Title

Genomics of Antarctic Cyanobacteria from Lakes Fryxell and Vanda: Sulfide Tolerant Oxygenic Photosynthesis, Implications of Polar Light Cycles, and Biogeography with Large-Scale k-mer Searching

Author

Lumian, Jessica Elizabeth Mizzi

Publication Date

2022

Peer reviewed|Thesis/dissertation

Genomics of Antarctic Cyanobacteria from Lakes Fryxell and Vanda: Sulfide Tolerant Oxygenic Photosynthesis, Implications of Polar Light Cycles, and Biogeography with Large-Scale k-mer Searching

By

JESSICA ELIZABETH MIZZI LUMIAN

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Dawn Sumner, Chair

_____

C. Titus Brown

_____

Patrick Shih

Committee in Charge

2022

ACKNOWLEDGEMENTS

First of all I would like to thank my advisor, Dr. Dawn Sumner, for all the support and guidance over the years that has shaped me into the scientist that I am today. Thank you to the entire Sumner lab for being such a brilliant and welcoming scientific environment, and I would like to specifically acknowledge Dr. Christy Grettenberger for all the assistance and support navigating the delightful (and sometimes chaotic) world of microbial ecology and Cyanobacteria phylogeny. I would also like to thank Dr. C. Titus Brown for changing my life in multiple ways: cultivating a wonderfully supportive learning environment to teach bioinformatics to an undergrad who had never heard of a terminal, and bringing me to Davis, CA with a job that kickstarted my computational adventures. Thank you to all of the members of the lab for Data Intensive Biology for teaching me so much, and continuing to be fantastic collaborators.

Thank you to my dissertation committee for the advice over the years, and thank you to the Microbiology Graduate Group and Department of Earth and Planetary Sciences for the microbiology and geobiology education. I am very appreciative of all my scientific collaborators for always being excited to share ideas and work together. I would also like to extend my gratitude to Dr. Julianne McCall, Dr. Shannon Muir, Megan Varvais, and the California Initiative to Advance Precision Medicine in the Governor's Office of Planning and Research for providing me with an enriching internship opportunity doing exciting and impactful work.

There aren't proper words to thank my family and friends for all the support over the years, but I'll try. Thank you to my parents for your unending support, thank you to my siblings for being hilarious and fun, and thank you to my friends for being downright peachy. I would like

to specifically thank Hyunsoo Gloria Kim for these past years of friendship, as I am convinced that I would have never made it through this program if not for us being two peas in a pod.

I can never thank Kevin Lumian enough for being the most supportive, empathetic, and thoughtful partner one could ask for. Once again: Kevin, thanks for being my husband! I'm so happy we're married. Finally, I would like thank Benjamin Andrew Bloom for teaching me what is most important in life and inspiring me to be the best person I can be.

ABSTRACT

Antarctic cyanobacteria form microbial mats in perennially ice-covered lakes in the McMurdo Dry Valleys, Antarctica. These mats demonstrate a variety of ecological and metabolic behaviors consistent with Antarctic conditions as well as specific challenges between and within the lakes. In this thesis, I explore the survival of cyanobacteria in Antarctic conditions focusing on sulfide stress, polar light availability, and their biogeographical distribution in other environments.

In Lake Fryxell, Antarctica, the benthic, filamentous cyanobacterium *Phormidium pseudopriestleyi* creates a 1-2 mm thick layer of 50 µmol L^{-1} O_2 in otherwise sulfidic water, demonstrating that it sustains oxygenic photosynthesis in the presence of sulfide. Sulfide inhibits oxygenic photosynthesis by blocking electron transfer between H_2O and the oxygen-evolving complex in the D1 protein of Photosystem II. The ability of cyanobacteria to counter this effect has implications for understanding the productivity of benthic microbial mats in sulfidic environments throughout Earth history. A metagenome-assembled genome (MAG) of *P. pseudopriestleyi* indicates a genetic capacity for oxygenic photosynthesis, including multiple copies of *psbA* (encoding the D1 protein of Photosystem II), and anoxygenic photosynthesis with a copy of *sqr* (encoding the sulfide quinone reductase protein that oxidizes sulfide). The genomic content of *P. pseudopriestleyi* is consistent with sulfide tolerance mechanisms including increasing *psbA* expression or directly oxidizing sulfide with sulfide quinone reductase. However, it is unknown whether the organism can perform anoxygenic photosynthesis using *sqr* and PS I while PS II is sulfide-inhibited.

The seasonal light availability of polar environments has implications for the functioning of circadian clocks in Antarctic cyanobacteria. However, polar cyanobacteria are underrepresented in available genomic data, limiting opportunities to study their genetic adaptations to this and other polar challenges. Chapter 2 presents four new Antarctic cyanobacteria MAGs, bringing the total number of polar MAGs and genomes to ten. These new cyanobacteria are from microbial mats in Lake Vanda located in the McMurdo Dry Valleys in Antarctica. The four MAGs were taxonomically categorized as a *Leptolyngyba*, *Pseudanabaena*, *Microcoleus*, and *Neosynechococcus*. The sizes of these MAGs range from 2.76 Mpb – 6.07 Mbp and the bin completion ranges from 74.2% – 92.57%. Furthermore, the four novel cyanobacteria have ANIs below 90% with each other and below 77% with existing polar cyanobacteria MAGs and genomes, which demonstrates genomic diversity among polar cyanobacteria genomes. Analysis of the gene content of all ten polar cyanobacteria demonstrates that they have metabolic capacity for photosynthesis and various cold tolerance mechanisms. Standard circadian rhythm genes are present in the majority of cyanobacteria studied with the *Leptolyngbya* and *Neosynechococcus* containing *kaiB3*, a divergent homolog of *kaiB*.

Although some work has been done on the biogeography of cyanobacteria within Antarctica, the global distribution of polar cyanobacteria is not well understood. It is not known if polar cyanobacteria are specialists of their environments or opportunists that tolerate survival in the cryosphere. These questions can be addressed by identifying the distribution of Antarctic cyanobacteria across global locations, but this is logistically complicated to do using 16S amplicon sequences, which is standard for biogeography studies. However, novel sourmash MAGsearch software based on large-scale k-mer searching allows for MAGs of interest to be identified in

publicly available metagenomic data. Chapter 3 used five Antarctic cyanobacteria MAGs from Lake Vanda (*Microcoleus*, *Neosynechococcus*, *Leptolyngbya*, and *Pseudanabaena*) and Lake Fryxell (*P. pseudopriestleyi*), as search queries. The sourmash MAGsearch revealed that the *Microcoleus* MAG was widely dispersed in a variety of environmental conditions, the *P. pseudopriestleyi* MAG was present in harsh environments, and the *Neosynechococcus, Leptolyngbya,* and *Pseudanabaena* MAGs were present in polar and a few non-polar environments. The search technique can be used to search for any organism present in metagenomes and has exciting implications for future biogeography studies.

INTRODUCTION

Cyanobacteria are responsible for oxygenating the atmosphere during the Great Oxygenation Event (GOE) as a result of evolving oxygenic photosynthesis (Kasting, 2013). The rise of oxygen in the atmosphere allowed for the evolution of aerobic life, which could not develop in the anoxic pre-GOE environments. This accumulation of O_2 in the atmosphere comes from the use of H_2O as an electron donor, which yields more energy than any other microbial metabolism electron donor (Vermaas, 2001). Although H_2O is an energy-rich compound, most bacteria are unable to biochemically split it because of the strong bonds between oxygen and hydrogen. Cyanobacteria are the only organisms in early Earth that evolved cellular machinery to split H_2O and thus produce O_2 as a by-product (Stal, 2007). Most O_2 is used in biochemical processes, but some of it persists in the surrounding environment. This accumulation led to the oxidation of environments where cyanobacteria performed oxygenic photosynthesis, and eventually gave rise of the oxygenation of Earth's atmosphere.

There are several factors that affect photosynthesis, and early Earth cyanobacteria faced many environmental challenges while performing oxygenic photosynthesis, such as light availability, sulfide, and cold stress (Canfield & Raiswell, 1999; Hoffman et al., 2017). The availability of light in an environment controls the cyanobacteria's ability to photosynthesize and oxygenate an environment. As a result of accumulated oxygen, sulfur cycling increased, which increased sulfide levels (Canfield & Raiswell, 1999). Sulfide normally inhibits oxygenic photosynthesis by blocking the electron donation step of H_2O molecules interacting with the oxygen evolving complex of the D1 protein in Photosystem II (Cardona et al., 2015). Furthermore, during the Cryogenian Snowball Earth period 720 – 635 million years ago, cyanobacteria had to

perform photosynthesis while experiencing cold stress (Hoffman et al., 2017; Hawes et al., 2018). Although it is difficult to categorize the geochemical interactions and stress responses of early Earth cyanobacteria, modern cyanobacteria living in extreme environments can be studied to better understand these processes.

Cyanobacteria experiencing these stresses grow in microbial mats in perennially ice-covered lakes in the McMurdo Dry Valleys, Antarctica (Sumner et al., 2015, 2016; Jungblut et al., 2016). In these mat communities, cyanobacteria drive primary productivity, which affects local ecology and geochemistry (Dillon et al., 2020). These mats face a variety of challenges specific to living in a polar environment. To reach Antarctica, organisms must survive freeze dry conditions while being transported through the atmosphere. Once in Antarctica, cyanobacteria must persist in cold temperatures and deal with variable light availability. Due to high latitude, polar environments receive three months of light in the summer and three months of darkness in the winter with transitional seasons in between. Furthermore, the light levels between lakes differ due to varying turbidity of the water column and thickness of the ice cover. Lake Vanda has the clearest water column of the McMurdo Dry Valleys lakes, while Lake Fryxell has a planktonic community that increases water column turbidity (Sumner et al., 2015; Castendyk et al., 2016). Furthermore, cyanobacteria receive varying amounts of light based on their proximity to the surface of a mat and their depth in the lake. The availability of light shapes the photosynthetic activity and ecology of these cyanobacteria.

In addition to light variance, Antarctic cyanobacteria deal with other challenges including cold stress and conditions specific to their local environment, like sulfide stress. In Lake Fryxell, *Phormidium pseudopriestleyi* creates a small amount of oxygen in anoxic waters despite the

presence of sulfide, which normally inhibits oxygenic photosynthesis (Sumner et al., 2015). *P. pseudopriestleyi* is the dominant organism on the top of the mat at this lake depth, but is less prominent at shallower depths (Jungblut et al., 2016; Dillon et al., 2020). This indicates that *P. pseudopriestleyi* can grow in sulfidic conditions, but other cyanobacteria are better equipped to grow at shallower and non-sulfidic depths (Dillon et al., 2020). Although all Antarctic cyanobacteria must deal with cold stress, different cyanobacteria persist in additional conditions specific to different parts of the lakes (Zhang et al., 2015). This leads to a variety of tolerance mechanisms and genomic diversity among Antarctic cyanobacteria.

Despite the genomic diversity of Antarctic cyanobacteria, these organisms remain underrepresented in public genomic data. At the time of this work, only five Antarctic cyanobacteria genomes and metagenome-assembled genomes (MAGs) were present on GenBank (Benson et al., 2013). One major challenge to acquiring new Antarctic cyanobacteria is that sampling in remote areas is expensive and logistically complicated. As a result, it is imperative to use existing data to answer multiple questions when possible.

The work presented in this thesis uses metagenomic high throughput sequencing to explore the survival of Antarctic cyanobacteria in different environments, which has implications for the ecology of early Earth cyanobacteria. The first chapter explores the metabolic potential of *P. pseudopriestleyi* and speculates how it can perform oxygenic photosynthesis in the presence of sulfide. In addition to the genomic content of this organism, the biogeochemistry of Lake Fryxell of the lake is explored, namely light and sulfide levels. In the second chapter, four new MAGs from Lake Vanda are presented, doubling the amount of public Antarctic cyanobacteria genomes and MAGs at the time of writing from five to ten. The implications of seasonal light

availability on all Antarctic cyanobacteria are analyzed in a comparative genomic study between the new five MAGs and the existing Antarctic MAGs and genome focusing on circadian clock, photosynthesis, and light harvesting genes. Finally, the biogeography of these five MAGs is explored using novel sourmash MAGsearch bioinformatics technology (Brown & Irber, 2016; Pierce et al., 2019; Brown, 2021; Irber, 2020a). This approach uses large-scale k-mer indexes to search for MAGs of interest in public metagenomic data on the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) (Leinonen et al., 2011).

Results demonstrate the diversity in metabolic potential, stress tolerance, and ecology of Antarctic cyanobacteria. A better understanding of the survival strategies of cyanobacteria in challenging environments will allow for the formation of hypotheses about early Earth cyanobacteria. Furthermore, understanding and categorizing the stress responses and ecological patterns of cyanobacteria can inform their use in bioremediation strategies for environmental changes due to climate change. Overall study of cyanobacteria in extreme environments remains an exciting area of research with implications for both the past and future.

CHAPTER 1

Metabolic Capacity of the Antarctic Cyanobacterium *Phormidium pseudopriestleyi* that Sustains Oxygenic Photosynthesis in the Presence of Hydrogen Sulfide

INTRODUCTION

Cyanobacterial production of O_2 from oxygenic photosynthesis oxidized the Earth's atmosphere during the Great Oxidation Event 2.4 billion years ago, which changed various elemental cycles, including the sulfur cycle (Kasting, 2013). Specifically, the Great Oxidation Event increased oxidative weathering, leading to a large flux of sulfate to the ocean, allowing more microbial sulfate reduction and increased sulfide concentrations in high productivity environments (Canfield, 1998). The biogeochemistry of oxygenic photosynthesis is influenced by the presence of sulfide, which normally inhibits oxygenic photosynthesis, presenting a challenge for cyanobacteria living in the presence of sulfide in diverse environments since the Great Oxidation Event. About 750 million years ago, a global "Snowball Earth" glaciation is associated with significant sulfate reduction (Dahl et al., 2011; Parnell & Boyce, 2017; Wang et al., 2019), suggesting that cyanobacteria may have sustained oxygenic photosynthesis in cold, sulfidic environments during at least one global glaciation. The only cold environment where sulfide-tolerant oxygenic photosynthesis has been described is Middle Island Sinkhole in Michigan, USA (8-10°C) (Voorhies et al., 2012), even though cyanobacteria are the dominant primary producers in many extreme cold ecosystems (Quesada & Vincent, 2012). Sulfide-tolerant O_2 production has also been documented in springs and sink holes in warmer environments, including Frasassi

springs in Italy (Macalady et al., 2006; Klatt et al., 2016) and Little Salt Spring sinkhole in Florida, USA (Beer et al., 2017; Hamilton et al., 2018).

Prior work identified a cyanobacterium, *P. pseudopriestleyi* in ice-covered Lake Fryxell, Antarctica, that creates a thin layer of O_2 in sulfidic pore water, an "oxygen oasis" (Jungblut et al., 2016; Sumner et al., 2015). Compared to other environments where sulfide-tolerant oxygenic photosynthesis has been observed, Lake Fryxell is colder (2.4–2.7°C), and its high latitude leads to months of continuous winter darkness. Based on sulfide and O_2 fluxes, the oxygen oasis is expected to disappear during the dark winter months when photosynthesis does not occur. The increasing availability of light in spring allows benthic mats dominated by *P. pseudopriestleyi* to transition from sulfidic to oxic conditions due to photosynthetic O_2 production (Sumner et al., 2015). However, the mechanism by which *P. pseudopriestleyi* tolerates sulfide and initiates this redox change is unknown.

Sulfide inhibits oxygenic photosynthesis by blocking electron donation interaction between H_2O and the Mn_4CaO_5 cluster at the oxygen-evolving complex (OEC) in the D1 protein of Photosystem II (PS II). Cyanobacteria respond to sulfide in one of four ways: (1) complete inhibition of oxygenic photosynthesis, (2) continued but partial inhibition of oxygenic photosynthesis, (3) simultaneous oxygenic photosynthesis and anoxygenic photosynthesis, or (4) shutting down of oxygenic photosynthesis and use of anoxygenic photosynthesis until enough sulfide is oxidized that oxygenic photosynthesis can start again (Oren et al., 1979; Cohen et al., 1986; S. R. Miller & Bebout, 2004) (Figure 1.1). Response 1 results in a cessation of the electron flow by blocking the OEC in PS II with no source of electrons for Photosystem I (PS I). Response 2 consists of the sulfide quinone reductase (SQR) protein oxidizing sulfide to elemental sulfur and

providing electrons to PS I to perform anoxygenic photosynthesis. Response 3 involves modification of the D1 protein by increasing *psbA* expression or switching to an alternate variant of the D1 protein allowing reduced O_2 production and electron flow to and functioning of PS I. Response 4 involves the mechanism from response 3 with SQR providing additional electrons to PS I (Klatt, Haas, et al., 2015; Beer et al., 2017; Dick et al., 2018).

The direct oxidation of sulfide to sulfur by SQR removes the oxygenic photosynthesis inhibiting chemical from the surrounding environment whether or not it provides electrons to fuel anoxygenic photosynthesis. Thus, its activity can eventually allow oxygenic photosynthesis. Alternatively, a cyanobacterium that can maintain oxygenic photosynthesis in sulfidic conditions produces O_2, which can react with the sulfide in its environment through the abiotic reaction:

$$H_2S + 2\ O_2 \rightarrow SO_4^{2-} + 2\ H^+ \qquad \text{(Reaction 1)}$$

SO_4^{2-} does not interfere with the OEC and does not affect oxygenic photosynthesis. Therefore, depending on the rate of oxygenic photosynthesis and kinetics of abiotic sulfide oxidation, which may be slow at low temperatures, the production of a small amount of O_2 might lead to a positive feedback loop by reducing oxygenic photosynthesis inhibition.

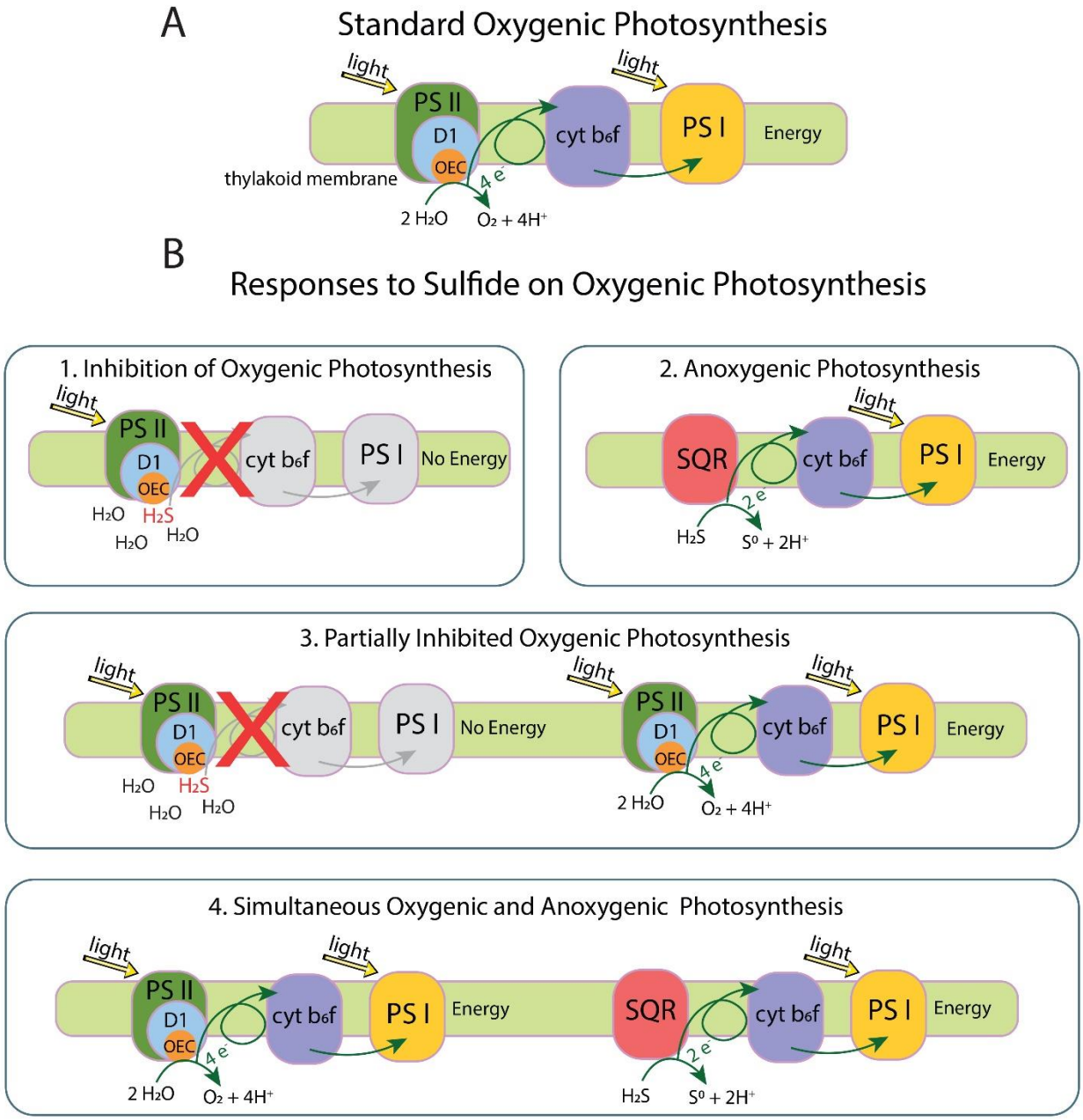


Figure 1.1 (A) standard electron flow in oxygenic photosynthesis. (B1) inhibition of oxygenic photosynthesis by sulfide blocking the oxygen-evolving complex (OEC) in the D1 protein, prohibiting H_2O from interacting with the OEC and from electrons to flow through the system. (B2) anoxygenic photosynthesis occurs when sulfide quinone reductase (SQR) extracts electrons from sulfide and passes them along the photosynthetic electron transport chain. Sulfide is oxidized to S^0, which does not interact with the OEC. (B3) partially inhibited oxygenic photosynthesis occurs when some D1 proteins are blocked by sulfide while others extract electrons from H_2O and pass them along to carry out oxygenic photosynthesis. Excess O_2 produced from oxygenic photosynthesis will oxidize sulfide to sulfate, removing sulfide from the

environment. Some cyanobacteria increase *psbA* expression to replace the D1 protein in response to stress, which may support this response. (B4) some cyanobacteria can do oxygenic photosynthesis and anoxygenic photosynthesis at the same time, or alternate between the two processes, until sulfide is fully depleted. Oxygenic photosynthesis protein complex image modified from the Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa & Goto, 2000; Kanehisa, 2019).

The D1 protein is directly affected by sulfide, and many cyanobacteria have multiple *psbA* genes encoding this protein. One model for reducing sulfide inhibition is altering the production of the D1 protein, which is essential for extracting electrons from H_2O for oxygenic photosynthesis. The effects of sulfide on D1 protein production have not been studied in cyanobacteria, although the effects of low light and 3-(3,4-dichlorophenyl)-1,1-dimehtylurea (DCMU) have (Bishop, 1958; Mulo et al., 2009, 2012). DCMU blocks the electron flow from PS II to PS I, which is similar to sulfide due to preventing electron flow to PS I by blocking the extraction of electrons from H_2O. Understanding how DCMU affects oxygenic photosynthesis may provide insights as to how cyanobacteria deal with sulfide. In response to inhibition from light or DCMU, some cyanobacteria increase their expression of *psbA* to support replacement of the D1 protein (Summerfield et al., 2008; Mulo et al., 2009, 2012). Cyanobacteria can also respond to some environmental stressors by using different types of D1 proteins. The D1 proteins are divided into four groups: One standard group present in all cyanobacteria that is used under normal conditions, a version specialized for microaerobic conditions, a version for red light, and a nonfunctional version to support nitrogen fixation (Cardona et al., 2015). To allow flexible responses to environmental conditions, most cyanobacteria contain multiple copies of *psbA*, which allows cyanobacteria to increase expression or use the version of the protein most appropriate for the environmental conditions (Mulo et al., 2009).

Instead of oxidizing sulfide with excess O_2, some cyanobacteria directly oxidize sulfide to sulfur with SQR (Garlick et al., 1977; Bronstein et al., 2000). Cyanobacteria able to do anoxygenic photosynthesis pass these electrons through the quinone pool to PS I (Cohen, Jørgensen, et al., 1975; Cohen, Padan, et al., 1975). Some cyanobacteria can switch between oxygenic photosynthesis and anoxygenic photosynthesis, while others can do both simultaneously, with both H_2O and sulfide donating electrons (Cohen, Jørgensen, et al., 1975; Cohen et al., 1986). However, some cyanobacteria, such as *Aphanothece halophytica*, that have *sqr* can survive in the presence of sulfide but not grow, suggesting that the electrons are not shuttled to an energy-producing pathway (Cohen, Padan, et al., 1975; Bronstein et al., 2000).

If *P. pseudopriestleyi* can create an oxygen oasis in the sulfidic benthic environment in Lake Fryxell (Sumner et al., 2015), then it must be able to perform photosynthesis in the presence of sulfide or oxidize the sulfide before producing O_2. To evaluate the genomic potential for tolerance mechanisms, we obtained a metagenome-assembled genome (MAG) from a natural sample and an enrichment culture of *P. pseudopriestleyi*. We use this MAG to evaluate the genomic potential for survival in elevated sulfide, low light, and cold temperatures and build on the phylogenetic characterization of the *P. pseudopriestleyi* 16S rRNA gene sequence in Jungblut et al. (Jungblut et al., 2016). This study presents an investigation into the metabolic potential of this MAG to gain a better understanding of the connection between metabolic potential and environmental function with implications for primary productivity in sulfidic environments throughout Earth history.

Site Description

Lake Fryxell is a perennially ice-covered lake located at 77° 36’ S 162° 6’ E in the McMurdo Dry Valleys of east Antarctica. It is 5 times 1.5 km with a maximum depth of ~20 m (Green & Lyons, 2009). The floor of Lake Fryxell is covered with photosynthetically active microbial mats to depths of ~10 m (Sumner et al., 2015). At the 9.8 m sampling depth, the lake floor is covered by flat prostrate mats dominated by a single diatom species and *P. pseudopriestleyi*.

The lake receives water from thirteen glacial meltwater streams (McKnight et al., 1999). Evaporation and ablation from the surface allow for water balance, as no streams flow out from the lake (Lawrence & Hendy, 1985). Salts remaining in the lake water and the historical balance of inflow and sublimation have led to density stratification of the lake water (Vincent, 1981; A. D. Jungblut et al., 2016). At the 9.8 m sampling depth, the salinity is approximately 4 mS cm^{-1} with 1.2 M NaCl (Laybourn-Parry et al., 1997; Lyons et al., 1999; Roberts et al., 2000), and sulfide (H_2S + HS^-) was present based on diver observations (rotten egg smell). Above 9 m, sulfide was undetectable, and below 10 m, sulfide concentration was 69.9 µM and increased to 1210 µM at the bottom of the lake (Green & Lyons, 2009).

Water temperature varies from 2.4 to 2.7°C, and pH varies from 7.50 to 7.52 along a dive transect established from 8.9 m to 11.0 m in depth (Jungblut et al., 2016). The water column has a sharp oxycline; dissolved oxygen is super saturated below the ice cover to a depth of 9.1 m where it decreases rapidly (Vincent, 1981; Spigel & Priscu, 2013). At 9.8 m depth, there is no O_2 in the water column, but a microlayer of 50 µmol O_2 L^{-1} is at least transiently present in the top ~1 mm of the mat [12].

Irradiance is highly seasonal. The lake experiences four months of darkness in the winter, followed by two months of diurnal light variations in spring, four months of continuous summer illumination, and two months of autumn diurnal light variations (McKnight et al., 1999). Even at peak illumination during the summer, only 0.5% - 3% of incident light penetrates the ice cover, and light is further attenuated by planktonic communities in the water column (Howard-Williams et al., 1998; Sumner et al., 2015). A daily average of 1–2 µmol photons m^{-2} s^{-1} reaches the mat at ~10 m water depth (Sumner et al., 2015). The most penetrating waveband at ~10 m in Lake Fryxell is 520–580 nm (Lizotte & Priscu, 1992).

MATERIALS AND METHODS

Sulfide Concentrations

Water samples (12 mL) for total sulfide analysis were collected on January 17, 2020 from within 10 mm of the lake floor by a diver using syringes. On return to the surface, samples were immediately injected into glass tubes and preserved for later analysis with 0.2 mL of 2 M zinc acetate. On return to New Zealand, samples were analyzed using the methylene blue method from Standard Methods for the Examination of Water and Wastewater (21st Edition) from the American Public Health Association (Miner, 2006).

Field Work

Samples of the microbial mat were collected in November 2012. Divers accessed the lake through a hole melted in the ice cover and sampled the microbial mat at 9.8 m depth (Hillman, 2013). Sampling and dissection of samples were performed using sterile technique. Briefly, divers used spatulas to cut samples of the mat and transfer them to plastic boxes under-water. In the field lab, mat samples were dissected according to morphology and pigmentation of layers. A blue-green biofilm from 9.8 m water depth was dominated by a single cyanobacterial morphotype based on field microscopy. Samples of this biofilm were peeled off the top of a prostrate microbial mat using sterile forceps (Jungblut et al., 2016). Samples for metagenomic sequencing were preserved in the field within a few hours of collection with Xpedition Soil/Fecal DNA MiniPrep kit (Zymo Research, Irvine, CA, USA) and stored on ice. They were shipped to UC Davis where they were stored at -80°C until they were processed for sequencing. One subsample for culturing was transferred to a sterile plastic vial filled with lake water that was filtered through a sterile syringe and 0.2 µm syringe filter. The cyanobacteria culture was stored at ambient indoor light at the lakeside laboratory for approximately ten days and then shipped to the Natural History Museum, London, UK where it was grown in BG11 liquid medium at 10°C, and 24 h light at an average of 9.25 µmol photons m^{-2} s^{-1} (Rippka et al., 1979).

DNA Extraction and Sequencing

DNA from the blue-green biofilm subsample was extracted from frozen samples using an Xpedition Soil/Fecal DNA MiniPrep kit (Zymo Research, Irvine, CA, USA) as per manufacturer

instructions. Metagenomic sequencing on mat samples was performed at the Genome Center DNA Technologies Core at the University of California using the Illumina HiSeq 2500, PE250 platform. Illumina's Nextera DNA Kit was used for library preparation (Oligonucleotide sequences (c) 2007–2013 Illumina, Inc., San Diego, CA USA, Country).

DNA was extracted from the cyanobacteria enrichment culture from Lake Fryxell using the MoBio Powerbio DNA extraction kit according to the manufacturer's instructions. The culture was sequenced on the Illumina HiSeq platform (2000 PE 100, Illumina, Inc., San Diego, CA, USA) at the University of Michigan DNA Sequencing Core.

Bioinformatics Analysis

The sequencing of the biofilm sample from 9.8 m depth resulted in 3,911,904 reads. The biofilm sample data were quality filtered to Q20, and forward and reverse reads were joined using PEAR v0.9.6 (Zhang et al., 2014). Singletons and replicates with fewer than 10,000 reads were removed from downstream analysis. Sequencing of the culture resulted in 47,243,886 reads. For the lab culture data, trimmomatic v0.36 (Bolger et al., 2014) was used to trim sequencing adaptors with a LEADING and TRAILING parameter of 3, a SLIDINGWINDOW parameter of 4:15, and a MINLEN parameter of 25. The interleave-reads.py script from khmer v2.1.2 (Crusoe et al., 2015) was used to interleave the reads. The biofilm sample and lab-cultured sample were assembled separately and by coassembly with MEGAHIT v1.1.2 (Li et al., 2015). QUAST v4.4 (Gurevich et al., 2013) was used to generate assembly statistics. Mapping of both sets of reads to the coassembly was done with bwa v2.3 and samtools v1.9 (Li et al., 2009, 2013).

Anvi'o v2.2.2 was used to bin and visualize the samples with the CONCOCT binning algorithm (Eren et al., 2015; Delmont & Eren, 2018). CheckM v1.0.7 (Parks et al., 2015) was used to assess the quality of the bins and assign phylogeny based on marker genes of interest. Taxonomy of the bins was assigned using the Genome Taxonomy Database (GTDB-tk) on KBase GhostKOALA v2.2 and Prokka v1.11 were used to annotate genes in the cyanobacterial bin of interest (Kanehisa & Goto, 2000; Matsen et al., 2010; Sukumaran & Holder, 2010; Hyatt et al., 2010; Price et al., 2010; Eddy, 2011; Ondov et al., 2016; Arkin et al., 2018; Jain et al., 2018; Harris et al., 2020; Chaumeil et al., 2020). To refine the bin, spacegraphcats was used to extract additional content of the bin with a k size of 21 (Brown et al., 2020). The code used for the analyses presented here is available at https://github.com/jessicalumian/fryxell-phormidium.

A custom Basic Local Alignment Search Tool (BLAST) database containing a reference amino acid sequence was constructed using the makeblastdb command in BLAST+, and D1 protein sequences from the *P. pseudopriestleyi* MAG were retrieved by using a blastx search with an e value of 1e-20 (Camacho et al., 2009). Subsequent analysis was performed on XSEDE Cipres Science Gateway (Miller et al., 2010). D1 protein sequence fragments were aligned to D1 and D2 protein sequences compiled by Cardona et al. (Cardona et al., 2015) with ClustalW v2.1 using standard parameters (Larkin et al., 2007). A best-fit model of evolution of LG + G4 was selected with ModelTest-ng v0.1.5 (Darriba et al., 2020) using maximum likelihood for the tree topology parameter and the discrete gamma rate categories option was selected for the candidate model's rate heterogeneity parameter. A phylogenetic tree was generated with RAxML-HPC2 v8.2.12 (Stamatakis, 2014) using a protein gamma model with an LG substitution matrix and 1000 bootstrap iterations.

To determine if the SQR in the *P. pseudopriestleyi* is type I or type II, the SQR amino acid sequence from the MAG was aligned to type I and type II references from Shahak and Hauska 2008 (Shahak & Hauska, 2008). The sequences were aligned with ClustalW v2.1 (Larkin et al., 2007) and then trimmed with TrimAl v1.2.59 (Capella-Gutiérrez et al., 2009) using standard parameters. A best-fit model of evolution of WAG + G4 was selected with ModelTest-ng v0.1.5 (Darriba et al., 2020) using maximum likelihood for the tree topology parameter and the discrete gamma rate categories option was selected for the candidate model's rate heterogeneity parameter. A phylogenetic tree was generated with RAxML-HPC2 v8.2.12 (Stamatakis, 2014) using a protein gamma model, WAG substitution matrix and 1000 bootstrap iterations.

The average nucleotide identity (ANI) was calculated between the *P. pseudopriestleyi* MAG and available Antarctic cyanobacteria genomes (*Leptolyngbya* sp. BC1307, accession number NRTA00000000.1, *Aurora vandensis*, accession number JAAXLU010000000, *Synechococcus* sp. CS-601, accession number CP018091, and *Phormidesmis priestleyi* ULC007, accession number MPPI01000000) (Lara et al., 2017; Chrismas, Williamson, et al., 2018a; Tang et al., 2019; Grettenberger et al., 2020). The ANI was also calculated between the MAG and an Arctic cyanobacterium closely related to an Antarctic strain (*Phormidesmis priestleyi* BC1401, accession number LXYR01000000) (Chrismas et al., 2016), and the closest related genome according to 16S rRNA gene sequence (*Oscillatoria acuminata* PCC 6304, accession number CP003607.1) (Shih et al., 2013). Calculations were performed using the ANI calculator from the Kostas lab using the default parameters (Rodriguez & Konstantinidis, 2016). The alignment options required a 700 bp minimum length, 70% minimum identity, and 50 minimum alignments. The fragment option window size was set to 1000 bp with a step size of 200 bp.

To identify the presence *P. pseudopriestleyi* in other locations, the 16S rRNA gene sequence reported in Jungblut et al. (Jungblut et al., 2016) (accession number KT347094) was aligned with blastn to sequences from Jungblut et al. (Jungblut et al., 2005) (accession numbers AY541534 and AY541575) and Taton et al. (Taton et al., 2006) (accession number DQ181670). The phylogenetic tree of the 16S rRNA gene sequence of *P. pseudopriestleyi* and the most closely related operational taxonomic units (OTUs) in Jungblut et al. (Jungblut et al., 2016) was used for context.

RESULTS

Sulfide Concentrations

Total sulfide ($[H_2S] + [HS^-]$) was measured to be <0.01 mg L^{-1} at 9.5 m and above. At 9.8 m, total sulfide was 0.091 mg L^{-1} and rose to 2.2 mg L^{-1} at 10.7 m depth. See Table 1.1 for all measurements.

Table 1.1 Total sulfide measurements on Lake Fryxell water samples collected on January 17, 2020.

Depth (m)	Total Sulfide (mg L^{-1})
8	<0.01
9	<0.01
9.5	<0.01
9.8	0.091
10.1	0.571
10.3	0.885
10.7	2.238

Assembly and Binning Statistics

Thirty-one bins were retrieved from the coassembly from the environmental biofilm samples and the strain *P. pseudopriestleyi* (see Table 1.2 for assembly statistics and Table 1.3 binning statistics), but there was only one bin belonging to the phylum Cyanobacteria based on comparison with GTDB-tk on KBASE database (Jungblut et al., 2016). This bin is 5.97 MB with 42,524 reads, 678 contigs, 4738 protein coding genes, and a GC content of 62.32%. The longest contig generated was 400,416 bp. The closest relative to *P. pseudopriestleyi* according to 16S rRNA OTUs, *Oscillatoria acuminata* PCC 6304 has a genome size of 7.7 MB and contains 5,687 protein coding genes. The *P. pseudopriestleyi* MAG is 91.73% complete, has 1.35% contamination and 8.33% strain heterogeneity. Because the bin used for analysis is incomplete, it may not contain genes that are present in the organism's genome. However, the MAG is sufficiently complete that it can be used to analyze the genomic potential for key metabolisms.

Table 1.2 Quality metrics of coassembly generated from sequencing of the mat and laboratory culture samples. All statistics are from QUAST v4.4, except for mapping statistics which were generated from samtools v1.9 using the flagstat parameter.

Metric	Coassembly
Total number of contigs	137,226
Longest contig length	453,833
Total length (bp)	175,382,959
GC content (%)	61.61
N50	1,441
Reads mapped from culture (%)	79.80
Reads mapped from mat (%)	54.27
Total reads mapped (%)	77.85
Number of contigs ≥ 0 bp	336,522
Number of contigs ≥ 1000 bp	41,881
Number of contigs ≥ 5000 bp	2,359
Number of contigs ≥ 10.000 bp	814
Number of contigs ≥ 25.000 bp	303
Number of contigs ≥ 50.000 bp	152

Table 1.3 Quality metrics of the *P. pseudopriestleyi* metagenome-assembled genome (MAG). All statistics are from QUAST v4.4, except for completion and contamination statistics which were generated from CheckM v1.0.7 and the number of protein coding genes from Prokka v1.11.

Metric	MAG
Total number of contigs	678
Longest contig length	44,245
Total length (bp)	5,965,908
GC content (%)	47.43
N50	10908
Completion (%)	91.73
Contamination (%)	1.35
Number of protein coding genes	4,738
Number of contigs ≥ 0 bp	678
Number of contigs ≥ 1000 bp	678
Number of contigs ≥ 5000 bp	458
Number of contigs ≥ 10.000 bp	203
Number of contigs ≥ 25.000 bp	20
Number of contigs ≥ 50.000 bp	0

Taxonomic Assignment of the MAG

The MAG obtained from *P. pseudopriestleyi* was classified as family Oscillatoriaceae and genus Oscillatoria based on comparison with GTDB-tk on KBASE database (Jungblut et al., 2016). The *P. pseudopriestleyi* MAG and *O. acuminata* PCC 6304 had an ANI of 88.99%, with a standard deviation of 3.25% based on 14,129 fragments. The *P. pseudopriestleyi* MAG and *P. priestleyi* BC1401 had an ANI of 75.69% with a standard deviation of 5.94% based on 75 fragments. The ANI score between the *P. pseudopriestleyi* MAG and *P. priestleyi* ULC007 was 72.03% with a standard deviation of 3.65% based on 72 fragments. Each of the ANIs between the *P. pseudopriestleyi* MAG and *A. vandensis*, *Synechococcus* sp. CS-601, and *Leptolyngbya* sp. BC1307 were less than 70%. These results suggest that the *P. pseudopriestleyi* MAG was most similar to *O. acuminata* (Rodriguez & Konstantinidis, 2016; Jain et al., 2018; Barco et al., 2020). However, additional cyanobacteria strains in Oscillatoriaceae need to be isolated and sequenced from

Antarctica to allow phylogenomic interference to better resolve the relationship between *Phormidium* and *Oscillatoria*.

Photosynthetic and Electron Transport Machinery

The *P. pseudopriestleyi* MAG contains genes for the pigments phycoerythrocyanin, phycocyanin, and allophycocyanin, which have absorption peaks at 575, 620, and 650 nm, respectively, and lacks the gene for phycoerythrin, the light-harvesting protein that absorbs wavelengths from 495 to 560 nm (Bogorad, 1975) (Table 1.4). The genes for the D1/D2 protein cluster in PS II (*psbA* and *psbD*) are present along with two genes for proteins that hold the Mn_4CaO_5 OEC cluster in place (*psbP* and *psbO*) (Cardona et al., 2015). The MAG contains four of the eight subunits that make up the cytochrome b_6f complex (*petB*, *petD*, *petA*, and *petC*). Notably, *petA* codes for apocytochrome f and *petC* codes for the iron-sulfur subunit within the b_6f complex. The genes for the remaining subunits, *petL*, *petM*, *petN*, and *petG*, were not identified in the bin (Table 1.4). The majority of PS I genes are present in the MAG, including the core chlorophyll dimer made up of *psaA* and *psaB* (Table 1.4). Genes for both ferredoxin (*petF*) and ferredoxin-NADP+ reductase (*petH*) are present. The MAG does not contain *petE*, which codes for the electron transport protein plastocyanin that connects PS II to PS I, however cytochrome c6 encoded by *petJ* can perform the same function (Durán et al., 2004). The MAG has all the genes necessary to create an F-type ATPase. Notably, the MAG has a type II *sqr* (Figure 1.2), which is necessary for sulfide oxidation in anoxygenic photosynthesis. The MAG also

contains genes encoding arsenic resistance (*arsR* and *acr3*) that may be involved with transcriptional regulation of *sqr* (Nagy et al., 2014).

The MAG contains all genes for a type 1 NADH dehydrogenase except for *ndhN*, subunit n of the complex. The succinate dehydrogenase gene *sdhD*, encoding a membrane anchor subunit, is absent but *sdhA*, *shdB*, and *sdhC* are present. The genes *ctaC*, *ctaD*, and *ctaE* for aa3-type cytochrome c oxidase genes are present but *ctaF* is not. The genes for cytochrome bd-quinol oxidase *cydA* and *cydB* are present, but *cydX* is not. Genes for alternative respiratory terminal oxidase and plastid terminal oxidase were not found in the MAG (Mullineaux, 2014; Lea-Smith et al., 2016).

Because of the importance of the D1 protein to oxygenic photosynthesis, we examined *psbA* sequence in more detail. Although spacegraphcats software was used to refine the MAG, a full copy of the gene for the D1 protein could not be obtained. Fragments of the D1 protein sequence are present, including the C-terminal and N-terminal portions ranging between 80 and 264 amino acids out of the full length 360 amino acids protein sequence. Additionally present are all seven of the amino acids involved in ligating with the Mn_4CaO_5 cluster in the OEC: Asp170, Glu189, His332, Glu333, His337, Asp342, and Ala344 (Vermaas et al., 1990; Murray, 2012). Each of these amino acids were present in fragments that contained the appropriate part of the protein sequence for *psbA*. A phylogenetic tree was constructed of the MAG's D1 fragments and reference sequences from all four D1 groups, and all fragments grouped closely with group 4 D1 proteins, demonstrating that the MAG does not contain an alternative version of the D1 protein (Figure 1.3). Where overlapping, the sequences of the fragments are not identical, indicating the presence of at least two copies of the D1 gene in the MAG.

Table 1.4 Phycobilisome, photosynthesis, and respiratory machinery genes present in the *P. pseudopriestleyi* MAG generated with GhostKoala v2.2.

Complex	Gene	Presence	Function
Allophycocyanin	*apcA*	Yes	Allophycocyanin α subunit
	apcB	Yes	Allophycocyanin β subunit
	apcC	Yes	Phycobilisome core linker protein
	apcD	Yes	Allophycocyanin-B
	apcE	Yes	Phycobilisome core-membrane linker protein
	apcF	Yes	Phycobilisome core component
Phycocyanin / Phycoerythrocyanin	*cpcA*	Yes	Phycocyanin α chain
	cpcB	Yes	Phycocyanin β chain
	cpcC	Yes	Phycocyanin-associated rod linker protein
	cpcD	No	Phycocyanin-associated, rod
	cpcE	Yes	Phycocyanobilin lyase α subunit
	cpcF	Yes	Phycocyanobilin lyase β subunit
	cpcG	Yes	Phycobilisome rod-core linker protein
Phycoerythrin	*cpeA*	No	Phycoerythrin α chain
	cpeB	No	Phycoerythrin β chain
	cpeC	No	Phycoerythrin-associated linker protein
	cpeD	No	Phycoerythrin-associated linker protein
	cpeE	No	Phycoerythrin-associated linker protein
	cpeR	No	Phycoerythrin-associated linker protein
	cpeS	No	Phycoerythrin-associated linker protein
	cpeT	No	CpeT protein
	cpeU	No	Billin biosynthesis protein
	cpeY	No	Billin biosynthesis protein
	cpeZ	No	Billin biosynthesis protein
Photosystem II	*psbA*	Yes	Photosystem II P680 reaction center D1 protein
	psbD	Yes	Photosystem II P680 reaction center D2 protein
	psbC	Yes	Photosystem II CP43 chlorophyll apoprotein
	psbB	Yes	Photosystem II CP47 chlorophyll apoprotein
	psbE	Yes	Photosystem II cytochrome b559 subunit α
	psbF	Yes	Photosystem II cytochrome b559 subunit β
	psbL	Yes	Photosystem II PsbL protein
	psbJ	Yes	Photosystem II PsbJ protein
	psbK	Yes	Photosystem II PsbK protein
	pskM	Yes	Photosystem II PsbM protein
	psbH	Yes	Photosystem II PsbH protein
	psbI	Yes	Photosystem II PsbI protein
	psbO	Yes	Photosystem II oxygen-evolving enhancer protein 1
	psbP	Yes	Photosystem II oxygen-evolving enhancer protein 2
	psbQ	No	Photosystem II oxygen-evolving enhancer protein 3
	psbR	No	Photosystem II 10 kDa protein
	psbS	No	Photosystem II 22kDa protein
	psbT	Yes	Photosystem II PsbT protein
	psbU	Yes	Photosystem II PsbU protein
	psbV	Yes	Photosystem II cytochrome c550

	psbW	No	Photosystem II PsbW protein
	psbX	Yes	Photosystem II PsbX protein
	psbY	Yes	Photosystem II PsbY protein
	psbZ	Yes	Photosystem II PsbZ protein
	Psb27	Yes	Photosystem II Psb27 protein
	psb28	Yes	Photosystem II 13kDa protein
	psb28-2	No	Photosystem II Psb28-2 protein
Photosystem I	*psaA*	Yes	Photosystem I P700 chlorophyll a apoprotein A1
	psaB	Yes	Photosystem I P700 chlorophyll a apoprotein A2
	psaC	Yes	Photosystem I subunit VII
	psaD	Yes	Photosystem I subunit II
	psaE	Yes	Photosystem I subunit IV
	psaF	Yes	Photosystem I subunit III
	psaG	No	Photosystem I subunit V
	psaH	No	Photosystem I subunit VI
	psaI	Yes	Photosystem I subunit VIII
	psaJ	No	Photosystem I subunit IX
	psaK	Yes	Photosystem I subunit X
	psaL	Yes	Photosystem I subunit XI
	psaM	Yes	Photosystem I subunit XII
	psaN	No	Photosystem I subunit PsaN
	psaO	No	Photosystem I subunit PsaO
	psaX	No	Photosystem I 4.8kDa protein
Cytochrome b_6f Complex	*petB*	Yes	cytochrome b_6
	petD	Yes	cytochrome b_6f complex subunit 4
	petA	Yes	apocytochrome f
	petC	Yes	cytochrome b_6f complex iron-sulfur subunit
	petL	No	cytochrome b_6f complex subunit 6
	petM	No	cytochrome b_6f subunit 7
	petN	No	cytochrome b_6f complex subunit 8
	petG	No	cytochrome b_6f complex subunit 5
Photosynthetic Electron Transport Chain	*petE*	No	plastocyanin
	petF	Yes	ferredoxin
	petH	Yes	ferredoxin-$NADP^+$ reductase
	petJ	Yes	cytochrome c6
F-type ATPase	*atpD*	Yes	H^+/Na^+ transporting ATPase subunit β
	atpA	Yes	F-type H^+/Na^+ transporting ATPase subunit α
	atpG	Yes	H^+ transporting ATPase subunit γ
	atpH	Yes	F-type H^+ transporting ATPase subunit δ
	atpC	Yes	F-type H^+ transporting ATPase subunit ε
	atpE	Yes	F-type H^+ transporting ATPase subunit c
	atpB	Yes	F-type H^+ transporting ATPase subunit a
	atpF	Yes	F-type H^+ transporting ATPase subunit b
	ndhC	Yes	NADH-quinone oxidoreductase subunit 3
	ndhK	Yes	NADH-quinone oxidoreductase subunit K
	ndhJ	Yes	NADH-quinone oxidoreductase subunit J
	ndhH	Yes	NADH-quinone oxidoreductase subunit H

NADH Dehydrogenase	*ndhA*	Yes	NADH-quinone oxidoreductase subunit 1
	ndhI	Yes	NADH-quinone oxidoreductase subunit I
	ndhG	Yes	NADH-quinone oxidoreductase subunit 6
	ndhE	Yes	NADH-quinone oxidoreductase subunit 4L
	ndhF	Yes	NADH-quinone oxidoreductase subunit 5
	ndhD	Yes	NADH-quinone oxidoreductase subunit 4
	ndhB	Yes	NADH-quinone oxidoreductase subunit 2
	ndhL	Yes	NADH-quinone oxidoreductase subunit L
	ndhM	Yes	NADH-quinone oxidoreductase subunit M
	ndhN	No	NADH-quinone oxidoreductase subunit N
	hoxE	Yes	bidirectional [NiFe] hydroganse diaphorase subunit
	hoxF	Yes	bidirectional [NiFe] hydroganse diaphorase subunit
	hoxU	Yes	bidirectional [NiFe] hydroganse diaphorase subunit
Succinate Dehydrogenase	*sdhC*	Yes	H^+/Na^+ transporting ATPase subunit β
	sdhD	No	F-type H^+/Na^+ transporting ATPase subunit α
	sdhA	Yes	H^+ transporting ATPase subunit γ
	sdhB	Yes	F-type H^+ transporting ATPase subunit δ
Cytochrome c oxidase	*ctaC*	Yes	cytochrome c oxidase subunit 2
	ctaD	Yes	cytochrome c oxidase subunit 1
	ctaE	Yes	cytochrome c oxidase subunit 3
	ctaF	No	cytochrome c oxidase subunit 4
Cytochrome bd complex	*cydA*	Yes	cytochrome bd ubiquinol oxidase subunit I
	cydB	Yes	cytochrome bd ubiquinol oxidase subunit II
	cydX	No	cytochrome bd ubiquinol oxidase subunit X

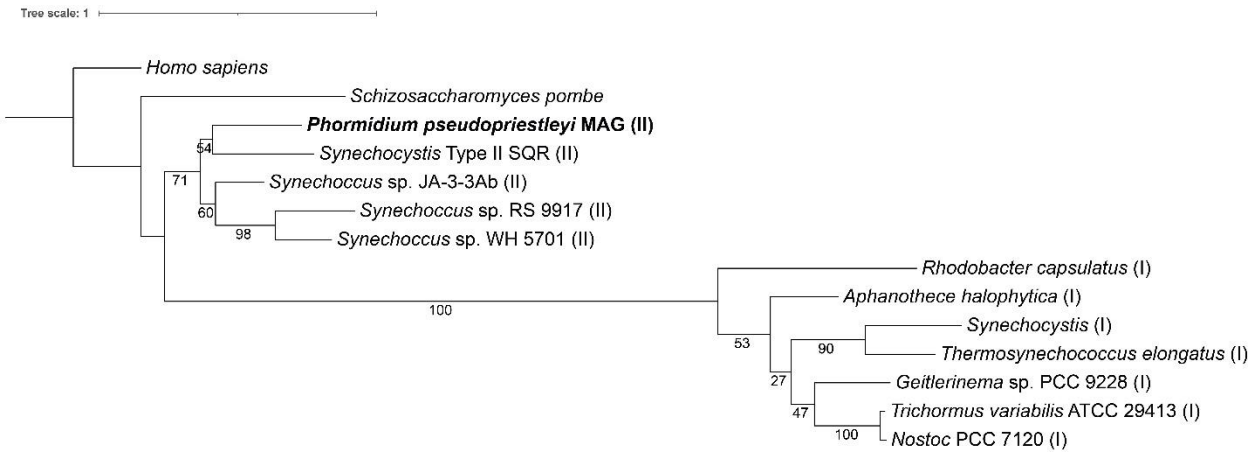


Figure 1.2 A maximum likelihood tree of type I and II SQR amino acid sequences from Shahak and Hauska, 2008. Homo sapiens (accession number AAH16836), *Schizosaccharomyces pombe* (accession number CAA21882), *Phormidium pseudopriestleyi* (presented in paper), *Synechocystis* SQR-type II (accession number WP_010872226), *Synechococcus* strain sp. JA-3-3Ab (accession number ABD00861), *Synechococcus* sp. RS 9917 (accession number EAQ69368), *Synechococcus* strain WH 5701 (accession number EAQ74835), *Rhodobacter capsulatus* (accession number

CAA66112), *Aphanothece halophytica* (accession number AAF72963), *Synechocystis* SQR-type I (accession number WP_011153573), *Thermosynechococcus elongatus* (accession number WP_011056143), *Geitlerinema* sp. PCC 9228 (formerly known as *Oscillatoria limnetica*, accession number AAF72962), *Trichormus variabilis* ATCC 29413 (formerly known as *Anabaena variabilis* accession number ABA22985), and *Nostoc* PCC 7120 (accession number WP_010998645). A type I SQR is indicated by (I) after the organism name, while SQR type II is indicated by (II).

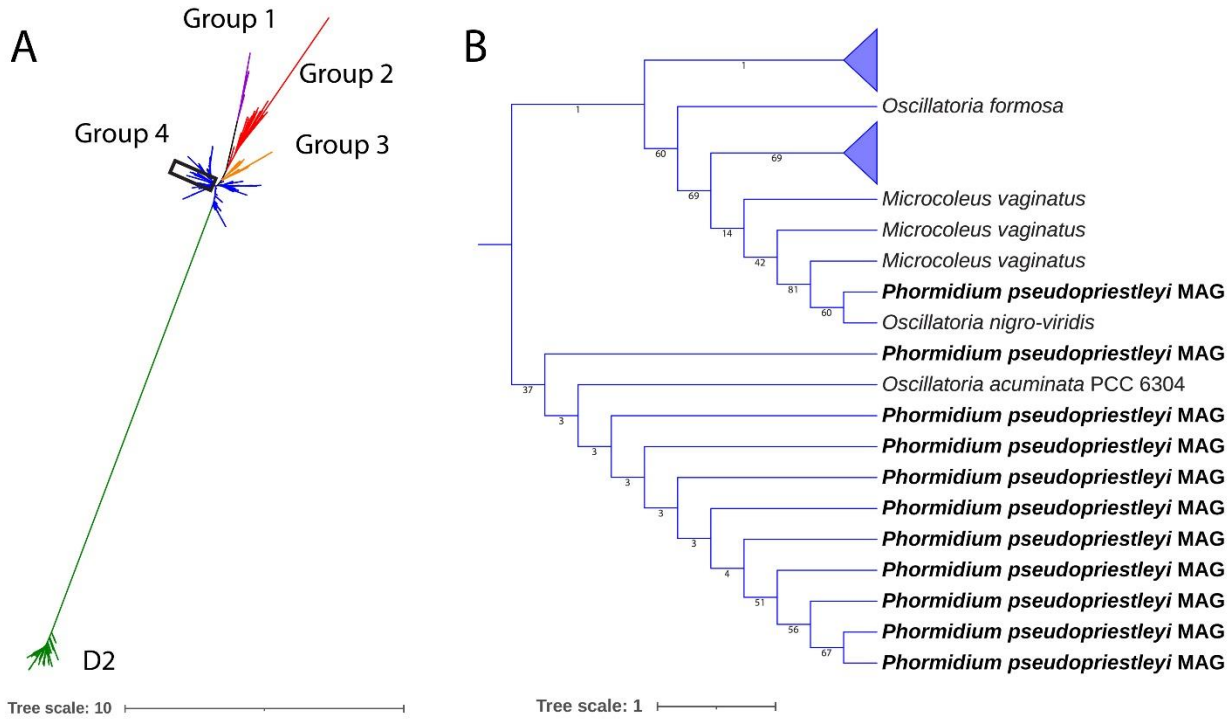


Figure 1.3 (a) a maximum likelihood tree of D1 and D2 proteins of sequences presented in Cardona et al. 2015 and the 11 D1 protein fragments in the *P. pseudopriestleyi* MAG. All *P. pseudopriestleyi* fragments (enclosed in the rectangle) grouped with group 4 D1 proteins. (b) region of the tree showing *P. pseudopriestleyi* and the most closely related D1 proteins.

Metabolic Pathways

The MAG contains genetic capacity for the Calvin cycle for carbon fixation. The MAG also contains genes for glycolysis via the Embden–Meyerhof pathway except for *tpiA* encoding for

triosephosphate isomerase. Capacity for the tricarboxylic acid cycle is present except for genes for 2-oxoglutarate dehydrogenase, which is absent in cyanobacteria, and fumarate hydratase. The MAG contains genes for the pentose phosphate pathway and glycogen and trehalose biosynthesis pathways. It also contains full capacity for the initiation, elongation, and β-oxidation of fatty acids (Table 1.5).

All genes necessary for assimilatory sulfate reduction are present in the MAG (*sat*, *cysNC*, *cysH*, and *sir*), but essential genes for dissimilatory sulfur metabolism are not (*aprAB* and *dsrAB*). Additionally, the MAG has the genes necessary for a sulfate ion transport system through a membrane (*cysPUWA*, *sbp*). The presence of *narB* and *nirA* indicate capacity for assimilatory nitrate reduction, but the MAG does not contain genes for dissimilatory nitrogen metabolism or nitrogen fixation (Varin et al., 2010). Although 19 genes associated with methane metabolism were found in the MAG, most of them are involved with various biosynthesis pathways, and there is not a full pathway for methanogenesis or methanotrophy. Notably, *hdrA2*, *hdrB2*, and *hdrC2*, are present, which code for heterodisulfide reductase, an enzyme typically found in methanogens. Previous work has found these genes at 9.8 m depth in Lake Fryxell, and consistent with the content of the MAG, the capacity for methanogenesis (*hdrD*) was absent (Dillon et al., 2020). The presence of *hdrB* may indicate capacity for flavin-based electron bifurcation (Ramos et al., 2015).

Table 1.5 Number of genes present in the *P. pseudopriestleyi* MAG in functional categories based on KEGG annotations.

Category	Complex or System	Number of Genes in MAG	Total Number of Genes in KEGG Category
Phycobilisome Antenna Proteins	Allophycocyanin	6	6
	Phycocyanin/Phycoerythrin	6	7
	Phycoerythrin	0	11
Photosynthesis Machinery	Photosystem II	22	27
	Photosystem I	10	16
	Cytochrome b_6f complex	4	8
	Photosynthetic electron transport	3	4
	F-type ATPase	8	8
Nitrogen Metabolism	Dissimilatory Nitrate Reduction	0	4
	Assimilatory Nitrate Reduction	2	5
	All Nitrogen Metabolism	9	35
Sulfur Metabolism	Assimilatory Sulfate Reduction	4	7
	Dissimilatory Sulfate Reduction and Oxidation	1	3
	All Sulfur Metabolism	10	54
Carbon Fixation	Carbon Fixation in Photosynthetic Organisms	9	23
Methane Metabolism	Methane Metabolism	19	79

Genes Implicated in the Adaption to Environmental Stress

P. pseudopriestleyi encodes some genes related to osmotic stress. Specifically, the MAG contains several genes related to sodium and potassium antiporters (*nhaS2*, *nhaS3*, *mrpA*, *mrpC*, *trk*, and *ktr*). Additionally, the MAG contains *treZ* and *treY*, which support a trehalose biosynthesis pathway, a compatible solute that has been found to have membrane protective features, particularly in filamentous, mat-forming cyanobacteria strains (Hincha & Hagemann, 2004). Trehalose has also been found in cyanobacteria tolerant of desiccation (Hershkovitz et al., 1991; Page-Sharp et al., 1999; Potts et al., 2005). The MAG does not contain genes related to glucosylglycercol or glycine betaine, which are compatible solutes that have been identified in

halotolerant and halophilic cyanobacteria strains (Hagemann, 2011). Sucrose can also be used as an osmolyte in cyanobacteria, and the MAG contains *spsA*, encoding sucrose phosphate synthase which supports the production of sucrose 6-phosphate, but not *spp*, encoding sucrose phosphate phosphatase, which is necessary to produce sucrose (Kirsch et al., 2019).

Some cyanobacteria synthesize the compounds scytonemin and mycosporine to overcome the harmful effects of long-term UV radiation exposure (Ehling-Schulz & Scherer, 1999). The MAG contains no genes related to the biosynthesis of these compounds, though the pathways are not fully understood (Sorrels et al., 2009; Wada et al., 2013). The photoprotective proteins such as orange carotenoid protein (*ocp*) and fluorescence recovery protein (*frp*) protect against high light stress by converting excess excitation energy to heat (Mullineaux, 2014), but the MAG does not contain either of these genes. Another method of dealing with high light stress is to divert electrons away from the photosynthetic electron transport chain using electron valves to prevent over-reduction of photosynthetic machinery. The MAG contains genetic capacity for cyanobacterial electron valves flavodiiron proteins Flv1-4 and another cyanobacterial bidirectional hydrogenase. Additionally, the MAG contains genes for terminal oxidases cytochrome bd-1 and cytochrome c oxidase, which can act as electron valves (Mullineaux, 2014). Carotenoids are another important molecule for photooxidative stress. The MAG contains *crtE*, *crtB*, *crtP*, *crtH*, *crtQ*, *cruF*, *cruG*, *crtR*, and *crtO*, allowing for carotenoid biosynthesis of myx-ol-2'-dimehtylfucoside, β-carotene, zeaxanthin, echinenone, and 3'-hydroxyechinenone. It is missing *cruE* and *cruH*, and thus does not demonstrate a capacity to produce the carotenoid synechoxanthin (Zhu et al., 2010; Falkowski & Raven, 2013; Mills et al., 2020). Additional carotenoid biosynthesis genes *crtI_1*, *crtI_2*, and *crtI_3* for lycopene and neurosporene phytoene

desaturase are present. Although chlorophyll F has been shown to support near-infrared oxygenic photosynthesis, the gene for chlorophyll F synthase (*chlf*) is not present in the MAG (Behrendt et al., 2015).

The MAG contains genes relating to replication, transcription, and translation associated with cold-tolerant organisms. It has the gene for DNA gyrase (*gyrA*) that helps uncoil DNA that is tightly wound at cold temperatures (Rodrigues & Tiedje, 2008). The MAG contains genes associated with cold-adapted ribosomal function, including ribosome-binding factor A (*rbfA*) as well as genes for translational factors Initiation Factor 1 and 2 (*infA*, *infB*) (Rodrigues & Tiedje, 2008). Genes for a ribosomal rescue system *ssrA* and *smpB* are also present. Genes for delta (12)-fatty-acid desaturase (*desA*) and NADPH-dependent stearoyl-CoA 9-desaturase (*desA3*) are present in the MAG. These produce unsaturated and branched fatty acids, which help organisms maintain membrane integrity at lower temperatures. None of the *csp* cold shock proteins are present in the MAG.

Cyanophycin is a copolymer of aspartic acid and arginine that stores nitrogen for when environmental nitrogen levels become deficient (Watzer & Forchhammer, 2018). The MAG contains genes for both cyanophycin synthetase to build the polymer (*cphA*) and cyanophycinase (*cphB*) to break it down.

DISCUSSION

Sulfide Resistance in P. pseudopriestleyi

Knowing the concentration of sulfide present in the mat at 9.8 m depth is important for understanding the extent of oxygenic photosynthesis inhibition experienced by *P. pseudopriestleyi* in Lake Fryxell during the winter to spring transition. The early spring concentration was likely higher than the 0.091 mg L^{-1} sulfide measured at this depth (Table 1.1), because water samples were collected in mid-January, several months after initiation of oxygenic photosynthesis, which results in sulfide oxidation. Thus, the sulfide concentrations reported here represent minimums for early spring oxygenic photosynthesis inhibition.

Both prior research (Sumner et al., 2015; Jungblut et al., 2016) and our field work demonstrate that *P. pseudopriestleyi* sustains oxygenic photosynthesis in an environment with at least 0.091 mg L^{-1} sulfide which suggests the presence of a tolerance mechanism for sulfide. One option for a sulfide tolerance mechanism is that the D1 protein is repaired and replaced while oxygenic photosynthesis occurs (response 3 or 4 in Figure 1.1). Another possible mechanism is that SQR production supports anoxygenic photosynthesis in the presence of sulfide (response 2 or 4 in Figure 1.1). Additionally, the low irradiance in Lake Fryxell may contribute to *P. pseudopriestleyi*'s sulfide tolerance by minimizing photodamage to the D1 protein and allowing oxygenic photosynthesis to occur at low rates (response 3 in Figure 1.1).

Sulfide inhibits water from interacting with the water-splitting Mn_4CaO_5 cluster of the OEC in the D1 protein of PS II (encoded by *psbA*), preventing the use of H_2O as an electron donor and consequently oxygenic photosynthesis (Miller & Bebout, 2004). Expression of *psbA* is

increased in cyanobacteria exposed to light stress or DCMU, which inhibits the quinone binding site of PS II and thus electron transfer to PS I (Bishop, 1958; Mulo et al., 2009, 2012). Similar increased expression of *psbA* may also occur in response to sulfide stress, although this effect has not been studied in cyanobacteria. *P. pseudopriestleyi* has multiple copies of the D1 protein, which may assist with increasing expression of *psbA*. Even though *P. pseudopriestleyi* grows in a low O_2 environment, the MAG appears to only contain genetic capacity for the standard group 4 D1 protein, and there is no evidence that the organism has a microaerobic D1 protein. Thus, although the MAG is incomplete, *P. pseudopriestleyi* does not appear to use an alternative D1 as part of its sulfide tolerance strategy. If its tolerance mechanism is related to the D1 protein, *P. pseudopriestleyi* likely overcomes sulfide inhibition by either increasing *psbA* expression (response 3 in Figure 1.1) or activating a mechanism that has not been identified. In this scenario, O_2 produced from oxygenic photosynthesis will oxidize sulfide in the abiotic reaction presented in reaction 1 at a quick enough rate to prevent sulfide inhibition. This oxidation is passive and may be kinetically very slow at Lake Fryxell temperatures (2°C) but includes positive feedback between decreasing sulfide and increasing capacity for O_2 production.

Alternatively, *P. pseudopriestleyi* may employ anoxygenic photosynthesis by directly oxidizing sulfide to S^0 with the membrane protein SQR (response 2 or 4 in Figure 1). This process would deplete sulfide, allowing oxygenic photosynthesis. Cyanobacteria capable of anoxygenic photosynthesis pass electrons from SQR oxidation through the quinone pool to PS I to harvest energy (Cohen et al., 1986). If *P. pseudopriestleyi* transfers electrons from SQR oxidation to PS I, it gains energy through anoxygenic photosynthesis, while sulfide is being depleted. The MAG contains the *sqr* gene and may be capable of this response. However, laboratory incubation

experiments are necessary to determine if the cyanobacterium performs anoxygenic photosynthesis, regardless of sqr in the genome (Bronstein et al., 2000; Klatt, Al-Najjar, et al., 2015; Beer et al., 2017).

Ecology of P. pseudopriestleyi in Lake Fryxell

The seasonality of Lake Fryxell controls how *P. pseudopriestleyi* shapes the redox potential of its environment by producing O_2 through oxygenic photosynthesis. Complete darkness from mid-April to mid-August allows sulfide to accumulate in the mats, creating a reduced, sulfidic environment (Sumner et al., 2015). As light becomes available starting in mid-August, oxygenic photosynthesis or anoxygenic photosynthesis may initiate, leading to the oxidation of sulfide. Constant summer irradiance starts at the end of October, with irradiance sufficient to allow oxygenic photosynthesis, resulting in the accumulation of O_2 the mats (Sumner et al., 2015). oxygenic photosynthesis slows down as light levels fall from mid-February to mid-April and then ceases in total darkness, and sulfide reaccumulates.

The low irradiance in *P. pseudopriestleyi*'s spring and summer environment may contribute to its sulfide resistance. Even during peak irradiance in the summer, the daily mean photon flux at 9.8 m depth is 1–2 µmol m^{-2} s^{-1}. Previous research demonstrates that low irradiance allowed the hot spring cyanobacteria *Planktothrix* str. FS34 to perform oxygenic photosynthesis uninhibited in up to 230 µM sulfide (or 7.83 mg L^{-1} sulfide) even though sulfide inhibited photosynthesis at higher light fluxes (Klatt, Haas, et al., 2015). Lower light levels may reduce photo-damage on the photosensitive D1 protein, allowing more D1 proteins in the

thylakoid membrane to perform sulfide-tolerant oxygenic photosynthesis. If sulfide levels are below the threshold for oxygenic photosynthesis inhibition, the O_2 from oxygenic photosynthesis will oxidize sulfide, eventually allowing O_2 to accumulate. If this mechanism is happening in Lake Fryxell, the balance of irradiance and sulfide levels in Lake Fryxell plays an important role in *P. pseudopriestleyi*'s ability to perform oxygenic photosynthesis in an extreme environment.

The effect of low irradiance on *P. pseudopriestleyi* is amplified by a partial mismatch between the wavelengths of light available and the pigments it can produce. The most penetrating waveband at 9 m in Lake Fryxell is 520–580 nm because the ice cover transmits blue light and absorbs wavelengths longer than 600 nm (Howard-Williams et al., 1989; Howard-Williams et al., 1998). The MAG contains genes for phycoerythrocyanin, phycocyanin, and allophycocyanin, which have peak absorptions at 575, 620, and 650 nm, respectively. Without the genes for phycoerythrin, absorption of wavelengths between 495–560 nm is limited. The combination of pigments and available light suggests that phycoerythrocyanin harvests most of the light for photosynthesis in the Lake Fryxell environment. Although there is a mismatch between available light and pigments for photosynthesis, the low irradiance of the environment is consistent with the absence of UV exposure genes in the MAG.

Continuous illumination during summer, even at low levels, requires a persistent nutrient source. The MAG has genes to create and break down cyanophycin, a nitrogen storage molecule synthesized by cyanobacteria. This may aid P. pseudopriestleyi in meeting peak nitrogen demands during the summer in Lake Fryxell, which is nitrogen limited (Priscu, 1995).

P. pseudopriestleyi in Other Environments

In addition to dominating a narrow sulfide-rich photic benthic zone of Lake Fryxell, *P. pseudopriestleyi* has been reported from several ponds and lakes across Antarctica, based on 16S rRNA gene analyses (Jungblut et al., 2005; Taton et al., 2006). *P. pseudopriestleyi*'s 16S gene sequence reported in Jungblut et al. (Jungblut et al., 2016) has a 99.45%, 91.28%, and 99.9% identity to 16S rRNA gene sequences from Antarctic Salt Pond, Fresh Pond, and Ace Lake, respectively (accession numbers AY541534, AY541575, and DQ181670) according to a blastn search. Salt and Fresh Ponds are meltwater ponds on the McMurdo Ice Shelf that experience high UV radiation (Roos & Vincent, 1998). Salt Pond is hypersaline with a conductivity of up to 52.9 mS cm^{-1}, with high salinity originating from diluted seawater or sulfate salts from chemical weathering of sedimentary material (Mora et al., 1991; Jungblut et al., 2005). Ace Lake is considered hyposaline with conductivity of 25.4–26.4 mS cm^{-1} and is permanently stratified (Hodgson et al., 2005; Taton et al., 2006). *P. pseudopriestleyi* is present in both locations, suggesting it is adapted to varying levels of salinity, UV, and high light stress in addition to sulfide. Future DNA sequencing beyond 16S amplicons in these environments may reveal whether or not *P. pseudopriestleyi* possesses additional stress tolerance genes related to these conditions that are absent in the Lake Fryxell MAG.

CONCLUSIONS

P. pseudopriestleyi is the first example of a cyanobacterium capable of sulfide-tolerant oxygenic photosynthesis in a cold, Antarctic environment. The MAG reconstructed for *P.*

pseudopriestleyi revealed a genome that is consistent with its ability to produce O2 in a sulfidic, cold, low-light environment of the perennially ice-covered Lake Fryxell, Antarctica. The MAG has a genomic capacity to deal with sulfide with multiple copies of a *psbA*, the D1 protein that is the site of water splitting, or by using SQR to deplete sulfide through anoxygenic photosynthesis or through sulfide oxidation. The low light levels at 9.8 m in Lake Fryxell may also contribute to its sulfide tolerance. Thus, there are likely several methods for dealing with sulfide stress on oxygenic photosynthesis in a low light environment. Sulfide tolerance varies widely among cyanobacteria, and the consideration of light level may have implications on a response to sulfide and should be studied further. Specifically, microelectrode measurements combined with gene ex-pression data are likely to uncover the molecular mechanism *P. pseudopriestleyi* uses to perform sulfide-tolerant oxygenic photosynthesis.

Besides Lake Fryxell, *P. pseudopriestleyi* has been found in shallow freshwater and hypersaline ice shelf melt water ponds and lakes, indicating a widespread distribution in Antarctica, and ability to thrive in a range of environmental conditions. If genomic data from *P. pseudopriestleyi* living in these environments can be obtained, a comparison with the MAG from Lake Fryxell can provide insight about the effects of various Antarctic conditions, such as UV exposure, high light levels, or salinity, on a genome. Additionally, the isolation and sequencing of other Antarctic cyanobacteria will allow for more in-depth genomic comparisons between Antarctic cyanobacteria genomes beyond the three genomes currently published.

A deeper understanding of the ecology of cold cyanobacteria ecosystems will provide insights into the production of O_2 through Earth history. Specifically, primary productivity during "Snowball Earth" glaciations was required to sustain the biosphere through these climatic crises.

In some cases, sulfide appears to have been abundant (Dahl et al., 2011; Parnell & Boyce, 2017; Hawes et al., 2018; Wang et al., 2019), suggesting that *P. pseudopriestleyi* may provide a model for how oxygenic photosynthesis persisted in some “Snowball Earth” ecosystems.

CHAPTER 2

Genomic Content of Four Novel Cyanobacteria MAGs from Lake Vanda, Antarctica and The Effect of Polar Light Cycles on Circadian Clocks

INTRODUCTION

Antarctic cyanobacteria form microbial mat communities that drive primary productivity in perennially ice-covered lakes (Jungblut et al., 2016; Sumner et al., 2016; Dillon et al., 2020). The microbial mats vary in texture and morphology among lakes and within different parts of the same lake. In Lake Vanda, mats comprised of cyanobacteria have grown into pinnacles ranging from millimeters to tens of centimeters (Love et al., 1983; Sumner et al., 2016; Wharton, 1994). The community composition of the mats also vary, reflecting various ecological processes (A. D. Jungblut et al., 2016; Sumner et al., 2016; Dillon et al., 2020). Dispersal of organisms to and between Antarctic lakes requires surviving freeze-drying and high UV exposure during transport from one habitable environment to the next. Once organisms reach the lake water, they are free from grazers, but they must compete for limited environmental resources. Selection pressures include sparse nutrients in the oligotrophic environment and low irradiance and spectral limits from ice cover which selects for organisms that are best suited to absorb available wavelengths (Jungblut et al., 2016; Sumner et al., 2016; Dillon et al., 2020). In addition, biochemical processes are affected by cold temperatures that slow metabolic rates and polar seasons which include extended periods of darkness during winter. Despite the strong selection pressures in Lake Vanda, the accumulation of biomass allows these ecosystems to become more complex.

Variations in the details of the selection create ecological drift between lakes as well as within different niches in the same lake.

In spite of environmental challenges, mats with pinnacles grow in Lake Vanda and have complex geochemical interactions. One of the prime drivers of genome diversification within the mats may be seasonal light distribution. Polar seasons have three months of constant light during summer and darkness during winter, which dictates when photosynthesis is possible. The variable light availability has implications for the metabolism of these cyanobacteria because they can only photosynthesize when light is available, and they must remain dormant or switch to another metabolic pathway during the dark part of the year. In non-polar environments, circadian clocks facilitate changes in metabolisms throughout the day by affecting gene expression, such as photosynthesis during the day and nitrogen fixation at night. Because circadian clocks usually use photosynthesis to synchronize with environmental conditions on a diel cycle (Swan et al., 2018), the long periods of darkness and light in Antarctica may cause the circadian clock to lose synchronization with a 24-hour cycle or affect the ability of the clock to influence gene expression. Some cyanobacteria possess diverged homologous versions of the standard *kaiABC* genes that encode the circadian clock. These *kaiA3B3C3* homologs may be linked to a metabolic switch in response to darkness (Aoki & Onai, 2009; Wiegard et al., 2020; Koebler et al., 2021), and they have not yet been identified in polar cyanobacteria.

Although cyanobacteria are responsible for a large portion of primary productivity in polar environments (Chrismas, Anesio, et al., 2018), cyanobacteria from Antarctica and the Arctic have received less attention compared to cyanobacteria in other areas (Anesio et al., 2009; Anesio & Laybourn-Parry, 2012; Chrismas, Williamson, et al., 2018b). Currently only six cyanobacteria

MAGs or closed genomes are publicly available: *Phormidium pseudopriestleyi* FRX01 (Lumian et al., 2021), *Aurora vandensis* (Grettenberger et al., 2020), *Synechococcus* sp. CS-601 (SynAce01) (Tang et al., 2019), *Leptolyngbya* BC 1307 (Chrismas, Williamson, et al., 2018b), *Phormidesmis priestleyi* BC 1401 (Chrismas et al., 2016), and *Phormidesmis priestleyi* ULC007 (Lara et al., 2017). This chapter presents four novel cyanobacterial MAGs from pinnacles in Lake Vanda, Antarctica and compares them to available polar cyanobacteria. Genes relating to photosynthesis and light harvesting, circadian clocks, and cold tolerance are of particular interest because of the unique challenges faced by polar organisms compared to organisms in other environments.

Lake Vanda Site Description

Lake Vanda is an ice-covered lake in the Wright Valley, one of the McMurdo Dry Valleys of Southern Victoria Land, Antarctica. It receives water from the Onyx River, which flows ~30 km inland from the Lower Wright Glacier (Castendyk et al., 2016). The balance of this water flow with water loss from ablation and evaporation determines the water level of the lake (Castendyk et al., 2016). There is a perennial ice cover of 3.5 – 4.0 m thickness, but during the summer, ice around the shore melts to produce an open-water moat (Sumner et al., 2016). Lake Vanda is well illuminated compared to other McMurdo Dry Valley lakes, with 15 – 20% of photosynthetically active ration (PAR) passing through the ice cover and a vertical extinction coefficient for downwelling PAR of 0.06 m^{-1} (Howard-Williams et al., 1989). Because ice transmits very little red light and water is transparent to blue-green light, the light spectrum is dominated by wavelengths shorter than 550 nm (Hawes & Schwarz, 2001). Most light around 455 nm is scattered at a depth

of 2.5 mm into the mat, with 16 – 90% of wavelengths ranging from 455 nm to 745 nm penetrating to that depth, with the exception of light around 600 nm because it is attenuated by the ice cover (Hawes & Schwarz, 2001). At 5 mm depth into the mat, only light around 745 nm was present due to other wavelengths being scattered or absorbed by the mat (Sumner et al., 2016).

Well-illuminated conditions and a lack of erosion, burrowing, and grazing has allowed the growth of extensive microbial mats in Lake Vanda. The mats extend from just under the ice cover to more than 50 m water depth (Wharton et al., 1983; Hawes & Schwarz, 2001). The upper part of the water column in Lake Vanda is well mixed, resulting in similar environmental conditions for mats from 4 – 26 m depth with the exception of irradiance. The temperature of this upper convection zone is ~4°C (Castendyk et al., 2016). Mats from greater than 10 m water depths have morphologies ranging from <1-mm-tall tufts to pinnacles that are centimeters tall. The largest pinnacles in the lake are up to ~30 cm tall with cylindrical shaped bases and cuspate tops (Sumner et al., 2016). Photosynthetic activity in the mats has been demonstrated starting from just below the ice cover to at least 40 m depth (Hawes & Schwarz, 2001). Mats are laminated on an annual basis, providing an internal record of mat growth and evolution in shape (Hawes et al., 2001, 2013; Sumner et al., 2016). A reconstruction of the growth history has shown that mat thickness and pinnacle heights have increased over time. There is a gradient of colonization in the lake with recently colonized mats just under the ice and mats colonized more than 100 years ago at greater depths (Hawes et al., 2013).

The first report of mats from 1980 described their composition as mostly cyanobacteria, specifically *Phormidium cf. fridium* (now *Leptolyngbya fridida*) (Komárek & Anagnostidis, 2005),

Lyngbya (now *Phormidium* and *Oscillatoria*), as well as pennate diatoms including *Navicula, Nitzschia, Caloneis*, and *Stauroneis*, and strands of moss (Kaspar et al., 1982; Love et al., 1983). Recent research on the Lake Vanda mats has shown them to be similar to the mats of other McMurdo Dry Valley lakes (Zhang et al., 2015). 16S sequencing has shown that the bacterial and archaeal biomass is dominated by *Leptolyngbya, Phormidium,* and *Tychonema* genomic groups (Sumner et al., 2016).

MATERIALS AND METHODS

Irradiance Data

PAR data was collected by Doran and Fountain, (2016). As part of the McMurdo Dry Valleys Long-Term Ecological Research effort data is continuously collected from the Lake Vanda meteorological station located near the mouth of the Onyx River. Every fifteen minutes, a signal is sent through sample sensors and recorded as summary statistics, including average PAR, minimum PAR, and maximum PAR. Information from July 2011 – July 2012 was accessed on June 23, 2021. Average PAR, minimum PAR, and maximum PAR were plotted using an R script available at: https://github.com/jessicalumian/vanda_mags.

Sample Collection and DNA Extraction

To obtain samples, SCUBA divers entered the lake through a large hole melted in the ice. They collected intact benthic microbial mats from 9 and 19 m lake depths using utensils sterilized

with alcohol wipes and brought samples to the surface in sterilized plastic containers. Collected pinnacles were dissected in a field lab based on mat color using sterile technique and within 12 hours of collection. Subsamples for genomic analysis were placed in Zymo Xpedition buffer (Zymo Research, Irvine, CA), and cells were lysed via bead beating. The stabilized samples were then frozen on dry ice and maintained frozen in the field. Upon leaving the field, samples were maintained at −20 °C during storage and transport to UC Davis by the US Antarctic Program. Samples were stored at -80 °C at UC Davis until DNA was extracted with the QuickDNA Fecal/Soil Microbe kit using the manufacturer's instructions (Zymo Research, Irvine, CA, USA). The extracted DNAs were quantified using Qubit (Life Technologies) and were concentrated via evaporation until the concentration was ≥ 10 ng μL^{-1}.

Sequencing

Sequencing was performed at the US Department of Energy Joint Genome Institute (JGI) using an Illumina HiSeq-2500 1 TB platform. An Illumina library was sequenced as 2 x 151 bp. BBDuk (v37.36) was used to remove common contaminants with the parameters removehuman=t, removedog=t, removecat=t, removemout=t, and removemicrobes=t. BBDuk was also used to trim reads containing adapter sequence and bases to the right of bases with a quality score of 0. It also removed reads containing 4 or more 'N' bases, with an average quality score less than 3, or had a minimum length less than 50 or 33% of the full read length. Reads that mapped to masked human, cat, dog, and mouse references at 93% identity or aligned to common microbial contaminants were also removed.

Bioinformatics Processing

Filtered and quality controlled raw data was retrieved from IMG Gold with JGI Gold IDs GP0191362 and Gp0191371. MEGAHIT v1.9.6 (Li et al., 2015) was used to assemble metagenomes with a minimum contig length of 500 bp and a paired end setting. Bowtie2 v1.2.2 (Langmead & Salzberg, 2012) was used to map reads back to the assembly. A depth file was generated using jgi_summarize_bam_contig_depths from MetaBAT v2.12.1 (Kang et al., 2015) and was used to generate bins with a minimum contig length of 2500 bp. Operon-mapper online web server was used to predict operons from the contigs in the MAGs (Taboada et al., 2018). All the bins obtained from the metagenome assembly were run through CheckM (Parks et al., 2015). The twenty bins with over 70% completeness were all annotated as cyanobacteria and then grouped into four taxa by GTDB-tk (Chaumeil et al., 2020). The bin with the highest completeness from each taxon was chosen as the representative bin to further analyze: BulkMat_35, MP8_15, MP8_171, and MP9_79.

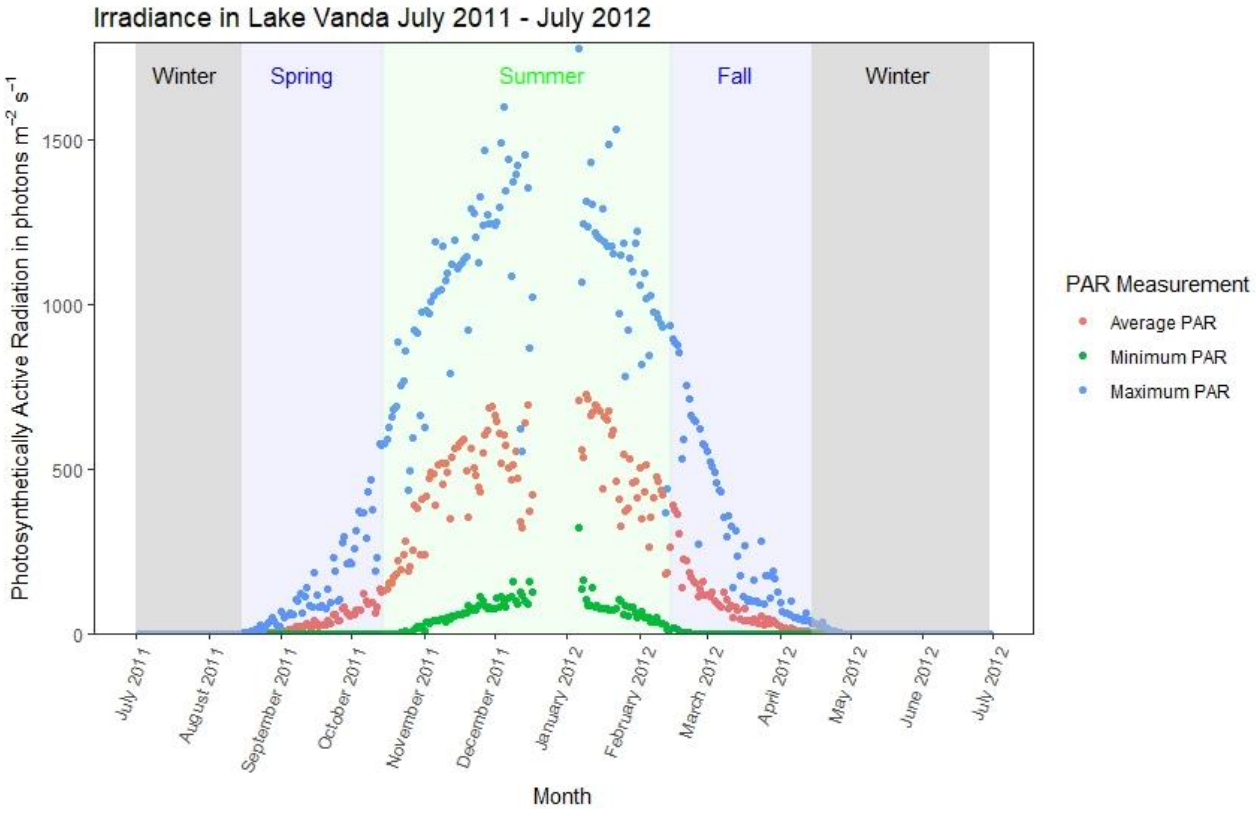
Other polar cyanobacteria were identified through a literature search: *Phormidium pseudopriestleyi* FRX01 (Lumian et al., 2021), *Aurora vandensis* (Grettenberger et al., 2020), *Synechococcus* (SynAce01) (Tang et al., 2019), *Leptolyngbya* BC 1307 (Chrismas, Williamson, et al., 2018b), *Phormidesmis priestleyi* BC 1401 (Chrismas et al., 2016), and *Phormidesmis priestleyi* ULC007 (Lara et al., 2017). Annotation was performed with custom blast databases and KEGG ghostkoala annotations (Kanehisa & Goto, 2000; Kanehisa et al., 2016; Camacho et al., 2009). Average nucleotide identity (ANI) was calculated using the ANI calculator from the Kostas lab using the default parameters (Rodriguez & Konstantinidis, 2016). The code for this analysis available here: https://github.com/jessicalumian/vanda_mags.

RESULTS

Irradiance Data

A graph of the average, minimum, and maximum daily PAR from July 2011 – July 2012 (Fig 2.1) shows that irradiance was absent during winter from mid-April to mid-August. More irradiance was present in the fall, summer, and spring, but daily irradiance depended on cloud cover and weather conditions. During fall and spring, diel cycles of light and darkness occurred, mimicking those at lower latitudes. As fall and spring progressed, the days got shorter or longer towards constant darkness or continuous light, respectively.

Fig 2.1 PAR in Lake Vanda from July 2011 - July 2012



Bin Statistics

Three of the four Vanda MAGs were present as bins over 70% complete in multiple metagenomic sequences, so the sample with the highest completion and lowest contamination was chosen for analysis (Table 2.1). The *Leptolyngbya* MAG (BulkMat_35) is from a green flat mat located in between pinnacles and was present in a total of eight samples. The *Pseudanabaena* MAG (MP8IB2_15) is only in one sample from the beige interior of a medium-sized pinnacle. The *Microcoleus* MAG (MP8IB2_171) is from the beige interior of a different medium-sized pinnacle and was present in seven samples. The *Neosynechococcus* MAG (MP9P1_79) is from a purple subsample of a medium-sized pinnacle and was present in four samples.

The novel MAGs from Lake Vanda were classified based on CheckM annotation as a *Pseudanabaena,* a *Microcoleus,* a *Leptolyngbya,* and a *Neosynechococcus* (Table 2.1). They have genome sizes ranging from 2.76 – 6.07 Mbp and GC content between 45.31 – 51.23%. Of all the MAGs and genomes, the *Pseudanabaena* MAG from Vanda has the second shortest length at 2,764,673 bp, just above SynAce01 with 2,750,634 bp. The *Microcoleus* MAG from Vanda has the longest total length with 6,072,304 bp, followed by the *Phormidium pseudopriestleyi* FRX01 MAG from Lake Fryxell with 5,965,908 bp. The *Microcoleus* and *Pseudanabaena* MAGs have the lowest GC content out of all cyanobacteria in this analysis with 45.47% and 45.31% respectively. Aside from the closed genome of SynAce01, the majority of contig lengths in the other cyanobacteria are less than 5,000 bp in length.

Table 2.1 Cyanobacteria bins over 70% complete from Lake Vanda mat samples. Bin completeness and contamination were determined by CheckM.

Organism	Mat Sample Characteristics	Bin Completeness	Bin Contamination
Leptolyngbya	Bulk green mat in between pinnacles (BulkMat_35)	92.57%	0.63%
	Medium-sized pinnacle, green color (MP6G1_181)	91.31%	1.1%
	Medium-sized pinnacle, green color (MP7G1_15)	89.86%	0.35%
	Medium-sized pinnacle, green color (MP5G1_15)	88.55%	0.12%
	Medium-sized pinnacle, purple color (MP7P2_86)	85.81%	0.47%
	Medium-sized pinnacle, purple color (MP9P1_182)	83.91%	0.24%
	Large-sized pinnacle (LV9_51)	82.73%	1.3%
	Small-sized pinnacle, green color (SP4G1_162)	78.25%	0.24%
Pseudanabaena	Medium-sized pinnacle, inner beige portion (MP8IB2_15)	74.2%	0.63%
Microcoleus	Medium-sized pinnacle, inner beige portion (MP8IB2_171)	88.88%	1.39%
	Medium-sized pinnacle, green color (MP5G1_130)	87.72%	1.09%
	Medium-sized pinnacle, purple color (MP9P1_1)	85.55%	1.09%
	Medium-sized pinnacle, green color (MP7G1_170)	84.72%	0%
	Medium-sized pinnacle, purple color (MP7P2_141)	81.66%	2.19%
	Medium-sized pinnacle, green color (MP6G1_116)	79.5%	0.55%
	Medium-sized pinnacle, inner beige portion (MP6IB1_67)	75.48%	0%
Neosynechococcus	Medium-sized pinnacle, purple color (MP9P1_79)	92.39%	0.24%
	Medium-sized pinnacle, green color (MP5G1_109)	92.24%	2.48%
	Large-sized pinnacle (LV9_51)	91.29%	1.65%
	Medium-sized pinnacle, inner beige portion (MP6IB_67)	75.48%	0%

Table 2.2 MAG and Genome Statistics from QUAST.

Statistics	*Leptolyngbya* (Vanda)	*Microcoleus* (Vanda)	*Neosynechococcus* (Vanda)	*Pseudanabaena* (Vanda)	*Aurora vandensis* (Vanda)
Total Number of Contigs	654	671	552	505	202
Longest contig (bp)	69,200	43,043	62,898	24,594	85,365
Total length (bp)	5,729,854	6,072,304	4.887,811	2,764,673	2,958,216
GC content (%)	51.23	45.47	50.03	45.31	55.41
N50	11,419	10,926	11,533	5,954	21,579
Number contigs >= 0 bp	654	671	552	505	202
Number contigs >= 1000 bp	654	671	552	505	202
Number contigs >= 5000 bp	406	476	349	232	164
Number contigs >= 10,000 bp	199	215	158	37	107
Number contigs >= 25,000 bp	24	19	25	0	32
Number contigs >= 50,000 bp	1	0	3	0	5

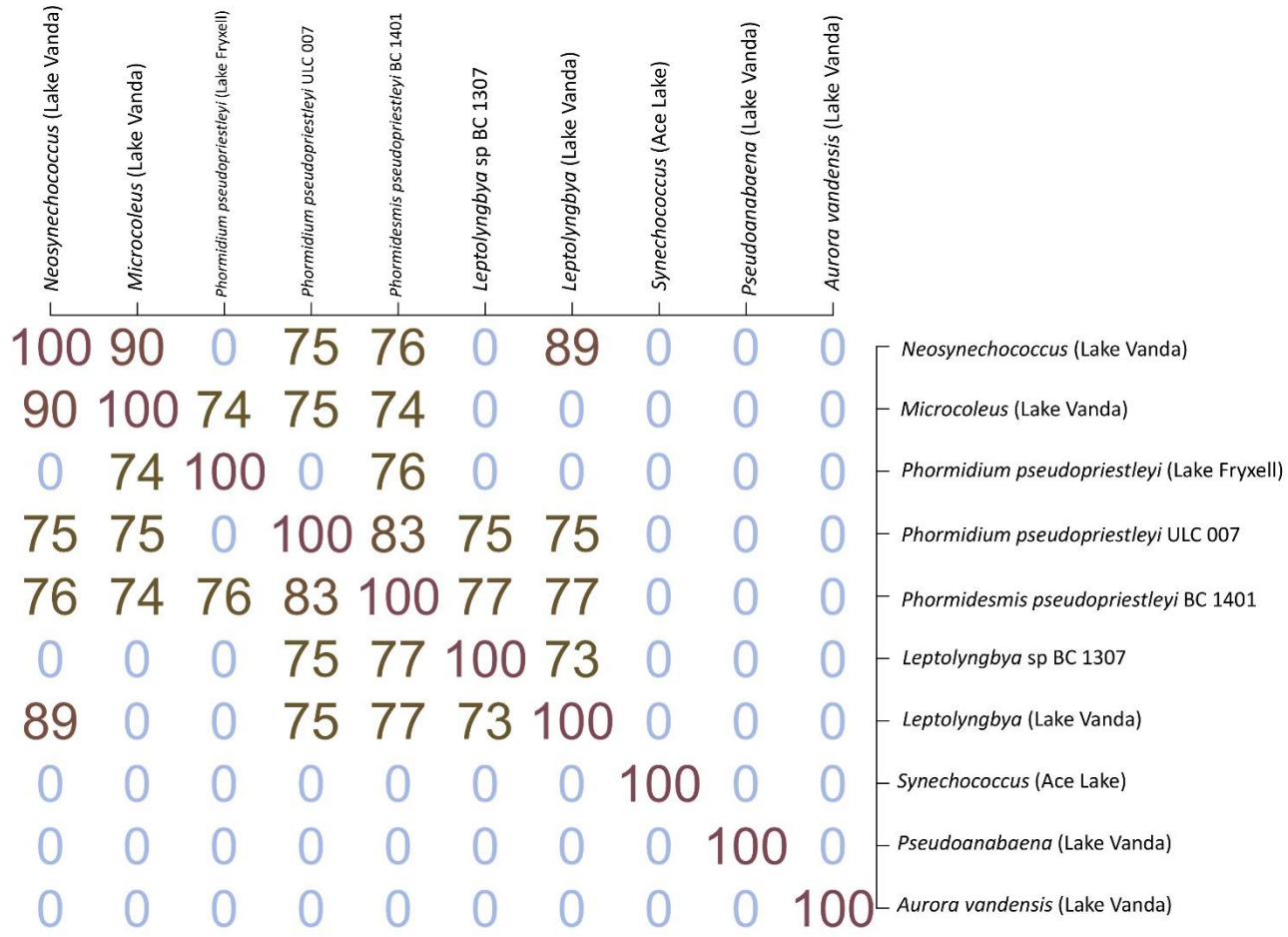
Statistics	*Synechococcus* SynAce01 (Ace Lake01)	*Leptolyngbya* 1307	*Phormidesmis priestleyi* 1401	*Phormidesmis priestleyi* 007	*Phormidium pseudopriestleyi* FRX01 (Fryxell)
Total Number of Contigs	1	222	213	118	678
Longest contig (bp)	2,750,634	294,697	275,582	497,214	44,245
Total length (bp)	2,750,634	4,916,582	5,546,500	5,684,389	5,965,908
GC content (%)	63.92	52.94	49.17	48.62	47.43
N50	2,750,634	82,349	79,760	152,457	10,908
Number contigs >= 0 bp	1	222	213	118	678
Number contigs >= 1000 bp	1	159	168	118	678
Number contigs >= 5000 bp	1	97	121	79	458
Number contigs >= 10,000 bp	1	82	96	72	203
Number contigs >= 25,000 bp	1	49	62	55	20
Number contigs >= 50,000 bp	1	32	37	36	0

ANI Results

The highest ANIs within all cyanobacteria in this analysis were 89 – 90% between the *Neosynechococcus*, *Microcoleus,* and *Leptolyngbya* MAGs from Lake Vanda, and the next highest ANI was between *Phormidesmis pseudopriestleyi* BC 1401 and *Phormidium pseudopriestleyi* ULC 007 MAGs. The *Pseudanabaena* MAG had an ANI under 70% with any other cyanobacterium, and the *Synechococcus* SynAce01 genome, *Pseudanabaena* and *Aurora vandensis* MAGs all had ANIs less than 70% with all other cyanobacteria included in this group. One reason for the low ANI between *Pseudanabaena* MAG and other cyanobacteria in this analysis could be that the *Pseudanabaena*'s MAG size is only 2.76 MB, while the other MAGs from Lake Vanda have a MAG size ranging from 4.88-6.07 MB.

The four cyanobacteria MAGs from Lake Vanda had ANIs under 90% with each other and ANIs under 77% with other cyanobacteria MAGs in this analysis. This indicates that the cyanobacteria from Lake Vanda provide additional diversity to the collection of publicly available polar cyanobacteria MAGs and genomes.

Figure 2.2 ANI matrix between novel Vanda MAGs and other polar cyanobacteria. ANIs under 70% are reported as 0.



Circadian Rhythm Genes

The posttranslational oscillator responsible for keeping time in cyanobacteria is made up of proteins encoded by *kaiA, kaiB,* and *kaiC*. (Table 2.3) These genes are present in the *Leptolyngbya*, *Microcoleus*, *Neosynechococcus*, and *Pseudanabaena* MAGs from Vanda, as well as all other cyanobacteria included in this analysis, except for *A. vandensis* and *Leptolyngbya* 1307 MAGs. Both the *Leptolyngbya* and the *Neosynechococcus* MAGs contain a copy of *kaiB3* in

addition to the standard *kaiB* gene. The *kaiB3* gene is a diverged homolog that is thought to play a role in the metabolic switch from light to darkness (Wiegard et al., 2020). The main proteins involved in output signaling to affect gene expression are encoded by *sasA, cikA,* and *rpaA*. All the cyanobacteria MAGs and genome included in this study contained multiple copies of these genes. (Table 2.3)

The *Leptolyngbya, Neosynechococcus,* and *Pseudanabaena* MAGs all contain *kaiABC* on the same operon. The *Microcoleus* MAG contains *kaiA* and *kaiB* at the end of a contig, and additional copy of *kaiA* on a different contig. *kaiC* was not identified on an operon, likely due to MAG incompleteness. The *Neosynechococcus* MAG also has a copy of *kaiB3* on an operon separate from the complete *kaiABC* group.

Table 2.3 Presence of circadian clock genes in cyanobacteria MAGs and genome in this study

Gene	***Leptolyngbya*** **(Vanda)**	***Microcoleus*** **(Vanda)**	***Neosynechococcus*** **(Vanda)**	***Pseudanabaena*** **(Vanda)**	***Aurora Vandensis*** **(Vanda)**
kaiA	1	1	1	1	0
kaiB	2	1	2	1	0
kaiC	1	1	1	1	0
sasA	4	3	0	1	>10
cikA	3	4	4	4	1
rpaA	3	3	3	3	2

Gene	***SynAce01*** **(Ace Lake)**	***Leptolyngbya*** **1307**	***Phormidesmis priestleyi*** **1401**	***Phormidesmis priestleyi*** **007**	***Phormidium pseudopriestleyi*** **FRX01**
kaiA	1	1	1	1	1
kaiB	1	0	1	2	3
kaiC	1	0	1	1	2
sasA	>10	>10	>10	>10	>10
cikA	3	2	0	0	2
rpaA	3	2	0	0	2

Photosynthesis Genes

Allophycocyanin, phycocyanin, and phycoerythrocyanin are phycobilisome proteins with absorption peaks at 650, 620, and 575 nm (Bogorad, 1975). All four novel MAGs from Lake Vanda contain the majority of genes for these pigments along with all other cyanobacteria included in this study. Phycoerythrin has an absorption peak ranging from 495 to 560 nm (Bogorad, 1975). The *Microcoleus* and *Pseudanabaena* MAGs from Vanda have all or some genes for phycoerythrin along with all other cyanobacteria in this analysis except for the *Leptolyngbya* and *Neosynechococcus* MAGs.

The D1/D2 protein cluster in photosystem II is coded by *psbA* and *psbD*. Out of the Vanda MAGs, only the *Leptolyngbya* MAG contains *psbA*, which encodes for the D1 protein and none of the MAGs contained *psbD*, which encodes for the D2 protein. Chlorophyll apoproteins cp43 and cp47 (*psbC*) and (*psbB*) are present in all four MAGs, as well as cytochrome b559 genes *psbE* and *psbF*, with the exception of the *Neosynechococcus* MAG, which is missing *psbF*. The oxygen evolving complex in photosystem II is stabilized by proteins encoded by *psbO* and *psbP*, which are present in all the MAGs with the exception of the *Pseudanabaena* and *Microcoleus* MAGs, which are missing *psbO*. Cytochrome b6f is encoded by a suite of genes: *petB, petD, petA, petC, petL, petM, petN,* and *petG*. All four MAGs have *petB, petD, petA,* and *petC* except for the *Pseudanabaena* MAG, which is missing *petA* and *petC*. Besides the *Neosynechococcus* MAG containing *petM* and the *Pseudanabaena* MAG containing *petN*, these genes along with *petL* and *petG* are missing from all MAGs. Plastocyanin (*petE*) is an electron transport protein that connects cytochrome b6f to photosystem I and is present in all MAGs except for the *Microcoleus* MAG.

Cytochrome c6 can act as an electron transport protein in the absence of plastocyanin (Bogorad, 1975) and is present in all MAGs except for the *Pseudanabaena* MAG. Genes for the chlorophyll protein dimer in photosystem I (*psaA* and *psaB*) are present in all the MAGs. Ferredoxin (*petF*) is present in all MAGs, and ferredoxin-NADP+ reductase (*petH*) is present in all MAGs except for the *Pseudanabaena* MAG. Genes for ATPase are present in all MAGs with the exception of *Microcoleus*, which is missing genes for the alpha (*atpA*), beta (atpD), and epsilon (atpC) proteins.

The presence and absence of phycobilisome and photosynthesis genes in the other polar cyanobacteria MAGs and genome are consistent with the novel Vanda MAGs. All cyanobacteria MAGs and genomes contain the majority of genes common to cyanobacteria and that are required for core cyanobacteria metabolism, such as oxygenic photosynthesis (Shi & Falkowski, 2008).

Table 2.4 Photosynthesis Genes in MAGs

Gene Group	***Leptolyngbya*** **(Vanda)**	***Microcoleus*** **(Vanda)**	***Neosynechococcus*** **(Vanda)**	***Pseudanabaena*** **(Vanda)**	***Aurora vandensis*** **(Vanda)**
Allophycocyanin	Yes	Yes	Most	Yes	Yes
Phycocyanin / Phycoerythrocyanin	Most	Yes	Yes	Yes	Most
Phycoerythrin	No	Most	No	Some	Most
Photosystem II	Most	Most	Some	Some	Some
Photosystem I	Some	Some	Some	Some	Some
Cytochrome b6f	Some	Some	Most	Some	Some
Photosynthetic electron transport	Yes	Most	Yes	Some	Yes
F-type ATPase	Yes	Most	Yes	Yes	Most

Gene Group	*SynAce* (Ace Lake)	*Leptolyngbya* 1307	*Phormidesmis priestleyi* 1401	*Phormidesmis priestleyi*	*Phormidium pseudopriestleyi* FRX01
Allophycocyanin	Yes	Yes	Most	Yes	Yes
Phycocyanin / Phycoerythrocyanin	Most	Yes	Yes	Yes	Most
Phycoerythrin	Yes	Most	Most	Most	Most
Photosystem II	Some	Most	Most	Most	Some
Photosystem I	Some	Some	Most	Most	Some
Cytochrome b6f	Some	Some	Most	Most	Some
Photosynthetic electron transport	Yes	Yes	Yes	Yes	Most
F-type ATPase	Yes	Yes	Yes	Yes	Yes

Yes = all genes present. Most = 2/3 or more genes present. Some = 1/3 or more genes present. No = No genes present

Cold Tolerance Genes

Cold shock genes allow cells to maintain normal functions, including transcription, translation, and energy production, at low temperatures. Chrismas et al. (2016) compiled a list of cold tolerance genes from Barria et al. (2013) and Varin et al. (2012) and analyzed their presence in cyanobacteria, finding all of them in at least one MAG (Table 2.5). However, none of the cyanobacteria MAGs or the genome in this analysis have *aceE*, the gene for pyruvate dehydrogenase. Most of the cyanobacteria MAGs have *aceF*, the gene for the S-acyltransferase component of the pyruvate dehydrogenase complex, but it is absent in the *Leptolyngbya* and *Pseudanabaena* MAGs. The only cold tolerance gene relating to carbohydrate transport and metabolism, *otsA*, α,α-trehalose-phosphate synthase, is only present in the *Synechococcus* SynAce01 genome and *Leptolyngbya* BC1307 MAG. Several genes related to replication, recombination, and repair are in all the cyanobacteria: *deaD* (RNA helicase DeaD box), *dnaA* (replication initiator protein), and *gyrA* (DNA topoisomerase II). All the cyanobacteria except for

the *Pseudanabaena* MAG have *recA*, which is involved with DNA cleavage. The gene for a histone-like DNA binding protein, *hupB,* is present in all the cyanobacteria except the *Pseudanabaena*, *Neosynechococcus*, *Microcoleus,* and *A. vandensis* MAGs. A trigger factor protein involved in protein export, encoded by *tig*, is present in all the cyanobacteria except for the *Pseudanabaena* MAG. Delta(12)-fatty-acid desaturase, encoded by *desA*, can change the degree of unsaturation of fatty acids at low temperatures (Wada et al., 1990), and it is present in all the cyanobacteria. The transcription related genes *rnr* (exoribonuclease II) and *nusA* (transcription termination) are present in all the cyanobacteria, but *csp* genes (cold shock proteins) are not present in any of the MAGs or the genome. Some cold tolerance genes are involved in translation, ribosome structure, and biogenesis, such as *pnp*, *yfiA, rbfA,* and *infABC*. All the cyanobacteria contain *pnp*, which codes for polyribonucleotide nucleotidyltransferase, and only the four cyanobacteria MAGs from Lake Vanda contain *yfiA*, which codes for ribosome-associated inhibitor A. The 30S ribosome-binding factor encoded by *rbfA* is not present in the *Neosynechococcus* MAG and *Synechococcus* SynAce01 genome but is present in all the other cyanobacteria. The translation initiation factors IF-1, IF-2, and IF-3 are encoded by *infABC* and are present in all MAGs and genome in this study except for the *Microcoleus* MAG, which is missing *infB,* and the *P. pseudopriestleyi* MAG, which is missing *infC*. There are two post-translational modification and chaperone cold shock genes: *dnaK* (chaperone protein DnaK) is present in all of the MAGs and genome, but *dnaJ* (chaperone protein DnaJ) is not present in any MAGs from Vanda or in the *Synechococcus* SynAce01 genome.

Table 2.5 Cold Tolerance Genes in MAGs

Gene Category	Genes	*Leptolyngbya* (Vanda)	*Microcoleus* (Vanda)	*Neosynechococcus* (Vanda)	*Pseudanabaena* (Vanda)	*Aurora vandensis* (Vanda)
Energy production and conversion	*aceE*	No	No	No	No	No
	aceF	Yes	No	No	Yes	Yes
Carbohydrate transport and metabolism	*otsA*	No	No	No	No	No
Replication, recombination, and repair	*deaD*	Yes	Yes	Yes	Yes	Yes
	recA	Yes	Yes	Yes	No	Yes
	dnaA	Yes	Yes	Yes	Yes	Yes
	gyrA	Yes	Yes	Yes	Yes	Yes
	hupB	Yes	No	No	No	No
Cell cycle control, cell division, chromosome partitioning	*tig*	Yes	Yes	Yes	No	Yes
Lipid transport and metabolism	*desA*	Yes	Yes	Yes	Yes	Yes
Transcription	*rnr*	Yes	Yes	Yes	Yes	Yes
	nusA	Yes	Yes	Yes	Yes	Yes
	csp	No	No	No	No	No
Translation, ribosomal structure, and biogenesis	*pnp*	Yes	Yes	Yes	Yes	Yes
	yfiA	Yes	Yes	Yes	Yes	No
	rbfA	Yes	Yes	No	Yes	Yes
	infA	Yes	Yes	Yes	Yes	Yes
	infB	Yes	No	Yes	Yes	Yes
	infC	Yes	Yes	Yes	No	Yes
Post-translational modification, protein turnover, and chaperones	*dnaK*	Yes	Yes	Yes	Yes	Yes
	dnaJ	No	No	No	No	Yes

Gene Category	Genes	*Synechococcus* SynAce01 (Ace Lake)	*Leptolyngbya* 1307	*Phormidesmis priestleyi* 1401	*Phormidesmis priestleyi* 007	*Phormidium pseudopriestleyi* FRX01
Energy production and conversion	*aceE*	No	No	No	No	No
	aceF	Yes	Yes	Yes	Yes	Yes
Carbohydrate transport and metabolism	*otsA*	Yes	Yes	No	No	No
Replication, recombination, and repair	*deaD*	Yes	Yes	Yes	Yes	Yes
	recA	Yes	Yes	Yes	Yes	Yes
	dnaA	Yes	Yes	Yes	Yes	Yes
	gyrA	Yes	Yes	Yes	Yes	Yes
	hupB	Yes	Yes	Yes	Yes	Yes
Cell cycle control, cell division, chromosome partitioning	*tig*	Yes	Yes	Yes	Yes	Yes
Lipid transport and metabolism	*desA*	Yes	Yes	Yes	Yes	Yes
Transcription	*rnr*	Yes	Yes	Yes	Yes	Yes
	nusA	Yes	Yes	Yes	Yes	Yes
	csp	No	No	No	No	No
Translation, ribosomal structure, and biogenesis	*pnp*	Yes	Yes	Yes	Yes	Yes
	yfiA	No	No	No	No	No
	rbfA	No	Yes	Yes	Yes	Yes
	infA	Yes	Yes	Yes	Yes	Yes
	infB	Yes	Yes	Yes	Yes	Yes
	infC	Yes	Yes	Yes	Yes	No
Post-translational modification, protein turnover, and chaperones	*dnaK*	Yes	Yes	Yes	Yes	Yes
	dnaJ	No	Yes	Yes	Yes	Yes

DISCUSSION

The genes present in the polar cyanobacteria MAGs and genome provide insights into the possible functions the polar cyanobacteria can perform. Survival mechanisms are required to deal with the various challenges faced by organisms that live in polar areas, such as light

availability and cold temperatures. Seasonal variations in light availability play a large role in controlling the metabolisms of photosynthetic organisms and is likely linked to the timing of the circadian clock. Cold temperatures also affect timing by slowing the rates of biochemical processes as well as making cellular structures more brittle. Although the specific mechanisms of cold tolerance cannot be ascertained from genomic data alone, examining gene content provides a foundational understanding of how polar cyanobacteria propagate and thrive in challenging environments.

Polar Light Availability and Cyanobacteria Metabolism

In polar cyanobacteria the circadian clock may maintain a 24-hour cycle into the summer because the posttranslational oscillator encoded by *kaiABC* runs independently of environmental conditions when not synchronized (Nakajima et al., 2005; Swan et al., 2018). The question remains as to how long Antarctic circadian clocks run on a ~24-hour cycle as light decays for winter or increases to be continuously present for summer. If cyanobacteria are photosynthesizing for part of the day, there will be differences in the oxidation state of the quinone pool and ATP/ADP ratios that will allow the circadian clock to sense the changes and synchronize to light levels. During the summer if light levels are low and photosynthesis slows down enough, the circadian clock may still synchronize to periods of higher and lower light availability. When light is completely gone during the winter there is no way for the clock to synchronize with a 24-hour cycle.

The presence of circadian clock genes in all of the polar cyanobacteria except for *A. vandensis* and *Leptolyngbya* BC 1307 suggests that the circadian clock plays an important role in regulating gene expression. However, seasonal light availability in polar light environments likely causes circadian clocks to function differently than those in non-polar environments. In non-polar environments, the circadian clock helps regulate metabolism on a daily cycle. In contrast, polar environments experience extended dark seasons which prevent the circadian clocks from synchronizing with a 24-hour day/night cycle. Winter extends from mid-April to mid-August when there is complete darkness, and the circadian clock may drift off the daily cycle. In contrast, from mid-October to mid-February there is sufficient light for photosynthesis 24 hours most days (Hawes et al., 2013) (Fig 2.1). The amount of light in the summer changes over a 24-hour period due to the height of the sun and topography. If differences in irradiance are large enough, the circadian clock may synchronize with a 24-hour light cycle during the summer, but it may also be influenced by cloud cover, which can cause lower irradiance than is available on a clear night. It is unclear how the circadian clock will synchronize with environmental conditions when cyanobacteria continuously photosynthesize with the constant presence of summer light (Hawes et al., 2013).

Synchronization similar to a non-polar environment is likely to occur during transitional seasons, mid-August to mid-October and mid-February to mid-April, which have dark periods within 24-hour day/night cycles. Periods of darkness increase each day during fall and the clock is expected to synchronize as it does in non-polar environments until winter arrives. The extended darkness ends in the spring as light begins to appear, triggering the cyanobacteria to photosynthesize. Because photosynthesis is linked to synchronizing the circadian clock to the

environment (Swan et al., 2018), the clock is expected to calibrate with light availability as soon as photosynthesis begins in the spring. Synchronization is expected to persist through spring conditions and may extend into the summer.

The calibration of the circadian clock with polar seasonality may affect cellular metabolism. If the output circadian proteins encoded by *sasA*, *cikA*, and *rpaA* are tied to the circadian clock (Swan et al., 2018), seasons will affect gene expression. Class 2 genes, which are transcribed at night in environments with diel cycles, would not be transcribed during the summer in favor of class 1 genes, which are transcribed during the day (Swan et al., 2018). If polar cyanobacteria treat summertime conditions like one continuous day, this could provide a challenge for processes that normally occur at night, like respiration and nitrogen fixation. During winter, genes associated with daytime conditions, such as photosynthesis genes, will not be expressed for long periods of time. After light becomes available, these genes will be expressed again, but it is not known if the commencement of photosynthesis in the spring is linked to the activity of the circadian clock.

During the austral winter, polar cyanobacteria must remain dormant or use alternative energy sources to survive complete darkness. Instead of photosynthesis, cyanobacteria may survive on heterotrophy, fermentation, or respire existing carbon storage molecules in the cell. Modeling tied to summer mat metabolism has suggested that mats grow and fix carbon during the summer and respire fixed carbon during the winter, but do not grow (Hawes et al., 2001). The cyanobacteria from Lake Vanda as well as the majority of other cyanobacteria included in this study have genes for storage molecules which may allow them to persist during periods of low metabolic activity: polyphosphates to store phosphates, cyanophycin and phycobilins to

store nitrogen, and glycogen to store carbon. The presence of these genes in polar cyanobacteria suggests that during periods of active photosynthesis and growth in the summer, storage molecules are synthesized. These molecules can be consumed during periods without photosynthesis, such as through the winter.

The seasonal metabolic activity of the cyanobacteria has implications for the ecology and geochemistry of the microbial mats in Lake Vanda. Summertime is when the mats are most metabolically active in Lake Vanda (Hawes et al., 2001) and this is consistent with measurements from Lake Hoare, another ice-covered lake in the McMurdo Dry Valley (Hawes et al., 2014). In Lake Vanda it is estimated that 390 mg of carbon m^{-2} y^{-1} is fixed each summer (Hawes et al., 2013) and net O_2 production was measured to be 120 µmol m^{-2} h^{-1} with PAR of 1 µmol quanta m^{-2} s^{-1} (Vopel and Hawes, 2006). Modeling on Lake Hoare mats predicted rates of O_2 production with saturated irradiance to be between 543.8 µmol m^{-2} h^{-1} – 537.5 µmol m^{-2} h^{-1}, which is 22 – 23% higher than O_2 flux based on *in situ* measurements, suggesting that photosynthesis in the Lake Hoare mats is light limited (Vopel and Hawes, 2006). Furthermore, Hawes et al. (2014) found that in Lake Hoare, photosynthesis rates were linearly proportional to photon flux in the lake. Because photon flux changes each day based on seasonality and cloud cover conditions, it is possible that the cyanobacteria circadian clocks are tied to photon flux, particularly when light is consistently available in the summer. Prior work has shown that the organic component of the laminae in Lake Vanda microbial mats is linked to photosynthesis, and thicker organic laminae is due to increased rates of photosynthesis (Hawes et al., 2013; Sumner et al., 2016). The thickness of the laminae may be tied to the circadian clock's effect on photosynthesis if the rate of photosynthesis is linearly proportional to the rate of photon flux in Lake Vanda like it is in Lake Hoare. Future

work involving *in situ* transcriptomics and PAR measurements can be done to determine how closely the circadian clock is tied to photon flux and seasonality.

Cold Tolerance Genes

All of the polar cyanobacteria in this study contain cold tolerance genes, which is not surprising because of their natural environments. Their ability to perform standard biochemical processes, like DNA replication, gene expression, and protein production are essential for their survival, and all the organisms studied have genes for these processes. However, each cyanobacterium MAG or genome in this study is missing at least some genes that are associated with cold tolerance. A study from Chrismas et al. found no patterns in cold tolerance genes between a cyanobacteria from a cold environment compared to relatives from warmer climates (Chrismas et al., 2016). The lack of correlation between cold tolerance gene content among polar cyanobacteria in this study is consistent with those findings, suggesting that these genes are not specific to cold environments, but instead are present in most cyanobacteria. Furthermore, the identification of the cold tolerance genes was based on the response of *Escherichia coli* to cold shock, not on cyanobacteria (Barria et al., 2013). Thus, it is possible that cyanobacteria possess additional genes and mechanisms associated with cold stress that have not yet been identified, particularly relating to photosynthesis machinery which is not present in *E. coli*.

Polar Cyanobacteria Adaptations

Polar cyanobacteria, which are interpreted as psychrotrophs rather than psychrophiles (Tang et al., 1997), must deal with cold stress challenges throughout a year with highly variable irradiance. Previous work based on 16S sequences, protein sequences, and Bayesian statistical analysis has suggested the existence of 20 cold tolerant cyanobacterial clades that live in polar or alpine regions (Chrismas et al., 2015). Furthermore, cold tolerance genes are not specific to polar cyanobacteria (Chrismas, Anesio, et al., 2018; Chrismas, Williamson, et al., 2018b). The lack of shared gene content between all cold tolerant cyanobacteria suggests that polar cyanobacteria use a variety of mechanisms to tolerate cold conditions, and these mechanisms could be dependent on the type of cold environment where they live. Although polar cyanobacteria generally grow better in moderate temperatures, their tolerances to harsher conditions may allow them to outperform cyanobacteria that would normally be prevalent in moderate climates but cannot survive colder temperatures. Previous work has identified polar cyanobacteria based on 16S sequences (Jungblut et al., 2005; Chrismas et al., 2015), but nothing can be said about the metabolic potential of these cyanobacteria compared to those in more temperate environments without genomic data. The acquisition of more polar cyanobacteria genomes is necessary to do comparative genomics and identify similarities and differences between polar cyanobacteria and closely related organisms in warmer environments to further elucidate signatures of cold tolerant genomes (Chrismas, Williamson, et al., 2018b).

CONCLUSIONS

Results provide insights into how polar seasonality impacts the circadian clock of cyanobacteria, particularly the extended darkness and continuous light of winter and summer. In order to build upon the foundational understanding laid out by analyzing genomic data, *in situ* experiments must be performed on the mats detailing the gene expression of polar cyanobacteria in different lighting conditions. This characterization of four new polar cyanobacteria MAGs brings the total number of MAGs and genomes to ten. The four novel MAGs in this chapter, as well as *A. vandensis* are from Lake Vanda in the McMurdo Dry Valley. *P. pseudopriestleyi* is from Lake Fryxell, also in the McMurdo Dry Valley. *Synechococcus* SynAce01 is from Ace Lake. Furthermore, there is genomic diversity among the six cyanobacteria MAGs and genome that come from the polar environments, which indicates the need for more sequencing of cyanobacteria genomes both from established sampling environments as well as under sampled environments.

CHAPTER 3

Biogeography of Five Antarctic Cyanobacteria Using Large-Scale k-mer Searching with sourmash MAGsearch

INTRODUCTION

Cyanobacteria are a diverse group of bacteria that are prevalent in a wide range of environments. In polar environments, cyanobacteria play an important part in shaping local ecology because of their role as primary producers (Stal, 2007; Quesada & Vincent, 2012; Chrismas et al., 2016). Cyanobacteria that live in Antarctica face many challenges including variable light availability, cold temperatures, and surviving freeze-drying conditions on the continent of Antarctica. To withstand these conditions, cyanobacteria may have tolerance mechanisms encoded in their genomes (Chrismas et al., 2015, 2016). However, the presence of tolerance genes in their genomes may make it more difficult for polar cyanobacteria to compete with other cyanobacteria in non-polar environments. To combat this, some polar cyanobacteria may only live in polar environments, while others may also live in environments that have similar conditions to the stresses they face in Antarctica, such as cold temperatures or light stress (Jungblut et al., 2016; Chrismas, Williamson, et al., 2018b; Lumian et al., 2021).

Currently, polar cyanobacteria are underrepresented in genomic data, despite the important role they play in primary productivity. One relatively well studied area for polar cyanobacteria is the McMurdo Dry Valleys in Antarctica. The McMurdo Dry Valleys are home to numerous perennially ice-covered lakes which contain vast microbial mats, including lakes Vanda and Fryxell (Jungblut et al., 2016; Sumner et al., 2016; Lumian et al., 2021). Due to a lack of

predators and limited water mixing, microbial mats prosper and sustain complex geochemical gradients in the lakes. These geochemical gradients structure competition within the communities, which are also dealing with challenging environmental conditions, such as highly seasonal light, nutrient limitation, and in part of Lake Fryxell, sulfidic water (Lumian et al., 2021).

The question of why Antarctic cyanobacteria can survive in challenging conditions and what other environments they live in can be addressed by biogeography studies. Most biogeography studies are based on 16S amplicon sequencing, with insights derived from the environmental conditions of sampling sites (Namsaraev et al., 2010; Bahl et al., 2011; Moreira et al., 2013; Harke et al., 2016; Ribeiro et al., 2018). Although 16S sequences are computationally easy to compare to each other, there are limitations to 16S-based biogeography studies. The 16S sequence does not assemble and bin well from metagenomes, which can prohibit MAGs from being compared with 16S-based biogeographical distributions. Because the MAGs presented in this work do not contain 16S sequences, it is not possible to analyze their presence in comparison to existing Antarctic biogeography work (Jungblut et al., 2010). An alternative to 16S-based biogeography is to search for genomic material of interest in high throughput sequencing data, but this is computationally more complicated. One option is to use an alignment-based approach in which the reads are aligned to reference genomes, which has been done for a large-scale viral discovery with Serratus (Edgar et al., 2022). Another option is large-scale k-mer matching, which is possible with sourmash MAGsearch (Brown & Irber, 2016; Irber, 2020a, 2020b; Brown, 2021). This technique opens the possibility of using metagenomic data for biogeography studies by searching all publicly available *unassembled* metagenomes on the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) (Leinonen et al., 2011) for

Antarctic MAGs of interest. In this chapter, sourmash MAGsearch was used to identify the presence of five Antarctic cyanobacteria MAGs in previously sampled environments across the globe using metagenome data sets on the NCBI SRA.

MATERIALS AND METHODS

Sourmash Methods and Application

The sourmash MAGsearch software used large-scale k-mer searching to search all metagenomes in the NCBI SRA as of September 2020 for matches with genomes of interest (Brown & Irber, 2016; Pierce et al., 2019). A specific signature file for each MAG was calculated based on k-mer abundance using sourmash with a k-mer size of 31. The signature files, which are much smaller than the MAG files, are used as queries to search signatures of publicly available metagenomes in the SRA. The use of k = 31 as a k-mer size enables detection of matches to ~91% average nucleotide identity (ANI) (Pierce-Ward, 2022). Matches are organized by containment percentage, which is the fraction of the k-mers from the query MAG of interest found in the subject metagenome. This technique can be applied to find any genome of interest in publicly available metagenomes (Pierce et al., 2019), which is a robust approach to expand biogeography studies.

Polar Cyanobacteria of Interest

Phormidium pseudopriestleyi is a well characterized cyanobacteria in Lake Fryxell, Antarctica (Lumian et al., 2021). Lake Fryxell is a perennially ice-covered lake located at 77.36° S, 162.6° E in the McMurdo Dry Valleys. The floor of the lake is covered with microbial mats, with *P. pseudopriestleyi* dominating the mats at 9.8 m in depth, where light levels are low (1-2 µmol photons m^{-2} s^{-1}) and sulfide is present in the water column (0.091 mg L^{-1}). *P. pseudopriestleyi* performs oxygenic photosynthesis in the presence of hydrogen sulfide, even though sulfide inhibits oxygenic photosynthesis (Sumner et al., 2015; Lumian et al., 2021). Lake conditions and sampling have been described in Jungblut et al. (2016), Dillon et al. (2020), and Lumian et al. (2021).

The *Neosynechococcus*, *Leptolyngbya*, *Microcoleus*, and *Pseudanabaena* MAGs are from microbial mats located in Lake Vanda, McMurdo Dry Valleys (Chapter 2). Lake Vanda is also a perennially ice-covered lake and is located at 77.53° S, 161.58° E. Microbial mats in Lake Vanda contain pinnacles that range from millimeters to centimeters tall. Unlike Lake Fryxell, there is no sulfide where sampled, and it is better illuminated at the sampled location than Lake Fryxell, though samples from the inside of pinnacles receive little light (Sumner et al., 2016). Sampling methods and lake conditions have previously been described in Sumner et al. (2016) and Chapter 2.

Bioinformatics to Obtain Antarctic Reference MAGs

Methods to obtain MAGs have been previously described in Lumian et al. (2021) for *P. pseudopriestleyi* MAG and Chapter 2 for the *Microcoleus, Neosynechococcus, Pseudanabaena,* and *Leptolyngbya* MAGs. Briefly, the *P. pseudopriestleyi* MAG was obtained from a microbial mat sample sequenced on an Illumina HiSeq 2500 PE250 platform and a laboratory culture was sequenced on an Illumina 2000 PE100 platform. The microbial mat sample was quality filtered, and forward and reverse reads were joined using PEAR v0.9.6 (Zhang et al., 2014). For the lab culture, trimmomatic v0.36 (Bolger et al., 2014) was used to trim sequencing adapters, and the interleave-reads.py script in khmer v2.1.2 (Crusoe et al., 2015) was used to interleave the reads. Both samples were assembled separately and together as a co-assembly by MEGAHIT v1.1.2 (Li et al., 2015) and mapped with bwa v2.3 (Li, 2013) and samtools v1.9 (Li et al., 2009). A single cyanobacteria bin was obtained using the CONCOCT binning algorithm in anvi'o and identified using CheckM (Eren et al., 2015; Delmont & Eren, 2018; Parks et al., 2015). The *P. pseudopriestleyi* bin was refined with spacegraphcats to extract additional content from the metagenomes with a k-mer size of 21 and a radius of 1 (Brown et al., 2020).

Methods to obtain the *Microcoleus, Pseudanabaena, Leptolyngyba*, and *Neosynechococcus* MAGs from Lake Vanda have been previously described in Grettenberger et al., (2020). Filtered and quality controlled raw data was retrieved from IMG Gold with JGI Gold IDs GP0191362 and Gp0191371. MEGAHIT v1.9.6 was used to assemble metagenomes with a minimum contig length of 500 bp and a paired end setting. Bowtie2 v1.2.2 and samtools v1.7 were used to map reads back to the assembly. A depth file was generated using jgi_summarize_bam_contig_depths from MetaBAT v2.12.1 (Kang et al., 2015), which was also

used to generate bins with a minimum contig length of 2500 bp. The lineage of the MAGs were classified with CheckM and GTDB-tk (Chaumeil et al., 2020).

Biogeography Analysis with sourmash

Signature files of the genomes of interest were generated using sourmash v3.5.0 (Brown & Irber, 2016) with k-mer sizes of 21, 31, 51, the --scaled parameter set to 1000, and abundance tracking. This generated a unique signature file specific to each of the five Antarctic MAGs. These signature files were searched against signature files previously generated for all 498,942 publicly available metagenome sets on the SRA as of September, 2020 using exact k-mer matching. Matches are organized by containment, which is the proportion of the query MAG k-mers found in the metagenome results at the specified ANI. Containment can be affected by a small metagenome size and even results with a low containment value may be relevant for identifying sequences related to a query MAG. The size of the Antarctic query MAGs ranged from 2.7 Mbp – 6.07 Mbp (Chapter 2), so a match with containment value of 5% implies 135,000 – 303,500 matching k-mers with k = 31 and 4,185,000 – 9,408,500 matching base pairs, which indicates significant shared genomic material between MAGs and metagenome matches.

Environmental metadata for the top hits of all MAGs with hits above 5% were recorded, with the exception of the *Microcoleus*, which had over 1,000 matches above that threshold. Select metagenomes with the highest containment values for each of the MAGs were downloaded from the SRA for further analysis. To represent geographical diversity, metagenomes were chosen from unique environments besides Lakes Vanda and Fryxell, with the

exception of a data set from Lake Fryxell mat lift-off and glacial meltwater (accession number SRR5468153) because it was the only environment with high containment matches for the *Neosynechococcus* and *Leptolyngbya* MAGs. These fifteen metagenomes were assembled, binned and annotated using the same method as the MAGs from Lake Vanda: MEGAHIT v1.9.6 was used for assembly, bowtie2 v1.2.2 and samtools v1.7 were used to map reads, MetaBAT v2.12.1 was used to bin the metagenomes. QUAST was used to generate assembly statistics.

To verify the results of the sourmash MAGsearch, the signatures of the query MAGs were searched for in the fifteen assembled and unassembled metagenomes downloaded from the SRA using sourmash search --containment with a k-mer size of 31. The code from this project is available at: https://github.com/jessicalumian/biogeo.

RESULTS

The five polar cyanobacteria MAGs used as search queries were found in a variety of non-polar metagenomic data sets in a range of environmental conditions (Table 3.1). The metagenome data sets with the highest containment of the MAGs are described in Tables 3.1 – 3.3 and 3.5. Information about additional environments where the *Microcoleus* was found with over 20% containment is displayed in Table 3.4. The SRA accession numbers of additional hits are available in Supplementary Tables 3.1.1 – 3.1.5.

The purpose of sourmash MAGsearch is to find shared k-mers between Antarctic MAGs and SRA metagenomes, and the presence of MAGs in these metagenomes may indicate the presence of the Antarctic organisms across the globe. A k-mer size of 31 indicates a ~91% ANI

between matched sequences with at least 5% containment. At 30% containment, this value increases to ~97% ANI (Pierce-Ward, 2022). Thus, a high containment value indicates the presence of the MAG in the metagenome, and supports the presence of the organism in the sampling location of that metagenome. Low containment values still represent shared genomic material with a k-mer size of 31, but cannot definitively support the presence of the Antarctic cyanobacteria in that environment. Containment, particularly at small values, can be affected by factors such as plasmids or small portions of shared contamination between the MAG and metagenome. As a validation for the sourmash MAGsearch k-mer method, assembled and unassembled SRA metagenomes that were selected for further bioinformatics analysis were mapped back to Antarctic MAGs (Table 3.5) using minimap2 v2.24 in genome-grist v0.8.3 (Li, 2018; Irber et al., 2022).

Table 3.1 Summary of sourmash MAGsearch hits

	Microcoleus	***Phormidium pseudopriestleyi***	***Pseudanabaena***	***Neosynechococcus***	***Leptolyngbya***
# Hits >75% Containment	6	30	6	3	5
# Hits >50% Containment	12	33	6	5	5
# Hits >25% Containment	119	38	10	6	6
# Hits >5% Containment	1,121	131	24	16	22
Total Hits	6,184	2,739	3,769	2,796	2,999
# Geographically Distinct Locations >25% Containment	27	3	3	1	1

Table 3.2 Metagenomes from the NCBI SRA with the Highest Containment Values

MAG	Containment (%)	Location	Accession Number
Microcoleus	99.18*	Mat lift-off from Lake Fryxell, Antarctica	SRR5468150
	65.02*	Polar Desert Sand Communities, Antarctica	SRR6266358
	57.50*	Moab Green Butte, Utah, USA	SRR5855414
	41.65*	Ningxia, China	SRR2952554
	41.10*	Sonoran Desert, Colorado Plateau, USA	SRR5247052
	40.61*	Pig Farm, UK	ERR3588763
	39.54*	Glacier Snow, China	SRR5891573
	38.36*	Mine Tailing Pool Sediment near Shaoyang, China	ERR1333181
	37.04*	Wastewater in Milwaukee, Wisconsin, USA	SRR5459769
	36.30*	Puca Glacier, Peru	SRR6048908
	35.71*	Negev Desert, Israel	SRR12473531
	33.58*	Southwest Germany	ERR192241
Phormidium pseudopriestleyi	98.49*	Microbial mat in Lake Fryxell	SRR7769747
	55.80*	Ace Lake, Antarctica	SRR528444
	23.54*	Rauer Islands, Antarctica	SRR5216658
	20.63*	Les Salins du Lion Bird Reserve, France	SRR7428116
	19.04*	Big Soda Lake, Nevada	SRR12522841
	18.25*	Étang de Berre Lagoon, France	SRR7428132
	11.99*	Sewage in Nairobi, Kenya	ERR3503286
	10.37	Wetland soil in Yanghu, China	SRR9691033
	8.98*	Salar del Huasco salt flat, Chile	SRR10186387
	8.48*	Simulated Metagenome	ERR738546
	7.61*	Human Gut	SRR6262267
Pseudanabaena	99.49*	Mat lift-off from Lake Fryxell, Antarctica	SRR5468149
	37.45	Dry Valley Sand Communities, Antarctica	SRR6266338
	33.54*	Nunavut, Canada	SRR5829599

	18.31*	Deception Island, Antarctica (Whaler's Bay Sediment)	ERR4192538
	7.45*	Microbial mat in Lake Fryxell	SRR7769784
	6.40*	Barataria Bay, Louisiana, USA	SRR2657229
	6.01	Sacramento Delta, California, USA	SRR5198900
Neosynechococcus	97.82*	Mat lift-off from Lake Fryxell, Antarctica	SRR5208701
	5.71*	Sacramento Delta, California, USA	SRR5198900
	5.31*	Amazon Forest	SRR490140
	5.19*	Barataria Bay, Louisiana, USA	SRR2657237
Leptolyngbya	98.72*	Mat lift-off from Lake Fryxell, Antarctica	SRR5468150
	8.40	Spitsbergen, Svalbard, Norway	SRR6683740
	6.56	Sacramento Delta, California, USA	SRR5198900

An asterisk (*) denotes where multiple samples from the same location above 5% containment were identified but are not shown in this table, which only shows distinct hits. The sample with the highest containment is shown. For an extended full list of hits, see Supplementary Tables 3.1.1 – 3.1.5.

Table 3.3 Environmental Conditions of Metagenomic Data Sampling Sites from Table 2

MAG	Location	Latitude and Longitude	BioSample Metadata
Microcoleus	Lake Fryxell, Antarctica	77.605 S, 163.1630 E	Isolation Source: Ice surface mat Collection Date: 2014-12-05
	Dry Valleys, Antarctica	78.0741 S, 163.8918 E	Isolation Source: Antarctic Sand Collection Date 2010-01-15
	Moab Green Butte, Utah, USA	38.42 N, 109.41 W	Isolation Source: biocrust samples from Green Butte Site near Canyonlands National Park along an apparent maturity gradient of Cyanobacteria-dominated biocrusts Collection Date: 2014-09 Sample Collection: soil coring with Petri dishes
	Ningxia, China	Not provided	Isolation Source: algae crusts Collection Date: 2013-04-15 Geographic Location: China: Ningxia

	Sonoran Desert, Colorado Plateau, USA	38.42 N, 109.4099 W	Isolation Source: Colorado Plateau and Sonoran Desert Plateau and Sonoran Desert
	Pig Farm, UK	55.95 N, -3.188 W	Environmental Context: Pig farm soil Collection Date: 2017-01-11 Project Name: The dynamics of antimicrobial resistance gene prevalence on a commercial pig farm: implications for policy
	Glacier Snow, China	38.2186 N, 81.120 E	Environmental Context: Glacier snow from glacier Collection Date: 2013-09 Depth: 0.01 m Elevation: 5800 m
	Mine Tailing Pool near Shaoyang, China	27.745 N, 111.46 E	Environmental Context: Mine tailing pool, sediment Collection Date: 2014-12-27 Depth: 0.1 m (?) Elevation: 286 m (?)
	Wastewater in Milwaukee, Wisconsin, USA	43.023 N, 87.895 W	Environmental Context: Wastewater communities Collection Date: 2014-07-17
	Puca Glacier, Peru	13.773 S, 71.071 W	Environmental Context: Early successional soil, N + P addition Collection Date: 2012
	Negev Desert, Israel	30.785 N, 34.767 E	Environmental Context: Temperate “Mediterranean” desert biome Collection Date: 2017-05-10 Depth: 0.2 cm Elevation: 0 m
	Southwest Germany	Not Provided	BioProject Information: Short read whole genome sequencing of 276 wild *Arabidopsis thaliana* rosettes from southwest Germany
Phormidium pseudopriestleyi	Lake Fryxell, Antarctica	77.6167 S, 163.1833 E	Collection Date: 2012-11 Isolation source: benthic surface of ice-covered lake
	Ace Lake, Antarctica	68.473 S, 78.188 E	Collection Date: 2014-02-15 Isolation Source: saline lake
	Rauer Islands, Antarctica	68.556 S, 78.191 E	Collection Date: 2015-01-11 Isolation Source: saline lake
	Bird Reserve, France	43.453 N, 5.230 E	Environmental Context: Microbial mat from brackish lagoon in a natural zone of ecological interest Collection Date: 2011-09 Depth: 0.2 cm Elevation: 0 m
	Big Soda Lake, Nevada, USA	39.523 N, 118.870 W	Collection Date: 2016-06-22
	Étang de Berre Lagoon, France	43.485 N, 5.188 E	Environmental Context: Microbial mat from hydrocarbon retention basic Collection Date: 2012-04 Depth: 0.2 cm Elevation: 2 m
	Sewage in Nairobi, Kenya	1.19 S, 36.47 E	Environmental Context: Stream collection for survey of infectious diseases and antimicrobial resistance Collection Date: 2014-07-28

	Wetland soil in Yanghu, China	29.19 N, 90.59 E	Collection Date: 2013-07-12
	Salar del Huasco salt flat, Chile	20.264 S, 68.875 W	Collection Date: 2018-06-02
	Simulated Metagenome	Not Applicable	Simulated metagenome based on real Illumina HiSeq 2000 data to benchmark metagenome analysis tools
	Human Gut	40.431 N, 79.959 W	Sample Context: Fecal sample from infant Collection Date: 2015
Pseudanabaena	Dry Valley Sand Communities, Antarctica	78.072 S, 163.719 E	Collection Date: 2009-01-15 Isolation Source: Antarctic Sand
	Borup Fiord, Nunavut, Canada	81.017 N, 81.583 W	Collection Date: 2015-07-18 Organism: Glacier metagenome
	Lake Fryxell, Antarctica	77.605 S, 163.1630 E	Isolation Source: Ice surface mat Collection Date: 2014-12-05
	Barataria Bay, Louisiana, USA	29.456 N, 89.887 W	Collection Date: 2012-06-06 Isolation Source: Marine sediment
	Sacramento Delta, California, USA	38.107 N, 121.649 W	Collection Date: 2011-08-18 Isolation Source: Wetland sediment
Neosynechococcus	Mat lift-off from Lake Fryxell, Antarctica	77.605 S, 163.163 E	Collection Date: 2014-12-05 Isolation source: Ice surface mat
	Sacramento Delta, California, USA	38.107 N, 121.649 W	Collection Date: 2011-08-18 Isolation Source: Wetland sediment
	Amazon Forest	Not Provided	Collection Date: 2010 Isolation Source: Soil metagenome
	Barataria Bay, Louisiana, USA	29.456 N, 89.887 W	Collection Date: 2012-06-06 Isolation Source: Marine sediment
Leptolyngbya	Mat lift-off from Lake Fryxell, Antarctica	77.605 S, 163.163 E	Collection Date: 2014-12-05 Isolation Source: Ice surface mat
	Spitsbergen, Svalbard, Norway	78.58 N, 12.05 E	Collection Date: 2016-07-01 Isolation Source: Soil and sediment
	Sacramento Delta, California, USA	38.107 N, 121.649 W	Collection Date: 2011-08-18 Isolation Source: Wetland sediment

Note: Question marks are used where units are assumed but not given in SRA metadata.

Table 3.4 Additional Metagenomes from Unique Locations >20% Containment for *Microcoleus* MAG

MAG	Containment (%)	Location	Latitude and Longitude	Accession Number
Microcoleus	32.14	Qing River, China	40.029 N, 116.368 E	SRR10571243
	31.77	Fecal Metagenome of Great Black-Headed Gulls around Qinghai Lake, China	36.78 N, 100.00 E	SRR10492798
	31.35	Agave Microbial Communities from Guanajuato, Mexico	21.766 N, 100.163 W	SRR4142282
	31.14	Miers Valley, Antarctica	78.160 S, 164.100 E	SRR3471615
	31.06	Glacial meltwater from Laohugou glacier, China	39.50 N, 96.52 E	SRR9965273
	28.87	Polar Desert Sand Communities, Antarctica	78.024 S, 163.917 E	SRR6266336
	28.77	Ace Lake, Antarctica	68.473 S, 78.188 E	SRR7528444
	28.41	Particulate Aerosol Particulate Matter, Beijing, China	40.01 N, 116.33 E	SRR10613504
	28.40	Soil crust, Chicken Creek, Germany	51.36 N, 14.15 E	SRR8357461
	28.17	Wetland soil, Lanzhou, China	36.09 N, 103.71 E	SRR9691044
	27.50	Terrestrial metagenome from Alberta, Canada	50.34 N, 113.77 W	SRR5678923
	27.15	Cave, Twin Sisters, Idaho	42.02 N, 113.72 W	SRR7774479
	25.88	Soil communities, Rifle, Colorado	39.53 N, 107.78 W	SRR3969602
	25.27	Semi-synthetic marine metagenomes from University of Algarve, Faro, Portugal	NA	ERR1992808
	25.25	Saskatchewan, Canada	50.28 N, 107.80 W	SRR7013884
	24.64	Wheat and chickpea soil microbiome, Australia	34.538 S, 138.690 E	ERR3029103
	23.58	Soil communities, Uluru, Australia	25.350 S, 131.052	ERR671932
	22.19	Rhizosphere soil, Mafikeng, South Africa	25.79 S, 25.61 E	SRR11128415
	20.79	Microbial mat, Eel River, California, USA	39.840 N, 123.710 W	SRR244337
	20.33	Biofilm in Wai-iti River, New Zealand	NA	SRR9948934
	20.29	Soil rhizosphere, Durango, Mexico	19.321 N, 99,194 W	SRR11092592

The *Microcoleus* was the most widely distributed MAG with 27 globally distinct locations above 25% containment (Tables 3.1 and 3.5), which is only a portion of meaningful hits. The *Microcoleus* and *P. pseudopriestleyi* MAGs were present in the most time series and subsamples from the same environmental location, which resulted in 1,121 hits above 5% for the *Microcoleus* MAG and 131 hits for *P. pseudopriestleyi* MAG (Table 3.1). The *Pseudanabaena* and *P. pseudopriestleyi* MAGs were found in three distinct locations above 25% containment while the *Neosynechococcus* and *Leptolyngbya* MAGs were only found in one location above 25% containment (Table 3.1). However, containment can be affected by small metagenomes which have low coverage of community members. Thus, even environmental hits with containment lower than 25% may indicate the presence of a genome related to the query MAG (Table 3.4).

The *Microcoleus* MAG was found in diverse environments with conditions ranging from hot to cold climates and including both arid and wet locations (Tables 3.1 and 3.5). Some environments are cold year-round such as Puca Glacier in Peru (36.30% containment), glacier snow in China (39.54% containment), and the ice-covered Lake Vanda, while others are temperate, like Wisconsin, USA (37.04% containment), or Southwest Germany (33.58% containment). *P. pseudopriestleyi* was found in three Antarctic metagenome data sets: Lake Fryxell mat samples (98.49% containment), Ace Lake (55.8% containment) and the Rauer Islands (23.54% containment). The highest 30 hits for the *P. pseudopriestleyi* MAG, including the three samples used to create the MAG, were from Lake Fryxell. This search revealed that *P. pseudopriestleyi* is present in other depths of Lake Fryxell in addition to the query MAG depth of 9.8 m despite not being prevalent at those depths based on 16S sequencing (Dillon et al., 2020). Besides Antarctica, the *P. pseudopriestleyi* MAG was found in a bird reserve next to a lagoon in

France called Les Salins du Lion (20.63% containment) as well as a hydrocarbon polluted saline lagoon called Étang de Berre (18.25% containment) which were part of a study on the effects of hydrocarbon pollution on microbial communities (Aubé et al., 2016). The *P. pseudopriestleyi* MAG was also found the Salar del Huasco salt flat in Chile (8.98% containment), antimicrobial treated sewage collected in Nairobi, Kenya (11.99% containment) and an infant gut fecal sample (7.61% containment). All these environments represent extreme conditions for cyanobacteria.

Although the *Microcoleus*, *Pseudanabaena*, *Neosynechococcus,* and *Leptolyngbya* MAGs were obtained from microbial mat pinnacles in Lake Vanda, they were all present in high containment (>97%) in mat lift-off samples from Lake Fryxell. The *Pseudanabaena* MAG was also present in a dry sand community in the McMurdo Dry Valleys (37.45% containment), where lakes Vanda and Fryxell are located, as well as Whaler's Bay on Deception Island in Antarctic (18.31 % containment) and Nunavut, Canada (33.54 % containment), which is cold but geographically distant from the Antarctic. Even though the top hits for the *Pseudanabaena, Neosynechococcus,* and *Leptolyngbya* were from cold environments, they were found in metagenomes from non-polar conditions with <10% containment: the *Pseudanabaena* and *Neosynechococcus* MAGs were found in data from Barataria Bay, Louisiana, USA; the *Pseudanabaena* and *Leptolyngbya* MAGs were found in data from the Sacramento Delta, California, USA; and the *Neosynechococcus* MAG was found in data from the Amazon rainforest.

Table 3.5 Quality Metrics of Metagenome Assemblies and Mapping Statistics

MAG	SRA Accession Number and Location	Number of Contigs (over 500 bp)	Total Length of Contigs (bp, over 500 bp)	Largest Contig (bp)	N50	% of Assembly Mapped to Reads	% of MAG Mapped to Unassembled Metagenome
Microcoleus	SRR5855414 Moab Green Butte, Utah, USA	779, 743	497,286,103	4,248	612	-	88
	SRR2952554 Ningxia, China	296,291	187,141,964	7,947	604	-	76
	SRR5247052 Sonoran Desert, Colorado Plateau, USA	781,528	498,867,695	3,982	613	0.18	74
	ERR3588763 Pig Farm, UK	402,432	243,270,443	3,395	581	22.39	77
	SRR5891573 Glacier Snow, China	979,446	619,490,337	3,614	606	22.22	77
	ERR1333181 Antimony Polluted Sediment, China	415,411	260,914,760	5,701	607	13.04	75
	SRR5459769 Wastewater in Milwaukee, Wisconsin USA	1,259,786	1,086,536,868	57,883	819	6.62	77
	SRR12473531 Negev Desert, Israel	297,517	187,591,328	4,696	600	10.67	76
	ERR192241 Southwest Germany	108,078	63,950,717	1,868	567	16.78	-
Phormidium pseudopriestleyi	SRR7528444 Ace Lake, Antarctica	218,966	195,047,488	62,020	822	7.05	62
	SRR5216658 Rauer Islands, Antarctica	250,605	212,798,142	64,599	794	-	28
	SRR7428116 Bird Reserve, France	462,930	371,142,339	36,453	146, 664	7.87	61
Pseudanabaena	SRR6266338 Dry Valley, Antarctica	651,239	424,874,538	9,934	624	-	44

	SRR5829599 Canada	704,963	564,862,171	27,496	751	6.48	79
Neosynechococcus	SRR5468153 Lake Fryxell, Antarctica	1,119,433	887,670,078	21,258	760	4.26	33
Leptolyngbya	SRR5468153 Lake Fryxell, Antarctica	1,119,433	887,670,078	21,258	760	4.26	26

Metagenomes representing geographically distinct locations were selected for further analysis to compare genomic data from different environments to the Antarctic MAGs. These data sets were run through an assembly and binning pipeline to obtain bins that could be compared to the Antarctic MAGs. However due to low coverage, metagenomes assemblies were poor quality with the majority of the N50s under 1,000 base pairs, which is the minimum contig length required to bin with MetaBAT. Thus, bins were not generated, and it would not have been possible to identify the presence of the MAGs in these metagenomes without using an assembly-independent technique.

DISCUSSION

Environmental Diversity of Microcoleus

The presence of the *Microcoleus* MAG in diverse environments indicates that it can survive in a variety of ecological conditions. In order to survive cold temperatures in Lake Vanda, the *Microcoleus* must deal with cellular membranes becoming brittle and slowed metabolism. However, some environments where the *Microcoleus* was found are only cold for part of the year (Moab Green Butte Desert; Ningxia, China; Southwest Germany; Milwaukee, Wisconsin; and the

UK) while other environments are cold year-round (Puca Glacier, Peru, and glacial snow in China). In contrast to cold conditions, hot temperatures can cause proteins to denature and prolonged exposure to sunlight can cause high light and UV stress. These conditions occur in the Moab Green Butte Desert, the Sonoran Desert, and the Negev Desert. Furthermore, the Moab Desert, and Sonoran Desert experience extreme temperature changes between morning and night (Turnage & Hinckley, 1938; Balling et al., 1998; McCann et al., 2018), forcing the *Microcoleus* to adapt to both conditions on a 24-hour cycle.

In addition to temperature range, the *Microcoleus* MAG was found in metagenomes from a variety of other conditions. Two locations came from dry soils (Moab and Negev Deserts), while other environments have seasonal precipitation (Shaoyang, China; the UK; Milwaukee, Wisconsin and Southwest, Germany) or continuous water (Qing River and Ace Lake). The *Microcoleus* MAG was also found in data from both high and low elevation environments (5800 m elevation in glacial snow in China and 0 m elevation in a bird reserve and lagoon in southern France). Interestingly, in Southwest Germany the MAG was found in metagenomic data of wild Arabidopsis plants. Overall, the variety of conditions where the *Microcoleus* MAG was found indicates that it may live in an impressive range of environments, such as extreme heat or cold, or moderate climates.

Environmental Diversity of Phormidium pseudopriestleyi

P. pseudopriestleyi is a sulfide-tolerant cyanobacteria found in a low light environment in Lake Fryxell, Antarctica. 16S sequencing has shown that it is present in other Antarctic

environments including Salt Pond and Fresh Pond (Jungblut et al., 2005; Lumian et al., 2021), so it is not surprising that the *P. pseudopriestleyi* MAG was found in saline Ace Lake and the Rauer islands metagenomes. Interestingly, a relative appears to be present in a pond at Les Salins du Lion, a bird reserve, and Étang de Berre, a hydrocarbon polluted saline lagoon, in southern France (Aubé et al., 2016). Four environmental conditions can be compared in these locations: irradiance, salinity, temperature, and sulfide concentrations. The amount of irradiance at Les Salins du Lion pond and Étang de Berre lagoon was not measured when environmental sampling occurred, but the elevation was recorded to be at 0 m, indicating that irradiance is higher at the surface of the pond than the low irradiance at the depth of sampling in Lake Fryxell (1-2 µmol/photon m^{-2} s^{-1}) (Sumner et al., 2015). Furthermore, Salt Pond and Fresh Pond have high illumination levels in the summer (Roos & Vincent, 1998; Jungblut et al., 2005), indicating that *P. pseudopriestleyi* can deal with light stress. Les Salins du Lion (14 g L^{-1} NaCl) and Étang de Berre (20 g L^{-1} NaCl) have a lower salinity than Lake Fryxell (70.13 g L^{-1} NaCl) and Salt Pond (~990 g L^{-1} NaCl), which is hypersaline (Jungblut et al., 2005; Aubé et al., 2016; Lumian et al., 2021). Previous work has showed that *P. pseudopriestleyi* increases the thickness of its extracellular polymeric substance layer in response to saline stress (Agrawal & Singh, 1999). Sulfide is also present in Les Salins du Lion, with a concentration of ~0.24 g L^{-1} at the time of sampling (Aubé et al., 2016), which was the highest value at any location or time sampled included in the study. This demonstrates a much higher sulfide tolerance than what was previously recorded in the Lake Fryxell sampling site, which was 9.8 x 10^{-5} g L^{-1} (Lumian et al., 2021).

In addition to Les Salins du Lion and Étang de Berre, the *P. pseudopriestleyi* MAG was found in globally distributed challenging environments such as a salt flat in Chile, antimicrobial

treated sewage in Kenya, and infant gut. The fact that *P. pseudopriestleyi* grows in difficult conditions suggests that it may specialize in extreme locations. In Lake Fryxell, *P. pseudopriestleyi* dominates microbial mats at 9.8 m depth in low light and sulfidic conditions but it is less abundant at shallower depths, even though there is more light availability and no sulfide (A. D. Jungblut et al., 2016; Dillon et al., 2020). Thus, *P. pseudopriestleyi* may grow slowly and find ecological success in environments that are too harsh for quicker growing cyanobacteria, which is consistent with the slow growth rate of *P. pseudopriestleyi* observed in unpublished lab data. The other environments where genomes similar to *P. pseudopriestleyi* were found may provide challenges that prohibit many other cyanobacteria from growing, allowing *P. pseudopriestleyi* to survive in a non-polar environment.

Environmental Diversity of Pseudanabaena, Neosynechococcus, and Leptolyngbya

The top matches for the *Pseudanabaena, Neosynechococcus*, and *Leptolyngbya* MAGs showed that they were also present in Lake Fryxell and that the *Pseudanabaena* MAG was in sediment in the McMurdo Dry Valleys. The presence of these MAGs in geographically distant locations at lower containment suggests they can survive transport to other locations. The mechanism of long-range distribution for this organism could be wind; atmospheric studies show bacteria from the Saharan desert are transported by wind throughout the Atlantic (Griffin et al., 2002; Gorbushina et al., 2007; A. D. Jungblut et al., 2010). A similar process is expected to allow Antarctic cyanobacteria to cross large distances and populate diverse geographic regions. The presence of the *Pseudanabaena* and *Neosynechococcus* MAGs in data from Barataria Bay,

Louisiana, USA and the *Pseudanabaena* and *Leptolyngbya* MAGs in data from the Sacramento Delta, California, USA indicate that these organisms may be transported globally and survive in non-polar environments. However, the low number of non-polar locations compared to the *Microcoleus* MAG suggests that they are not as successful at integrating into non-polar environments. In these favorable environments, the *Pseudanabaena*, *Leptolyngbya* and *Neosynechococcus* are likely out competed by other organisms who are more effective at growing faster and persisting in non-polar communities.

Implications for Biogeographic Distributions

The perceived distributions of organisms in biogeography studies are affected by sampling and publishing biases. Sampling in remote locations is logistically difficult and is often centered around established sampling locations which may be near research stations and infrastructures. This results in many studies and publications from established sampling locations and a deeper understanding of local ecology and geochemical processes in these environments. Biogeography studies, however, benefit from widespread sampling in many locations. Conducting widespread ecological sampling is expensive and can be impractical, so it is advantageous to search existing data sets for as much information as possible. Using sourmash MAGsearch to search public metagenomes makes the most out of data from remote areas by revealing previously unknown locations of organisms of interest. Furthermore, results from this analysis included remote areas, including various sites in Antarctica, which may not have otherwise been identified as locations of the query MAGs.

Despite being affected by sampling bias like all biogeography studies, the results showed presence of the *Microcoleus* MAG globally distributed over a wide variety of environments, the *P. pseudopriestleyi* MAG in predominantly in harsh environments, and the *Pseudanabaena, Neosynechococcus,* and *Leptolyngbya* MAGs in both polar and non-polar environments, but much less widespread than the *Microcoleus* MAG. The numerous sites with the *Microcoleus* MAG suggest that it contains genetic capacity to adapt to many types of environments. It may also have a faster growth rate than an extreme conditions specialist, like *P. pseudopriestleyi*, which would allow it to compete in a variety of ecological communities, some of which experience stressful conditions.

Although the *Microcoleus* MAG is by far the most globally diverse cyanobacteria in this study, there is variety in the distributions of the other four MAGs. The prevalence of the *P. pseudopriestleyi* MAG in harsh environments indicates that it finds ecological successful in stressful environments, and it is likely outperformed by other organisms in moderate environments. The *Pseudanabaena*, *Neosynechococcus*, and *Leptolyngbya* MAGs were found in a few non-polar environments, indicating they can survive in moderate environments. A genomic analysis of photosynthesis and circadian rhythm genes in chapter 2 did not reveal a reason why the *Microcoleus* is so prevalent compared to the other MAGs in this study, though previous work has shown *Microcoleus* to be a cosmopolitan genus (Garcia-Pichel et al., 1996, 2001). Diving deeper into the metabolic potential of each organism and interactions between metagenome community members may offer insights as to why some organisms are prevalent in a multitude of environments while others are prevalent in only certain conditions.

CONCLUSIONS

This chapter presents the first biogeography study using the large-scale k-mer-based sourmash MAGsearch software. This search revealed the *Microcoleus* MAG is present in a variety of environments, the *P. pseudopriestleyi* MAG is present in harsh environments, and the *Neosynechococcus, Leptolyngbya,* and *Pseudanabaena* MAGs were found in polar environments. Further study of these MAGs may reveal why the *Microcoleus* is so pervasive compared to the other cyanobacteria in this study.

The ability to use sourmash MAGsearch to find genomes similar to query MAGs in publicly available metagenomic data sets has important implications for biogeography studies, which have been predominantly based on 16S sequencing due of the prevalence of data and ease of comparison. The ability to search public metagenomes for organisms of interest greatly increases the amount of data that can be used for biogeography studies. This technique is especially helpful for organisms that are in remote locations and underrepresented in genomic data, such as polar cyanobacteria, by increasing the number of known environments much faster than would be possible with targeted field studies. Additionally, sourmash MAGsearch can be used to identify accessible sampling locations of organisms from remote environments, such as the *Pseudanabaena* and *Leptolyngbya* MAGs from Lake Vanda being identified in the Sacramento Delta, though further validation should be done for matches with low containment. As more data sets are made publicly available on the NCBI SRA, more information about the distribution of cryosphere cyanobacteria can be attained.

References

Agrawal, S. C., & Singh, V. (1999). Viability of dried vegetative trichomes, formation of akinetes and heterocysts and akinete germination in some blue-green algae under water stress. *Folia Microbiologica*, *44*(4), 411–418. https://doi.org/10.1007/BF02903715

Anesio, A. M., Hodson, A. J., Fritz, A., Psenner, R., & Sattler, B. (2009). High microbial activity on glaciers: Importance to the global carbon cycle. *Global Change Biology*, *15*(4), 955–960. https://doi.org/10.1111/j.1365-2486.2008.01758.x

Anesio, A. M., & Laybourn-Parry, J. (2012). Glaciers and ice sheets as a biome. *Trends in Ecology & Evolution*, *27*(4), 219–225. https://doi.org/10.1016/j.tree.2011.09.012

Aoki, S., & Onai, K. (2009). Circadian Clocks of Synechocystis sp. Strain PCC 6803, Thermosynechococcus elongatus, Prochlorococcus spp., Trichodesmium spp. And Other Species. In J. L. Ditty, S. R. Mackey, & C. H. Johnson (Eds.), *Bacterial Circadian Programs* (pp. 259–282). Springer. https://doi.org/10.1007/978-3-540-88431-6_15

Arkin, A. P., Cottingham, R. W., Henry, C. S., Harris, N. L., Stevens, R. L., Maslov, S., Dehal, P., Ware, D., Perez, F., Canon, S., Sneddon, M. W., Henderson, M. L., Riehl, W. J., Murphy-Olson, D., Chan, S. Y., Kamimura, R. T., Kumari, S., Drake, M. M., Brettin, T. S., … Yu, D. (2018). KBase: The United States Department of Energy Systems Biology Knowledgebase. *Nature Biotechnology*, *36*(7), 566–569. https://doi.org/10.1038/nbt.4163

Aubé, J., Senin, P., Pringault, O., Bonin, P., Deflandre, B., Bouchez, O., Bru, N., Biritxinaga-Etchart, E., Klopp, C., Guyoneaud, R., & Goñi-Urriza, M. (2016). The impact of long-term hydrocarbon exposure on the structure, activity, and biogeochemical functioning of microbial mats. *Marine Pollution Bulletin*, *111*(1), 115–125. https://doi.org/10.1016/j.marpolbul.2016.07.023

Bahl, J., Lau, M. C. Y., Smith, G. J. D., Vijaykrishna, D., Cary, S. C., Lacap, D. C., Lee, C. K., Papke, R. T., Warren-Rhodes, K. A., Wong, F. K. Y., McKay, C. P., & Pointing, S. B. (2011). Ancient origins determine global biogeography of hot and cold desert cyanobacteria. *Nature Communications*, *2*(1), 163. https://doi.org/10.1038/ncomms1167

Balling, R. C., Klopatek, J. M., Hildebrandt, M. L., Moritz, C. K., & Watts, C. J. (1998). Impacts of Land Degradation on Historical Temperature Records from the Sonoran Desert. *Climatic Change*, *40*(3), 669–681. https://doi.org/10.1023/A:1005370115396

Barco, R. A., Garrity, G. M., Scott, J. J., Amend, J. P., Nealson, K. H., & Emerson, D. (2020). A Genus Definition for Bacteria and Archaea Based on a Standard Genome Relatedness Index. *MBio*, *11*(1). https://doi.org/10.1128/mBio.02475-19

Barria, C., Malecki, M., & Arraiano, C. M. (2013). Bacterial adaptation to cold. *Microbiology, 159*(Pt_12), 2437–2443. https://doi.org/10.1099/mic.0.052209-0

Beer, D. de, Weber, M., Chennu, A., Hamilton, T., Lott, C., Macalady, J., & Klatt, J. M. (2017). Oxygenic and anoxygenic photosynthesis in a microbial mat from an anoxic and sulfidic spring. *Environmental Microbiology*, *19*(3), 1251–1265. https://doi.org/10.1111/1462-2920.13654

Behrendt, L., Brejnrod, A., Schliep, M., Sørensen, S. J., Larkum, A. W., & Kühl, M. (2015). Chlorophyll f -driven photosynthesis in a cavernous cyanobacterium. *The ISME Journal*, *9*(9), 2108–2111. https://doi.org/10.1038/ismej.2015.14

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. https://doi.org/10.1093/nar/gks1195

Bishop, N. I. (1958). The influence of the herbicide, DCMU, on the oxygen-evolving system of photosynthesis. *Biochimica Et Biophysica Acta*, *27*(1), 205–206. https://doi.org/10.1016/0006-3002(58)90313-5

Bogorad, L. (1975). Phycobiliproteins and Complementary Chromatic Adaptation. *Annual Review of Plant Physiology*, *26*(1), 369–401. https://doi.org/10.1146/annurev.pp.26.060175.002101

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bronstein, M., Schütz, M., Hauska, G., Padan, E., & Shahak, Y. (2000). Cyanobacterial Sulfide-Quinone Reductase: Cloning and Heterologous Expression. *Journal of Bacteriology*, *182*(12), 3336–3344. https://doi.org/10.1128/JB.182.12.3336-3344.2000

Brown, C. T. (2021, June 8). Searching all public metagenomes with sourmash. *Living in an Ivory Basement*. http://ivory.idyll.org/blog/2021-MAGsearch.html

Brown, C. T., & Irber, L. (2016). sourmash: A library for MinHash sketching of DNA. *Journal of Open Source Software*, *1*(5), 27. https://doi.org/10.21105/joss.00027

Brown, C. T., Moritz, D., O'Brien, M. P., Reidl, F., Reiter, T., & Sullivan, B. D. (2020). Exploring neighborhoods in large metagenome assembly graphs using spacegraphcats reveals hidden sequence diversity. *Genome Biology*, *21*(1), 164. https://doi.org/10.1186/s13059-020-02066-4

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Canfield, D. E. (1998). A new model for Proterozoic ocean chemistry. *Nature*, *396*(6710), 450–453. https://doi.org/10.1038/24839

Canfield, D. E., & Raiswell, R. (1999). *The Evolution of the Sulfur Cycle*.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Cardona, T., Murray, J. W., & Rutherford, A. W. (2015). Origin and Evolution of Water Oxidation before the Last Common Ancestor of the Cyanobacteria. *Molecular Biology and Evolution*, *32*(5), 1310–1328. https://doi.org/10.1093/molbev/msv024

Castendyk, D. N., Obryk, M. K., Leidman, S. Z., Gooseff, M., & Hawes, I. (2016). Lake Vanda: A sentinel for climate change in the McMurdo Sound Region of Antarctica. *Global and Planetary Change*, *144*, 213–227. https://doi.org/10.1016/j.gloplacha.2016.06.007

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2020). GTDB-Tk: A toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*, *36*(6), 1925–1927. https://doi.org/10.1093/bioinformatics/btz848

Chrismas, N. A. M., Anesio, A. M., & Sánchez-Baracaldo, P. (2015). Multiple adaptations to polar and alpine environments within cyanobacteria: A phylogenomic and Bayesian approach. *Frontiers in Microbiology*, *6*, 1070. https://doi.org/10.3389/fmicb.2015.01070

Chrismas, N. A. M., Anesio, A. M., & Sánchez-Baracaldo, P. (2018). The future of genomics in polar and alpine cyanobacteria. *FEMS Microbiology Ecology*, *94*(4). https://doi.org/10.1093/femsec/fiy032

Chrismas, N. A. M., Barker, G., Anesio, A. M., & Sánchez-Baracaldo, P. (2016). Genomic mechanisms for cold tolerance and production of exopolysaccharides in the Arctic cyanobacterium Phormidesmis priestleyi BC1401. *BMC Genomics*, *17*. https://doi.org/10.1186/s12864-016-2846-4

Chrismas, N. A. M., Williamson, C. J., Yallop, M. L., Anesio, A. M., & Sánchez-Baracaldo, P. (2018a). Photoecology of the Antarctic cyanobacterium Leptolyngbya sp. BC1307 brought to light through community analysis, comparative genomics and in vitro photophysiology. *Molecular Ecology*, *27*(24), 5279–5293. https://doi.org/10.1111/mec.14953

Chrismas, N. A. M., Williamson, C. J., Yallop, M. L., Anesio, A. M., & Sánchez-Baracaldo, P. (2018b). Photoecology of the Antarctic cyanobacterium Leptolyngbya sp. BC1307 brought to light

through community analysis, comparative genomics and in vitro photophysiology. *Molecular Ecology*, *27*(24), 5279–5293. https://doi.org/10.1111/mec.14953

Cohen, Y., Jørgensen, B. B., Padan, E., & Shilo, M. (1975). Sulphide-dependent anoxygenic photosynthesis in the cyanobacterium Oscillatoria limnetica. *Nature*, *257*(5526), 489. https://doi.org/10.1038/257489a0

Cohen, Y., Jørgensen, B. B., Revsbech, N. P., & Poplawski, R. (1986). Adaptation to Hydrogen Sulfide of Oxygenic and Anoxygenic Photosynthesis among Cyanobacteria. *Appl. Environ. Microbiol.*, *51*(2), 398–407.

Cohen, Y., Padan, E., & Shilo, M. (1975). Facultative anoxygenic photosynthesis in the cyanobacterium Oscillatoria limnetica. *Journal of Bacteriology*, *123*(3), 855–861.

Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., Cartwright, R., Charbonneau, A., Constantinides, B., Edvenson, G., Fay, S., Fenton, J., Fenzl, T., Fish, J., Garcia-Gutierrez, L., Garland, P., Gluck, J., González, I., Guermond, S., Guo, J., … Brown, C. T. (2015). The khmer software package: Enabling efficient nucleotide sequence analysis. *F1000Research*, *4*. https://doi.org/10.12688/f1000research.6924.1

Dahl, T. W., Canfield, D. E., Rosing, M. T., Frei, R. E., Gordon, G. W., Knoll, A. H., & Anbar, A. D. (2011). Molybdenum evidence for expansive sulfidic water masses in ~750Ma oceans. *Earth and Planetary Science Letters*, *311*(3), 264–274. https://doi.org/10.1016/j.epsl.2011.09.016

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Delmont, T. O., & Eren, A. M. (2018). Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ*, *6*, e4320. https://doi.org/10.7717/peerj.4320

Dick, G. J., Grim, S. L., & Klatt, J. M. (2018). Controls on O2 Production in Cyanobacterial Mats and Implications for Earth's Oxygenation. *Annual Review of Earth and Planetary Sciences*, *46*(1), null. https://doi.org/10.1146/annurev-earth-082517-010035

Dillon, M. L., Hawes, I., Jungblut, A. D., Mackey, T. J., Eisen, J. A., Doran, P. T., & Sumner, D. Y. (2020). Energetic and Environmental Constraints on the Community Structure of Benthic Microbial Mats in Lake Fryxell, Antarctica. *FEMS Microbiology Ecology*, *96*(2). https://doi.org/10.1093/femsec/fiz207

Doran, P., & Fountain, A. (2016). *McMurdo Dry Valleys Lake Vanda Meteorological Station Daily Averages* [Data set]. Environmental Data Initiative. https://doi.org/10.6073/PASTA/5BD6F1B72FB87ACCE25D2B56646964EB

Durán, R. V., Hervás, M., De la Rosa, M. A., & Navarro, J. A. (2004). The Efficient Functioning of Photosynthesis and Respiration in Synechocystis sp. PCC 6803 Strictly Requires the Presence of either Cytochrome c6 or Plastocyanin*. *Journal of Biological Chemistry*, *279*(8), 7229–7233. https://doi.org/10.1074/jbc.M311565200

Eddy, S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, *7*(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Edgar, R. C., Taylor, J., Lin, V., Altman, T., Barbera, P., Meleshko, D., Lohr, D., Novakovsky, G., Buchfink, B., Al-Shayeb, B., Banfield, J. F., de la Peña, M., Korobeynikov, A., Chikhi, R., & Babaian, A. (2022). Petabase-scale sequence alignment catalyses viral discovery. *Nature*, *602*(7895), 142–147. https://doi.org/10.1038/s41586-021-04332-2

Ehling-Schulz, M., & Scherer, S. (1999). UV protection in cyanobacteria. *European Journal of Phycology*, *34*(4), 329–338. https://doi.org/10.1080/09670269910001736392

Eren, A. M., Esen, Ö. C., Quince, C., Vineis, J. H., Morrison, H. G., Sogin, M. L., & Delmont, T. O. (2015). Anvi'o: An advanced analysis and visualization platform for 'omics data. *PeerJ*, *3*, e1319. https://doi.org/10.7717/peerj.1319

Falkowski, P. G., & Raven, J. A. (2013). *Aquatic Photosynthesis: Second Edition*. Princeton University Press.

Garcia-Pichel, F., López-Cortés, A., & Nübel, U. (2001). Phylogenetic and Morphological Diversity of Cyanobacteria in Soil Desert Crusts from the Colorado Plateau. *Applied and Environmental Microbiology*. https://doi.org/10.1128/AEM.67.4.1902-1910.2001

Garcia-Pichel, F., Prufert-Bebout, L., & Muyzer, G. (1996). Phenotypic and phylogenetic analyses show Microcoleus chthonoplastes to be a cosmopolitan cyanobacterium. *Applied and Environmental Microbiology*. https://journals.asm.org/doi/abs/10.1128/aem.62.9.3284-3291.1996

Garlick, S., Oren, A., & Padan, E. (1977). Occurrence of facultative anoxygenic photosynthesis among filamentous and unicellular cyanobacteria. *Journal of Bacteriology*, *129*(2), 623–629.

Gorbushina, A. A., Kort, R., Schulte, A., Lazarus, D., Schnetger, B., Brumsack, H.-J., Broughton, W. J., & Favet, J. (2007). Life in Darwin's dust: Intercontinental transport and survival of microbes in the nineteenth century. *Environmental Microbiology*, *9*(12), 2911–2922. https://doi.org/10.1111/j.1462-2920.2007.01461.x

Green, W. J., & Lyons, W. B. (2009). The Saline Lakes of the McMurdo Dry Valleys, Antarctica. *Aquatic Geochemistry*, *15*(1), 321–348. https://doi.org/10.1007/s10498-008-9052-1

Grettenberger, C. L., Sumner, D. Y., Wall, K., Brown, C. T., Eisen, J. A., Mackey, T. J., Hawes, I., Jospin, G., & Jungblut, A. D. (2020). A phylogenetically novel cyanobacterium most closely related to Gloeobacter. *The ISME Journal*, *14*(8), 2142–2152. https://doi.org/10.1038/s41396-020-0668-5

Griffin, D. W., Kellogg, C. A., Garrison, V. H., & Shinn, E. A. (2002). The Global Transport of Dust: An Intercontinental river of dust, microorganisms and toxic chemicals flows through the Earth's atmosphere. *American Scientist*, *90*(3), 228–235.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hagemann, M. (2011). Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiology Reviews*, *35*(1), 87–123. https://doi.org/10.1111/j.1574-6976.2010.00234.x

Hamilton, T. L., Klatt, J. M., de Beer, D., & Macalady, J. L. (2018). Cyanobacterial photosynthesis under sulfidic conditions: Insights from the isolate *Leptolyngbya* sp. strain hensonii. *The ISME Journal*, *12*(2), 568–584. https://doi.org/10.1038/ismej.2017.193

Harke, M. J., Steffen, M. M., Gobler, C. J., Otten, T. G., Wilhelm, S. W., Wood, S. A., & Paerl, H. W. (2016). A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae*, *54*, 4–20. https://doi.org/10.1016/j.hal.2015.12.007

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hawes, I., Giles, H., & Doran, P. T. (2014). Estimating photosynthetic activity in microbial mats in an ice-covered Antarctic lake using automated oxygen microelectrode profiling and variable chlorophyll fluorescence. *Limnology and Oceanography*, *59*(3), 674–688. https://doi.org/10.4319/lo.2014.59.3.0674

Hawes, I., Jungblut, A. D., Matys, E. D., & Summons, R. E. (2018). The "Dirty Ice" of the McMurdo Ice Shelf: Analogues for biological oases during the Cryogenian. *Geobiology*, *16*(4), 369–377. https://doi.org/10.1111/gbi.12280

Hawes, I., Moorhead, D., Sutherland, D., Schmeling, J., & Schwarz, A.-M. (2001). Benthic primary production in two perennially ice-covered Antarctic lakes: Patterns of biomass accumulation with a model of community metabolism. *Antarctic Science; Cambridge*, *13*(1), 18–27.

Hawes, I., & Schwarz, A.-M. J. (2001). Absorption and Utilization of Irradiance by Cyanobacterial Mats in Two Ice-Covered Antarctic Lakes with Contrasting Light Climates. *Journal of Phycology*, *37*(1), 5–15. https://doi.org/10.1046/j.1529-8817.1999.014012005.x

Hawes, I., Sumner, D. Y., Andersen, D. T., Jungblut, A. D., & Mackey, T. J. (2013). Timescales of Growth Response of Microbial Mats to Environmental Change in an Ice-Covered Antarctic Lake. *Biology*, *2*(1), 151–176. https://doi.org/10.3390/biology2010151

Hershkovitz, N., Oren, A., & Cohen, Y. (1991). Accumulation of Trehalose and Sucrose in Cyanobacteria Exposed to Matric Water Stress. *Applied and Environmental Microbiology*, *57*(3), 645–648.

Hillman, C. (2013). *Structure of benthic microbial mat assemblages in Lake Fryxell, Antarctica.* https://ir.canterbury.ac.nz/handle/10092/8737

Hincha, D. K., & Hagemann, M. (2004). Stabilization of model membranes during drying by compatible solutes involved in the stress tolerance of plants and microorganisms. *Biochemical Journal*, *383*(2), 277–283. https://doi.org/10.1042/BJ20040746

Hodgson, D. A., Verleyen, E., Sabbe, K., Squier, A. H., Keely, B. J., Leng, M. J., Saunders, K. M., & Vyverman, W. (2005). Late Quaternary climate-driven environmental change in the Larsemann Hills, East Antarctica, multi-proxy evidence from a lake sediment core. *Quaternary Research*, *64*(1), 83–99. https://doi.org/10.1016/j.yqres.2005.04.002

Hoffman, P. F., Abbot, D. S., Ashkenazy, Y., Benn, D. I., Brocks, J. J., Cohen, P. A., Cox, G. M., Creveling, J. R., Donnadieu, Y., Erwin, D. H., Fairchild, I. J., Ferreira, D., Goodman, J. C., Halverson, G. P., Jansen, M. F., Le Hir, G., Love, G. D., Macdonald, F. A., Maloof, A. C., … Warren, S. G. (2017). Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Science Advances*, *3*(11), e1600983. https://doi.org/10.1126/sciadv.1600983

Howard-Williams, C., Pridmore, R., Downes, M. T., & Vincent, W. F. (1989). Microbial biomass, photosynthesis and chlorophyll a related pigments in the ponds of the McMurdo Ice Shelf, Antarctica. *Antarctic Science*, *1*(2), 125–131. https://doi.org/10.1017/S0954102089000192

Howard-Williams, C., Schwarz, A.-M., Hawes, I., & Priscu, J. C. (1998). Optical Properties of the Mcmurdo Dry Valley Lakes, Antarctica. In *Ecosystem Dynamics in a Polar Desert: The Mcmurdo Dry Valleys, Antarctica* (pp. 189–203). American Geophysical Union (AGU). https://doi.org/10.1029/AR072p0189

Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., & Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*, *11*, 119. https://doi.org/10.1186/1471-2105-11-119

Irber, L. C. (2020a). Decentralizing Indices for Genomic Data [Ph.D., University of California, Davis]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/2503641751/abstract/7B8543548D284D81PQ/1

Irber, L. C. (2020b, July 24). MinHashing all the things: A quick analysis of MAG search results. *Gabbleblotchits*. https://blog.luizirber.org/2020/07/24/mag-results/

Irber, L. C., Brooks, P. T., Reiter, T. E., Pierce-Ward, N. T., Hera, M. R., Koslicki, D., & Brown, C. T. (2022). Lightweight compositional analysis of metagenomes with FracMinHash and minimum metagenome covers. *BioRxiv*, 2022.01.11.475838. https://doi.org/10.1101/2022.01.11.475838

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114. https://doi.org/10.1038/s41467-018-07641-9

Jr, R. A. W., Parker, B. C., & Jr, G. M. S. (1983). Distribution, species composition and morphology of algal mats in Antarctic dry valley lakes. *Phycologia*, *22*(4), 355–365. https://doi.org/10.2216/i0031-8884-22-4-355.1

Jungblut, A. D., Hawes, I., Mackey, T. J., Krusor, M., Doran, P. T., Sumner, D. Y., Eisen, J. A., Hillman, C., & Goroncy, A. K. (2016). Microbial Mat Communities along an Oxygen Gradient in a Perennially

Ice-Covered Antarctic Lake. *Applied and Environmental Microbiology*, *82*(2), 620–630. https://doi.org/10.1128/AEM.02699-15

Jungblut, A. D., Lovejoy, C., & Vincent, W. F. (2010). Global distribution of cyanobacterial ecotypes in the cold biosphere. *The ISME Journal*, *4*(2), 191–202. https://doi.org/10.1038/ismej.2009.113

Jungblut, A.-D., Hawes, I., Mountfort, D., Hitzfeld, B., Dietrich, D. R., Burns, B. P., & Neilan, B. A. (2005). Diversity within cyanobacterial mat communities in variable salinity meltwater ponds of McMurdo Ice Shelf, Antarctica. *Environmental Microbiology*, *7*(4), 519–529. https://doi.org/10.1111/j.1462-2920.2005.00717.x

Kanehisa, M. (2019). Toward understanding the origin and evolution of cellular organisms. *Protein Science: A Publication of the Protein Society*, *28*(11), 1947–1951. https://doi.org/10.1002/pro.3715

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, *28*(1), 27–30. https://doi.org/10.1093/nar/28.1.27

Kanehisa, M., Sato, Y., & Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, *428*(4), 726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165. https://doi.org/10.7717/peerj.1165

Kaspar, M., Simmons, G. M., Parker, B. C., Seaburg, K. G., Wharton, R. A., & Smith, R. I. L. (1982). Bryum Hedw. Collected from Lake Vanda, Antarctica. *The Bryologist*, *85*(4), 424–430. https://doi.org/10.2307/3242912

Kasting, J. F. (2013). What caused the rise of atmospheric O2? *Chemical Geology*, *362*, 13–25. https://doi.org/10.1016/j.chemgeo.2013.05.039

Kirsch, F., Klähn, S., & Hagemann, M. (2019). Salt-Regulated Accumulation of the Compatible Solutes Sucrose and Glucosylglycerol in Cyanobacteria and Its Biotechnological Potential. *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02139

Klatt, J. M., Al-Najjar, M. A. A., Yilmaz, P., Lavik, G., Beer, D. de, & Polerecky, L. (2015). Anoxygenic Photosynthesis Controls Oxygenic Photosynthesis in a Cyanobacterium from a Sulfidic Spring. *Applied and Environmental Microbiology*, *81*(6), 2025–2031. https://doi.org/10.1128/AEM.03579-14

Klatt, J. M., Haas, S., Yilmaz, P., de Beer, D., & Polerecky, L. (2015). Hydrogen sulfide can inhibit and enhance oxygenic photosynthesis in a cyanobacterium from sulfidic springs. *Environmental Microbiology*, *17*(9), 3301–3313. https://doi.org/10.1111/1462-2920.12791

Klatt, J. M., Meyer, S., Häusler, S., Macalady, J. L., de Beer, D., & Polerecky, L. (2016). Structure and function of natural sulphide-oxidizing microbial mats under dynamic input of light and chemical energy. *The ISME Journal*, *10*(4), 921–933. https://doi.org/10.1038/ismej.2015.167

Koebler, C., Schmelling, N. M., Pawlowski, A., Spaet, P., Scheurer, N. M., Berwanger, L., Macek, B., Axmann, I. M., & Wilde, A. (2021). A chimeric KaiA-like regulator extends the nonstandard KaiB3-KaiC3 clock system in bacteria. *BioRxiv*, 2021.07.20.453058. https://doi.org/10.1101/2021.07.20.453058

Komárek, J., & Anagnostidis, K. (2005). Cyanoprokaryota II. Teil Oscillatoriales. *SüBwasserflora von Mitteleuropa*, *19*, 1–759.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Lara, Y., Durieu, B., Cornet, L., Verlaine, O., Rippka, R., Pessi, I. S., Misztak, A., Joris, B., Javaux, E. J., Baurain, D., & Wilmotte, A. (2017). Draft Genome Sequence of the Axenic Strain

Phormidesmispriestleyi ULC007, a Cyanobacterium Isolated from Lake Bruehwiler (Larsemann Hills, Antarctica). *Genome Announcements*, *5*(7). https://doi.org/10.1128/genomeA.01546-16

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., & Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. https://doi.org/10.1093/bioinformatics/btm404

Lawrence, M. J. F., & Hendy, C. H. (1985). Water column and sediment characteristics of Lake Fryxell, Taylor Valley, Antarctica. *New Zealand Journal of Geology and Geophysics*, *28*(3), 543–552. https://doi.org/10.1080/00288306.1985.10421206

Laybourn-Parry, J., James, M. R., McKnight, D. M., Priscu, J., Spaulding, S. A., & Shiel, R. (1997). The microbial plankton of Lake Fryxell, southern Victoria Land, Antarctica during the summers of 1992 and 1994. *Polar Biology*, *17*(2), 54–61. https://doi.org/10.1007/s003000050104

Lea-Smith, D. J., Bombelli, P., Vasudevan, R., & Howe, C. J. (2016). Photosynthetic, respiratory and extracellular electron transport pathways in cyanobacteria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1857*(3), 247–255. https://doi.org/10.1016/j.bbabio.2015.10.007

Leinonen, R., Sugawara, H., Shumway, M., & on behalf of the International Nucleotide Sequence Database Collaboration. (2011). The Sequence Read Archive. *Nucleic Acids Research*, *39*(suppl_1), D19–D21. https://doi.org/10.1093/nar/gkq1019

Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, *31*(10), 1674–1676. https://doi.org/10.1093/bioinformatics/btv033

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lizotte, M. P., & Priscu, J. C. (1992). Photosynthesis-Irradiance Relationships in Phytoplankton from the Physically Stable Water Column of a Perennially Ice-Covered Lake (lake Bonney, Antarctica)1. *Journal of Phycology*, *28*(2), 179–185. https://doi.org/10.1111/j.0022-3646.1992.00179.x

Love, F. G., Jr, G. M. S., Parker, B. C., Jr, R. A. W., & Seaburg, K. G. (1983). Modern conophyton-like microbial mats discovered in Lake Vanda, Antarctica. *Geomicrobiology Journal*, *3*(1), 33–48. https://doi.org/10.1080/01490458309377782

Lumian, J. E., Jungblut, A. D., Dillion, M. L., Hawes, I., Doran, P. T., Mackey, T. J., Dick, G. J., Grettenberger, C. L., & Sumner, D. Y. (2021). Metabolic Capacity of the Antarctic Cyanobacterium Phormidium pseudopriestleyi That Sustains Oxygenic Photosynthesis in the Presence of Hydrogen Sulfide. *Genes*, *12*(3), 426. https://doi.org/10.3390/genes12030426

Lyons, W. B., Frape, S. K., & Welch, K. A. (1999). History of McMurdo Dry Valley lakes, Antarctica, from stable chlorine isotope data. *Geology*, *27*(6), 527–530. https://doi.org/10.1130/0091-7613(1999)027<0527:HOMDVL>2.3.CO;2

Macalady, J. L., Lyon, E. H., Koffman, B., Albertson, L. K., Meyer, K., Galdenzi, S., & Mariani, S. (2006). Dominant Microbial Populations in Limestone-Corroding Stream Biofilms, Frasassi Cave System, Italy. *Applied and Environmental Microbiology*, *72*(8), 5596–5609. https://doi.org/10.1128/AEM.00715-06

Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, *11*, 538. https://doi.org/10.1186/1471-2105-11-538

McCann, R. B., Lynch, J., & Adams, J. (2018). Mitigating Projected Impacts of Climate Change and Building Resiliency Through Permaculture. In *Addressing Climate Change at the Community Level in the United States*. Routledge.

McKnight, D. M., Niyogi, D. K., Alger, A. S., Bomblies, A., Conovitz, P. A., & Tate, C. M. (1999). Dry Valley Streams in Antarctica: Ecosystems Waiting for Water. *BioScience*, *49*(12), 985–995. https://doi.org/10.1525/bisi.1999.49.12.985

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. https://doi.org/10.1109/GCE.2010.5676129

Miller, S. R., & Bebout, B. M. (2004). Variation in Sulfide Tolerance of Photosystem II in Phylogenetically Diverse Cyanobacteria from Sulfidic Habitats. *Applied and Environmental Microbiology*, *70*(2), 736–744. https://doi.org/10.1128/AEM.70.2.736-744.2004

Mills, L. A., McCormick, A. J., & Lea-Smith, D. J. (2020). Current knowledge and recent advances in understanding metabolism of the model cyanobacterium Synechocystis sp. PCC 6803. *Bioscience Reports*, *40*(4). https://doi.org/10.1042/BSR20193325

Miner, G. (2006). Standard Methods for the Examination of Water and Wastewater, 21st Edition. *American Water Works Association. Journal*, *98*(1), 130.

Mora, S. J. D., Whitehead, R. F., & Gregory, M. (1991). Aqueous geochemistry of major constituents in the Alph River and tributaries in Walcott Bay, Victoria Land, Antarctica. *Antarctic Science*, *3*(1), 73–86. https://doi.org/10.1017/S0954102091000111

Moreira, C., Vasconcelos, V., & Antunes, A. (2013). Phylogeny and Biogeography of Cyanobacteria and Their Produced Toxins. *Marine Drugs*, *11*(11), 4350–4369. https://doi.org/10.3390/md11114350

Mullineaux, C. W. (2014). Electron transport and light-harvesting switches in cyanobacteria. *Frontiers in Plant Science*, *5*. https://doi.org/10.3389/fpls.2014.00007

Mulo, P., Sakurai, I., & Aro, E.-M. (2012). Strategies for psbA gene expression in cyanobacteria, green algae and higher plants: From transcription to PSII repair. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1817*(1), 247–257. https://doi.org/10.1016/j.bbabio.2011.04.011

Mulo, P., Sicora, C., & Aro, E.-M. (2009). Cyanobacterial psbA gene family: Optimization of oxygenic photosynthesis. *Cellular and Molecular Life Sciences*, *66*(23), 3697. https://doi.org/10.1007/s00018-009-0103-6

Murray, J. W. (2012). Sequence variation at the oxygen-evolving centre of photosystem II: A new class of 'rogue' cyanobacterial D1 proteins. *Photosynthesis Research*, *110*(3), 177–184. https://doi.org/10.1007/s11120-011-9714-5

Nagy, C. I., Vass, I., Rákhely, G., Vass, I. Z., Tóth, A., Duzs, Á., Peca, L., Kruk, J., & Kós, P. B. (2014). Coregulated Genes Link Sulfide:Quinone Oxidoreductase and Arsenic Metabolism in Synechocystis sp. Strain PCC6803. *Journal of Bacteriology*, *196*(19), 3430–3440. https://doi.org/10.1128/JB.01864-14

Nakajima, M., Imai, K., Ito, H., Nishiwaki, T., Murayama, Y., Iwasaki, H., Oyama, T., & Kondo, T. (2005). Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation in vitro. *Science (New York, N.Y.)*, *308*(5720), 414–415. https://doi.org/10.1126/science.1108451

Namsaraev, Z., Mano, M.-J., Fernandez, R., & Wilmotte, A. (2010). Biogeography of terrestrial cyanobacteria from Antarctic ice-free areas. *Annals of Glaciology*, *51*(56), 171–177. https://doi.org/10.3189/172756411795931930

Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., & Phillippy, A. M. (2016). Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biology*, *17*(1), 132. https://doi.org/10.1186/s13059-016-0997-x

Oren, A., Padan, E., & Malkin, S. (1979). Sulfide inhibition of photosystem II in cyanobacteria (blue-green algae) and tobacco chloroplasts. *Biochimica Et Biophysica Acta*, *546*(2), 270–279.

Page-Sharp, M., Behm, C. A., & Smith, G. D. (1999). Involvement of the compatible solutes trehalose and sucrose in the response to salt stress of a cyanobacterial Scytonema species isolated from desert soils. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1472*(3), 519–528. https://doi.org/10.1016/S0304-4165(99)00155-5

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. https://doi.org/10.1101/gr.186072.114

Parnell, J., & Boyce, A. J. (2017). Microbial sulphate reduction during Neoproterozoic glaciation, Port Askaig Formation, UK. *Journal of the Geological Society*, *174*, 850–854. https://doi.org/10.1144/jgs2016-147

Pierce, N. T., Irber, L., Reiter, T., Brooks, P., & Brown, C. T. (2019). *Large-scale sequence comparisons with* sourmash (8:1006). F1000Research. https://doi.org/10.12688/f1000research.19675.1

Pierce-Ward, N. T. (2022, February 7). *Personal Communication, in prep*.

Potts, M., Slaughter, S. M., Hunneke, F.-U., Garst, J. F., & Helm, R. F. (2005). Desiccation Tolerance of Prokaryotes: Application of Principles to Human Cells. *Integrative and Comparative Biology*, *45*(5), 800–809. https://doi.org/10.1093/icb/45.5.800

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE*, *5*(3). https://doi.org/10.1371/journal.pone.0009490

Priscu, J. C. (1995). Phytoplankton nutrient deficiency in lakes of the McMurdo dry valleys, Antarctica. *Freshwater Biology*, *34*(2), 215–227. https://doi.org/10.1111/j.1365-2427.1995.tb00882.x

Quesada, A., & Vincent, W. F. (2012). Cyanobacteria in the Cryosphere: Snow, Ice and Extreme Cold. In B. A. Whitton (Ed.), *Ecology of Cyanobacteria II: Their Diversity in Space and Time* (pp. 387–399). Springer Netherlands. https://doi.org/10.1007/978-94-007-3855-3_14

Ramos, A. R., Grein, F., Oliveira, G. P., Venceslau, S. S., Keller, K. L., Wall, J. D., & Pereira, I. A. C. (2015). The FlxABCD-HdrABC proteins correspond to a novel NADH dehydrogenase/heterodisulfide reductase widespread in anaerobic bacteria and involved in ethanol metabolism in Desulfovibrio vulgaris Hildenborough. *Environmental Microbiology*, *17*(7), 2288–2305. https://doi.org/10.1111/1462-2920.12689

Ribeiro, K. F., Duarte, L., & Crossetti, L. O. (2018). Everything is not everywhere: A tale on the biogeography of cyanobacteria. *Hydrobiologia*, *820*(1), 23–48. https://doi.org/10.1007/s10750-018-3669-x

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M., & Stanier, R. Y. (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Microbiology, 111*(1), 1–61. https://doi.org/10.1099/00221287-111-1-1

Roberts, E. C., Laybourn-Parry, J., McKnight, D. M., & Novarino, G. (2000). Stratification and dynamics of microbial loop communities in Lake Fryxell, Antarctica. *Freshwater Biology*, *44*(4), 649–661. https://doi.org/10.1046/j.1365-2427.2000.00612.x

Rodrigues, D. F., & Tiedje, J. M. (2008). Coping with Our Cold Planet. *Applied and Environmental Microbiology*, *74*(6), 1677–1686. https://doi.org/10.1128/AEM.02000-07

Rodriguez-R, L. M., & Konstantinidis, K. T. (2016). *The enveomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes* (e1900v1). PeerJ Inc. https://doi.org/10.7287/peerj.preprints.1900v1

Roos, J. C., & Vincent, W. F. (1998). Temperature Dependence of Uv Radiation Effects on Antarctic Cyanobacteria. *Journal of Phycology*, *34*(1), 118–125. https://doi.org/10.1046/j.1529-8817.1998.340118.x

Shahak, Y., & Hauska, G. (2008). Sulfide Oxidation from Cyanobacteria to Humans: Sulfide–Quinone Oxidoreductase (SQR). In *Sulfur Metabolism in Phototrophic Organisms* (pp. 319–335). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-6863-8_16

Shi, T., & Falkowski, P. G. (2008). Genome evolution in cyanobacteria: The stable core and the variable shell. *Proceedings of the National Academy of Sciences*, *105*(7), 2510–2515. https://doi.org/10.1073/pnas.0711165105

Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., … Kerfeld, C. A. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 1053–1058. https://doi.org/10.1073/pnas.1217107110

Sorrels, C. M., Proteau, P. J., & Gerwick, W. H. (2009). Organization, Evolution, and Expression Analysis of the Biosynthetic Gene Cluster for Scytonemin, a Cyanobacterial UV-Absorbing Pigment. *Applied and Environmental Microbiology*, *75*(14), 4861–4869. https://doi.org/10.1128/AEM.02508-08

Spigel, R. H., & Priscu, J. C. (2013). Physical Limnology of the Mcmurdo Dry Valleys Lakes. In *Ecosystem Dynamics in a Polar Desert: The Mcmurdo Dry Valleys, Antarctica* (pp. 153–187). American Geophysical Union (AGU). https://doi.org/10.1029/AR072p0153

Stal, L. J. (2007). Cyanobacteria. In J. Seckbach (Ed.), *Algae and Cyanobacteria in Extreme Environments* (pp. 659–680). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6112-7_36

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Sukumaran, J., & Holder, M. T. (2010). DendroPy: A Python library for phylogenetic computing. *Bioinformatics*, *26*(12), 1569–1571. https://doi.org/10.1093/bioinformatics/btq228

Summerfield, T. C., Toepel, J., & Sherman, L. A. (2008). Low-Oxygen Induction of Normally Cryptic psbA Genes in Cyanobacteria. *Biochemistry*, *47*(49), 12939–12941. https://doi.org/10.1021/bi8018916

Sumner, D. Y., Hawes, I., Mackey, T. J., Jungblut, A. D., & Doran, P. T. (2015). Antarctic microbial mats: A modern analog for Archean lacustrine oxygen oases. *Geology*, *43*(10), 887–890. https://doi.org/10.1130/G36966.1

Sumner, D. Y., Jungblut, A. D., Hawes, I., Andersen, D. T., Mackey, T. J., & Wall, K. (2016). Growth of elaborate microbial pinnacles in Lake Vanda, Antarctica. *Geobiology*, *14*(6), 556–574. https://doi.org/10.1111/gbi.12188

Swan, J. A., Golden, S. S., LiWang, A., & Partch, C. L. (2018). Structure, function, and mechanism of the core circadian clock in cyanobacteria. *Journal of Biological Chemistry*, *293*(14), 5026–5034. https://doi.org/10.1074/jbc.TM117.001433

Taboada, B., Estrada, K., Ciria, R., & Merino, E. (2018). Operon-mapper: A web server for precise operon identification in bacterial and archaeal genomes. *Bioinformatics*, *34*(23), 4118–4120. https://doi.org/10.1093/bioinformatics/bty496

Tang, E. P. Y., Tremblay, R., & Vincent, W. F. (1997). Cyanobacterial Dominance of Polar Freshwater Ecosystems: Are High-Latitude Mat-Formers Adapted to Low Temperature?1. *Journal of Phycology*, *33*(2), 171–181. https://doi.org/10.1111/j.0022-3646.1997.00171.x

Tang, J., Du, L.-M., Liang, Y.-M., & Daroch, M. (2019). Complete Genome Sequence and Comparative Analysis of Synechococcus sp. CS-601 (SynAce01), a Cold-Adapted Cyanobacterium from an Oligotrophic Antarctic Habitat. *International Journal of Molecular Sciences*, *20*(1), 152. https://doi.org/10.3390/ijms20010152

Taton, A., Grubisic, S., Balthasart, P., Hodgson, D. A., Laybourn-Parry, J., & Wilmotte, A. (2006). Biogeographical distribution and ecological ranges of benthic cyanobacteria in East Antarctic lakes. *FEMS Microbiology Ecology*, *57*(2), 272–289. https://doi.org/10.1111/j.1574-6941.2006.00110.x

Turnage, W. V., & Hinckley, A. L. (1938). Freezing Weather in Relation to Plant Distribution in the Sonoran Desert. *Ecological Monographs*, *8*(4), 529–550. https://doi.org/10.2307/1943083

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., & Corbeil, J. (2010). Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems. *Limnology and Oceanography*, *55*(5), 1901–1911. https://doi.org/10.4319/lo.2010.55.5.1901

Varin, T., Lovejoy, C., Jungblut, A. D., Vincent, W. F., & Corbeil, J. (2012). Metagenomic Analysis of Stress Genes in Microbial Mat Communities from Antarctica and the High Arctic. *Applied and Environmental Microbiology*, *78*(2), 549–559. https://doi.org/10.1128/AEM.06354-11

Vermaas. (2001). Photosynthesis and Respiration in Cyanobacteria. In *ELS*. American Cancer Society. https://doi.org/10.1038/npg.els.0001670

Vermaas, W., Charité, J., & Shen, G. Z. (1990). Glu-69 of the D2 protein in photosystem II is a potential ligand to Mn involved in photosynthetic oxygen evolution. *Biochemistry*, *29*(22), 5325–5332. https://doi.org/10.1021/bi00474a017

Vincent, W. F. (1981). Production Strategies in Antarctic Inland Waters: Phytoplankton Eco-Physiology in a Permanently Ice-Covered Lake. *Ecology*, *62*(5), 1215–1224. https://doi.org/10.2307/1937286

Voorhies, A. A., Biddanda, B. A., Kendall, S. T., Jain, S., Marcus, D. N., Nold, S. C., Sheldon, N. D., & Dick, G. J. (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*(3), 250–267. https://doi.org/10.1111/j.1472-4669.2012.00322.x

Vopel. (2006). *Photosynthetic performance of benthic microbial mats in Lake Hoare, Antarctica*. https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.2006.51.4.1801

Wada, H., Combos, Z., & Murata, N. (1990). Enhancement of chilling tolerance of a cyanobacterium by genetic manipulation of fatty acid desaturation. *Nature*, *347*(6289), 200–203. https://doi.org/10.1038/347200a0

Wada, N., Sakamoto, T., & Matsugo, S. (2013). Multiple Roles of Photosynthetic and Sunscreen Pigments in Cyanobacteria Focusing on the Oxidative Stress. *Metabolites*, *3*(2), 463–483. https://doi.org/10.3390/metabo3020463

Wang, P., Algeo, T. J., Zhou, Q., Yu, W., Du, Y., Qin, Y., Xu, Y., Yuan, L., & Pan, W. (2019). Large accumulations of 34S-enriched pyrite in a low-sulfate marine basin: The Sturtian Nanhua Basin, South China. *Precambrian Research*, *335*, 105504. https://doi.org/10.1016/j.precamres.2019.105504

Watzer, B., & Forchhammer, K. (2018). Cyanophycin Synthesis Optimizes Nitrogen Utilization in the Unicellular Cyanobacterium Synechocystis sp. Strain PCC 6803. *Appl. Environ. Microbiol.*, *84*(20), e01298-18. https://doi.org/10.1128/AEM.01298-18

Wharton, R. A. (1994). *Sediment oxygen profiles in a super-oxygenated antarctic lake*. https://aslopubs.onlinelibrary.wiley.com/doi/abs/10.4319/lo.1994.39.4.0839

Wiegard, A., Köbler, C., Oyama, K., Dörrich, A. K., Azai, C., Terauchi, K., Wilde, A., & Axmann, I. M. (2020). Synechocystis KaiC3 Displays Temperature- and KaiB-Dependent ATPase Activity and Is Important for Growth in Darkness. *Journal of Bacteriology*, *202*(4), e00478-19. https://doi.org/10.1128/JB.00478-19

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*(5), 614–620. https://doi.org/10.1093/bioinformatics/btt593

Zhang, L., Jungblut, A. D., Hawes, I., Andersen, D. T., Sumner, D. Y., & Mackey, T. J. (2015). Cyanobacterial diversity in benthic mats of the McMurdo Dry Valley lakes, Antarctica. *Polar Biology*, *38*(8), 1097–1110. https://doi.org/10.1007/s00300-015-1669-0

Zhu, Y., Graham, J. E., Ludwig, M., Xiong, W., Alvey, R. M., Shen, G., & Bryant, D. A. (2010). Roles of xanthophyll carotenoids in protection against photoinhibition and oxidative stress in the

cyanobacterium Synechococcus sp. Strain PCC 7002. *Archives of Biochemistry and Biophysics*, *504*(1), 86–99. https://doi.org/10.1016/j.abb.2010.07.007

APPENDIX

Supplemental Table S3.1.1 Extended Matches from sourmash MAGsearch *Neosynechococcus* MAG Hits >5% Containment

MATCHES	CONTAINMENT	LOCATION
SRR5208701	97.82%	Lake Fryxell liftoff and glacier meltwater
SRR5468149	97.16%	Lake Fryxell liftoff and glacier meltwater
SRR5208700	81.58%	Lake Fryxell liftoff and glacier meltwater
SRR5208699	74.52%	Lake Fryxell liftoff and glacier meltwater
SRR5468150	65.46%	Lake Fryxell liftoff and glacier meltwater
SRR5468153	36.51%	Lake Fryxell liftoff and glacier meltwater
SRR5198900	5.71%	Sacramento Delta
SRR490140	5.31%	Amazon Forest
SRR2657237	5.19%	Barataria Bay
SRR490130	5.19%	Amazon Forest
SRR2657229	5.08%	Barataria Bay
SRR490139	5.08%	Barataria Bay
SRR490127	5.02%	Barataria Bay

Supplemental Table S3.1.2 Extended Matches from sourmash MAGsearch *Leptolyngbya* MAG Hits >5% Containment

MATCHES	CONTAINMENT	LOCATION
SRR5468150	98.72%	Lake Fryxell liftoff and glacier meltwater
SRR5468149	98.58%	Lake Fryxell liftoff and glacier meltwater
SRR5208699	98.35%	Lake Fryxell liftoff and glacier meltwater
SRR5208701	97.98%	Lake Fryxell liftoff and glacier meltwater
SRR5208700	97.70%	Lake Fryxell liftoff and glacier meltwater
SRR5468153	25.16%	Lake Fryxell liftoff and glacier meltwater
SRR6683740	8.40%	Arctic Lake Metagenome
SRR5198900	6.56%	Sacramento Delta

Supplemental Table S3.1.3 Extended Matches from sourmash MAGsearch *Pseudanabaena* MAG Hits >5% Containment

MATCHES	CONTAINMENT	LOCATION
SRR5468149	99.49%	Lake Fryxell liftoff and glacier meltwater
SRR5468150	99.46%	Lake Fryxell liftoff and glacier meltwater
SRR5208701	99.20%	Lake Fryxell liftoff and glacier meltwater
SRR5468153	99.13%	Lake Fryxell liftoff and glacier meltwater
SRR5208700	98.91%	Lake Fryxell liftoff and glacier meltwater
SRR5208699	98.12%	Lake Fryxell liftoff and glacier meltwater
SRR6266338	37.45%	Polar Desert Sand Communities
SRR5829599	33.54%	Nunavut, Canada
SRR5215118	30.57%	Nunavut, Canada
SRR5829597	26.16%	Nunavut, Canada
ERR4192538	18.31%	Deception Island, Antarctica (Whaler's Bay Sediment)
ERR4192539	16.43%	Deception Island, Antarctica (Whaler's Bay Sediment)
SRR7769784	7.45%	Antarctic Microbial Mat
SRR7769706	6.84%	Antarctic Microbial Mat
SRR2657229	6.40%	Barataria Bay
SRR5198900	6.01%	Sacramento, Delta
SRR2657549	5.72%	Barataria Bay
SRR5198902	5.64%	Barataria Bay
SRR2657237	5.61%	Barataria Bay
SRR7769810	5.46%	Antarctic Microbial Mat
SRR8842248	5.39%	Cryoconite from Svalbard
SRR7769748	5.21%	Antarctic Microbial Mat
SRR2657347	5.14%	Barataria Bay

Supplemental Table S3.1.4 Extended Matches from sourmash MAGsearch *Phormidium* MAG Hits >5% Containment

MATCHES	CONTAINMENT	LOCATION
SRR7769578	99.39%	Antarctic Microbial Mat
SRR7769621	99.32%	Antarctic Microbial Mat
SRR7769581	98.49%	Antarctic Microbial Mat
SRR7769747	98.49%	Antarctic Microbial Mat
SRR7769746	98.44%	Antarctic Microbial Mat
SRR7769635	98.25%	Antarctic Microbial Mat
SRR7769622	98.18%	Antarctic Microbial Mat
SRR7769634	98.11%	Antarctic Microbial Mat
SRR7769582	98.07%	Antarctic Microbial Mat
SRR7769583	97.79%	Antarctic Microbial Mat
SRR7769576	97.74%	Antarctic Microbial Mat
SRR7769579	97.60%	Antarctic Microbial Mat
SRR7769748	97.15%	Antarctic Microbial Mat
SRR7769793	96.56%	Antarctic Microbial Mat
SRR7769792	95.86%	Antarctic Microbial Mat

SRR7769639	94.70%	Antarctic Microbial Mat
SRR7769754	93.97%	Antarctic Microbial Mat
SRR7769518	92.65%	Antarctic Microbial Mat
SRR7769683	92.01%	Antarctic Microbial Mat
SRR7769753	91.90%	Antarctic Microbial Mat
SRR7769616	91.83%	Antarctic Microbial Mat
SRR7769558	91.00%	Antarctic Microbial Mat
SRR7769636	90.18%	Antarctic Microbial Mat
SRR7769554	88.88%	Antarctic Microbial Mat
SRR7769620	85.02%	Antarctic Microbial Mat
SRR7769624	83.72%	Antarctic Microbial Mat
SRR7769514	82.76%	Antarctic Microbial Mat
SRR7769557	81.55%	Antarctic Microbial Mat
SRR7769751	80.54%	Antarctic Microbial Mat
SRR7769794	77.45%	Antarctic Microbial Mat
SRR7769755	61.13%	Antarctic Microbial Mat
SRR7528444	55.79%	Ace Lake Saline
SRR7769788	54.53%	Antarctic Microbial Mat
SRR7529760	47.77%	Ace Lake Saline
SRR7769790	42.68%	Antarctic Microbial Mat
SRR7769643	39.13%	Antarctic Microbial Mat
SRR7769519	29.52%	Antarctic Microbial Mat
SRR7769812	25.39%	Antarctic Microbial Mat
SRR5216658	23.54%	Rauer Islands, Antarctica (saline)
SRR7769513	22.92%	Antarctic Microbial Mat
SRR6129595	22.38%	Rauer Islands, Antarctica (saline)
SRR7769618	22.10%	Antarctic Microbial Mat
SRR6185695	21.16%	Rauer Islands, Antarctica (saline)
SRR7529754	20.83%	Ace Lake Saline
SRR7428116	20.63%	Brackish Lagoon (SL)
SRR7428117	20.33%	Brackish Lagoon (SL)
SRR7428114	20.28%	Brackish Lagoon (SL)
SRR7769787	19.27%	Antarctic Microbial Mat
SRR7428121	19.17%	Brackish Lagoon (SL)
SRR7428120	19.10%	Brackish Lagoon (SL)
SRR12522841	19.04%	Big Soda Lake, Nevada
SRR7428115	18.87%	Brackish Lagoon (SL)
SRR7769528	18.47%	Antarctic Microbial Mat
SRR7428132	18.25%	Brackish Lagoon (EBD)
SRR12522839	17.71%	Big Soda Lake, Nevada
SRR7769662	16.79%	Antarctic Microbial Mat
SRR7769657	16.75%	Antarctic Microbial Mat
SRR7769512	16.66%	Antarctic Microbial Mat
SRR12522840	16.52%	Big Soda Lake, Nevada
SRR7769693	16.11%	Antarctic Microbial Mat
SRR7529732	15.95%	Ace Lake Saline
SRR7769584	15.20%	Antarctic Microbial Mat
SRR7428131	14.98%	Brackish Lagoon (EBD)
SRR7769531	14.14%	Antarctic Microbial Mat
SRR7769802	14.11%	Antarctic Microbial Mat

SRR7769623	13.08%	Antarctic Microbial Mat
SRR7769574	13.01%	Antarctic Microbial Mat
SRR7769801	12.98%	Antarctic Microbial Mat
SRR7769678	12.70%	Antarctic Microbial Mat
SRR7769659	12.16%	Antarctic Microbial Mat
ERR3503286	11.99%	Nairobi, Kenya (sewage, antimicrobial resistance)
SRR7769811	11.76%	Antarctic Microbial Mat
SRR7529753	11.73%	Ace Lake Saline
SRR7769810	11.28%	Antarctic Microbial Mat
SRR7769559	11.26%	Antarctic Microbial Mat
SRR7769735	11.14%	Antarctic Microbial Mat
SRR7769804	10.96%	Antarctic Microbial Mat
SRR7769696	10.95%	Antarctic Microbial Mat
SRR7769515	10.89%	Antarctic Microbial Mat
SRR7769776	10.50%	Antarctic Microbial Mat
SRR9691033	10.37%	Yanghu, China (wetland soil)
SRR7769805	10.17%	Antarctic Microbial Mat
SRR7769660	10.11%	Antarctic Microbial Mat
SRR7769650	10.03%	Antarctic Microbial Mat
SRR7769752	10.03%	Antarctic Microbial Mat
SRR7769808	10.01%	Antarctic Microbial Mat
SRR7769734	9.47%	Antarctic Microbial Mat
SRR7769773	9.42%	Antarctic Microbial Mat
SRR7769575	9.07%	Antarctic Microbial Mat
SRR10186387	8.98%	Salar de Huasco, Chile (sediment)
ERR3503282	8.83%	Nairobi, Kenya (sewage, antimicrobial resistance)
SRR7769615	8.69%	Antarctic Microbial Mat
SRR7769580	8.64%	Antarctic Microbial Mat
SRR7769695	8.62%	Antarctic Microbial Mat
SRR7769775	8.51%	Antarctic Microbial Mat
SRR12522836	8.48%	Big Soda Lake, Nevada
ERR738546	8.48%	Simulated Metagenome
SRR7769553	8.41%	Antarctic Microbial Mat
SRR7769642	8.36%	Antarctic Microbial Mat
ERR738544	8.10%	Simulated Metagenome
ERR738545	7.87%	Simulated Metagenome
SRR7769664	7.87%	Antarctic Microbial Mat
SRR6262267	7.61%	Human Gut
SRR7428125	7.30%	Brackish Lagoon (EBD)
SRR7769736	7.05%	Antarctic Microbial Mat
SRR7769570	6.95%	Antarctic Microbial Mat
SRR7769803	6.48%	Antarctic Microbial Mat
SRR5198900	6.38%	Sacramento Delta
SRR11412982	6.31%	Human Gut

Supplemental Table S3.1.5 Extended Matches from sourmash MAGsearch *Microcoleus* MAG Hits >30% Containment

MATCHES	CONTAINMENT	LOCATION
SRR5468150	99.18%	Lake Fryxell liftoff and glacier meltwater
SRR5468153	99.18%	Lake Fryxell liftoff and glacier meltwater
SRR5208700	98.66%	Lake Fryxell liftoff and glacier meltwater
SRR5468149	98.58%	Lake Fryxell liftoff and glacier meltwater
SRR5208699	86.19%	Lake Fryxell liftoff and glacier meltwater
SRR5208701	84.99%	Lake Fryxell liftoff and glacier meltwater
SRR6266358	65.02%	Polar Desert Sand Communities
SRR5855414	57.50%	Moab Soil Crust
SRR5855413	54.76%	Moab Soil Crust
SRR5855418	53.50%	Moab Soil Crust
SRR5855417	52.39%	Moab Soil Crust
SRR5855428	52.34%	Moab Soil Crust
SRR5855424	48.82%	Moab Soil Crust
SRR5855412	47.99%	Moab Soil Crust
SRR5855429	43.55%	Moab Soil Crust
SRR5855432	42.39%	Moab Soil Crust
SRR2952554	41.65%	Ningxia, China (soil crust)
SRR2954705	41.24%	Ningxia, China (soil crust)
SRR5247052	41.10%	Sonoran Desert
ERR3588763	40.61%	UK Pig Farm
SRR5855420	40.52%	Moab Soil Crust
SRR3439671	40.10%	Ningxia, China (soil crust)
SRR5830676	39.54%	Polar Desert Sand Communities
SRR5891573	39.54%	Glacial Snow, China
ERR1333181	38.36%	Mine Tailing Pool in China
SRR5459769	37.04%	Wastewater, Wisconsin
SRR6048908	36.30%	Puca Glacier, Peru
SRR12473531	35.71%	Mediterranean Desert Community
SRR12473532	35.68%	Mediterranean Desert Community
SRR5855438	34.78%	Moab Soil Crust
SRR12473534	34.30%	Mediterranean Desert Community
ERR3192241	33.57%	Southwest Germany (Arabidopsis community)