Spatiotemporal Shifts in Cyanobacterial Communities in a Northern Temperate Watershed – Applications of Next-Generation Sequencing and Implications for Monitoring and Climate Change Adaptation

by

Ellen S. Cameron

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Steven Wilhelm Professor, University of Tennessee
Supervisor(s)	Dr. Kirsten Müller Professor, University of Waterloo
	Dr. Monica Emelko Professor, University of Waterloo
Internal Members	Dr. Laura Hug Assistant Professor, University of Waterloo
	Dr. Jonathan Witt Associate Professor, University of Waterloo
Internal-External Member	Dr. Merrin Macrae Associate Professor, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

Ellen Cameron was the primary author for the work conducted in this thesis. Chapter 1 and 5 were not written for publication. The main body of this thesis consists of three chapters prepared as manuscripts with intention of publication. Exceptions to sole authorship of material are highlighted below.

Chapter 2

This research was conducted at the University of Waterloo using example data taken from subsequent thesis chapters. Dr. Philip Schmidt provided guidance on the statistical limitations of different normalization methods and aided in the development of repeatedly rarefying as a normalization technique. Ellen Cameron conducted data analysis and prepared the manuscript with Dr. Philip Schmidt under the supervision of Dr. Monica Emelko and Dr. Kirsten Müller. Benjamin Tremblay assisted in the design of functions and provided technical support in the development of the R package, *mirlyn*, presented in this chapter. A version of this work is available as a pre-print on the bioRxiv server as:

Cameron, E.S., Schmidt, P.J., Tremblay, B. J.-M., Emelko, M.B., Müller, K.M. (2020). To rarefy or not to rarefy: Enhancing microbial community analysis through next-generation sequencing. bioRxiv, doi: https://doi.org/10.1101/2020.09.09.290049

Chapter 3

This research was conducted at the Turkey Lakes Watershed Research Station with the support of Environment and Climate Change Canada in August 2018. Ellen Cameron conducted sample collection, sample processing, subsequent DNA extractions and data analysis. Ellen Cameron prepared the manuscript under the supervision of Dr. Kirsten Müller and Dr. Monica Emelko.

This work is in preparation for submission as:

Cameron, E.S., Emelko, M.B., Müller, K.M. Cyanobacterial populations in lakes are dynamic and their characterization requires consideration of diurnal and spatial variation.

Chapter 4

This research was conducted at the Turkey Lakes Watershed Research Station with the support of Environment and Climate Change Canada (July 2018 – May 2019) and Natural Resources Canada (June 2019 – January 2020) with arrangements for site access in 2019 by Dr. Kara Webster (Natural Resources Canada). Ellen Cameron conducted sampling in the spring and summer periods. Ellen Cameron processed all seasonal samples, conducted all other DNA extractions excluding Big Turkey 2019 samples, and data analysis. Dr. Mike Stone (University of Waterloo) and Dr. Jim Buttle (Trent University) provided valuable perspectives and context on the impact hydrology and watershed processes on bacterial community composition. Ellen Cameron prepared the manuscript with the support of Dr. Monica Emelko in development of a conceptual model and under the supervision of Dr. Kirsten Müller.

This work is in preparation for submission as:

Cameron, E.S., Emelko, M.B., Stone, M., Buttle., J., Webster, K.L., Müller, K.M. Early Seasonal Increases in Relative Abundance of Potentially Toxic Cyanobacteria: A concerning impact of Climate Change Protracting the Vernal Window in Northern Temperate Lakes?

Abstract

Cyanobacteria, a group of photosynthetic bacteria, threaten water quality and drinking water resources globally through the production of potent toxins and the formation of dense surface blooms. These bloom events are increasing in intensity, frequency, and duration due to warming climates and anthropogenic land use and require monitoring programs for water quality management. However, cyanobacteria vary both spatially and temporally and if sampling efforts do not reflect this variation, potentially toxic organisms may be undetected or underestimated. This thesis explores the spatiotemporal trends of cyanobacterial communities in a series of interconnected, oligotrophic lakes in a northern temperate watershed (Turkey Lakes Watershed; North Part, ON) using next-generation sequencing (NGS).

Next-generation sequencing of marker genes allows for rapid characterization of environmental communities and has become increasingly accessible, allowing for interdisciplinary applications. Optimal approaches in data handling and analysis are debated due to key challenges arising due to the data structure. Amplicon sequencing samples will vary in library sizes—the total number of reads—but this variation is not biologically meaningful and library sizes must be normalized to account for these differences. Rarefying, the process of subsampling to a normalized size, is frequently used to account for this variation but has been highly criticized due to the omission of valid data. To address the concerns of data omission, repeated iterations of rarefying were evaluated as a normalization technique in diversity analyses (Chapter 2). Repeatedly rarefying was demonstrated to characterize variation introduced through subsampling for applications in diversity analyses. This technique was implemented in the subsequent analysis of cyanobacterial communities in this thesis.

Cyanobacterial communities are dynamic exhibiting heterogeneity in their spatial and temporal distribution in lakes. This spatiotemporal variation is driven by environmental conditions and physical characteristics (e.g., cell size, cell density) of taxa and can

subsequently create challenges in monitoring. The spatiotemporal variation of cyanobacterial communities was characterized on both a diurnal scale (Chapter 3) and seasonal scale (Chapter 4) through amplicon sequencing of the V4 region of the 16S rRNA gene. Although the lakes in this study did not have visible bloom biomass, cyanobacterial sequences comprised up to 56% of the bacterial community and were frequently dominated by sequences classified as picocyanobacterial genera, which range from 0.2 – 2.0 µm in diameter. This dominance exemplifies the inability to rely on visual detection as a monitoring technique. In both studies, trends in the spatiotemporal variation varied between the lake sites due to differences in morphometry, thermal stratification and surrounding landscape processes demonstrating the impact of system specific characteristics on cyanobacterial dynamics. In combination with warming climates in temperate zones, cyanobacterial growth habits may change and appear as significant components of the bacterial community as early as May in oligotrophic lakes contrasting the previous perception of peak occurrence in the late summer requiring monitoring protocols to re-evaluate appropriate sampling time frames in temperate systems.

The research conducted in this thesis identifies key areas for developing ecologically relevant sampling guidelines for cyanobacterial monitoring in lakes. Monitoring protocols are frequently developed from characteristics of common bloom forming taxa resulting in reliance on visual observation of biomass at the surface of the water and focusing sampling efforts to the summer months when blooms typically occur. This research demonstrated the flaws in these assumptions and provides a discussion on appropriate recommendations. Specifically, cyanobacterial community dynamics were demonstrated to be impacted by system specific characteristics and sampling protocols must be tailored to reflect the (i) physicochemical characteristics of the system, and (ii) ecological community structure. The research presented herein demonstrates the need for re-evaluation of current guidelines due to shifts in cyanobacterial growth habits in response to warming climates, and the reported dominance of picocyanobacteria which may impose toxicity risks despite the absence of visible biomass.

Acknowledgements

I would like to acknowledge the funding for this research which was supported by the NSERC Network for Forested Drinking Water Source Protection Technologies (*forWater Network*) and two Queen Elizabeth II Graduate Scholarships in Science and Technology. The work conducted at Turkey Lakes Watershed was further supported by staff from Environment and Climate Change Canada (Mike McAulay) and Natural Resources Canada (Dr. Kara Webster, Jamie Broad, Stephanie Nelson) allowing for site access and providing assistance in sample collection.

Throughout this research, I have been provided with ample support and opportunities from my co-supervisors. I would like to thank Dr. Kirsten Müller for her continued support from my initial introduction to research as an undergraduate thesis student. I also extend my gratitude towards my co-supervisor, Dr. Monica Emelko for immersing me to the world of interdisciplinary collaboration in research. Under their guidance, I have been able to develop as an independent and confident researcher.

I would like to thank Dr. Philip Schmidt who traversed the treacherous territories of statistics in amplicon sequencing and for providing guidance in high quality manuscript preparation and the peer-review process.

I would like to extend my appreciation to the members of the Müller lab past and present. The emotional and intellectual support that this group of scientists provides is more than one could ever ask for. Specifically, I must express sincere gratitude to Timothy Shardlow, Vanessa Poletto-Borges, and Emilie Power for assistance in field sample collection and I can only hope that the drone of a vacuum pump no longer haunts their dreams. I am also grateful for the assistance of Yaryna Kudla in the DNA extraction of the Big Turkey summer 2019 samples.

Finally, I must express sincere gratitude to Benjamin Tremblay for his constant support. Tolerating my chaos is a greater challenge than graduate school itself.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

1.1.1 Cyanobacteria: An Overview

Lakes naturally have characteristics that can broadly be classified into chemical (e.g., pH, dissolved oxygen levels and nutrient content) and physical characteristics (e.g., temperature, Secchi depth [a measure of water clarity; Bukata et al., 1988], mean depth, volume and surface area; Järvinen et al., 2002; Quinlan et al., 2003). In combination, physicochemical characteristics set constraints on present aquatic biota by influencing environmental conditions and resource availability (Davison, 1991; Quinlan et al., 2003; Rhee and Gotham, 1981) which subsequently drives food web structure and ecosystem dynamics (Zadereev, 2017). These characteristics are influenced by external factors such as surrounding landscape (e.g., mineralogy, elevation, soil composition; Mountain et al., 2015) and hydrologic connectivity (Lapierre et al., 2015; Sass et al., 2007; Webster et al., 2000). Additionally, latitudinal placement within geographical regions further impact the environmental conditions (Cunha et al., 2016; Gillooly and Dodson, 2000; Smol et al., 2005) with higher latitudes expected to exhibit more severe changes as a result of climate change (Smol et al., 2005). These factors impact environmental conditions driving the physicochemical profiles of lakes observed seasonally (Cunha et al., 2016; Smol et al., 2005) and will subsequently drive the ecological trends observed in aquatic biological communities, including phytoplankton.

Within aquatic ecosystems, phytoplankton constitute a critical component contributing to approximately 50% of global primary production (Guschina and Harwood, 2006). These

organisms also synthesize essential compounds, including long-chain polyunsaturated fatty acids which contribute to necessary physiological requirements of many other organisms (Gladyshev et al., 2013; Masclaux et al., 2012). Phytoplankton constitute a diverse group of organisms with the shared feature of being located in suspension within the water column and having a phototrophic metabolism (Reynolds, 2006). This assemblage of organisms includes representatives from the bacterial and eukaryotic kingdoms (Reynolds, 2006) that constitute a range of sizes (Callieri and Stockner, 2002). Size classifications initially included macroplankton (200-2000 µm), microplankton (20-200 µm), nanoplankton (2-20 µm), but were revised in 1978 to include: picoplankton (0.2-2 µm) and femtoplankton (0.02-0.2 µm; Callieri and Stockner, 2002; Sieburth et al., 1978). While phytoplankton are critical to aquatic ecosystem structure and function, the work within this thesis will focus on the bacterial component of the phytoplankton, the Cyanobacteria.

Cyanobacteria, a group of photosynthetic bacteria, have a long evolutionary history resulting in the oxygenation of the Earth's atmosphere approximately 2.4 billion years ago (Demoulin et al., 2019) with reported microfossils and stromatolites in Australia dating back approximately 3.5 billion years (Demoulin et al., 2019; Van Kranendonk et al., 2003). However, the exact occurrence of these organisms during the Archean period is frequently debated due to difficulty in microfossil identification, and the potential for other microbial processes to result in the formation of stromatolites which are frequently associated with cyanobacteria (Demoulin et al., 2019). These organisms occupy a broad range of habitats globally ranging from terrestrial to aquatic ecosystems, and tropical to polar regions (Paerl, 2014; Paerl and Huisman, 2009). Cyanobacteria have a variety of adaptations that enable them to respond to environmental stress including the formation of tolerant resting cells, the

presence of photoprotective cellular pigments, nitrogen fixation pathways and the regulation of buoyancy in response to light and nutrient gradients (Huisman et al., 2018; Paerl, 2014). In addition, this group is known to form blooms, a term which refers to the visual accumulation of biomass resulting in discoloration of waters (Huisman et al., 2018). These blooms pose a variety of negative impacts to aquatic ecosystems and water quality including toxin production (Paerl and Huisman, 2009), reductions in light penetration through aggregation at the surface of waters (Anderson, 2009), alteration of food web dynamics, and development of anoxic conditions upon bloom decay (Huisman et al., 2018) making the research on cyanobacterial dynamics critical for water quality management.

Due to the range of negative impacts these organisms impose to water quality, research frequently focuses on the characterization of environmental conditions that promote growth and bloom formation. However, the environmental conditions influencing cyanobacterial blooms and toxin production are still not well understood and there is mixed consensus on the causes of such blooms, likely driven by the taxonomic diversity present within this group (Bertani et al., 2017; Griffith and Loik, 2010). For example, nutrient composition and quantity are key factors for cyanobacterial growth (Heisler et al., 2008; Paterson et al., 2017). Specifically, growth is frequently dependent on phosphorus availability (Schindler, 1977), a nutrient which is often limiting in freshwater systems (Hao et al., 2012) and in temperate regions (Meerhoff et al., 2012). Despite the common hypothesis that reductions in phosphorus will limit cyanobacterial growth and bloom formation, lakes that experience phosphorus limitation still experience bloom events (Paerl et al., 2016; Paterson et al., 2017) with increased frequency of reports in oligotrophic temperate lakes in Ontario (Winter et al., 2011). The increased reporting of cyanobacterial blooms being reported in oligotrophic lakes in Ontario

(Winter et al., 2011) requires further investigation and characterization of cyanobacterial community dynamics in low nutrient, temperate systems.

Lakes are exposed to many temporal changes (daily, seasonally and annually) in environmental conditions which collectively influences growth, physiology and distribution of species (Andersen et al., 2013; Davison, 1991; Moisan et al., 2002). These conditions include nutrient availability, lake morphometry, light availability, water column stability and water temperature (Dokulil, 2003). Community composition changes rapidly in response to environmental variation and can be observed through the seasonal succession of phytoplankton populations (Andersen et al., 2013; Andersen, 1992; Jaworska and Zdanowski, 2012) and heterogeneity in spatial distribution (Cyr, 2017; Pick and Agbeti, 1991). Temperature is a key driving factor in community structure due to physiological optima resulting in seasonal population succession in phytoplankton communities (Butterwick et al., 2005). However, characteristic seasonal community dynamics may be at risk due to climate change resulting in elevated water temperatures and an earlier onset of summer stratification promoting the occurrence of cyanobacterial populations (Jaworska and Zdanowski, 2012) due to elevated growth rates at higher temperatures (Yang et al., 2017). However, in addition to temperature, nutrient availability drives community composition (Andersson et al., 2015), with small sized picocyanobacteria thriving in low nutrient environments due to rapid nutrient uptake (Callieri and Stockner, 2002; Collos et al., 2009). These shifts in composition in response to environmental conditions demonstrates the potential for spatiotemporal variability in population abundances and community composition manifested in response to spatial, seasonal, or diurnal gradients of resource availability including nutrients and light. This

dynamic fluctuation in response to environmental gradients requires characterization to identify the potential implications for detection in monitoring protocols.

Cyanobacteria threaten drinking water source quality and the provision of safe drinking water through the production of compounds associated with unpleasant taste and odor, potent toxins (Burkholder et al., 2010). Additionally, these organisms may disrupt treatment processes by increasing coagulant demand, increasing sludge production and clogging filters and thus reducing filter run times (Burkholder et al., 2010). Certain cyanobacterial species are toxic and produce cyanotoxins, (e.g., microcystins, nodularins and anatoxins; Carmichael, 1994; Vasconcelos, 2001) while the reason for the production is not understood, it may include grazing deterrents (Schatz et al., 2007) to cellular communication and colony formation (Harke et al., 2016). Water sources containing these toxins put humans at risk from exposure through participation in recreational water activities or consumption of contaminated water (Graham et al., 2008) with the first acute cyanotoxin poisoning reported in scientific literature in 1878 and the anecdotal reports of toxic populations reported throughout history (Chorus and Bartram, 1999). Critically, toxic and non-toxic species cannot be visually differentiated (Gallina et al., 2017) but can be distinguished genetically due to the presence of synthetase genes, which may be detected using molecular methods (Christensen et al., 2021). Although toxicity within this group is frequently highlighted in common bloom forming taxa (e.g., *Aphanizomenon, Dolichospermum, Microcystis*; Huisman et al., 2018), toxin production has been previously reported in picocyanobacterial taxa with microcystin including *Synechocystis*, *Synechococcus*, *Aphanocapsa cumulus*, and *Cyanobium rubescens* (Śliwińska-Wilczewska et al., 2018) indicating the potential for water quality concerns in systems dominated by these organisms such as oligotrophic lakes.

In addition to cyanotoxins, cyanobacteria produce other compounds associated with water quality including lipopolysaccharides (Stewart et al., 2006) or compounds associated with taste and odor problems in drinking water sources (e.g., geosmin, 2-methylisoborneol; Watson, 2003). While lipopolysaccharides have demonstrated pathogenic effects in other gram-negative bacteria, the lipopolysaccharides of cyanobacterial taxa are weakly toxic comparatively with differences arising due to the taxonomic distance and differences between the groups (Stewart et al., 2006). Contrary to the toxic compounds, compounds associated with taste and odor problems in drinking water are not harmful (Watson, 2003). However, these compounds often result in consumer complaints (Graham et al., 2008; Ministry for the Environment and Ministry of Health., 2009) and potential association of taste and odor in drinking water with toxicity (McGuire, 1995). The management of these compounds is difficult without advanced treatment processes (Chapman, 2010), which are not commonly used at most water treatment plants, and the occurrence of these events is not well understood creating further challenges in control (Watson, 2010).

1.1.2 Monitoring Protocols

The potential impact of cyanobacterial populations on water quality and drinking water resources is increasingly necessitating the establishment of monitoring programs (Graham et al., 2008). However, consistent guidelines for monitoring are not readily available (Graham et al., 2008). These guidelines are frequently produced to aid in the development of cyanobacteria monitoring protocols by providing a general summary on these organisms (Graham et al., 2008) but may provide inaccurate or misleading information on the ecology and physiology of cyanobacteria (Colorado Lake and Reservoir Management, 2015). For example, monitoring guidance may falsely state that all cyanobacteria are toxic, blooms only occur in warmer

months and that cyanobacteria are the only algae capable of forming blooms (Colorado Lake and Reservoir Management, 2015). Misconceptions regarding when, where and why cyanobacteria are found and how they can be characterized can lead to incorrect ecosystem characterization subsequently generating a false sense of security regarding algae proliferation and associated risks to drinking water treatment and public health protection.

The utility of monitoring protocols is contingent upon sample collection. Water samples may be collected using a variety of equipment including plankton nets, water collection bottles and pumps (Ehrlich, 2010) and should not be prefiltered to prevent systematic losses (Callieri et al., 2012). Using these techniques, sampling may rely on surface samples, discrete depth samples or depth integrated samples (Graham et al., 2008). While surface sampling is typically used to sample surface scums (Graham et al., 2008), this approach ignores spatial heterogeneity over the depth of the water column and will easily cause underestimation or missed detection of potentially toxic cyanobacterial populations. A focus on sampling only the water surface is regularly perpetuated due to the common misconception of cyanobacterial abundances being maximal at the surface arising from applying characteristics of bloom forming taxa, such as *Microcystis*, to this diverse group of organisms as a homogeneous entity (Freeman et al., 2020). Integrated depth sampling provides an overall characterization of communities by accounting for the vertical variability in spatial distribution but does not provide spatial resolution as discrete depth sampling can (Ehrlich, 2010; Graham et al., 2008). Discrete depth sampling is typically employed only when the distribution of populations has been established or when a structure of interest such as water intake occurs at a specified depth but can easily become logistically intensive (Graham et al., 2008). Despite the intensive nature of discrete depth sampling, the level of spatial resolution that this technique

can provide is critical for being able to better predict periods of higher risk and treatment challenges. The identification of spatial distribution in systems of interest is critical for developing baseline knowledge which can be accomplished through the characterization of cyanobacteria at depths throughout the water column using discrete depth sampling.

Routine sampling of recreational water bodies and drinking water supplies regularly includes daily and weekly sampling focused on peak usage and when cyanobacterial events have previously occurred (Graham et al., 2008). Frequent sampling is required due to the highly dynamic nature of phytoplankton communities (Ehrlich, 2010). To gain optimal detailed information representing this dynamic variability, daily sample collection with multiple time points at multiple depths would be utilized (Ehrlich, 2010) to account for diurnal (Visser et al., 2005) and spatial (Hunter et al., 2008) variability. However, realistically, logistically intensive sampling programs are not always possible resulting in sampling being conducted weekly, biweekly, monthly, quarterly, or in a mixed program which focuses sampling efforts to high-risk periods (Ehrlich, 2010). Without previous knowledge on cyanobacterial population in the system of study, development of sampling protocols may not reflect the diurnal and spatial variability in this dynamic group of organisms, resulting in vast underestimation or completed missed detection of populations.

Characteristically in temperate ecosystems, cyanobacterial growth is associated with the mid-to-late summer and early fall resulting in sampling efforts focused to these periods (Chorus et al., 2000; Graham et al., 2008). However, some systems have reported cyanobacterial blooms under the ice during the winter (Wejnerowski et al., 2018). Despite these winter cyanobacterial blooms, monitoring is typically reduced or absent during the winter months (Ehrlich, 2010) arising due to misconceptions of winter ecosystems entering a state of

dormancy and due to the challenges associated with winter sampling (Felföldi et al., 2016; Hampton et al., 2015; Powers and Hampton, 2016). The absence of knowledge on cyanobacterial dynamics during periods of ice cover and known bloom events requires inclusion of ice-cover months in monitoring programs to explore system dynamics and advance understanding on winter cyanobacterial community processes.

Following sample collection, water samples may be used for microscopic identification and cell enumeration (Colorado Lakes and Reservoir Management, 2015) or photosynthetic pigment concentrations can be used to estimate algal biomass using spectrophotometry (Ehrlich, 2010). Although microscopy allows for rapid identification of taxonomic composition of phytoplankton communities, this technique relies on trained personnel and is further limited due to the inability to visually differentiate between toxic and non-toxic organisms (Westrick et al., 2010). The implementation of modern molecular methods, including PCR and DNA sequencing, allows for rapid and sensitive detection of organisms of interest in environmental samples (Burkholder et al., 2010) with potential utility in applications for water quality monitoring. Next-generation sequencing has revolutionized the ability to study DNA collected from environmental samples (Bartram et al., 2011; Hugerth and Andersson, 2017; Shokralla et al., 2012). While shotgun sequencing allows for the characterization of the entire community, including both taxonomic composition and functional gene profiles, it is not widely accessible due to the high sequencing cost and high computational power required for analysis (Clooney et al., 2016; Langille et al., 2013). In contrast, the relatively low cost of amplicon sequencing has increased accessibility and popularity of this technique (Clooney et al., 2016; Langille et al., 2013) in interdisciplinary fields including microbial ecology, food safety, wastewater remediation, forensics and

medicine (McLaren et al., 2019). Specifically, within the water industry, amplicon sequencing has been used to characterize and predict cyanobacterial blooms (Tromas et al., 2017), evaluate groundwater vulnerability to pathogen intrusion (Chik et al., 2020) and monitor treatment performance in diverse settings (Vierheilig et al., 2015) showing the potential utility of this technique in applied settings.

Amplicon sequencing provides the opportunity to rapidly distinguish between diverse species composition with no phenotypic differences (McQuillan and Robidart, 2017), as is the case for the cryptic picocyanobacteria (Callieri et al., 2012). Community structure and composition is highly dynamic and the use of genetic approaches in community analysis can contribute to the knowledge on how community composition changes in response to environmental conditions (Anantharaman et al., 2016) and the ecological function associated with biodiversity (Bohmann et al., 2014). For amplicon sequencing, the polymerase chain reaction (PCR) can be used to amplify chosen segments of a genome (Girones et al., 2010) using primers designed to target the genetic sequence for specific species and lineages or for universal taxa identification (Bohmann et al., 2014; Sherwood and Presting, 2007). For example, ribosomes are present in all living organisms, excluding viruses, (Tsukuda et al., 2017) and the small subunit rRNA genes are highly conserved, rarely experience horizontal gene transfer and contain both conserved and hypervariable genes providing phylogenetic and evolutionary information on organisms (Weisburg et al., 1991) and taxonomic classification (Quast et al., 2013). Specifically, the 16S rRNA gene is the standard used for cyanobacteria and bacterial identification (Genuário et al., 2016; Yang et al., 2016) with well-developed databases available (Hugerth and Anderson, 2017).

Although amplicon sequencing provides the opportunity to characterize microbial communities without previous challenges of lab cultivation and microscopic identification (Girones et al., 2010; Yang et al., 2016), amplicon sequencing data is statistically complicated (Weiss et al., 2017). Amplicon sequencing datasets will have a total number of sequencing reads, known as the library size, which represents a fixed size random sample of amplified DNA fragments and does not provide absolute abundance of sequence variants (Gloor et al., 2016; Gloor et al., 2017). Library sizes between different samples in a single sequencing run can vary widely and is not representative of biological variation (McMurdie and Holmes, 2014) disallowing for raw sequence reads to be compared directly and requiring library size normalization prior to analysis (Gloor et al., 2016). While 16S rRNA sequencing has been accepted as a gold standard in taxonomic marker gene analysis, the complimentary data handling and statistical analysis of amplicon data has significantly lagged. Researchers are presented with the challenge of navigating often confounding literature in determining the most appropriate analysis option. As amplicon sequencing continues to traverse interdisciplinary boundaries, it is critical to realize that obtaining sequence data is only the first step towards microbial community characterization and that data handling may impact downstream analyses and data interpretation.

1.2 Study Site: The Turkey Lakes Watershed

The research conducted in this thesis was conducted at the Turkey Lakes Watershed (Figure 1.1). The Turkey Lakes Watershed (TLW) study began in 1980 with the initial purpose of exploring the impact of atmospheric deposition of acidifying substances on aquatic and terrestrial habitats with collaborations between Environment Canada, Natural Resources Canada, Fisheries and Oceans Canada and several universities (Jeffries et al., 1988; Jeffries

and Foster, 2001). The watershed is located approximately 50 km north of Sault Ste Marie, Ontario on the northern margin of the Great Lakes-St Lawrence Forest region and is 10.5 km^2 in area (Jeffries et al., 1988; Jeffries and Foster 2001). Four interconnected oligotrophic lakes are located within the watershed: Batchwana Lake, Wishart Lake, Little Turkey Lake and Big Turkey Lake (Jeffries et al., 1988).

This chain of lakes exhibits a gradient in environmental conditions within a single watershed and has provided previous research opportunities for various research including studies exploring the relationship between primary production and chemical composition and production of fish in a cascading lake system (Jeffries et al., 1988). The site is relatively undisturbed apart from logging activity that occurred in the 1950s (Jeffries et al., 1988) and a controlled forest harvesting experiment conducted in 1997 (Lindsay et al., 2004). Additionally, the watershed is largely unoccupied and is not subjected to the impacts of human land use (Jeffries et al., 1988). The relief ranges widely within the watershed with the lowest point at 340 m to the highest point atop Batchwana Mountain measuring 630 m and an average relief of 290 m (Jeffries et al., 1988). The watershed is underlain by Precambrian silicate greenstone and is located on the Canadian Shield with varying amounts of glacial till (Jeffries et al., 1988; Jeffries and Foster, 2001). The watershed is occupied by an uneven aged mature to overmature forest with old growth hardwood system dominated by sugar maple and yellow birch (Jeffries and Foster 2001).

The headwaters of Batchwana Lake are divided into a distinct northern and southern basin. The four lakes have different characteristics with varying drainage areas (24.0 – 803 Ha), lake surface area (5.88 – 52 ha), maximum depth (4.5 – 37 m), mean depth (2.19 – 12.2 m), volume (1.90 – 63.4 $m^3 * 10^5$) and water renewal time (0.15 – 1.3 yr) as summarized in

Table 1.1 (Jeffries et al., 1988). The outflow of Batchwana Lake is Norberg Creek which experiences a rapid change in elevation from 497 m to 388 m prior to entering Wishart Lake (Jeffries et al., 1988). Water flows from Wishart to Little Turkey and finally enters Big Turkey Lake where the outflow enters Batchwana River and subsequently into Lake Superior (Jeffries et al., 1988). The high precipitation causes high flushing and short water renewal times of the lakes with the shortest water renewal time observed in Wishart and the longest in Big Turkey (Jeffries et al., 1988). Typically, the lakes experience two periods of thermal stratification with direct water column stratification occurring from mid-May to October (Figure A1), and inverse stratification occurring during periods of ice cover from December to April (Jeffries et al., 1988). In addition to this, Wishart Lake is frequently well-mixed through the ice-free period due to the shallow depth profile (Jeffries et al., 1988). This study specifically explores the cyanobacterial communities in Big Turkey, Little Turkey and Wishart Lake to contrast the impacts of water column stability, maximum depth, drainage area and hydrologic position on community composition.

Figure 1.1: The Turkey Lakes Watershed location and lake sites. The watershed is located approximately 50 km north of Sault Ste Marie, Ontario as indicated by the marker on the map. The watershed consists of 4 interconnected basins visualized here using topographic maps generated through the Ontario Ministry of Natural Resources & Forestry.

Table 1.1 Summary of characteristics of the lakes of Turkey Lakes Watershed adapted from Jeffries et al. (1988).

Lake	Drainage Basin Area (ha)	Lake Surface Area (ha)	Maximum Depth (m)	Mean Depth (m)	Lake Volume (m^3 * 10^5)	Water Renewal Time (yr)
Upper Batchwana	24.0	5.88	11.3	3.87	2.27	1.3
Lower Batchwana	85.6	5.82	10.9	3.27	1.90	0.30
Wishart	337	19.2	4.5	2.19	4.21	0.15
Little Turkey	491	19.2	13.0	6.04	11.6	0.25
Big Turkey	803	52.0	37.0	12.2	63.4	0.94

Median total phosphorus concentrations have previously ranged from 0.16 to 0.19 µmol/L and total nitrogen from 29 to 39 µmol/L making phosphorus the limiting nutrient in this system (Jeffries et al., 1988). The low phosphorus of these lakes classifies them as oligotrophic, typical of lakes located on the Canadian Shield (Jeffries et al., 1988). Previous phytoplankton community characterization performed in 1980 revealed that cyanobacteria were the dominant algal species in all lakes (Jeffries et al., 1988). Specifically, *Merismopedia punctata* was the major taxa in Batchwana Lake contributing solely to the peak observed in the summer (Jeffries et al., 1988). Similarly, in Wishart Lake, *M. punctata* was also abundant but an increase in *Microcystis flos-aquae* was also observed in August (Jeffries et al., 1988). Big Turkey and Little Turkey Lake showed similar composition in phytoplankton

communities. *Chroococcus dispersus* always composed a large portion of phytoplankton communities. However seasonal variation in cyanobacterial communities were observed with high abundances of *Aphanothece* in July, *Microcystis* in August and *Coelosphaerium* in September (Jeffries et al., 1988). In all lakes, few organisms were present in the colder months with <1000 cells/mL detected, but these communities commonly included representatives of chryosphytes, diatoms, green algae, dinoflagellates and cyanobacteria (Jeffries et al., 1988). The previous reported dominance of cyanobacteria in this series of interconnected lakes provides the opportunity to further explore the spatiotemporal dynamics of cyanobacterial populations in oligotrophic systems. The characterization of phytoplankton conducted in 1980 likely relied on microscopic identification and may have utilized collection with plankton nets both resulting in the potential for underestimation and non-identification of picocyanobacterial species which may dominate in oligotrophic lakes such as these. The use of NGS presents a rapid and sensitive technique to characterize the genetic diversity present within the modern-day cyanobacterial communities of the lakes of the TLW.

As a result of climate warming, changes in the hydrological cycle and vernal and autumnal windows have been observed at the TLW. The end of snowpack was recorded to range from April 2 to May 3 with the start of snowpack ranging from October 31 to December 11 (Creed et al., 2015). The initiation of spring greening was observed from April 27 to May 17 with the end of season senescence in autumn ranging from September 25 to October 28 (Creed et al., 2015). The vernal window ranged from 8 to 37 days in length and the autumnal window from 3 to 62 days (Creed et al., 2015). The growing season length has been observed to increase with climate warming lasting later into the year, but width of the vernal and autumnal windows was not observed in the study (Creed et al., 2015). Significant increases in

annual air temperature have been observed at a rate of 0.6°C per decade and have been related to changes observed in the hydrological cycle including declines in precipitation at a rate of 82.2 mm per decade and total annual discharge decreasing at a rate of 109.6 mm per decade (Creed et al., 2015). Climate warming has been observed across all months but changes in the precipitation were specifically observed in the autumn with decreased precipitation occurring in August and September and increased precipitation in October (Creed et al., 2015). The impacts of climate warming in temperature changes and influences on the hydrological cycle within the watershed that have previously been reported will allow for a discussion on the subsequent impacts of climate change on cyanobacterial community structure.

1.3 Research Objectives

The increasing occurrence, intensity and duration of cyanobacterial blooms globally warrants further investigation into the dynamic variability present in this group of organisms. While cyanobacterial growth dynamics have largely been associated with warm temperature, high light availability, and high nutrient eutrophic systems, it is critical to further develop our understanding on the dynamics of these populations in nutrient limited, oligotrophic systems due to the increasing frequency of cyanobacterial blooms in oligotrophic lakes in Ontario (Winter et al., 2011). Building upon the previous research conducted at the TLW, these lakes provide a unique opportunity to explore cyanobacterial community composition in a series of interconnected lakes which have previously shown cyanobacteria dominance despite their oligotrophic status. The overarching goal of this thesis is to characterize the spatiotemporal trends in cyanobacterial communities in northern temperate lakes of the TLW. Characterization of spatiotemporal variability in cyanobacterial communities will advance

knowledge on system specific responses which can be utilized in forwarding the developments of guidelines for monitoring.

This research was developed in a hierarchical approach to initially evaluate data handling techniques and to finally build up to long-term seasonal trends observed within interconnected lakes within the same watershed. Three independent but complimentary manuscripts which are currently in submission or in preparation for submission to peer-reviewed journals served to address the following research questions herein this thesis:

RQ1: *What is the impact of rarefying as an amplicon sequencing library normalization technique in diversity analyses?*

First, rarefying as a library normalization tool was evaluated. While previous literature has criticized the use of rarefying due to the omission of valid data (McMurdie and Holmes, 2013), other techniques are associated with the challenge of artificially augmenting data with pseudocounts to correct for the high frequency of zero counts (Gloor et al., 2016). Although rarefying has received criticism as a normalization technique, the full utility of the technique through use of repeated iterations and the impact of library size selection has not been previously explored. Prior to delving into the characterization of cyanobacterial communities within the lakes of the TLW, Chapter 2 evaluates the application of repeatedly rarefying as a library normalization technique to explore RQ1. The normalization techniques developed in this work were also prepared as an R package, *mirlyn*, to increase accessibility within the scientific community. Following the analytical review conducted on rarefying as a normalization technique, the remainder of this thesis serves to characterize the spatiotemporal variation in cyanobacterial community composition within the lakes of the TLW.

RQ2: *How does the distribution of planktonic cyanobacteria fluctuate within a stratified and non-stratified lake in the TLW and consequently, how will sampling time impact detection in monitoring protocols?*

The diurnal migrations of cyanobacteria have previously been characterized in limnological studies. However, studies characterizing diurnal variation using next-generation sequencing techniques have been limited in scope (e.g., inclusion of only one sampling day per period; Shahraki et al., 2020) or number. In Chapter 3, the diurnal variation of cyanobacterial communities in lakes with varying water column stability was evaluated using amplicon sequencing. Due to the direct implications that diurnal migrations may impose to monitoring protocols, common monitoring recommendations were evaluated to identify limitations of current guidelines due to cyanobacterial ecology. Through characterization of diurnal variation and critical review on available monitoring guidelines, RQ2 was able to be explored fully, identifying the potential impacts of sampling time in detection and protocol design.

RQ3: *Do seasonal variation and spatial distribution vary between lakes within the same watershed as a result of abiotic characteristics of the system (e.g., depth, drainage area, water renewal time)?*

In addition to the diurnal variation present in cyanobacterial communities, these populations also undergo seasonal fluctuations driven by changing environmental conditions associated with meteorological conditions (Fanesi et al., 2016). Furthermore, the physicochemical characteristics of lakes are largely impacted by their interaction with the terrestrial landscape (Mountain et al., 2015) and the hydrologic connectivity (Lapierre et al., 2015; Sass et al., 2007; Webster et al., 2000). However, environmental conditions and

hydrological cycling are changing in response to climate change (Creed et al., 2015). In Chapter 4, a spatial and seasonal profile of the cyanobacterial communities in the downstream lakes of the TLW were characterized. This study uniquely included sampling during ice-cover to further advance understanding on the winter dynamics of cyanobacteria. The lakes of the TLW provided the unique opportunity to examine three lakes with varying physicochemical characteristics including depth, drainage area and water renewal time to identify the impacts of lake morphometry on cyanobacterial community structure.

The research conducted in these chapters serves to advance the knowledge on cyanobacterial dynamics in oligotrophic lakes and contributes to the research program at the TLW. However, in addition to the ecological insights that this research provides, there are direct applications for using these findings for further development of monitoring guidelines and the applications of NGS for cyanobacterial community characterization. A summary of this work and the significance of this research for applications within water quality monitoring is presented in Chapter 5.

Chapter 2
Enhancing Diversity Analysis of Microbial Communities Through Next-Generation Sequencing and Rarefying Repeatedly

2.1 Introduction

Next-generation sequencing (NGS) has revolutionized the analysis of environmental systems through the characterization of microbial communities and their function by using DNA collected from samples that contain mixed assemblages of organisms (Bartram et al., 2011; Hugerth and Andersson, 2017; Shokralla et al., 2012). Fewer than 1% of species in the environment can be isolated and cultured, limiting the ability to identify rare and difficult-to-cultivate members of the community (Bodor et al., 2020; Cho and Giovannoni, 2004; Ferguson et al., 1984). In addition to the limitations of culturing, microscopic evaluation of environmental samples remains of limited utility because of challenges in high-resolution taxonomic identification and the inability to infer function from morphology (Hugerth and Andersson, 2017). Metagenomic studies employ next-generation sequencing (NGS) technology to analyze large quantities of diverse environmental DNA (Thomas et al., 2012) and have reduced the challenges associated with culturing and microscopic identification in these contexts (McMurdie and Holmes, 2014).

Metagenomics encompasses a conglomerate of different sequencing experimental designs, including amplicon sequencing (sequencing of amplified genes of interest) and shotgun sequencing (sequencing of fragments of present genetic material). While shotgun sequencing allows characterization of the entire community, including both taxonomic composition and functional gene profiles, it is not widely accessible due to high sequencing costs and

computational requirements for analysis (Bartram et al., 2011; Clooney et al., 2016; Langille et al., 2013). In contrast, the relatively low cost of amplicon sequencing has made it an increasingly popular technique (Clooney et al., 2016; Langille et al., 2013). The amplification and sequencing of specific genes (e.g., taxonomic marker genes) enables characterization of microbial community composition (Hodkinson and Grice, 2015); as a result, it has been successfully applied in many areas of water research. For example, amplicon sequencing has been used to characterize and predict cyanobacteria blooms (Tromas et al., 2017), describe microbial communities found in aquatic ecosystems (Zhang et al., 2020), and evaluate groundwater vulnerability to pathogen intrusion (Chik et al., 2020). It has also been applied to water quality and treatment performance monitoring in diverse settings (Vierheilig et al., 2015), including drinking water distribution systems (Perrin et al., 2019; Shaw et al., 2015), drinking water biofilters (Kirisits et al., 2019), anaerobic digesters (Lam et al., 2020), and cooling towers (Paranjape et al., 2020).

Processing and analysis of amplicon sequencing data is statistically complicated for a number of reasons (Weiss et al., 2017). For example, library sizes (i.e., the total number of sequencing reads within a sample) can vary widely among different samples, even within a single sequencing run. The disparity in library sizes between samples may not represent actual differences in microbial communities (McMurdie and Holmes, 2014) and cannot be compared directly. For example, two replicate samples with 5,000 and 20,000 sequence reads, respectively, are likely to have different read counts for specific sequence variants simply due to the difference in library size. While parametric tools such as generalized linear modelling (McMurdie and Holmes, 2014) can provide a statistically sound framework for differential abundance analysis, drawing biologically meaningful diversity analysis conclusions from

amplicon sequencing data requires library sizes of data to be normalized to account for the artificial variation in counts between samples due to differences in library sizes (McKnight et al., 2019). Notably, a variety of normalization techniques that may affect the analysis and interpretation of results have been suggested including expressing counts as proportions of the total library size (McMurdie and Holmes, 2014), upper quartile log fold changes (e.g., *edgeR*; Robinson et al., 2009), centered log-ratio transformations (Gloor et al., 2017), geometric mean pairwise ratios (Chen et al., 2018), variance stabilizing transformations (e.g., *DESeq2*; Love et al., 2014), relative log expressions (Badri et al., 2018), and rarefaction (i.e., the process of rarefying libraries to a common size).

Rarefaction is a normalization tool initially developed for ecological diversity analyses to allow for sample comparison without associated bias from differences in sample size (Sanders, 1968). Rarefaction normalizes samples of differing sample size by subsampling each to a shared threshold, often equal to the smallest sample size (Willis, 2019). For samples that are larger than the threshold, data are randomly subsampled until the normalized library size is achieved. Although initially developed for use in ecological studies, rarefaction is a commonly used library size normalization technique for amplicon sequencing data but is frequently criticized (Gloor et al., 2017; McMurdie and Holmes, 2014; McKnight et al., 2019). Similar to the original employment in ecological studies, libraries are subsampled to create "rarefied" libraries of a consistent size among samples (Gloor et al., 2017; McMurdie and Holmes, 2014). Despite the prevalence of this technique, rarefying has been critiqued due to the artificial variation introduced through subsampling and the omission of valid data through loss of sequence counts or exclusion of samples with small library sizes (McMurdie and Holmes, 2014). However, rarefying is typically conducted in a single iteration and only provides a

snapshot of the community at the smaller normalized library size and is incapable of assessing the variability introduced through subsampling. Repeatedly rarefying has the potential to address the statistical concerns associated with omission of data and would provide a more statistically acceptable technique than performing a single iteration of rarefying for diversity analyses. Despite the criticism of rarefying, it is frequently used as a normalization technique, requiring further discussion on statistically appropriate approaches for analysis and interpretation of amplicon sequencing data.

Here, the application of rarefying as a library size normalization technique for diversity analyses is investigated to determine if criticisms of the technique remain when rarefying is implemented repeatedly, which allows evaluation of the variability introduced in subsampling from the original library size to a lower rarefied library size shared among all samples. While rarefying repeatedly has been explored superficially in previous research, this paper seeks to fully analyze the utility of repeated rarefaction and impacts of parameter selection on interpretation of diversity analyses. Specifically, this chapter addresses (i) appropriate usage of rarefying to characterize variation introduced through random subsampling, (ii) the impact of subsampling with or without replacement on diversity analysis, and (iii) the impact of library size selection on diversity analyses such as the Shannon index and Bray-Curtis dissimilarity ordinations. Different scenarios were generated to evaluate the impacts of rarefying library sizes on the interpretation of community diversity analyses.

2.2 Theory

2.2.1 An Overview of Amplicon Sequencing Analysis

Due to the inevitable interdisciplinarity of water research and the complexity and novelty of NGS relative to traditional microbiological methods used in water quality analyses, further detail on amplicon sequencing is provided. Amplification and sequencing of taxonomic marker genes has been used extensively to examine phylogeny, evolution and taxonomic classification of numerous groups across the three domains of life (Quast et al., 2013; Weisburg et al., 1991; Woese et al., 1990). Taxonomic marker genes include the 16S rRNA gene found in mitochondria, chloroplasts, bacteria and archaea (Case et al., 2007; Tsukuda et al., 2017; Weisburg et al., 1991; Yang et al., 2016), or the 18S rRNA gene within the nucleus of eukaryotes (Field et al., 1988). Widely-used reference databases have been developed containing marker gene sequences across numerous phyla (Hugerth and Andersson, 2017).

The 16S rRNA gene consists of nine highly conserved regions separated by nine hypervariable regions (V1-V9; Gray et al., 1984) and is approximately 1,540 base pairs in length (Kim et al., 2011; Schloss and Handelsman, 2004). While sequencing of the full 16S rRNA gene provides the highest taxonomic resolution (Johnson et al., 2019), many studies only utilize partial sequences due to limitations in read length of NGS platforms (Kim et al., 2011). Next-generation sequencing on Illumina platforms (Illumina Inc., San Diego, California) produces reads that are up to 350 base pairs in length, requiring selection of an appropriate region of the 16S rRNA gene to amplify and sequence for optimal taxonomic resolution (Bukin et al., 2019; Kim et al., 2011). Sequencing the more conservative regions of the 16S rRNA gene may be limited to resolution of higher levels of taxonomy, while more variable regions can provide higher resolution for the classification of sequences to the genus

and species levels in bacteria and archaea (Bukin et al., 2019; Kim et al., 2011; Yang et al., 2016).

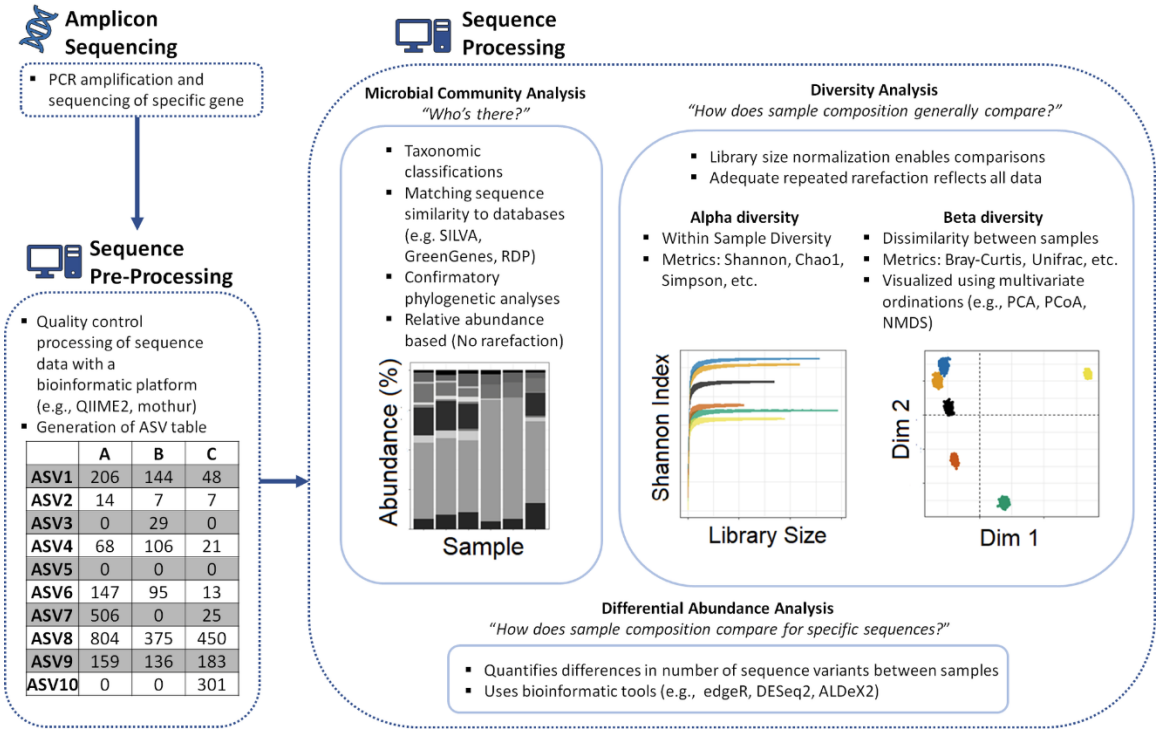
Different variable regions of the 16S rRNA gene may be biased towards different taxa (Johnson et al., 2019) and be preferred for different ecosystems (Escapa et al., 2020). For example, the V4 region has been shown to strongly differentiate taxa from the phyla Cyanobacteria, Firmicutes, Fusobacteria, Plantomycetes, and Tenericutes but the V3 region best differentiates taxa from the phyla Proteobacteria (e.g., *Escherichia coli*, *Salmonella*, *Campylobacter*), Acidobacteria, Bacteroidetes, Chloroflexi, Gemmatimonadetes, Nitrospirae, and Spirochaetae (Zhang et al., 2018). The V4 region of the 16S rRNA gene is frequently targeted using specific primers designed to minimize phylum amplification bias while accounting for common aquatic bacteria (Walters et al., 2015) and is frequently used in aquatic studies (Zhang et al., 2018). It is important to consider suitability of a 16S rRNA region for the habitat (Escapa et al., 2020) and the taxa present in the microbial community due to potential bias of analyzing differing subregions of the 16S rRNA gene (Johnson et al., 2019; Zhang et al., 2018).

The use of amplicon sequencing of partial sequences of the 16S rRNA gene allows examination of microbial community composition and the exploration of shifts in community structure in response to environmental conditions (Hodkinson and Grice, 2015), and identification of differentially abundant taxa between samples (Hugerth and Andersson, 2017). Amplicon sequencing datasets can be analyzed using a variety of bioinformatics tools for sequence analysis (e.g., sequence denoising, taxonomic classification, diversity analysis) including *mothur* (Schloss et al., 2009) and *QIIME2* (Bolyen et al., 2019). Previously, sequencing analysis involved the creation of operational taxonomic units (OTUs), dataset

dependent features, by clustering sequences into groups that met a certain similarity threshold, resulting in a loss of representation of variation in sequences and precluding cross-study comparison (Callahan et al., 2017). Advances in computational power have allowed a shift from use of OTUs to amplicon sequence variants (ASVs) representative of each unique sequence in a sample, which allows for the comparison of sequence variants generated in different studies and retains the representation of biological variation (Callahan et al., 2017). The implementation of tools included in bioinformatics pipelines, such as *DADA2* (Callahan et al., 2016) or *Deblur* (Amir et al., 2017), allows quality control of sequencing through the removal of sequencing errors and for the creation of ASVs (Amir et al., 2017; Callahan et al., 2016).

Taxonomic classification of 16S rRNA sequences using rRNA databases including SILVA (Quast et al., 2013), the Ribosomal Database Project (Cole et al., 2014) and GreenGenes (DeSantis et al., 2006) allows for construction of taxonomic community profiles (Bartram et al., 2011). Quality controlled sequencing data for a particular run is then organized into large matrices where columns represent experimental samples and rows contain counts for different ASVs (Weiss et al., 2017). Amplicon sequencing samples have a total number of sequencing reads known as the library size (McMurdie and Holmes, 2014), but do not provide information on the absolute abundance of sequence variants (Gloor et al., 2017, 2016). This data can be used for studies on taxonomic composition, differential abundance analysis and diversity analyses (Figure 2.1). Taxonomic composition analysis allows for characterization of microbial communities by classifying sequence variants based on similarities to sequences in online databases. The creation of taxonomic composition graphs frequently expresses community composition in proportions. Differential abundance analysis is utilized to explore

whether specific sequence variants are found in significantly different proportions between samples (Weiss et al., 2017) to identify potential biological drivers for these differences. This application is outside the scope of this article and is frequently performed using programs initially designed for transcriptomics such as *DESeq2* (Love et al., 2014) and *edgeR* (Robinson et al., 2009), or programs designed to account for the compositional structure of sequence data *ALDeX2* (Fernandes et al., 2014). The final potential application of this data, is diversity analyses which can be evaluated on varying scales from within sample (alpha) to between samples(beta; Sepkoski, 1988) but is associated with the challenge of the true diversity of environmental samples largely remaining unknown (Hughes et al., 2001).



	A	B	C
ASV1	206	144	48
ASV2	14	7	7
ASV3	0	29	0
ASV4	68	106	21
ASV5	0	0	0
ASV6	147	95	13
ASV7	506	0	25
ASV8	804	375	450
ASV9	159	136	183
ASV10	0	0	301

Figure 2.1 Schematic of general workflow in amplicon sequencing of samples. Amplicon sequencing utilizes PCR amplification of a specific gene of interest. Prior to conducting downstream analyses including microbial community analysis, differential abundance analysis and diversity analysis sequences must be pre-processed to generate the ASV feature table and taxonomic

classifications. This data can be used for downstream analyses to examine questions such as who's there, how does sample composition compare and how do counts for specific sequences compare.

Alpha diversity serves to identify richness (e.g., number of observed ASVs) and evenness (e.g., allocation of read counts across observed ASVs) within a sample (Willis, 2019). Comparison of alpha diversity among samples of differing library sizes may result in inherent biases, with samples having larger library sizes appearing more diverse due to the potential presence of more sequence variants in samples with larger libraries (Willis, 2019). This requires samples to have equal library sizes before comparison to prevent bias fabricated only from differences in initial library size (Willis, 2019). Diversity indices used to characterize the alpha diversity of samples include but are not limited to the Shannon index (Shannon, 1948), Chao1 index (Chao and Bunge, 2002), and the Simpson index (Simpson, 1949), but unique details of such indices should be understood for correct usage. For example, Chao1 relies on the observation of singletons in data to estimate diversity (Chao and Bunge, 2002), but denoising processes for sequencing data may remove singleton reads making the Chao1 estimator invalid for accurate analysis. The Shannon index, used in this study, is affected by differing library sizes because the contribution of rare sequences to total diversity is progressively lost with smaller library sizes.

Similar to alpha diversity, samples with differing library sizes in beta diversity analyses may produce erroneous results due to the potential for samples with larger library sizes to have more unique sequences simply due to the presence of more sequence variants (Weiss et al., 2017). A variety of different beta-diversity metrics can be used to compare sequence variant composition between samples including Bray-Curtis (Bray and Curtis, 1957) or Unifrac (Lozupone and Knight, 2007) distances and then visualized using ordination techniques (e.g.,

PCA, PCoA, NMDS). Bray-Curtis dissimilarity, used in this study, includes pairwise comparison of the numbers for each ASV between two samples, which are expected to be quite dissimilar if library sizes vary substantially.

Diversity analysis requires library size normalization to account for bias introduced through varying read counts in samples. For example, samples with larger library sizes may appear more diverse simply due to the presence of more sequences. Normalization methods include a variety of approaches ranging from rarefying to statistical transformations as discussed previously. However, McKnight et al. (2019) noted that the failure of most normalization techniques to transform data to equal library sizes "is discouraging, as standardizing read depths are the initial impetus for normalizing the data (i.e., if all samples had equal read depths after sequencing, there would be no need to normalize". The limitations of these normalization techniques in the application to diversity analyses are discussed in further detail in Section 2.2.2.

2.2.2 Limitations of Library Normalization Techniques

Diversity analysis, as it is presently applied, usually requires library size normalization to account for bias introduced through varying read counts in samples. For example, samples with larger library sizes may appear more diverse simply due to the presence of more sequences. Normalization techniques that feature various statistical transformations have been proposed for use in place of rarefying or proportions (McKnight et al., 2019), including upper-quartile log fold change (e.g., Robinson et al., 2009), centered log-ratio transformations (e.g., Gloor et al., 2017), geometric mean pairwise ratios (e.g., Chen et al., 2018), variance stabilizing transformations (e.g., Love et al., 2014) or relative log expressions (e.g., Badri et al., 2018). McKnight et al. (2019) noted that the failure of most

normalization techniques to transform data to equal library sizes for diversity analysis "is discouraging, as standardizing read depths are the initial impetus for normalizing the data (i.e., if all samples had equal read depths after sequencing, there would be no need to normalize".

These proposed alternatives to rarefying are also often compromised by the presence of large proportions of zero count data in tabulated amplicon sequencing read counts. Zero counts represent a lack of information (Silverman et al., 2018) and may arise from true absence of the sequence variant in the sample or a loss resulting in it not being detected when it was actually present (Tsilimigras and Fodor, 2016; Wang and LêCao, 2019). Nonetheless, many normalization procedures for amplicon sequencing datasets require zero counts to be omitted or modified, especially when applying transformations that utilize logarithms (e.g., centered log-ratio, relative log expressions, geometric mean pairwise ratios). Methods that utilize logarithms involve fabricating count values (pseudocounts) for the many zeros of which amplicon sequencing datasets are comprised and selecting a pseudocount value is an additional challenge (Weiss et al., 2017) that may be accomplished using probabilistic arguments (Gloor et al., 2016; 2017). Zeros are a natural occurrence in discrete, count-based data such as the counting of microorganisms or amplicon sequences and adjusting or omitting them can introduce substantial bias into microbial analyses (Chik et al., 2018).

McMurdie and Holmes (2014) noted that use of proportions is problematic due to heteroscedasticity: for example, one sequence read in a library size of 100 is a far less precise representation of source composition than 100 sequence reads in a library size of 10,000, even though both comprise 1% of the observed sequences. McKnight et al. (2019) favour use of proportions in diversity analysis without noting how precision of proportions, and the

degree to which alpha diversity in the source is reflected (Willis, 2019), varies with library size. Willis (2019) also points towards a conceptually better approach to diversity analysis that accounts for measurement error and the difference between the sample data and the population (environmental source) of which the sample data are only a partial representation. Diversity analysis in general does not do this, as it applies a set of calculations to sample data (or some transformation thereof) to obtain one value of alpha diversity or one point on an ordination plot. Pending further development of such approaches, this study revisits rarefying because of the practical simplicity of comparing diversity among samples of equal library size (Schmidt et al., 2021).

McMurdie and Holmes (2014) propose that rarefying is not a statistically valid normalization technique due to the omission of valid data, which may be resolved for the purposes of diversity analysis by rarefying repeatedly to represent all sequences in the proportions with which they were observed and compare sample-level microbial community diversity at a particular library size. In addition, McMurdie and Holmes (2014) dismissed repeatedly rarefying as a normalization technique, in part because repeatedly rarefying an artificial library consisting of a 50:50 ratio of two sequence variants does not yield a 50:50 ratio at the rarefied library size and this added noise could affect downstream analyses. However, such error is inherent to subsampling, whether from a population or from a larger sequence library and has thus already affected samples with smaller library sizes; it is the reason why simple proportions are less precise in samples with smaller library sizes.

McMurdie and Holmes (2014), also cited the investigation of Navas-Molina et al. (2013) as an example of repeatedly rarefying to normalize library sizes and used it to support their dismissal of this technique due to the omission of valid data and added variability. However,

it is critical to note that the work in Navas-Molina et al. (2013) reported using jackknife resampling of sequences, which cannot be equated to repeatedly rarefying (random resampling with or without replacement). Hence, it is necessary to build upon preliminary analysis of repeatedly rarefying as a normalization technique and to explore the impact of subsampling approach and normalized library size on diversity analysis results.

2.3 Methods

2.3.1 Example Data – DNA Extraction & Amplicon Sequencing

Samples used in the analyses are part of a larger study at Turkey Lakes Watershed (North Part, ON) but only an illustrative subset of samples is considered for the purpose of evaluating rarefaction and not for ecological interpretation. Further detail on the collection of these samples is presented in subsequent chapters in this thesis. The use of the subset of samples from a larger study to evaluate rarefying as a normalization technique avoids utilizing simulated data. DNA extracts isolated from environmental samples were submitted for amplicon sequencing using the Illumina MiSeq platform (Illumina Inc., San Diego, California) at the commercial laboratory Metagenom Bio Inc. (Waterloo, Ontario). Primers designed to target the 16S rRNA gene V4 region [515FB (GTGYCAGCMGCCGCGGTAA) and 806RB (GGACTACNVGGGTWTCTAAT; Walters et al., 2015)] were used for PCR amplification.

2.3.2 Sequence Processing & Library Normalization

The program *QIIME2* (v. 2019.10; Bolyen et al., 2019) was used for bioinformatic processing of sequence reads. Demultiplexed paired-end sequences were trimmed and denoised, including the removal of chimeric sequences and singleton sequence variants to avoid sequences that may not be representative of real organisms, using *DADA2* (Callahan et al., 2016) to construct the ASV table. Zeroing all singleton sequences could erroneously

remove legitimate sequences, particularly if the sequence in question is detected in large numbers in other similar samples; however, the potential effect of such error upon diversity analysis is beyond the scope of this work. Output files from *QIIME2* were imported into R (v. 4.0.1; R Core Team, 2020) for community analyses using *qiime2R* (v. 0.99.23; Bisanz, 2018). Initial sequence libraries were further filtered using *phyloseq* (v. 1.32.0; McMurdie and Holmes, 2013) to exclude amplicon sequence variants that were taxonomically classified as mitochondria or chloroplast sequences. We developed a package called *mirlyn* (Multiple Iterations of Rarefaction for Library Normalization; Cameron and Tremblay, 2020) that facilitates implementation of techniques used in this study built from existing R packages (Appendix 2). Using the output from *phyloseq*, *mirlyn* was used to (1) generate rarefaction curves, (2) repeatedly rarefy libraries to account for variation in library sizes among samples, and (3) plot diversity metrics given repeated rarefaction.

2.3.3 Community Diversity Analyses on Normalized Libraries

The impact of normalized library size on the Shannon index, an alpha diversity metric, was evaluated. Normalized libraries were also used for beta diversity analysis. Hellinger-transformed data was used to calculate Bray-Curtis distances (Bray and Curtis, 1957). Principal component analysis (PCA) was conducted on the Bray-Curtis distance matrices.

2.3.4 Study Approach

Typically, rarefaction has only been conducted a single time in microbial community analyses, and this omits a random subset of observed sequences, introducing a possible source of error. To examine this error, samples were repeatedly rarefied 1000 times. This repetition provides a representative suite of rarefied samples capturing the randomness in sequence variant composition imposed by rarefying. The sections below address the various decisions

that must be made by the analyst and factors affecting reliability of results when rarefaction is used.

2.3.4.1 The Effects of Subsampling Approach – With or Without Replacement

Rarefying library sizes may be performed with or without replacement. To evaluate the effects of subsampling replacement approaches, we repeatedly rarefied filtered sequence libraries to varying depths with and without replacement. Results of the two approaches were contrasted in diversity analyses to evaluate the impact of subsampling approach on interpretation of results.

2.3.4.2 The Effects of Normalized Library Size Selection

Rarefying involves the selection of an appropriate sampling depth to be shared by each sample. To evaluate the effects of different rarefied library sizes, filtered sequence libraries were rarefied repeatedly to varying depths. Results for various sampling depths were contrasted in diversity analyses to evaluate the impact of normalized library size selection on interpretation of results.

2.4 Results & Discussion

2.4.1 Use of Rarefaction Curves to Explore Suitable Normalized Library Sizes

Rarefying requires the selection of a potentially arbitrary normalized library size, which can impact subsequent community diversity analyses and therefore presents users with the challenge of making an appropriate decision of what size to select (McMurdie and Holmes, 2014). Suitable sampling depths for groups of samples can be determined through the examination of rarefaction curves (Figure 2.2). By selecting a library size that encompasses the flattening portion of the curve for each sample, it is generally assumed that the normalized library size will adequately capture the diversity within the samples despite the exclusion of

sequence variants during the rarefying process (i.e., there are progressively diminishing returns in including more of the observed sequence variants as the rarefaction curve flattens).

Suggestions have previously been made encouraging selection of a normalized library size that is encompassing of most samples (e.g., 10,000 sequences) and advocation against rarefying below certain depths (e.g., 1,000 sequences) due to decreases in data quality (Navas-Molina et al., 2013). However, generic criteria may not be applicable to all datasets and exploratory data analysis is often required to make informed and appropriate decisions on the selection of a normalized library size. Although previous research advises against rarefying below certain thresholds, users may be presented with the dilemma of selecting a sampling depth that either does not capture the full diversity of a sample depicted in the rarefaction curve (Figure 2.2 – I) or would require the omission of entire samples with smaller library sizes (Figure 2.2 – III). The implementation of multiple iterations of rarefying library sizes will aid in alleviating this dilemma by capturing the potential losses in community diversity for samples that are rarefied to lower than ideal depth. Doing so with two or more normalized library sizes may reveal differences in diversity attributable to relatively rare variants that could be suppressed by normalizing to too small of a library size.

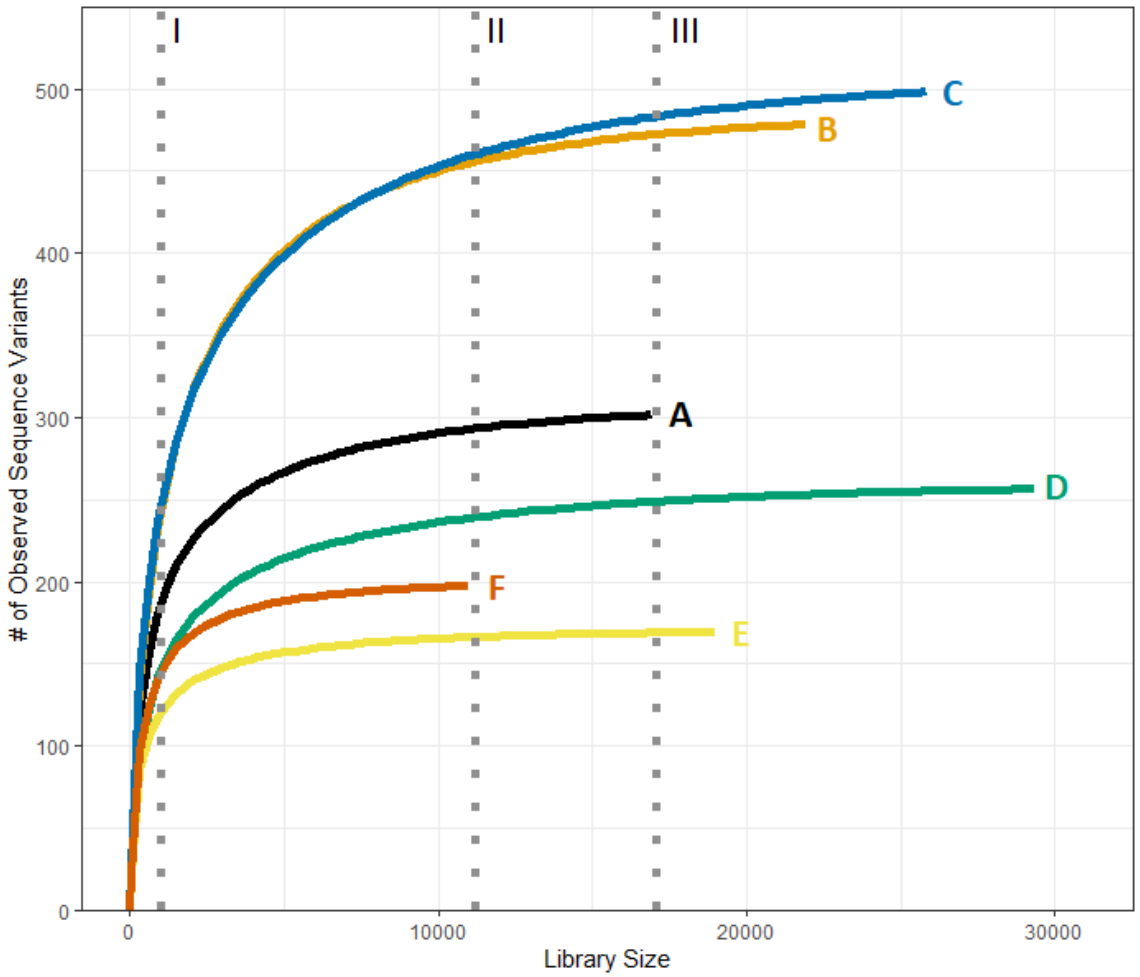


Figure 2.2 Rarefaction curves showing the number of unique sequence variants as a function of normalized library size for six samples (labelled A – F) collected from a freshwater northern temperate oligotrophic lake of varying diversity and initial library size. Selection of unnecessarily small library sizes (I) omits many sequence variants. Rarefying to the smallest library size (II) omits fewer sequences and variants. While selection of a larger normalized library size (III) would omit even less sequences, it is necessary to omit entire samples (e.g., Sample F) that have too few sequences).

2.4.2 The Effects of Subsampling Approach & Normalized Library Size Selection on Alpha Diversity Analyses

The differences in input parameters for rarefying samples requires users to be diligent in the selection of appropriate tools and commands for their analysis. The R package *phyloseq*,

a popular tool for microbiome analyses, has default settings for rarefying including sampling with replacement to optimize computational run time and memory usage (McMurdie and Holmes, 2013), although sampling without replacement is more appropriate statistically. Sampling without replacement draws a subset from the observed set of sequences (as though the sample had yielded only the specified library size), whereas sampling with replacement fabricates a set of sequences in similar proportions to the observed set of sequences (Figure 2.3). Sampling with replacement can potentially cause a rare sequence variant to appear more frequently in the rarefied sample than it was actually observed in the original library.

Rarefying libraries with or without replacement was not found to substantially impact the Shannon index in the scenarios considered in this study (Figure 2.4A), but users should still be aware of potential implications of sampling with or without replacement when rarefying libraries. Libraries rarefied with replacement are observed to have a slightly reduced Shannon index relative to libraries rarefied without replacement at many library sizes.

The conservation of larger normalized library sizes allows detection of more diversity with minimal variation observed between the iterations of rarefaction (Figure 2.4A). The largest considered normalized library size (the sample with the smallest library size has 11,213 sequences) captured the highest Shannon index values, while the Shannon index diminishes for all samples at lower normalized library sizes. The use of repeated iterations of rarefying allows variation introduced through subsampling to be represented in the diversity metric, which is small at larger library sizes. While there was only slight disparity in the Shannon index values between the largest library size and unnormalized data, this may not always be the case and is dependent on the sequence variant composition of the samples. Samples dominated by a large number of low-abundance sequence variants are more likely to have a

substantially reduced Shannon index value at a larger normalized library size. Alternatively, samples dominated by only several highly abundant sequence variants will be comparatively robust to rarefying. A plot of the Shannon index as a function of rarefied library size (Figure 2.4B) demonstrates the overall robustness of the Shannon index of these samples for larger library sizes (e.g., > 5,000 sequences) and the increased variation and diminishing values when proceeding to smaller rarefied library sizes. When the normalized library size was decreased to 5,000, the Shannon index is still only slightly reduced by the rarefaction but there is greater variability introduced from rarefying.

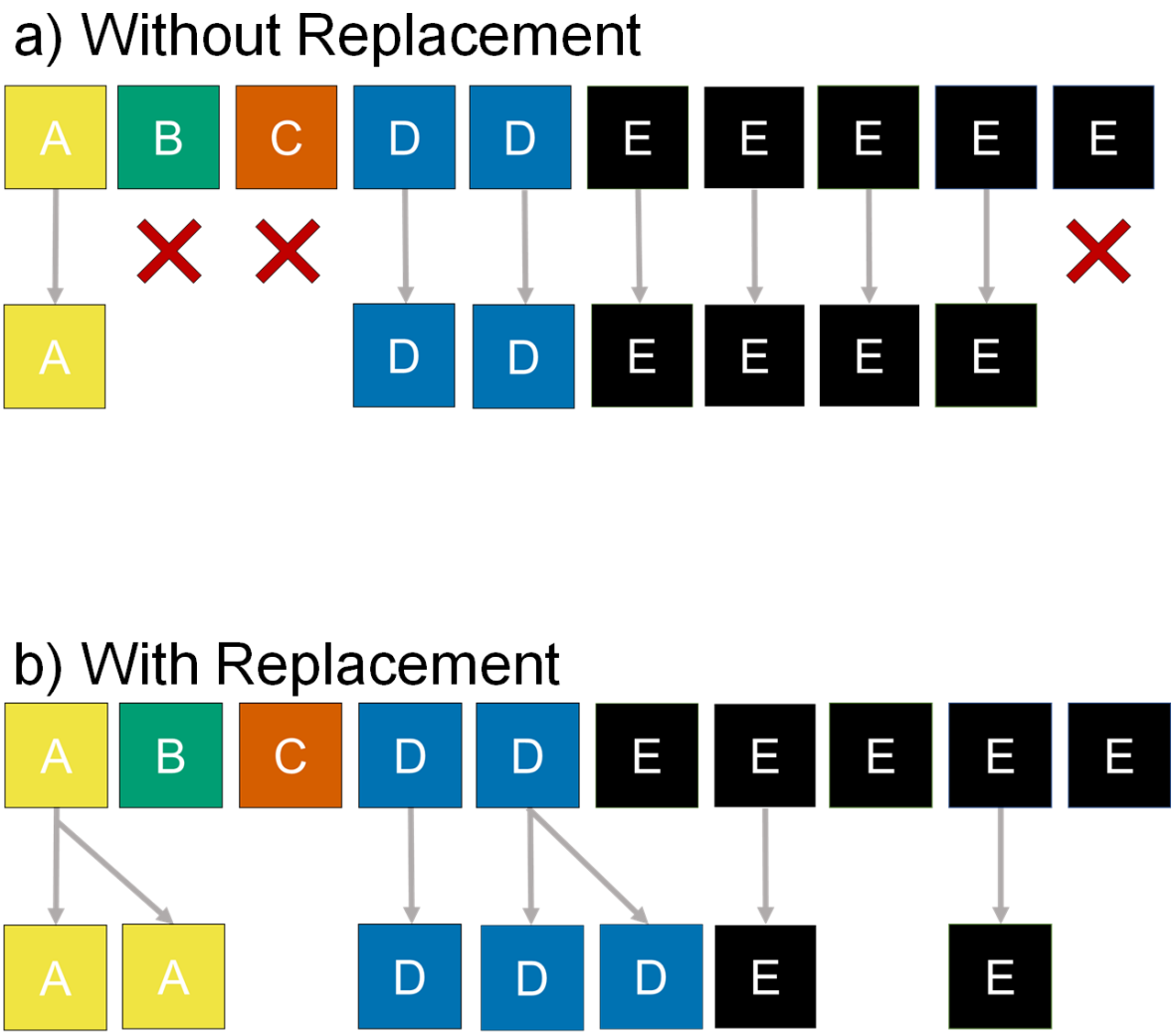


Figure 2.3 The mechanics of rarefying with or without replacement for a hypothetical sample with a library size of ten composed of five sequence variants (A – E). Rarefying without replacement (a) draws a subset from the observed library excluding the complementary subset, while rarefying with replacement (b) has the potential to artificially inflate the numbers of some sequence variants beyond what was observed.

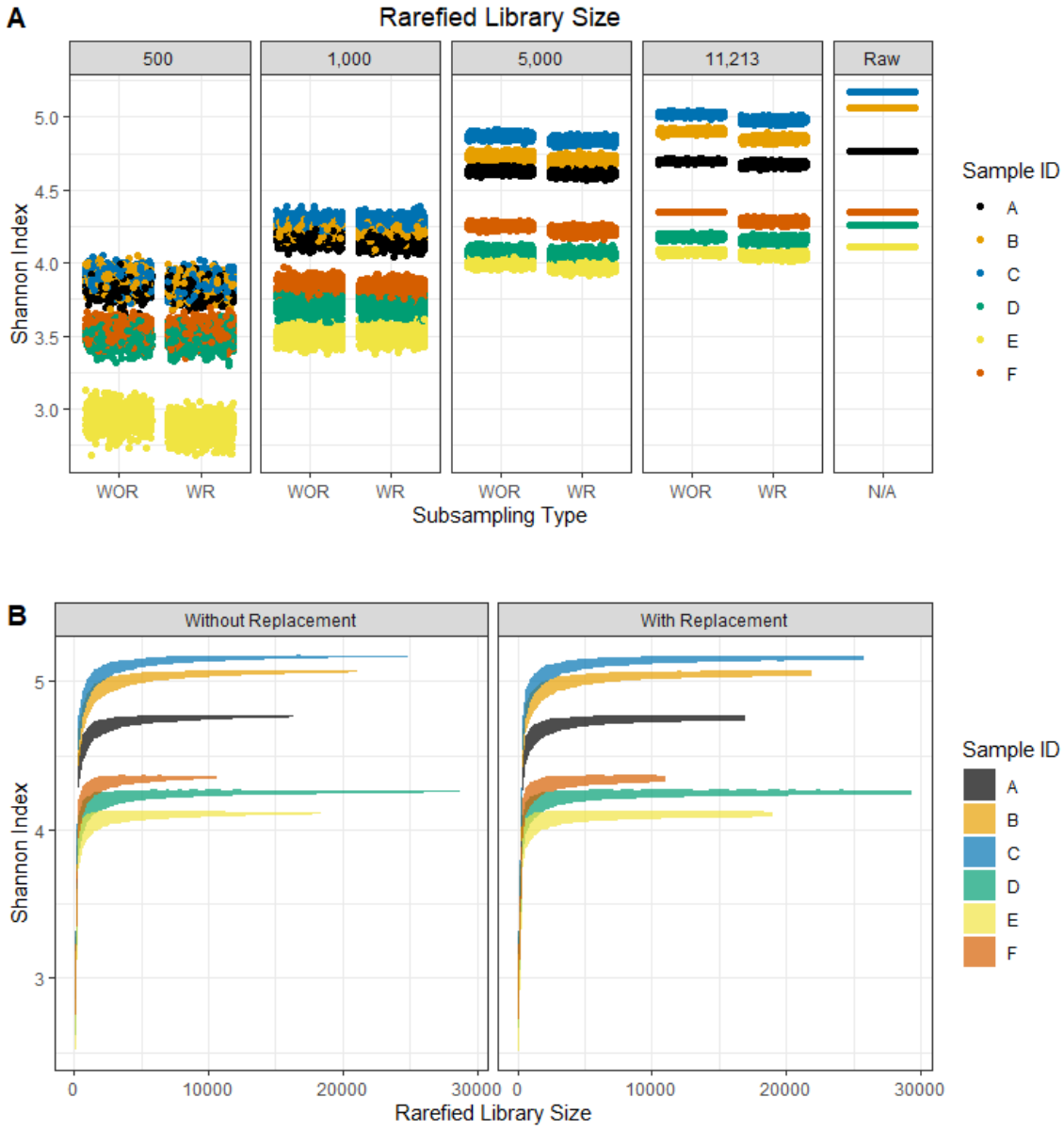


Figure 2.4 Effect of chosen rarefied library size and sampling with (WR) or without (WOR) replacement upon the Shannon Diversity Index. Six microbial communities were rarefied repeatedly (A) at specific rarefied library sizes of 11,213 sequences, 5,000 sequences, 1,000 sequences, and 500 sequences and (B) to evaluate the Shannon Index as a function of rarefied library size.

The consistency of the diversity metric when rarefying repeatedly is extremely degraded when libraries were rarefied to the smallest considered library size of 500 sequences. It illustrates the potential to reach incorrect conclusions if rarefying is completed only once. When rarefying repeatedly to a small library size, however, diversity index values that are highly inconsistent and suppressed relative to the diversity of the unrarefied data may lead to inappropriate claims of identical diversity values between samples. The extreme reduction and introduced variation of the Shannon index suggests that the selection of smaller rarefied library sizes should be approached with caution when using alpha diversity metrics, while larger normalized library sizes prevent loss of precision and reduction of the Shannon index value. However, as previously noted, the reduction in the value of the Shannon index will be dependent on the sequence variant composition of the samples.

Previous research evaluating normalization techniques has only focused on beta diversity analysis and differential abundance analysis (Gloor et al., 2017; McMurdie and Holmes, 2014; Weiss et al., 2017), but the appropriateness of library size normalization techniques for alpha diversity metrics must be evaluated due to the prerequisite of having equal library sizes for accurate calculation. Utilization of unnormalized library sizes with alpha diversity metrics may generate bias due to the potential for samples with larger library sizes to inherently reflect more of the diversity in the source than a sample with a small library size. The repeated iterations of rarefying library sizes allow characterization of the variability introduced to sample diversity by rarefying at any rarefied library size (Figure 2.3).

2.4.3 The Effects of Subsampling Approach & Normalized Library Size Selection on Beta Diversity Analysis

When samples were repeatedly rarefied to a common normalized library size with and without replacement, similar amounts of variation in the Bray-Curtis PCA ordinations were

observed between the sampling approaches (Figure 2.5). This indicates that although rarefying with replacement seems potentially erroneous due to the fabrication of count values that are not representative of actual data, the impact on the variation introduced into the Bray-Curtis dissimilarity distances is not large and will likely not interfere with the interpretation of results. However, rarefying without replacement should be encouraged because it is more theoretically correct, and it has not been comprehensively demonstrated that sampling with replacement is a valid approximation for all types of diversity analysis or library compositions.

When larger normalized library sizes are maintained through rarefaction, there is less potential variation introduced into beta diversity analyses, including Bray-Curtis dissimilarity PCA ordinations. For example, in the largest normalized library size possible for these data (Figure 2.5A), a minimal amount of variation was observed within each community, indicating that the preservation of higher sequence counts minimizes the amount of artificial variation introduced into datasets by rarefaction (including no variation for Sample F because it is not actually rarefied in this scenario). For this reason, rarefying to the smallest library size of a set of samples is a sensible guideline. Although, a normalized library size of 5,000 is lower than the flattening portion of the rarefaction curve for samples A, B, and C (Figure 2.2), the selection of this potentially inappropriate normalized library size (Figure 2.5C) can still accurately reflect the diversity between samples without excess artificial variation introduced through rarefaction. Due to the variation introduced to the Bray-Curtis dissimilarity ordinations in the smaller rarefied library sizes (Figure 2.5E, G), it is critical to include computational replicates of rarefied libraries to fully characterize the introduced variation in communities. As discussed above, it has been suggested that repeatedly rarefying is inappropriate due to the introduction of "added noise". However, as demonstrated, the maintenance of larger rarefied library sizes

when repeatedly rarefying does not impact interpretation of beta-diversity analysis results. Without this replication, rarefaction to small, normalized library sizes could result in artificial similarity or dissimilarity identified between samples.

Beta diversity analysis of very small, rarefied library sizes (Figure 2.6A, B, C) can still reflect similar clustering patterns observed in larger library sizes but with a much lower resolution of clusters. Rarefying has previously been shown to be an appropriate normalization tool for samples with low sequence counts (e.g., <1,000 sequences per sample) by Weiss et al. (2017), which is promising for datasets containing samples with small initial library sizes or potentially subsetting data to explore diversity within specific phyla (e.g., cyanobacteria). Caution must be taken to avoid selection of an excessively small, normalized library size due to the introduction of extreme levels of artificial variation that compromises accurate depiction of diversity (Figure 2.6D) and only reflects small portions of the sequence variants from samples with large library sizes. The tradeoff between rarefying to a smaller than advisable library size or excluding entire samples with small library sizes remains and can possibly be resolved by analyzing results with all samples and a small, rarefied library size as well as with some omitted samples and a larger rarefied library size.

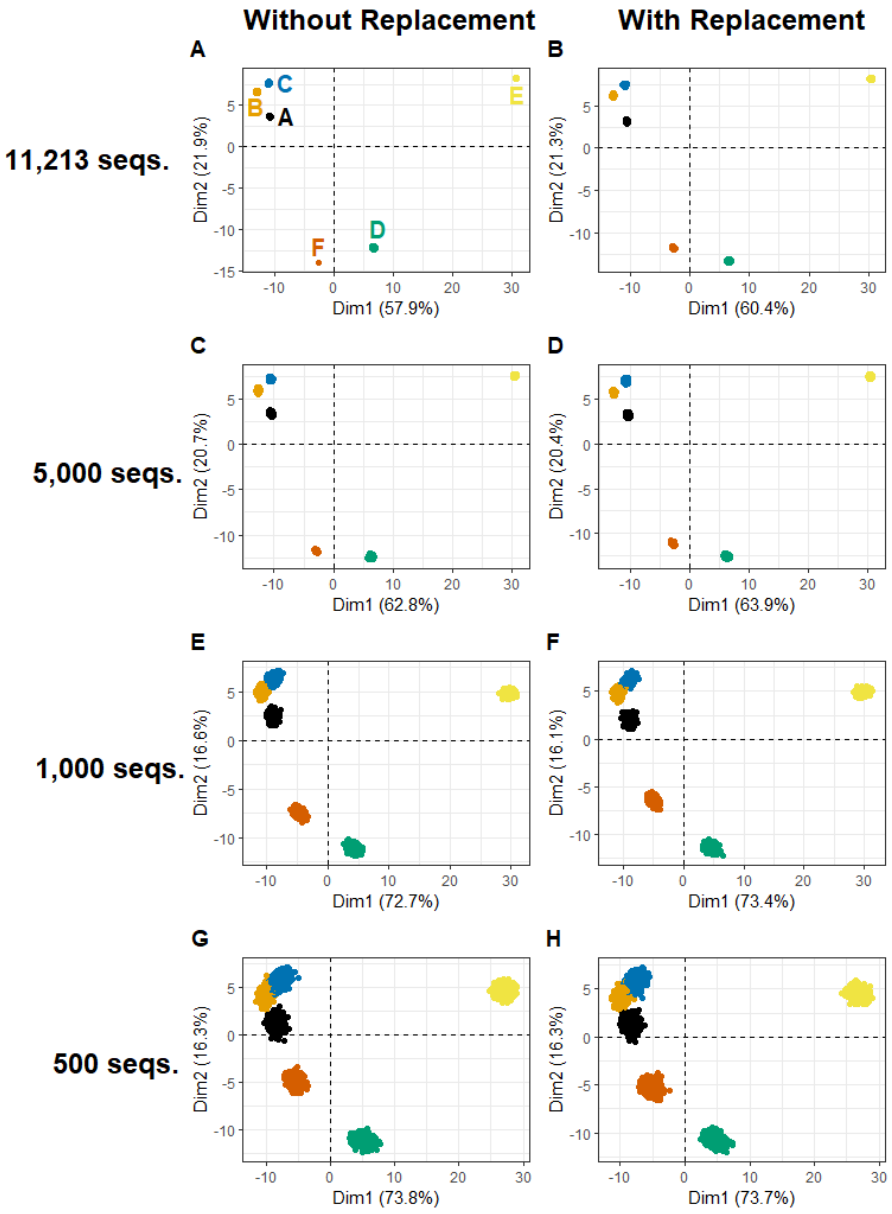


Figure 2.5 Variation in PCA ordinations (using the Bray-Curtis dissimilarity on Hellinger transformed rarefied libraries) of six microbial communities repeatedly rarefied with and without replacement to varying library sizes.

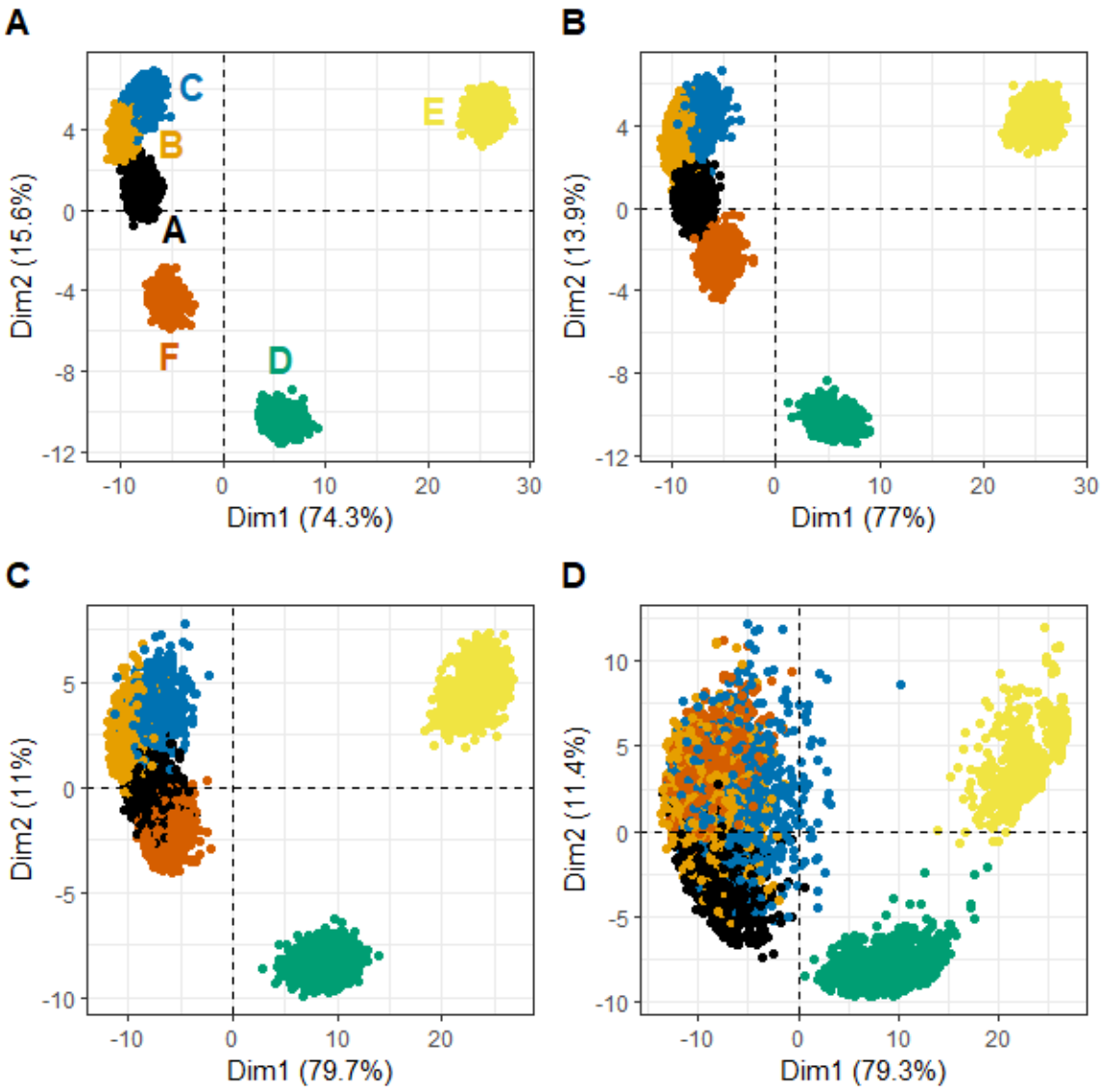


Figure 2.6 Variation in PCA ordinations (using the Bray-Curtis dissimilarity on Hellinger transformed rarefied microbial communities) of six microbial communities repeatedly rarefied to very small library sizes of (A) 400, (B) 300, (C) 200 and (D) 100 sequences.

Although rarefying has the potential to introduce artificial variation into data used in beta diversity analyses, these results suggest that it does not become problematic until rarefying to normalized library sizes that are very small (e.g., 500 sequences or less) for the samples considered. While we saw a degradation of the consistency and value of the alpha diversity Shannon index at 500 sequences, beta diversity analyses may be more robust to rarefaction and capable of reflecting qualitative clusters in ordination as previously discussed in Weiss et al. (2017). The artificial variation introduced to beta diversity analyses by rarefaction could lead to erroneous interpretation of results, but the implementation of multiple iterations of rarefying library sizes allows a full representation of this variation to aid in determining if apparent similarity or dissimilarity is a chance result of rarefying.

The use of non-normalized data has been shown to be more susceptible to the generation of artificial clusters in ordinations, and rarefying has been demonstrated to be an effective normalization technique for beta diversity analyses (Weiss et al., 2017). However, the use of a single iteration of rarefying does result in the omission of valid data (McMurdie and Holmes, 2014). Repeated iterations of rarefying in this study demonstrated that rarefying repeatedly does not substantially impact the output and interpretation of beta diversity analyses unless rarefying to sizes that are inadvisably small to begin with. McMurdie and Holmes (2014) were dismissive of rarefying repeatedly, but their analysis of such repetition was not evaluated in beta-diversity or differential abundance analysis. In the case of differential abundance analysis, the added variability of rarefying would be statistically inappropriate relative to generalized linear modelling that can account for varying library sizes. Additionally, repeatedly rarefying allows for characterization of variation introduced through subsampling while accounting for discrepancies in library size, supporting the potential utility of the

normalization technique for beta diversity analyses. McKnight et al. (2019) preferred use of proportions in diversity analysis over rarefying (arguing that both were superior to other normalization approaches). While proportions normalize the sum of the ASV weights to one for each sample, we note that the approach does not normalize the library size in terms of sequence counts. This is important because sample proportions will provide a more precise reflection of the true proportions of which the set of sequences is believed to be representative in samples with larger libraries than in samples with smaller libraries. Using proportions of unnormalized sequence count libraries in beta diversity analysis overlooks the loss of alpha diversity associated with smaller library sizes when comparing samples with different library sizes.

2.4.4 The Need for Library Size Normalization

The increasing popularity and accessibility of amplicon sequencing has enabled the scientific community to gain access to a wealth of microbial community data that would otherwise not have been accessible. However, despite amplicon sequencing of taxonomic marker genes being the gold standard approach for microbial community analysis, the data handling and statistical analysis is still in the early stages of development. The diversity analyses that the scientific community desires to perform on amplicon sequencing data require library sizes to be normalized across samples, which creates the challenge of determining appropriate normalization techniques. New normalization techniques and tools are constantly being developed and released to the community with claims that the newest technique is the best and only solution that should be utilized for analysis but may be associated with data handling limitations or be too specifically tailored to a particular type of analysis or desired property. For example, the centered-log ratio transformation (Gloor et al., 2016) cannot be

used on zero count data and amplicon sequencing datasets must be augmented with an artificial pseudocount to apply the normalization technique. The limitations of normalization techniques may affect downstream analyses, making it critical to understand the implications of the technique chosen. Further discussion within the scientific community is needed to ensure rigorous interpretation of amplicon sequencing data without unwarranted bias introduced by the normalization technique. Despite the potential limitations, rarefying remains common in current research requiring library size normalization, especially for diversity analysis. The implementation of a single iteration of rarefying is problematic due to the omission of valid data and should not be used for library size normalization. Conducting repeated iterations of rarefying allows for the characterization of variation introduced through random subsampling in diversity analyses but would be inappropriate for differential abundance analysis where generalized linear modelling of non-normalized data is possible (McMurdie and Holmes, 2014).

2.5 Conclusions

- Rarefying with or without replacement did not substantially impact the interpretation of alpha (Shannon index) or beta (Bray-Curtis dissimilarity) diversity analyses considered in this study but rarefying without replacement is theoretically appropriate and will provide more accurate reflection of sample diversity.
- To avoid the arbitrary loss of available information through rarefaction to a common library size, the random error introduced through rarefaction should be evaluated by repeating rarefaction multiple times.
- Rarefying repeatedly statistically describes possible realizations of the data if the number of sequences read had been limited to the normalized library size, thus allowing

diversity analysis using samples of equal library size in a way that accounts for the data loss in rarefying.

- The use of larger normalized library sizes when rarefying minimizes the amount of artificial variation introduced into diversity analyses but may necessitate omission of samples with small library sizes (or analysis at both inclusive low library sizes and restrictive higher library sizes).
- Ordination patterns are relatively well preserved down to small, normalized library sizes with increasing variation shown by repeatedly rarefying, whereas the Shannon index is very susceptible to being impacted by small, normalized library sizes both in declining values and variability introduced through rarefaction.
- Even though repeated rarefaction can characterize the error introduced by excluding some fraction of the sequence variants, rarefying to extremely small sizes (e.g., 100 sequences) is inappropriate because the substantial introduced variation leads to an inability to differentiate between sample clusters.
- Further development of strategies (e.g., data handling, library size normalization for diversity analyses) for ensuring rigorous interpretation of amplicon sequencing data is required.

Chapter 3
Cyanobacterial Populations are Dynamic and their Characterization Requires Consideration of Diurnal and Spatial Variation

3.1 Introduction

Cyanobacteria are recognized as a threat to surface water quality through the formation of dense blooms and the production of secondary metabolites including taste and odor compounds (e.g., geosmin, 2-methyl isoborneol) and potent cyanotoxins (Huisman et al., 2018; Paerl, 2014; Vu et al., 2020). The potential risk that these organisms impose to ecosystems, recreational use and water treatment necessitates water quality monitoring programs. Traditionally, these programs rely on visual identification of cyanobacterial genera based on morphological characteristics observed using microscopy. Quantitative measurements may include direct enumeration of cells or indirect estimation using chlorophyll-*a* quantification (Chorus and Bartram, 1999). Although these methods are frequently utilized, reliance on microscopic identification of taxa is limited due to the impossibility of visually differentiating between toxic and non-toxic cyanobacteria (Gallina et al., 2017). The implementation of modern molecular methods, such as next-generation sequencing, provides a rapid and sensitive technique that can be applied to characterize both taxonomic composition and functional potential (e.g., toxicity; Clooney et al., 2016; Langille et al., 2013) of cyanobacterial populations. Although next-generation sequencing data can be used to rapidly characterize communities and predict the potential for toxicity, it is critical that sampling protocols are designed to ensure the collection of ecologically relevant data.

As a result of anthropogenic activity, cyanobacterial bloom frequency and intensity are increasing globally (Huisman et al., 2018) supporting the need for development of ecologically relevant sampling protocols. They have a variety of adaptations to environmental stress including the formation of tolerant resting cells, the presence of photoprotective cellular pigments, and the regulation of buoyancy in response to light and nutrient gradients (Huisman et al., 2018; Paerl, 2014). Specifically, cyanobacteria buoyancy is varied in response to light irradiance levels (Visser et al., 2005) resulting in oscillatory diurnal migration patterns allowing for growth under advantageous conditions (Chien et al., 2013; Chu et al., 2007) and is largely driven by cellular density (Naselli-Flores et al., 2021) and water column stability (Walsby et al., 1997) as summarized in Table 3.1.

Although cyanobacteria are known to experience oscillatory diurnal migrations, monitoring guidelines frequently do not provide recommendations for optimal sampling times (Chorus et al., 2000; Chorus and Bartram, 1999; Sarnelle et al., 2010). Typically, during the day with high light availability, cyanobacteria move downwards in the water column (Frempong, 1981; Visser et al., 2005) as carbohydrates accumulate with high rates of photosynthesis (Ibelings et al., 1991). In overnight periods, they experience upward migrations as carbohydrates are utilized and photosynthetic rates are limited (Ibelings et al., 1991; Visser et al., 2005). If sampling is conducted arbitrarily or at inconsistent time points across multiple sampling events, collected data may not be ecologically representative due to these diurnal migrations. However, to contribute further to the complexities in cyanobacterial distribution, the diurnal migrations and spatial distribution are impacted by water column stability including thermal stratification or weather induced mixing events (Walsby et al., 1997). Due to the threats that cyanobacteria impose on water quality, it is critical for monitoring programs to be

designed to allow for collection of data that is representative of cyanobacterial community structure by sampling at an ecologically relevant time of day.

The impact of sampling time on cyanobacterial community composition was evaluated using amplicon sequencing of the V4 region in the 16S rRNA gene. Taxonomic composition and community diversity analyses generated from amplicon sequencing data provided insights into diurnal trends in distribution. Specifically, the fluctuations in cyanobacterial community composition were evaluated (i) over a multi-time point sampling period, and (ii) spatially within the water column of a stratified lake. The evaluation of spatial and temporal trends present in cyanobacterial communities provides a comprehensive evaluation on the potential impact of sampling time and system specific conditions on detection and will provide critical insight into the development of monitoring programs.

Table 3.1 A summary of characteristics of cyanobacteria and environmental conditions impacting buoyancy and spatial distribution in the water column.

Characteristics of Cyanobacteria		
Characteristic	**Example Taxa**	**Impact on Buoyancy**
Cell Ballast Content	-	High photosynthetic rates result in accumulation of carbohydrates increasing cell density resulting in downward migration (Chien et al., 2013; Hunter et al., 2008; Li et al., 2016; Westwood and Ganf, 2004).
Gas Vacuolate	*Microcystis, Aphanizomenon, Nostoc, Anabeana, Oscillatoria, Coelosphaerium* (Staley, 1980; Wallsby, 1981)	Provides positive buoyancy allowing for maintenance of position within the photic zone (Walsby et al., 1997). Decreased gas vacuole content after exposure to high light irradiance results in loss of buoyancy (Westwood and Ganf, 2004)
Small & Unicellular	*Synechococcus, Cyanobium,Synechocystis, Cyanobacterium* (Sliwinska-Wilczeska et al., 2018)	Small cell size allows for maintenance of water column position (Reynolds et al., 1987; Śliwińska-Wilczewska et al., 2018)
Small & Colonial	*Aphanocapsa, Aphanothece, Chroococcus, Coelosphaerium, Cyanobium, Cyanodictyon, Merismopedia, Romeira, Snowella, Tetracerus* (Sliwinska-Wilczeska et al., 2018)	Smaller colonies exhibit more random spatial movement with no clear diurnal pattern (Chien et al., 2013).
Large Colonial & Filamentous Forms	*Dolichospermum circinale* (Westwood and Ganf, 2004)	Larger colonies move more rapidly allowing for migration of greater depths (Reynolds et al., 1987; Westwood and Ganf, 2004). Sinking rates are also faster (Ganf, 1974).

Table 3.1 Continued

Environmental Conditions		
Condition	**Examples**	**Impact on Distribution**
Water Column Stability	Thermal Stratification	Creates zonation in the water column frequently with nutrient depleted, light rich water surface and light-limited, nutrient rich deep waters (Chien et al., 2013). Vertical migrations allow for access to optimal environmental conditions (Chien et al., 2013).
	Non-Stratified Water Columns	Wind induced mixing of the water column may result in homogeneous distributions (Frempong, 1981; Hunter et al., 2008; Walsby et al., 1997).
	External Mixing Events – Storms	Storm events may result in downward mixing of communities (Walsby et al., 1997)
Light Availability	Daytime	Exposure to high light in the daytime results in loss in buoyancy with high photosynthetic rate and accumulation of carbohydrates resulting in downward migration (Ibelings et al., 1991).
	Nighttime	With light limitation and decreased photosynthetic rates, cellular carbohydrates are utilized, resulting in decreased density and upward migration for light access in daytime (Ibelings et al., 1991).

3.2 Methods

3.2.1 Study Site: Turkey Lakes Watershed

The Turkey Lakes Watershed Study was established in 1980 to investigate ecosystem effects of acidic atmospheric deposition—Jeffries et al. (1988) provided a comprehensive description of the physical characteristics of the watershed. The Turkey Lakes Watershed (TLW) is approximately 50 km north of Sault Ste Marie, Ontario on the Canadian Shield in an uneven-aged tolerant hardwood and mixed conifer forest landscape (Jeffries et al., 1988). It consists of 4 interconnected lakes fed by both first order streams and groundwater: Batchwana Lake, Wishart Lake, Little Turkey Lake and Big Turkey Lake (Jeffries et al., 1988; Figure 1.1). These lakes thermally stratify during summer and winter annually, with the exception of Wishart Lake. Wind-induced mixing of this shallow lake generally prevents thermal stratification. These lakes are classified oligotrophic to mesotrophic, and cyanobacteria are the dominant members of phytoplankton communities (Jeffries et al., 1988). Water samples for daily variation analysis were collected from the deepest point in Little Turkey Lake (47°02’37.2”N 84°24’24.4”W) and Wishart Lake (47°03’00.0”N 84°23’58.3”W). The well-mixed water column of Wishart Lake and thermally stratified water column of Little Turkey Lake (Figure A1) allowed for contrast between lakes with different water column stability.

3.2.2 Sample Collection

Water samples were collected over a two-day period in August 2018 (August 22 and August 23). They were collected at Secchi depth across three time points on both days: morning (8-9 a.m.), midday (12-1 p.m.) and afternoon (4-5 p.m.). This is the water depth at which light penetration is approximately 1% of surface illumination; it is considered to be the maximum depth at which there is generally sufficient light for photosynthesis (Bukata et al., 1988). Secchi

depth was measured in a standardized manner from the shaded side of the boat and by the same individual. Samples for Wishart Lake were collected from near the bottom (4 m) due to high water clarity. Samples were collected from Secchi depths ranging from 4-5.25 m in Little Turkey Lake depending on time of collection and sampling day. Water samples were also collected at the surface (0 m) in Little Turkey Lake on the first sampling day to identify potential correlations in cyanobacterial communities between depths as a function of time.

Water samples collected using a Masterflex E/S portable sampler peristaltic pump were serially filtered through a 47 mm GF/C filter (Whatman plc, Buckinghamshire, United Kingdom). After vacuum filtration, 250 mL of filtered water was then filtered through a 0.22 µm Sterivex™ filter to collect additional microbes. Whatman GF/C and Sterivex™ filters were stored at -20°C prior to DNA extraction. Sampling details are provided in Table A1.

3.2.3 DNA Extraction, 16S rRNA Gene Amplicon Sequencing

DNA extraction was performed using the DNeasy PowerSoil Kit (QIAGEN Inc., Venlo, Netherlands) following the manufacturer's protocol. Elution buffer was added to the spin columns for 15 minutes prior to elution of the DNA extract. DNA was quantified using a NanoDrop spectrophotometer when possible (Table A1). Although the detection limit is reported as 2ng/µl (Thermo Fisher Scientific, 2009), absolute values have been shown to be less accurate for DNA concentrations less than 10ng/µl (Khetan et al., 2019) resulting in potentially erroneous reads for low concentration environmental samples used in this study. The DNA extracts were submitted for amplicon sequencing using the Illumina MiSeq platform (Illumina Inc., San Diego, United States) at a commercial laboratory (Metagenom Bio Inc.,Waterloo, ON) using primers designed to target the 16S rRNA gene V4 region [515FB

(GTGYCAGCMGCCGCGGTAA) and 806RB (GGACTACNVGGGTWTCTAAT)] (Walters et al., 2015).

3.2.4 Sequence Processing & Library Size Normalization

The program QIIME2 (v. 2019.10; Bolyen et al., 2019) was used for bioinformatic processing. Demultiplexed paired-end sequences were trimmed and denoised, including the removal of chimeric sequences and singleton sequence variants, using DADA2 (Callahan et al., 2016) to construct the amplicon sequence variant (ASV) table. Taxonomic classification was performed using a Naïve-Bayes taxonomic classifier trained using the SILVA138 database (Quast et al., 2013; Yilmaz et al., 2014). Taxonomic assignments for amplicon sequence variants (ASV) classified as Cyanobacteria at the phylum level were manually curated to reflect taxonomic assignments above the genus level according to AlgaeBase. Files from QIIME2 were imported into R (v. 4.0.1) for downstream analyses using *qiime2R* (v. 0.99.23) (Bisanz, 2018). Initial sequence libraries were filtered to exclude ASVs that were taxonomically classified as mitochondria or chloroplast sequences using *phyloseq* (v. 1.32.0; McMurdie and Holmes, 2012). For cyanobacterial community analysis, ASVs classified as Cyanobacteria at the phylum level were filtered to create libraries consisting of only cyanobacterial sequences. Samples were repeatedly rarefied without replacement to a normalized library size of 370 reads for community diversity analyses using *mirlyn* (Cameron and Tremblay, 2020).

3.2.5 Cyanobacterial Communities - Taxonomic Composition & Diversity Analyses

The composition of communities was assessed at the taxonomic order level and relative abundances were visualized using a heatmap produced with *mirlyn*. Relative abundances were randomized across phyla within samples to identify significantly dominant groups within the

bacterial community in relation to sampling time with a Bonferroni correction. While a p-value of 0.05 is frequently viewed as the threshold for significance, due to the heterogeneity and high variability present in environmental systems, a p-value of 0.2 has been selected as the threshold for significance in these communities.

To confirm the taxonomic classification performed by the Naïve-Bayes classifier, a phylogenetic tree was constructed in MEGA X (Kumar et al., 2018) using cyanobacterial reference sequences and sequences from samples (Figure C6). Sequences classified to the genera highlighted in guidelines and resources for sampling protocol deigns (Graham et al., 2008; Vidal et al., 2021) including *Microcystis*, *Anabaena*, *Aphanizomenon*, *Pseudanabaena*, and *Synechococcus*, were selected for further evaluation. Notably, other taxa contributed to the compositional structure at the order level but were excluded from this analysis due to the frequent focus on specific bloom forming and toxic taxa for water quality management. However, in addition to the aforementioned genera, sequences classified to the following were also included in this analysis. *Radiocystis* has been shown to have identical 16S rRNA genes as *Microcystis* (Vidal et al., 2021) and toxicity (Vieira et al., 2003). *Cyanobium* is another potentially toxic picocyanobacterial genera (Śliwińska-Wilczewska et al., 2018) detected in high relative abundances in these samples. Relative abundances of selected cyanobacterial sequences were visualized using a heatmap and were evaluated based on unicellular, filamentous, or colonial morphologies to characterize taxa specific diurnal trends.

Community diversity analyses were performed using *mirlyn* on repeatedly rarefied libraries (Cameron et al., 2020). The Shannon Index (Shanon, 1948), an alpha diversity metric, was analyzed for sample comparison to identify trends in sample diversity as a function of time. Rarefied libraries were also used for beta-diversity analyses. Rarefied libraries were

transformed using a Hellinger transformation. Hellinger transformed data was used to calculate Bray-Curtis distances (Bray and Curtis, 1957) used in principal component analysis (PCA).

3.3 Results & Discussion

3.3.1 System Specific Diurnal Trends in Cyanobacterial Communities

Relative abundances of cyanobacterial sequences were higher in Little Turkey Lake than Wishart Lake (Figure 3.1A; Table 3.2). In Little Turkey Lake at Secchi depth, the increase in relative abundance of cyanobacteria within our libraries was observed to exhibit diurnal trends with significant representation in the bacterial community on both sampling days during the afternoon sampling time point ($p=0.011$, $p=0.001$; Table C3) and during the midday on the second sampling day ($p=0.0006$). However, they were not significantly represented in the community at the midday timepoint of the first sampling day ($p=0.27$). Further deviation between the representation of cyanobacteria in the bacterial community of Little Turkey Lake was observed in the morning sampling time points between the first sampling day ($p = 1$) and second sampling day ($p = 0.28$) indicating the potential for dynamic shifts in community composition on the diurnal scale. Unlike Little Turkey Lake, Wishart Lake did not exhibit a consistent and recurring increase in abundances from morning to afternoon and relative abundances remained more constant across sampling time (Figure 3.1B; Table 3.2). Additionally, within the bacterial community of Wishart Lake, cyanobacterial abundances did not exhibit temporal trends and showed no significant representation across sampling times and sampling days ($p=1$).

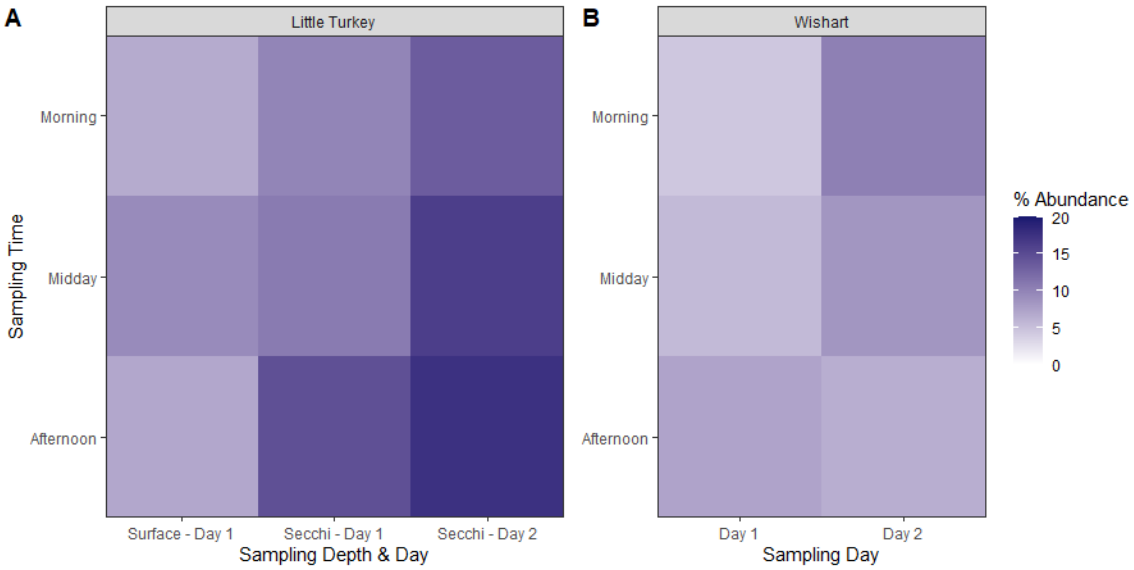


Figure 3.1 Heatmap depicting the relative abundances of amplicon sequence variants classified to the phyla Cyanobacteria within the bacterial community across a multi-time point sampling series in a stratified (Little Turkey) and non-stratified (Wishart) lake. Amplicon sequence variants of the V4 region of the 16S rRNA gene classified to the phylum Cyanobacteria were selected to examine the contribution of cyanobacterial communities to the bacterial community at (A) surface and Secchi depth in Little Turkey Lake, and (B) Secchi depth in Wishart Lake. At Secchi depth, cyanobacteria exhibited increased abundances later in the day in the stratified lake but no consistent diurnal trend in the non-stratified lake.

Table 3.2 Relative abundances as percentages of cyanobacteria within the bacterial community and subsequent composition of cyanobacterial communities. Values were rounded to two decimal points and excluded groups that were present at less than 1% abundance (-).

Taxonomic Group	Day 1			Day2		Sampling Time
	Little Turkey		Wishart Secchi	Little Turkey Secchi	Wishart Secchi	
	Secchi	Surface				
Cyanobacteria	9.91	6.67	4.39	13.42	10.34	Morning
	10.74	9.30	5.47	16.28	8.47	Midday
	14.62	6.97	7.31	17.53	6.46	Afternoon
Chroococcales	33.55	26.98	2.43	15.66	2.69	Morning
	10.18	26.13	3.19	12.54	4.03	Midday
	16.27	23.41	1.87	16.30	3.59	Afternoon
Chroococcaceae	12.00	7.13	-	2.85	1.41	Morning
	2.20	9.35	1.09	3.93	1.04	Midday
	4.99	8.58	-	5.52	1.79	Afternoon
Microcystaceae	15.33	16.25	1.08	7.62	-	Morning
	4.41	13.26	-	4.01	1.55	Midday
	5.52	10.66	1.09	5.18	1.79	Afternoon
Nostocales	1.33	1.81	-	-	-	Morning
	0.72	1.08	-	-	0.22	Midday
	0.70	2.20	-	0.17	-	Afternoon
Oscillatoriales	-	-	-	-	-	Morning
	-	-	-	-	-	Midday
	-	-	-	-	-	Afternoon

Table 3.2 Continued

Taxonomic Group	Day			Day 2		Sampling Time
	Little Turkey		Wishart Secchi	Little Turkey Secchi	Wishart Secchi	
	Secchi	Surface				
Synechococcales	62.03	66.09	91.89	82.00	90.28	Morning
	87.77	69.93	93.10	85.70	89.67	Midday
	81.19	71.24	89.86	81.39	90.54	Afternoon
Coelosphaeriaceae	2.93	1.24	1.35	3.20	1.29	Morning
	2.36	1.92	-	2.32	-	Midday
	4.29	3.34	-	2.82	-	Afternoon
Merismopediaceae	1.07	1.41	-	-	-	Morning
	-	1.09	-	-	-	Midday
	-	-	-	1.11	-	Afternoon
Pseudanabaenaceae	-	-	4.32	-	2.58	Morning
	-	-	14.37	-	1.85	Midday
	-	-	2.34	-	8.46	Afternoon
Synechococcaceae	61.87	66.42	87.57	82.35	87.70	Morning
	88.05	70.47	78.74	85.86	87.82	Midday
	81.44	71.37	87.52	81.60	82.11	Afternoon

The significant representation of cyanobacterial sequences in the afternoon at Secchi depth demonstrates the occurrence of daily water column migrations in Little Turkey Lake. Previous research has exhibited diurnal cycling arising from daily fluctuations in buoyancy causing oscillatory vertical migration patterns observed in populations (Howard, 2001). The water column stratification of Little Turkey Lake may create light limited environments at deeper depths and light rich environments at the surface. The increase in relative abundances of cyanobacterial sequences observed at Secchi depth could be a result of upward migration of

populations from deep, light-limited environments to exploit light availability or the downward migration of water surface populations for avoidance of high light irradiance (Olli, 1999). While the migration to Secchi depth cannot be confirmed without additional sampling depths, it is expected that this increase is driven by the downward migration for avoidance of high light irradiance (Olli, 1999) and as a result of increased cellular density through accumulation of photosynthetic products (Chu et al., 2007; Xiao et al., 2012). The observed phenomena in Little Turkey Lake demonstrates the importance of incorporating sampling time into monitoring protocols that do not utilize depth integrated sampling. Consequently, the restriction of sampling to a single time point or a single discrete depth in stratified lakes may result in vast underestimation of cyanobacterial abundances.

Dissimilar to Little Turkey Lake, Wishart Lake exhibited similar relative abundances of cyanobacterial sequences throughout the multi-time point sampling series due to the shallow non-stratified water column. Vertical distribution of cyanobacteria in shallow lakes that are not stratified have been shown to be less dependent on the light cycle (Ibelings et al., 1991) supporting the non-temporal response observed in Wishart Lake. Although Wishart and Little Turkey Lake are interconnected and are located within the same watershed, both lakes exhibited differing abundances and diurnal trends indicating the dynamic nature of cyanobacterial communities in individual systems. The dynamic nature of communities across sites warrants the development of unique sampling protocols and supports the notion that universal sampling strategies are near impossible to design (Pobel et al., 2011).

In addition to examining the daily trends of cyanobacterial abundances at Secchi depth, samples collected from the surface in Little Turkey Lake were assessed over time points in a single day to identify diurnal trends arising between different depths within a stratified water

column. At the surface, abundances fluctuated across time (6.67 – 9.30%; Figure 3.1A; Table 3.2) but were consistently lower than that detected at Secchi depth. Unlike the Secchi depth community, the relative abundances of cyanobacterial sequences at the surface were not significantly represented in the bacterial community across sampling times ($p = 1$; Table C3) indicating no significant increase as time progressed. The differing abundances detected at surface and Secchi depth justifies the incorporation of additional depth sampling into monitoring protocols because if sampling efforts are restricted to the surface, cyanobacterial populations may be severely underestimated.

It was initially hypothesized that a decrease in the relative abundance of cyanobacteria at the surface would correlate with increases at Secchi depth due to downward water column migration. However, the depth difference and amount of elapsed time between sampling points suggests otherwise. Rapid vertical migrations of cyanobacteria have previously been demonstrated (Hunter et al., 2008) with velocities typically ranging from 0.01 to 0.40 metres/hour, and for large filaments of *Dolichospermum circinale* (as *Anabaena circinalis*) being as high as 2.0 metres/hour (Westwood and Ganf, 2004). Basing migration velocities on previously reported values, surface cyanobacterial populations in Little Turkey Lake could theoretically migrate between 0.04 and 1.60 meters or up to 8.0 meters for large filaments in the elapsed time (~4 hours) between the morning and midday sampling points. However, the depth difference between the surface and Secchi depth was ~5 m indicating that the observed change in relative abundances is not a result of discrete population migration from surface to Secchi depth. Instead, these fluctuations in abundances may be caused by dynamic migrations of cyanobacteria from non-sampled depths including depths between surface and Secchi (*i.e.,* 0.01- 5.25 m) or depths deeper than Secchi depth (*i.e.*, >5.25 m) demonstrating the need for

inclusion of discrete depth sampling throughout the entirety of the water column to fully characterize diurnal migrations. While discrete depth sampling can provide highly specific details on the spatial distribution of populations, elucidating true population size, diversity and distribution is near impossible in the absence of a full depth profile.

Cyanobacterial sequence libraries were repeatedly rarefied to a normalized library size of 370 to explore diurnal trends in community diversity (Figure 3.2). The diversity of the cyanobacterial community in Little Turkey Lake at Secchi depth experienced small fluctuations which were not linked to specific diurnal responses. For example, on the first sampling day, the Shannon Index decreased from morning to afternoon but on the second sampling day it increased. Alternatively, Wishart Lake exhibited a peak in diversity at the midday time point (Figure 3.2A). Across sampling depths in Little Turkey Lake, communities at the surface were found to be less diverse than Secchi depth communities during the morning (Figure 3.2B). However, throughout the day, the Shannon Index approached equivalent values. While the diversity of communities at the surface remained relatively consistent, the cyanobacteria diversity at Secchi depth was observed to experience a slight decrease by afternoon of the first sampling day indicating the potential for further downward migration in the stratified layer during the afternoon sampling period. Modifying cyanobacteria monitoring protocols to be time sensitive and collected at appropriate depths tailored to individual systems will allow for the collection of data that is reflective of the full genetic and functional diversity of cyanobacterial communities.

To assess similarity between communities at differing sampling times, lake site and sampling depth, the Bray-Curtis dissimilarity was used and visualized using a PCA ordination on the rarefied data (Figure 3.2C). Cyanobacterial communities within lakes were found to be

more similar within lakes than between lakes with distinct clusters formed for Little Turkey and Wishart Lake. Additional similarity was observed within lakes between sampling time points across sampling days with the exception of the morning sampling time points for the first sampling day, discussed in further detail in Section 3.3.3. The innate dissimilarity between sampling depths and sampling sites located within the same watershed further demonstrates the need for development of system specific sampling protocols for collection of ecologically accurate data.

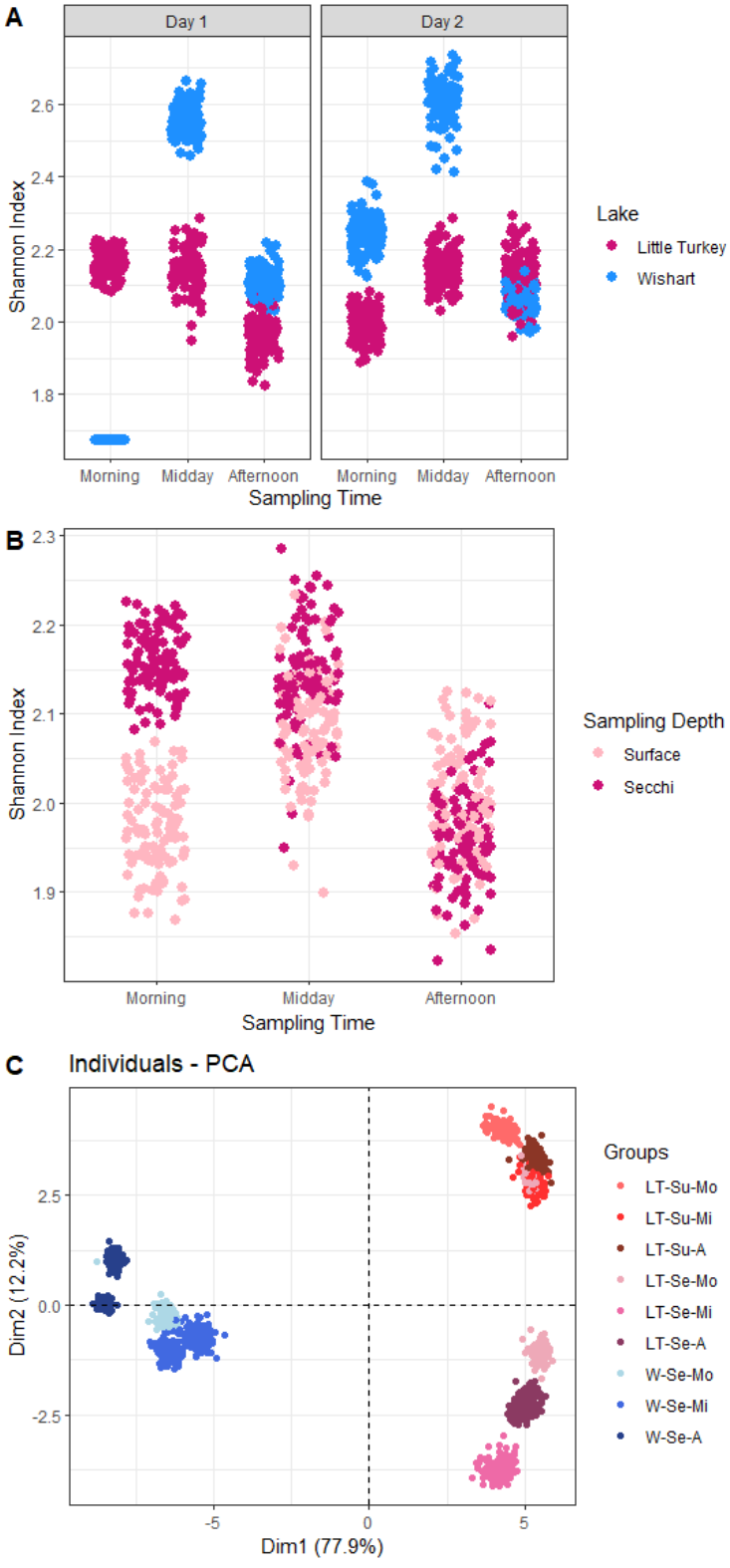

A
Day 1
Day 2
Shannon Index
2.6
2.4
2.2
2.0
1.8
Morning
Midday
Afternoon
Sampling Time
Lake
Little Turkey
Wishart
B
2.3
2.2
2.1
2.0
1.9
Sampling Depth
Surface
Secchi
C
Individuals - PCA
Dim2 (12.2%)
2.5
0.0
-2.5
-5
0
5
Dim1 (77.9%)
Groups
LT-Su-Mo
LT-Su-Mi
LT-Su-A
LT-Se-Mo
LT-Se-Mi
LT-Se-A
W-Se-Mo
W-Se-Mi
W-Se-A

Figure 3.2 Alpha and beta-diversity analyses of cyanobacterial communities collected across a multi-time point sampling series in a stratified (Little Turkey) and non-stratified lake (Wishart). Amplicon sequence variants classified to the phylum Cyanobacteria were filtered to characterize diversity within cyanobacterial communities. The Shannon Index was calculated on rarefied libraries to evaluate the effects of (A) lake site and sampling time on community diversity and (B) sampling depth and sampling time on community diversity. (C) The Bray-Curtis dissimilarity metric was used to explore similarities in communities between sampling times (Morning = Mo, Midday = Mi, Afternoon = A), lake site (Little Turkey = LT, Wishart = W) and sampling depth (Su = Surface, Se = Secchi) demonstrating unique communities between sampling depth and lake site.

3.3.2 Diurnal Trends of Bloom Forming & Toxic Cyanobacterial Taxa

To explore the impact of taxa specific diurnal responses, the taxonomic and ASV composition of cyanobacterial communities were assessed. Across samples, a total of 41 ASVs classified to the phylum Cyanobacteria were identified including taxa belonging to the cyanobacterial orders Chroococcales, Nostocales, and Synechococcales (Table 3.3). For analogous comparatives to information provided in guidelines for monitoring, cyanobacterial orders were grouped by morphology as follows: (i) Unicellular Taxa – Synechococcales (*Cyanobium*, *Synechococcus*), (ii) Colonial – Nostocales (*Microcystis, Radiocystis*), and (iii) Filamentous – Nostocales (*Anabaena, Aphanizomenon*) and *Pseudanabaena* (Order – Synechococcales). Notably, the genus *Pseudanabaena* initially was classified as Oscillatoriales based on the filamentous morphology but recent genomic sequencing and examination of ultrastructural characteristics has resulted in reclassification into the order Synechococcales (Komárek et al., 2014; Vidal et al., 2021). For the purpose of this investigation, ASVs classified to the genus *Pseudanabaena* were included with other filamentous taxa.

3.3.2.1 Unicellular Taxa

Both lakes were dominated across sampling times by sequences classified to the order Synechococcales (Figure 3.3). Specifically, these sequences were classified to the family Synechococcaceae (Table 3.2) which includes unicellular picocyanobacterial genera such as *Synechococcus* and *Cyanobium*. This taxonomic classification was affirmed through the creation of a phylogenetic tree (Figure C.6). The dominance of Synechococcaceae classified ASVs across sampling times indicates a lack of distinct diurnal migrations due to the ability of smaller sized taxa to constantly maintain water column position better than larger colonial and filamentous taxa (Śliwińska-Wilczewska et al., 2018; Yamamoto and Nakahara, 2006). Spatially within Little Turkey Lake, Synechococcaceae ASVs were found in higher abundances at Secchi depth which may be due to the higher abundance of gas vacuolate taxa at the surface which is which is discussed in further detail in Section 3.3.2.2. However, the high abundances of Synechococcaceae ASVs at Secchi depth and the potential toxicity in picocyanobacterial taxa requires monitoring protocols to not restrict sampling efforts to the water surface alone.

While the cyanobacteria communities in both lakes were dominated by the ASVs classified to the family Synechococcaceae, large differences were observed in the composition between Little Turkey (Figure 4A) and Wishart Lake (Figure 4B). Specifically, Little Turkey Lake included 7 ASVs and Wishart Lake included 17 ASVs classified to the family Synechococcaceae, demonstrating the difference in ASV composition between systems within the same watershed. ASV848 was found consistently in high abundances in both lakes and ASV846 was found in high abundances in Little Turkey Lake. Despite belonging to the same taxonomic family, both ASVs exhibited distinct diurnal trends. For example, ASV848

exhibited a decrease in abundance at the midday sampling time point but ASV846 experienced a peak in relative abundances at the midday indicating the potential for unique diurnal responses within individual cyanobacterial populations. Although Wishart Lake exhibited homogeneity in diurnal variation when examining cyanobacteria abundances broadly, individual ASVs exhibited unique diurnal responses. For example, ASV855, ASV859, ASV850, and ASV849 were absent in morning periods on both sampling days but were present in the midday or afternoon. The consistent and ephemeral occurrence of different Synechococcaceae ASVs shows the potential ecological variation present in individual taxa and highlights the requirement for characterization of cyanobacterial communities within systems to optimize monitoring efforts.

Table 3.3 Taxonomic classification of potentially bloom-forming and toxic cyanobacteria classified amplicon sequence variants in the diurnal sampling series.

ASV ID	Taxonomic Classification	Lake Site
ASV819	Nostocales	Little Turkey
ASV822	Nostocales	Little Turkey
ASV824	Nostocales	Little Turkey
ASV838	Synechococcaceae	Wishart
ASV839	Synechococcaceae	Little Turkey & Wishart
ASV842	Synechococcaceae	Wishart
ASV843	Synechococcaceae	Wishart
ASV846	Synechococcaceae	Little Turkey & Wishart
ASV848	Synechococcaceae	Little Turkey & Wishart
ASV849	Synechococcaceae	Wishart
ASV850	Synechococcaceae	Little Turkey & Wishart
ASV853	Synechococcaceae	Wishart

Table 3.3 Continued

ASV ID	Taxonomic Classification	Lake Site
ASV855	Synechococcaceae	Little Turkey & Wishart
ASV857	Synechococcaceae	Wishart
ASV858	Synechococcaceae	Wishart
ASV859	Synechococcaceae	Wishart
ASV865	Synechococcaceae	Little Turkey & Wishart
ASV866	Synechococcaceae	Little Turkey & Wishart
ASV869	Synechococcaceae	Wishart
ASV921	Synechococcaceae	Wishart
ASV919	*Radiocystis*	Little Turkey & Wishart
ASV913	*Microcystis*	Little Turkey
ASV 914	*Microcystis*	Little Turkey & Wishart
ASV806	*Pseudanabaena*	Wishart

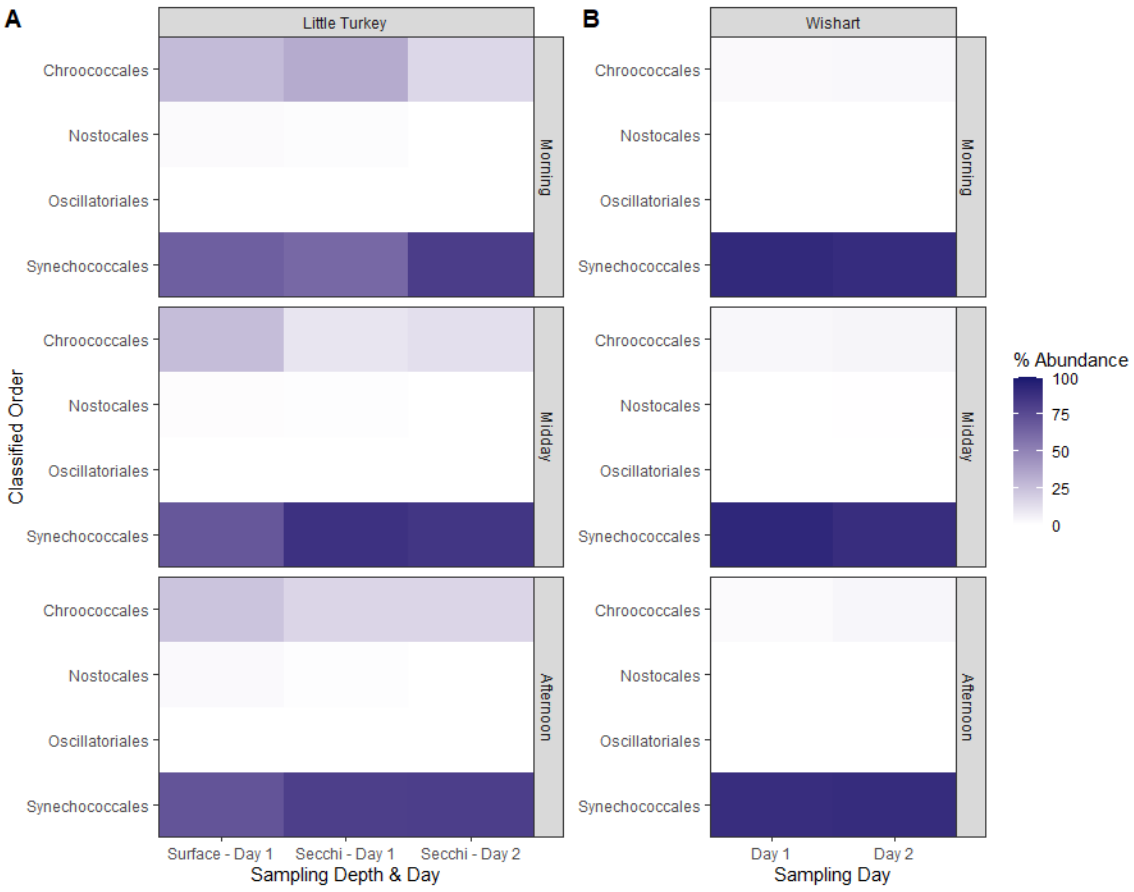


Figure 3.3 Heatmap depicting the composition of cyanobacterial communities at the order level across a multi-time point sampling series in a stratified (Little Turkey) and non-stratified (Wishart) lake. Amplicon sequence variants of the V4 region of the 16S rRNA gene were classified to the phylum Cyanobacteria were selected to examine the taxonomic composition of cyanobacterial communities at (A) surface and Secchi depth in Little Turkey Lake, and (B) Secchi depth in Wishart Lake. Cyanobacterial communities in both lakes were consistently dominated by the order Synechococcales, which contains potentially toxic picocyanobacterial genera.

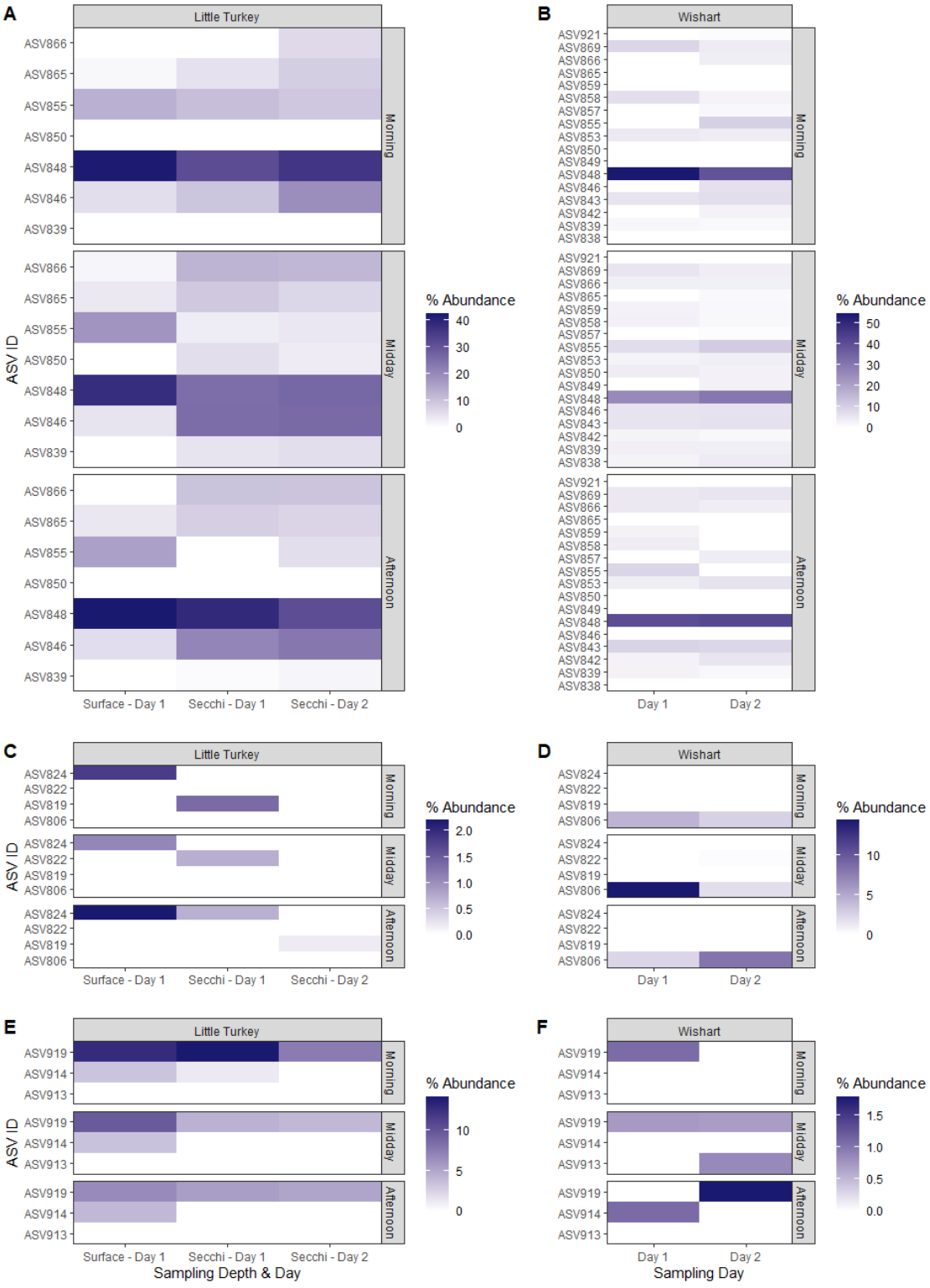

A
Little Turkey
Morning
Midday
Afternoon
ASV ID
ASV866
ASV865
ASV855
ASV850
ASV848
ASV846
ASV839
% Abundance
40
30
20
10
0
Surface - Day 1
Secchi - Day 1
Secchi - Day 2
B
Wishart
ASV921
ASV869
ASV866
ASV865
ASV859
ASV858
ASV857
ASV855
ASV853
ASV850
ASV849
ASV848
ASV846
ASV843
ASV842
ASV839
ASV838
50
40
30
20
10
0
Day 1
Day 2
C
ASV824
ASV822
ASV819
ASV806
2.0
1.5
1.0
0.5
0.0
D
10
5
0
E
ASV919
ASV914
ASV913
Sampling Depth & Day
F
1.5
1.0
0.5
0.0
Sampling Day

Figure 3.4 Heatmap depicting the relative abundances of individual amplicon sequence variants of interest across a multi-time point sampling series in a stratified (Little Turkey) and non-stratified (Wishart) lake. Amplicon sequence variants (ASV) of the V4 region of the 16S rRNA gene classified to the phylum Cyanobacteria to highlight diurnal responses of common toxic and bloom forming genera including Synechococcaceae (*Synechococcus, Cyanobium*), Nostocales (*Anabaena*, *Aphanizomenon*), *Pseudanabaena* and Microcystaceae (*Microcystis*, *Radiocystis*). Relative abundances of individual ASVs in cyanobacterial community composition for (A/B) unicellular taxa (Synechococcaceae), (C/D) filamentous taxa (Nostocales & *Pseudanabaena*), and (E/F) colonial taxa (Microcystaceae) to identify taxa specific diurnal responses.

3.3.2.2 Colonial & Filamentous Taxa

Chroococcales and Nostocales were observed in lower abundances in Little Turkey (Figure 3.3A) and Wishart Lake (Figure 3.3B) but continued to demonstrate system specific differences in the compositional structure of cyanobacterial communities. Little Turkey Lake had a larger Chroococcales population than observed in Wishart Lake (Figure 3.2; Table 3.2). Additionally, Nostocales sequences were detected in Little Turkey Lake but were largely undetected in Wishart Lake. Further uniqueness of the communities between systems was observed at the ASV level. Specifically, ASV806 (classified to the genus *Pseudanabaena*) was only detected in Wishart Lake (Figure 3.4D) and Nostocales ASVs (ASV824, 822, 819) were only detected in Little Turkey Lake (Figure 3.4C) exhibiting uniquity in cyanobacterial community structure between systems. Further unique ASV composition was observed with the detection of ASV913 (classified to the genus *Microcystis*) in Wishart Lake at very low abundances (Figure 3.4F) but larger abundances of ASV914 (classified to the genus *Microcystis*) and ASV919 (classified to the genus *Radiocystis*) in Little Turkey Lake (Figure 3.4E). These differences observed in the compositional structure of cyanobacterial

communities between lakes further supports the necessity for unique sampling protocols for different systems discussed in detail in Section 3.3.4.

Within samples, sequences attributed to taxa that have previously been reported to contain gas vesicles, including *Microcystis, Anabaena,* and *Aphanizomenon,* were detected. Sequences classified to these genera including ASV824 (classified to the genus *Anabaena*), and ASV914 (classified to the genus *Microcystis*), were consistently detected at the surface in Little Turkey Lake, potentially because of positive buoyancy regulation due to the presence of gas vacuoles. Diurnal variation associated with downward migration during the daylight was exhibited in the surface populations of ASV919 (classified to the genus *Radiocystis*) as a consistent decrease in abundance was observed from morning to afternoon exhibiting the expected diurnal migration trend. While taxa with gas vesicles were detected regularly in surface samples of Little Turkey Lake, ASV819 (classified to the genus *Anabaena*) and ASV822 (classified to the order Nostocales) were only detected at Secchi depth indicating that gas vacuolate taxa may be found deeper in the water column and supporting the distribution of cyanobacteria through the photic zone and not only as surface accumulations (Graham et al., 2008). The presence of taxa with gas vesicles outside of the water surface diurnally signifies the importance of conducting sampling at additional depths for accurate monitoring.

3.3.3 External Disturbances Impact Cyanobacteria Distribution

Evaluation of community dissimilarity between samples revealed that the cyanobacterial communities at Secchi depth in the morning of the first sampling day were highly similar to the surface communities in Little Turkey Lake, a phenomenon that was not observed subsequently (Figure 3.2C). Although the taxonomic composition of the surface and Secchi depth cyanobacterial communities showed similar order level composition across

sampling times (Figure 3.3A), the relative abundance of the order Chroococcales, was found in higher abundances at Secchi depth in the morning of the first sampling day resembling values detected at the surface. The ASV composition revealed similar results with ASV919 (classified to the genus *Radiocystis*) and ASV914 (classified to the genus *Microcystis*) detected at higher abundances in the morning period of the first sampling day (Figure 3.4E). In the overnight period preceding the morning of the first sampling day a heavy rainfall event occurred. Storm events have previously been shown to result in downward mixing of cyanobacterial populations (Walsby et al., 1997) which likely resulted in redistribution of water surface populations throughout the water column of Little Turkey Lake. While the composition of the Secchi depth community in the morning period of the first sampling day did not resemble subsequent Secchi depth communities, the water column and community composition restabilized by midday. Due to the impact of external disturbances on water column mixing and cyanobacteria distribution, sampling should be temporally delayed following weather induced mixing events to allow for water column re-stabilization and redistribution of cyanobacterial populations.

Unlike Little Turkey Lake, Wishart Lake did not exhibit obvious impacts in the distribution of cyanobacteria following the rainfall event. High dissimilarity in community composition was not observed between the morning of the first sampling day and subsequent time points (Figure 3.2C) and taxonomic composition of communities at the order level were consistent across sampling times (Figure 3.3B). Previous diurnal studies on freshwater microbial communities did not show an impact of meteorological conditions on bacterial abundances (Filippini et al., 2008) indicating that the response to meteorological conditions may be system specific. The potential for system specific responses to external mixing events

as demonstrated in Little Turkey and Wishart Lake further signifies the importance of utilizing unique protocols tailored to system dynamics.

3.3.4 Implications for Cyanobacteria Monitoring Sampling Protocols

Sampling protocols designed to reflect cyanobacterial ecology is critical for accurate representation of the community. Sampling protocols and guidelines frequently provide guidance on appropriate sampling time and sampling depth to ensure accurate monitoring (Table 3.4). However, these recommendations can range widely from providing design features that rely on user interpretation (e.g., "Later in the day", "Surface"), generalized sampling conditions (e.g., "10 a.m. – 3 p.m.", depth integrated sampling) or complete absence of recommendations as is the case for appropriate sampling times. In this study, cyanobacterial abundances continued to fluctuate between the hours of 9 a.m. and 4 p.m. demonstrating that large, generalized time frames are too broad to apply universally to varying aquatic systems. The reliance of depth integrated sampling allows for comprehensive characterization of the cyanobacterial communities throughout the entire water column with a single sample (Ministry for the Environment and Ministry of Health., 2009; Newcombe, 2009; Ohio Environmental Protection Agency, 2013; Sarnelle et al., 2010; University of New Hampshire - Center for Freshwater Biology, 2010). Alternatively, discrete depth sampling can characterize trends in spatial distribution at critical depths in the water (e.g., water intake pipes; Graham et al., 2008) but if this sampling is restricted to a single depth it cannot account for heterogeneity in cyanobacterial distribution (Vidal et al., 2014). Optimally, full discrete depth profiles would be conducted to characterize spatial distribution of cyanobacterial communities. However, with limitations of time and resources, depth integrated sampling is an appropriate technique for providing an overview of cyanobacterial populations that encompasses diurnal migrations.

Table 3.4 A summary of cyanobacteria monitoring sampling protocol recommendations for sampling time and sampling depth and the potential impact on detection arising due to diurnal migration cycles.

Sampling Protocol Design Feature	Impact on Detection	References
Recommended Sampling Time		
No optimal sampling time advisory	Inconsistent and/or arbitrary sampling time will result in variation in the detection of cyanobacterial populations.	Chorus et al., 2000; Chorus and Bartram, 1999; Graham et al., 2008; Sarnelle et al., 2010; Board of Directors -Colorado Lake & Reservoir Management Association, 2015
10 a.m. – 3 p .m.	Cyanobacterial populations will experience fluctuations during the large timeframe. Samples collected at 10 a.m. will differ in composition from samples collected at 3 p.m. due to water column migration resulting in the potential for missed detection or underrepresentation of cyanobacterial populations.	Klamath river blue green algae working group, 2009; University of New Hampshire - Center for Freshwater Biology, 2010
"Later in the day"	Sample collection time is dependent on user interpretation of "later in the day" which creates bias from the interpretation of sampling protocol.	Newcombe, 2009
Morning (for surface sampling only)	Sampling at the surface in the morning period is appropriate due to known trends in water column migration of cyanobacterial populations.	Ministry for the Environment and Ministry of Health., 2009
Recommended Sampling Depth		
Integrated water column depth sample	Integrated depth sampling accounts for the potential of daily variation in cyanobacterial populations.	Ministry for the Environment and Ministry of Health., 2009; Newcombe, 2009; Ohio Environmental Protection Agency, 2013; Sarnelle et al., 2010; University of New Hampshire - Center for Freshwater Biology, 2010; Board of Directors -Colorado Lake & Reservoir Management Association, 2015; Klamath river blue green algae working group, 2009

Traditional cyanobacteria monitoring is performed using microscope counts or chlorophyll-*a* analysis (Chorus et al., 2000). However, advances in molecular techniques and computational analysis have provided new avenues to gain further insight on cyanobacterial populations. Amplicon sequencing of taxonomic marker genes, such as the 16S rRNA gene, is a rapid and sensitive technique to characterize microbial communities, including cyanobacterial community composition. Additionally, information on potential for toxicity can be obtained through amplicon sequencing of toxin genes or metagenomic sequencing. Gaining further insight into the potential for toxicity prior to toxin detection allows for proactive management rather than reactive management of source waters (Chapman, 2010). Although amplicon sequencing has revolutionized our ability to study microbial communities, amplicon sequencing data cannot provide absolute quantification of abundances and is inherently compositional (Gloor et al., 2017), which introduces challenges in the interpretation of changes in community structure. For example, an observed decrease in in relative abundances of cyanobacteria, may only be an artifact of increased sequence abundances of other taxonomic groups and not be representative of an actual absolute decrease in cyanobacterial populations. The inherent compositional nature of amplicon sequencing data and the inability to provide information on absolute abundances for monitoring purposes would require further support from traditional cell enumeration techniques or other molecular techniques (e.g., flow cytometry (Patel et al., 2019), qPCR (Chiu et al., 2017)) to ensure accurate quantification of cyanobacterial populations.

Insights into cyanobacterial dynamics are becoming increasingly accessible with the application of modern molecular techniques but these techniques are limited to the accuracy of samples collected for analysis. In addition to when and where samples are collected, how

samples are collected can also significantly impact detection capabilities. The collection of raw unfiltered water, as performed in this study, allows for detection of small sized taxa which may otherwise be excluded through use of equipment such as plankton nets. The use of non-size exclusionary sampling equipment (e.g., Van Dorn samplers, peristaltic pumps, Kemmerer bottles; Graham et al., 2008) must be standardized in sampling protocols to ensure the collection of accurate data. While guidelines and resources are available to assist in the development of sampling protocols for monitoring, these should be viewed as starting points to characterize cyanobacterial communities and should be adapted to the dynamics of the system of interest. The efficacy of cyanobacteria monitoring protocols is dependent on understanding the system including physiochemical characteristics (e.g., thermal stratification) and cyanobacterial community structure (Pobel et al., 2011; Welker et al., 2021) which can only be achieved through more intensive monitoring efforts. To optimize cyanobacteria monitoring, protocols must be tailored to specific systems, even when sites are located within the same watershed, and must incorporate sampling time to reflect diurnal variability. Initial characterization of cyanobacterial communities should account for spatial and temporal distribution through inclusion of discrete depth sampling across depth profiles over a multi-time point sampling series using non-size exclusionary sampling equipment. Detailed knowledge on the community composition and spatiotemporal variability of systems is necessary for development of ecologically accurate sampling protocols. While sampling efforts may be limited due to time and resources, the reduction in sampling efforts may result in missed changes in cyanobacterial communities which may impose threats to water quality.

3.4 Conclusions

- The potential for diurnal migration should be reflected in cyanobacterial monitoring programs through the inclusion of multiple sampling times or at minimum, conducting sampling at an ecologically significant time of day.
- Sampling of lakes must not be restricted to water surface and should incorporate multiple depths to reflect spatial heterogeneity in the water column.
- The positive buoyancy of cyanobacteria with gas vesicles may frequently be concentrated at the surface but is not limited to the surface with occurrence at deeper depths in the water column.
- Rainfall or wind induced mixing events such as those observed in this investigation may significantly impact cyanobacterial community composition and distribution.
- Cyanobacterial monitoring may be enhanced through the incorporation of system characteristics (e.g., thermal stratification) and characterization of communities for design of sampling protocols that are system specific.

Chapter 4
Early Seasonal Increases in Relative Abundance of Potentially Toxic Cyanobacteria: A concerning impact of Climate Change Protracting the Vernal Window in Northern Temperate Lakes?

4.1 Introduction

Steady increases in harmful algal blooms (HABs) have been observed in freshwater environments for decades as a result of increased nutrient availability associated with changing climate and anthropogenic activities, including watershed and land development, construction of dams/impoundments, river diversion, deforestation, and environmental discharges acting as point and non-point sources (Creed et al., 2015; Emelko et al., 2016; Winter et al., 2011). While the term 'algae' is utilized frequently when discussing aquatic systems, differentiation of the types of organisms present within this diverse group is critical for understanding their roles in aquatic ecosystems and potential impacts to water quality. Photosynthetic cyanobacteria included in this group threaten water quality through the formation of dense blooms and the production of secondary metabolites including taste and odor compounds (e.g., geosmin, 2-methylisoborneol[MIB]) and several cyanotoxins of human and environmental health concern (Harke et al., 2016; Huisman et al., 2018; Paerl, 2014; Vu et al., 2020). These include hepatotoxins (e.g., microcystins, nodularin, cylindrospermopsin), neurotoxins (e.g., saxitoxins, anatoxin-a, anatoxin-a(s), homoanatoxin-a), cytotoxins (e.g., aplysiatoxin, debromoaplysiatoxin, lingbyatoxin, lipopolysaharide endotoxin), and skin and gastrointestinal irritants (Cheung et al., 2013). A non-photosynthetic sister-clade of cyanobacteria—Melainabacteria—was recently identified in groundwater, tap water,

municipal wastewater stabilization ponds, and the human gut (Di Rienzi et al., 2013; Soo et al., 2014) which may hypothetically serve as a novel contributor to toxicity (Nunes-Costa et al., 2020).

Fundamental physiological differences found between different algal taxa influence their distribution (Irwin et al., 2012, 2006) with growth being limited by temperature resulting in specific seasonal and geographical distributions of populations. (Butterwick et al., 2005; Dell et al., 2011). Lakes in temperate regions exhibit seasonal succession of algal populations driven by changes in water temperature (Butterwick et al., 2005; Dell et al., 2011), light availability, water column stratification, and nutrient availability which are influenced by meteorological conditions (Rusak et al., 2018). While exceptions exist to 'typical' seasonal succession of algal populations (Fanesi et al., 2016), typical trends in temperate systems consist of winter communities dominated by cryptophytes, chrysophytes and diatoms, (Beall et al., 2016; Felföldi et al., 2016; Phillips and Fawley, 2002), high abundances of diatoms in the spring (Jaworska and Zdanowski, 2012; Winder et al., 2009), followed by the dominance of green algae (Staehr and Birkeland, 2006; Winder et al., 2009; Winder and Hunter, 2008) and cyanobacteria in the summer (Staehr and Birkeland, 2006; Winder et al., 2009; Winder and Hunter, 2008).

Although algal populations are present year-round, research and sampling of aquatic ecosystems has largely focused on the spring and summer resulting in a critical gap in knowledge and understanding on winter limnological processes (Felföldi et al., 2016; Wilhelm et al., 2014). Winter sampling can often be logistically complicated (e.g., difficult to access sampling sites, safety) and traditionally, ice-covered systems have been viewed as 'dormant' further resulting in a lack of interest in winter processes (Felföldi et al., 2016; Hampton et al.,

2015; Powers and Hampton, 2016). However, recent research has revealed the presence of unique niches in microbial and phytoplankton communities adapted to low temperatures and low light during the winter (Phillips and Fawley, 2002; Tran et al., 2018). Of further concern, winter cyanobacterial blooms have been detected under ice in systems (Wejnerowski et al., 2018) demonstrating the importance of developing seasonally inclusive monitoring designs to advance knowledge on winter community dynamics.

In addition to the seasonal occurrence and dominance of taxa within the phytoplankton community, organisms are rarely homogeneously distributed within lakes (Cyr, 2017; Vidal et al., 2014). Water currents results in patchy distribution across different areas of aquatic systems (Cyr, 2017) but organisms may also be vertically distributed in response to gradients in temperature, light, and nutrient availability (Jakubowska and Szeląg-Wasielewska, 2015; Vidal et al., 2014) or as a result of cell size (Chien et al., 2013; Śliwińska-Wilczewska et al., 2018). Picocyanobacteria, cells ranging in 0.2 – 2.0 μm in size, are able to maintain their water column position (Śliwińska-Wilczewska et al., 2018) and other cyanobacteria may have gas vacuoles which allow for buoyancy regulation and subsequently determination of water column position (Pfeifer, 2012). For example, picocyanobacteria can grow in low light intensity environments at deeper depths (Jakubowska and Szeląg-Wasielewska, 2015) while large *Microcystis* colonies may be found closer to the surface due to positive buoyancy regulation (Visser et al., 2005). The association of cyanobacteria with the formation of blooms at the surface has resulted in sampling protocols focused on visual detection of biomass (Chorus and Bartram, 1999; Newcombe, 2009). Cyanobacteria have typically been associated with the ability to withstand higher levels of light irradiance, further supporting the notion to focus sampling efforts at the water surface, but recent research has revealed higher light

sensitivity in cyanobacteria resulting in cellular damage (Beecraft et al., 2019) suggesting that populations may not be concentrated at the surface as traditionally expected. Additionally, picocyanobacteria have been detected in higher abundances in the upper hypolimnion (Pick and Agbeti, 1991) further demonstrating the need to sample deeper depths.

Although previous research has characterized the spatial and seasonal variability in phytoplankton communities, these characterized trends may be challenged as a result of shifting environmental conditions imposed by climate change. Specifically, in northern temperate systems, temperatures have significantly increased, growing seasons have lengthened, and precipitation has declined and shifted such that additional peaks in precipitation and discharge are observed in the late fall (Creed et al., 2015). Due to the interaction and linkage between terrestrial and aquatic ecosystems, factors such as the vernal and autumnal windows may significantly impact cyanobacterial dynamics (Creed et al., 2015). The vernal window has been previously defined as the "period that marks the end of winter and start of the growing season during which rapid shifts in ecosystem energy, water, nutrient and carbon dynamics occur" by Contosta et al., 2017, and the autumnal window as the period between the canopy leaf fall and onset of snowpack by Creed et al., 2015. Both windows signify periods of high flow associated with spring snow melts and autumnal storms resulting in high periods of terrestrial runoff (Creed et al., 2015) directly impacting the physical characteristics and subsequently the ecological characteristics of lakes (Contosta et al., 2017). While these terms are typically used in reference to terrestrial systems, these periods of rapid shifts in energy, nutrients and carbon are also applicable to freshwater vernal and autumnal windows as evident by the prevalence of cyanobacterial blooms into the late fall in oligotrophic lakes in Ontario (Winter et al., 2011). Although previous research in this realm has been

conducted, it is of note that studies on algal blooms in North American temperate systems have not examined the (i) direct impacts on lakes located within a northern temperate forested watershed and (ii) cyanobacterial community dynamics in absence of visual biomass warranting further evaluation of shifting trends in spatiotemporal variation as a result of climate change.

The spatial and seasonal trends in cyanobacterial community composition were evaluated using 16S rRNA gene amplicon sequencing to explore cyanobacterial dynamics in oligotrophic lakes within a northern temperate forested watershed and to identify the potential downstream effects of shifting vernal windows on these systems. Specifically, depth profiles were used to evaluate vertical distribution patterns during the summer, and seasonal trends were evaluated over an interannual period in three oligotrophic lakes to elucidate spatiotemporal trends in cyanobacterial community composition. The exploration of spatiotemporal trends present in cyanobacteria communities provides a comprehensive evaluation on the potential impact of sampling month and depth on water quality sampling protocols and will provide critical insight on developing ecologically meaningful sampling protocols for cyanobacteria in oligotrophic lakes. Additionally, the inclusion of seasonal sampling in the winter will continue to advance the fields of winter limnology and winter cyanobacterial dynamics.

4.2 Methods

4.2.1 Study Site: Turkey Lakes Watershed

The Turkey Lakes Watershed (TLW) Study was established in 1980 to investigate ecosystem effects of acidic atmospheric deposition—Jeffries et al. (1988) provided a comprehensive description of the physical characteristics of the watershed. In brief, the TLW is situated about 50 km north of Sault Ste. Marie, Ontario on the Canadian Shield in an uneven-

aged tolerant hardwood and mixed conifer forest landscape. Geological parent materials in the watershed consist of Precambrian silicate greenstone (i.e., metamorphosed basalt) (Semkin and Jeffries 1983)—glacial till overlies the bedrock. At higher elevations (i.e., Wishart Lake), the till thickness is less than 1 m, with frequent surface exposure of bedrock; there is substantially more till (1 to 2 m) at lower elevation (i.e., Big Turkey Lake, Little Turkey Lake) (Jeffries et al., 1988). It consists of four interconnected lakes that are classified as oligotrophic to mesotrophic and fed by both first order streams and groundwater: Batchawana Lake, Wishart Lake, Little Turkey Lake and Big Turkey Lake (Figure 1.1). Except for Wishart, each of these lakes thermally stratify during summer and winter annually (Figure A1). Wind-induced mixing in this shallow lake generally prevents thermal stratification. During periods of stratification, oxygen is depleted in at lower lake depths; zones of anoxia sometimes develop in Batchawana and Little Turkey Lakes. Macrophytes are abundant along the margins of the lakes (Smokorowski et al., 2021) and cyanobacteria were historically identified as dominant members of the phytoplankton communities (Jeffries et al., 1988).

4.2.2 Sample Collection

To explore seasonal and inter-annual variability of bacterial communities (especially cyanobacteria), water samples were collected between July 2018 and January 2020 at Secchi depth (during ice-free periods) and 0.25 m below ice (during periods of ice cover), at the deepest locations in Big Turkey, Little Turkey, and Wishart Lakes. Specifically, samples were collected once a month in July-August 2018, October 2018, February-March 2019, May-August 2019 and January 2020. This is the water depth at which light penetration is approximately 1% of surface illumination; it is considered to be the maximum depth at which there is generally sufficient light for photosynthesis (Bukata et al., 1988). Secchi depth was

measured in a standardized manner from the shaded side of a boat by two individuals. In addition to the long-term seasonal sampling, a second more detailed sampling program was undertaken in July and August of 2018. Here, samples were collected from the water surface, Secchi depth and one meter below Secchi depth at each lake site to describe the vertical distribution of cyanobacterial communities within the water column, including depths of especially low light intensity that have not been widely investigated. Details regarding the sampling program are provided in Table A2.

4.2.3 DNA Extraction & 16S rRNA Gene Amplicon Sequencing

DNA extraction was performed using the DNeasy PowerSoil Kit (QIAGEN Inc., Venlo, Netherlands) following the manufacturer's protocol. Elution buffer was added to the spin columns for 15 minutes prior to elution of the DNA extract. DNA was quantified using a NanoDrop spectrophotometer when possible (Table A2). Although the detection limit is reported as 2ng/µl (Thermo Fisher Scientific, 2009) absolute values have been shown to be less accurate for DNA concentrations less than 10ng/µl (Khetan et al., 2019) resulting in potentially erroneous quantification for low concentration environmental samples used in this study. The DNA extracts were submitted for amplicon sequencing using the Illumina MiSeq platform (Illumina Inc., San Diego, United States) at a commercial laboratory (Metagenom Bio Inc.,Waterloo, ON). Primers designed to target the 16S rRNA gene V4 region [515FB (GTGYCAGCMGCCGCGGTAA) and 806RB (GGACTACNVGGGTWTCTAAT)] (Walters et al., 2015) were used for PCR amplification.

4.2.4 Sequence Processing & Library Size Normalization

The program QIIME2 (v. 2019.10; Bolyen et al., 2019) was used for bioinformatic processing. Demultiplexed paired-end sequences were trimmed and denoised, including the

removal of chimeric sequences and singleton sequence variants, using DADA2 (Callahan et al., 2016) to construct the amplicon sequence variant (ASV) table. Taxonomic classification was performed using a Naïve-Bayes taxonomic classifier trained using the SILVA138 database (Quast et al., 2013; Yilmaz et al., 2014). Taxonomic assignments for amplicon sequence variants (ASV) classified as Cyanobacteria at the phylum level were manually curated to reflect taxonomic assignments above the genus level according to AlgaeBase. In addition to the manual curation of the higher taxonomic levels of cyanobacterial sequences, sequences classified as Melainabacteria are discussed herein with other sequences classified as photosynthetic cyanobacterial genera. While Melainabacteria have previously been included as an order within the phylum Cyanobacteria within databases such as SILVA (Quast et al,, 2013; Yilmaz et al., 2014), more recently, it has been proposed the phylum Cyanobacteria is restricted to only include "organisms in the domain bacteria able to carry out oxygenic photosynthesis with water as an electron donor and to reduce carbon dioxide as a source of carbon, or those secondarily evolved from such organisms" (Garcia-Pichel et al., 2019) resulting in the segregation of Cyanobacteria and Melainabacteria to separate taxonomic groups. However, for the purpose of this research, reflecting the classifications presented in the SILVA database (Quast et al., 2013; Yilmaz et al., 2014), sequences classified as Melainabacteria are presented herein with other photosynthetic cyanobacterial sequences for brevity and to highlight the impact of non-photosynthetic organisms in microbial processes. As more research is conducted on these organisms, this decision may not be reflective of current accepted taxonomic status and should be noted.

Files from QIIME2 were imported into R (v. 4.0.1) for downstream analyses using *qiime2R* (v. 0.99.23; Bisanz, 2018). Initial sequence libraries were filtered to exclude amplicon

sequence variants that were taxonomically classified as mitochondria or chloroplast sequences using *phyloseq* (v. 1.32.0; McMurdie and Holmes, 2012). For cyanobacterial community analysis, sequence variants classified as Cyanobacteria at the phylum level were filtered to create libraries consisting of only cyanobacteria classified sequences. Samples were repeatedly rarefied without replacement to normalized library sizes using *mirlyn* (Cameron and Tremblay, 2020). Cyanobacterial communities were repeatedly rarefied to 824 reads for community analysis of the long-term seasonal samples and summer 2018 depth profile samples. Long-term seasonal samples were repeatedly rarefied to a normalized library size of 824 reads for cyanobacterial community analysis. Samples with less than 500 cyanobacteria sequences were excluded from the seasonal diversity analyses.

4.2.5 Taxonomic Composition & Community Diversity Analyses

Stacked bar charts were created to visualize the bacterial community composition as relative abundances of major phyla across the sampling regimes using *mirlyn*. The composition of cyanobacterial communities was further assessed at the taxonomic order level, and relative abundances were visualized using a heatmap. Relative abundances were randomized across phyla within samples to identify significantly dominant groups within the bacterial community in relation to sampling conditions with a Bonferroni correction. While a p-value of 0.05 is frequently viewed as the threshold for significance, due to the heterogeneity and high variability present in environmental systems, a p-value of 0.2 has been selected as the threshold for significance in these communities.

To confirm the taxonomic classification performed by the Naïve-Bayes classifier, a phylogenetic tree was constructed in MEGA X using cyanobacterial reference sequences and sequences from samples (Figure D.4). Sequences classified to the genera highlighted in

guidelines for sampling protocol deigns (Graham et al., 2008; Vidal et al., 2021) including *Microcystis*, *Anabaena*, *Aphanizomenon*, *Pseudanabaena*, and *Synechococcus* were evaluated further because of the importance of bloom- and toxin-forming taxa for water quality management, though notably, other taxa contributed to the compositional structure at the order level. Sequences classified to two additional genera were also included herein. *Radiocystis* can carry 16S rRNA genes that are identical *Microcystis* (Vidal et al., 2014) and can produce toxin (Vieira et al., 2003). *Cyanobium* is a potentially toxic picocyanobacteria that is increasingly implicated as a contributor to phytoplankton blooms (Śliwińska-Wilczewska et al., 2018). Relative abundances of selected cyanobacterial sequences were visualized using a heatmap and were evaluated based on unicellular, filamentous, or colonial morphologies to characterize taxa-specific trends.

Community diversity analyses were performed using *mirlyn* on repeatedly rarefied libraries. The Shannon Index (Shannon, 1948) was analyzed to identify trends in alpha diversity as a function of sampling depth or sampling month. To explore similarity in community composition, rarefied libraries were transformed using a Hellinger transformation and then used to calculate Bray-Curtis distances (Bray and Curtis, 1957) used in principal component analysis (PCA).

4.3 Results

4.3.1 Spatiotemporal Variation of Cyanobacterial Communities

Within the bacterial community, compositional shifts were observed seasonally with distinct fluctuations in the relative abundances of cyanobacterial sequences (0 – 56.31%; Figure 4.1; Table 4.1). In Big Turkey Lake, these sequences were significantly represented in the bacterial community in all ice-free sampling months ($p < 0.05$; Table D3) except for August

2018 where they were significantly represented ($p = 0.1$). During the summer months (June, July, August) in Little Turkey and Wishart Lake, cyanobacteria were significantly represented in the bacterial communities ($p < 0.009$), apart from August 2018 for both lakes ($p = 1$). However, unlike the bacterial communities observed in Big Turkey Lake, cyanobacterial sequences were not significantly represented in October and May in both Little Turkey and Wishart Lake ($p = 1$) indicating system specific seasonal trends. Cyanobacterial sequences were not significantly represented during the ice-covered months of February 2019, March 2019, and January 2020 in Big Turkey ($p > 0.96$; Table D3), Little Turkey ($p = 1$), and Wishart Lake ($p = 1$) coinciding with the detection of minimal abundances or complete absence. Although cyanobacteria comprised less than 1% of the total bacterial community during ice-cover months, the detected sequences were largely composed of ASVs classified to the poorly studied, non-photosynthetic basal lineages of cyanobacteria, Melainabacteria and Sericytochromatia (Figure D1; Table 4.1).

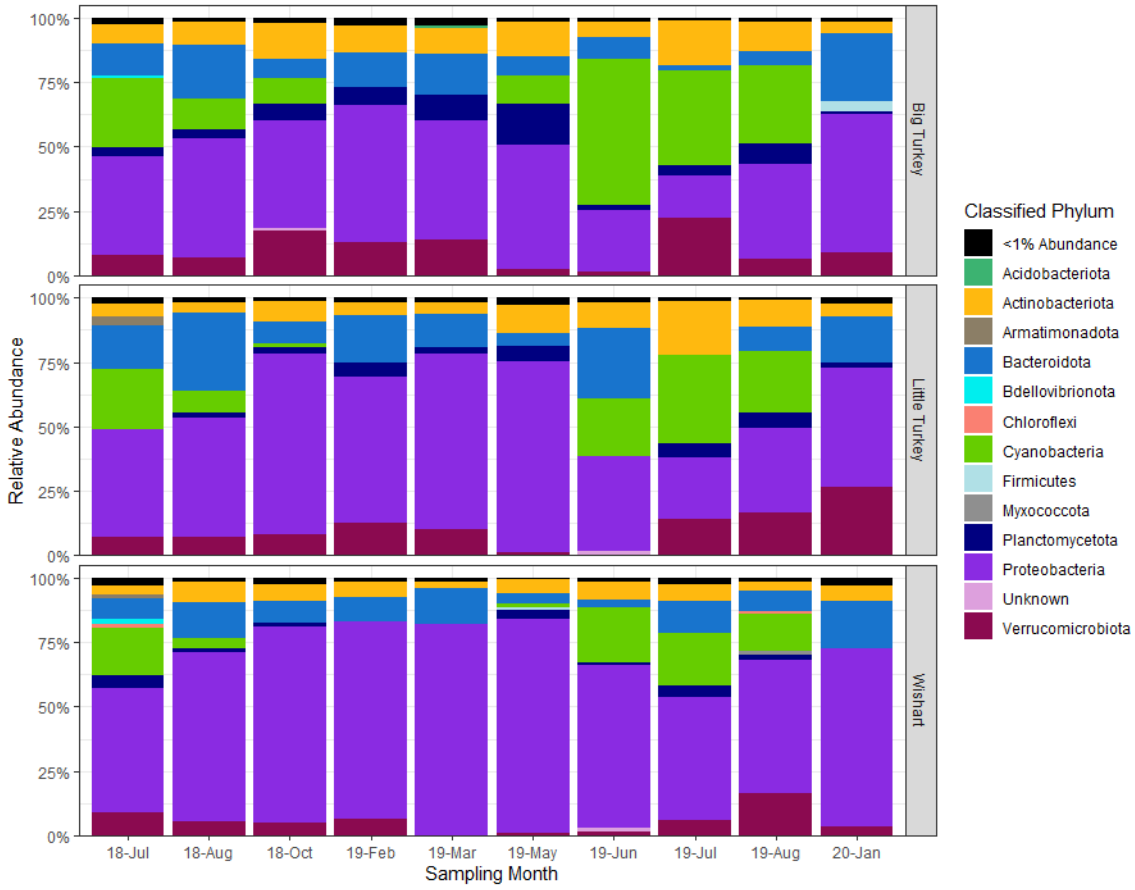


Figure 4.1 Stacked bar charts depicting the relative abundances of major bacterial phyla identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a multi-seasonal timeframe in a non-stratified lake (Wishart), a mid-sized stratified lake (Little Turkey) and deep stratified lake (Big Turkey).

Table 4.1 Relative abundances as percentages of cyanobacterial sequences represented within the bacterial community, and the relative abundances of taxonomic orders that compose the cyanobacterial community across a seasonal period.

Taxonomic Level	Big Turkey	Little Turkey	Wishart	Sampling Month
Cyanobacteria	26.99	23.57	18.46	18-Jul
	11.81	8.54	3.84	18-Aug
	10.10	1.20	0.905	18-Oct
	0.166	0	0.0104	19-Feb
	0.181	0	0	19-Mar
	11.07	0.77	1.51	19-May
	56.31	22.04	21.14	19-Jun
	36.60	34.23	20.46	19-Jul
	30.35	23.99	14.37	19-Aug
	0	0.0651	0.112	20-Jan
Chroococcales	3.19	10.33	5.17	18-Jul
	14.00	11.22	0	18-Aug
	8.60	0	1.50	18-Oct
	0	N/A	0	19-Feb
	0	N/A	N/A	19-Mar
	0.22	0	0	19-May
	0.53	3.75	1.00	19-Jun
	11.70	10.49	1.58	19-Jul
	29.14	11.87	2.80	19-Aug
	N/A	42.86	0	20-Jan
Nostocales	0	0.43	0.12	18-Jul
	0	0	0	18-Aug
	1.73	1.79	0.75	18-Oct
	0	N/A	0	19-Feb
	0	N/A	N/A	19-Mar
	0	0	0	19-May
	0	1.20	0.71	19-Jun
	0.09	0	0	19-Jul
	0.28	0.22	0	19-Aug
	N/A	0	0	20-Jan
Synechococcales	96.41	87.25	92.53	18-Jul
	84.36	87.96	93.20	18-Aug
	89.03	93.91	90.23	18-Oct
	100	N/A	0	19-Feb
	100	N/A	N/A	19-Mar
	99.78	100	100	19-May
	99.47	94.56	97.27	19-Jun
	87.95	89.17	96.03	19-Jul

Table 4.1 Continued

Taxonomic Level	Big Turkey	Little Turkey	Wishart	Sampling Month
Synechococcales	69.04	87.44	93.47	19-Aug
	N/A	0	25.58	20-Jan
Non-Photosynthetic Orders				
Caenarcaniphilales	0	0	0	18-Jul
	0.13	0	0	18-Aug
	0.07	0	0	18-Oct
	0	N/A	0	19-Feb
	0	N/A	N/A	19-Mar
	0	0	0	19-May
	0	0	0	19-Jun
	0	0	0	19-Jul
	0	0	0	19-Aug
	N/A	57.14	11.63	20-Jan
Obscuribacterales	0.04	0	0	18-Jul
	0.06	0	0	18-Aug
	0	0	0	18-Oct
	0	N/A	100	19-Feb
	0	N/A	N/A	19-Mar
	0	0	2.78	19-May
	0	0	0	19-Jun
	0	0	0	19-Jul
	0.18	0	0	19-Aug
	N/A	0	39.53	20-Jan
Sericytochromatia	0	0	0	18-Jul
	0	0	0	18-Aug
	0	0	0	18-Oct
	0	N/A	0	19-Feb
	0	N/A	N/A	19-Mar
	0	0	0	19-May
	0	0	0	19-Jun
	0	0	0	19-Jul
	0	0	0	19-Aug
	N/A	0	18.6	20-Jan
Vampirovibrionales	0.1	0.08	0	18-Jul
	0	0	0	18-Aug
	0	0	0	18-Oct
	0	N/A	0	19-Feb
	0	N/A	N/A	19-Mar
	0	0	0	19-May
	0	0	0	19-Jun

Table 4.1 Continued

Taxonomic Level	Big Turkey	Little Turkey	Wishart	Sampling Month
	0	0	0	19-Jul
	0.27	0	0	19-Aug
	N/A	0	0	20-Jan

To further evaluate the dynamics of cyanobacterial populations during the summer months when cyanobacteria were abundant, the lakes were sampled across a depth profile in July and August of 2018 to identify spatial variation in the water column distribution of populations. Varying spatial trends were observed in cyanobacterial abundances (3.84 – 27.27%) in the bacterial communities between lakes sites (Figure 4.2; Table 4.2). Cyanobacterial sequences were significantly represented in the bacterial community across all sampling depths in Big Turkey and Wishart Lake in July 2018 ($p < 0.00002$; Table D4) indicating the presence of high abundances distributed throughout the water column. Although sequences were significantly represented across all sampling depths in Big Turkey Lake, abundances were observed to increase across depth with a maximum observed at the deepest sampled point (45.53%). Alternatively, sequence abundance in Wishart Lake did not experience large fluctuations across depths and exhibited a more homogeneous distribution throughout the water column (3.84 – 4.77%). In Little Turkey Lake, cyanobacterial sequences were only significantly represented in the bacterial community at Secchi depth in July 2018 ($p = 5.44e-05$) and not significantly represented at the surface ($p = 0.29$) and below Secchi depth ($p = 0.53$). Similar to the results obtained in the seasonality study, sequences were not significantly represented across sampling depths in Wishart Lake ($p = 1$) depths and at the surface, and Secchi depth in Little Turkey Lake ($p = 1$) in August 2018, corresponding to the seasonal decrease in cyanobacterial populations within the bacterial community. Alternatively,

they were not significantly represented in the bacterial communities of Big Turkey Lake across sampling depths in August 2018 ($p < 0.45$) and at the deepest sampling point in Little Turkey Lake ($p = 0.44$). Due to the diversity in distribution arising from cellular morphology, specific spatial distribution trends of cyanobacterial communities are discussed in further detail in Section 4.3.2.

Sequence libraries were repeatedly rarefied to a normalized library size of 824 sequences to evaluate the spatiotemporal variation observed in cyanobacterial community diversity. It is of note that the ability to examine seasonal trends in diversity were limited due to the low sequence counts present in samples outside of traditional field seasons (e.g., October – May). However, of the data that could be analyzed, seasonal trends in community diversity were revealed between ice-free sampling months. In Big Turkey Lake and Little Turkey Lake, similar Shannon Index values were observed in July and August (Figure 4.3A). In Wishart Lake, communities were more diverse in July than August. In addition to trends between sampling months, interannual variation between sampling years was detected in Big Turkey and Wishart Lake. In July and August 2019, Big Turkey Lake exhibited higher community diversity and Wishart Lake exhibited lower community diversity comparatively to the 2018 sampling months.

Cyanobacterial communities were observed to show both dissimilarity between lakes and between sampling months (Figure 4.3B). Communities were unique between the interconnected lakes located within the same watershed. Within Wishart Lake and Little Turkey Lake, there was high similarity in cyanobacterial communities between sampling months in different years showing similar community composition occurring annually. In contrast, the communities of Big Turkey Lake showed more variation in composition with the

May sample being most dissimilar from other sampling months. High similarity was observed between August communities of differing years in Big Turkey Lake but more dissimilarity in July of different sampling years further demonstrating the occurrence of interannual variation in community composition. Although Big Turkey, Little Turkey and Wishart Lake are interconnected, unique cyanobacterial community composition was observed between lakes. This supports the need for system-specific monitoring to ensure accurate detection of cyanobacteria communities.

In addition to the seasonal trends observed in cyanobacterial diversity, spatial trends were also identified. In August of 2018, the cyanobacterial community diversity was highest at the surface of all three lakes with comparable values of the Shannon Index detected (Figure 4.4A). Additionally, the diversity values of Big Turkey and Little Turkey Lake in August were comparable. In July, the Shannon Index values were similar across all three sampling depths in Wishart Lake supporting that well-mixed water columns results in homogenous distributions creating similar diversity values at all sampling depths. In Big Turkey Lake, similar diversity values were detected at both surface and the deepest point, with highest values measured at Secchi. Comparable values were also detected in cyanobacteria diversity at Secchi depth between Big Turkey and Little Turkey Lake in July and comparable values to those detected in August. Little Turkey Lake showed similar trends in diversity between sampling months. Wishart experienced a decrease in cyanobacteria diversity from July to August while alternatively, the surface and deep sampling points in Big Turkey experienced marked increases in diversity from July to August.

Cyanobacterial communities showed distinct compositional similarities within lakes. Wishart and Little Turkey Lake communities showed high similarity between sampling months

(Figure 4.4B). Within Little Turkey Lake, communities collected at the same sampling depth were found to be more similar. However, the surface communities of Little Turkey Lake between sampling months showed higher dissimilarity than observed at Secchi depth and below. Similar to the seasonal variation, cyanobacterial communities in Big Turkey Lake showed the largest dissimilarity between sampling points. Communities sampled during the same month, but different depths showed more similarity than communities sampled at the same depth in different sampling months. However, communities were still distinct at different depths. This suggests that in deep, stratified, oligotrophic lakes, distinct communities occur spatially and seasonally and may create challenges in monitoring with the conjunction of diurnal variation (Chapter 3). Alternatively, in the shallow, non-stratified lake, seasonally and spatially unique communities did not occur.

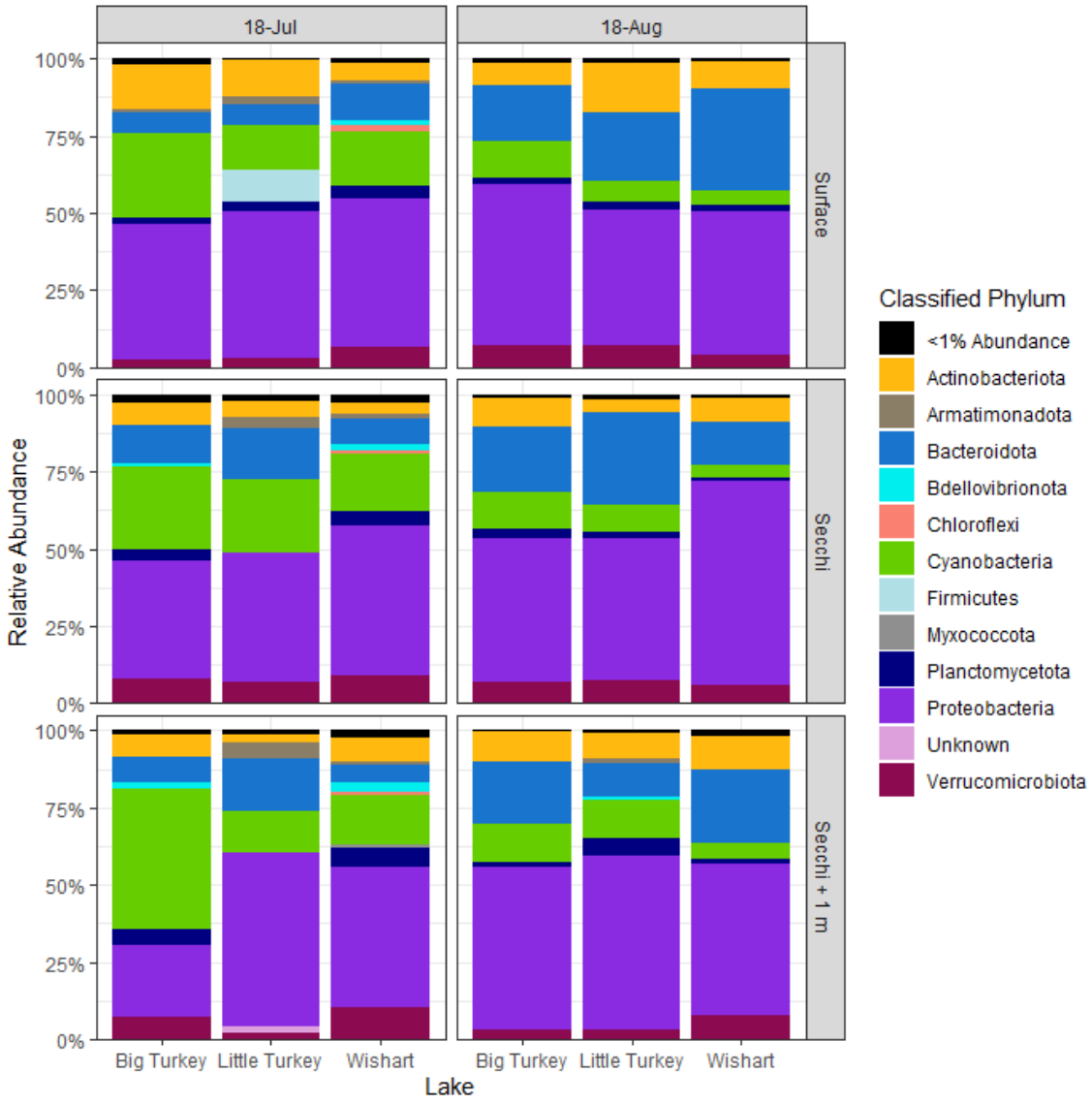
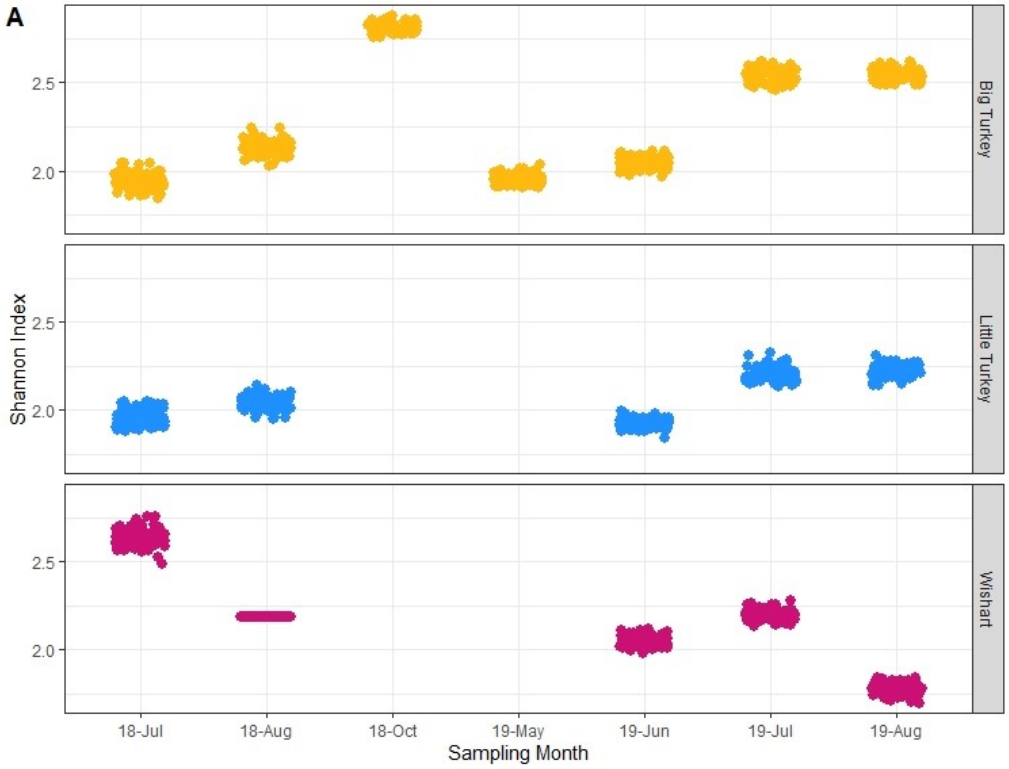


Figure 4.2 Stacked bar charts depicting the relative abundances of major bacterial phyla identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a depth profile during the summer of 2018 in a non-stratified lake (Wishart), a mid-sized stratified lake (Little Turkey) and deep stratified lake (Big Turkey).

Table 4.2 Relative abundances of cyanobacterial sequences represented within the bacterial community, and the relative abundances of taxonomic orders that comprise >1% of the cyanobacterial community across a depth profile.

Taxonomic Level	July 2018			August 2018			Sampling Depth
	Big Turkey	Little Turkey	Wishart	Big Turkey	Little Turkey	Wishart	
Cyanobacteria	27.33	14.35	17.57	11.52	6.82	4.53	Surface
	27.10	23.57	18.62	11.82	8.54	3.84	Secchi
	45.53	13.86	16.03	12.79	12.61	4.77	Secchi + 1 m
Chroococcales	6.96	9.13	12.04	9.28	30.76	3.30	Surface
	3.20	10.34	5.18	14.00	11.22	0	Secchi
	0.75	6.26	5.69	6.98	12.22	2.43	Secchi + 1 m
Nostocales	1.38	0.68	0	0	0	0	Surface
	0	0.43	0.12	0.75	0	0	Secchi
	0	2.31	0	0	0	0	Secchi + 1 m
Synechococcales	90.16	86.65	86.48	88.97	65.18	89.39	Surface
	96.41	87.26	92.53	84.37	87.96	93.20	Secchi
	99.15	90.65	92.92	92.39	86.92	91.30	Secchi + 1 m


A
Big Turkey
Little Turkey
Wishart
Shannon Index
2.5
2.0
18-Jul
18-Aug
18-Oct
19-May
19-Jun
19-Jul
19-Aug
Sampling Month

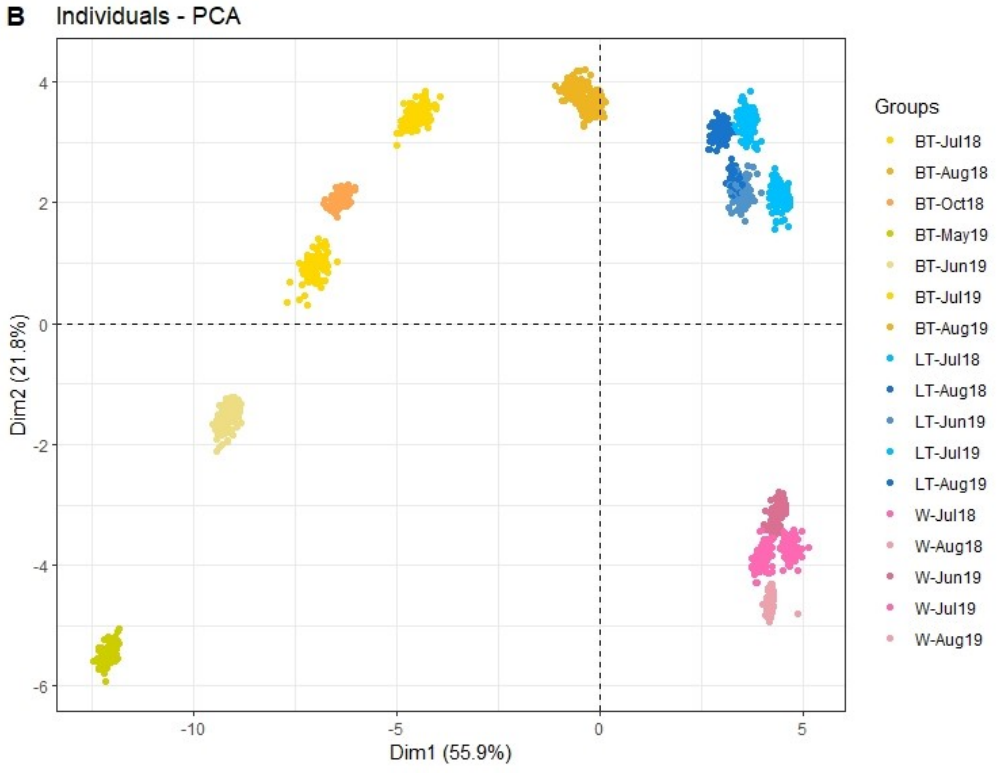
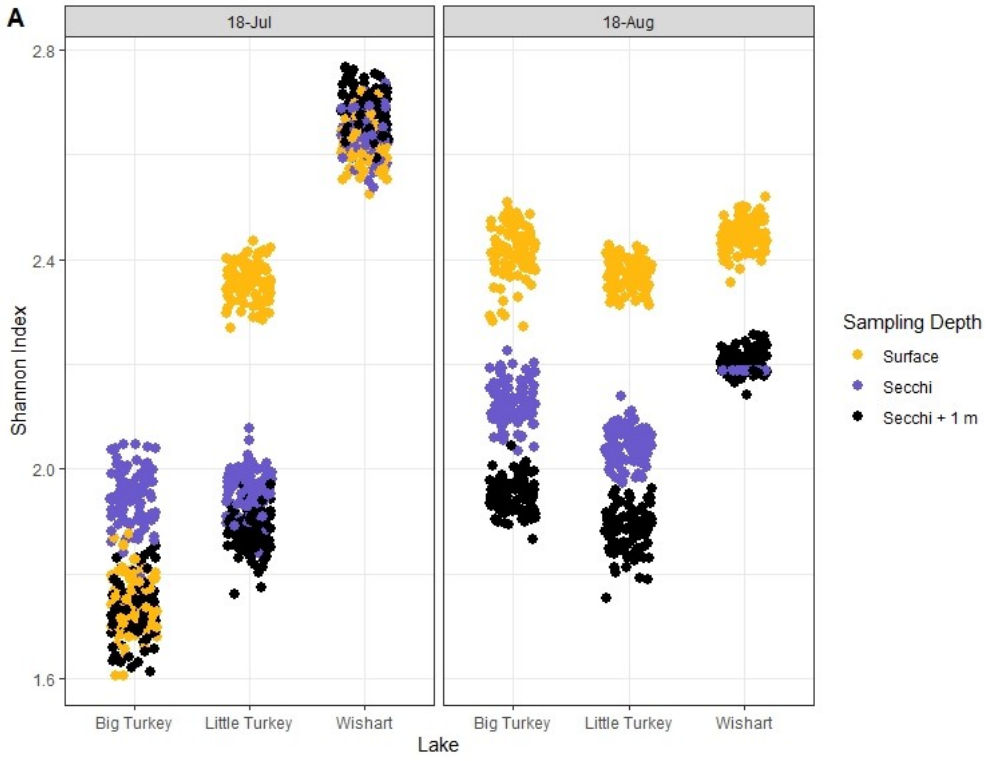

B
Individuals - PCA
Dim2 (21.8%)
Dim1 (55.9%)
Groups
BT-Jul18
BT-Aug18
BT-Oct18
BT-May19
BT-Jun19
BT-Jul19
BT-Aug19
LT-Jul18
LT-Aug18
LT-Jun19
LT-Jul19
LT-Aug19
W-Jul18
W-Aug18
W-Jun19
W-Jul19
W-Aug19

Figure 4.3 Alpha and beta-diversity analyses of cyanobacterial communities collected across the ice-free seasonal sampling series in a deep stratified (Big Turkey), mid-sized stratified (Little Turkey) and shallow non-stratified lake (Wishart). Amplicon sequence variants classified to the phylum Cyanobacteria were filtered to characterize diversity within cyanobacterial communities. The Shannon Index was calculated on rarefied libraries to evaluate the effects of (A) lake site and sampling month on community diversity. (B) The Bray-Curtis dissimilarity metric was used to explore similarities in communities between lake site (Little Turkey = LT, Wishart = W) and sampling month demonstrating unique communities between lakes.


A
18-Jul
18-Aug
Shannon Index
2.8
2.4
2.0
1.6
Big Turkey
Little Turkey
Wishart
Lake
Sampling Depth
Surface
Secchi
Secchi + 1 m

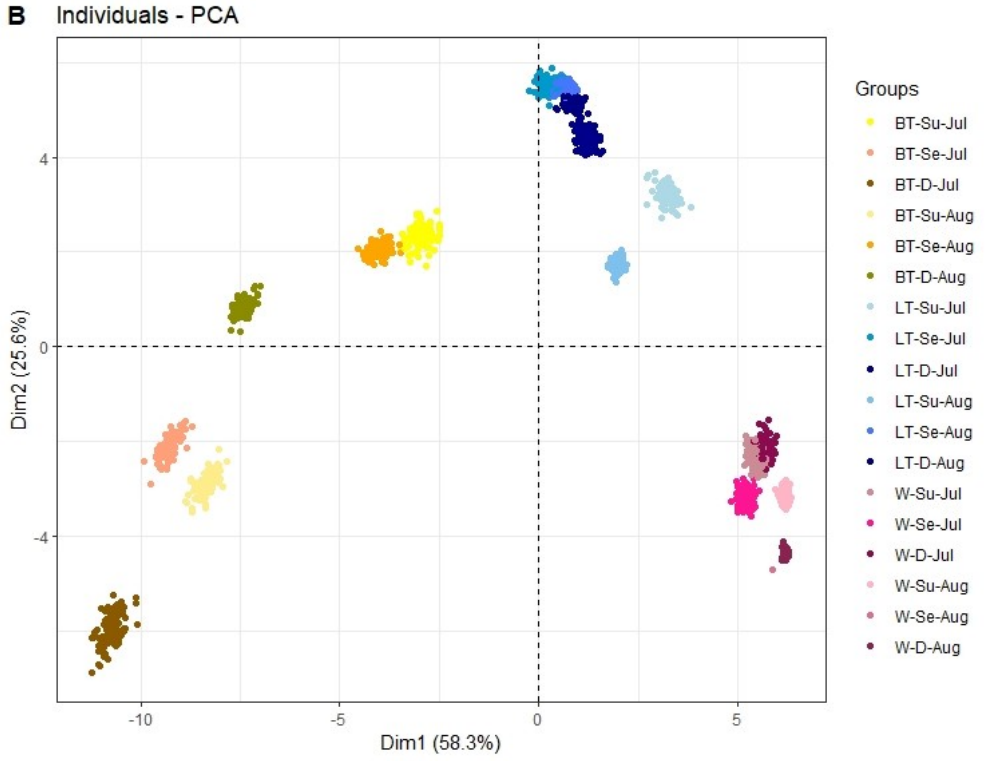

B
Individuals - PCA
Dim2 (25.6%)
Dim1 (58.3%)
4
0
-4
-10
-5
0
5
Groups
BT-Su-Jul
BT-Se-Jul
BT-D-Jul
BT-Su-Aug
BT-Se-Aug
BT-D-Aug
LT-Su-Jul
LT-Se-Jul
LT-D-Jul
LT-Su-Aug
LT-Se-Aug
LT-D-Aug
W-Su-Jul
W-Se-Jul
W-D-Jul
W-Su-Aug
W-Se-Aug
W-D-Aug

Figure 4.4 Alpha and beta-diversity analyses of cyanobacterial communities collected across a depth profile in the summer of 2018 in a deep stratified (Big Turkey), mid-sized stratified (Little Turkey) and shallow non-stratified lake (Wishart). Amplicon sequence variants classified to the phylum Cyanobacteria were filtered to characterize diversity within cyanobacterial communities. The Shannon Index was calculated on rarefied libraries to evaluate the effects of (A) lake site, summer sampling month and sampling depth on community diversity. (B) The Bray-Curtis dissimilarity metric was used to explore similarities in communities between lake site (Little Turkey = LT, Wishart = W), sampling month and sampling depth (Su = Surface, Se = Secchi, D = Secchi + 1 m) demonstrating unique communities between lakes.

4.3.2 Spatiotemporal Variation of Potentially Bloom Forming & Toxic Taxa

To explore the impact of spatial distribution on taxa, the taxonomic and ASV composition of cyanobacterial communities were assessed. Across samples, a total of 97 ASVs classified to the phylum Cyanobacteria were identified including taxa belonging to the cyanobacterial orders Chroococcales, Nostocales, and Synechococcales (Table 4.3). For analogous comparatives to information provided in guidelines for monitoring, cyanobacterial orders were grouped by morphology as follows: (i) Unicellular Taxa – Synechococcales (*Cyanobium*, *Synechococcus*), (ii) Colonial – Nostocales (*Microcystis, Radiocystis*), and (iii) Filamentous – Nostocales (*Anabaena, Aphanizomenon*) and *Pseudanabaena* (Order – Synechococcales). Notably, the genus *Pseudanabaena* initially was classified as Oscillatoriales based on the filamentous morphology but recent genomic sequencing and examination of ultrastructural characteristics has resulted in reclassification into the order Synechococcales (Vidal et al., 2021; Komárek et al., 2014). For the purpose of this study, *Pseudanabaena* classified sequence variants were included with other filamentous taxa from the order Nostocales. To further explore seasonal trends in cyanobacterial populations, the taxonomic composition of cyanobacteria communities was assessed.

4.3.2.1 Unicellular Taxa

Sequences classified to the order Synechococcales (84.42 – 99.15%) consistently dominated cyanobacterial communities across sampling depths (Figure 4.5A; Table 4.2). Specifically, the majority of these sequences were classified to the family Synechococcaceae which includes unicellular picocyanobacterial genera such as *Synechococcus* and *Cyanobium*. While sequence variants classified to the family Synechococceae were revealed to dominate the cyanobacterial community, examination of the ASV composition revealed the dominance of individual Synechococcaceae classified ASVs contributing up to 56.87% of the cyanobacterial community.

The spatial distribution varied across ASVs with some being found exclusively at one depth, but others found distributed throughout the water column (Figure 4.5B). The ASV abundances across depths in Wishart Lake were more homogeneous than observed in Little Turkey and Big Turkey Lake further supporting the impact of water column stratification on population distribution. For example, in Wishart Lake in July and August 2018, ASV848 was found in similar abundances across sampling depths. However, in Little Turkey Lake, ASV848 was detected across sampling depths but was found at varying abundance with highest values detected below Secchi depth, demonstrating the impact of thermal stratification on distribution of cyanobacterial taxa. Further specialized spatial distribution of ASVs was observed in the stratified water column of Big Turkey Lake. For example, in July 2018, ASV844 was found at low abundances at the surface in Big Turkey Lake but was found at higher abundances at Secchi depth and below Secchi. However, in August 2018, the spatial distribution of ASV844 was opposite to that observed in July 2018 with higher abundances detected at the surface indicating that spatial distribution is further impacted by seasonality in addition to thermal stratification.

Cyanobacterial communities across the seasonal sampling period were also dominated by sequences classified as Synechococcales (69.04 – 99.78%; Figure 4.6A Table 4.1) and specifically the family, Synechococcaceae (25.58 – 100%). The reappearance of cyanobacteria in the bacterial community observed in Big Turkey Lake in May 2019 was exclusively due to ASVs classified as Synechococcaceae. Certain Synechococcaceae ASVs were found to be seasonally restricted but others were ubiquitous across seasons. For example, ASV867 was only detected during October and May in Big Turkey Lake indicating the presence of unique populations occupying environmental conditions with lower water temperatures. This phenomenon was further seen with ASV 866 which was detected at higher abundances in all three lakes in October, and in Wishart and Big Turkey Lake in May. ASV859 was also detected in higher abundances in this period in Wishart and Little Turkey Lake. Alternatively, other ASVs, such as ASV846, were present across seasonal periods but experienced fluctuations with highest abundances detected during the summer months. This is further exhibited with ASV848 in Wishart Lake, which was present in high abundances across all ice-free sampling months demonstrating the diversity present within picocyanobacterial taxa and their ecological niches.

4.3.2.2 Colonial & Filamentous Taxa

Sequences classified to the order Nostocales were detected at very low abundances seasonally (<1 – 1.79%; Table 1; Figure 4.6D) and across the depth profile (<1 – 2.31%; Table 4.2; Figure 4.5D). Across the depth profile, ASVs classified to gas vacuolate Nostocales taxa including ASV 822, 820 and 815 were consistently found at the surface due to positive buoyancy and the sampling conducted during the morning period at this site in support of the diurnal trends discussed in Chapter 3. The changes in ASV composition observed between

studies demonstrates the importance of performing frequent sampling to encompass the variability and fluxes in community composition and the need to include multiple sampling points to account for potential heterogeneity in distribution.

Sequences classified to the order Chroococcales (2.43 – 30.76%) and specifically the family Microcystaceae (2.11 – 16.04%) were detected at lower abundances within the cyanobacterial community across sampling depths (Figure 4.5C; Table 4.2). Higher abundances of ASVs classified to the family of Microcystaceae were detected at the surface in Little Turkey Lake in August of 2018 which may be attributed to positive buoyancy due to the potential presence of gas vacuoles. The distribution of gas vacuolate taxa has also been demonstrated to be dependent on sampling time in these systems (Chapter 3). Similar to trends observed between stratified and non-stratified systems, the distribution of ASV919 (classified as *Radiocystis*) was observed to be homogeneously distributed in Wishart lake in July 2018 comparative to Big Turkey and Little Turkey Lake. ASV919 was observed to have spatial distribution with peaks in abundance occurring at Secchi depth or at the surface depending on the system and sampling month. The disparity in distribution between lakes and sampling months demonstrates the combined interaction of lake characteristic and seasonality on water column distribution of populations.

Seasonally, sequences classified to the order Chroococcales (1.00 – 29.14%) and specifically the family Microcystaceae (1.11 – 21.60%) were found in higher percentages in Big Turkey and Little Turkey Lake during summer sampling periods (Figure 4.6C; Table 4.1). Populations of Microcystaceae classified ASVs were detected into late October in Big Turkey Lake indicating the potential for bloom forming taxa growth to occur into the late fall. The relative abundance of Microcystaceae in the cyanobacterial community varied between

sampling years. Wishart Lake did not have a consistently detected population of Chroococcales across the seasonal sampling period. Specifically, ASV919 (classified as *Radiocystis*) was observed to show distinct seasonal trends in Big Turkey and Little Turkey Lake but varied between systems. In Little Turkey Lake, ASV919 was observed to reach maximal abundances in July of both sampling years and were not detected in the fall and spring. However, in Big Turkey Lake, peak abundances were detected later in the season in August with a population that persisted into the fall but was not detected during the spring. In addition to the increase in abundances observed seasonally, interannual variation was observed in the abundances of Microcystaceae. In Big Turkey Lake, relative abundance varied in July (10.69%) and August (21.60%) of 2019 versus July (1.11%) and August (10.96%) of 2018 demonstrating the potential for interannual variation in cyanobacterial community structure, which may be influenced by nutrient availability, water turbidity and lake morphometry.

Table 4.3 Taxonomic classification of potentially toxic bloom-forming amplicon sequence variants.

ASV ID	Taxonomic Classification
ASV912	*Microcystis*
ASV913	*Microcystis*
ASV914	*Microcystis*
ASV919	*Radiocystis*
ASV920	*Radiocystis*
ASV837	Synechococcaceae
ASV838	Synechococcaceae
ASV839	Synechococcaceae
ASV840	Synechococcaceae
ASV841	Synechococcaceae
ASV842	Synechococcaceae
ASV843	Synechococcaceae
ASV844	Synechococcaceae
ASV845	Synechococcaceae
ASV846	Synechococcaceae
ASV847	Synechococcaceae
ASV848	Synechococcaceae
ASV849	Synechococcaceae
ASV850	Synechococcaceae
ASV851	Synechococcaceae
ASV853	Synechococcaceae
ASV855	Synechococcaceae
ASV857	Synechococcaceae
ASV858	Synechococcaceae
ASV859	Synechococcaceae
ASV860	Synechococcaceae

Table 4.3 Continued

ASV ID	Taxonomic Classification
ASV862	Synechococcaceae
ASV864	Synechococcaceae
ASV865	Synechococcaceae
ASV866	Synechococcaceae
ASV867	Synechococcaceae
ASV868	Synechococcaceae
ASV869	Synechococcaceae
ASV806	*Pseudanabaena*
ASV809	*Pseudanabaena*
ASV808	*Pseudanabaena*
ASV810	*Pseudanabaena*
ASV815	Nostocales
ASV819	Nostocales
ASV820	Nostocales
ASV821	Nostocales
ASV822	Nostocales
ASV823	Nostocales
ASV824	Nostocales

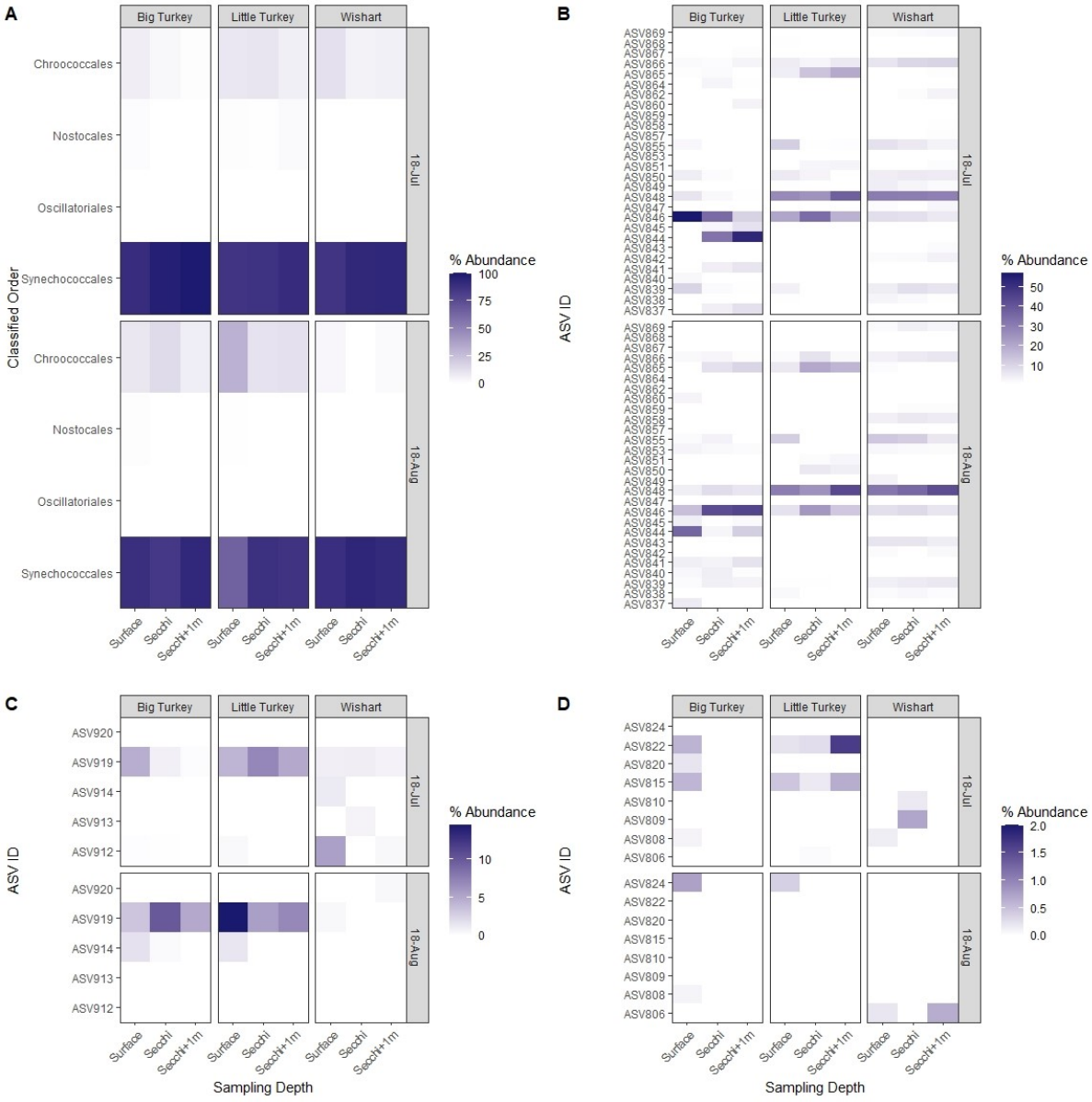


Figure 4.5 Heatmap depicting the relative abundances of (A) cyanobacterial community composition at the order level and (B, C, D) individual amplicon sequence variants of interest across a depth profile in the summer of 2018 in a non-stratified lake (Wishart), a mid-sized stratified lake (Little Turkey) and deep stratified lake (Big Turkey). Water column distribution varied between lakes depending on stratification and depth. To further explore the spatial distribution of individual taxa, amplicon sequence variants (ASV) of the V4 region of the 16S rRNA gene classified to common toxic and bloom forming genera. To further explore the spatial distribution of individual taxa, amplicon sequence variants (ASV) of the V4 region of the 16S rRNA gene classified to common toxic and bloom forming genera including Synechococcaceae (*Synechococcus, Cyanobium*),

Nostocales (*Anabaena*, *Aphanizomenon*), *Pseudanabaena* and Microcystaceae (*Microcystis*, *Radiocystis*). Relative abundances of individual ASVs in cyanobacterial community composition for (B) unicellular taxa (Synechococcaceae), (C) colonial taxa (Microcystaceae) and (D) filamentous taxa (Nostocales & *Pseudanabaena*).

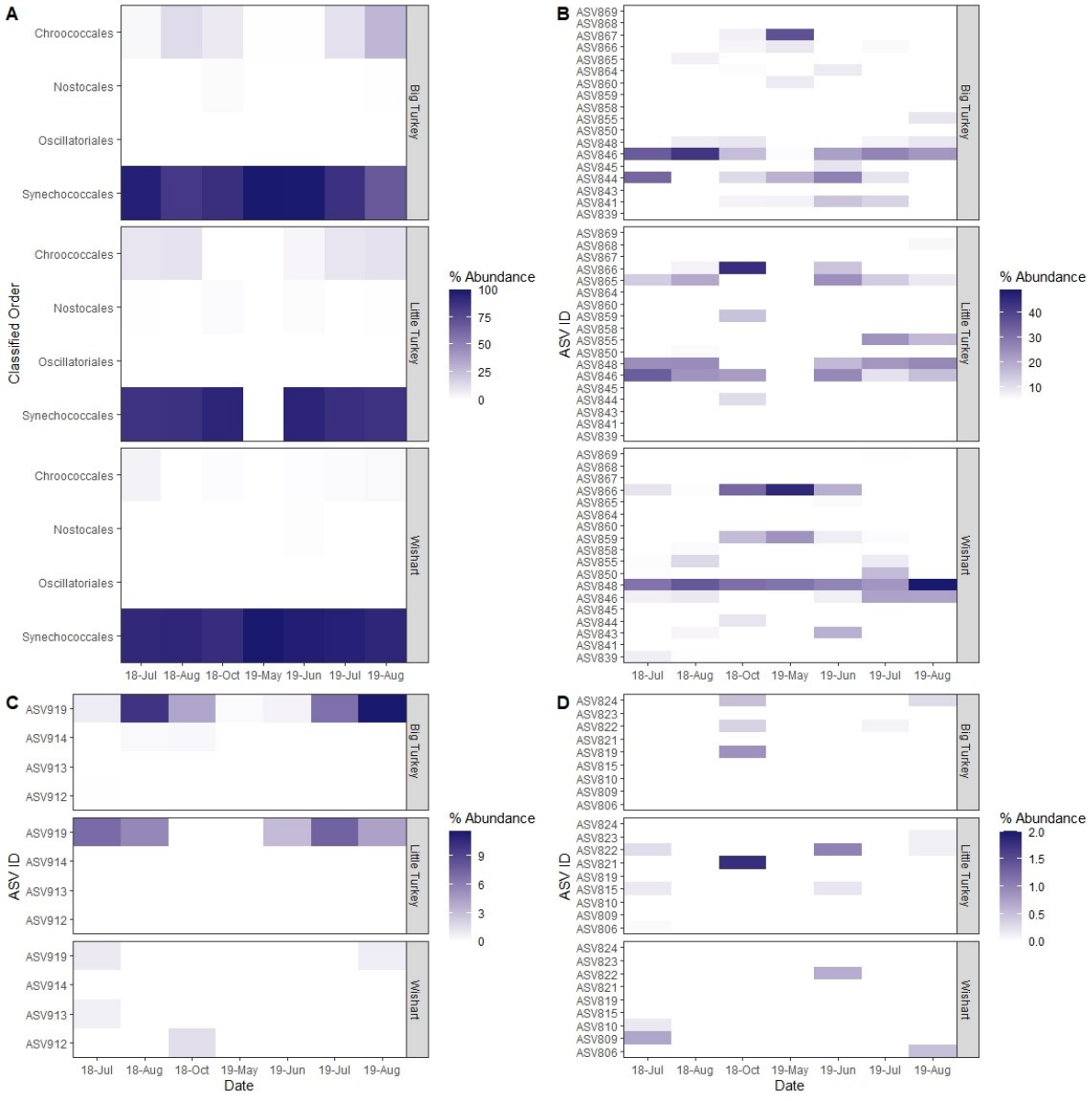


Figure 4.6 Heatmap depicting the relative abundances of (A) cyanobacterial community composition at the order level and (BC, D) individual amplicon sequence variants of interest across a seasonal ice-free period in a non-stratified lake (Wishart), a mid-sized stratified lake (Little Turkey) and deep stratified lake (Big Turkey). Cyanobacterial communities were

consistently dominated by the order Synechococcales which includes picocyanobacterial genera. To further explore the seasonal distribution of individual taxa, amplicon sequence variants (ASV) of the V4 region of the 16S rRNA gene classified to common toxic and bloom forming genera including Synechococcaceae (*Synechococcus, Cyanobium*), Nostocales (*Anabaena*, *Aphanizomenon*), *Pseudanabaena* and Microcystaceae (*Microcystis*, *Radiocystis*). Relative abundances of individual ASVs in cyanobacterial community composition for (B) unicellular taxa (Synechococcaceae), (C) colonial taxa (Microcystaceae) and (D) filamentous taxa (Nostocales & *Pseudanabaena*).

4.4 Discussion, Conclusions & Implications

Consistent with global trends (Carey et al., 2008; Wells et al., 2020) the number of algal blooms reported in Ontario, Canada has been significantly increasing, especially in lakes on the Canadian Shield where these increases are predominantly comprised of potentially toxin-producing cyanobacteria (Winter et al., 2011). The multi-year bacterial community analysis conducted across three lakes in the TLW and reported herein aligns with those observations. It demonstrated that contrary to lingering beliefs regarding winter limnology that often dismiss winter periods (especially ice cover) as ecologically unimportant relative to the summer "growing season" (Powers and Hampton, 2016), evaluation of broader seasonal variation in lake microbial, and especially—but not exclusively—cyanobacterial communities, can provide critical insights regarding climate change impacts on oligotrophic, northern temperate lake ecosystems and the associated implications to human and environmental health. This study provided six important observations:

(1) cyanobacteria persisted year-round in the oligotrophic, northern temperate lakes of the TLW,

(2) cyanobacterial communities during ice-covered months included sequences classified to the recently identified non-photosynthetic, potentially toxic basal lineage, Melainabacteria,

(3) cyanobacteria comprised a significant portion of the bacterial communities in the study lakes as early as May and persisted into late October,

(4) picocyanobacteria were especially dominant during ice-free periods,

(5) picocyanobacterial populations shifted seasonally—while certain sequences were dominant during ice-free months, other sequences were restricted to either (i) the shoulder seasons of the vernal window (i.e., spring and fall) or (ii) only during periods of winter ice cover, and

(6) lakes with lower depth ratios and longer water renewal times (i.e., Big Turkey Lake) had higher relative abundances of cyanobacteria.

The presence and persistence of potentially toxic picocyanobacteria within the lakes of the temperate forest biome of Canada have not been previously reported. In the lakes of the TLW, picocyanobacterial taxa dominated the cyanobacterial communities seasonally (Figure 4.6) and spatially (Figure 4.5). Although picocyanobacteria are abundant in diverse freshwater and marine environments (Cai et al., 2010; Chen et al., 2006; Collos et al., 2009; Felföldi et al., 2016; Gin et al., 2021) sufficient understanding of the occurrence and characterization of their blooms, toxicity, and allelopathic activity is lacking (Śliwińska-Wilczewska et al., 2018), especially in freshwater environments. Freshwater assemblages of picocyanobacteria are complex and dynamic because they have developed many evolutionary mechanistic adaptations (e.g., small size, ability to grow in low light intensity environments, rapid nutrient uptake, ability to maintain water column position, etc.) and interactions (e.g., allelopathy) with larger primary producers and predators, which enable their exploitation of environmental variability (Callieri, 2008; Śliwińska-Wilczewska et al., 2018). Abiotic and biotic factors such as lake morphometry, thermal regime, and trophic state influence picocyanobacterial

dynamics—it has been suggested they may be at least as important as nutrients in affecting the structure of freshwater picocyanobaterial communities (Callieri, 2007).

These organisms are common components of the photic zone but distribution and abundance can range widely depending on the system conditions (Callieri and Stockner, 2002). Previous reports including a survey of 43 lakes and ponds indicated that picocyanobacteria prefer large, deep lakes with high hydrologic retention times and incomplete mixing due to vertical density differences (Callieri, 2008; Camacho et al., 2003). This was also exhibited in the lakes of the TLW with higher relative abundances of cyanobacteria detected in the stratified lakes, Big Turkey and Little Turkey Lake comparative to the shallow, well-mixed lake, Wishart Lake (Figure 4.1; Table 4.1). Although previous research has suggested that picocyanobacteria reach peak abundances prior to the onset of thermal stratification (Callieri and Stockner, 2000; Fahnenstiel et al., 1991; Li et al., 2020), it was observed that Synechococcaceae temporally dominated the communities independent of stratification in the lakes of the TLW suggesting other environmental factors are influencing cyanobacterial community structure.

Thermal stratification may impact the spatial distribution of cyanobacterial populations (Pick and Agbeti, 1991). Stratified lakes were expected to show more consistent ASV composition across depths located within the same thermal layer. However, sampling conducted at multiple depths within the metalimnion of Big Turkey and Little Turkey Lake identified the heterogeneity in abundance of picocyanobacterial ASVs corresponding to the highly variable vertical distribution which has been reported previously in other stratified lakes (Hall and Vincent, 1994; Stockner et al., 2006). Within larger lakes, including Lake Huron and Michigan, peak abundances of picocyanobacteria have been detected in the lower metalimnion

and upper hypolimnion (Stockner et al, 2006) with higher abundances occurring under low light intensity (Jakubowska and Szeląg-Wasielewska, 2015; Pick and Agbeti, 1991). The highest proportion of cyanobacterial sequences within the bacterial community was detected in Big Turkey Lake 1 meter below Secchi depth (Figure 4.2; Table 4.2) which supports the previously reported occurrence in the lower metalimnion with low light intensity (Stockner et al., 2006).

In these environments, a spring or early summer peak and second autumnal peak have been observed (Callieri, 2008; Stockner et al., 2002). In temperate freshwater and marine environments, picocyanobacteria are typically more abundant in the warm season than in the cold season, during which cell density decreases of approximately three orders of magnitude (Postius and Ernst, 1999; Waterbury and Valois, 1993) and shifts to completely different populations (Cai et al., 2010) have been reported. Similar bimodal patterns and shifts between summer and winter relative abundance of picocyanobacteria were observed in the TLW, as reported herein. These seasonal shifts in populations were observed in the ASV composition of cyanobacterial communities (Figure 4.5; Figure 4.6) due to specific populations or subclades being more adapted to lower temperatures (Cai et al., 2010) resulting in the non-ubiquitous occurrence of ASVs. Seasonal trends were also observed in the relative abundances of the cyanobacterial order, Chroococcales, which includes potentially toxic bloom-forming genera including *Microcystis*. While picocyanobacterial sequences were consistently high during the ice-free sampling months, Chroococcales-classified sequences were found in higher abundances in the late summer in Big Turkey and Little Turkey Lake demonstrating seasonal shifts between different cyanobacterial taxa that occupy different ecological niches. However,

the absence or low abundances of Chroococcales in Wishart Lake in these months indicates more complex processes shaping the seasonal shifts in cyanobacterial community composition.

The significant representation of picocyanobacteria in the bacterial community of Big Turkey Lake shortly after the spring snowmelt in May 2019 is especially notable. The warming climate has led to increased temperatures in the TLW (Creed et al., 2014) which is frequently associated with increased cyanobacterial occurrence due to elevated growth rates at warmer temperatures and prolonged thermal stratification. However, due to the complexity of aquatic ecosystems, this single factor cannot be attributed to the significant spring proliferation of cyanobacteria observed in Big Turkey Lake. The relatively greater availability of glacial till surrounding the lower elevation lakes of the watershed, the abundance of macrophytes at the margins of all lakes and the specific lake morphometry of Big Turkey Lake as described by Jeffries et al. (1988) may explain the significant spring proliferation (Genkai-Kato and Carpenter, 2005; Carpenter, 1983). The depth ratio (i.e., the ratio of mean to maximum lake depth) is substantially lower in Big Turkey lake than in the other study lakes. In lakes such as Big Turkey Lake (Figure A1; Table 1.1), in which the thermocline is shallower than approximately one to two times the mean depth, the epilimnion's sediment surface area to volume ratio declines with depth ratio. The potential nutrient recycling from the sediment surface, productivity, and sediment accretion rates are expected to increase as depth ratio decreases (Carpenter, 1983), suggesting higher nutrient recycling present in Big Turkey Lake comparatively to the other lakes in the watershed. As well, more sediment can be eroded during runoff from the Big Turkey Lake watershed during precipitation events or snowmelt periods because there is more available sediment on the surrounding landscape, relative to Wishart Lake (Jeffries et al., 1988). The delivery of available sediment to oligotrophic lakes such as

Big Turkey Lake may also result in the release of phosphorus to the water column (Withers and Jarvie, 2008; Froelich, 1989), and contribute to the spring proliferation of picocyanobacteria (Passoni and Callieri, 2000) due to their efficient nutrient utilization (Śliwińska-Wilczewska et al., 2018). Additionally, the abundance of macrophytes in the lakes of the TLW (Jeffries et al., 1988) may further modulate or even limit phosphorus recycling from sediments (Genkai-Kato and Carpenter, 2005; Scheffer, 1998) and availability to broader microbial communities. Changes in the biotic or abiotic factors that would alter this complex balance at the watershed-scale cannot be described at present and warrant further investigation.

Melainabacteria and Sericytochromatia have previously been found in aphotic environments (Monchamp et al., 2019; Soo et al., 2014). Although the biogeography and ecology of those organisms remains understudied, genomes have previously been isolated from varying aquatic sources including lakes (Monchamp et al., 2019) and engineered aquatic systems including water treatment facilities and water distribution systems (Ling et al., 2018; Zamyadi et al., 2019). The presence of Melainabacteria and Sericytochromatia sequences when lakes were covered with up to 0.81 m of ice and additional snow coverage limiting light availability aligns with the aphotic environments these lineages were previously detected in. The detection of this non-phototrophic basal lineage of Cyanobacteria during periods of ice cover suggests the potential role of these organisms in winter microbial community processes warranting further investigation into these lineages. While the ecological function and role of these organisms is not well understood (Montchamp et al., 2019), synthesis of the neurotoxin, β-N-methylamino-L-alanine (BMAA), which is associated with neurodegenerative diseases such as Parkinson's and Alzheimer's (Cervantes Cianca et al., 2012) has been hypothesized in this group of organisms (Nunes-Costa et al., 2020). The potential for toxin production in this

understudied group of organisms that have been demonstrated to be present in natural aquatic systems, such as those in this study, and engineered aquatic systems (Ling et al., 2018; Zamyadi et al., 2019) necessitates further studies on the distribution and function of these organisms to identify novel threats to human health in drinking water sources.

The presence and dominance of picocyanobacteria in the TLW is notable due to the potential production of several metabolites with significant human and environmental health concern or aesthetic significance. They may produce microcystin and nodularin, which are both hepatotoxins (Chorus et al., 2000; Jakubowska and Szeląg-Wasielewska, 2015; Vareli et al., 2013). Several species of picocyanobacteria may also potentially synthesize BMAA (Cervantes Cianca et al., 2012; Cox et al., 2003). Recent studies on tropical taxa have also noted the potential for cylindrospermopsin and anatoxin production with toxicity is often underestimated due to limitations in sensitivity of assays (Gin et al., 2021). In addition to toxins, these organisms may also produce geosmin and MIB (Graham et al., 2008; Jakubowska and Szeląg-Wasielewska, 2015; Watson, 2003), which are taste and odor forming compounds that commonly result in customer complaints when present in drinking water (McGuire, 1995; Suffet et al., 1996). The abundance of potentially toxic picocyanobacteria warrants further study due to the human and environmental health risks that these organisms impose. While it is known that they contribute to a significant fraction of the total primary productivity in freshwater and marine environments (Stockner et al., 2002; Waterbury et al., 1986), the biotic and abiotic factors that drive their proliferation and potential toxin production are not well understood (Sliwijnska-Wilczewska et al., 2018). This critical gap in knowledge thereby precludes climate change adaptation for communities whose drinking water supplies, livelihoods, recreation, and spirituality may be impacted by changes in these ecosystems.

Given the human and environmental health implications of these compounds and the significant costs associated with their removal from drinking water (Emelko et al., 2011), it is critical to explore the possibility of picocyanobacteria-associated toxin production in freshwaters, which may be undetected because of reliance on visual observation of accumulated cyanobacterial biomass and traditional foci on monitoring of colonial and filamentous bloom forming cyanobacteria (Chorus et al., 2000; Newcombe, 2009). Accordingly, broader and more comprehensive monitoring is required (Chapter 3; Pobel et al., 2011; Welker et al., 2021) to advance understanding of picocyanobacterial dynamics in response to local biotic and abiotic drivers, some of which are impacted by changing climate (Callieri, 2008; Drakare and Liess, 2010).

While climate change is not being proposed herein as a driver of the persistence and dominance of picocyanobacteria in the TLW, it emphasizes the pressing need to better understand the potential roles that picocyanobacteria and non-photosynthetic Melainabacteria (such as those associated with sequences that were observed in the TLW) may play in toxin production and trophic status modulation, especially in oligotrophic lakes and reservoirs that are relied upon for the provision of drinking water. Climate warming is resulting in earlier spring snowmelt discharges and extending the vernal window by delaying the onset in timing and magnitude of autumnal storms within the temperate forest biome of Canada (Creed et al., 2015). It is generally understood that algal blooms tend to occur at the height of summer and in early fall. Recently, it has been reasonably suggested that significant shifts in algal bloom initiation and persistence to *later* in the autumn season (from September to November) in oligotrophic, northern temperate lakes (Winter et al., 2011) may be attributed to changes in nutrient loading resulting from autumn storms shifting to the post-canopy leaf fall period in

these biomes (Creed et al., 2015). In contrast, this complementary works suggests that the convergence of key biotic factors—climate forcing of hydrological and biogeochemical processes, and intrinsic landscape features such as lake morphometry—may create conditions that lead to *early* seasonal increases in the relative abundance of potentially toxic cyanobacteria (i.e., picocyanobacteria) within the temperate forest biome of Canada. While the occurrence of blooms of potentially toxic cyanobacteria later in the fall may be one concerning implication of the extension of the autumnal window in northern temperate lakes, it may be possible that the earlier opening of the vernal window as a result of climate warming may promote these concerns earlier in the spring and/or exacerbate them later in the fall.

Chapter 5
Research Summary, Implications & Recommendations

5.1 Research Summary

Cyanobacteria and their associated impacts on water quality and treatability are frequently associated with eutrophic systems dominated by visible blooms occurring at the water surface (Paerl et al., 2016). However, the dynamic nature of cyanobacterial taxa may allow for these organisms to dominate bacterial and phytoplankton communities despite the absence of visible biomass. Although nutrient loading, with focus on phosphorus (Schindler, 1977), is frequently purveyed as the main source of concern for the dominance of these organisms (Heisler et al., 2008; Paterson et al., 2017), lakes that have experienced phosphorus limitation still experience bloom events (Paerl et al., 2016; Paterson et al., 2017) with community composition experiencing changes in response to nutrient availability (Andersson et al., 2015). The increased occurrence of cyanobacterial blooms in oligotrophic lakes in Ontario (Winter et al., 2011) warrants further investigation into the community dynamics present in low-nutrient systems. Specifically, the research conducted in this thesis demonstrated the potential applications of amplicon sequencing for characterizing the spatiotemporal variation present within cyanobacterial communities in oligotrophic lakes where visible biomass is absent.

Although amplicon sequencing provides a rapid and sensitive technique for analyzing environmental samples, this technique does not come without challenges of its own. One of the main challenges arises prior to conducting downstream bioinformatic and statistical analyses including diversity and differential abundance analyses due to the cruciality of library size normalization to account for size bias introduced through differing total read counts

between samples. Previous research has criticized rarefaction (McMurdie and Holmes, 2013), the process of subsampling to a normalized library size, due to data being omitted. Notably, the research conducted in Chapter 2 demonstrated an appropriate way of rarefying libraries for applications in diversity analyses. While rarefying samples in a single iteration results in the loss of data, conducting repeated iterations allows for the characterization of the variability introduced through subsampling. This research also evaluated the impact of normalized library size and subsampling style on diversity analyses results demonstrating the potential impacts of parameter selection on analysis outputs. The demonstration of the utility of repeatedly rarefying in diversity analyses allowed for it to be applied within the remainder of the research conducted in this thesis to explore trends in cyanobacterial community diversity in the lakes of Turkey Lakes Watershed (TLW).

Within the TLW, cyanobacteria were previously shown to dominate the phytoplankton communities in 1980 (Jeffries et al., 1988). The use of amplicon sequencing allowed for identification of the genetic diversity present within the lakes. Although cyanobacteria have been frequently reported at the surface and in the late summer in temperate regions (Chorus et al., 2000; Graham et al., 2008), the communities within these lakes are highly dynamic exhibiting both spatial and temporal variation. The research conducted in this thesis served to identify the spatiotemporal variation in cyanobacterial communities across sampling regimes to examine (i) short-term diurnal variation, (ii) long-term seasonal variation and (iii) spatial variation within the column. Although distinct spatiotemporal sampling regimes were developed, across all the studies in this thesis, picocyanobacterial classified sequences dominated the cyanobacterial communities. The competitive advantage of picocyanobacteria in oligotrophic systems previously shown supports this dominance (Sliwinska-Wilczewska et

al., 2018) and reveals the potential for high abundances of these organisms even in low nutrient systems.

Cyanobacteria are known to experience oscillatory diurnal migrations in response to light availability arising from buoyancy regulation mechanisms (Visser et al., 2005) which were further supported and characterized using amplicon sequencing in Chapter 3. Within Little Turkey Lake, a thermally stratified lake, and Wishart Lake, a non-stratified lake, distinct diurnal fluctuations in cyanobacterial abundances were observed. The well-mixed water column of Wishart Lake demonstrated the homogeneous distributions of cyanobacteria through non-significant fluctuations of abundances as a function of time. In contrast, Little Turkey Lake demonstrated increasing cyanobacterial abundances as the day progressed indicating the potential for diurnal variability and migrations present in stratified lakes. Further impacts of water column stability were observed after an external mixing event caused by heavy precipitation prior to sampling resulted in mixing in the stratified water column and redistribution of water surface communities to deeper depths in the water column. While water column stability and stratification were observed to impact the trends in diurnal variability, these system-specific responses are further driven by taxonomic composition of communities. Although gas vacuolate taxa are well characterized for buoyancy regulation and diurnal water column migration (Staley, 1980; Walsby, 1981), different diurnal variability was observed between sequences classified as gas vacuolate taxa, showing the inherent complexity present in these systems driven by both physicochemical and ecological characteristics.

Further demonstrating the impact of water column stability on cyanobacterial distribution, distinct spatial trends were observed between Big Turkey, Little Turkey, and Wishart Lake in Chapter 4. Similar to the diurnal study in Chapter 3, Wishart Lake continued

to show homogeneous distributions of cyanobacterial abundances across the shallow, well-mixed water column. In contrast, Big Turkey Lake exhibited higher cyanobacterial relative abundances at deeper depths, likely driven by the dominance of picocyanobacterial taxa that have previously been detected in higher abundances in the lower metalimnion and upper hypolimnion (Pick and Agbeti, 1991; Stockner et al., 2006). Large-scale spatial trends in the relative abundances were detected between lake sites demonstrating the unique community dynamics present within each site despite being in close proximity and being hydrologically interconnected. Wishart Lake consistently showed lower cyanobacterial relative abundances than Big Turkey and Little Turkey Lake, contrasting previous research suggesting higher algal biomass and productivity in shallow lakes (Staehr et al., 2012). The results from the lakes of the TLW contrasts previous lake morphometric studies (Staehr et al., 2012) highlighting the complexity of interactions that shape cyanobacterial growth and communities in aquatic ecosystems. Greater availability of glacial till, abundances of macrophytes and lake morphometry may contribute to cyanobacterial community structure. However, in northern temperate regions, these factors may be impacted by warming climates.

In addition to the spatial and diurnal variability exhibited in these lakes, cyanobacterial communities showed seasonal trends in abundances as discussed in Chapter 4. Cyanobacterial sequences were detected at >1% relative abundance as early as May in Big Turkey Lake indicating the early re-emergence of cyanobacteria within the bacterial community potentially due to climate warming in northern temperate regions. The peak relative abundance was reached by June suggesting that the generalization of the late summer occurrence of cyanobacteria is not universally applicable to all systems (Chorus et al., 2000; Graham et al., 2008). Cyanobacterial sequences were detected at extremely low relative abundances (<1%)

and inconsistently during the ice-cover months, paralleling the low cellular concentrations observed in 1980 ice-cover months (Webster et al., 2021). However, the presence during the ice-cover periods demonstrated the consistent occurrence of these organisms throughout the year albeit at significantly lower levels. Additionally, under periods of ice-cover, cyanobacterial communities were regularly comprised of sequences classified as Melainabacteria. The detection of these organisms during the winter, in temperate oligotrophic lakes furthers our understanding on the potential distribution and ecological niches of these organisms.

The spring cyanobacterial communities were exclusively due to picocyanobacterial classified sequences corresponding to previous studies showing *Cyanobium* blooms occurring in the spring (Callieri and Stockner, 2000; Li et al., 2020). Additionally, some of these sequences showed seasonal occurrence supporting that some clades and taxa may be more well adapted to lower temperatures than others, as also previously observed within larger systems such as the Chesapeake Bay (Cai et al., 2010). Although Synechococcaceae-classified sequences belonging to picocyanobacterial taxa consistently dominated the cyanobacterial communities, Microcystaceae-classified sequences were also detected. Specifically, the Microcystaceae sequences were detected in the summer (July and August) in higher abundances in Big Turkey and Little Turkey Lake, and into October for Big Turkey Lake demonstrating the seasonal succession of cyanobacterial taxa within the community.

5.2 Implications

As amplicon sequencing becomes an increasingly available analysis tool for application in interdisciplinary fields, appropriate data handling is critical to maintain the integrity of downstream analyses interpretation including diversity analyses. For example, if only a single

iteration of rarefying is conducted, there is no way to tell whether that single normalized library is representative of data or has been impacted by high levels of variability. Although subsampling with or without replacement did not significantly impact the results, selection of smaller than necessary library sizes will introduce excess variability through data exclusion impacting interpretation of results as examined in Chapter 2. This excess variability has the potential to lose the ability to differentiate between samples as the spread of data increases as library sizes are decreased. It is critical for amplicon sequencing studies to understand the limitations and potential implications on parameter selection on the analysis and consequently the data interpretation to prevent mishandling and misinterpretation. In this study, the use of amplicon sequencing allowed for the characterization of the genetic diversity present within the picocyanobacterial community which would have been immensely difficult if relying on morphological differentiation.

Amplicon sequencing allowed for characterization of the genetic diversity present within picocyanobacterial taxa. Previous phytoplankton community surveys conducted in 1980 in the lakes of Turkey Lakes Watershed did not show high abundances of picocyanobacterial taxa. However, the sample collection and organism identification may have significantly impacted the differentiation in cyanobacterial community structure. It is likely that a plankton net was used; importantly, smaller sized taxa can pass through these nets and thus may have been excluded from analyses of cyanobacterial abundance and diversity (Ehrlich, 2010). Additionally, microscopic identification of picocyanobacterial taxa is frequently challenging due to the lack of distinct morphological characteristics (Jakubowska and Szelag-Wasielewska, 2015). Underestimation of these organisms due to sampling bias or limitations in identification methods can impose risks to water quality due to the potential toxicity

demonstrated in some taxa (Sliwinska-Wilczewska et al., 2018). The identification of high abundances of picocyanobacterial sequences with amplicon sequencing from water samples collected with a peristaltic pump demonstrated the importance of sample collection and identification techniques to ensure accuracy.

The dominance of potentially toxic picocyanobacterial taxa within oligotrophic lakes challenges the traditional approaches in monitoring. For example, the reliance on visual observation of cyanobacterial biomass for increasing sampling efforts does not reflect the highly abundant non-visible, potentially toxic picocyanobacterial populations present at greater depths in the water column. While some cyanobacterial taxa form dense blooms at the surface, the diversity present in size and cellular morphology within this group of organisms cannot be characterized through direct visual observation of water surface independently. Reliance on visual observation or restricting sampling to the surface will result in underestimation or missed detection of cyanobacterial populations due to the potential for high abundances of picocyanobacterial populations located at deeper depth within the water column as demonstrated in this study and others (Pick and Agbeti, 1991; Stockner et al., 2006). In this research, potentially toxic-bloom-forming-taxa-classified sequences (e.g., Microcystaceae) were frequently found in higher abundances at the surface supporting common generalizations of cyanobacteria being found at the water surface. Nonetheless, it is critical to expand the views on cyanobacteria to encompass the spatiotemporal diversity present within this group of organisms (Freeman et al., 2020). Exclusive focus on common bloom-forming taxa, such as *Microcystis*, as the basis for the characteristics of all Cyanobacteria will continue to ignore the potential impacts to water quality associated with picocyanobacterial taxa and underestimate population sizes.

Diurnal variability can further create challenges in the development of monitoring protocols. Diurnal migrations of cyanobacteria can substantially impact the detection, especially if using discrete depth sampling. Despite the potential impact of sampling time on accurate detection and characterization, sampling time is frequently overlooked in monitoring protocol guidelines. For example, the absence of specified sampling windows or very wide windows (e.g., 10AM – 3PM) in guidelines can result in biased or non-representative community characterization. Large time frames do not avoid the diurnal migrations with the potential for abundances detected at 10AM to vary significantly from the abundances detected at 3PM as demonstrated in the research conducted in Chapter 3. Similarly limiting sample collection to one depth of the water column (e.g., 50 cm below the surface), will not reflect both spatial and diurnal variability in the distribution of cyanobacterial populations further contributing to biased or non-representative community characterization. As presented in Chapter 4, lakes with stratified water columns have different cyanobacteria community composition at varying depths of the water column. In conjunction, these results show the inherent complexities present in cyanobacterial communities which must be accounted for with monitoring efforts.

While water column stability has previously been demonstrated to impact cyanobacterial distribution, external mixing events caused by inclement weather are frequently overlooked. A mixing event caused by heavy rainfall prior to sample collection in the diurnal study in Chapter 3 demonstrated the impact of weather induced mixing, which resulted in redistribution of water surface communities to deeper depths of the water column. If sampling is conducted as per usual following inclement weather events, there is the potential for underestimation of cyanobacterial abundances due to disturbed distribution within the water

column. With increasing extreme precipitation arising as a result of climate change this will continue to impact lake ecosystems (Woolway et al., 2020) through potential increased nutrient loading from surface run off creating optimal environmental conditions for cyanobacterial growth. However, increased precipitation will also create challenges in accurate detection and quantification of cyanobacterial abundances due to weather induced mixing.

Manifestations of climate change including warming temperatures and altered precipitation patterns are observed in both terrestrial (Kharin et al., 2007) and aquatic ecosystems (Woolway et al., 2020). Warming water temperatures may result in earlier onset and prolonged periods of thermal stratification (Woolway et al., 2020). Due to this warming and alteration of the vernal period (Creed et al., 2015; Contosta et al., 2017), cyanobacteria may occur at higher abundances within the water column earlier in the season and may reach peak abundances earlier. Previous studies have demonstrated the persistence of cyanobacteria into the late fall within other oligotrophic lakes in Ontario (Winter et al., 2011) but this may depend on system dynamics. For the TLW, with drier growing seasons (Creed et al., 2015) and the onset of thermal stratification as early as the end of May, cyanobacteria were demonstrated in this research to be able to thrive in the bacterial community earlier in the season. Early resurgence of cyanobacteria in temperate watersheds may be associated with the massive influx of nutrients during the spring melt period (Creed et al., 2015; Lindsay et al., 2004) demonstrating the complex interaction of hydrologic regimes, landscape processes, lake morphometry and climate on cyanobacterial community dynamics. This early resurgence of cyanobacterial populations may impose critical risks to water quality and water security at an earlier than expected period.

5.3 Recommendations

I) Limitations of amplicon sequencing must be known in interdisciplinary studies but can be augmented with additional techniques.

Amplicon sequencing is an invaluable technique that can be applied in an interdisciplinary sector of water-related research areas with direct applications to water quality and management. However, to preserve the integrity of this technique, researchers must be equipped with appropriate background knowledge and versed in the benefits and limitations of the analysis to make appropriate selections for analysis. While no gold star standard currently exists, it is critical that amplicon sequencing studies fully evaluate the limitations of varying normalization strategies to be able to identify the potential implications in the subsequent data analysis and to allow researchers to make appropriate selections for their data. The awareness on limitations and potential implications of normalization techniques must continue to be discussed to ensure rigorous data interpretation by all researchers. Although amplicon sequencing of the taxonomic marker genes (e.g., 16S rRNA gene) allows for rapid characterization of cyanobacterial communities and the potential for toxicity could be obtained through sequencing of toxin genes (e.g., *mcy*E), this technique cannot provide absolute quantification (Gloor et al., 2017). For absolute quantification of populations, amplicon sequencing projects can be augmented with the inclusion of other quantitative molecular techniques including flow cytometry (Patel et al., 2019) or qPCR (Chiu et al., 2017).

II) The absence of visible biomass cannot be equated to absence of water quality concerns.

Due to the dominance of picocyanobacteria in systems, including those that are low nutrient as shown in Chapter 3 and 4, guidelines and protocols must be re-evaluated to ensure that they encompass the size variability observed in cyanobacterial communities. Some

guidelines (e.g., Ministry for the Environment and Ministry of Health, 2009) have shifted their protocols to measure biovolume instead of cellular concentration to account for the size variability within this group of organisms, demonstrating the ability to adjust protocols for detection and measurement of picocyanobacteria. Of greater concern is the reliance on visual detection of cyanobacterial biomass for increasing sampling efforts. The absence of visible biomass cannot be equated to absence of water quality concerns and guidelines must be updated to be inclusive of these often-overlooked picocyanobacteria due to the potential threats that they pose to water quality.

III) Monitoring must not be restricted to the water surface and requires knowledge on physicochemical and ecological characteristics of the study system.

Current monitoring protocols frequently employ the use of integrated depth sampling which provides a comprehensive view on the cyanobacterial population within the entire sampled water column. While this technique reduces logistics of sampling effort and processing time through the collection of a single sample, it is done at the cost of spatial distribution resolution. In systems where there is not a previously established characterization of the spatial distribution, when possible, discrete depth profiling sampling should be conducted to increase understanding on the system dynamics and spatial distribution of organisms. However, at minimum, sampling cannot be restricted to the surface due to the high relative abundances detected at lower depths in the water column as observed in Chapter 3 and 4. The fundamentals of sampling in monitoring protocols are dependent on system dynamics with understanding the system being vital to successful execution. Lakes are highly dynamic and varying trends in cyanobacterial abundances and community composition were observed in interconnected lakes located within the same watershed, further demonstrating the need for

system specific monitoring. To further aid in the development of ecologically meaningful protocols, cyanobacteria cannot continue to be viewed as a homogeneous entity with generalizations based on characteristics of common freshwater bloom-forming taxa such as *Microcystis* (Freeman et al., 2020). Continuing with this homogeneous view on cyanobacteria will severely neglect the diversity present within these organisms and further contribute to common misconceptions that interfere with water quality monitoring and management.

IV) Sampling must not be restricted to traditional mid-summer to early fall periods.

Further stemming from the homogeneous view on cyanobacteria, in temperate zones, cyanobacterial growth is generalized to occur from the mid-summer to early fall periods (Chorus et al., 2000; Graham et al., 2008). Restricting sampling to these periods has resulted in a lack in understanding and knowledge on winter cyanobacterial dynamics despite reported bloom events under ice (Wejnerowski et al., 2018) . It is critical for more studies and sampling to occur that are inclusive of ice-covered periods to identify the novel trends present in these systems. The prevalence of these organisms throughout the year warrants consistent monitoring. Although this may not be logistically possible due to constraints to time and resources, or challenges associated with winter sampling, the lack of winter limnological studies cannot continue. Absence of information on how cyanobacteria are distributed within systems due to ice cover severely impacts the ability to collect samples meaningfully. In addition to the need to advance knowledge on winter limnological processes, as a result of climate change, cyanobacterial blooms are being reported later into the fall (Winter et al., 2011). However, depending on the system, cyanobacteria may occur earlier in the season, as seen in Big Turkey Lake in Chapter 4, requiring monitoring to be implemented in earlier seasonal periods or risk failure of early detection of potentially toxic organisms.

References

Amir, A., Daniel, M., Navas-Molina, J., Kopylova, E., Morton, J., Xu, Z.Z., Eric, K., Thompson, L., Hyde, E., Gonzalez, A., Knight, R., 2017. Deblur rapidly resolves single-nucleotide community sequence patterns. mSystems 2:e00191, e00191-16.

Anantharaman, K., Brown, C.T., Hug, L.A., Sharon, I., Castelle, C.J., Probst, A.J., Thomas, B.C., Singh, A., Wilkins, M.J., Karaoz, U., Brodie, E.L., Williams, K.H., Hubbard, S.S., Banfield, J.F., 2016. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. Nat. Commun. 7, 1–11.

Andersen, G.S., Pedersen, M.F., Nielsen, S.L., 2013. Temperature acclimation and heat tolerance of photosynthesis in Norwegian Sachharina latissima (Laminariales, Phaeophyceae). J. Phycol. 700, 689–700.

Andersen, R.A., 1992. Diversity of eukaryotic algae. Biodivers. Conserv. 1, 267–292.

Anderson, 2009. Approaches to monitoring, control and management of harmful algal blooms (HABs). Ocean Coast Manag 52, 1–13.

Andersson, A., Höglander, H., Karlsson, C., Huseby, S., 2015. Key role of phosphorus and nitrogen in regulating cyanobacterial community composition in the northern Baltic Sea. Estuar. Coast. Shelf Sci. 164, 161–171.

Badri, M., Kurtz, Z., Muller, C., Bonneau, R., 2018. Normalization methods for microbial abundance data strongly affect correlation estimates. bioRxiv 406264.

Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G., Neufeld, J.D., 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. Appl. Environ. Microbiol. 77, 3846–3852.

Beall, B.F.N., Twiss, M.R., Smith, D.E., Oyserman, B.O., Rozmarynowycz, M.J., Binding, C.E., Bourbonniere, R.A., Bullerjahn, G.S., Palmer, M.E., Reavie, E.D., Waters, L.M.K., Woityra, L.W.C., McKay, R.M.L., 2016. Ice cover extent drives phytoplankton and bacterial community structure in a large north-temperate lake: implications for a warming climate. Environ. Microbiol. 18, 1704–1719.

Beecraft, L., Watson, S.B., Smith, R.E.H., 2019. Innate resistance of PSII efficiency to sunlight stress is not an advantage for cyanobacteria compared to eukaryotic phytoplankton. Aquat. Ecol. 53, 347–364.

Bertani, I., Steger, C.E., Obenour, D.R., Fahnenstiel, G.L., Bridgeman, T.B., Johengen, T.H., Sayers, M.J., Shuchman, R.A., Scavia, D., 2017. Tracking cyanobacteria blooms: Do different monitoring approaches tell the same story? Sci. Total Environ. 575, 294–308.

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

Bodor, A., Bounedjoum, N., Vincze, G.E., Erdeiné Kis, Á., Laczi, K., Bende, G., Szilágyi, Á., Kovács, T., Perei, K., Rákhely, G., 2020. Challenges of unculturable bacteria: environmental perspectives. Rev. Environ. Sci. Biotechnol. 19, 1–22.

Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Yu, D.W., de Bruyn, M., 2014. Environmental DNA for wildlife biology and biodiversity monitoring. Trends Ecol. Evol. 29, 358–367.

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., Cope, E.K., Da Silva, R., Diener, C., Dorrestein, P.C., Douglas, G.M., Durall, D.M., Duvallet, C., Edwardson, C.F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J.M., Gibbons, S.M., Gibson, D.L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G.A., Janssen, S., Jarmusch, A.K., Jiang, L., Kaehler, B.D., Kang, K. Bin, Keefe, C.R., Keim, P., Kelley, S.T., Knights, D., Koester, I., Kosciolek, T., Kreps, J., Langille, M.G.I., Lee, J., Ley, R., Liu, Y.X., Loftfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B.D., McDonald, D., McIver, L.J., Melnik, A. V., Metcalf, J.L., Morgan, S.C., Morton, J.T., Naimey, A.T., Navas-Molina, J.A., Nothias, L.F., Orchanian, S.B., Pearson, T., Peoples, S.L., Petras, D., Preuss, M.L., Pruesse, E., Rasmussen, L.B., Rivers, A., Robeson, M.S., Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S.J., Spear, J.R., Swafford, A.D., Thompson, L.R., Torres, P.J., Trinh, P., Tripathi, A., Turnbaugh, P.J., Ul-Hasan, S., van der Hooft, J.J.J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von

Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K.C., Williamson, C.H.D., Willis, A.D., Xu, Z.Z., Zaneveld, J.R., Zhang, Y., Zhu, Q., Knight, R., Caporaso, J.G., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat. Biotechnol. 37, 852–857.

Bray, J.R., Curtis, J.T., 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecol. Monogr. 27, 325–349. https://doi.org/10.2307/1942268

Bukata, R.P., Jerome, J.H., Bruton, J.E., 1988. Relationships Among Secchi Disk Depth, Beam Attenuation Coefficient, and Irradiance Attenuation Coefficient for Great Lakes Waters. J. Great Lakes Res. 14, 347–355.

Burkholder, J., Frazier, W., Rothenberger, M.B., 2010. Source Water Assessment and Control/Treatment Strategies for Harmful and Noxious Algae. in American Water Works Association, Algae Source to Treatment - Manual of Water Supply Practices M57 First Edition, pp. 299-328.

Bukin, Y.S., Galachyants, Y.P., Morozov, I. V., Bukin, S. V., Zakharenko, A.S., Zemskaya, T.I., 2019. The effect of 16S rRNA region choice on bacterial community metabarcoding results. Sci. Data 6, 1–14.

Butterwick, C., Heaney, S.I., Talling, J.F., 2005. Diversity in the influence of temperature on the growth rates of freshwater algae, and its ecological relevance. Freshw. Biol. 50, 291–300.

Cai, H., Wang, K., Huang, S., Jiao, N., Chen, F., 2010. Distinct patterns of picocyanobacterial communities in winter and summer in the chesapeake bay. Appl. Environ. Microbiol. 76, 2955–2960.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639–2643.

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. Nat. Methods 13, 581–583.

Callieri, C., 2008. Picophytoplankton in Freshwater Ecosystems: The Importance of Small-

Sized Phototrophs. Freshw. Rev. 1, 1–28.

Callieri, C., Cronberg, G., Stockner, J.G., 2012. Freshwater Picocyanobacteria: Single Cells, Microcolonies and Colonial Forms, in: Whitton, B.A. (Ed.), Ecology of Cyanobacteria II: Their Diversity in Space and Time. Springer Netherlands, Dordrecht, pp. 229–269.

Callieri, C., Stockner, J., 2000. Picocyanobacteria success in oligotrophic lakes: Fact or fiction? J. Limnol. 59, 72–76.

Callieri, C., Stockner, J.G., 2002. Freshwater autotrophic picoplankton: A review. J. Limnol. 61, 1–14.

Camacho, A., Miracle, M.R., Vicente, E., 2003. Which factors determine the abundance and distribution of picocyanobacteria in inland waters? A comparison among different types of lakes and ponds. Arch. fur Hydrobiol. 157, 321–338.

Cameron, E.S., Tremblay, B.J-M., 2020. *mirln*: Multiple iterations of rarefying for library normalization. http://github.com/escamero/mirlyn

Cameron, E.S., Schmidt, P.J., Tremblay, B. J.-M., Emelko, M.B., Müller, K.M., 2020. To rarefy or not to rarefy: Enhancing microbial community analysis through next-generation sequencing. bioRxiv.

Carey, C.C., Weathers, K.C., Cottingham, K.L., 2008. Gloeotrichia echinulata blooms in an oligotrophic lake: Helpful insights from eutrophic lakes. J. Plankton Res. 30, 893–904.

Carmichael, W.W., 1994. The toxins of cyanobacteria. Sci. Am. 270, 78–86.

Carpenter, S. R., 1983. Lake geometry: implications for production and sediment accretion rates. J. Theor. Biol., 105, 273-286.

Case, R.J., Boucher, Y., Dahllöf, I., Holmström, C., Doolittle, W.F., Kjelleberg, S., 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl. Environ. Microbiol. 73, 278–288.

Cervantes Cianca, R.C., Baptista, M.S., Lopes, V.R., Vasconcelos, V.M., 2012. The non-protein amino acid β-N-methylamino-l-alanine in Portuguese cyanobacterial isolates. Amino Acids 42, 2473–2479.

Chao, A., Bunge, J., 2002. Estimating the number of species in a stochastic abundance

model. Biometrics 58, 531–539.

Chapman, 2010. Cyanobacteria, in American Water Works Association, Algae Source to Treatment - Manual of Water Supply Practices M57 First Edition, pp. 125-146.

Chen, F., Wang, K., Kan, J., Suzuki, M.T., Wommack, K.E., 2006. Diverse and unique picocyanobacteria in Chesapeake Bay, revealed by 16S-23S rRNA internal transcribed spacer sequences. Appl. Environ. Microbiol. 72, 2239–2243.

Chen, L., Reeve, J., Zhang, L., Huang, S., Wang, X., Chen, J., 2018. GMPR: A robust normalization method for zero-inflated count data with application to microbiome sequencing data. PeerJ 2018, 1–20.

Cheung, M.Y., Liang, S., Lee, J., 2013. Toxin-producing cyanobacteria in freshwater: A review of the problems, impact on drinking water safety, and efforts for protecting public health. J. Microbiol. 51, 1–10.

Chien, Y.C., Wu, S.C., Chen, W.C., Chou, C.C., 2013. Model simulation of diurnal vertical migration patterns of different-sized colonies of microcystis employing a particle trajectory approach. Environ. Eng. Sci. 30, 179–186.

Chik, A.H.S., Emelko, M.B., Anderson, W.B., O'Sullivan, K.E., Savio, D., Farnleitner, A.H., Blaschke, A.P., Schijven, J.F., 2020. Evaluation of groundwater bacterial community composition to inform waterborne pathogen vulnerability assessments. Sci. Total Environ. 743, 140472.

Chik, A.H.S., Schmidt, P.J., Emelko, M.B., 2018. Learning something from nothing: The critical importance of rethinking microbial non-detects. Front. Microbiol. 9, 1–9.

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. Int. J. Environ. Res. Public Health 14.

Cho, J.C., Giovannoni, S.J., 2004. Cultivation and Growth Characteristics of a Diverse Group of Oligotrophic Marine Gammaproteobacteria. Appl. Environ. Microbiol. 70, 432–440.

Chorus, I., Bartram, J. (Eds.), 1999. Toxic cyanobacteria in water: A guide to their public

health consequences, monitoring and management. E & FN Spon, London.

Chorus, I., Falconer, I.R., Salas, H.J., Bartram, J., 2000. Health risks caused by freshwater cyanobacteria in recreational waters. J. Toxicol. Environ. Heal. - Part B Crit. Rev. 3, 323–347.

Christensen, V.G., Stelzer, E.A., Eikenberry, B.C., Olds, H.T., LeDuc, J.F., Maki, R.P., Saley, A.M., Norland, J., Khan, E., 2021. Cyanotoxin mixture models: Relating environmental variables and toxin co-occurrence to human exposure risk. J. Hazard. Mater. 415.

Chu, Z., Jin, X., Yang, B., Zeng, Q., 2007. Buoyancy regulation of *Microcystis flos-aquae* during phosphorus-limited and nitrogen-limited growth. J. Plankton Res. 29, 739–745.

Clooney, A.G., Fouhy, F., Sleator, R.D., O'Driscoll, A., Stanton, C., Cotter, P.D., Claesson, M.J., 2016. Comparing apples and oranges?: Next generation sequencing and its impact on microbiome analysis. PLoS One 11, 1–16.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: Data and tools for high throughput rRNA analysis. Nucleic Acids Res. 42, 633–642.

Collos, Y., Bec, B., Jauzein, C., Abadie, E., Laugier, T., Lautier, J., Pastoureaud, A., Souchu, P., Vaquer, A., 2009. Oligotrophication and emergence of picocyanobacteria and a toxic dinoflagellate in Thau lagoon, southern France. J. Sea Res. 61, 68–75.

Contosta, A.R., Adolph, A., Burchsted, D., Burakowski, E., Green, M., Guerra, D., Albert, M., Dibb, J., Martin, M., McDowell, W.H., Routhier, M., Wake, C., Whitaker, R., Wollheim, W., 2017. A longer vernal window: the role of winter coldness and snowpack in driving spring transitions and lags. Glob. Chang. Biol. 23, 1610–1625.

Cox, P.A., Banack, S.A., Murch, S.J., 2003. Biomagnification of cyanobacterial neurotoxins and neurodegenerative disease among the Chamorro people of Guam. Proc. Natl. Acad. Sci. U. S. A. 100, 13380–13383.

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. Hydrol. Process. 29, 3519–3534.

Creed, I.F., Spargo, A.T., Jones, J.A., Buttle, J.M., Adams, M.B., Beall, F.D., Booth, E.G., Campbell, J.L., Clow, D., Elder, K., Green, M.B., Grimm, N.B., Miniat, C., Ramlal, P., Saha, A., Sebestyen, S., Spittlehouse, D., Sterling, S., Williams, M.W., Winkler, R., Yao, H., 2014. Changing forest water yields in response to climate warming: Results from long-term experimental watershed sites across North America. Glob. Chang. Biol. 20, 3191–3208.

Cunha, D.G.F., Sabogal-Paz, L.P., Dodds, W.K., 2016. Land use influence on raw surface water quality and treatment costs for drinking supply in São Paulo State (Brazil). Ecol. Eng. 94, 516–524.

Cyr, H., 2017a. Winds and the distribution of nearshore phytoplankton in a stratified lake. Water Res. 122, 114–127.

Davison, I.R., 1991. Environmental effects on algal photosynthesis: temperature. J. Phycol. 27, 2–8.

Dell, A.I., Pawar, S., Savage, V.M., 2011. Systematic variation in the temperature dependence of physiological and ecological traits. Proc. Natl. Acad. Sci. 108, 10591–10596.

Demoulin, C.F., Lara, Y.J., Cornet, L., François, C., Baurain, D., Wilmotte, A., Javaux, E.J., 2019. Cyanobacteria evolution: Insight from the fossil record. Free Radic. Biol. Med. 140, 206–223.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L., 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72, 5069–5072.

Di Rienzi, S.C., Sharon, I., Wrighton, K.C., Koren, O., Hug, L.A., Thomas, B.C., Goodrich, J.K., Bell, J.T., Spector, T.D., Banfield, J.F., Ley, R.E., 2013. The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. Elife 2, 1–25.

Dokulil, M.T., 2003. Chapter 9 Algae as ecological bio-indicators. Trace Met. other Contam. Environ. 6, 285–327.

Drakare, S., Liess, A., 2010. Local factors control the community composition of cyanobacteria in lakes while heterotrophic bacteria follow a neutral model. Freshw. Biol. 55, 2447–2457.

Ehrlich, L. Sampling and Identification: Methods and Strategies, in American Water Works Association, Algae Source to Treatment - Manual of Water Supply Practices M57 First Edition, pp. 25-69.

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. Glob. Chang. Biol. 22, 1168–1184.

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 Res., 45, 461-472.

Escapa, I. F., Huang, Y., Chen, T., Lin, M., Kokaras, A., Dewhirst, F.E., Lemon, K.P., 2020. Construction of habitat-specific training sets to achieve species-level assignment in 16S rRNA gene datasets. Microbiome 8, 65.

Fahnenstiel, G.L., Carrick, H.J., Rogers, C.E., Sicko-Goad, L., 1991. Red Fluorescing Phototrophic Picoplankton in the Laurentian Great Lakes: What Are They and What Are They Doing? Int. Rev. der gesamten Hydrobiol. und Hydrogr. 76, 603–616.

Fanesi, A., Wagner, H., Becker, A., Wilhelm, C., 2016. Temperature affects the partitioning of absorbed light energy in freshwater phytoplankton. Freshw. Biol. 61, 1365–1378.

Felföldi, T., Ramganesh, S., Somogyi, B., Krett, G., Jurecska, L., Szabó, A., Vörös, L., Márialigeti, K., Máthé, I., 2016. Winter Planktonic Microbial Communities in Highland Aquatic Habitats. Geomicrobiol. J. 33, 494–504.

Ferguson, R.L., Buckley, E.N., Palumbo, A. V., 1984. Response of marine bacterioplankton to differential filtration and confinement. Appl. Environ. Microbiol. 47, 49–55.

Fernandes, A.D., Reid, J.N.S., Macklaim, J.M., McMurrough, T.A., Edgell, D.R., Gloor, G.B., 2014. Unifying the analysis of high-throughput sequencing datasets: Characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. Microbiome 2, 1–13.

Field, K.G., Olsen, G.J., Lane, D.J., Giovannoni, S.J., Ghiselin, M.T., Raff, E.C., Pace, N.R., Raff, R.A., 1988. Molecular phylogeny of the animal kingdom. Science, 80, 239, 748–753.

Filippini, M., Buesing, N., Gessner, M.O., 2008. Temporal dynamics of freshwater bacterio- and virioplankton along a littoral-pelagic gradient. Freshw. Biol. 53, 1114–1125.

Freeman, E.C., Creed, I.F., Jones, B., Bergström, A.K., 2020. Global changes may be promoting a rise in select cyanobacteria in nutrient-poor northern lakes. Glob. Chang. Biol. 26, 4966–4987.

Frempong, E., 1981. Diel Variation in the Abundance, Vertical Distribution, and Species Composition of Phytoplankton in a Eutrophic English Lake. J. Ecol. 69, 919.

Gallina, N., Beniston, M., Jacquet, S., 2017. Estimating future cyanobacterial occurrence and importance in lakes: a case study with *Planktothrix rubescens* in Lake Geneva. Aquat. Sci. 79, 249–263.

Ganf, G.G., 1974. Diurnal Mixing and the Vertical Distribution of Phytoplankton in a Shallow Equatorial Lake (Lake George , Uganda). J. Ecol. 62, 611–629.

Garcia-Pichel, F., Zehr, J. P., Bhattacharya, D., Pakrasi, H. B., 2019. What's in a name? The case of cyanobacteria. J. Phycol. 56, 1-5.

Genkai-Kato, M., Carpenter, S. R., 2005. Eutrophication due to phosphorus recycling in relation to lake morphometry, temperature, and macrophytes. Ecology, 86, 1, 210-219.

Genuário, D.B., Lorenzi, A.S., Agujaro, L.F., Isaac, R. de L., Azevedo, M.T. de P., Cantúsio Neto, R., Fiore, M.F., 2016. Cyanobacterial community and microcystin production in a recreational reservoir with constant *Microcystis* blooms. Hydrobiologia 779, 105–125.

Gillooly, J., Dodson, S., 2000. Latitudinal patterns in the size distribution and seasonal dynamics of new world, freshwater cladocerans. Limnol. Oceanogr. 45, 22–30.

Gin, K.Y.H., Sim, Z.Y., Goh, K.C., Kok, J.W.K., Te, S.H., Tran, N.H., Li, W., He, Y., 2021. Novel cyanotoxin-producing Synechococcus in tropical lakes. Water Res. 192, 116828.

Girones, R., Ferrús, M.A., Alonso, J.L., Rodriguez-Manzano, J., Calgua, B., de Abreu Corrêa, A., Hundesa, A., Carratala, A., Bofill-Mas, S., 2010. Molecular detection of pathogens in water - The pros and cons of molecular techniques. Water Res. 44, 4325–4339.

Gladyshev, M.I., Sushchik, N.N., Makhutova, O.N., 2013. Production of EPA and DHA in aquatic ecosystems and their transfer to the land. Prostaglandins Other Lipid Mediat. 107, 117–126. https://doi.org/10.1016/j.prostaglandins.2013.03.002

Gloor, G.B., Macklaim, J.M., Pawlowsky-Glahn, V., Egozcue, J.J., 2017. Microbiome datasets are compositional: And this is not optional. Front. Microbiol. 8, 1–6.

Gloor, G.B., Macklaim, J.M., Vu, M., Fernandes, A.D., 2016. Compositional uncertainty should not be ignored in high-throughput sequencing data analysis. Austrian J. Stat. 45, 73–87.

Graham, J.L., Loftin, K.A., Ziegler, A.C., Meyer, M.T., 2008. Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs, Scientific investigations report 2008-5038.

Gray, M.W., Sankoff, D., Cedergren, R.J., 1984. On the evolutionary descent of organisms and organelles: A global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res. 12, 5837–5852.

Griffith, A., Loik, M., 2010. Effects of climate and snow depth on Bromus tectorum population dynamics at high elevation. Oecologia 164, 821–832.

Guschina, I. a., Harwood, J.L., 2006. Lipids and lipid metabolism in eukaryotic algae. Prog. Lipid Res. 45, 160–186.

Hall, J.A., Vincent, W.F., 1994. Vertical and horizontal structure of the picophytoplankton community in a stratified coastal system off New Zealand. New Zeal. J. Mar. Freshw. Res. 28, 299–308.

Hall, M., Beiko, R.G., 2018. 16S rRNA Gene Analysis with QIIME2, in: Methods in

Molecular Biology.

Hampton, S.E., Moore, M. V., Ozersky, T., Stanley, E.H., Polashenski, C.M., Galloway, A.W.E., 2015. Heating up a cold subject: Prospects for under-ice plankton research in lakes. J. Plankton Res. 37, 277–284.

Hao, Z., Li, Y., Cai, W., Wu, P., Liu, Y., Wang, G., 2012. Possible nutrient limiting factor in long term operation of closed aquatic ecosystem. Adv. Sp. Res. 49, 841–849.

Harke, M.J., Steffen, M.M., Gobler, C.J., Otten, T.G., Wilhelm, S.W., Wood, S.A., Paerl, H.W., 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. Harmful Algae 54, 4–20.

Heisler, J., Glibert, P.M., Burkholder, J.M., Anderson, D.M., Cochlan, W., Dennison, W.C., Dortch, Q., Gobler, C.J., Heil, C.A., Humphries, E., Lewitus, A., Magnien, R., Marshall, H.G., Sellner, K., Stockwell, D.A., Stoecker, D.K., Suddleson, M., 2008. Eutrophication and harmful algal blooms: A scientific consensus. Harmful Algae 8, 3–13.

Hodkinson, B.P., Grice, E.A., 2015. Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research. Adv. Wound Care 4, 50–58.

Howard, A., 2001. Modeling Movement Patterns of the Cyanobacterium , Microcystis. Ecol. Appl. 11, 304–310.

Hugerth, L.W., Andersson, A.F., 2017. Analysing microbial community composition through amplicon sequencing: From sampling to hypothesis testing. Front. Microbiol. 8, 1561.

Hughes, J.B., Hellmann, J.J., Ricketts, T.H., Bohannan, B.J.M., 2001. Counting the Uncountable: Statistical Approaches to Estimating Microbial Diversity. Appl. Environ. Microbiol. 67, 4399–4406.

Huisman, J., Codd, G.A., Paerl, H.W., Ibelings, B.W., Verspagen, J.M.H., Visser, P.M., 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. 16, 471–483.

Hunter, P.D., Tyler, A.N., Willby, N.J., Gilvear, D.J., 2008. The spatial dynamics of vertical migration by Microcystis aeruginosa in a eutrophic shallow lake: A case study using high spatial resolution time-series airborne remote sensing. Limnol. Oceanogr. 53, 2391–2406.

Ibelings, B.W., Mur, L.R., Walsby, A.E., 1991. Diurnal changes in buoyancy and vertical distribution in populations of Microcystisin two shallow lakes. J. Plankton Res. 13, 419–436.

Irwin, A.J., Finkel, Z. V., Schofield, O.M.E., Falkowski, P.G., 2006. Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. J. Plankton Res. 28, 459–471.

Irwin, A.J., Nelles, A.M., Finkel, Z. V., 2012. Phytoplankton niches estimated from field data. Limnol. Oceanogr. 57, 787–797.

Jakubowska, N., Szeląg-Wasielewska, E., 2015. Toxic picoplanktonic cyanobacteria - Review. Mar. Drugs 13, 1497–1518.

Järvinen, M., Rask, M., Ruuhijärvi, J., Arvola, L., 2002. Temporal coherence in water temperature and chemistry under the ice of boreal lakes (Finland). Water Res. 36, 3949–3956.

Jaworska, B., Zdanowski, B., 2012. Patterns of seasonal phytoplankton dynamics as the element of ecological successional changes proceeding in a lake (Lake Kortowskie, northern Poland). Limnol. Rev. 11, 105–112.

Jeffries, D.S., Foster, N.W., 2001. The Turkey Lakes Watershed Study: Milestones and prospects. Ecosystems 4, 501–502.

Jeffries, D.S., Kelso, J.R.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.

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M., Sodergren, E., Weinstock, G.M., 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nat. Commun. 10, 1–11.

Kharin, V. V., Zwiers, F.W., Zhang, X., Hegerl, G.C., 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J. Clim. 20, 1419–1444.

Khetan, D., Gupta, N., Chaudhary, R., Shukla, J.S., 2019. Comparison of UV spectrometry and fluorometry-based methods for quantification of cell-free DNA in red cell components. Asian J. Transfus. Sci. 13, 95–99.

Kim, M., Morrison, M., Yu, Z., 2011. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J. Microbiol. Methods 84, 81–87.

Kirisits, M.J., Emelko, M.B., Pinto, A.J., 2019. Applying biotechnology for drinking water biofiltration: advancing science and practice. Curr. Opin. Biotechnol. 57, 197–204.

Komárek, J., Kaštovský, J., Mareš, J., & Johansen, J. R., 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. Preslia, 86, 295-33.

Colorado Lakes and Reservoir Management, 2015. Guidance Document for Harmful Algal Blooms in Colorado.

Lam, T.Y.C., Mei, R., Wu, Z., Lee, P.K.H., Liu, W.T., Lee, P.H., 2020. Superior resolution characterisation of microbial diversity in anaerobic digesters using full-length 16S rRNA gene amplicon sequencing. Water Res. 178, 115815.

Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A., Clemente, J.C., Burkepile, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G., Huttenhower, C., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814–821.

Lapierre, J.F., Seekell, D.A., del Giorgio, P.A., 2015. Climate and landscape influence on indicators of lake carbon cycling through spatial patterns in dissolved organic carbon. Glob. Chang. Biol. 21, 4425–4435.

Li, H., Barber, M., Lu, J., Goel, R., 2020. Microbial community successions and their dynamic functions during harmful cyanobacterial blooms in a freshwater lake. Water Res. 185, 116292.

Li, M., Zhu, W., Guo, L., Hu, J., Chen, H., Xiao, M., 2016. To increase size or decrease

density? Different *Microcystis* species has different choice to form blooms. Sci. Rep. 6, 1–10.

Lindsay, J.B., Creed, I.F., Beall, F.D., 2004. Drainage basin morphometrics for depressional landscapes. Water Resour. Res. 40, 1–9.

Ling, F., Whitaker, R., LeChevallier, M.W., Liu, W.T., 2018. Drinking water microbiome assembly induced by water stagnation. ISME J. 12, 1520–1531.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550.

Lozupone, C.A., Knight, R., 2007. Global patterns in bacterial diversity. Proc. Natl. Acad. Sci. U. S. A. 104, 11436–11440.

Manoylov, K.M., 2014. Taxonomic identification of algae (morphological and molecular): species concepts, methodologies, and their implications for ecological bioassessment. J. Phycol. 50, 409–424.

Masclaux, H., Bec, A., Kainz, M.J., Perrière, F., Desvilettes, C., Bourdier, G., 2012. Accumulation of polyunsaturated fatty acids by cladocerans: Effects of taxonomy, temperature and food. Freshw. Biol. 57, 696–703.

McGuire, M.J., 1995. Off-Flavor as the Consumer's Measure of Drinking Water Safety. Water Sci. 31, 1–8.

McKnight, D.T., Huerlimann, R., Bower, D.S., Schwarzkopf, L., Alford, R.A., Zenger, K.R., 2019. Methods for normalizing microbiome data: An ecological perspective. Methods Ecol. Evol. 10, 389–400.

McLaren, M.R., Willis, A.D., Callahan, B.J., 2019. Consistent and correctable bias in metagenomic sequencing experiments. Elife 8, 1–31.

McMurdie, P.J., Holmes, S., 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. PLoS Comput. Biol. 10.

McMurdie, P.J., Holmes, S., 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One 8, e61217.

McMurdie, P.J., Holmes, S., 2012. *Phyloseq*: A bioconductor package for handling and

analysis of high-throughput phylogenetic sequence data, in: Pacific Symposium on Biocomputing.

McQuillan, J.S., Robidart, J.C., 2017. Molecular-biological sensing in aquatic environments: recent developments and emerging capabilities. Curr. Opin. Biotechnol. 45, 43–50.

Meerhoff, M., Teixeira-de Mello, F., Kruk, C., Alonso, C., González-Bergonzoni, I., Pacheco, J.P., Lacerot, G., Arim, M., Beklioğlu, M., Brucet, S., Goyenola, G., Iglesias, C., Mazzeo, N., Kosten, S., Jeppesen, E., 2012. Environmental Warming in Shallow Lakes. A Review of Potential Changes in Community Structure as Evidenced from Space-for-Time Substitution Approaches, Advances in Ecological Research.

Ministry for the Environment and Ministry of Health., 2009. New Zealand Guidelines for Cyanobacteria in Recreational Fresh Waters. Ministry for the Environment, Wellington.

Moisan, J.R., Moisan, T.A., Abbott, M.R., 2002. Modelling the effect of temperature on the maximum growth rates of phytoplankton populations. Ecol. Modell. 153, 197–215.

Monchamp, M.E., Spaak, P., Pomati, F., 2019. Long term diversity and distribution of non-photosynthetic cyanobacteria in peri-alpine lakes. Front. Microbiol. 10, 1–11.

Mountain, N., James A.L., Chutko, K., 2015. Groundwater and surface water influences on streamflow in a mesoscale Precambrian Shield catchment. Hydrol. Proccess. 29, 941-3953.

Naselli-Flores, L., Zohary, T., Padisák, J., 2021. Life in suspension and its impact on phytoplankton morphology: an homage to Colin S. Reynolds. Hydrobiologia 848, 7–30.

Navas-Molina, J.A., Peralta-Sánchez, J.M., González, A., McMurdie, P.J., Vázquez-Baeza, Y., Xu, Z., Ursell, L.K., Lauber, C., Zhou, H., Song, S.J., Huntley, J., Ackermann, G.L., Berg-Lyons, D., Holmes, S., Caporaso, J.G., Knight, R., 2013. Advancing our understanding of the human microbiome using QIIME. Methods Enzymol. 531, 371–444.

Newcombe, G. (Ed.), 2009. International Guidane Manual for the Management of Toxic Cyanobacteria. Global Water Research Coalition.

Nunes-Costa, D., Magalhães, J.D., G-Fernandes, M., Cardoso, S.M., Empadinhas, N., 2020.

Microbial BMAA and the Pathway for Parkinson's Disease Neurodegeneration. Front. Aging Neurosci. 12, 1–9.

Ohio Environmental Protection Agency, 2013. Inland Lakes Sampling Procedure Manual.

Olli, K., 1999. Diel vertical migration of phytoplankton and heterotrophic flagellates in the Gulf of Riga. J. Mar. Syst. 23, 145–163.

Paerl, H.W., 2014. Mitigating harmful cyanobacterial blooms in a human- and climatically-impacted world. Life 4, 988–1012.

Paerl, H.W., Huisman, J., 2009. Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. Environ. Microbiol. Rep. 1, 27–37.

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. Environ. Sci. Technol. 50, 10805–10813.

Paranjape, K., Bédard, É., Whyte, L.G., Ronholm, J., Prévost, M., Faucher, S.P., 2020. Presence of Legionella spp. in cooling towers: the role of microbial diversity, Pseudomonas, and continuous chlorine application. Water Res. 169, 115252.

Passoni, S., Callieri, C., 2000. Picocyanobacteria single forms, aggregates and microcolonies: survival strategy or species succession? SIL Proceedings, 1922-2010 27, 1879–1883.

Patel, R., de Oliveira, A., Newby, R., Chu, T., 2019. Flow cytometric analysis of freshwater cyanobacteria: A case study. Water (Switzerland) 11.

Paterson, A.M., Rühland, K.M., Anstey, C. V., Smol, J.P., 2017. Climate as a driver of increasing algal production in Lake of the Woods, Ontario, Canada. Lake Reserv. Manag. 33, 403–414.

Perrin, Y., Bouchon, D., Delafont, V., Moulin, L., Héchard, Y., 2019. Microbiome of drinking water: A full-scale spatio-temporal study to monitor water quality in the Paris distribution system. Water Res. 149, 375–385.

Pfeifer, F., 2012. Distribution, formation and regulation of gas vesicles. Nat. Rev. Microbiol. 10, 705–715.

Phillips, K.A., Fawley, M.W., 2002. Winter phytoplankton community structure in three shallow temperate lakes during ice cover. Hydrobiologia 470, 97–113.

Pick, F.R., Agbeti, M., 1991. The Seasonal Dynamics and Composition of Photosynthetic Picoplankton Communities in Temperate Lakes in Ontario, Canada. Int. Rev. der gesamten Hydrobiol. und Hydrogr. 76, 565–580.

Pobel, D., Robin, J., Humbert, J.F., 2011. Influence of sampling strategies on the monitoring of cyanobacteria in shallow lakes: Lessons from a case study in France. Water Res. 45, 1005–1014.

Postius, C., Ernst, A., 1999. Mechanisms of dominance: Coexistence of picocyanobacterial genotypes in a freshwater ecosystem. Arch. Microbiol. 172, 69–75.

Powers, S.M., Hampton, S.E., 2016. Winter Limnology as a New Frontier. Limnol. Oceanogr. Bull. 25, 103–108.

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 Res. 41, 590–596.

Quinlan, R., Paterson, A.M., Hall, R.I., Dillon, P.J., Wilkinson, A.N., Cumming, B.F., Douglas, M.S. V, Smol, J.P., Dillion, P.J., 2003. A landscape approach to examining spatial patterns of limnological variables and long-term environmental change in a southern Canadian lake district. Freshw. Biol. 48, 1676–1697.

R Core Team 2020. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Reynolds, C.S., 2006. Ecology of Phytoplankton. Cambridge University Press, New York.

Reynolds, C.S., Oliver, R.L., Walsby, A.E., 1987. Cyanobacterial dominance: The role of buoyancy regulation in dynamic lake environments. New Zeal. J. Mar. Freshw. Res. 21, 379–390.

Rhee, G.-Y., Gotham, I.J., 1981. The effect of environmental factors on phytoplankton growth: Temperature and the interactions of temperature with nutrient limitation1. Limnol. Oceanogr. 26, 635–648.

Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140.

Rusak, J.A., Tanentzap, A.J., Klug, J.L., Rose, K.C., Hendricks, S.P., Jennings, E., Laas, A., Pierson, D., Ryder, E., Smyth, R.L., White, D.S., Winslow, L.A., Adrian, R., Arvola, L., de Eyto, E., Feuchtmayr, H., Honti, M., Istvánovics, V., Jones, I.D., McBride, C.G., Schmidt, S.R., Seekell, D., Staehr, P.A., Zhu, G., 2018. Wind and trophic status explain within and among-lake variability of algal biomass. Limnol. Oceanogr. Lett. 3, 409–418.

Sanders, H.L., 1968. Marine Benthic Diversity : A Comparative Study. Am. Nat. 102, 243–282.

Sarnelle, O., Morrison, J., Kaul, R., Horst, G., Wandell, H., Bednarz, R., 2010. Citizen monitoring: Testing hypotheses about the interactive influences of eutrophication and mussel invasion on a cyanobacterial toxin in lakes. Water Res. 44, 141–150.

Sass, G.Z., Creed, I.F., Bayley, S.E., Devito, K.J., 2007. Understanding variation in trophic status of lakes on the Boreal Plain: A 20 year retrospective using Landsat TM imagery. Remote Sens. Environ. 109, 127–141.

Schatz, D., Keren, Y., Vardi, A., Sukenik, A., Carmeli, S., Börner, T., Dittmann, E., Kaplan, A., 2007. Towards clarification of the biological role of microcystins, a family of cyanobacterial toxins. Environ. Microbiol. 9, 965–970.

Schindler, D.W., 1977. Evolution of Phosphorus Limitation in Lakes. Science (80-.). 195, 260–262.

Schloss, P.D., Handelsman, J., 2004. Status of the Microbial Census. Microbiol. Mol. Biol. Rev. 64, 686–691.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 75, 7537–7541.

Schmidt, P. J., Cameron, E. S., Müller, K. M., Emelko, M. B., 2021. Ensuring that fundamentals of quantitative microbiology are reflected in microbial diversity analyses based on next-generation sequencing. bioRxiv.

Sepkoski, J.J., 1988. Alpha, beta, or gamma: Where does all the diversity go? Paleobiology 14, 221–234.

Shahraki, A.H., Chaganti, S.R., Heath, D.D., 2020. Diel Dynamics of Freshwater Bacterial Communities at Beaches in Lake Erie and Lake St. Clair, Windsor, Ontario. Microb. Ecol. 81, 1–13.

Shannon, C.E., 1948. A mathematical theory of communication. The Bell System Techinical Journal, 27:369-423, 623-656.

Sharma, N.K., Rai, A.K., 2011. Biodiversity and biogeography of microalgae: progress and pitfalls. Environ. Rev. 19, 1–15.

Shaw, J.L.A., Monis, P., Weyrich, L.S., Sawade, E., Drikas, M., Cooper, A.J., 2015. Using amplicon sequencing to characterize and monitor bacterial diversity in drinking water distribution systems. Appl. Environ. Microbiol. 81, 6463–6473.

Sherwood, A.R., Presting, G.G., 2007. Universal primers amplify a 23S rDNA plastid marker in eukaryotic algae and cyanobacteria. J. Phycol. 43, 605–608.

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecol. 21, 1794–1805.

Sieburth, J.M., Smetacek, V., Lenz, J., 1978. Pelagic ecosystem structure: Heterotrophic compartments of the plankton and their relationship to plankton size fractions. Limnol. Oceanogr. 23, 1256–1263.

Silverman, J., Roche, K., Mukherjee, S., David, L., 2018. Naught all zeros in sequence count data are the same. bioRxiv 477794.

Simpson, E.H., 1949. Measurment of Diversity. Nature 688, 688.

Śliwińska-Wilczewska, S., Maculewicz, J., Felpeto, A.B., Latała, A., 2018. Allelopathic and bloom-forming picocyanobacteria in a changing world. Toxins (Basel). 10, 1–20.

Smokorowski, K.E., Pearce, J.L., Geiling, W.D., Pratt, T.C., 2021. Wood Removals from

Lakes May Not Necessarily Elicit Fish Population Responses. North Am. J. Fish. Manag. 41, 142–157.

Smol, J.P., Wolfe, A.P., Birks, H.J.B., Douglas, M.S. V., Jones, V.J., Korhola, A., Pienitz, R., Ruhland, K., Sorvari, S., Antoniades, D., Brooks, S.J., Fallu, M.-A., Hughes, M., Keatley, B.E., Laing, T.E., Michelutti, N., Nazarova, L., Nyman, M., Paterson, A.M., Perren, B., Quinlan, R., Rautio, M., Saulnier-Talbot, E., Siitonen, S., Solovieva, N., Weckstrom, J., 2005. Climate-driven regime shifts in the biological communities of arctic lakes. Proc. Natl. Acad. Sci. 102, 4397–4402.

Soo, R., Skennerton, C., Sekiguchi, Y., Imelfort, M., Paech, S., Dennis, P., Steen, J., Parks, D., Tyson, G., Hugenholtz, P., 2014. Photosynthesis is not a universal feature of the phylum Cyanobacteria.

Staehr, P.A., Baastrup-Spohr, L., Sand-Jensen, K., Stedmon, C., 2012. Lake metabolism scales with lake morphometry and catchment conditions. Aquat. Sci. 74, 155–169.

Staehr, P.A., Birkeland, M.J., 2006. Temperature acclimation of growth, photosynthesis and respiration in two mesophilic phytoplankton species. Phycologia 45, 648–656.

Staley, J.T., 1980. The gas vacuole : an early organelle of prokaryote motility? Orig. Life 10, 111–116.

Stewart, I., Schluter, P.J., Shaw, G.R., 2006. Cyanobacterial lipopolysaccharides and human health - A review. Environ. Heal. A Glob. Access Sci. Source 5, 1–23.

Stockner, J., Callieri, C., Cronberg, G., 2006. Picoplankton and Other Non-Bloom-Forming Cyanobacteria in Lakes. Ecol. Cyanobacteria 195–231.

Stockner, J., Callieri, C., Cronberg, G., 2002. Picoplankton and Other Non-Bloom-Forming Cyanobacteria in Lakes, in: Whitton, B.A., Potts, M. (Eds.), The Ecology of Cyanobacteria: Their Diversity in Time and Space. Springer Netherlands, Dordrecht, pp. 195–231.

Suffet, I.H., Khiari, D., Mallevialle, J., 1996. Taste and odor problems in drinking water, in: Saxby, M.J. (Ed.), Food Taints and Off-Flavours. Springer US, Boston, MA, pp. 107–138.

Thermo Fisher Scientific, 2009. NanoDrop 2000/2000c Spectrophotometer V1.0 User Manual.

Thomas, T., Gilbert, J., Meyer, F., 2012. Metagenomics - a guide from sampling to data analysis. Microb. Inform. Exp. 2, 3.

Tran, P., Ramachandran, A., Khawasik, O., Beisner, B.E., Rautio, M., Huot, Y., Walsh, D.A., 2018. Microbial life under ice: Metagenome diversity and in situ activity of Verrucomicrobia in seasonally ice-covered Lakes. Environ. Microbiol. 20, 2568–2584.

Tromas, N., Fortin, N., Bedrani, L., Terrat, Y., Cardoso, P., Bird, D., Greer, C.W., Shapiro, B.J., 2017. Characterising and predicting cyanobacterial blooms in an 8-year amplicon sequencing time course. ISME J. 11, 1746–1763.

Tsilimigras, M.C.B., Fodor, A.A., 2016. Compositional data analysis of the microbiome: fundamentals, tools, and challenges. Ann. Epidemiol. 26, 330–335.

Tsukuda, M., Kitahara, K., Miyazaki, K., 2017. Comparative RNA function analysis reveals high functional similarity between distantly related bacterial 16 S rRNAs. Sci. Rep. 7, 1–8.

University of New Hampshire - Center for Freshwater Biology, 2010. Standard Operating Procedure for Field Sampling of Cyanobacteria in Lakes.

Van Kranendonk, M.J., Webb, G.E., Kamber, B.S., 2003. Geological and trace element evidence for a marine sedimentary environment of deposition and biogenicity of 3.45 Ga stromatolitic carbonates in the Pilbara Craton, and support for a reducing Archaean ocean. Geobiology 1, 91–108.

Vareli, K., Jaeger, W., Touka, A., Frillingos, S., Briasoulis, E., Sainis, I., 2013. Hepatotoxic seafood poisoning (HSP) due to microcystins: A threat from the ocean? Mar. Drugs 11, 2751–2768.

Vasconcelos, V., 2001. Cyanobacteria toxins: Diversity and ecological effects. Limnetica 20, 45–58.

Vidal, J., Rigosi, A., Hoyer, A., Escot, C., Rueda, F.J., 2014. Spatial distribution of phytoplankton cells in small elongated lakes subject to weak diurnal wind forcing.

Aquat. Sci. 76, 83–99.

Vidal, L., Ballot, A., Azevedo, S.M.F.O., Padisák, J., Welker, M., 2021. Introduction to cyanobacteria, in: Chorus, I., Welker, M. (Eds.), Toxic Cyanobacteria in Water - A Guide to Their Public Health Consequences, Monitoring and Management. CRC Press, pp. 163–211.

Vieira, J.M.D.S., Azevedo, M.T.D.P., De Oliveira Azevedo, S.M.F., Honda, R.Y., Corrêa, B., 2003. Microcystin production by Radiocystis fernandoi (Chroococcales, Cyanobacteria) isolated from a drinking water reservoir in the city of Belém, PA, Brazilian Amazonia region. Toxicon 42, 709–713.

Vierheilig, J., Savio, D., Farnleitner, A.H., Reischer, G.H., Ley, R.E., Mach, R.L., Farnleitner, A.H., Reischer, G.H., 2015. Potential applications of next generation DNA sequencing of 16S rRNA gene amplicons in microbial water quality monitoring. Water Sci. Technol. 72, 1962–1972.

Visser, P.M., Ibelings, B.W., Mur, L.R., Walsby, A.E., 2005. The Ecophysiology of the Harmful Cyanobacterium Microcystis, in: Huisman, J., Matthijs, H.C.P., Visser, P.M. (Eds.), Harmful Cyanobacteria. Springer, pp. 109–142.

Vu, H.P., Nguyen, L.N., Zdarta, J., Nga, T.T.V., Nghiem, L.D., 2020. Blue-Green Algae in Surface Water: Problems and Opportunities. Curr. Pollut. Reports 6, 105–122.

Walsby, A.E., 1981. Cyanobacteria: Planktonic Gas-Vacuolate Forms, in: Starr, M.P., Stolp, H., Trüper, H.G., Balows, A., Schlegel, H.G. (Eds.), The Prokaryotes: A Handbook on Habitats, Isolation, and Identification of Bacteria. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 224–235.

Walsby, A.E., Hayes, P.K., Boje, R., Stal, L.J., 1997. The selective advantage of buoyancy provided by gas vesicles for planktonic cyanobacteria in the Baltic Sea. New Phytol. 136, 407–417.

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, e0009-15.

Wang, Y., LêCao, K.-A., 2019. Managing batch effects in microbiome data. Brief. Bioinform. https://doi.org/10.1093/bib/bbz105

Waterbury, J.B., Valois, F.W., 1993. Resistance to co-occurring phages enables marine Synechococcus communities to coexist with cyanophages abundant in seawater. Appl. Environ. Microbiol. 59, 3393–3399.

Waterbury, J.B., Watson, S.W., Valois, F.W., Franks, D.G., 1986. Biological and ecological characterization of the marine unicellular cyanobacterium Synechococcus, in: Platt, T., Li, W.K.W. (Eds.), Photosynthetic Picoplankton. p. 71-120.

Watson, S.B., 2003. Cyanobacterial and eukaryotic algal odour compounds: Signals or by-products? A review of their biological activity. Phycologia 42, 332–350.

Watson, S.B., 2010. Algal Taste and Odor, in American Water Works Association, Algae Source to Treatment - Manual of Water Supply Practices M57 First Edition, pp. 329-376.

Webster, K.E., Soranno, P.A., Baines, S.B., Kratz, T.K., Bowser, C.J., Dillon, P.J., Campbell, P., Fee, E.J., Hecky, R.E., 2000. Structuring features of lake districts: Landscape controls on lake chemical responses to drought. Freshw. Biol. 43, 499–515.

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. Hydrol. Process. 3, 1–16.

Weisburg, W.G., Barns, S.M., Pelletier, D.A., Lane, D.J., 1991. 16S Ribosomal DNA Amplification for Phylogenetic Study. J. Bacteriol. 173, 697–703.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J.R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E.R., Knight, R., 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. Microbiome 5, 1–18.

Wejnerowski, Ł., Rzymski, P., Kokociński, M., Meriluoto, J., 2018. The structure and

toxicity of winter cyanobacterial bloom in a eutrophic lake of the temperate zone. Ecotoxicology 27, 752–760.

Welker, M., Chorus, I., Schaeffer, B.A., Urquhart, E., 2021. Planning monitoring programmes for cyanobacteria and cyanotoxins, in: Chorus, I., Welker, M. (Eds.), Toxic Cyanobacteria in Water-A Guide to Their Public Health Consequences, Monitoring and Management. CRC Press, pp. 641–668.

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.

Westrick, J.A., Southwell, B., Szlag, D., Zimba, P.V., 2010. Detection of cyanotoxins during potable water treatment, in American Water Works Association, Algae Source to Treatment - Manual of Water Supply Practices M57 First Edition, pp. 299-328.

Westwood, K.J., Ganf, G.G., 2004. Effect of mixing patterns and light dose on growth of Anabaena circinalis in a turbid, lowland river. River Res. Appl. 20, 115–126.

Wilhelm, S.W., Lecleir, G.R., Bullerjahn, G.S., Mckay, R.M., Saxton, M.A., Twiss, M.R., Bourbonniere, R.A., 2014. Seasonal changes in microbial community structure and activity imply winter production is linked to summer hypoxia in a large lake. FEMS Microbiol. Ecol. 87, 475–485.

Willis, A.D., 2019. Rarefaction, alpha diversity, and statistics. Front. Microbiol. 10.

Winder, M., Hunter, D.A., 2008. Temporal organization of phytoplankton communities linked to physical forcing. Oecologia 156, 179–192.

Winder, M., Reuter, J.E., Schladow, S.G., 2009. Lake warming favours small-sized planktonic diatom species. Proc. R. Soc. B Biol. Sci. 276, 427–435.

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 Reserv. Manag. 27, 107–114.

Woese, C.R., Kandler, O., Wheelis, M.L., 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. Proc. Natl. Acad. Sci. U. S. A. 87, 4576–4579.

Woolway, R.I., Kraemer, B.M., Lenters, J.D., Merchant, C.J., O'Reilly, C.M., Sharma, S., 2020. Global lake responses to climate change. Nat Rev Earth Environ **1,** 388–403.

Xiao, Y., Gan, N., Liu, J., Zheng, L., Song, L., 2012. Heterogeneity of buoyancy in response to light between two buoyant types of cyanobacterium Microcystis. Hydrobiologia 679, 297–311.

Yamamoto, Y., Nakahara, H., 2006. Seasonal variations in the diel vertical distribution of phytoplankton and zooplankton in a shallow pond. Phycol. Res. 54, 280–293.

Yang, B., Wang, Y., Qian, P.Y., 2016. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. BMC Bioinformatics 17, 1–8.

Yang, J., Tang, H., Zhang, X., Zhu, X., Huang, Y., Yang, Z., 2017. High temperature and pH favor *Microcystis aeruginosa* to outcompete *Scenedesmus obliquus*.

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 Res. 42, D643–D648.

Zadereev, E., 2017. Ecology of Meromictic Lakes 228. https://doi.org/10.1007/978-3-319-49143-1

Zamyadi, A., Romanis, C., Mills, T., Neilan, B., Choo, F., Coral, L.A., Gale, D., Newcombe, G., Crosbie, N., Stuetz, R., Henderson, R.K., 2019. Diagnosing water treatment critical control points for cyanobacterial removal: Exploring benefits of combined microscopy, next-generation sequencing, and cell integrity methods. Water Res. 152, 96–105.

Zhang, J., Ding, X., Guan, R., Zhu, C., Xu, C., Zhu, B., Zhang, H., Xiong, Z., Xue, Y., Tu, J., Lu, Z., 2018. Evaluation of different 16S rRNA gene V regions for exploring bacterial diversity in a eutrophic freshwater lake. Sci. Total Environ. 618, 1254–1267.

Zhang, L., Fang, W., Li, X., Lu, W., Li, J., 2020. Strong linkages between dissolved organic matter and the aquatic bacterial community in an urban river. Water Res.

Appendices

Appendix A

Turkey Lakes Watershed: Sampling Summary, Water Temperature Profiles & Water Chemistry

All samples were collected at the deepest point in each lake found at the following coordinates: Big Turkey Lake (47°02'54.7"N 84°25'19.3"W), Little Turkey Lake (47°02'37.2"N 84°24'24.4"W) and Wishart Lake (47°03'00.0"N 84°23'58.3"W).

Table A.1 Sample summary for diurnal bacterial community characterization (Chapter 3). All water samples were collected using a peristaltic pump (Masterflex E/S Portable Sampler), then vacuum filtered through a 47 mm GF/C filter (Whatman, plc, Buckinghamshire, United Kingdom), and filtered again through a 0.22 µm Sterivex™ filter. Whatman GF/C and Sterivex™ filters were stored at -20°C prior to DNA extraction.

Sample ID	Lake	Sampling Depth (m)	Sampling Time	Sampling Date	Volume Filtered (mL)	DNA Concentration (ng/µL) [GF + Sterivex]
TLW94	Wishart	4	9 :30 A.M.	August 22, 2018	600	3.3 + 1
TLW103	Wishart	4	5 :00 P.M.	August 22, 2018	400	3.6 + 1
TLW106	Wishart	4	1 :00 P.M.	August 22, 2018	450	4.3 + 0
TLW112	Wishart	4	8 :30 A.M.	August 23, 2018	500	5.3 + 1
TLW121	Wishart	4	4 :45 P.M.	August 23, 2018	400	3.9 + 1.5
TLW124	Wishart	4	12 :00 P.M.	August 23, 2018	500	4.6 + 1.9

Table A.1 Continued

Sample ID	Lake	Sampling Depth (m)	Sampling Time	Sampling Date	Volume Filtered (mL)	DNA Concentration (ng/µL) [GF + Sterivex]
TLW127	Little Turkey	0	8 :30 A.M.	August 22, 2018	1000	2.4 + 1.3
TLW130	Little Turkey	5	8 :30 A.M.	August 22, 2018	1000	3.9 + 1
TLW133	Little Turkey	0	12 :00 P.M.	August 22, 2018	1000	2.7 + 2.4
TLW136	Little Turkey	0	4 :30 P.M.	August 22, 2018	1000	2.7 + 0
TLW139	Little Turkey	5.25	4 :30 P.M.	August 22, 2018	800	3.6 + 2.7
TLW142	Little Turkey	5.25	12 :00 P.M.	August 22, 2018	700	3.7 + 0
TLW145	Little Turkey	4.75	9 :00 A.M.	August 23, 2018	500	3.8 + 1
TLW157	Little Turkey	5	4 :15 P.M.	August 23, 2018	500	3 + 1
TLW160	Little Turkey	5.25	12 :45 P.M.	August 23, 2018	750	2.7 + 0

Table A.2 Sample summary for seasonality and depth profile for bacterial community characterization (Chapter 4). All water samples were collected using a peristaltic pump (Masterflex E/S Portable Sampler), then vacuum filtered through a 47 mm GF/C filter (Whatman, plc, Buckinghamshire, United Kingdom), and stored at -20°C prior to DNA extraction. To ensure sufficient biomass for analysis, 350 to 1000 mL of water were filtered.

Sample ID	Lake	Sampling Depth (m)	Sampling Date	Volume Filtered (mL)	DNA Concentration (ng/µL)
TLW37	Big Turkey	0	July 18, 2018	950	2.2
TLW42	Big Turkey	7	July 18, 2018	800	12.4
TLW43	Big Turkey	8	July 18, 2018	800	11.7
TLW82	Big Turkey	0	August 13, 2018	750	8
TLW85	Big Turkey	5	August 13, 2018	600	12.1
TLW88	Big Turkey	6	August 13, 2018	500	8.7
TLW171	Big Turkey	4	October 25, 2018	1000	7.9
TLW181	Big Turkey	0.85 (Under Ice)	February 19 ,2019	1000	10.3
TLW191	Big Turkey	1.06 (Under Ice)	March 25, 2019	1000	12.7
TLW232	Big Turkey	5.25	May 23, 2019	1000	7.4
TLW277	Big Turkey	5.75	June 28, 2019	1000	0.9
TLW322	Big Turkey	5.25	July 24, 2019	1000	0.7
TLW325	Big Turkey	6.25	July 24, 2019	1000	1.6
TLW367	Big Turkey	5	August 21, 2019	1000	0.9
TLW407	Big Turkey	N/A (Under Ice)	January 23, 2020	1000	1.2
TLW28	Little Turkey	0	July 18, 2018	1000	3.3
TLW31	Little Turkey	5.25	July 18, 2018	400	2.8

Table A.2 Continued

Sample ID	Lake	Sampling Depth (m)	Sampling Date	Volume Filtered (mL)	DNA Concentration (ng/µL)
TLW36	Little Turkey	6.25	July 18, 2018	500	2.8
TLW73	Little Turkey	0	August 13, 2018	1000	4.1
TLW76	Little Turkey	6	August 13, 2018	550	5.7
TLW79	Little Turkey	7	August 13, 2018	600	4.7
TLW169	Little Turkey	3	October 25, 2018	1000	3.3
TLW179	Little Turkey	0.89 (Under Ice)	February 19, 2019	1000	4.4
TLW189	Little Turkey	1.03 (Under Ice)	March 25, 2019	1000	2.4
TLW223	Little Turkey	4	May 23, 2019	1000	6
TLW268	Little Turkey	4	June 28, 2019	1000	2.2
TLW313	Little Turkey	4	July 24, 2019	1000	1.4
TLW358	Little Turkey	5	August 21, 2019	1000	2.2
TLW409	Little Turkey	0.76 (Under Ice)	January 22, 2020	1000	1.8
TLW19	Wishart	0	July 16, 2018	950	3
TLW22	Wishart	3.25	July 16, 2018	400	3.5
TLW26	Wishart	4.25	July 16, 2018	250	2.6
TLW64	Wishart	0	August 14, 2018	600	0.8
TLW67	Wishart	3.5	August 14, 2018	450	0*
TLW70	Wishart	4.5	August 14, 2018	350	3.2
TLW167	Wishart	3	October 25, 2018	1000	2.5
TLW177	Wishart	0.87 (Under Ice)	February 19, 2019	1000	2.6
TLW187	Wishart	0.96 (Under Ice)	March 27, 2019	1000	5.5
TLW214	Wishart	3.5	May 22, 2019	1000	3.1
TLW259	Wishart	3	June 28, 2019	600	6.3
TLW304	Wishart	2.5	July 25, 2019	500	3.5

Table A.2 Continued

Sample ID	Lake	Sampling Depth (m)	Sampling Date	Volume Filtered (mL)	DNA Concentration (ng/µL)
TLW349	Wishart	3	August 21, 2019	500	5.4
TLW405	Wishart	0.76 (Under Ice)	January 22, 2020	1000	2.2

Table A.3 Sampling conditions for sample collection as prepared by Environment and Climate Change Canada and Natural Resources Canada field technicians. Conditions not provided for dates sampled are not available due to missing field data sheets.

Lake	Sampling Date	Sampling Time	Cloud Coverage (%)	Wind	Air Temperature (°C)	Ice Cover
Wishart	July 16, 2018	2 :16 P.M.	0	Light	24.3	N/A
Little Turkey	July 16, 2018	10 :00 A.M.	0	Very Light	17.5	N/A
Big Turkey	July 18, 2018	11 :15 A.M.	10	Light	23.4	N/A
Wishart	August 14, 2018	11 :00 A.M.	5	Calm	28.3	N/A
Little Turkey	August 13, 2018	12 :40 P.M.	0	Light	28.9	N/A
Big Turkey	August 13, 2018	10 :15 A.M.	0	Light	24.1	N/A
Wishart	October 24, 2018	10 :15 A.M.	10	Light	6.8	Ice cover in AM
Little Turkey	October 25, 2018	12 :55 P.M.	100	Light	4.8	N/A
Big Turkey	October 25, 2018	10 :20 A.M.	100	Light	5.4	N/A
Little Turkey	June 28, 2019	12 :11 P.M.	25	Moderate	26.4	N/A

Table A.3 Continued

Lake	Sampling Date	Sampling Time	Cloud Coverage (%)	Wind	Air Temperature (°C)	Ice Cover
Big Turkey	June 28, 2019	9 :39 A.M.	95	None	19.3	N/A
Wishart	July 25, 2019	9 :40 A.M.	0	None	21.2	N/A
Little Turkey	July 24, 2019	12 :45 P.M.	5	Light	23.1	N/A
Big Turkey	July 24, 2019	10 :17 A.M.	95%	None	20.0	N/A
Wishart	August 22, 2019	12 :30 P.M.	65	Light	19.9	N/A
Little Turkey	August 21, 2019	1 :00 P.M.	40	Moderate	21.5	N/A
Big Turkey	August 21, 2019	10 :37 A.M.	10	Moderate	21.2	N/A

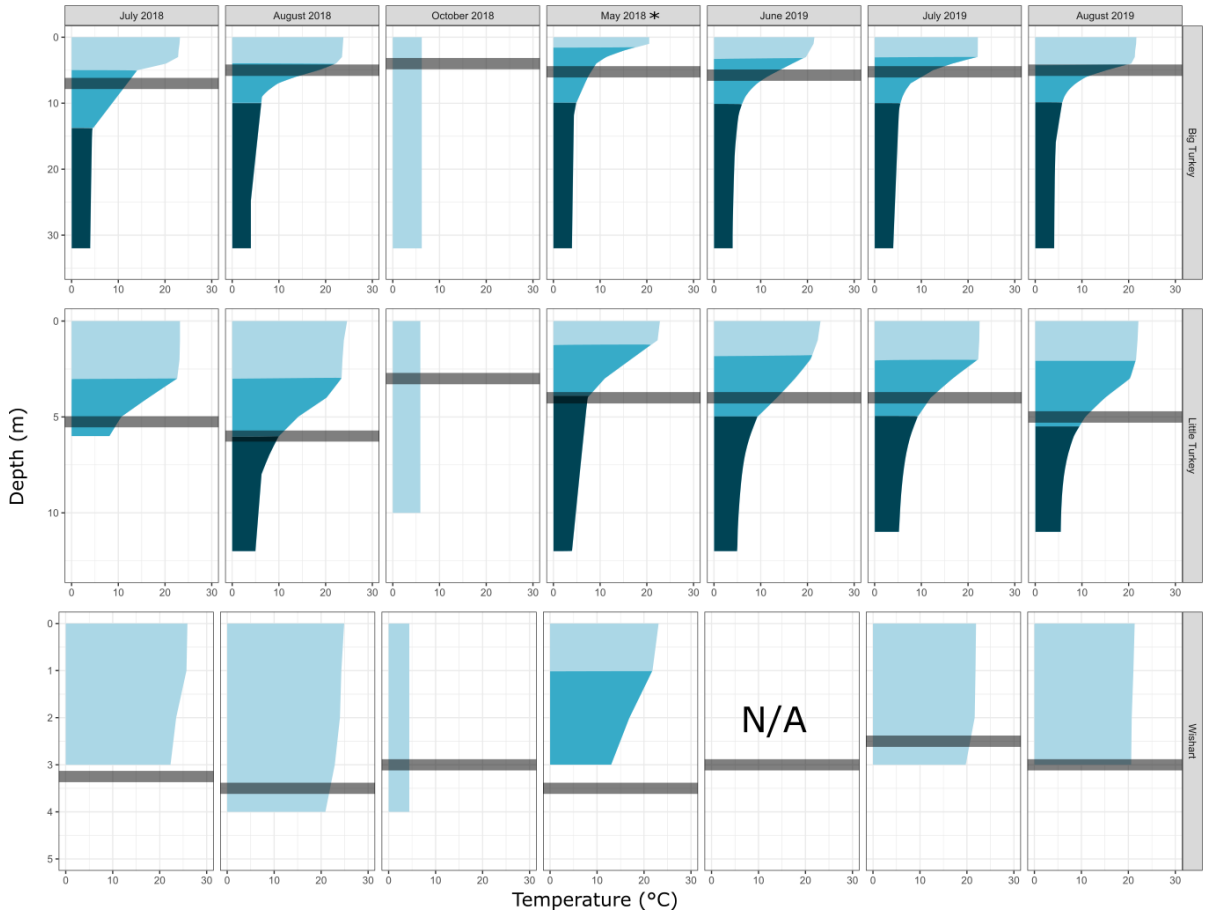


Figure A.1 Water column temperature profiles in ice-free months. Water column temperature profiles were collected during ice-months in Big Turkey (Max depth = 37 m), Little Turkey (Max depth = 13 m) and Wishart Lake (Max depth = 4.5 m). Secchi depth, which is used as a sampling depth in the studies in this thesis, is indicated with the shaded grey bar. Thermally stratified layers are identified as epilimnion (light shade), metalimnion (medium shade) and hypolimnion (dark shade). Notably, water column profiles for June 2019 in Wishart Lake were not available. May 2019 data in all three lake sites was also unavailable but to demonstrate previous thermal stratification trends, data from May 2018 have been provided.

Table A.4 Bulk water chemistry data for Wishart Lake as prepared by Environment and Climate Change Canada and Natural Resources Canada (Webster et al., 2021). Water chemistry parameters at the Turkey Lakes Watershed were monitored by Environment and Climate Change Canada and Natural Resources Canada including: pH, conductivity (Con.), alkalinity (Alk.), calcium (Ca), potassium (K), magnesium (Mg), sodium (Na), sulfate (SO_4), chloride (Cl), silicon dioxide (SiO_2), nitrite and nitrate ($NO_2 + NO_3$), ammonium (NH_4), total organic carbon (TOC), total inorganic carbon (TIC), aluminum (Al), iron (Fe), manganese (Mn), and zinc (Zn). At the time of preparation of this thesis, phosphorus and total nitrogen data was not available. Nutrient, ions and metals are presented in ppm.

Date	16-Jul-18	14-Aug-18	24-Oct-18	18-Feb-19	24-May-19	24-Jun-19	25-Jul-19	24-Aug-19
pH	6.951	7.047	6.501	6.415	6.654	6.918	6.793	6.725
Con.	22.700	23.900	17.270	24.800	17.580	18.850	20.100	21.600
Alk.	0.141	0.239	0.079	0.151	0.083	0.107	0.118	0.118
Ca	3.380	3.822	2.478	3.411	2.547	2.631	2.837	2.930
K	0.213	0.253	0.248	0.234	0.146	0.180	0.213	0.189
Mg	0.377	0.455	0.300	0.378	0.283	0.296	0.328	0.348
Na	0.584	0.690	0.468	0.551	0.452	0.483	0.511	0.535
SO_4	2.159	2.282	2.021	2.373	2.065	2.130	2.247	2.424
Cl	0.145	0.119	0.156	0.157	0.100	0.110	0.104	0.104
SiO_2	2.670	2.170	3.600	4.990	3.590	2.760	2.570	2.560
$NO_2 + NO_3$	0.113	-0.001	0.041	0.175	0.215	0.051	-0.001	-0.004
NH_4	0.020	0.002	0.019	0.088	0.013	0.009	0.004	0.002
TOC	4.993	4.433	6.554	5.071	3.431	4.095	4.214	4.055
TIC	1.777	1.962	1.068	2.373	1.251	1.374	1.586	1.564
Al	0.052	0.042	0.144	0.095	0.088	0.064	0.049	0.033
Fe	0.036	0.040	0.053	0.054	0.032	0.039	0.047	0.038
Mn	0.011	0.011	0.011	0.020	0.008	0.010	0.015	0.015
Zn	0.003	0.002	0.005	0.006	0.004	0.003	0.003	0.002

Table A.5 Bulk water chemistry data for Little Turkey Lake as prepared by Environment and Climate Change Canada and Natural Resources Canada (Webster et al., 2021). Water chemistry parameters at the Turkey Lakes Watershed were monitored by Environment and Climate Change Canada and Natural Resources Canada including: pH, conductivity (Con.), alkalinity (Alk.), calcium (Ca), potassium (K), magnesium (Mg), sodium (Na), sulfate (SO_4), chloride (Cl), silicon dioxide (SiO_2), nitrite and nitrate ($NO_2 + NO_3$), ammonium (NH_4), total organic carbon (TOC), total inorganic carbon (TIC), aluminum (Al), iron (Fe), manganese (Mn), and zinc (Zn). At the time of preparation of this thesis, phosphorus and total nitrogen data was not available. Nutrient, ions and metals are presented in ppm.

Date	16-Jul-18	14-Aug-18	24-Oct-18	18-Feb-19	24-May-19	24-Jun-19	25-Jul-19	24-Aug-19
pH	6.968	6.953	6.787	6.614	6.657	6.714	6.987	6.886
Con.	29.000	28.300	25.400	27.700	28.000	24.300	25.600	26.200
Alk.	0.193	0.186	0.153	0.178	0.164	0.139	0.158	0.170
Ca	4.455	4.769	3.954	3.990	4.096	3.783	3.776	3.832
K	0.213	0.258	0.264	0.251	0.248	0.205	0.228	0.237
Mg	0.414	0.469	0.391	0.399	0.409	0.358	0.356	0.365
Na	0.587	0.660	0.650	0.565	0.578	0.525	0.505	0.531
SO_4	2.661	2.633	2.418	2.692	2.714	2.455	2.505	2.571
Cl	0.158	0.142	0.304	0.164	0.167	0.126	0.127	0.132
SiO_2	3.620	3.370	3.420	4.090	4.330	3.880	3.410	3.190
$NO_2 + NO_3$	0.217	0.158	0.050	0.113	0.142	0.185	0.113	0.074
NH_4	0.019	0.019	0.034	0.038	0.034	0.016	0.026	0.031
TOC	3.974	3.517	5.726	5.043	4.462	3.735	3.870	3.922
TIC	2.299	2.436	2.134	2.568	2.427	2.147	2.093	2.078
Al	0.049	0.048	0.090	0.075	0.071	0.073	0.058	0.048
Fe	0.026	0.030	0.064	0.044	0.058	0.041	0.039	0.022
Mn	0.010	0.014	0.024	0.017	0.025	0.012	0.012	0.008
Zn	0.004	0.003	0.004	0.004	0.003	0.005	0.003	0.003

Table A.6 Bulk water chemistry data for Wishart Lake as prepared by Environment and Climate Change Canada and Natural Resources Canada (Webster et al., 2021). Water chemistry parameters at the Turkey Lakes Watershed were monitored by Environment and Climate Change Canada and Natural Resources Canada including: pH, conductivity (Con.), alkalinity (Alk.), calcium (Ca), potassium (K), magnesium (Mg), sodium (Na), sulfate (SO_4), chloride (Cl), silicon dioxide (SiO_2), nitrite and nitrate ($NO_2 + NO_3$), ammonium (NH_4), total organic carbon (T.O.C), total inorganic carbon (T.I.C), aluminum (Al), iron (Fe), manganese (Mn), and zinc (Zn). At the time of preparation of this thesis, phosphorus and total nitrogen data was not available. Nutrient, ions and metals are presented in ppm.

Date	16-Jul-18	14-Aug-18	24-Oct-18	18-Feb-19	24-May-19	24-Jun-19	25-Jul-19	24-Aug-19
pH	7.006	7.057	6.924	6.826	6.945	6.873	7.150	6.911
Con.	33.500	33.100	31.700	32.800	33.000	31.100	31.300	31.500
Alk.	0.227	0.261	0.210	0.221	0.210	0.199	0.200	0.203
Ca	5.312	5.610	5.088	4.935	4.957	4.956	4.719	4.820
K	0.205	0.246	0.216	0.226	0.227	0.212	0.221	0.243
Mg	0.416	0.460	0.415	0.413	0.415	0.392	0.381	0.385
Na	0.586	0.644	0.559	0.579	0.581	0.562	0.542	0.547
SO_4	2.923	2.883	2.785	2.852	2.886	2.701	2.719	2.779
Cl	0.143	0.132	0.148	0.154	0.158	0.133	0.130	0.138
SiO_2	3.660	3.320	3.400	3.760	3.840	3.740	3.550	3.450
$NO_2 + NO_3$	0.238	0.215	0.172	0.216	0.222	0.214	0.195	0.174
NH_4	0.011	0.009	0.015	0.005	0.002	0.008	0.008	0.013
TOC	3.376	3.246	3.954	3.784	3.416	3.319	3.391	3.368
TIC	2.873	3.053	3.035	3.179	3.072	2.929	2.807	2.812
Al	0.032	0.032	0.043	0.034	0.032	0.039	0.032	0.032
Fe	0.013	0.017	0.034	0.020	0.021	0.022	0.015	0.014
Mn	0.005	0.005	0.014	0.007	0.006	0.008	0.004	0.004
Zn	0.003	0.004	0.003	0.003	0.002	0.005	0.003	0.003

Table A.7 Library sizes for diurnal samples collected across a multi-time point sampling series (Chapter 3).

Sample	Lake	Depth	Day	Sampling Time	Library Size
TLW127	Little Turkey	Surface	Day 1	Morning	18076
TLW130	Little Turkey	Secchi	Day 1	Morning	7567
TLW133	Little Turkey	Surface	Day 1	Midday	16781
TLW142	Little Turkey	Secchi	Day 1	Midday	30000
TLW136	Little Turkey	Surface	Day 1	Afternoon	22768
TLW139	Little Turkey	Secchi	Day 1	Afternoon	7809
TLW145	Little Turkey	Secchi	Day 2	Morning	10471
TLW160	Little Turkey	Secchi	Day 2	Midday	29718
TLW157	Little Turkey	Secchi	Day 2	Afternoon	13333
TLW94	Wishart	Secchi	Day 1	Morning	8429
TLW106	Wishart	Secchi	Day 1	Midday	25173
TLW103	Wishart	Secchi	Day 1	Afternoon	8765
TLW112	Wishart	Secchi	Day 2	Morning	8258
TLW124	Wishart	Secchi	Day 2	Midday	31869
TLW121	Wishart	Secchi	Day 2	Afternoon	9523

Table A.8 Library sizes for long-term samples collected across a seasonal and spatial depth profile sampling series (Chapter 4). Samples marked with a * indicate example samples utilized in the development of the R package, *mirlyn* (Chapter 2).

Sample	Lake	Sampling Depth	Sampling Date	Library Size
TLW407	Big Turkey	Ice	20-Jan	10130
TLW181	Big Turkey	Ice	19-Feb	34969
TLW191	Big Turkey	Ice	19-Mar	24334
TLW232	Big Turkey	Secchi	19-May	33248
TLW277	Big Turkey	Secchi	19-Jun	17582
TLW37	Big Turkey	Surface	18-Jul	36611
TLW43	Big Turkey	Secchi+1m	18-Jul	31609
TLW42	Big Turkey	Secchi	18-Jul	45202
TLW325	Big Turkey	Secchi + 1 m	19-Jul	25916
TLW322	Big Turkey	Secchi	19-Jul	21642
TLW82	Big Turkey	Surface	18-Aug	40589
TLW88	Big Turkey	Secchi+1m	18-Aug	28478
TLW85	Big Turkey	Secchi	18-Aug	39113
TLW367	Big Turkey	Secchi	19-Aug	19881
TLW171	Big Turkey	Secchi	18-Oct	28026
TLW409	Little Turkey	Ice	20-Jan	10750

Table A.8 Continued

Sample	Lake	Sampling Depth	Sampling Date	Library Size
TLW179	Little Turkey	Ice	19-Feb	17434
TLW189	Little Turkey	Ice	19-Mar	17554
TLW223	Little Turkey	Secchi	19-May	10099
TLW268	Little Turkey	Secchi	19-Jun	21042
TLW28	Little Turkey	Surface	18-Jul	40210
TLW36	Little Turkey	Secchi+1m	18-Jul	40710
TLW31	Little Turkey	Secchi	18-Jul	32267
TLW313	Little Turkey	Secchi	19-Jul	35558
TLW73	Little Turkey	Surface	18-Aug	39842
TLW79	Little Turkey	Secchi+1m	18-Aug	44951
TLW76	Little Turkey	Secchi	18-Aug	35893
TLW358	Little Turkey	Secchi	19-Aug	21963
TLW169	Little Turkey	Secchi	18-Oct	23482
TLW405	Wishart	Ice	20-Jan	38530
TLW177*	Wishart	Ice	19-Feb	19145
TLW187	Wishart	Ice	19-Mar	10344
TLW214	Wishart	Secchi	19-May	9534
TLW259	Wishart	Secchi	19-Jun	22562
TLW19*	Wishart	Surface	18-Jul	17048
TLW22*	Wishart	Secchi+1m	18-Jul	25947
TLW22*	Wishart	Secchi	18-Jul	22037
TLW304	Wishart	Secchi	19-Jul	18252
TLW64	Wishart	Surface	18-Aug	33521
TLW70	Wishart	Secchi+1m	18-Aug	22421
TLW67	Wishart	Secchi	18-Aug	21481
TLW349*	Wishart	Secchi	19-Aug	11213
TLW167*	Wishart	Secchi	18-Oct	29615

Appendix B
Bioinformatic Processing – Example Workflows & mirlyn Functionality

B1. Example QIIME2 Workflow

Analyses in this thesis were conducted using QIIME2 v. 2019.10 (Bolyen et al., 2019). The following is an example of the workflow conducted in the analyses.

a. Data Import

Sequence files obtained from Metagenom Bio Inc. (Waterloo, ON) were demultiplexed paired end reads that included two fastq.gz files for the forward and reverse reads of each sample. These files were in the Casava 1.8 demultiplexed format.

```
qiime tools import \
     --type 'SampleData[PairedEndSequencesWithQuality]' \
     --input-path sequencefiles \
     --input-format CasavaOneEightSingeLanePerSampleDirFmt \
     --output-path demux-paired-end-sequences.qza
```

b. Quality Control & ASV Table Generation

Sequence quality control was performed using DADA2 (Callahan et al., 2016). For paired end read joining, reads must be long enough that they overlap but also removes low quality reads.

```
qiime dada2 denoise-paired \
     --i-demultiplexed-seqs demux-paired-end-sequences.qza \
     --p-trim-left-f 19 \
     --p-trim-left-r 250 \
     --p-trunc-len-f 20 \
     --p-trunc-len-r 225 \
     --o-table asv-table.qza \
     --o-representative-sequences rep-seqs.qza \
     --o-denoising-stats dada2-denoise-stats.qza
```

c. Taxonomic Classification

Taxonomic classifications in this thesis were performed using a Naïve Bayes probabilistic classifier trained with the SILVA138 reference database for the 515F and 806R V4 16S rRNA primers.

```
qiime feature-classifier classify-sklearn \
     --i-classifier SILVA138_classifier.qza \
     --i-reads rep-seqs.qza \
     --o-classification tax-file.qza
```

d. Downstream Analyses

While QIIME2 hosts a variety of functions for taxonomic community composition and diversity analyses, the analyses conducted throughout this thesis was performed using R.

B2. Example R Workflow Using *mirlyn*

a. Data Import

QIIME2R (Bisanz, 2013) was used to import .qza files into R as a *phylsoeq* (McMurdie and Holmes, 2013) objects. This generated phyloseq object was then used for community diversity analyses using *mirlyn*, an R package developed in Chapter 2 including various functions for library normalization and diversity analyses.

```
library(mirlyn)
library(phyloseq)
library(qiime2R)
phyloseq_object<-qza_to_phyloseq("asv-table.qza",
"rooted-tree.qza","tax-file.qza", "metadata.txt")
```

b. Data Handling

Amplicon sequencing datasets are large including a high number of unique amplicon sequence variants, sample metadata and taxonomic classification for amplicon sequence variants. To ease in the handling of this data, functions have been created to create compiled tables including sample metadata, ASV abundances and taxonomic classification. These data frames can be exported as a CSV file using *write.csv()*.

1. **Assigning ASV Identifiers to Sequence Variants**

   During the creation of the ASV table in QIIME2, each unique sequence is assigned an identifier consisting of a string of characters (e.g., 88b44c11059bcf2950ca0ac50f3eb08f). To improve readability, the *asv_rename()* function codes these character string identifiers to a new identifier in the form "ASV###". While this step is not mandatory, it allows for easy reference to specific ASVs of interest.

```
asv_rename(example)
```

2. **Generation of Data Frame from *phyloseq* Object**

   Data is initially imported into R as a *phyloseq* object. The *phyloseq* object is critical for subsequently performing diversity analysis. However, for plotting options or

subsequent export as a CSV file, the *phyloseq_to_df()* function will convert the *phyloseq* object to a data frame containing the ASV counts, taxonomic classification and metadata.

```
example_df <- phyloseq_to_df(example)
```

3. Generation of Compiled ASV Table

The data frame generated using *phyloseq_to_df()* can be further organized to focus on the read counts of each ASV across the different samples. The *get_asv_table()* also includes the taxonomic classification of the ASV but does not include sample metadata.

```
example_df_asv <- get_asv_table(example_df)
```

a. Taxonomic Composition

1. Visualization

mirlyn provides two visualization options for taxonomic communities including stacked bar plots and heatmaps. Heatmaps are optimal to use when interested in exploring the trends in relative abundances of one taxonomic group (e.g., Cyanobacteria). Alternatively, the stacked bar charts can be used to identify overall composition of communities.

```
# Stacked Barcharts at the Phylum Rank
cols <- c("black", "darkgoldenrod1", "dodgerblue", "deeppink4",
"chartreuse3", "burlywood4", "navy", "blueviolet", "tan2",
"lavenderblush3", "cyan4")

example_barchart <- bartax(example, "Sample", taxrank =
"Phylum", cols = cols)

# Heatmap of Cyanobacterial Abundances in the Bacterial
Community
example_df_phylum <- example_df %>% group_by(Sample, Id,
Phylum) %>% summarise(abaundance = (sum(abundance)) %>%
mutate(Proportion = abundance/sum(abundance)*100)

plot_heat(example_df_phylum, taxlevel = "Phylum", taxaname =
"Cyanobacteria", xvar = "sample", yvar = "Id", fillvar =
"Proportion")+scae_fill_gradient(low = "white", high =
"midnightblue")
```

2. Compositional Significance Testing

Amplicon sequencing data is inherently compositional (Gloor et al., 2016). The composition of these communities is reported in relative abundance but raises the question of when is a group statistically abundant within the community. The *randomseqsig()* function will identify whether a taxonomic group of interest is significantly dominant in the community. This can be used to identify conditions where a taxonomic group of interest (e.g., Cyanobacteria) are in significantly higher abundances.

```
# Calculate significance of Phylum: Cyanobacteria
compsig_example <- randomseqsig(example, taxlevel = "Phylum",
group = "Cyanobacteria", nshuff = 1000)
```

b. Diversity Analyses

Prior to conducting diversity analysis, libraries must be normalized to account for variation in library sizes. A variety of techniques are available each with their own benefits and limitations and researchers are encouraged to evaluate the effectiveness of these techniques for their data. However, for this research, *mirlyn* utilizes repeated iterations of rarefying, the process of subsampling to a user specified library size.

1. Library Normalization

To identify appropriate library sizes to rarefy to, a rarefaction curve can be generated to provide an overview of the observed ASV in samples corresponding to different rarefied library sizes. Theoretically, samples that have a plateau in the curve have reached maximal observed diversity. This visualization should be used to select an appropriate library size which encompasses maximal diversity while being inclusive of samples.

```
# Creation of rarefaction curve data frame
Rarefy_whole_rep_example <- rarefy_whole_rep(example,
rep = 100)

#Visualization of rarefaction curve
Rarecurve_ex <- rarecurve(rarefy_whole_rep_ex,
sample = "Sample")
```

2. Multiple Iterations of Rarefying Libraries

After generating rarefaction curves, users may select an appropriate rarefied library size for their analysis. Users should aim to select a library size that represents maximal diversity and is inclusive of all samples. In the case where users must make the decision between losing samples or drastically reducing the represented diversity, users may opt to conduct analyses at the lower library size inclusive of all samples at the loss of diversity in some samples in addition to a larger rarefied library size which results in exclusion of small library size samples. Depending on the data structure, users may choose to include a different number of repeated iterations. For example, if the repeated iterations do not result in highly variable outputs in the diversity analyses, the number of iterations may be reduced. However, if large variation is present, users should aim to include a larger number of iterations to allow for better characterization of variation introduced through random subsampling. The mirl_object will be used in the subsequent analyses.

```
# Creation of mirl object - Repeatedly rarefy 100 times

mirl_object <- mirl(example, libsize = 10000, rep = 100,
set.seed = 120)
```

3. Alpha Diversity

mirlyn contains two visualization options for alpha-diversity analyses. Both implement the use of a diversity metric (e.g., Shannon diversity index). The *alphadivDF()* function utilizes the *mirl_object* generated in the previous step and is only applicable to the diversity metric at the specified library size used with *mirl()*. The *alphacone()* function generates a distribution of the diversity metric across different rarefied library sizes providing users with a comprehensive view of the diversity metric as a function of rarefied library size.

```
# Alphawich Functions
# Generates dataframe of alpha-diversity metric from mirl_object
alphadiv_df <- alphadivDF(mirl_object)

# Generates visualization from alphadiv_df. Substitute xvar for
your own metadata column.
alphawichVis(example, xvar = "Sample")

#Alphacone Functions
# Load example data from mirlyn.
data(example)

# Generates dataframe of alpha-diversity metric across all
library sizes.
alphacone_example <- alphacone(example, rep = 100)

# Generates distribution plot of alpha-diversity metric across
library sizes.
alphaconeVis(alphacone_example, "Sample")
```

4. Beta-Diversity

Currently, mirlyn only supports the use of PCA for beta-diversity analyses. Future ordination techniques such as PCoA and NMDS may be implemented in future versions. A Hellinger transformation is recommended to apply to sequence count data prior to conducting PCA to account for the arch-effect regularly seen in ecological data. The beta-diversity functions utilize the *mirl_object* generated previously.

```
# Generation of PCA object
betamatPCA_object <- betamatPCA(mirl_object)

# Ordination Visualization
betamatPCAvis(betamatPCA_object, groups = c("A", "B", "C",
"D","E","F"), reps = 10, colours = c("#000000", "#E69F00",
"#0072B2", "#009E73", "#F0E442", "#D55E00"))
```

Table B.1 Functions from other R packages used in mirlyn. *mirlyn* is an R package developed for library normalization and diversity analyses of amplicon sequencing and is available at www.github.com/escamero/mirlyn.

***mirlyn* function**	**Description**	**Functions used from other packages**	**Citations**
bartax()	Generate taxonomic composition barcharts from taxonomic abundance data.	microbiome::transform(): generate compositional data from abundance data.	Leo Lahti et al. microbiome R package. URL: http://microbiome.github.io
		phyloseq::tax_glom(): combine compositional data by desired taxonomic level. ggplot2::ggplot(): plotting engine for all visualization.	Paul J. McMurdie and Susan Holmes (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4):e61217.
alphawhichDF()	Calculate alpha diversity values from sequence count data.	vegan::diversity(): used for alpha diversity calculation.	Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan

Table B.1 Continued

mirlyn function	Description	Functions used from other packages	Citations
alphacone()	Calculate alpha diversity values at different increments of library rarefaction for sequence data.	phyloseq::rarefy_even_depth(): rarefy taxonomic abundance data from multiple samples to an equal depth.	Paul J. McMurdie and Susan Holmes (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4):e61217.
		vegan::diversity(): used for alpha diversity calculation.	Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan
betamatPCA()	Principle component analysis of beta diversity values calculated from sequence count data.	vegan::vegdist(): calculate dissimilarity indices from taxonomic abundance data.	Jari Oksanen, F. Guillaume Blanchet, Michael Friendly, Roeland Kindt, Pierre Legendre, Dan McGlinn, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs and Helene Wagner (2019). vegan: Community Ecology Package. R package version 2.5-6. https://CRAN.R-project.org/package=vegan
		vegan::decostand(): apply desired standardization to taxonomic abundance data prior to calculation of dissimilarity indices.	
betamatPCA()		stats::prcomp(): principle component analysis.	R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
mirl()	Repeated rarefaction of	phyloseq::rarefy_even_depth(): rarefy	Paul J. McMurdie and Susan Holmes (2013). phyloseq: An R

Table B.1 Continued

mirlyn function	Description	Functions used from other packages	Citations
	sequence count data.	taxonomic abundance data from multiple samples to an equal depth.	package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4):e61217.
rarefy_whole_rep()	Repeated rarefaction of taxonomic abundance data at incremental library sizes.	phyloseq::rarefy_even_depth(): rarefy taxonomic abundance data from multiple samples to an equal depth.	Paul J. McMurdie and Susan Holmes (2013). phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. PLoS ONE 8(4):e61217.
rarecurve()	Visualize observed ASV count from rarefy_whole_rep() output.	ggplot2::ggplot(): plotting engine for all visualization.	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
alphawhichVis()	Visualize alpha diversity results from alphawhichDF().	ggplot2::ggplot(): plotting engine for all visualization.	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
betamatPCAvis() betamatPCAvis()	Plot principle component analysis from betamatPCA().	ggplot2::ggplot(): plotting engine for all visualization.	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
		factoextra::fviz_pca_ind(): wrapper for ggplot2 visualization of PCA data.	Alboukadel Kassambara and Fabian Mundt (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7. https://CRAN.R-project.org/package=factoextra
phyloseqtodf()	Creation of a dataframe from a *phyloseq* object including taxonomy, ASV read	tidyr::gather(): gathers columns for data frame reorganization.	Hadley Wickham and Lionel Henry (2020). tidyr: Tidy Messy Data. R package version 1.1.0. https://CRAN.R-project.org/package=tidyr

Table B.1 Continued

mirlyn function	Description	Functions used from other packages	Citations
	counts and metadata.		
get_asv_table()	Generates a compiled ASV table with counts and taxonomic classification of individual amplicon sequence variants.	dplyr::mutate_if(): apply transformation to variables dplyr::distinct(): subsets unique rows from dataframe.	Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2020). dplyr: A Grammar of Data Manipulation. R package version 1.0.0. https://CRAN.R-project.org/package=dplyr
randomseqsig()	Identification of whether a taxonomic group is significantly dominant in the community using data shuffling.	reshape2::melt(): generates molten dataframe.	Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL: http://www.jstatsoft.org/v21/i12/.
plot_heat()	Generates heat maps visualizing the relative abundance of a taxonomic group of interest from a dataframe.	ggplot2::ggplot(): plotting engine for all visualization.	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
asv_rename()	Assigns unique ASV I.D.'s to sequence variants.	readr::write_tsv(): generates tab delimited file from data frame object.	Hadley Wickham and Jim Hester (2020). readr: Read Rectangular Text Data. R package version 1.4.0. https://CRAN.R-project.org/package=readr
fasta_rename()	Assigns correspondin g unique ASV	Biostrings::readDNAStringSet(): reads FASTA format file.	H. Pagès, P. Aboyoun, R. Gentleman and S. DebRoy (2020). Biostrings: Efficient manipulation of biological

Table B.1 Continued

***mirlyn* function**	**Description**	**Functions used from other packages**	**Citations**
	identifiers to a FASTA file.	Biostrings::writeXStringSet(): generates FASTA file. readr::read_tsv(): reads tab delimited file into data frame.	strings. R package version 2.58.0. https://bioconductor.org/packages/Biostrings
fullbartax()	Generates taxonomic composition bar charts for specified taxonomic levels.	ggplot2::ggplot(): plotting engine for all visualization.	H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Appendix C
Supplementary Materials for Chapter 3

Table C.1 Relative abundances of major bacterial phyla rounded to two decimal points across a diurnal multi-time point sampling series. Phyla present at less than 1% abundance across all samples were excluded from this table.

Classified Phylum	Day 1			Day2		Sampling Time
	Little Turkey		Wishart	Little Turkey	Wishart	
	Secchi	Surface	Secchi	Secchi	Secchi	
Actinobacteriota	16.20	17.94	7.79	22.80	8.40	Morning
	11.02	17.03	9.87	13.67	10.69	Midday
	11.63	13.95	15.74	16.3	8.42	Afternoon
Bacteroidota	12.23	13.95	14.9	9.71	13.42	Morning
	17.27	14.89	12.43	6.86	12.01	Midday
	13.65	16.94	9.22	6.82	12.96	Afternoon
Bdellovibrionota	0.26	0.077	0.94	0.98	0.92	Morning
	0.78	0.16	1.12	0.79	0.78	Midday
	1.41	0.18	1.07	0.71	1.61	Afternoon
Cyanobacteria	9.91	6.67	4.39	13.42	10.34	Morning
	10.74	9.30	5.47	16.28	8.47	Midday
	14.62	6.97	7.31	17.53	6.46	Afternoon
Myxococcota	0.21	0.11	0.87	0.38	0.54	Morning
	1.45	0.23	0.85	0.88	0.66	Midday
	0.56	0.079	0.55	0.12	0.63	Afternoon
Planctomycetota	4.31	3.54	2.07	5.76	6.10	Morning
	4.19	4.45	2.34	5.93	3.41	Midday
	6.13	3.12	4.43	6.58	3.03	Afternoon
Proteobacteria	49.42	49.54	50.32	40.70	45.25	Morning
	40.83	47.16	54.34	41.40	51.18	Midday
	38.94	50.79	49.86	42.20	53.49	Afternoon

Table C.1 Continued

Classified Phylum	Day 1			Day 2		Sampling Time
	Little Turkey		Wishart	Little Turkey	Wishart	
	Secchi	Surface	Secchi	Secchi	Secchi	
Verrucomicrobiota	6.10	7.19	17.37	5.55	12.92	Morning
	12.85	5.31	11.36	12.77	11.03	Midday
	12.27	6.90	9.40	8.97	11.00	Afternoon

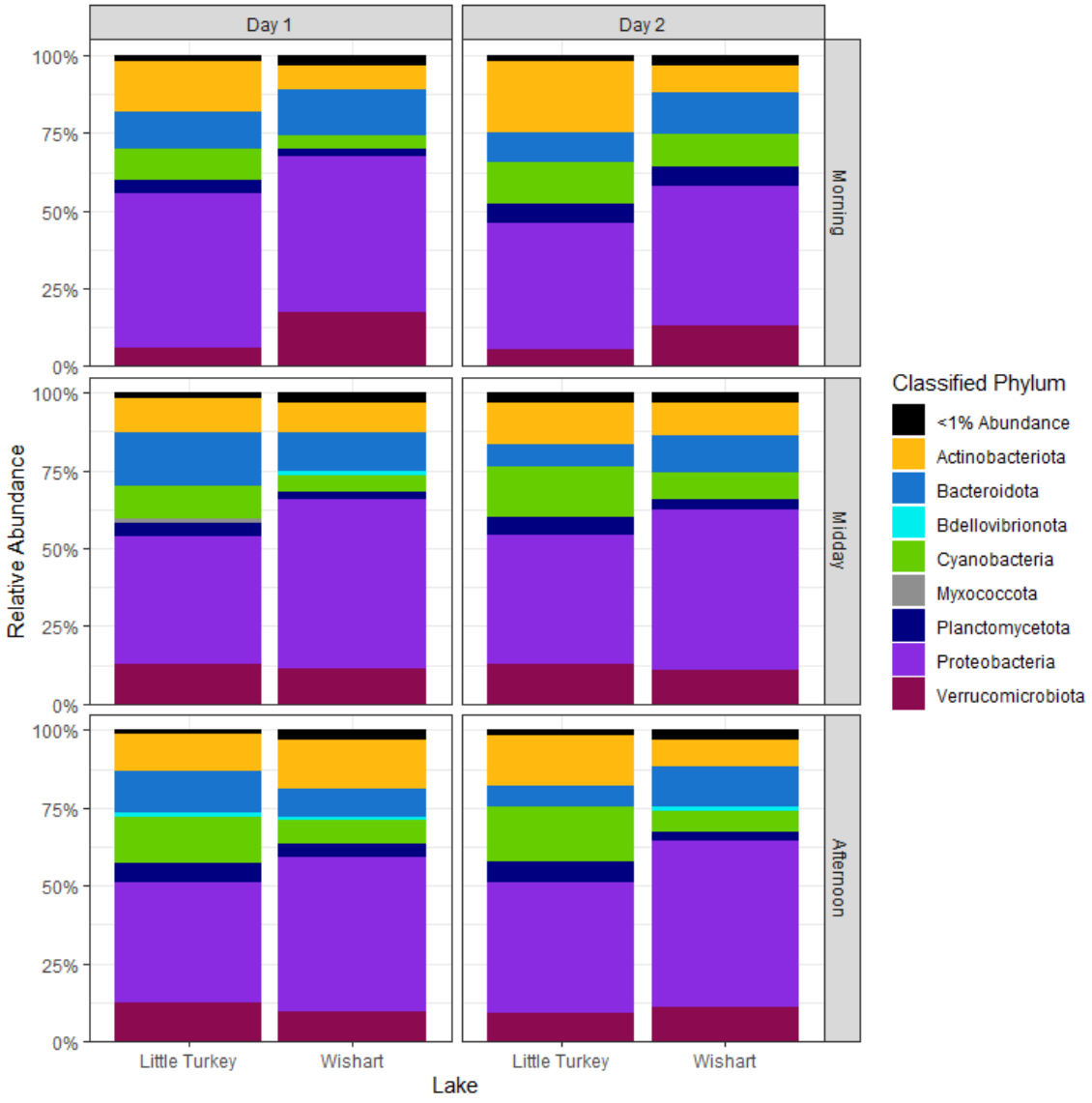


Figure C.1 Stacked bar charts depicting the relative abundances of major bacterial phyla identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a multi-time point sampling series in a non-stratified lake (Wishart) and a mid-sized stratified lake (Little Turkey).

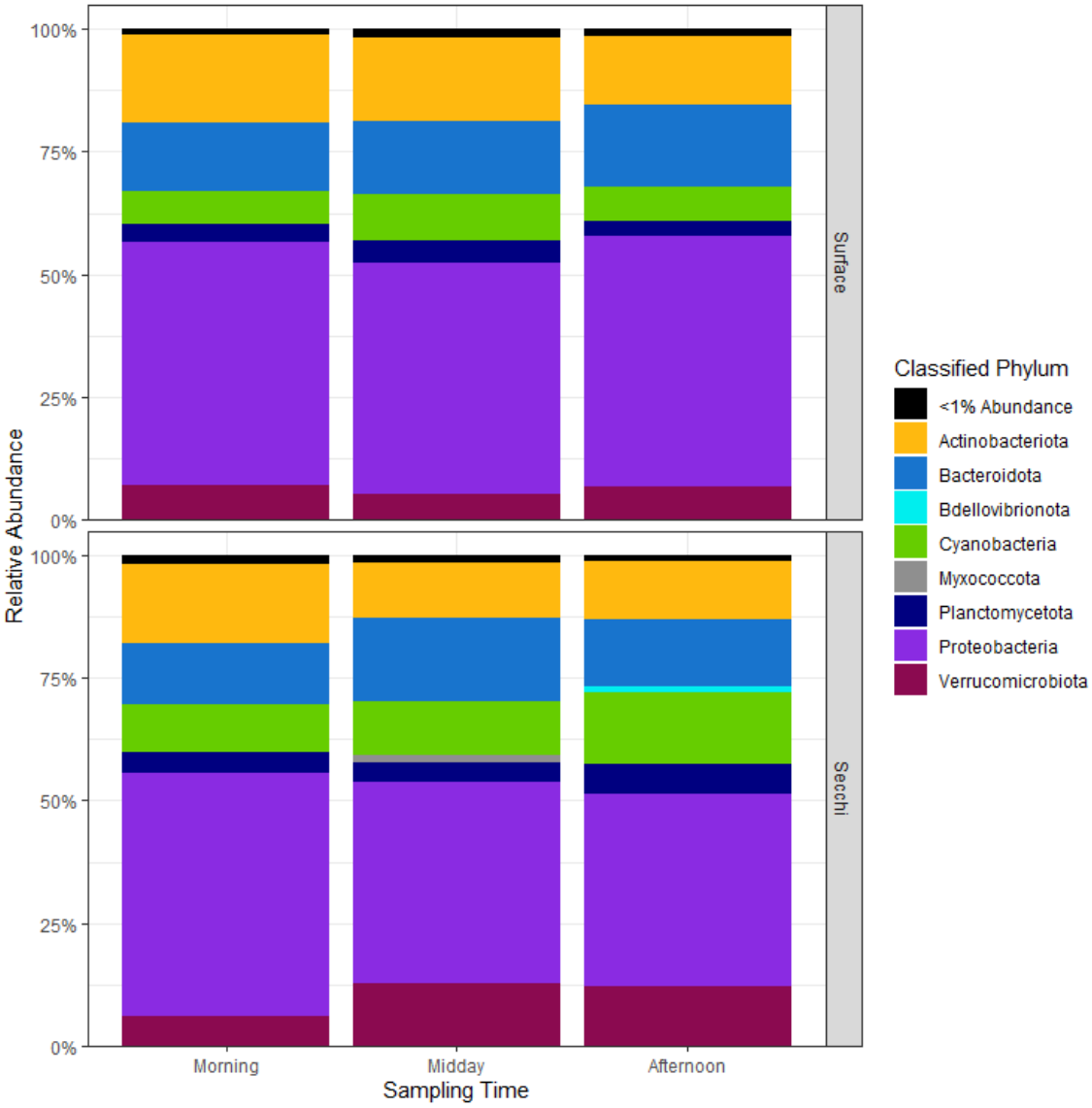


Figure C. 2 Stacked bar charts depicting the relative abundances of major bacterial phyla identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a multi-time point sampling series between two sampling depths in a stratified lake (Little Turkey).

Table C.2 Relative abundances of cyanobacterial families rounded to two decimal points. Phyla present at less than 1% abundance were excluded from this table.

Family	Day 1			Day2		Sampling Time
	Little Turkey		Wishart	Little Turkey	Wishart	
	Secchi	Surface	Secchi	Secchi	Secchi	
Aphanizomenonaceae	1.33	1.82	-	-	-	Morning
	-	1.09	-	-	-	Midday
	-	2.21	-	-	-	Afternoon
Chroococcaceae	12.00	7.13	-	2.85	1.41	Morning
	2.20	9.35	1.09	3.93	1.04	Midday
	4.99	8.58	-	5.52	1.79	Afternoon
Coelosphaeriaceae	2.93	1.24	1.35	3.20	1.29	Morning
	2.36	1.92	-	2.32	-	Midday
	4.29	3.34	-	2.82	-	Afternoon
Cyanothecaceae	2.13	1.08	-	1.42	-	Morning
	1.24	-	-	1.61	-	Midday
	1.49	-	-	1.71	-	Afternoon
Merismopediaceae	1.07	1.41	-	-	-	Morning
	-	1.09	-	-	-	Midday
	-	-	-	1.11	-	Afternoon
Microcystaceae	15.33	16.25	1.08	7.62	-	Morning
	4.41	13.26		4.01	1.55	Midday
	5.52	10.66	1.09	5.18	1.79	Afternoon
Pseudanabaenaceae	-	-	4.32	-	2.58	Morning
	-	-	14.37	-	1.85	Midday
	-	-	2.34	-	8.46	Afternoon
Synechococcaceae	61.87	66.42	87.57	82.35	87.70	Morning
	88.05	70.47	78.74	85.86	87.82	Midday
	81.44	71.37	87.52	81.60	82.11	Afternoon
Unknown	3.07	4.64	5.14	1.92	6.09	Morning
	1.02	2.11	3.48	1.57	5.70	Midday
	1.58	2.59	8.27	1.71	5.85	Afternoon

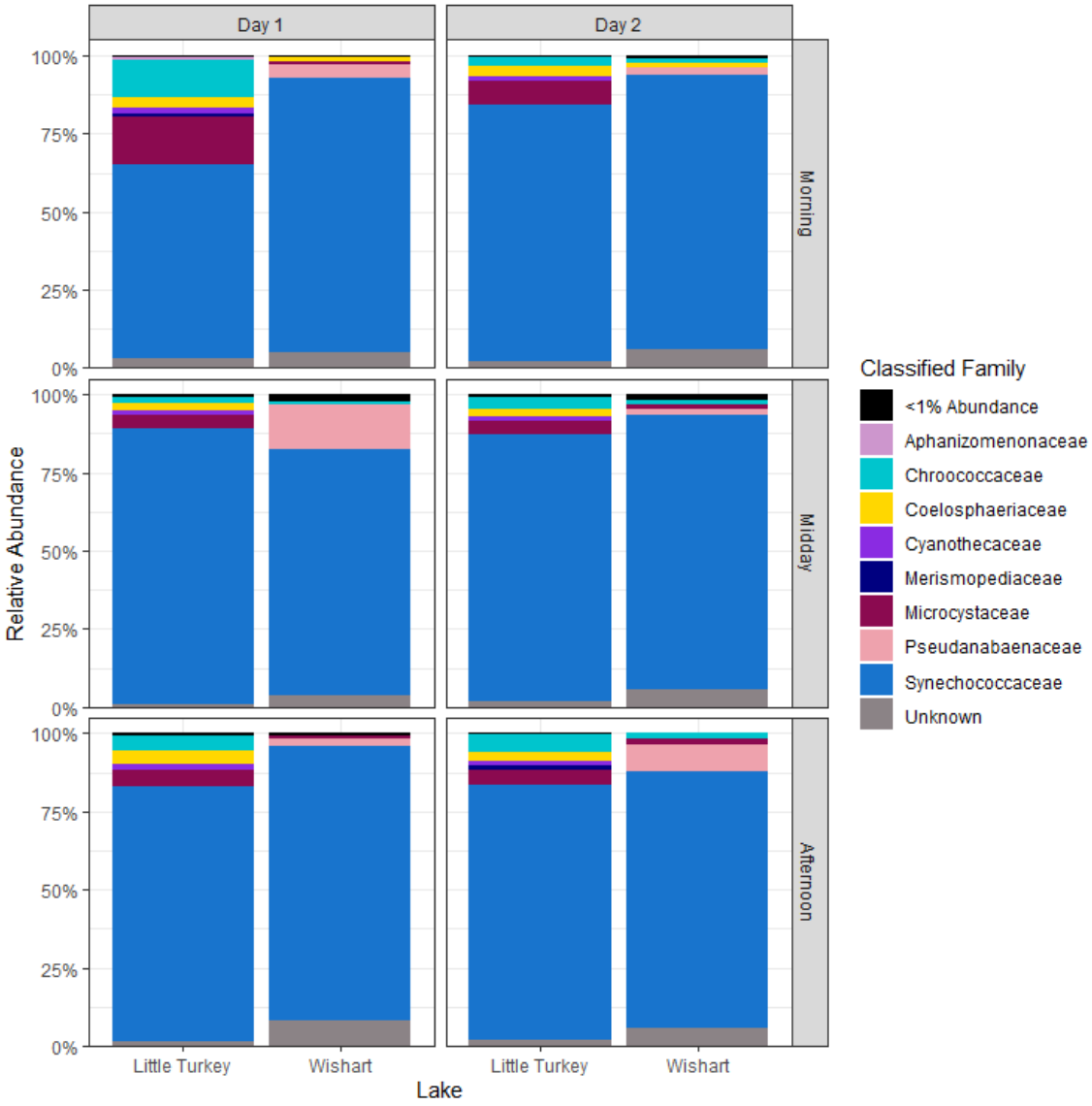


Figure C.3 Stacked bar charts depicting the relative abundances of cyanobacterial families comprising the cyanobacterial communities identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a multi-time point sampling series in a non-stratified lake (Wishart) and a mid-sized stratified lake (Little Turkey).

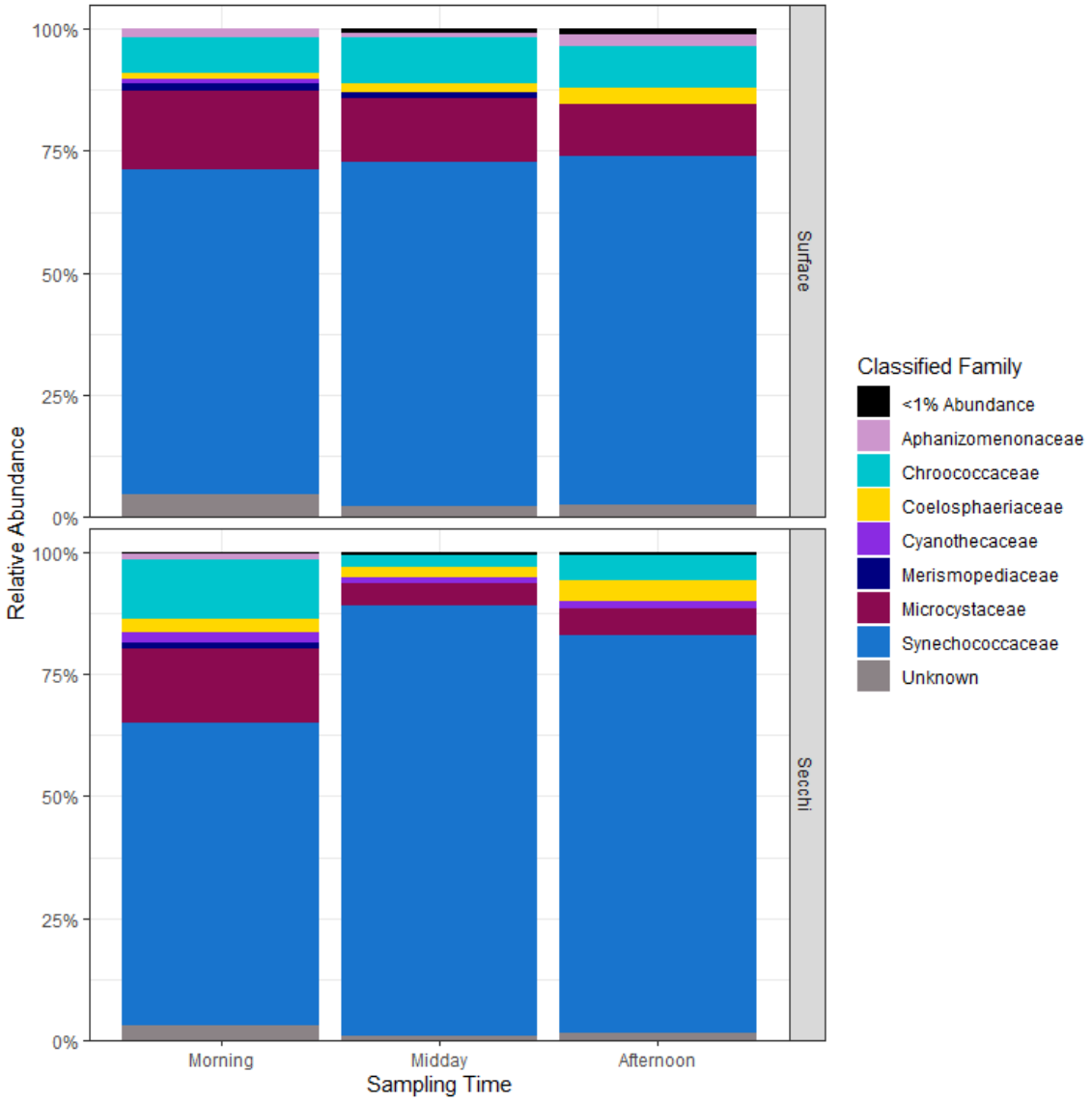


Figure C.4 Stacked bar charts depicting the relative abundances of cyanobacterial families comprising the cyanobacterial communities identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a multi-time point sampling series between two sampling depths in a stratified lake (Little Turkey).

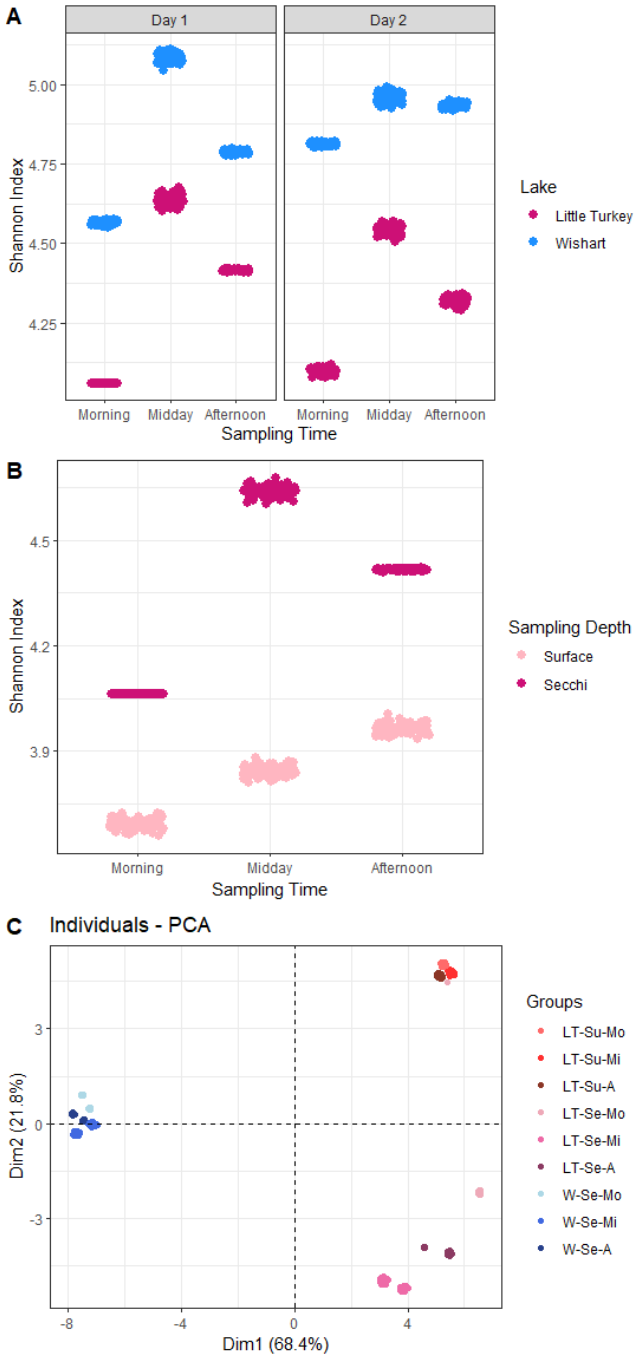


Figure C.5 Alpha and beta-diversity analyses of bacterial communities across a diurnal sampling series in a stratified and non-stratified lake. Amplicon sequence variants classified to the phylum Cyanobacteria were filtered to characterize diversity within cyanobacterial communities. The Shannon Index was calculated on rarefied libraries to evaluate the effects of (A) sampling time and lake, (B) sampling time and depth and (C) The Bray-Curtis dissimilarity metric was used to explore similarities in communities between lake site (Little Turkey = LT, Wishart = W).


% Bootstrap Support
90 - 100
89 - 80
79 - 70
Nostocales Clade
ASV819 ASV822 ASV824
Synechococcaceae Clade
ASV838 ASV839 ASV842
ASV843 ASV846 ASV848
ASV849 ASV850 ASV853
ASV855 ASV857 ASV858
ASV859 ASV865 ASV866
ASV869 ASV921
Radiocystis Clade
ASV919
Microcystis Clade
ASV913 ASV914
Pseudanabaenaceae Clade
ASV806
0.10

Figure C.6 Phylogenetic tree indicating the taxonomic affiliation of amplicon sequence variants classified as phototrophic cyanobacteria inferred using the General Time Reversible substitution model with Gamma invariant sites included. 1000 boot strap replicates were performed to estimate support for clades. Reference sequences for common freshwater cyanobacteria were included to confirm the taxonomic classification of the Naïve-Bayes classifier. Genus and species names were not modified from the reference entries in NCBI and may not reflect current accepted nomenclature. For the purpose of this research, taxonomic classifications were only examined at higher taxonomic levels of order and family.

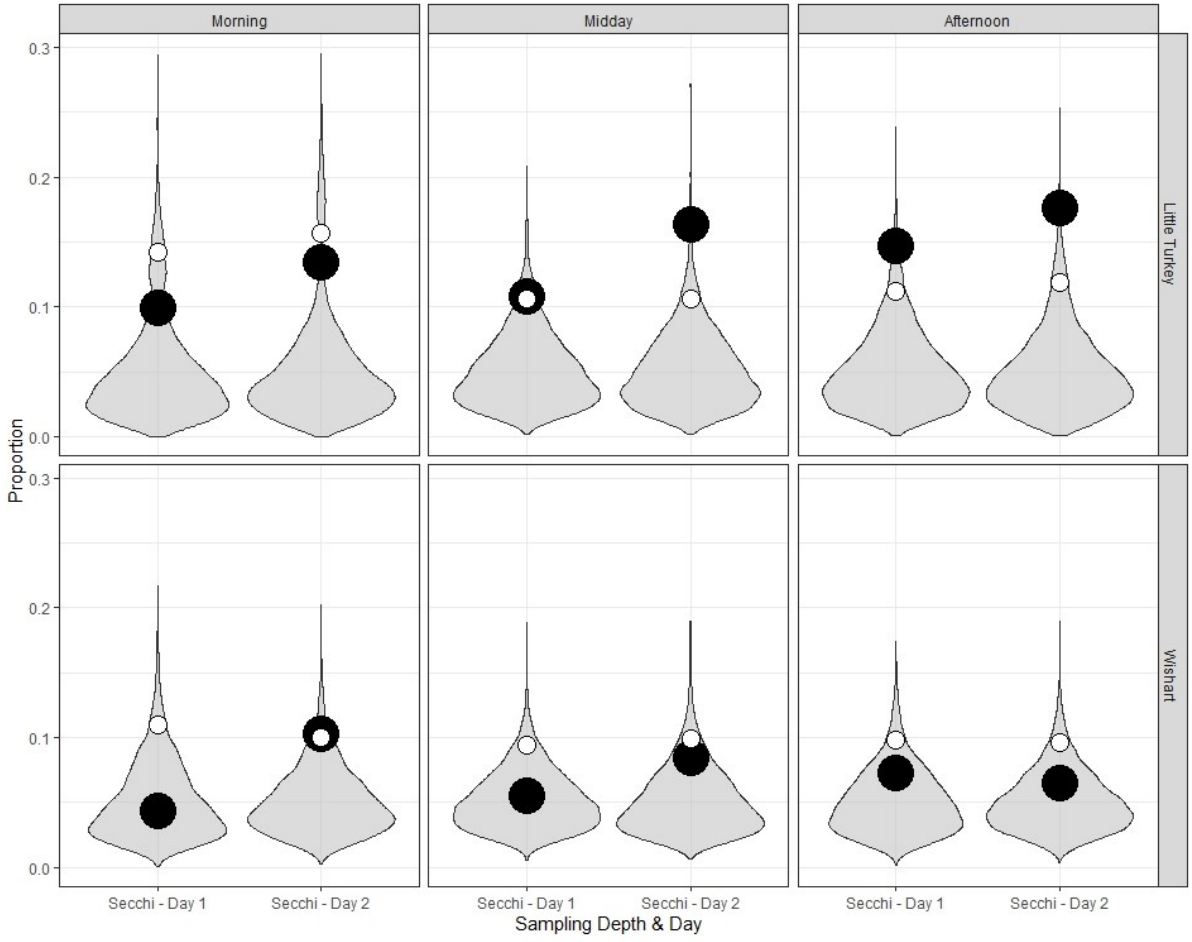


Figure C.7 Violin plots visualizing the distribution of randomized total phylum counts and the specific relative abundance of cyanobacteria relative to the 95th percentile across sampling times at Secchi depth. The observed cyanobacterial values are visualized with the solid black circle, and the

95th percentile with the white circle. Black dots present above the white circle represents an enrichment of cyanobacteria present in the bacterial community.

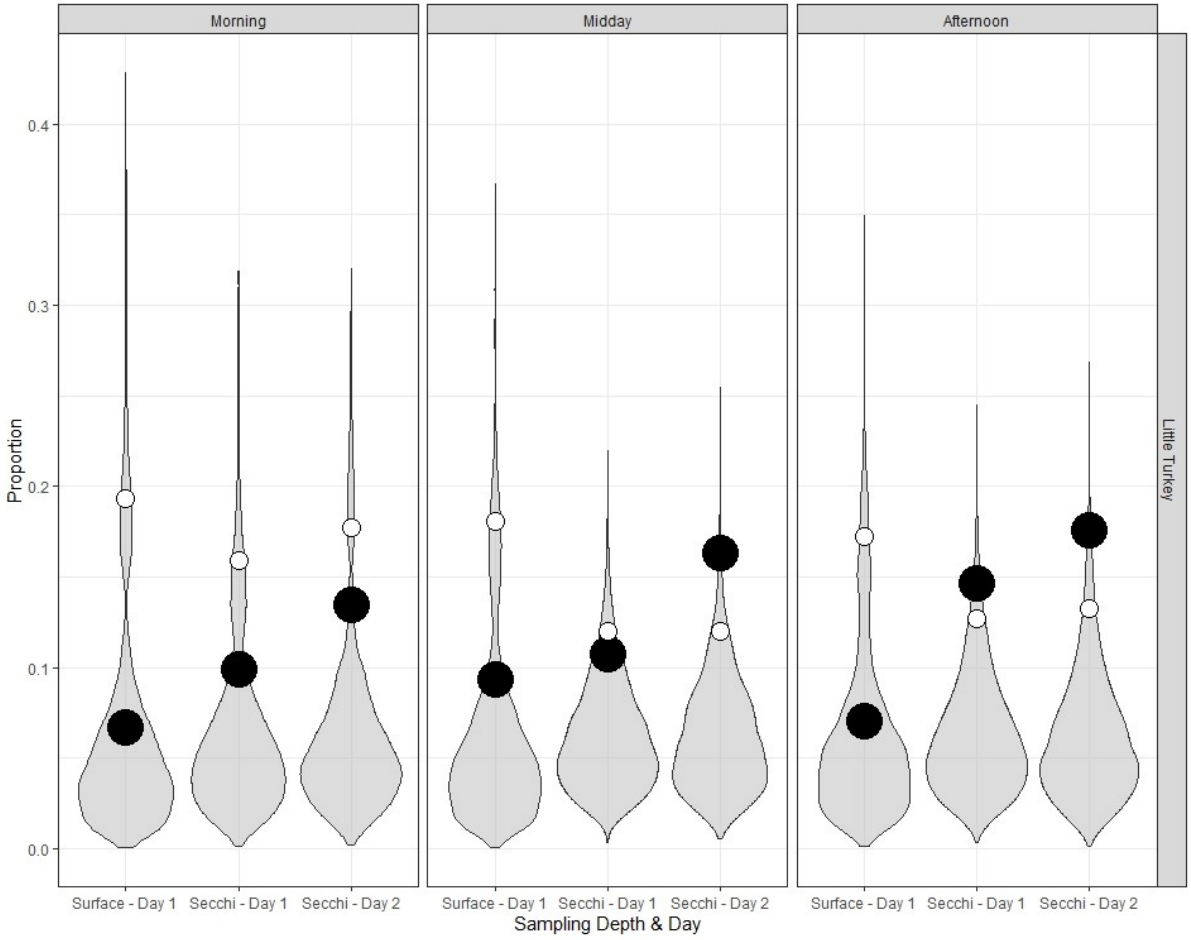


Figure C.8 Violin plots visualizing the distribution of randomized total phylum counts and the specific relative abundance of cyanobacteria relative to the 95th percentile across sampling times at Secchi depth and the water surface in Little Turkey Lake. The observed cyanobacterial values are visualized with the solid black circle, and the 95th percentile with the white circle. Black dots present above the white circle represents an enriched abundance of cyanobacteria present in the bacterial community.

Table C.3 Probability values for cyanobacterial enrichment within the randomized total phylum counts per sample. The mean and standard deviation of the randomized counts were used to calculate Z-scores. From the Z-scores, probability values were obtained and adjusted with a Bonferroni correction.

Lake	Sampling Time	Sampling Day	Mean	Standard Deviation	Observed	p-value	Adjusted p-value
Little Turkey	Morning	Day 1	0.05	0.04	0.10	0.13	1.00
Little Turkey	Morning	Day2	0.05	0.04	0.13	0.02	0.28
Little Turkey	Midday	Day 1	0.05	0.03	0.11	0.02	0.28
Little Turkey	Midday	Day 2	0.05	0.03	0.16	0.00	0.00
Little Turkey	Afternoon	Day 2	0.05	0.03	0.18	0.00	0.00
Little Turkey	Afternoon	Day1	0.05	0.03	0.15	0.00	0.01
Wishart	Morning	Day 1	0.05	0.03	0.04	0.40	1.00
Wishart	Morning	Day 2	0.05	0.03	0.10	0.02	0.25
Wishart	Midday	Day 1	0.05	0.02	0.05	0.44	1.00
Wishart	Midday	Day 2	0.05	0.03	0.08	0.10	1.00
Wishart	Afternoon	Day 1	0.05	0.03	0.07	0.20	1.00
Wishart	Afternoon	Day 2	0.05	0.02	0.06	0.29	1.00
Little Turkey (Surface)	Morning	Day 1	0.067	0.059	0.067	0.50	1.00
Little Turkey (Surface)	Midday	Day 1	0.067	0.053	0.094	0.31	1
Little Turkey (Surface)	Afternoon	Day 1	0.067	0.049	0.07	0.48	1

Table C.4 Relative abundances of ASVs in the cyanobacterial community across a diurnal multi-time point sampling series in a stratified lake (Little Turkey = LT) and non-stratified lake (Wishart = W). Samples in Little Turkey Lake were collected at surface (Su) and Secchi depth (Se). Samples in Wishart Lake were only collected at Secchi depth. Sampling was performed across a multi time point sampling series with sampling at 8:30-9:30 A.M. (Mo), 12:30-1:30 P.M. (Mid), and 4:30-5:30 P.M. (Aft).

ASV ID	LT-Su-Mo1	LT-Su-Mid1	LT-Su-Aft1	LT-Se-Mor1	LT-Se-Mid1	LT-Se-Aft1	LT-Se-Mor2	LT-Se-Mid2	LT-Se-Aft2	W-Mor1	W-Mid1	W-Aft1	W-Mor2	W-Mid2	W-Aft2
Chroococcales															
ASV884	0.41	1.02	0.88	1.47	0.40	0.61	0.99	0.85	0.68	0.00	0.00	0.00	0.00	0.00	0.00
ASV887	6.02	7.76	7.34	10.53	1.02	2.79	1.84	1.79	3.71	0.00	0.00	0.00	0.00	0.00	0.00
ASV890	0.00	0.70	0.44	0.00	1.27	1.92	1.84	1.71	2.05	1.35	0.58	0.00	1.29	0.56	0.00
ASV893	1.24	1.21	2.88	2.93	1.08	2.36	1.35	0.60	0.77	0.00	0.00	0.00	0.00	0.00	0.00
ASV904	1.07	0.32	0.00	0.00	0.00	0.00	0.64	0.00	0.00	0.00	0.00	0.78	0.00	0.30	0.00
ASV905	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV906	0.33	0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV908	0.00	0.00	0.00	1.07	0.00	0.00	0.00	0.70	1.11	0.00	0.00	0.00	0.00	0.00	0.00
ASV910	0.66	0.51	0.31	0.00	0.77	1.57	0.00	1.28	1.11	0.00	1.09	0.00	1.41	1.04	1.79
ASV913	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.85	0.00
ASV914	3.30	3.43	3.95	1.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.09	0.00	0.00	0.00
ASV915	0.00	0.00	0.00	0.00	0.00	1.48	0.00	0.31	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV916	0.00	0.00	0.00	0.00	0.00	0.00	1.42	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV917	1.07	0.70	0.88	2.13	1.24	0.00	0.00	0.00	1.70	0.00	0.80	0.00	0.00	0.59	0.00

Table C.4 Continued

ASV ID	LT-Su-Mo1	LT-Su-Mid1	LT-Su-Aft1	LT-Se-Mor1	LT-Se-Mid1	LT-Se-Aft1	LT-Se-Mor2	LT-Se-Mid2	LT-Se-Aft2	W-Mor1	W-Mid1	W-Aft1	W-Mor2	W-Mid2	W-Aft2
ASV919	12.87	9.73	6.65	14.13	4.39	5.50	7.58	4.00	5.16	1.08	0.72	0.00	0.00	0.70	1.79
Gloeobacterales															
ASV929	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.00	0.00	0.94	0.22	0.00
ASV930	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.15	0.00
ASV931	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Nostocales															
ASV819	0.00	0.00	0.00	1.33	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00
ASV822	0.00	0.00	0.00	0.00	0.71	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00
ASV824	1.82	1.08	2.19	0.00	0.00	0.70	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Oscillatoriales															
ASV831	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV880	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00
Synechococcales															
ASV806	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.32	14.34	2.34	2.58	1.85	8.13
ASV829	0.00	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ASV832	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33
ASV838	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.61	0.00	0.00	4.18	0.00
ASV839	0.00	0.00	0.00	0.00	4.64	0.79	0.00	5.53	1.70	1.89	3.55	2.81	1.41	3.44	1.79
ASV842	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.61	2.65	2.81	1.70	5.53

Table C.4 Continued

ASV ID	LT-Su-Mo1	LT-Su-Mid1	LT-Su-Aft1	LT-Se-Mor1	LT-Se-Mid1	LT-Se-Aft1	LT-Se-Mor2	LT-Se-Mid2	LT-Se-Aft2	W-Mor1	W-Mid1	W-Aft1	W-Mor2	W-Mid2	W-Aft2
ASV843	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	5.95	6.01	9.36	7.14	6.37	8.62
ASV846	5.45	4.64	5.58	9.73	25.47	20.96	19.42	26.18	23.43	0.00	5.94	0.00	6.67	6.11	0.00
ASV848	42.16	38.33	42.45	31.87	25.32	39.04	36.57	26.66	31.66	54.86	26.07	41.03	39.34	30.32	42.11
ASV849	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.92	0.00
ASV850	0.00	0.00	0.00	0.00	5.51	0.00	0.00	3.26	0.00	0.00	3.98	0.00	0.00	2.92	0.00
ASV853	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	4.59	2.61	3.59	3.98	3.41	6.18
ASV855	13.28	18.31	16.18	10.80	3.16	0.00	9.43	3.90	5.54	0.00	7.24	8.89	10.07	11.18	0.00
ASV857	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.64	0.78	4.07
ASV858	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.03	3.04	4.06	2.58	1.78	0.00
ASV859	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.68	2.81	0.00	1.63	0.00
ASV865	1.32	3.81	4.01	4.93	9.25	7.95	8.22	6.83	7.16	0.00	0.00	0.00	0.00	1.44	0.00
ASV866	0.00	1.65	0.00	0.00	12.75	10.13	6.24	12.13	9.80	0.00	3.40	4.84	3.63	3.11	3.74
ASV869	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.92	5.43	4.84	4.22	3.74	5.69
ASV897	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.55	0.00	0.00	0.00	0.00	0.00	0.00
ASV898	0.00	2.35	0.75	0.00	0.00	1.22	2.13	0.00	0.00	0.00	0.00	0.00	3.63	0.00	0.00
ASV899	3.88	0.83	2.01	4.53	1.67	0.96	0.00	1.22	1.41	4.32	3.40	2.65	0.00	2.78	4.39
ASV921	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.59	0.00	0.00
Caenarcaniphilales															
ASV803	0.00	0.45	0.13	0.00	0.25	0.00	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table C.4 Continued

ASV ID	LT-Su-Mo1	LT-Su-Mid1	LT-Su-Aft1	LT-Se-Mor1	LT-Se-Mid1	LT-Se-Aft1	LT-Se-Mor2	LT-Se-Mid2	LT-Se-Aft2	W-Mor1	W-Mid1	W-Aft1	W-Mor2	W-Mid2	W-Aft2
Gastranaerophilales															
ASV933	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.00	0.00	0.00	0.00
Vampirovibrionales															
ASV800	0.50	0.32	0.44	0.00	0.06	0.26	0.43	0.14	0.43	0.00	0.00	0.00	0.00	0.00	0.00
Unknown															
ASV909	4.62	2.10	2.57	3.07	1.02	1.57	1.91	1.57	1.70	5.14	3.48	8.27	6.09	5.70	5.85

Appendix D
Supplementary Materials for Chapter 4

Table D.1 Relative abundances of major bacterial phyla rounded to two decimal points across a depth profile in the summer of 2018. Phyla present at less than 1% abundance across all samples were excluded from this table.

Classified Phylum	July 2018			August 2018			Lake
	Surface	Secchi	Secchi + 1 m	Surface	Secchi	Secchi + 1 m	
Actinobacteria	14.67	7.38	7.00	7.17	9.01	9.61	Big Turkey
	11.92	5.07	2.49	5.07	8.10	16.35	Little Turkey
	5.81	3.49	7.76	8.73	7.72	11.02	Wishart
Armatimonadota	1.09	0.40	0.041	0.61	0.44	0.16	Big Turkey
	2.15	3.81	4.84	0.36	0.65	1.73	Little Turkey
	1.30	1.35	1.32	0.18	0.14	0.25	Wishart
Bacteroidetota	6.54	12.23	8.46	12.23	20.01	18.42	Big Turkey
	6.90	16.60	17.02	21.79	30.06	10.44	Little Turkey
	11.68	8.22	5.33	33.44	14.02	23.99	Wishart
Bdellovibrionota	1.00	1.27	1.79	0.44	0.44	0.19	Big Turkey
	0.11	0.70	0.47	0.39	0.46	1.04	Little Turkey
	1.56	2.15	3.27	0.092	0.31	0.28	Wishart

Table D.1 Continued

Classified Phylum	July 2018			August 2018			Lake
	Surface	Secchi	Secchi + 1 m	Surface	Secchi	Secchi + 1 m	
Chloroflexi	0.044	0.50	0.76	0.10	0.072	0.077	Big Turkey
	0.007	0.034	0.064	0.038	0.21	0.73	Little Turkey
	1.89	1.28	1.14	0.16	0.35	0.42	Wishart
Cyanobacteria	27.27	26.96	45.44	11.51	11.78	12.78	Big Turkey
	14.33	23.55	13.85	6.78	8.54	12.60	Little Turkey
	17.44	18.46	15.87	4.53	3.84	4.77	Wishart
Firmicutes	0.20	0.022	0	0.025	0.015	0	Big Turkey
	10.28	0.44	0.15	0.010	0	0	Little Turkey
	0.076	0.041	0.24	0.021	0.037	0.013	Wishart
Myxococcota	0.10	0.83	0.47	0.12	0.22	0.063	Big Turkey
	0.19	0.022	0	0.15	0.21	0.031	Little Turkey
	0.49	0.88	1.21	0.10	0.098	0.20	Wishart

Table D.2 Relative abundances of major bacterial phyla rounded to two decimal points across a multi-seasonal profile. Phyla present at less than 1% abundance across all samples were excluded from this table.

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Acidobacteriota	0.071	0	0.40	18-Jul
	0.0128	0	0	18-Aug
	0.085	0.13	0.054	18-Oct
	0.49	0.025	0.073	19-Feb
	1.05	0.83	0.50	19-Mar
	0.27	0.83	0	19-May
	0.028	0	0.022	19-Jun
	0	0	0.22	19-Jul
	0.010	0.0091	0	19-Aug
	0	0.25	0.37	20-Jan
Actinobacteriota	7.38	5.07	3.49	18-Jul
	9.01	4.24	7.72	18-Aug
	14.03	7.55	6.31	18-Oct
	10.65	5.17	6.34	19-Feb
	9.90	4.41	2.36	19-Mar
	13.46	10.96	5.43	19-May
	5.99	9.66	7.06	19-Jun
	17.36	20.63	6.67	19-Jul
	11.33	10.52	3.53	19-Aug
	4.72	5.01	5.97	20-Jan
Armatimonadota	0.40	3.81	1.35	18-Jul
	0.44	0.65	0.25	18-Aug
	0.14	0.064	0.054	18-Oct

Table D.2 Continued

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Armatimonadota	0.21	0.14	0.084	19-Feb
	0.21	0.17	0	19-Mar
	0.21	0.48	0	19-May
	0.30	0.085	0.60	19-Jun
	0.24	0.11	0.15	19-Jul
	0.31	0.23	0.054	19-Aug
	0.32	0.54	0.062	20-Jan
Bacteroidota	12.23	16.60	8.22	18-Jul
	21.16	30.06	14.02	18-Aug
	7.16	8.81	8.80	18-Oct
	12.94	18.15	9.41	19-Feb
	16.01	12.94	14.05	19-Mar
	7.23	4.95	3.85	19-May
	8.75	27.67	2.96	19-Jun
	2.19	0.80	12.16	19-Jul
	5.73	9.19	8.09	19-Aug
	26.27	18.10	18.21	20-Jan
Bdellovibrionota	1.27	0.70	2.15	18-Jul
	0.44	0.45	0.31	18-Aug
	0.24	0.034	0.064	18-Oct
	0.22	0.33	0.037	19-Feb
	0.55	0.13	0.019	19-Mar
	0.24	0.11	0	19-May
	0.12	0.052	0.47	19-Jun

Table D.2 Continued

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Bdellovibrionota	0.070	0.0056	0.45	19-Jul
	0.34	0.20	0.71	19-Aug
	0.089	0.21	0.62	20-Jan
Chloroflexi	0.50	0.034	1.28	18-Jul
	0.35	0.21	0.35	18-Aug
	0.48	0.35	0.37	18-Oct
	0.54	0.08	0.031	19-Feb
	0.46	0.017	0.029	19-Mar
	0.50	0.15	0	19-May
	0.76	0.23	0.23	19-Jun
	0.23	0.076	0.33	19-Jul
	0.12	0.087	1.17	19-Aug
	0.089	0.26	0.10	20-Jan
Cyanobacteria	26.99	23.57	18.46	18-Jul
	11.81	8.54	3.84	18-Aug
	10.10	1.20	0.90	18-Oct
	0.17	N/A	0.01	19-Feb
	0.18	0	0	19-Mar
	11.07	0.77	1.51	19-May
	56.31	22.04	21.14	19-Jun
	36.60	34.23	20.46	19-Jul
	30.35	23.99	14.37	19-Aug
	0	0.0.065	0.11	20-Jan
Firmicutes	0.022	0.44	0.041	18-Jul
	0.015	0	0.037	18-Aug

Table D.2 Continued

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Firmicutes	0.057	0.16	0	18-Oct
	0.054	0.14	0	19-Feb
	0.062	0	0	19-Mar
	0.057	0	1.10	19-May
	0	0.0095	0.031	19-Jun
	0.092	0.053	0.027	19-Jul
	0.050	0.027	0	19-Aug
	3.76	0.13	0.32	20-Jan
Myxococcota	0.83	0.0.22	0.88	18-Jul
	0.22	0.021	0.098	18-Aug
	0.31	0.0.34	0.12	18-Oct
	0.15	0.057	0.031	19-Feb
	0.21	0.16	0.19	19-Mar
	0.060	0.079	0.21	19-May
	0.11	0	0.13	19-Jun
	0.83	0.022	0.87	19-Jul
	0.11	0	1.17	19-Aug
	0.030	0.16	0.18	20-Jan
Planctomycetota	3.37	0.80	4.79	18-Jul
	3.44	2.06	1.38	18-Aug
	6.58	2.51	1.09	18-Oct
	7.10	5.77	0.76	19-Feb
	9.75	2.71	0.058	19-Mar
	15.64	6.06	3.40	19-May
	1.96	0.69	1.25	19-Jun

Table D.2 Continued

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Planctomycetota	4.09	5.68	4.84	19-Jul
	7.59	6.30	2.28	19-Aug
	1.21	2.12	0.33	20-Jan
Proteobacteria	38.11	41.83	48.29	18-Jul
	46.28	46.23	65.63	18-Aug
	41.55	70.31	76.33	18-Oct
	53.55	56.49	76.52	19-Feb
	46.31	67.84	82.08	19-Mar
	48.20	74.28	82.90	19-May
	24.11	36.95	62.78	19-Jun
	16.64	24.04	47.70	19-Jul
	37.27	32.49	51.78	19-Aug
	53.44	46.05	69.35	20-Jan
Unknown	0.35	0.11	0.24	18-Jul
	0.030	0.039	0.22	18-Aug
	1.09	0.66	0.92	18-Oct
	0.12	0.14	0.063	19-Feb
	0.15	0.24	0	19-Mar
	0.32	0.49	0.59	19-May
	0.16	1.74	1.72	19-Jun
	0.22	0	0.45	19-Jul
	0.061	0.096	0.31	19-Aug
	0.80	0.24	0.53	20-Jan
Verrucomicrobiota	8.02	7.03	9.03	18-Jul
	6.87	7.27	5.65	18-Aug

Table D.2 Continued

Classified Phylum	Big Turkey	Little Turkey	Wishart	Sampling Month
Verrucomicrobiota	17.41	8.12	4.87	18-Oct
	12.68	12.68	6.38	19-Feb
	14.02	10.29	0.66	19-Mar
	2.65	1.15	1.14	19-May
	1.40	0.88	1.44	19-Jun
	22.16	13.96	5.77	19-Jul
	6.23	16.75	16.27	19-Aug
	9.19	26.56	3.27	20-Jan

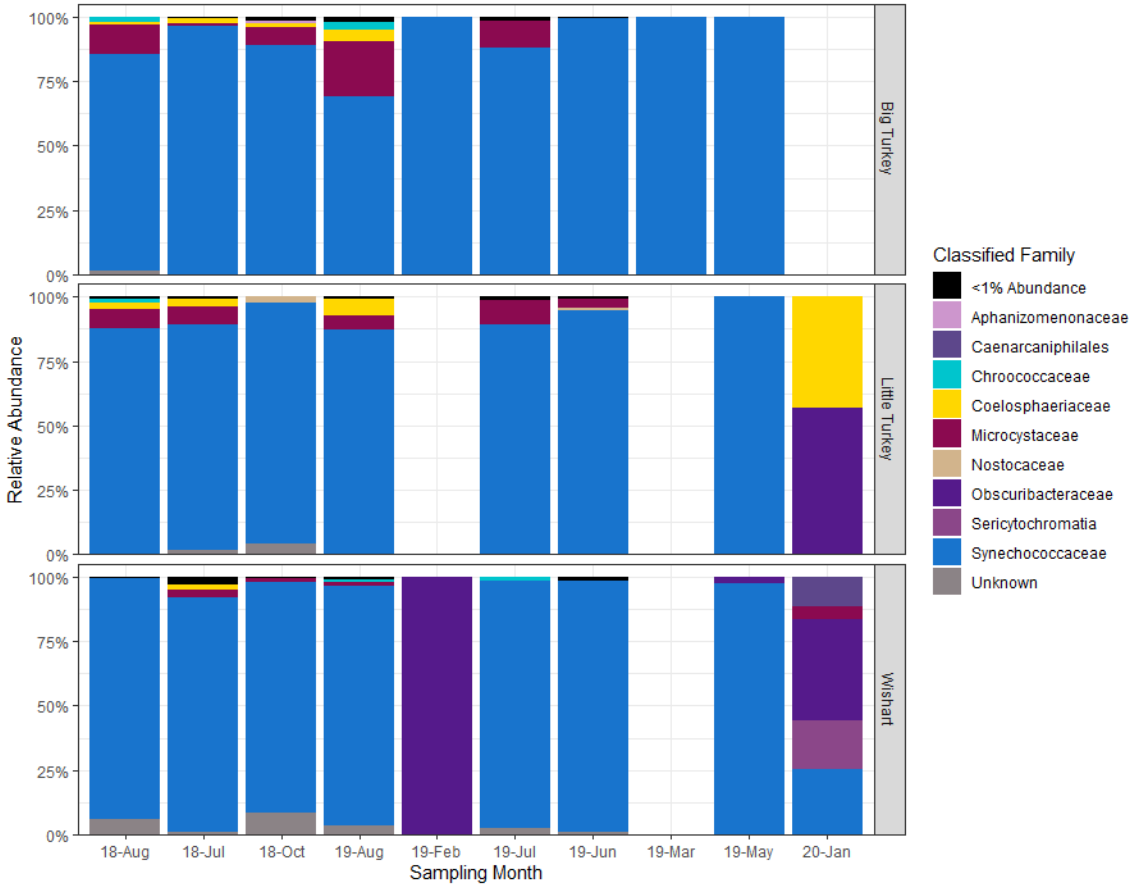


Figure D.1 Stacked bar charts depicting the relative abundances of cyanobacterial families composing the cyanobacterial communities identified from amplicon sequencing of the V4 region of the 16S rRNA gene across a seasonal sampling series in a shallow lake (Wishart), mid-sized lake (Little Turkey) and deep lake (Big Turkey). Notably, cyanobacteria were present at less than 1% abundance in the bacterial communities during ice-covered months. Sampling months with no bar signify the absence of cyanobacterial sequences.

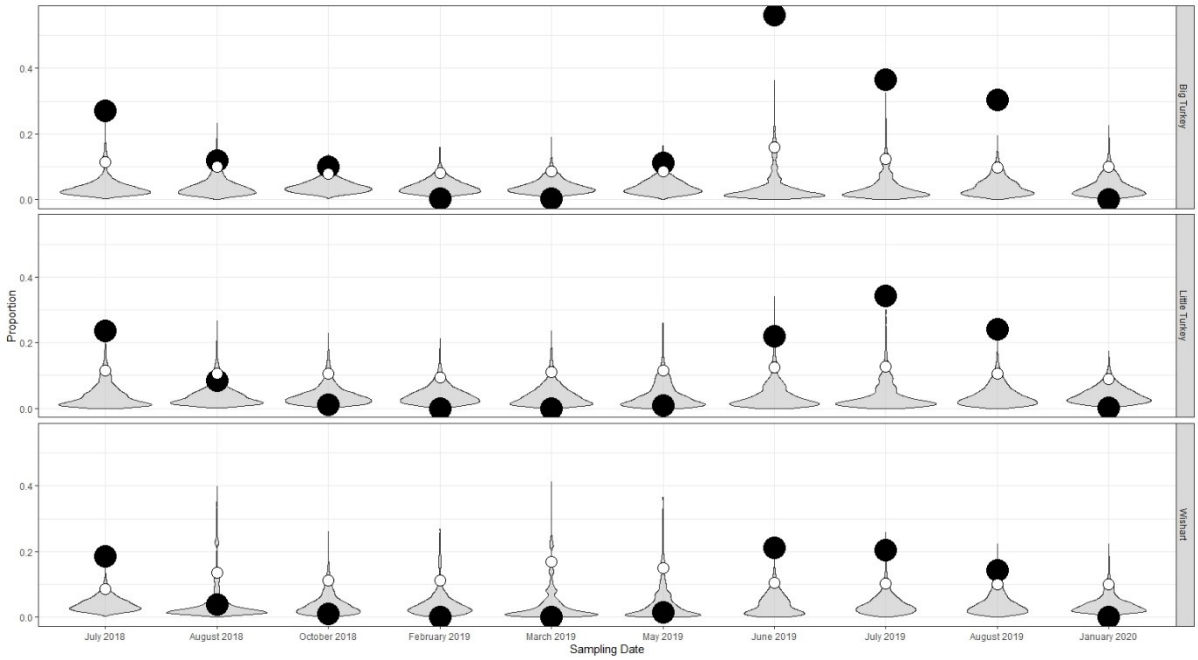


Figure D.2 Violin plots visualizing the distribution of randomized total phylum counts and the specific relative abundance of cyanobacteria relative to the 95th percentile across sampling months in a multi-seasonal period in Big Turkey, Little Turkey and Wishart Lake. The observed cyanobacterial values are visualized with the solid black circle, and the 95th percentile with the white circle. Black dots present above the white circle represents an enriched abundance of cyanobacteria present in the bacterial community.

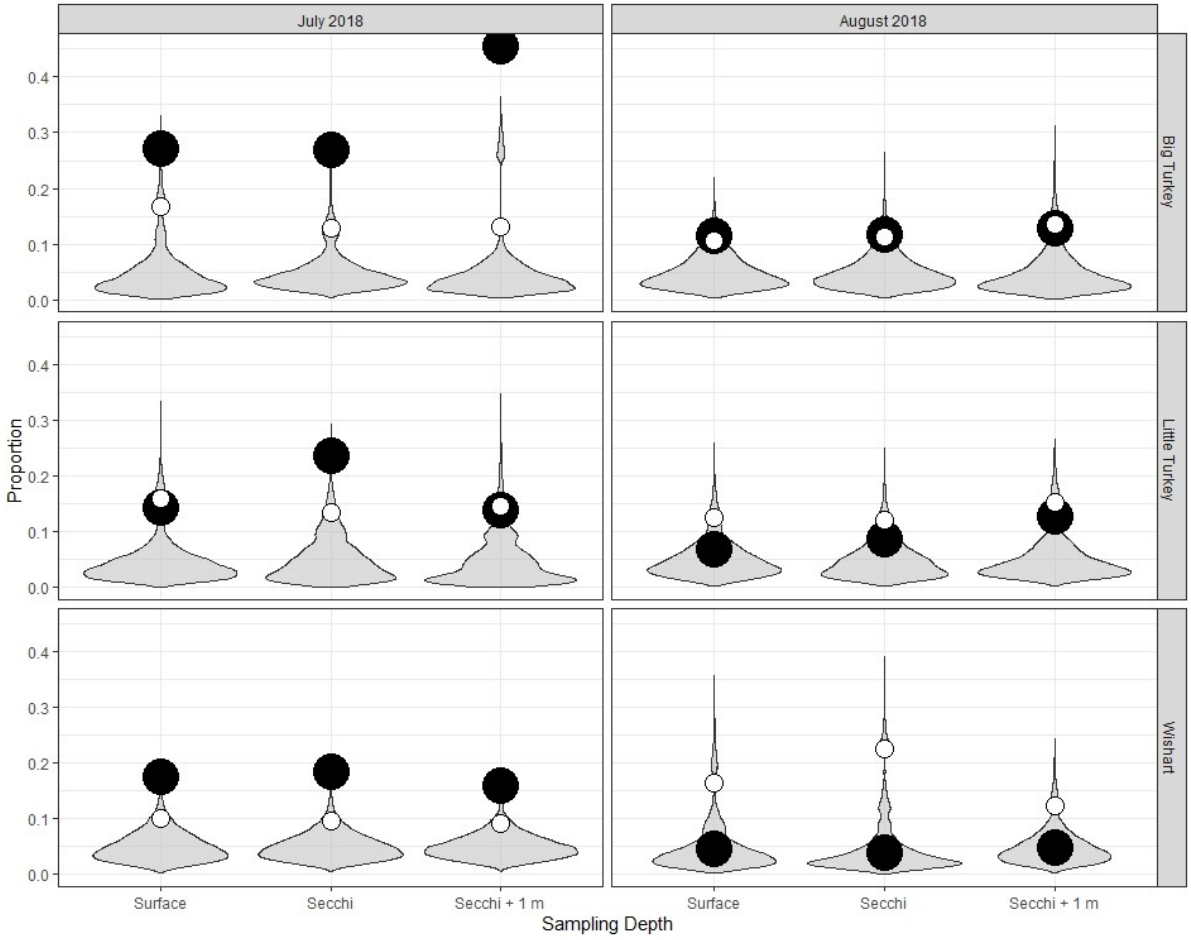


Figure D.3 Violin plots visualizing the distribution of randomized total phylum counts and the specific relative abundance of cyanobacteria relative to the 95th percentile across sampling depths in Big Turkey, Little Turkey and Wishart Lake. The observed cyanobacterial values are visualized with the solid black circle, and the 95th percentile with the white circle. Black dots present above the white circle represents an enriched abundance of cyanobacteria present in the bacterial community

Table D.3 Probability values for cyanobacterial enrichment within the randomized total phylum counts per sample across a seasonal series. The mean and standard deviation of the randomized counts were used to calculate Z-scores. From the Z-scores, probability values were obtained and adjusted with a Bonferroni correction.

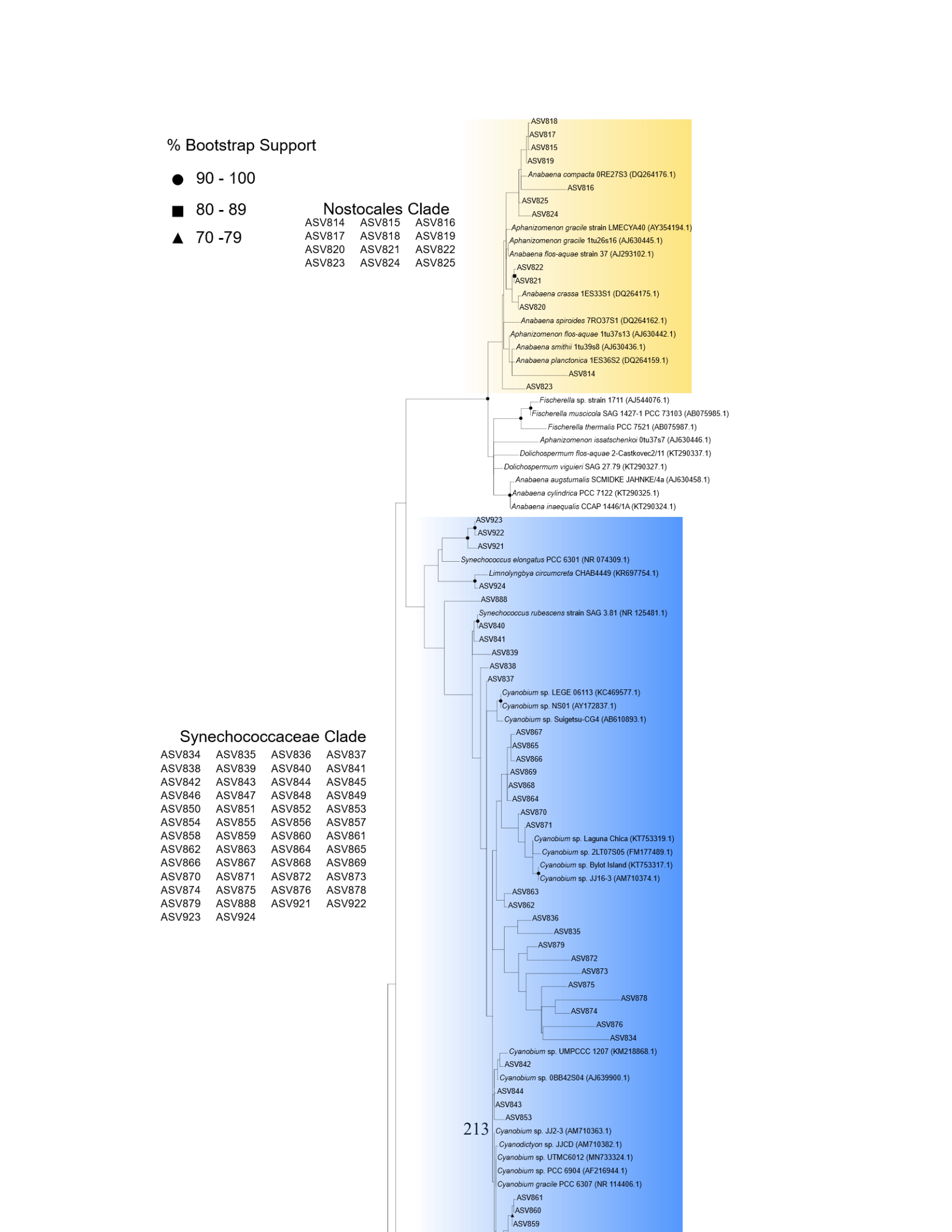
Lake	Sampling Date	Mean	Standard Deviation	Observed	p-value	Adjusted p-value
Big Turkey	20-Jan	0.04	0.03	0.00	0.08	1.00
Big Turkey	19-Feb	0.04	0.02	0.00	0.03	0.96
Big Turkey	19-Mar	0.04	0.02	0.00	0.03	1.00
Big Turkey	19-May	0.04	0.02	0.11	0.00	0.03
Big Turkey	19-Jun	0.04	0.05	0.56	0.00	0.00
Big Turkey	18-Jul	0.04	0.03	0.27	0.00	0.00
Big Turkey	19-Jul	0.04	0.04	0.37	0.00	0.00
Big Turkey	18-Aug	0.04	0.03	0.12	0.00	0.10
Big Turkey	19-Aug	0.04	0.03	0.30	0.00	0.00
Big Turkey	18-Oct	0.04	0.02	0.10	0.00	0.03
Little Turkey	20-Jan	0.04	0.02	0.00	0.05	1.00
Little Turkey	19-Feb	0.04	0.03	0.00	0.07	1.00
Little Turkey	19-Mar	0.04	0.03	0.00	0.10	1.00
Little Turkey	19-May	0.04	0.04	0.01	0.19	1.00
Little Turkey	19-Jun	0.04	0.04	0.22	0.00	0.00
Little Turkey	18-Jul	0.04	0.04	0.24	0.00	0.00
Little Turkey	19-Jul	0.04	0.04	0.34	0.00	0.00
Little Turkey	18-Aug	0.04	0.03	0.09	0.08	1.00
Little Turkey	19-Aug	0.04	0.03	0.24	0.00	0.00
Little Turkey	18-Oct	0.04	0.03	0.01	0.18	1.00
Wishart	20-Jan	0.04	0.03	0.00	0.07	1.00
Wishart	19-Feb	0.04	0.04	0.00	0.14	1.00
Wishart	19-Mar	0.04	0.06	0.00	0.24	1.00
Wishart	19-May	0.04	0.05	0.02	0.30	1.00
Wishart	19-Jun	0.04	0.03	0.21	0.00	0.00
Wishart	18-Jul	0.04	0.02	0.18	0.00	0.00
Wishart	19-Jul	0.04	0.03	0.20	0.00	0.00

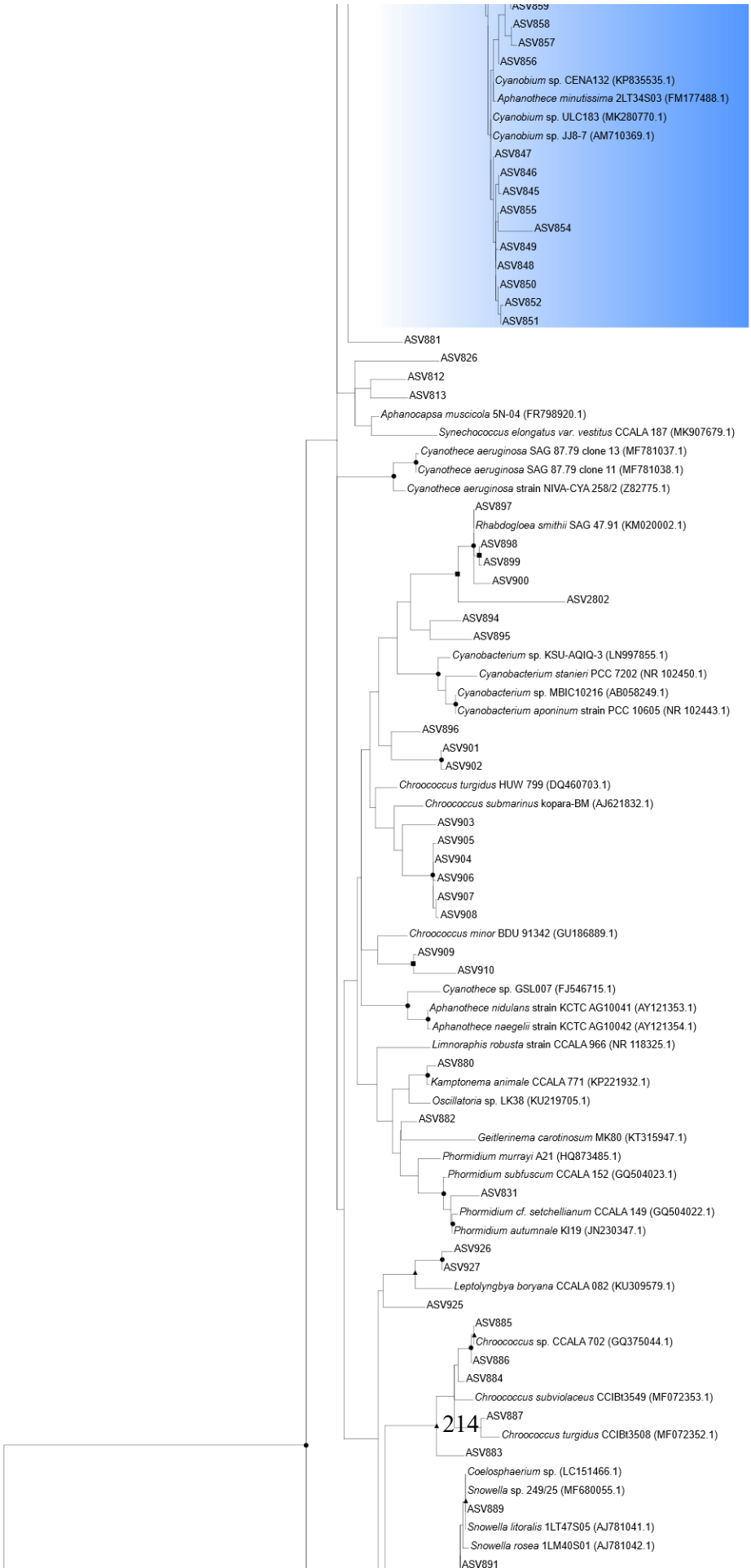
Table D.3 Continued

Lake	Sampling Date	Mean	Standard Deviation	Observed	p-value	Adjusted p-value
Wishart	18-Aug	0.04	0.05	0.04	0.49	1.00
Wishart	19-Aug	0.04	0.03	0.14	0.00	0.01
Wishart	18-Oct	0.04	0.03	0.01	0.17	1.00

Table D.4 Probability values for cyanobacterial enrichment within the randomized total phylum counts per sample across a depth profile. The mean and standard deviation of the randomized counts were used to calculate Z-scores. From the Z-scores, probability values were obtained and adjusted with a Bonferroni correction.

Lake	Month	Depth	Mean	Standard Deviation	Observed	p-value	Adjusted p-value
Big Turkey	August	Surface	0.05	0.03	0.12	0.02	0.31
Big Turkey	August	Secchi	0.05	0.03	0.12	0.01	0.22
Big Turkey	August	Deep	0.05	0.04	0.13	0.03	0.45
Big Turkey	July	Surface	0.05	0.05	0.27	0.00	0.00
Big Turkey	July	Secchi	0.05	0.03	0.27	0.00	0.00
Big Turkey	July	Deep	0.05	0.06	0.45	0.00	0.00
Little Turkey	August	Surface	0.05	0.03	0.07	0.30	1.00
Little Turkey	August	Secchi	0.05	0.03	0.09	0.15	1.00
Little Turkey	August	Deep	0.05	0.04	0.13	0.02	0.45
Little Turkey	July	Surface	0.05	0.04	0.14	0.02	0.30
Little Turkey	July	Secchi	0.05	0.04	0.24	0.00	0.00
Little Turkey	July	Deep	0.05	0.05	0.14	0.03	0.53
Wishart	August	Surface	0.05	0.05	0.05	0.46	1.00
Wishart	August	Secchi	0.05	0.06	0.04	0.42	1.00
Wishart	August	Deep	0.05	0.03	0.05	0.47	1.00
Wishart	July	Surface	0.05	0.03	0.17	0.00	0.00
Wishart	July	Secchi	0.05	0.02	0.18	0.00	0.00
Wishart	July	Deep	0.05	0.02	0.16	0.00	0.00


% Bootstrap Support
90 - 100
80 - 89
70 -79
Nostocales Clade
ASV814 ASV815 ASV816
ASV817 ASV818 ASV819
ASV820 ASV821 ASV822
ASV823 ASV824 ASV825
Synechococcaceae Clade
ASV834 ASV835 ASV836 ASV837
ASV838 ASV839 ASV840 ASV841
ASV842 ASV843 ASV844 ASV845
ASV846 ASV847 ASV848 ASV849
ASV850 ASV851 ASV852 ASV853
ASV854 ASV855 ASV856 ASV857
ASV858 ASV859 ASV860 ASV861
ASV862 ASV863 ASV864 ASV865
ASV866 ASV867 ASV868 ASV869
ASV870 ASV871 ASV872 ASV873
ASV874 ASV875 ASV876 ASV878
ASV879 ASV888 ASV921 ASV922
ASV923 ASV924
ASV818
ASV817
ASV815
ASV819
Anabaena compacta 0RE27S3 (DQ264176.1)
ASV816
ASV825
ASV824
Aphanizomenon gracile strain LMECYA40 (AY354194.1)
Aphanizomenon gracile 1tu26s16 (AJ630445.1)
Anabaena flos-aquae strain 37 (AJ293102.1)
ASV822
ASV821
Anabaena crassa 1ES33S1 (DQ264175.1)
ASV820
Anabaena spiroides 7RO37S1 (DQ264162.1)
Aphanizomenon flos-aquae 1tu37s13 (AJ630442.1)
Anabaena smithii 1tu39s8 (AJ630436.1)
Anabaena planctonica 1ES36S2 (DQ264159.1)
ASV814
ASV823
Fischerella sp. strain 1711 (AJ544076.1)
Fischerella muscicola SAG 1427-1 PCC 73103 (AB075985.1)
Fischerella thermalis PCC 7521 (AB075987.1)
Aphanizomenon issatschenkoi 0tu37s7 (AJ630446.1)
Dolichospermum flos-aquae 2-Castkovec2/11 (KT290337.1)
Dolichospermum viguieri SAG 27.79 (KT290327.1)
Anabaena augstumalis SCMIDKE JAHNKE/4a (AJ630458.1)
Anabaena cylindrica PCC 7122 (KT290325.1)
Anabaena inaequalis CCAP 1446/1A (KT290324.1)
ASV923
ASV922
ASV921
Synechococcus elongatus PCC 6301 (NR 074309.1)
Limnolyngbya circumcreta CHAB4449 (KR697754.1)
ASV924
ASV888
Synechococcus rubescens strain SAG 3.81 (NR 125481.1)
ASV840
ASV841
ASV839
ASV838
ASV837
Cyanobium sp. LEGE 06113 (KC469577.1)
Cyanobium sp. NS01 (AY172837.1)
Cyanobium sp. Suigetsu-CG4 (AB610893.1)
ASV867
ASV865
ASV866
ASV869
ASV868
ASV864
ASV870
ASV871
Cyanobium sp. Laguna Chica (KT753319.1)
Cyanobium sp. 2LT07S05 (FM177489.1)
Cyanobium sp. Bylot Island (KT753317.1)
Cyanobium sp. JJ16-3 (AM710374.1)
ASV863
ASV862
ASV836
ASV835
ASV879
ASV872
ASV873
ASV875
ASV878
ASV874
ASV876
ASV834
Cyanobium sp. UMPCCC 1207 (KM218868.1)
ASV842
Cyanobium sp. 0BB42S04 (AJ639900.1)
ASV844
ASV843
ASV853
Cyanobium sp. JJ2-3 (AM710363.1)
Cyanodictyon sp. JJCD (AM710382.1)
Cyanobium sp. UTMC6012 (MN733324.1)
Cyanobium sp. PCC 6904 (AF216944.1)
Cyanobium gracile PCC 6307 (NR 114406.1)
ASV861
ASV860
ASV859
213


ASV859
ASV858
ASV857
ASV856
Cyanobium sp. CENA132 (KP835535.1)
Aphanothece minutissima 2LT34S03 (FM177488.1)
Cyanobium sp. ULC183 (MK280770.1)
Cyanobium sp. JJ8-7 (AM710369.1)
ASV847
ASV846
ASV845
ASV855
ASV854
ASV849
ASV848
ASV850
ASV852
ASV851
ASV881
ASV826
ASV812
ASV813
Aphanocapsa muscicola 5N-04 (FR798920.1)
Synechococcus elongatus var. vestitus CCALA 187 (MK907679.1)
Cyanothece aeruginosa SAG 87.79 clone 13 (MF781037.1)
Cyanothece aeruginosa SAG 87.79 clone 11 (MF781038.1)
Cyanothece aeruginosa strain NIVA-CYA 258/2 (Z82775.1)
ASV897
Rhabdogloea smithii SAG 47.91 (KM020002.1)
ASV898
ASV899
ASV900
ASV2802
ASV894
ASV895
Cyanobacterium sp. KSU-AQIQ-3 (LN997855.1)
Cyanobacterium stanieri PCC 7202 (NR 102450.1)
Cyanobacterium sp. MBIC10216 (AB058249.1)
Cyanobacterium aponinum strain PCC 10605 (NR 102443.1)
ASV896
ASV901
ASV902
Chroococcus turgidus HUW 799 (DQ460703.1)
Chroococcus submarinus kopara-BM (AJ621832.1)
ASV903
ASV905
ASV904
ASV906
ASV907
ASV908
Chroococcus minor BDU 91342 (GU186889.1)
ASV909
ASV910
Cyanothece sp. GSL007 (FJ546715.1)
Aphanothece nidulans strain KCTC AG10041 (AY121353.1)
Aphanothece naegelii strain KCTC AG10042 (AY121354.1)
Limnoraphis robusta strain CCALA 966 (NR 118325.1)
ASV880
Kamptonema animale CCALA 771 (KP221932.1)
Oscillatoria sp. LK38 (KU219705.1)
ASV882
Geitlerinema carotinosum MK80 (KT315947.1)
Phormidium murrayi A21 (HQ873485.1)
Phormidium subfuscum CCALA 152 (GQ504023.1)
ASV831
Phormidium cf. setchellianum CCALA 149 (GQ504022.1)
Phormidium autumnale KI19 (JN230347.1)
ASV926
ASV927
Leptolyngbya boryana CCALA 082 (KU309579.1)
ASV925
ASV885
Chroococcus sp. CCALA 702 (GQ375044.1)
ASV886
ASV884
Chroococcus subviolaceus CCIBt3549 (MF072353.1)
214
ASV887
Chroococcus turgidus CCIBt3508 (MF072352.1)
ASV883
Coelosphaerium sp. (LC151466.1)
Snowella sp. 249/25 (MF680055.1)
ASV889
Snowella litoralis 1LT47S05 (AJ781041.1)
Snowella rosea 1LM40S01 (AJ781042.1)
ASV891

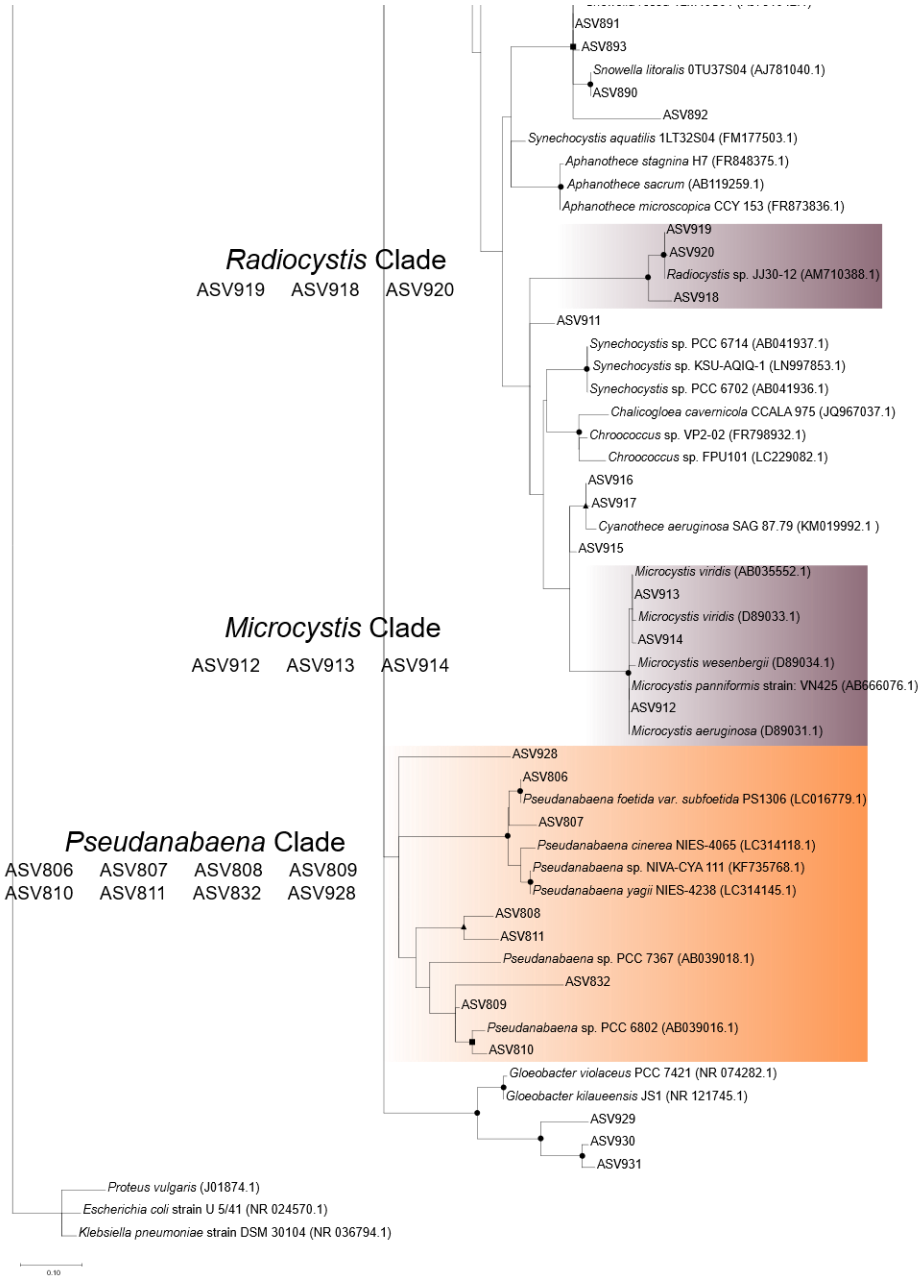


Figure D.4 Phylogenetic tree indicating the taxonomic affiliation of amplicon sequence variants classified as phototrophic cyanobacteria inferred using the General Time Reversible substitution model with Gamma invariant sites included. 1000 boot strap replicates were performed to estimate support for clades. Reference sequences for common freshwater cyanobacteria were included to confirm the taxonomic classification of the Naïve-Bayes classifier. Genus and species names were not modified from the reference entries in NCBI and may not reflect current accepted nomenclature. For the purpose of this research, taxonomic classifications were only examined at higher taxonomic levels of order and family.

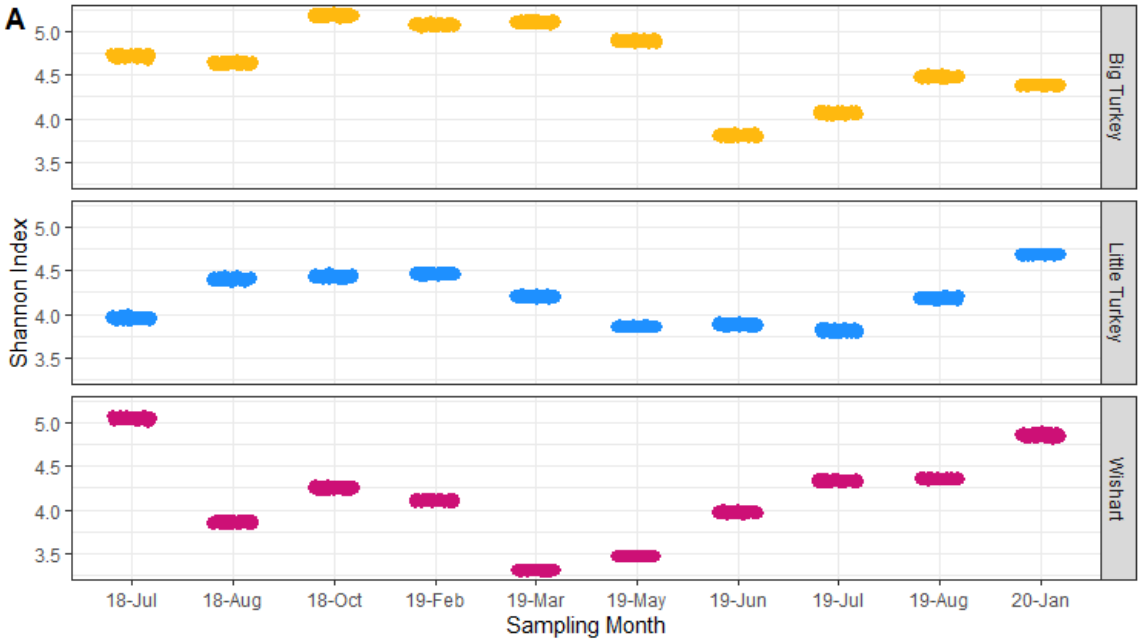


Figure D.5 Alpha diversity analysis of bacterial communities collected across a multi-seasonal period in Big Turkey, Little Turkey and Wishart Lake. The Shannon Index was calculated on libraries that were repeatedly rarefied to a normalized size of 9,534 reads to identify seasonal variation in bacterial community diversity values.

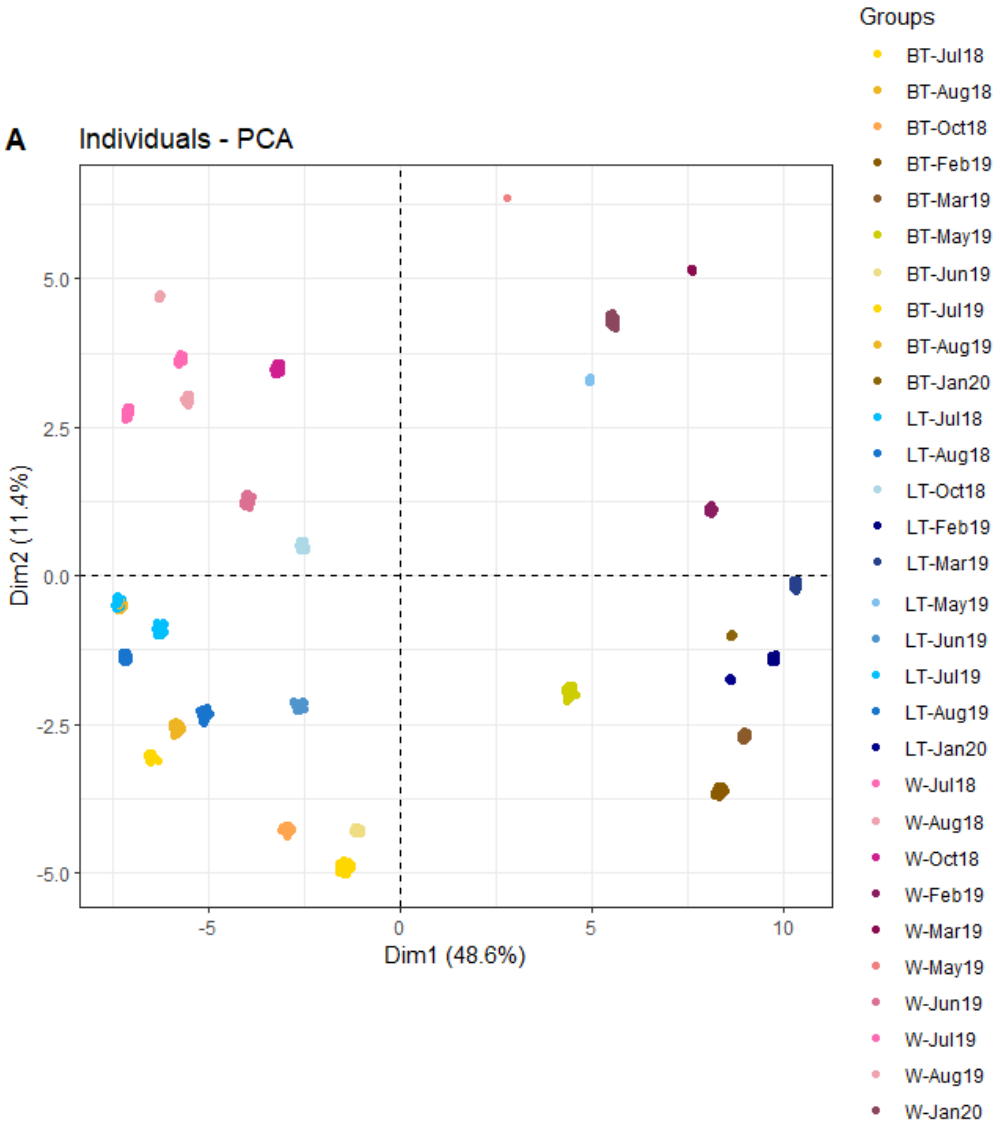


Figure D.6 Beta-diversity analysis on bacterial community similarity of seasonal samples in Big Turkey (BT), Little Turkey (LT) and Wishart (W) Lake. Sequence libraries were repeatedly rarefied to a normalized library size of 9,534. Rarefied libraries were transformed using a Hellinger transformation. The Bray-Curtis dissimilarity metric was calculated on transformed data and visualized using a PCA ordination to identify seasonal trends in bacterial community structure.

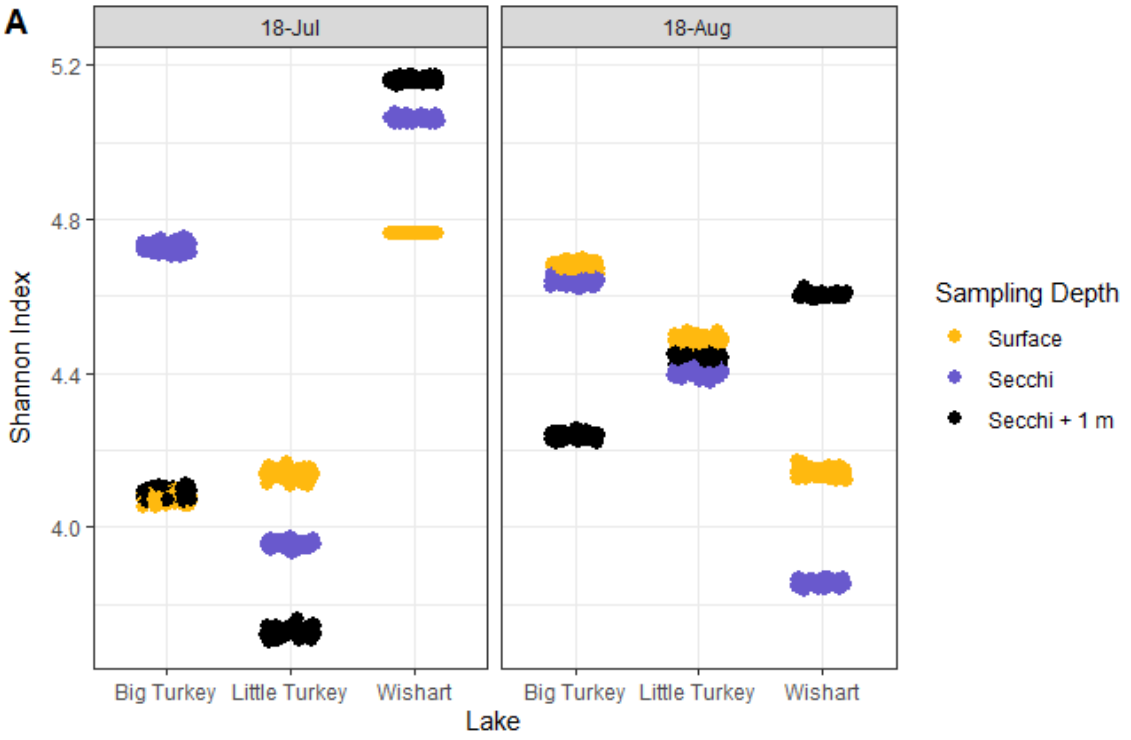


Figure D.7 Alpha diversity analysis of bacterial communities collected across a depth profile in the summer of 2018 in Big Turkey, Little Turkey and Wishart Lake. The Shannon Index was calculated on libraries that were repeatedly rarefied to a normalized size of 17,048 reads to identify spatial variation in bacterial community diversity values.

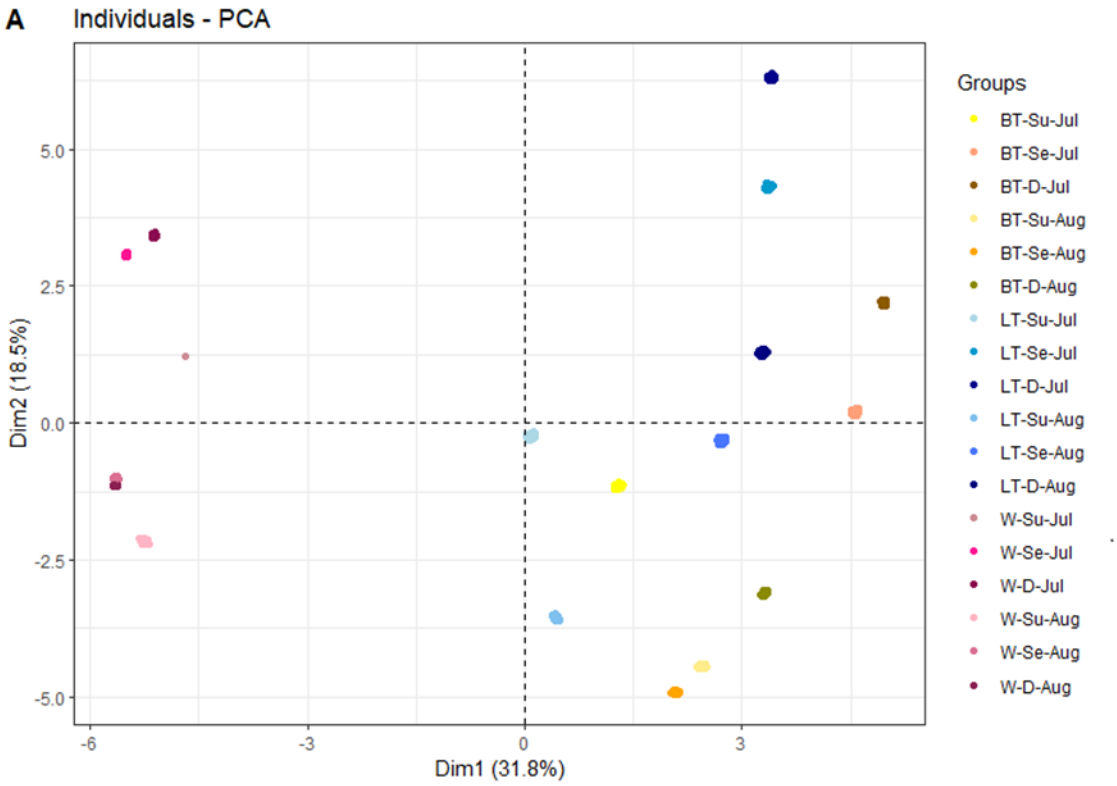


Figure D.8 Beta-diversity analysis on bacterial community similarity of seasonal samples in Big Turkey (BT), Little Turkey (LT) and Wishart (W) Lake. Sequence libraries were repeatedly rarefied to a normalized library size of 17,048. Rarefied libraries were transformed using a Hellinger transformation. The Bray-Curtis metric was calculated on transformed data and visualized using a PCA ordination to identify spatial trends in bacterial community structure.

Table D.5 Relative abundances of ASVs in the cyanobacterial community across a depth profile in a deep, stratified lake (Big Turkey = BT) mid-sized, stratified lake (Little Turkey = LT) and non-stratified lake (Wishart = W) in the summer 0f 2018. Samples were collected at surface (Su), Secchi depth (Se) and 1 meter below Secchi depth (D).

ASV ID	BT-Su-Jul18	BT-Se-Jul18	BT-D-Jul18	BT-Su-Aug18	BT-Se-Aug18	BT-D-Aug18	LT-Su-Jul18	LT-Se-Jul18	LT-D-Jul18	LT-Su-Aug18	LT-Se-Aug18	LT-D-Aug18	W-Su-Jul18	W-Se-Jul18	W-D-Jul18	W-Su-Aug18	W-Se-[illegible]	W-D-Aug18
Chroococcales																		
ASV883	0	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0	0
ASV884	0.2	0	0	0.6	0.97	0.27	0	0	0	1.18	0.16	0.09	0	0	0	0	0	0
ASV885	0	0	0	0.24	0.17	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV887	0.45	0	0	0.39	0.48	0	0.43	0.12	0.07	5.27	0.72	0.62	0	0	0	0	0	0
ASV910	0.37	0.34	0.04	0.39	0.26	0	0.24	0.09	0.12	1.47	0.62	0.37	0.44	0.37	0.44	0.59	0	0.65
ASV889	0	0	0	0	0	0	0	0.36	0	0	0	0	0	0	0	0	0	0
ASV890	0.45	0.96	0.29	0.68	1.15	0.93	1.11	1.32	0.35	1.84	0.78	0.76	2.15	2.07	2.62	1.12	0	0.94
ASV892	0	0	0	0	0	0.05	0	0	0	0	0	0	0	0	0	0	0	0
ASV893	0.21	0.79	0.22	0.24	0	0.3	2.36	1.14	0.21	1.33	1.5	1.32	0	0	0	0	0	0
ASV894	0	0	0	0	0	0	0	0	0	0	0	0	0	0.12	0	0	0	0
ASV895	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0
ASV903	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV904	0	0.22	0	0	0	0	0	0	0	0	1.17	0	0	0	0	0.4	0	0.37
ASV906	0	0	0	1.37	0	0	0	0	0	0	0	0	1.58	0.61	0.66	0	0	0
ASV907	0	0	0	0	0	0	0	0	0	0	0	1.18	0	0	0	0	0	0
ASV908	0	0	0	0	0	0	0	0.26	0.41	1.47	0	0	0	0	0	0	0	0
ASV911	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0
ASV912	0.14	0.03	0	0	0	0	0.38	0	0	0	0	0	5.51	0	0.46	0	0	0

Table D.5 Continued

ASV ID	BT-Su-Jul18	BT-Se-Jul18	BT-D-Jul18	BT-S8-Aug18	BT-Se-Aug18	BT-D-Aug18	LT-Su-Jul18	LT-Se-Jul18	LT-D-Jul18	LT-Su-Aug18	LT-Se-Aug18	LT-D-Aug18	W-Su-Jul18	W-Se-Jul18	W-D-Jul18	W-Su-Aug18	W-Se-Aug18	W-D-Aug18
ASV913	0	0	0	0	0	0	0	0	0	0	0	0	0	0.66	0	0	0	0
ASV914	0	0	0	1.69	0.3	0	0	0	0	1.44	0	0	1.21	0	0	0	0	0
ASV915	0.15	0	0	0	0	0	0	0	0.14	0.33	0	0.35	0	0	0	0	0	0
ASV916	0.31	0	0	0.39	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV917	0	0	0	0	0.63	0.52	0.35	0	0	1.92	0.69	0.6	0.2	0	0.78	0.73	0	0
ASV918	0	0	0	0.09	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV919	4.67	0.86	0.2	3.29	10.03	4.95	4.25	7.05	4.95	14.52	5.58	6.92	0.94	0.98	0.73	0.46	0	0
ASV920	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.47
Chroococcidiopsidales																		
ASV828	0	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cyanobacteriales																		
ASV901	0	0	0	0	0	0	0	0	0	0	0	0	0	0.22	0	0	0	0
ASV896	0	0	0	0	0	0	0	0	0	0	0	0	0	0.3	0	0	0	0
Gloeobacterales																		
ASV929	0	0	0	0	0	0	0	0	0	0	0	0	0	0.15	0.22	1.19	0.73	1.03
ASV930	0	0	0	0	0	0	0	0	0	0	0	0	0	0.61	0.05	0	0	0
Nostocales																		
ASV824	0	0	0	0.75	0	0	0	0	0	0.37	0	0	0	0	0	0	0	0
ASV815	0.58	0	0	0	0	0	0.45	0.17	0.66	0	0	0	0	0	0	0	0	0
ASV820	0.2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV822	0.6	0	0	0	0	0	0.23	0.26	1.65	0	0	0	0	0	0	0	0	0
ASV825	0	0	0	0	0	0	0	0	0	0	0	0	0	0.12	0	0	0	0
Oscillatoriales																		

Table D.5 Continued

ASV ID	BT-Su-Jul18	BT-Se-Jul18	BT-D-Jul18	BT-S8-Aug18	BT-Se-Aug18	BT-D-Aug18	LT-Su-Jul18	LT-Se-Jul18	LT-D-Jul18	LT-Su-Aug18	LT-Se-Aug18	LT-D-Aug18	W-Su-Jul18	W-Se-Jul18	W-D-Jul18	W-Su-Aug18	W-Se-Aug18	W-D-Aug18
ASV881	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Synechococcales																		
ASV812	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.15	0	0	0
ASV813	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0
ASV924	0	0	0	0	0	0	0	0	0	0	0	0	0	0.12	0	0	0	0
ASV925	0	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0
ASV927	0	0	0	0	0	0	0	0	0	0	0	0	0	0.52	0	0	0	0
ASV806	0	0	0	0	0	0	0	0.04	0	0	0	0	0	0	0	0.2	0	0.65
ASV807	0	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0	0
ASV809	0	0	0	0	0	0	0	0	0	0	0	0	0	0.71	0	0	0	0
ASV810	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0
ASV928	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV808	0.09	0	0	0.09	0	0	0	0	0	0	0	0	0.13	0	0	0	0	0
ASV811	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	0	0	0	0
ASV835	0	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0
ASV837	0	4.64	7.85	5.24	0	0.66	0	0	0	0	0	0	0	0	0	0	0	0
ASV838	0	0	0	0	0	0	1.16	0	0	2.62	0	0	3.19	2.36	1.19	2.11	2.67	1.78
ASV839	10.49	1.92	1.18	2.01	4.22	3.6	3.94	0.62	0.64	1.18	1.31	1.02	4.73	7.97	6.75	4.22	5.34	6.08
ASV840	2.77	0.4	0.19	2.7	4.48	1.48	0	0	0	0	0	0	0	0	0	0	0	0
ASV841	0.33	4.59	6.75	4.47	3.68	7.39	0	0	0	0	0	0	0	0	0	0	0	0
ASV842	0	0	0	0	0	0	0.19	0.12	0.14	0	0.29	0.6	1.91	2.29	3.71	1.91	0	2.43
ASV843	0	0	0	0	0	0	0.45	0	0	0	0	0	1.01	1.08	1.92	6.59	6.8	4.77
ASV844	0.99	33	52.99	36.47	3.16	11.73	0	0	0	0	0	0	0	0	0	0	0	0

Table D.5 Continued

ASV ID	BT-Su-Jul18	BT-Se-Jul18	BT-D-Jul18	BT-S8-Aug18	BT-Se-Aug18	BT-D-Aug18	LT-Su-Jul18	LT-Se-Jul18	LT-D-Jul18	LT-Su-Aug18	LT-Se-Aug18	LT-D-Aug18	W-Su-Jul18	W-Se-Jul18	W-D-Jul18	W-Su-Aug18	W-Se-Aug18	W-D-Aug18
ASV845	0	3.02	7.23	4.43	0	2.14	0	0	0	0	0	0	0	0	0	0	0	0
ASV846	56.87	34.92	10.88	14.99	42.75	45.14	21.04	34.55	18.28	7.59	24.18	13.06	8.19	7.7	5.85	7.91	8.98	6.55
ASV847	0	0	0	0	0	0	0	0	0	0	0	0	1.81	0	2.94	0	0	0
ASV848	6.6	2.44	1.13	4.49	8.42	6.37	27.17	25.2	37.85	30.36	25.48	43.71	30.68	30.19	29.5	32.17	35.56	42.38
ASV849	0	0	0	0	0	0	1.2	0	0	0	0	0	4.93	3.25	2.11	3.89	0	0
ASV850	4.76	1.84	0	0	0	0	5.5	3.31	0	0	5.97	4.4	4.83	5.53	5.34	0	0	0
ASV851	0	0	0	0	0	0	0.85	3.01	3.46	0	1.76	2.97	0	0	1.77	0	0	0
ASV852	0	0	0.15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV853	0.26	1.03	0.09	3.29	2.49	2.06	0	0	0	0	0	0	0.64	0.61	0.44	2.7	2.31	1.96
ASV854	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV855	2.63	0	0	2.05	3.81	0	11.55	1.2	1.56	11.75	0	0	7.62	5.83	3.67	13.05	11.65	6.27
ASV856	0	0	0	0	0	0	0	0	0	0	0	0	0	0.47	0	0	0	0
ASV857	0	0	0	0	0	0	0	0	0	0	0	0	0.77	0.64	1.36	0	0	0
ASV858	0	0	0	0	0	0	0	0	0	0	0	0	0.64	0.49	1.21	4.28	5.95	5.24
ASV859	0	0	0	0	0	0	0.4	0	0	0	0	0	0.3	0.98	0.75	1.05	1.7	1.59
ASV860	0	0.97	3.55	3.29	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV861	0	0	0.54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV862	0	0	0	0	0	0	0	0	0	0	0	0	0.77	1.6	3.52	0	0	0
ASV863	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.32	0	0	0
ASV864	0	3.17	1.24	1.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV865	1.34	1.98	0.2	0.88	7.58	10.05	4.91	13.89	19.88	6.01	20.1	17.14	0.97	0.61	1.51	1.78	0	0
ASV866	2.21	1.77	3.43	2.48	3.07	1.48	5.09	3.18	5.83	2.03	7.18	2.44	6.65	9.81	10.68	3.96	5.46	6.45
ASV867	0	0.35	1.53	0.86	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table D.5 Continued

ASV ID	BT-Su-Jul18	BT-Se-Jul18	BT-D-Jul18	BT-S8-Aug18	BT-Se-Aug18	BT-D-Aug18	LT-Su-Jul18	LT-Se-Jul18	LT-D-Jul18	LT-Su-Aug18	LT-Se-Aug18	LT-D-Aug18	W-Su-Jul18	W-Se-Jul18	W-D-Jul18	W-Su-Aug18	W-Se-Aug18	W-D-Aug18
ASV868	0	0	0	0	0	0	1.3	0	0	0	0	0	0	0	0	0	0	0
ASV869	0	0	0	0	0	0	0	0.63	1.08	0	0	0	1.68	2.14	2.69	2.04	4.25	3.18
ASV870	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0
ASV871	0	0.16	0.17	0.06	0	0.22	0	0	0	0	0	0	0	0	0	0	0	0
ASV897	0.41	0.19	0.03	0	0.58	0	0.69	0.79	0.57	0	0	0	0	0	0	0	0	0
ASV899	0.33	0	0	0	0	0	1.2	0.72	1.33	3.65	1.7	1.57	4.87	6.88	5.41	1.52	2.55	1.96
ASV922	0	0	0	0	0	0	0	0	0	0	0	0	0	0.17	0	0	0	0
ASV923	0.02	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Caenarcaniphilales																		
ASV803	0	0	0	0.13	0.13	0	0	0	0	0.52	0	0	0.13	0	0	0	0	0
ASV805	0	0	0	0	0	0	0	0	0	0	0	0	0.07	0	0	0	0	0
Obscuribacterales																		
ASV935	0	0.04	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ASV938	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.05	0	0	0
ASV939	0	0	0	0	0.06	0	0	0	0	0	0	0	0	0	0	0	0	0
Unknown Order																		
ASV909	1.49	0.25	0.1	0.88	1.49	0.63	3.54	1.89	0.78	3.17	0.82	0.87	1.34	0.89	1.12	6.13	6.07	5.24

Table D.6 Relative abundances of ASVs in the cyanobacterial community in Big Turkey Lake across a multi-seasonal period.

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
Chroococcales										
ASV884	0.00	0.97	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV885	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV886	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.45	N/A
ASV887	0.00	0.48	0.00	0.00	0.00	0.00	0.00	0.49	2.62	N/A
ASV890	0.96	1.15	0.46	0.00	0.00	0.00	0.00	0.27	2.90	N/A
ASV891	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24	0.40	N/A
ASV893	0.79	0.00	0.74	0.00	0.00	0.00	0.00	0.00	0.86	N/A
ASV903	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV904	0.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV906	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83	1.74	N/A
ASV908	0.00	0.00	1.27	0.00	0.00	0.00	0.00	1.97	6.71	N/A
ASV910	0.34	0.26	0.32	0.00	0.00	0.00	0.00	0.00	0.28	N/A
ASV912	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV914	0.00	0.30	0.35	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV915	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.43	N/A
ASV916	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.90	1.18	N/A
ASV917	0.00	0.63	1.31	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV919	0.86	10.03	4.13	0.00	0.00	0.22	0.53	6.78	11.54	N/A
Chroococcidiopsidales										
ASV828	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Cyanobacteriales										
ASV882	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	N/A
Nostocales										
ASV819	0.00	0.00	0.88	0.00	0.00	0.00	0.00	0.00	0.00	N/A

Table D.6 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV822	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.09	0.00	N/A
ASV824	0.00	0.00	0.49	0.00	0.00	0.00	0.00	0.00	0.28	N/A
Synechococcales										
ASV2802	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	N/A
ASV834	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	N/A
ASV836	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	N/A
ASV837	4.64	0.00	1.77	0.00	0.00	4.70	3.25	1.26	0.00	N/A
ASV839	1.92	4.22	2.15	0.00	0.00	0.00	0.90	2.88	1.11	N/A
ASV840	0.40	4.48	3.14	0.00	0.00	0.00	0.00	0.24	0.36	N/A
ASV841	4.59	3.68	7.10	25.86	38.64	7.42	15.24	12.73	0.71	N/A
ASV842	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV844	33.00	3.16	11.51	51.72	56.82	18.66	27.49	9.91	0.00	N/A
ASV845	3.02	0.00	4.98	0.00	0.00	3.72	9.83	4.09	0.00	N/A
ASV846	34.92	42.75	16.60	0.00	0.00	5.81	21.78	26.98	23.29	N/A
ASV848	2.44	8.42	9.36	0.00	0.00	0.00	3.45	7.08	9.18	N/A
ASV850	1.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV852	0.00	0.00	0.00	0.00	0.00	0.84	0.00	0.00	0.00	N/A
ASV853	1.03	2.49	0.85	0.00	0.00	0.00	0.00	0.00	1.06	N/A
ASV855	0.00	3.81	0.00	0.00	0.00	0.00	0.00	3.37	9.50	N/A
ASV859	0.00	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	N/A
ASV860	0.97	0.00	5.23	0.00	0.00	8.56	0.87	0.51	0.00	N/A
ASV861	0.00	0.00	0.00	0.00	0.00	2.01	0.00	0.00	0.00	N/A
ASV864	3.17	0.00	5.61	0.00	0.00	1.66	8.80	3.45	0.41	N/A
ASV865	1.98	7.58	3.14	0.00	0.00	0.00	0.53	4.79	3.68	N/A

Table D.6 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV866	1.77	3.07	6.85	22.41	0.00	9.10	4.67	6.05	2.09	N/A
ASV867	0.35	0.00	7.73	0.00	0.00	36.92	0.99	0.47	0.00	N/A
ASV869	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61	0.00	N/A
ASV871	0.16	0.00	0.32	0.00	0.00	0.22	0.91	0.37	0.00	N/A
ASV872	0.00	0.00	0.18	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV873	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	N/A
ASV874	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	N/A
ASV875	0.00	0.00	0.00	0.00	4.55	0.00	0.00	0.00	0.00	N/A
ASV878	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	N/A
ASV897	0.19	0.58	1.91	0.00	0.00	0.00	0.00	2.01	13.53	N/A
ASV898	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.83	N/A
ASV899	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.11	0.00	N/A
ASV900	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	N/A
ASV924	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	N/A
ASV928	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Caenarcaniphilales										
ASV803	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV804	0.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Obscuribacterales										
ASV935	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
ASV937	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	N/A
ASV939	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Vampirovibrionales										
ASV801	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.15	N/A
ASV802	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.50	N/A

Table D.6 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV932	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	N/A
Unknown										
ASV909	0.25	1.49	0.64	0.00	0.00	0.00	0.00	0.27	0.71	N/A

Table D.7 Relative abundances of ASVs in the cyanobacterial community in Little Turkey Lake across a multi-seasonal period.

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
Chroococcales										
ASV884	0.00	0.16	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV887	0.12	0.72	0.00	N/A	N/A	0.00	0.00	0.12	0.09	0.00
ASV889	0.36	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV890	1.32	0.78	0.00	N/A	N/A	0.00	0.35	0.47	2.32	0.00
ASV893	1.14	1.50	0.00	N/A	N/A	0.00	0.00	0.00	4.02	42.86
ASV904	0.00	1.17	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV908	0.26	0.00	0.00	N/A	N/A	0.00	0.00	1.65	1.10	0.00
ASV910	0.09	0.62	0.00	N/A	N/A	0.00	0.15	0.39	0.00	0.00
ASV915	0.00	0.00	0.00	N/A	N/A	0.00	0.13	0.00	0.00	0.00
ASV916	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.35	0.00	0.00
ASV917	0.00	0.69	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV919	7.05	5.58	0.00	N/A	N/A	0.00	3.13	7.51	4.33	0.00
Nostocales										
ASV815	0.17	0.00	0.00	N/A	N/A	0.00	0.19	0.00	0.00	0.00

Table D.7 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV821	0.00	0.00	1.78	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV822	0.26	0.00	0.00	N/A	N/A	0.00	1.01	0.00	0.11	0.00
ASV823	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.11	0.00
Synechococcales										
ASV806	0.04	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV839	0.62	1.31	0.00	N/A	N/A	0.00	3.56	2.79	1.77	0.00
ASV842	0.12	0.29	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV843	0.00	0.00	0.00	N/A	N/A	0.00	2.54	1.39	0.00	0.00
ASV844	0.00	0.00	11.03	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV846	34.55	24.18	22.42	N/A	N/A	0.00	26.09	10.23	16.21	0.00
ASV848	25.20	25.48	0.00	N/A	N/A	0.00	17.34	23.53	26.15	0.00
ASV849	0.00	0.00	0.00	N/A	N/A	0.00	0.00	3.65	0.00	0.00
ASV850	3.31	5.97	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
ASV851	3.01	1.76	0.00	N/A	N/A	0.00	0.93	0.00	0.00	0.00
ASV853	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.19	4.48	0.00
ASV854	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.03	0.00	0.00
ASV855	1.20	0.00	0.00	N/A	N/A	0.00	3.43	24.27	17.31	0.00
ASV859	0.00	0.00	15.66	N/A	N/A	7.69	0.32	0.58	0.00	0.00
ASV865	13.89	20.10	0.00	N/A	N/A	0.00	24.67	14.05	9.05	0.00
ASV866	3.18	7.18	44.13	N/A	N/A	92.31	15.25	2.85	0.00	0.00
ASV868	0.00	0.00	0.00	N/A	N/A	0.00	0.00	1.89	6.32	0.00
ASV869	0.63	0.00	0.00	N/A	N/A	0.00	0.00	0.98	0.00	0.00
ASV879	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.11	0.00	0.00
ASV888	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.06	0.00
ASV897	0.79	0.00	0.00	N/A	N/A	0.00	0.17	1.18	2.96	0.00

Table D.7 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV898	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.00	2.62	0.00
ASV899	0.72	1.70	0.00	N/A	N/A	0.00	0.26	1.45	0.53	0.00
Obscuribacterales										
ASV3420	0.00	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.00	57.14
Vampirovibrionales										
ASV801	0.08	0.00	0.00	N/A	N/A	0.00	0.00	0.00	0.00	0.00
Unknown										
ASV909	1.89	0.82	4.27	N/A	N/A	0.00	0.47	0.34	0.46	0.00

Table D.8 Relative abundances of ASVs in the cyanobacterial community in Wishart Lake across a multi-seasonal period.

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
Chroococcales										
ASV890	2.07	0.00	0.00	0.00	N/A	0.00	0.31	0.00	0.00	0.00
ASV894	0.12	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV895	0.17	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV904	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.19	0.81	0.00
ASV906	0.61	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV910	0.37	0.00	0.00	0.00	N/A	0.00	0.69	1.39	1.18	0.00
ASV911	0.17	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV912	0.00	0.00	1.49	0.00	N/A	0.00	0.00	0.00	0.00	0.00

Table D.8 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV913	0.66	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	4.65
ASV919	0.98	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.81	0.00
Cyanobacteriales										
ASV896	0.30	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV901	0.22	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
Gloeobacterales										
ASV929	0.15	0.73	0.00	0.00	N/A	0.00	0.00	0.08	0.31	0.00
ASV930	0.61	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
Nostocales										
ASV816	0.00	0.00	0.00	0.00	N/A	0.00	0.06	0.00	0.00	0.00
ASV817	0.00	0.00	0.75	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV822	0.00	0.00	0.00	0.00	N/A	0.00	0.65	0.00	0.00	0.00
ASV825	0.12	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
Synechococcales										
ASV806	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.50	0.00
ASV807	0.05	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV809	0.71	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV810	0.17	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV813	0.07	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV835	0.07	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00

Table D.8 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV838	2.36	2.67	0.00	0.00	N/A	0.00	0.00	0.00	1.92	0.00
ASV839	7.97	5.34	0.00	0.00	N/A	0.00	4.88	0.43	1.74	0.00
ASV842	2.29	0.00	0.00	0.00	N/A	0.00	0.27	0.96	2.17	0.00
ASV843	1.08	6.80	0.00	0.00	N/A	0.00	18.78	1.34	2.61	0.00
ASV844	0.00	0.00	9.70	0.00	N/A	0.00	0.00	0.27	0.00	0.00
ASV846	7.70	8.98	0.00	0.00	N/A	0.00	7.61	21.47	21.29	0.00
ASV848	30.19	35.56	30.97	0.00	N/A	29.17	27.02	23.72	48.85	25.58
ASV849	3.25	0.00	0.00	0.00	N/A	0.00	0.00	3.03	0.00	0.00
ASV850	5.53	0.00	0.00	0.00	N/A	0.00	2.03	16.95	4.90	0.00
ASV853	0.61	2.31	0.00	0.00	N/A	0.00	0.00	3.03	0.00	0.00
ASV855	5.83	11.65	0.00	0.00	N/A	0.00	0.00	8.57	0.00	0.00
ASV856	0.47	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV857	0.64	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV858	0.49	5.95	0.00	0.00	N/A	0.00	0.00	0.56	0.00	0.00
ASV859	0.98	1.70	16.79	0.00	N/A	23.61	8.07	5.81	5.15	0.00
ASV862	1.60	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV865	0.61	0.00	0.00	0.00	N/A	0.00	5.89	0.00	0.00	0.00
ASV866	9.81	5.46	32.09	0.00	N/A	44.44	19.29	3.99	1.06	0.00
ASV869	2.14	4.25	0.00	0.00	N/A	0.00	2.79	5.44	2.36	0.00
ASV870	0.17	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00

Table D.8 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV876	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.05	0.00	0.00
ASV898	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.74	0.00
ASV899	6.88	2.55	0.00	0.00	N/A	0.00	0.52	0.43	0.00	0.00
ASV922	0.17	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV924	0.12	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV925	0.05	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV926	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.19	0.00
ASV927	0.52	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	0.00
Caenarcaniphilales										
ASV3419	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	11.63
Obscuribacterales										
ASV3421	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	18.60
ASV3422	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	20.93
ASV934	0.00	0.00	0.00	100.00	N/A	0.00	0.00	0.00	0.00	0.00
ASV936	0.00	0.00	0.00	0.00	N/A	2.78	0.00	0.00	0.00	0.00
Sericytochromatia										
ASV3373	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	13.95
ASV3374	0.00	0.00	0.00	0.00	N/A	0.00	0.00	0.00	0.00	4.65
Unknown										
ASV826	0.00	0.00	0.00	0.00	N/A	0.00	0.04	0.00	0.00	0.00

Table D.8 Continued

ASV ID	Jul-18	Aug-18	Oct-18	Feb-19	Mar-19	May-19	Jun-19	Jul-19	Aug-19	Jan-20
ASV909	0.89	6.07	8.21	0.00	N/A	0.00	1.07	2.30	3.41	0.00