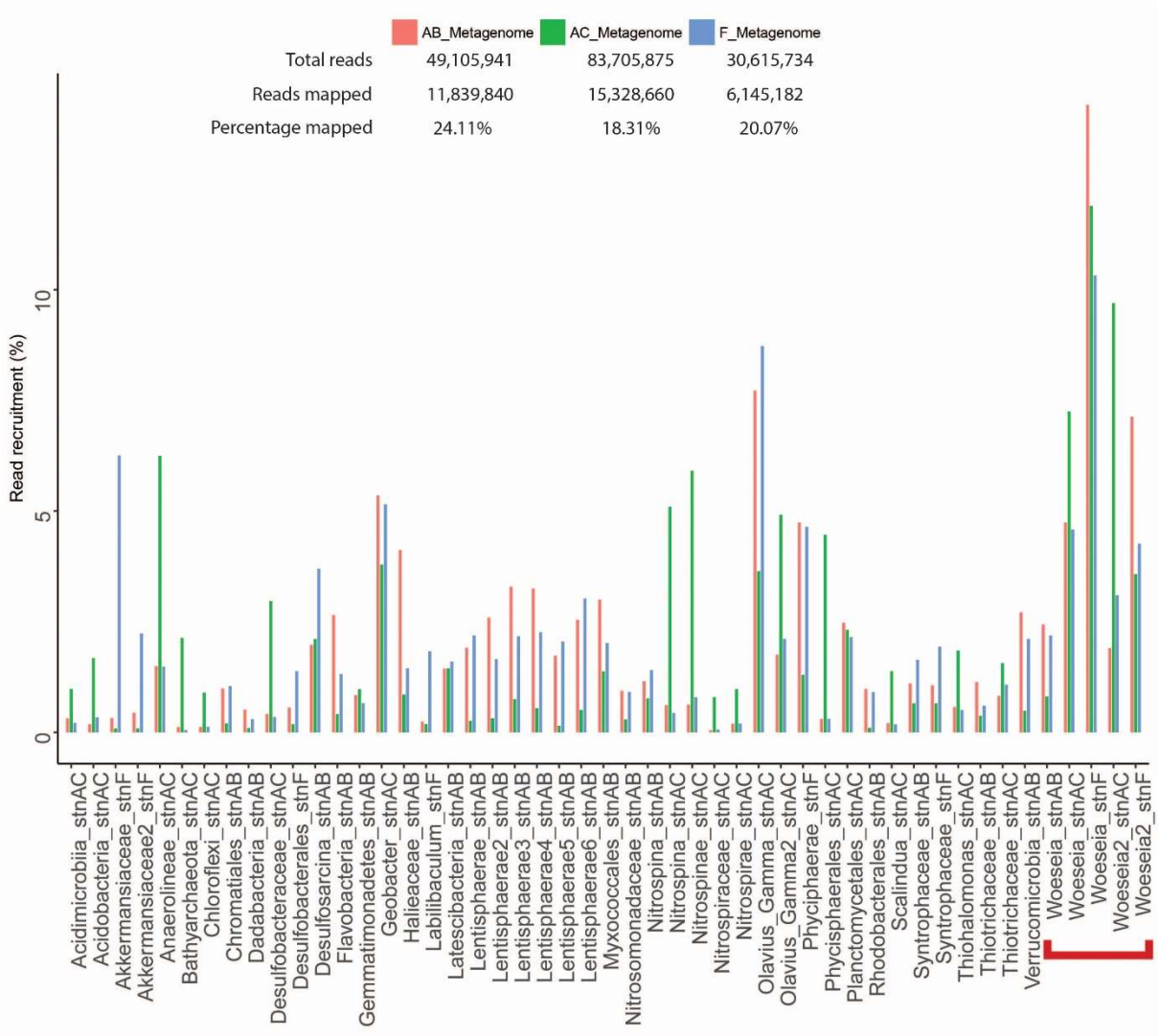

Figure 5.5. Read recruitment for each reconstructed MAG. Woeseiaceae genomes are highlighted with a red bracket. The size of each metagenome is indicated below each item in the legend, as well as number and percentage of reads mapped to all MAGs combined.

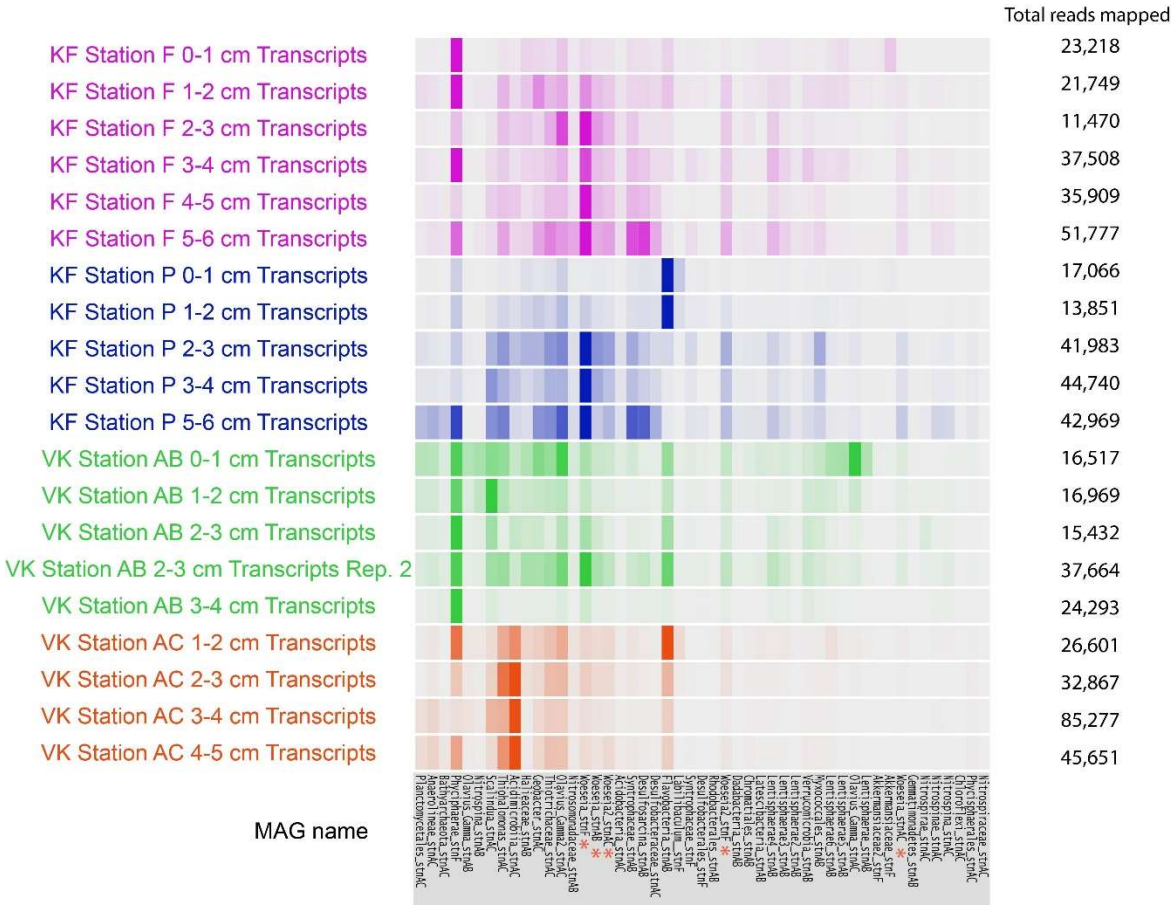

Figure 5.6. Visualization of transcript abundance for each MAG. Kongsfjorden (KF) samples include transcripts from stations F shown in pink and P shown in blue. Van Keulenfjorden (VK) samples include transcripts from stations AB shown in green and AC shown in orange. MAGs are ordered by abundance and Woeseiaceae genomes are highlighted with orange asterisk. The total number of reads mapped is noted on the right.

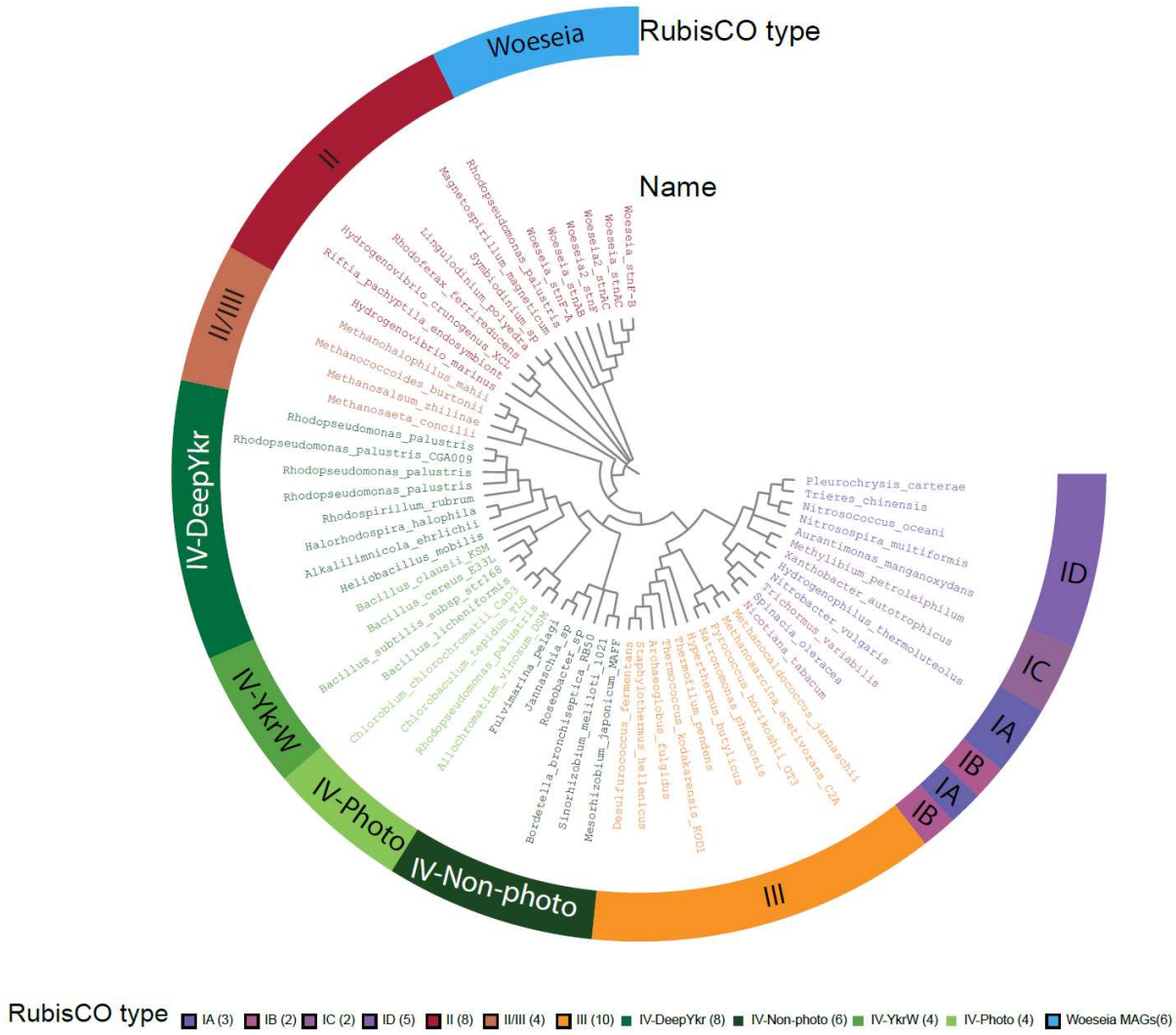

Figure 5.7. Sequence analysis of RubisCO in Woeseiaceae MAGs. Tree was built with full length (> 480 aa) sequences for large and small chains of ribulose bisphosphate carboxylase (RubisCO) were downloaded from NCBI and compared with sequences annotated in our MAGs. Alignments were conducted in Mega v. 7 using Clustal. Mega was also used to construct the neighbor-joining tree and topologies shown have >80% support after 1000 bootstraps.

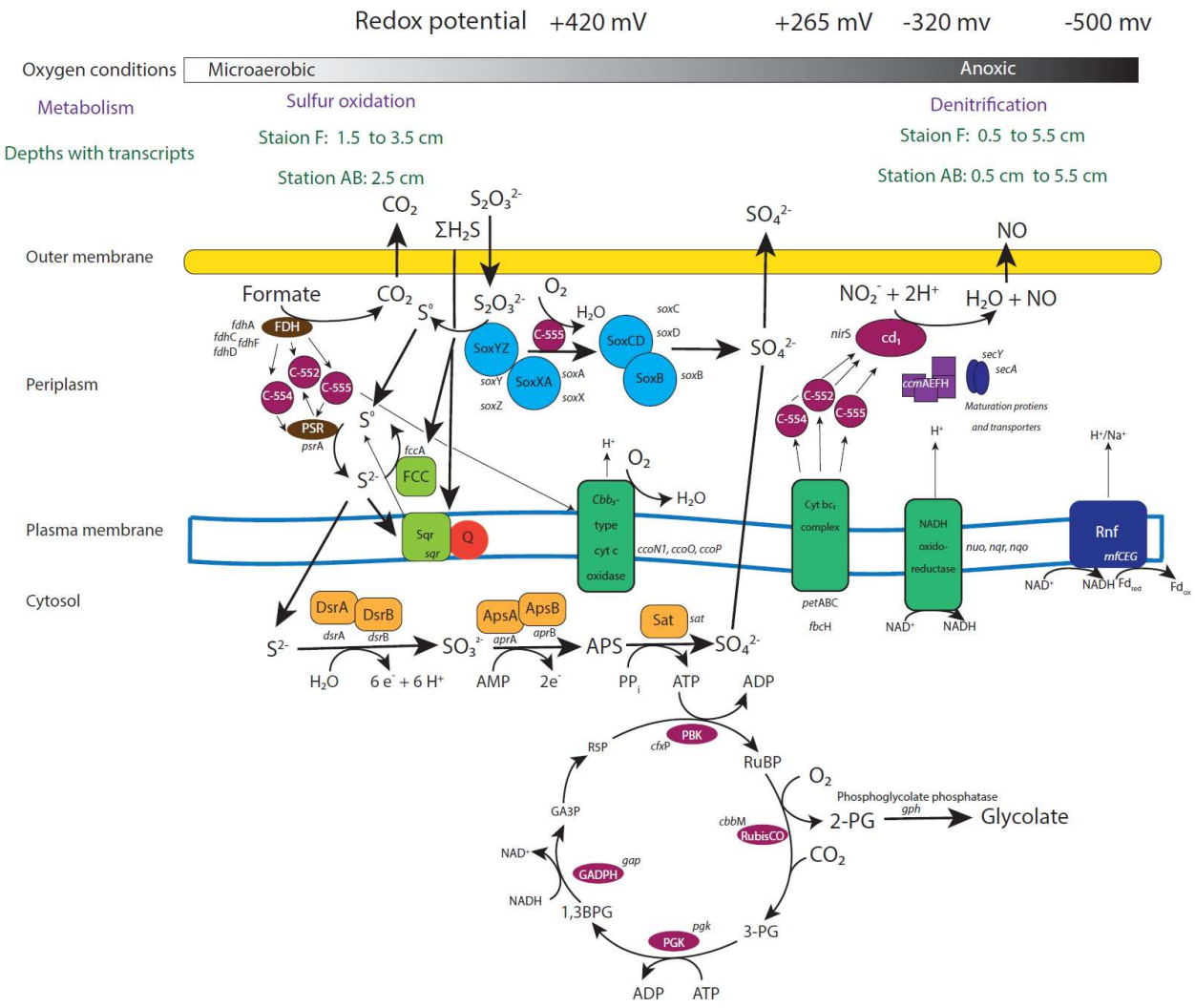

Figure 5.8. Metabolic cartoon of sulfide oxidation coupled to the Calvin Benson Cycle and denitrification. Redox potential indicated at the top taken from Rauhamaki et al., 2009 (cbb_3-type cyt c = +420 mV), Trumpower 1990 (cyt bc_1 complex = +265 mV), and Biegal et al., 2011 (Rnf = -500 mV).

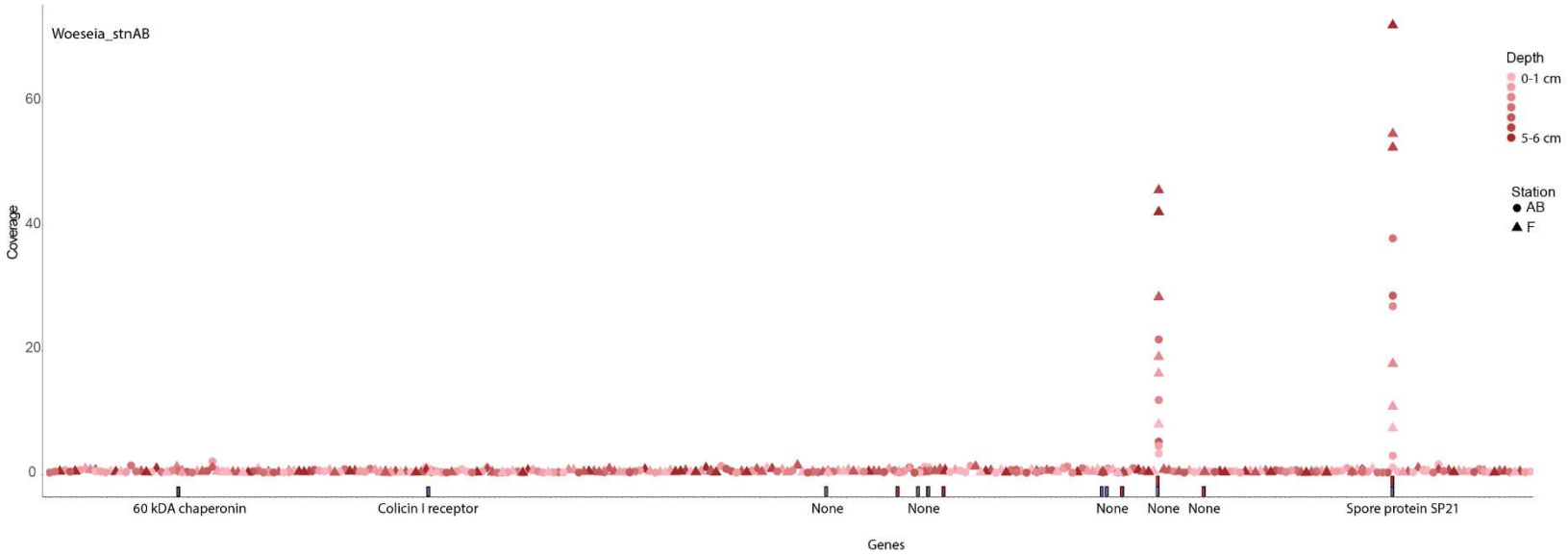

Figure 5.9. Transcriptional coverage across the genome of Woeseia_stnAB at each depth interval in stations AB (circles) and F (triangles). The y-axis is the coverage value and discreet genes are positioned across the x-axis. Only genes with a significant increase or decrease in transcriptional coverage as detected with regression analysis are annotated. Rectangles above each gene indicates that an $R^2 > 0.1$ was detected for station AB (blue) or F (red). Details about regression statistics can be found in Table 5.5.

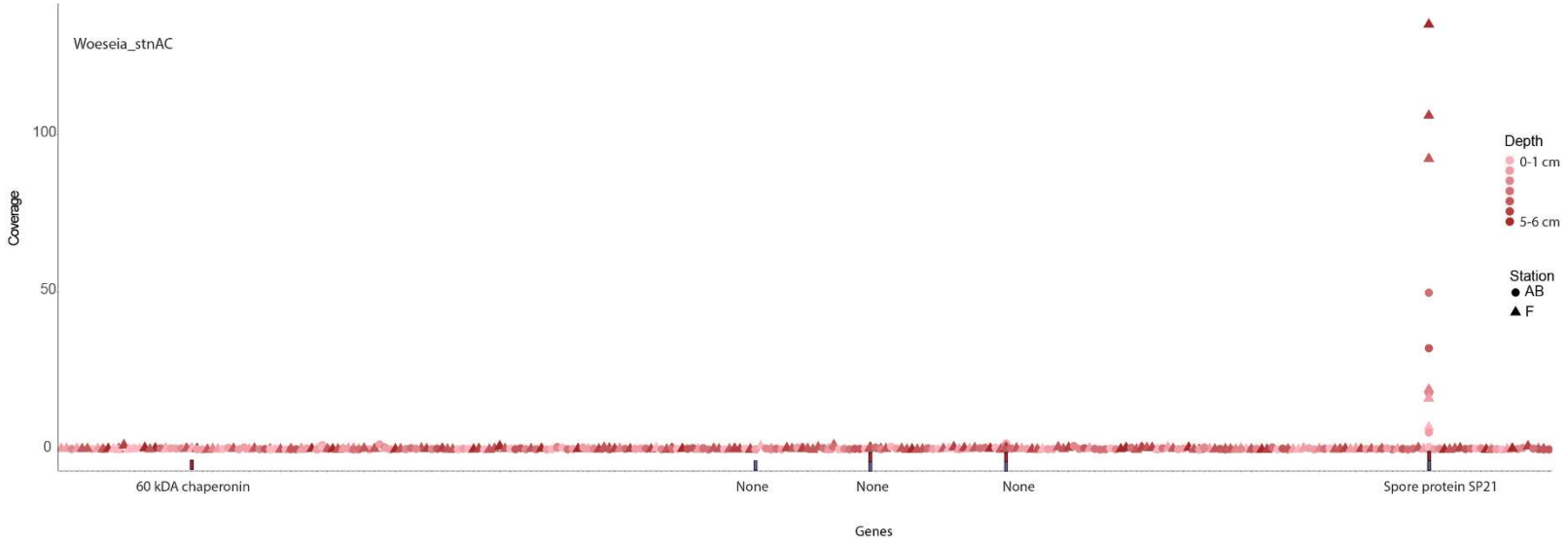

Figure 5.10. Transcriptional coverage across the genome of Woeseia_stnAC at each depth interval in stations AB (circles) and F (triangles). The y-axis is the coverage value and discreet genes are positioned across the x-axis. Only genes with a significant increase or decrease in transcriptional coverage as detected with regression analysis are annotated. Rectangles above each gene indicates that an $R^2 > 0.1$ was detected for station AB (blue) or F (red). Details about regression statistics can be found in Table 5.6.

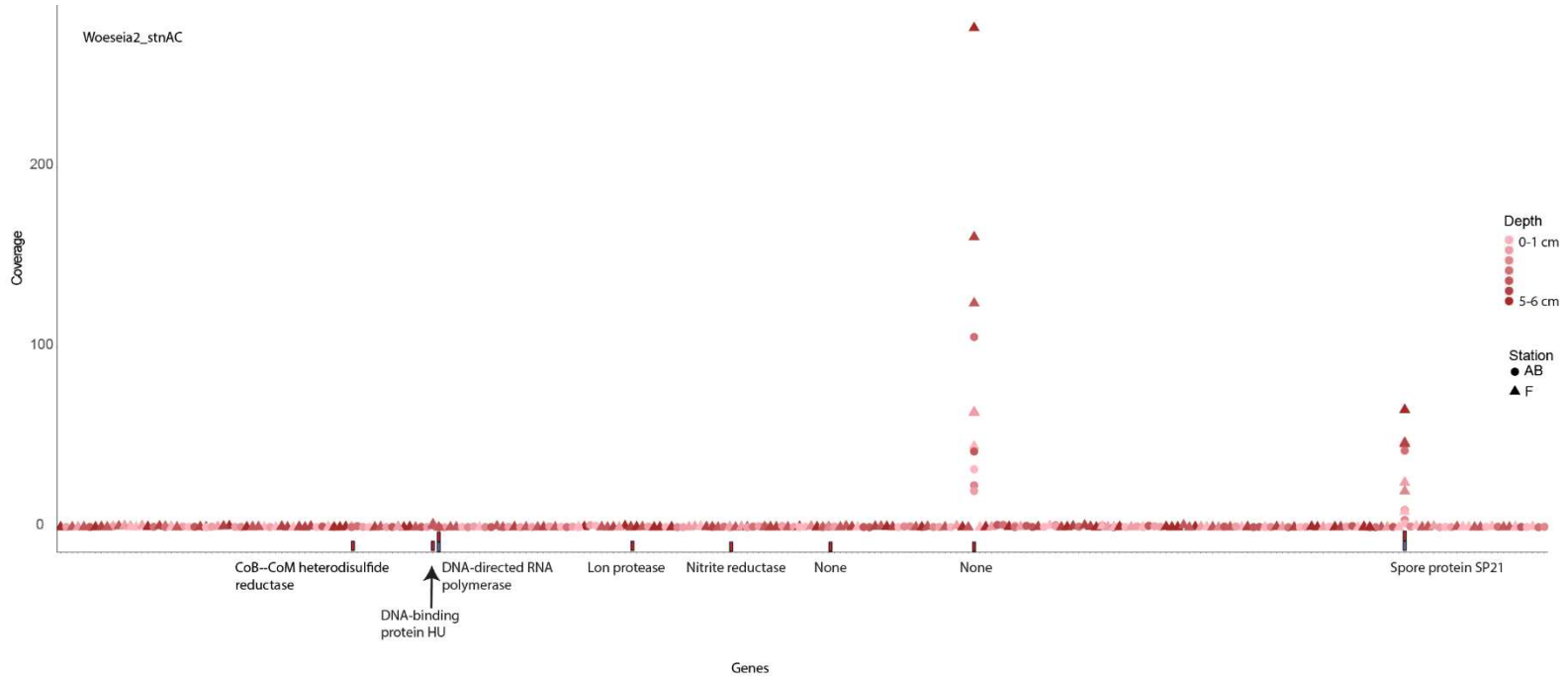

Figure 5.11.Transcriptional coverage across the genome of Woeseia2_stnAC at each depth interval in stations AB (circles) and F (triangles). The y-axis is the coverage value and discreet genes are positioned across the x-axis. Only genes with a significant increase or decrease in transcriptional coverage as detected with regression analysis are annotated. Rectangles above each gene indicates that an $R^2 > 0.1$ was detected for station AB (blue) or F (red). Details about regression statistics can be found in Table 5.7.

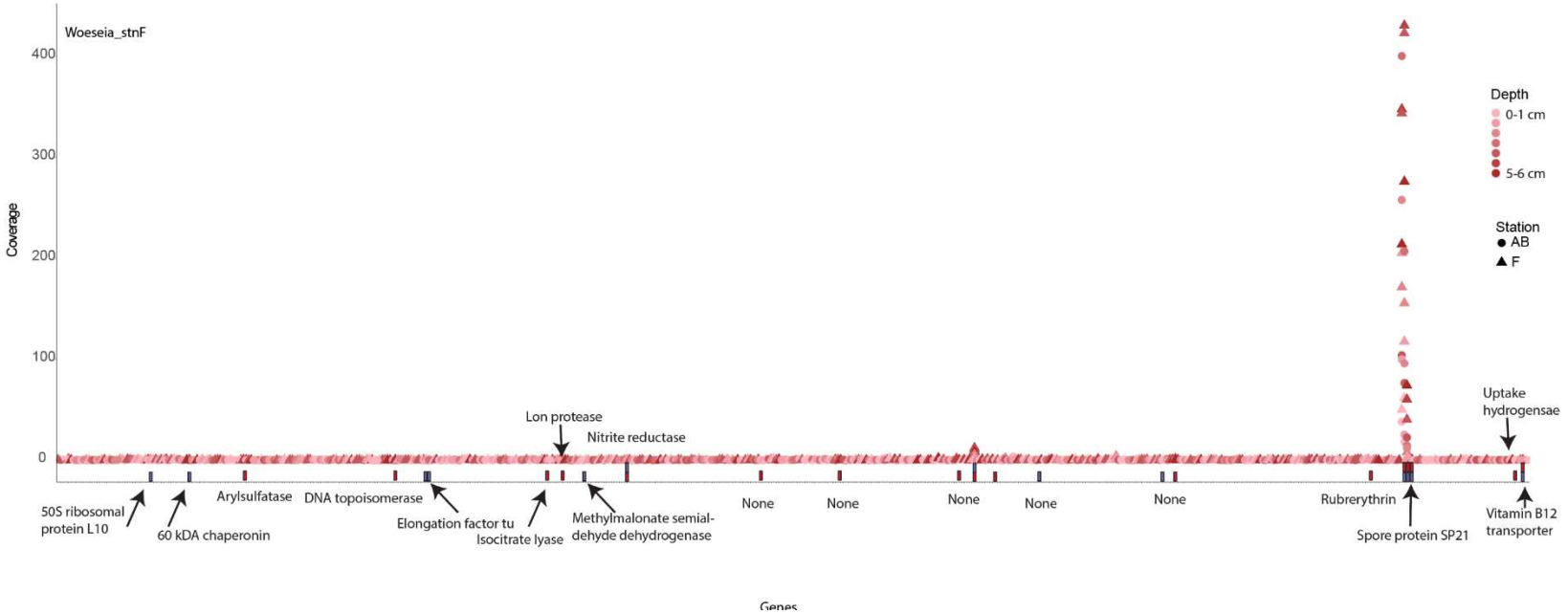

Figure 5.12.Transcriptional coverage across the genome of Woeseia_stnF at each depth interval in stations AB (circles) and F (triangles). The y-axis is the coverage value and discreet genes are positioned across the x-axis. Only genes with a significant increase or decrease in transcriptional coverage as detected with regression analysis are annotated. Rectangles above each gene indicates that an $R^2 > 0.1$ was detected for station AB (blue) or F (red). Details about regression statistics can be found in Table 5.8.

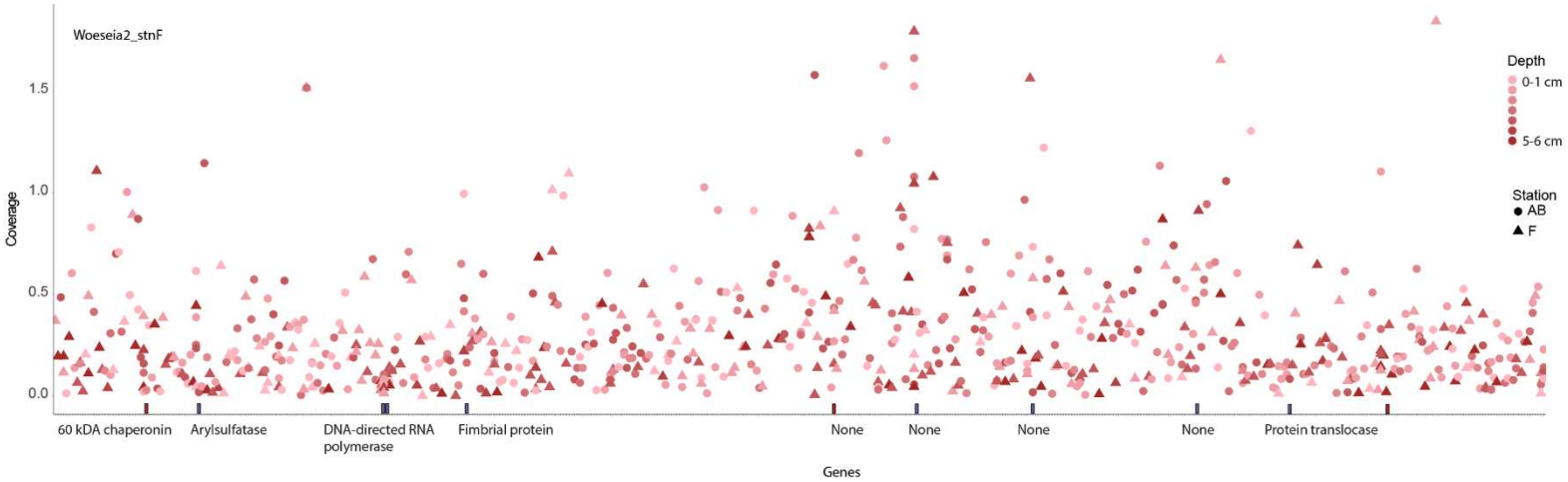

Figure 5.13. Transcriptional coverage across the genome of Woeseia2_stnF at each depth interval in stations AB (circles) and F (triangles). The y-axis is the coverage value and discreet genes are positioned across the x-axis. Only genes with a significant increase or decrease in transcriptional coverage as detected with regression analysis are annotated. Rectangles above each gene indicates that an $R^2 > 0.1$ was detected for station AB (blue) or F (red). Details about regression statistics can be found in Table 5.9.

Appendix II: Script for Metabat Binning

Script for metagenomic binning with MetaBat:

```
MetaBat steps:

# run ridiculously extreme specific binning
minLength=2000
minSamples=5
p1=99
p2=98
minProb=99
minCorr=99
minBinned=40

# run a slightly less ridiculous extreme specific binning
minLength=2000
minSamples=5
p1=98
p2=96
minProb=99
minCorr=99
minBinned=40

# run extremely specific binning
minLength=2000
minSamples=5
p1=95
p2=90
minProb=99
minCorr=99
minBinned=40
```

Appendix III: R Code

R script for relative abundance by 16S rRNA gene libraries

```
#JTB255 relabund


library(plyr)

library(dplyr)

library(ggplot2)

library(tidyr)


JTB255<-read.csv("JTB255_abund.csv")

head(JTB255)

JTB255_tidy<-gather(JTB255, depth, abundance, X0.5:X19.5)

write.csv(JTB255_tidy, "JTB255_tidy.csv") # remove X's in front of depths and add
replicate numbers in new column

JTB255_tidy_fixed<-read.csv("JTB255_tidy_fixed.csv")


library(scales)

show_col(hue_pal()(4)) #make a palette

cols<-c("AB" = "#7CAE00", "AC" = "#C77CFF")


d <- JTB255_tidy_fixed

head(d)

make_plot <- function(d, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...) {

  p <- ggplot(d, aes(x=depth, y=abundance, color=Site, shape=Replicate)) +

    geom_point(aes(fill=Site), colour="black", size=1, stroke=1) +

    scale_fill_manual(values=cols) +

    scale_colour_manual(values=cols) +

    scale_shape_manual(values = c(21,22)) +

    theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),

          panel.background = element_blank(), axis.line = element_line(colour =
"black")) +

    geom_line(size=0.5) +

    scale_x_reverse() +

    coord_flip() +

    ylim(0,3)+

    labs(y="Relative abundance (%)", x = "Depth (cmbsf)")
```

```
  # Do you want to print the plot to the screen?
  if(print_plot) {
    print(p)
  }

  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(d$Site[1], ".png")
    }
    ggsave(filename, p, ...)
  }

  p
}
# Test this function on one Station that I pull out manually
test_set <- d[d$Site== unique(d$Site)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(d, c("Site"), make_plot, save_plot=TRUE, print_plot=FALSE,
height=4, width=3, units="in", dpi=300)

# Later on you can do other things, like print them
l_ply(plot_list, print)
```

R script for regression analysis of transcripts:

```
library(plyr)

library(dplyr)

library(ggplot2)

#### regression function

regression=function(df){

  #setting the regression function.

  reg_fun<-lm(formula=df$Value ~ df$Depth) #regression function

  #getting the slope, intercept, R square and adjusted R squared of

  #the regression function (with 3 decimals).

  slope<-round(coef(reg_fun)[2],3)

  intercept<-round(coef(reg_fun)[1],3)

  R2<-round(as.numeric(summary(reg_fun)[8]),3)

  R2.Adj<-round(as.numeric(summary(reg_fun)[9]),3)

  c(slope,intercept,R2,R2.Adj)

}

############# Genome: Woeseia_stnAB

# Station AB


WoeseiaAB<-read.csv("Woeseia_both_sorted.csv")

WoesAB_AB<-subset(WoeseiaAB, Station=="AB")

WoesAB_AB<-subset(WoesAB_AB, Type=="Coverage")

head(WoesAB_AB)


WoesAB_regression <- ddply(WoesAB_AB, ~ Gene_name, regression)

colnames(WoesAB_regression)<-c ("value","slope","intercept","R2","R2.Adj")


write.csv(WoesAB_regression, "WoesAB_regression_stationAB.csv")


make_plot <- function(WoesAB_AB, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{

  p <- WoesAB_AB %>%

    ggplot(aes(x = Depth, y = Value)) +

    geom_point() +
```

```
    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }


  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(WoesAB_AB$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }


  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesAB_AB[WoesAB_AB$Gene_name == unique(WoesAB_AB$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)


# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesAB_AB, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)


# Station F
WoesAB_F<-subset(WoeseiaAB, Station=="F")
WoesAB_F<-subset(WoesAB_F, Type=="Coverage")


head(WoesAB_F)


WoesF_regression <- ddply(WoesAB_F, ~ Gene_name, regression)
```

```
colnames(WoesF_regression)<-c ("value","slope","intercept","R2","R2.Adj")


write.csv(WoesF_regression, "WoesAB_regression_stationF.csv")


make_plot <- function(WoesAB_F, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- WoesAB_F %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }


  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(WoesAB_F$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }


  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesAB_F[WoesAB_F$Gene_name == unique(WoesAB_F$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)


# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesAB_F, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)
```

```
############Woeseia_stnAC#########

#Station AB

Woeseia_AC<-read.csv("WoeseiaAC_both_sorted.csv")

WoesAC_AB<-subset(Woeseia_AC, Station=="AB")

WoesAC_AB<-subset(WoesAC_AB, Type=="Coverage")

head(WoesAC_AB)


WoesAC_regression <- ddply(WoesAC_AB, ~ Gene_name, regression)

colnames(WoesAC_regression)<-c ("value","slope","intercept","R2","R2.Adj")


write.csv(WoesAC_regression, "WoesAC_regression_stationAB.csv")


make_plot <- function(WoesAC_AB, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- WoesAC_AB %>%

    ggplot(aes(x = Depth, y = Value)) +

    geom_point() +

    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?

  if(print_plot) {

    print(p)

  }


  # Do you want to save the plot?

  if(save_plot) {

    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png

      filename <- paste0(WoesAC_AB$Gene_name[1], ".png")

    }

    ggsave(filename, p, ...)

  }
```

```
  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesAC_AB[WoesAC_AB$Gene_name == unique(WoesAC_AB$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesAC_AB, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)

#Station F
WoesAC_F<-subset(Woeseia_AC, Station=="F")
WoesAC_F<-subset(WoesAC_F, Type=="Coverage")
head(WoesAC_F)

WoesAC_F_regression <- ddply(WoesAC_F, ~ Gene_name, regression)
colnames(WoesAC_F_regression)<-c ("value","slope","intercept","R2","R2.Adj")
head(WoesAC_F_regression)
write.csv(WoesAC_F_regression, "WoesAC_regression_stationF.csv")

make_plot <- function(WoesAC_F, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- WoesAB_F %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)

  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }
```

```
  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(WoesAB_F$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }

  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesAB_F[WoesAB_F$Gene_name == unique(WoesAB_F$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesAB_F, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)

#############Woeseia2_stnAC
#Station AB
Woes2AC<-read.csv("WoeseiaAC2_both_sorted.csv")
Woes2AC_AB<-subset(Woes2AC, Station=="AB")
Woes2AC_AB<-subset(Woes2AC_AB, Type=="Coverage")
head(Woes2AC_AB)

WoesAC2_regression_AB <- ddply(Woes2AC_AB, ~ Gene_name, regression)
colnames(WoesAC2_regression_AB)<-c ("value","slope","intercept","R2","R2.Adj")

write.csv(WoesAC2_regression_AB, "WoesAC2_regression_stationAB.csv")

make_plot <- function(Woes2AC_AB, save_plot=TRUE, print_plot=FALSE, filename=NULL,
...) {
  p <- Woes2AC_AB %>%
```

```
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)

  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }

  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(Woes2AC_AB$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }

  p
}
# Test this function on one GENE that I pull out manually
test_set <- Woes2AC_AB[Woes2AC_AB$Gene_name == unique(Woes2AC_AB$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(Woes2AC_AB, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)

#Station F
Woes2AC_F<-subset(Woes2AC, Station=="F")
Woes2AC_F<-subset(Woes2AC_F, Type=="Coverage")
head(Woes2AC_F)
```

```
WoesAC2_regression_F <- ddply(Woes2AC_F, ~ Gene_name, regression)
colnames(WoesAC2_regression_F)<-c ("value","slope","intercept","R2","R2.Adj")

write.csv(WoesAC2_regression_F, "WoesAC2_regression_stationF.csv")

make_plot <- function(Woes2AC_F, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- Woes2AC_F %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)

  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }

  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(Woes2AC_F$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }

  p
}
# Test this function on one GENE that I pull out manually
test_set <- Woes2AC_F[Woes2AC_F$Gene_name == unique(Woes2AC_F$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
```

```
plot_list <- dlply(Woes2AC_F, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)


#########Woeseia_stnF

Woeseia_F<-read.csv("WoeseiaF_both_sorted.csv")

WoesF_AB<-subset(Woeseia_F, Station=="AB")

WoesF_AB<-subset(WoesF_AB, Type=="Coverage")

head(WoesF_AB)


WoesF_regression_AB <- ddply(WoesF_AB, ~ Gene_name, regression)

colnames(WoesF_regression_AB)<-c ("value","slope","intercept","R2","R2.Adj")


write.csv(WoesF_regression_AB, "WoesF_regression_stationAB.csv")


make_plot <- function(WoesF_AB, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- WoesF_AB %>%

    ggplot(aes(x = Depth, y = Value)) +

    geom_point() +

    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?

  if(print_plot) {

    print(p)

  }


  # Do you want to save the plot?

  if(save_plot) {

    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png

      filename <- paste0(WoesF_AB$Gene_name[1], ".png")

    }

    ggsave(filename, p, ...)

  }
```

```
  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesF_AB[WoesF_AB$Gene_name == unique(WoesF_AB$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesF_AB, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)

#Station F
WoesF_F<-subset(Woeseia_F, Station=="F")
WoesF_F<-subset(WoesF_F, Type=="Coverage")
head(WoesF_F)

WoesF_regression_F <- ddply(WoesF_F, ~ Gene_name, regression)
colnames(WoesF_regression_F)<-c ("value","slope","intercept","R2","R2.Adj")

write.csv(WoesF_regression_F, "WoesF_regression_stationF.csv")

make_plot <- function(WoesF_F, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...) {
  p <- WoesF_F %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)

  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }

  # Do you want to save the plot?
  if(save_plot) {
```

```
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(WoesF_F$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }

  p
}
# Test this function on one GENE that I pull out manually
test_set <- WoesF_F[WoesF_F$Gene_name == unique(WoesF_F$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)

# Use dlply to make a list of each data set, and save them
plot_list <- dlply(WoesF_F, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)

######Woeseia2_stnF
#Station AB
Woeseia2_F<-read.csv("WoeseiaF2_both_sorted.csv")
Woes2F_AB<-subset(Woeseia2_F, Station=="AB")
Woes2F_AB<-subset(Woes2F_AB, Type=="Coverage")
head(Woes2F_AB)

Woes2F_regression_AB <- ddply(Woes2F_AB, ~ Gene_name, regression)
colnames(Woes2F_regression_AB)<-c ("value","slope","intercept","R2","R2.Adj")

write.csv(Woes2F_regression_AB, "Woes2F_regression_stationAB.csv")

make_plot <- function(Woes2F_AB, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- Woes2F_AB %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
```

```
    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }


  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(Woes2F_AB$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }


  p
}
# Test this function on one GENE that I pull out manually
test_set <- Woes2F_AB[Woes2F_AB$Gene_name == unique(Woes2F_AB$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)


# Use dlply to make a list of each data set, and save them
plot_list <- dlply(Woes2F_AB, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)


#Station F
Woes2F_F<-subset(Woeseia2_F, Station=="F")
Woes2F_F<-subset(Woes2F_F, Type=="Coverage")
head(Woes2F_F)


Woes2F_regression_F <- ddply(Woes2F_F, ~ Gene_name, regression)
colnames(Woes2F_regression_F)<-c ("value","slope","intercept","R2","R2.Adj")
```

```
write.csv(Woes2F_regression_F, "Woes2F_regression_stationF.csv")


make_plot <- function(Woes2F_F, save_plot=TRUE, print_plot=FALSE, filename=NULL, ...)
{
  p <- Woes2F_F %>%
    ggplot(aes(x = Depth, y = Value)) +
    geom_point() +
    geom_smooth(method = "lm", se = F)


  # Do you want to print hte plot to the screen?
  if(print_plot) {
    print(p)
  }


  # Do you want to save the plot?
  if(save_plot) {
    if(is.null(filename)) { # create a filename for the plot automatically, if one
hasn't been specified, and add .png
      filename <- paste0(Woes2F_F$Gene_name[1], ".png")
    }
    ggsave(filename, p, ...)
  }


  p
}
# Test this function on one GENE that I pull out manually
test_set <- Woes2F_F[Woes2F_F$Gene_name == unique(Woes2F_F$Gene_name)[1], ]
test_plot <- make_plot(test_set)
print(test_plot)


# Use dlply to make a list of each data set, and save them
plot_list <- dlply(Woes2F_F, c("Gene_name"), make_plot, save_plot=TRUE,
print_plot=FALSE, height=10, width=8, units="in", dpi=300)
```

R script for metabolite heatmaps

```
library(gplots)

library(ggplot)

library(RColorBrewer)

Metabo_palette<-colorRampPalette(c("white", "lightpink", "black"))(n=299)

#across all samples

Metabo_all<-read.csv("Metabolites_all.csv")

Metabo_all_names<-Metabo_all[,1] #assign metabolite names as labels

Metabo_all_matrix<-data.matrix(Metabo_all[,2:ncol(Metabo_all)]) #make the dataframe
into a matrix

rownames(Metabo_all_matrix) <-Metabo_all_names #assign row names for matrix



heatmap.2(Metabo_all_matrix,

          main = "All sites", # heat map title

          notecol="black",      # change font color of cell labels to black

          density.info="none",  # turns off density plot inside color legend

          trace="none",         # turns off trace lines inside the heat map

          scale = c("row"), na.rm=TRUE,

          margins =c(12,9),     # widens margins around plot

          col=Metabo_palette,  # use on color palette defined earlier

          dendrogram='none',

          Rowv=FALSE,

          Colv="NA")            # turn off column clustering



# KF stnF


Metabo<-read.csv("Metabolites_F.csv")

Metabo_names<-Metabo[,1] #assign metabolite names as labels

Metabo_matrix<-data.matrix(Metabo[,2:ncol(Metabo)]) #make the dataframe into a matrix

rownames(Metabo_matrix) <-Metabo_names #assign row names for matrix
```

```
heatmap.2(Metabo_matrix,
          main = "Kongsfjord Stn F metabolites", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          scale = c("row"), na.rm=TRUE,
          margins =c(12,9),     # widens margins around plot
          col=Metabo_palette,  # use on color palette defined earlier
          dendrogram='none',
          Rowv=FALSE,
          Colv="NA")            # turn off column clustering

# VK stnAB

Metabo_AB<-read.csv("Metabolites_AB.csv")
Metabo_AB_names<-Metabo_AB[,1]
Metabo_AB_matrix<-data.matrix(Metabo_AB[,2:ncol(Metabo_AB)])
rownames(Metabo_AB_matrix) <- Metabo_AB_names

heatmap.2(Metabo_AB_matrix,
          main = "Van Keulenfjord Stn AB metabolites", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          scale = c("row"), na.rm=TRUE,
          margins =c(12,9),     # widens margins around plot
          col=Metabo_palette,  # use on color palette defined earlier
          dendrogram='none',
          Rowv=FALSE,
          Colv="NA")            # turn off column clustering

# VK stnAC
```

```
Metabo_AC<-read.csv("Metabolites_AC.csv")
Metabo_AC_names<-Metabo_AC[,1]
Metabo_AC_matrix<-data.matrix(Metabo_AC[,2:ncol(Metabo_AC)])
rownames(Metabo_AC_matrix) <- Metabo_AC_names

heatmap.2(Metabo_AC_matrix,
          main = "Van Keulenfjord Stn AC metabolites", # heat map title
          notecol="black",      # change font color of cell labels to black
          density.info="none",  # turns off density plot inside color legend
          trace="none",         # turns off trace lines inside the heat map
          scale = c("row"), na.rm=TRUE,
          margins =c(12,9),     # widens margins around plot
          col=Metabo_palette,  # use on color palette defined earlier
          dendrogram='none',
          Rowv=FALSE,
          Colv="NA")            # turn off column clustering
```

Appendix IV: KEGG Pathways

Peptide and carbohydrate metabolism in Woeseiaceae MAGs

The evidence we found for sulfur oxidation and denitrification support recent work on the metabolic versatility in Woeseiaceae, which highlighted its ability to function both as a chemolithoautotroph (1) and as a key driver of carbon fixation in marine sediments (2). However, because the only cultured representative of this clade was identified as a chemoheterotroph (3), we searched for genes related to carbon and peptide oxidation metabolism in the genomes (4). We examined our genomes with Kegg Decoder, which takes amino acid sequences as input and annotates these sequences with KO ontology. These KO annotations were placed into a published pipeline for the reconstruction of key metabolic KEGG pathways. A full description of the KOs used for each pathway can be found at https://github.com/bjtully/BioData/tree/master/KEGGDecoder.

All genomes encoded a diversity of peptidase genes (Figures A3-1, A3-2), including peptides within the family M28 containing aminopeptidases and carboxypeptidases, and M50, which is a family of metalloendopeptidases with a subclass (M50B) related to sporulation factors. In addition, all or most genomes encode genes for di- and tripeptidases, oligopeptidase F, phosphoserine aminotransferase, and/or peptidase S26. Transcription for peptidases was restricted to aminopeptidase N by Woeseia2_stnAC at Kongsfjorden, Xaa-Pro aminopeptidase by Woeseia2_stnF and phosphoserine aminopeptidase by Woeseia_stnAC in Van Keulenfjorden. Consistent with our findings, the isolate *Woeseia oceani* XK5, displayed proteolytic enzyme activity in culture (3).

Annotation with dbCAN2 (5) showed the presence of diverse carbohydrate-active enzymes, including glycosyltransferases, as well as auxiliary activity enzymes and carbohydrate binding proteins (Table 5A-III.1). This agrees with the genomic contents of other Woeseiaceae

representatives, although our genomes appear to be less enriched with respect to glysoside hydrolases compared to those genomes (1). Woeseia2_stnF also contained polysaccharide lyases. Transporters for organic molecules were only encoded in MAGs recovered from station F and beta-glucanase in the Woeseia2_stnF and Woeseia_stnAB.

References

1. Mußmann M, Pjevac P, Krüger K, Dyksma S. 2017. Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. The ISME journal 11:1276.

2. Dyksma S, Bischof K, Fuchs BM, Hoffmann K, Meier D, Meyerdierks A, Pjevac P, Probandt D, Richter M, Stepanauskas R. 2016. Ubiquitous Gammaproteobacteria dominate dark carbon fixation in coastal sediments. The ISME journal 10:1939.

3. Du Z-J, Wang Z-J, Zhao J-X, Chen G-J. 2016. Woeseia oceani gen. nov., sp. nov., a chemoheterotrophic member of the order Chromatiales, and proposal of Woeseiaceae fam. nov. International journal of systematic and evolutionary microbiology 66:107-112.

4. Graham E, Heidelberg J, Tully B. 2018. Potential for primary productivity in a globally-distributed bacterial phototroph. The ISME journal:1.

5. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Research 46:W95-W101.

Table 5A-III.1. Results from dbCAN2 annotation of genomes for carbohydrate-active enzymes.

Genome	AA	CBM	CE	GH	GT	PL
Woeseia_stnAB	9	5	10	14	12	0
Woeseia_stnAC	5	3	5	10	7	0
Woeseia2_stnAC	7	2	7	14	7	0
Woeseia_stnF	11	9	15	27	21	0
Woeseia2_stnF	10	5	7	16	10	6

AA = Auxiliary activities

CBM = Carbohydrate binding modules

CE = Carbohydrate esterases

GH = Glycoside hydrolases

GT = Glycosyltransferases

PL = Polysaccharide lyases

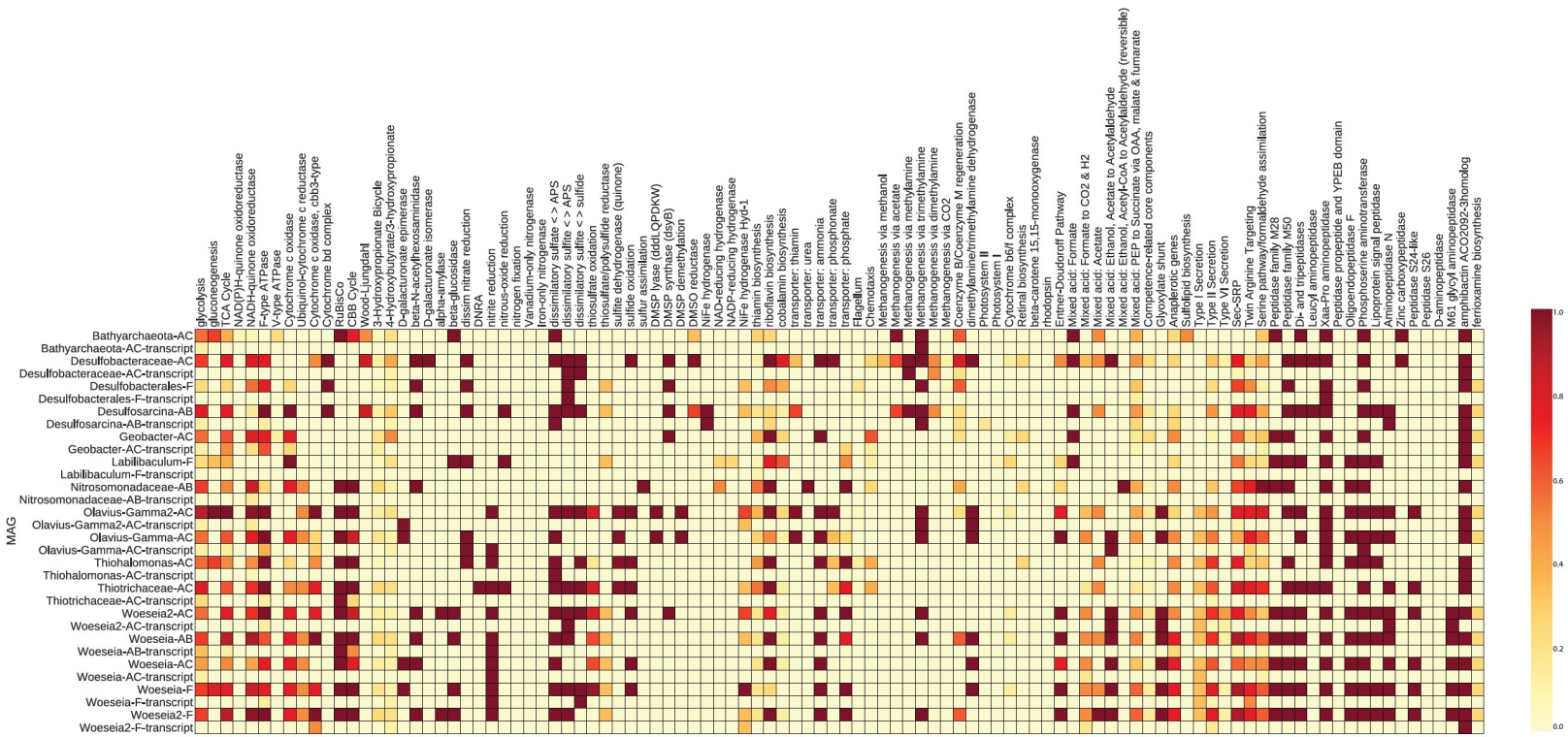

Figure 5A-III.1. Heatmap of genomic contents and expression of genes in Kongsfjorden (all depths) across all MAGs reconstructed in this study.

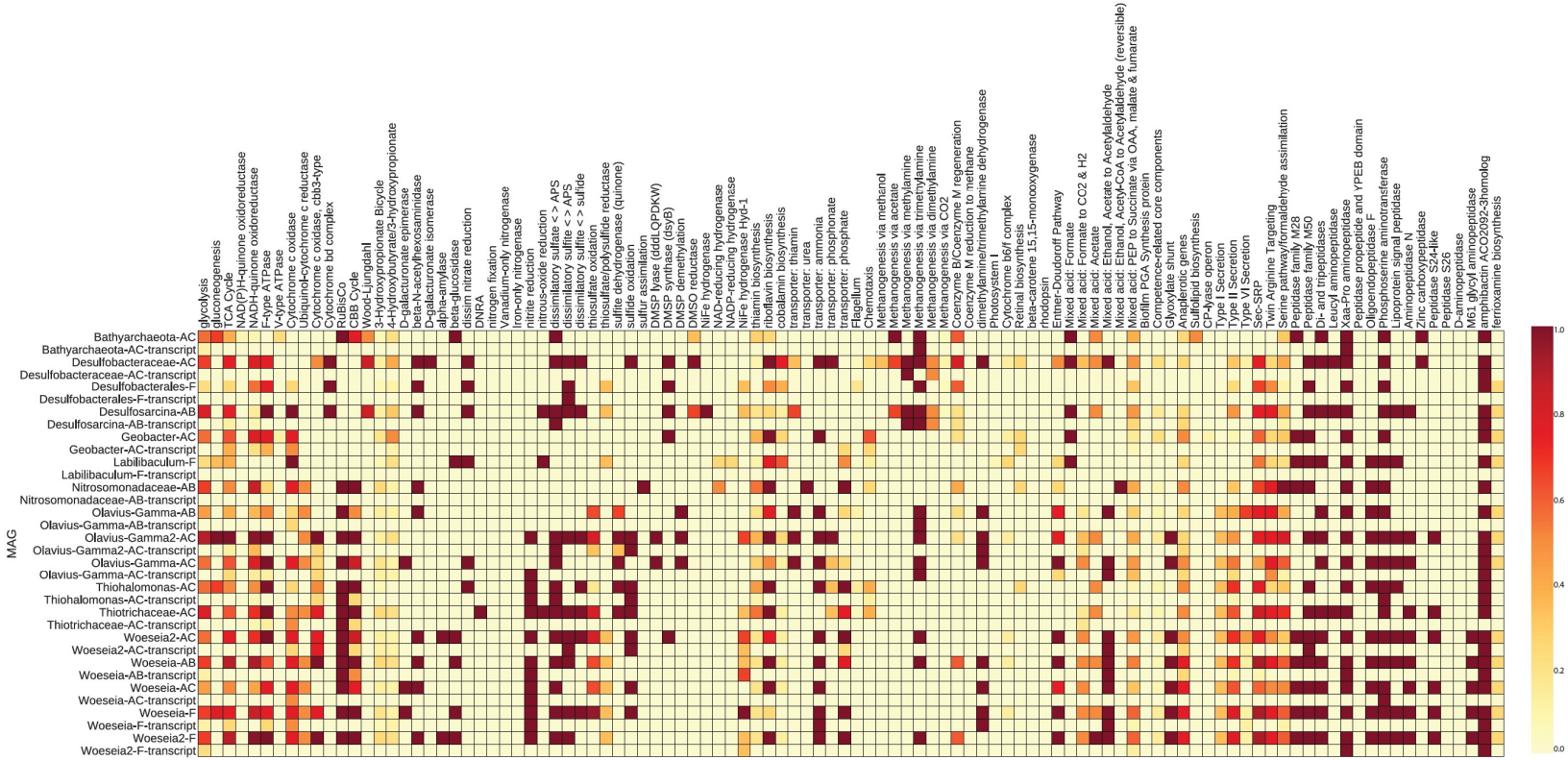

Figure 5A-III.2. Heatmap of genomic contents and expression of genes in Van Keulenfjorden (all depths) across all MAGs reconstructed in this study.

Appendix V: Pangenomic Analysis

Pangenomic analysis and enriched functions

Comparisons of the presence or absence of gene clusters was determined for our 5 Woeseiaceae genomes alongside the 5 publicly available genomes from NCBI and IMG (Figure 5A-IV.1). Genome statistics can be found in Table 5.3 in the main text. We used the pangenomics workflow in Anvi'o (1) and implemented the 'anvi-get-enriched-functions-per-pan-group' program to identify the enrichment or relative depletion of genes in our genomes compared to those in the database. Each gene was assigned an enrichment score, which is a metric for determining how unique a function/gene call is to the genomes that are assigned to a specific group compared all other genomes included in the pangenome analysis. Specifically, the enrichment score is "the test statistic for a two sample Z-test for proportions. It takes the proportion of times the function is observed in the group, subtracts the proportion of times the function is observed outside the group, and re-scales this difference to reflect the number of samples observed in each group. The adjustment for group size means that larger scores are given when groups are larger – essentially, a difference between groups can be considered more robust when there are more representatives of each group" (http://merenlab.org/2016/11/08/pangenomics-v2/).

We found that among the core, highly-enriched functions in our genomes were genes encoding for cytochrome b6. (Table 5A-IV.1). Cytochrome b6 is used as a means for passing electrons between photosystem II and I in chloroplasts of plants and cyanobacteria (2, 3) as well as in green sulfur bacteria (4). Our genomes do not contain strong evidence for functional photosystems I and/or II, and because the gene encoding for cytochrome b6, *pet*B, also annotates in our genomes as cytochrome b (part of the respiratory complex III or cytochrome bc_1 complex),

due to its structural similarity and sequence homology (5), we interpret the cytochrome b6 annotation to be incorrect.

Another highly-enriched gene was ornithine carbamoyltransferase, which is part of the detoxifying urea cycle and involved in the *de novo* synthesis of arginine through the conversion of ornithine and carbamoyl phosphate (CP) into citrulline and inorganic phosphate. This enzyme has been found to contain species-specific structural adaptations to allow function at both high (6) and low temperatures (7). No transcripts were found for the gene encoding ornithine carbamoyltransferase (*arg*F) in our genomes, making it hard to interpret the cause for such enrichment of this gene.

Depleted genes

Among the depleted genes in our genomes, nitric oxide reductase subunits b and c (*nor*B and *nor*C) were at the top of the list. This is interesting, as nitric oxide reductase would continue the process of intracellular denitrification that is begun by NirS encoded and transcribed in our genomes. However, the notable lack of NorB and NorC within all of our genomes suggests that nitric oxide is transferred out of the cell instead of being dealt with internally.

It is also worth noting that several copies of genes/functions may be found in the genomes, and so enrichment scores may vary according to which copy the program is testing. For example, cytochrome c-554(548) is listed as being depleted in our genomes (Table 5A-IV.2), with 0/5 MAGs in this study encoding for it; however, as written in the main text, nearly all MAGs do in fact have cytochrome c-554(548) encoded in their genome. Another point for concern is that this program does not take the completeness of each genome into account when performing enrichment analysis. These points, coupled with the possibility of misannotations

discussed above, indicates that data resulting from enrichment analysis within Anvi'o should be interpreted with caution.

References

1. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, Delmont TO. 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3:e1319.

2. Stroebel D, Choquet Y, Popot J-L, Picot D. 2003. An atypical haem in the cytochrome b 6 f complex. Nature 426:413.

3. Hasan SS, Stofleth JT, Yamashita E, Cramer WA. 2013. Lipid-Induced Conformational Changes within the Cytochrome b(6)f Complex of Oxygenic Photosynthesis. Biochemistry 52:2649-2654.

4. Gregersen LH, Bryant DA, Frigaard N-U. 2011. Mechanisms and Evolution of Oxidative Sulfur Metabolism in Green Sulfur Bacteria. Frontiers in Microbiology 2:116.

5. Widger WR, Cramer W, Herrmann R, Trebst A. 1984. Sequence homology and structural similarity between cytochrome b of mitochondrial complex III and the chloroplast b6-f complex: position of the cytochrome b hemes in the membrane. Proceedings of the National Academy of Sciences 81:674-678.

6. Villeret V, Clantin B, Tricot C, Legrain C, Roovers M, Stalon V, Glansdorff N, Van Beeumen J. 1998. The crystal structure of Pyrococcus furiosus ornithine carbamoyltransferase reveals a key role for oligomerization in enzyme stability at extremely high temperatures. Proceedings of the National Academy of Sciences 95:2801-2806.

7. Xu Y, Feller G, Gerday C, Glansdorff N. 2003. Metabolic Enzymes from Psychrophilic Bacteria: Challenge of Adaptation to Low Temperatures in Ornithine Carbamoyltransferase from Moritella abyssi. Journal of Bacteriology 185:2161-2168

Table 5A-IV.1. Enriched genes in our Woeseiaceae MAGs according to pangenomic analysis in Anvi'o. WES = Weighted enrichment score. Core genes are genes that are found in all 5 of the MAGs in this study.

Prokka:Prodigal call	WES	Wilcoxon p-val	Occurrence in our genomes (out of 5)	Occurrence outside of our genomes (out of 5)	Core in our genomes?
Cytochrome b6-f complex iron-sulfur subunit 1	10	0.01	5	0	TRUE
Cytochrome b6	10	0.01	5	0	TRUE
Ornithine carbamoyltransferase	10	0.01	5	0	TRUE
Dehydrosqualene desaturase	8	0.04	4	0	FALSE
HTH-type transcriptional regulator CdhR	8	0.04	4	0	FALSE
O-acetyltransferase OatA	8	0.04	4	0	FALSE
Polysulfide reductase chain A	8	0.04	5	1	TRUE
Aldehyde oxidoreductase	8	0.04	5	1	TRUE
NADPH-Fe(3+) oxidoreductase subunit beta	8	0.04	5	1	TRUE
D-hydantoinase/dihydropyrimidinase	8	0.04	5	1	TRUE
Thiamine kinase	8	0.04	5	1	TRUE
3-hydroxy-5-methyl-1-naphthoate 3-O-methyltransferase	8	0.04	4	0	FALSE
ABC transporter permease YtrF	8	0.04	5	1	TRUE
Pyruvate kinase	8	0.04	5	1	TRUE
Sulfate/thiosulfate import ATP-binding protein CysA	8	0.04	5	1	TRUE
Ribulose bisphosphate carboxylase	8	0.04	5	1	TRUE
Nicotinate dehydrogenase FAD-subunit	8	0.04	5	1	TRUE
Hdr-like menaquinol oxidoreductase iron-sulfur subunit 2	8	0.04	5	1	TRUE
Cytochrome b561	8	0.04	5	1	TRUE

Table 5A-IV.1 continued.

Prokka:Prodigal call	WES	Wilcoxon p-val	Occurrence in our genomes (out of 5)	Occurrence outside of our genomes (out of 5)	Core in our genomes?
SCO1 protein	8	0.04	5	1	TRUE
Purine catabolism protein PucG	8	0.04	4	0	FALSE
Carboxypeptidase G2	8	0.04	5	1	TRUE
putative iron export permease protein FetB	8	0.04	5	1	TRUE
Carbon dioxide concentrating mechanism protein CcmL	8	0.04	4	0	FALSE
putative protease YhbU	8	0.04	4	0	FALSE
Sporulation initiation phosphotransferase F	8	0.04	4	0	FALSE
Electron transport complex subunit RnfC	8	0.04	5	1	TRUE
Adenylylsulfate reductase subunit alpha	8	0.04	5	1	TRUE
Histone deacetylase-like amidohydrolase	8	0.04	5	1	TRUE

Table 5A-IV.2. Depleted genes in our Woeseiaceae MAGs according to pangenomic analysis in Anvi'o compared to other genomes. WES = Weighted enrichment score.

Prokka:Prodigal call	WES	Wilcoxon p-val	Occurrence in our genomes (out of 5)	Occurrence outside of our genomes (out of 5)
Nitric oxide reductase subunit B	10	0.01	0	5
Superoxide dismutase [Fe]	10	0.01	0	5
Glutamate--tRNA ligase	10	0.01	0	5
Phage shock protein B	10	0.01	0	5
Asparagine synthetase [glutamine-hydrolyzing] 1	10	0.01	0	5
tRNA 2-thiocytidine biosynthesis protein TtcA	8	0.04	0	4
putative Ni/Fe-hydrogenase B-type cytochrome subunit	8	0.04	1	5
Poly-beta-1,6-N-acetyl-D-glucosamine N-deacetylase	8	0.04	0	4
Universal stress protein	8	0.04	0	4
Alpha-agarase	8	0.04	0	4
HTH-type transcriptional repressor CarH	8	0.04	0	4
Major cardiolipin synthase ClsA	8	0.04	0	4
Biosynthetic arginine decarboxylase	8	0.04	0	4
Na(+)/H(+) antiporter subunit C1	8	0.04	0	4
Cytochrome c'	8	0.04	0	4
N-acetylglucosaminyldiphosphoundecaprenol N-acetyl-beta-D-mannosaminyltransferase	8	0.04	0	4

Table 5A-IV.2 continued.

Prokka:Prodigal call	WES	Wilcoxon p-val	Occurrence in our genomes (out of 5)	Occurrence outside of our genomes (out of 5)
Cytochrome c-554(548)	8	0.04	0	4
Glutaredoxin 3	8	0.04	1	5
Na(+)/H(+) antiporter subunit G	8	0.04	0	4
Deoxyribodipyrimidine photo-lyase	8	0.04	1	5
Amino-acid carrier protein AlsT	8	0.04	1	5
Glutaredoxin arsenate reductase	8	0.04	0	4
L-arginine-specific L-amino acid ligase	8	0.04	0	4
Putative aldehyde dehydrogenase AldA	8	0.04	0	4
Alanine racemase, biosynthetic	8	0.04	1	5
DNA protection during starvation protein 2	8	0.04	0	4
Polysialic acid transport protein KpsD	8	0.04	1	5
Diguanylate cyclase DosC	8	0.04	0	4
Carbon storage regulator	8	0.04	0	4
High-potential iron-sulfur protein isozyme 2	8	0.04	0	4
Undecaprenyl-phosphate alpha-N-acetylglucosaminyl 1-phosphate transferase	8	0.04	0	4
Na(+)/H(+) antiporter subunit D	8	0.04	0	4
Putative agmatine deiminase	8	0.04	0	4
Bifunctional transcriptional activator/DNA repair enzyme Ada	8	0.04	0	4
Endonuclease 8	8	0.04	0	4
Immunogenic protein MPB70	8	0.04	0	4
Nitric oxide reductase subunit C	8	0.04	0	4
Inner membrane protein YecN	8	0.04	0	4
L-2-hydroxyglutarate oxidase LhgO	8	0.04	0	4
putative FAD-linked oxidoreductase	8	0.04	1	5

Table 5A-IV.2 continued.

Prokka:Prodigal call	WES	Wilcoxon p-val	Occurrence in our genomes (out of 5)	Occurrence outside of our genomes (out of 5)
Protease HtpX	8	0.04	1	5
Oxygen sensor protein DosP	8	0.04	1	5
Cycloserine biosynthesis protein DcsG	8	0.04	0	4
Ferric enterobactin receptor	8	0.04	0	4
Deoxyribose-phosphate aldolase	8	0.04	0	4
Renalase	8	0.04	0	4
Zinc transporter ZupT	8	0.04	0	4

Figure 5A-IV.1. Pangenomic analysis of gene clusters in genomes from this study (pink) and those in the database (black). The tree on the right is a neighbor-joining tree using a concatenated alignment of 49 ribosomal proteins. The presence of carbon fixing and sulfur oxidizing genes is noted in green, as is the presence of high spore protein SP21 expression in our transcriptomes. The red lines at the bottom indicate the locations of highly enriched genes, including ornithine carbamoyltransferase and carbon dioxide concentrating mechanisms.

Appendix VI: Metabolites

Overall results

Metabolite data were generated in collaboration with Dr. Campagna's lab. The results presented in this appendix are average data from three technical replicates performed on two to three biological replicates from sediment samples collected during the 2016 campaign. The co-extraction of DNA, RNA, and metabolite data from the same samples allowed us to compare side-by-side the 16S rRNA gene amplicon library data, transcriptional coverage, and metabolite abundances for sediment from sites AC, AB, and F down to ~ 6 cm depth.

Over 50 metabolites were detected across all sites, with most metabolites represented in all sites. The metabolites detected were involved in amino acid metabolism (Figures 5A-VI.1 – 5A-VI.5), nucleic acid biosynthesis (Figure 5A-VI.6 – 5A-VI.7), the TCA cycle (Figure 5A-VI.8), signaling (Figure 5A-VI.9), iron uptake, sporulation, oxidative stress (Figure 5A-VI.10), vitamins (Figure 5A-VI.11) and uric acid (Figure 5A-VI.12). Heatmaps were generated to view global trends in metabolites with depth at each site (Figures 5A-VI.13 – 5A-VI 15).

DHPS and sulfolactate

Site AB had clear depth trends for some metabolites (Figure 5A-VI.13), including 2,3-dihydroxypropane-1-sulfonate (DHPS) and sulfolactate. DHPS has been shown to have a role in the Roseobacter marine food web (1) and can be used as the sole carbon source for *Ruegeria pomeroyi* DSS-3 (2). In addition, DHPS is an immediate precursor to sulfolactate, which can be excreted by cells and then remineralized by other bacteria through the Entner–Doudoroff pathway for sulfoquinovose degradation (3) or an alternative bifurcated pathway (4), representing a potential crucial and often overlooked link in the sulfur cycle (5). Sulfolactate has also been shown to play a role in the formation of spores in *Bacillus subtilis* and up to 5% of the dry weight of spores can be accumulated sulfolactic acid (6). The structure of sulfolactic acid suggests high chelating capacity, much like dipicolinic acid, also involved in spore formation (6).

Because our metabolite data does not discriminate between extracellular and intracellular metabolites, it is difficult to interpret the role of sulfolactate and DHPS in our sediments. Networks were built to understand the affiliation of metabolites with transcripts and members of the microbial community (Figure 5A-VI.16). Interestingly, DHPS is only significantly connected in the AB network and sulfolactate is only significantly connected in the AC network. All connections are negative, indicating consistent anti-correlations between all nodes in Figure 5A-VI.16. In AB, DHPS is negatively connected to transcript nodes related to energy and solute transport, including a type *c* cytochrome and a TRAP transporter which allows the uptake of succinate and malate. The negative connection to the TRAP transporter is interesting, as co-localization of genes for DHPS catabolism and transport (including a TRAP transporter) found previously in *R. pomeroyi* DSS-3 suggested co-transcription (2). Sulfolactate in the AC network is only negatively connected with a laccase domain protein, which potentially is involved in lignin degradation. Our previous observation of enhanced spore protein transcription with depth (Chapter 5) will provide a framework for future work aimed at untangling the connections between DHPS, energy metabolism, sulfur cycling, and spore formation in these sediments.

Glutathione disulfide redox coupling

Glutathione disulfide, an abundant thiol in proteobacteria, was detected at both F and AC (but not AB). Within cells under normal conditions, this metabolite nearly exclusively exists in its reduced form called glutathione. However, under conditions of oxidative stress wherein reactive oxygen species threaten damage to cellular components, glutathione acts as an antioxidant, neutralizing free radicals (7). It is because of this, the ratio of the oxidized form (glutathione disulfide) to the reduced form (glutathione) of this thiol acts a redox sensor for cells (7-9). Our metabolite data suggest that the glutathione redox couple for dealing with oxidative

stress is being induced at sites AC and F, although we were unable to find transcripts for glutathione peroxidase at either site. Likewise, we were unable to detect transcripts for glutathione reductase, which brings oxidized glutathione disulfide to its reduced state. Network analysis showed that in AC, glutathione disulfide is positively connected to kynurenic acid (Figure 5A-VI.17), a metabolite generated through tryptophan degradation that has been shown to also have antioxidant properties against reactive oxygen species in the mammal model (10). In addition, glutathione disulfide is also positively connected in our network to the family Bdellvibrionaceae, obligate aerobic bacterial parasites (11). The association between this group and mechanisms for free radical detoxification suggest that the Bdellvibrionaceae contain adaptive metabolic responses to changing redox conditions in these sediments. Future work will address the novelty of the glutathione redox couple in Bdellvibrionaceae and the detection of evidence for this in published Bdellvibrionaceae genomes.

Salicylate and iron uptake

Salicylate was detected at all sites. This metabolite is a precursor to siderophores in *Pseudomonas aeruginosa* (12, 13) and *Mycobacterium tuberculosis* (14). It has also been shown to have siderophore activity on its own (15), although its low binding constant at physiological pH makes it unlikely to compete successfully with naturally-occurring iron-scavenging ions, such as phosphate (16). Microorganisms have several iron-chelating strategies for acquiring the iron necessary for proper enzymatic function, and the genes related to these iron-chelating molecules are tightly regulated according to environmental concentrations of soluble iron (Reviewed in (17, 18)). Sediments at Kongsfjorden and Van Keulenfjorden are heavily influenced by bedrock lithology, with iron-rich conglomerates, sandstones, and shales in Kongsfjorden and iron-rich sandstones and sub-glacial pyrite oxidation in Van Keulenfjorden

supplying oxidized iron to the seabed (19). Across sites, we were not able to identify a clear relationship between transcriptional coverage for hemin receptors, siderophores, and ferrienterobactins and environmental concentrations of dissolved iron, ranging from 20 to ~200 μM with depth and across sites (unpublished data, (20)). Transcriptional coverage for the siderophore-encoding gene *ton*B was restricted to the Woeseiaceae, and although network analysis did not show a connection between this clade's relative abundance by 16S, it did show that salicylate is negatively correlated with spore protein SP21 (Figure 5A-VI.17). This protein is highly transcribed nearly exclusively by the Woeseiaceae at all sites examined (Chapter 5) and suggests a connection between spore formation and iron scavenging that will need to be explored in future work. Ongoing analysis includes identifying statistical correlations between transcriptional coverage of iron-chelating genes and environmental measurements of iron, phylogenetic affiliation of iron-chelating molecules, and understanding the role that salicylate may play in iron-scavenging within these sediments.

References

1. Durham BP, Sharma S, Luo H, Smith CB, Amin SA, Bender SJ, Dearth SP, Van Mooy BAS, Campagna SR, Kujawinski EB, Armbrust EV, Moran MA. 2015. Cryptic carbon and sulfur cycling between surface ocean plankton. Proceedings of the National Academy of Sciences 112:453-457.

2. Mayer J, Huhn T, Habeck M, Denger K, Hollemeyer K, Cook AM. 2010. 2, 3-Dihydroxypropane-1-sulfonate degraded by Cupriavidus pinatubonensis JMP134: purification of dihydroxypropanesulfonate 3-dehydrogenase. Microbiology 156:1556-1564.

3. Felux A-K, Spiteller D, Klebensberger J, Schleheck D. 2015. Entner–Doudoroff pathway for sulfoquinovose degradation in Pseudomonas putida SQ1. Proceedings of the National Academy of Sciences 112:E4298-E4305.

4. Denger K, Mayer J, Buhmann M, Weinitschke S, Smits TH, Cook AM. 2009. Bifurcated degradative pathway of 3-sulfolactate in Roseovarius nubinhibens ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfolyase. Journal of bacteriology 191:5648-5656.

5. Roy AB, Hewlins MJ, Ellis AJ, Harwood JL, White GF. 2003. Glycolytic breakdown of sulfoquinovose in bacteria: a missing link in the sulfur cycle. Applied and environmental microbiology 69:6434-6441.

6. Bonsen PP, Spudich JA, Nelson DL, Kornberg A. 1969. Biochemical studies of bacterial sporulation and germination XII. A sulfonic acid as a major sulfur compound of Bacillus subtilis spores. Journal of bacteriology 98:62-68.

7. Schafer FQ, Buettner GR. 2001. Redox environment of the cell as viewed through the redox state of the glutathione disulfide/glutathione couple. Free radical biology and medicine 30:1191-1212.

8. Townsend DM, Tew KD, Tapiero H. 2003. The importance of glutathione in human

disease. Biomedicine & Pharmacotherapy 57:145-155.

9. Lushchak VI. 2011. Adaptive response to oxidative stress: Bacteria, fungi, plants and animals. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology 153:175-190.

10. Lugo-Huitrón R, Blanco-Ayala T, Ugalde-Muniz P, Carrillo-Mora P, Pedraza-Chaverri J, Silva-Adaya D, Maldonado P, Torres I, Pinzon E, Ortiz-Islas E. 2011. On the antioxidant properties of kynurenic acid: free radical scavenging activity and inhibition of oxidative stress. Neurotoxicology and teratology 33:538-547.

11. Kuever J, Rosenberg E, DeLong E, Lory S, Stackebrandt E, Thompson F. 2014. The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria.

12. Ankenbauer RG, Cox CD. 1988. Isolation and characterization of Pseudomonas aeruginosa mutants requiring salicylic acid for pyochelin biosynthesis. Journal of bacteriology 170:5364-5367.

13. Ankenbauer R, Toyokuni T, Staley A, Rinehart K, Cox CD. 1988. Synthesis and biological activity of pyochelin, a siderophore of Pseudomonas aeruginosa. Journal of bacteriology 170:5344-5351.

14. De Voss JJ, Rutter K, Schroeder BG, Su H, Zhu Y, Barry CE. 2000. The salicylate-derived mycobactin siderophores of Mycobacterium tuberculosis are essential for growth in macrophages. Proceedings of the National Academy of Sciences 97:1252-1257.

15. Sokol P, Lewis C, Dennis J. 1992. Isolation of a novel siderophore from Pseudomonas cepacia. Journal of medical microbiology 36:184-189.

16. Chipperfield JR, Ratledge C. 2000. Salicylic acid is not a bacterial siderophore: a theoretical study. Biometals 13:165-168.

17. Wandersman C, Delepelaire P. 2004. Bacterial iron sources: from siderophores to hemophores. Annu Rev Microbiol 58:611-647.

18. Escolar L, Pérez-Martín J, De Lorenzo V. 1999. Opening the iron box: transcriptional metalloregulation by the Fur protein. Journal of bacteriology 181:6223-6229.

19. Wehrmann LM, Formolo MJ, Owens JD, Raiswell R, Ferdelman TG, Riedinger N, Lyons TW. 2014. Iron and manganese speciation and cycling in glacially influenced high-latitude fjord sediments (West Spitsbergen, Svalbard): Evidence for a benthic recycling-transport mechanism. Geochim Cosmochim Acta 141:628-655.

20. Wehrmann LM, Riedinger N, Brunner B, Kamyshny Jr A, Hubert CR, Herbert LC, Brüchert V, Jørgensen BB, Ferdelman TG, Formolo MJ. 2017. Iron-controlled oxidative sulfur cycling recorded in the distribution and isotopic composition of sulfur species in glacially influenced fjord sediments of west Svalbard. Chemical Geology 466:678-695.

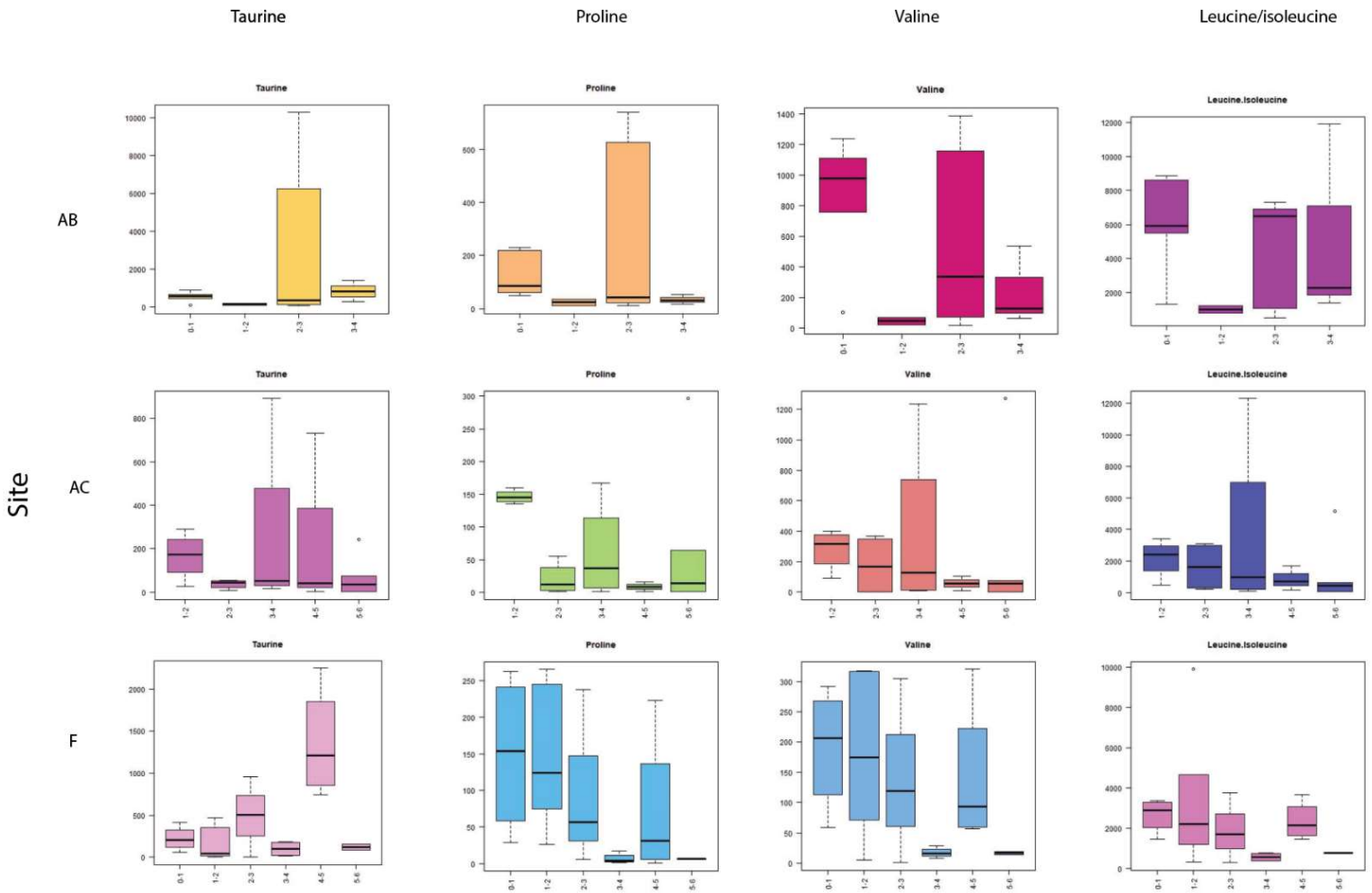

Figure 5A-VI.1. Amino acids detected across sites AB, AC, and F. Depth increases across the x-axis.

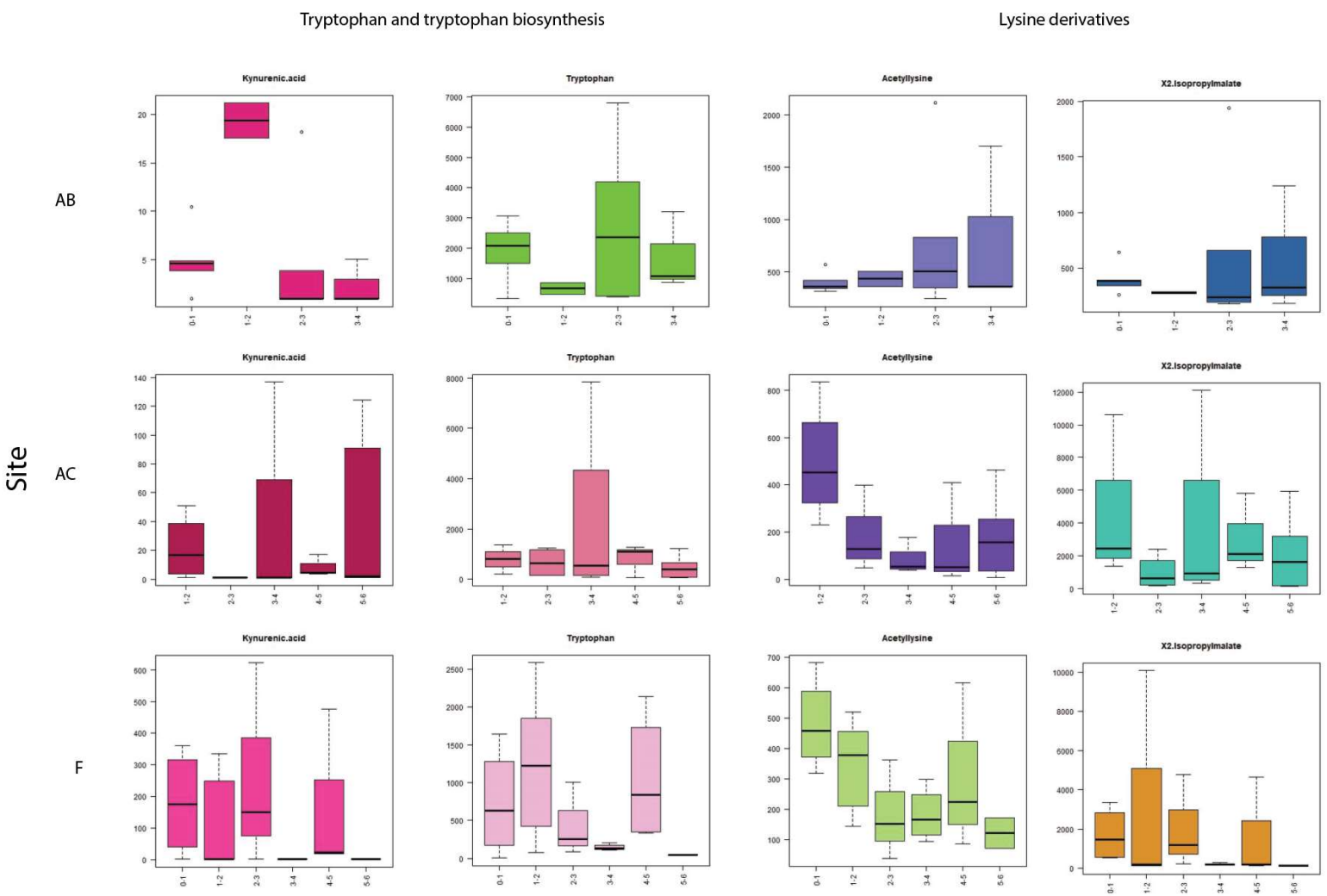

Figure 5A-VI.2. Amino acid derivatives detected in sites AB, AC, and F. Depth increases across the x-axis.

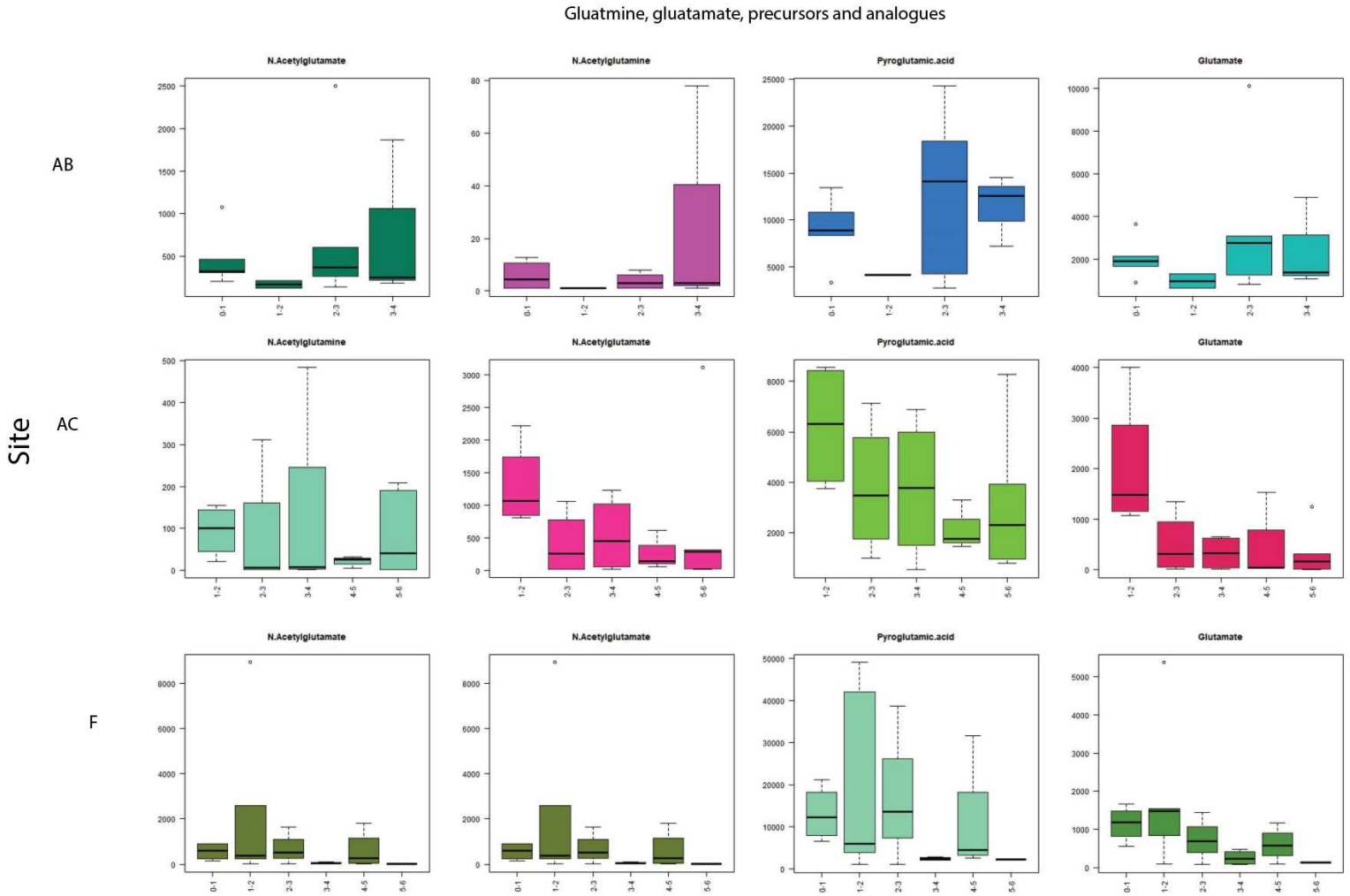

Figure 5A-VI.3. Precursors and analogues for glutamate and glutamine detected in metabolite data for sites AB, AC, and F. Depth increases along the x-axis.

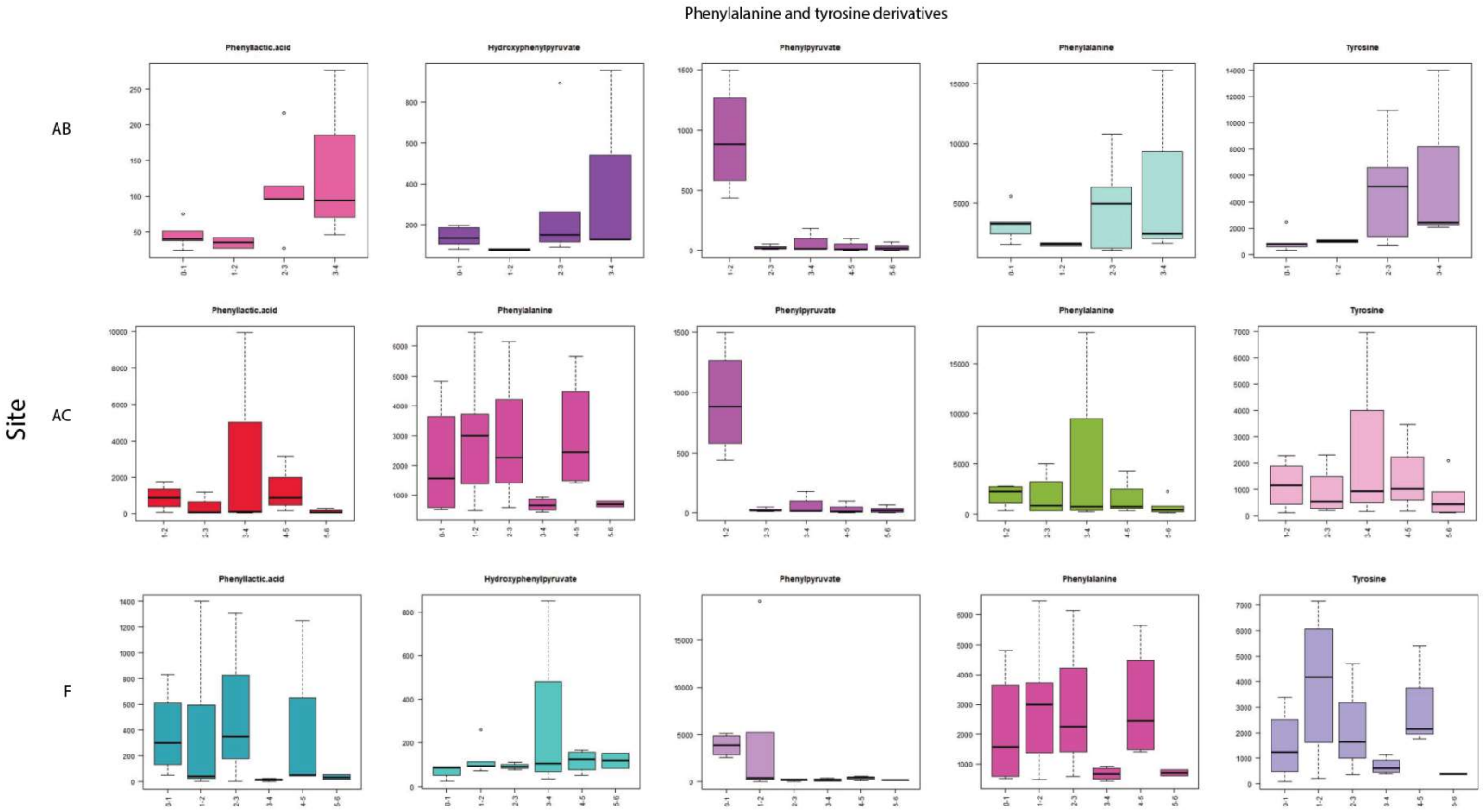

Figure 5A-VI.4. Phenylalanine and tyrosine derivatives detected in metabolite data for sites AB, AC, and F. Depth increases along the x-axis.

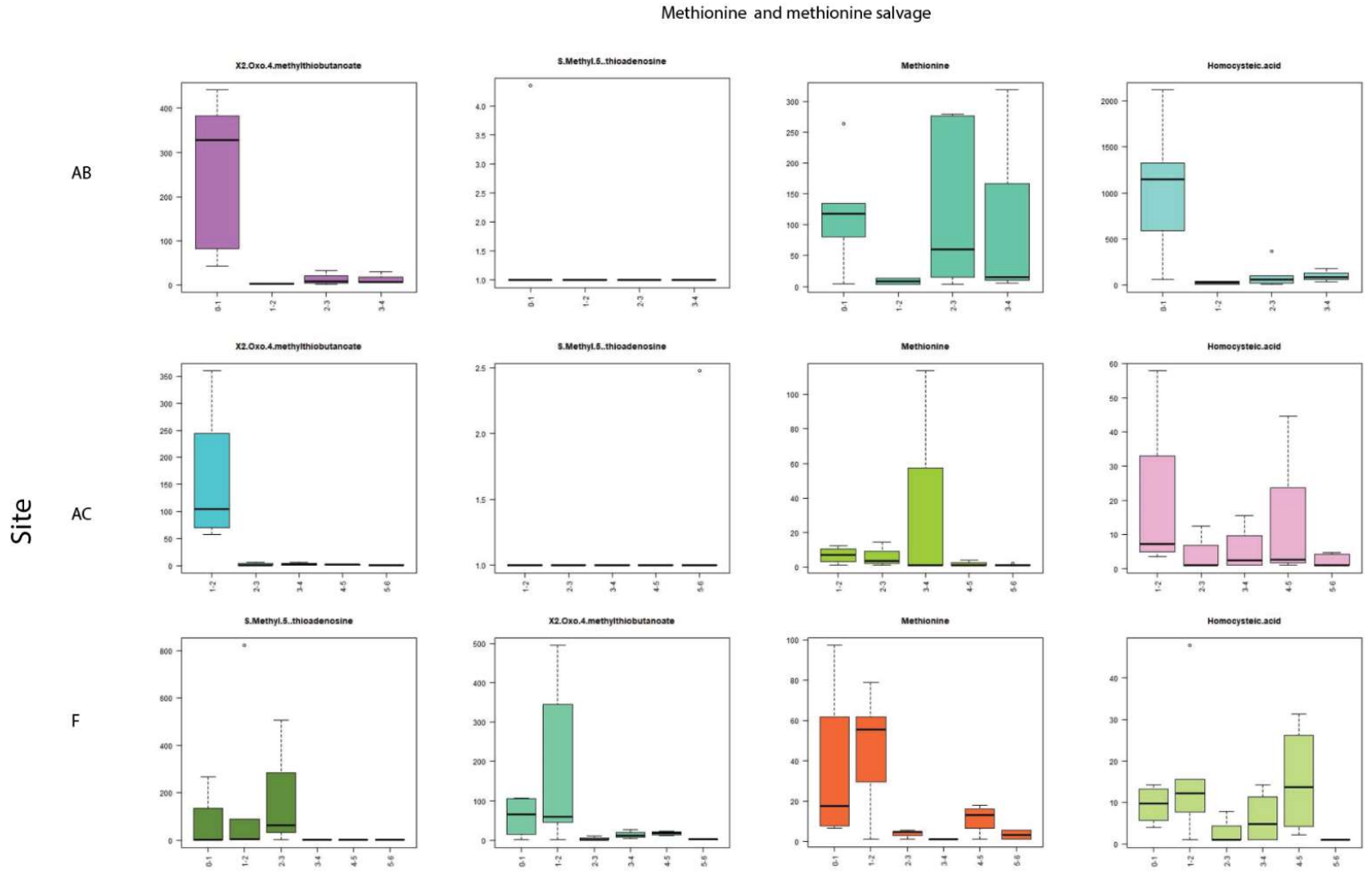

Figure 5A-VI.5. Methionine biosynthesis and scavenging metabolites detected in sites AB, AC, and F. Depth increases along the x-axis.

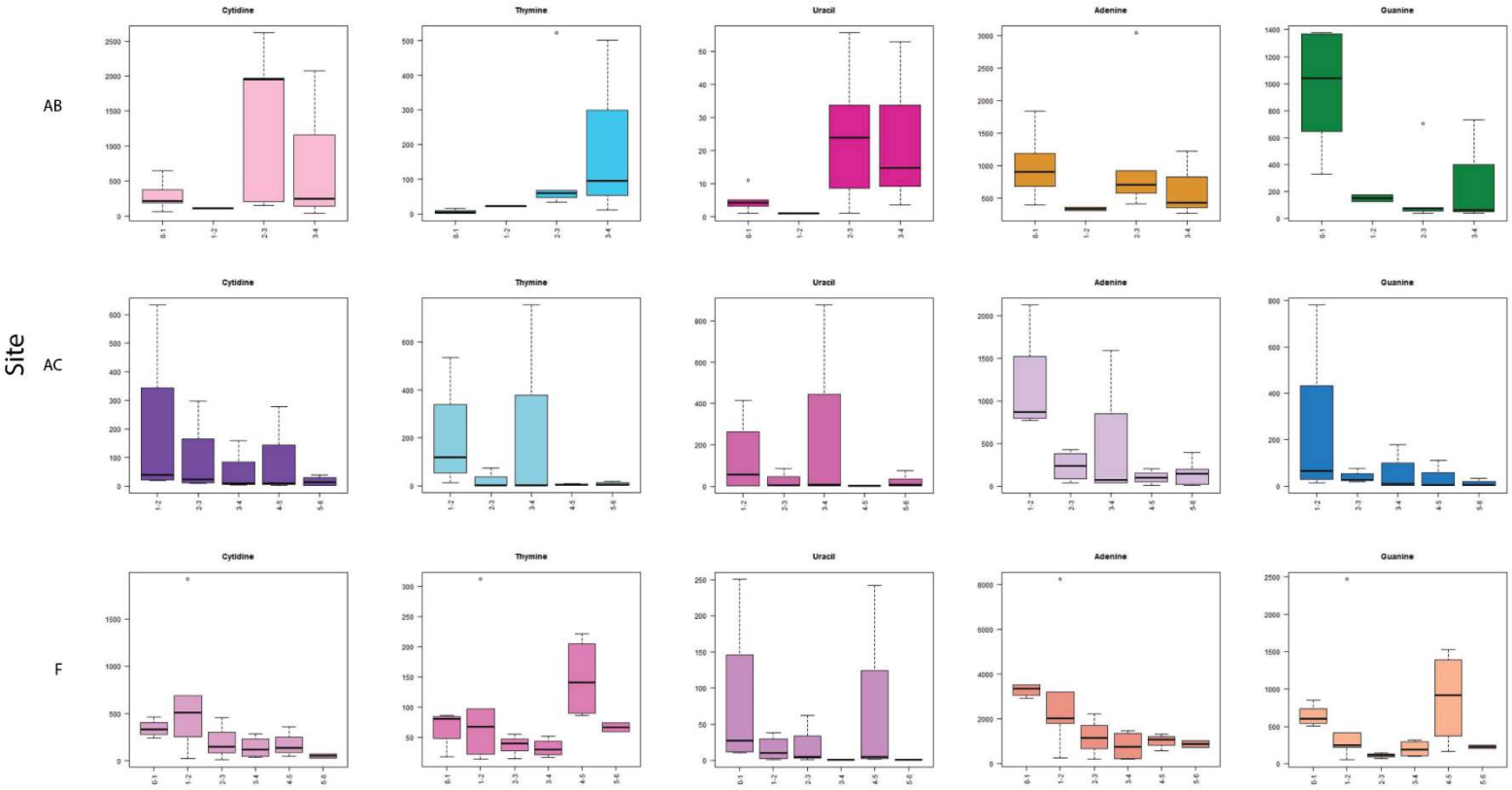

Figure 5A-VI.6. Nucleic acid metabolites detected in sites AB, AC, and F. Depth increases along the x-axis.

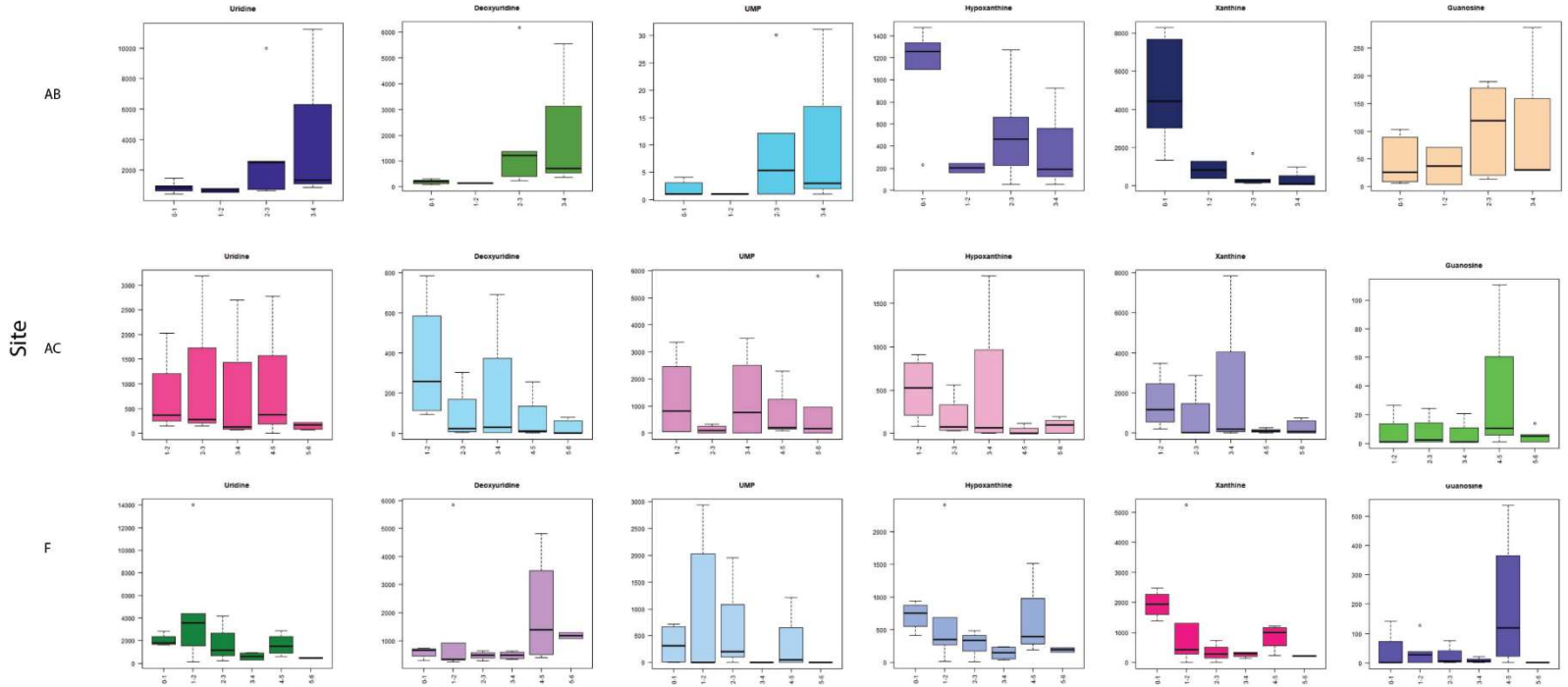

Figure 5A-VI.7. Nucleic acid derivatives and intermediates detected in sites AB, AC, and F. Depth increases along the x-axis.

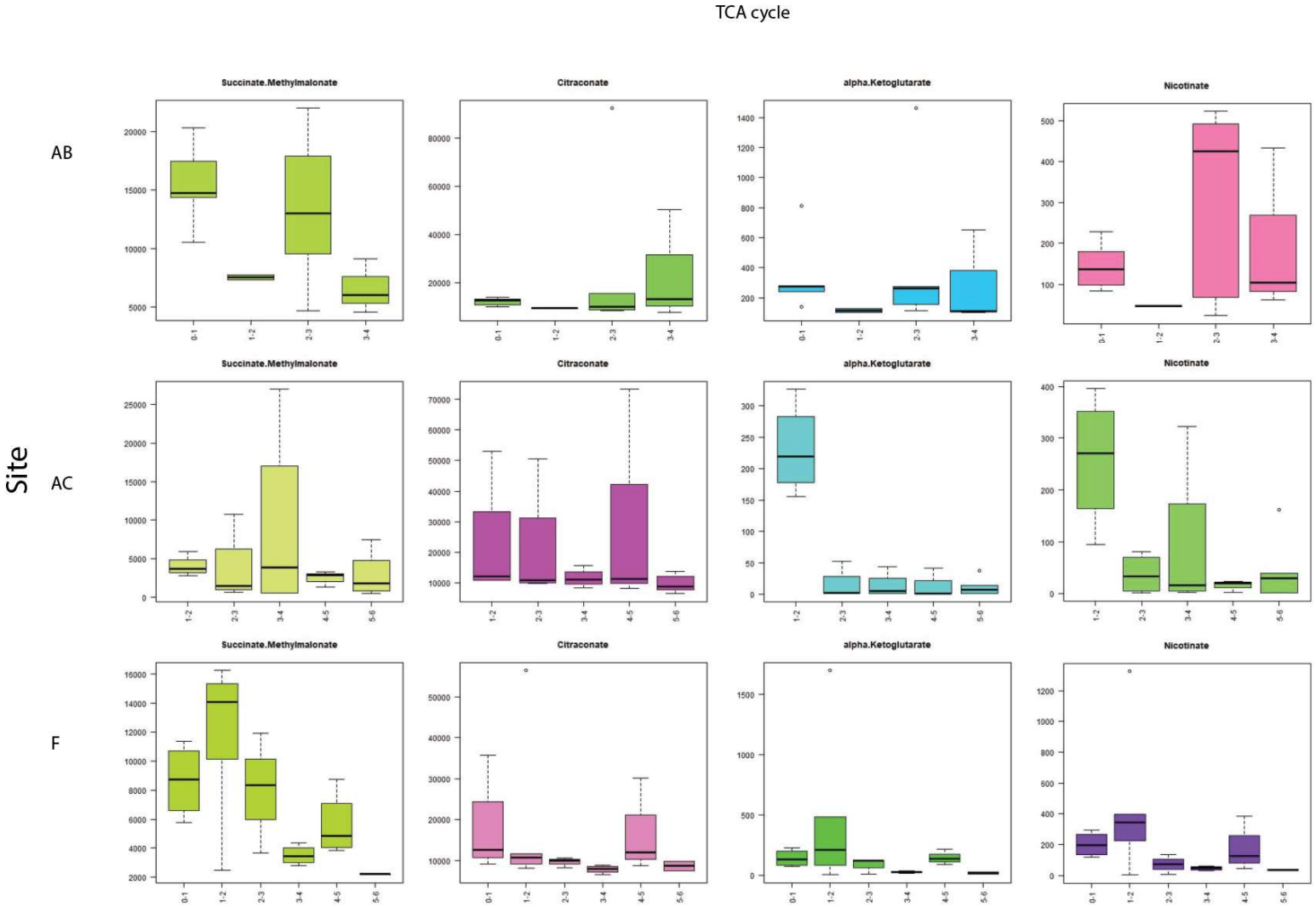

Figure 5A-VI.8. TCA cycle metabolites detected in sites AB, AC, and F. Depth increases along the x-axis.

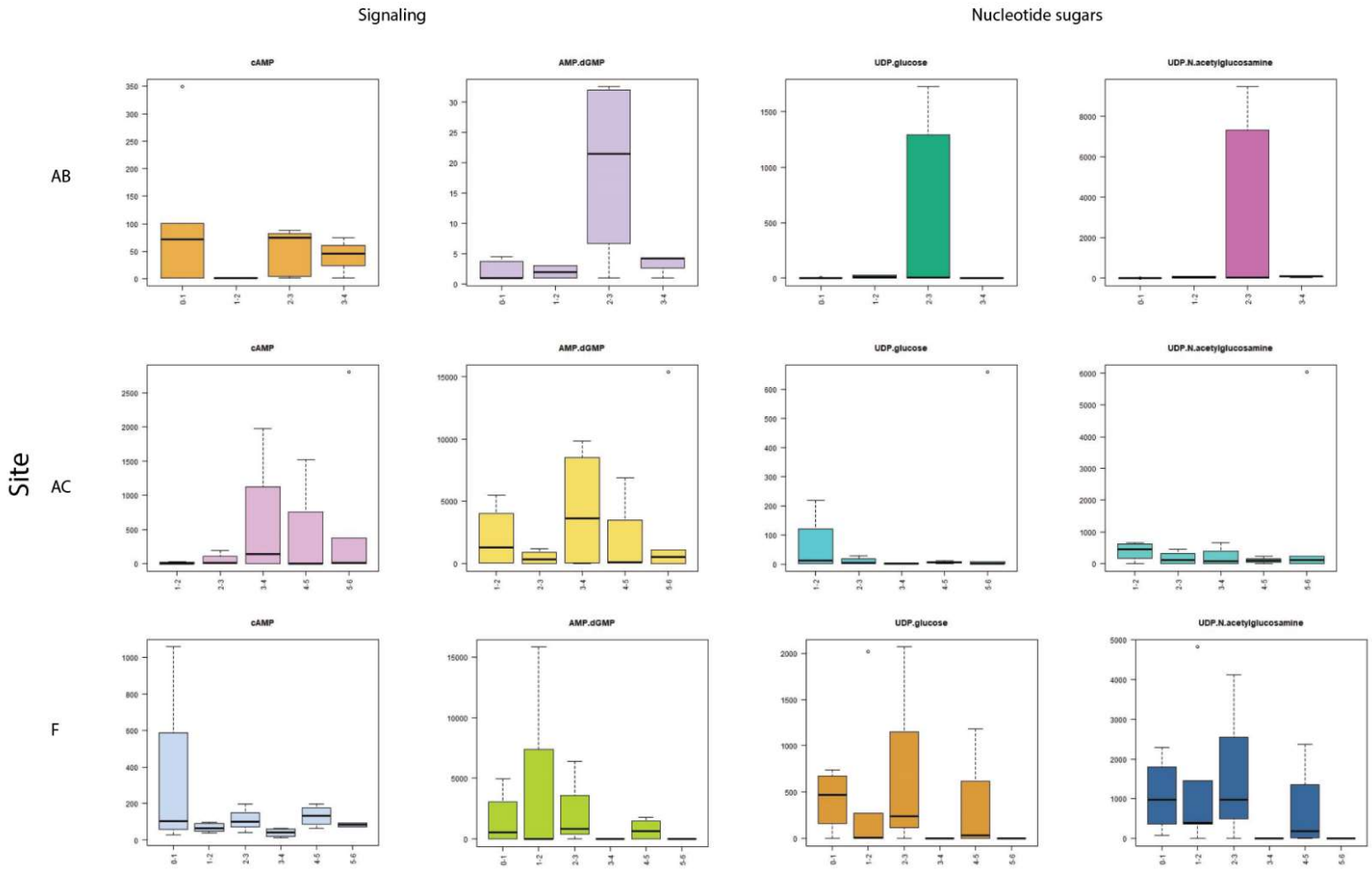

Figure 5A-VI.9. Signaling metabolites (cAMP and AMP/dGMP) and nucleotide sugar metabolites (UDP glucose and UDP N-acetylglucosamine) in sites AB, AC, and F. Depth increases along the x-axis.

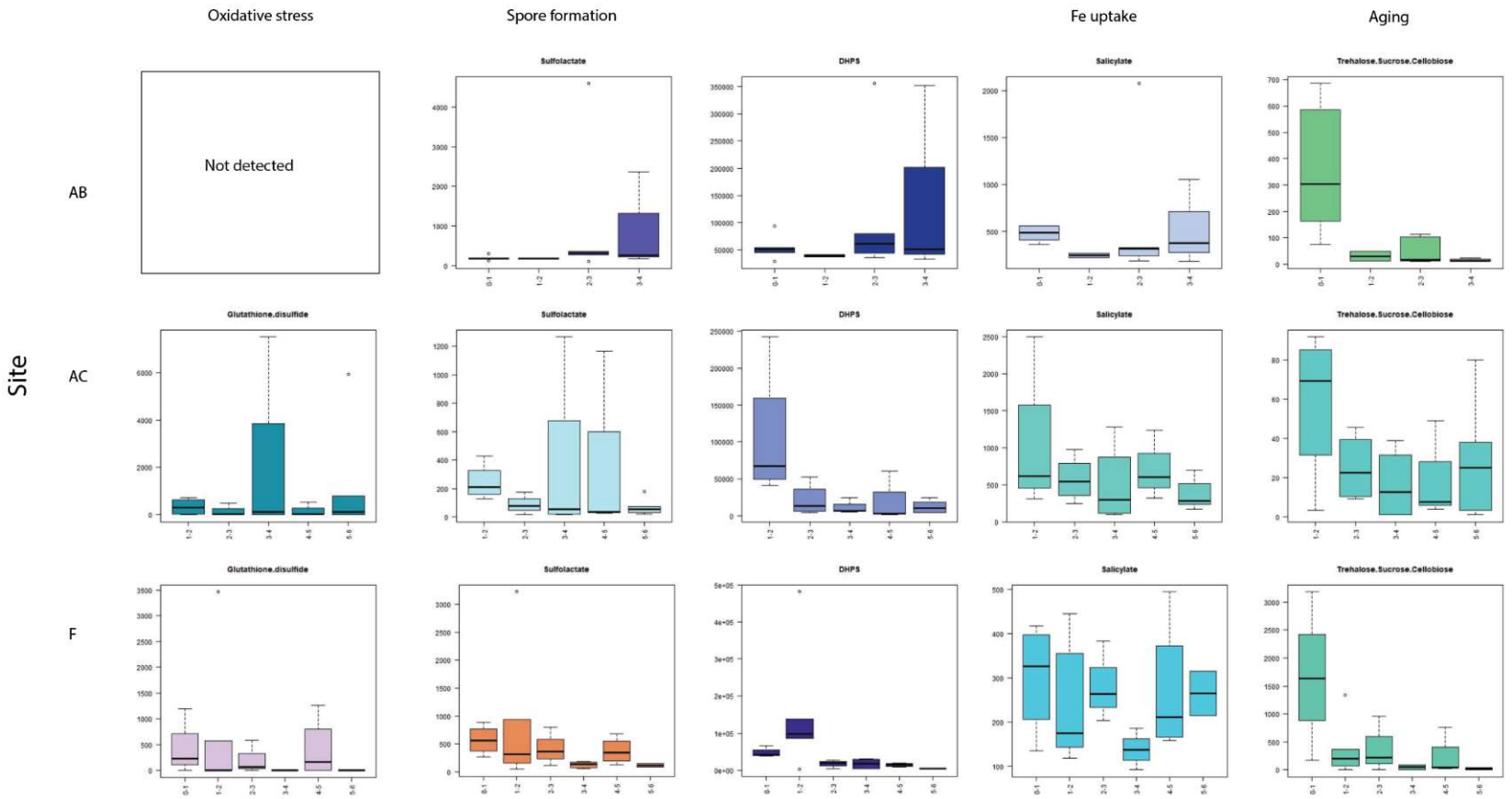

Figure 5A-VI.10. Oxidative stress, spore formation, iron uptake, and aging metabolites in sites AB, AC, and F. Depth increases along the x-axis.

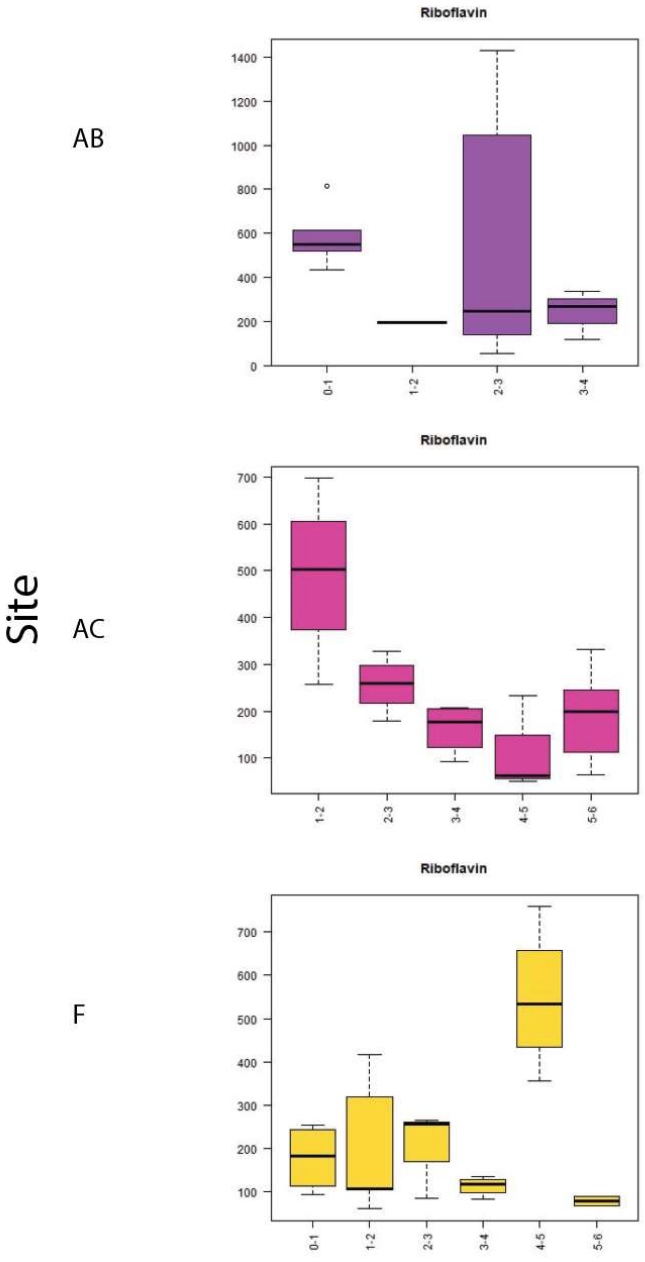

Figure 5A-VI. 11. Riboflavin detected in sites AB, AC and F. Depth increases along the x-axis.

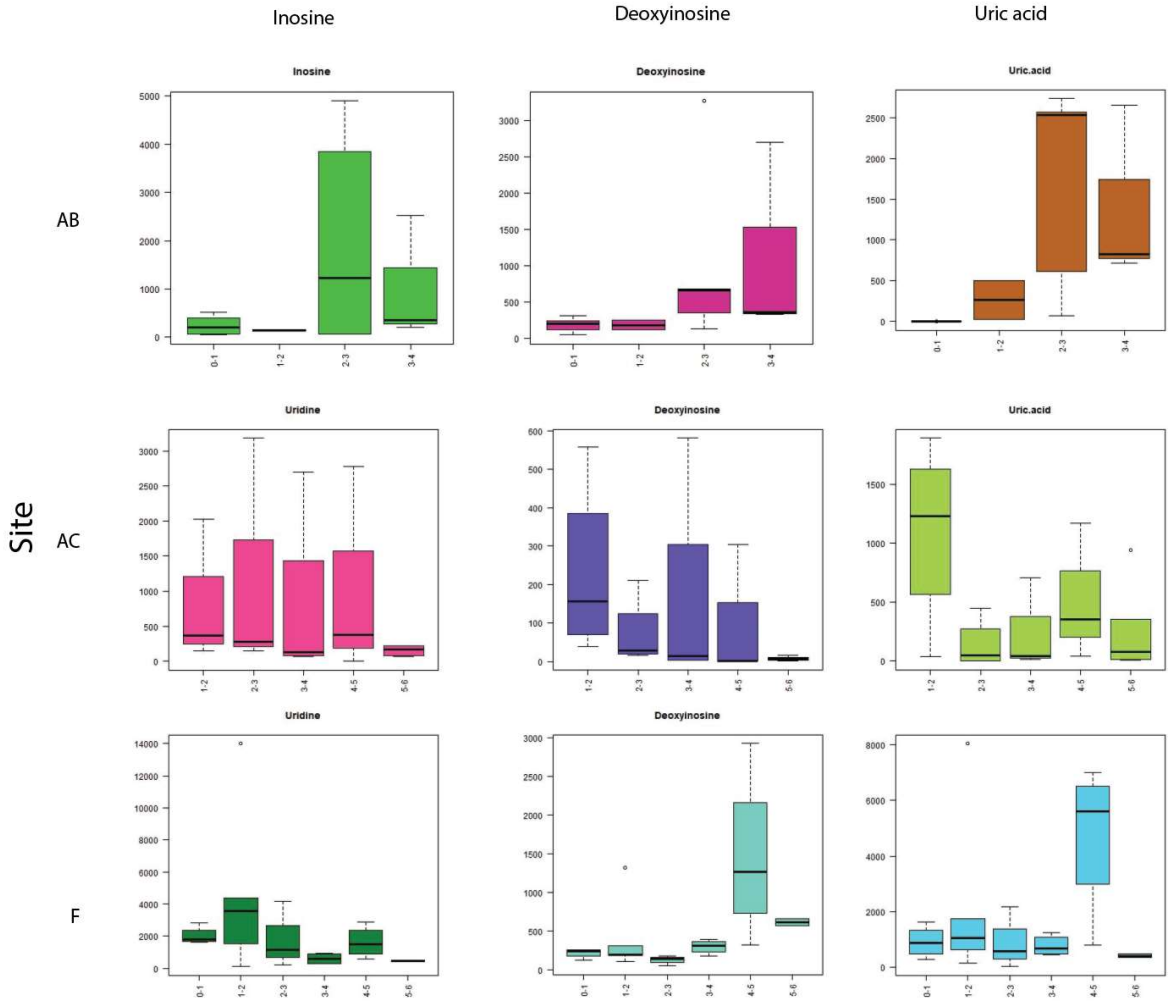

Figure 5A-VI.12. Uric acid generation metabolites for sites AB, AC, and F. Depth increases along x-axis.

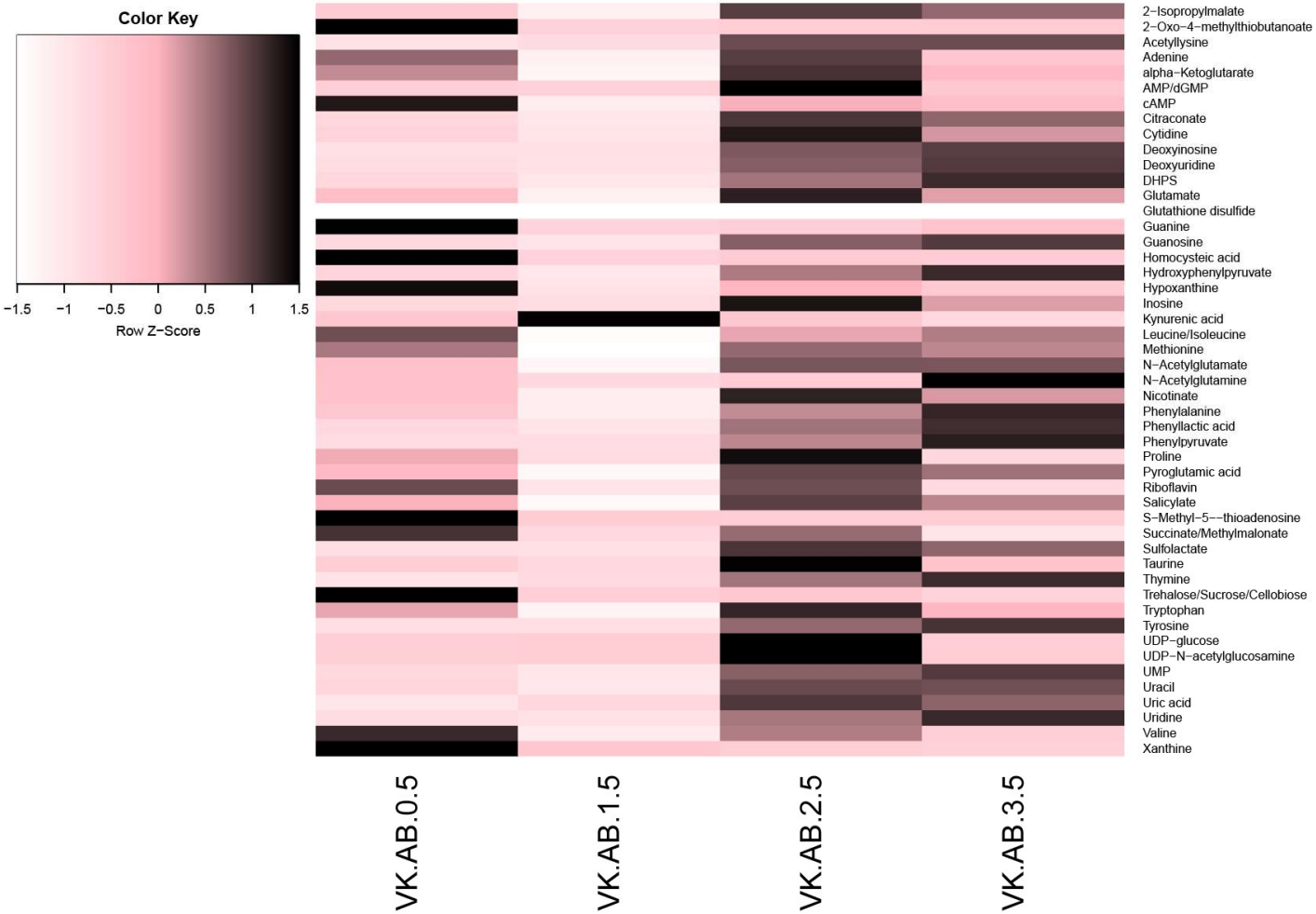

Figure 5A-VI.13. Heatmap of metabolite distribution in site AB. Depth increases along the x-axis. Z-scores were calculated for each row and indicate the standard deviation from the mean value for each metabolite.

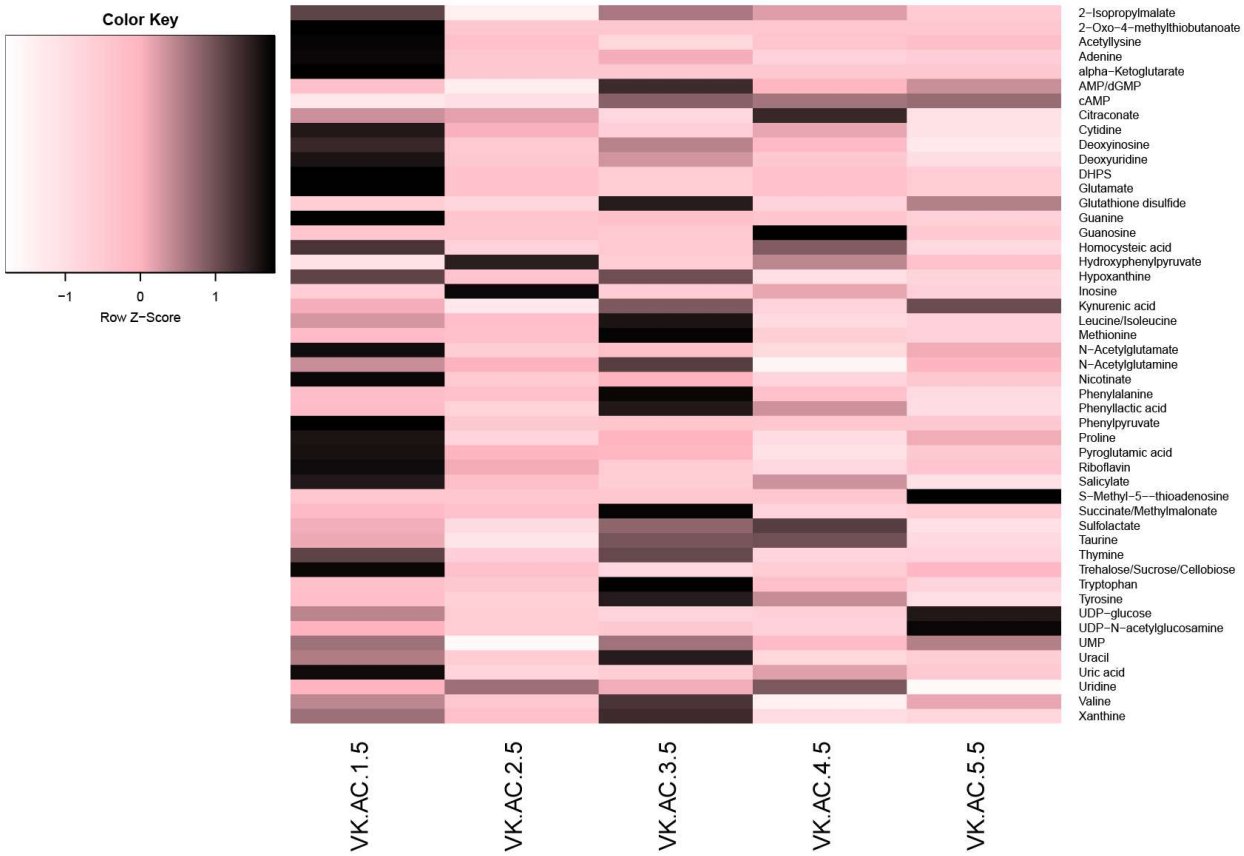

Figure 5A-VI.14. Heatmap of metabolite distribution in site AC. Depth increases along the x-axis. Z-scores were calculated for each row and indicate the standard deviation from the mean value for each metabolite.

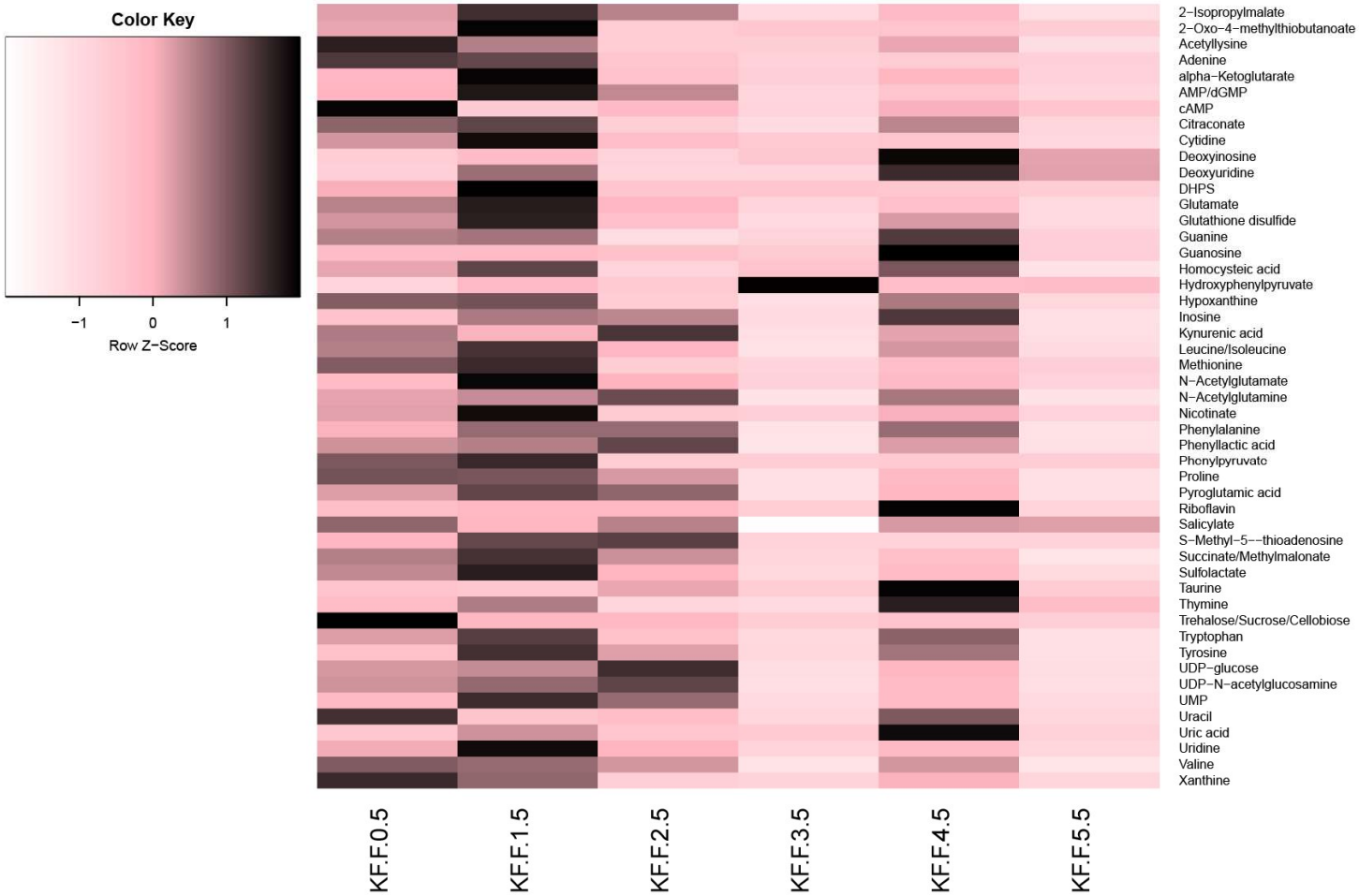

Figure 5A-VI.15. Heatmap of metabolite distribution in site F. Depth increases along the x-axis. Z-scores were calculated for each row and indicate the standard deviation from the mean value for each metabolite.

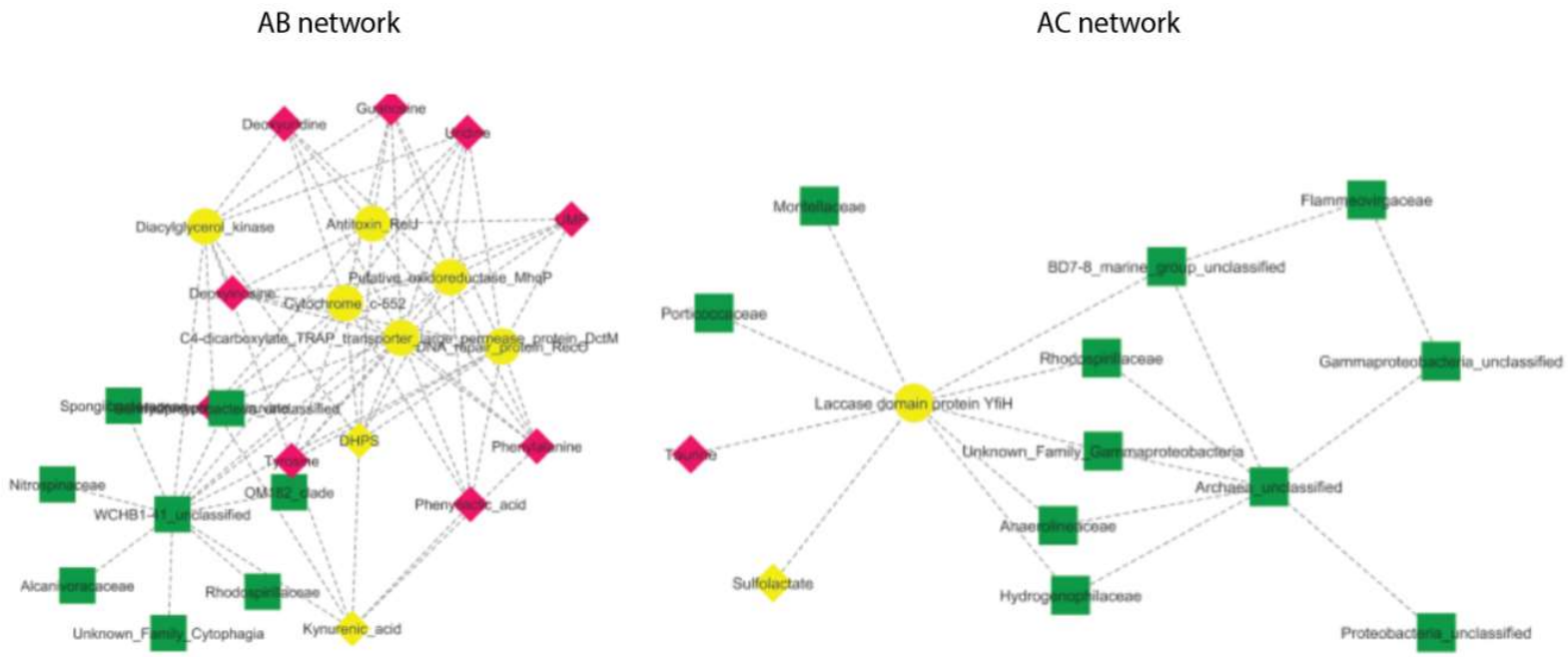

Figure 5A-VI.16. Network associations of DHPS and sulfolactate in sites AB and AC. Networks were built with eLSA using transcript coverage, metabolite peak data, and 16S relative abundances at the family level. Transcripts are represented as circles, metabolites as pink diamonds, and microbial families are green squares. The first neighbors of DHPS (for AB network) and sulfolactate (for AC network) are highlighted in yellow. All edges (lines connecting nodes) represent negative relationships.

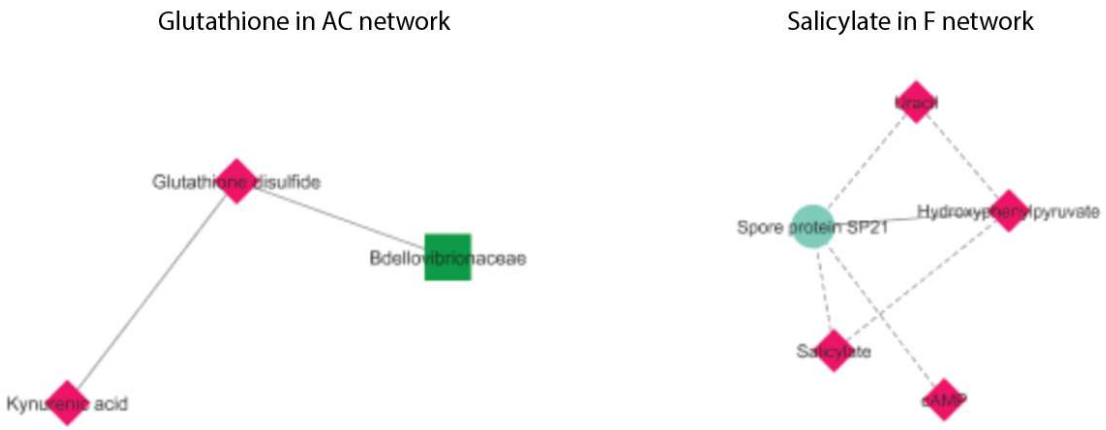

Figure 5A-VI.17. Networks for glutathione in AC and salicylate in F.

Chapter 6: Conclusion

Within the past few years, advances in molecular and sequencing techniques have allowed us glimpses into the vast subsurface biosphere. Uncovering the abundance, identity, metabolic potential, and transcriptional activity of buried microbial populations has applications for unlocking the secrets behind microbial subsistence in deeply buried sediment on geological times scales, modeling biogeochemical cycling, and predicting the trajectory of ecosystem dynamics with continued climate warming.

We have shown here that quantitative PCR (qPCR) is a reliable means for estimating numbers of bacteria and archaea within marine sediments in which fluorescent microscopy techniques, such as catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) may be technically challenging. In addition to being higher throughput compared to microscopy, qPCR is a relatively reproducible means for estimating abundance if best practices are followed.

Understanding the metabolic and physiological capabilities of subsurface microorganisms is often hampered by limited success in culturing efforts. Where isolation techniques failed, sequencing and bioinformatics tools succeeded in allowing us to delve into the genomic repertoire and adaptations within a member of the methanogenic genus *Methanosarcina* of Archaea enriched from permafrost from the McMurdo Dry Valley. Genome analysis showed that *Methanosarcina lacustris* sp. Ant1 can generate methane from diverse substrates and has structural adaptations that allow it to remain viable after being locked away in permafrost on geological time scales. Our work provides evidence that the climate-change induced deepening of permafrost active layers may provide an additional source of the greenhouse gas methane.

Climate change is a severe threat to Arctic ecosystems, in particular. Marine terminating glaciers in Svalbard serve as conduits for highly oxidized terrestrial material that make for a unique sediment environment. In this environment, differences in the ratio of autochthonous (surface water primary productivity) to allochthonous (terrigenous ancient organics and coal) material results in different oxidizing conditions for carbon-degrading microorganisms along the long axis of fjords. Low ratios of fresh organics to organic-poor sediment at the head of Van Keulenfjord, Svalbard provides the suboxic milieu that permits microbial partitioning according to redox zonation with depth, with a relatively deep zone of iron reducers and sulfur oxidizers. This iron reducing/sulfur oxidizing zone shallows as the organic matter content and lability increases at the fjord mouth. Our work served as the first high throughput cross-fjord investigation into the sequence abundance of iron and sulfur cycling clades in Van Keulenfjorden and lays the groundwork for future work aimed at predicting how glacial retreat may affect microbial community structure, metabolic function, and interaction with the carbon cycle.

The Woeseiaceae/JTB255 clade within the Steroidobacterales is a key driver of dark carbon fixation in marine sediments worldwide. In Van Keulenfjorden, 16S rRNA gene library analysis showed that this group is highly abundant, especially within the first 5 cm of sediment. Genomic reconstruction and transcriptome analysis showed that in addition to having broad respiratory flexibility under microaerobic to anoxic conditions, the Woeseiaceae transcribe stress-mitigation and spore-forming proteins with continued burial. The ability to form spores, likely a type of myxospore, is perhaps a clue to the secret behind its high relative abundance across a vast, global biogeographical range.

Vita

Joy "her-parents-didn't-bother-to-give-her-a-middle-name" Buongiorno Altom was born in Brooklyn, New York, but soon traded concrete and subway trains for creeks and trailer parks after relocating to middle Tennessee when she was 10 years of age. Here, she spent her childhood catching red ear slider turtles and beating boys in bicycle races. In high school, she was in the top 10% of her class, president of the honor society, and the highest-achieving JROTC goth that Hendersonville High School has ever seen. Like the microorganisms she studies, she took advantage of the few resources she was lucky to have and decided to make a better life for herseslf. At 17, she left home and took several jobs on campus to pay her way through a Bachelor's degree at Tennessee Technological University in Cookeville, TN. It was here she cultivated her love for herpetology and geology, and a certain guy that she would marry (only 7) years later. She relocated to Knoxville, TN for her Master's degree in geoscience (the herpetologist wasn't taking new graduate students). Upon completion of her degree—still utterly confused as to what a zoologist/microbialite specialist is supposed to do with this skillset—she decided to take a hard left and get her Ph.D. in microbiology. She has met every challenge with optimism, excitement, and drive, surprising even herself when she learned Unix. Her science has allowed her to explore Martian analogues in the high Andes mountains, geobiologically-magical hot springs in California, sleepy mountains of Colorado (that soon after flooded—climate change is real!), and the Spitzbergen Archipelago, which is unmatched in its beautiful scenery and people. In her spare time, Joy dances with swords in a local tribal fusion belly dancing troupe, watches videos of metal concerts while running on the treadmill pretending she's in a mosh pit, provides educational outreach to underserved communities, serves as the coordinator of a local chapter of an international nonprofit organization for scientific literacy, writes letters to her state

representatives, and tries to train her Pitbull mix to "shake" (it isn’t going well). After graduation, Joy will get another tattoo (with the money her parents gifted to her to get some of her tattoos removed) and move to Washington, D.C to work as a NASA Astrobiology Institute fellow at Carnegie Institute.