INVESTIGATION OF MICROBIAL INTERACTIONS AND ECOSYSTEM DYNAMICS IN A LOW O_2 CYANOBACTERIAL MAT

by

Alexander A. Voorhies

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Earth and Environmental Sciences)
in The University of Michigan
2014

Doctoral Committee:

Assistant Professor Gregory J. Dick, Chair
Associate Professor Matthew R. Chapman
Assistant Professor Vincent J. Denef
Professor Daniel C. Fisher
Associate Professor Nathan D. Sheldon

To my wife Hannah

ACKNOWLEDGEMENTS

Funding for the research presented here was provided by the National Science Foundation, the University of Michigan CCMB Pilot Grant, and a Scott Turner research award from the University of Michigan Earth and Environmental Sciences Department. I am grateful for the opportunities to explore my scientific interests this funding has made possible.

I would like to acknowledge my co-authors and collaborators, who offered advice, guidance and immeasurable assistance throughout this process. Gregory J. Dick, Bopi Biddanda, Scott T. Kendall, Sunit Jain, Daniel N. Marcus, Stephen C. Nold and Nathan D. Sheldon are co-authors on CHAPTER II, which was published in Geobiology in 2012; Gregory J. Dick was a co-author on CHAPTER III, which is in preparation for publication; and Gregory J. Dick, Sarah D. Eisenlord, Daniel N. Marcus, Melissa B. Duhaime, Bopaiah A. Biddanda and James D Cavalcoli are co-authors on Chapter IV, which is in preparation for publication.

I thank my committee members: Matt Chapman, Nathan Sheldon, Vincent Denef and Dan Fisher. Their input and guidance throughout my graduate studies has kept me on track and made significant enhancements to this dissertation. Specifically I would like to thank my advisor and chair, Greg Dick. I first met Greg working on tangential projects at UC Berkeley, and came to the University of Michigan specifically to work with him. He has given me enough support and guidance to succeed in my graduate studies, but also enough room to grow as a scientist and follow my own scientific interests. I will be forever grateful for his understanding, patience and mentorship.

I would like to thank the National Oceanic and Atmospheric Administration's Thunder Bay National Marine Sanctuary for logistical and sampling assistance. Specifically I wish to thank Jeff Grey, Russ Green, Wayne Lusardi, Joe Hoyt and Tane Casserly, and the many crews of the *R/V Storm* for their sampling assistance, observations of the Middle Island Sinkhole and the great deal of care and effort they put into assisting us. This research would not have been possible without their assistance.

Thank you to my lab mates and fellow graduate students who were always available to help sort out issues that arose or commiserate upon the life of a graduate student. Specifically I would like to thank Sunit Jain for extensive assistance with script writing and DNA assembly theory. Thank you to Melissa Duhaime for help with viral identification and excellent advice. Thank you to Jim Cavalcoli for de novo genome assembly assistance, and Ryan Lesniewski for metagenome assembly assistance. Thank you to Sarah Eisenlord for statistical analysis, editing, and company on long runs that helped to keep me sane through this process.

I wish to thank my family and friends who have supported me financially and emotionally during my years at UM. Thank you to my parents Dave and Laurie for the nudge to do something different with my life that came at just the right time, and their support throughout this process. Thank you to my Mother-Outlaw Jean for the bags of coffee, chocolate and editing assistance on multiple occasions. Finally, thank you to my Wife Hannah, for following me into the cold Midwest, and her tireless support and love.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Cyanobacteria are believed to be responsible for the oxygenation of the Earth's atmosphere and oceans, which enabled the evolution of metabolisms that depend on O_2. Little is known about cyanobacteria adapted to low-O_2, sulfidic conditions, which dominated the oceans when oxygenic photosynthesis first evolved. To better understand how such cyanobacteria function and contribute to biogeochemistry, metagenomics and metatranscriptomics were used to characterize modern cyanobacterial mats that thrive under low-O_2, sulfidic conditions in the Middle Island Sinkhole (MIS) of Lake Huron.

Metagenomics revealed a consortium of microorganisms that regulate biogeochemical cycling at the sediment/water interface. The mats were dominated by *Phormidium,* a cyanobacterium that was inferred to perform anoxygenic photosynthesis in the presence of sulfide based on (i) primary production rate experiments, (ii) expression of sulfide quinone reductase, and (iii) a high ratio of transcripts for photosystem I to photosystem II. Combined with excess organic matter, chemical reductants and rapid utilization of O_2 by respiration, this anoxygenic photosynthesis makes the MIS mats a net sink for O_2. Such anoxygenic cyanobacterial mats were likely widespread under the low-O_2 conditions of the Proterozoic, and may help to explain why atmospheric O_2 levels remained low for much of Earth's history.

Genome sequences were reconstructed for the dominant mat organisms, and transcript abundance was used to identify organisms expressing metabolic pathways that regulate geochemical cycling at MIS. *Desulfobacterales* were responsible for mediating production of sulfide, which likely contributes to hypoxia at MIS and regulates oxygenic versus anoxygenic

photosynthesis by *Phormidium*. Members of the Proteobacteria were found to perform aerobic oxidation of various sulfur species, H_2 and CO. Viral predation was detected by two way exchange of DNA between *Phormidium* and PhV1, an abundant virus at MIS. *Phormidium* used viral DNA within a CRISPR system to defend itself, while PhV1 was found to possess a host derived *nblA* gene, which breaks down photosynthetic pigments. Overall, this work suggests that ancient cyanobacterial mats were not necessarily a source for O_2, and that sulfide concentration, metabolic products from other organisms, viral predation, and light availability could all influence cyanobacterial production of O_2 in low-O_2 environments.

CHAPTER I

Introduction

1.1 Microbial mediation of geochemical cycles

In recent years, there has been growing recognition of the extent to which microbes are responsible for mediating Earth's geochemical cycling of major elements such as carbon and nitrogen as well as less abundant elements like trace metals (Falkowski et al 2008). Particularly important is the fact that microbes obtain energy for metabolism and growth by catalyzing redox reactions. This changes the redox state of molecules, which can affect inherent properties such as solubility and volatility and influence the mobility of elements in the environment. One of the best examples of how microbes can influence global biogeochemistry is oxygenic photosynthesis, during which sunlight is used to oxidize water, releasing O_2 as a byproduct. This process transformed the atmosphere and oceans from an anoxic state on the early Earth to the oxic planet now inhabited by oxygen-requiring plants and animals.

This chapter reviews current knowledge of the cyanobacteria, the group of organisms likely responsible for the largest redox shift in Earth's history, the oxygenation of the Earth. The Middle Island Sinkhole (MIS), a modern analog of ancient cyanobacterial mats, is introduced as our sampling site, followed by a review of modern molecular genomics techniques and goals. Finally, there is a summary of the research chapters that follow, the major findings of each, and how they fit together to improve our understanding of these organisms and their environment.

1.2 Cyanobacteria and the oxygenation of the Earth

Cyanobacteria are hailed as the originators of oxygenic photosynthesis, and are largely thought responsible for the oxygenation of the Earth's atmosphere. Earth's first long-term oxygenation was long thought to have begun some 2.4 billion years ago (Ga) with the Great Oxidation Event (Farquhar et al 2011, Holland 2006), but recent evidence suggests that there were appreciable levels of O_2 much earlier (Crowe et al 2013). Oxygenic photosynthesis is the process of harvesting light energy to split water for the purpose of energy production, creating O_2 gas as a byproduct. Anoxygenic photosynthesis is believed to have existed before the evolution of the cyanobacteria, and does not create oxygen as a byproduct, using only one photosystem (I), instead of the two photosystems (I and II) required for oxygenic photosynthesis (Blankenship et al 2007).

Though most studied cyanobacteria can only perform oxygenic photosynthesis, require O_2, and are inhibited by sulfide, some cyanobacteria can withstand low O_2 and high sulfide and still perform photosynthesis (Buhring et al 2011, Cohen et al 1986, Jorgensen et al 1986b). Cohen et al. (1986) describe four responses to sulfide by cyanobacteria. Most cyanobacteria are sulfide sensitive, and photosynthesis completely shuts down in the presence of sulfide. Some cyanobacteria are sulfide tolerant, and able to continue performing oxygenic photosynthesis under sulfidic conditions. The last two groups perform anoxygenic photosynthesis in the presence of sulfide, either in addition to oxygenic photosynthesis or by completely switching over to anoxygenic photosynthesis (Cohen et al 1986).

The oxygenation of the earth was a series of redox shifts that slowly added oxygen to the oceans and atmosphere and took almost two billion years (Farquhar et al 2011, Holland 2006).

A rise in atmospheric O_2 is detected in the geologic record around 2.4 Ga, but remained at roughly 1% of current atmospheric levels of O_2 in an intermediate stage of oxygenation until around 0.8 Ga when it rose to near current levels (Holland 2006, Kump 2008). This period was characterized by low levels of oxygen in the atmosphere, pervasive ocean anoxia, and localized or intermittent euxina (Lyons et al 2009, Reinhard et al 2013). These variable redox conditions favor cyanobacteria that can switch between oxygenic and anoxygenic photosynthesis depending on the redox state of their environment, and can tolerate varying levels of O_2 and sulfide (Johnston et al 2009a). Such organisms have been implicated in prolonging the intermediate stage of atmospheric oxygenation by performing a combination of oxygenic and anoxygenic photosynthesis. During anoxygenic photosynthesis, metabolically versatile cyanobacteria could have produced elemental sulfur instead of O_2 as a waste product, thus reducing net O_2 production and preventing it from accumulating in higher concentrations (Johnston et al 2009a). While H_2S is a better electron acceptor than water, H_2S became scarce as O_2 built up in the oceans, and water was available in almost unlimited supply (Cohen et al 1986, Jorgensen et al 1986b).

Cyanobacteria capable of photosynthesis under sulfidic conditions are rare in the modern world, but were likely ubiquitous in ancient oceans and instrumental in the evolution of more complex life on Earth by providing oxygen as a widely available electron acceptor. Metabolically versatile cyanobacteria can still be found in sulfidic environments, including ice covered lakes in Antarctica (Andersen et al 2011), sinkholes in Lake Huron (Biddanda et al 2006), hot springs in Yellowstone Park (Castenholz 1977b), and they have been implicated in emerging coral diseases around the world (Myers et al 2007, Myers and Richardson 2009). These environments have redox-stratified conditions similar to those thought to exist during the

Proterozoic (Johnston et al 2009a, Shen et al 2003, Walter and Bauld 1983), and support complex cyanobacterial mats that show similarity to ancient mats found in the geologic record.

Fossilized microbial mats are some of our earliest evidence for the presence of life on Earth, dating back as far as 3.4 Ga (Allwood et al 2006, Walter and Paterson 1994) or 3.5 Ga (Noffke et al 2013). These fossilized remains of microbial communities are abundant in the rock record, and suggest that microbial mats likely dominated microbial mediation of geochemical cycling until around 0.6 Ga, when multicellular life becomes abundant (Anbar and Knoll 2002, Hayes and Waldbauer 2006). Though mats preserved in the rock record are often interpreted as cyanobacteria, distinguishing them from mats formed by other organisms on the basis of morphology can be difficult. Stable organic molecules (or the degraded remains of those molecules) often preserved in the rock record (Welander et al 2009) may provide methods to distinguish groups of microbes based on the unique molecules they produce (Summons et al 1999). Studying modern cyanobacterial mats that are analogs of ancient systems could increase our understanding of ancient cyanobacteria by documenting the conditions under which metabolically versatile cyanobacteria succeed, the other microbes with which they co-occur, how these organisms affect their environment, and the unique molecular products they produce and that could be preserved in the rock record.

1.3 The Middle Island Sinkhole

The research presented here was initiated by the discovery of brilliant purple microbial mats living under low-O_2, sulfidic conditions in a karst sinkhole located in the Thunder Bay National Marine Sanctuary, in Lake Huron, MI. The mats were discovered by scuba divers from the National Oceanic and Atmospheric Administration who were working with a team of researchers

searching for shipwrecks in the Great Lakes, and investigated the 23m depression in the seafloor. MIS and its inhabitants have been a curiosity to local fisherman and scuba divers for some years. Divers noted that a meter of dense salty water reeking of sulfide (even through waterproof dry suits) covered the mats of bacteria. The groundwater at MIS was found to be different than normal lake water, with lower temperature, pH, and concentration of dissolved oxygen, as well as higher salt concentrations from various sulfate salts.

Phormidium was identified as the dominant organism in the mats by 16s ribosomal RNA marker gene studies (Biddanda et al 2006), and additional research indicated that cyanobacteria at MIS were likely performing anoxygenic photosynthesis, or significant amounts of chemosynthesis, making them a unique subject for study (Voorhies et al 2012). The combination of environmental redox conditions with the presence of cyanobacteria potentially capable of anoxygenic photosynthesis, suggested MIS and *Phormidium* would make an informative model of ancient cyanobacterial, organisms we know very little about. In order to characterize low O_2, sulfide tolerant cyanobacteria and their interactions with other members of the community, including viruses, samples from the mat were collected from MIS and DNA and RNA from the mats were sequenced for genomic analysis.

1.4 Microbial genomics and the age of the genome

The field of microbial genomics is a relatively new one, and uses DNA and RNA sequencing to study organisms based on the content of their genome and the expression of their genes (Allen and Banfield 2005, Wilmes et al 2009). The advancement of high-throughput shotgun sequencing of environmental DNA (metagenomics) (Hurwitz and Sullivan 2013, Wrighton et al 2012) and RNA (metatranscriptomics) (Gilbert et al 2008, Shi et al 2011) has allowed the study

of entire microbial communities, often consisting of organisms which had previously only been identified by studies of environmental 16s rRNA marker genes (Rajendhran and Gunasekaran 2011, Tajima et al 1999). This allows us to gauge the metabolic potential and gene expression for environmentally important organisms, without the need to first bring them into laboratory culture. While growing microbes in the lab can provide a wealth of knowledge about that organism, many organisms are resistant to cultivation. This leaves the vast majority of microbial diversity as a proverbial black box, where we see the reactants and products of microbial metabolisms in the environment, but the mechanisms driving that chemistry remain a mystery. By rebuilding the genomes of the organisms performing this chemistry, we can assign specific functions to groups of organisms, and better understand how different groups of microbes interact with their neighbors and shape the environment around them.

Genomics can help us understand how microbes influence their environment and neighbors. Sequencing genomes allows individual genes to be compared to those of known and cultured organisms, thereby assigning function to those genes. When gene content and expression for a community of microbes is examined in this way, complex relationships become evident where the products of one microbial metabolism become the reactants for other microbial metabolisms. For example, sulfate might be reduced to sulfide by one organism, while another organism could oxidize the sulfide back to sulfate. Investigating only one of these organisms leaves out half of the story, but by investigating entire communities we can assess interactions between microorganisms and the environment.

Metagenomics and metatranscriptomics provide snapshots of nucleic acids from all community members present in high enough abundance to detect, including viruses. While bacteria and archaea are estimated to be the most populous cellular organisms on the planet,

viruses that infect them are estimated to be an order of magnitude more abundant (Rohwer and Thurber 2009, Suttle 2007). Viruses are a prevalent source of microbial mortality (Avrani et al 2011), and can shape the function, diversity, and evolution, of microbial communities (Lindell et al 2004). Viruses are obligate parasites comprised of short segments of DNA or RNA packaged in a coat of protein or simple lipid membrane. Their genomes usually encode a handful of genes required to reproduce viral particles, however viruses require the protein machinery of a host cell in order to reproduce themselves, often destroying the host in the process of viral reproduction. Despite the ubiquitous nature of viruses, their representation in genome and gene databases is a small fraction of that available for bacteria, archaea or even eukaryotes. Viral influence on microbial communities is poorly understood, due in part to poor genomic databases and difficulty in isolating viruses, but also because viral genomes are very small by comparison to microbes and complex eukaryotes, so environmental DNA sequencing recovers proportionally less of their DNA. This makes environments like MIS special, because having a dominant organism yields higher genome coverage for viruses attacking that organism. This also allows virus and host to be sequenced in the same sample, introducing less procedural bias than if they were sampled and processed separately.

As an indicator of the impact of viruses on microbial communities, many bacteria and archaea use the CRISPR/CAS system (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR associated genes) to defend themselves from invading forms of nucleic acids such as viruses (and plasmids) by incorporating small segments of DNA from the invader into their genomes for recognition and destruction of subsequent invaders (Makarova et al 2011a, Makarova et al 2011b, Sorek et al 2008). By investigating the CRISPR region of a genome, viruses that have previously attacked that organism can be identified (Tyson and Banfield 2008).

By the same token, viruses often incorporate genes from their hosts into their own genome, allowing a virus to be linked to its host even if the host does not contain a CRISPR system (Hurwitz and Sullivan 2013, Lindell et al 2004, Lindell et al 2007).

Investigating all members of a microbial community based on their genome, and the genomes of those in close proximity to them can help expose complex interactions that would be otherwise undetectable. Many of these interactions have large impacts on the environment in modern systems, which can give us a better understanding of ancient systems under similar conditions.

1.5 Organization of the dissertation

This dissertation comprises three research chapters which characterize a cyanobacterial mat community thriving under low-O_2 sulfidic conditions. In Chapter II, genomes of two cyanobacterial genera, *Phormidium* and *Oscillatoria*, were recovered from 2007 MIS samples, and are described and characterized along with analysis of process rates, mineralogy, and isotopic composition of the mats. The mats were found to be a net sink for O_2, and genes encoding sulfide quinone oxidoreductase (SQR) were found in the genome of both *Phormidium* and *Oscillatoria*, suggesting their involvement in anoxygenic photosynthesis.

In Chapter III, genomes were recovered for 32 groups of microbes at MIS sampled at seven time points between 2007 and 2012, and transcripts of functional genes were analyzed to assess the roles of individual community members in mat biogeochemistry. Gene expression indicates that Phormidium is performing anoxygenic photosynthesis and that photosynthetic gene regulation differs from previously studied cyanobacteria. Microbes responsible for geochemical cycling at MIS are identified and their contributions to redox chemistry are measured by

transcript abundance of metabolic marker genes. Finally, genomes for three organisms of novel phyla with no cultured representative are reported, constituting the first genomic data available for two of them.

In Chapter IV, two genotypes of a virus ("PhV1") infecting *Phormidium* are described and tracked over the course of five years, and interactions between virus and host are detected based on exchange of genetic material. Viral DNA from PhV1 was found in the CRISPR system encoded on the *Phormidium* genome, and PhV1 was found to encode a *Phormidium* derived *nblA* gene, which encodes an enzyme that breaks down phycobilisomes, a major component of the photosynthetic apparatus. Viral and host abundances were tracked over time, and correlations of abundance suggest that viruses play a major role in ecosystem dynamics. Chapter V summarizes the results of these studies and places them in the context of understanding ancient cyanobacterial mats that lived under low-O_2 sulfidic conditions.

1.6 References

Allen EE, Banfield JF (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology* **3:** 489-498.

Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW (2006). Stromatolite reef from the Early Archaean era of Australia. *Nature* **441:** 714-718.

Anbar AD, Knoll AH (2002). Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297:** 1137-1142.

Andersen DT, Sumner DY, Hawes I, Webster-Brown J, Mckay CP (2011). Discovery of large conical stromatolites in Lake Untersee, Antarctica. *Geobiology* **9:** 280-293.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011). Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474:** 604-608.

Biddanda BA, Coleman DF, Johengen TH, Ruberg SA, Meadows GA, VanSumeran HW *et al* (2006). Exploration of a submerged sinkhole ecosystem in Lake Huron. *Ecosystems* **9:** 828-842.

Blankenship R, Sadekar S, Raymond J (2007). The evolutionary transition from anoxygenic to oxygenic photosynthesis. In: Falkowski PG, Knoll AH (eds). *Evolution of primary producers in the sea*. Elsevier. pp 21-35.

Buhring SI, Sievert SM, Jonkers HM, Ertefai T, Elshahed MS, Krumholz LR *et al* (2011). Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems. *Geobiology* **9:** 166-179.

Castenholz RW (1977). Effect of sulfide on blue-green-algae of hot springs .2. Yellowstone-national-park. *Microbial Ecology* **3:** 79-105.

Cohen Y, Jorgensen BB, Revsbech NP, Poplawski R (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl Environ Microbiol* **51:** 398-407.

Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R *et al* (2013). Atmospheric oxygenation three billion years ago. *Nature* **501:** 535-+.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320:** 1034-1039.

Farquhar J, Zerkle AL, Bekker A (2011). Geological constraints on the origin of oxygenic photosynthesis. *Photosynth Res* **107:** 11-36.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al* (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *Plos One* **3**.

Hayes JM, Waldbauer JR (2006). The carbon cycle and associated redox processes through time. *Philosophical Transactions of the Royal Society B-Biological Sciences* **361:** 931-950.

Holland HD (2006). The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci* **361:** 903-915.

Hurwitz BL, Sullivan MB (2013). The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *Plos One* **8**.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106:** 16925-16929.

Jorgensen BB, Cohen Y, Revsbech NP (1986). Transition from Anoxygenic to Oxygenic Photosynthesis in a Microcoleus-Chthonoplastes Cyanobacterial Mat. *Appl Environ Microbiol* **51:** 408-417.

Kump LR (2008). The rise of atmospheric oxygen. *Nature* **451:** 277-278.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci U S A* **101:** 11013-11018.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al* (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449:** 83-86.

Lyons TW, Anbar AD, Severmann S, Scott C, Gill BC (2009). Tracking Euxinia in the Ancient Ocean: A Multiproxy Perspective and Proterozoic Case Study. *Annual Review of Earth and Planetary Sciences* **37:** 507-534.

Makarova KS, Aravind L, Wolf YI, Koonin EV (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct* **6**.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P *et al* (2011b). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467-477.

Myers JL, Sekar R, Richardson LL (2007). Molecular detection and ecological significance of the cyanobacterial genera Geitlerinema and Leptolyngbya in black band disease of corals. *Appl Environ Microbiol* **73:** 5173-5182.

Myers JL, Richardson LL (2009). Adaptation of cyanobacteria to the sulfide-rich microenvironment of black band disease of coral. *Fems Microbiology Ecology* **67:** 242-251.

Noffke N, Christian D, Wacey D, Hazen RM (2013). Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the ca. 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia. *Astrobiology* **13:** 1103-1124.

Rajendhran J, Gunasekaran P (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research* **166:** 99-110.

Reinhard CT, Planavsky NJ, Robbins LJ, Partin CA, Gill BC, Lalonde SV *et al* (2013). Proterozoic ocean redox and biogeochemical stasis. *Proc Natl Acad Sci U S A* **110:** 5357-5362.

Rohwer F, Thurber RV (2009). Viruses manipulate the marine environment. *Nature* **459:** 207-212.

Shen Y, Knoll AH, Walter MR (2003). Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423:** 632-635.

Shi YM, Tyson GW, Eppley JM, DeLong EF (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *Isme Journal* **5:** 999-1013.

Sorek R, Kunin V, Hugenholtz P (2008). CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology* **6:** 181-186.

Summons RE, Jahnke LL, Hope JM, Logan GA (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400:** 554-557.

Suttle CA (2007). Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H *et al* (1999). Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *Fems Microbiology Ecology* **29:** 159-169.

Tyson GW, Banfield JF (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10:** 200-207.

Voorhies AA, Biddanda BA, Kendall ST, Jain S, Marcus DN, Nold SC *et al* (2012). Cyanobacterial life at low O2: community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. *Geobiology* **10:** 250-267.

Walter GH, Paterson HEH (1994). The implications of paleontological evidence for theories of ecological communities and species richness. *Australian Journal of Ecology* **19:** 241-250.

Walter MR, Bauld J (1983). The association of sulphate evaporites, stromatolitic carbonates and glacial sediments: examples from the Proterozoic of Australia and the Cainzoic of Antarctica. *Precambrian Research* **21:** 129-148.

Welander PV, Hunter RC, Zhang L, Sessions AL, Summons RE, Newman DK (2009). Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **191:** 6145-6156.

Wilmes P, Simmons SL, Denef VJ, Banfield JF (2009). The dynamic genetic repertoire of microbial communities. *Fems Microbiology Reviews* **33:** 109-132.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al* (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337:** 1661-1665.

CHAPTER II

Cyanobacterial life at low O_2: Community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat

Alexander A. Voorhies[1], Bopi Biddanda[4], Scott T. Kendall[4], Sunit Jain[1,3], Daniel N. Marcus[1], Stephen C. Nold[5], Nathan D. Sheldon[1], and Gregory J. Dick[1,2,3,*]

[1]Deptartment of Earth and Environmental Sciences
[2]Department of Ecology and Evolutionary Biology
[3]Center for Computational Medicine and Bioinformatics
[4]Annis Water Resources Institute, Grand Valley State University, Muskegon, MI, USA,
[5]Biology Department, University of Wisconsin-Stout, Menomonie, WI 54751, USA

*correspondence: gdick@umich.edu; ph: 734.763.3228; fax: 734.763.4690;

Published in Geobiology, 10: 250–267. doi: 10.1111/j.1472-4669.2012.00322.x
Article first published online: 8 Mar 2012

Abstract

Cyanobacteria are renowned as the mediators of Earth's oxygenation. However, little is known about the cyanobacterial communities that flourished under the low-O_2 conditions that characterized most of their evolutionary history. Microbial mats in the submerged Middle Island Sinkhole (MIS) of Lake Huron provide opportunities to investigate cyanobacteria under such persistent low-O_2 conditions. Here venting groundwater rich in sulfate and low in O_2 supports a unique benthic ecosystem of purple-colored cyanobacterial mats. Beneath the mat is a layer of carbonate that is enriched in calcite and to a lesser extent dolomite. *In situ* benthic metabolism

chambers revealed that the mats are net sinks for O_2, suggesting primary production mechanisms other than oxygenic photosynthesis. Indeed, ^{14}C-bicarbonate uptake studies of autotrophic production show variable contributions from oxygenic and anoxygenic photosynthesis and chemosynthesis, presumably due to supply of sulfide. These results suggest the presence of either facultatively anoxygenic cyanobacteria or a mix of oxygenic/anoxygenic types of cyanobacteria. Shotgun metagenomic sequencing revealed a remarkably low-diversity mat community dominated by just one genotype most closely related to the cyanobacterium *Phormidium autumnale*, for which an essentially complete genome was reconstructed. Also recovered were partial genomes from a second genotype of *Phormidium* and several *Oscillatoria*. Despite the taxonomic simplicity, diverse cyanobacterial genes putatively involved in sulfur oxidation were identified, suggesting a diversity of sulfide physiologies. The dominant *Phormidium* genome reflects versatile metabolism and physiology that is specialized for a communal lifestyle under fluctuating redox conditions and light availability. Overall, this study provides genomic and physiologic insights into low-O_2 cyanobacterial mat ecosystems that played crucial geobiological roles over long stretches of Earth history.

2.1 Introduction

Cyanobacteria mediated Earth's oxygenation and thus played a central role in geochemical and biological evolution. They are widely recognized as the innovators of oxygenic photosynthesis, in which water provides electrons for photosynthesis and O_2 is released as a byproduct (Blankenship et al 2007). This cyanobacterial metabolism is thought to have driven a significant increase in atmospheric-O_2 concentration ~2.4 billion years ago known as the great oxidation event (GOE) (Bekker et al 2004). Conversely, recent work also suggests that cyanobacteria capable of anoxygenic photosynthesis may have subsequently perpetuated an extended low-O_2

phase of Earth's history (Johnston et al 2009b). This intermediate stage of Earth's redox history lasted for at least a billion years and was characterized by a low-O_2 atmosphere and redox-stratified oceans, where sulfide-O_2 interfaces would have been prevalent (Johnston et al 2009b, Lyons et al 2009). Despite the large portion of cyanobacterial evolution that occurred while O_2 was scarce, and the critical geobiological turning points that occurred under such conditions (Falkowski et al 2008, Johnston et al 2009b), little is known about the genetic or physiological characteristics of cyanobacteria that thrive under persistent sulfide-rich and/or O_2-limited conditions.

Modern cyanobacteria exhibit a range of physiologies in the presence of sulfide. Most are highly sensitive to sulfide due to irreversible blockage of the H_2O-splitting component of photosystem II (Cohen et al 1986, Miller and Bebout 2004a). However, cyanobacteria inhabiting anoxic or hypoxic environments that are regularly exposed to sulfide have developed strategies for sulfide tolerance and even utilization (Buhring et al 2011, Castenholz 1976, Castenholz 1977a, Garlick et al 1977b, Oren et al 1977). Cohen et al. described several adaptations to sulfide, ranging from sulfide-resistant oxygenic photosynthesis to the ability to utilize sulfide as the electron donor for anoxygenic photosynthesis (Cohen et al 1986). Different cyanobacterial species sharing the same oxic/anoxic interfacial environment often exhibit distinct sulfide physiologies, with different sulfide optima and tolerance (Garlick et al 1977b, Jorgensen et al 1986a). Such cyanobacteria that are capable of tolerating sulfide or using it for anoxygenic photosynthesis are phylogenetically diverse and spread throughout the phylum Cyanobacteria (Miller and Bebout 2004a).

Despite the phylogenetically widespread nature of anoxygenic photosynthesis amongst the cyanobacteria, the biochemical mechanisms of this process and its genetic underpinnings

have been studied in just a few cyanobacterial strains, primarily *Geitlerinema* sp. PCC 9228 (formerly *Oscillatoria limnetica*). When confronted with sulfide in the presence of light, this organism rapidly switches from oxygenic to anoxygenic photosynthesis by an inducible process that requires protein synthesis (Cohen et al 1975b, Oren and Padan 1978). Photosystem I receives electrons from sulfide and transfers them to the electron transport chain to drive proton pumping (Belkin and Padan 1978), and extracellular globules of elemental sulfur are generated as an end-product (Cohen et al 1975a). Biochemical and genetic methods have identified genes encoding the enzyme that oxidizes sulfide, sulfide quinone reductase (SQR), which transfers electrons from sulfide to the quinone pool (Arieli et al 1994b, Bronstein et al 2000a, Schutz et al 1997). In *Geitlerinema* sp. 9228, these sulfide-derived electrons are used for photosynthesis or nitrogen fixation, whereas in the cyanobacterium *Aphanothece halophytica* and the yeast *Schizosaccharomyces pombe* SQR is thought to oxidize sulfide for the purpose of detoxification (Bronstein et al 2000a).

Investigation of modern cyanobacteria inhabiting low-O_2 environments can provide insights into the biological processes that influenced Earth's oxygenation. Particularly relevant to understanding the rise of O_2 on Earth are questions surrounding the evolution of anoxygenic photosynthesis in the cyanobacteria, mechanisms by which versatile cyanobacteria regulate oxygenic vs. anoxygenic photosynthesis, and factors that affect competition between versatile cyanobacteria and other anoxygenic bacteria. Many studies have focused on stratified cyanobacterial mat communities where O_2 and sulfide concentrations fluctuate on diel cycles; often cyanobacteria are exposed to sulfidic conditions at night and oxic conditions during the day (Richardson and Castenholz 1987). However, few studies have focused on cyanobacteria that thrive under persistent low-O_2 conditions.

Here we investigate cyanobacterial mats inhabiting such a persistently low-O_2 environment, the submerged Middle Island Sinkhole (MIS) in Lake Huron (Figure 2.1). Groundwater that gently vents into the MIS bottom (water depth 23m) has significantly different physical and chemical properties than Lake Huron water, with a lower temperature (7–9 vs. 4–25^{0}C), lower pH (~7.1 vs. 8.3), lower concentrations of dissolved oxygen (0–2 vs. 5–11 mg L^{-1}), lower oxidation-reduction potential (-134 vs. 500 mV) and higher specific conductivity (~2.3 vs. 0.3 mS cm^{-1}) (Biddanda et al 2009). The high conductivity of venting groundwater is attributable to high concentrations of dissolved sulfate (1,250 mg L^{-1}), carbonate (48 mg L^{-1}), and chloride (25 mg L^{-1}) ions derived from interactions with subsurface Devonian evaporites (Black 1983, Ruberg et al 2008). This dense groundwater forms a thin (~1 m), visibly stratified benthic layer that persists perennially except for brief disruptions due to major storms (Ruberg et al 2008). The low-O_2 conditions of the groundwater inhibit typical Lake Huron biological communities, favoring purple-colored cyanobacterial mats with finger-like protrusions (Figure 2.2) that have not been found elsewhere in the Great Lakes (Biddanda et al 2009). Beneath the mats are stratified microbial layers of sulfide-oxidizing bacteria, sulfate-reducing bacteria, and methanogens (Nold et al 2010a). Molecular diversity studies (Nold et al 2010a) have revealed that they are dominated by cyanobacteria remarkably similar (>98% 16S rRNA gene sequence identity) to *Phormidium autumnale* isolates from the Arctic and Antarctic (Comte et al 2007, Taton et al 2006a, Taton et al 2006b), and host to distinctive archaeal and eukaryotic communities (Nold et al 2010c). Furthermore, carbon, nitrogen and sulfur stable isotopic signatures of sinkhole chemistry have been detected in the surrounding environment and food web (Sanders et al 2011).

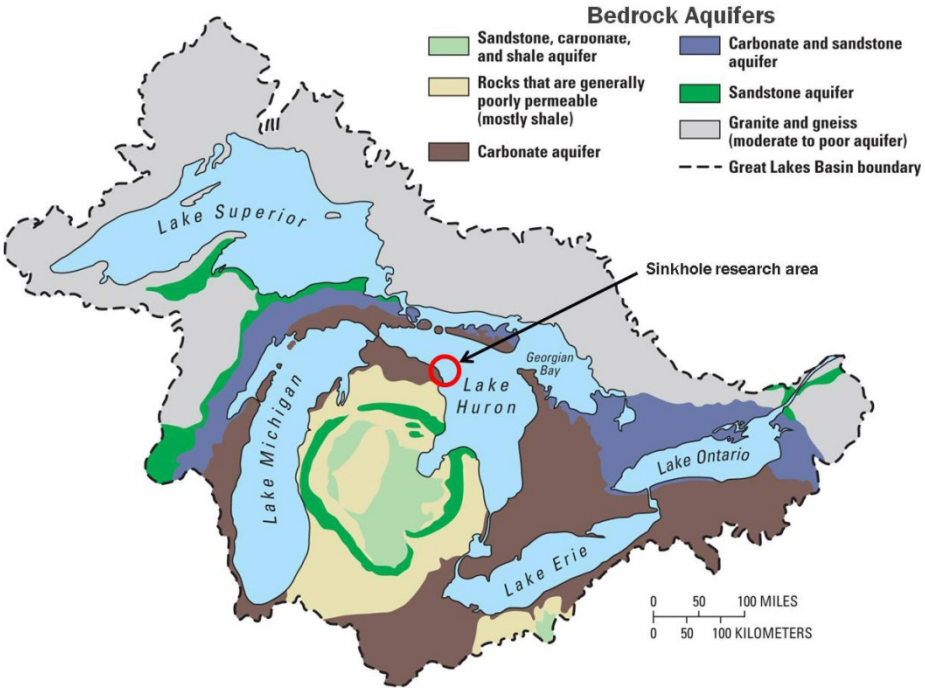


Figure 2.1 Location map of the study area and geologic map of bedrock aquifers of the Great Lakes Basin. The focus of this study is the Middle Island Sinkhole, one of many submerged karst sinkholes in the Thunder Bay National Marine Sanctuary, Lake Huron (modified from Ruberg et al. 2008, Biddanda et al. 2009).

Little attention has been paid to the potential for the MIS mats as analogs for ancient microbial mat ecosystems. Microbial mats are widespread throughout the Precambrian rock record, often as carbonate-associated stromatolites (Grotzinger and Knoll 1999). In addition to their low-O_2 habitat and facultatively anoxygenic metabolism (Biddanda et al 2009), several other features of the MIS mats make them excellent and novel analogs of Precambrian cyanobacterial mats (Biddanda et al in press). The MIS mats are bathed in groundwater that is constantly cold (7–9°C year-long) and thus representative of low-temperature stromatolite settings that were common in the Paleoproterozoic (Kopp et al 2005, Walter and Bauld 1983). Sulfate concentrations are intermediate between freshwater and seawater (Ruberg et al 2008), similar to those of the Proterozoic oceans (Shen et al 2003). Underlying the cyanobacterial mat

layer is a mineral layer thought to be rich in carbonate (Nold et al 2010a). Finally, the raised finger-like mat features at MIS (Figure 2.2) are similar to the conical mat structures produced by *P. autumnale* in an Antarctic lake, which were recently highlighted as analogs of stromatolites (Andersen et al 2011). Here we present genomic and functional insights into the MIS cyanobacterial mats and highlight their value as novel analogs of ancient anoxygenic phototrophic ecosystems.

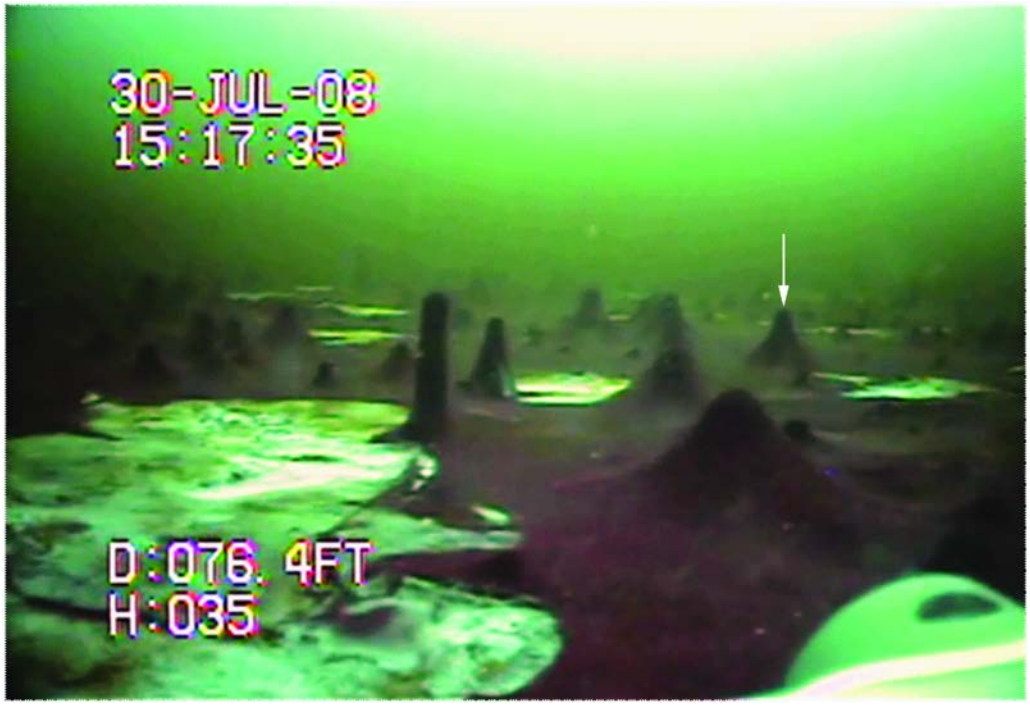


Figure 2.2 Remotely operated vehicle image of the sinkhole bottom, showing cyanobacterial mats, including purple-colored prostrate mat and raised conical structures we refer to as "fingers", and exposed white areas that have lost cyanobacterial mat cover. "Fingers" average 10-15 centimeters in height; an example is indicated in the figure by a white arrow. Photo credit: R. Paddock and V. Klump, University of Wisconsin-Milwaukee, USA).

2.2 Methods

Field work and sampling

Field work was conducted in the spring, summer, and fall from May 2007 to June 2009 at the Middle Island Sinkhole (N 45.19843°N, W 083.32721°W) near Alpena, MI (Figure 2.1). Water samples were collected from above and below the near-bottom chemocline (lake water and groundwater, respectively) by divers using Niskin bottles and then dispensed into 10L collapsible poly cubitainers. Sediment cores with intact mats, underlying sediments, and overlying water were hand collected by divers using plexiglas tubes (7.5 cm dia. x 20 cm tall). Cores were capped with rubber stoppers and kept upright in a core rack that was raised to the surface. Water and cores were kept in iced coolers in the field and then refrigerated in the lab. Water was stored at 4^{o}C and cores were maintained at *in situ* temperatures of ~9.5^{o}C. A subset of sediment cores were extruded and sectioned for microscopy, isotopic, and mineralogical analysis, according to visual cues in the sediment profile, including thin (0–0.2 cm) prostrate cyanobacterial mat, the underlying mineral-rich layer (0.2–0.5 cm), and sections of the thick organic-rich sediment. Sediment samples were stored in 2 ml plastic tubes. The "finger" used for metagenomics was collected on 6/14/2007, separated from the sediment, and transferred to a 50-ml polypropylene tube. Once collected, all samples were stored frozen at -20°C until processing.

Microscopic studies of mat structure and composition

Filamentous cyanobacteria from the surface of the mats were gently suctioned into eye droppers, fixed with 2% formaldehyde, imaged by differential interference contrast (DIC) with a Nikon eclipse 80i microscope, and photographed with a QIClick Qimaging digital camera. In order to obtain a cross-sectional image of the mat-sediment continuum, intact mats were carefully peeled from the surface of sediment cores and placed on a Petri dish over a thin layer of groundwater.

Portions of the mat were sectioned and photographed using a Nikon SMZ-2T Binocular Microscope equipped with a Micropublisher 5.0 RTV QImaging digital camera.

Benthic Chamber Studies of Dissolved Oxygen

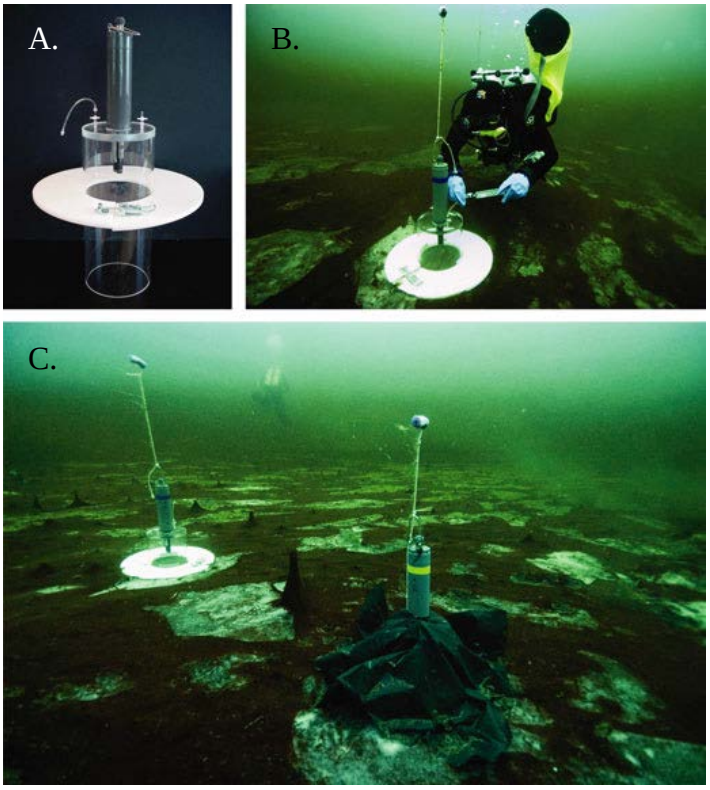


Figure 2.3 (A) Custom benthic metabolism chamber equipped with YSI-sonde sensors for dissolved oxygen, temperature and conductivity. (B) Diver deployment of chamber over benthic mat. (C) light and dark chamber deployments, respectively, at the Middle Island Sinkhole (Photo credit: B. Biddanda, Grand Valley State University, and T. Casserly, NOAA, USA).

Benthic metabolic studies were performed as time series using custom diver-deployed acrylic benthic chambers to evaluate potential for *in situ* net O_2 production via oxygenic photosynthesis and net O_2 consumption via respiration. Each chamber consisted of a tube (21 cm dia. x 50 cm length), adjustable plastic collar, and a removable cap equipped with an YSI 6920 sonde (Figure

2.3). Divers first pushed the tube/collar through the water column and into the mat/sediments, ensuring the tube contained representative groundwater and not any overlying lake water. The collar allowed precise control over chamber volume and sensor position before clamping the cap onto the chamber tube. These procedures allowed us to measure metabolic processes *in situ* without disturbing the intact microbial communities with overlying groundwater and underlying sediments. Sondes were configured to record data every hour for each sensor including temperature, conductivity, pH, oxidative-reductive potential, and O_2. From the measured changes in dissolved O_2, estimates of photosynthesis and respiration of carbon were made using either a photosynthetic quotient of 1.0 or a respiratory quotient of 1.0, respectively (Biddanda et al 1994). Typically, triplicate light and dark (covered with opaque dark plastic sheets) chambers were deployed for a period of 24–48 hrs and changes in dissolved oxygen tracked as an index of net carbon metabolism. During chamber studies (June 14–15, 2007; July 24–26, 2007; August 14–15, 2007; September 18–19, 2008), hourly PAR measurements at the mat surface were recorded from a LICOR LS-193 spherical bulb to a Nexsens SDL-500 data logger. Chamber deployment studies were conducted five times during 2007 (May 17–18; June 14–15; July 24–26; August 11–14; August 14–15), twice during 2008 (June 18–19; September 3–5) and once during 2009 (June 3–5 2009).

Autotrophic process measurements by ^{14}C bicarbonate uptake

Laboratory ^{14}C-bicarbonate uptake experiments were conducted on mats from sediment cores within 24 hours of collection using two methods: 1) mats were left intact with sediments in cores to simulate processes occurring at the mat/sediment interface in the sinkhole (e.g. sulfide production), and 2) mats were peeled away from the surface of the sediment cores and

homogenized in groundwater prior to incubation to simulate conditions of oxic groundwater. ^{14}C-sodium bicarbonate was added at a final specific activity of 2µCi ml^{-1}, and *in vitro* experiments were performed under simulated in situ conditions of temperature and light for 6–8 hrs under the following conditions: 1) light vs. dark treatments to distinguish between photosynthesis and dark chemosynthesis, and 2) treatments with and without DCMU, an inhibitor of photosystem II and hence oxygenic photosynthesis (Biddanda et al 2006, Pedros-Alio et al 1993). Each treatment had a parallel "killed" treatment where the samples were pre-treated with 2% formaldehyde for 3 hrs prior to label spike and incubation. At the end of the incubations, any unassimilated inorganic carbon was liberated from the samples by acidification with 1N HCL for 16 hrs, and the radioactivity of the remaining assimilated organic carbon alone was determined in a Beckman LS6500 Liquid Scintillation Counter. Killed controls accounted for 2–10% of the radiolabel found in live samples. All autotrophic production estimates (oxygenic photosynthesis, anoxygenic photosynthesis and chemosynthesis) were done after correcting for the radioactivity in parallel killed controls (Biddanda et al 2006, Casamayor et al 2008). Incubations were conducted in a temperature controlled dry incubator at ~9.5^{0}C, the average temperature of the groundwater at MIS. Light source was provided by a 75 watt 120 volt Sylvania Halogen lamp with a light output of 1100 lumens, and regulated by one layer of blue film and one neutral density filter to approximate the light climate available at the bottom of the Middle Island Sinkhole (~5% of surface irradiance) (Biddanda et al 2009, Ruberg et al 2008).

Stable Isotope Analyses

Samples (n = 5) collected from a microbial "finger" were placed in a freeze-dryer for 24 hours and then decarbonated in weak HCl (2% solution) and re-dried. The resulting powders were

weighed on a microbalance (~150 μg) and placed in tin capsules. The capsules were combusted in a Costech elemental analyzer attached to a Delta V+ isotope-ratio mass spectrometer for isotopic analysis. Results were calibrated using standards IAEA600 and IAEA-CH-6 and are reported as $\delta^{13}C_{org}$ values in per mil notation relative to the VPDB scale. Analytical precision was maintained at better than 0.1 ‰ during the run.

X-ray Diffraction (XRD)

Samples (n = 18) were collected as a depth profile through a cyanobacterial mat into the underlying sediments to a depth of 23.5 cm. X-ray diffraction (XRD) spectra were measured on powders using a 2.2 kW Cu-Kα Rigaku Ultima IV XRD (40 kV, 44 mA beam) with a Theta/Theta wide angle goniometer from 2–70° with a 0.05 2θ step-size. Measurements of peak position were made using PDXL software from Rigaku.

DNA extraction, Genome Sequencing Annotation, and Phylogenetic Analyses

DNA was extracted from 1g of mat material using the Fast DNA spin kit for soil (MP Biomedicals), a Fastprep-24 Bead Beater (MP Biomedicals), and a DNA Clean and Concentrator-5 kit (Zymo Research) according to the manufacturer, except that only 0.3g of beads were used for bead-beating. DNA was quantified using the Quant-IT PicoGreen dsDNA reagent and kit (Invitrogen) and submitted to the University of Michigan DNA Sequencing Core for one plate of 454 Titanium pyrosequencing. Genomic assembly was performed using MIRA (Chevreux et al 2004), and contigs binned using emergent self-organizing maps (ESOM) of tetranucleotide frequency patterns, whereby contiguous sequences were chopped into 5kb sequences for which tetranucleotide frequencies were calculated and then clustered by ESOM

(Dick et al 2009a). Further binning was done manually using BLAST. Annotation was done through the Joint Genome Institute's (JGI) Integrated Microbial Genomes Expert Review (IMG-ER) portal (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi), where gene and protein sequences are publicly available (see Supplemental Table 2.2 in Appendix A for accession numbers). Nucleotide sequences were submitted to GenBank under BioProject ID PRJNA72255. Unless noted otherwise, all BLAST analyses were performed using an e-value cutoff of 1e-5. Genes for sulfur oxidation were identified through BLAST with queries described in detail in Supplemental Table 2.1 in Appendix A. Universally conserved genes used to evaluate genome completeness were identified via BLAST with cutoffs of 1e-30 and 60% sequence identity. Sequences for phylogenetic analysis were aligned with ClustalW (Larkin et al 2007). Phylogenetic analysis was performed with MEGA 4 (Tamura et al 2007) for minimum evolution and maximum parsimony trees and RAxML (Stamatakis 2006) for maximum likelihood trees. All three approaches were used and found to yield consistent results for each phylogenetic analysis. All trees were bootstrapped 5,000 times; only bootstrap values > 70 are reported on the trees.

2.3 Results and Discussion

Mat structure and microscopy

The benthic environment of the Middle Island Sinkhole (MIS) is impacted by hypoxic, saline groundwater where dense cyanobacterial mats thrive (Figure 2.2). In addition to prostrate purple mats, there are occasional white patches where the cyanobacteria are not growing, exposing sediment or white layers of the microbial mat, as well as variably shaped features (conical to columnar) of raised cyanobacterial mat, which we designate "fingers" (Figure 2.2). These

fingers contain gas bubbles of methane and sulfide derived from microbial metabolism in underlying sediments. The structures are similar to those observed in ice-covered Antarctic lakes, where they have been observed to "lift-off" due to buoyant microbial gases (Andersen et al 2011, Cowan and Tow 2004, Hawes and Schwarz 1999, Wharton et al 1983).

Microscopic examination of the MIS mats showed predominately filamentous cells with straight, unbranched trichomes lacking heterocysts that fall into two main groups on the basis of trichome width: ~12-16 μm thick trichomes and ~6 μm thin trichomes (Figure 2.4). Sheaths have been observed in both types but their presence appears to be variable. Thick trichomes are consistently observed with rectangular cells that are shorter than one half cell width and terminate with rounded apical cells (Figure 2.4C), common of the genus *Oscillatoria*. Thin trichomes have cell shapes that are less consistent but are typically longer than their width and terminate with rounded apical cells (Figure 2.4B) common in the genus *Phormidium* (Komarek et al 2003, Vincent 2000). Despite advances made using modern genotypic and phenotypic approaches, there remains considerable uncertainty regarding the classification of these cyanobacteria at the species level (Marquardt and Palinska, 2007; Strunecky, 2010). Although trichomes of *Phormidium* sp. and *Oscillatoria* sp. were the most abundant, we have observed other types of less common cyanobacterial filaments that are spiral-shaped and tall-celled (Figure 2.4A). Their identities are unclear due to lack of coverage in both the metagenome and clone libraries performed previously (Nold et al 2010a). Clearly, additional molecular and taxonomic studies are needed.

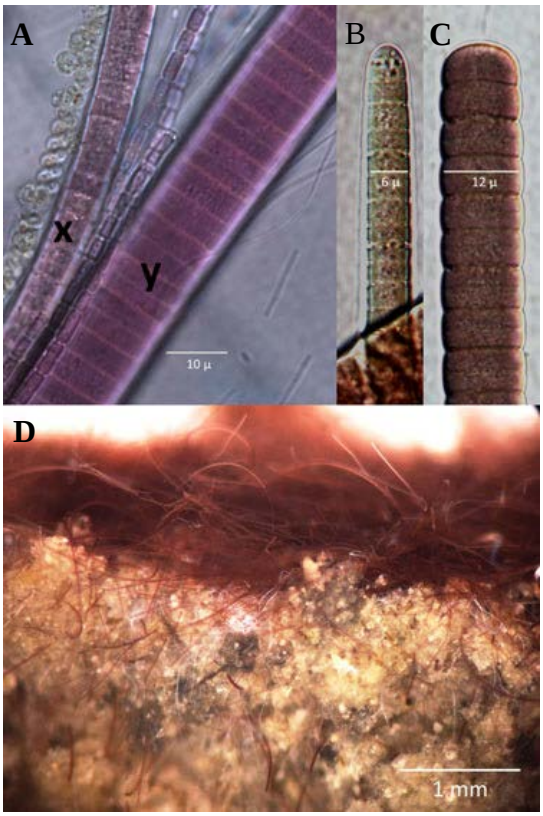


Figure 2.4 (A) Bright-field microscope images of dominant thin (x) and thick (y) cyanobacterial trichomes of *Phormidium* sp. and *Oscillatoria* sp., respectively. The spiral-shaped filaments and tall-celled filaments are far less common and have not been identified. (B) and (C) show details of rounded apical cells in *Phormidium* sp. and *Oscillatoria* sp. trichomes, respectively. (D) Cross-sectional stereo microscopic image of purple microbial mat at the MIS showing layers of motile purple cyanobacterial trichomes on top and motile white filamentous sulfur oxidizing bacteria and carbonate crystals below. Underlying dark organic sediment is not pictured.

Under the fluorescence microscope, the filamentous cyanobacteria autofluoresce purple-red when exited by green light, suggesting the presence of phycobiliproteins. Vertical cross section of the mat and sediment revealed a thick layer of woven cyanobacterial trichomes over a white crystalline sediment matrix interspersed with unidentified white filamentous bacteria (Figure 2.4D). Trichomes exhibited remarkable motility under light, being able to rapidly re-aggregate or climb over small pebbles in minutes or hours. When placed in darkness, the white filaments were observed to migrate to the mat surface, consistent with behavior of mat-associated sulfur-oxidizing chemosynthetic bacteria. Determining whether such diel migrations

occur *in situ* at MIS and the biogeochemical consequences of any such spatio-temporal dynamics is worthy of future investigation.

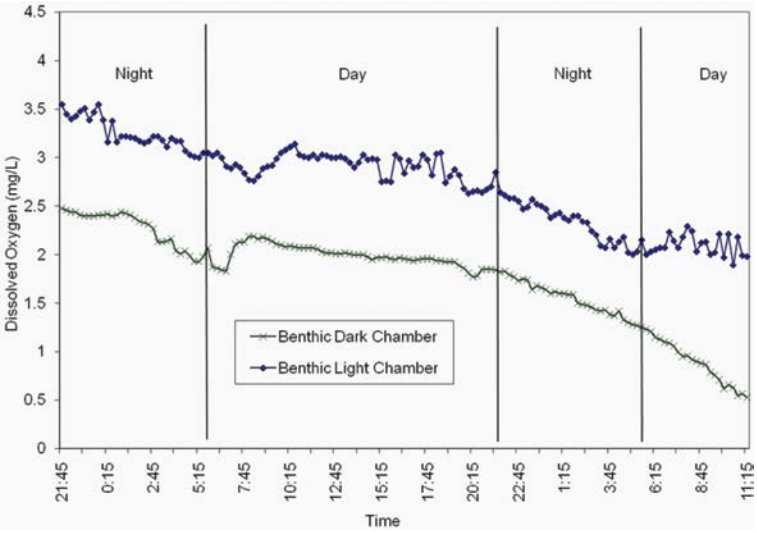


Figure 2.5 Dissolved O_2 concentration measured in benthic metabolism chambers over a 36-hour period (July 24-26, 2007). Changes in dissolved O_2 concentration in light and dark chambers (both decreasing at nearly identical slopes) suggest that alternative production mechanisms such as anoxygenic photosynthesis and chemosynthesis must be prevalent at sinkholes to explain how such a prolific mat community could sustain itself. Nearly identical rates and trends were measured in different years from 2006-2009.

Carbon metabolism and respiration

Two approaches were taken to investigate carbon metabolism in the MIS mats. First, benthic metabolism chambers were deployed *in situ* under light and dark conditions (Figure 2.3) to measure changes in dissolved O_2 as an index of O_2 production/respiration. Second, lab experiments were performed on mat/sediment cores to track autotrophic ^{14}C-bicarbonate incorporation into biomass. Five benthic chamber studies conducted in 2007 and 2008 consistently showed there was net consumption of O_2, with dissolved O_2 decreasing similarly in both light and dark treatments (Figure 2.5). However, the rates of net O_2 consumption were significantly different (paired t-test, $p=0.025$), being ~30% greater in the dark (~17 mg C m^{-2} d^{-1})

than in the light (~12 mg C m^{-2} d^{-1}) (Table 2.1). This net consumption of O_2 shows that any O_2 produced via oxygenic photosynthesis is quickly consumed either by aerobic respiration or by reduction with sulfide (chemical or chemosynthetic) (Supplemental Figure 2.2 in Appendix A). Such tight coupling of O_2 production and consumption has been observed in cyanobacterial mats previously (Canfield and Des Marais 1993). Calculations based on these observed rates of O_2 consumption and using a respiratory quotient of 1.0 suggest that ~17 mg C m^{-2} d^{-1} of autotrophic carbon synthesis is required to balance the system (Table 2.1). These results also indicate an excess of carbon synthesis relative to O_2 production, and suggest that primary production mechanisms that do not produce O_2, such as anoxygenic photosynthesis and chemosynthesis, play significant roles in the carbon balance of the MIS habitat (Supplemental Figure 2.2 in Appendix A).

Table 2.1

Carbon consumption in in-situ light and dark benthic chambers estimated from decline in dissolved oxygen over 24-48 hr deployments at MIS during 5 experiments in 2007 and 2008. A typical example of time-series data (July 2007 study) is shown in Figure 5.

Date of Benthic Chamber Study	Light Chambers (mg C/m^2/d)	Dark Chambers (mg C/m^2/d)
June 14-15, 2007	7	11
July 24-26, 2007	11	14
August 11-14, 2007	12	17
August 14-15, 2007	11	22
September 18-19, 2008	19	22
Mean ± S.D. (n=5) **Mat/Sediment Carbon Consumption**	12.0 (4.4)	17.2 (4.8)

Table 2.2
Autotrophic production processes in intact MIS cyanobacterial mat + sediment (cores), and mat filaments in groundwater (suspension). Mean ± S.D. (n=3).

Production Process	Intact Cores (low/no O_2) (mg/C/m2/d)	Mat suspension (oxygenated) (mg C/L/d)
Oxygenic PS	not detected	4.0 (2.3)
Anoxygenic PS	8.4 (2.4)	0.9 (0.4)
Chemosynthesis	8.6 (2.1)	0.2 (0.03)
Total autotrophic production	17	5.1

^{14}C-bicarbonate incorporation studies of mats in laboratory incubations were used to further assess contributions to primary production from oxygenic photosynthesis, anoxygenic photosynthesis, and chemosynthesis. Experiments were conducted under two conditions designed to assess the effect of sulfide and O_2 on autotrophic production processes by the mats: (1) mats were left intact with sediments (sulfide present; low O_2), and (2) mats were removed from the sediment and suspended in groundwater (sulfide absent; high O_2). For the intact cores, primary production was dominated by anoxygenic photosynthesis and chemosynthesis, and no oxygenic photosynthesis was detected (Table 2.2). In contrast, oxygenic photosynthesis was the main mode of primary production in the oxygenated mat suspension. Interestingly, if we take the difference in net O_2 consumption between day and night benthic chambers (5.2 mg C m^{-2} d^{-1}) as a rough estimate of oxygenic photosynthesis (Table 2.1), this rate is approximately matched by the measured rate of oxygenic photosynthesis in ^{14}C tracer studies (4.0 mg C m^{-2} d^{-1}) – albeit under oxygenated and sulfide-free conditions (Table 2.2). Based on these findings we draw the following conclusions. First, the mats are metabolically versatile, being capable of significant primary production through oxygenic or anoxygenic photosynthesis or chemosynthesis. Second, the balance of oxygenic versus anoxygenic photosynthesis depends on the presence of sulfide

and/or O_2. Third, measured rates of anoxygenic photosynthesis and chemosynthesis of ~ 17 mg C m^{-2} d^{-1} from ^{14}C tracer studies of intact sediment cores are collectively sufficient to satisfy the carbon-deficit of ~ 17 mg C m^{-2} d^{-1} estimated from O_2 consumption observed in the benthic chamber studies (Tables 2.1 and 2.2) – suggesting that the carbon cycle in the MIS mat is well-balanced (Supplemental Figure 2.2 in Appendix A). In terms of the quantity of primary production, rates of anoxygenic photosynthesis and chemosynthesis measured in the MIS mat-sediment complex are comparable to those made for aquatic microbial mat communities in Solar Lake, Israel, and other similar mat-dominated habitats (Bachar et al 2007, Casamayor et al 2008, Cohen et al 1986, Fontes et al 2011, Jorgensen et al 1979, Jorgensen et al 1983, Overmann et al 1991, Revsbech et al 1983, Stal 2000).

Microbial "finger" $\delta^{13}C_{org}$ values ranged from -27.25 to -28.43‰, with a mean of -27.88‰ (± 0.47 1σ). The C wt. % was variable within a relatively narrow range, and averaged 44.88%. Groundwater in the MIS system has a dissolved inorganic carbon (DIC) isotopic composition of -4.1‰ (Sanders et al 2011), so fractionation of carbon by the mat due to photosynthesis was -24‰. These $\delta^{13}C_{org}$ values are consistent with typical autotrophic carbon fixation by the Calvin-Benson cycle (which is present in the MIS-Ph1 genome – see below). While cyanobacterial systems in nature may display a wide array of $\delta^{13}C_{org}$ values, both the mean value of the Lake Huron mats themselves and the degree of C fractionation are consistent with previous results from both marine and freshwater cyanobacterial systems (Schidlowski 2000). Thus, while there are relatively few Precambrian lacustrine systems that have been recognized, results from this system are also relevant for understanding similar marine settings with photic-zone benthic microbial mats. The MIS mat $\delta^{13}C_{org}$ values are consistent with reported terrestrial Precambrian organic matter (Horodyski and Knauth 1994, Imbus et al 1992, Retallack and

Mindszenty 1994, Rye and Holland 2000). Most Precambrian kerogen (fossilized organic matter that is insoluble in solvents) ranges between -25 and -40‰ (Pavlov et al 2001), where the lightest values are from systems thought to have been influenced by methanotrophy. Microbial decomposition of C in sediments typically changes $\delta^{13}C_{org}$ values by +2‰ or more relative to the original value of the organic matter (Walter et al 2007), so if organic matter from a Precambrian system similar to the Middle Island Sinkhole ($\delta^{13}C_{org}$ value of -27.9‰) was buried, the preserved signal would be near the heavy end of the Precambrian kerogen range, appropriately representing photosynthetic systems without significant methanotrophic influence. Precambrian microbial mats are most often recognized on the basis of gross morphological features (e.g., stromatolites) or of microbially induced sedimentary structures (Noffke, 2009; Sheldon, in press), neither of which are obvious in the MIS system. Our results showing an anoxygenic cyanobacterial mat with $\delta^{13}C_{org}$ values that are indistinguishable from those of oxygenic cyanobacteria highlights the difficulty in relating $\delta^{13}C_{org}$ values to specific metabolic or biogeochemical function. Further work is needed on the isotopic and elemental composition of the carbonates in the system, and on the $\delta^{15}N$ values of the mats, to determine whether the combined isotopic results will provide a biosignature for similar systems in the geologic record.

Identification of minerals associated with mats and underlying sediments

The unique geochemical setting of the MIS mats presents opportunities to investigate cyanobacterial calcification under conditions relevant to the Precambrian, which may inform longstanding questions regarding the distribution of calcified stromatolites through geologic time (Grotzinger and Knoll 1999, Riding 2006). The MIS mats are not lithified, but a carbonate-rich layer just beneath the mat has been observed (Nold et al 2010a). XRD identified quartz, calcite,

and dolomite as the three major mineral phases associated with the mat and underlying sediments. For the purposes of comparing normalized abundance, we present the ratio of XRD primary peak intensity of calcite ($2\theta = 29.5°$) and dolomite ($2\theta = 31°$) to quartz ($2\theta = 26.7°$). This calcite/quartz ratio (C/Q) ranged with depth in mat/sediment from 0.39 to 2.31, with a mean of 0.57 ($\pm$ 12 1σ; excluding the high value). The layer immediately beneath the mat is characterized by a > 14σ increase in the C/Q ratio, indicating a prominent enrichment of calcite and confirming the presence of a carbonate-rich sedimentary layer immediately beneath the cyanobacterial mat (Nold et al 2010a). The dolomite/quartz ratio (D/Q) ranged from 0.58–1.12 with a mean of 0.79 ($\pm$ 0.13 1σ). The D/Q ratio was elevated both deep in the core and near the surface, including a >2 σ enrichment immediately underneath the cyanobacterial mat, at the same level as the calcite enrichment.

There are several possible mechanisms by which the microbial mat and/or sediment communities may influence carbonate precipitation in this environment. First, cyanobacteria produce extracellular polymeric substances (EPS), which can nucleate carbonate mineralization and influence the type of mineral produced (Riding 2006, Riding 2011). Heterotrophic bacteria have also been shown to catalyze carbonate formation through heterogeneous nucleation (Bosak and Newman 2003). Second, carbonate precipitation can be promoted by increases in alkalinity driven by photosynthesis or heterotrophic sulfate reduction (Dupraz et al 2009). In particular, carbon concentration mechanisms possessed by the dominant MIS organism (see *carbon acquisition and metabolism* section below) are thought to induce calcification (Riding 2011). Sulfate reduction observed in the carbonate-rich layer of MIS has been linked to sulfate reducing bacteria (Nold et al 2010a) and interestingly, the cyanobacteria themselves also show some evidence for sulfur reduction (see below) and thus could play a role in carbonate precipitation

through this mechanism. However, it is unclear whether sulfur/sulfate metabolism is directly involved in carbonate precipitation in the MIS sinkhole, and the relevance of this process under Precambrian conditions is also questionable (Bosak and Newman 2003). The formation of dolomite in modern, freshwater, low temperature settings is rare, but microbial mediation of dolomite precipitation has also been linked to sulfate reduction (Van Lith et al 2003, Vasconcelos et al 1995, Warthmann et al 2000) and as well as methanogenesis (Kenward et al 2009). Both of these metabolisms are active in MIS sediments but further investigation is required to determine the specific chemical and biological factors that influence the formation of the observed calcite and dolomite.

Table 2.3 Summary of metagenomic assembly.

Genomic Bin	No. of reads	No. of contigs	Avg. contig length (bp)	Avg. coverage	Avg. % GC	Total consensus sequence (MB)
Metagenome	827,593	19,463	1,357	4.4x	43	26.7
MIS-Ph1	577,920	555	11,251	38.4x	45.1	6.4
MIS-Ph2	14,400	254	3,991	5.9x	44.4	1
Os	94,238	748	6,210	7.9x	38.9	4.6
Unassigned	141,035	17,906	819	3.2x	43.1	14.7

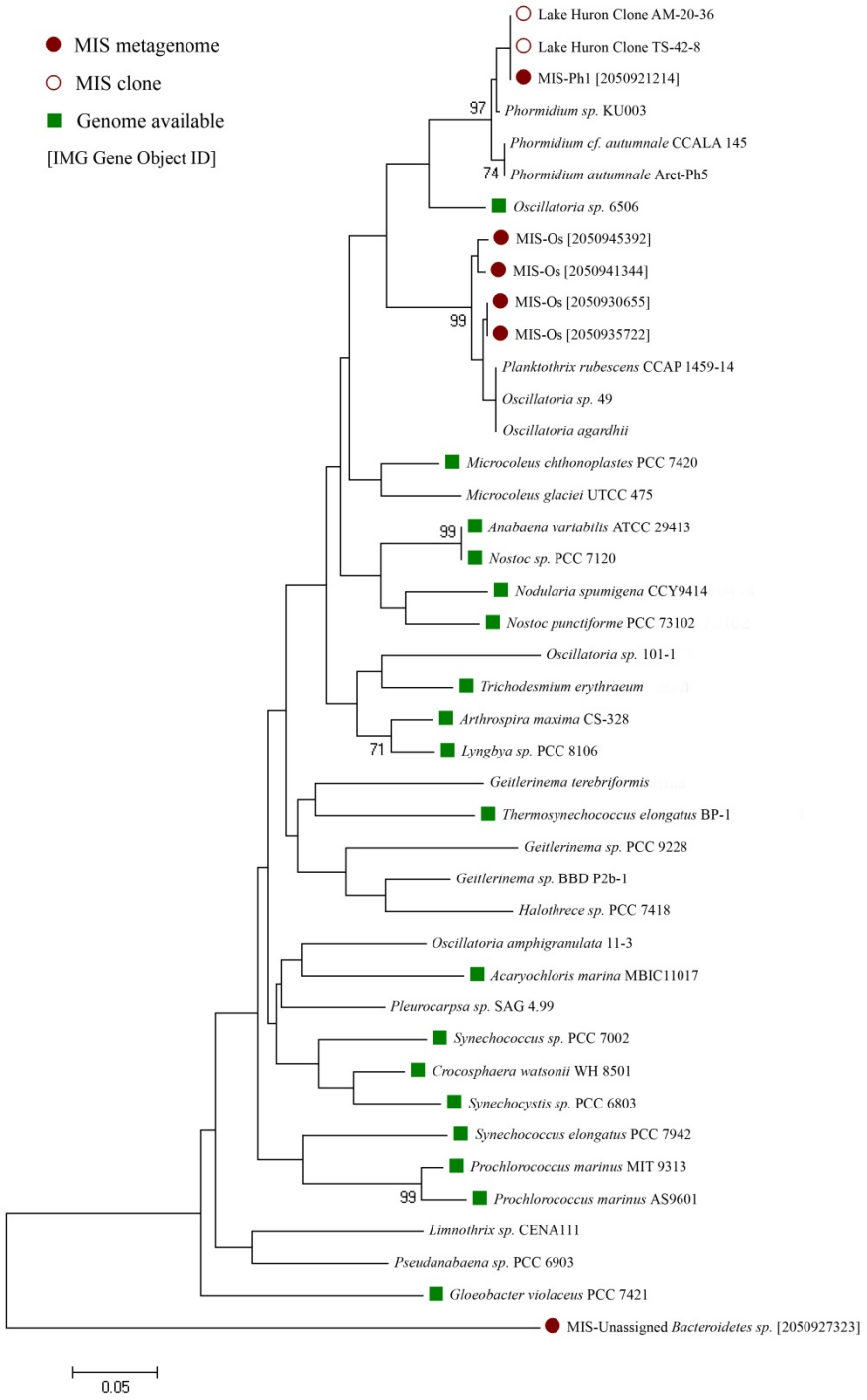


Figure 2.6 Phylogenetic tree of the 16S rRNA gene of selected cyanobacteria. Metagenomic contigs from this study (closed red circles), clones from Nold et al. (2010a) (open red circles), and species for which genome sequences are available (green squares) are indicated. Bootstrap values are the results of 5000 iterations; values < 70 are not shown.

Metagenomic sequencing, assembly, and binning

To explore the genetic diversity and metabolic capability of the MIS microbial mat, community genomic DNA of a mat finger was shotgun sequenced, producing 827,593 DNA sequencing reads that assembled into 19,463 contiguous sequences (contigs) (Table 2.3). 16S rRNA gene sequences in this metagenome were dominated by two genera of cyanobacteria, *Phormidium* and *Oscillatoria*. The dominant 16S rRNA gene sequence (38x average coverage) is closely related to *Phormidium autumnale*-like sequences retrieved previously from prostrate MIS mats and from a variety of Arctic and Antarctic environments (Figure 2.6) (Nold et al 2010a). Of the 45 sequences obtained from the MIS mat previously, 31 came from the same dominant *Phormidium* operational taxonomic unit (OTU), indicating that this organism dominates both prostrate and raised mats in the MIS. Also present in the MIS metagenome were four contigs (2–13X coverage) with 16S rRNA gene sequences that cluster tightly with *Planktothrix rubescens* CCAP 1459-14, *Oscillatoria agardhii*, and *Oscillatoria* sp. 49. A partial *Bacteroidetes*-like 16S rRNA (2x genomic coverage) was also recovered.

To better understand the genetic potential underpinning the physiology and metabolism of the different MIS mat community members, contigs were assigned to taxonomic groups, or "genomic bins", based on emergent self-organizing maps of tetranucleotide frequency (Dick et al 2009a). Two bins were apparent, one containing contigs with the *Phormidium* 16S rRNA gene (MIS-Ph1), and the other containing the *Oscillatoria* 16S contigs. Contigs in the *Phormidium* bin showed a bimodal distribution of genomic coverage with approximately half at 38x coverage and half at less than 10x coverage; we designate these as two separate bins, MIS-Ph1 and MIS-Ph2 respectively. The MIS-Ph2 bin contains a partial 16s rRNA gene that is identical to the 16S rRNA gene from MIS-Ph1, thus we infer that these bins represent two closely-related but distinct

genotypes, one high abundance and one low abundance. The *Oscillatoria* bin contains contigs from at least four different genotypes that are present at similar abundance (2–10x coverage); we refer to these populations collectively as MIS-Os. The 4.6 megabases of DNA recovered in this bin represent partial genomes of the *Oscillatoria* populations. 70% of the total DNA sequence reads assembled into the dominant MIS-Ph1 genotype, and just 17% of DNA sequence reads fell outside of the dominant *Phormidium* and *Oscillatoria* organisms (Table 2.3). These results show that MIS mats have remarkably low species and genomic diversity, containing just a few dominant cyanobacteria and several lower abundance bacteria.

Putative genes for anoxygenic photosynthesis

In order to investigate the potential for anoxygenic photosynthesis among members of the MIS mat community, we searched the metagenome for genes known to be involved in sulfur oxidation. Genes with sequence homology to SQR were found in the genomic bins of all three dominant mat cyanobacteria as well as in unassigned contigs (Table 2.4; accession numbers and annotations of all genes discussed in the text are provided in Supplemental Table 2.2 in Appendix A). The *Oscillatoria* bin contains an SQR with 56% amino acid identity to the SQR from the cyanobacterium *Geitlerinema* sp. PCC 9228, which is involved in sulfide-dependent anoxygenic photosynthesis (Bronstein et al. 2000). The *Phormidium* and unassigned bins contain homologs of SQR that are much more divergent. Four genes with limited sequence similarity to SQR in the unassigned contigs are most closely related to oxidoreductases of unknown function from *Oscillatoria* sp. 6506 and *Lyngbya* sp. PCC 8106 (15–25% amino acid identity). These unassigned SQR genes are on contigs with very low genomic coverage (1–2X) and thus derive from very low abundance organisms, so it is highly unlikely that they contribute

significantly to the anoxygenic photosynthesis reported in Table 2.2. A phylogenetic tree of SQR shows that the MIS-Os SQR falls into a well-resolved cluster of SQRs including two with experimentally verified sulfide-oxidizing activity (*Geitlerinema* sp. PCC 9228 and *A. halophytica*) and eight others that are present in sequenced cyanobacterial genomes (Figure 2.7). Putative SQR sequences from MIS-Ph1 and MIS-Ph2 fall outside of the main clade of cyanobacterial SQR sequences but within the broader family, which includes an SQR from the eukaryote *Arenicola marina* that oxidizes sulfide for the purpose of detoxification (Bronstein et al 2000a).

Table 2.4. Occurrence of dissimilatory sulfur metabolism genes in the MIS finger metagenome.

Organism	**SQR**	**fccA**	**fccB**	**sox**	**dsr**	**sorA**	**sorB**	**aprA**	**aprB**	**qmoA**	**qmoB**	**qmoC**
LHS-Ph1	+	-	-	-	+	-	-	+	-	-	+	-
LHS-Ph2	+	-	+	-	+	-	-	-	-	+	+	+
LHS-Os	+	-	+	-	+	+	-	-	-	-	-	-
Unassigned	+	-	+	-	+	+	-	-	-	-	+	+

BLAST survey after Frigaard and Dahl (2008), using parameters and queries described in *Materials and Methods*. Note that the dsr column represents the presence of any of the 13 dsr genes (dsrABCEFHLNMKJOP). None of the bins has a full set of dsr genes, but all bins contain BLAST hits to some dsr genes.

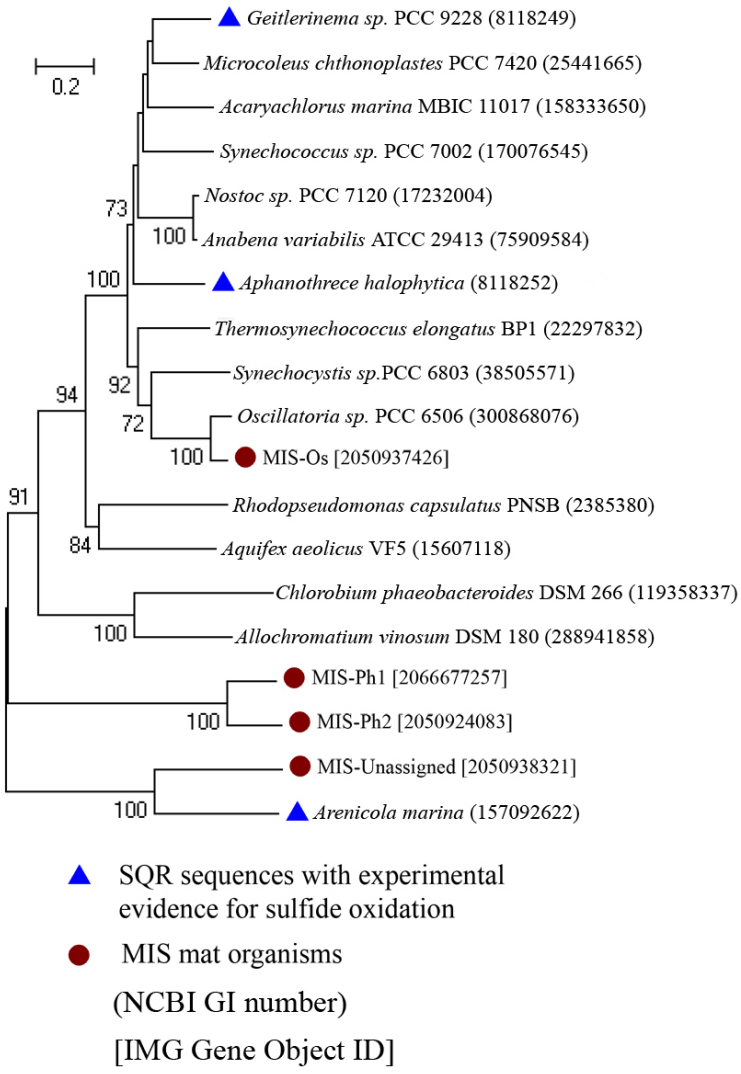


Figure 2.7 Phylogenetic tree of sulfide quinone oxidoreductase (SQR). Sequences from the MIS mat are indicated with red circles. Genes that have been experimentally verified to oxidize sulfide are shown with a blue triangle.

While the MIS mat SQR sequences that we report do indeed likely encode functional sulfide-oxidizing enzymes, the actual physiological role that this sulfide oxidation plays is difficult to discern based solely on sequence similarity. *Geitlerinema* sp. PCC 9228 and *Aphanothece halophytica* encode SQR enzymes that have been experimentally shown to oxidize sulfide (Bronstein et al 2000a), but they perform different physiological functions. The SQR from *Geitlerinema* sp. PCC 9228 can be used for sulfide-dependent anoxygenic photosynthesis,

whereas the SQR from *A. halophytica* is used for detoxification of sulfide and is not linked to cell growth (Bronstein et al 2000a, Oren and Shilo 1979). Further complicating the functional role of this enzyme in the cyanobacteria, the SQR from *Geitlerinema* sp. PCC 9228 can also be involved in anaerobic respiration (Oren and Shilo 1979). Clearly, there is much that remains to be learned about the function of SQR-like genes in the cyanobacteria.

Although SQR is the only enzyme that has been implicated in anoxygenic photosynthesis in the cyanobacteria to date, sulfide-dependent anoxygenic photosynthesis is phylogenetically widespread throughout the cyanobacteria (Miller and Bebout 2004a). The vast majority of this diversity has not been explored with genetic or biochemical tools, therefore it is feasible that there are novel sulfur oxidation pathways in the cyanobacteria. Homologs of several genes known to be involved in sulfur oxidation in the MIS metagenome were identified (Table 2.4), including flavocytochrome c (*fccB*), various dissimilatory sulfur reductase (*dsr*) genes, sulfite:cytochrome c oxidoreductase (*sor*), adenosince-5'-phosphosulfate reductase (*apr*), and the quinone-interacting membrane bound oxidoreductase (*qmo*) (Frigaard and Dahl, 2008). Many of these genes are only distantly related (<30% AA identity) to known sulfur-oxidizing enzymes, and in many cases only a subset of subunits are present, thus we cannot ascribe function or taxonomy (in the case of unassigned contigs) to these genes with confidence.

Also notable was the absence of certain sulfur oxidizing pathways typically prevalent in bacterial sulfur oxidation, including anoxygenic photosynthesis. No *sox* genes (Friedrich et al., 2005; Ludwig et al., 2006) were identified, and while some *dsr* genes were found in the metagenome, they were typically quite divergent from known genes, and no bin had a full complement. Finally, SQR is the only gene implied in anoxygenic photosynthesis associated with sulfide oxidation found in MIS-Ph1 (Frigaard et al 2008). Taken together, our results

suggest that the anoxygenic photosynthesis observed in the MIS mat is conducted by cyanobacteria via genes and biochemical pathways that are not yet well-characterized.

Evidence for a complete genome of Phormidium sp. MIS-Ph1

The MIS-Ph1 genome has ~40X genomic coverage and few polymorphisms (Supplemental Figure 2.1 in Appendix A), suggesting an essentially complete genome from a near-clonal population. To evaluate genome completeness, we searched the MIS-Ph1 genome for 40 universally conserved housekeeping genes that are not often duplicated or horizontally transferred (Raes et al 2007). All 40 genes are present in the MIS-Ph1 genome; 36 are present in one copy and four were present in 2 copies (Supplemental Table 2.3 in Appendix A). The occurrence of multiple copies of these genes is not uncommon amongst sequenced cyanobacteria; every cyanobacterial genome available on the JGI Integrated Microbial Genomes website (60 total as of May, 2011) has at least two of these 40 genes in multiple copies, and eight of the genomes are missing at least one of the 40 genes. The presence of all 40 universal housekeeping genes suggests that the gene content of the MIS-Ph1 genome is very close to complete. The completeness of the MIS-Ph1 genome is also supported by the presence of genes encoding complete photosynthetic machinery and pathways of energy and carbon metabolism; we identified genes for biosynthesis of chlorophyll, for proteins of photosystems I and II, the cytochrome b6/f complex, a complete ATP synthase, NADH dehydrogenase complex, and a Heme-Cu-type cytochrome/quinol oxidase.

A total of 5824 genes were identified in the *Phormidium sp.* MIS-Ph1 genome, including 5774 protein coding genes and 50 RNA (rRNA and tRNA) coding genes. 1095 of these genes are present in the vast majority of cyanobacteria genomes sequenced to date (>59 of 62), whereas

602 of the MIS-Ph1 genes are not present in any other cyanobacterial genomes. Overall, the majority of MIS-Ph1 genes could not be assigned specific functions (Figure 2.8). Genes that are absent or uncommon in previously sequenced cyanobacterial genomes are especially poorly defined in terms of function (Figure 2.8).

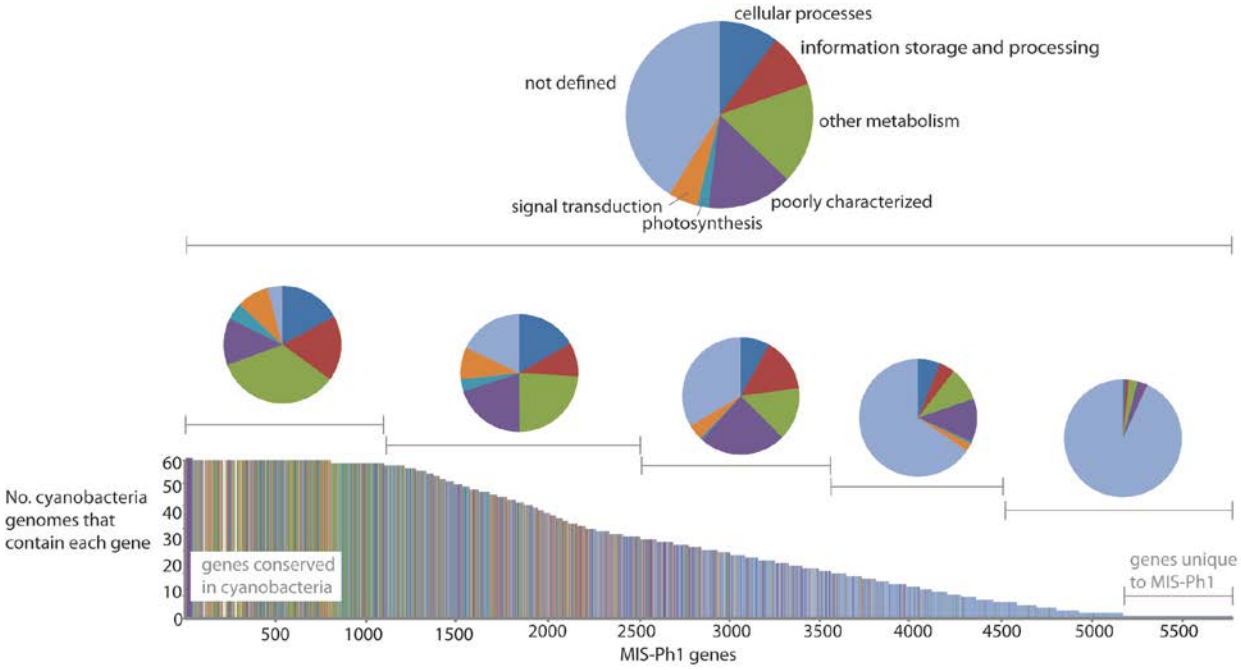


Figure 2.8 Function and distribution of *Phormidium sp.* MIS-Ph1 genes across the 62 cyanobacteria genome sequences currently publicly available. The top pie chart shows COG functional categories for the entire MIS-Ph1 genome. Pie graphs below show composition of COG functional categories for the indicated fractions. "Cellular Processes" includes genes for cell cycle control, cell division, and cell motility. "Poorly Characterized" indicates that the genes are of unknown function.

Carbon acquisition and metabolism

The MIS-Ph1 genome encodes a complete Calvin-Benson cycle, consistent with the $\delta^{13}C_{org}$ values (-27.25 to -28.43‰) reported above and its role as the major primary producer of the finger community. The genome also has genes for carboxysome shell proteins and carbonic anhydrase, which constitute a carbon concentrating mechanism (CCM) and suggest that CO_2 can become limiting in the mat environment. This CCM is also thought to induce carbonate precipitation by cyanobacteria (Riding 2006).

Genes for glycogen synthase and carbohydrate branching and debranching enzymes reveal a mechanism of carbon and energy storage and subsequent utilization. The complete genes for glucose degradation via the pentose phosphate pathway are present, as are two genes that allow *Phormidium* to perform acetate fermentation; pyruvate:ferredoxin oxidoreductase and acetate kinase (Stal and Moezelaar 1997). Fermentation may sustain MIS-Ph1 at night when sunlight is no longer available for photosynthesis. During the day, mat-forming cyanobacteria store carbon and energy in polyglucose reserves and ferment those reserves at night (Nold and Ward 1996, Stal 1995), creating fermentation products that are cross-fed to heterotrophic community partners.

In addition to genes for autotrophy, the MIS-Ph1 genome also contains genes for the acquisition and utilization of organic carbon (carbohydrates and amino acids) from the environment (Supplemental Table 2.2 in Appendix A). Interestingly, genes encoding beta-galactosidase/beta-glucuronidases are not found in any of the other 60 cyanobacteria genomes sequenced to date except the recently sequenced closest relative, *Oscillatoria* sp. PCC 6506 (Mejean et al 2010). There are close homologs of these proteins (> 50% amino acid ID) in other phyla such as Proteobacteria, Verrucomicrobia, Firmicutes, and Lentisphaerae, raising the

intriguing possibility of a horizontal gene transfer origin of these genes in the lineage shared by MIS-1 and *Oscillatoria* sp. PCC 6506.

Oxygen sensing, regulation, and respiratory metabolism

The MIS-Ph1 genome contains a host of genes dedicated to sensing and metabolizing O_2 and regulating O_2-sensitive pathways. There are two different electron transport chain terminal reductases: a cytochrome c oxidase (subunits I, II, and III) that is present in nearly all currently available cyanobacteria genomes, and a cytochrome bd plastoquinol oxidase (subunits 1 and 2) that is less widely distributed in cyanobacteria and which is known to operate under low-O_2 conditions (Kana et al 2001). There are also hints of anaerobic respiration; the SQR described above has been reported to be involved in reduction of elemental sulfur, and there is also a gene annotated as a sulfite reductase. The presence of these genes raises the intriguing possibility that MIS-Ph1 could be involved in dissimilatory sulfur reduction that has been observed in the MIS mat (Nold et al 2010a). However, these genes are distantly related to genes of known function, hence experimental evidence is required to test this possibility. Although anaerobic respiration of elemental sulfur has been observed in cyanobacteria, it remains poorly understood (Stal and Moezelaar 1997), and dissimilatory reduction of sulfite or sulfate by cyanobacteria has not been described. The presence of two terminal O_2 reductases as well as a putative mechanism for sulfur reduction may provide versatility in the stratified redox environment in which MIS-Ph1 thrives. This versatility would also be important in a microbial mat ecosystem where oxygen and sulfide concentrations likely vary on a diel cycle (Stal 2000).

MIS-Ph1 is also well-equipped for detoxification of reactive oxygen species (ROS) with genes for superoxide dismutase, cytochrome c peroxidase, glutathione peroxidase, and

peroxiredoxin. These genes likely play important roles in protecting the MIS-Ph1 biomolecules, including the photosynthetic apparatus, from ROS that are commonly produced at both photosynthetic reaction centers (He and Hader 2002).

Hopanoid biosynthesis

Genes for squalene-hopene cyclase (*shc*) and a radical SAM methylase (*hpnP*) (Welander et al 2010) have been implicated in biosynthesis of 2-methylhopanoids, which have been used as a biomarker of cyanobacteria and oxygenic photosynthesis in the geologic record (Summons et al 1999). However, recent evidence that 2-methylhopaoids and their biosynthetic genes are not present in all cyanobacteria and are present in certain non-cyanobacteria questions the reliability of this biomarker (Rashby et al 2007, Welander et al 2009, Welander et al 2010). We were unable to detect *shc* genes within the MIS-Ph1 genome, and homologs of *hpnP* are present but at only at such low similarity (<35% amino acid identity) that their function is uncertain. The absence of hopanoid biosynthesis genes in a cyanobacterium that thrives under persistent low-O_2 conditions, which were likely common in cyanobacterial habitats through certain periods of the Precambrian, casts further doubt on the utility of hopanoids as a biomarker of cyanobacteria in the geologic record.

Genomic insights into interactions of MIS-Ph1 with the mat community

The MIS-Ph1 genome contains genes that reflect a communal lifestyle in which interactions with other community members are prevalent. First, there are many genes reflective of viral predation pressure (Supplemental Table 2.2 in Appendix A), including seven CRISPR sequences, which are thought to provide adaptive immunity to genetic elements such as viruses and plasmids

(Makarova et al 2011b). Second, there are many genes that appear to be involved in antagonistic chemical interactions with other community members including the toxins colicin D and hemolysin, non-ribosomal peptide synthetases (NRPS) commonly involved in the production of bioactive compounds, and antibiotic synthesis and drug resistance. The function of genes for antibiotics is uncertain given the diverse physiological functions recently attributed to them such as electron transfer, signaling, and community development (Dietrich et al 2008, Wang et al 2010). Third, we detected a large number of genes involved in communication and mat construction (Supplemental Table 2.2 in Appendix A). Overall, the communal lifestyle encoded by the MIS-Ph1 genome offers significant benefits including antibiotic resistance, predation deterrence, and facilitation of nutrient acquisition (Blenkinsopp and Costerton 1991).

Environmental sensing, regulation, and nutrient acquisition

The MIS-Ph1 genome includes many genes dedicated to interfacing with the environment. Genes encoding a number of light antenna proteins, including allophycocyanin, phycoerythrin, and phycocyanin/phycoerythrocyanin, indicate an ability to efficiently capture energy over a wide spectrum of light (DeRuyter and Fromme 2008) which would be useful over daily and seasonal light fluctuations. There are also four genes for bacteriophytochromes, which are used for sensing light and regulating light-dependent cellular processes. Thus the MIS-Ph1 genome is well-equipped to optimize photosynthetic machinery according to prevailing light availability. There are many more sensing and regulatory genes for light and motility for which only general functional predication can be made (Supplemental Table 2.2 in Appendix A).

The MIS-Ph1 genome encodes numerous transport systems for efficient acquisition of nutrients from the environment. Genes are present for high-affinity transporters of iron (Fe^{2+},

Fe^{3+} and heme), cobalt, nickel, manganese, phosphate, nitrate, and sulfate and for regulation of cellular processes governing homeostasis of these nutrients. The only authentic nitrogen fixation gene identified is present on an unassigned contig from a low abundance member of the community. An exceptional number of genes (5) for cobalt-containing cobalamin (Vitamin B_{12}) biosynthesis are present; this micronutrient limits primary production in pelagic marine environments (Bertrand et al 2007) but its significance in mats is unknown.

Finally, genes encoding biosynthesis of betaine and trehalose suggest that these organic compatible solutes are used to maintain cellular turgor pressure in the face of osmotic stress. The presence of both trehalose, which has been associated with halotolerance in freshwater cyanobacteria, and betaine, which has been associated with halotolerance in hypersaline cyanobacteria (Stal, 2000), suggests that MIS-Ph1 is prepared to survive a broad range of salinities.

2.4 Conclusions

The Middle Island Sinkhole (MIS) hosts cyanobacterial mats that thrive under persistent low-O_2 conditions, thus providing a novel modern system for investigating geobiological processes that were critical in geochemical and biological evolution. Cyanobacterial mat systems are typically considered as a source of O_2, for example for the great oxidation event (Holland 2006) or for O_2 oases that fostered development of early animal life (Gingras 2011). In contrast, we find that the MIS mats can be sinks for O_2 due to significant primary production of carbon via anoxygenic photosynthesis and chemosynthesis. Recognition of this modern microbial mat community that is dominated by cyanobacteria and has $\delta^{13}C_{org}$ values indistinguishable from oxygenic cyanobacteria, yet functions as a sink for O_2, underscores the need for caution in inferring

metabolic functions of filamentous cyanobacteria (or of sedimentary structures putatively made by cyanobacteria) preserved in the geologic record. More broadly, these results suggest that the biogeochemical function of cyanobacterial communities through Earth history should be considered more carefully, for example as done by Johnston et al (2009).

Remarkably, the MIS mat community is dominated by just one genotype of the cyanobacterium *Phormidium sp.* MIS-Ph1, which enabled cultivation-independent insights through genomic reconstruction. This genome sequence reveals many adaptations for life in the hostile MIS environment, including metabolic versatility (autotrophy and heterotrophy) and the ability to facultatively switch between oxygenic and anoxygenic photosynthesis. In addition, it also possesses systems to sense environmental conditions and optimize cellular machinery according to light and redox conditions. Also encoded in the genome are tools to sustain a dense mat community while contending with other microorganisms, grazers, and viruses.
Although these findings provide the first genomic insights into life within cyanobacterial mats under low-O_2 concentrations, the preponderance of novel genes of unknown function also highlights critical gaps in understanding of the genetic underpinnings of these systems, especially with regards to anoxygenic photosynthesis. Further insights into the relationship between genetic diversity and community function await deeper sequencing and tracking of physiology and gene expression over gradients of light and redox chemistry. The metagenome also holds clues to nutrient requirements that will guide cultivation of the key organisms so that genome-generated hypotheses can be tested experimentally. Overall, the MIS mat system offers a promising natural laboratory for determining the factors that control processes of geobiological interest such as O_2 production, and for evaluating mineralogical, isotopic, and organic signatures relevant to interpretation of the geologic record.

The dominance of the MIS mats by one genotype points to exceptionally low microbial diversity; such unevenness of microbial community structure is rivaled by just a few microbial communities in extreme subsurface environments (Chivian et al 2008, Denef et al 2010). On the one hand, this lack of diversity likely reflects the incredible versatility of MIS-Ph1; through its ability to thrive under a wide range of conditions it appears to have simultaneously occupied several key niches in the MIS environment. On the other hand, the uneven, low-diversity nature of the community indicates a lack of redundancy that could signal a fragility of the MIS community (Wittebolle et al 2009). This potential fragility should be considered in efforts to protect and preserve these unique ecosystems in the face of environmental change.

Acknowledgements

We gratefully acknowledge the NOAA Thunder Bay National Marine Sanctuary (particularly Jeff Grey, Russ Green, Wayne Lusardi, Joe Hoyt and Tane Casserly) for their assistance in field operations, the University of Michigan DNA Sequencing Core for DNA sequencing, Anja Schleicher for assistance with XRD, Steve Long and Michael Snider for assistance with microscopy and microphotography, and Ryan Lesniewski for assistance with metagenomic assembly. The authors also thank students in the BIO 370 *Biotechnology* course at UW-Stout who identified and described genes unique to the *P. autumnale* MIS-Ph1 genome as part of the course requirements. This work was supported by NSF grants MCB0603944 and MCB0604158 to BB and SN, and EAR1035955 to GJD, NDS, and BB.

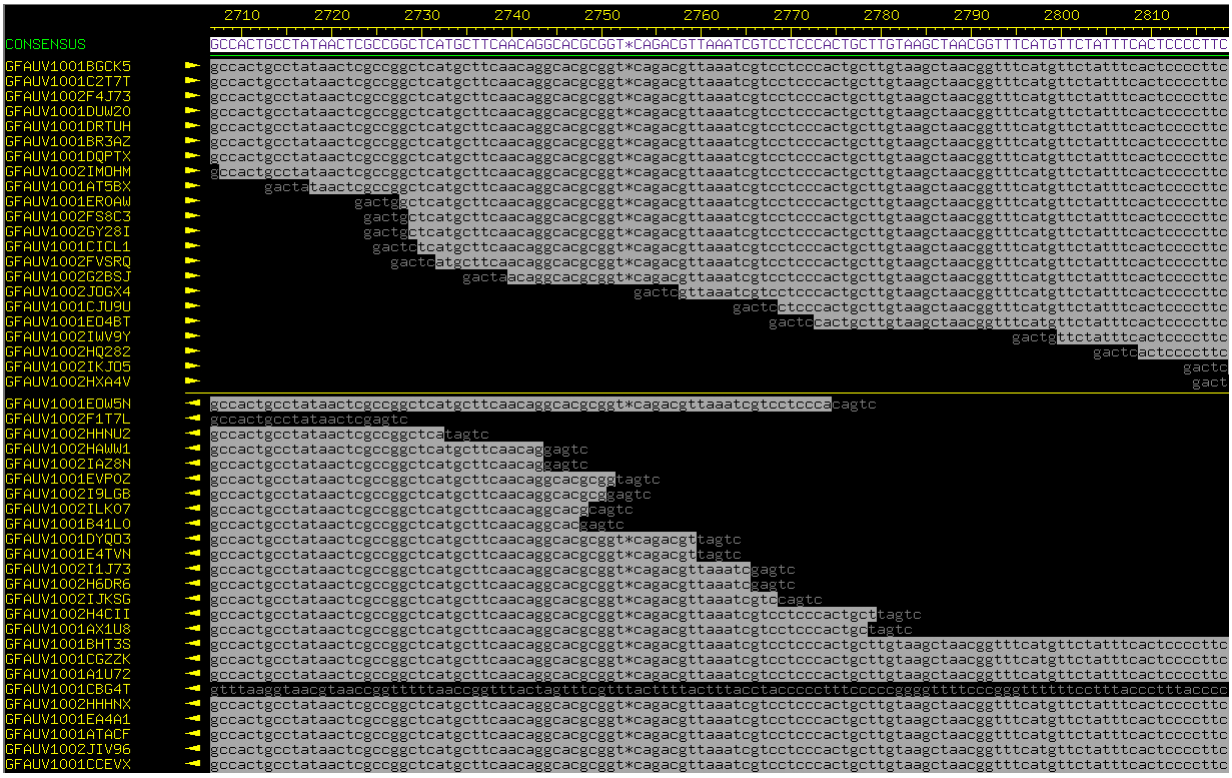
2.5 Appendix A

CHAPTER II Supplemental Material

Supplemental Information

Contents:

Supplemental Figure 1: Screen shot of a typical MIS-PH1 contig viewed through consed. Each line represents a different DNA sequencing read likely derived from a different MIS-Ph1 cell. Note the showing deep ~40X genomic coverage and the absence of polymorphisms, suggesting a clonal population.



Supplemental Table 2.1: Genes used as queries for sulfur oxidation genes.

Gene	Organism	NCBI Accession#
FccA	Alc. vinosum	AAA23316
FccB	Alc. vinosum	AAB86576
SorA	Sta. novella	AAF64400
SorB	Sta. novella	AAF64401
Dsr (All)	Alc. vinosum	U84760
SoxA	Par. denitrificans	CAA55827
SoxB	Par. denitrificans	CAA55824
SoxC	Par. denitrificans	CAA55829
SoxX	Par. denitrificans	CAB94379
SoxY	Par. denitrificans	CAB94380
SoxZ	Par. denitrificans	CAB94381
AprAB	Alc. vinosum	U84759
QmoA	Cba. tepidum	CT0866
QmoB	Cba. tepidum	CT0867
QmoC	Cba. tepidum	CT0868
Sqr	Geitlerinema sp. PCC 9228	AF242368
Sqr	Aph. Halophytica	AF242369
Sqr	Rho. capsulatus	CAA66112

Supplemental Table 2.2: List of genes and accession numbers

MIS genes of interest with IMG identifiers

Putative Gene	Bin	IMG Identifier
Putative genes for anoxygenic photosynthesis		
Sqr	MIS-Os	2050937426
Sqr	MIS-Ph1	2066677257
Sqr	MIS-Ph2	2050924083
Carbon acquisition and metabolism		
Glycogen Synthase	MIS-Ph1	2066679283
Carbohydrate Branching	MIS-Ph1	2066678045
Carbohydrate Branching	MIS-Ph1	2066677165
Carbohydrate Debranching	MIS-Ph1	2066679940
Glucokinase	MIS-Ph1	2066679941
Ferredoxin Oxidoreductase	MIS-Ph1	2066680410
Acetate Kinase	MIS-Ph1	2066678572
OprB	MIS-Ph1	2066680010
ABC-type Transporters	MIS-Ph1	2066676523- 2066676525
ABC-type Transporters	MIS-Ph1	2066678479- 2066678484
Carbohydrate Transporter	MIS-Ph1	2066678493
Sugar Catabolism	MIS-Ph1	2066681927
Sugar Catabolism	MIS-Ph1	2066681203
Sugar Catabolism	MIS-Ph1	2066681702
Sugar Catabolism	MIS-Ph1	2066676806
Oxygen sensing, regulation, and metabolism		
Superoxide dismutase	MIS-Ph1	2066676810
Cytochrome c peroxidase	MIS-Ph1	2066679523
Cytochrome c peroxidase	MIS-Ph1	2066679959
Glutathione peroxidase	MIS-Ph1	2066679615
Peroxiredoxin	MIS-Ph1	2066676812
Genomic insights into P. autumnale interactions with the mat community		
Eco57I restriction endonuclease	MIS-Ph1	2066678847
Eco57I restriction endonuclease	MIS-Ph1	2066680058
Restriction-modification system	MIS-Ph1	2066677275
Restriction-modification system	MIS-Ph1	2066678508

HNH endonuclease	MIS-Ph1	2066679190
Exonuclease	MIS-Ph1	2066678883
Restriction endonuclease	MIS-Ph1	2066677343
Outer membrane protective protein	MIS-Ph1	2066678472
DNA methylases	MIS-Ph1	2066679231, 2066678696
AIPR	MIS-Ph1	2066676575
Aspartyl protease	MIS-Ph1	2066679955
Phage integrase	MIS-Ph1	2066676918
Colicin D	MIS-Ph1	2066676804
Vancomycin resistance gene	MIS-Ph1	2066676642
β-lactamase	MIS-Ph1	2066677179
Hemaglutinin-like domain	MIS-Ph1	2066679448
Hemolysin	MIS-Ph1	2066677404
Hemolysin	MIS-Ph1	2066677627
Mechanoreceptors	MIS-Ph1	2066677234
Mechanoreceptors	MIS-Ph1	2066680012
Mechanoreceptors	MIS-Ph1	2066676983
Signal transduction pathways	MIS-Ph1	
Signal peptidases	MIS-Ph1	2066679126
Signal transducing histidine kinase	MIS-Ph1	2066677562
Extracellular signaling molecule	MIS-Ph1	2066676908
Biofilm-forming	MIS-Ph1	2066678737
Biofilm-forming	MIS-Ph1	2066676776

Environmental sensing, regulation, and adaptation

Signal transduction	MIS-Ph1	2066677572
Signal transduction	MIS-Ph2	2066678224
Signal transduction	MIS-Ph3	2066677573

Evolution and genetic diversity

Competence protein	MIS-Ph1	2066679070

Supplemental Table 2.3: Single copy COG genes used to quantify genome completeness

			Out of 60 cyanos	
COG	**Description**	**Count**	**2+/genome**	**absent**
COG0012	GTP-binding protein YchF	1	0	0
COG0016	Phenylalanyl-tRNA synthetase, alpha subunit	1	1	0
COG0018	Arginyl-tRNA synthetase	1	14	0
COG0048	Ribosomal protein S12P	1	0	0
COG0049	Ribosomal protein S7P	1	0	1
COG0052	Ribosomal protein S2P	1	0	0
COG0060	Isoleucyl-tRNA synthetase	2	3	0
COG0080	Ribosomal protein L11P	1	0	1
COG0081	Ribosomal protein L1P	1	0	1
COG0085	DNA-directed RNA polymerase subunit beta	2	2	1
COG0087	Ribosomal protein L3P	1	1	0
COG0088	Ribosomal protein L4P	1	0	0
COG0090	Ribosomal protein L2P	1	0	0
COG0091	Ribosomal protein L22	1	0	0
COG0092	Ribosomal protein S3P	1	0	0
COG0093	Ribosomal protein L14P	1	1	0
COG0094	Ribosomal protein L5P	1	0	0
COG0096	Ribosomal protein S8P	1	0	0
COG0097	Ribosomal protein L6P	1	0	0
COG0098	Ribosomal protein S5P	1	0	0
COG0099	Ribosomal protein S13P	1	0	0
COG0100	Ribosomal protein S11P	1	0	0
COG0102	Ribosomal protein L13P	1	0	0
COG0103	Ribosomal protein S9P	1	0	0
COG0124	Histidyl-tRNA synthetase	2	22	0
COG0143	Methionyl-tRNA synthetase	1	1	0
COG0172	Seryl-tRNA synthetase	1	0	0
COG0184	Ribosomal protein S15P/S13E	1	1	0
COG0185	Ribosomal protein S19P	1	0	1
COG0186	Ribosomal protein S17P	1	1	0
COG0197	Ribosomal protein L16P	1	2	0
COG0200	Ribosomal protein L15P	1	1	0
COG0201	Protein translocase subunit secY alpha	1	0	1
COG0202	DNA-directed RNA polymerase subunit alpha	1	0	0

COG0256	Ribosomal protein L18	1	0	0
COG0495	Leucyl-tRNA synthetase	2	1	0
COG0522	Ribosomal protein S4P	1	0	0
COG0525	Ribosomal proteinalyl-tRNA synthetase	1	1	0
COG0533	O-sialoglycoprotein endopeptidase	1	1	1
COG0541	Signal recognition particle subunit FFH	1	1	1

2.6 References

Andersen DT, Sumner DY, Hawes I, Webster-Brown J, Mckay CP (2011). Discovery of large conical stromatolites in Lake Untersee, Antarctica. *Geobiology* 9: 280-293.

Arieli B, Shahak Y, Taglicht D, Hauska G, Padan E (1994). Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. *Journal of Biological Chemistry* 269: 5705-5711.

Bachar A, Omoregie E, de Wit R, Jonkers HM (2007). Diversity and function of Chloroflexus-like bacteria in a hypersaline microbial mat: phylogenetic characterization and impact on aerobic respiration. *Appl Environ Microbiol* 73: 3975-3983.

Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL *et al* (2004). Dating the rise of atmospheric oxygen. *Nature* 427: 117-120.

Belkin S, Padan E (1978). Hydrogen metabolism in the facultative anoxygenic Cyanobacteria (Blue-Green Algae) *Oscillatoria limnetica* and *Aphanothece halophytica*. *Archives of Microbiology* 116: 109-111.

Bertrand EM, Saito MA, Rose JM, Riesselman CR, Lohan MC, Noble AE *et al* (2007). Vitamin B-12 and iron colimitation of phytoplankton growth in the Ross Sea. *Limnology and Oceanography* 52: 1079-1093.

Biddanda BA, Opsahl S, Benner R (1994). Plankton respiration and carbon flux through bacterioplankton on the Louisiana Shelf. *Limnology and Oceanography* 39: 1259-1275.

Biddanda BA, Coleman DF, Johengen TH, Ruberg SA, Meadows GA, VanSumeran HW *et al* (2006). Exploration of a submerged sinkhole ecosystem in Lake Huron. *Ecosystems* 9: 828-842.

Biddanda BA, Nold SC, Ruberg SA, Kendall ST, Sanders TG, Gray JJ (2009). Great lakes sinkholes: a microbiogeochemical frontier. *Eos* 90: 61-68.

Biddanda BA, Nold SC, Dick GJ, Kendall ST, Vail JH, Ruberg SA *et al* (in press). Rock, water, microbes: Underwater sinkholes in Lake Huron are habitats for ancient microbial life. *Nature Education*.

Black TJ (1983). Selected views of the tectonics, structure and karst in northern lower Michigan. In: Kimmel RE (ed). *Michigan Basin Geological Society Field Conference Proceedings*. Michigan Basin Geological Society: Lansing, MI. pp 11-35.

Blankenship R, Sadekar S, Raymond J (2007). The evolutionary transition from anoxygenic to oxygenic photosynthesis. In: Falkowski PG, Knoll AH (eds). *Evolution of primary producers in the sea*. Elsevier. pp 21-35.

Blenkinsopp SA, Costerton JW (1991). Understanding bacterial biofilms. *Trends in Biotechnology* 9: 138-143.

Bosak T, Newman DK (2003). Microbial nucleation of calcium carbonate in the Precambrian 31: 577-580.

Bronstein M, Schutz M, Hauska G, Padan E, Shahak Y (2000). Cyanobacterial sulfide-quinone reductase: cloning and heterologous expression. *J Bacteriol* 182: 3336-3344.

Buhring SI, Sievert SM, Jonkers HM, Ertefai T, Elshahed MS, Krumholz LR *et al* (2011). Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems. *Geobiology* 9: 166-179.

Canfield DE, Des Marais DJ (1993). Biogeochemical Cycles of Carbon, Sulfur, and Free Oxygen in a Microbial Mat. *Geochim Cosmochim Ac* 57: 3971-3984.

Casamayor EO, Garcia-Cantizano J, Pedros-Alio C (2008). Carbon dioxide fixation in the dark by photosynthetic bacteria in sulfide-rich stratified lakes with oxic-anoxic interfaces. *Limnology and Oceanography* 53: 1193-1203.

Castenholz RW (1976). Effect of sulfide on blue-green algae of hot springs .1. New-Zealand *Journal of Phycology* 12: 54-68.

Castenholz RW (1977). Effect of sulfide on blue-green algae of hot springs .2. Yellowstone National Park. *Microbial Ecology* 3: 79-105.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T *et al* (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* 14: 1147-1159.

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, Desantis TZ *et al* (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* 322: 275-278.

Cohen Y, Jorgensen BB, Padan E, Shilo M (1975a). Sulphide-dependent anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. *Nature* 257: 489-492.

Cohen Y, Padan E, Shilo M (1975b). Facultative anoxygenic photosynthesis in the cyanobacterium Oscillatoria limnetica. *J Bacteriol* 123: 855-861.

Cohen Y, Jorgensen BB, Revsbech NP, Poplawski R (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl Environ Microbiol* 51: 398-407.

Comte K, Sabacka M, Carre-Mlouka A, Elster J, Komarek J (2007). Relationships between the Arctic and the Antarctic cyanobacteria; three Phormidium-like strains evaluated by a polyphasic approach. *Fems Microbiology Ecology* 59: 366-376.

Cowan DA, Tow LA (2004). Endangered antarctic environments. *Annu Rev Microbiol* 58: 649-690.

Denef VJ, Mueller RS, Banfield JF (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *Isme Journal* 4: 599-610.

DeRuyter YS, Fromme P (2008). Molecular structure of the photosynthetic apparatus. In: Herrero A, Flores E (eds). *The Cyanobacteria, molecular biology, genomics, and evolution*. Caister Academic Press: Norfolk, UK. pp 217-269.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al* (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* 10: R85.

Dietrich LEP, Teal TK, Price-Whelan A, Newman DK (2008). Redox-active antibiotics control gene expression and community behavior in divergent bacteria. *Science* 321: 1203-1206.

Dupraz C, Reid RP, Braissant O, Decho AW, Norman RS, Visscher PT (2009). Processes of carbonate precipitation in modern microbial mats. *Earth-Science Reviews* 96: 141-162.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* 320: 1034-1039.

Fontes MLS, Suzuki MT, Cottrell MT, Abreu PC (2011). Primary Production in a Subtropical Stratified Coastal Lagoon-Contribution of Anoxygenic Phototrophic Bacteria. *Microbial Ecology* 61: 223-237.

Frigaard N-U, Dahl C, Robert KP (2008). Sulfur Metabolism in Phototrophic Sulfur Bacteria. *Advances in Microbial Physiology*. Academic Press. pp 103-200.

Garlick S, Oren A, Padan E (1977). Occurrence of facultative anoxygenic photosynthesis among filamentous and unicellular cyanobacteria. *J Bacteriol* 129: 623-629.

Gingras M, J.W. Hagadorn, A. Seilacher, S.V. Lalonde, E. Pecoits, D. Petrash, and K.O. Konhauser (2011). Possible evolution of mobile animals in association with microbial mats. *Nature Geoscience*.

Grotzinger JP, Knoll AH (1999). Stromatolites in Precambrian carbonates: Evolutionary mileposts or environmental dipsticks? *Annual Review of Earth and Planetary Sciences* 27: 313-358.

Hawes I, Schwarz AM (1999). Photosynthesis in an extreme shade environment: Benthic microbial mats from Lake Hoare, a permanently ice-covered Antarctic lake. *Journal of Phycology* 35: 448-459.

He Y-Y, Hader D-P (2002). Reactive oxygen species and UV-B: effect on cyanobacteria. *Photochemical & Photobiological Sciences* 1: 729-736.

Holland HD (2006). The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci* 361: 903-915.

Horodyski RJ, Knauth LP (1994). Life on land in the precambrian. *Science* 263: 494-498.

Imbus SW, Macko SA, Elmore RD, Engel MH (1992). Stable isotope (C, S, N) and molecular studies on the Precambrian Nonesuch Shale (Wisconsin-Michigan, USA) - evidence for differential preservation rates, depositional environment and hydrothermal influence. *Chem Geol* 101: 255-281.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* 106: 16925-16929.

Jorgensen BB, Revsbech NP, Blackburn TH, Cohen Y (1979). Diurnal cycle of oxygen and sulfide microgradients and microbial photosynthesis in a cyanobacterial mat sediment. *Appl Environ Microbiol* 38: 46-58.

Jorgensen BB, Revsbech NP, Cohen Y (1983). Photosynthesis and structure of benthic microbial mats - microelectrode and sem studies of 4 cyanobacterial communities. *Limnology and Oceanography* 28: 1075-1093.

Jorgensen BB, Cohen Y, Revsbech NP (1986). Transition from anoxygenic to oxygenic photosynthesis in a *Microcoleus chthonoplastes* cyanobacterial mat. *Appl Environ Microbiol* 51: 408-417.

Kana BD, Weinstein EA, Avarbock D, Dawes SS, Rubin H, Mizrahi V (2001). Characterization of the cydAB-Encoded Cytochrome bd Oxidase from Mycobacterium smegmatis. *J Bacteriol* 183: 7076-7086.

Kenward PA, Goldstein RH, Gonzalez LA, Roberts JA (2009). Precipitation of low-temperature dolomite from an anaerobic microbial consortium: the role of methanogenic Archaea. *Geobiology* 7: 556-565.

Komarek J, Kling H, Komarkova J (2003). Filamentous cyanobacteria. In: Wehr JD, Sheath R (eds). *Freshwater algae of North America: Ecology and classification*. Elsevier Science: San Diego, CA, USA. pp 117-196.

Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ (2005). The paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* 102: 11131-11136.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948.

Lyons TW, Anbar AD, Severmann S, Scott C, Gill BC (2009). Tracking Euxinia in the Ancient Ocean: A Multiproxy Perspective and Proterozoic Case Study. *Annual Review of Earth and Planetary Sciences* 37: 507-534.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P *et al* (2011). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9: 467-477.

Mejean A, Mazmouz R, Mann S, Calteau A, Medigue C, Ploux O (2010). The Genome Sequence of the Cyanobacterium Oscillatoria sp. PCC 6506 Reveals Several Gene Clusters Responsible for the Biosynthesis of Toxins and Secondary Metabolites. *J Bacteriol* 192: 5264-5265.

Miller SR, Bebout BM (2004). Variation in sulfide tolerance of photosystem II in phylogenetically diverse cyanobacteria from sulfidic habitats. *Appl Environ Microbiol* 70: 736-744.

Nold S, Ward D (1996). Photosynthate Partitioning and Fermentation in Hot Spring Microbial Mat Communities. *Appl Environ Microbiol* 62: 4598-4607.

Nold SC, Pangborn JB, Zajack HA, Kendall ST, Rediske RR, Biddanda BA (2010a). Benthic bacterial diversity in submerged sinkhole ecosystems. *Appl Environ Microbiol* 76: 347-351.

Nold SC, Zajack HA, Biddanda BA (2010b). Eukaryal and archaeal diversity in a submerged sinkhole ecosystem influenced by sulfur-rich, hypoxic groundwater. *Journal of Great Lakes Research* 36: 366-375.

Oren A, Padan E, Avron M (1977). Quantum yields for oxygenic and anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. *Proceedings of the National Academy of Sciences of the United States of America* 74: 2152-2156.

Oren A, Padan E (1978). Induction of anaerobic, photoautotrophic growth in the cyanobacterium *Oscillatoria limnetica*. *Journal of Bacteriology* 133: 558-563.

Oren A, Shilo M (1979). Anaerobic heterotrophic dark metabolism in the cyanobacterium <i>Oscillatoria limnetica: Sulfur respiration and lactate fermentation. *Arch Microbiol* 122: 77-84.

Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991). Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography* 36: 846-859.

Pavlov AA, Kasting JF, Eigenbrode JL, Freeman KH (2001). Organic haze in Earth's early atmosphere: source of low-13C late Archean kerogens? *Geology* 29: 1003-1006.

Pedros-Alio C, Garcia-Cantizano J, Calderon J (1993). Bacterial production in anaerobic water columns. In: Kemp P, Sherr B, Sherr B, Cole J (eds). *Handbook of methods in aquatic microbial ecology*. Lewis Publishers: Boca Raton. pp 519-530.

Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* 8: R10.

Rashby SE, Sessions AL, Summons RE, Newman DK (2007). Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci U S A* 104: 15099-15104.

Retallack GJ, Mindszenty A (1994). Well preserved late Precambrian Paleosols from Northwest Scotland. *Journal of Sedimentary Research* 64: 264-281.

Revsbech NP, Jorgensen BB, Blackburn TH, Cohen Y (1983). Microelectrode studies of the photosynthesis and O_2, H_2S, and pH profiles of a microbial mat. *Limnology and Oceanography* 28: 1062-1074.

Richardson LL, Castenholz RW (1987). Diel Vertical Movements of the Cyanobacterium Oscillatoria terebriformis in a Sulfide-Rich Hot Spring Microbial Mat. *Appl Environ Microbiol* 53: 2142-2150.

Riding R (2006). Cyanobacterial calcification, carbon dioxide concentrating mechanisms, and Proterozoic-Cambrian changes in atmospheric composition. *Geobiology* 4: 299-316.

Riding R (2011). Calcified cyanobacteria. In: Reitner J, Thiel V (eds). *Encyclopedia of Geobiology*. Springer: Heidelberg. pp 211-223.

Ruberg SA, Kendall ST, Biddanda BA, Black T, Nold SC, Lusardi WR *et al* (2008). Observations of the Middle Island Sinkhole in Lake Huron - A unique hydrogeologic and glacial creation of 400 million years. *Marine Technology Society Journal* 42: 12-21.

Rye R, Holland HD (2000). Life associated with a 2.76 Ga ephemeral pond?: Evidence from Mount Roe #2 paleosol. *Geology* 28: 483-486.

Sanders JTG, Biddanda BA, Stricker CA, Nold SC (2011). Stable Isotope Analysis Reveals Benthic Macroinvertebrate and Fish Communities Linked to Submerged Groundwater Vents in Lake Huron. *Aquatic Biology* 12: 1-11.

Schidlowski M (ed) (2000) *Carbon isotopes and microbial sediments*. Springer-Verlag: Berlin.

Schutz M, Shahak Y, Padan E, Hauska G (1997). Sulfide-quinone reductase from Rhodobacter capsulatus. Purification, cloning, and expression. *J Biol Chem* 272: 9890-9894.

Shen Y, Knoll AH, Walter MR (2003). Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* 423: 632-635.

Stal LJ (1995). Tansley Review No. 84. Physiological Ecology of Cyanobacteria in Microbial Mats and Other Communities. *New Phytologist* 131: 1-32.

Stal LJ, Moezelaar R (1997). Fermentation in cyanobacteria. *FEMS Microbiology Reviews* 21: 179-211.

Stal LJ (2000). Cyanobacterial Mats and Stromatolites. In: Whitton B, Potts M (eds). *The Ecology of Cyanobacteria*. Kluwer Academic Publishers: Dordrecht/London/Boston. pp 61-120.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.

Summons RE, Jahnke LL, Hope JM, Logan GA (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400: 554-557.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596-1599.

Taton A, Grubisic S, Balthasart P, Hodgson DA, Laybourn-Parry J, Wilmotte A (2006a). Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. *Fems Microbiology Ecology* 57: 272-289.

Taton A, Grubisic S, Ertz D, Hodgson DA, Piccardi R, Biondi N *et al* (2006b). Polyphasic study of Antarctic cyanobacterial strains. *Journal of Phycology* 42: 1257-1270.

Van Lith Y, Warthmann R, Vasconcelos C, Mckenzie JA (2003). Sulphate-reducing bacteria induce low-temperature Ca-dolomite and high Mg-calcite formation. *Geobiology* 1: 71-79.

Vasconcelos C, Mckenzie JA, Bernasconi S, Grujic D, Tien AJ (1995). Microbial Mediation as a Possible Mechanism for Natural Dolomite Formation at Low-Temperatures. *Nature* 377: 220-222.

Vincent WF (2000). Cyanobacterial dominance in the Polar regions. In: Potts BAWaM (ed). *The Ecology of Cyanobacteria*. Kluver Academic Publishers: Amsterdam. pp 321-340.

Walter LM, Ku TCW, Muehlenbacks K, Patterson WP, Bonnell L (2007). Controls on δ13C of dissolved inorganic carbon in marine pore waters: an integrated case study of isotope exchange during syndepositional recrystallization of biogenic carbonate sediments (South Florida Platform, USA). . *Deep Sea Research Part II: Topical Studies in Oceanography* 54: 1163-1200.

Walter MR, Bauld J (1983). The association of sulphate evaporites, stromatolitic carbonates and glacial sediments: examples from the Proterozoic of Australia and the Cainzoic of Antarctica. *Precambrian Research* 21: 129-148.

Wang Y, Kern SE, Newman DK (2010). Endogenous phenazine antibiotics promote anaerobic survival of *Pseudomonas aeruginosa* via extracellular electron transfer. *Journal of Bacteriology* 192: 365-369.

Warthmann R, van Lith Y, Vasconcelos C, McKenzie JA, Karpoff AM (2000). Bacterially induced dolomite precipitation in anoxic culture experiments. *Geology* 28: 1091-1094.

Welander PV, Hunter RC, Zhang L, Sessions AL, Summons RE, Newman DK (2009). Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* 191: 6145-6156.

Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc Natl Acad Sci U S A* 107: 8537-8542.

Wharton RA, Parker BC, Simmons GM (1983). Distribution, Species Composition and Morphology of Algal Mats in Antarctic Dry Valley Lakes. *Phycologia* 22: 355-365.

Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K *et al* (2009). Initial community evenness favours functionality under selective stress. *Nature* 458: 623-626.

CHAPTER III

Metabolic function and microbial mediation of geochemical cycling revealed by community genomic analysis and gene expression of a low-O_2 cyanobacterial mat

Alexander A. Voorhies[1], and Gregory J. Dick[1,2]

[1] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, 48109

[2] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109

Abstract

Metagenomic and metatranscriptomic analysis was conducted on 15 samples collected over five years from a benthic cyanobacterial mat in a low-O_2 sulfidic environment to investigate phototrophic, lithotrophic, and heterotrophic metabolisms. Random shotgun sequencing, de novo assembly and binning yielded 32 genomic bins, including genomes for the dominant cyanobacteria, three bacterial candidate divisions, and for numerous Proteobacteria involved in sulfate reduction, sulfide oxidation, and heterotrophy. A near-complete genome was recovered for candidate division WS3, as well as partial genomes for SM2F11 and RF3. This is the first substantial genomic and transcriptomic data to be reported for SM2F11 and RF3, both of which appear to be novel with regard to gene content. Mapping of transcripts to the WS3 genome

revealed expression of genes involved in chemotaxis, H_2 oxidation and aerobic respiration using a cytochrome C oxidase/reductase, and gene content suggests WS3 is a versatile aerobic oxidizer of H_2 and sulfur species. Genomic bins were also recovered from several *Betaproteobacteria*, *Epsilonproteobacteria*, and *Gammaproteobacteria*. *Thiotrichales* were found to primarily express genes for sulfur oxidation. *Desulfobacterales* genomes recruited the majority of transcripts for sulfate reduction, suggesting that these organisms are chiefly responsible for production of sulfide at MIS.

Higher abundance of transcripts was observed for the photosynthetic reaction core genes *psaA* and *psbA* from cyanobacteria in the dark than in samples collected during the day. This is in contrast to previously studied cyanobacteria that show the peak of *psaA* and *psbA* transcription during the day, coincident with photosynthesis. *Phormidium* transcripts were recovered for sulfide quinone oxidoreductase, which is known to oxidize sulfide during anoxygenic photosynthesis by cyanobacteria. We also observed a high ratio of transcripts for photosystem I to photosystem II genes, consistent with a genetic regulatory shift towards photosystem I for anoxygenic photosynthesis in the presence of sulfide. We infer that *Phormidium* populations within the mat are likely performing a combination of anoxygenic and oxygenic photosynthesis. Overall, these genomic and transcriptomic results link specific groups of microbes to observed redox chemistry that underpins biogeochemical cycling at MIS.

3.1 Introduction

Recent advances in environmental shotgun sequencing (metagenomics) (Wrighton et al 2012) and single cell genomics (Rinke et al 2013) have begun to shed light on the functions and potential environmental impacts of microbes that have remained unstudied because of their

resistance to isolation and culture. While a small number of model cultured microbes have been well studied and characterized, they represent only a fraction of microbial diversity in nature as revealed by culture-independent 16s small subunit ribosomal RNA (16s) studies (Rajendhran and Gunasekaran 2011, Rappe and Giovannoni 2003). Culture independent genomic approaches have enabled the reconstruction of genomes and prediction of metabolic potential for microbes in diverse environments, and added to our understanding of biogeochemical cycles in these systems (Castelle et al 2013, Falkowski et al 2008). Shotgun sequencing of environmentally expressed RNA (metatranscriptomics) allows us to take culture independent snapshots of transcriptional activity, which can illuminate the dynamics of major metabolic pathways influencing biogeochemical cycling in time and space (Frias-Lopez et al 2008).

Oxygenic photosynthesis, the cellular ability to harvest light energy using water as the electron donor and producing O_2 as a byproduct, is thought to be largely responsible for the oxygenation of Earth's atmosphere and the resulting shift in redox chemistry. Cyanobacteria are thought to have first evolved oxygenic photosynthesis by coupling together two different photosystems derived from anoxygenic bacteria (Blankenship et al 2007). Transcription of photosynthesis genes by cyanobacteria has been shown to be regulated on a day/night cycle (Ito et al 2009, Stoeckel et al 2008), peaking during the day in both laboratory pure cultures (Stockel et al 2011) and environmental samples (Frias-Lopez et al 2008). However, these transcriptomic studies have focused on unicellular cyanobacteria that are ubiquitous in surface freshwater and marine waters, but may not be representative of filamentous cyanobacterial expression in benthic microbial mats.

While most cyanobacteria typically conduct oxygenic photosynthesis and are inhibited by sulfide, some cyanobacteria have been shown to perform anoxygenic photosynthesis in the

presence of sulfide (Cohen et al 1986). During oxygenic photosynthesis, photosystem II (PS II) and photosystem I (PS I) operate in tandem. If PS II is inhibited by sulfide, cyanobacteria capable of anoxygenic photosynthesis shut down PS II and utilize sulfide as an electron donor for PS I (Bronstein et al 2000b, Schutz et al 1997). Sulfide-quinone oxidoreductase (SQR) is a key enzyme for anoxygenic photosynthesis by cyanobacteria (Bronstein et al 2000b) that oxidizes sulfide and transfers electrons to PS I through the quinone pool, effectively bypassing PS II (Arieli et al 1994a, Bronstein et al 2000b, Schutz et al 1997). SQR is a diverse protein family that has also been linked to sulfide detoxification in cyanobacteria and other phototrophs (Bronstein et al 2000b). Although laboratory studies have elucidated the physiological responses of cyanobacteria to sulfide and the role of SQR in anoxygenic photosynthesis, little is known about how anoxygenic photosynthesis by cyanobacteria operates at the genomic or transcriptomic level, or in natural communities of microorganisms.

The Middle Island Sinkhole (MIS) in Lake Huron, MI is host to cyanobacterial mats living in low-O_2, sulfidic conditions. Although low-O_2conditions are often inhibitory to modern cyanobacteria, they were common throughout much of their evolutionary history (Johnston et al 2009a), thus MIS provides a valuable functional analog of ancient cyanobacteria. The mats sit atop anoxic, organic-rich sediments where microbial methanogenesis and sulfate reduction produce methane and sulfide, leading to sharp and dynamic redox gradients that favor organisms with versatile metabolisms (Nold et al 2010b, Nold et al 2010d, Voorhies et al 2012). Previous studies of MIS have indicated that the mats are metabolically versatile, having the ability to conduct oxygenic photosynthesis, anoxygenic photosynthesis, and chemosynthesis, but are dominated by cyanobacteria (Voorhies et al 2012).

The MIS mats were shown to be a net sink for O_2, suggesting a significant contribution to primary production by anoxygenic photosynthesis, likely by *Phormidium* or *Oscillatoria*, whose genomes revealed SQR genes that could be used to perform anoxygenic photosynthesis (Voorhies et al 2012). In order to investigate metabolic gene expression in low O_2 cyanobacterial mats, and pair organisms to specific redox chemistry at MIS, we used metagenomics and metatranscriptomics to reconstruct genomes for the most abundant organisms from fifteen samples collected at seven time points collected between 2007 and 2012. Transcript abundance showed that Phormidium is performing both oxygenic and anoxygenic photosynthesis, likely using SQR. Various Proteobacteria were responsible for sulfur cycling at MIS, and genomes and expression are reported for three novel bacterial candidate divisions, WS3, SM2F11 and RF3.

3.2 Materials and Methods

Sample collection and sequencing

15 environmental mat samples were collected by scuba divers aboard the *R/V Storm* between 2007 and 2012 (Supplemental Table 3.1 in Appendix B). Samples were collected from within a 100m area by hand cores containing sediments, mat and overlying groundwater. Cores were rapidly transferred to the surface and mats were extracted from the cores and submerged in RNA Later immediately shipboard, with less than five minutes passing from collection to preservation. Before preservation, mat samples were quickly rinsed in ground water at the top of the core to remove as much underlying sediment as possible, but some sediment entrainment was unavoidable. These samples are diverse in nature, from microbial structures we refer to as 'fingers', to mat lying flat on the overlying sediments, and contain varying levels of sediment,

especially the two samples from 2010 which were heavily entrained with sediment. In order to maximize reconstruction of genomes from low abundance members, all 15 samples were used in a combined metagenomic assembly.

DNA was extracted and processed for shotgun metagenomic sequencing (without amplification) as previously described (Voorhies et al 2012). Samples were sequenced using an Illumina Hi Seq 2000 instrument producing paired end reads at the University of Michigan DNA Sequencing Core. In 2012, three samples of mat were collected at approximately 1pm (day), and 1am (night) from within a 9 m^2 sampling area. RNA was extracted from these six samples, randomly amplified with the MessageAmp II-Bacteria Kit (Ambion), and converted to complementary (cDNA) using the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen), both as previously described (Frias-Lopez et al 2008). No ribosomal RNA (rRNA) removal methods were used and cDNA was sequenced at the University of Michigan DNA Sequencing Core on an Illumina Hi Seq 2000 instrument producing paired end reads. The 2012 samples were collected with the intention of comparing RNA transcript abundance of photosynthetic (and various other) metabolic genes in day versus night samples.

Assembly and genomic analysis

A total of 922 million reads from all genomic DNA (gDNA) samples were combined and co-assembled using IDBA UD (Peng et al 2012). Contigs greater than 4kb were sorted into genomic bins of varying genome completeness (Supplemental Figure 3.1 in Appendix B) using emergent self-organizing maps of tetranucleotide frequencies as described previously (Dick et al 2009b). *Phormidium* contigs between 1kb-4kb were identified by BLASTp to previous assemblies and added to the *Phormidium* bin and designated as 'short contigs'. Taxonomy of genomic bins was determined by BLASTn of 16s genes against the Silva Small Subunit RNA

database release 115 (Quast et al 2013), or in the absence of a 16s gene, by tallying best BLAST hits to a certain taxonomic group where at least half of all genes in a bin got their best hit to that group. Genus level assignment was only made for bins with a 16s gene with greater than 97% nucleotide identity to a known organism.

Assembly was validated against previous assemblies of MIS using multiple sequencing platforms and assembly programs including a previously published metagenome (Voorhies et al 2012), Illumina assemblies of individual samples created with Velvet (Zerbino and Birney 2008), and by manual curation using the genomic viewers IGV (Robinson et al 2011) and Geneious (Biomatters 2013) to visualize reads mapped to contigs and genes using BWA (Li and Durbin 2009). Gene calling and annotation was performed by the Joint Genome Institute's Integrated Microbial Genomes Expert Review portal (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) and annotation of specific genes of interest were verified using BLASTp with cutoffs of 150 bitscore and ID of 30% to NR, COG, KEGG and Pfam.

Read coverage for cDNA and gDNA from each sample was assessed by mapping reads to contigs and genes using BWA using default settings. Read coverage was normalized for the length of the contig, gene or bin of interest and sequencing effort based on formula 3.1. Day and night cDNA read coverage was calculated by averaging the three day and three night samples, and error bars represent standard deviation of coverage across samples.

$$\text{avg. read coverage} = \left(\frac{\text{Reads mapped} * \text{Average read length (100bp)}}{\text{Total length in bp of gene, contig or genome}}\right) * \text{Sequencing effort normalization}$$

Formula 3.1 Formula used for calculating average coverage of a gene, contig or bin of interest. Sequencing effort normalization is a ratio that accounts for variable sequencing effort between samples.

3.3 Results and Discussion

Community composition and function

Metagenomic assembly produced 420,571 contiguous sequences (contigs) greater than 1kb, comprising 1,140,878,834 bp of consensus sequence and 1,478,052 predicted genes. Genomic binning by emergent self-organizing maps of tetranucleotide frequency (Supplemental Figure 3.1 in Appendix B) yielded 32 genomic bins and 137 estimated genomes representing 12 phyla (Table 3.1). Cyanobacteria were the most abundant organisms in the mat, with *Phormidium* being present at nearly 4000x genome coverage, and *Oscillatoria* present at approximately 500x genome coverage. Together the *Phormidium* and *Oscillatoria* make up more than 50% of the mat community in eleven out of fifteen samples, as well as total gDNA from all samples combined (Figure 3.1). While community membership is dynamic across time and space, *Phormidium* is consistently the dominant organism in the MIS mats.

Genomic bins were also recovered for Acidobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria, Spirochaetes, Verrucomicrobia and eukaryotes, as well as a Bacillariophyta (diatom) chloroplast, which encodes its own genome. Many of these groups are commonly found in anaerobic or hypoxic sediments and other environments (Tsertova et al 2011, Yau et al 2013), and several of the low abundance genomic bins have been shown by 16s survey to be enriched in the sediments below the mats at MIS (Nold et al 2010b). Among the genomic bins containing 16s genes (Supplemental Table 3.2 in Appendix B) were representatives of three bacterial candidate divisions: a near complete genome for WS3 (3.9Mbp), a partial genome of SM2F11 (1.2Mbp), and a partial genome of RF3 (170Kbp). Genome completeness was estimated based on the presence of 36 single copy house-keeping genes (Ciccarelli et al 2006), with WS3 containing a single copy of all 36 genes. The SM2F11

bin contains 2 partial genomes with an estimated completeness of 81%, while the RF3 bin has an estimated completeness of 17% (Table 3.1).

Table 3.1 Genomic bin statistics and metabolism

	Bin Statistics						C Metabolism				N Cycling	Photosynthesis		O_2 Resp	Ox Pathways			S Cycling					
Classification	Total Length of Consensus Sequence	Number of Genes	Estimated Genomes per Bin	Estimated Max Completeness	Total Avg gDNA Coverage	Total Avg cDNA Coverage	Autotrophy (RuBisCo)	Autotrophy rTCA (*aclB*)	Autotrophy WL Pathway (CO dehydrogenase)	Heterotrophy (C transport)	Denitrification (*nirK*)	Photosystem I (*psaA*)	Photosystem II (*psbA*)	Oxygen respiration (Cyt C, cbb3 or bd1 type oxidases)	CO oxidation (*coxL*)	Hydrogen oxidation (*hupL/hydB*)	Methane/Ammonia oxidation (*mmoC*)	S Reduction (*dsrA*)	S Oxidation (*rdsrA*)	Sulfide Redox (SQR)	Sulfide Redox (fcc)	Sulfite Oxidation (*aprA*)	Thiosulfate Oxidation (*soxZ*)
Acidobacteria Holophagae	8,803,791	9,168	4	92%	26	0.1		X		X				X	X				X				
Bacteroidetes Bacteroidia	2,042,926	1,727	2	75%	39	6.6								X					X			X	
Bacteroidetes Amoebophilus	517,169	676	1	72%	116	5.0																	
Bacteroidetes Cytophagia	443,445	390	1	3%	28	0.0																X	
Bacteroidetes Flavobacteriia	3,350,921	3,826	2	58%	66	0.2				X				X								X	
Bacteroidetes Sphingobacteriia	737,284	668	1	3%	20	0.0																	
Can. Div. RF3	172,673	214	2	17%	13	0.1																	
Can. Div. SM2F11	1,227,879	1,569	2	81%	19	2.3					?												
Can. Div. WS3	3,953,597	3,582	1	100%	37	1.5		X		X				X	X	X			X	X		X	
Chloroflexi	25,318,483	25,578	10	100%	46	0.1	X	X		X	?			X	X	X			X			X	
Cyanobacteria Oscillatoria	5,894,342	6,710	3	86%	553	1.3	X			X		X	X	X								X	
Cyanobacteria Phormidium	2,565,191	3,007	1	8%	1595	11.0	X			X		X	X	X						X			
Cyanobacteria Phormidium (short contigs)	3,966,550	6,351	1	64%	4884	33.7	X			X		X	X	X						X		X	
Cyanobacteria Oscillatoriales	7,802,043	9,123	6	100%	164	2.2	X			X		X	X	X	X							X	
Firmicutes Clostridia	4,792,235	4,744	4	94%	18	0.0			X	X					?				X			X	
Betaproteobacteria Burkholderiales	48,036,329	53,710	15	100%	67	0.4	X	X		X	X			X	X	X			X	X	X	X	X
Betaproteobacteria	32,854,738	37,034	15	100%	36	0.2	X	X		X	X			X	X	X			X		X	X	X
Deltaproteobacteria Desulfobacterales	14,987,426	16,532	4	97%	77	0.6			X	X				X	X	X		X				X	
Deltaproteobacteria Desulfovibrionales	2,359,490	2,441	1	31%	19	0.1			X	X				X	X	X		X					
Deltaproteobacteria Syntrophobacterales	136,929	156	0	0%	39	14.4																	
Deltaproteobacteria	28,366,627	27,476	10	100%	46	0.1			X	X				X				X		X		X	
Epsilonproteobacteria Campylobacterales	9,415,889	10,771	8	97%	35	0.4		X		X	X			X	?	X			X	X		X	X
Epsilonproteobacteria	470,829	564	1	19%	12	7.6		?															
Gammaproteobacteria Chromatiales	10,837,393	12,090	2	72%	61	0.7		X		X	X					X						X	
Gammaproteobacteria Methylococcales	4,968,049	5,677	2	42%	24	0.6		X		X				X		X	X					X	X
Gammaproteobacteria Thiotrichales	8,586,237	10,421	5	78%	124	0.3	X	X		X	X			X	X	X			X		X	X	
Proteobacteria	2,391,567	2,645	2	72%	43	0.0										X							
Spirochaetes Spirochaetales	25,556,253	26,605	10	100%	36	0.2		X	X	X				X	X				X			X	
Tenericutes Mollicutes	381,123	463	1	28%	15	0.0																	
Verrucomicrobia Opitutae	3,065,910	2,963	2	81%	17	0.0				X				X		X						X	
Verrucomicrobia	2,164,122	2,173	2	97%	20	0.3								X								X	
Eukaryote	41,816,827	65,030	13	92%	18	0.2				X				X	X								
Diatom Chloroplast	384,170	588	4	67%	728	37.7	X					X	X										
Unclassified	199,891,643	232,746	125				X	X	X	X	X	X		X	X	X	X		X		X	X	
	508,260,080	587,418	266																				

* These contigs are less than 4kb and were binned to the Phormidium based on BLASTp and gDNA coverage.

1 This belongs to the chloroplast of a Bacillariophyta Diatom.

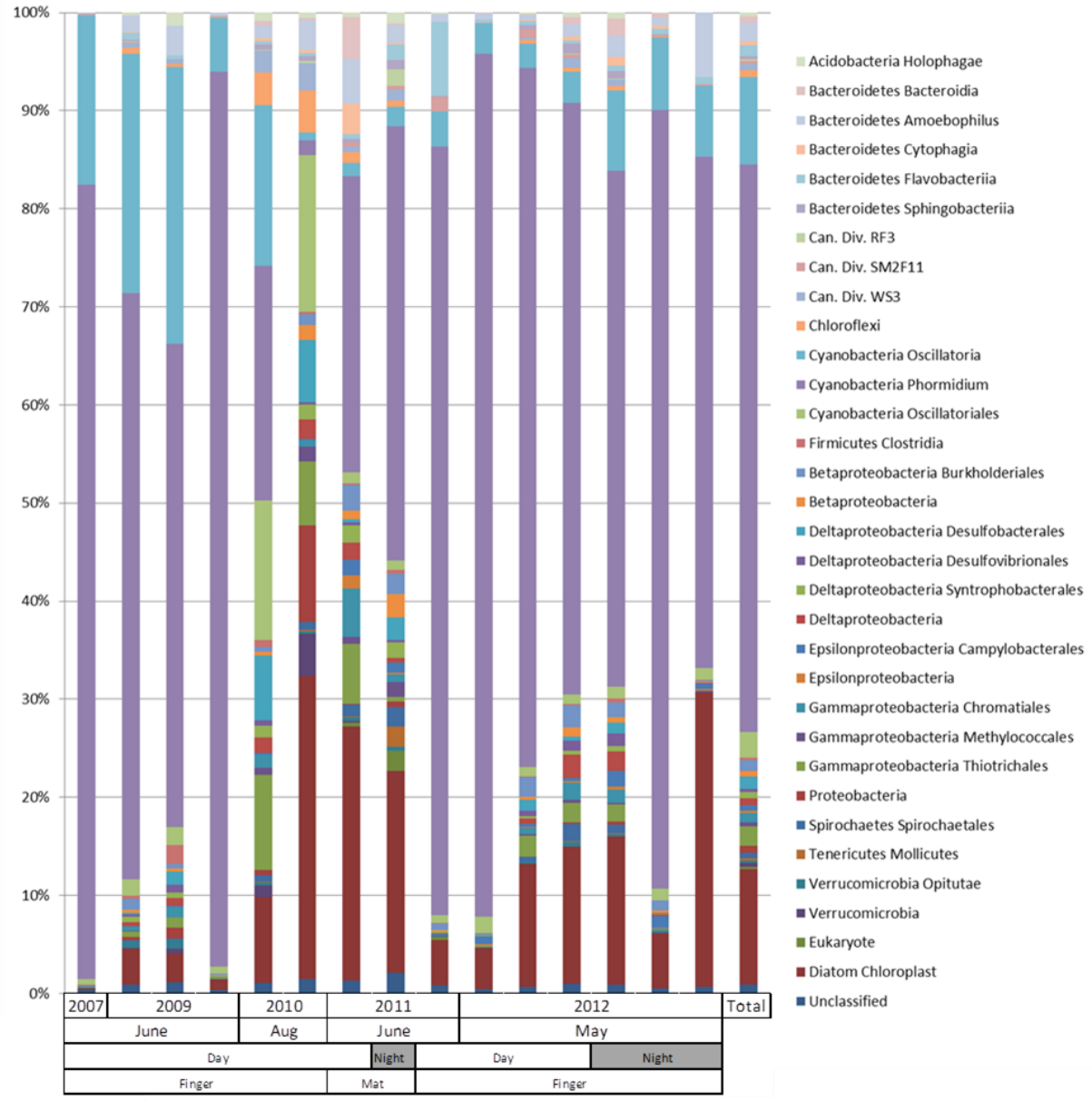


Figure 3.1 Abundance of each genomic bin by sample as a percentage of the total community. Finger samples were microbial mat structures raised off the sediment. Mat samples were in contact with the sediment.

BLASTp was used to identify potential metabolic pathways of interest in the genomic bins based on hits to NR, COG, KEGG and Pfam. The WS3 genome contains genes for autotrophy by the reverse TCA cycle as well as heterotrophy using branch chain ABC-type amino acid transporters, and citrate transporters. Genes for aerobic respiration by cytochrome C-

type, cytochrome bd-type and heme/copper-type cytochrome quinol oxidases were recovered along with a host of oxidative pathways, including those for H_2, CO, sulfur, sulfide, and sulfite oxidation (Table 3.1). These findings indicate that the MIS type WS3 is metabolically versatile and adapted to using O_2 as an electron acceptor, and able to utilize a wide range of electron donors. A single cell genome for a member of WS3 was recently recovered from soil (Rinke et al 2013), but it did not contain genes for electron transport or sulfur utilization genes that are present in the MIS WS3 metagenome, suggesting varying metabolic potential within the WS3 candidate division.

The data presented here is the first genomic sequence reported for candidate divisions SM2F11 and RF3, which have been previously identified only by 16s rRNA gene sequence (Castelle et al 2013, Tajima et al 1999, Tsertova et al 2011). The genome of SM2F11 is remarkably novel, with 812 of 1569 genes having no significant matches to the NR, COG, KEGG or Pfam databases, making identification of metabolic genes difficult. SM2F11 codes for a Rhodanese-related sulfurtransferase and coupled thioredoxin reductase, which can be used for intracellular thiosulfate oxidation, or other functions including sulfur detoxification (Ray et al 2000). It also has a predicted protein with low sequence similarity (23% amino acid sequence ID) to COG3256 (Nitric oxide reductase large subunit) which can be involved in denitrification. No identifiable metabolic genes were found in the RF3 genome (Table 3.1).

Analysis of genomic bins from proteobacteria identified organisms involved in sulfur cycling at MIS. A full set of dissimilatory sulfite reductase genes was identified within the bin for the *Desulfobacterales* (*Deltaproteobacteria*), indicating it is capable of sulfate reduction. Potential for oxidation of elemental sulfur using reverse dissimilatory sulfite reductase (*rdsrAB*) was detected in the *Burkholderiales* (*Betaproteobacteria*), *Campylobacterales*

(*Epsilonproteobacteria*) and *Methylococcales* (*Gammaproteobacteria*), and *soxABZ* genes for sulfur oxidation were identified in multiple members of the *Gammaproteobacteria* and *Epsilonproteobacteria*. Finally, methane oxidation genes *pmmoCAB* were identified within the *Methylococcales* genomic bin.

Whole-community transcriptomics

Metatranscriptomic sequencing was conducted on three samples collected at approximately 1pm (day) and three collected at approximately 1am (night). All of these samples were collected within an area with a radius of three meters, and the day/night samples were collected twelve hours apart. Overall, the majority of the 500 genes with the highest transcript abundance at MIS show relatively equal abundance of transcripts between day and night samples, with most organisms showing no significant difference in transcript abundance (Figure 3.2). However, several of the organisms expressing the most abundant genes show subtle trends in day versus night transcript abundance.

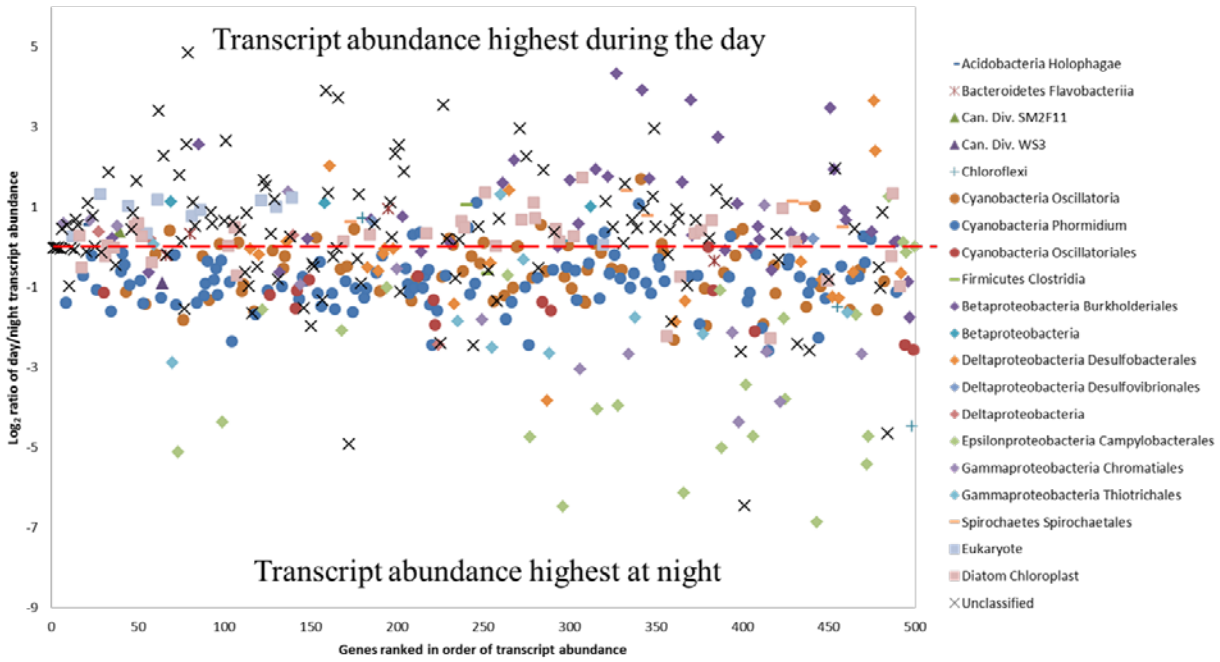


Figure 3.2 Rank abundance plot of the 500 most abundant mRNA transcripts at MIS. Genes with a positive Y axis value had higher transcript abundance during the day, while genes with negative values indicate higher transcript abundance at night. The red dashed line indicates equal abundance of transcripts in day/night samples. Transcript abundance is the average of three samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Cyanobacterial genes were found to dominate the 500 genes with highest transcript abundance, with many genes having a higher abundance of transcripts at night. *Phormidium* was responsible for 108 of the 500 most transcriptionally abundant genes, with 92% of those genes having higher transcript abundance at night. Diel regulation of gene expression is not uncommon amongst cyanobacteria studied previously, with as much as 80% of *Prochlorococcus* genes showing gene regulation based on light availability (Zinser et al 2009). *Oscillatoria* was responsible for 68 transcriptionally abundant genes, with 74% showing higher transcript levels at night. By contrast, the chloroplast belonging to a Bacillariophyta diatom showed higher day transcript abundance in 71% of the thirty eight genes shown here. The *Campylobacterales* (*Epsilonproteobacteria*) were responsible for twenty five of the most transcriptionally abundant genes, with 92% having more expression at night, consistent with these putative sulfur oxidizing

bacteria (Table 3.1) migrating vertically to track shoaling sulfide at night as has been observed previously in redox stratified microbial mat systems (Garcia-Pichel et al 1994). The *Burkholderiales* (*Betaproteobacteria*) showed higher transcript abundance in 80% of its forty highest expressed genes during the day. All five groups responsible for the majority of the top 500 transcriptionally abundant genes (279/500 genes) show higher transcript abundance in either day or night samples, indicating temporal fluctuations in gene expression play an important role at MIS.

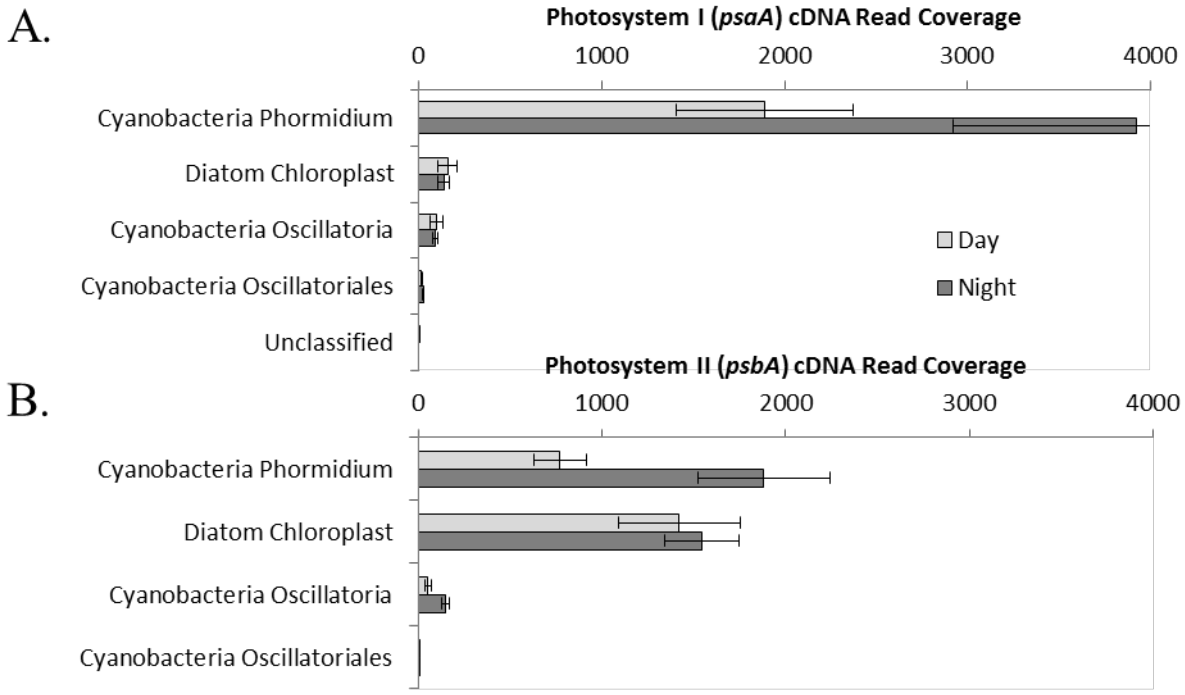


Figure 3.3 Transcript abundance of photosynthesis genes. (A) Transcript abundance of photosystem I reaction core gene *psaA*; (B) Transcript abundance of photosystem II reaction core gene *psbA*. Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Analysis of transcript data for genes encoding photosystem I *psaA* (PS I) and photosystem II *psbA* (PS II) show that *Phormidium* had significantly higher abundance of

transcripts at night than in day samples (Figure 3.3). In contrast, previous studies of marine cyanobacterial transcription have shown highest expression of photosynthesis reaction core genes (*psaA* and *psbA*) during the day, when photosynthesis takes place (Shi et al 2010, Stockel et al 2011, Zinser et al 2009). The diatom chloroplast showed no significant difference in transcript abundance of either PS I or PS II genes between day and night samples.

A large number of transcripts for *Phormidium* PS II genes were obtained (>2000x cDNA coverage), indicating that it is performing oxygenic photosynthesis at least some of the time. However, *Phormidium* had two times higher abundance of transcripts for PS I genes than PS II genes (Figure 3.3). In contrast, the diatom chloroplast expressed about six times more PS II genes than PS I genes, and a high ratio PS II to PS I normalized transcripts for *psbA* to *psaA* was also found in *Prochlorococcus*, a planktonic unicellular marine cyanobacteria (Zinser et al 2009). Assuming that the 6:1 ratio of transcripts for PS II to PS I genes is normal for oxygenic photosynthesis, we infer that the elevated level of PS I transcripts for *Phormidium* reflects transcriptional regulation of photosystem genes to down-regulate PS II genes and/or up-regulate PS I genes to conduct anoxygenic photosynthesis in the presence of sulfide. Although to our knowledge this is the first transcriptional data from anoxygenic cyanobacteria, it is consistent with the physiological shift towards PS II-independent anoxygenic photosynthesis that has been reported previously (Cohen et al 1986, Jorgensen et al 1986b), and with genes for anoxygenic photosynthesis being inducible and under the control of a transcriptional regulator (Bronstein et al 2000b).

The key enzyme for transfer of electrons from sulfide to PS I during anoxygenic photosynthesis by cyanobacteria is sulfide quinone oxidoreductase (SQR) (Bronstein et al 2000b). Although there are homologous genes of SQR in six different MIS genomic bins (Table

3.1), the *Phormidium* SQR had by far the highest abundance of transcripts in the mat community (Figure 3.4). The *Phormidium* SQR showed higher transcript abundance at night than during the day (Figure 3.4). A *Phormidium* gene for sulfite oxidation (*aprA*) also recruited transcripts, though at lower levels than SQR and in equal abundance between day and night. Previous work indicates that the oxidation product of anoxygenic photosynthesis by cyanobacteria is elemental sulfur (Garlick et al 1977a), the expected product of sulfide oxidation by SQR, thus *aprA* is not necessarily involved in anoxygenic photosynthesis.

Of the five sulfur oxidizing metabolisms investigated (Figure 3.4), thiosulfate oxidation using the *sox* genes showed the highest transcript abundance and was dominated by the *Epsilonproteobacteria, Gammaproteobacteria,* and *Betaproteobacteria*. *Betaproteobacteria* showed the highest sulfide oxidation transcript abundance using the reverse dissimilatory sulfite reductase pathway (*rdsrA/B*) while sulfide oxidation using flavocytochrome c sulfide dehydrogenase (fcc) observed significant transcriptional abundance from the *Betaproteobacteria* and *Gammaproteobacteria*.

Transcripts from *Desulfobacterales* (*Deltaproteobacteria*) for sulfur reduction using the dissimilarity sulfite reductase pathway (dsr) were highly abundant with over 90x total cDNA read coverage (Figure 3.5). Reduction of sulfate to elemental sulfur and hydrogen sulfide is well characterized amongst the *Desulfobacterales* and many Deltaproteobacteria, and sulfate reducing bacteria are believed to be major players in carbon cycling in sediments and other anaerobic environments (Muyzer and Stams 2008). The only other *dsr* genes that showed expression do not belong to a classified genomic bin, and together they have only one fifth of the transcript abundance of *Desulfobacterales* (Figure 3.5), indicating this deltaproteobacterium is responsible for the majority of sulfate reduction at MIS.

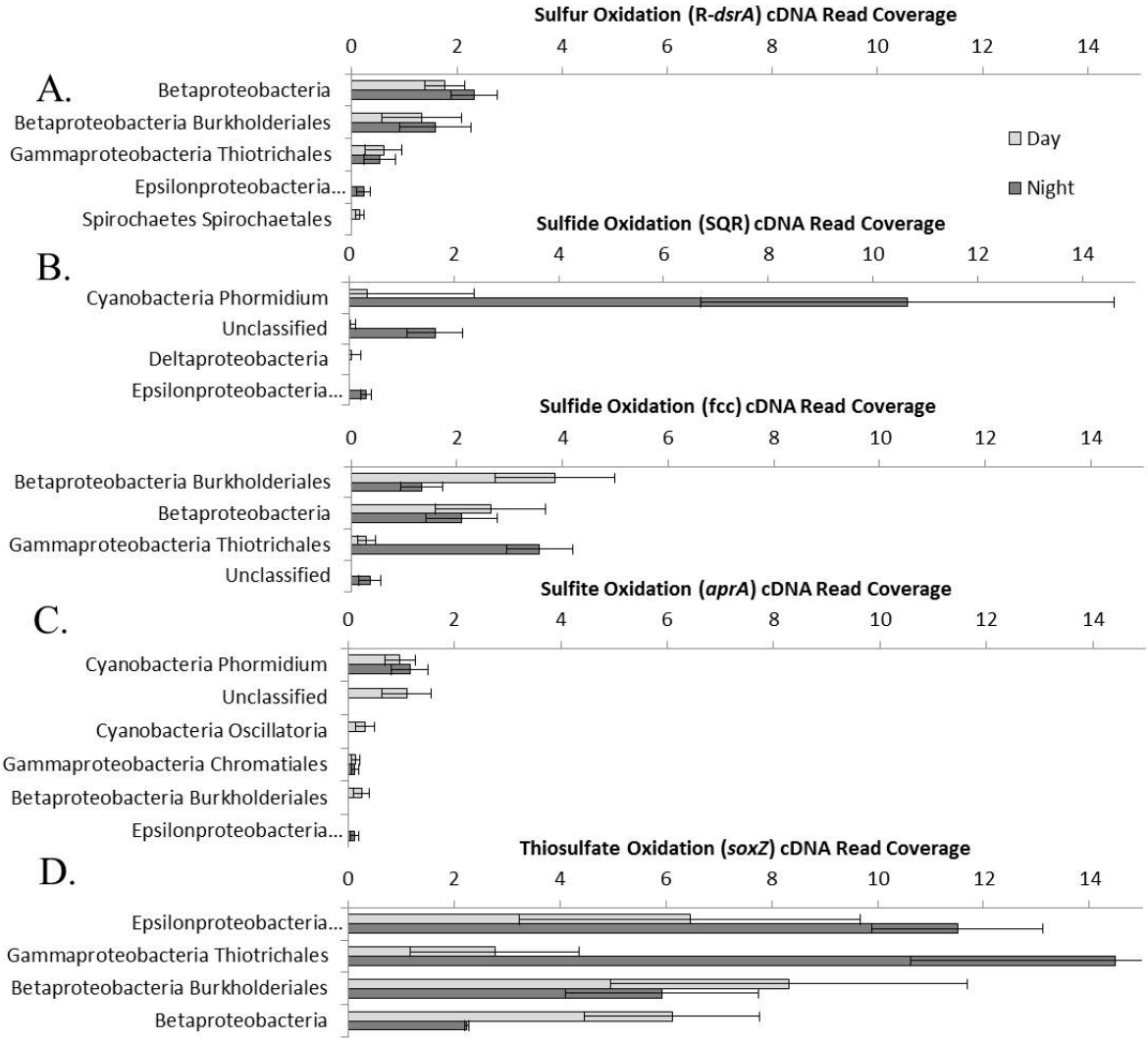


Figure 3.4 Transcript abundance of sulfur cycling genes. (A) Sulfur oxidation using R-*dsrA*; (B) sulfide oxidation using sulfide quinone oxidoreductase (SQR); (C) sulfide oxidation using flavocytochrome C sulfide dehydrogenase (fcc); (D) sulfite oxidation using *aprA*; (E) thiosulfate oxidation using *soxZ* (only bins with *soxABZ* included). Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

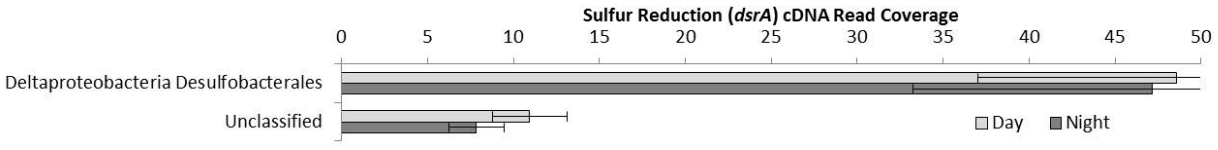


Figure 3.5 Transcript abundance of sulfur reduction using dissimilatory sulfite reductase (*dsrA*). Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated

as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Transcript abundance for three autotrophic pathways was measured to assess sources of primary production at MIS. Autotrophy using ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCo) for the Calvin cycle was measured, as was ATP citrate lyase (*aclB*) for the reverse tricarboxylic acid cycle, and CO dehydrogenase/acetyl-CoA synthase was measured for the Wood Ljungdahl pathway (Figure 3.6). RuBisCo showed by far the highest transcript abundance of any autotrophic pathway, but low overall expression of autotrophy genes could be due to input of organic carbon from higher in the water column (Nold et al 2013). The diatom chloroplast displayed the highest cDNA read coverage of RuBisCo, with no significant difference between day and night transcript abundance. Both *Phormidium* and *Thiotrichales* (*Gammaproteobacteria*) showed highest transcript abundance of RuBisCo at night, and had the second and third highest autotrophy contributions respectively. *Burkholderiales* (*Betaproteobacteria*) had the highest reverse TCA cycle transcript abundance and *Desulfobacterales* (*Deltaproteobacteria*) had the highest WLP transcript abundance, though neither showed significant difference in day/night expression.

Transcriptional evidence of H_2, CO, and methane oxidation as well as denitrification by *nirK* were all detected at reduced levels (< 15x gene coverage) compared to other metabolisms at MIS, and these metabolisms were restricted to the proteobacteria and chloroflexi (Figure 3.7). While a *nifH* homolog was found in the *Phormidium* genome, the rest of the pathway for nitrogen fixation was lacking, and significant numbers of transcripts for nitrogen fixation were not detected in any bin.

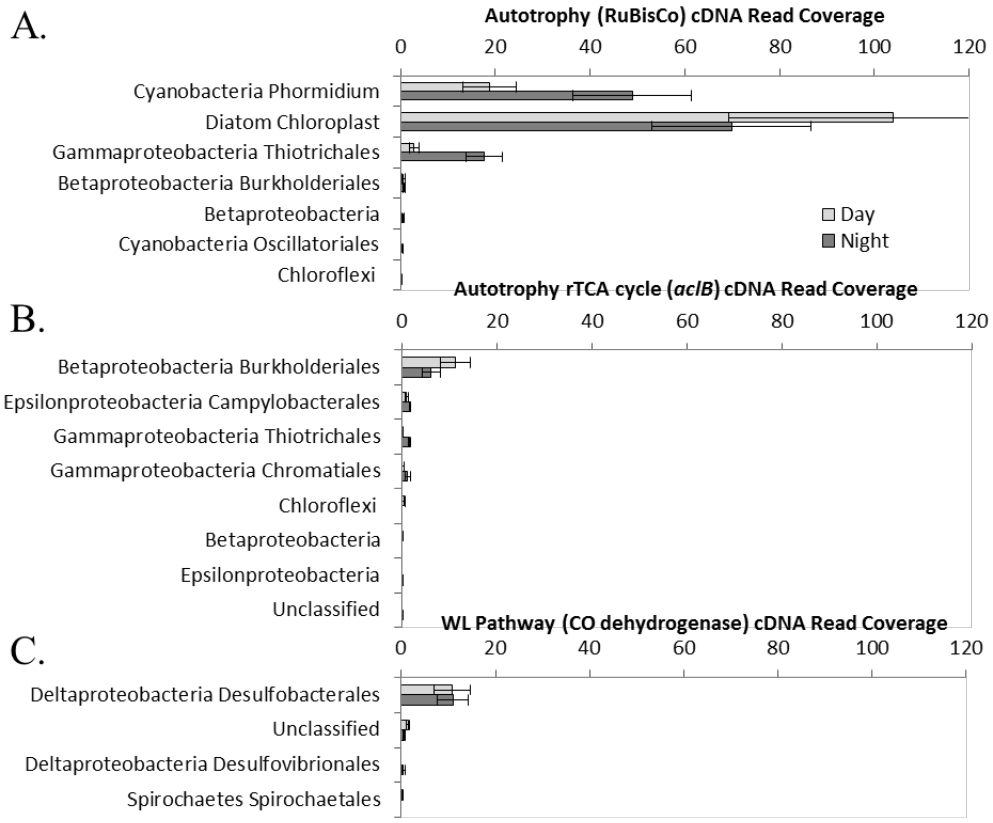


Figure 3.6 Transcript abundance of autotrophy genes. (A) RuBisCo; (B) reverse TCA cycle; (C) Wood Ljungdahl pathway. Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

A small number of transcripts were observed for WS3 genes related to H_2 oxidation (Figure 3.7), aerobic respiration and chemotaxis (Supplemental Table 3.3 in Appendix B). No expression of functional genes was found in SM2F11 or RF3, and both showed expression of relatively small amounts of rRNA and mRNA compared to the rest of the community (Supplemental Figure 3.2 in Appendix B).

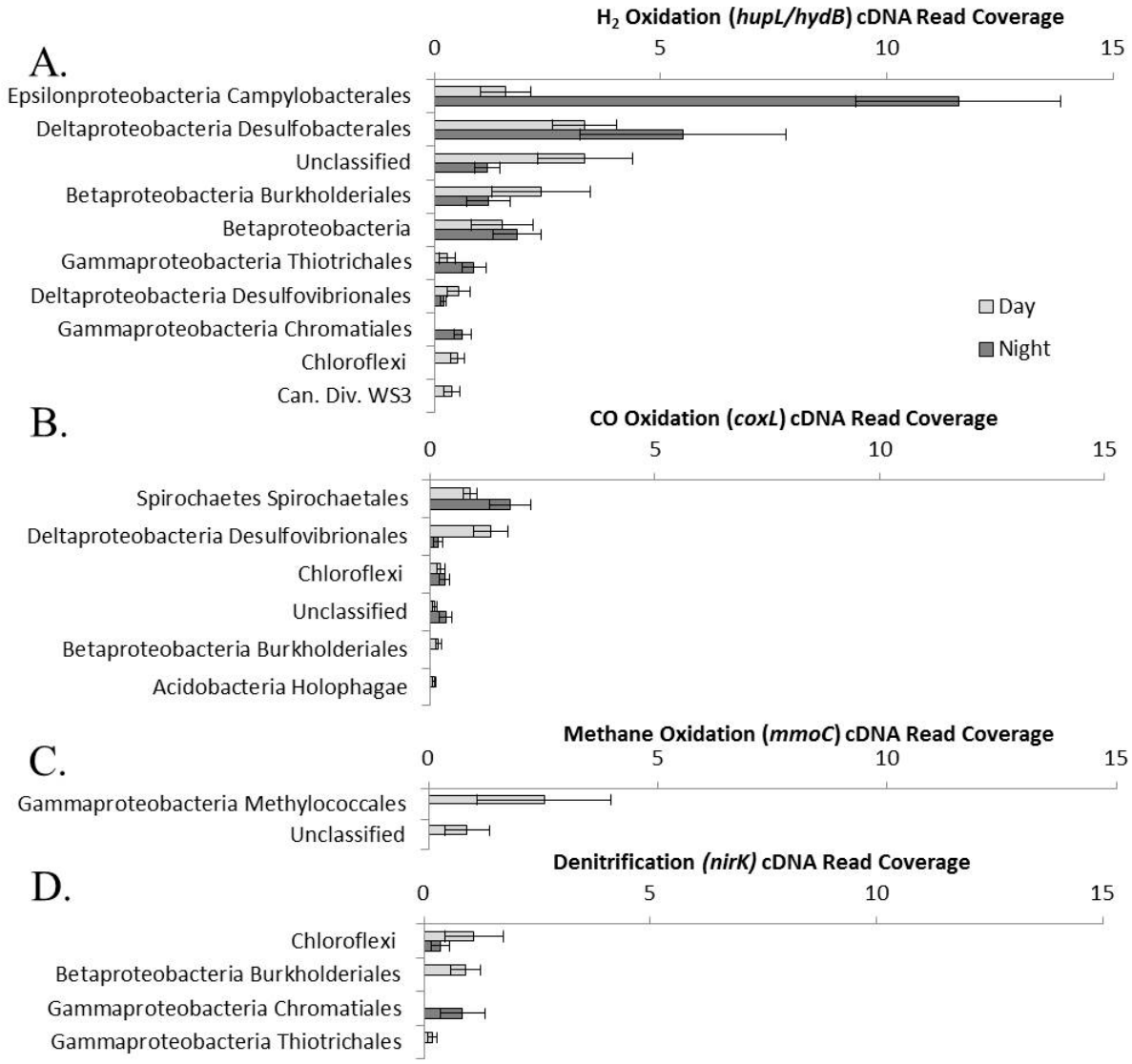


Figure 3.7 Transcript abundance for functional genes of interest. (A) Hydrogen oxidation using *hupL*/*hydB*; (B) carbon monoxide oxidation using *coxL*; (C) methane/ammonia oxidation using *mmoC*; (D) denitrification using *nirK*. Error bars based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Phormidium expression

Comparison of the average *Phormidium* transcript abundance for day and night samples shows that *Phormidium* had significantly higher transcript abundance of mRNA (rRNA removed) at

night, with twice the gene coverage as during the day (Figure 3.8A). Of the metabolic and functional genes examined, photosynthesis genes showed higher transcript abundance by orders of magnitude and totaled over 5000x average gene coverage of *psaA* and 300x average gene coverage of *psbA* (Figure 3.8). In contrast, other functional genes including those for autotrophy and oxygen respiration were not detected above 100x coverage (Figure 3.8B). Both *psaA* and *psbA* code for the reaction centers of their respective photosystem, which are known to denature at an enhanced rate compared to other proteins due to absorption of excess light energy from photosynthesis (Mattoo et al 1984, Soitamo et al 1996). This higher cellular demand for protein likely explains the higher abundance of transcripts observed for these genes.

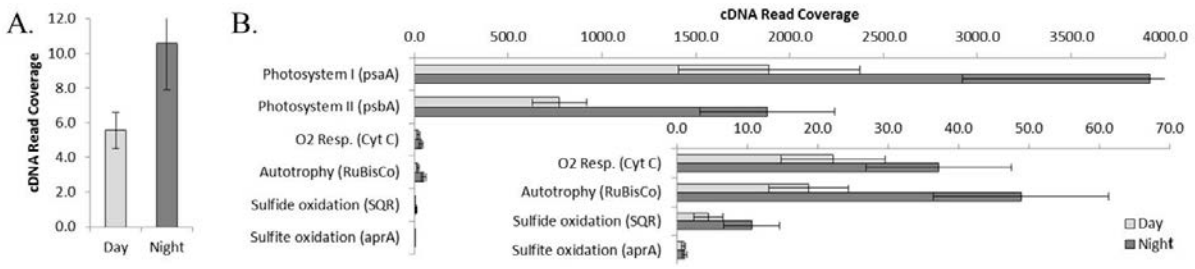


Figure 3.8 Transcript abundance of *Phormidium* genes. (A) Transcript abundance for the whole *Phormidium* genome; (B) *Phormidium* metabolic and functional genes. Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Phormidium photosynthesis genes consistently show significantly higher transcript abundance at night (Figure 3.9) with the exceptions of *psbL/U/P*, which are involved in energy transfer or the oxygen evolving complex of PS II. Of note is the fact that SQR had higher transcript abundance than PS I genes *psaL* and *psaX*, indicating that transcript abundance levels for SQR are on par with other *Phormidium* genes potentially involved in anoxygenic photosynthesis. The ability to do anoxygenic photosynthesis using SQR would be consistent

with other members of the Oscillatoriales, such as *Geitlerinema* sp. PCC9228 (formerly *Oscillatoria limnetica*) which has been shown to do anoxygenic photosynthesis in the presence of sulfide using SQR (Cohen et al 1975c).

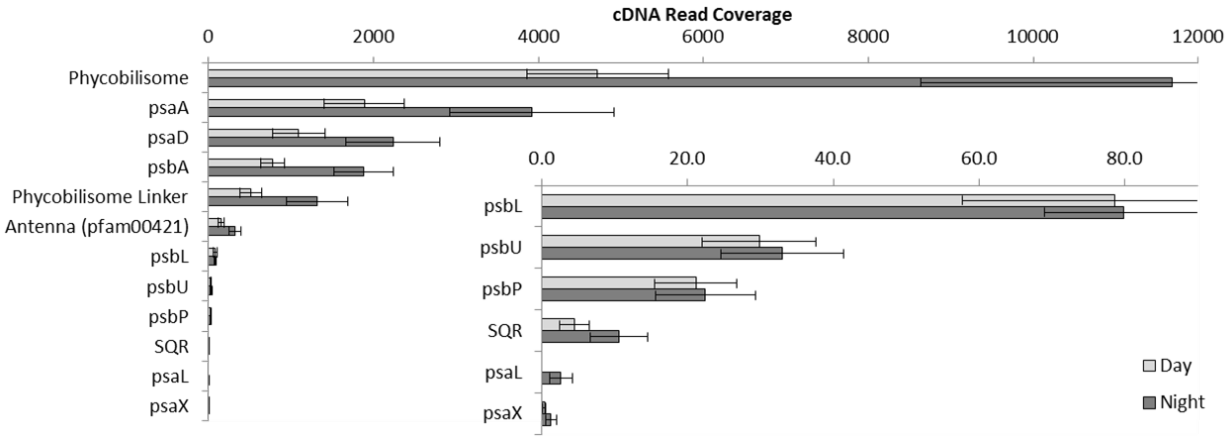


Figure 3.9 Transcript abundance of *Phormidium* photosynthesis genes. Error bars are based on standard deviation of normalized and averaged day and night samples. Coverage is calculated as the average transcript abundance of 3 samples and accounts for gene length in different organisms and varying sampling effort between samples. See *Methods* for a description of normalization procedures.

Overall, the metatranscriptomic data presented here is consistent with the *Phormidium* population being metabolically versatile, capable of both oxygenic and anoxygenic photosynthesis. It is not clear whether this phototrophic versatility in the *Phormidium* stems from niche adaptation amongst closely related ecotypes (e.g., differential activity of strains that are oxygenic and anoxygenic specialists) or true cellular versatility in which *Phormidium* cells switch pathways depending on sulfide concentration. Further, our bulk sampling of mats was not sensitive to the vertical gradients of sulfide concentration, so we are unable to evaluate potential vertical stratification of oxygenic/anoxygenic photosynthesis in the mat.

3.4 Conclusions

Culture independent analysis of microbial communities has greatly expanded our understanding of environments like MIS, where cultivation and isolation efforts have been unsuccessful. Further, investigating an entire community can highlight interactions between community members that are only present in a mixed assemblage of microbes. Here we report partial and near-complete genomes and metabolic gene expression for the most abundant organisms responsible for mediating microbial biogeochemistry in MIS microbial mats and surface sediments. Several of these organisms belong to novel bacterial phyla with no cultured representatives, and a partial genome for SM2F11 and a second genome for WS3 along with expression data constitute a significant advancement in our understanding of un-cultured and un-explored branches of the tree of life.

Analysis of metabolic and functional gene expression at MIS shows that photosynthesis by diatoms and cyanobacteria is the highest expressed metabolic pathway investigated, with *Phormidium* responsible for the majority of primary production by photosynthesis, sulfide/sulfite oxidation, and a significant portion of carbon fixation, highlighting the essential role *Phormidium* plays in ecosystem processes. Proteobacteria were largely responsible for microbially mediated sulfur cycling with the *Burkholderiales* (*Betaproteobacteria*), *Thiotrichales* and *Chromatiales* (*Gammaproteobacteria*), and *Epsilonproteobacteria* responsible for the majority of sulfur oxidation, and the *Desulfobacterales* (*Deltaproteobacteria*) responsible for the majority of sulfur reduction. The *Desulfobacterales* were also found to oxidize hydrogen, which appears to be coupled to sulfate reduction, and are likely the source of the hydrogen sulfide at MIS. While all three novel bacterial phyla showed some gene expression (Supplemental Figure 3.2 in Appendix B), WS3 transcripts for hydrogen oxidation, aerobic respiration and chemotaxis genes showed the only identifiable metabolic or functional gene

expression amongst these phyla. With more than half the gene content of both SM2F11 and RF3 unidentifiable by NR, KEGG, COG or Pfam, and the small portion of the RF3 genome recovered (~17%), lack of identifiable metabolic or functional genes is not surprising.

Based on expression ratios of PS II genes to PS I genes, *Phormidium* appears to be performing anoxygenic photosynthesis, and all lines of evidence indicate it is using SQR to obtain electrons from sulfide rather than using PS II to obtain electrons from water, consistent with other members of the Oscillatoriales (Bronstein et al 2000b, Cohen et al 1975c). We observe that *Phormidium* shows significantly higher transcript abundance at night, including genes for metabolic pathways such as photosynthesis, autotrophy and sulfur oxidation. This is in contrast to previous cyanobacterial transcriptomic analysis that shows higher transcript abundance of photosynthesis reaction center genes (Colon-Lopez and Sherman 1998, Shi et al 2010, Stockel et al 2011, Straub et al 2011, Zinser et al 2009) during the time of day when Phormidium's photosynthetic transcript abundance is lower. It is not clear if this difference is due to fundamental differences in life cycle, environmental or ecological factors, ability to do anoxygenic photosynthesis, or some other factor. Previous reports focused on unicellular cyanobacteria, which have not been shown to be capable of anoxygenic photosynthesis. Perhaps more importantly, these unicellular cyanobacteria typically undergo rapid cell division every 24 hours (Vijayan et al 2009, Zinser et al 2009), whereas *Phormidium* species are filamentous and much slower growing (0.07-0.5 d^{-1}, depending on light and nutrient availability) (Litchman 2000, Litchman et al 2003). While the driving factor for the difference in gene regulation is unclear (and likely a combination of factors), these results highlight the important differences between MIS benthic microbial mats and better studied cyanobacteria.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

We would like to thank NOAA and the Thunder Bay National Marine Sanctuary, Russ Green and the divers and crew of the *R/V Storm* for sampling and logistical assistance; Sunit Jain for assembly strategy and bioinformatics assistance; and the University of Michigan DNA sequencing core for DNA sequencing. This work was supported by NSF grant EAR1035955 to GJD, the University of Michigan CCMB Pilot Grant to JDC and GJD, and the Scott Turner Award to AAV.

Data Access

Sequences from this study are available from NCBI (http://www.ncbinlm.nih.gov/bioproject) under BioProject PRJNA72255. Reads from all 15 metagenomes and 6 metatranscriptomes are available on NCBI's Sequence Read Archive, and accession numbers can be found on Supplemental Table 3.4.

3.5 Appendix B

CHAPTER III Supplemental Material

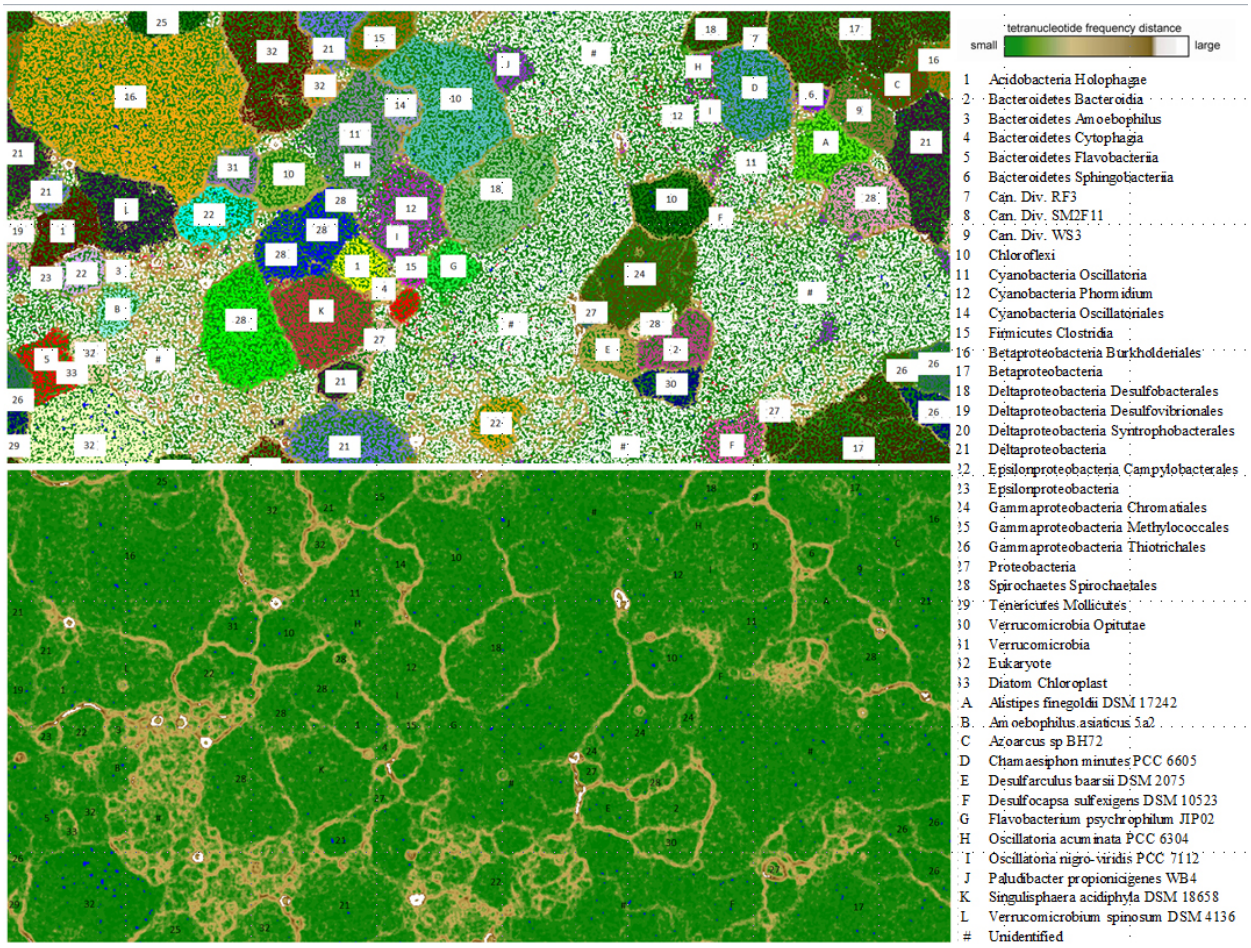
Supplemental Table 3.1 Metagenome sample summary

Sample	Collection Month	Latitude	Longitude	Finger[1]	Prostrate Mat[2]	Day	Night	Illumina gDNA Reads[3]	Illumina cDNA Reads[3]
2007-1D	June	45.1984	83.32721°	x		x		55,742,528	none
2009-1D	June	45.1984	83.32721°		x	x		56,610,020	none
2009-2D	June	45.1984	83.32721°	x		x		58,811,950	none
2009-3D	June	45.1984	83.32721°	x		x		25,542,490	none
2010-1D	Aug	45.1984	83.32721°	x		x		69,716,858	none
2010-2D	Aug	45.1984	83.32721°		x	x		67,190,598	none
2011-1D	June	45.1984	83.32721°		x	x		98,430,946	none
2011-2N	June	45.1984	83.32721°		x		x	211,878,358	none
2011-3D	June	45.1984	83.32721°	x		x		59,608,800	none
2012-1D	May	45.1984	83.32721°	x		x		25,666,608	11,997,534
2012-2D	May	45.1984	83.32721°	x		x		41,595,684	14,577,110
2012-3D	May	45.1984	83.32721°	x		x		41,318,538	11,524,300
2012-4N	May	45.1984	83.32721°	x			x	42,876,682	12,125,414
2012-5N	May	45.1984	83.32721°	x			x	38,335,454	12,100,438
2012-6N	May	45.1984	83.32721°	x			x	28,207,614	13,079,448

1 Microbial mat structure that is raised off the lake sediments

2 Microbial mat that lies flat on lake sediments

3 Numbers reflect de-replicated and trimmed read totals used for mapping



Supplemental Figure 3.1 Emergent self-organizing maps of tetranucleotide frequencies used to sort sequences into genomic bins. (A) map with 5kb sequences shown; (B) map with sequences removed showing green background representing similar tetranculeotide frequency, with brown lines distinguishing groups of like sequences. Numbers represent bins of sequences from this study; letters mark genomes available in public databases.

Phylum	Class	Order	Genus	16s Accession Number	Annotated Bin Accession Number	Contig	Taxon ID	Gene Length	%ID	Silva hit Lineage
obacteria	Holophagae	Holophagales	unclassified			scaffold_43476	MIS_100068986	240	99.17	Bacteria;Acidobacteria;Holophagae;Holophagales;Holophagaceae;uncultured;uncultured Acidobacteria bacterium
eroidetes	Bacteroidia	Bacteroidales	unclassified			scaffold_63470	MIS_5k.100387871	372	88.37	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;CR-115;uncultured bacterium
	Bacteroidia	Bacteroidales	unclassified			scaffold_34787	MIS_100000041	368	95.65	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Porphyromonadaceae;Paludibacter;uncultured bacterium
	Bacteroidia	Bacteroidales	unclassified			scaffold_13040	MIS_5k.100130411	376	92.06	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;Rs-E47 termite group;uncultured bacterium
	Bacteroidia	Bacteroidales	unclassified			scaffold_1687	MIS_5k.100016882 7	599	90.85	Bacteria;Bacteroidetes;Bacteroidia;Bacteroidales;S24-7;uncultured bacterium
	Cytophagia	Cytophagales	Amoebophilus			scaffold_96302	MIS_100562933	1481	98.25	Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Flammeovirgaceae;Candidatus Amoebophilus;uncultured Bacteroidetes bacterium
	Cytophagia	Cytophagales	unclassified			scaffold_16286	MIS_5k.100162873	1497	85.01	Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Cyclobacteriaceae;uncultured;uncultured bacterium
	Cytophagia	Cytophagales	unclassified			scaffold_166776	MIS_101264242	973	89.89	Bacteria;Bacteroidetes;Cytophagia;Cytophagales;Cytophagaceae;Flexibacter;Flexibacter ruber
	Flavobacteriia	Flavobacteriales	unclassified							
	Sphingobacteriia	Sphingobacteriales	unclassified							
Div. RF3	unclassified	unclassified	unclassified			scaffold_76773	MIS_100372431	387	93.28	Bacteria;RF3;uncultured bacterium
Div. SM2-F11	unclassified	unclassified	unclassified			scaffold_5199	MIS_5k.100052003	1355	85.28	Bacteria;SM2F11;uncultured bacterium
Div. WS3	unclassified	unclassified	unclassified			scaffold_959	MIS_5k.100009602 1	983	97.97	Bacteria;Candidate division WS3;uncultured bacterium
roflexi	Chloroflexi	Chloroflexales	unclassified			scaffold_4221	MIS_5k.100042223	1480	92.91	Bacteria;Chloroflexi;Chloroflexia;Chloroflexales;Roseiflexaceae;Roseiflexus;uncultured bacterium
nobacteria	unclassified	Oscillatoriales	Oscillatoria							
		Oscillatoriales	Phormidium							
		Oscillatoriales	Phormidium*							
		Oscillatoriales	unclassified							
micutes	Clostridia	Clostridiales	unclassified			scaffold_65397	MIS_100266735	1471	93.95	Bacteria;Firmicutes;Clostridia;Clostridiales;Lachnospiraceae;uncultured;uncultured bacterium
eobacteria	Betaproteobacteria	Burkholderiales	unclassified			scaffold_10470	MIS_5k.100104711	340	100	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Hydrogenophaga;uncultured bacterium
		unclassified	unclassified			scaffold_3869	MIS_5k.100038702 4	306	99.02	Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Gallionellaceae;Sideroxydans;uncultured bacterium
						scaffold_39012	MIS_5k.100357601	363	98.07	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales;Rhodocyclaceae;Sulfuritalea;uncultured Rhodocyclaceae bacterium
	Deltaproteobacteria	Desulfobacterales	unclassified							
		Desulfovibrionales	unclassified			scaffold_10696	MIS_5k.100106974	1542	99.68	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfomicrobiaceae;Desulfomicrobium;Desulfomicrobium baculatum DSM 4028
						scaffold_41448	MIS_100032201	1515	99.54	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio;uncultured bacterium
		Syntrophobacterales	Smithella			scaffold_21895	MIS_5k.100218963	1523	98.23	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophaceae;Smithella;uncultured delta proteobacterium
		unclassified	unclassified							
	Epsilonproteobacteria	Campylobacterales	unclassified			scaffold_59101	MIS_5k.100384531	1457	97.7	Bacteria;Proteobacteria;Epsilonproteobacteria;Campylobacterales;Campylobacteraceae;Arcobacter;uncultured bacterium
		unclassified	unclassified							
	Gammaproteobacteria	Chromatiales	unclassified							
		Methylococcales	unclassified							
		Thiotrichales	unclassified							
	unclassified	unclassified	unclassified							
ochaetes	Spirochaetales	Spirochaetales	unclassified			scaffold_41090	MIS_5k.100361693	1508	99.14	Bacteria;Spirochaetae;Spirochaetes;Spirochaetales;Leptospiraceae;uncultured;uncultured bacterium
eriicutes	Mollicutes	Acholeplasmatales	unclassified			scaffold_25923	MIS_5k.100239242	1477	99.38	Bacteria;Tenericutes;Mollicutes;Acholeplasmatales;Acholeplasmataceae;Acholeplasma;uncultured bacterium
rucomicrobia	Opitutae	unclassified	unclassified			scaffold_33224	MIS_5k.100332255	397	92.23	Bacteria;Verrucomicrobia;Class Incertae Sedis;Order Incertae Sedis;Family Incertae Sedis;Candidatus Methylacidiphilum;uncultured bacterium
	unclassified	unclassified	unclassified			scaffold_15623	MIS_5k.100156248	376	98.4	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae;uncultured;uncultured bacterium
assified	unclassified	unclassified	unclassified							
llariophyta[1]	Coscinodiscophyceae	unclassified	Thalassiosiraceae							
assified	unclassified	unclassified	unclassified							

based on BLASTp and gDNA coverage.

Classification	Length	2012-1D	2012-2D	2012-3D	2012-4N	2012-5N	2012-6N	Total cDNA Coverage	
Acidobacteria Holophagae	8803791	0.01	0.03	0.02	0.02	0.02	0.01	0.11	x
Bacteroidetes Bacteroidia	2042926	2.10	0.41	0.61	2.33	0.90	0.26	6.61	x
Bacteroidetes Amoebophilus	517169	0.77	1.30	0.59	0.82	0.73	0.78	4.98	x
Bacteroidetes Cytophagia	443445	0.01	0.00	0.01	0.01	0.02	0.00	0.05	x
Bacteroidetes Flavobacteriia	3350921	0.07	0.02	0.01	0.02	0.03	0.05	0.20	x
Bacteroidetes Sphingobacteriia	737284	0.00	0.00	0.00	0.00	0.00	0.00	0.01	x
Can. Div. RF3	172673	0.01	0.03	0.02	0.00	0.01	0.01	0.07	x
Can. Div. SM2F11	1227879	0.52	0.51	0.37	0.12	0.11	0.63	2.25	x
Can. Div. WS3	3953597	0.05	0.15	0.35	0.58	0.06	0.29	1.48	x
Chloroflexi	25318483	0.01	0.01	0.01	0.02	0.02	0.02	0.09	x
Cyanobacteria Oscillatoria	5894342	0.34	0.17	0.03	0.19	0.39	0.16	1.29	x
Cyanobacteria Phormidium	6531741	4.39	2.69	2.04	1.24	5.97	8.49	24.81	x
Cyanobacteria Oscillatoriales	7802043	0.46	0.82	0.18	0.07	0.13	0.53	2.20	x
Firmicutes Clostridia	4792235	0.00	0.01	0.00	0.01	0.01	0.01	0.04	x
Betaproteobacteria Burkholderiales	48036329	0.14	0.06	0.02	0.04	0.09	0.05	0.40	x
Betaproteobacteria	32854738	0.04	0.03	0.02	0.02	0.04	0.02	0.17	x
Deltaproteobacteria Desulfobacterales	14987426	0.04	0.24	0.08	0.18	0.02	0.05	0.62	x
Deltaproteobacteria Desulfovibrionales	2359490	0.04	0.03	0.02	0.02	0.03	0.01	0.14	x
Deltaproteobacteria Syntrophobacterales	136929	0.74	5.01	2.34	4.05	0.81	1.43	14.38	x
Deltaproteobacteria	28366627	0.02	0.01	0.03	0.03	0.03	0.01	0.14	x
Epsilonproteobacteria Campylobacterales	9415889	0.05	0.01	0.01	0.19	0.07	0.10	0.42	x
Epsilonproteobacteria	470829	1.22	0.74	1.24	1.94	1.61	0.89	7.65	x
Gammaproteobacteria Chromatiales	10837393	0.07	0.14	0.09	0.18	0.16	0.06	0.70	x
Gammaproteobacteria Methylococcales	4968049	0.03	0.22	0.26	0.06	0.06	0.03	0.65	x
Gammaproteobacteria Thiotrichales	8586237	0.01	0.02	0.06	0.05	0.01	0.16	0.30	x
Proteobacteria	2391567	0.00	0.00	0.00	0.00	0.00	0.00	0.01	x
Spirochaetes Spirochaetales	25556253	0.05	0.02	0.02	0.02	0.05	0.02	0.17	x
Tenericutes Mollicutes	381123	0.00	0.00	0.00	0.00	0.00	0.00	0.01	x
Verrucomicrobia Opitutae	3065910	0.00	0.00	0.00	0.00	0.00	0.00	0.01	x
Verrucomicrobia	2164122	0.03	0.07	0.10	0.06	0.05	0.03	0.33	x
Eukaryote	41816827	0.04	0.01	0.08	0.02	0.04	0.04	0.23	x
Diatom Chloroplast	384170	7.14	7.64	5.91	5.31	7.19	4.53	37.72	x
Unclassified	832617639	0.50	0.42	0.37	0.51	0.36	0.42	2.58	x

Supplemental Figure 3.2 Transcript coverage of contigs for each bin. Green indicates the highest transcript coverage, yellow the median and red no transcript coverage. Transcript abundance normalized for total length of unique sequence per bin, not normalized for sequencing effort.

Supplemental Table 3.3 Bacterial Candidate Division expression

WS3 Expression			Raw number of transcripts							
Locus_Tag	gene length	Annotation	2012-1D	2012-2D	2012-3D	2012-4N	2012-5N	2012-6N	total	Coverage
MIS_5k.100000481	404	23S rRNA. Bacterial LSU	1545	4845	12346	21387	2040	10425	52588.0	13016.8
MIS_5k.1000096021	971	16S rRNA. Bacterial SSU	210	985	550	426	137	269	2577.0	265.4
MIS_5k.1000004765	1613	Flagellar hook-length control protein	21	136	396	619	48	387	1607.0	99.6
MIS_5k.1000012627	203	tRNA_Arg_CCG	10	4	8	8	4	1	35.0	17.2
MIS_5k.100007384	638	TonB-dependent Receptor Plug Domain	4	0	2	3	8	9	26.0	4.1
MIS_5k.100000482	116	5S rRNA. Bacterial TSU	0	0	2	1	0	0	3.0	2.6
MIS_5k.100005152	506	Single-stranded DNA-binding protein	0	0	0	7	4	0	11.0	2.2
MIS_5k.1000775516	470	Ribosomal protein S7	0	0	4	0	0	2	6.0	1.3
MIS_5k.1000201714	314	Protein of unknown function (DUF1180)	0	0	0	2	0	2	4.0	1.3
MIS_5k.100008165	3041	Unknown Function	10	12	8	0	0	6	36.0	1.2
MIS_5k.1000111919	377	Predicted DNA-binding proteins - COG1342	0	4	0	0	0	0	4.0	1.1
MIS_5k.1000164614	386	Ribosomal protein S9	2	0	2	0	0	0	4.0	1.0
MIS_100176751	215	Flagellin and related hook-associated proteins	0	0	2	0	0	0	2.0	0.9
MIS_5k.1000125426	701	Unknown Function	0	0	0	0	6	0	6.0	0.9
MIS_5k.100000996	269	Ribosomal protein S15P/S13E	0	0	0	0	2	0	2.0	0.7
MIS_5k.100060992	1199	Ornithine/acetylornithine aminotransferase	0	0	0	0	0	8	8.0	0.7
MIS_5k.100032788	611	Unknown Function	0	0	0	0	2	2	4.0	0.7
MIS_5k.1000173613	308	1,4-alpha-glucan branching enzyme	0	0	0	2	0	0	2.0	0.6
MIS_5k.100158831	308	NADH:ubiquinone oxidoreductase 24 kD subunit	0	0	0	0	2	0	2.0	0.6
MIS_5k.100007386	644	Unknown Function	0	4	0	0	0	0	4.0	0.6
MIS_5k.1000055076	674	Flagellin and related hook-associated proteins	0	0	0	0	0	4	4.0	0.6
MIS_100291774	677	Ni,Fe-hydrogenase maturation factor	0	4	0	0	0	0	4.0	0.6
MIS_5k.1000052116	689	Deoxynucleoside kinases	0	0	0	0	0	4	4.0	0.6
MIS_5k.100112016	1463	Cytochrome c bacterial	0	0	0	0	8	0	8.0	0.5
MIS_5k.1000775517	920	Translation elongation factors (GTPases)	3	0	2	0	0	0	5.0	0.5
MIS_5k.1000045342	197	Ribosomal protein L29	0	0	1	0	0	0	1.0	0.5
MIS_100296862	1202	Phosphate-selective porin O and P	0	3	3	0	0	0	6.0	0.5
MIS_5k.1000045323	401	Ribosomal protein S11	0	0	0	2	0	0	2.0	0.5
MIS_5k.100012642	410	Unknown Function	0	0	2	0	0	0	2.0	0.5
MIS_5k.1000045343	413	Ribosomal protein L16/L10E	0	0	2	0	0	0	2.0	0.5
MIS_5k.100021533	1298	Major Facilitator Superfamily	0	6	0	0	0	0	6.0	0.5
MIS_100016422	656	Unknown Function	3	0	0	0	0	0	3.0	0.5
MIS_5k.1000385212	941	Glycine cleavage system H protein (lipoate-binding)	0	0	0	0	4	0	4.0	0.4
MIS_5k.1000004844	980	Acetyl-CoA carboxylase alpha subunit	0	0	0	0	0	4	4.0	0.4
MIS_5k.100096404	986	Pseudouridine synthase	0	0	4	0	0	0	4.0	0.4
MIS_5k.100004534	1001	Nucleoside-diphosphate-sugar epimerases	0	0	4	0	0	0	4.0	0.4
MIS_5k.100051924	503	Formate hydrogenlyase subunit 6	0	0	0	0	2	0	2.0	0.4

MIS_100291772	1793	Ni,Fe-hydrogenase I large subunit	1	6	0	0	0	0	7.0	0.4
MIS_5k.1000045347	830	Ribosomal protein L2	3	0	0	0	0	0	3.0	0.4
MIS_5k.1000079520	1697	Unknown Function	0	0	6	0	0	0	6.0	0.4
MIS_5k.1000155612	572	Unknown Function	0	0	0	0	0	2	2.0	0.3
MIS_5k.1000005491	1145	S-adenosylmethionine synthetase	0	4	0	0	0	0	4.0	0.3
MIS_5k.100007385	581	Unknown Function	0	0	2	0	0	0	2.0	0.3
MIS_5k.1000055023	584	Protein of unknown function (DUF3359)	0	0	0	0	2	0	2.0	0.3
MIS_5k.1000388715	1187	Glycosyltransferase	0	0	4	0	0	0	4.0	0.3
MIS_5k.1000073841	599	Uncharacterized membrane protein	2	0	0	0	0	0	2.0	0.3
MIS_5k.100021538	1220	Aspartate/tyrosine/aromatic aminotransferase	0	0	0	0	0	4	4.0	0.3
MIS_5k.10000031112	1241	AAA domain (Cdc48 subfamily)	4	0	0	0	0	0	4.0	0.3
MIS_5k.100044144	674	Uncharacterized conserved protein, COG4359	0	0	2	0	0	0	2.0	0.3
MIS_5k.100012545	1358	Glucose/sorbosone dehydrogenases	0	0	0	0	0	4	4.0	0.3
MIS_5k.100041885	683	Riboflavin synthase alpha chain	0	0	2	0	0	0	2.0	0.3
MIS_5k.100158832	689	AT-rich DNA-binding protein	0	0	0	0	2	0	2.0	0.3
MIS_5k.100202267	1412	F0F1-type ATP synthase, beta subunit	0	4	0	0	0	0	4.0	0.3
MIS_5k.1000045324	368	Ribosomal protein S13	0	0	0	0	0	1	1.0	0.3
MIS_5k.100202265	1505	F0F1-type ATP synthase, alpha subunit	0	0	0	0	0	4	4.0	0.3
MIS_5k.1000440111	776	Ribosomal protein L25 (general stress protein Ctc)	0	0	0	0	0	2	2.0	0.3
MIS_5k.1000201711	398	Predicted transcriptional regulator - COG1959	0	1	0	0	0	0	1.0	0.3
MIS_5k.100015009	1214	Uncharacterized conserved protein - COG2966	0	0	2	0	1	0	3.0	0.2
MIS_5k.1000020150	1235	Unknown Function	3	0	0	0	0	0	3.0	0.2
MIS_5k.1000547312	1247	Glycine/serine hydroxymethyltransferase	0	0	0	0	3	0	3.0	0.2
MIS_5k.1000012630	1289	Glutamate dehydrogenase/leucine dehydrogenase	3	0	0	0	0	0	3.0	0.2
MIS_5k.1000004817	1727	NADH:ubiquinone oxidoreductase 49 kD subunit 7	4	0	0	0	0	0	4.0	0.2
MIS_5k.1000684319	902	rRNA methylase	0	0	0	2	0	0	2.0	0.2
MIS_5k.100008333	2711	Predicted Zn-dependent peptidases - COG0612	0	0	3	0	3	0	6.0	0.2
MIS_5k.1000005446	905	Ketopantoate hydroxymethyltransferase	0	0	2	0	0	0	2.0	0.2
MIS_5k.1000004720	1364	Uncharacterized conserved protein - COG1354	3	0	0	0	0	0	3.0	0.2
MIS_5k.1000057530	2771	Methylase of chemotaxis methyl-accepting proteins	3	3	0	0	0	0	6.0	0.2
MIS_5k.1000009931	932	Unknown Function	0	2	0	0	0	0	2.0	0.2
MIS_5k.1000079116	1913	Deoxyxylulose-5-phosphate synthase	0	0	0	4	0	0	4.0	0.2
MIS_5k.1000012631	2960	Phosphoenolpyruvate synthase	0	0	6	0	0	0	6.0	0.2
MIS_5k.1000020118	998	Unknown Function	0	0	0	0	0	2	2.0	0.2
MIS_5k.1000045321	998	DNA-directed RNA polymerase, alpha subunit	2	0	0	0	0	0	2.0	0.2
MIS_5k.1000440115	1001	Predicted membrane protein (DUF2232)	0	0	0	2	0	0	2.0	0.2
MIS_5k.100037096	1544	Repeat of unknown function (DUF346)	0	0	3	0	0	0	3.0	0.2
MIS_5k.1000370913	1046	Predicted nucleoside-diphosphate sugar epimerases	0	2	0	0	0	0	2.0	0.2
MIS_5k.1000445011	2141	Cell division GTPase	0	0	0	4	0	0	4.0	0.2
MIS_5k.100005055	2183	Alpha-glucosidases, family 31 of glycosyl hydrolases	4	0	0	0	0	0	4.0	0.2
MIS_5k.1000020127	1142	Gram-negative bacterial tonB protein	0	0	0	0	2	0	2.0	0.2
MIS_5k.1000083316	2297	Putative Ser protein kinase	0	0	0	0	0	4	4.0	0.2
MIS_5k.100009855	2309	Unknown Function	0	0	4	0	0	0	4.0	0.2
MIS_5k.1000045352	1190	GTPases - translation elongation factors	2	0	0	0	0	0	2.0	0.2
MIS_5k.100170552	1202	Predicted signal transduction protein - COG1639	0	2	0	0	0	0	2.0	0.2

MIS_5k.1000215320	611	WD40-like Beta Propeller Repeat	0	0	0	0	1	0	1.0	0.2
MIS_5k.1000025513	1835	Bacterial Ig-like domain (DUF1927)	0	2	0	1	0	0	3.0	0.2
MIS_5k.1000164620	1844	Single-stranded DNA-specific exonuclease	0	0	0	0	0	3	3.0	0.2
MIS_5k.1001121711	1280	Unknown Function	0	0	0	2	0	0	2.0	0.2
MIS_5k.1000025514	1286	Unknown Function	0	0	0	2	0	0	2.0	0.2
MIS_5k.100134091	2699	Uncharacterized conserved protein	0	0	0	0	0	4	4.0	0.1
MIS_5k.1000009920	1367	Glycine cleavage system protein P (pyridoxal-binding)	2	0	0	0	0	0	2.0	0.1
MIS_5k.1000091120	1445	Glutamine phosphoribosylpyrophosphate amidotransferase	0	0	0	0	0	2	2.0	0.1
MIS_5k.100087906	2981	Signal transduction histidine kinase	0	0	0	0	4	0	4.0	0.1
MIS_5k.100096408	1493	Transcription elongation factor	0	0	2	0	0	0	2.0	0.1
MIS_100018691	1505	FOG: PKD repeat	0	2	0	0	0	0	2.0	0.1
MIS_5k.1000012632	1550	NAD-dependent aldehyde dehydrogenases	0	0	0	0	0	2	2.0	0.1
MIS_5k.1000091122	1571	Ribonucleases G and E	0	0	0	0	0	2	2.0	0.1
MIS_5k.1000017234	1616	Preprotein translocase subunit SecD	0	0	0	2	0	0	2.0	0.1
MIS_5k.100000548	1622	Acyl-CoA dehydrogenases	0	2	0	0	0	0	2.0	0.1
MIS_5k.1000079125	818	Sugar transferases involved in lipopolysaccharide synthesis	0	0	0	1	0	0	1.0	0.1
MIS_5k.1000012757	1664	Unknown Function	0	0	0	0	0	2	2.0	0.1
MIS_5k.100021537	1664	Phosphoenolpyruvate carboxykinase (ATP)	0	0	0	0	2	0	2.0	0.1
MIS_5k.100008167	2501	Unknown Function	0	3	0	0	0	0	3.0	0.1
MIS_5k.100012691	2546	Unknown Function	0	0	0	0	3	0	3.0	0.1
MIS_5k.100087907	1712	Unknown Function	0	0	0	0	0	2	2.0	0.1
MIS_5k.1000003162	1769	Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase	0	0	2	0	0	0	2.0	0.1
MIS_5k.100173795	1880	Predicted symporter	0	0	2	0	0	0	2.0	0.1
MIS_5k.100183332	3860	Cellobiose phosphorylase	0	0	0	0	0	4	4.0	0.1
MIS_5k.1000025516	1952	Gamma-glutamyltransferase	0	0	2	0	0	0	2.0	0.1
MIS_5k.100322755	1961	Methyl-accepting chemotaxis protein	0	2	0	0	0	0	2.0	0.1
MIS_5k.1000007041	1007	Unknown Function	0	0	0	0	1	0	1.0	0.1
MIS_5k.1000109117	2057	Urocanate hydratase	2	0	0	0	0	0	2.0	0.1
MIS_5k.1000201715	3110	Sigma-54 interaction domain	0	0	3	0	0	0	3.0	0.1
MIS_5k.100012645	1058	ABC-type multidrug transport system, ATPase component	1	0	0	0	0	0	1.0	0.1
MIS_5k.100279953	2126	Response regulators consisting of a CheY-like domain	2	0	0	0	0	0	2.0	0.1
MIS_5k.100017383	2150	Polyphosphate kinase	0	0	2	0	0	0	2.0	0.1
MIS_5k.10000054104	1088	Lauroyl/myristoyl acyltransferase	0	0	0	0	0	1	1.0	0.1
MIS_5k.100008164	2183	Glycyl-tRNA synthetase, beta subunit	0	0	0	0	0	2	2.0	0.1
MIS_5k.1000061510	2267	Glycosidases	2	0	0	0	0	0	2.0	0.1
MIS_5k.1000012720	1139	DNA polymerase sliding clamp subunit (PCNA homolog)	1	0	0	0	0	0	1.0	0.1
MIS_5k.1000852610	2345	Transketolase	0	0	0	2	0	0	2.0	0.1
MIS_5k.100065842	2375	Unknown Function	0	2	0	0	0	0	2.0	0.1
MIS_5k.100033692	4757	Regulatory P domain	0	0	0	0	4	0	4.0	0.1
MIS_5k.1000007026	2399	3-hydroxyacyl-CoA dehydrogenase	0	2	0	0	0	0	2.0	0.1
MIS_5k.100335782	2411	Unknown Function	0	0	2	0	0	0	2.0	0.1
MIS_5k.1000055048	3152	Cation/multidrug efflux pump	0	0	2	0	0	0	2.0	0.1
MIS_5k.1000025520	3491	Predicted glutamine amidotransferase COG0311	2	0	0	0	0	0	2.0	0.1
MIS_5k.1000005437	1817	Parvulin-like peptidyl-prolyl isomerase	0	0	0	0	0	1	1.0	0.1

SM2F11 Expression			Raw number of transcripts							
Locus_Tag	gene length	Annotation	2012-1D	2012-2D	2012-3D	2012-4N	2012-5N	2012-6N	total	Coverage
MIS_5k.100018238	3038.0	23S rRNA. Bacterial LSU	5329.0	5049.0	3374.0	1117.0	1010.0	5294.0	21173.0	696.9
MIS_5k.10005200.8_exon	1331.0	Unknown Function	563.0	559.0	565.0	164.0	121.0	1038.0	3010.0	226.1
MIS_5k.100052003	1581.0	16S rRNA. Bacterial SSU	521.0	553.0	538.0	161.0	165.0	1104.0	3042.0	192.4
MIS_5k.1002343518	89.0	Unknown Function	3.0	0.0	1.0	0.0	1.0	4.0	9.0	10.1
MIS_5k.100018237	485.0	Ribosomal protein S7	13.0	6.0	2.0	0.0	0.0	7.0	28.0	5.8
MIS_5k.100371722	983.0	Rhoptry-associated protein 1 (RAP-1)	4.0	4.0	4.0	0.0	0.0	10.0	22.0	2.2
MIS_5k.100246773	165.0	Unknown Function	0.0	0.0	0.0	0.0	2.0	0.0	2.0	1.2
MIS_5k.1000675514	185.0	Unknown Function	0.0	0.0	0.0	0.0	2.0	0.0	2.0	1.1
MIS_5k.100281276	743.0	PspA/IM30 family	0.0	0.0	0.0	0.0	0.0	8.0	8.0	1.1
MIS_5k.1001165710	212.0	Unknown Function	0.0	0.0	2.0	0.0	0.0	0.0	2.0	0.9
MIS_5k.100049449	425.0	Ribosomal protein L16/L10E	0.0	0.0	0.0	0.0	0.0	4.0	4.0	0.9
MIS_5k.100253273	449.0	Unknown Function	0.0	0.0	2.0	0.0	1.0	1.0	4.0	0.9
MIS_5k.100337843	230.0	Preprotein translocase SecG subunit	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.9
MIS_5k.1000182319	233.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.9
MIS_100033151	761.0	Bacterial SH3 domain	2.0	2.0	0.0	0.0	0.0	2.0	6.0	0.8
MIS_5k.100274233	257.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.8
MIS_5k.1000675519	260.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.8
MIS_5k.1002467715	311.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.6
MIS_5k.1000983711	629.0	Superoxide dismutase	0.0	0.0	0.0	0.0	0.0	4.0	4.0	0.6
MIS_5k.100177625	950.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	6.0	6.0	0.6
MIS_5k.100064688	317.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.6
MIS_5k.100251219	800.0	Ribosomal protein S5	0.0	0.0	0.0	0.0	0.0	5.0	5.0	0.6
MIS_5k.100197803	320.0	Bacterial nucleoid DNA-binding protein	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.6
MIS_5k.100188458	977.0	Lactate dehydrogenase and related dehydrogenases	0.0	6.0	0.0	0.0	0.0	0.0	6.0	0.6
MIS_5k.100116579	710.0	Protease subunit of ATP-dependent Clp proteases	0.0	0.0	0.0	0.0	0.0	4.0	4.0	0.6
MIS_5k.100291202	533.0	Hemolysin-type calcium-binding repeat (2 copies)	3.0	0.0	0.0	0.0	0.0	0.0	3.0	0.6
MIS_5k.100112637	182.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	1.0	1.0	0.5
MIS_5k.100107866	380.0	Ribosomal protein S13	0.0	2.0	0.0	0.0	0.0	0.0	2.0	0.5
MIS_5k.100061957	383.0	PRA1 family protein	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.5
MIS_5k.100220618	386.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.5
MIS_5k.100052007	428.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.5
MIS_5k.100107042	1136.0	Unknown Function	0.0	4.0	0.0	0.0	0.0	1.0	5.0	0.4
MIS_5k.100134358	458.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.4
MIS_5k.1000387118	1637.0	Unknown Function	0.0	1.0	0.0	0.0	0.0	6.0	7.0	0.4
MIS_5k.100093917	482.0	Unknown Function	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.4
MIS_5k.100168716	500.0	Ribosomal protein L24	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.4
MIS_5k.100052002	527.0	Uncharacterized protein conserved in bacteria	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.4
MIS_100081633	539.0	Protease subunit of ATP-dependent Clp proteases	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.4
MIS_5k.100246771	275.0	Cell division control protein 14, SIN component	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.4
MIS_5k.100222872	620.0	Predicted membrane protein - COG2717	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.3
MIS_100035223	647.0	RecF/RecN/SMC N terminal domain	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.3

MIS_100072062	677.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.3
MIS_5k.1000297912	1382.0	UME (NUC010) domain	0.0	4.0	0.0	0.0	0.0	0.0	4.0	0.3
MIS_5k.1000573111	1199.0	GTPases - translation elongation factors	0.0	0.0	0.0	0.0	0.0	3.0	3.0	0.3
MIS_5k.100096006	1256.0	Periplasmic serine proteases (ClpP class)	3.0	0.0	0.0	0.0	0.0	0.0	3.0	0.2
MIS_5k.100200072	2219.0	Domain of unknown function DUF11	5.0	0.0	0.0	0.0	0.0	0.0	5.0	0.2
MIS_5k.100154632	905.0	Pyruvate:ferredoxin oxidoreductase	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.2
MIS_5k.100318411	947.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.2
MIS_5k.100318284	950.0	Malate/lactate dehydrogenases	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.2
MIS_5k.1000182316	1979.0	Bacterial capsule synthesis protein PGA_cap	4.0	0.0	0.0	0.0	0.0	0.0	4.0	0.2
MIS_100121642	2117.0	Unknown Function	4.0	0.0	0.0	0.0	0.0	0.0	4.0	0.2
MIS_5k.100096002	1106.0	Unknown Function	0.0	0.0	0.0	2.0	0.0	0.0	2.0	0.2
MIS_5k.100044758	1178.0	Predicted periplasmic solute-binding protein - COG1559	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.2
MIS_5k.100291209	1832.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	3.0	3.0	0.2
MIS_5k.1000528912	1235.0	Putative translation factor (SUA5)	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.2
MIS_5k.100129781	1910.0	Intermediate filament tail domain	3.0	0.0	0.0	0.0	0.0	0.0	3.0	0.2
MIS_5k.100280172	692.0	Ribosomal protein S3	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.1
MIS_5k.100220615	2078.0	Glycosyltransferase	3.0	0.0	0.0	0.0	0.0	0.0	3.0	0.1
MIS_100120067	1400.0	Excalibur calcium-binding domain	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.100339322	1418.0	Bacterial Ig-like domain (group 3)	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1
MIS_5k.100033507	1586.0	S-adenosylmethionine-dependent methyltransferase COG0275	0.0	2.0	0.0	0.0	0.0	0.0	2.0	0.1
MIS_100072063	2420.0	Type IIA topoisomerase, A subunit	0.0	0.0	0.0	0.0	0.0	3.0	3.0	0.1
MIS_5k.100050266	815.0	Unknown Function	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.1
MIS_5k.100105104	1688.0	Preprotein translocase subunit SecY	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.100378955	1697.0	GTPase of unknown function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.100354965	1745.0	Unknown Function	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.1000511710	1943.0	Hydrolase of the metallo-beta-lactamase superfamily - COG0595	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.100107044	2141.0	Unknown Function	0.0	0.0	0.0	0.0	2.0	0.0	2.0	0.1
MIS_5k.100349173	2180.0	Domain of unknown function (DUF3357)	0.0	0.0	0.0	0.0	0.0	2.0	2.0	0.1
MIS_5k.100107043	2306.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1
MIS_5k.100253272	2390.0	Type IIA topoisomerase, B subunit	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1
MIS_5k.100277045	1196.0	Domain of unknown function (DUF1978)	1.0	0.0	0.0	0.0	0.0	0.0	1.0	0.1
MIS_5k.100064684	2450.0	Unknown Function	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1
MIS_5k.100061952	1355.0	Unknown Function	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.1
MIS_5k.1000528914	3737.0	Isoleucyl-tRNA synthetase	0.0	0.0	2.0	0.0	0.0	0.0	2.0	0.1
MIS_5k.100050269	3935.0	Leucyl-tRNA synthetase	2.0	0.0	0.0	0.0	0.0	0.0	2.0	0.1

RF3 Expression			Raw number of transcripts							
Locus_Tag	gene length	Annotation	2012-1D	2012-2D	2012-3D	2012-4N	2012-5N	2012-6N	total	Coverage
MIS_100372431	375	16S rRNA. Bacterial SSU	13	48	28	7	13	5	114.0	30.4
MIS_5k.100397601	101	Lysyl-tRNA synthetase (class II)	4	0	0	0	0	0	4.0	4.0
MIS_100441878	350	Protein of unknown function (DUF1475)	0	0	0	0	0	4	4.0	1.1
MIS_5k.1000316721	953	Unknown Function	2	0	0	0	0	0	2.0	0.2
MIS_5k.100397602	917	Unknown Function	1	0	0	0	0	0	1.0	0.1

Supplemental Table 3.4 NCBI SRA[1] accession information

Sample Alias	Accession #
MISgDNA_2007-1D	SRS455884
MISgDNA_2009-1D	SRS455885
MISgDNA_2009-2D	SRS455886
MISgDNA_2009-3D	SRS455906
MISgDNA_2010-1D	SRS455888
MISgDNA_2011-1D	SRS455889
MISgDNA_2011-2N	SRS455890
MISgDNA_2011-3D	SRS455891
MISgDNA_2012-1D	SRS455892
MISgDNA_2012-2D	SRS455893
MISgDNA_2012-3D	SRS455894
MISgDNA_2012-4N	SRS455895
MISgDNA_2012-5N	SRS455896
MISgDNA_2012-6N	SRS455898
MIScDNA_2012-1D	SRS455900
MIScDNA_2012-2D	SRS455901
MIScDNA_2012-3D	SRS455902
MIScDNA_2012-4N	SRS455903
MIScDNA_2012-5N	SRS455904
MIScDNA_2012-6N	SRS455905

1 Sequence Read Archive

3.6 References

Arieli B, Shahak Y, Taglicht D, Hauska G, Padan E (1994). Purification and Characterization of Sulfide-Quinone Reductase, a Novel Enzyme Driving Anoxygenic Photosynthesis in Oscillatoria-Limnetica. *Journal of Biological Chemistry* **269:** 5705-5711.

Biomatters (2013). Geneious, 6.1.5 edn.

Blankenship R, Sadekar S, Raymond J (2007). The evolutionary transition from anoxygenic to oxygenic photosynthesis. In: Falkowski PG, Knoll AH (eds). *Evolution of primary producers in the sea*. Elsevier. pp 21-35.

Bronstein M, Schutz M, Hauska G, Padan E, Shahak Y (2000). Cyanobacterial sulfide-quinone reductase: Cloning and heterologous expression. *J Bacteriol* **182:** 3336-3344.

Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu DY *et al* (2013). Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nature Communications* **4**.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283-1287.

Cohen Y, Padan E, Shilo M (1975). Facultative anoxygenic photosynthesis in cyanobacterium oscillatoria-limnetica. *J Bacteriol* **123:** 855-861.

Cohen Y, Jorgensen BB, Revsbech NP, Poplawski R (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl Environ Microbiol* **51:** 398-407.

Colon-Lopez MS, Sherman LA (1998). Transcriptional and translational regulation of photosystem I and II genes in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. *J Bacteriol* **180:** 519-526.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Yelton AP, Banfield JF (2009). Community-wide analysis of microbial genome sequence signatures. *Genome Biology* **10**.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320:** 1034-1039.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105:** 3805-3810.

Garcia-Pichel F, Mechling M, Castenholz RW (1994). Diel migrations of microorganisms within a benthic, hypersaline mat community. *Appl Environ Microbiol* **60:** 1500-1511.

Garlick S, Oren A, Padan E (1977). Occurrence of facultative anoxygenic photosynthesis among filamentous and unicellular cyanobacteria. *J Bacteriol* **129:** 623-629.

Ito H, Mutsuda M, Murayama Y, Tomita J, Hosokawa N, Terauchi K *et al* (2009). Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc Natl Acad Sci U S A* **106:** 14168-14173.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106:** 16925-16929.

Jorgensen BB, Cohen Y, Revsbech NP (1986). Transition from Anoxygenic to Oxygenic Photosynthesis in a Microcoleus-Chthonoplastes Cyanobacterial Mat. *Appl Environ Microbiol* **51:** 408-417.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-1760.

Litchman E (2000). Growth rates of phytoplankton under fluctuating light. *Freshwater Biology* **44:** 223-235.

Litchman E, Steiner D, Bossard P (2003). Photosynthetic and growth responses of three freshwater algae to phosphorus limitation and daylength. *Freshwater Biology* **48:** 2141-2148.

Mattoo AK, Hoffmanfalk H, Marder JB, Edelman M (1984). Regulation of protein-metabolism - coupling of photosynthetic electron-transport to in-vivo degradation of the rapidly metabolized 32-kilodalton protein of the chloroplast membranes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81:** 1380-1384.

Muyzer G, Stams AJM (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* **6:** 441-454.

Nold SC, Pangborn JB, Zajack HA, Kendall ST, Rediske RR, Biddanda BA (2010a). Benthic Bacterial Diversity in Submerged Sinkhole Ecosystems. *Appl Environ Microbiol* **76:** 347-351.

Nold SC, Zajack HA, Biddanda BA (2010b). Eukaryal and archaeal diversity in a submerged sinkhole ecosystem influenced by sulfur-rich, hypoxic groundwater. *Journal of Great Lakes Research* **36:** 366-375.

Nold SC, Bellecourt MJ, Kendall ST, Ruberg SA, Sanders TG, Klump JV *et al* (2013). Underwater sinkhole sediments sequester Lake Huron's carbon. *Biogeochemistry* **115:** 235-250.

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

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41:** D590-D596.

Rajendhran J, Gunasekaran P (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research* **166:** 99-110.

Rappe MS, Giovannoni SJ (2003). The uncultured microbial majority. *Annual Review of Microbiology* **57:** 369-394.

Ray WK, Zeng G, Potters MB, Mansuri AM, Larson TJ (2000). Characterization of a 12-kilodalton rhodanese encoded by glpE of Escherichia coli and its interaction with thioredoxin. *J Bacteriol* **182:** 2277-2284.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F *et al* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499:** 431-437.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G *et al* (2011). Integrative genomics viewer. *Nature Biotechnology* **29:** 24-26.

Schutz M, Shahak Y, Padan E, Hauska G (1997). Sulfide-quinone reductase from Rhodobacter capsulatus. Purification, cloning, and expression. *J Biol Chem* **272:** 9890-9894.

Shi T, Ilikchyan I, Rabouille S, Zehr JP (2010). Genome-wide analysis of diel gene expression in the unicellular N-2-fixing cyanobacterium Crocosphaera watsonii WH 8501. *Isme Journal* **4:** 621-632.

Soitamo AJ, Zhou G, Clarke AK, Oquist G, Gustafsson P, Aro EM (1996). Over-production of the D1:2 protein makes Synechococcus cells more tolerant to photoinhibition of Photosystem II. *Plant Molecular Biology* **30:** 467-478.

Stockel J, Jacobs JM, Elvitigala TR, Liberton M, Welsh EA, Polpitiya AD *et al* (2011). Diurnal Rhythms Result in Significant Changes in the Cellular Protein Complement in the Cyanobacterium Cyanothece 51142. *Plos One* **6**.

Stoeckel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, Pakrasi HB (2008). Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci U S A* **105:** 6156-6161.

Straub C, Quillardet P, Vergalli J, de Marsac NT, Humbert JF (2011). A Day in the Life of Microcystis aeruginosa Strain PCC 7806 as Revealed by a Transcriptomic Analysis. *Plos One* **6**.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H *et al* (1999). Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *Fems Microbiology Ecology* **29:** 159-169.

Tsertova N, Kisand A, Tammert H, Kisand V (2011). Low seasonal variability in community composition of sediment bacteria in large and shallow lake. *Environmental Microbiology Reports* **3:** 270-277.

Vijayan V, Zuzow R, O'Shea EK (2009). Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* **106:** 22564-22568.

Voorhies AA, Biddanda BA, Kendall ST, Jain S, Marcus DN, Nold SC *et al* (2012). Cyanobacterial life at low O2: community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. *Geobiology* **10:** 250-267.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al* (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337:** 1661-1665.

Yau S, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, Rich J *et al* (2013). Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *Isme Journal* **7:** 1944-1961.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:** 821-829.

Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML *et al* (2009). Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, Prochlorococcus. *Plos One* **4**.

CHAPTER IV

Two-way genetic exchange underpins cyanobacteria-virus interactions in a low-O_2 mat community

Alexander A. Voorhies[1], Sarah D. Eisenlord[2], Daniel N. Marcus[1], Melissa B. Duhaime[3], Bopaiah A. Biddanda[4], James D Cavalcoli[5], and Gregory J. Dick[1,3]

[1] Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, Michigan, 48109
[2] School of Natural Resources and Environment, University of Michigan, Ann Arbor, Michigan, 48109
[3] Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan, 48109
[4] Annis Water Resources Institute, Grand Valley State University, Muskegon, Michigan, 49441
[5] University of Michigan Bioinformatics Core, University of Michigan, Ann Arbor, Michigan, 48109

Abstract

The Middle Island Sinkhole (MIS) in Lake Huron, MI is home to cyanobacterial mats with extremely low species diversity, and the dominant cyanobacteria show extensive genomic evidence of viral predation pressure, making it an excellent model community for studying the interactions of viruses that infect bacteria. Metagenomic and metatranscriptomic sequencing of mat samples collected at five time points between 2007 and 2012 enabled the recovery of complete genomes for two novel genotypes of a virus, designated as PhV1.TypeA and PhV1.TypeB. Both viral genotypes code for and express a host derived *nblA* gene responsible for degradation of phycobilisomes in the host cyanobacteria. Two complete CRISPR (clustered

regularly interspaced short palindromic repeats) subtype III-B loci, thought to be microbial defense mechanisms against viruses and plasmids,were recovered from the putative cyanobacterial host, *Phormidium* sp. MIS-PhA. Genes from both genotypes of PhV1 as well as the CRISPR loci were expressed in all six 2012 samples, showing that both viral predation and cyanobacterial defense are transcriptionally active in the environment. One 45bp CRISPR spacer from *Phormidium* had 100% nucleotide identity to the PhV1.TypeB genome, but this region was absent from the PhV1.TypeA genome, rendering it immune to host CRISPR defense. Analysis of viral abundance over time shows that while PhV1.TypeB (which was dominant in 2007) was targeted by the host CRISPR spacer, its abundance dropped significantly over the time of this study, and by 2012 it was displaced by PhV1.TypeA. These results reveal extensive genetic exchange and interactions between virus and host, highlighting the value of parallel analysis of viral and host genomes in natural microbial communities.

Keywords

Phage, Virus, CRISPR, cas, Cyanobacteria, Metagenomic, Metatranscriptomics.

4.1 Introduction

Bacteria and archaea are some of the most diverse and abundant organisms on the planet, yet it is estimated that there are as many as ten times more viruses (Weinbauer 2004). Viruses evolve at a rapid rate, forcing bacteria and archaea to adapt just as quickly, driving the evolution of microbial defense mechanisms in what is often described as an evolutionary arms race (Avrani et al 2011, Banfield and Young 2009, Heidelberg et al 2009, Rodriguez-Valera et al 2009, Weinbauer 2004). Genome sequences are available for only a small portion of known viruses,

largely from cultured viruses, leaving the interactions between bacteria and their viruses in the natural environment largely unexplored (Suttle 2007). Understanding the fundamental dynamics of the interactions between microbial defense systems and viruses has major implications for medicine, industry, and environmental microbiology.

Bacteria and archaea utilize an adaptive immunity mechanism known as the CRISPR/*cas* (CRISPR-associated sequences) system to defend themselves from viruses and plasmids (Barrangou et al 2007, Makarova et al 2006). CRISPRs have been found in 40% of sequenced bacteria and 90% of sequenced archaea (Grissa et al 2007a), and are thought to have originated in archaea and spread to bacteria through lateral gene transfer (Godde and Bickerton 2006, Makarova et al 2006). CRISPRs consist of short sequences of viral, plasmid or host DNA ("spacers") (Barrangou et al 2007) interleaved between conserved repeat sequences and accompanied by an adjacent set of CRIPR-associated (*cas*) genes (Haft et al 2005). *Cas* genes code for the proteins needed to acquire new spacers from invading viruses or plasmids (Barrangou et al 2007, Horvath and Barrangou 2010, Paez-Espino et al 2013, Sun et al 2013), and prevent the production of proteins from invading sources of DNA (Garneau et al 2010, Haft et al 2005, Horvath and Barrangou 2010, Sorek et al 2008, van der Oost et al 2009). New spacers are acquired from a region of viral or plasmid DNA referred to as the proto-spacer and are added rapidly to one end of the CRISPR (the "leader" end) upon introduction to viral or plasmid DNA. Spacers on the "trailer end" are conserved, and may guard against reinfection by persistent viruses (Andersson and Banfield 2008, Heidelberg et al 2009, Tyson and Banfield 2008, Weinberger et al 2012). Spacer content changes rapidly (though the mechanisms of change are still poorly understood), reflecting constantly changing dynamics in the evolution of

viruses and microbial defenses against them (Andersson and Banfield 2008, Paez-Espino et al 2013, Tyson and Banfield 2008).

While viruses are estimated to be an order of magnitude more abundant than microbes in most environments, their specific roles in ecosystem dynamics remain poorly understood (Rohwer and Thurber 2009, Suttle 2007). Classification of viruses by phylogenetic marker genes, which is typically used for ecological metrics of diversity and community structure, remains challenging because no single gene is conserved amongst all viruses (such as the 16s rRNA for bacteria). Further, classifying natural viral assemblages is made more complicated by extensive genome rearrangements; viral populations are in constant states of genomic flux (Emerson et al 2012). As a consequence, viruses have been shown to contribute to host genome diversification through lateral gene transfer and hijacking of host genes, dramatically shifting microbial community diversity on short time scales (Lindell et al 2004, Lindell et al 2005, Mann et al 2003). Finally, confounding our understanding of the role of viruses in overall ecosystem functioning, it is often difficult to link novel viruses to their hosts, which is essential to gauge the impact of viral activity on key microbial populations. Towards this goal, recently developed methods, such as phageFISH (Allers et al 2013) and digital PCR (Tadmor et al 2011), can be used to successfully identify virus-host interactions, while whole community sequencing data can offer a more high-throughput approach. For instance, the sequence-based linkage between virus-derived spacer and host, which is intrinsic to active CRISPR/cas systems as described above, has been used successfully to link hosts and viruses *in silico* (Andersson and Banfield 2008). Viral populations in environmental samples can be difficult to track, often requiring far deeper sequencing than their microbial counterparts due to their greater genome diversity (Duhaime et al 2012). Therefore, low diversity environments where viral strain variation can be

teased apart with sequence based techniques without the need for isolation and pure culture are important systems for studying predator-prey relations between virus and host.

The Middle Island Sinkhole (MIS) in Lake Huron, MI hosts cyanobacterial mats in a low oxygen, sulfidic habitat (Ruberg et al 2008) that are dominated by a single genus of cyanobacteria that shows evidence of viral predation pressure (Voorhies et al 2012). Here we reconstruct two genome sequences of the virus PhV1 and document two directional exchange of DNA between the virus PhV1 and its host *Phormidium*. *Phormidium* encodes multiple CRISPR systems, one of which contains a DNA sequence (CRISPR spacer) derived from PhV1, and PhV1 contains a host derived AMG (*nblA*) which breaks down photosynthetic pigments.

4.2 Results

Recovery of two circular viral genomes from MIS metagenomes

Fourteen MIS cyanobacterial mat samples were collected at seven time points in five different years between 2007 and 2012, including day and night samples in 2011 and 2012. Random shotgun sequencing and *de novo* genome assembly was used to recover complete genomes for two genomic variants of a novel virus, designated as PhV1.TypeA and PhV1.TypeB (Figure 4.1). Only two of fourteen samples independently yielded complete genomes for both TypeA and TypeB, likely due to the significantly larger number of reads available in those samples (Table 4.1). Separation of PhV1 genotypes in remaining samples was achieved by mapping reads from each sample to template sequences and visually verifying the reads matched to the correct variant (Supplemental Figure 4.1 in Appendix C). By accurately identifying reads belonging to TypeA or TypeB, consensus sequences were created and used to evaluate genome completeness for each sample (Supplemental Figure 4.2 in Appendix C).

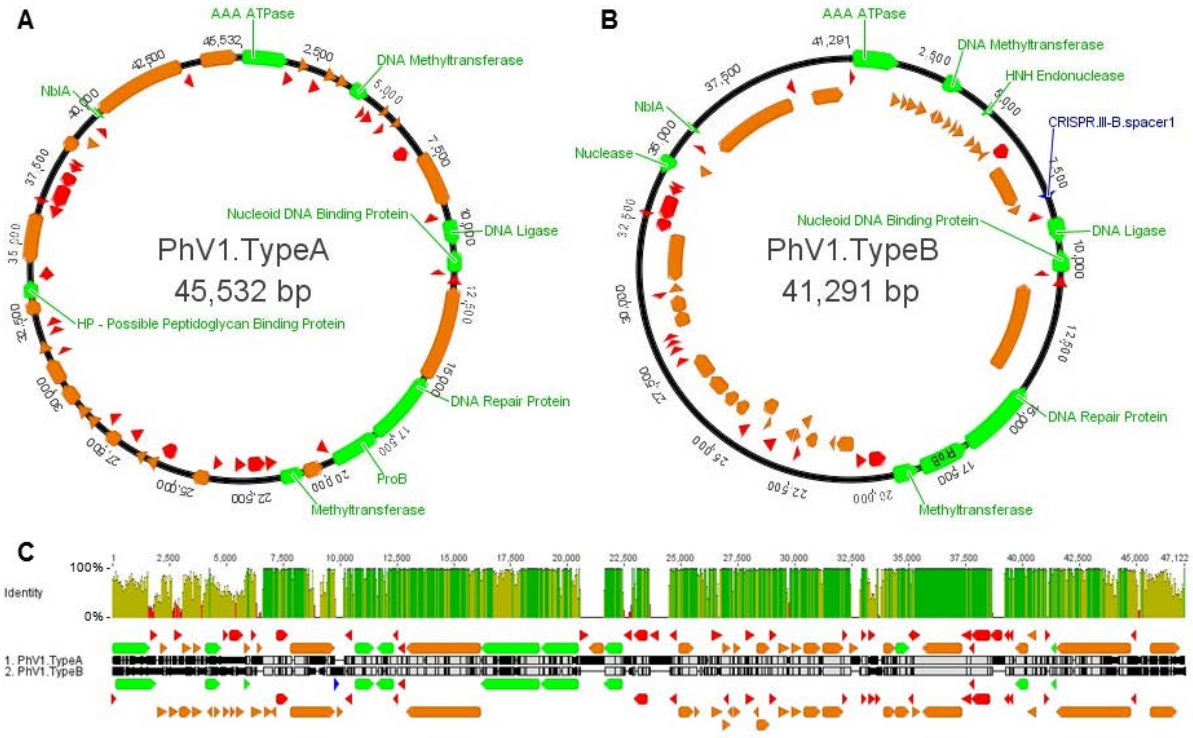


Figure 4.1 *Phormidium* phage PhV1 genotypes. (A) Genome map of PhV1.TypeA with gene annotations; (B) Genome map of PhV1.TypeB with gene annotations; (C) Nucleotide alignment of TypeA and TypeB. Green: genes with annotation; orange: hypothetical genes present in NR with no annotation; red: novel genes with no homologs in NR. The nucleotide alignment is represented visually with grey areas showing nucleotide agreement and black regions indicate nucleotide disagreement, and thin line indicating gaps in the alignment. Nucleotide identity (tops of (C)) is represented by green (100%), gold (99-30%), and red (less than 30%).

PhV1 genomes assembled into circular chromosomes, with genome sizes of 45kb for PhV1.TypeA and 41kb for PhV1.TypeB, encoding 62 and 57 predicted genes respectively. Annotation of these genomes (Supplemental Tables 4.1 and 4.2 in Appendix C) revealed primarily conserved genes of unknown function or completely novel coding regions with no known homologs, with only 9 genes in TypeA and 10 genes in TypeB having predicted functions. The two genomes share 81% nucleotide identity, but due to extensive insertions and deletions (Figure 4.1) they only share 44 genes. The genome size and number of coding genes

present in PhV1 is consistent with cyanophage Pf-WMP3, a member of the T7 supergroup phages that has a 43kb genome with 41 genes and infects *Phormidium foveolarum* (Liu et al 2008). However, it should be noted that PhV1 and cyanophage Pf-WMP3 do not share any genes, and are thus genetically unrelated. While only two genotypes of PhV1 were abundant enough to rebuild complete genomes, there is extensive evidence that other low abundance variants of PhV1 exist at MIS (Supplemental Figure 4.3 in Appendix C).

BLASTp analysis showed that thirteen PhV1 genes had predicted protein sequences with greater than 30% amino acid identity to predicted proteins from a custom database containing all phage genomes available in the NCBI RefSeq database. Twenty three PhV1 predicted proteins had greater than 30% ID BLASTp similarity to NCBI's viral database, while two PhV1 genes were similar to known prophage (viral DNA integrated within a host genome) in cyanobacterial genomes available in NCBI. The possibility that PhV1 is a plasmid is unlikely based on lack of signature plasmid replication or pilus genes (Ma et al 2012), homologs of which were found on plasmids at MIS but not in either PhV1 genome. No ubiquitous plasmid signature genes have been identified, so NCBI's plasmid DB was searched for PhV1 gene homologs. Only seven PhV1 genes had homology via BLASTp, but had similarity scores of less than 50%, and the homology was to genes of unknown function. Taken together, these results indicate that PhV1 is a virus and not a plasmid.

PhV1 appears to target the dominant cyanobacterium at MIS, *Phormidium* sp. MIS-PhA, based on two main lines of evidence. First, PhV1 and *Phormidium* share a tetranucleotide frequency (Supplemental Figure 4.4 in Appendix C), which has been shown to accurately distinguish between *Phormidium* and other cyanobacteria at MIS (Voorhies et al 2012). It has been shown that viruses often display a similar tetranucleotide frequency to their bacterial hosts

(Andersson and Banfield 2008, Dick et al 2009b, Pride and Schoenfeld 2008). Second, the *Phormidium* sp. MIS-PhA genome contains a 45bp segment of DNA matching to the PhV1.TypeB genome with 100% nucleotide identity that is flanked by CRISPR subtype III-B repeats and is part of a complete CRISPR locus (Figure 4.2).

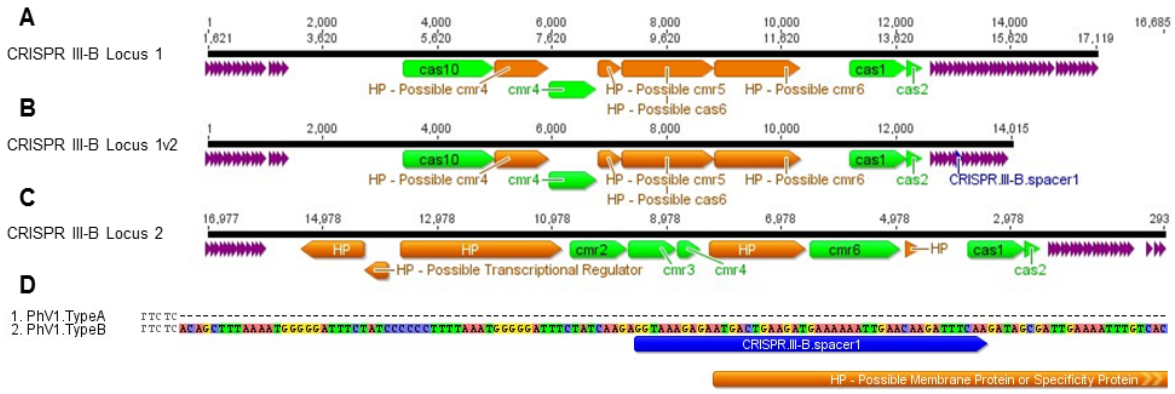


Figure 4.2 Gene maps of CRISPR subtype III-B loci belonging to *Phormidium* sp. MIS-PhA. (A) CRISPR III-B.Locus1 with the most abundant spacer order; (B) CRISPR III-B.Locus1.v2 shows alternative spacers and contains CRISPR.III-B.spacer1; (C) CRISPR III-B.Locus2 with the most abundant spacer order; (D) CRISPR spacer from III-B.Locus1.v2 aligned between PhV1 genotypes; dashes indicate missing sequence in that region of the genome. Green: genes which could be annotated; orange: genes of unknown function; purple: CRISPR III-B repeat; blue: region that matches a CRISPR III-B spacer.

Although most PhV1 genes have no predicted function, or are completely unrepresented in public gene databases, both genotypes encode a phycobilisome degradation protein NblA, which has been found in two other cyanophage genomes (Gao et al 2012, Yoshida-Takashima et al 2012). NblA is thought to improve viral replication by breaking down the major photosynthetic apparatus, thereby providing the virus with raw materials or avoiding photo-inhibition (Gao et al 2012, Yoshida-Takashima et al 2012). In addition to being present in the viral genomes, *nblA* is also present in the cyanobacterial host. Four distinct versions of NblA are encoded by *Phormidium* at MIS, two of which are 69% and 67% similar at the amino acid level to PhV1 (Figure 4.3).

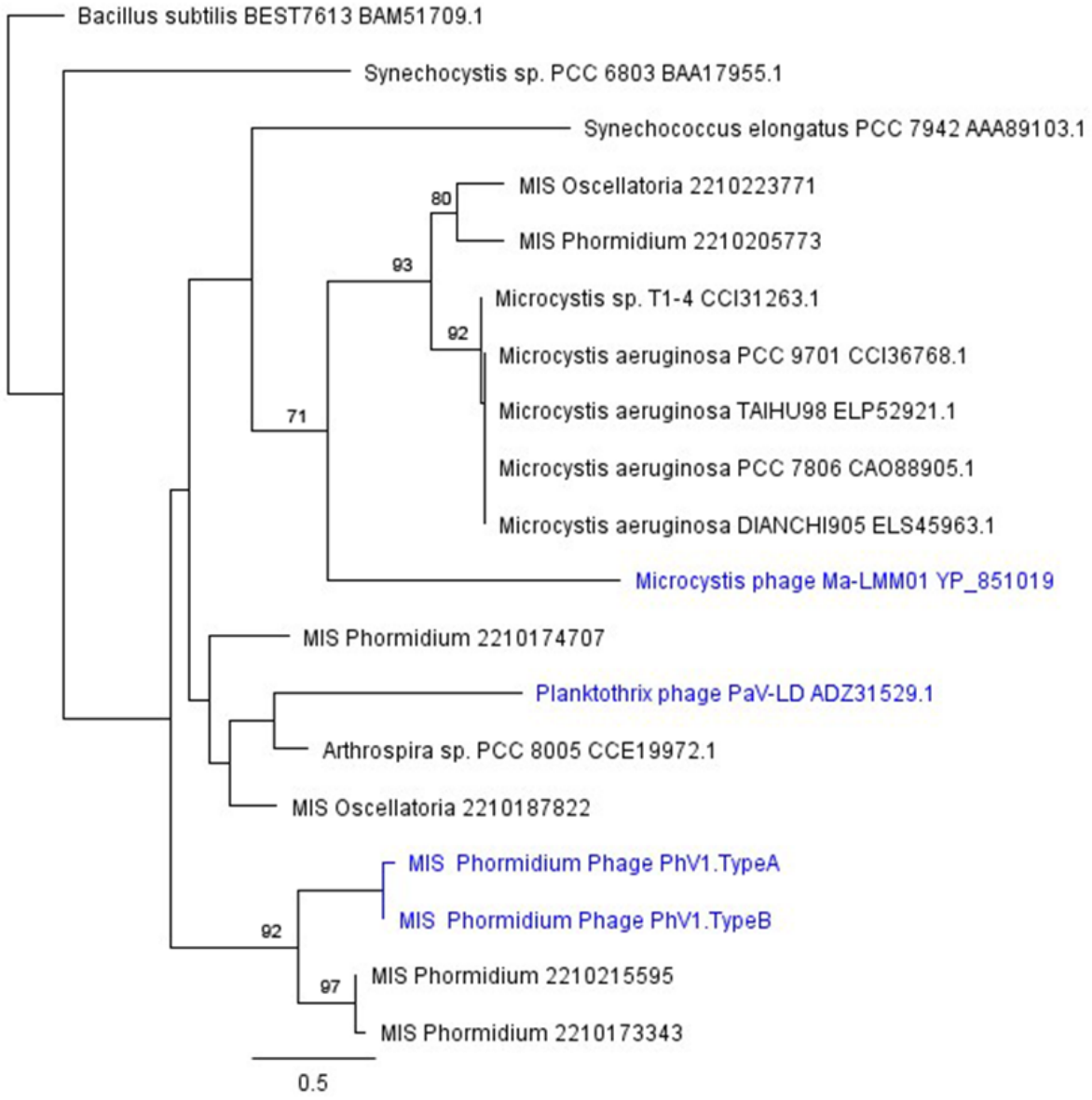


Figure 4.3 Maximum likelihood tree of cyanobacterial and viral phycobilisome degradation protein NblA. Bootstrap values below 70 have been removed. Blue text indicates NblA sequences from viral genomes.

Host defense

Phormidium sp. MIS-PhA is the dominant organism at MIS and the only cyanobacterium present in all samples analyzed here. It encodes for multiple CRISPR systems (Voorhies et al 2012),

which could provide defense mechanisms against viruses and plasmids. *Phormidium* sp. MIS-PhA encodes two distinct subtype III-B CRISPR loci (as classified by Makarova et al. (Makarova et al 2011a, Makarova et al 2011c) and designated here as Locus1 and Locus2) based on the presence of conserved *cas* genes (Figure 4.2) and five conserved signature repeats (Supplemental Table 4.3 in Appendix C). While these two CRISPR loci share repeats and homologous *cas* genes, the *cas1* genes share only 88.7% nucleotide identity. Spacers, which may contain viral or plasmid DNA, are distinct between the two CRISPR loci (Supplemental Figure 4.5 in Appendix C), and 33 contigs containing different arrangements of 225 unique spacers were recovered. *Cas* genes and the most abundant spacer combinations ("spacer contigs") are conserved and were reconstructed for all samples (Supplemental Figure 4.6 in Appendix C). Less abundant spacer contigs (representing less frequent arrangements of spacers) varied in both presence and abundance between samples, showing a gradual shift in spacer content over time (Figure 4.4). Average read abundance of "spacer contigs" (calculated from assembled contigs of spacers) was not significantly different to average read abundance calculated from single spacers (Figure 4.5). This pattern suggests that the majority of spacers are commonly found in the conserved order represented by our designated spacer contigs. For most spacers, we do not see a discordance between "conserved order" spacer abundance and total spacer abundance, which would suggest prevalent exchange of single spacers through lateral gene transfer, or multiple acquisitions of the same spacer.

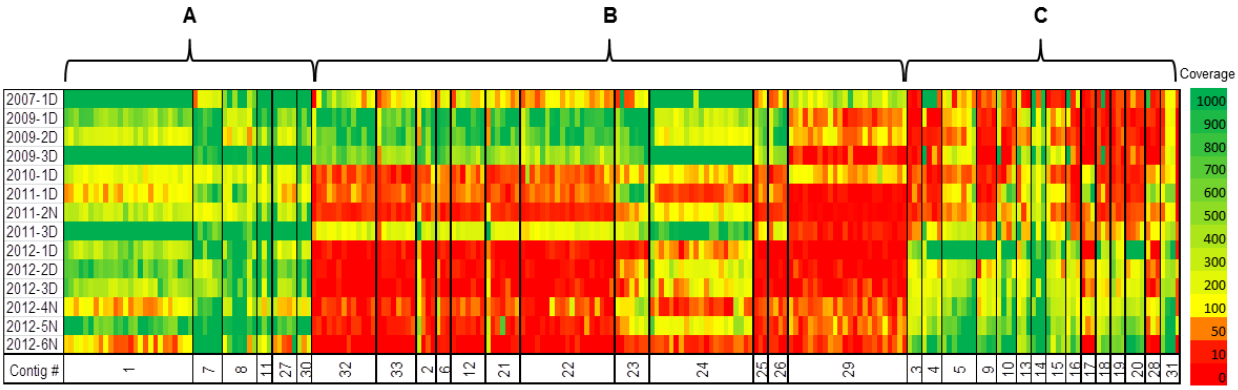


Figure 4.4 Heat map of CRISPR III-B spacer abundance across 14 samples from 2007 to 2012. Year and type of sample (day(D)/night(N)) are indicated on the far left. Columns represent individual spacers, and black vertical lines distinguish groups of spacers with conserved order assembled to the same "spacer contig". (A) spacers and conserved order contigs with relatively even abundance across most samples; (B) spacers and conserved order contigs more abundant in earlier samples; (C) spacers and conserved order contigs more abundant in later samples.

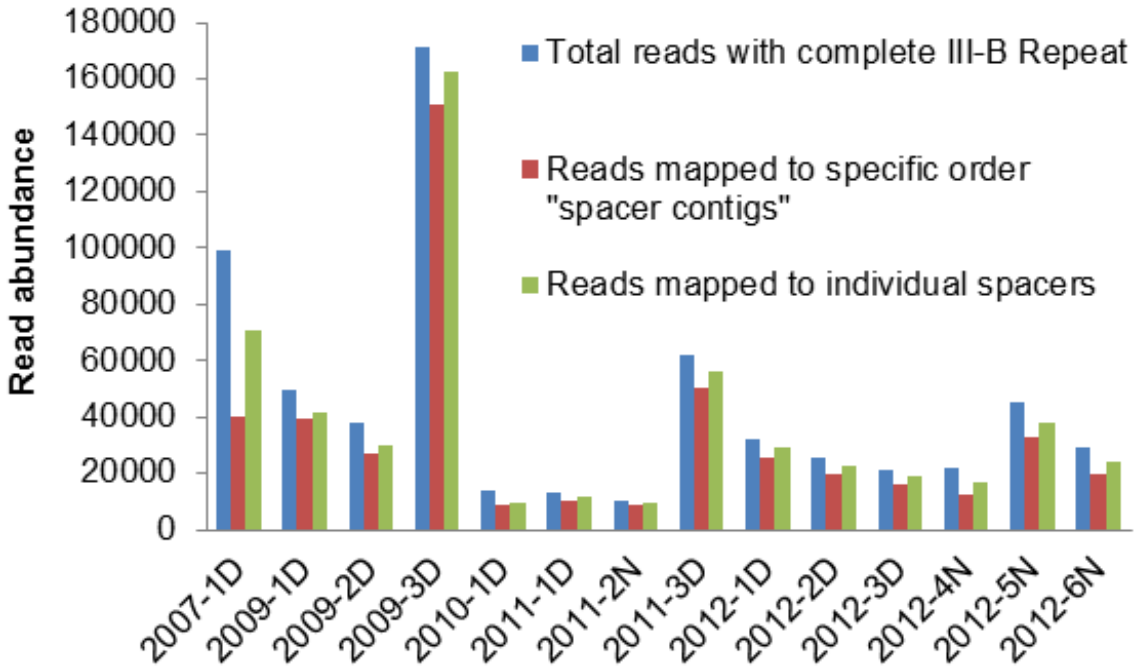


Figure 4.5 CRISPR subtype III-B individual spacer and conserved order "spacer contigs" normalized average read abundance.

In order to determine the degree to which spacers maintain the same spacer order (e.g., by vertical inheritance) versus having a different spacer order (e.g., by horizontal transfer, multiple acquisitions, recombination, etc.) the composition of spacer contigs was compared to that of individual spacers over the course of five years using ANOSIM. The R statistic from ANOSIM ranges from 0-1; the closer to 1, the more distinct the gene assemblages are by year, the closer to 0, the more similar. The composition of individual spacer abundances are more similar across sample years (ANOSIM R statistic of 0.668, 0.1% significant), than conserved order spacer contig abundances (R statistic of 0.824, 0.1% significant). In agreement with the ANOSIM results, CLUSTER analysis of spacer composition between samples indicated that samples were more similar to each other when comparing the composition of individual spacers (60% similarity) instead of spacer contigs (40% similarity) (Figure 4.6). This implies that spacer order differs from sample to sample more than overall spacer content. PERMANOVA results showed the composition of both individual spacer and conserved order spacer contig abundances are conserved within each sampling year (P=0.005; and 0.001, respectively), making samples collected in the same year more similar to each other than samples from other years. Principal coordinate analyses (PCO; Supplemental Figure 4.7 in Appendix C) also showed that samples are more similar when examining individual spacers than conserved order spacer contigs, with 58.6% and 53.8% of the variation between samples explained by the first PCO axis (correlated with sample year) for individual spacers and conserved order spacer contigs respectively (Supplemental Figure 4.7 in Appendix C).

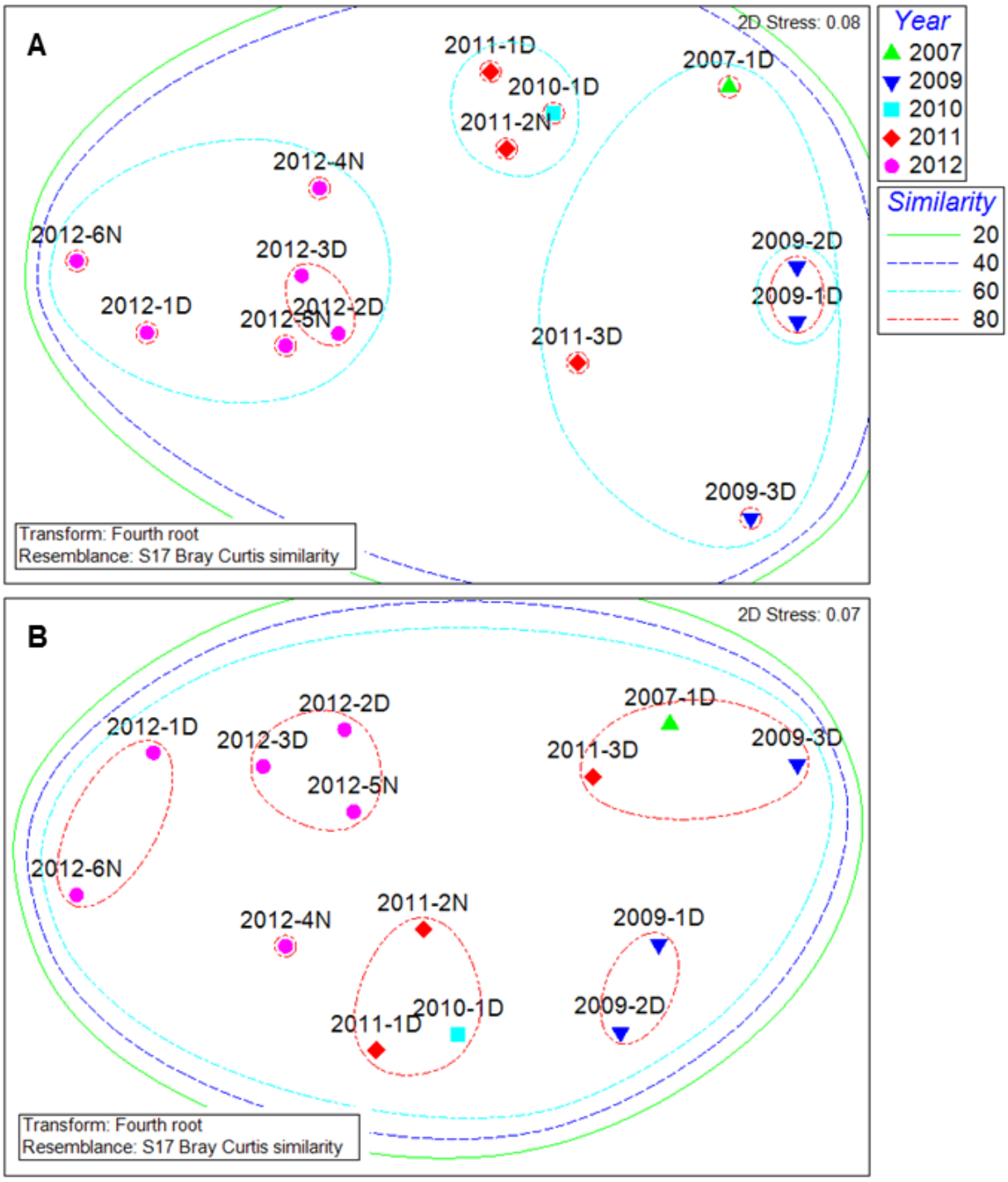


Figure 4.6 Non-metric multidimensional scaling (nMDS) plots of normalized average CRISPR spacer read abundance. (A) Similarity of abundance of CRISPR conserved order "spacer contigs"; (B) similarity of abundance of individual CRISPR spacers.

BLASTn was used to compare the 225 unique spacers from CRISPR subtype III-B against both genotypes of PhV1, revealing a single spacer that shares 100% nucleotide identity with the proto-spacer region of PhV1.TypeB, and is present in half the samples analyzed. This spacer, designated CRISPR.III-B.spacer1, belongs to the *Phormidium* sp. MIS-PhA CRISPR Locus1, whereas Locus2 does not possess any spacers matching to either genotype of PhV1. The PhV1.TypeB proto-spacer that matches CRISPR.III-B.spacer1 is not present in PhV1.TypeA due to a gap in that region of the TypeA genome (Figure 4.2D).

Gene expression of PhV1 and host CRISPR loci

Metatranscriptomic sequencing was performed on six samples collected in 2012. cDNA reads were mapped to template sequences from PhV1 and the *Phormidium* CRISPR loci to determine if these regions were actively expressing RNA in the environment at the time of collection. The most highly expressed PhV1 genes were from genes of unknown function, while *nblA* and a gene with a predicted peptidoglycan binding domain were also expressed in multiple samples, albeit at low levels (Supplemental Figure 4.8 in Appendix C). These results indicate that the PhV1 viruses were transcriptionally active at the time of sampling.

All six metatranscriptomes from 2012 contained transcripts from both *cas* genes and spacer regions in CRISPR Locus1 (Supplemental Figure 4.9 in Appendix C), demonstrating that *Phormidium* was actively transcribing its CRISPR loci in the environment at the time of collection. A non-coding region directly adjacent to the upstream series of spacers (located at the ~1500bp mark in Supplemental Figure 4.9 in Appendix C), is the most highly expressed area of Locus1. This region may represent a regulatory non-protein-coding small RNA, which have been identified at high frequencies in other environmental microbial metatranscriptomes and

have been implicated in regulatory functions based on their proximity to regulatory genes (e.g., carbon metabolism, nutrient acquisition) and lack of open reading frames (Frias-Lopez et al 2008, Gilbert et al 2008, Shi et al 2009). It should also be noted that the spacer/repeat regions show significantly more expression than the *cas* genes themselves in all samples. As half of the 2012 samples were collected during the day and half at night, RNA expression was examined for day/night trends, but none were detected.

Abundance of viral genotypes and PhV1 targeting CRISPR spacer across samples

The CRISPR.III-B.spacer1 that matches PhV1.TypeB was recovered from seven out of fourteen samples taken from MIS between 2007 and 2012, including the 2007 sample, one of the three 2011 samples, and five out of six samples from 2012 (Figure 4.7). Normalized (to account for varied sequencing effort) and averaged read abundance calculated for both genotypes of PhV1 show a decline in TypeB abundance and subsequent rise of TypeA abundance over time. In the 2007 sample, PhV1.TypeB is over three-fold more abundant than PhV1.TypeA, a trend that continues until 2011 when the viral community shifts to PhV1.TypeA being more abundant. During the same period, the abundance of *Phormidium* (based on read abundance of 4 single copy housekeeping genes; *L2P*, *L4P*, *secY* and *ychF*) is relatively high in 2007-2010, but declines in 2011-2012. Of the nine samples from 2011-2012, only 2 samples have *Phormidium* abundances equivalent to those detected in 2007-2010. Both samples in which *Phormidium* has a higher read abundance also contain CRISPR.III-B.spacer1 at relatively high levels (Supplemental Figure 4.10 in Appendix C) compared with other samples containing that spacer.

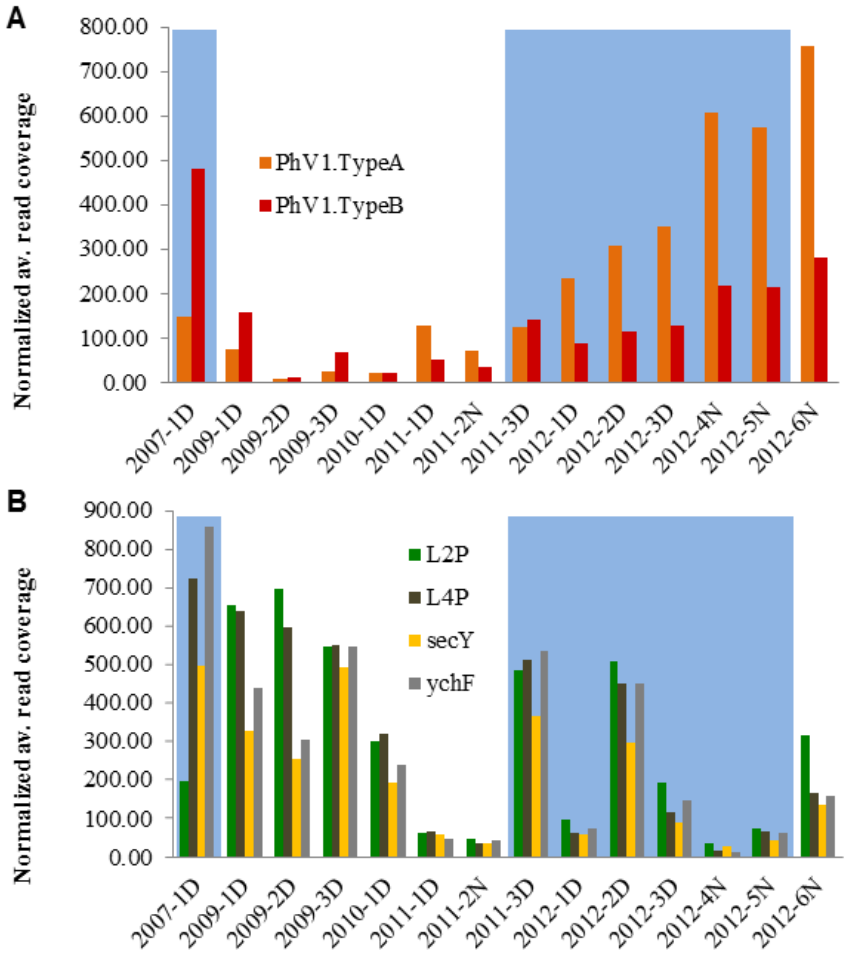


Figure 4.7 Normalized gDNA read abundance of *Phormidium* and its viruses plotted for each sample and averaged over the length of the gene/genome. (A) virus PhV1.TypeA and PhV1.TypeB read abundance; (B) *Phormidium* read abundance based on four different single copy housekeeping genes. Blue background shows samples that contain CRISPR.III-B.spacer1 that matches PhV1.TypeB. Note that samples collected during the day are followed by D and samples collected at night are followed by N.

4.3 Discussion

We have used a metagenomic and metatranscriptomic approach to link the cyanobacterium that dominates a microbial mat and a highly abundant virus that preys upon it. Genome analysis shows that *Phormidium* phage MIS-PhV1 contains genes that are mostly of unknown function, and shares no gene sequence with viruses known to infect *Phormidium* in other environments (Liu et al 2008), highlighting the novelty of PhV1. When PhV1 genes were compared to

publicly available viral databases, no structural gene homologs were identified, making further phylogenetic classification impossible. Two distinct genotypes of PhV1 contain large genome rearrangements consisting of insertions, deletions and SNPs, and there is extensive evidence that many low abundance genotypes coexist in the environment.

Both genotypes of PhV1 encode a phycobilisome degradation protein NblA, which is used by cyanobacteria to break down the major photosynthetic apparatus and by viruses attacking cyanobacteria (Gao et al 2012, Yoshida-Takashima et al 2012). Though the benefit to viruses encoding *nblA* is unknown, it has been speculated that NblA could be used to obtain amino acids for viral replication and to reduce photo-damage to viral particles during infection (Gao et al 2012, Yoshida-Takashima et al 2012). While doubling times for *Phormidium* species at MIS are unknown, their psychrophilic nature (growing at 9° C) and the doubling times of known *Phormidium* (0.07-0.5 d^{-1}, depending on light and nutrient availability) (Litchman 2000, Litchman et al 2003) suggest the MIS strains are relatively slow growing. As slower host growth is correlated with longer infection periods (Middelboe 2000), MIS *Phormidium* viruses likely experience long latent periods, exposing them to photo-oxidative damage for longer periods of time in the intracellular environment. *Microcystis* phages also encode an NblA that is expressed during infection of a relatively slow growing host (0.12-0.45 d^{-1}) (Van Der Westhuizen and Eloff 1985, Yoshida-Takashima et al 2012). Dampening hosts' photosynthetic capacity and reducing potential buildup of reactive oxygen species by degrading host phycobilisomes may be a strategy employed by viruses infecting slow-growing cyanobacteria to improve fitness.

The *nblA* encoded by PhV1 clusters phylogenetically with *nblA* encoded by the host *Phormidium* sp. MIS-PhA, suggesting that *nblA* was acquired by PhV1 from its host instead of other viral lineages (Figure 4.3). This pattern is consistent with other viral encoded auxiliary

metabolic genes (AMGs) (Breitbart et al 2007), which are primarily derived from host or other microbial origins (Ignacio-Espinoza and Sullivan 2012). However, current lack of available host and viral genomes for these groups make definitive determination of the phylogenetic relationships and evolutionary history of *nblA* difficult. More sequences of both host and viral encoded *nblA* are required to obtain finer phylogenetic resolution of this gene.

As has been noted in other environmental studies of virus-host dynamics (Andersson and Banfield 2008, Pride et al 2011, Tyson and Banfield 2008), extensive evidence from the current study suggests that PhV1 relies upon rearrangements of its genome to evade CRISPR defenses. This form of viral evasion of host defenses differs from isolate experiments that have shown rapid point mutation in regions targeted by CRISPRs (Sun et al 2013). Indeed, gDNA sequence reads belonging to PhV1.TypeB found in 2011-2012 contain no conserved mutations in the region targeted by CRISPR.III-B.spacer1, even though TypeB persists in the environment (at reduced abundance) in all 2011-2012 samples. It is unclear if the lack of mutation in the proto-spacer region of PhV1.TypeB is due to strong positive selection on that viral gene, reflects the relatively slow cyanobacterial growth rates at MIS compared to most viral/CRISPR studies in laboratory settings, or some other factor.

Given the high abundance and persistence of PhV1, it is surprising that only one spacer was found matching TypeB and no spacers were found matching TypeA. This differs from previous studies which show rapid spacer acquisition upon exposure to new viruses (Barrangou et al 2007, Sapranauskas et al 2011, Tyson and Banfield 2008). There is no clear explanation for the lack of observed mutations in the proto-spacer region of PhV1, or the lack of spacers targeting the very abundant virus PhV1.TypeA, but these observations imply a more stable host-

prey dynamic with slower adaptation rates than has been observed in culture (Barrangou et al 2007, Sun et al 2013) or other environments (Andersson and Banfield 2008).

Spacer content of the CRISPR subtype III-B loci was found to be relatively consistent within samples collected from the same year, but distinct between sampling years. A core set of spacers are present in all samples and probably provide resistance to ancestral or persistently occurring viruses (Tyson and Banfield 2008), whereas less abundant spacers change in abundance over time and probably reflect recently acquired spacers. Similarity in read abundance between CRISPR spacer contigs and individual spacers implies most spacers maintain their order over the time scale of the current study, with a few high abundance spacers responsible for the majority of difference in spacer content from sample to sample and year to year. Recent work by Paez-Espino et al. has shown a strong bias for re-acquisition of spacers from previously sampled proto-spacers (Paez-Espino et al 2013), which may account for the observed highly abundant spacers at MIS.

During the course of this study we observed the abundance of PhV1 genotypes, as well as its *Phormidium* host, shift in a consistent fashion. In 2007 *Phormidium* abundance is at its highest, PhV1.TypeB abundance is at its highest, and PhV1.TypeA abundance is low (Figure 4.7). From 2009 to 2011, PhV1 abundance for both genotypes is quite low, while *Phormidium* abundance appears to undergo a steady decline. By 2012 PhV1.TypeA abundance is greater than PhV1.TypeB, and *Phormidium* remains lower abundance than previous years. The correlation of abundance reversals of PhV1 genotypes, and overall drop in *Phormidium* abundance with the appearance of a CRISPR spacer matching PhV1.TypeB in 2007 implies that *Phormidium* was able to successfully defend itself from PhV1.TypeB, causing the observed drop in TypeB abundance. The decline in PhV1.TypeB abundance correlates well with the rise of

PhV1.TypeA, suggesting TypeA was able to evade *Phormidium's* CRISPR defense due to absence of the proto-spacer region. Indeed, increased abundance of PhV1.TypeA in 2011 and 2012 correlates with a severe decrease in abundance of *Phormidium* in seven out of nine samples. These correlations suggest a 'Kill-the-Winner' (KtW) scenario (Rodriguez-Valera et al 2009, Thingstad 2000) in which viruses target the dominant organism until its abundance decreases. Such viral activity may promote increased diversity of cyanobacteria within the MIS microbial mats.

4.4 Conclusions

We have used metagenomics and metatranscriptomics to document exchange of DNA between a virus and its host. Two genomes for the virus *Phormidium* phage MIS-PhV1 were recovered and linked to the dominant Cyanobacteria at MIS, *Phormidium* sp. MIS-PhA. PhV1 was linked to *Phormidium* using a host derived AMG (*nblA*) encoded by the virus, which breaks down cyanobacterial photosynthetic pigments attuned to low light, called phycobilisomes. A CRISPR spacer in the *Phormidium* genome matches a sequence in the PhV1.TypeB genome, and illuminates one potential reason for observed shifts in viral abundance from TypeB being dominant in 2007 (and targeted by a *Phormidium* CRISPR spacer), to TypeA (which is not susceptible to host CRISPR defense due to a deletion of the protospacer) being dominant in 2012.

This study highlights some of the advantages of culture independent community analysis by linking a high abundance virus to its host using metagenomic sequence data, inviting inferences about viral ecology and host interactions. Assembly of time-series metagenomes revealed multiple lines of evidence that imply that viral-host adaptation rates may be slower than

those observed in other studies, including slow addition of CRISPR spacers by *Phormidium*, and lack of conserved mutations in the protospacer region of PhV1.TypeB. It is not clear if this apparently slower adaptation rate is due to host physiology (e.g., slow growth rates of *Phormidium*), environmental factors (e.g., low temperature, low light), or some combination of these and other factors.

It is clear that multiple exchanges of DNA between virus and host have occurred in the low oxygen cyanobacterial mats at the Middle Island Sinkhole, reflecting viral pressure on the dominant community member, *Phormidium*. With current viral gene databases severely under-representing true viral diversity in nature, discovery of novel viruses such as PhV1 is to be expected. Therefore, linking viruses to their hosts and to possible ecological strategies can be invaluable for improving our understanding of viral influences on microbial mortality, and repercussions for ecosystem dynamics.

4.5 Materials and Methods

Sampling and sample preparation

Microbial mat samples were collected aboard the *R/V Storm* by NOAA divers from MIS between 2007 and 2012. DNA was extracted and processed (without amplification) for metagenomic shotgun pyrosequencing as previously described (Voorhies et al 2012).

RNA was extracted from 2012 samples, converted to cDNA, and amplified for sequencing as previously described (Frias-Lopez et al 2008). All samples available and preserved in RNA later from MIS were used for metagenomic analysis to maximize sample time points.

Sequencing and assembly

Please note, specific parameters for all bioinformatics software can be found in Supplemental Table 4 in Appendix C.

Fourteen environmental samples (Table 4.1) were shotgun sequenced on an Illumina Hi Seq 2000 instrument producing paired end reads at the University of Michigan DNA Sequencing Core. Assembly of Illumina shotgun reads was achieved by running three independent Velvet (Zerbino and Birney 2008) assemblies for each sample with kmer values of 91, 75 and 61 which were then edited with MetaVelvet (Namiki et al 2012) and combined using the SAMtools (Li et al 2009) function minimus2. Reads files were dereplicated prior to assembly to remove reads with 100% ID and overlap, and trimmed based on quality scores before assembly.

Table 4.1 Metagenome sample summary

Sample	Finger[1]	Prostrate Mat[2]	Day	Night	Illumina gDNA Reads	gDNA Normalization Factor	Illumina cDNA Reads	cDNA Normalization Factor
2007-1D	x		x		55,742,528	3.80	none	none
2009-1D		x	x		56,610,020	3.74	none	none
2009-2D	x		x		58,811,950	3.60	none	none
2009-3D	x		x		25,542,490	8.30	none	none
2010-1D	x		x		69,716,858	3.04	none	none
2011-1D		x	x		98,430,946	2.15	none	none
2011-2N		x		x	211,878,358	1.00	none	none
2011-3D	x		x		59,608,800	3.55	none	none
2012-1D	x		x		25,666,608	8.26	11,997,534	1.22
2012-2D	x		x		41,595,684	5.09	14,577,110	1.00
2012-3D	x		x		41,318,538	5.13	11,524,300	1.26
2012-4N	x			x	42,876,682	4.94	12,125,414	1.20
2012-5N	x			x	38,335,454	5.53	12,100,438	1.20
2012-6N	x			x	28,207,614	7.51	13,079,448	1.11

1 Microbial mat structure that is raised off the lake sediments

2 Microbial mat that lies flat on lake sediments

Separate assemblies were performed for each of the 14 samples, and contiguous sequences from all 14 samples were combined using minimus2 to identify regions of interest such as CRISPR loci or viral genomes that were present in multiple samples. Reads from each sample were then mapped back onto the templates using Geneious (Biomatters 2013) (www.geneious.com) to create sample specific sequences from the consensus. Assemblies were validated against the previous 454Ti assembly (Voorhies et al 2012), and through extensive manual curation using the genomic viewers IGV (Robinson et al 2011) and Geneious to visualize reads mapped onto contigs using BWA (Li and Durbin 2009). Gene calling and annotation was performed by the Joint Genome Institute's Integrated Microbial Genomes Expert Review portal (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi) confirmed using Prodigal (Hyatt et al 2010). The annotations of specific genes of interest were verified using BLASTp with cutoffs of 150 bitscore and ID of 30% to NR, COG, KEGG and Pfam.

Estimating viral and bacterial abundance

Genomic DNA and cDNA read abundance was calculated by first mapping reads to regions of interest in BWA, and then extracting those reads into a new fastq file. The selected reads were then competitively mapped to ROI DNA templates using Geneious (reads with multiple 100% matches were mapped to all possible regions). Separation of different strains of PhV1, and CRISPRs in the *Phormidium* genome, were performed visually in Geneious based on read coverage and single nucleotide polymorphism (SNP) patterns. Read abundances were normalized for the number of reads available in a sample, and then averaged over the length of the ROI or genome, see Table 4.1 for normalization factors and sample sizes.

Multivariate Statistics

Multivariate statistics were conducted in PRIMER v6 (Clarke 2006) to visualize and analyze the differences in spacer and contig composition in samples from 2007 through 2012. Normalized spacer abundances and normalized contig abundances were 4^{th} root transformed to de-emphasize the most abundant sequences, and then rank similarity matrices were calculated with the Bray-Curtis similarity metric. A one way analysis of similarity (ANOSIM) with year as the main factor was then calculated with 999 permutations to retrieve the R statistic for community similarity between years (Clarke 1993). Hierarchical clustering, using the group average method, was calculated to visualize the natural grouping of samples. Non-metric multidimensional scaling (NMDS) plots and Principal-coordinates analyses were created to further visualize differences in the composition of spacers and contigs within and between sampling dates. The multivariate adaptation of ANOVA, PERMANOVA (Anderson 2001) was then calculated with year as the main factor and Monte Carlo simulations to test the difference between individual spacer and spacer contig abundance across years. To examine which genes contributed most to the difference in the composition, the SIMPER tool in PRIMER v6 was used to calculate the contribution of each spacer or contig to the average Bray-Curtis dissimilarity between years.

ML tree

Protein sequences were globally aligned using the Geneious Alignment tool using a Blosum62 cost matrix. A Maximum Likelihood tree was created using PhyML (Guindon and Gascuel 2003) using the JTT substitution model and were bootstrapped 5000 times, only values >70 are

reported on the tree. An alternate tree was created using the MUSCLE aligner (Edgar 2004) and the LG+I+G substitution model in PhyML to verify the topology of the tree.

BLAST

Three main types of BLAST analysis were performed. Searches for function were conducted using BLASTp with an identity cutoff of 60%. Homology searches were conducted using BLASTp with an identity cutoff of 30%. Searches for CRISPR repeats and spacers were conducted using BLASTn with an Identity cutoff of 100% over the length of the repeat or spacer, though lower stringency was used to search for spacers containing point mutations.

Viral identification

Potential viral sequences were searched for homology against NCBI's viral DB, and two custom databases. One of the custom databases was created using all available complete phage genomes available on NCBI's RefSeq, with all non-phage sequences (mostly eukaryotic viruses) removed. The second was a curated database of genes found in viral metagenomes titled "The Pacific Ocean Virome" and curated by Hurwitz et. al. (Hurwitz and Sullivan 2013). MIS sequences were also compared to a custom database created using all plasmid sequences available on NCBI's RefSeq.

CRISPR identification

CRISPR sequences were identified using CRISPR Finder Online (Grissa et al 2007b) and *cas* genes were identified using a combination of BLAST and CRISPRdb (Grissa et al 2007a).

CRISPR repeats on contigs with *cas* genes were searched using BLAST against all metagenomic samples to identify contigs with valid CRISPRs that contain valid spacer sequences, but lacked *cas* genes due to incomplete assembly. *Cas* genes were identified using BLASTp against a custom database comprising known *cas* genes from NCBI's RefSeq and *cas* genes described by Makarova et al. (Makarova et al 2011c).

Data access

The sequence data from this study has been submitted to NCBI (http://www.ncbinlm.nih.gov/bioproject) under BioProject identifier PRJNA72255. Individual accession numbers for reads from the 14 metagenomic samples deposited in NCBI's Sequence Read Archive can be found in Supplemental Table 4.5 in Appendix C. Accession numbers for assembled and annotated sequences from MIS can be found in Supplemental Table 4.6 in Appendix C. Accession numbers for non-MIS *NblA* sequences can be found in Supplemental Table 4.7 in Appendix C.

Abbreviations used

Middle Island Sinkhole (MIS)

CRISPR (clustered regularly interspaced short palindromic repeats)

CRISPR/*cas* (CRISPR-associated sequences)

auxiliary metabolic genes (AMGs)

single nucleotide polymorphism (SNP)

Non-metric multidimensional scaling (NMDS)

Competing interests

The authors declare that they have no competing interests.

Author's contributions

AV and GD conceived the study. AV carried out sampling and sample prep for sequencing, carried out data analysis and drafted the manuscript. SE carried out the multi-dimensional analysis and helped draft the manuscript. DM carried out NblA analysis, assisted in sample collection and helped draft the manuscript. MD participated in viral analysis and helped draft the manuscript. JC participated in data analysis and helped draft the manuscript. BB coordinated sampling, assisted in sampling and helped draft the manuscript. GD assisted in sample collection, participated in the analysis and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

The following additional data are available with the online version of this paper.

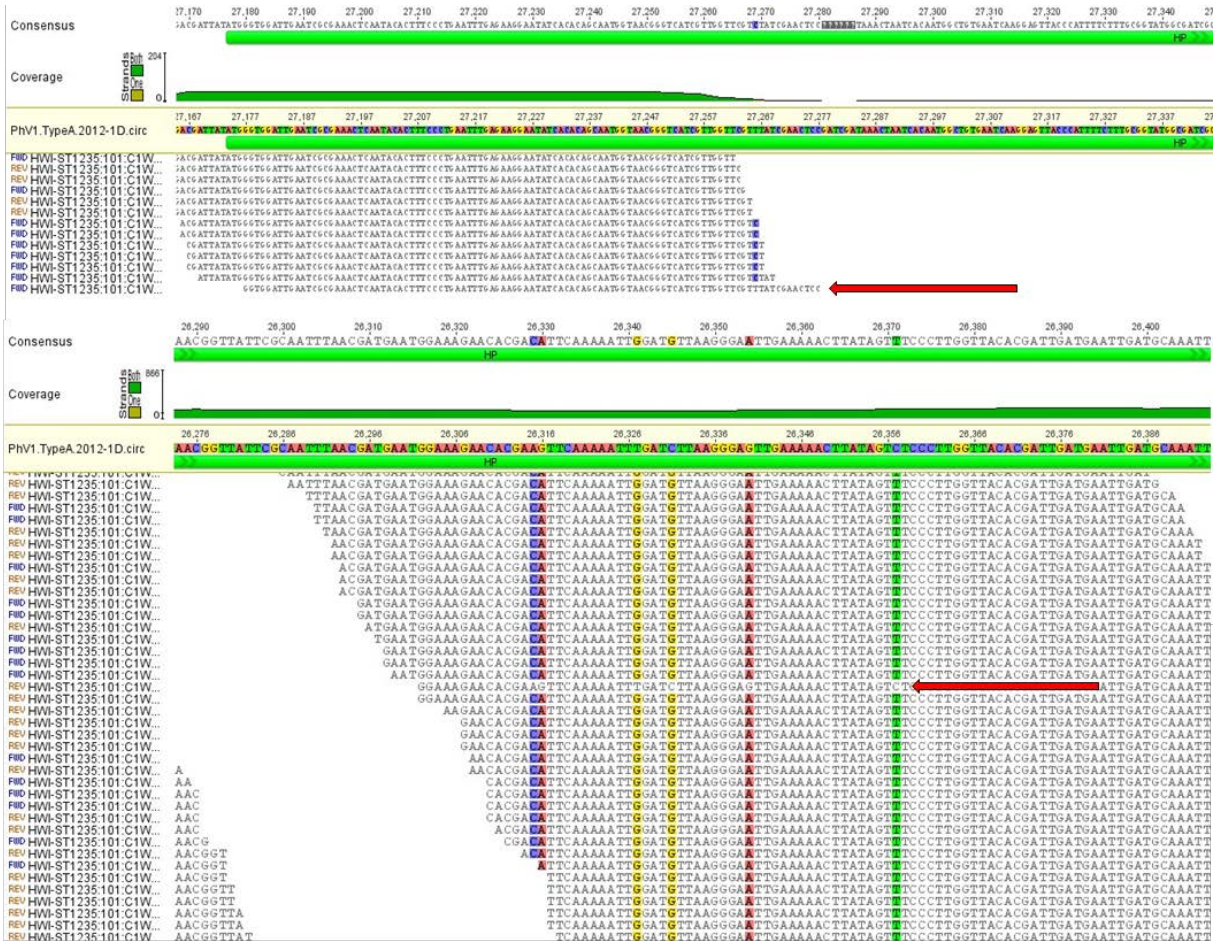
Appendix C. Supplemental Figures 4.1-4.10 with legends and Supplemental Tables 4.1-4.7

Acknowledgments

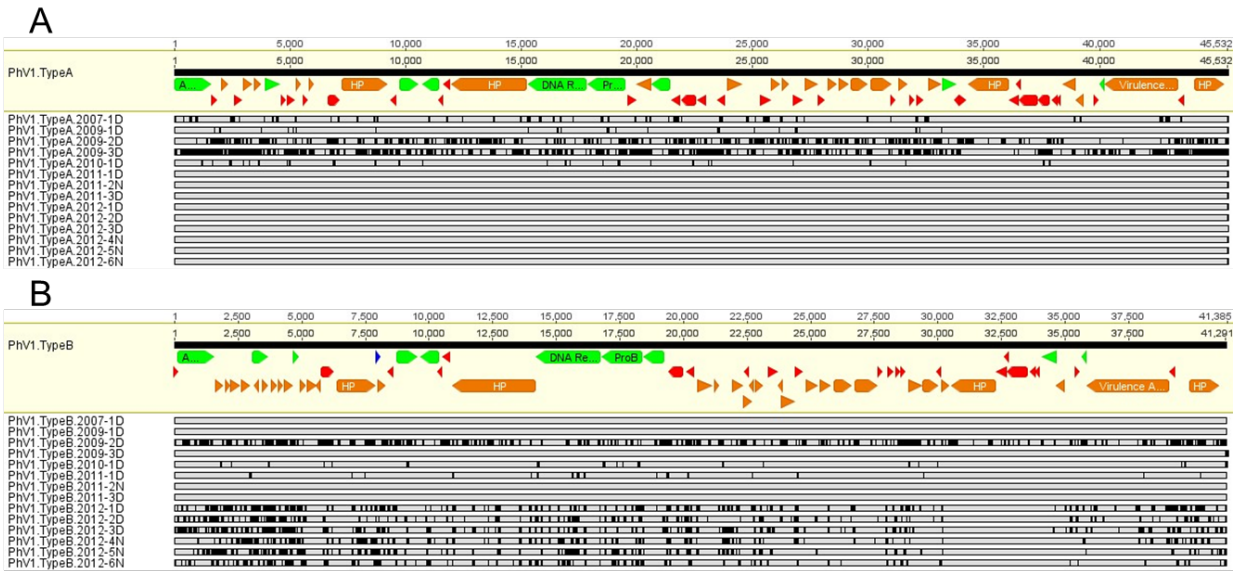
We would like to thank NOAA, Russ Green, and the Thunder Bay National Marine Sanctuary for sampling and logistical assistance, the University of Michigan DNA Sequencing Core for DNA sequencing, and Sunit Jain for assembly strategy and bioinformatics help. This work was supported by NSF grant EAR1035955 to GJD, EAR 1035957 to BAB, the University of Michigan CCMB Pilot Grant to JDC and GJD, and the Scott Turner Award to AAV.

4.6 Appendix C

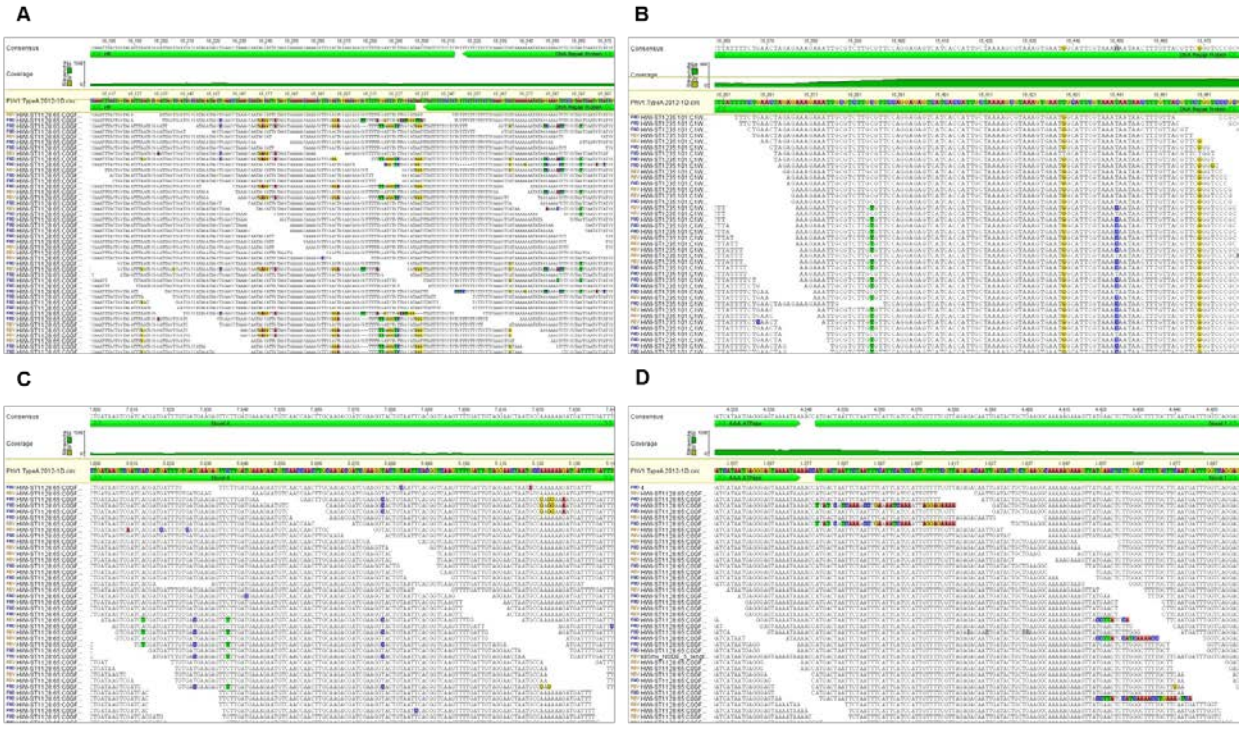
CHAPTER IV Supplemental Materials



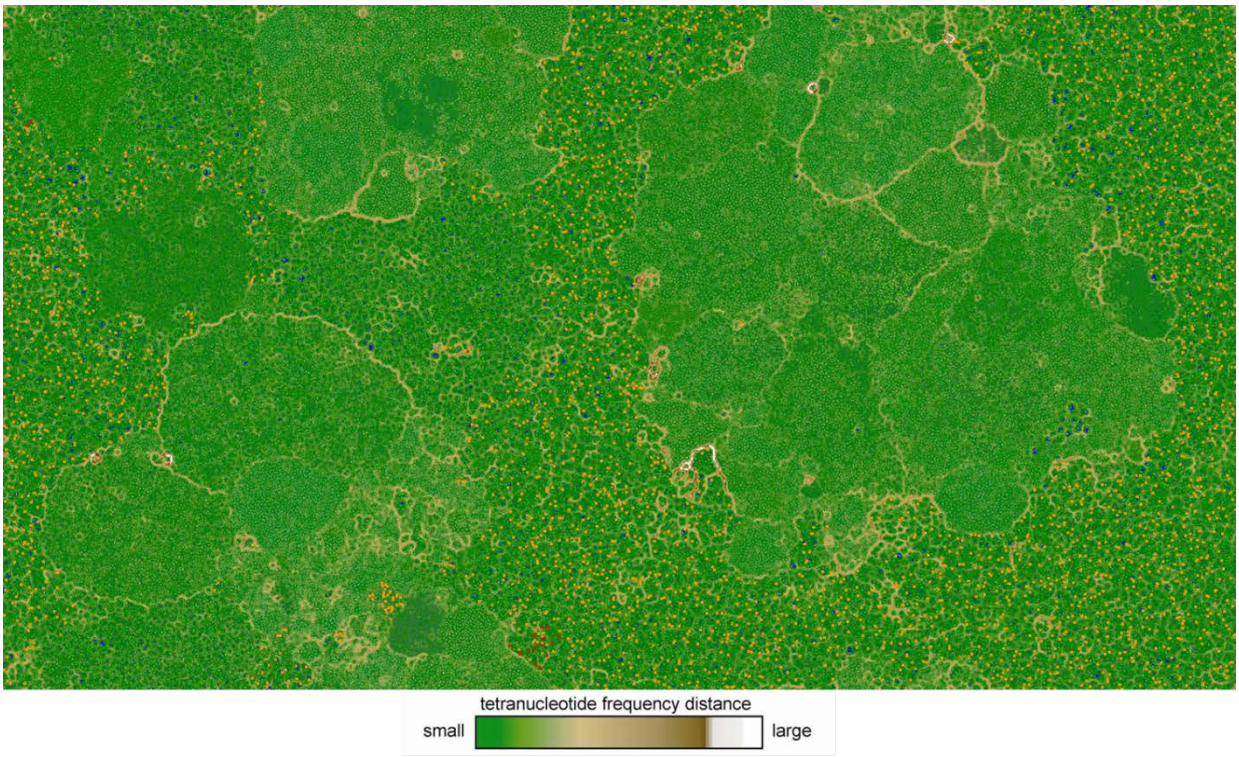
Supplemental Figure 4.1 Screen captures from Geneious showing reads mapped to PhV1.TypeA. (Highlighted bases) denote deviation from the template sequence; (red arrow) denotes low abundance genomic variant matching the template.



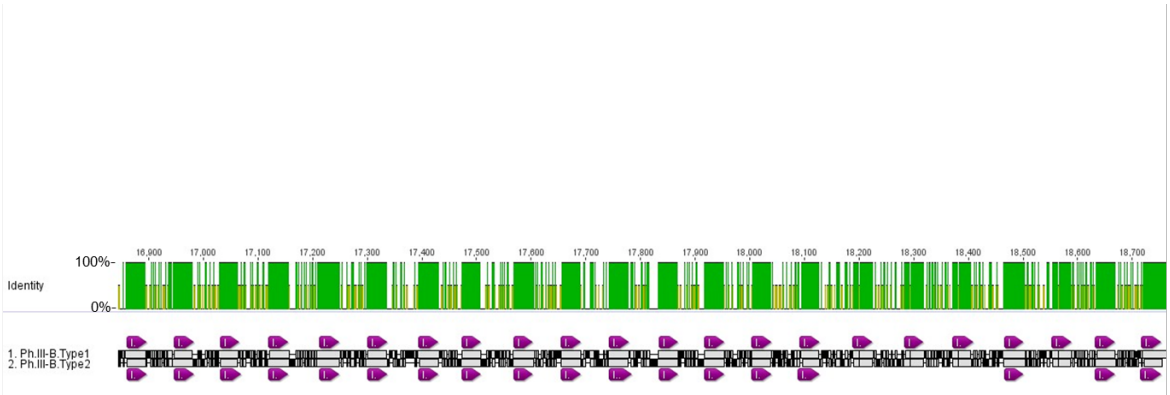
Supplemental Figure 4.2 Reads from each sample were mapped to PhV1 genotypes and consensus sequences show genome completeness for each sample. (A) reads mapped to PhV1.TypeA; (B) reads mapped to PhV1.TypeB. (Green) genes which could be annotated; (orange) hypothetical genes present in NR with no annotation; (blue) region that matches a CRISPR III-B spacer; (grey) regions with read coverage; (black) regions lacking read coverage.



Supplemental Figure 4.3 Screen captures from Geneious showing multiple genomic variants whos reads mapped to PhV1.TypeA. (A-C) Multiple SNP patterns; (D) insertions or deletions. (Highlighted bases) Denote deviation from the template sequence.

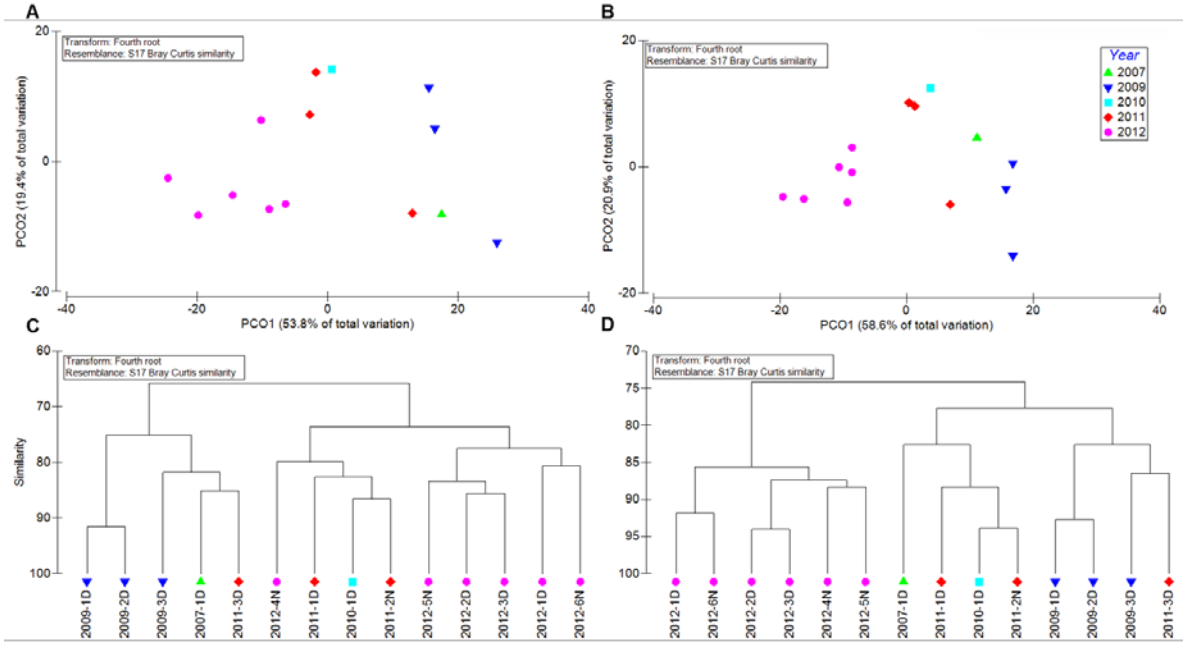


Supplemental Figure 4.4 Self organizing ESOM map showing *Phormidium* and PhV1 share a tetranucleotide frequency. 5kb sequences (represented by dots) with similar tetranucleotide frequencies are located closer together. Note the right and left edges, as well as the top and bottom edges are contiguous. (Orange dots) sequences from *Phormidium*; (red dots) sequences from PhV1.



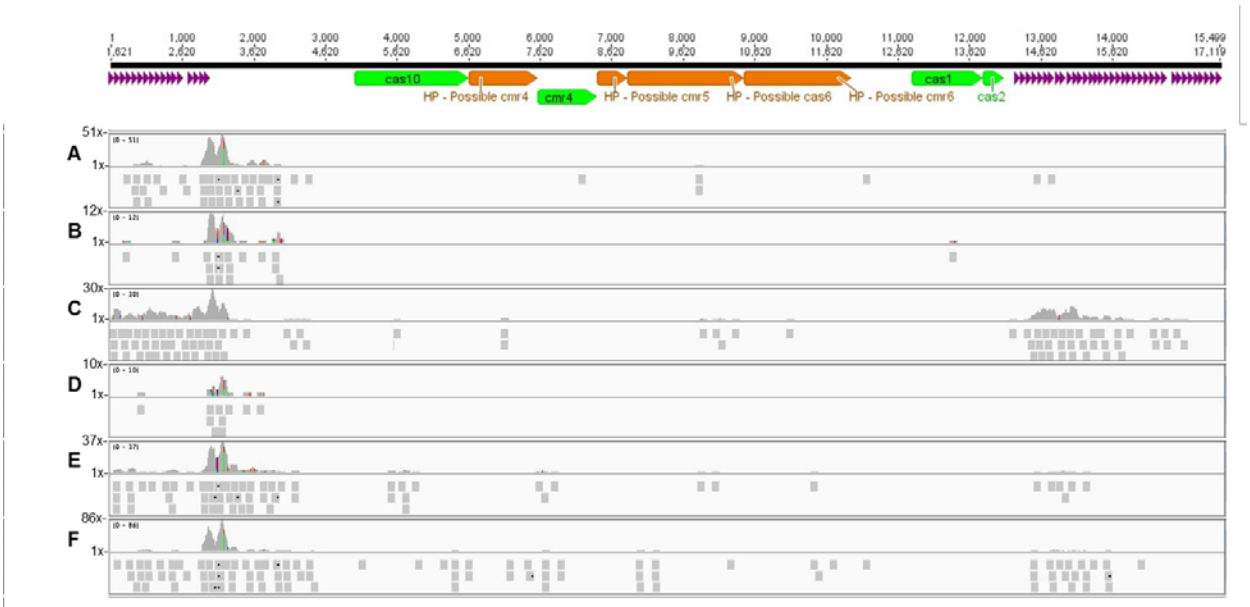
Supplemental Figure 4.5 Nucleotide alignment of the Phormidium Type III-B CRISPR region from Type1 and Type2. (Purple) CRISPR repeats with spacers located between repeats; (grey) areas of nucleotide agreement; (black) nucleotide disagreement.

Supplemental Figure 4.6 Nucleotide alignment of Phormidium CRISPR subtype III-B Locus1 showing sequence conservation in each of the 14 samples. (Green) genes which could be annotated; (orange) hypothetical genes present in NR with no annotation; (purple) CRISPR repeats with spacers located between repeats; (grey) regions of nucleotide agreement.

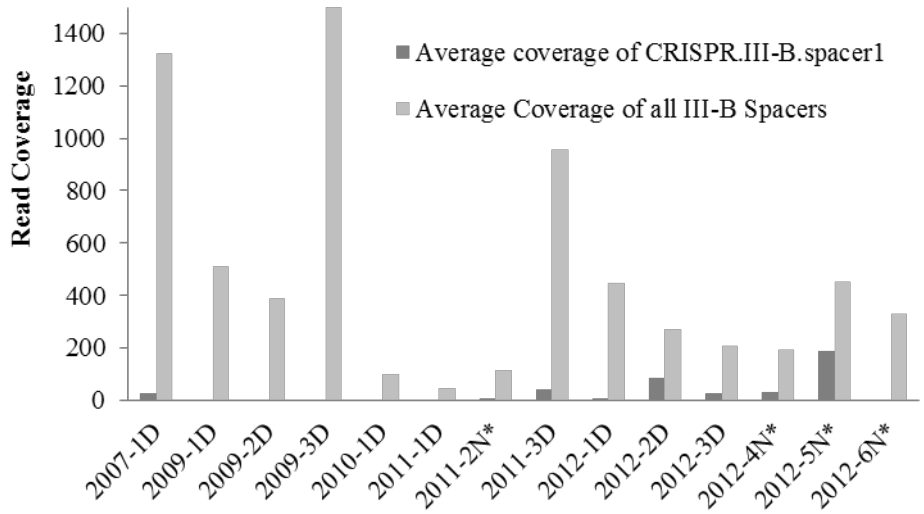


Supplemental Figure 4.7 Multivariate analysis of CRISPR spacer contigs and individual spacers. (A) Principal coordinate analyses of CRISPR spacer contigs by year; (B) principal coordinate analyses of individual CRISPR spacers by year; (C) group average analysis of CRISPR spacer contigs by year: (D) group average analysis of individual CRISPR spacers by year.

Supplemental Figure 4.8 cDNA from 2012 samples mapped to PhV1. Note that peaks range from 1x-65x coverage. (A) PhV1.TypeA; (B) PhV1.TypeB. (Green) genes which could be annotated; (orange) genes with no known function; (red) novel genes; (blue) region that matches a CRISPR III-B spacer.



Supplemental Figure 4.9 cDNA from 2012 samples mapped to CRISPR Locus1 . (A) Sample 2012-1D; (B) sample 2012-2D; (C) sample 2012-3D; (D) sample 2012-4N; (E) sample 2012-5N; (F) sample 2012-6N; (Green) genes which could be annotated; (orange) hypothetical genes; (purple) CRISPR repeats, with spacers located between repeats. (Grey boxes under peaks) individual reads with window cut off at 3x read depth.



Supplemental Figure 4.10 Normalized read abundance of CRISPR type III-B spacers averaged over the length of the spacer.

Supplemental Table 4.1. PhV1.TypeA Annotations

Start	Stop	Length	Annotation	Direction
1	1,575	1,575	AAA ATPase	forward
1,580	1,873	294	Novel 1	forward
2,017	2,319	303	HP	forward
2,600	2,932	333	Novel 2	forward
2,952	3,350	399	HP	forward
3,429	3,764	336	HP	forward
3,927	4,571	645	DNA Methyltransferase	forward
4,652	4,816	165	Novel 3	forward
4,860	5,210	351	Novel 4	forward
5,250	5,483	234	HP	forward
5,568	5,774	207	Novel 5	forward
5,823	6,041	219	HP	forward
6,643	7,149	507	Novel 6	forward
7,256	9,217	1,962	HP	forward
9,329	9,601	273	Novel 7	reverse
9,728	10,558	831	DNA Ligase	forward
10,713	11,459	747	Nucleoid DNA Binding Protein	reverse
11,456	11,614	159	Novel 8	reverse
11,614	11,919	306	Novel 9	reverse
11,989	15,270	3,282	HP	reverse
15,274	17,829	2,556	DNA Repair Protein	reverse
17,865	19,526	1,662	ProB	reverse
19,596	19,967	372	Novel 10	forward
19,978	20,634	657	HP	reverse
20,638	21,453	816	Methyltransferase	reverse
21,516	21,878	363	Novel 11	reverse

21,946	22,566	621	Novel 12	reverse
22,623	23,015	393	Novel 13	reverse
23,497	23,823	327	Novel 14	reverse
23,925	24,545	621	HP	forward
25,327	25,809	483	Novel 15	forward
25,806	26,171	366	HP	forward
26,267	26,581	315	HP	forward
26,768	27,172	405	Novel 16	forward
27,287	27,838	552	HP	forward
27,841	28,140	300	Novel 17	forward
28,256	28,645	390	HP	forward
28,723	29,172	450	HP	forward
29,266	30,003	738	HP	forward
30,110	31,009	900	Hypothetical Phage Gene	forward
31,006	31,209	204	Novel 18	forward
31,302	31,688	387	HP	forward
31,806	32,021	216	Novel 19	forward
32,098	32,403	306	Novel 20	forward
32,629	33,183	555	HP	forward
33,193	33,801	609	HP - Possible Peptidoglycan Binding Protein	forward
33,796	33,906	111	Novel 21	reverse
33,939	34,262	324	Novel 22	forward
34,351	36,096	1,746	HP	reverse
36,093	36,515	423	Novel 23	reverse
36,472	36,579	108	Novel 24	reverse
36,524	37,327	804	Novel 25	reverse
37,358	37,855	498	Novel 26	reverse
37,946	38,191	246	Novel 27	reverse
38,192	38,311	120	Novel 28	reverse
38,431	38,994	564	HP	reverse
38,975	39,334	360	HP	reverse
39,740	39,979	240	Novel 29	forward
40,058	40,219	162	NblA	reverse
40,230	43,418	3,189	Virulence Associated E Protein	reverse
43,415	43,666	252	Novel 30	reverse
44,133	45,413	1,281	HP	forward

Supplemental Table 4.2 PhV1.TypeB Annotations

Start	Stop	Length	Name	Direction
1	161	161	Novel 31	forward
106	1,575	1,470	AAA ATPase	forward
1,597	1,923	327	HP	forward
1,999	2,196	198	HP - Possible Hydrolase	forward
2,207	2,599	393	HP	forward
2,628	2,989	362	HP	forward

3,058	3,696	639	DNA Methyltransferase	forward
3,161	3,340	180	HP	reverse
3,430	3,696	267	HP	forward
3,804	4,037	234	HP	forward
4,086	4,286	201	HP	forward
4,322	4,645	324	HP	forward
4,642	4,878	237	HNH Endonuclease	forward
4,936	5,170	235	HP	forward
5,217	5,588	372	HP	forward
5,592	5,750	159	HP	reverse
5,768	6,274	507	Novel 6	forward
6,381	7,910	1,530	HP	forward
7,994	8,281	288	HP - Possible Membrane Protein or Specificity Protein	forward
8,376	8,627	252	Novel 7	reverse
8,730	9,560	831	DNA Ligase	forward
9,679	10,398	720	Nucleoid DNA Binding Protein	reverse
10,395	10,550	156	Novel 8	reverse
10,550	10,855	306	Novel 9	reverse
10,925	14,206	3,282	HP	reverse
14,210	16,768	2,559	DNA Repair Protein	reverse
16,804	18,396	1,593	ProB	reverse
18,460	19,275	816	Methyltransferase	reverse
19,437	20,015	579	Novel 12	reverse
20,107	20,433	327	Novel 14	reverse
20,537	21,157	621	HP	forward
21,284	21,391	108	HP	forward
21,925	22,354	430	HP - Phage Protein 7.7 Like	forward
22,351	22,716	366	HP	forward
22,413	22,577	165	Novel 32	reverse
22,574	22,771	198	HP	reverse
22,832	23,146	315	HP	forward
23,334	23,738	405	Novel 16	forward
23,735	23,911	177	HP	reverse
23,852	24,403	552	HP	forward
24,406	24,705	300	Novel 17	forward
24,821	25,282	462	HP	forward
25,360	25,809	450	HP	forward
25,903	26,640	738	HP	forward
26,733	27,632	900	HP - Possible gp37/gp68 Phage Gene	forward
27,629	27,832	204	Novel 18	forward
28,039	28,254	216	Novel 19	forward
28,344	28,532	189	Novel 20	forward
28,594	28,701	108	CDS	forward
28,852	29,406	555	HP - Possible Endonuclease	forward
29,416	30,024	609	HP - Possible Peptidoglycan Binding Protein	forward
30,019	30,114	96	Novel 21	reverse
30,147	30,470	324	HP	forward
30,559	32,304	1,746	HP	reverse
32,301	32,723	423	Novel 23	reverse
32,680	32,787	108	Novel 24	reverse

32,732	33,535	804	Novel 25	reverse
33,626	33,862	237	Novel 27	reverse
33,863	33,982	120	Novel 28	reverse
34,111	34,674	564	Nuclease	reverse
34,655	35,014	360	HP	reverse
35,419	35,547	129	Novel 29	forward
35,684	35,845	162	NblA	reverse
35,856	39,104	3,249	Virulence Associated E Protein	reverse
39,101	39,352	252	Novel 30	reverse
39,908	41,080	1,173	HP	forward

Supplemental Table 3. III-B CRISPR repeats

	Repeat Sequence
III-B.Repeat.A	CTCCCCACTCGTTGGGGAAACTAATTGAATGGAAAC
III-B.Repeat.B	CTCCCCACTCGCTGGGGAAACTAATTGAATGGAAAC
III-B.Repeat.C	GTCGCTTCTATTCGTAGAAGTGGAATTAATGGAAAC
III-B.Repeat.D	GTCCCCACTCGCTGGGGAAACTAATTGAATGGAAAC
III-B.Repeat.E	CCCTACCGATGGGTTTAAATCGGATTAGTTGGAAAC

Supplemental Table 4. Parameters used for bioinformatics software

Program Name: Geneious v6.1.5

Function	Algorythem	Cost Matrix	Gap Open	Gap extend	Type	Iterations
Multiple Alignment	Geneious Alignment	93%	12	3	Global alignment with free end gaps	3

Function	Sensitivity	Trim	Max Gap	Max Gap Size	Min Overlap	Overlap ID	Max Mismatches
Map to Reference	Custom	None	10%	10bp	60	100%	2%

Function	Sensitivity	Trim	Max Gap	Max Gap Size	Min Overlap	Overlap ID	Max Mismatches
De Novo Assemble	Custom	None	10%	10bp	60	100%	2%

Program Name: Velvet v1.2.08

Function	kmer values	Read Type
Assembly	91, 75, 61	Paired

Program Name: MetaVelvet v1.2.02

Function	Paramaters
Correct Assembly	Default

Program Name: BWA v0.7.4

Function	Paramaters
sampe	Default

Program Name: Samtools v0.1.19

Function	OVERLAP	MINID
minimus2	200bp	98%

Program Name: BLAST v2.2.27

Function	Output format	Min ID	Min Bitscore	e value	Target
blastp	6	60%	50	1.00E-05	Function

Function	Output format	Min ID	Min Bitscore	e value	Target
blastp	6	30%	50	1.00E-05	Homology

Supplemental Table 5. Sample read accession information

Sample Alias	Sample Title	BioProject ID	BioSample ID	Accession #
MISgDNA_2007-1D	Finger 2007-1D gDNA	PRJNA72255	SAMN02228699	SRS455884
MISgDNA_2009-1D	Prostrate Mat 2009-1D gDNA	PRJNA72255	SAMN02228700	SRS455885
MISgDNA_2009-2D	Finger 2009-2D gDNA	PRJNA72255	SAMN02228701	SRS455886
MISgDNA_2009-3D	Finger 2009-3D gDNA	PRJNA72255	SAMN02230180	SRS455906
MISgDNA_2010-1D	Finger 2010-1D gDNA	PRJNA72255	SAMN02228702	SRS455888
MISgDNA_2011-1D	Prostrate Mat 2011-1D gDNA	PRJNA72255	SAMN02228703	SRS455889
MISgDNA_2011-2N	Prostrate Mat 2011-2N gDNA	PRJNA72255	SAMN02228704	SRS455890
MISgDNA_2011-3D	Finger 2011-3D gDNA	PRJNA72255	SAMN02228705	SRS455891
MISgDNA_2012-1D	Finger 2012-1D gDNA	PRJNA72255	SAMN02228706	SRS455892
MISgDNA_2012-2D	Finger 2012-2D gDNA	PRJNA72255	SAMN02228707	SRS455893
MISgDNA_2012-3D	Finger 2012-3D gDNA	PRJNA72255	SAMN02228708	SRS455894
MISgDNA_2012-4N	Finger 2012-4N gDNA	PRJNA72255	SAMN02228709	SRS455895
MISgDNA_2012-5N	Finger 2012-5N gDNA	PRJNA72255	SAMN02228710	SRS455896
MISgDNA_2012-6N	Finger 2012-6N gDNA	PRJNA72255	SAMN02228711	SRS455898
MIScDNA_2012-1D	Finger 2012-1D cDNA	PRJNA72255	SAMN02230071	SRS455900
MIScDNA_2012-2D	Finger 2012-2D cDNA	PRJNA72255	SAMN02230072	SRS455901
MIScDNA_2012-3D	Finger 2012-3D cDNA	PRJNA72255	SAMN02230073	SRS455902
MIScDNA_2012-4N	Finger 2012-4N cDNA	PRJNA72255	SAMN02230074	SRS455903
MIScDNA_2012-5N	Finger 2012-5N cDNA	PRJNA72255	SAMN02230075	SRS455904
MIScDNA_2012-6N	Finger 2012-6N cDNA	PRJNA72255	SAMN02230076	SRS455905

Supplemental Table 6. Sequence accession information

Sequence Title	BioProject ID	Accession #
Phormidium Phage MIS-PhV1.TypeA	PRJNA72255	KF437907
Phormidium Phage MIS-PhV1.TypeB	PRJNA72255	KF437908
Spacer Contigs 1-33	PRJNA72255	KF487036 - KF487068
CRISPR III-B Type1	PRJNA72255	KF487069
CRISPR III-B Type1 v2	PRJNA72255	KF487070
CRISPR III-B Type2	PRJNA72255	KF487071
L2P	PRJNA72255	KF487072
L4P	PRJNA72255	KF487073
ychF	PRJNA72255	KF487074
secY	PRJNA72255	KF487075
nblA	PRJNA72255	KF487076 - KF487081

Sequence Title	BioProject ID	BioSample ID
2007-2012 Metagenomes	PRJNA72255	SAMN02231594
Phormidium sp. MIS-PhA	PRJNA72255	SAMN02265253

Supplemental Table 7. Accession numbers for NblA tree

nblA source	NCBI accession number
Bacillus subtilis BEST7613	BAM51709.1
Microcystis aeruginosa PCC 7806	CAO88905.1
Arthrospira sp. PCC 8005	CCE19972.1
Synechocystis sp. PCC 6803	BAA17955.1
Synechococcus elongatus PCC 7942	AAA89103.1
Microcystis aeruginosa DIANCHI905	ELS45963.1
Microcystis aeruginosa TAIHU98	ELP52921.1
Microcystis aeruginosa PCC 9701	CCI36768.1
Microcystis sp. T1-4	CCI31263.1
Planktothrix phage PaV-LD	ADZ31529.1
Microcystis phage Ma-LMM01	YP_851019
Phormidium Phage PhV1.TypeA	KF437907
Phormidium Phage PhV1.TypeB	KF437908
MIS Phormidium 1	KF487076
MIS Phormidium 2	KF487077
MIS Phormidium 3	KF487078
MIS Phormidium 4	KF487079
MIS Oscellatoria 1	KF487080
MIS Oscellatoria 2	KF487081

4.7 References

Allen EE, Banfield JF (2005). Community genomics in microbial ecology and evolution. *Nature Reviews Microbiology* **3:** 489-498.

Allers E, Moraru C, Duhaime MB, Beneze E, Solonenko N, Barrero-Canosa J *et al* (2013). Single-cell and population level viral infection dynamics revealed by phageFISH, a method to visualize intracellular and free viruses. *Environ Microbiol* **15:** 2306-2318.

Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW (2006). Stromatolite reef from the Early Archaean era of Australia. *Nature* **441:** 714-718.

Anbar AD, Knoll AH (2002). Proterozoic ocean chemistry and evolution: A bioinorganic bridge? *Science* **297:** 1137-1142.

Andersen DT, Sumner DY, Hawes I, Webster-Brown J, Mckay CP (2011). Discovery of large conical stromatolites in Lake Untersee, Antarctica. *Geobiology* **9:** 280-293.

Anderson MJ (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology* **26:** 32-46.

Andersson AF, Banfield JF (2008). Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320:** 1047-1050.

Arieli B, Shahak Y, Taglicht D, Hauska G, Padan E (1994a). Purification and Characterization of Sulfide-Quinone Reductase, a Novel Enzyme Driving Anoxygenic Photosynthesis in Oscillatoria-Limnetica. *Journal of Biological Chemistry* **269:** 5705-5711.

Arieli B, Shahak Y, Taglicht D, Hauska G, Padan E (1994b). Purification and characterization of sulfide-quinone reductase, a novel enzyme driving anoxygenic photosynthesis in *Oscillatoria limnetica*. *Journal of Biological Chemistry* **269:** 5705-5711.

Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D (2011). Genomic island variability facilitates Prochlorococcus-virus coexistence. *Nature* **474:** 604-608.

Bachar A, Omoregie E, de Wit R, Jonkers HM (2007). Diversity and function of Chloroflexus-like bacteria in a hypersaline microbial mat: phylogenetic characterization and impact on aerobic respiration. *Appl Environ Microbiol* **73:** 3975-3983.

Banfield JF, Young M (2009). Variety-the Splice of Life-in Microbial Communities. *Science* **326:** 1198-1199.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S *et al* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709-1712.

Bekker A, Holland HD, Wang PL, Rumble D, Stein HJ, Hannah JL *et al* (2004). Dating the rise of atmospheric oxygen. *Nature* **427:** 117-120.

Belkin S, Padan E (1978). Hydrogen metabolism in the facultative anoxygenic Cyanobacteria (Blue-Green Algae) *Oscillatoria limnetica* and *Aphanothece halophytica*. *Archives of Microbiology* **116:** 109-111.

Bertrand EM, Saito MA, Rose JM, Riesselman CR, Lohan MC, Noble AE *et al* (2007). Vitamin B-12 and iron colimitation of phytoplankton growth in the Ross Sea. *Limnology and Oceanography* **52:** 1079-1093.

Biddanda BA, Opsahl S, Benner R (1994). Plankton respiration and carbon flux through bacterioplankton on the Louisiana Shelf. *Limnology and Oceanography* **39:** 1259-1275.

Biddanda BA, Coleman DF, Johengen TH, Ruberg SA, Meadows GA, VanSumeran HW *et al* (2006). Exploration of a submerged sinkhole ecosystem in Lake Huron. *Ecosystems* **9:** 828-842.

Biddanda BA, Nold SC, Ruberg SA, Kendall ST, Sanders TG, Gray JJ (2009). Great lakes sinkholes: a microbiogeochemical frontier. *Eos* **90:** 61-68.

Biddanda BA, Nold SC, Dick GJ, Kendall ST, Vail JH, Ruberg SA *et al* (in press). Rock, water, microbes: Underwater sinkholes in Lake Huron are habitats for ancient microbial life. *Nature Education*.

Biomatters (2013). Geneious, 6.1.5 edn.

Black TJ (1983). Selected views of the tectonics, structure and karst in northern lower Michigan. In: Kimmel RE (ed). *Michigan Basin Geological Society Field Conference Proceedings*. Michigan Basin Geological Society: Lansing, MI. pp 11-35.

Blank CE (2004). Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2:** 1-20.

Blankenship R, Sadekar S, Raymond J (2007). The evolutionary transition from anoxygenic to oxygenic photosynthesis. In: Falkowski PG, Knoll AH (eds). *Evolution of primary producers in the sea*. Elsevier. pp 21-35.

Blenkinsopp SA, Costerton JW (1991). Understanding bacterial biofilms. *Trends in Biotechnology* **9:** 138-143.

Bosak T, Newman DK (2003). Microbial nucleation of calcium carbonate in the Precambrian **31:** 577-580.

Breitbart M, Thompson LR, Suttle CA, Sullivan MB (2007). Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20:** 135-139.

Bronstein M, Schutz M, Hauska G, Padan E, Shahak Y (2000a). Cyanobacterial sulfide-quinone reductase: cloning and heterologous expression. *J Bacteriol* **182:** 3336-3344.

Bronstein M, Schutz M, Hauska G, Padan E, Shahak Y (2000b). Cyanobacterial sulfide-quinone reductase: Cloning and heterologous expression. *J Bacteriol* **182:** 3336-3344.

Buhring SI, Sievert SM, Jonkers HM, Ertefai T, Elshahed MS, Krumholz LR *et al* (2011). Insights into chemotaxonomic composition and carbon cycling of phototrophic communities in an artesian sulfur-rich spring (Zodletone, Oklahoma, USA), a possible analog for ancient microbial mat systems. *Geobiology* **9:** 166-179.

Canfield DE, Des Marais DJ (1993). Biogeochemical Cycles of Carbon, Sulfur, and Free Oxygen in a Microbial Mat. *Geochim Cosmochim Ac* **57:** 3971-3984.

Canfield DE, Farquhar J (2009). Animal evolution, bioturbation, and the sulfate concentration of the oceans. *Proc Natl Acad Sci U S A* **106:** 8123-8127.

Casamayor EO, Garcia-Cantizano J, Pedros-Alio C (2008). Carbon dioxide fixation in the dark by photosynthetic bacteria in sulfide-rich stratified lakes with oxic-anoxic interfaces. *Limnology and Oceanography* **53:** 1193-1203.

Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu DY *et al* (2013). Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nature Communications* **4**.

Castenholz RW (1976). Effect of sulfide on blue-green algae of hot springs .1. New-Zealand *Journal of Phycology* **12:** 54-68.

Castenholz RW (1977a). Effect of sulfide on blue-green algae of hot springs .2. Yellowstone National Park. *Microbial Ecology* **3:** 79-105.

Castenholz RW (1977b). Effect of sulfide on blue-green-algae of hot springs .2. Yellowstone-national-park. *Microbial Ecology* **3:** 79-105.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Muller WEG, Wetter T *et al* (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. *Genome Research* **14:** 1147-1159.

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, Desantis TZ *et al* (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322:** 275-278.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283-1287.

Clarke K, Gorley, RN (2006). Primer v6: User Manual/Tutorial, v6 edn. PRIMER-E: Plymouth.

Clarke KR (1993). Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18:** 117-143.

Cohen Y, Jorgensen BB, Padan E, Shilo M (1975a). Sulphide-dependent anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. *Nature* **257:** 489-492.

Cohen Y, Padan E, Shilo M (1975b). Facultative anoxygenic photosynthesis in the cyanobacterium Oscillatoria limnetica. *J Bacteriol* **123:** 855-861.

Cohen Y, Padan E, Shilo M (1975c). Facultative anoxygenic photosynthesis in cyanobacterium oscillatoria-limnetica. *J Bacteriol* **123:** 855-861.

Cohen Y, Jorgensen BB, Revsbech NP, Poplawski R (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl Environ Microbiol* **51:** 398-407.

Colon-Lopez MS, Sherman LA (1998). Transcriptional and translational regulation of photosystem I and II genes in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. *J Bacteriol* **180:** 519-526.

Comte K, Sabacka M, Carre-Mlouka A, Elster J, Komarek J (2007). Relationships between the Arctic and the Antarctic cyanobacteria; three Phormidium-like strains evaluated by a polyphasic approach. *Fems Microbiology Ecology* **59:** 366-376.

Cowan DA, Tow LA (2004). Endangered antarctic environments. *Annu Rev Microbiol* **58:** 649-690.

Crowe SA, Dossing LN, Beukes NJ, Bau M, Kruger SJ, Frei R *et al* (2013). Atmospheric oxygenation three billion years ago. *Nature* **501:** 535-+.

Denef VJ, Mueller RS, Banfield JF (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *Isme Journal* **4:** 599-610.

DeRuyter YS, Fromme P (2008). Molecular structure of the photosynthetic apparatus. In: Herrero A, Flores E (eds). *The Cyanobacteria, molecular biology, genomics, and evolution*. Caister Academic Press: Norfolk, UK. pp 217-269.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Thomas BC, Yelton AP *et al* (2009a). Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.

Dick GJ, Andersson AF, Baker BJ, Simmons SL, Yelton AP, Banfield JF (2009b). Community-wide analysis of microbial genome sequence signatures. *Genome Biology* **10**.

Dietrich LEP, Teal TK, Price-Whelan A, Newman DK (2008). Redox-active antibiotics control gene expression and community behavior in divergent bacteria. *Science* **321:** 1203-1206.

Duhaime MB, Deng L, Poulos BT, Sullivan MB (2012). Towards quantitative metagenomics of wild viruses and other ultra-low concentration DNA samples: a rigorous assessment and optimization of the linker amplification method. *Environ Microbiol* **14:** 2526-2537.

Dupraz C, Reid RP, Braissant O, Decho AW, Norman RS, Visscher PT (2009). Processes of carbonate precipitation in modern microbial mats. *Earth-Science Reviews* **96:** 141-162.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5:** 1-19.

Emerson JB, Thomas BC, Andrade K, Allen EE, Heidelberg KB, Banfield JF (2012). Dynamic Viral Populations in Hypersaline Systems as Revealed by Metagenomic Assembly. *Appl Environ Microbiol* **78:** 6309-6320.

Falkowski PG, Fenchel T, Delong EF (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320:** 1034-1039.

Farquhar J, Zerkle AL, Bekker A (2011). Geological constraints on the origin of oxygenic photosynthesis. *Photosynth Res* **107:** 11-36.

Fontes MLS, Suzuki MT, Cottrell MT, Abreu PC (2011). Primary Production in a Subtropical Stratified Coastal Lagoon-Contribution of Anoxygenic Phototrophic Bacteria. *Microbial Ecology* **61:** 223-237.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105:** 3805-3810.

Frigaard N-U, Dahl C, Robert KP (2008). Sulfur Metabolism in Phototrophic Sulfur Bacteria. *Advances in Microbial Physiology*. Academic Press. pp 103-200.

Gao EB, Gui J-F, Zhang Q-Y (2012). A Novel Cyanophage with a Cyanobacterial Nonbleaching Protein A Gene in the Genome. *Journal of Virology* **86:** 236-245.

Garcia-Pichel F, Mechling M, Castenholz RW (1994). Diel migrations of microorganisms within a benthic, hypersaline mat community. *Appl Environ Microbiol* **60:** 1500-1511.

Garlick S, Oren A, Padan E (1977a). Occurrence of facultative anoxygenic photosynthesis among filamentous and unicellular cyanobacteria. *J Bacteriol* **129:** 623-629.

Garlick S, Oren A, Padan E (1977b). Occurrence of facultative anoxygenic photosynthesis among filamentous and unicellular cyanobacteria. *J Bacteriol* **129:** 623-629.

Garneau JE, Dupuis ME, Villion M, Romero DA, Barrangou R, Boyaval P *et al* (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468:** 67-+.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al* (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *Plos One* **3**.

Gingras M, J.W. Hagadorn, A. Seilacher, S.V. Lalonde, E. Pecoits, D. Petrash, and K.O. Konhauser (2011). Possible evolution of mobile animals in association with microbial mats. *Nature Geoscience.*

Godde JS, Bickerton A (2006). The repetitive DNA elements called CRISPRs and their associated genes: Evidence of horizontal transfer among prokaryotes. *Journal of Molecular Evolution* **62:** 718-729.

Grissa I, Vergnaud G, Pourcel C (2007a). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *Bmc Bioinformatics* **8**.

Grissa I, Vergnaud G, Pourcel C (2007b). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research* **35:** W52-W57.

Grotzinger JP, Knoll AH (1999). Stromatolites in Precambrian carbonates: Evolutionary mileposts or environmental dipsticks? *Annual Review of Earth and Planetary Sciences* **27:** 313-358.

Guindon S, Gascuel O (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52:** 696-704.

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *Plos Computational Biology* **1:** 474-483.

Hawes I, Schwarz AM (1999). Photosynthesis in an extreme shade environment: Benthic microbial mats from Lake Hoare, a permanently ice-covered Antarctic lake. *Journal of Phycology* **35:** 448-459.

Hayes JM, Waldbauer JR (2006). The carbon cycle and associated redox processes through time. *Philosophical Transactions of the Royal Society B-Biological Sciences* **361:** 931-950.

He Y-Y, Hader D-P (2002). Reactive oxygen species and UV-B: effect on cyanobacteria. *Photochemical & Photobiological Sciences* **1:** 729-736.

Heidelberg JF, Nelson WC, Schoenfeld T, Bhaya D (2009). Germ Warfare in a Microbial Mat Community: CRISPRs Provide Insights into the Co-Evolution of Host and Viral Genomes. *Plos One* **4**.

Holland HD (2006). The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci* **361:** 903-915.

Horodyski RJ, Knauth LP (1994). Life on land in the Precambrian. *Science* **263:** 494-498.

Horvath P, Barrangou R (2010). CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science* **327:** 167-170.

Hurwitz BL, Sullivan MB (2013). The Pacific Ocean Virome (POV): A Marine Viral Metagenomic Dataset and Associated Protein Clusters for Quantitative Viral Ecology. *Plos One* **8**.

Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *Bmc Bioinformatics* **11**.

Ignacio-Espinoza JC, Sullivan MB (2012). Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ Microbiol* **14:** 2113-2126.

Imbus SW, Macko SA, Elmore RD, Engel MH (1992). Stable isotope (C, S, N) and molecular studies on the Precambrian Nonesuch Shale (Wisconsin-Michigan, USA) - evidence for differential preservation rates, depositional environment and hydrothermal influence. *Chem Geol* **101:** 255-281.

Ito H, Mutsuda M, Murayama Y, Tomita J, Hosokawa N, Terauchi K *et al* (2009). Cyanobacterial daily life with Kai-based circadian and diurnal genome-wide transcriptional control in Synechococcus elongatus. *Proc Natl Acad Sci U S A* **106:** 14168-14173.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009a). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106:** 16925-16929.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009b). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106:** 16925-16929.

Jorgensen BB, Revsbech NP, Blackburn TH, Cohen Y (1979). Diurnal cycle of oxygen and sulfide microgradients and microbial photosynthesis in a cyanobacterial mat sediment. *Appl Environ Microbiol* **38:** 46-58.

Jorgensen BB, Revsbech NP, Cohen Y (1983). Photosynthesis and structure of benthic microbial mats - microelectrode and SEM studies of 4 cyanobacterial communities. *Limnology and Oceanography* **28:** 1075-1093.

Jorgensen BB, Cohen Y, Revsbech NP (1986a). Transition from anoxygenic to oxygenic photosynthesis in a *Microcoleus chthonoplastes* cyanobacterial mat. *Appl Environ Microbiol* **51:** 408-417.

Jorgensen BB, Cohen Y, Revsbech NP (1986b). Transition from Anoxygenic to Oxygenic Photosynthesis in a Microcoleus-Chthonoplastes Cyanobacterial Mat. *Appl Environ Microbiol* **51:** 408-417.

Kana BD, Weinstein EA, Avarbock D, Dawes SS, Rubin H, Mizrahi V (2001). Characterization of the cydAB-Encoded Cytochrome bd Oxidase from Mycobacterium smegmatis. *J Bacteriol* **183:** 7076-7086.

Kenward PA, Goldstein RH, Gonzalez LA, Roberts JA (2009). Precipitation of low-temperature dolomite from an anaerobic microbial consortium: the role of methanogenic Archaea. *Geobiology* **7:** 556-565.

Komarek J, Kling H, Komarkova J (2003). Filamentous cyanobacteria. In: Wehr JD, Sheath R (eds). *Freshwater algae of North America: Ecology and classification*. Elsevier Science: San Diego, CA, USA. pp 117-196.

Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ (2005). The paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci U S A* **102:** 11131-11136.

Kulkarni G, Wu CH, Newman DK (2013). The General Stress Response Factor EcfG Regulates Expression of the C-2 Hopanoid Methylase HpnP in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **195:** 2490-2498.

Kump LR (2008). The rise of atmospheric oxygen. *Nature* **451:** 277-278.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H *et al* (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23:** 2947-2948.

Li H, Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al* (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078-2079.

Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW (2004). Transfer of photosynthesis genes to and from Prochlorococcus viruses. *Proc Natl Acad Sci U S A* **101:** 11013-11018.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438:** 86-89.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al* (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449:** 83-86.

Litchman E (2000). Growth rates of phytoplankton under fluctuating light. *Freshwater Biology* **44:** 223-235.

Litchman E, Steiner D, Bossard P (2003). Photosynthetic and growth responses of three freshwater algae to phosphorus limitation and daylength. *Freshwater Biology* **48:** 2141-2148.

Liu X, Kong S, Shi M, Fu L, Gao Y, An C (2008). Genomic Analysis of Freshwater Cyanophage Pf-WMP3 Infecting Cyanobacterium Phormidium foveolarum: The Conserved Elements for a Phage. *Microbial Ecology* **56:** 671-680.

Lyons TW, Anbar AD, Severmann S, Scott C, Gill BC (2009). Tracking Euxinia in the Ancient Ocean: A Multiproxy Perspective and Proterozoic Case Study. *Annual Review of Earth and Planetary Sciences* **37:** 507-534.

Ma Y, Paulsen IT, Palenik B (2012). Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ Microbiol* **14:** 453-466.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**.

Makarova KS, Aravind L, Wolf YI, Koonin EV (2011a). Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biology Direct* **6**.

Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P *et al* (2011b). Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* **9:** 467-477.

Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P *et al* (2011c). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology* **9:** 467-477.

Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003). Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* **424:** 741-741.

Mattoo AK, Hoffmanfalk H, Marder JB, Edelman M (1984). Regulation of protein-metabolism - coupling of photosynthetic electron-transport to in-vivo degradation of the rapidly metabolized 32-kilodalton protein of the chloroplast membranes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **81:** 1380-1384.

Mejean A, Mazmouz R, Mann S, Calteau A, Medigue C, Ploux O (2010). The Genome Sequence of the Cyanobacterium Oscillatoria sp. PCC 6506 Reveals Several Gene Clusters Responsible for the Biosynthesis of Toxins and Secondary Metabolites. *J Bacteriol* **192:** 5264-5265.

Middelboe M (2000). Bacterial growth rate and marine virus-host dynamics. *Microbial Ecology* **40:** 114-124.

Miller SR, Bebout BM (2004a). Variation in sulfide tolerance of photosystem II in phylogenetically diverse cyanobacteria from sulfidic habitats. *Appl Environ Microbiol* **70:** 736-744.

Miller SR, Bebout BM (2004b). Variation in sulfide tolerance of photosystem II in phylogenetically diverse cyanobacteria from sulfidic habitats. *Appl Environ Microbiol* **70:** 736-744.

Muyzer G, Stams AJM (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* **6:** 441-454.

Myers JL, Sekar R, Richardson LL (2007). Molecular detection and ecological significance of the cyanobacterial genera Geitlerinema and Leptolyngbya in black band disease of corals. *Appl Environ Microbiol* **73:** 5173-5182.

Myers JL, Richardson LL (2009). Adaptation of cyanobacteria to the sulfide-rich microenvironment of black band disease of coral. *Fems Microbiology Ecology* **67:** 242-251.

Namiki T, Hachiya T, Tanaka H, Sakakibara Y (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Research* **40**.

Noffke N, Christian D, Wacey D, Hazen RM (2013). Microbially Induced Sedimentary Structures Recording an Ancient Ecosystem in the ca. 3.48 Billion-Year-Old Dresser Formation, Pilbara, Western Australia. *Astrobiology* **13:** 1103-1124.

Nold S, Ward D (1996). Photosynthate Partitioning and Fermentation in Hot Spring Microbial Mat Communities. *Appl Environ Microbiol* **62:** 4598-4607.

Nold SC, Pangborn JB, Zajack HA, Kendall ST, Rediske RR, Biddanda BA (2010a). Benthic bacterial diversity in submerged sinkhole ecosystems. *Appl Environ Microbiol* **76:** 347-351.

Nold SC, Pangborn JB, Zajack HA, Kendall ST, Rediske RR, Biddanda BA (2010b). Benthic Bacterial Diversity in Submerged Sinkhole Ecosystems. *Appl Environ Microbiol* **76:** 347-351.

Nold SC, Zajack HA, Biddanda BA (2010c). Eukaryal and archaeal diversity in a submerged sinkhole ecosystem influenced by sulfur-rich, hypoxic groundwater. *Journal of Great Lakes Research* **36:** 366-375.

Nold SC, Zajack HA, Biddanda BA (2010d). Eukaryal and archaeal diversity in a submerged sinkhole ecosystem influenced by sulfur-rich, hypoxic groundwater. *Journal of Great Lakes Research* **36:** 366-375.

Nold SC, Bellecourt MJ, Kendall ST, Ruberg SA, Sanders TG, Klump JV *et al* (2013). Underwater sinkhole sediments sequester Lake Huron's carbon. *Biogeochemistry* **115:** 235-250.

Oren A, Padan E, Avron M (1977). Quantum yields for oxygenic and anoxygenic photosynthesis in the cyanobacterium *Oscillatoria limnetica*. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 2152-2156.

Oren A, Padan E (1978). Induction of anaerobic, photoautotrophic growth in the cyanobacterium *Oscillatoria limnetica*. *Journal of Bacteriology* **133:** 558-563.

Oren A, Shilo M (1979). Anaerobic heterotrophic dark metabolism in the cyanobacterium <i>Oscillatoria limnetica: Sulfur respiration and lactate fermentation. *Arch Microbiol* **122:** 77-84.

Overmann J, Beatty JT, Hall KJ, Pfennig N, Northcote TG (1991). Characterization of a dense, purple sulfur bacterial layer in a meromictic salt lake. *Limnology and Oceanography* **36:** 846-859.

Paez-Espino D, Morovic W, Sun CL, Thomas BC, Ueda K-i, Stahl B *et al* (2013). Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nature Communications* **4**.

Pavlov AA, Kasting JF, Eigenbrode JL, Freeman KH (2001). Organic haze in Earth’s early atmosphere: source of low-13C late Archean kerogens? *Geology* **29:** 1003-1006.

Pedros-Alio C, Garcia-Cantizano J, Calderon J (1993). Bacterial production in anaerobic water columns. In: Kemp P, Sherr B, Sherr B, Cole J (eds). *Handbook of methods in aquatic microbial ecology*. Lewis Publishers: Boca Raton. pp 519-530.

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

Pride DT, Schoenfeld T (2008). Genome signature analysis of thermal virus metagenomes reveals Archaea and thermophilic signatures. *Bmc Genomics* **9**.

Pride DT, Sun CL, Salzman J, Rao N, Loomer P, Armitage GC *et al* (2011). Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Research* **21:** 126-136.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41:** D590-D596.

Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* **8:** R10.

Rajendhran J, Gunasekaran P (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research* **166:** 99-110.

Rappe MS, Giovannoni SJ (2003). The uncultured microbial majority. *Annual Review of Microbiology* **57:** 369-394.

Rashby SE, Sessions AL, Summons RE, Newman DK (2007). Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci U S A* **104:** 15099-15104.

Ray WK, Zeng G, Potters MB, Mansuri AM, Larson TJ (2000). Characterization of a 12-kilodalton rhodanese encoded by glpE of Escherichia coli and its interaction with thioredoxin. *J Bacteriol* **182:** 2277-2284.

Reinhard CT, Planavsky NJ, Robbins LJ, Partin CA, Gill BC, Lalonde SV *et al* (2013). Proterozoic ocean redox and biogeochemical stasis. *Proc Natl Acad Sci U S A* **110:** 5357-5362.

Retallack GJ, Mindszenty A (1994). Well preserved late Precambrian Paleosols from Northwest Scotland. *Journal of Sedimentary Research* **64:** 264-281.

Revsbech NP, Jorgensen BB, Blackburn TH, Cohen Y (1983). Microelectrode studies of the photosynthesis and O_2, H_2S, and pH profiles of a microbial mat. *Limnology and Oceanography* **28:** 1062-1074.

Richardson LL, Castenholz RW (1987). Diel Vertical Movements of the Cyanobacterium Oscillatoria terebriformis in a Sulfide-Rich Hot Spring Microbial Mat. *Appl Environ Microbiol* **53:** 2142-2150.

Riding R (2006). Cyanobacterial calcification, carbon dioxide concentrating mechanisms, and Proterozoic-Cambrian changes in atmospheric composition. *Geobiology* **4:** 299-316.

Riding R (2011). Calcified cyanobacteria. In: Reitner J, Thiel V (eds). *Encyclopedia of Geobiology*. Springer: Heidelberg. pp 211-223.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F *et al* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499:** 431-437.

Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G *et al* (2011). Integrative genomics viewer. *Nature Biotechnology* **29:** 24-26.

Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F *et al* (2009). Opinion: Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7:** 828-836.

Rohwer F, Thurber RV (2009). Viruses manipulate the marine environment. *Nature* **459:** 207-212.

Ruberg SA, Kendall ST, Biddanda BA, Black T, Nold SC, Lusardi WR *et al* (2008). Observations of the Middle Island Sinkhole in Lake Huron - A unique hydrogeologic and glacial creation of 400 million years. *Marine Technology Society Journal* **42:** 12-21.

Rye R, Holland HD (2000). Life associated with a 2.76 Ga ephemeral pond?: Evidence from Mount Roe #2 paleosol. *Geology* **28:** 483-486.

Sanders JTG, Biddanda BA, Stricker CA, Nold SC (2011). Stable Isotope Analysis Reveals Benthic Macroinvertebrate and Fish Communities Linked to Submerged Groundwater Vents in Lake Huron. *Aquatic Biology* **12:** 1-11.

Sapranauskas R, Gasiunas G, Fremaux C, Barrangou R, Horvath P, Siksnys V (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Research* **39:** 9275-9282.

Schidlowski M (ed) (2000) *Carbon isotopes and microbial sediments*. Springer-Verlag: Berlin.

Schutz M, Shahak Y, Padan E, Hauska G (1997). Sulfide-quinone reductase from Rhodobacter capsulatus. Purification, cloning, and expression. *J Biol Chem* **272:** 9890-9894.

Shen Y, Knoll AH, Walter MR (2003). Evidence for low sulphate and anoxia in a mid-Proterozoic marine basin. *Nature* **423:** 632-635.

Shi T, Ilikchyan I, Rabouille S, Zehr JP (2010). Genome-wide analysis of diel gene expression in the unicellular N-2-fixing cyanobacterium Crocosphaera watsonii WH 8501. *Isme Journal* **4:** 621-632.

Shi Y, Tyson GW, DeLong EF (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-U154.

Shi YM, Tyson GW, Eppley JM, DeLong EF (2011). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *Isme Journal* **5:** 999-1013.

Soitamo AJ, Zhou G, Clarke AK, Oquist G, Gustafsson P, Aro EM (1996). Over-production of the D1:2 protein makes Synechococcus cells more tolerant to photoinhibition of Photosystem II. *Plant Molecular Biology* **30:** 467-478.

Sorek R, Kunin V, Hugenholtz P (2008). CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature Reviews Microbiology* **6:** 181-186.

Stal LJ (1995). Tansley Review No. 84. Physiological Ecology of Cyanobacteria in Microbial Mats and Other Communities. *New Phytologist* **131:** 1-32.

Stal LJ, Moezelaar R (1997). Fermentation in cyanobacteria. *FEMS Microbiology Reviews* **21:** 179-211.

Stal LJ (2000). Cyanobacterial Mats and Stromatolites. In: Whitton B, Potts M (eds). *The Ecology of Cyanobacteria*. Kluwer Academic Publishers: Dordrecht/London/Boston. pp 61-120.

Stamatakis A (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22:** 2688-2690.

Stockel J, Jacobs JM, Elvitigala TR, Liberton M, Welsh EA, Polpitiya AD *et al* (2011). Diurnal Rhythms Result in Significant Changes in the Cellular Protein Complement in the Cyanobacterium Cyanothece 51142. *Plos One* **6**.

Stoeckel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, Pakrasi HB (2008). Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci U S A* **105:** 6156-6161.

Straub C, Quillardet P, Vergalli J, de Marsac NT, Humbert JF (2011). A Day in the Life of Microcystis aeruginosa Strain PCC 7806 as Revealed by a Transcriptomic Analysis. *Plos One* **6**.

Summons RE, Jahnke LL, Hope JM, Logan GA (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400:** 554-557.

Sun CL, Barrangou R, Thomas BC, Horvath P, Fremaux C, Banfield JF (2013). Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* **15:** 463-470.

Suttle CA (2007). Marine viruses - major players in the global ecosystem. *Nature Reviews Microbiology* **5:** 801-812.

Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R (2011). Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR. *Science* **333:** 58-62.

Tajima K, Aminov RI, Nagamine T, Ogata K, Nakamura M, Matsui H *et al* (1999). Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *Fems Microbiology Ecology* **29:** 159-169.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* **24:** 1596-1599.

Taton A, Grubisic S, Balthasart P, Hodgson DA, Laybourn-Parry J, Wilmotte A (2006a). Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. *Fems Microbiology Ecology* **57:** 272-289.

Taton A, Grubisic S, Ertz D, Hodgson DA, Piccardi R, Biondi N *et al* (2006b). Polyphasic study of Antarctic cyanobacterial strains. *Journal of Phycology* **42:** 1257-1270.

Thingstad TF (2000). Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnology and Oceanography* **45:** 1320-1328.

Tsertova N, Kisand A, Tammert H, Kisand V (2011). Low seasonal variability in community composition of sediment bacteria in large and shallow lake. *Environmental Microbiology Reports* **3:** 270-277.

Tyson GW, Banfield JF (2008). Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10:** 200-207.

van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJJ (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences* **34:** 401-407.

Van Der Westhuizen AJ, Eloff JN (1985). Effect of temperature and light on the toxicity and growth of the blue-green alga Microcystis-aeruginosa uv-006. *Planta (Heidelberg)* **163:** 55-59.

Van Lith Y, Warthmann R, Vasconcelos C, Mckenzie JA (2003). Sulphate-reducing bacteria induce low-temperature Ca-dolomite and high Mg-calcite formation. *Geobiology* **1:** 71-79.

Vasconcelos C, Mckenzie JA, Bernasconi S, Grujic D, Tien AJ (1995). Microbial Mediation as a Possible Mechanism for Natural Dolomite Formation at Low-Temperatures. *Nature* **377:** 220-222.

Vijayan V, Zuzow R, O'Shea EK (2009). Oscillations in supercoiling drive circadian gene expression in cyanobacteria. *Proc Natl Acad Sci U S A* **106:** 22564-22568.

Vincent WF (2000). Cyanobacterial dominance in the Polar regions. In: Potts BAWaM (ed). *The Ecology of Cyanobacteria*. Kluver Academic Publishers: Amsterdam. pp 321-340.

Voorhies AA, Biddanda BA, Kendall ST, Jain S, Marcus DN, Nold SC *et al* (2012). Cyanobacterial life at low O2: community genomics and function reveal metabolic versatility and extremely low diversity in a Great Lakes sinkhole mat. *Geobiology* **10:** 250-267.

Walter GH, Paterson HEH (1994). The implications of paleontological evidence for theories of ecological communities and species richness. *Australian Journal of Ecology* **19:** 241-250.

Walter LM, Ku TCW, Muehlenbacks K, Patterson WP, Bonnell L (2007). Controls on δ13C of dissolved inorganic carbon in marine pore waters: an integrated case study of isotope exchange during syndepositional recrystallization of biogenic carbonate sediments (South Florida Platform, USA). . *Deep Sea Research Part II: Topical Studies in Oceanography* **54:** 1163-1200.

Walter MR, Bauld J (1983). The association of sulphate evaporites, stromatolitic carbonates and glacial sediments: examples from the Proterozoic of Australia and the Cainzoic of Antarctica. *Precambrian Research* **21:** 129-148.

Wang Y, Kern SE, Newman DK (2010). Endogenous phenazine antibiotics promote anaerobic survival of *Pseudomonas aeruginosa* via extracellular electron transfer. *Journal of Bacteriology* **192:** 365-369.

Warthmann R, van Lith Y, Vasconcelos C, McKenzie JA, Karpoff AM (2000). Bacterially induced dolomite precipitation in anoxic culture experiments. *Geology* **28:** 1091-1094.

Weinbauer MG (2004). Ecology of prokaryotic viruses. *Fems Microbiology Reviews* **28:** 127-181.

Weinberger AD, Sun CL, Plucinski MM, Denef VJ, Thomas BC, Horvath P *et al* (2012). Persisting Viral Sequences Shape Microbial CRISPR-based Immunity. *Plos Computational Biology* **8**.

Welander PV, Hunter RC, Zhang L, Sessions AL, Summons RE, Newman DK (2009). Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **191:** 6145-6156.

Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc Natl Acad Sci U S A* **107:** 8537-8542.

Wharton RA, Parker BC, Simmons GM (1983). Distribution, Species Composition and Morphology of Algal Mats in Antarctic Dry Valley Lakes. *Phycologia* **22:** 355-365.

Wilmes P, Simmons SL, Denef VJ, Banfield JF (2009). The dynamic genetic repertoire of microbial communities. *Fems Microbiology Reviews* **33:** 109-132.

Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K *et al* (2009). Initial community evenness favours functionality under selective stress. *Nature* **458:** 623-626.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al* (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337:** 1661-1665.

Yau S, Lauro FM, Williams TJ, DeMaere MZ, Brown MV, Rich J *et al* (2013). Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *Isme Journal* **7:** 1944-1961.

Yoshida-Takashima Y, Yoshida M, Ogata H, Nagasaki K, Hiroishi S, Yoshida T (2012). Cyanophage Infection in the Bloom-Forming Cyanobacteria Microcystis aeruginosa in Surface Freshwater. *Microbes and Environments* **27:** 350-355.

Zerbino DR, Birney E (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:** 821-829.

Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML *et al* (2009). Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, Prochlorococcus. *Plos One* **4**.

CHAPTER V

Conclusions

5.1 Introduction

While low-O_2 sulfidic environments are rare on the modern Earth, they were pervasive before the evolution of oxygenic photosynthesis by early cyanobacteria, and even after the great oxygenation event (GOE) there is evidence that the oceans remained low-O_2 sulfidic environments for almost two billion years before rising to modern levels of O_2 (Canfield and Farquhar 2009, Holland 2006). Although cyanobacteria are recognized as the agents of Earth's oxygenation, little is known about the factors that control oxygen production by cyanobacterial ecosystems under low-O_2 or redox-stratified conditions. A critical aspect of this issue is the regulation of oxygenic versus anoxygenic photosynthesis, at both the cellular level (i.e., regulation of metabolism) and community level (i.e., competition between various metabolisms and ecotypes). Fossils of microbial mats are preserved in the geologic record as some of the earliest evidence of life on Earth, but unfortunately beyond cell morphology and chemical/isotopic signatures of the surrounding rock, very little are left of those organisms for us to study, and there is limited information regarding photosynthetic metabolism of these early microbial mats. Here I have presented metagenomic and metatranscriptomic analysis of a modern low-O_2 sulfidic ecosystem that is well suited as a model of Precambrian cyanobacterial mats. The conclusions are broken into sections where I first address the overall environment at

MIS, the most abundant and transcriptionally active organisms, and their potential contributions to geochemistry. I then focus on conclusions regarding *Phormidium*, the dominant organism at MIS, and potential model of ancient cyanobacteria. In closing I briefly address potential directions for future research.

5.2 The Middle Island Sinkhole, a modern analog of ancient microbial mats

Conditions at MIS

The microbial mats at MIS are one of the better analogs of ancient redox-stratified cyanobacterial mats that have been found in the modern world. Sharp chemical gradients and microbially mediated biogeochemistry result from immersion in a meter of dense sulfidic water low in O_2 bathing the mats (Chapter II), creating an environment similar to what microbial mats in Proterozoic oceans might have experienced (Canfield and Farquhar 2009, Holland 2006). The MIS mat community is composed primarily of *Phormidium*, a cyanobacterium capable of facultative oxygenic/anoxygenic photosynthesis, a versatile metabolism that would have given it great advantages in the fluctuating redox conditions of the early Earth (Chapters II and III). Less abundant community members were found to perform a wide range of metabolic biogeochemistry including oxidation and reduction reactions (Chapter III), helping to make the mats a net sink for O_2 (Chapter II). This finding is surprising in light of cyanobacteria's well-recognized role in oxygen production, and provides new perspectives on the role of cyanobacteria in Earth's redox evolution. Though it is widely agreed that the evolution of oxygenic photosynthesis probably took place shortly before the GOE (Blank 2004, Blankenship et al 2007), the reason for the long delay in the rise of atmospheric and aquatic O_2 to modern levels is still poorly understood. Most modern cyanobacterial mats are sources of O_2 (Gingras

2011), indicating substantial contributions to primary production by anoxygenic photosynthesis or chemosynthesis at MIS.

When oxygenic photosynthesis is used to create biomass by converting CO_2 to organic carbon, O_2 is produced and can be used by heterotrophs to break that organic matter back down into CO_2, releasing it back to the atmosphere, completing the cycle. However, if carbon is fixed into organic matter by anoxygenic photosynthesis or chemosynthesis, O_2 is not produced. In anoxic, organic carbon-rich environments any O_2 that is produced by oxygenic photosynthesis will be quickly used before it can escape to the atmosphere. Thus, significant contributions to primary production by anoxygenic photosynthesis, including those by metabolically versatile cyanobacteria, could slow the rise of oxygen, even after the evolution and proliferation of cyanobacteria capable of oxygenic photosynthesis (Johnston et al 2009a). Our finding of a modern cyanobacterial mat system in low-O_2 sulfidic conditions that serves as a sink for O_2 helps to demonstrate at least one of the potential reasons why the Earth's atmosphere may have remained at around 1% of its current levels for over a billion years (Holland 2006), and highlights the importance of the balance between oxygenic and anoxygenic photosynthesis in controlling the redox impact of microbial communities.

MIS community composition and metabolism

The mats at MIS are largely dominated by *Phormidium,* with over 4000x genomic DNA coverage in metagenomic datasets from fifteen samples collected between 2007 and 2012. The second most abundant organism (*Oscillatoria*) had 553x average genome coverage, and most genomes recovered had less than 100x total coverage (Chapter III). This makes for a community with low diversity and evenness that is quite rare in modern systems, which usually have more

species present, and a more even spread of genomic coverage (Chivian et al 2008, Denef et al 2010). *Phormidium* also performs the majority of photosynthesis and makes a significant contribution to carbon fixation along with the chloroplast of a Bacillariophyta diatom (Chapter II and III). The dominance of *Phormidium* indicates that the MIS mats lack functional redundancy and suggests that they could be susceptible to changing environmental conditions (Wittebolle et al 2009).

Several other groups of bacteria play key metabolic roles that influence the geochemistry of MIS. Five genomic bins comprising partial genomes for seven organisms from the phylum Bacteroidetes were recovered (Chapter III). Based on their abundance and gene content, these Bacteroidetes populations are probably performing a combination of sulfite oxidation and aerobic heterotrophy at MIS, though only aerobic respiration genes showed expression (Chapter III). Members of this phylum have been well documented as heterotrophs, and the MIS mats are a carbon rich environment (Nold et al 2013). Many microbes at MIS were also found to code for and express ABC transporters of the branch chain amino acid and polar amino acid variety, as well as di and tricarboxylate transporters, all of which have been implicated in heterotrophy.

Members of the *Desulfobacterales* (*Deltaproteobacteria*) were found to perform the majority of sulfate reduction using the DSR system, converting sulfate into hydrogen sulfide (Chapter II and III). They were also found to express hydrogen oxidation genes that are usually coupled to sulfate reduction in other *Deltaproteobacteria*. *Desulfobacterales* coded for and expressed genes for both autotrophy and heterotrophy, but transcript coverage was significantly higher for autotrophy genes (Chapter III). Sulfate reduction has been well documented within the *Deltaproteobacteria*, which are believed to be influential in carbon cycling in anaerobic environments such as sediments (Muyzer and Stams 2008). Based on expression of sulfate

reducing genes, the *Desulfobacterales* are converting significant amounts of the sulfate at MIS into sulfide. Sulfide serves as an energy source for lithotrophy performed by *Gammaproteobacteria* and *Epsilonproteobacteria*, and freely reacts with available O_2, contributing to hypoxia at MIS.

Betaproteobacteria, *Epsilonproteobacteria*, and *Gammaproteobacteria* are responsible for the majority of sulfur oxidation at MIS including thiosulfate oxidation using the *sox* system. *Betaproteobacteria* had the most transcripts for sulfide oxidation using the reverse dissimilatory sulfite reductase pathway, while the *Betaproteobacteria* and *Gammaproteobacteria* used flavocytochrome C sulfide dehydrogenase to oxidize sulfide. *Phormidium* also made contributions to sulfide oxidation using sulfide quinone oxidoreductase (SQR) and sulfite oxidation using *aprA*. Sulfur oxidizing and reducing bacteria play roles in modulating concentrations of sulfide at MIS, which is likely a key factor governing the form of photosynthesis conducted by the overlying cyanobacteria.

MIS is host to novel organisms

MIS has proved to be a fruitful source of genomic data for several phyla that have no cultured representatives. The ARB-Silva database currently lists 57 phyla (Quast et al 2013), 30 of which are referred to as candidate divisions because they contain no cultured representatives. Because these microbial groups are unknown in terms of metabolism and function, they have been referred to as "Microbial Dark Matter" (Castelle et al 2013, Rinke et al 2013). While a few organisms have been cultured and very well characterized, less than 1% of organisms can be cultured in the lab, with entire phyla represented only by 16s rRNA marker gene surveys (Rajendhran and Gunasekaran 2011, Rappe and Giovannoni 2003). In addition, almost all phyla

have entire classes with no cultured representatives or environmental genomes available, and a major push to fill in representative genomes on the tree of life is currently underway (Castelle et al 2013, Rinke et al 2013, Wrighton et al 2012).

Significant genomic and transcriptomic information on bacterial candidate divisions WS3, SM2F11 and RF3 was retrieved from MIS. I have presented a near-complete genome and transcriptomic data for WS3, which codes for several forms of sulfur, hydrogen and carbon monoxide oxidation, as well as autotrophy using the reverse TCA cycle. Transcripts of genes relating to hydrogen oxidation, aerobic respiration and chemotaxis were recovered (Chapter III). This is the second genomic and first transcriptomic data available for WS3, and shows that WS3 codes for a wide range of functions that were not found in the previously published single cell genome, such as sulfur oxidation and electron transport genes (Rinke et al 2013) (Chapter III).

Partial genomes were recovered for SM2F11 and RF3, both of which are novel in terms of gene content (Chapter III). The SM2F11 genome is estimated to be 81% complete, yet 812 of 1569 genes have no known homologs, making identification of functional or metabolic genes difficult. Low similarity (23% amino acid identity) was detected to genes involved in denitrification, as well as a Rhodanese-related sulfurtransferase and coupled thioredoxin reductase, which is a large family of genes involved in thiosulfate oxidation for a variety of cellular and extracellular functions (Ray et al 2000). Only a small portion of the RF3 genome was recovered (~18%), and no metabolic genes were identified (Chapter III).

The unique microbial community and geochemistry at MIS make it a valuable resource for trying to understand ancient cyanobacterial mat systems. Our results highlight the large diversity of organisms to be found in modern microbial systems, and challenges in understanding

ancient microbial systems when our knowledge of modern microbial diversity and contributions to biogeochemistry are so incomplete.

5.3 Phormidium: A metabolically versatile cyanobacterium

Phormidium as a model of low-O_2 sulfide tolerant cyanobacteria

Phormidium is a filamentous cyanobacterium (Figure 2.4) commonly able to survive in low-O_2, sulfidic conditions that usually inhibit cyanobacteria from performing photosynthesis (Cohen et al 1986, Miller and Bebout 2004b). Measurements of primary production process rates and oxygen concentrations, gene content, and photosynthetic gene expression all indicate that the MIS *Phormidium* are capable of conducting anoxygenic photosynthesis (Chapter II and III) when photosystem II is inhibited by sulfide, as has been observed in some other cyanobacteria of the *Oscillatoriales* (Bronstein et al 2000b). Microprobe measurements of sulfide and oxygen at MIS indicate fluctuating redox conditions that could inhibit photosystem II during periods of high sulfide (Kinsman-Costello et al., unpublished data). Periodic inhibition of photosystem II by sulfide (Cohen et al 1986) may explain why *Phormidium* appears to use both anoxygenic and oxygenic photosynthesis (Figure 3.3). It is not clear if all *Phormidium* cells at MIS are capable of anoxygenic photosynthesis or if different strains are adapted to different niches. It is also not clear if MIS Phormidium use a combination of oxygenic and anoxygenic photosynthesis simultaneously, or if they only use anoxygenic photosynthesis when PS II is inhibited. Closed genomes or pure cultures of *Phormidium* strains will be required to make these distinctions.

While the high ratio of *Phormidium* PS I versus PS II transcript abundance is consistent with anoxygenic photosynthesis (Figure 3.3), many aspects of the mechanism remain uncertain. *Phormidium* does encode and heavily express SQR, which some cyanobacteria can use to

perform anoxygenic photosynthesis (Bronstein et al 2000b, Cohen et al 1986). SQR is a diverse family of proteins that can also be used for sulfide detoxification (Bronstein et al 2000b), and the SQR encoded by *Phormidium* is too divergent from any SQR with experimentally determined function to be certain of the *Phormidium* encoded SQRs function. SQR in *Phormidium* is more heavily expressed than some PS I genes used in anoxygenic photosynthesis (Figure 3.9), consistent with its involvement in anoxygenic photosynthesis.

Metatranscriptomic analysis of *Phormidium* revealed significantly higher transcript abundance in night samples compared to day samples (Figure 3.8), in contrast to unicellular cyanobacteria that reach peak transcript abundance for photosynthesis core genes during the day (Colon-Lopez and Sherman 1998, Shi et al 2010, Stockel et al 2011, Straub et al 2011, Zinser et al 2009). These model cyanobacteria are unicellular, and have been shown to express a portion of their genes (20-80%) on a diel cycle, which always includes photosynthesis genes (Shi et al 2010, Stockel et al 2011, Zinser et al 2009). We do not have enough temporal data points to resolve *Phormidium's* potential diel cycle of gene expression, but from the day/night data it is clear that MIS *Phormidium* transcriptome patterns are distinct from those observed for cyanobacteria previously. During the day when model cyanobacterial transcript abundance is at its highest point for photosynthesis reaction core genes (Zinser et al 2009), *Phormidium* average transcript abundance for photosynthesis genes is significantly lower than during the night. It is not clear when *Phormidium's* peak transcript abundance is, only that it is significantly higher at the night sampling point than the day. Transcriptomic analysis of MIS mats at higher temporal resolution is required to further investigate this discrepancy. However, it is clear that *Phormidium* deviates from expression patterns observed in model unicellular cyanobacteria, and

care should be taken when using unicellular cyanobacteria as models of all cyanobacterial photosynthesis.

Using genomics of modern communities to inform studies of ancient systems

Studying the gene content and expression of modern low-O_2 sulfide tolerant cyanobacteria can inform our search for biomarkers of ancient cyanobacteria. Genes for *shc* and *hpnP* have been implicated in the biosynthesis of 2-methylhopanoids (Welander et al 2009), which have been used as markers of cyanobacteria (and by proxy oxygenic photosynthesis) in fossilized microbial mats (Summons et al 1999). However, doubt has recently been cast on using these molecules as markers of cyanobacteria, as not all cyanobacteria code for them, and they have been found in the genomes of organisms other than cyanobacteria (Rashby et al 2007, Welander et al 2009, Welander et al 2010). Likewise, these genes are not encoded by *Phormidium* at MIS (Chapter II), and have now been shown to be a response to stress, and are not specific to cyanobacteria or oxygenic photosynthesis (Kulkarni et al 2013). Genomic analysis of modern systems can be a powerful tool in determining the utility of biomarkers of ancient systems.

The role of viral predation on modern systems

Viral predation has been shown to play a role in microbial community abundance and structure in many environments, and arguments could be made that it affects every environment where life is found (Banfield and Young 2009, Rodriguez-Valera et al 2009, Weinbauer 2004). Viruses have been estimated to be ten times more abundant in the oceans than microbes, yet our understanding of their impact on host mortality and repercussions for environmental metabolic function is still poorly characterized (Weinbauer 2004). Microbes encode a defense mechanism

called the CRISPR/cas system that has been shown to provide some level of immunity against recurring invading nucleic acids like viruses and plasmids (Barrangou et al 2007, Makarova et al 2006). Three CRISPR systems were recovered from the *Phormidium* genome (Figure 4.2), and one of these CRISPR loci targets an extremely abundant virus at MIS, *Phormidium* phage PhV1 (Figure 4.1), indicating that the MIS cyanobacteria face viral predation pressure.

Complete circular genomes were reconstructed for two genotypes of the virus PhV1 found at MIS (Figure 4.1), only one of which was targeted by *Phormidium's* CRISPR system due to genome rearrangements (Figure 4.2D). In addition to *Phormidium* containing a small portion of the PhV1 genome, PhV1 encodes a phycobilisome degradation protein *nblA* that is derived from *Phormidium* and breaks down phycobilisomes, large protein complexes that absorb wavelengths of light that are inaccessible to chlorophyll. Two way transfer of DNA links Phormidium and PhV1 using sequence data alone, and invites inferences about viral physiology based on host physiology (Chapter IV).

Over the five years of the study, we observed opposite trends in the abundance of the two PhV1 genotypes, with PhV1.TypeB (the genotype targeted by *Phormidium's* CRISPR and dominant in 2007) dropping in abundance, and PhV1.TypeA (immune to *Phormidium's* CRISPR due to genome rearrangement and dominant in 2012) showing a marked increase during that time (Figure 4.7A). During the study, the relative abundance of *Phormidium* in the metagenome drops, correlating with the rise in abundance of PhV1.TypeA (Figure 4.7B). Though this relationship does not show causation, as there could be other reasons for the drop in *Phormidium* abundance (e.g., fluctuations in environmental conditions, stochastic variability between samples), two directional exchange of DNA by virus and host does indicate that viruses have the potential to play a major role in ecosystem dynamics of modern cyanobacterial mat systems.

Better understanding of the role of viruses in modern systems could inform our understanding of their role in ancient systems.

5.4 Potential directions for future investigations

Our studies of the cyanobacterial mats at MIS have raised nearly as many questions as they have answered, with advancements in our understanding of *Phormidium* and MIS hinting at multiple potential lines of further investigation. Here I present potential directions for further research at MIS that could provide insight into modern and ancient cyanobacterial mat systems, and warrant further scientific inquiry.

Phormidium, SQR and anoxygenic photosynthesis

Several important questions regarding *Phormidium* still remain unanswered. While several lines of evidence indicate that *Phormidium* is performing anoxygenic photosynthesis, likely using SQR, culture based genomic investigations will be required to determine the details of photosynthetic gene regulation in *Phormidium*. It is not clear if the PS II genes encoded by Phormidium are resistant enough to sulfide to perform oxygenic and anoxygenic photosynthesis at the same time, or if the presence of sulfide totally inhibits PS II, forcing *Phormidium* to switch to anoxygenic photosynthesis alone. In addition, multiple genotypes of *Phormidium* exist at MIS, and resolving the differences in gene content and function between these variants may inform our understanding of niche differentiation and the ecological role of diversity in the MIS system.

Efforts to bring *Phormidium* from MIS into laboratory culture have been ongoing for some years and have been unsuccessful. Though we can get small pieces of mat to grow in the

lab, microscopy images indicate the community shifts during this time to no longer reflect the community at MIS, and the mats eventually deteriorate and die. Isolation or long-term enrichment of *Phormidium* cultures would be a critical step in understanding the mechanics of cyanobacterial tolerance to sulfide and their ability to perform anoxygenic photosynthesis. A wide range of experiments could be designed to better understand the genetic basis for this metabolism, as well as verify *Phormidium's* use of SQR for anoxygenic photosynthesis. Isolates of *Phormidium* could also be used to investigate the diel cycle of photosynthetic gene expression, placing it in context of other transcriptionally studied cyanobacteria.

Efforts to assemble closed genomes for *Phormidium* strains from metagenomes have been unsuccessful despite nearly 4000x genome coverage. This inability to close the genome is likely due to the high incidence (~200) of transposons, mobile genetic elements that can move DNA sequences around within one genome or from one genome to another. Transposons with identical sequences ranging from 500 to 1000bp occur in multiple places in the *Phormidium* genome, making the exact sequence of genes impossible to determine because the transposons are five to ten times longer than the short read DNA sequences available. Efforts to use PacBio sequencing (which produces reads >5000bp and could span these transposons) failed for unknown reasons, preventing closure of the *Phormidium* genome. Efforts to overcome sequencing failure are underway currently, and closed genomes for *Phormidium* strains would be a valuable tool for determining the different rolls *Phormidium* plays in biogeochemistry.

Spatial and temporal resolution of MIS

All of the samples discussed in this dissertation were collected from the same area, with a diameter of about 50-100 meters in the "arena" portion of MIS. Sampling focused only on the

bulk cyanobacterial mats, not the sediments beneath or the water above. Spatially resolved sampling (e.g., across vertical redox gradients in the mat or across horizontal gradients along groundwater flow) coupled to geochemical measurements could provide insight into community composition and dynamics, especially the links between hydrology, chemistry, and biology at MIS. In addition, visual observations of the mats show weak and incomplete mats in May, dense healthy looking mats in June/July, and weak and incomplete mats in September, suggesting a yearly cycle of growth and deterioration for the mats at MIS that is poorly understood. Sampling coupled to diver observations concerning mat health taken regularly over a two year period could help determine the nature of annual microbial cycling at MIS.

Filling in the Tree of Life

Metagenomic sampling of the MIS mats recovered genomes for three uncultured bacterial phyla believed to reside in the sediments below the mats. These sediments could be a fruitful source of genomic and transcriptomic data for uncultured groups of microbes we currently know nothing about. 16s rRNA sequences assembled from MIS mats indicate that members of at least 10 bacterial phyla with no cultured representatives can be found at MIS, and deep sequencing of MIS sediments could greatly expand our knowledge of the un-cultured and un-sequenced branches of the tree of life.

5.5 References

Banfield JF, Young M (2009). Variety-the Splice of Life-in Microbial Communities. *Science* **326:** 1198-1199.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S *et al* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709-1712.

Blank CE (2004). Evolutionary timing of the origins of mesophilic sulphate reduction and oxygenic photosynthesis: a phylogenomic dating approach. *Geobiology* **2:** 1-20.

Blankenship R, Sadekar S, Raymond J (2007). The evolutionary transition from anoxygenic to oxygenic photosynthesis. In: Falkowski PG, Knoll AH (eds). *Evolution of primary producers in the sea*. Elsevier. pp 21-35.

Bronstein M, Schutz M, Hauska G, Padan E, Shahak Y (2000). Cyanobacterial sulfide-quinone reductase: Cloning and heterologous expression. *J Bacteriol* **182:** 3336-3344.

Canfield DE, Farquhar J (2009). Animal evolution, bioturbation, and the sulfate concentration of the oceans. *Proc Natl Acad Sci U S A* **106:** 8123-8127.

Castelle CJ, Hug LA, Wrighton KC, Thomas BC, Williams KH, Wu DY *et al* (2013). Extraordinary phylogenetic diversity and metabolic versatility in aquifer sediment. *Nature Communications* **4**.

Chivian D, Brodie EL, Alm EJ, Culley DE, Dehal PS, Desantis TZ *et al* (2008). Environmental genomics reveals a single-species ecosystem deep within Earth. *Science* **322:** 275-278.

Cohen Y, Jorgensen BB, Revsbech NP, Poplawski R (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl Environ Microbiol* **51:** 398-407.

Colon-Lopez MS, Sherman LA (1998). Transcriptional and translational regulation of photosystem I and II genes in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. *J Bacteriol* **180:** 519-526.

Denef VJ, Mueller RS, Banfield JF (2010). AMD biofilms: using model communities to study microbial evolution and ecological complexity in nature. *Isme Journal* **4:** 599-610.

Gingras M, J.W. Hagadorn, A. Seilacher, S.V. Lalonde, E. Pecoits, D. Petrash, and K.O. Konhauser (2011). Possible evolution of mobile animals in association with microbial mats. *Nature Geoscience*.

Holland HD (2006). The oxygenation of the atmosphere and oceans. *Philos Trans R Soc Lond B Biol Sci* **361:** 903-915.

Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009). Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci U S A* **106:** 16925-16929.

Kulkarni G, Wu CH, Newman DK (2013). The General Stress Response Factor EcfG Regulates Expression of the C-2 Hopanoid Methylase HpnP in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **195:** 2490-2498.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology Direct* **1**.

Miller SR, Bebout BM (2004). Variation in sulfide tolerance of photosystem II in phylogenetically diverse cyanobacteria from sulfidic habitats. *Appl Environ Microbiol* **70:** 736-744.

Muyzer G, Stams AJM (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* **6:** 441-454.

Nold SC, Bellecourt MJ, Kendall ST, Ruberg SA, Sanders TG, Klump JV *et al* (2013). Underwater sinkhole sediments sequester Lake Huron's carbon. *Biogeochemistry* **115:** 235-250.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* **41:** D590-D596.

Rajendhran J, Gunasekaran P (2011). Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond. *Microbiological Research* **166:** 99-110.

Rappe MS, Giovannoni SJ (2003). The uncultured microbial majority. *Annual Review of Microbiology* **57:** 369-394.

Rashby SE, Sessions AL, Summons RE, Newman DK (2007). Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci U S A* **104:** 15099-15104.

Ray WK, Zeng G, Potters MB, Mansuri AM, Larson TJ (2000). Characterization of a 12-kilodalton rhodanese encoded by glpE of Escherichia coli and its interaction with thioredoxin. *J Bacteriol* **182:** 2277-2284.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F *et al* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499:** 431-437.

Rodriguez-Valera F, Martin-Cuadrado A-B, Rodriguez-Brito B, Pasic L, Thingstad TF, Rohwer F *et al* (2009). OPINION Explaining microbial population genomics through phage predation. *Nature Reviews Microbiology* **7:** 828-836.

Shi T, Ilikchyan I, Rabouille S, Zehr JP (2010). Genome-wide analysis of diel gene expression in the unicellular N-2-fixing cyanobacterium Crocosphaera watsonii WH 8501. *Isme Journal* **4:** 621-632.

Stockel J, Jacobs JM, Elvitigala TR, Liberton M, Welsh EA, Polpitiya AD *et al* (2011). Diurnal Rhythms Result in Significant Changes in the Cellular Protein Complement in the Cyanobacterium Cyanothece 51142. *Plos One* **6**.

Straub C, Quillardet P, Vergalli J, de Marsac NT, Humbert JF (2011). A Day in the Life of Microcystis aeruginosa Strain PCC 7806 as Revealed by a Transcriptomic Analysis. *Plos One* **6**.

Summons RE, Jahnke LL, Hope JM, Logan GA (1999). 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* **400:** 554-557.

Weinbauer MG (2004). Ecology of prokaryotic viruses. *Fems Microbiology Reviews* **28:** 127-181.

Welander PV, Hunter RC, Zhang L, Sessions AL, Summons RE, Newman DK (2009). Hopanoids play a role in membrane integrity and pH homeostasis in Rhodopseudomonas palustris TIE-1. *J Bacteriol* **191:** 6145-6156.

Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK (2010). Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc Natl Acad Sci U S A* **107:** 8537-8542.

Wittebolle L, Marzorati M, Clement L, Balloi A, Daffonchio D, Heylen K *et al* (2009). Initial community evenness favours functionality under selective stress. *Nature* **458:** 623-626.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al* (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337:** 1661-1665.

Zinser ER, Lindell D, Johnson ZI, Futschik ME, Steglich C, Coleman ML *et al* (2009). Choreography of the Transcriptome, Photophysiology, and Cell Cycle of a Minimal Photoautotroph, Prochlorococcus. *Plos One* **4**.