Analysis of Cyanobacteria and its Abundance in the Turkey Lakes Watershed using Molecular Techniques

by

Anjali Krishna

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Toxic cyanobacterial blooms continue to pose a threat to the quality and safety of drinking water globally by producing toxins and forming dense surface blooms. Forested watersheds naturally provide high quality drinking water to various communities but are threatened by bloom events that are increasing due to warming climates and anthropogenic land use. Monitoring programs utilized in drinking water sources are required to adapt to the changing intensity and frequency of these blooms where observation of cyanobacterial composition and abundance may vary based on sampling efforts. However, due to the variation and adaptability of these organisms spatially and temporally, cyanobacteria are often overlooked if surface blooms are not visualized, where these organisms may be present and abundant throughout the water column at different depths and vary throughout the day. The undetected organisms may release potent toxins that are threats to drinking water security if left untreated. The harmful toxic blooms comprise of cyclic hepatotoxins, involved in causing severe liver damage and affecting human and aquatic health. The aim of this study was to identify and quantify the cyanobacterial community composition and abundance of potential toxin producing genes in an oligotrophic northern forested watershed (Turkey Lakes Watershed, Ontario, Canada).

To evaluate the composition and abundance, water samples were collected from Little Turkey Lake in May, June, July, and August 2022 at integrated and varying depths to determine variability over a summer season and at different timepoints in a single day. Microbial DNA was extracted from the water samples for 16S rRNA gene sequencing where data was obtained for bioinformatic and phylogenetic analyses. Extracted samples underwent quantitative PCR analysis for identification of gene copy numbers of cyanobacteria and potential microcystin producing organisms. With the extension of the ice-free season through warmer temperatures, and changes in environmental parameters, cyanobacteria and potential cyanotoxin producers appeared as early as May in this oligotrophic lake system. Peak abundances of cyanobacterial and potential cyanotoxin producing gene copy numbers were observed in the months of July and August, without visible blooms during sampling.

Cyanobacterial composition had variability between the months, days, and timepoints when sampling, demonstrating the importance of consisting monitoring and sampling efforts due to the changing composition and abundance. Variation was observed among the depths within the water column, where integrated sampling provided a snapshot of the water system and can be useful for efficient analysis of the system, but multiple depth sampling is more representative of the community composition and abundance of cyanobacteria. This illustrates that monitoring protocols for drinking water sources require evaluation for the appropriate sampling protocol, timepoints, and location of the water column as each water system is unique.

This research provides insight into cyanobacterial emergence in earlier summer months in an oligotrophic water system. It is applicable for the development of monitoring and drinking water treatment protocols for toxin-producing cyanobacteria, where analysis of the full water column is required with consistent sampling and integrated sampling is efficient, especially when there is an absence of a visible surface bloom. The inclusion of molecular characterization (amplicon sequencing and qPCR) is a valuable tool that can be cost efficient and effective ways to analyze samples. This research can then be expanded to other toxins and secondary metabolites produced by cyanobacteria.

Acknowledgements

I would first like to acknowledge the funding for this research was provided by the NSERC Network for Forested Drinking Water Source Protection Technologies (*for*Water Network).

I would like to thank my supervisor, Dr. Kirsten Müller, for presenting me with this opportunity and your support throughout this project. You have been a wonderful role model and have provided me with guidance to push through any obstacles that came my way during these past few years. I couldn't have done it without your constant encouragement and feedback. I would also like to thank my committee members, Dr. Monica Emelko and Dr. Jozef Nissimov, for your support throughout this study.

I would like to thank Kara Webster and Jamie Broad in Sault Ste. Marie, Ontario, for your help at the Turkey Lakes Watershed and ensuring my sampling went smoothly. A special thank you to my wonderful team who made conducting field work such an amazing experience. I appreciate all your help and thank you for the memories we made along the way (Harshina Brijlall, Timothy Shardlow, Hadeel Abed, Isaac Meza Padilla, Emmanuelle Roy, and Jessica Panchaud).

I would like to express my gratitude to the Müller labs members, past and present, I would not have been able to achieve this without them. Thank you, Dr. Ellen Cameron, for paving the way and building the foundation for research at Turkey Lakes Watershed and for your guidance throughout this project. To Harshina Brijlall, I am so thankful we started our degrees at the same time, organized field work together, and became such close friends through it all. I couldn't have asked for a better field sampling partner, so thank you so much. Thank you, Timothy Shardlow, for your help with bioinformatics and answering any questions I had. Thank you to both Hadeel Abed and Anna Valero, I am so grateful for your friendships and am very glad I was able to meet you throughout this degree.

Thank you to my closest friend Brianna McDonnell for always being there and I am forever grateful for our friendship. Thank you to all my friends for their constant encouragement and reassurance throughout my degree.

I would like to thank my parents for their constant support throughout my life. I would not have been able to do this without you and I appreciate everything you have done for me.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AMT	Aminotransferase domain
ASV	Amplicon Sequence Variant
BLAST	Basic Local Alignment Search Tool
bp	Base pairs (nucleotides)
DADA2	Divisive Amplicon Denoising Algorithm 2
DNA	Deoxyribonucleic acid
HAB	Harmful Algal Bloom
kb	Kilobase (1000 base pairs of DNA)
***mcy*E**	Gene encoding non-ribsomonal peptide and polyketide synthase fusion protein
MEGA	Molecular Evolutionary Genetic Analysis
mirlyn	Multiple Iterations of Rarefaction for Library Normalization
MSA	Multiple Sequence Alignment
N	Nitrogen
NCBI	National Center for Biotechnology Information
NRPS	Non-ribosomal peptide synthetase
NSERC	Natural Sciences and Engineering Research Council (Canada)
NTC	No Template Control
P	Phosphorus
PCR	Polymerase Chain Reaction
PKS	Polyketide synthetase
QIIME2	Quantitative Insights Into Microbial Ecology 2
qPCR	Quantitative Polymerase Chain Reaction
RDA	Redundancy Analysis
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
TDS	Total Dissolved Solids
TLW	Turkey Lakes Watershed
V3	Variable region 3 of the 16S rRNA gene
V4	Variable region 4 of the 16S rRNA gene

Chapter 1 – Introduction

1.1 Overview

Accessibility of clean and safe drinking water is essential for human health as exposure to inadequate quality of water can be harmful to health (R. H. Lopes et al., 2022). Data collected in different countries vary by being scarce or unavailable due to low monitoring capacity of low- or middle-income places and rural areas, increasing health risks based on unknown water conditions (R. H. Lopes et al., 2022). Microbial contamination is a large factor that is involved in changing water quality levels through the presence of disease-causing organisms in drinking water (Ashbolt, 2004; R. H. Lopes et al., 2022). The importance of safe drinking water has influenced the increase of research to fill knowledge gaps in areas like microbial ecology to understand the presence and abundance of organisms in freshwater oligotrophic systems that are typically understudied and under evaluated (Douterelo et al., 2014; Russo et al., 2016).

Cyanobacteria, also known as blue-green algae, are the oldest known oxygen-producing organisms on Earth with fossil records dating back about 3.5 billion years ago (Paerl & Otten, 2013). The exploitation of nutrient-scarce and nutrient-enriched terrestrial and aquatic environments by many genera occur through the ability to fix atmospheric nitrogen and store phosphorus (Paerl & Otten, 2013). These abilities have increased the frequency of cyanobacterial populations resulting in the rise of toxic, food web altering, and hypoxia-generating harmful cyanobacterial blooms (Paerl, 2018; Paerl & Otten, 2013). With climate change-exacerbated disturbances (i.e., wildfires, stormwater) and human-induced

environmental changes, increased frequency of harmful cyanobacterial blooms are affecting the quality of drinking water through the production of cyanotoxins that are linked to various human diseases, neurological impairment, and death (Emelko et al., 2011; Ho & Michalak, 2020; Paerl & Paul, 2012; Wells et al., 2020). Forested watersheds supply high-quality drinking water to local communities and can be susceptible to these blooms, where the level of water purification can play a role in the local aquatic ecology and human health (Emelko et al., 2011; A. F. Lopes et al., 2019). This study focused on examining the Turkey Lakes Watershed in Northern Ontario, Canada to determine community composition and abundance of cyanobacteria present, with a focus on potential microcystin toxin-producing organisms in an oligotrophic system.

1.2 Forested Watersheds

The regulation of the quality of surface and groundwater in forested watersheds help provide a source of high-quality drinking water (A. F. Lopes et al., 2019). Forested headwaters are involved in providing most of the surface water that is usable, safe and clean to downstream regions that may require them (Emelko et al., 2011). Forests are involved in reducing soil run-off, mitigating stormflows, and moderating excess nutrient input that could increase pollutant levels (Caldwell et al., 2023; A. F. Lopes et al., 2019). However, forested watersheds are vulnerable to climate change-exacerbated disturbances such as wildfire, insect diseases, and land conversion for agriculture and urban land (Emelko et al., 2011; A. F. Lopes et al., 2019). Changes in these watersheds, especially with forest cover can affect

water treatment costs as forest cover helps with raw water quality, decreasing operational costs and large capital costs can be avoided (A. F. Lopes et al., 2019).

It is important to protect watersheds from these changes and potential disturbances as watersheds are valuable for providing high quality drinking water for various communities (Postel & Thompson, 2005). Policy makers and land managers utilize data linking forests with water quality to make informed decisions about water treatment infrastructure and forest management in critical water supply watersheds (Caldwell et al., 2023). Cities in North America have implemented successful watershed protection programs where millions of dollars were saved by avoiding the construction of filtration plants and annual operating costs (Blackburn et al., 2021; A. F. Lopes et al., 2019). Watershed protection ultimately provides high-quality source water and investing in watershed management and monitoring can help with drinking water treatment resilience (Blackburn et al., 2021).

1.2.1 Turkey Lakes Watershed

Collaborators within the *for*Water NSERC Network proposed research across major Canadian ecozones to study climate-change exacerbated disturbances on these areas to develop and investigate source water protection technologies and forested watershed management. Major study sites were chosen through long-term data present on the effects of forest management technologies on water that can be incorporated for drinking water treatability. One of these areas includes the Turkey Lakes Watershed, where studies began in the 1980s to explore the impact of long-range transport of air pollutants, mainly acidic deposition, at this site of the Canadian Shield that has vulnerable terrain that is undisturbed

(Jeffries et al., 1988). This research site is mainly unoccupied and not affected by the impact of human activities, and no significant natural disturbances have impacted the watershed (forest fires, insect outbreaks), where the most recent operational logging occurred in the 1950s, making this site relatively undisturbed (Jeffries et al., 1988; Webster et al., 2021). Since the conclusion of the initial deposition study, a whole-ecosystem approach was used for multidisciplinary analysis and data collection for understanding long-term environmental trends and provided contributions to advancement of fundamental watershed theory (Webster et al., 2021).

The main study site for this project is the Turkey Lakes Watershed (TLW) located in central Ontario approximately 50 km north of Sault St. Marie (Jeffries et al., 1988). It is on the Canadian Shield in the northern region of the Great Lakes-St. Lawrence forests (Jeffries & Foster, 2001). The forest consists of an uneven-aged hardwood system with the main tree species being sugar maple (Jeffries & Foster, 2001). The watershed contains four oligotrophic lake systems, Batchawana Lake, Wishart Lake, Little Turkey Lake, and Big Turkey Lake that are interconnected (**Figure 1.1**) (Jeffries et al., 1988). All the lakes, except Wishart, undergo thermal stratification from December to April and mid-May to the end of October, and in the autumn and spring, this stability of thermal stratification breaks down (Jeffries et al., 1988). The deeper depths of the water in Batchawana and Little Turkey Lake also become anoxic due to oxygen depletion (Jeffries et al., 1988). Research conducted in this thesis focused on Little Turkey Lake, with characteristics of this lake having a lake surface area of 19.2 ha, a maximum depth of 13.0 m, and a mean depth of 6.04 m (Jeffries et

al., 1988). Work conducted by Ellen Cameron, a former PhD student in the Müller lab, focused on collecting samples from the Turkey Lakes Watershed, with an emphasis on community composition over different depths and multiple years (Cameron, 2021). Following the previous work conducted, community composition over one seasonal period and multiple depths to determine abundance of cyanobacteria present was analyzed at Little Turkey Lake. Across most of the lakes at TLW, cyanobacterial genera are the dominant algae present, with *Chroococcus dispersus* and *Microcystis flos-aquae* largely being observed in Little Turkey Lake (Jeffries et al., 1988).

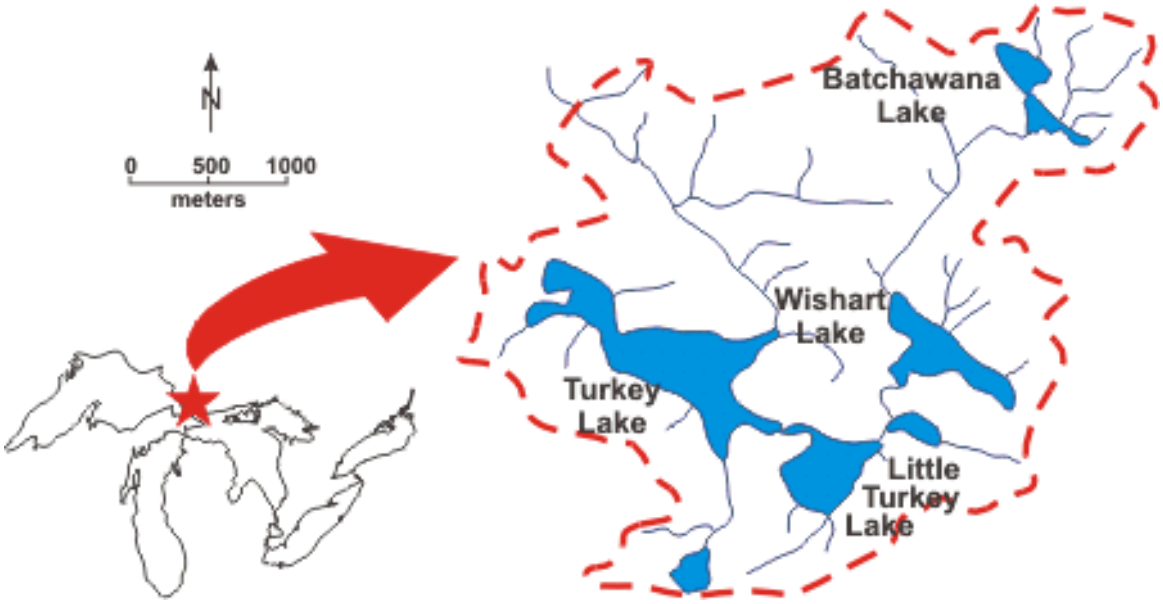


Figure 1.1 Map of Turkey Lakes Watershed located in the Algoma District of Ontario approximately 50 km north of Sault Ste. Marie (Government of Canada, 2014).

1.2.2 Algal Bloom Formation

Forests have a large influence on temperatures and how heat flows as forests can remain cool due to shade and canopy cover (Ellison et al., 2017). Deforestation and anthropogenic land use of forests can affect the canopy cover which allows for change in sunlight access to the water, increasing water temperatures (Ellison et al., 2017). The easy access to sunlight for photosynthesis by cyanobacteria stimulates bloom formation at the surface due to optimal growth conditions (Huisman et al., 2018). In forested watersheds, this can become an issue that affects drinking water quality due to the increase in blooms and toxin production (Huisman et al., 2018).

1.3 Harmful Algal Blooms

Algal blooms are a natural phenomenon that has expanded into a global issue due to the increase of reports on toxic algal blooms in freshwater systems (Favot et al., 2023; Winter et al., 2011). Harmful algal blooms are most observed and studied in coastal marine environments but can affect the open ocean and brackish or freshwater systems (Anderson et al., 2012). Harmful algal blooms (HABs) can be described in multiple ways with an example being microscopic algae accumulation and proliferation on the surface and within the water column that can be toxic even at low cell concentrations (Anderson et al., 2012; P. Zhang et al., 2022). This description can be applied to both highly toxic planktonic species and accumulation of cells on benthic substrates (Anderson et al., 2012).

Environmental and ecological changes in food web dynamics, nutrient levels, water column structure, carbon dioxide, and pH in marine and freshwater ecosystems can

exacerbate blooms increasing HAB formation (Emelko et al., 2016; Stauffer et al., 2019). Along with increasing temperatures, some HABs are predicted to proliferate with the ongoing changing climate in combination with differences in the eutrophication and coastal development (Stauffer et al., 2019). Eutrophication (over-enrichment) and development through anthropogenic nutrient loading and the favourable climatic condition for the microscopic algae promotes the expansion of the HABs at a global scale (Paerl, 2018). The microscopic algae blooms are comprised of certain cyanobacteria (blue-green algae), where biological interactions, dynamics, and toxicity within these HABs are not well known, but represents an important research gap that requires understanding (Anderson et al., 2012). The toxicity arises from the occurrence of cyanobacteria that causes water quality issues through the production of a wide range of odours and potent toxins (Bláha et al., 2009).

1.3.1 Cyanobacterial Harmful Algal Blooms

Cyanobacteria are photosynthetic prokaryotes that were involved in Earth's transition from a carbon dioxide-rich atmosphere to an oxygen-rich atmosphere through photosynthesis (Mazard et al., 2016). Cyanobacterial habitats include terrestrial and aquatic habitats including deserts, freshwater systems, and marine systems, with genera observed in eutrophic and oligotrophic conditions (Mazard et al., 2016). The ecological success of cyanobacteria in freshwater systems is visible through harmful cyanobacterial blooms that causes the loss of water clarity and quality (Paerl & Otten, 2013). The harmful aspect of the bloom is through the production of secondary metabolites, like certain toxins, that can cause gastrointestinal illnesses and liver and kidney damage in humans (Mazard et al., 2016). The accumulation of

these toxic blooms can vary depending on the environment and the adaptations made by specific species (Paerl & Otten, 2013).

Cyanobacteria have adapted to various stressors that help provide an advantage over other organisms to thrive and proliferate in certain environmental conditions. An ecological advantage of certain cyanobacterial species includes specific nitrogen fixing genera not needing external inputs of nitrogen for survival (Paerl & Otten, 2013). Cyanobacteria can also exploit stratified environments by producing gas vesicles that are involved in controlling buoyancy in the water column helping the organism stay within the euphotic zone and near the surface (Paerl & Otten, 2013). Bloom forming cyanobacteria optimize access to light and nutrients through the possession of heat and desiccation tolerant cells, photoprotective pigments, and alterations in buoyancy to travel throughout the water column (Paerl, 2018). These adaptations provide resistance to environmental extremes increasing bloom formation (Paerl, 2018). Studying cyanobacteria in freshwater systems and the role played by the organisms in bloom formation and toxin production can help narrow down ways to improve drinking water quality.

1.3.2 Cyanotoxins

Toxins produced by cyanobacteria, cyanotoxins, affect human health and can also affect processes in freshwater systems (Winter et al., 2011). Cyanotoxins are secondary metabolites produced by bloom forming cyanobacteria (Huisman et al., 2018). Cyanotoxins fall into the group of peptides, alkaloids, or lipidic compounds based on chemical structures where cyanobacterial lipopolysaccharides (components of the cell wall) can be labelled as

cyanotoxins (Bláha et al., 2009). The impact on production of cyanotoxins by environmental factors are currently unknown, but nutrients, temperature, competition, or all these factors and more could play a role in the release of cyanotoxins into water systems (Paerl & Otten, 2013). Some common N_2-fixing toxin producing genera are *Anabaena*, *Cylindrosperopsis*, and *Nodularia* (Paerl & Otten, 2013). The most common non-N_2-fixing toxin producing cyanobacteria are *Microcystis and Planktothrix* (Paerl & Otten, 2013). Some examples of cyanotoxins include microcystins, nodularins, cylindrospermopsins, anatoxin-a, saxitoxins, and β- *N*-methylamino-L-alanine (BMAA) (Huisman et al., 2018).

Cyanotoxins can be classified as neurotoxins, hepatotoxins, dermatotoxins, irritant toxins, and cytotoxins which affect animals, and can be released when cells die and lysis occurs (Bláha et al., 2009; Paerl, 2018; Winter et al., 2011). If a bloom collapse were to occur, large quantities of dissolved toxins could be released into the water column (Paerl & Otten, 2013). It is possible that certain toxins may be physiologically or ecologically functional in cyanobacteria, like the binding of microcystin protecting the carbon-fixing enzyme to adapt to oxidative stress (Huisman et al., 2018). However, without a clear indication on why toxins are produced, or what role toxins play in cyanobacteria processes, it is difficult to control toxin production in drinking water, which affects drinking water quality.

1.3.3 Microcystin

In the environment, microcystin is present in high amounts in cyanobacterial biomass, but overall research conducted on the function of microcystin is lacking (Bláha et al., 2009).

Microcystin belongs to the family of monocyclic heptapeptides with the most common toxin variation being microcystin-LR, where the LR represents the presence of the amino acids, leucine (L) and arginine (R) (Bláha et al., 2009; Du et al., 2019). Of the cyanotoxins, microcystin-LR is the most regulated cyanotoxin with provisional treated drinking water guideline values established by the World Health Organization at 1 µg/L and a maximum acceptable concentration set by Health Canada at 1.5 µg/L (Wert et al., 2014). Microcystins are involved in causing severe liver damage and can also be tumour-promoting by affecting mammalian tissues (Huisman et al., 2018). Exposure to large doses of microcystin can cause cytoskeletal disorganization that may lead to hemorrhagic shock causing liver damage (Müller et al., 2017). Microcystin producing genera of cyanobacteria are listed in **Table 1.1**. These health risks associated with high concentrations of microcystin present in drinking water can be harmful to humans and is a public health concern, where research conducted on the potential toxin-producing organisms can help provide more input into understanding the drinking water sources (Bláha et al., 2009; Du et al., 2019).

Table 1.1 Microcystin producing genera of cyanobacteria.

Toxin	**Genera**
Microcystin	*Anabaenopsis, Annamia, Aphanocapsa, Arthrospira, Calothrix, Dolichospermum (Anabaena), Fischerella, Haphalosiphon, Leptolyngbya, Merismopedia, Microcystis, Nostoc, Oscillatoria, Phormidium, Planktothrix, Plectonema, Pseudanabaena, Radiocystis, Synechococcus*

Table adapted from (Chorus & Welker, 2021).

1.3.4 Environmental Factors that Influence Algal Blooms

There are various triggers for blooms with climate change and anthropogenic activities having a large impact on algal blooms (Nwankwegu et al., 2019). With the rising of sea levels, global warming, and changes in the pattern of precipitation, the frequency and proliferation of blooms has increased (Nwankwegu et al., 2019). Extreme hydrological events, like storm-related flooding increases nutrient loading, and droughts decrease water flow and increase residence time for blooms to expand (Nwankwegu et al., 2019; Paerl, 2018). Warmer surface waters lead to intensified vertical stratification as the rise in temperature causes ice cover to melt earlier and freeze later, increasing the period of time for potential bloom formation (Paerl, 2018). Bloom forming cyanobacteria can survive and adapt to changes in climate, while also showing optimal growth in higher temperatures (Nwankwegu et al., 2019).

Anthropogenic activities are human activities that can favour the expansion of HABs, with some triggers including the construction of dams, deforestation, and land developments (Nwankwegu et al., 2019). The excessive anthropogenic nutrient loading give rise to toxic cyanobacterial harmful algal blooms due to tolerance and adaptation by certain taxa to these man-made and environmental changes (Paerl, 2018). As mentioned previously, the favourable conditions for proliferation include warmer temperatures and nutrient rich concentrations (Winter et al., 2011).

1.3.4.1 Changes in Temperature

Bloom-forming cyanobacteria are generally favoured by rising temperatures and with climate change, this can be seen as a catalyst for bloom formation (Winter et al., 2011). The warmer temperatures increase growth rate of the bloom forming cyanobacteria and provides greater water column stability as cyanobacteria can migrate vertically to find optimal environments in the photic zone, which is advantageous against organisms that cannot migrate (Schmale et al., 2019). Many bloom forming bacteria reach maximal growth rates at higher temperatures and growth rates appear to increase faster with temperature in comparison to eukaryotic phytoplankton (Huisman et al., 2018). Warmer temperatures have been an advantage to cyanobacterial genera that can exploit the stratification through the formation of gas vesicles that provide buoyancy to travel the water column (Paerl & Otten, 2013). There is a higher rate of photosynthesis at the surface due to warmer temperatures, and this plays a role in blocking the non-buoyant phytoplankton in the deeper parts of the water column which limits growth (Paerl & Otten, 2013). Climate change has also lengthened the ice-free season with warming beginning earlier and lasting longer from mid-spring to fall (Paerl, 2018; Winter et al., 2011). In Ontario, rising temperatures have shown an increase in the relative abundance of green algae due to warmer climates playing a large role in unwanted bloom formation (Winter et al., 2011).

1.3.4.2 Nutrient Input

Nutrient-rich freshwater environments have increased cyanobacterial harmful algal bloom (CyanoHABs) formation through the influence of phosphorus (P) and nitrogen (N)

(Paerl & Otten, 2013). Widespread eutrophication of freshwater systems has increased harmful algal blooms through N and P runoff due to anthropogenic influences like intensified agriculture practices (Huisman et al., 2018; Schmale et al., 2019). While chemical composition of fertilizer has changed over time, urea is currently the most commonly used N fertilizer as it is a biologically available form of N that can initiate bloom formation (Paerl et al., 2016).

In marine waters, nitrogen is the limiting macronutrient for primary production of photosynthetic organisms (Anderson et al., 2002). In freshwater lakes, P is observed in the least abundance among the nutrients that are needed by photosynthetic organisms (Anderson et al., 2002). The more common form of inorganic P is orthophosphate (PO_4^{3-}) that CyanoHABs can accumulate and store intracellularly when there is a decrease in external P (Paerl, 2018). The larger abundance of P relative to N can favour the growth of CyanoHABs that are specific for N_2-fixation which can supply N for primary production (Paerl & Otten, 2013). This occurs through enzymatically converting atmospheric nitrogen (N_2) to ammonia (NH_3) that is biologically available (Paerl & Otten, 2013).

The increase in N fertilizer has changed the N to P ratio in many water sources with increased nitrogen loading changing the species composition (Huisman et al., 2018). Non-nitrogen-fixing cyanobacteria like *Microcystis spp.* have been observed in higher abundances than nitrogen-fixing cyanobacteria which may also stimulate the production of the cyanotoxin microcystin (Huisman et al., 2018). Nutrient rich water systems play a large role

in bloom proliferation leading to overall low drinking water quality through the expansion of toxin producing cyanobacteria.

1.4 Molecular Analysis for Understanding Microbial Communities

Exposure to cyanotoxins occur through ingestion of contaminated drinking water or recreational exposure (i.e., swimming in contaminated water), where is it important to study levels of microcystin present in watersheds and the effect on drinking water quality (Greer et al., 2018). The production of these toxins contribute to the reason why algal blooms are harmful, with many environmental factors playing a role in causing the harmful algal blooms (HABs). Although HABs form naturally, there are many conditions that are involved in bloom expansion including climate change and anthropogenic activities (Nwankwegu et al., 2019). There are various molecular tools that are effective and can be used in combination to monitor cyanobacterial communities in different regions that are susceptible to toxic cyanobacterial blooms (Müller et al., 2017). To better understand community composition and abundance in an oligotrophic freshwater system, these molecular tools can be utilized to analyze samples efficiently, and can be translated to field monitoring to overcome limitations of conventional methods of identification.

1.4.1 16S rRNA Gene Sequencing

This study focused on completing a community analysis of bacterial species present in Little Turkey Lake using molecular tools and methods to determine if there are bacteria that could affect water quality. This was conducted using high throughput 16S rRNA gene

amplicon sequencing and bioinformatics (Straub et al., 2020). Amplicon sequencing helps eliminate the challenges associated with using microscopy for analysis and provides a rapid and sensitive method to study DNA (Cameron et al., 2021). The 16S rRNA gene is present in all microorganisms and it is a phylogenetic marker with highly conserved and variable regions that evolved independently which can now be used to differentiate taxa (**Figure 1.2**) (Straub et al., 2020). This gene can be amplified using universal primers 515F (5'-GTGYCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACNVGGGTWTCTAAT-3') (W. Walters et al., 2015).

After amplification, next-generation sequencing (NGS) technology, Illumina MiSeq, can be used at a lower cost for high-throughput in genomic studies (McElhoe et al., 2014). Once sequencing is conducted, resulting sequences will provide meaningful insight into the evolutionary patterns and relationships, and phylogenetic trees can be constructed to describe the evolutionary relatedness of the communities (McCormack et al., 2013; Sharma & Lal, 2017). It is valuable to use these molecular methods to analyze microbial communities from watershed samples collected to determine what organisms are present that can pose threats to water quality.

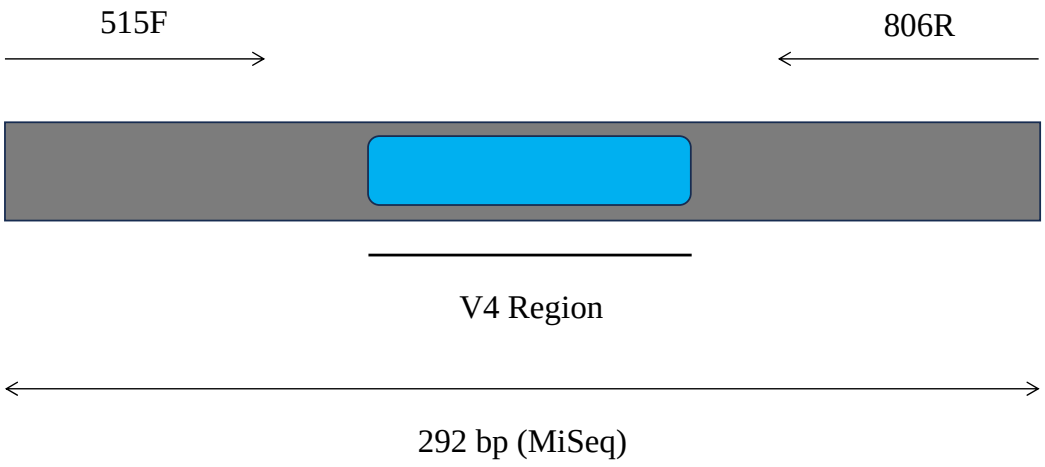


Figure 1.2 16S rRNA gene conserved, variable, and hypervariable regions with primer pairs used for sequencing. Conserved regions are represented in grey and variable region 4 (V4) is represented in blue. Arrows represent forward and reverse primers respectively. Image adapted from Shahi et al., 2017.

1.4.2 Cyanobacterial Classification

Using 16S rRNA gene sequencing, taxonomic classification can be conducted and can help with understanding organism diversity (Komárek, 2016). With a wide range of forms of cyanobacteria that are difficult to differentiate and diversify, evaluation of taxonomy and classification proves to be difficult (Komárek, 2016). Previously, taxonomy has been based on morphological characteristics, however, with the recent inclusions of molecular analyses, cyanobacterial taxonomy has made rapid changes through sequencing and established workflows (Komárek, 2016; Strunecký et al., 2023). The increase in cyanobacterial genera have occurred through a combination of morphological taxonomy and genomic sequencies of well-defined reference strains to help construct robust phylogenetic

trees (Strunecký et al., 2023). However, with this comes limitations as algae have variation and creating a database that comprises all aspects is a difficult and long-term process, where revisions are always required.

1.4.3 Microcystin Producing Cyanobacteria

Molecular methods can be used to monitor the presence of the species of cyanobacteria that release these microcystin (Jungblut & Neilan, 2006). Microcystin synthesis involves the *mcy* gene cluster involving the *mcy*E gene that codes for microcystin synthetase, important for the formation of microcystin (**Figure 1.3**) (Pearson et al., 2016). A molecular method was developed to be included in a PCR reaction to detect microcystin producing strains through the aminotransferase (AMT) domain, which is located between the non-ribosomal peptide synthetase (NRPS) and polyketide synthetase (PKS) domains of the *mcy*E gene (Jungblut & Neilan, 2006). The AMT region is conserved in the biosynthetic pathway and is important for microcystin synthesis, where this gene is conserved in multiple genera of cyanobacteria (Beversdorf et al., 2015; Jungblut & Neilan, 2006). Amplification can be conducted using HEPF (5'-TTTGGGGTTAACTTTTTTGGGCATAGTC-3' and HEPR (5'-AATTCTTGAGGCTGTAAATCGGGTTT-3') primers that target the AMT region (Jungblut & Neilan, 2006). Isolation and amplification of this gene can help with identifying potential toxin producing cyanobacterial species present in the water samples collected.

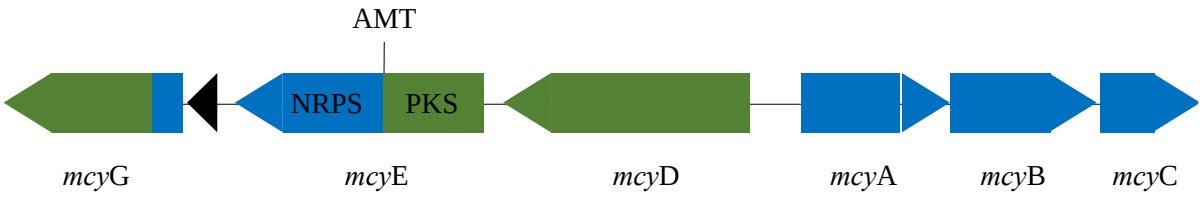


Figure 1.3 *Microcystis aeruginosa mcy* gene cluster. Polyketide synthetase (PKS) represented in green, non-ribosomal peptide synthetase (NRPS) represented in blue, and domain coding for biosynthesis enzymes indicated in black. Image adapted from Pearson et al., 2016.

1.4.4 Quantitative Polymerase Chain Reaction

Quantitative polymerase chain reaction (qPCR) molecular techniques have been used to determine presence of genes efficiently due to the quick turnaround time (Chiu et al., 2017). The qPCR technique can be used to amplify and quantify the target gene through the measurement of fluorescence during the PCR cycle which is sensitive and accurate (Pinto et al., 2012). This molecular method can be used to quantify the abundance of bacterial gene copy number present in the sample using universal primers and toxin gene specific primers (Chiu et al., 2017; Pinto et al., 2012; Walther et al., 2020). This study focused on the specific amplification and quantification of the copy number for the cyanobacterial 16S rRNA gene V3 region (**Figure 1.4)**, where nucleotide sequence data provides information on a large diversity of cyanobacterial species (Zwart et al., 2005).

Microscopy methods can be limiting as monitoring focuses on risks based on biomass, but qPCR can detect a higher content of genes present even when cyanobacterial

abundance is lower (Pacheco et al., 2016). This technology is valuable for detecting and quantifying potential toxin producers and can be used as a predictive tool for evaluating the susceptibility of source waters to blooms for preparing for future harmful algal blooms (Al-Tebrineh et al., 2012). Quantitative PCR utilizes identical conditions in its assays, along with high analytical sensitivity and specificity, high throughput, providing high confidence of gene identity (Borchardt et al., 2021; Rocha & Manaia, 2020). Previously, qPCR data has been routinely used to inform public policy and health for recreational water quality and epidemiological surveillance (Borchardt et al., 2021), and would be extremely valuable in drinking water monitoring protocols. Producing high quality qPCR data for understanding microbial communities to analyze abundance can be beneficial to optimizing drinking water treatment protocols (Borchardt et al., 2021). By optimizing the qPCR protocol for conducting research on samples taken from Little Turkey Lake, cyanobacterial community analysis and abundance of toxin gene copy numbers were conducted to further analyze research gaps in cyanobacterial molecular analysis.

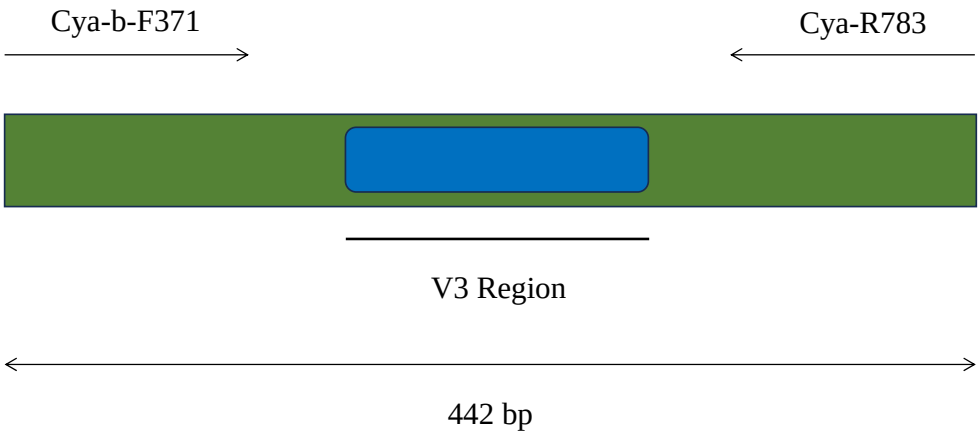


Figure 1.4 Cyanobacterial 16S rRNA gene V3 region conserved, variable, and hypervariable regions with primer pairs used for sequencing. Conserved regions are represented in green and variable region 3 (V3) is represented in blue. Arrows represent forward and reverse primers respectively. Image adapted from Zwart et al., 2005.

1.5 Research Objectives

With the increase of nuisance blooms and cyanobacterial harmful algal blooms proliferation globally, research into the variability of organisms present in freshwater oligotrophic systems is required. Changes in environmental factors like increasing temperatures, changes in pH, and variation in nutrient levels impacts the community composition variability in oligotrophic lakes in Ontario. Turkey Lakes Watershed provides a model system to determine cyanobacterial community composition and any variability that might occur based on sampling due to minimal developmental activity in surrounding areas and the long-term research that occurs there. The objectives of this study were to:

1. **Identify the cyanobacterial community composition at multiple sampling timepoints and depths at the Turkey Lakes Watershed to determine variation and the impact of the environment.**
2. **Quantify the abundance of cyanobacteria and potential toxin-producing organisms present using quantitative PCR to determine gene presence and variation over multiple timepoints in a day, depths, and a season.**

The purpose of this study was to provide insight into molecular techniques that can be utilized effectively and efficiently to analyze water systems for community composition and abundance to understand variation that may occur based on the sampling protocol. This study focused on observing amplicon sequence variants from samples sequenced and comparing it to reference strains from databases to understand the microbial communities as there have been recent advances in sequencing technologies for cyanobacteria (Lefler et al., 2023). These objectives provide methodology for a combination of efficient and effective molecular analysis to analyze high quality data for improvement of drinking water monitoring protocols.

Monitoring protocols being developed currently do not include depth variability and do not take into consideration changes that may affect the vernal window that is optimal for cyanobacterial growth. This thesis looked to expand knowledge collected from Dr Ellen Cameron's study of the Turkey Lakes Watershed (Cameron, 2021) through sampling multiple depths to observe any differences in communities present at the various depths. The seasonal aspect of adding in the months of May to August provided analysis into the

variation and abundance of cyanobacterial communities present in Little Turkey Lake over a period. It is important to observe the cyanobacterial communities, along with potential toxin producers present, and this study filled in the gap of knowledge about the abundance of toxin-producing organisms that may be observed. By analyzing samples at various depths, sampling protocols can be updated and developed to include important information about oligotrophic lakes throughout multiple months.

Chapter 2 – Methods

2.1 Study Site

The main study site for this project was the Turkey Lakes Watershed (TLW) located in central Ontario, Canada approximately 50 km north of Sault St. Marie (Jeffries et al., 1988). This watershed is located on the Canadian Shield in the northern region of the Laurentian Great Lakes and the St. Lawrence forest (Jeffries & Foster, 2001). The forest consists of an uneven-aged hardwood system with the predominant tree species being sugar maple (*Acer saccharum*) (Jeffries & Foster, 2001). The watershed contains four lakes, Batchawana Lake, Wishart Lake, Little Turkey Lake, and Big Turkey Lake that are interconnected (**Figure 1.1**) (Jeffries et al., 1988). All the lakes, excluding Wishart, undergo thermal stratification from December to April and mid-May to the end of October, and in the autumn and spring, this stability of thermal stratification breaks down (Jeffries et al., 1988). The deeper depths of the water in Batchawana and Little Turkey Lake also become anoxic due to oxygen depletion (Jeffries et al., 1988). Across most of the lakes, cyanobacterial genera are the phytoplankton present, with *Chroococcus dispersus* and *Microcystis flos-aquae* largely being observed in Little Turkey Lake (Jeffries et al., 1988). The research at TLW has provided data that has guided policy issues, drinking water treatability, and changes to ecosystems due to climate change-exacerbated disturbances, making it a useful system to study (Webster et al., 2021).

2.2 Sample Collection

To aid determination of cyanobacterial community composition and abundance in an oligotrophic water system (TLW) over a summer season, sampling took place over two-day periods during the months of May to August 2022. All water samples were taken near the deepest point of Little Turkey Lake, which had been previously marked by an anchored buoy. In addition, sampling also occurred at different timepoints during the day consisting of morning (9-10am), afternoon (11am-12pm), and evening (2-3pm) sample times. Water samples were taken from three different depths in duplicate: surface, Secchi, and 1 m below Secchi from May to August to account for variable distribution of organisms in the water column. Secchi depth is an estimate of water transparency and is measured by lowering a black and white Secchi disk into the water column until it is no longer visible and then raised to where it is just seen (X. Liu et al., 2020), and a sample was then taken at that depth and the depth recorded. Beginning in June, integrated samples were also taken alongside multiple depths that can be utilized to analyze the entire water column substituting sampling analysis at specific depths.

For all timepoints in the day and throughout the months of May to August, a Masterflex E/S portable sampler peristaltic pump was used for the collection of duplicate 1 L water samples at the different depths. These water samples were then filtered through a 1.2 μm pore size, 47 mm GF/C filter Whatman filter (Whatman plc, Buckinghamshire, United Kingdom) and stored in coolers until transported back to Waterloo where the samples were

then stored at -20°C until DNA extraction. Further sampling and environmental details are illustrated in **Table A1**, **Table A2, Table A3** and **Table A4**.

2.3 DNA Extraction and 16S rRNA Gene Amplicon Sequencing

DNA extraction was conducted on the filters using the DNeasy PowerSoil Kit (QIAGEN Inc., Venlo, Netherlands). Filters were initially cut using scissors and ground down using a Bead RuptorElite Homogenizer (Omni International, USA) set to 2 mins for 1 cycle after adding the first solution and vortexing. Samples were then centrifuged following the protocol included with the DNeasy Powersoil kit. A NanoDrop 2000 spectrophotometer (Thermo Scientific Inc, USA) was used to determine quantity and purity of the extracted DNA. Sample were incubated at -20°C until further molecular analysis was conducted. The DNA extracts were then diluted (if required) to have concentrations within the range of 1-10 ng/μL to present for amplicon sequencing. Extracted samples of 20 μL were submitted for 16S rRNA gene V4 region amplicon sequencing using the Illumina Miseq Platform (Illumina Inc., San Diego, United States) conducted at a commercial laboratory, Metagenom Bio Inc. (Waterloo, Ontario, Canada). Samples sent for sequencing include three different timepoints throughout a day (morning, afternoon, and evening) at Secchi depth in May and integrated depth for samples from June – August. Sequencing protocols were conducted by Metagenom Bio Inc. (https://www.metagenom.com) with amplification using universal 16S V4 rRNA primers (515FB: 5'- GTGYCAGCMGCCGCGGTAA and 806RB: 5'- GGACTACNVGGGTWTCTAAT) (W. Walters et al., 2015).

2.3.1 Sequencing Analysis

Bioinformatic analyses were conducted on the paired end demultiplexed forward and reverse sequences obtained from Metagenom Bio Inc. using QIIME2 (v. 2022.8) (Bolyen et al., 2019). The sequences were imported into the QIIME2 pipeline and DADA2 (Callahan et al., 2016) was used for trimming and truncating sequences to 250 base pairs for filtration of lower quality reads. The DADA2 pipeline was used to remove Illumina sequencing errors and following quality control, the amplicon sequence variant (ASV) table was generated. This pipeline does not infer sequences of single reads, where diversity analysis may be impacted by singleton omission that can lead to potential sources of bias (Chik et al., 2018; Schmidt et al., 2022). This analysis and the potential effects of the errors on diversity is outside the scope of this work and was not analyzed. The SILVA 138 database (Quast et al., 2013; Yilmaz et al., 2014) was used for taxonomic classification and assignment of the ASVs using the Naïve-Bayes taxonomic classifier. ASVs were then filtered to remove any taxonomic classifications that were assigned to mitochondria and chloroplasts. Filtered sequences were then used for further ASV analysis.

2.3.2 Community and Diversity Analysis

QIIME2 generated files were imported into R (v. 4.2.2) for community and diversity analyses (R Core Team, 2022). Downstream analyses were conducted using *qiime2R* (v. 0.99.6) (Bisanz, 2018) and *phyloseq* (v. 1.42.0) (McMurdie & Holmes, 2013) objects were created using QIIME2 files. The R package, *ggplot2* (Wickham, 2009) was utilized to produce relative abundance of ASVs to create taxonomic bar plots using the frequency of an

ASV in comparison to the total number of ASVs observed in the sample. Following this, *mirlyn* was used to repeatedly rarefying to normalize library sizes for further diversity analyses with a rarefaction curve output to determine normalized library sizes (Cameron et al., 2021). The rarefaction curves generated for bacterial communities (consisting of cyanobacteria) and cyanobacteria separately provided normalized library size that can be used for normalizing the samples. The bacterial communities were rarified to 18000 sequences and the cyanobacterial communities were rarified to 4500 sequences. With the library normalization, further diversity analysis within samples were conducted using alpha diversity metrics (Shannon diversity index) using 100 iterations of rarefying library sizes to consider community diversity loss and artificial variation that may occur due to subsampling (Cameron et al., 2021). A redundancy analysis (RDA) was conducted using RStudio and the package *vegan* to determine significant relationships between environmental parameters (i.e., temperature, pH, conductivity, and total dissolved solids) and the sample diversity (Barbosa et al., 2022).

2.3.3 Phylogenetic Analysis of Cyanobacterial Sequences

A reference database comparison was conducted to analyze the 16S rRNA gene sequences that were assigned to cyanobacteria from the samples collected. The Basic Local Alignment Search Tool (BLAST) (https://blast.ncbi.nlm.nih.gov/Blast.cgi) associated with the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov) helped with analysis of sequence similarities to previously characterized and published cyanobacterial 16S rRNA gene sequences. The BLAST

algorithm was set to exclude uncultured sequences and reference sequences were chosen based on a query cover larger than 70% and percent identity above 90%. Reference sequences from published papers were obtained from GenBank to include within the multiple sequence alignment to aid in the identification of unknown taxa. Cyanoseq, that has been curated to contain cyanobacterial 16S rRNA gene sequences to analyze phylogenetic relationships (Lefler et al., 2023), was used as a secondary database to include reference sequences from other published sources down to genus and species level. This database aided with acquiring reference sequences outside of the BLAST results to incorporate more reference databases for this phylogenetic analysis. The databases that provided the reference sequences, accession numbers, and references are provided in **Table B1**.

Reference sequences and the 16S rRNA gene sequences from the samples were combined and, using MEGA X (Molecular Evolutionary Genetics Analysis), a multiple sequence alignment (MSA) was conducted with a ClustalW algorithm (Kumar et al., 2018). An initial alignment was conducted, where references sequences were then trimmed to match the length of the gene sequences containing just the 16S rRNA gene V4 region of interest. The sequences were re-aligned for alignment accuracy and a Maximum Likelihood (ML) method was used to produce the most probable phylogenetic tree of where a bootstrap value of 1000 was used for resampling the data to determine confidence of the clusters. The Kimura-2 parameter model with a discrete gamma distribution (G) and rate invariable site (I) model was chosen as the best model using the MEGA X parameters. The phylogenetic tree with reference sequences and the 16S rRNA gene V4 region gene sequences from samples

collected provides an inference of evolutionary relatedness based on alignment similarity, taxonomic resolution, and clustering patterns observed within the tree. The phylogenetic tree has been coloured based on inference of clustering of the various cyanobacterial Orders according to the assigned orders of the published reference sequences obtained from the NCBI and Cyanoseq databases and bootstrap values above 70% are represented as symbols.

2.4 DNA Extraction and Quantitative PCR on samples from Little Turkey Lake

Quantitative polymerase chain reaction (qPCR) was conducted on samples collected from Little Turkey Lake, after DNA extraction, to quantitatively determine the DNA copies/µL of the cyanobacterial 16S rRNA gene V3 region and *mcy*E gene. Amplification and quantification were performed using a PowerTrack™ SYBR Green Master Mix (Thermo Fisher Scientific Inc, USA) compatible with the QuantStudio™ 6 Pro qPCR machine (Thermo Fisher Scientific Inc, USA). All analyses were conducted and run through the Design and Analysis Software (v. 2.6.0) (Thermo Scientific Inc, USA) associated with the qPCR machine. Cyanobacteria-specific 16S rRNA gene primers Cya-b-F371 (5'-CCTACGGGAGGCAGCAGTGGGGAATTTTCCG-3') and Cya-R783 (5'GACTACWGGGGTATCTAATCCCW-3') were used in the qPCR reaction for analysis of cyanobacterial gene copy number abundance (Zwart et al., 2005). The primers for the AMT region of the *mcy*E gene include HEPF (5'-TTTGGGGTTAACTTTTTTGGGCATAGTC-3') and HEPR (5'-AATTCTTGAGGCTGTAAATCGGGTTT-3'), utilized for the measurement of the

abundance of gene copy numbers of the potential microcystin toxin producers (Jungblut & Neilan, 2006). Assays were conducted using a final volume of 20 µL per single reaction including a 1X final concentration of 2X master mix (10 µL), 1X yellow sample buffer (0.5 µL) for loading samples, 0.4 µM forward primer (10 µM stock), 0.4 µM reverse primer (10 µM stock), 2 µL of DNA from environmental samples, and topped off with PCR grade water. A no template control (NTC) was included as a negative control. The samples were analyzed in triplicate for each environmental sample with a positive control of known concentration of the gene of interest (gBlocks™, Integrated DNA Technologies) (Otten et al., 2016), and a NTC were included in each run. The cycling program for the cyanobacterial-specific 16S rRNA gene of interest consisted of an initial preheating step of 95°C for 3 minutes, followed by 40 cycles of 95°C for 20 s, 52°C for 20 s, and 72°C for 20 s, ending with an optional hold at 4°C (Müller et al., 2017). The cycling program for the AMT region for the *mcy*E gene of interest consisted of an initial preheating step of 95°C for 3 minutes, followed by 40 cycles of 95°C for 20 s, 56°C for 30 s, and 72°C for 1 min, ending with an optional hold at 4°C (Müller et al., 2017). For both analyses, a melt (dissociation) curve was performed by gradually increasing the temperature from 56°C to 95°C with 1°C/s increments and a hold time of 2 s at each increment (Müller et al., 2017).

To determine if the gene of interest amplicons were amplified, qPCR products were run on a 1% agarose gel with a GeneRuler 1KB+ DNA ladder (Thermo Scientific Inc, USA) as a standard comparison. The estimated product size for the cyanobacteria-specific 16S rRNA gene is 442 bp (Zwart et al., 2005) and the estimated product size for the AMT region

of the *mcy*E gene is 472 bp (Jungblut & Neilan, 2006). The data from the qPCR analysis underwent ANOVA (analysis of variance) statistics to determine the significant differences between the different depths of the water column using RStudio (RStudio Team, 2020).

2.4.1 Quantitative PCR Standard Curve

A standard curve was produced and utilized to compare the environmental data to determine abundance of gene copy numbers. Known concentrations of the genes of interest were obtained through synthetic sequences (gBlocks™, Integrated DNA Technologies) (Otten et al., 2016). The gBlocks™ were analyzed using 10-fold dilutions to create a standard curve with known concentrations for comparison with the environmental samples. The same cycling programs performed for the environmental samples were conducted for the standard curves for both genes of interest (see **Section 2.4**). The standard curve was used to convert the obtained Ct values into gene copy numbers using the equation of the line obtained (DNA copies/µL vs. average Ct value) (Ahmed & Kim, 2018). Presentation of the standard curves were done using *pcr*, a R package that can produce standard curves with input of Ct values using RStudio (Ahmed & Kim, 2018; RStudio Team, 2020).

Chapter 3 – Results

3.1 Bacterial Community Composition

Bacterial community composition profiles using the V4 region of the 16S rRNA gene, were analyzed through assigning taxonomy using a bioinformatic pipeline (**Figure 3.1**). A total of 12 water samples were sequenced with outputs of amplicon sequence variants (ASVs) from 5,313 bacterial and 3 archaeal phyla, including 31 unassigned phyla presented as unclassified and 11 ASVs assigned to chloroplast and mitochondrial hits. A sum of 5,358 unique ASVs with a total frequency of 533,254 were observed across all the samples. After chloroplast and mitochondrial hits were filtered and removed, 4,300 unique ASVs were observed with a total frequency of 482,372 across all samples. The dominant bacterial phyla which composed of 95% of the reads observed through the months of May to August are listed in descending order (**Table 3.1**).

Table 3.1 Relative abundance of bacterial phyla observed from all samples.

Phylum	**Relative Abundance (%)**
Proteobacteria	35.3
Bacteroidetes	16.3
Actinobacteriota	12.8
Verrucomicrobiota	7.1
Cyanobacteria	4.0
Planctomycetota	3.5
Bdellovibrionota	2.1
Firmicutes	1.4
Chloroflexi	1.1
“Bacteria” or “Unassigned”	11.4
Other	4.9

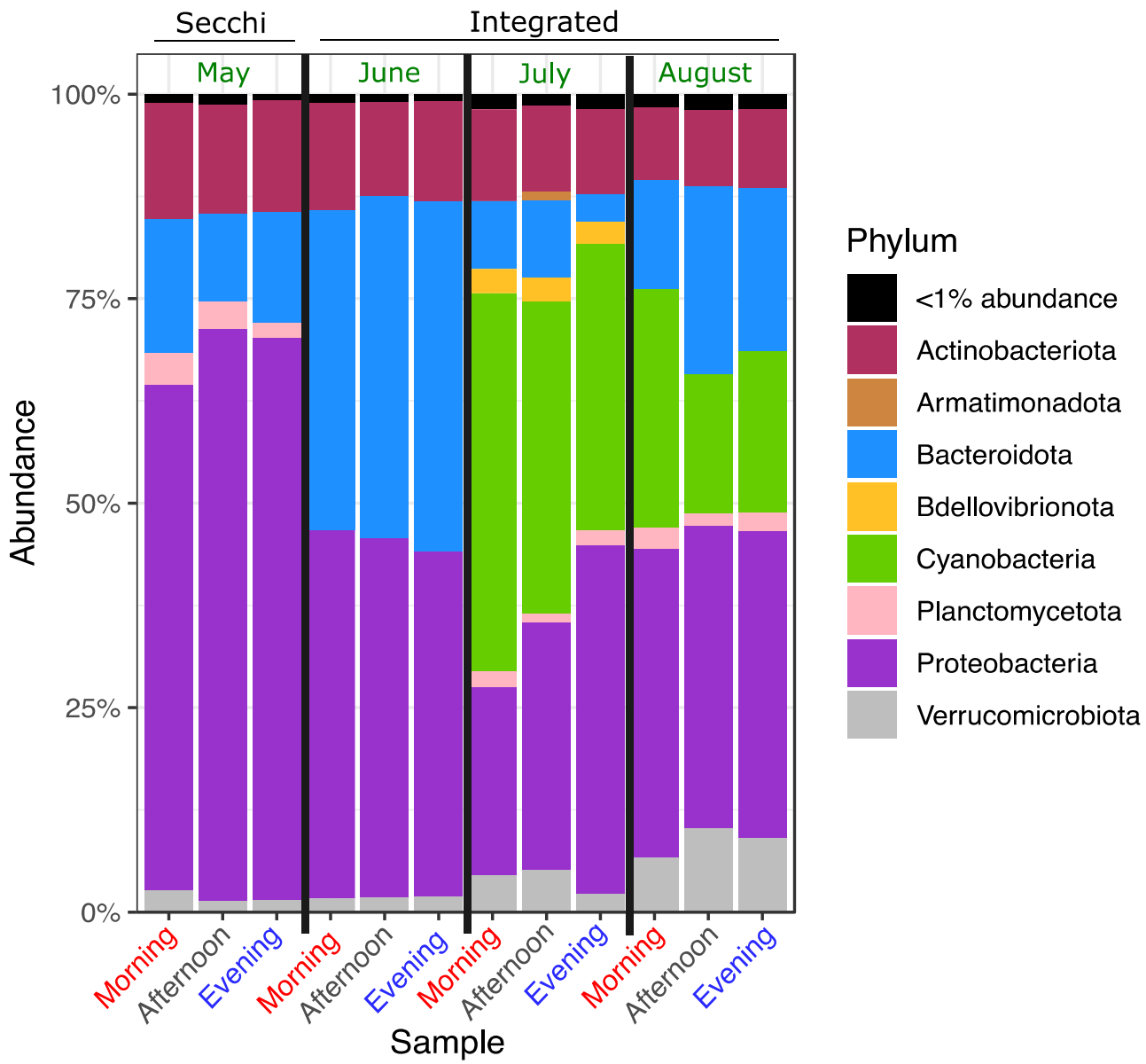


Figure 3.1 Relative abundance chart of bacterial phyla from Little Turkey Lake samples collected from May to June 2022. Samples are ordered based on time of collection and labeled with a specific depth where samples differentiate.

3.2 Cyanobacterial Community Composition

An output of 173 unique cyanobacterial ASVs were analyzed from the samples with a total frequency of 70,681 (**Table 3.2**). Using the SILVA taxonomic classification, 165 ASVs were resolved down to the genus-species level. For assignment, 72 reads were assigned to the sequence *Cyanobium* PCC-6307 (NR_114406.1), with this genus making up 42% of the reads and 21 reads being assigned to *Microcystis* PCC-7914 (Pasteur Culture Collection, Paris, France). Within the cyanobacteria phylum, these two genera comprise 54% of the assignments with the rest of the reads of multiple different organisms making up 46% observed across the samples (**Figure 3.2**).

Table 3.2 Cyanobacterial amplicon sequence variants (ASVs) with total number of reads per month at Little Turkey Lake over three time points.

Sample Site	Month	Time	Cyanobacterial Reads	Total Reads
Little Turkey Lake	May	Morning	27	55461
		Afternoon	140	42721
		Evening	23	26854
	June	Morning	91	54642
		Afternoon	63	41380
		Evening	58	41502
	July	Morning	22877	50392
		Afternoon	6830	18470
		Evening	22172	65095
	August	Morning	7515	26122
		Afternoon	4607	27403
		Evening	6278	32330

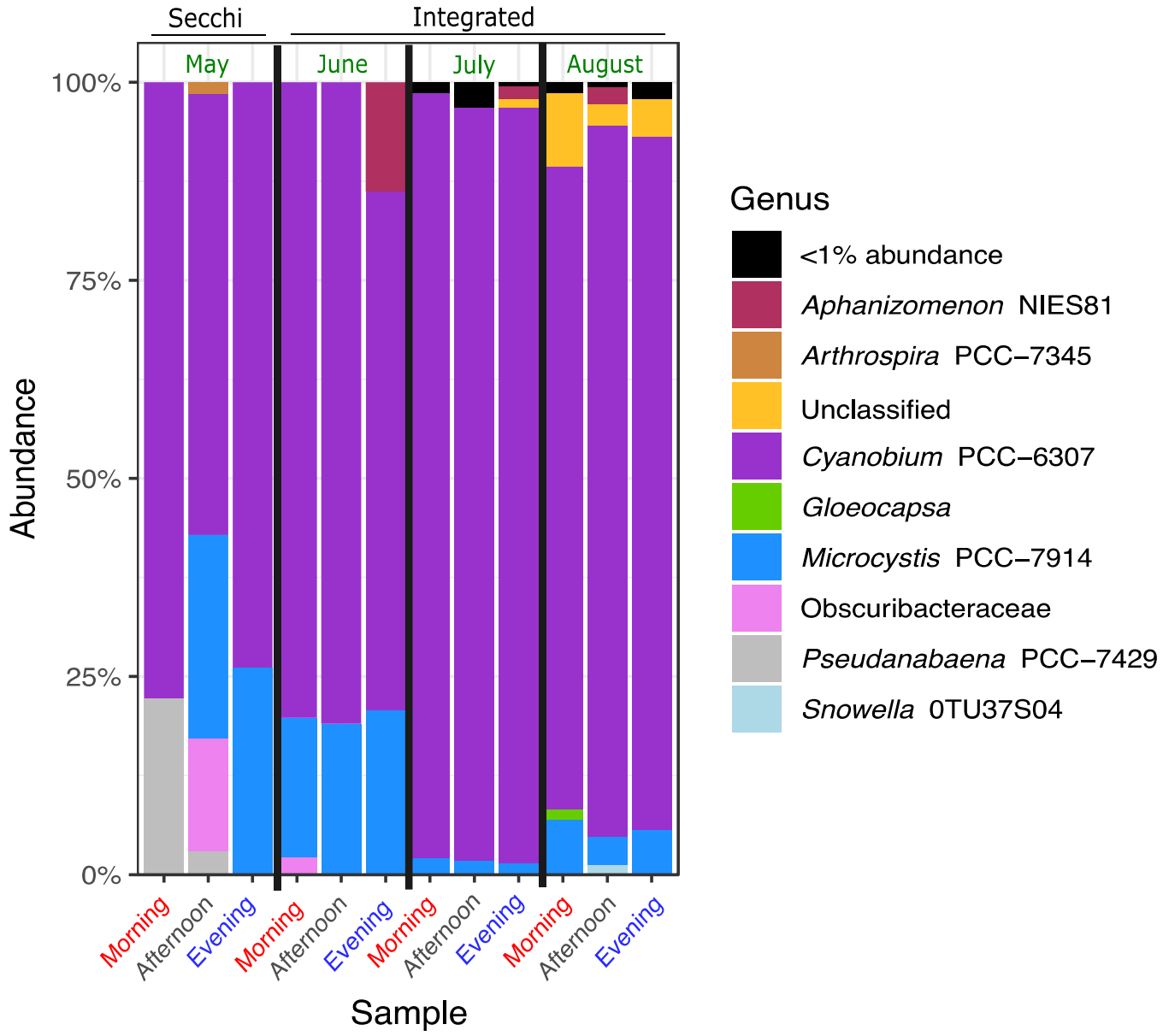


Figure 3.2 Relative abundance chart of cyanobacterial genera from Little Turkey Lake samples collected from May to June 2022. Samples are ordered based on time of collection and labeled with a specific depth where samples differentiate.

3.2.1 Cyanobacterial Phylogenetic Tree

A maximum likelihood phylogenetic tree was constructed using the 174 cyanobacteria ASVs analyzed with the inclusion of reference sequences obtained from the NCBI and CyanoSeq databases to observe evolutionary relationships of the sequences. Reference sequences were selected after conducting a BLAST search with the ASVs and chosen through genus names in CyanoSeq. Published reference sequences were selected, with accession numbers, and are provided in **Table B1**. The length of the alignment was 250 nucleotides to capture the sequences obtained from both the environmental samples and the reference sequences. The phylogenetic tree was divided by clustering of cyanobacterial orders according to reference sequence classification in the NCBI database through published articles, and not the assigned taxonomy from the SILVA database (**Figure 3.3**). The six orders observed across the months of May to August 2022 for Little Turkey Lake include Synechococcales, Nostocales, Oscillatoriales, Pseudanabaenales, Chroococcales, and Gloeobacterales. Bootstrap values are represented by shapes that illustrate levels of support with a circle for strongly supported (90-100%), a square for moderately supported (80-89%), and a star for weakly supported (70-79%). The outgroup for this tree was represented through various bacterial species that do not fall within the cyanobacteria phylum.

Moving from the top of the tree on page 40 and following on to page 41, the second largest number of clusters represented fall within the order Synechococcales, with 4 separate clusters highlighted in purple, making this order non-monophyletic. The first large clade consists of reference sequences for Synechococcales, with one moderately supported (80-

89%) monophyletic group consisting of the *Synechococcus* PCC9005 (AS216950.1) reference sequence and 8 ASVs. Within this cluster, there are multiple monophyletic and paraphyletic groups that have good support (>70%) and are also not well supported (<70%), that do not associate with any reference sequences. The second Synechococcales cluster has a moderately supported (80-89%) combination of reference sequences and 2 ASVs. The third Synechococcales cluster is not well supported (<70%), but there is a monophyletic relationship between ASV136 and the *Trichocoleus caatingensis* CATCD2 (MT311250.1) reference sequence. The fourth Synechococcales cluster consists of 5 ASVs and the *Snowella litoralis* 0TU35S07 (AJ781039.1) reference sequence which is strongly supported (90-100%).

On page 41, the Nostocales order had 2 clusters represented on the tree, highlighted in green, illustrating a non-monophyletic relationship. The first cluster portrayed a strong paraphyletic group consisting of 5 ASVs and the *Dolichospermum lemmermannii* NIVA-CYA 697 (LT670892.1) reference sequence. Within this cluster, there are also weakly supported (70-79%) paraphyletic branches consisting of reference sequences for species in the Nostocales order, but also groups for ASVs without reference sequences that are supported (70-100%). The second cluster for Nostocales consists of a weakly supported (70-79%) monophyletic relationship between ASV20 and *Calothrix* sp. Su 37 (MH271078.1).

Sequences that are classified in the order Pseudanabaenales appeared in 5 clusters, highlighted in beige, within the phylogenetic tree and are therefore non-monophyletic. Moving from the top of the tree on page 41 and following onto page 42, the first clade of

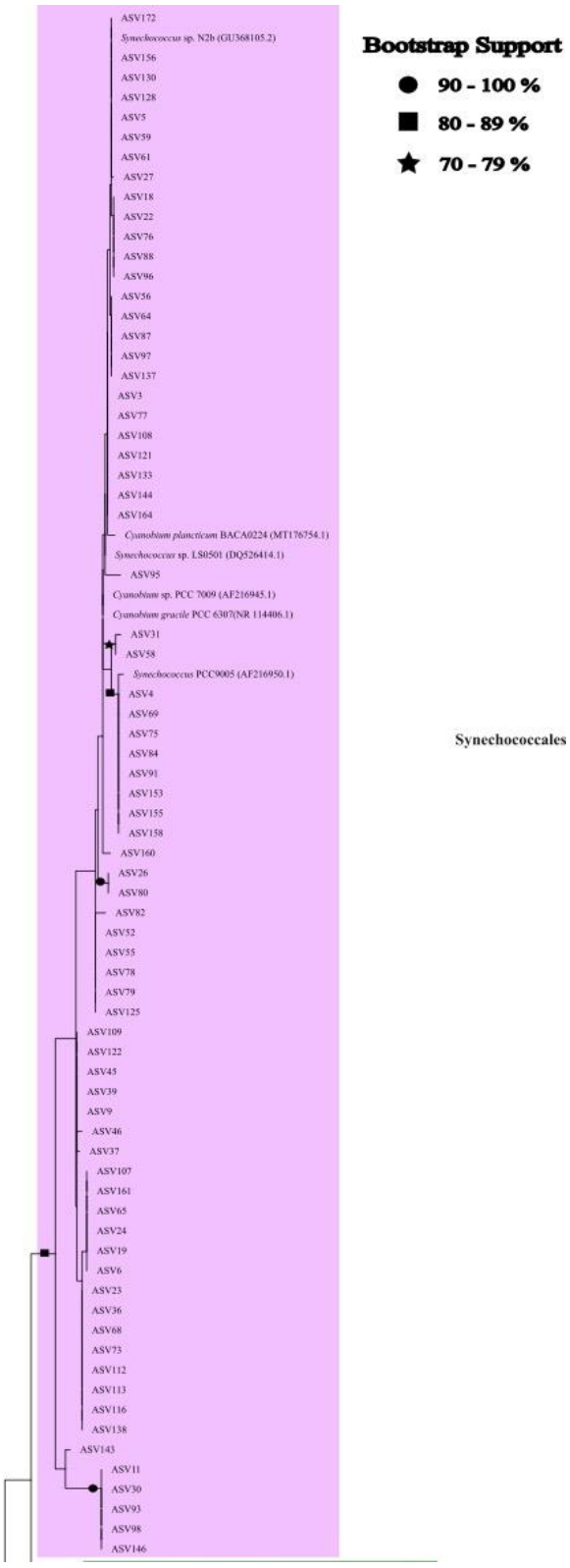
Pseudanabaenales consisted of the reference sequence from *Pegethrix indistincta* strain GSE-TBC-7GA (NR 172691.1) and ASV141, which was well supported by bootstrap resampling (90-100%). The second Pseudanabaenales clade was not well supported (<70%) and included the reference sequence *Chamaethrix vaginata* (NR 177769.1) which was well supported clustering with ASV154. The third Pseudanabaenales clade consisted of two strongly supported (90-100%) monophyletic groups with ASV13 and reference sequence from *Leptolyngbya* sp. CYN82 (F925324.1), and ASV34 with *Leptodesmis alaskaensis* V20 (MK861877.1), along with one strongly supported (90-100%) paraphyletic group. The fourth clade is comprised of two strongly supported (90-100%) monophyletic groups that separate with ASV135 and *Limnothrix mirabilis* AWQC-LIM001 (KY550441.1) on one branch and ASV14, ASV29 and ASV54 on a separate branch. The final cluster for the Pseudanabaenales order on page 42 consists of a monophyletic relationship that is strongly supported (90-100%) with 3 ASVs and 2 reference sequences including *Pseudanabaena foetida* var. intermedia NIES-512 gene (LC153790.1) and the *Pseudanabaena limnetica* gene (LC016776.1) reference sequences. There is a weakly supported (70-79%) paraphyletic relationship between ASV50 and the *Pseudanabaena* sp. PCC 7403 (AB075995.1) and *Phormidium mucicola* IAM M-221 gene (AB003165.1) reference sequences.

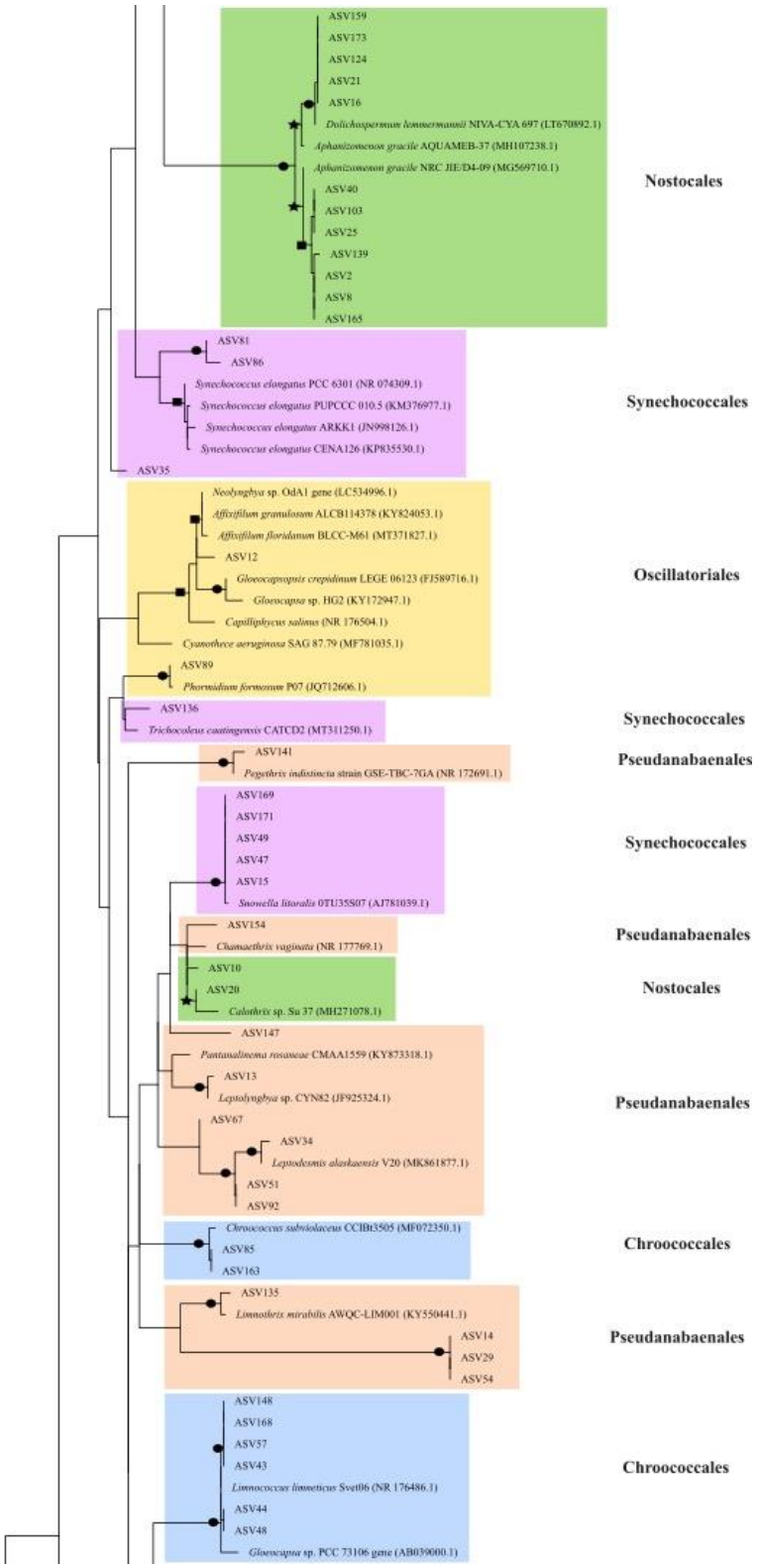
The Chroococcales order had 3 clusters on pages 41-42, highlighted in blue within the tree, and is therefore non-monophyletic. The first cluster illustrates a paraphyletic relationship with a strong support (90-100%) consisting of 2 ASVs and the *Chroococcus subviolaceus* CCIBt3505 (MF072350.1) reference sequences. The second cluster consists of

a monophyletic relationship between 4 ASVs and *Limnococcus limneticus* Svet06 (NR 176486.1), along with a paraphyletic relationship between 2 ASVs and *Gloeocapsa* sp. PCC 73106 gene (AB039000.1) with strong support (90-100%). The last cluster of the Chroococcales order consists of multiple branches that have good bootstrap support (70-100%), with one paraphyletic relationship between 4 ASVs and *Microcystis aeruginosa* PCC 7941 (AJ133171.1), along with a monophyletic relationship between 10 ASVs and the *Radiocystis geminata* TAU-MAC 1214 (OL310656.1) and *Radiocystis* sp. JJ30 3 (AM710389.1) reference sequences with strong support (90-100%).

On pages 41-42, the Oscillatoriales order had 2 clusters represented in yellow, making this order non-monophyletic. The first cluster illustrates a moderately supported (80-89%) paraphyletic relationship between multiple reference sequences and ASV12. There is also a strongly supported (90-100%) monophyletic relationship between the *Phormidium formosum* P07 (JQ712606.1) and ASV89. The second cluster for the Oscillatoriales order consists of a strongly supported (90-100%) group of ASVs with an unsupported association with the *Ammassolinea attenuata* (NR 177760.1) reference sequence.

The order with the least number of clusters is Gloeobacterales with 1 cluster group on page 42, represented and highlighted in grey, making this order monophyletic. There is a moderately supported (80-89%) paraphyletic relationship between 4 ASVs and *Anthocerotibacter panamensis* strain C109 (NR 172653.1). Other branches consisting of ASVs represented in this cluster have good support (70-100%), but are not directly associated with reference sequences.


ASV172
Synechococcus sp. N2b (GU368105.2)
ASV156
ASV130
ASV128
ASV5
ASV59
ASV61
ASV27
ASV18
ASV22
ASV76
ASV88
ASV96
ASV56
ASV64
ASV87
ASV97
ASV137
ASV3
ASV77
ASV108
ASV121
ASV133
ASV144
ASV164
Cyanobium plancticum BACA0224 (MT176754.1)
Synechococcus sp. LS0501 (DQ526414.1)
ASV95
Cyanobium sp. PCC 7009 (AF216945.1)
Cyanobium gracile PCC 6307(NR 114406.1)
ASV31
ASV58
Synechococcus PCC9005 (AF216950.1)
ASV4
ASV69
ASV75
ASV84
ASV91
ASV153
ASV155
ASV158
ASV160
ASV26
ASV80
ASV82
ASV52
ASV55
ASV78
ASV79
ASV125
ASV109
ASV122
ASV45
ASV39
ASV9
ASV46
ASV37
ASV107
ASV161
ASV65
ASV24
ASV19
ASV6
ASV23
ASV36
ASV68
ASV73
ASV112
ASV113
ASV116
ASV138
ASV143
ASV11
ASV30
ASV93
ASV98
ASV146
Bootstrap Support
90 - 100 %
80 - 89 %
70 - 79 %
Synechococcales


ASV159
ASV173
ASV124
ASV21
ASV16
Dolichospermum lemmermannii NIVA-CYA 697 (LT670892.1)
Aphanizomenon gracile AQUAMEB-37 (MH107238.1)
Aphanizomenon gracile NRC JIE/D4-09 (MG569710.1)
ASV40
ASV103
ASV25
ASV139
ASV2
ASV8
ASV165
Nostocales
ASV81
ASV86
Synechococcus elongatus PCC 6301 (NR 074309.1)
Synechococcus elongatus PUPCCC 010.5 (KM376977.1)
Synechococcus elongatus ARKK1 (JN998126.1)
Synechococcus elongatus CENA126 (KP835530.1)
Synechococcales
ASV35
Neolyngbya sp. OdA1 gene (LC534996.1)
Affixifilum granulosum ALCB114378 (KY824053.1)
Affixifilum floridanum BLCC-M61 (MT371827.1)
ASV12
Gloeocapsopsis crepidinum LEGE 06123 (FJ589716.1)
Gloeocapsa sp. HG2 (KY172947.1)
Capilliphycus salinus (NR 176504.1)
Cyanothece aeruginosa SAG 87.79 (MF781035.1)
Oscillatoriales
ASV89
Phormidium formosum P07 (JQ712606.1)
ASV136
Trichocoleus cautingensis CATCD2 (MT311250.1)
Synechococcales
ASV141
Pegethrix indistincta strain GSE-TBC-7GA (NR 172691.1)
Pseudanabaenales
ASV169
ASV171
ASV49
ASV47
ASV15
Snowella litoralis 0TU35S07 (AJ781039.1)
Synechococcales
ASV154
Chamaethrix vaginata (NR 177769.1)
Pseudanabaenales
ASV10
ASV20
Calothrix sp. Su 37 (MH271078.1)
Nostocales
ASV147
Pantanalinema rosaneae CMAA1559 (KY873318.1)
ASV13
Leptolyngbya sp. CYN82 (JF925324.1)
ASV67
ASV34
Leptodesmis alaskaensis V20 (MK861877.1)
ASV51
ASV92
Pseudanabaenales
Chroococcus subviolaceus CCIBt3505 (MF072350.1)
ASV85
ASV163
Chroococcales
ASV135
Limnothrix mirabilis AWQC-LIM001 (KY550441.1)
ASV14
ASV29
ASV54
Pseudanabaenales
ASV148
ASV168
ASV57
ASV43
Limnococcus limneticus Svet06 (NR 176486.1)
ASV44
ASV48
Gloeocapsa sp. PCC 73106 gene (AB039000.1)
Chroococcales

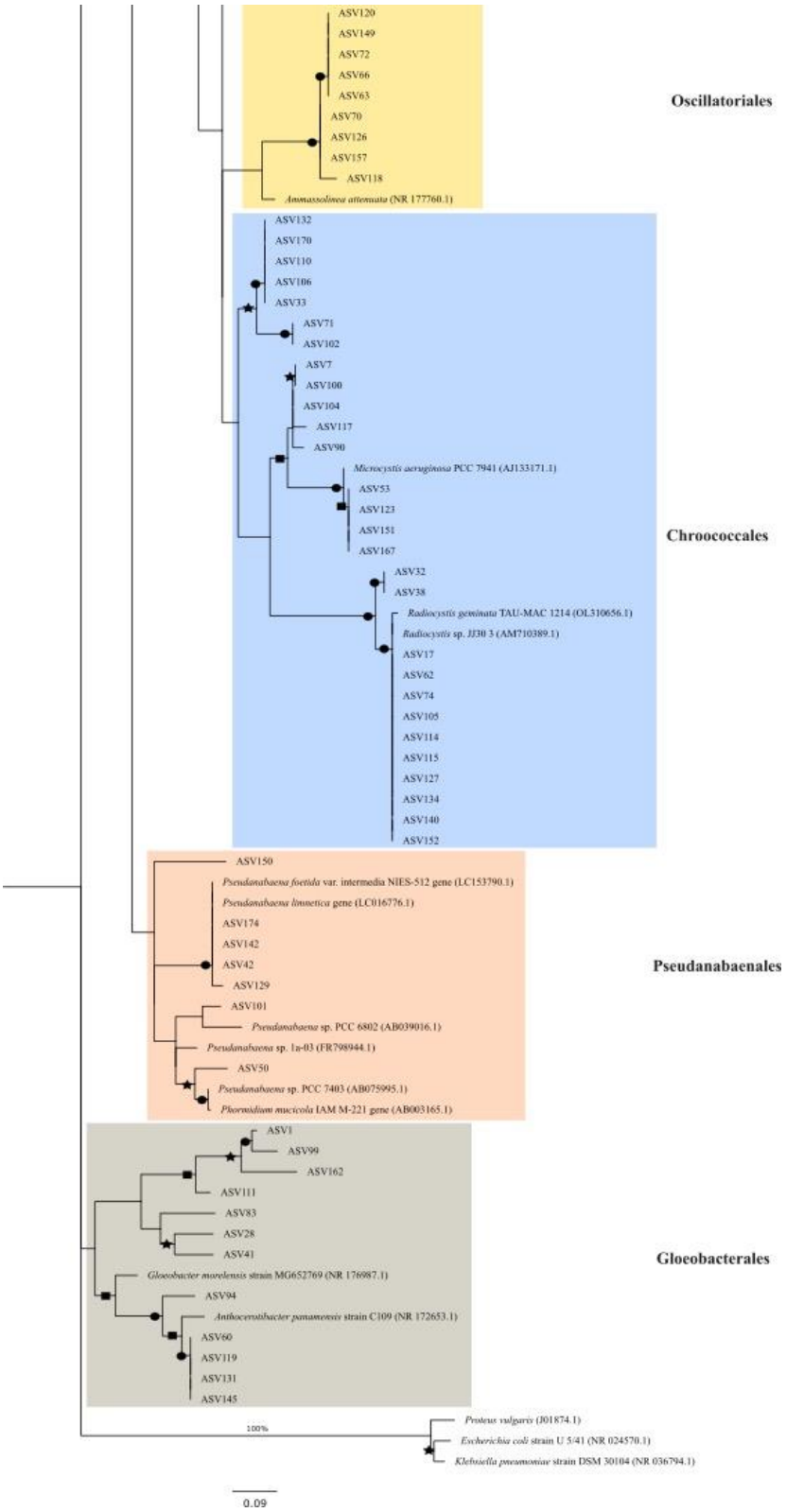

ASV120
ASV149
ASV72
ASV66
ASV63
ASV70
ASV126
ASV157
ASV118
Ammassolinea attenuata (NR 177760.1)
Oscillatoriales
ASV132
ASV170
ASV110
ASV106
ASV33
ASV71
ASV102
ASV7
ASV100
ASV104
ASV117
ASV90
Microcystis aeruginosa PCC 7941 (AJ133171.1)
ASV53
ASV123
ASV151
ASV167
ASV32
ASV38
Radiocystis geminata TAU-MAC 1214 (OL310656.1)
Radiocystis sp. JJ30 3 (AM710389.1)
ASV17
ASV62
ASV74
ASV105
ASV114
ASV115
ASV127
ASV134
ASV140
ASV152
Chroococcales
ASV150
Pseudanabaena foetida var. intermedia NIES-512 gene (LC153790.1)
Pseudanabaena limnetica gene (LC016776.1)
ASV174
ASV142
ASV42
ASV129
ASV101
Pseudanabaena sp. PCC 6802 (AB039016.1)
Pseudanabaena sp. 1a-03 (FR798944.1)
ASV50
Pseudanabaena sp. PCC 7403 (AB075995.1)
Phormidium mucicola IAM M-221 gene (AB003165.1)
Pseudanabaenales
ASV1
ASV99
ASV162
ASV111
ASV83
ASV28
ASV41
Gloeobacter morelensis strain MG652769 (NR 176987.1)
ASV94
Anthocerotibacter panamensis strain C109 (NR 172653.1)
ASV60
ASV119
ASV131
ASV145
Gloeobacterales
100%
Proteus vulgaris (J01874.1)
Escherichia coli strain U 5/41 (NR 024570.1)
Klebsiella pneumoniae strain DSM 30104 (NR 036794.1)
0.09

Figure 3.3 Phylogenetic tree of ASVs from Little Turkey Lake for the months of May to August 2022 assigned to cyanobacteria. Reference sequences obtained through BLAST and the NCBI database and included for taxonomic resolution. The phylogenetic tree was constructed using MEGA X through the Maximum Likelihood method with a bootstrap of 1000. Bootstrap values above 70% are assigned a shape (circle – 90-100%, square – 80-89%, star – 70-79%). Branches lacking shapes illustrate less than 70% bootstrap support. Colours are assigned to orders of Cyanobacteria that sequences clustered within.

3.3 Alpha Diversity of Communities

Alpha diversity was analyzed for both bacterial communities and filtered for cyanobacterial communities within all the samples. The R package *mirlyn* was used for rarefaction and normalizing the library size (Cameron et al., 2021). Using a rarefied library size of 18000 for bacterial communities, all samples from May to August fell above the rarefied library size and were included, separated by time points in a day. Through analysis using the Shannon Index, bacterial community diversity within samples fell in the range of 4-5 (**Figure 3.4**). Cyanobacterial community analysis was conducted using a rarefied library size of 4500 that excluded the months of May and June due to low number of reads. Through analysis using the Shannon Index, cyanobacterial community diversity within samples fell in the range of 1.5-2.5 (**Figure 3.5**). There were more monthly variations observed for whole bacterial community diversity within samples in comparison to the cyanobacterial community diversity between months and timepoints within a day. For the month of May in the bacterial community diversity, the depth of analysis was Secchi, and integrated samples were analyzed for June – August for both analyses.

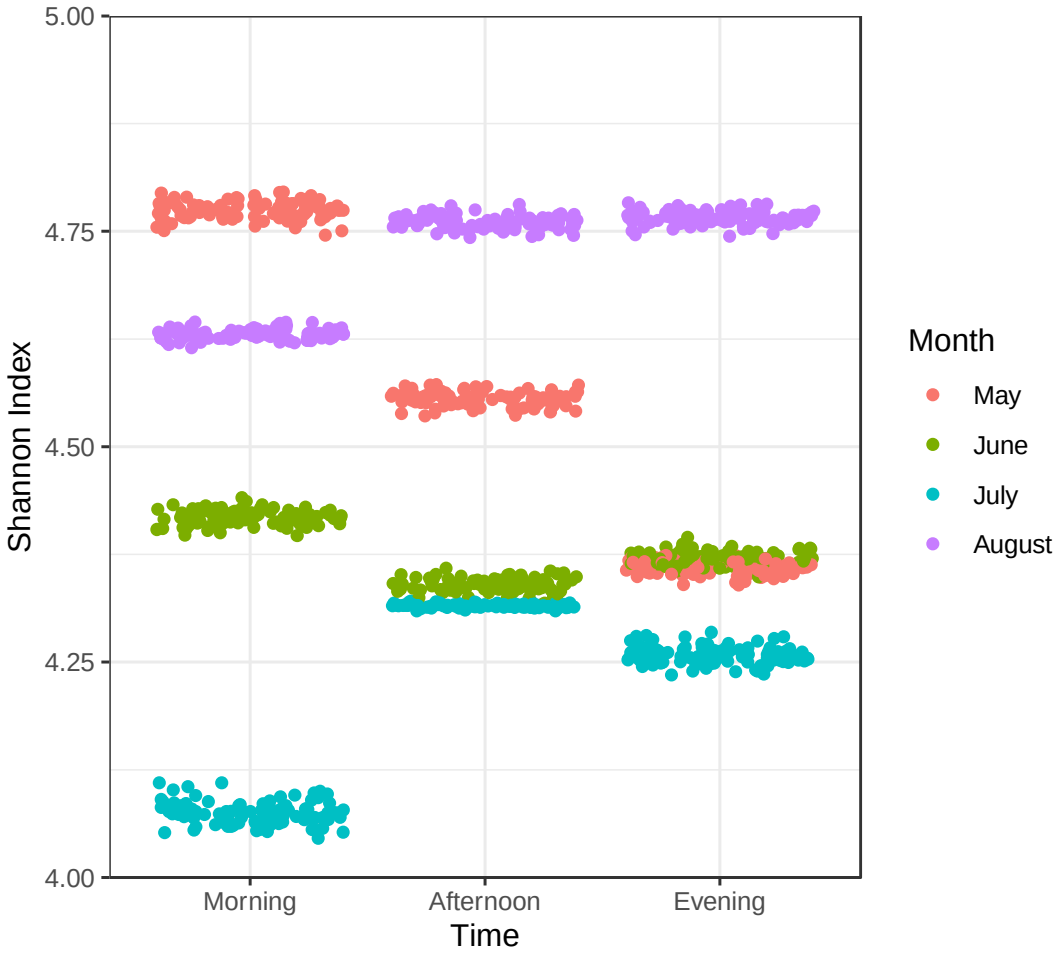


Figure 3.4 Alpha diversity of bacterial communities from Little Turkey Lake sampled from May to August 2022. Bacterial community alpha diversity plot constructed with a rarefied library size of 18000 using the Shannon diversity index through the R package *mirlyn*. No samples were excluded after rarefying the library size.

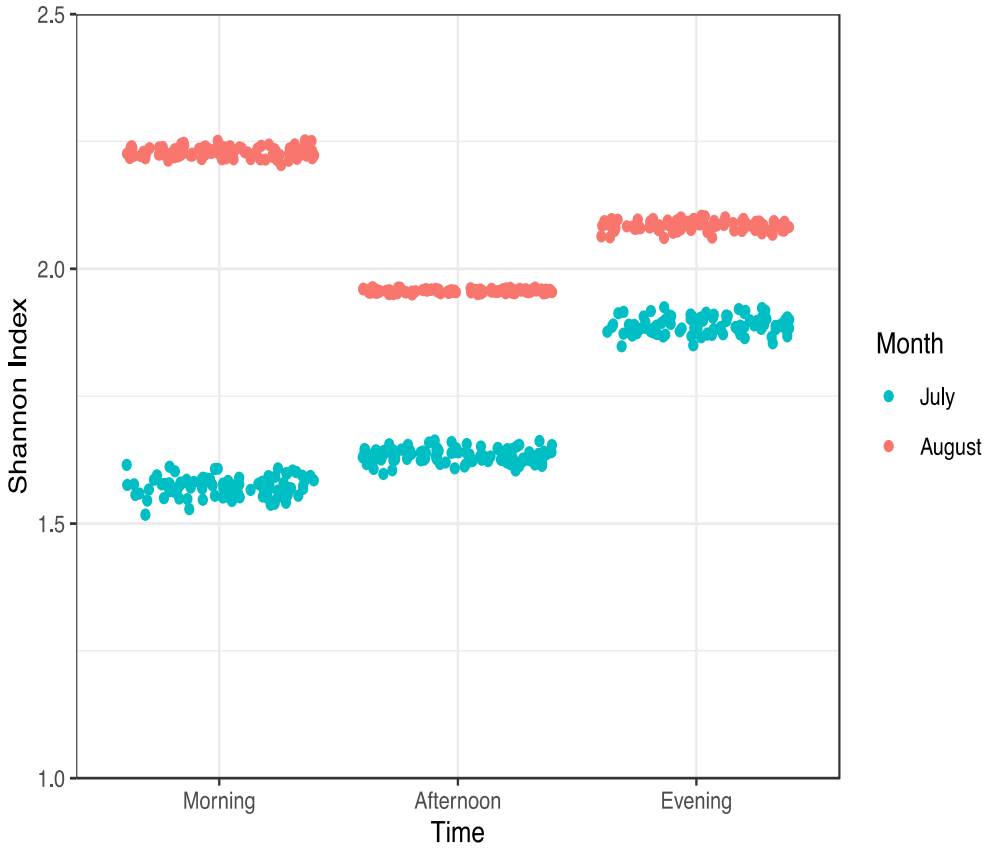


Figure 3.5 Alpha diversity of cyanobacterial communities from Little Turkey Lake sampled from May to August 2022. Cyanobacterial community alpha diversity plot constructed with a rarefied library size of 4500 using the Shannon diversity index through the R package *mirlyn*. May and June samples were excluded from analysis due to low cyanobacterial reads.

3.4 Redundancy Analysis

A redundancy analysis was conducted to determine significant relationships with the environmental variables (i.e., temperature, pH, conductivity, and total dissolved solids) and the taxonomy of samples in the different months and timepoints. Majority of the samples were not significantly affected by the environmental variables, with few outliers being

present. The July evening and August morning samples were significantly driven by the pH and temperature at the time of sampling. The parameters did not have a significant impact on the morning May and morning July samples visible through placement of the points on the graph. Dispersion of most samples fell within the middle of the graph (**Figure 3.6**).

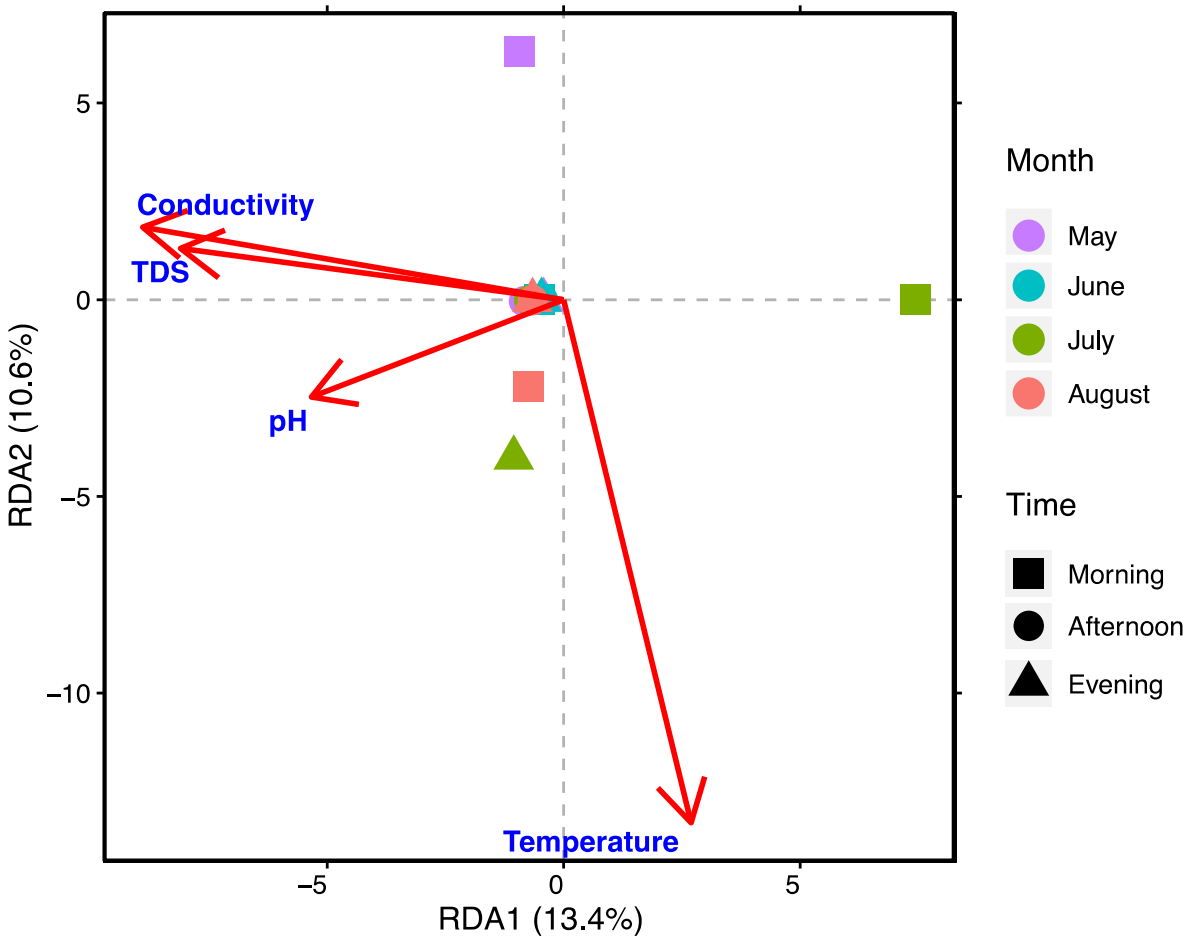


Figure 3.6 Redundancy analysis ordination plot for analysis of environmental parameters impacts on samples collected from May to August 2022 at Little Turkey Lake, outlined with three different timepoints. Legend outlines colour variation by month and shape variation for timepoints in a singular day. Significant environmental parameters are illustrated by red arrows.

3.5 Standard Curve for Quantitative PCR

3.5.1 Cyanobacterial V3 Region of the 16S rRNA Gene Standard Curve Analysis

A standard curve was produced using a dilution series of template DNA sequences of the gene of interests using known concentrations (gBlocks™, Integrated DNA Technologies) (Otten et al., 2016). Six dilutions were conducted to determine abundance of gene copy numbers ranging from 2.20E+09 – 2.20E+04 DNA copies/µL. The Cya-b-F371 and Cya-R783 primer set (Zwart et al., 2005) was used for producing the V4 region of the 16S rRNA gene standard curve. The regression equation was $y = 63.181 – 6.1438x$ with an $R^2 = 0.9974$ and an efficiency of 45.46% (**Figure 3.7**). In the equation, *y* is threshold cycle (Ct value) at a set fluorescence threshold level of 0.164 and *x* is the amount of DNA that is represented as a $\log_{10}$ DNA copies/µL. A melt curve analysis was conducted from the qPCR products where all outputs illustrated a similar peak at 85°C (**Figure 3.8**).

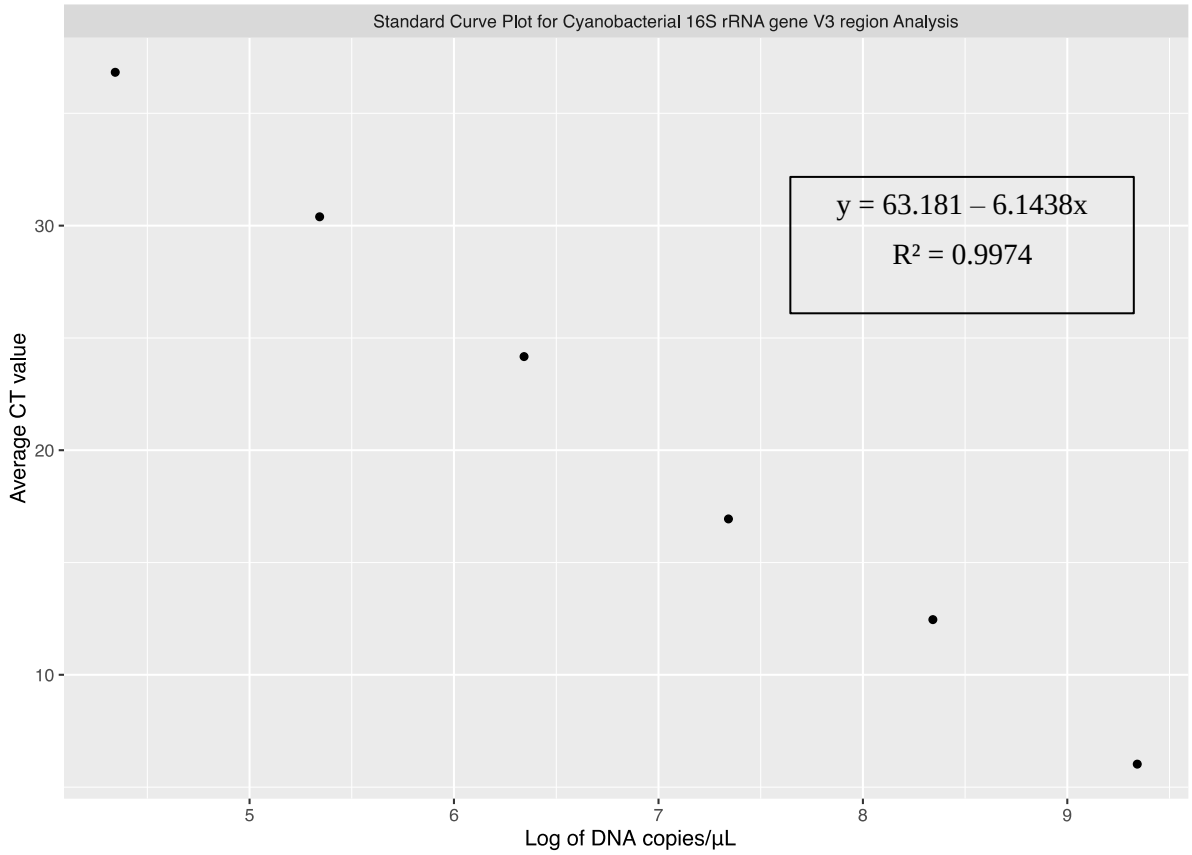


Figure 3.7 Cyanobacterial 16S rRNA gene V3 region amplicon standard curve constructed using known concentrations of DNA copies/µL in a 10-fold dilution series using the Cya-b-F371 and Cya-R783 primer set (Zwart et al., 2005)**.** All samples were conducted in triplicates with the average determination illustrated in the curve.

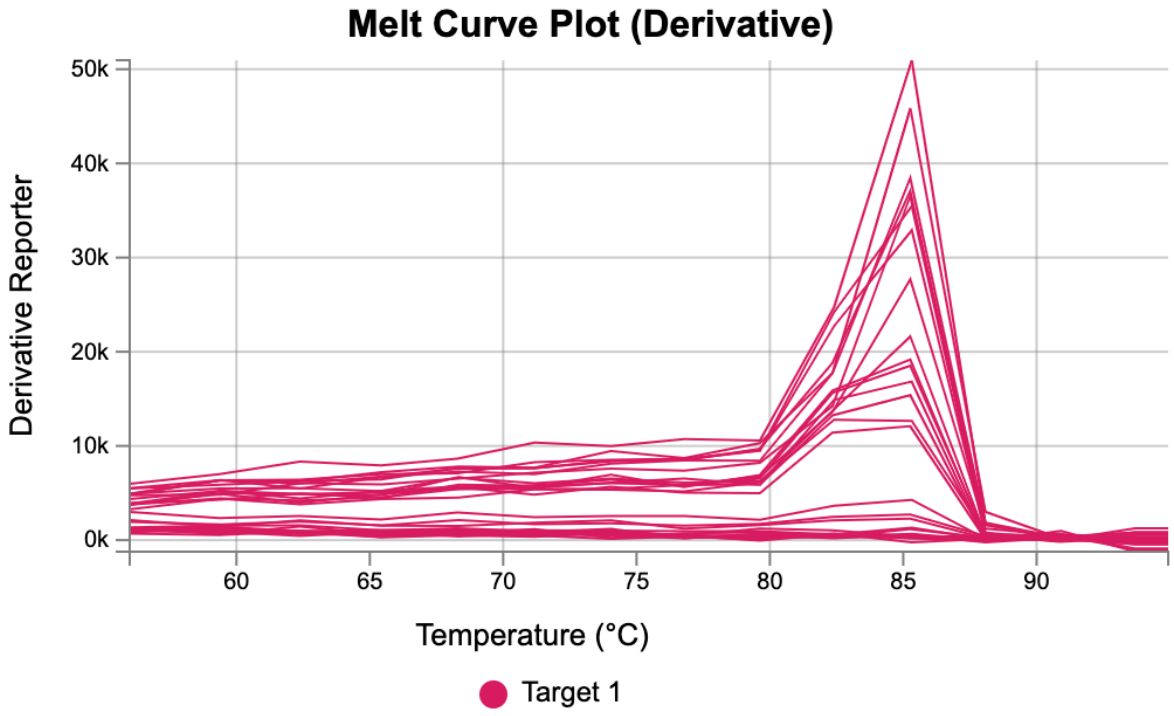


Figure 3.8 Cyanobacterial 16S rRNA gene V3 region amplicon melting curve constructed using known concentrations of DNA copies/µL in a 10-fold dilution series using the Cya-b-F371 and Cya-R783 primer set (Zwart et al., 2005)**.** All samples were conducted in triplicates.

*3.5.2 mcy*E *Gene Standard Curve Analysis*

A standard curve was produced using a dilution series of template DNA sequences of the gene of interests using known concentrations (gBlocks™, Integrated DNA Technologies) (Otten et al., 2016). Six dilutions were conducted to determine abundance of gene copy numbers ranging from 2.08E+08 – 2.08E+01 DNA copies/µL. The HEPF and HEPR primer set (Jungblut & Neilan, 2006) for the AMT region of the *mcy*E gene was used for producing the standard curve. The regression equation was y = 31.545 – 3.2276x with an R^2 = 0.9967

and an efficiency of 118.72% (**Figure 3.9**). In the equation, *y* is threshold cycle (Ct value) at a set fluorescence threshold level of 0.087 and *x* is the amount of DNA that is represented as a $\log_{10}$ DNA copies/µL. A melt curve analysis was conducted from the qPCR products where all outputs illustrated a similar peak at ~83°C (**Figure 3.10**). No template control peaks were present at ~74°C.

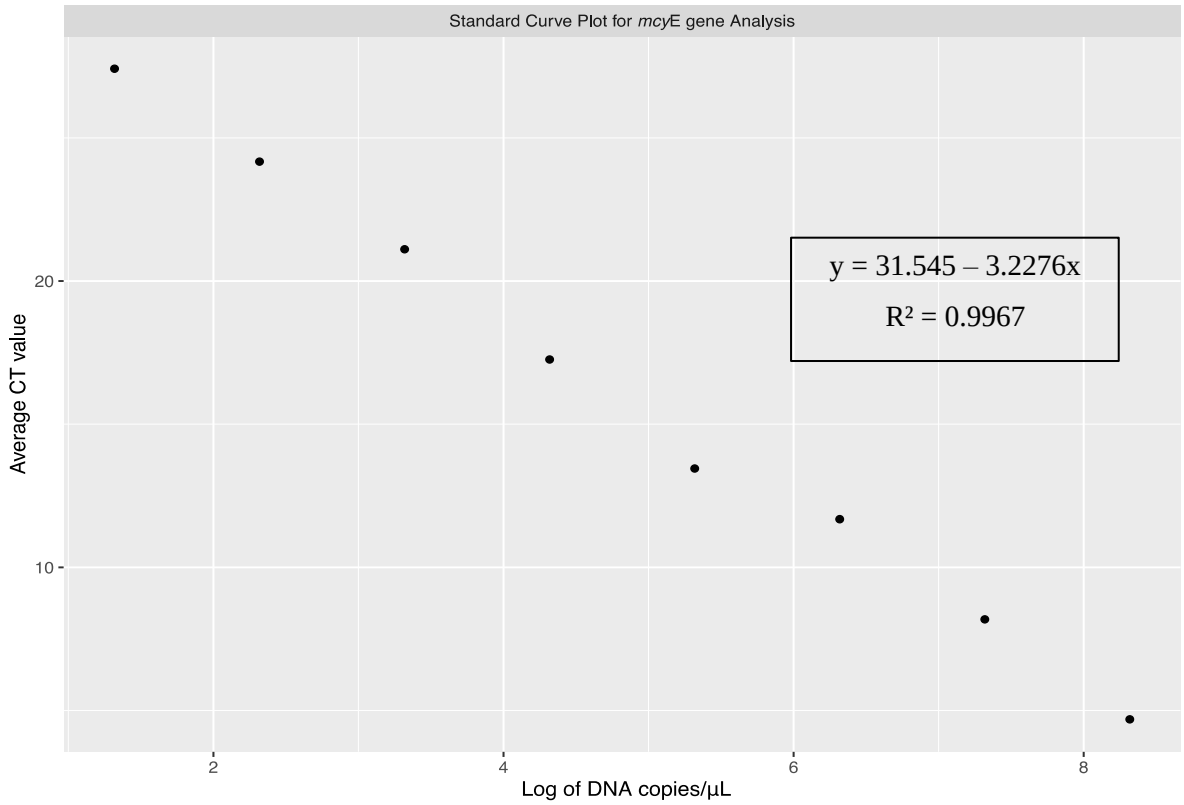


Figure 3.9 *mcy*E gene AMT region amplicon standard curve constructed using known concentrations of DNA copies/µL in a 10-fold dilution series using the HEPF and HEPR primer set (Jungblut & Neilan, 2006). All samples were conducted in triplicates with the average determination illustrated in the curve.

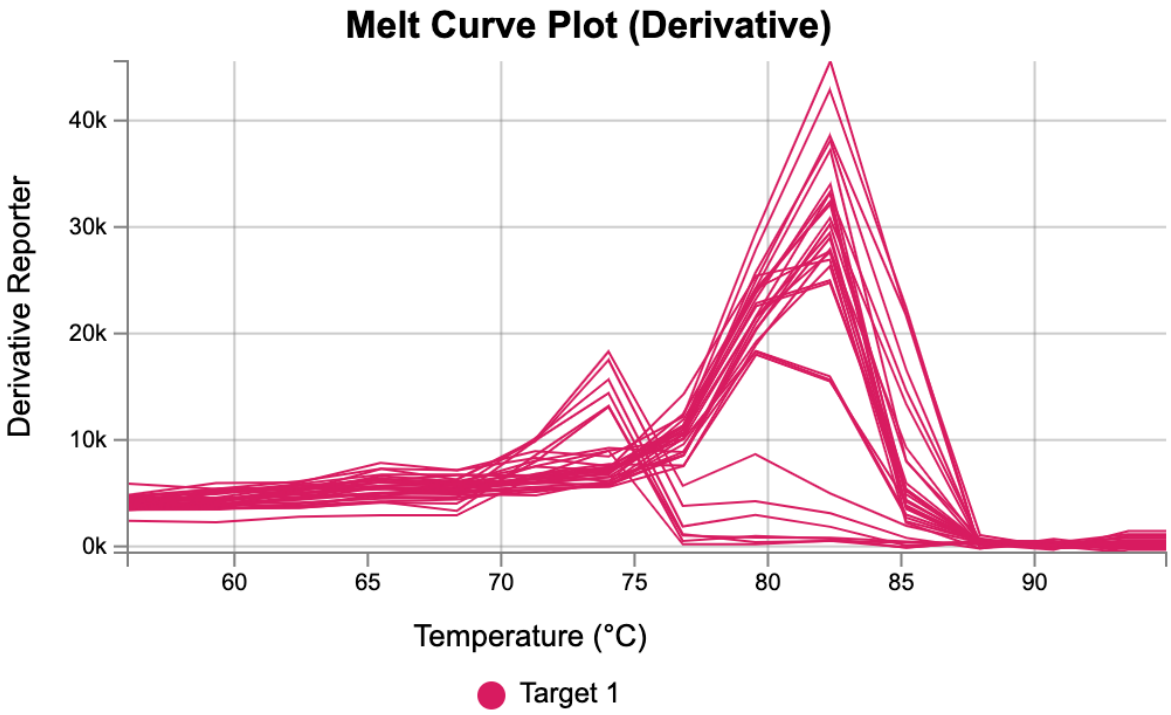


Figure 3.10 Melting curve for the *mcy*E gene amplicon AMT region constructed using known concentrations of DNA copies/µL in a 10-fold dilution series using the HEPF and HEPR primer set (Jungblut & Neilan, 2006). All samples were conducted in triplicates.

3.6 Sample Analysis for Quantitative PCR

Water samples were collected through the months of May to August 2022, filtered for biomass through a GF/C filter, where DNA was extracted, and quantitative PCR (qPCR) was conducted on the samples. Environmental analysis was compared to a standard curve for each gene of interest to compare abundances for all the months analyzed. The cyanobacterial

16S rRNA gene V3 region and the microcystin AMT region of the *mcy*E gene (*mcy*E gene) were compared per month and throughout the summer season. Analysis on DNA copies/µL for the cyanobacterial 16S rRNA gene V3 region abundance were normalized by 4 due to an average gene copy number of four observed per cell (Schirrmeister et al., 2012). Statistical analysis was conducted using a one-way ANOVA ($p < 0.05$) to determine statistical significance at each depth compared to each timepoint in a single day.

3.6.1 qPCR Analysis on Samples Collected for May 2022

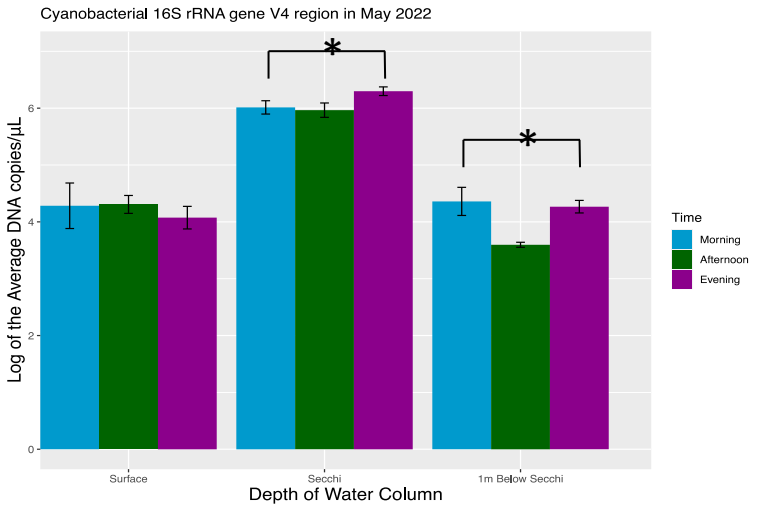
Analysis on samples from the month of May were compared to a standard curve and DNA copies/µL was calculated. Abundances ranged from 3.597 – 6.299 DNA copies/µL for cyanobacterial 16S rRNA gene V3 region and 0.030 – 7.437 DNA copies/µL for *mcy*E (**Table 3.3**). Values for the cyanobacterial 16S rRNA gene V3 region are represented as $\log_{10}$ values due to large amounts of copy number in comparison to the *mcy*E gene of interest. Statistical analysis using a one-way ANOVA illustrated that Secchi and 1m below Secchi were significantly different in the cyanobacteria 16S rRNA gene analysis and all depths were significantly different for the *mcy*E gene analysis. Comparing the different depths sampled at different timepoints, there is variation between abundances for both genes of interest, with an emphasis on the potential toxin producing gene copy number being present as early as May within this oligotrophic system (**Figure 3.11**).

Table 3.3 Quantitative PCR data of total number of cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and total potential microcystin producing cyanobacteria (*mcy*E) represented as DNA copies/µL for May 2022 from samples collected at Little Turkey Lake.

No. of Cyanobacteria (Cyano-16S rRNA gene) and No. of Potential Microcystin Producing Cyanobacteria (*mcy*E) represented as DNA copies/µL for May 2022						
	Morning		Afternoon		Evening	
Depth	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E
Surface	4.284 (± 0.400)	7.437 (± 0.891)	4.307 (± 0.158)	0.195 (± 0.104)	4.075 (± 0.198)	0.985 (±0.091)
Secchi	6.014 (± 0.117)	0.030 (± 0.017)	5.966 (± 0.126)	0.165 (± 0.130)	6.299 (± 0.076)	1.788 (± 0.248)
1m Below Secchi	4.360 (± 0.247)	2.941 (± 0.710)	3.597 (± 0.044)	0.988 (± 0.375)	4.267 (± 0.111)	0.127 (± 0.039)

Abundance was compared to the standard curves for the cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and the *mcy*E gene, respectively. The values represent the average of triplicates ± standard deviation. Values for the cyanobacterial 16S rRNA gene V3 region are represented as $\log_{10}$ values.

A)



B)

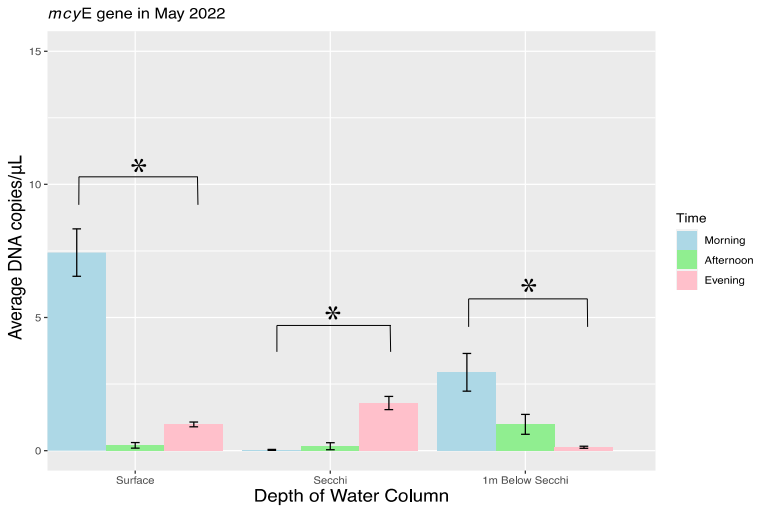


Figure 3.11 Quantitative PCR data of total number of cyanobacteria (16S rRNA gene V3 region) and total potential microcystin producing (*mcy*E) cyanobacteria represented as DNA copies/µL for May 2022 from samples collected at Little Turkey Lake. The values represent the average of triplicates ± standard deviation. A one-way ANOVA ($p<0.05$) was conducted for each depth and statistical significance is outlined with a (*). (A) Graphical illustration of abundance of 16S rRNA gene DNA copies/µL compared to the standard curve for the cyanobacterial 16S rRNA gene V3 region represented as $\log_{10}$ values. (B) Graphical illustration of abundance of *mcy*E gene DNA copies/µL compared to the standard curve for the *mcy*E gene.

3.6.2 qPCR Analysis on Samples Collected for June 2022

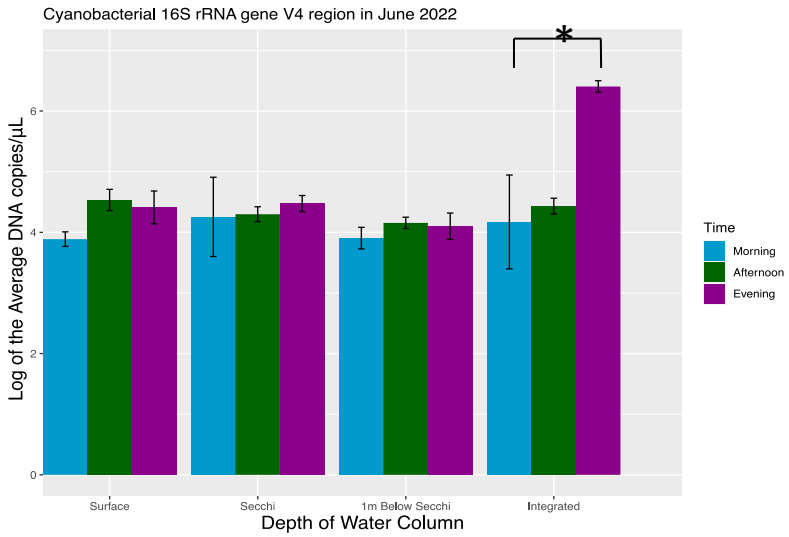
Analysis on samples from the month of June were compared to a standard curve and DNA copies/µL was calculated. Abundances ranged from 3.887 – 6.404 DNA copies/µL for cyanobacterial 16S rRNA gene V3 region and 0.113 – 15.002 DNA copies/µL for *mcy*E (**Table 3.4**). Values for the cyanobacterial 16S rRNA gene V3 region are represented as $\log_{10}$ values due to large amounts of copy number in comparison to the *mcy*E gene of interest. Statistical analysis using a one-way ANOVA illustrated that Integrated depth was significantly different in the cyanobacterial 16S rRNA gene V3 region analysis and Surface and Integrated depths were significantly different for the *mcy*E gene analysis. In comparison, over the two months there is an increase of overall abundance of the genes of interest from May to June with variations between depths and timepoints within a singular day (**Figure 3.12**).

Table 3.4 Quantitative PCR data of total number of cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and total potential microcystin producing cyanobacteria (*mcy*E) represented as DNA copies/µL for June 2022 from samples collected at Little Turkey Lake.

No. of Cyanobacteria (Cyano-16S rRNA gene) and No. of Potential Microcystin Producing Cyanobacteria (*mcy*E) represented as DNA copies/µL for June 2022						
	Morning		Afternoon		Evening	
Depth	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E
Surface	3.887 (± 0.120)	15.002 (± 1.639)	4.534 (± 0.175)	2.828 (± 2.328)	4.410 (± 0.271)	6.466 (± 1.861)
Secchi	4.254 (± 0.654)	2.686 (± 0.809)	4.298 (± 0.123)	4.832 (± 0.609)	4.473 (± 0.133)	5.111 (± 2.265)
1m Below Secchi	3.903 (± 0.178)	5.200 (± 1.434)	4.155 (± 0.093)	5.909 (± 0.940)	4.101 (± 0.217)	4.127 (± 0.514)
Integrated	4.171 (± 0.774)	5.930 (± 0.777)	4.432 (± 0.131)	4.916 (± 1.607)	6.404 (± 0.096)	9.616 (± 4.358)

Abundance was compared to the standard curves for the cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and the *mcy*E gene, respectively. The values represent the average of triplicates ± standard deviation. Values for the cyanobacterial 16S rRNA gene V3 region are represented as $\log_{10}$ values.

A)



B)

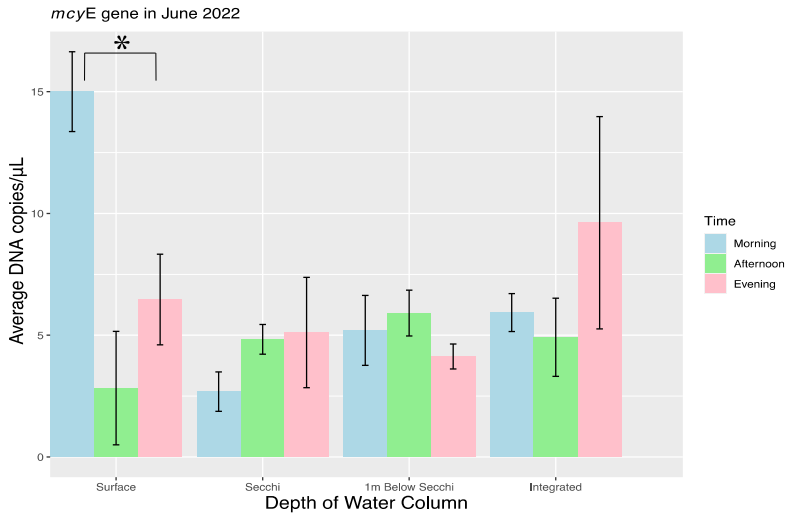


Figure 3.12 Quantitative PCR data of total number of cyanobacteria (16S rRNA gene) and total potential microcystin producing (*mcy*E) cyanobacteria represented as DNA copies/µL for June 2022 from samples collected at Little Turkey Lake. The values represent the average of triplicates ± standard deviation. A one-way ANOVA ($p<0.05$) was conducted for each depth and statistical significance is outlined with a (*). (A) Graphical illustration of abundance of 16S rRNA gene DNA copies/µL compared to the standard curve for the cyanobacterial 16S rRNA gene V3 region represented as $\log_{10}$ values. (B) Graphical illustration of abundance of *mcy*E gene DNA copies/µL compared to the standard curve for the *mcy*E gene.

3.6.3 qPCR Analysis on Samples Collected for July 2022

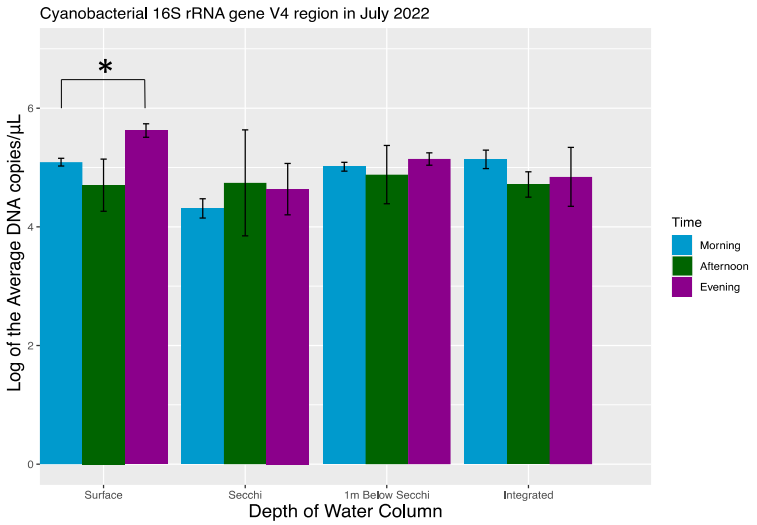
Analysis on samples from the month of June were compared to a standard curve and DNA copies/µL was calculated. Abundances ranged from 4.311 – 5.624 DNA copies/µL for cyanobacterial 16S rRNA gene V3 region and 0.238 – 10.732 DNA copies/µL for *mcy*E (**Table 3.5**). Values for the cyanobacterial 16S rRNA gene V3 region are represented as a $\log_{10}$ value due to large amounts of copy number in comparison to the *mcy*E gene of interest. Statistical analysis using a one-way ANOVA illustrated that Surface depth was significantly different in the cyanobacterial 16S rRNA gene V3 region analysis and all the depths were significantly different for the *mcy*E gene analysis. There is an increase of overall abundance of the cyanobacterial 16S rRNA gene V3 region from May and June to July with variations between depths and timepoints within a singular day. For the month of July, there was an overall decrease of abundance from June to July for the *mcy*E gene, but there was still a large variation between the depths and timepoints (**Figure 3.13**).

Table 3.5 Quantitative PCR data of total number of cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and total potential microcystin producing cyanobacteria (*mcy*E) represented as DNA copies/µL for July 2022 from samples collected at Little Turkey Lake.

No. of Cyanobacteria (Cyano-16S rRNA gene) and No. of Potential Microcystin Producing Cyanobacteria (*mcy*E) represented as DNA copies/µL for July 2022						
	Morning		Afternoon		Evening	
Depth	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E
Surface	5.090 (± 0.066)	9.970 (± 2.377)	4.703 (± 0.440)	2.185 (± 0.595)	5.624 (± 0.113)	10.732 (± 2.510)
Secchi	4.311 (± 0.163)	0.373 (± 0.128)	4.741 (± 0.894)	0.860 (± 0.403)	4.635 (± 0.433)	1.087 (± 0.068)
1m Below Secchi	5.012 (± 0.075)	0.819 (± 0.263)	4.880 (± 0.491)	0.238 (± 0.175)	5.143 (± 0.103)	0.457 (± 0.216)
Integrated	5.138 (± 0.155)	1.528 (± 0.111)	4.715 (± 0.214)	0.789 (± 0.175)	4.842 (± 0.496)	5.366 (± 0.935)

Abundance was compared to the standard curves for the cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and the *mcy*E gene, respectively. The values represent the average of triplicates ± standard deviation. Values for the cyanobacterial 16S rRNA gene V3 region are represented as $\log_{10}$ values.

A)



B)

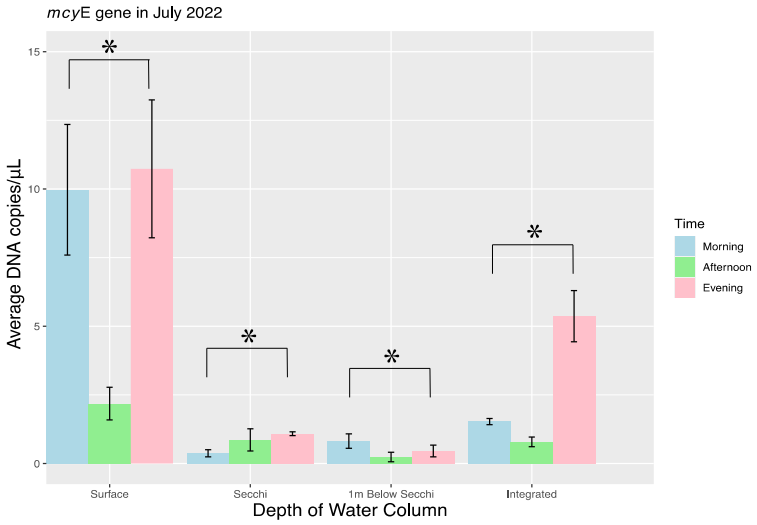


Figure 3.13 Quantitative PCR data of total number of cyanobacteria (16S rRNA gene) and total potential microcystin producing (*mcy*E) cyanobacteria represented as DNA copies/µL for July 2022 from samples collected at Little Turkey Lake. The values represent the average of triplicates ± standard deviation. A one-way ANOVA ($p<0.05$) was conducted for each depth and statistical significance is outlined with a (*). (A) Graphical illustration of abundance of 16S rRNA gene DNA copies/µL compared to the standard curve for the cyanobacterial 16S rRNA gene V3 region represented as $\log_{10}$ values. (B) Graphical illustration of abundance of *mcy*E gene DNA copies/µL compared to the standard curve for the *mcy*E gene.

3.6.4 qPCR Analysis on Samples Collected for August 2022

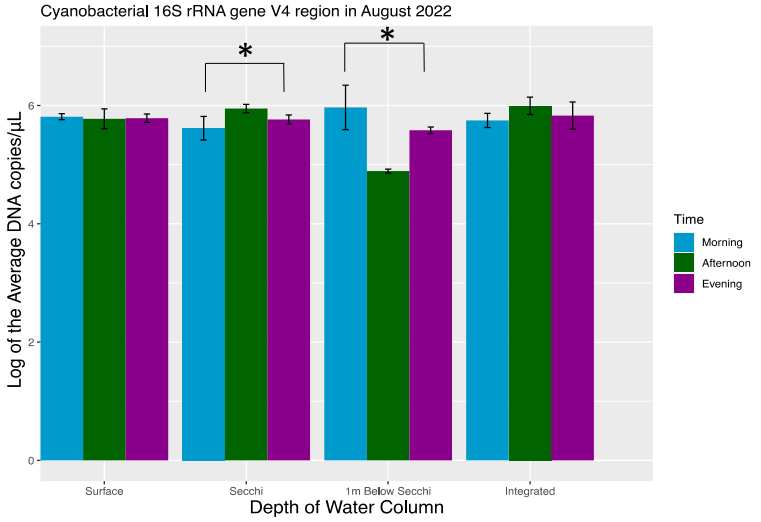
Analysis on samples from the month of June were compared to a standard curve and DNA copies/µL was calculated. Abundances ranged from 4.892 – 5.996 DNA copies/µL for cyanobacterial 16S rRNA gene V3 region and -0.300 – 2.786 DNA copies/µL for *mcy*E (**Table 3.6**). Values for the cyanobacterial 16S rRNA gene V3 region and the *mcy*E gene are represented as $\log_{10}$ values due to large amounts of copy number in comparison to the other months, resulting in a change of scale for the graphical analysis. Statistical analysis using a one-way ANOVA illustrated that Surface depth was significantly different in the cyanobacterial 16S rRNA gene V3 region analysis and all the depths were significantly different for the *mcy*E gene analysis. There is an increase of overall abundance of the cyanobacterial 16S rRNA gene V3 region from May to August with variations between depths and timepoints within a singular day. For the month of August, there was an overall increase in abundance for both genes with a large variation between the depths and timepoints (**Figure 3.14**).

Table 3.6 Quantitative PCR data of total number of cyanobacteria 16S rRNA gene V3 region (Cyano-16S rRNA gene) and total potential microcystin producing cyanobacteria (*mcy*E) represented as DNA copies/µL for August 2022 from samples collected at Little Turkey Lake.

No. of Cyanobacteria (Cyano-16S rRNA gene) and No. of Potential Microcystin Producing Cyanobacteria (*mcy*E) represented as DNA copies/µL for August 2022						
	Morning		Afternoon		Evening	
Depth	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E	Cyano-16S rRNA gene	*mcy*E
Surface	5.812 (± 0.051)	2.160 (± 0.021)	5.775 (± 0.169)	2.639 (± 0.133)	5.785 (± 0.071)	2.786 (± 0.146)
Secchi	5.617 (± 0.199)	0.905 (± 0.022)	5.949 (± 0.070)	0.327 (± 0.047)	5.765 (± 0.075)	0.540 (± 0.117)
1m Below Secchi	5.969 (± 0.376)	1.177 (± 0.160)	4.892 (± 0.032)	-0.300 (± 0.137)	5.581 (± 0.054)	0.411 (± 0.104)
Integrated	5.749 (± 0.120)	1.880 (± 0.049)	5.996 (± 0.147)	1.733 (± 0.123)	5.831 (± 0.230)	1.019 (± 0.094)

Abundance was compared to the standard curves for the cyanobacterial 16S rRNA gene V3 region (Cyano-16S rRNA gene) and the *mcy*E gene, respectively. The values represent the average of triplicates ± standard deviation. Values for the Cyano-16S rRNA gene and the *mcy*E gene are represented as $\log_{10}$ values.

A)



B)

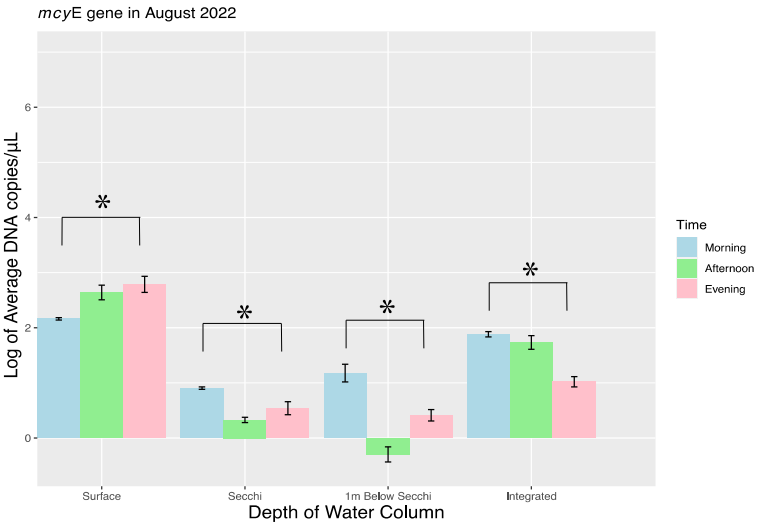


Figure 3.14 Quantitative PCR data of total number of cyanobacteria (16S rRNA gene) and total potential microcystin producing (*mcy*E) cyanobacteria represented as DNA copies/µL for August 2022 from samples collected at Little Turkey Lake. The values represent the average of triplicates ± standard deviation. A one-way ANOVA ($p<0.05$) was conducted for each depth and statistical significance is outlined with a (*). (A) Graphical illustration of abundance of 16S rRNA gene DNA copies/µL compared to the standard curve for the cyanobacterial 16S rRNA gene V3 region represented as $\log_{10}$ values. (B) Graphical illustration of abundance of *mcy*E gene DNA copies/µL compared to the standard curve for the *mcy*E gene.

Chapter 4 – Discussion

4.1 The Impact of a Changing Bacterial Community in an Oligotrophic Freshwater Forested Watershed

Safe drinking water is important for public health and the implementation of water security is an ecological and environmental issue that consistently requires attention (Zhou et al., 2021). Forested watersheds provide usable water supply that is high quality and quantity, including drinking water, where land disturbances (e.g., wildfire, agricultural development) can have large effects on the source water quality and drinking water treatment (Emelko et al., 2011). Turkey Lakes Watershed (Ontario, Canada) is an example of a forested watershed that is the model study system for analysis on the effect of the natural environment on a freshwater oligotrophic system due to it being an untouched water source with minimal human recreation and urban development around it and has decades of research associated with it (Jeffries et al., 1988; Webster et al., 2021). Within the Turkey Lakes Watershed, there have been multiple studies focused on forestry, water chemistry, hydrological systems, and other ecosystem analyses, with minimal research on the microbial communities present in this water source (Webster et al., 2021).

4.1.1 Bacterial Relative Abundance Analysis in an Oligotrophic Watershed

Microbial pollution directly impacts the deterioration of drinking water quality caused by the proliferation of infectious, toxin and odour producing microorganisms that can be transmitted through diet, aerosols, and contact which can endanger human health (Zhou et

al., 2021). Typically, culture-based methods are used to analyze microbial quantity in drinking water, but with the wide range of non-culturable bacteria, this is an underestimation of what is present in the water system (Zhou et al., 2021). Culture-independent methods like 16S rRNA gene sequencing and metagenomic surveys are useful to detect lower concentrations of microorganisms, with the advantage of obtaining microbial diversity data in drinking water systems (Van Assche et al., 2019; Zhou et al., 2021). The diversity and communities are sensitive to differences in the environment and can change the structure and function of the aquatic ecosystem, where different water sources and implementation of various sampling times may have different compositions of microbial communities (Zhou et al., 2021). This research looked at expanding previous work conducted Dr. Ellen Cameron by what was previously reported at Little Turkey Lake through the incorporation of multiple depth sampling (Secchi and integrated depths) at multiple timepoints in a singular day, over four months in one year (Cameron, 2021).

Analysis of bacterial community composition of the Turkey Lakes Watershed, with a focus on Little Turkey Lake, from May to August 2022 was conducted through 16S rRNA gene sequencing and sequencing reads for each month illustrated specific differences in abundance at various timepoints in a singular day for total reads. The most abundant phylum detected across all samples was Proteobacteria and was observed across all depths and timepoints in during a single day (**Figure 3.1**). Proteobacteria was noted to be the most abundant in samples collected at Little Turkey Lake in 2019 as well (Cameron, 2021). In drinking water reservoirs, this phylum has also been observed to be the most abundant

(Bautista-De los Santos et al., 2016; Feng et al., 2019; Van Assche et al., 2019). In addition, Proteobacteria are also the most dominant phylum in reference 16S rRNA gene databases as some genera within this phylum include opportunistic pathogens which can be transmitted by water, like *Achromobacter* and *Burkholderia* (Bautista-De los Santos et al., 2016; Ferro et al., 2019). Analyzing the community composition and the observance of Proteobacteria as a dominant phylum that contains opportunistic pathogens puts into perspective the importance of knowing bacterial diversity, especially with monitoring and treatment processes of drinking water sources.

In addition, other studies have observed the bacterial community composition of surface waters to consist of large percentages of other phyla, including Actinobacteriota, Firmicutes, Cyanobacteria, and Bacteroidetes (Davenport et al., 2019; Guan et al., 2018; Van Assche et al., 2019). This was also noted in Little Turkey Lake in 2022 where the phyla Bacteroidetes, Actinobacteriota, and Verrucomicrobiota were also observed across all samples and timepoints. An overall community composition was previously reported by Dr. Ellen Cameron at the Turkey Lakes Watershed observed that the Proteobacteria, Bacteroidota, Actinobacteriota, and Cyanobacteria were the main phyla present throughout the water column and in the month of August 2019 (Cameron, 2021). This holistic view of the same sampling site after two years has provided an in-depth analysis of the bacterial community composition and the variation based on time and depth of sampling.

For environmental based studies, typically eutrophic water systems are studied (Cakmak et al., 2022; Feng et al., 2019; Yang et al., 2008), where it is essential that other

trophic systems are further analyzed and can help with providing valuable information for protection and remediation of the lakes important for providing safe drinking water. A study conducted by Feng et al., 2019 compared the differences in community composition between eutrophic and oligotrophic lakes, where composition varied based on trophic type through 16S rRNA gene sequencing using Illumina MiSeq (Feng et al., 2019). In the oligotrophic water system, the most abundant phylum was Proteobacteria, with Chloroflexi, Acidobacteria, Actinobacteria, and Bacteroidetes also dominating the system (Feng et al., 2019), similar to the community composition observed at Little Turkey Lake. These orders were also observed in eutrophic water systems emphasizing the importance of this overlap between communities in eutrophic and oligotrophic systems where conducting further research on oligotrophic systems and their environmental factors that help shape the taxon composition is required (Ji et al., 2019).

4.1.2 Bacterial Taxonomic Richness and Evenness Variation based on Sampling Times

Understanding the diversity of microbial organisms in drinking water sources helps with identifying the potential microorganisms that may pose a risk to public health, where multiple techniques of disinfection is an effective strategy to employ if required based on community and diversity (Jing et al., 2021). Determining sampling richness and evenness throughout the samples collected helps with analyzing taxonomic diversity of each sample and if variation occurs to evaluate the impact of sampling time of day and month on species diversity (Aloraini et al., 2023). This approach provides a more in-depth analysis of each sampling point within samples to compare the potential of more sequence variants (Willis,

2019). This can be analyzed through alpha diversity and Shannon diversity index where samples with higher taxonomic diversity typically have higher Shannon diversity indices (Tiwari et al., 2021).

Alpha diversity of samples collected from Little Turkey Lake illustrated variations of community composition between the months of May to August 2022 and sampling times within each day. The Shannon index was used in this study following library normalization to observe the diversity indices with the samples The normalization technique covers potential biases that may arise from smaller sample sizes, and with a larger sample size for bacterial community composition, diversity in all the samples can be observed (Cameron et al., 2021). Through the diversity variation at different points, changes in sampling efforts and sample analysis are required when analyzing each unique drinking water source, along with updating monitoring protocols.

A study conducted by Zheng et al., illustrated that freshwater Shannon indices range from 4.16 to 4.55 for bacterial diversity (Zheng et al., 2022). Similar to this study, the observed alpha diversity range observed with the samples are observed within the range of 4 to 5 for the Shannon index, illustrating increased taxonomic diversity between the ASVs (Tiwari et al., 2021; K. E. Walters & Martiny, 2020). Variations in diversity among the months were observed as each sample at the various timepoints were present based on the reads. Similar values as observed in **Figure 3.4** were also noted in previous studies conducted on alpha diversity at Little Turkey Lake (Cameron, 2021). The diversity index changed in May from the morning sampling timepoint to the evening timepoint decreased

with the index but increased in August. Between the months of June and July, the diversity index fluctuates from 4.0 – 4.5 throughout the timepoints, providing evidence that no sampling timepoint can be predicted to increase or decrease, and monitoring should consistently be conducted. The change in alpha diversity throughout timepoints in a day and between sampling months further emphasizes the importance of multiple sampling times and days to understand the water system that is being studied and what organisms are present at the various points.

4.1.3 The Impact of Environmental Parameters on Bacterial Diversity

Ecological parameters affect bacterial diversity in different water systems, an example being surface waters experiencing changes in temperature and light in the different seasons of the year, where focus on changes in taxonomic and functional diversity helps with understanding each unique water system (Tiwari et al., 2021). Significant relationships between the environmental variables and taxonomy of samples were analyzed at different sampling timepoints for the months of May to August 2022 at Little Turkey Lake. A redundancy analysis (RDA) was conducted on physiochemical variables like temperature, pH, total dissolved solids, and conductivity of the sample collected before filtering to test for effects on the bacterial reads and composition (Nwosu et al., 2021).

Lake stratification plays a role in the transition between colder and warmer layers of the water column, where with increasing global temperatures, there may be changes in the periods of stratification which may be extended and changes the structure of the bacterial community composition affected by temperature (Nwosu et al., 2021). Similarly, for pH, the

stratification can play a role is increased or decreased pH that may be suitable for proliferation of the organisms (Nwosu et al., 2021). The samples that were observed in the center of the plot may be equally driven by the environmental factors, or there may be other variables that are drivers for the change in community, outside the scope of this research requiring further analysis (Paliy & Shankar, 2016). Most of the samples were not significantly impacted by the variables measured as they fell in the middle of the graph analysis, however temperature and pH did significantly impact the July evening and August morning sample diversity (**Figure 3.6**). Previous studies conducted have illustrated the impact of pH level changes and temperature being major drivers on bacterial community structure (S. Liu et al., 2015; M. Zhang et al., 2012). The RDA analysis has illustrated the unpredictability of sampling and how environmental parameters may or may not play a role in bacterial dispersion.

Overall, there is similarity illustrated with relative abundances and diversity of bacterial community composition as similar phyla are present throughout the freshwater, oligotrophic system, but variation occurs throughout the season with different phyla emerging and relative abundances changing. Different source water quality, trophic levels, and community composition plays a role in the application of drinking water treatment processes where the biological diversity of the drinking water source can affect the microbial diversity of the drinking water after the biological treatment technology (Zhou et al., 2021). This puts emphasis on the details required in sampling and monitoring protocols of water systems due to variation in bacterial community composition based on time of day and

season, where each water system is unique and requires specific protocols based on the different systems, that can vary based on sampling time. Many water-borne pathogens can go undetected due to lack of easy and reliable methods, where the inclusion of sequencing and molecular analysis can provide an in-depth review of what is observed in the water (Saxena et al., 2015). Through this, monitoring programs and treatment processes vary based on the system and each system requires different analysis to fully understand the presence and diversity of its microorganisms.

4.2 Quantitative PCR Analysis to Quantify the Presence of Cyanobacterial Gene and Microcystin Gene Copy Numbers

Harmful cyanobacteria are present in drinking water sources and are involved in deteriorating drinking source water quality (Chiu et al., 2017). The use of many molecular techniques has been developed and applied to the toxigenic genes and cyanotoxins to complete multiple sample analysis in an effective and efficient manner (Chiu et al., 2017). The development of this efficient molecular tool to determine the potential cyanobacterial conserved 16S rRNA gene abundance and hepatotoxin toxin-producing gene presence provides insight into determining if there is variation of samples collected and how that plays a role in understanding the water system.

4.2.1 Standard Curve Analysis

Quantitative PCR (qPCR) was used to determine the abundance of cyanobacterial and hepatotoxic gene copy numbers were present in samples collected from May to August 2022

at Little Turkey Lake (**Figure 3.7 & 3.9**). Using a dilution series of template DNA sequences for the genes of interest, a standard curve was obtained to compare to environmental samples (Otten et al., 2016). PCR amplification efficiency was below the ideal range for the cyanobacterial 16S rRNA gene V3 region analysis and above for the ideal range for the *mcy*E analysis (Chiu et al., 2017). However, the quantification of the gene copy number was still possible due to the standard curves having high R^2 values that illustrates good correlation between the DNA template and Ct values (Chiu et al., 2017).

For the cyanobacterial 16S rRNA gene and the *mcy*E analysis, all samples illustrated one peak at 85 °C for the melting curve where no primer peaks were observed in the samples with DNA templates, but there were small primer-dimer peaks in the negative control (NTC) samples (**Figure 3.8 & 3.10**), which were smaller in comparison to the template samples (Müller et al., 2017). With biological bias, copy number does vary in cyanobacteria, where some may contain up to five 16S rRNA gene copies, and the high copy number can lead to overestimation in PCR-based methods (MacKeigan et al., 2022). Throughout the analysis, this has been accounted for during the qPCR analysis to find an overall normalization between the samples. Previous studies have noted that within the cyanobacterial genome, there is only one cluster of the *mcy*E gene and therefore does not require the normalization (Müller et al., 2017). PCR products were run on an agarose gel where expected sequence lengths were obtained to further support the evidence that target sequences were amplified.

Quantitative PCR (qPCR) analysis was conducted on samples from Little Turkey Lake from the months of May 2022 to August at various timepoints in a day and at various

depths within the water column. This provides an analysis on specific gene copy number counts of cyanobacterial organisms present within the water column, and potential hepatotoxin producing organisms, which was compared to the standard curve equations obtained (Müller et al., 2017; Zwart et al., 2005).

*4.2.2 Cyanobacterial 16S rRNA and mcy*E *Gene Quantification*

Analysis conducted for the month of May 2022 included the depths of surface, Secchi, and 1 meter below Secchi. On average, the gene copy numbers for the cyanobacterial 16S rRNA gene were higher at Secchi depth with peaks reached for the evening timepoint and on average, higher at surface level for the *mcy*E gene with peaks reached at the morning timepoint (**Figure 3.11 A&B**). For the months of June to August 2022, analysis was conducted for surface, Secchi, 1 meter below Secchi, and an integrated sample of the water column. On average for June 2022, the gene copy numbers for the cyanobacterial 16S rRNA gene were higher in the integrated sample with peaks reached at the evening timepoint and on average, higher at surface level for the *mcy*E gene with peaks reached at the morning timepoint (**Figure 3.12 A&B**). On average for July 2022, the gene copy numbers for the cyanobacterial 16S rRNA gene were higher in the surface sample with peaks reached at the evening timepoint and on average, higher at surface level for the *mcy*E gene with peaks reached at the evening timepoint (**Figure 3.13 A&B**). On average for August 2022, the gene copy numbers for the cyanobacterial 16S rRNA gene were similar at all depths with peaks reached at all timepoints and on average, higher at surface and integrated depths for the *mcy*E gene with peaks reached at the morning and evening timepoints (**Figure 3.14 A&B**).

Overall, for August 2022, there was a significant increase in gene copy number for both analyses, similar to what was observed in the relative abundances analyzed for Little Turkey Lake.

The analysis on abundance for gene copy number presence has provided evidence of re-emergence of cyanobacteria at various depths in the water column can occur as early as May in the season, with an increase of abundance throughout the season in July and August. With the incorporation of an integrated sample, this provides an effective and efficient way to sample and analyze the water system without visible blooms as the numbers at the various depths are indicative of the presence of cyanobacteria at all depths of the water column. Similar trends were observed between the months as abundance changed between timepoints, and depths. As cyanobacteria are involved in diurnal migration patterns to find advantageous conditions to proliferate, the change in abundance at the different depths and timepoints are not surprising (Gray et al., 2019).

Overall, with qPCR analysis, the higher presence of gene copy numbers for both cyanobacterial and potential hepatotoxin producing organisms illustrates the importance of analyzing water systems thoroughly. Even without a visible bloom, potential bloom-forming harmful cyanobacteria may be present in the system earlier in the season and throughout the column, with variations in a singular day. Through this, monitoring protocols need to be updated based on each water system to understand what is present and in what quantity.

4.3 Taxonomy of Cyanobacteria in an Oligotrophic Freshwater Forested Watershed

Cyanobacteria are photosynthetic prokaryotes that can be referred to as blue-green algae, and include a wide range of morphology types like unicellular, filamentous colony-and mat-forming genera and that inhabit terrestrial and aquatic habitats (Mazard et al., 2016). As previously mentioned in section 4.1, the analysis of these bacterial organisms plays a role in understanding the freshwater system and the impact on the deterioration of drinking water quality due to their proliferation (Zhou et al., 2021). Similar to bacterial community composition, the analysis of cyanobacteria and their dominance in freshwater systems provides information for monitoring protocols and drinking water treatment as these organisms can produce toxins that affect human health, and more attention and research is required to understand them (Mazard et al., 2016; Zhou et al., 2021).

4.3.1 Analysis of Cyanobacterial Relative Abundance in a Forested Watershed

Bacterial community composition from surface waters have been widely studied, with a focus on cyanobacterial populations at the immediate subsurface, or sediment for cyanobacterial experimental research and monitoring protocols (Rogers et al., 2021; Van Assche et al., 2019). Relative abundance analysis conducted at the Turkey Lakes Watershed provides information on dominant phyla based on sampling time and depth, similar to information gathered in section 4.1. In Little Turkey Lake, the relative abundance for cyanobacteria varied with a large observation of reads in the months of July and August at all timepoints, with smaller number of reads observed in May and June. With the initiation of

snowmelt and the impact of climate warming on temperature in early May at Turkey Lakes Watershed (Creed et al., 2015), this impacts the emergence of cyanobacteria earlier in the season and provides opportunity for potential toxic bloom formation. Cyanobacterial emergence and dominance are associated with nutrient availability, but the presence earlier in the season has been connected to the climatic shifts that have occurred through warmer water temperature, giving these organisms a competitive advantage over others (Larsen et al., 2020). The presence of cyanobacteria in the earlier season has illustrated that there has been a potential impact of warming climates changing seasons on this oligotrophic northern lake and its microbial ecosystems.

Integrated depth sampling was conducted throughout the months of June – August at Little Turkey Lake. The Secchi depth for the month of May was analyzed to determine community composition below surface due to integrated sampling initiating in June 2022, giving insight into multiple depth sampling and presence of cyanobacterial reads below surface level. This provides insight into the vertical variation of microbial abundances and allows for monitoring of the entire water column measured outside of just surface water sampling to fully characterize community composition (Selbig et al., 2012). This approach to sampling can increase the probability of detection of potential toxin producing organisms as only surface populations do not comprise of the entire community composition. Integrated depth sampling also covers the effects of direct water column stratification that occurs in Little Turkey Lake from mid-May to October that changes light and nutrient environments for these organisms (Gray et al., 2019; Jeffries et al., 1988). Variation in phytoplankton and

its abundance at various levels of the water column can be sampled and studied through integrated sampling, where sinking phytoplankton are not overshadowed by dominant buoyant cyanobacteria and easily detected (Gray et al., 2019; Huisman et al., 2018).

Within a sampling month, there was variation of cyanobacterial reads (**Table 3.2**) within a single day based off three different timepoints. The measurement of various timepoints can help with temporal resolution and high frequency monitoring of the water system, important for detection of potential toxin producing cyanobacteria (Huisman et al., 2018; Vilmin et al., 2018). Similar to integrated depth sampling, the time of day used for sampling can be important for collecting a representative sample of the water column. The changes in cyanobacterial distribution due to migrations within the water column will be neglected if only minimal sampling is conducted at the surface, once a day, or with a visible bloom (Gray et al., 2019). Through the incorporation of multiple depth and timepoint sampling, cyanobacterial community composition and relative abundance was analyzed with emphasis provided on earlier emergence of cyanobacteria, and the importance of variation in the water column throughout the season.

4.3.2 Cyanobacteria Diversity Among Months and Sampling Times

The alpha diversity of samples collected from Little Turkey Lake were observed to determine potential variations of cyanobacterial community composition between the months of July to August 2022 and sampling times within each day. The normalization technique was also performed for the cyanobacterial diversity, similar to the bacterial diversity analysis, and used a rarefied library size of 4500 that excluded the months of May and June due to

lower number of reads. This covers potential biases that may arise from smaller sample sizes and can avoid the bias that larger samples can appear to be more diverse due to larger volumes of reads, where cyanobacterial community composition diversity in most of the samples can be observed (Cameron et al., 2021). The observed alpha diversity range observed within the samples were observed to be in the range of 1.5 to 2.5 for the Shannon index, illustrating decreased diversity between the reads in comparison to bacterial diversity (K. E. Walters & Martiny, 2020). The change in July from the morning sampling timepoint to the evening timepoint increases with the index but fluctuates in August as the value for the afternoon index goes below 2.0 and rises above 2.0 for the evening timepoint. Cyanobacterial indices also fell within this range from samples collected at Little Turkey Lake previously, illustrating consisted diversity, but variations at the different timepoints (Cameron, 2021). Similar to bacterial diversity, this provides evidence that no sampling timepoint can be predicted to increase or decrease and change in number of reads may affect the diversity of the sample (Cameron et al., 2021).

4.3.3 Analysis of Cyanobacterial Community Composition using a Maximum Likelihood Phylogenetic Tree

Morphological methods of cyanobacterial taxonomic classification are insufficient tools for the detection, profiling, and identification of genetic diversity, where the use of 16S rRNA gene sequencing and phylogenetics helps advance the molecular methods for cyanobacterial and cyanotoxin research (Komárek, 2016; Moreira et al., 2013). The combination of molecular gene targets, morphological and biochemical potential leads to

establishing classification systems that ultimately present as monophyletic groupings to illustrate taxonomy that correlates with phylogeny (Cheng et al., 2022; Strunecký et al., 2023). A maximum likelihood phylogenetic tree was constructed to observe evolutionary relationships between the obtained cyanobacterial sequences to reference sequences provided through online databases (**Figure 3.3**). These published reference sequences helped with assigning clusters of taxonomy to determine what sequences were similar to the classifications. The phylogenetic tree consists of all sequenced samples and identified genomes from the May to August 2022 sampling season at Little Turkey Lake (Table A1). Within the tree, Synechococcales, Pseudanabaenales, and Chroococcales were the most abundant cyanobacterial orders due to a larger number of observed amplicon sequence variants (ASVs).

Within the phylogenetic tree, the ASVs associated with Synechococcales is the most abundant with a variety of well-supported ASVs that clustered together with reference sequences. The reference sequences obtained and observed in the tree are dominated by *Synechococcus* and *Cyanobium*, which are picocyanobacteria that take up about 50 – 75% of cyanobacterial community composition relative abundances in mesotrophic and oligotrophic lake systems (MacKeigan et al., 2022). The diversity of the genera in freshwater systems is much lower in comparison to the genomes sequenced for marine strains and studying the evolution of these organisms helps with filling gaps in the genetic diversity research (Cabello-Yeves et al., 2017). In most clusters associated with the tree, there are groups of ASVs that link to one or two reference sequences, or none at all, and this is observed

throughout the tree within this cyanobacterial order. By regularly isolating, identifying, sequencing and updating reference libraries, more representative species of picocyanobacteria can be resolved through the increase of more readily available genetic information (MacKeigan et al., 2022). The clustering of the unknown AVS is valid and significant through the higher bootstrap values, but further research and sequencing is required to identify and present the evolutionary significance of the unknown ASVs.

Synechococcales are typically cocci cyanobacteria with parietal thylakoids with polyphyletic lineages (Strunecký et al., 2023). Many of the reference genera consist of picocyanobacteria, which are globally observed in marine systems and are poorly studied at the genomic level in freshwater systems in comparison to the marine environment (Cabello-Yeves et al., 2017). These picocyanobacteria range from 0.2-3 μm oval to elongated cells and are single or colonial rod-shaped cells that can form dense populations and the dominant (Cabello-Yeves et al., 2017; Komárek et al., 2020; Strunecký et al., 2023). These organisms have advantageous adaptations like floating ability, increased growth rates, and tolerance of lower light levels and higher temperatures (Sim et al., 2023). Within this order, there are more than 100 genera that were classified based on simple morphology, but through the inclusion of 16S rRNA gene sequencing, 40 newly established genera with 88 species were added (Strunecký et al., 2023).

The order Nostocales contains well-defined colourless cells that are specialized for dinitrogen fixation through the formation of heterocytes (Strunecký et al., 2023). Spore-like cells are formed by many genera and some cyanobacteria within this order are filamentous

where multiple traits like branching, polarity, position of heterocytes, and other characteristics provide an array of phenotypic diversity in comparison to other orders (Strunecký et al., 2023). However, there is a limited range of genomic diversity when comparing relationships between reference sequences and the ASVs observed through the 16S rRNA gene within this order (Strunecký et al., 2023), where the typical monophyly is not observed in the phylogenetic tree composed from ASVs from Little Turkey Lake. Organisms within this order inhabit terrestrial biotopes or the benthos and periphyton of the freshwater systems and can invade and proliferate into new habitats and can modify the ecosystem function (Strunecký et al., 2023; Sukenik et al., 2012). Through the analysis of morphological versus phylogenetic criteria, organisms within this order have gaps at the phylogenetic level where a combination of these two studies helps with genomic data integration with traditional morphological features into a conclusive taxonomic placement (Strunecký et al., 2023).

The order Pseudanabaenales consists of filamentous cyanobacteria formed by short to long oval or barrel-shaped cells that do not contain heterocytes and akinetes and have parietal thylakoid organization (Strunecký et al., 2023). These organisms are observed within freshwater systems and are poorly studied bloom-forming cyanobacteria (Acinas et al., 2009; Strunecký et al., 2023). Genera within this order are typically polyphyletic however, with the current taxonomic state, the sequences accumulated in GenBank is unsatisfactory and fall into various descriptions (Strunecký et al., 2023). Within the phylogenetic tree, the Pseudanabaenales order is well supported with clusters of reference sequences and ASVs, but

there are still gaps in the tree that require filling when some ASVs are clustered together without reference sequences.

The order Chroococcales has been subject to multiple taxonomic reorganizations due to its complexity (Strunecký et al., 2023). This order is morphologically variable and comprises of simple spherical or hemispherical cells that can develop into colonies, unicellular rod, or disk-like cells contain parietal thylakoids (Strunecký et al., 2023). Many genera within Chroococcales can be observed in environments with higher mineral contents like brackish or saline waters, and some strains that are involved in phytoplankton blooms can be observed in freshwater systems, like *Microcystis* and *Snowella* (Lefler et al., 2020; Strunecký et al., 2023). Cultivation of pure cyanobacteria and obtaining data from the number of genera observed within this order has been proven to be difficult, where in the incorporation of phylogeny based on multilocus analysis has helped with defining orders and maintaining monophyly (Strunecký et al., 2023). However, within the constructed phylogenetic tree, this order illustrates polyphyly, including gaps where some ASVs are not associated with reference sequences, or multiple ASVs group with few reference sequences. With the lack of genomic data, more information is required to incorporate reference sequences that link to ASVs.

The last two orders observed within the phylogenetic tree consists of Oscillatoriales, a polyphyletic order and Gloeobacterales, a monophyletic order (Strunecký et al., 2023). Oscillatoriales is include irregular thylakoid structure and has cell diameters varying from 2 – 115 µm, where this order still combines major filamentous genera that was not included in

Synechococcales, but there is a striking phylogenetic distance that requires further studies to define a narrower core cluster (Strunecký et al., 2023). Gloeobacterales consist of genera that have the absence of thylakoids but harbour the photosynthetic and respiratory complexes to produce energy-rich photosynthetic products (Strunecký et al., 2023). Within cyanobacterial phylogeny, the genera identified within this order are typically observed at the basal position in the phylogenetic tree, similar to what was observed in the constructed tree (Strunecký et al., 2023).

To overcome limitations of microscopy and to introduce high throughput sample processing and detection of rare species, molecular genetic techniques are useful and informative using a taxonomically informative DNA region and comparing it to a reference sequence library and helping differentiate toxic and non-toxic strains (Li et al., 2019; MacKeigan et al., 2022). Through further sequencing efforts and analysis and taxonomic identification of cyanobacterial strains that may have gaps in research, especially ASVs that have strong bootstrap support without associated reference sequences, current classification systems and databases can be upgraded to cover both morphological identity with genomic identity (Strunecký et al., 2023).

4.3.4 Analysis of Toxin-Producing Reference Sequences Observed in the Phylogenetic Tree

Cyanobacteria produce multiple toxins that are harmful to human and aquatic health (MacKeigan et al., 2022), with this study emphasizing focus on the hepatotoxin, microcystin and its presence in this environment. Throughout the phylogenetic tree, multiple potential

toxin-producing cyanobacteria were observed through well-supported branches of the reference sequences and associated ASVs. The presence of these organisms in an oligotrophic freshwater system throughout the months of May to August at Little Turkey Lake were analyzed and represented in the phylogenetic tree.

Reference sequences presented in the phylogenetic tree included sequences from toxin producing strains of picocyanobacteria like *Synechococcus* and *Cyanobium* involved in secreting secondary compounds like microcystin (Śliwińska-Wilczewska et al., 2018). Picocyanobacteria toxicity and their dominance during bloom and non-bloom periods are typically neglected from microscopic studies due to their relatively small size (Li et al., 2019), but the analysis of community composition through sequencing has provided a holistic view of the presence of these organisms. With the increase of the incidence and intensity of cyanobacterial blooms, the influence of environmental factors on picocyanobacteria and toxicity needs further studying, including the proliferation of these organisms in freshwater oligotrophic systems (Śliwińska-Wilczewska et al., 2018).

Nostocales within freshwater systems can be invasive and alter the structure of the native community by influencing the water quality through release of toxin compounds while having advantages characteristics like increased growth rate and resource use efficiency methods (Sukenik et al., 2012). Potential toxin producing genera observed within this order included in the phylogenetic tree through reference sequences contains *Aphanizomenon*, *Calothrix*, and *Dolichospermum* (Nowruzi & Porzani, 2021; Sukenik et al., 2012). *Aphanizomenon* is involved in producing cylindrospermopsin, a cyanotoxin known to show

effects in zooplankton, invertebrates and certain plant species, and blooms of *Dolichospermum* (*Anabaena*) and *Calothrix* are known to produce microcystin and are planktonic (Chorus & Welker, 2021; Kinnear, 2010; Nowruzi & Porzani, 2021). Microcystin producing organisms can be observed in high-level taxa of cyanobacteria, but distribution of the occurrence between genera and species is not consistent, as presented in the Nostocales order with some genera having the potential to produce microcystin but does not encompass all of them (Chorus & Welker, 2021)

Multiple species of *Pseudanabaena* within the order Pseudanabaenales and observed within the phylogenetic tree were associated with reference sequences and are involved in producing cyanotoxins (Chorus & Welker, 2021). This planktonic cyanobacterium genus produces surface blooms under its favourable conditions and have both toxin producing and non-producing strains observed in freshwater systems (Chorus & Welker, 2021; MacKeigan et al., 2022). The genus *Phormidium* is also a potential microcystin toxin producer and is able to dominate throughout season, with significant toxic activity in water systems (Chorus & Welker, 2021; Teneva et al., 2005). The genus *Leptolyngbya* has genetic information to potentially produce microcystin (Frazão et al., 2010). With morphological similarity to other phylogenetically unrelated filamentous taxa, further studies should include description and taxonomic assignment of these strains for better recognition and should implement genomic data to identify potential microcystin producing genera (Strunecký et al., 2023).

Within the Chroococcales order, two reference sequences belonging to freshwater bloom forming cyanobacterial genera, *Microcystis* and *Radiocystis* were observed (Lefler et

al., 2020; Lv et al., 2022). Many studies testing microcystin production and regulation use *Microcystis aeruginosa* PCC 7806 as a model organism to help with understanding cyanotoxins in cyanobacteria and the mechanisms used (Pimentel & Giani, 2014). Molecular characterization of the genus *Radiocystis* is rarely analyzed, but further studies have noted similarities with morphology, formation of blooms, and production of cyanotoxins in comparison to *Microcystis aeruginosa* (Lv et al., 2022). The frequency and intensity of these genera are directly related to the formation of blooms that can be harmful in eutrophic systems (Lv et al., 2022), but with these genera being observed in an oligotrophic system like Little Turkey Lake, this further illustrates that more research and attention is required in these systems.

Through the diversity of organisms present in the oligotrophic water system being able to produce microcystin, further insight into genomic information is required to fill gaps in research and incorporation of multiple molecular techniques will help provide a holistic approach to understanding potential toxin-producing organisms. Analyzing the system to understand the risk of the presence of toxin-producing strains of cyanobacteria provides more specific insight into understanding the water system and how to proceed with monitoring and treatment processes, especially for drinking water sources.

4.4 Conclusions, Implications, and Future Directions

4.4.1 Community Composition Variation and the Impact of a Changing Environment

The first aspect of this research looked at identifying cyanobacterial community composition at differing timepoints and depths at Turkey lakes Watershed to determine if

there is variation and the impact of the environment that may influence this variation. This research provided a snapshot of the microbial community in an oligotrophic system through integrative sampling of the water column and gained insight on the seasonal and diurnal variation of cyanobacterial organisms. There was a large presence of cyanobacteria observed in July and August, with a highlight of cyanobacteria emerging earlier in the season in this oligotrophic system. A report from the Turkey Lakes Watershed illustrated a slight extension of the growing season and changes in the vernal window that may affect cyanobacterial dynamics (Creed et al., 2015). This may occur due to variation in temperatures and changes in precipitation resulting in high areas of terrestrial runoff that can directly impact the characteristics of the lake (Creed et al., 2015). With the long-term data that has been collected at Turkey Lakes, this emphasizes the need for further analysis on a northern temperate lake and changes that may be occurring over time.

Cyanobacterial blooms are increasing worldwide through changes in temperature, pH, and nutrient loading through anthropogenic land use (Nwosu et al., 2021; Paerl, 2018). Changes in temperature may prolong periods of stratification in lakes that can change the community structure of cyanobacteria (Nwosu et al., 2021). Variation seasonally was observed in Little Turkey Lake with cyanobacterial re-emergence in the months of May and June, reaching peak abundances in July and August potentially due to the climate warming in the northern temperate area. These peak abundances were reached throughout the water column at various depths and timepoints without a visible surface bloom, illustrating the changes in distribution of these organisms and how unpredictable they are. This was also

observed in an oligotrophic system, where similar community compositions were noted in eutrophic freshwater systems, making this a comparable observation (Feng et al., 2019).

Within Turkey Lakes Watershed, cyanobacteria have previously been noted to dominate the phytoplankton communities in 1980 (Jeffries et al., 1988). The communities within these lakes were identified through variation diurnally, seasonally, and spatially throughout the water column. The changes in composition may be explained by various environmental factors that can affect cyanobacterial proliferation. In addition to the environment, cyanobacteria have adapted to seasonal changes and are able to use cellular advantages to access light and nutrient availability to their preference (Paerl, 2018). The mechanisms for these adaptations and environmental factors in combination require further analysis to understand the variation in community composition in each unique water system.

4.4.2 Quantification of Overall Cyanobacterial and Potential Toxin-Producing Organism Presence

Quantitative PCR is a valuable tool that can be used for detection of genes of interests and providing an amount present in the sample (Pinto et al., 2012). The role of qPCR in a potential predictive role for cyanobacterial bloom and toxin risk is extremely important and an effective way to conduct analysis of samples to quantify the genes of interest (Lu et al., 2020). This study conducted illustrated the use of qPCR as a molecular tool to quantify abundance of cyanobacterial organisms and potential microcystin producing organisms present in an oligotrophic system. Throughout the season, gene copy numbers varied based on sampling time and depth of the water column, with integrated sampling providing a

generally good overall snapshot analysis of the depths. Peak abundances were observed in August 2022 for both genes of interest. However, with a large rise in gene copy numbers, there was no visible bloom observed when sampling. This illustrates the need to emphasize that the absence of an observed bloom is not correlated with the absence of cyanobacterial organisms in larger abundances. Along with variation in abundance, the increase of gene copy numbers throughout the summer season without a visible bloom is indicative of the importance of knowing the water system and what may impact the variation.

The decrease in peak average abundance from June 2022 to July 2022 for the microcystin gene of interest was observed and was compared to the general ratio of overall cyanobacterial 16S rRNA gene V3 region analysis. This decrease in gene copy number may be attributed to multiple factors that can impact the community composition of potential toxin-producing organisms where there may have been a change in the genera and strains that were being detected throughout the summer season for the *mcy*E gene (Lu et al., 2020). Further validation on the multiple strains and genes of potential toxin producing cyanobacteria is required, similar to amplicon sequencing, to ensure the full genome involved in toxin production is captured (Lu et al., 2020). The presence and variation of genera in different environmental conditions and zones of the water column may change based on temperature changes, and light and nutrient availability (Gray et al., 2019). As previously mentioned, there are multiple variables that may play a role in the change in abundance of the genes of interest, with further analysis required over multiple years to determine any patterns that may be occurring.

The in-depth analysis using quantitative PCR conducted on multiple depth sampling was informative on the community composition and changes in abundance within the water column specifically for cyanobacteria and potential toxin producing organisms. Reactive sampling at the surface after a visible bloom has occurred is useful for microscopy and culturing identification but does not provide opportunity to have a proactive approach to analysis. This study focused on integrative sampling and multiple depth sampling for analysis as these methods can be useful for continuous monitoring (Kalkhajeh et al., 2019). Multiple depth sampling is intensive but can providing an in-depth analysis on the water system being studied. Integrative sampling is useful and efficient and can be general snapshot of the water column for continuous monitoring (Kalkhajeh et al., 2019). However, the changes in sampling within different protocols and the decision for the correct technique to use is dependent on the research question. Due to this variability, multiple depth sampling is initially recommended to understand the system, and then integrative sampling may be used for continuous monitoring over time, as it may not be enough to analyze the abundance of organisms present within the system. Through this, an overall proactive approach may be used for further analysis of water systems that may be vulnerable to harmful algal blooms.

4.4.3 Amplicon Sequencing and Gaps in Cyanobacterial Taxonomy

The use of amplicon sequencing helped with understanding the cyanobacterial genetic presence and diversity to determine community composition. Throughout the community analysis, picocyanobacterial reference sequences were dominant as these organisms can thrive in lower nutrient level environments (Batista & Giani, 2019). Similar community

compositions were observed previously at the Turkey Lakes Watershed (Cameron, 2021), with many amplicon sequence variants associated with multiple reference sequences belonging to picocyanobacteria. However, after consistent research in cyanobacterial taxonomy, there are still gaps in picocyanobacterial research with large taxonomical characterizations and diversity missing in reference sequence databases (MacKeigan et al., 2022). With constant improvements on molecular techniques, taxonomy of cyanobacteria is continuously undergoing revisions and development of valid reference sequences from isolated strains are still required (MacKeigan et al., 2022).

4.4.4 Future Directions on Cyanobacterial Community and Abundance Analysis

The presence of cyanobacteria and potential microcystin producing organisms in the oligotrophic watershed at higher abundances illustrates the need for continuous monitoring and detection of potential harmful algal blooms that may arise. Molecular techniques are effective and efficient tools but have limitations like understanding the complete taxonomical diversity and limitations to grasp the entire genotypic community and quantity. As previously mentioned, there are limitations to amplicon sequencing and cyanobacterial taxonomy with the prevalence of picocyanobacteria (Batista & Giani, 2019). This requires further culturing for isolation, identification, and sequencing of cyanobacterial strains to fill knowledge gaps within reference databases.

The analysis completed in this study would require other tools like toxin, nutrient, and metagenomic analyses to have a holistic understanding of the cyanobacterial community and to help predict causes of blooms and why toxins are produced (Chorus & Welker, 2021).

Through the incorporation of toxin and protein analysis, concentrations can be measured and analyzed instead of the potential to produce toxins. This would be a valuable addition to any sampling protocols to understand the full mechanisms occurring in each unique water system. Updating monitoring protocols specifically in oligotrophic water systems is required to understand baseline trends over multiple years. This can be through the incorporation of various molecular techniques and a combination of environmental metadata, where analysis of the community and potential environmental impacts on the community can be conducted (Paerl et al., 2016). This in turn can be useful for incorporating watershed protection and management to reduce further downstream drinking water treatment costs as these watersheds naturally provide a source of high-quality drinking water (Emelko et al., 2011).

The research conducted and the results presented provide a significant and conceptualized framework for drinking water monitoring through the analysis of bacterial community composition and abundance, and its impacts on an oligotrophic freshwater system. Identifying and classifying taxonomy of cyanobacteria that may be harmful helps with understanding the presence of these organisms in the drinking water systems that can lead to providing in-depth analysis for monitoring and water treatment programs. Typically, cyanobacterial research is associated with eutrophic systems and visible blooms on the water surface (Paerl et al., 2016). The research conducted in this thesis looked at molecular tools that can be used in combination to characterize and quantify the spatial and temporal variation of cyanobacterial communities in oligotrophic systems when visible biomass is absent.

References

Acinas, S. G., Haverkamp, T. H. A., Huisman, J., & Stal, L. J. (2009). Phenotypic and genetic diversification of Pseudanabaena spp. (cyanobacteria). *ISME Journal*, *3*(1), 31–46. https://doi.org/10.1038/ismej.2008.78

Ahmed, M., & Kim, D. R. (2018). pcr: An R package for quality assessment, analysis and testing of qPCR data. *PeerJ*, *2018*(3). https://doi.org/10.7717/peerj.4473

Aloraini, S., Alum, A., & Abbaszadegan, M. (2023). Impact of Pipe Material and Temperature on Drinking Water Microbiome and Prevalence of Legionella, Mycobacterium, and Pseudomonas Species. *Microorganisms*, *11*(2). https://doi.org/10.3390/microorganisms11020352

Al-Tebrineh, J., Pearson, L. A., Yasar, S. A., & Neilan, B. A. (2012). A multiplex qPCR targeting hepato- and neurotoxigenic cyanobacteria of global significance. *Harmful Algae*, *15*, 19–25. https://doi.org/10.1016/j.hal.2011.11.001

Anderson, D. M., Cembella, A. D., & Hallegraeff, G. M. (2012). Progress in understanding harmful algal blooms: Paradigm shifts and new technologies for research, monitoring, and management. *Annual Review of Marine Science*, *4*, 143–176. https://doi.org/10.1146/annurev-marine-120308-081121

Anderson, D. M., Glibert, P. M., & Burkholder, J. M. (2002). Harmful Algal Blooms and Eutrophication Nutrient Sources, Composition, and Consequences. *Estuaries*, *25*(4b), 704–726. https://doi.org/https://doi.org/10.1007/BF02804901

Ashbolt, N. J. (2004). Microbial contamination of drinking water and disease outcomes in developing regions. *Toxicology*, *198*(1–3), 229–238. https://doi.org/10.1016/j.tox.2004.01.030

Barbosa, M., Lefler, F. W., Berthold, D. E., Briggs-Gonzalez, V. S., Mazzotti, F. J., & Laughinghouse, H. D. (2022). Trophic State Drives the Diversity of Protists in a

Tropical River (New River, Belize). *Microorganisms*, *10*(12). https://doi.org/10.3390/microorganisms10122425

Batista, A. M. M., & Giani, A. (2019). Spatiotemporal variability of cyanobacterial community in a Brazilian oligomesotrophic reservoir: The picocyanobacterial dominance. *Ecohydrology and Hydrobiology*, *19*(4), 566–576. https://doi.org/10.1016/j.ecohyd.2019.03.002

Bautista-De los Santos, Q. M., Schroeder, J. L., Sevillano-Rivera, M. C., Sungthong, R., Ijaz, U. Z., Sloan, W. T., & Pinto, A. J. (2016). Emerging investigators series: Microbial communities in full-scale drinking water distribution systems-a meta-analysis. *Environmental Science: Water Research and Technology*, *2*(4), 631–644. https://doi.org/10.1039/c6ew00030d

Beversdorf, L. J., Chaston, S. D., Miller, T. R., & McMahon, K. D. (2015). Microcystin *mcy*A and *mcy*E gene abundances are not appropriate indicators of microcystin concentrations in lakes. *PLoS ONE*, *10*(5). https://doi.org/10.1371/journal.pone.0125353

Bisanz, J. E. (2018). *qiime2R: Importing QIIME2 artifacts and associated data into R sessions*. https://github.com/jbisanz/qiime2R

Blackburn, E. A. J., Emelko, M. B., Dickson-Anderson, S., & Stone, M. (2021). Advancing on the promises of techno-ecological nature-based solutions: A framework for green technology in water supply and treatment. *Blue-Green Systems*, *3*(1), 81–94. https://doi.org/10.2166/bgs.2021.008

Bláha, L., Babica, P., & Maršálek, B. (2009). Toxins produced in cyanobacterial water blooms - toxicity and risks. *Interdisciplinary Toxicology*, *2*(2), 36–41. https://doi.org/10.2478/v10102-009-0006-2

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and

extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857. https://doi.org/10.1038/s41587-019-0209-9

Borchardt, M. A., Boehm, A. B., Salit, M., Spencer, S. K., Wigginton, K. R., & Noble, R. T. (2021). The Environmental Microbiology Minimum Information (EMMI) Guidelines: QPCR and dPCR Quality and Reporting for Environmental Microbiology. In *Environmental Science and Technology* (Vol. 55, Issue 15, pp. 10210–10223). American Chemical Society. https://doi.org/10.1021/acs.est.1c01767

Cabello-Yeves, P. J., Haro-Moreno, J. M., Martin-Cuadrado, A. B., Ghai, R., Picazo, A., Camacho, A., & Rodriguez-Valera, F. (2017). Novel Synechococcus genomes reconstructed from freshwater reservoirs. *Frontiers in Microbiology*, *8*(JUN). https://doi.org/10.3389/fmicb.2017.01151

Cakmak, E. K., Hartl, M., Kisser, J., & Cetecioglu, Z. (2022). Phosphorus mining from eutrophic marine environment towards a blue economy: The role of bio-based applications. In *Water Research* (Vol. 219). Elsevier Ltd. https://doi.org/10.1016/j.watres.2022.118505

Caldwell, P. V., Martin, K. L., Vose, J. M., Baker, J. S., Warziniack, T. W., Costanza, J. K., Frey, G. E., Nehra, A., & Mihiar, C. M. (2023). Forested watersheds provide the highest water quality among all land cover types, but the benefit of this ecosystem service depends on landscape context. *Science of the Total Environment*, *882*. https://doi.org/10.1016/j.scitotenv.2023.163550

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Cameron, E. S. (2021). *Spatiotemporal Shifts in Cyanobacterial Communities in a Northern Temperate Watershed - Applications of Next-Generation Sequencing and Implications for Monitoring and Climate Change Adaptation*.

Cameron, E. S., Schmidt, P. J., Tremblay, B. J. M., Emelko, M. B., & Müller, K. M. (2021). Enhancing diversity analysis by repeatedly rarefying next generation sequencing data describing microbial communities. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-01636-1

Cheng, Y., Geng, R., Shan, L., Liu, Y., Zhang, H., Xiao, P., Ma, Z., & Li, R. (2022). Taxonomic Discussion on Cyanobacterial Systematics at Family Level, with Special Regards to Phormidiaceae by Using the Strains of Chinese Newly Recorded Genera Ancylothrix and Potamolinea. *Diversity*, *14*(4). https://doi.org/10.3390/d14040301

Chik, A. H. S., Schmidt, P. J., & Emelko, M. B. (2018). Learning something from nothing: The critical importance of rethinking microbial non-detects. *Frontiers in Microbiology*, *9*(OCT). https://doi.org/10.3389/fmicb.2018.02304

Chiu, Y. T., Chen, Y. H., Wang, T. S., Yen, H. K., & Lin, T. F. (2017). A qPCR-based tool to diagnose the presence of harmful cyanobacteria and cyanotoxins in drinking water sources. *International Journal of Environmental Research and Public Health*, *14*(5). https://doi.org/10.3390/ijerph14050547

Chorus, I., & Welker, M. (2021). *Toxic Cyanobacteria in Water; A Guide to Their Public Health Consequences, Monitoring and Management; Second Edition* (Second Edition, Vol. 2). CRC Press. https://doi.org/https://doi.org/10.1201/9781003081449

Creed, I. F., Hwang, T., Lutz, B., & Way, D. (2015). Climate warming causes intensification of the hydrological cycle, resulting in changes to the vernal and autumnal windows in a northern temperate forest. *Hydrological Processes*, *29*(16), 3519–3534. https://doi.org/10.1002/hyp.10450

Davenport, E. J., Neudeck, M. J., Matson, P. G., Bullerjahn, G. S., Davis, T. W., Wilhelm, S. W., Denney, M. K., Krausfeldt, L. E., Stough, J. M. A., Meyer, K. A., Dick, G. J., Johengen, T. H., Lindquist, E., Tringe, S. G., & McKay, R. M. L. (2019). Metatranscriptomic Analyses of Diel Metabolic Functions During a Microcystis Bloom

in Western Lake Erie (United States). *Frontiers in Microbiology*, *10*. https://doi.org/10.3389/fmicb.2019.02081

Douterelo, I., Boxall, J. B., Deines, P., Sekar, R., Fish, K. E., & Biggs, C. A. (2014). Methodological approaches for studying the microbial ecology of drinking water distribution systems. In *Water Research* (Vol. 65, pp. 134–156). Elsevier Ltd. https://doi.org/10.1016/j.watres.2014.07.008

Du, X., Liu, H., Yuan, L., Wang, Y., Ma, Y., Wang, R., Chen, X., Losiewicz, M. D., Guo, H., & Zhang, H. (2019). The diversity of cyanobacterial toxins on structural characterization, distribution and identification: A systematic review. *Toxins*, *11*(9). https://doi.org/10.3390/toxins11090530

Ellison, D., Morris, C. E., Locatelli, B., Sheil, D., Cohen, J., Murdiyarso, D., Gutierrez, V., Noordwijk, M. van, Creed, I. F., Pokorny, J., Gaveau, D., Spracklen, D. V., Tobella, A. B., Ilstedt, U., Teuling, A. J., Gebrehiwot, S. G., Sands, D. C., Muys, B., Verbist, B., … Sullivan, C. A. (2017). Trees, forests and water: Cool insights for a hot world. *Global Environmental Change*, *43*, 51–61. https://doi.org/10.1016/j.gloenvcha.2017.01.002

Emelko, M. B., Silins, U., Bladon, K. D., & Stone, M. (2011). Implications of land disturbance on drinking water treatability in a changing climate: Demonstrating the need for " source water supply and protection" strategies. *Water Research*, *45*(2), 461–472. https://doi.org/10.1016/j.watres.2010.08.051

Emelko, M. B., Stone, M., Silins, U., Allin, D., Collins, A. L., Williams, C. H. S., Martens, A. M., & Bladon, K. D. (2016). Sediment-phosphorus dynamics can shift aquatic ecology and cause downstream legacy effects after wildfire in large river systems. *Global Change Biology*, *22*(3), 1168–1184. https://doi.org/10.1111/gcb.13073

Favot, E. J., Holeton, C., DeSellas, A. M., & Paterson, A. M. (2023). Cyanobacterial blooms in Ontario, Canada: continued increase in reports through the 21st century. *Lake and Reservoir Management*, *39*(1), 1–20. https://doi.org/10.1080/10402381.2022.2157781

Feng, C., Jia, J., Wang, C., Han, M., Dong, C., Huo, B., Li, D., & Liu, X. (2019). Phytoplankton and bacterial community structure in two chinese lakes of different trophic status. *Microorganisms*, *7*(12). https://doi.org/10.3390/microorganisms7120621

Ferro, P., Vaz-Moreira, I., & Manaia, C. M. (2019). Betaproteobacteria are predominant in drinking water: are there reasons for concern? In *Critical Reviews in Microbiology* (Vol. 45, Issues 5–6, pp. 649–667). Taylor and Francis Ltd. https://doi.org/10.1080/1040841X.2019.1680602

Frazão, B., Martins, R., & Vasconcelos, V. (2010). Are known cyanotoxins involved in the toxicity of picoplanktonic and filamentous north atlantic marine cyanobacteria? *Marine Drugs*, *8*(6), 1908–1919. https://doi.org/10.3390/md8061908

Government of Canada. (2014). *Characteristics of the Turkey Lakes watershed study site*. Government of Canada.

Gray, E., Elliott, J. A., Mackay, E. B., Folkard, A. M., Keenan, P. O., & Jones, I. D. (2019). Modelling lake cyanobacterial blooms: Disentangling the climate-driven impacts of changing mixed depth and water temperature. *Freshwater Biology*, *64*(12), 2141–2155. https://doi.org/10.1111/fwb.13402

Greer, B., Meneely, J. P., & Elliott, C. T. (2018). Uptake and accumulation of Microcystin-LR based on exposure through drinking water: An animal model assessing the human health risk. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-018-23312-7

Guan, Y., Jia, J., Wu, L., Xue, X., Zhang, G., & Wang, Z. (2018). Analysis of Bacterial Community Characteristics, Abundance of Antibiotics and Antibiotic Resistance Genes Along a Pollution Gradient of Ba River in Xi'an, China. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.03191

Ho, J. C., & Michalak, A. M. (2020). Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. *Limnology and Oceanography*, *65*(5), 992–1009. https://doi.org/10.1002/lno.11365

Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. *Nature Reviews Microbiology*, *16*(8), 471–483. https://doi.org/10.1038/s41579-018-0040-1

Jeffries, D. S., & Foster, N. W. (2001). The Turkey Lakes watershed study: milestones and prospects. *Ecosystems*, *4*, 501–502. https://doi.org/https://doi.org/10.1007/s10021-001-0023-2

Jeffries, D. S., M Kelso, J. R., Morrison, I. K., Kelso, M., & Morrison, I. K. (1988). Physical, Chemical, and Biological Characteristics of the Turkey Lakes Watershed, Central Ontario, Canada. *Can. J. Fish. Aquat. Sci.*, *45*, 3–13. https://doi.org/https://doi.org/10.1139/f88-262

Ji, B., Liang, J., Ma, Y., Zhu, L., & Liu, Y. (2019). Bacterial community and eutrophic index analysis of the East Lake. *Environmental Pollution*, *252*, 682–688. https://doi.org/10.1016/j.envpol.2019.05.138

Jing, Z., Lu, Z., Mao, T., Cao, W., Wang, W., Ke, Y., Zhao, Z., Wang, X., & Sun, W. (2021). Microbial composition and diversity of drinking water: A full scale spatial-temporal investigation of a city in northern China. *Science of the Total Environment*, *776*. https://doi.org/10.1016/j.scitotenv.2021.145986

Jungblut, A. D., & Neilan, B. A. (2006). Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. *Archives of Microbiology*, *185*(2), 107–114. https://doi.org/10.1007/s00203-005-0073-5

Kalkhajeh, Y. K., Amiri, B. J., Huang, B., Khalyani, A. H., Hu, W., Gao, H., & Thompson, M. L. (2019). Methods for sample collection, storage, and analysis of freshwater phosphorus. In *Water (Switzerland)* (Vol. 11, Issue 9). MDPI AG. https://doi.org/10.3390/w11091889

Kinnear, S. (2010). Cylindrospermopsin: A decade of progress on bioaccumulation research. In *Marine Drugs* (Vol. 8, Issue 3, pp. 542–564). MDPI AG. https://doi.org/10.3390/md8030542

Komárek, J. (2016). A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. *European Journal of Phycology*, *51*(3), 346–353. https://doi.org/10.1080/09670262.2016.1163738

Komárek, J., Johansen, J. R., Šmarda, J., & Strunecký, O. (2020). Phylogeny and taxonomy of synechococcus–like cyanobacteria. *Fottea*, *20*(2), 171–191. https://doi.org/10.5507/fot.2020.006

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Larsen, M. L., Baulch, H. M., Schiff, S. L., Simon, D. F., Sauvé, S., & Venkiteswaran, J. J. (2020). Extreme rainfall drives early onset cyanobacterial bloom. *Facets*, *5*(1), 899–920. https://doi.org/10.1139/FACETS-2020-0022

Lefler, F. W., Barbosa, M., Berthold, D. E., Dail, H., & Iv, L. (2020). *Genome Sequences of Two Microcystis aeruginosa (Chroococcales, Cyanobacteria) Strains from Florida (United States) with Disparate Toxigenic Potentials*. https://doi.org/10.1128/MRA.00844

Lefler, F. W., Berthold, D. E., & Laughinghouse, H. D. (2023). Cyanoseq: A database of cyanobacterial 16S rRNA gene sequences with curated taxonomy. *Journal of Phycology*, *59*(3), 470–480. https://doi.org/https://doi.org/10.1111/jpy.13335

Li, H., Alsanea, A., Barber, M., & Goel, R. (2019). High-throughput DNA sequencing reveals the dominance of pico- and other filamentous cyanobacteria in an urban freshwater Lake. *Science of the Total Environment*, *661*, 465–480. https://doi.org/10.1016/j.scitotenv.2019.01.141

Liu, S., Ren, H., Shen, L., Lou, L., Tian, G., Zheng, P., & Hu, B. (2015). pH levels drive bacterial community structure in the Qiantang River as determined by 454 pyrosequencing. *Frontiers in Microbiology*, *6*(MAR). https://doi.org/10.3389/fmicb.2015.00285

Liu, X., Meng, X., Wang, X., Bi, D., Chen, L., & Lou, Q. (2020). Using a semi-analytical model to retrieve Secchi depth in coastal and estuarine waters. *Acta Oceanologica Sinica*, *39*(8), 103–112. https://doi.org/10.1007/s13131-020-1620-2

Lopes, A. F., Macdonald, J. L., Quinteiro, P., Arroja, L., Carvalho-Santos, C., Cunha-e-Sá, M. A., & Dias, A. C. (2019). Surface vs. groundwater: The effect of forest cover on the costs of drinking water. *Water Resources and Economics*, *28*. https://doi.org/10.1016/j.wre.2018.06.002

Lopes, R. H., Silva, C. R. D. V., Salvador, P. T. C. de O., Silva, Í. de S., Heller, L., & Uchôa, S. A. da C. (2022). Surveillance of Drinking Water Quality Worldwide: Scoping Review Protocol. *International Journal of Environmental Research and Public Health*, *19*(15). https://doi.org/10.3390/ijerph19158989

Lu, J., Struewing, I., Wymer, L., Tettenhorst, D. R., Shoemaker, J., & Allen, J. (2020). Use of qPCR and RT-qPCR for monitoring variations of microcystin producers and as an early warning system to predict toxin production in an Ohio inland lake. *Water Research*, *170*. https://doi.org/10.1016/j.watres.2019.115262

Lv, X., Cheng, Y., Zhang, S., Hu, Z., Xiao, P., Zhang, H., Geng, R., & Li, R. (2022). Polyphasic Characterization and Taxonomic Evaluation of a Bloom-Forming Strain Morphologically Resembling Radiocystis fernandoi (Chroococcales, Cyanobacteria) from Lake Erhai, China. *Diversity*, *14*(10). https://doi.org/10.3390/d14100816

MacKeigan, P. W., Garner, R. E., Monchamp, M. È., Walsh, D. A., Onana, V. E., Kraemer, S. A., Pick, F. R., Beisner, B. E., Agbeti, M. D., da Costa, N. B., Shapiro, B. J., & Gregory-Eaves, I. (2022). Comparing microscopy and DNA metabarcoding techniques

for identifying cyanobacteria assemblages across hundreds of lakes. *Harmful Algae*, *113*. https://doi.org/10.1016/j.hal.2022.102187

Mazard, S., Penesyan, A., Ostrowski, M., Paulsen, I. T., & Egan, S. (2016). Tiny microbes with a big impact: The role of cyanobacteria and their metabolites in shaping our future. *Marine Drugs*, *14*(5). https://doi.org/10.3390/md14050097

McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, *66*(2), 526–538. https://doi.org/10.1016/j.ympev.2011.12.007

McElhoe, J. A., Holland, M. M., Makova, K. D., Su, M. S. W., Paul, I. M., Baker, C. H., Faith, S. A., & Young, B. (2014). Development and assessment of an optimized next-generation DNA sequencing approach for the metagenome using the Illumina MiSeq. *Forensic Science International: Genetics*, *13*, 20–29. https://doi.org/10.1016/j.fsigen.2014.05.007

McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4). https://doi.org/10.1371/journal.pone.0061217

Moreira, C., Vasconcelos, V., & Antunes, A. (2013). Phylogeny and biogeography of cyanobacteria and their produced toxins. In *Marine Drugs* (Vol. 11, Issue 11, pp. 4350–4369). MDPI AG. https://doi.org/10.3390/md11114350

Müller, K. M., Chhun, A., Guildford, S. J., Yakobowski, S. J., & Jonlija, M. (2017). Molecular and ecological characterization of toxic cyanobacteria from the Bay of Quinte (Lake Ontario) and Maumee Bay (Lake Erie). *Journal of Great Lakes Research*, *43*(6), 1067–1083. https://doi.org/10.1016/j.jglr.2017.03.021

Nowruzi, B., & Porzani, S. J. (2021). Toxic compounds produced by cyanobacteria belonging to several species of the order Nostocales: A review. In *Journal of Applied*

Toxicology (Vol. 41, Issue 4, pp. 510–548). John Wiley and Sons Ltd. https://doi.org/10.1002/jat.4088

Nwankwegu, A. S., Li, Y., Huang, Y., Wei, J., Norgbey, E., Sarpong, L., Lai, Q., & Wang, K. (2019). Harmful algal blooms under changing climate and constantly increasing anthropogenic actions: the review of management implications. *3 Biotech*, *9*(12). https://doi.org/10.1007/s13205-019-1976-1

Nwosu, E. C., Roeser, P., Yang, S., Pinkerneil, S., Ganzert, L., Dittmann, E., Brauer, A., Wagner, D., & Liebner, S. (2021). Species-Level Spatio-Temporal Dynamics of Cyanobacteria in a Hard-Water Temperate Lake in the Southern Baltics. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.761259

Otten, T. G., Graham, J. L., Harris, T. D., & Dreher, T. W. (2016). Elucidation of tasteand odor-producing bacteria and toxigenic cyanobacteria in a Midwestern drinking water supply reservoir by shotgun metagenomic analysis. *Applied and Environmental Microbiology*, *82*(17), 5410–5420. https://doi.org/10.1128/AEM.01334-16

Pacheco, A. B. F., Guedes, I. A., & Azevedo, S. M. F. O. (2016). Is qPCR a reliable indicator of cyanotoxin risk in freshwater? *Toxins*, *8*(6). https://doi.org/10.3390/toxins8060172

Paerl, H. W. (2018). Mitigating toxic planktonic cyanobacterial blooms in aquatic ecosystems facing increasing anthropogenic and climatic pressures. *Toxins*, *10*(2). https://doi.org/10.3390/toxins10020076

Paerl, H. W., & Otten, T. G. (2013). Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microbial Ecology*, *65*(4), 995–1010. https://doi.org/10.1007/s00248-012-0159-y

Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. *Water Research*, *46*(5), 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002

Paerl, H. W., Scott, J. T., McCarthy, M. J., Newell, S. E., Gardner, W. S., Havens, K. E., Hoffman, D. K., Wilhelm, S. W., & Wurtsbaugh, W. A. (2016). It Takes Two to Tango: When and Where Dual Nutrient (N & P) Reductions Are Needed to Protect Lakes and Downstream Ecosystems. *Environmental Science and Technology*, *50*(20), 10805–10813. https://doi.org/10.1021/acs.est.6b02575

Paliy, O., & Shankar, V. (2016). Application of multivariate statistical techniques in microbial ecology. In *Molecular Ecology* (Vol. 25, Issue 5, pp. 1032–1057). Blackwell Publishing Ltd. https://doi.org/10.1111/mec.13536

Pearson, L. A., Dittmann, E., Mazmouz, R., Ongley, S. E., D'Agostino, P. M., & Neilan, B. A. (2016). The genetics, biosynthesis and regulation of toxic specialized metabolites of cyanobacteria. *Harmful Algae*, *54*, 98–111. https://doi.org/10.1016/j.hal.2015.11.002

Pimentel, J. S. M., & Giani, A. (2014). Microcystin production and regulation under nutrient stress conditions in toxic Microcystis strains. *Applied and Environmental Microbiology*, *80*(18), 5836–5843. https://doi.org/10.1128/AEM.01009-14

Pinto, F., Pacheco, C. C., Ferreira, D., Moradas-Ferreira, P., & Tamagnini, P. (2012). Selection of suitable reference genes for RT-qPCR analyses in cyanobacteria. *PLoS ONE*, *7*(4). https://doi.org/10.1371/journal.pone.0034983

Postel, S. L., & Thompson, B. H. (2005). Watershed protection: Capturing the benefits of nature's water supply services. *Natural Resources Forum*, *29*(2), 98–108. https://doi.org/10.1111/j.1477-8947.2005.00119.x

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., & Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1). https://doi.org/10.1093/nar/gks1219

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. https://www.r-project.org/

Rocha, J., & Manaia, C. M. (2020). Cell-based internal standard for qPCR determinations of antibiotic resistance indicators in environmental water samples. *Ecological Indicators*, *113*. https://doi.org/10.1016/j.ecolind.2020.106194

Rogers, J. E., Devereux, R., James, J. B., George, S. E., & Forshay, K. J. (2021). Seasonal Distribution of Cyanobacteria in Three Urban Eutrophic Lakes Results from an Epidemic-like Response to Environmental Conditions. *Current Microbiology*, *78*(6), 2298–2316. https://doi.org/10.1007/s00284-021-02498-6

RStudio Team. (2020). *RStudio: Integrated Development for R*.

Saxena, G., Bharagava, R. N., Kaithwas, G., & Raj, A. (2015). Microbial indicators, pathogens and methods for their monitoring in water environment. *Journal of Water and Health*, *13*(2), 319–339. https://doi.org/10.2166/wh.2014.275

Schirrmeister, B. E., Dalquen, D. A., Anisimova, M., & Bagheri, H. C. (2012). Gene copy number variation and its significance in cyanobacterial phylogeny. *BMC Microbiology*, *12*, 177. https://doi.org/10.1186/1471-2180-12-177

Schmale, D. G., Ault, A. P., Saad, W., Scott, D. T., & Westrick, J. A. (2019). Perspectives on harmful algal blooms (HABs) and the cyberbiosecurity of freshwater systems. *Frontiers in Bioengineering and Biotechnology*, *7*(JUN). https://doi.org/10.3389/fbioe.2019.00128

Schmidt, P. J., Cameron, E. S., Müller, K. M., & Emelko, M. B. (2022). Ensuring That Fundamentals of Quantitative Microbiology Are Reflected in Microbial Diversity Analyses Based on Next-Generation Sequencing. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.728146

Selbig, W. R., Cox, A., & Bannerman, R. T. (2012). Verification of a depth-integrated sample arm as a means to reduce solids stratification bias in urban stormwater sampling. *Journal of Environmental Monitoring*, *14*(4), 1138–1144. https://doi.org/10.1039/c2em10999a

Shahi, S. K., Freedman, S. N., & Mangalam, A. K. (2017). Gut microbiome in multiple sclerosis: The players involved and the roles they play. *Gut Microbes*, *8*(6), 607–615. https://doi.org/10.1080/19490976.2017.1349041

Sharma, A., & Lal, R. (2017). Survey of (Meta)genomic Approaches for Understanding Microbial Community Dynamics. *Indian Journal of Microbiology*, *57*(1), 23–38. https://doi.org/10.1007/s12088-016-0629-x

Sim, Z. Y., Goh, K. C., He, Y., & Gin, K. Y. H. (2023). Present and future potential role of toxin-producing Synechococcus in the tropical region. *Science of the Total Environment*, *896*. https://doi.org/10.1016/j.scitotenv.2023.165230

Śliwińska-Wilczewska, S., Maculewicz, J., Felpeto, A. B., & Latała, A. (2018). Allelopathic and bloom-forming picocyanobacteria in a changing world. In *Toxins* (Vol. 10, Issue 1). MDPI AG. https://doi.org/10.3390/toxins10010048

Stauffer, B. A., Bowers, H. A., Buckley, E., Davis, T. W., Johengen, T. H., Kudela, R., McManus, M. A., Purcell, H., Smith, G. J., Woude, A. Vander, & Tamburri, M. N. (2019). Considerations in harmful algal bloom research and monitoring: Perspectives from a consensus-building workshop and technology testing. In *Frontiers in Marine Science* (Vol. 6, Issue JUL). Frontiers Media S.A. https://doi.org/10.3389/fmars.2019.00399

Straub, D., Blackwell, N., Langarica-Fuentes, A., Peltzer, A., Nahnsen, S., & Kleindienst, S. (2020). Interpretations of Environmental Microbial Community Studies Are Biased by the Selected 16S rRNA (Gene) Amplicon Sequencing Pipeline. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.550420

Strunecký, O., Ivanova, A. P., & Mareš, J. (2023). An updated classification of cyanobacterial orders and families based on phylogenomic and polyphasic analysis. In *Journal of Phycology* (Vol. 59, Issue 1, pp. 12–51). John Wiley and Sons Inc. https://doi.org/10.1111/jpy.13304

Sukenik, A., Hadas, O., Kaplan, A., & Quesada, A. (2012). Invasion of Nostocales (cyanobacteria) to subtropical and temperate freshwater lakes - physiological, regional, and global driving forces. *Frontiers in Microbiology*, *3*(MAR). https://doi.org/10.3389/fmicb.2012.00086

Teneva, I., Dzhambazov, B., Koleva, L., Mladenov, R., & Schirmer, K. (2005). Toxic potential of five freshwater Phormidium species (Cyanoprokaryota). *Toxicon*, *45*(6), 711–725. https://doi.org/10.1016/j.toxicon.2005.01.018

Tiwari, A., Hokajärvi, A. M., Domingo, J. S., Elk, M., Jayaprakash, B., Ryu, H., Siponen, S., Vepsäläinen, A., Kauppinen, A., Puurunen, O., Artimo, A., Perkola, N., Huttula, T., Miettinen, I. T., & Pitkänen, T. (2021). Bacterial diversity and predicted enzymatic function in a multipurpose surface water system – from wastewater effluent discharges to drinking water production. *Environmental Microbiomes*, *16*(1). https://doi.org/10.1186/s40793-021-00379-w

Van Assche, A., Crauwels, S., De Brabanter, J., Willems, K. A., & Lievens, B. (2019). Characterization of the bacterial community composition in water of drinking water production and distribution systems in Flanders, Belgium. *MicrobiologyOpen*, *8*(5). https://doi.org/10.1002/mbo3.726

Vilmin, L., Flipo, N., Escoffier, N., & Groleau, A. (2018). Estimation of the water quality of a large urbanized river as defined by the European WFD: what is the optimal sampling frequency? *Environmental Science and Pollution Research*, *25*(24), 23485–23501. https://doi.org/10.1007/s11356-016-7109-z

Walters, K. E., & Martiny, J. B. H. (2020). Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. *PLoS ONE*, *15*(9 September). https://doi.org/10.1371/journal.pone.0233872

Walters, W., Hyde, E. R., Berg-Lyons, D., Ackermann, G., Humphrey, G., Parada, A., Gilbert, J. A., Jansson, J. K., Caporaso, J. G., Fuhrman, J. A., Apprill, A., & Knight, R. (2015). Improved Bacterial 16S rRNA Gene (V4 and V4-5) and Fungal Internal

Transcribed Spacer Marker Gene Primers for Microbial Community Surveys. *MSystems*, *1*(1). https://doi.org/10.1128/msystems.00009-15

Walther, J., Schwarz, A., Witthohn, M., Strieth, D., Muffler, K., & Ulber, R. (2020). A qPCR method for distinguishing biomass from non-axenic terrestrial cyanobacteria cultures in hetero- or mixotrophic cultivations. *Journal of Applied Phycology*, *32*(6), 3767–3774. https://doi.org/10.1007/s10811-020-02282-2

Webster, K. L., Leach, J. A., Hazlett, P. W., Fleming, R. L., Emilson, E. J. S., Houle, D., Chan, K. H. Y., Norouzian, F., Cole, A. S., O'Brien, J. M., Smokorowski, K. E., Nelson, S. A., & Yanni, S. D. (2021). Turkey Lakes Watershed, Ontario, Canada: 40 years of interdisciplinary whole-ecosystem research. *Hydrological Processes*, *35*(4). https://doi.org/10.1002/hyp.14109

Wells, M. L., Karlson, B., Wulff, A., Kudela, R., Trick, C., Asnaghi, V., Berdalet, E., Cochlan, W., Davidson, K., De Rijcke, M., Dutkiewicz, S., Hallegraeff, G., Flynn, K. J., Legrand, C., Paerl, H., Silke, J., Suikkanen, S., Thompson, P., & Trainer, V. L. (2020). Future HAB science: Directions and challenges in a changing climate. *Harmful Algae*, *91*. https://doi.org/10.1016/j.hal.2019.101632

Wert, E. C., Korak, J. A., Trenholm, R. A., & Rosario-Ortiz, F. L. (2014). Effect of oxidant exposure on the release of intracellular microcystin, MIB, and geosmin from three cyanobacteria species. *Water Research*, *52*, 251–259. https://doi.org/10.1016/j.watres.2013.11.001

Wickham, H. (2009). ggplot2: Elegant Graphics for Data Analysis. In *ggplot2*. Springer New York. https://doi.org/10.1007/978-0-387-98141-3

Willis, A. D. (2019). Rarefaction, alpha diversity, and statistics. *Frontiers in Microbiology*, *10*(OCT). https://doi.org/10.3389/fmicb.2019.02407

Winter, J. G., Desellas, A. M., Fletcher, R., Heintsch, L., Morley, A., Nakamoto, L., & Utsumi, K. (2011). Algal blooms in Ontario, Canada: Increases in reports since 1994.

Lake and Reservoir Management, *27*(2), 107–114. https://doi.org/10.1080/07438141.2011.557765

Yang, X. E., Wu, X., Hao, H. L., & He, Z. L. (2008). Mechanisms and assessment of water eutrophication. In *Journal of Zhejiang University: Science B* (Vol. 9, Issue 3, pp. 197–209). https://doi.org/10.1631/jzus.B0710626

Yilmaz, P., Parfrey, L. W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, J., Ludwig, W., & Glöckner, F. O. (2014). The SILVA and "all-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Research*, *42*(D1). https://doi.org/10.1093/nar/gkt1209

Zhang, M., Yu, N., Chen, L., Jiang, C., Tao, Y., Zhang, T., Chen, J., & Xue, D. (2012). Structure and seasonal dynamics of bacterial communities in three urban rivers in China. *Aquatic Sciences*, *74*(1), 113–120. https://doi.org/10.1007/s00027-011-0201-z

Zhang, P., Peng, C., Zhang, J., Zhang, J., Chen, J., & Zhao, H. (2022). Long-Term Harmful Algal Blooms and Nutrients Patterns Affected by Climate Change and Anthropogenic Pressures in the Zhanjiang Bay, China. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.849819

Zheng, F., Zhang, T., Yin, S., Qin, G., Chen, J., Zhang, J., Zhao, D., Leng, X., An, S., & Xia, L. (2022). Comparison and interpretation of freshwater bacterial structure and interactions with organic to nutrient imbalances in restored wetlands. *Frontiers in Microbiology*, *13*. https://doi.org/10.3389/fmicb.2022.946537

Zhou, W., Li, W., Chen, J., Zhou, Y., Wei, Z., & Gong, L. (2021). Microbial diversity in full-scale water supply systems through sequencing technology: a review. In *RSC Advances* (Vol. 11, Issue 41, pp. 25484–25496). Royal Society of Chemistry. https://doi.org/10.1039/d1ra03680g

Zwart, G., Kamst-Van Agterveld, M. P., Van Der Werff-Staverman, I., Hagen, F., Hoogveld, H. L., & Gons, H. J. (2005). Molecular characterization of cyanobacterial diversity in a

shallow eutrophic lake. *Environmental Microbiology*, *7*(3), 365–377. https://doi.org/10.1111/j.1462-2920.2005.00715.x

Appendix A: Sampling Data for 16S rRNA Gene Sequencing and qPCR

Table A1. Sampling data for samples sent for sequencing from Little Turkey Lake from May to August 2022.

Sample Code	Date	Time of Day	Depth (categorical)	Depth (value, m)
LT15	27-May	Morning	Secchi	6
LT21	27-May	Afternoon	Secchi	4
LT27	27-May	Evening	Secchi	5
LT53	21-Jun	Morning	Integrated	0-5.75
LT61	21-Jun	Afternoon	Integrated	0-6
LT69	21-Jun	Evening	Integrated	0-5.5
LT133	22-Jul	Morning	Integrated	0-6.5
LT141	22-Jul	Afternoon	Integrated	0-5.25
LT149	22-Jul	Evening	Integrated	0-5
LT181	24-Aug	Morning	Integrated	0-6
LT189	24-Aug	Afternoon	Integrated	0 - 6.25

Table A2. Environmental data for samples sent for sequencing from Little Turkey Lake from May to August 2022.

Sample Code	pH	Temperature (°C)	Conductivity (ps/cm)	TDS (ppm)	Weather Conditions
LT15	6.78	13.8	28.8	20.2	grey skies, overcast, some sun, 7.6°C in Sault Ste. Marie
LT21	7.07	15.2	27.2	19.3	slightly overcast with blue patches, calm water, little wind, some sun, 9.6°C in Sault Ste. Marie
LT27	6.8	15.3	16.3	10.5	slight breeze, cloudy, overcast, rocky water, 13.3°C in Sault Ste. Marie
LT53	6.69	14.5	28.3	20.4	cool, overcast, calm current, soft wind, 18.9°C in Sault Ste. Marie
LT61	6.59	18.4	29.4	20.9	sunny, little cloudy, humid, 25.3°C in Sault Ste. Marie
LT69	6.18	16.9	29.8	21.2	sunny, no wind, humid, 29.1°C in Sault Ste. Marie
LT133	6.5	19	21.4	15.1	sunny (no clouds), no wind, warm, calm water, 21.9°C in Sault Ste. Marie
LT141	6.62	20.6	27.2	19.3	sunny, white clouds, calm water, slight wind, warm, 24.9°C in Sault Ste. Marie
LT149	6.78	22.1	27.1	19.3	sunny, clouds, windy, 22.6°C in Sault Ste. Marie
LT181	7.19	18.4	27.8	19.6	sunny, slight wind, calm, some clouds, blue sky, calm water, 19.3°C in Sault Ste. Marie
LT189	6.36	19.4	28.8	19.7	cloudy, no wind, calm, some clouds, blue sky, calm water, 23.8°C in Sault Ste. Marie
LT197	6.97	19.1	27.2	19.5	cloudy, calm water, slight wind, more darker clouds, slight drizzle before getting on boat, 23.7°C in Sault Ste. Marie

Table A3. Sampling data for samples tested in quantitative PCR analysis from Little Turkey Lake from May to August 2022.

Sample Code	Date	Time	Depth (categorical)	Depth (value, m)
LT13	27-May	Morning	Surface	0
LT15	27-May	Morning	Secchi	6
LT17	27-May	Morning	1m below Secchi	7
LT19	27-May	Afternoon	Surface	0
LT21	27-May	Afternoon	Secchi	4
LT23	27-May	Afternoon	1m below Secchi	5
LT25	27-May	Evening	Surface	0
LT27	27-May	Evening	Secchi	5
LT29	27-May	Evening	1m below Secchi	6
LT47	21-Jun	Morning	Surface	0
LT49	21-Jun	Morning	Secchi	4.75
LT51	21-Jun	Morning	1m below Secchi	5.75
LT53	21-Jun	Morning	Integrated	0-5.75
LT55	21-Jun	Afternoon	Surface	0
LT57	21-Jun	Afternoon	Secchi	5
LT59	21-Jun	Afternoon	1m below Secchi	6
LT61	21-Jun	Afternoon	Integrated	0-6
LT63	21-Jun	Evening	Surface	0
LT65	21-Jun	Evening	Secchi	4.5
LT67	21-Jun	Evening	1m below Secchi	5.5
LT69	21-Jun	Evening	Integrated	0-5.5
LT127	22-Jul	Morning	Surface	0
LT129	22-Jul	Morning	Secchi	5.5
LT131	22-Jul	Morning	1m below Secchi	6.5
LT133	22-Jul	Morning	Integrated	0-6.5

LT135	22-Jul	Afternoon	Surface	0
LT137	22-Jul	Afternoon	Secchi	4.25
LT139	22-Jul	Afternoon	1m below Secchi	5.25
LT141	22-Jul	Afternoon	Integrated	0-5.25
LT143	22-Jul	Evening	Surface	0
LT145	22-Jul	Evening	Secchi	4
LT147	22-Jul	Evening	1m below Secchi	5
LT149	22-Jul	Evening	Integrated	0-5
LT175	24-Aug	Morning	Surface	0
LT177	24-Aug	Morning	Secchi	5
LT179	24-Aug	Morning	1m below Secchi	6
LT181	24-Aug	Morning	Integrated	0-6
LT183	24-Aug	Afternoon	Surface	0
LT185	24-Aug	Afternoon	Secchi	5.25
LT187	24-Aug	Afternoon	1m below Secchi	6.25
LT189	24-Aug	Afternoon	Integrated	0 - 6.25
LT191	24-Aug	Evening	Surface	0
LT193	24-Aug	Evening	Secchi	5
LT195	24-Aug	Evening	1m below Secchi	6
LT197	24-Aug	Evening	Integrated	0-6

Table A4. Environmental data for samples tested in quantitative PCR analysis from Little Turkey Lake from May to August 2022.

Sample Code	pH	Temperature (°C)	Conductivity (ps/cm)	TDS (ppm)	Weather Conditions
LT13	6.93	13.9	30.5	21.7	grey skies, overcast, some sun, 7.6°C in Sault Ste. Marie
LT15	6.78	13.8	28.8	20.2	grey skies, overcast, some sun, 7.6°C in Sault Ste. Marie
LT17	6.69	13.6	31.6	22.7	grey skies, overcast, some sun, 7.6°C in Sault Ste. Marie
LT19	7.56	16.6	25.5	18.2	slightly overcast with blue patches, calm water, little wind, some sun, 9.6°C in Sault Ste. Marie
LT21	7.07	15.2	27.2	19.3	slightly overcast with blue patches, calm water, little wind, some sun, 9.6°C in Sault Ste. Marie
LT23	6.68	13.3	29.1	20.5	slightly overcast with blue patches, calm water, little wind, some sun, 9.6°C in Sault Ste. Marie
LT25	7.53	16.9	27.8	13.1	slight breeze, cloudy, overcast, rocky water, 13.3°C in Sault Ste. Marie
LT27	6.8	15.3	16.3	10.5	slight breeze, cloudy, overcast, rocky water, 13.3°C in Sault Ste. Marie
LT29	6.81	15	29,.6	20.9	slight breeze, cloudy, overcast, rocky water, 13.3°C in Sault Ste. Marie
LT47	7.85	19.5	29.9	21.5	cool, overcast, calm current, soft wind, 18.9°C in Sault Ste. Marie

LT49	6.97	16.7	29	20.6	cool, overcast, calm current, soft wind, 18.9°C in Sault Ste. Marie
LT51	6.71	15	29.3	20.3	cool, overcast, calm current, soft wind, 18.9°C in Sault Ste. Marie
LT53	6.69	14.5	28.3	20.4	cool, overcast, calm current, soft wind, 18.9°C in Sault Ste. Marie
LT55	7.46	21.7	26.7	19	sunny, little cloudy, humid, 25.3°C in Sault Ste. Marie
LT57	6.79	18.7	28.1	19.7	sunny, little cloudy, humid, 25.3°C in Sault Ste. Marie
LT59	6.63	17.8	30	21.5	sunny, little cloudy, humid, 25.3°C in Sault Ste. Marie
LT61	6.59	18.4	29.4	20.9	sunny, little cloudy, humid, 25.3°C in Sault Ste. Marie
LT63	7.31	22.9	26.5	19	sunny, no wind, humid, 29.1°C in Sault Ste. Marie
LT65	6.32	19.1	28.3	19.7	sunny, no wind, humid, 29.1°C in Sault Ste. Marie
LT67	6.23	17.1	29.2	21	sunny, no wind, humid, 29.1°C in Sault Ste. Marie
LT69	6.18	16.9	29.8	21.2	sunny, no wind, humid, 29.1°C in Sault Ste. Marie
LT127	7.31	20.8	39.8	28.4	sunny (no clouds), no wind, warm, calm water, 21.9°C in Sault Ste. Marie
LT129	6.49	18.3	29	20.4	sunny (no clouds), no wind, warm, calm water, 21.9°C in Sault Ste. Marie
LT131	6.32	16.7	27.8	19.9	sunny (no clouds), no wind, warm, calm water, 21.9°C in Sault Ste. Marie

LT133	6.5	19	21.4	15.1	sunny (no clouds), no wind, warm, calm water, 21.9°C in Sault Ste. Marie
LT135	7.08	22.8	55.8	32	sunny, white clouds, calm water, slight wind, warm, 24.9°C in Sault Ste. Marie
LT137	6.15	22.3	29.8	21.1	sunny, white clouds, calm water, slight wind, warm, 24.9°C in Sault Ste. Marie
LT139	6.58	20.8	27.2	19.1	sunny, white clouds, calm water, slight wind, warm, 24.9°C in Sault Ste. Marie
LT141	6.62	20.6	27.2	19.3	sunny, white clouds, calm water, slight wind, warm, 24.9°C in Sault Ste. Marie
LT143	7.16	23.4	28.7	20.5	sunny, clouds, windy, 22.6°C in Sault Ste. Marie
LT145	6.67	22.1	34.3	24	sunny, clouds, windy, 22.6°C in Sault Ste. Marie
LT147	6.43	20.6	29.1	20.2	sunny, clouds, windy, 22.6°C in Sault Ste. Marie
LT149	6.78	22.1	27.1	19.3	sunny, clouds, windy, 22.6°C in Sault Ste. Marie
LT175	7.81	21.6	29.1	20.8	sunny, slight wind, calm, some clouds, blue sky, calm water, 19.3°C in Sault Ste. Marie
LT177	7.5	20.4	27.8	19.8	sunny, slight wind, calm, some clouds, blue sky, calm water, 19.3°C in Sault Ste. Marie
LT179	7.14	18.3	28.9	20.6	sunny, slight wind, calm, some clouds, blue sky, calm water, 19.3°C in Sault Ste. Marie
LT181	7.19	18.4	27.8	19.6	sunny, slight wind, calm, some clouds, blue sky, calm

					water, 19.3°C in Sault Ste. Marie
LT183	7.6	22.7	27.3	19.4	cloudy, no wind, calm, some clouds, blue sky, calm water, 23.8°C in Sault Ste. Marie
LT185	6.94	20.4	28.8	20.4	cloudy, no wind, calm, some clouds, blue sky, calm water, 23.8°C in Sault Ste. Marie
LT187	6.8	21	29.3	20.8	cloudy, no wind, calm, some clouds, blue sky, calm water, 23.8°C in Sault Ste. Marie
LT189	6.36	19.4	28.8	19.7	cloudy, no wind, calm, some clouds, blue sky, calm water, 23.8°C in Sault Ste. Marie
LT191	7.74	23.1	28.3	20.3	cloudy, calm water, slight wind, more darker clouds, slight drizzle before getting on boat, 23.7°C in Sault Ste. Marie
LT193	7.22	20.7	28	19.9	cloudy, calm water, slight wind, more darker clouds, slight drizzle before getting on boat, 23.7°C in Sault Ste. Marie
LT195	6.95	19.6	29.3	20.4	cloudy, calm water, slight wind, more darker clouds, slight drizzle before getting on boat, 23.7°C in Sault Ste. Marie
LT197	6.97	19.1	27.2	19.5	cloudy, calm water, slight wind, more darker clouds, slight drizzle before getting on boat, 23.7°C in Sault Ste. Marie

Appendix B: Reference Sequences and Taxonomic Assignments from Bioinformatic Analyses

Table B1. Reference sequences with associated accession numbers included in the cyanobacterial phylogenetic tree.

Accession Number	Assigned taxonomy	Database
MG569710.1	*Aphanizomenon gracile* NRC JIE/D4-09	NCBI
MH107238.1	*Aphanizomenon gracile* AQUAMEB-37	Cyanoseq
DQ526414.1	*Synechococcus* sp. LS0501	NCBI
AF216950.1	*Synechococcus* PCC9005	NCBI
JN998126.1	*Synechococcus elongatus* ARKK1	Cyanoseq
NR_074309.1	*Synechococcus elongatus* PCC 6301	Cyanoseq
AF216945.1	*Cyanobium* sp. PCC 7009	Cyanoseq
NR_114406.1	*Cyanobium gracile* PCC 6307	Cyanoseq
GU368105.2	*Synechococcus* sp. N2b	NCBI
MT176754.1	*Cyanobium plancticum* BACA0224	NCBI
MF781035.1	*Cyanothece aeruginosa* SAG 87.79	Cyanoseq
AJ133171.1	*Microcystis aeruginosa* PCC 7941	Cyanoseq
NR_177769.1	*Chamaethrix vaginata*	NCBI
MH271078.1	*Calothrix* sp. Su 37	NCBI
LC534996.1	*Neolyngbya* sp. OdA1 gene	NCBI
KY824053.1	*Affixifilum granulosum* ALCB114378	NCBI

JF925324.1	*Leptolyngbya* sp. CYN82	NCBI
AJ781039.1	*Snowella litoralis* 0TU35S07	NCBI
LT670892.1	*Dolichospermum lemmermannii* NIVA-CYA 697	NCBI
OL310656.1	*Radiocystis geminata* TAU-MAC 1214	NCBI
AM710389.1	*Radiocystis sp* JJ30 3	NCBI
MT371827.1	*Affixifilum floridanum* BLCC-M61	NCBI
FJ589716.1	*Gloeocapsopsis crepidinum* LEGE 06123	NCBI
MK861877.1	*Leptodesmis alaskaensis* V20	NCBI
LC153790.1	*Pseudanabaena foetida* var. intermedia NIES-512 gene	NCBI
LC016776.1	*Pseudanabaena limnetica* gene	NCBI
NR_176486.1	*Limnococcus limneticus* Svet06	NCBI
AB039000.1	*Gloeocapsa* sp. PCC 73106 gene	NCBI
KY172947.1	*Gloeocapsa* sp. HG2	Cyanoseq
AB075995.1	*Pseudanabaena* sp. PCC 7403	NCBI
AB003165.1	*Phormidium mucicola* IAM M-221 gene	NCBI
NR_172653.1	*Anthocerotibacter panamensis* strain C109	NCBI
NR_176987.1	*Gloeobacter morelensis* strain MG652769	NCBI
NR_177760.1	*Ammassolinea attenuata*	NCBI
KY873318.1	*Pantanalinema rosaneae* CMAA1559	NCBI
NR_176504.1	*Capilliphycus salinus*	NCBI
MF072350.1	*Chroococcus subviolaceus* CCIBt3505	NCBI

KM376977.1	*Synechococcus elongatus* PUPCCC 010.5	NCBI
KP835530.1	*Synechococcus elongatus* CENA126	NCBI
JQ712606.1	*Phormidium formosum* P07	NCBI
AB039016.1	*Pseudanabaena* sp. PCC 6802	NCBI
FR798944.1	*Pseudanabaena* sp. 1a-03	NCBI
NR_177760.1	*Ammassolinea attenuata*	NCBI
KY550441.1	*Limnothrix mirabilis* AWQC-LIM001	NCBI
MT311250.1	*Trichocoleus caatingensis* CATCD2	NCBI
NR_172691.1	*Pegethrix indistincta* strain GSE-TBC-7GA	NCBI
J01874.1	*Proteus vulgaris*	NCBI
NR_024570.1	*Escherichia coli* strain U 5/41	NCBI
NR_036794.1	*Klebsiella pneumoniae* strain DSM 30104	NCBI

Table B2. Taxonomic assignment of cyanobacteria amplicon sequence variants (ASVs) using SILVA.

ASV ID	Family	Genus	Species
ASV1	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV2	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV3	Nostocaceae	*Aphanizomenon* NIES81	Unclassified
ASV4	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV5	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV6	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV7	Nostocaceae	*Aphanizomenon* NIES81	Unclassified
ASV8	Nostocaceae	*Aphanizomenon* NIES81	Unclassified
ASV9	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV10	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV11	Nostocaceae	*Aphanizomenon* NIES81	Unclassified
ASV12	Nostocaceae	*Aphanizomenon* NIES81	*Anabaena* sp.
ASV13	Phormidiaceae	*Arthrospira* PCC-7345	*Neolyngbya granulosa*
ASV14	Caenarcaniphilales	*Caenarcaniphilales*	Unclassified
ASV15	Caenarcaniphilales	*Caenarcaniphilales*	Unclassified
ASV16	Caenarcaniphilales	*Caenarcaniphilales*	Unclassified
ASV17	Microcystaceae	*Chalicogloea* CCALA 975	Unclassified
ASV18	Microcystaceae	*Chalicogloea* CCALA 975	Unclassified
ASV19	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV20	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV21	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV22	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*

ASV23	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV24	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV25	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV26	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV27	Cyanobacteriia (Unclassified)	*Cyanobacteriia* (Unclassified)	*Rhabdogloea smithii*
ASV28	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV29	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV30	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV31	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV32	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV33	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV34	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV35	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV36	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV37	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV38	Cyanobiaceae	*Cyanobium* PCC-6307	*Cyanobium* sp.
ASV39	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV40	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV41	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV42	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV43	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV44	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV45	Cyanobiaceae	*Cyanobium* PCC-6307	*Cyanobium* sp.
ASV46	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV47	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV48	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified

ASV49	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV50	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV51	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV52	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV53	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV54	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV55	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV56	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV57	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV58	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV59	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV60	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV61	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV62	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV63	Cyanobiaceae	*Cyanobium* PCC-6307	*Cyanobium* sp.
ASV64	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV65	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV66	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV67	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV68	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV69	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV70	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV71	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV72	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV73	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV74	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV75	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV76	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV77	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified

ASV78	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV79	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV80	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV81	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV82	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV83	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV84	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV85	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV86	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV87	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV88	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV89	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV90	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV91	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV92	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV93	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV94	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV95	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV96	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV97	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV98	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV99	Cyanobiaceae	*Cyanobium* PCC-6307	Unclassified
ASV100	Gastranaerophilales	*Gastranaerophilales*	Unclassified
ASV101	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV102	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV103	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV104	Gloeocapsaceae	*Gleocapsa*	*Chroococcus subviolaceus*
ASV105	Gloeocapsaceae	*Gleocapsa*	*Chroococcus subviolaceus*

ASV106	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV107	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV108	Gloeocapsaceae	*Gleocapsa*	*Limnococcus limneticus*
ASV109	Gloeobacteraceae	*Gloeobacter* PCC-7421	Unclassified
ASV110	Gloeobacteraceae	*Gloeobacter* PCC-7421	Unclassified
ASV111	Gloeobacteraceae	*Gloeobacter* PCC-7421	Unclassified
ASV112	Gloeobacteraceae	*Gloeobacter* PCC-7421	Unclassified
ASV113	Gloeobacteraceae	*Gloeobacter* PCC-7421	Unclassified
ASV114	Phormidiaceae	*Kamptonema* PCC-6407	Unclassified
ASV115	Unknown Family	*Leptolyngbya* ANT.L52.2	*Leptolyngbya* sp.
ASV116	Leptolyngbyaceae	*Leptolyngbya* BN43	*Leptolyngbya* sp.
ASV117	Unknown Family	*Leptolyngbya* EcFYyyy-00	Unclassified
ASV118	Leptolyngbyaceae	*Leptolyngbyaceae*	unclassified cyanounclassified
ASV119	Leptolyngbyaceae	*Leptolyngbyaceae*	unclassified cyanounclassified
ASV120	Limnotrichaceae	*Limnothrix*	Unclassified
ASV121	Microcystaceae	*Microcystis* PCC-7914	Unclassified
ASV122	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV123	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV124	Microcystaceae	*Microcystis* PCC-7914	*Cyanothece aeruginosa*
ASV125	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV126	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV127	Microcystaceae	*Microcystis* PCC-7914	*Cyanothece aeruginosa*
ASV128	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV129	Microcystaceae	*Microcystis* PCC-7914	*Cyanothece aeruginosa*
ASV130	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV131	Microcystaceae	*Microcystis* PCC-7914	*Cyanothece aeruginosa*
ASV132	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV133	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.

ASV134	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV135	Microcystaceae	*Microcystis* PCC-7914	*Cyanothece aeruginosa*
ASV136	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV137	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV138	Microcystaceae	*Microcystis* PCC-7914	Unclassified
ASV139	Microcystaceae	*Microcystis* PCC-7914	Unclassified
ASV140	Microcystaceae	*Microcystis* PCC-7914	*Radiocystis* sp.
ASV141	Microcystaceae	*Microcystis* PCC-7914	Unclassified
ASV142	Leptolyngbyaceae	MIZ36	Unclassified
ASV143	Obscuribacteraceae	*Obscuribacteraceae*	Unclassified
ASV144	Obscuribacteraceae	*Obscuribacteraceae*	Unclassified
ASV145	Obscuribacteraceae	*Obscuribacteraceae*	Unclassified
ASV146	Pseudanabaenaceae	*Pseudanabaena* PCC-7429	Unclassified
ASV147	Pseudanabaenaceae	*Pseudanabaena* PCC-7429	*Pseudanabaena foetida*
ASV148	Pseudanabaenaceae	*Pseudanabaena* PCC-7429	*Pseudanabaena foetida*
ASV149	Pseudanabaenaceae	*Pseudanabaena* PCC-7429	*Pseudanabaena foetida*
ASV150	RD011	RD011	Unclassified
ASV151	RD017	RD017	Unclassified
ASV152	Microcystaceae	*Snowella* 0TU37S04	*Snowella litoralis*
ASV153	Microcystaceae	*Snowella* 0TU37S04	*Snowella litoralis*
ASV154	Microcystaceae	*Snowella* 0TU37S04	*Snowella litoralis*
ASV155	Microcystaceae	*Snowella* 0TU37S04	*Snowella litoralis*
ASV156	Microcystaceae	*Snowella* 0TU37S04	*Snowella litoralis*
ASV157	Pseudanabaenaceae	*Synechococcus* PCC-7502	Unclassified
ASV158	Leptolyngbyaceae	TG-45	Unclassified
ASV159	Synechococcaceae	Unclassified	Unclassified
ASV160	Synechococcaceae	Unclassified	Unclassified
ASV161	Pseudanabaenaceae	Unclassified	Unclassified
ASV162	Vampirovibrionaceae	*Vampirovibrionaceae*	Unclassified

ASV163	Vampirovibrionales	*Vampirovibrionales*	Unclassified
ASV164	Vampirovibrionales	*Vampirovibrionales*	Unclassified
ASV165	Leptolyngbyaceae	YT-3	Unclassified
ASV166	Unclassified	Unclassified	Unclassified
ASV167	Unclassified	Unclassified	Unclassified
ASV168	Unclassified	Unclassified	Unclassified
ASV169	Leptolyngbyaceae	Unclassified	Unclassified
ASV170	Unclassified	Unclassified	Unclassified
ASV171	Unclassified	Unclassified	Unclassified
ASV172	Leptolyngbyaceae	Unclassified	Unclassified
ASV173	Pseudanabaenaceae	Unclassified	Unclassified
ASV174	Pseudanabaenaceae	Unclassified	Unclassified