THE DISTRIBUTION AND PREDICTORS OF CYANOBACTERIA AND THEIR TOXINS ACROSS A WIDE RANGE OF CANADIAN LAKES

Paul William MacKeigan

Department of Biology

McGill University, Montréal

May 2023

A thesis submitted to McGill University in partial fulfillment of requirements of the degree of

Doctor of Philosophy

TABLE OF CONTENTS

ACKNOWLEDMENTS

Words cannot express my gratitude and appreciation to my supervisor, Dr. Irene Gregory-Eaves for her motivation and support over the last few years. I could not have undertaken taken this journey without your patience, direction and encouragement. Thank you for being as involved as you were and having such a hands-on approach. And thank you for providing me with numerous opportunities to present at conferences, give guest lectures and attend workshops, all of which have given me new skills and contributed to my growth as a researcher. I'll truly miss our weekly meetings and your advice. Thank you for everything.

I would like to express my deepest appreciation to my co-supervisor, Dr. Beatrix Beisner. Thank you for your mentorship, feedback, and rapid turn-around on everything I've sent you. It was always constructive and truly valuable. Thank you for your continuous support. I would also like the thank my committee members, Dr. Frances Pick and Dr. Jennifer Sunday. It was an honor to communicate and work with such esteemed scientists. I always appreciated our annual meetings, critical questions, thoughtful discussions and prompt feedback whenever I needed it. Your ideas and suggestions were always insightful and greatly improved my work. Thank you for taking the time.

To the past and present members of the Gregory-Eaves lab group, it was a pleasure working alongside all of you. Thank you for your friendship and creating a positive, constructive and fun learning environment. I'm glad to have met all of you. In particular, I would like to acknowledge Rebecca Garner for not only teaching me all things molecular but answering all of my technical questions with bioinformatic processing. This thesis would not have been possible without all your help and dedication. I am forever grateful.

I truly believe Montréal and Québec are the best places on earth to study limnology. Thank you to the several funding sources made widely available to students in this province, including NSERC, FRQNT and Écolac. Thank you to the GRIL for fostering an incredibly collaborative learning environment for young professionals in aquatic ecology. Their conferences and workshops were some of my favorite moments during my PhD.

This endeavor would not have been possible without the LakePulse network. I am grateful to Yannick Huot for spearheading this remarkable lake sampling campaign and it was a pleasure to be a part of. Thank you to my and all of the sampling teams. The efforts of so many students will undoubtably have a long-lasting impact on the study of Canadian lakes. I'll forever have memories of visiting some of Canada's most remarkable landscapes.

Thank you to all of my co-authors, especially Zofia Taranu. Thank you for always taking the time to chat stats, it's been an honor to learn from the best. I am also thankful to the many collaborators I got the opportunity to work with on several projects. In particular, thank you to Dr. Isabelle Domaizon for welcoming me to her facilities on the banks of Lake Geneva. Living in such a picturesque part of the world and working with a close team of dedicated scientists was one of the best experiences of my life. I'd like to recognize Cécile Chardon for taking the time to teach me qPCR and various molecular techniques, as well as ameliorating my French speaking skills. At the Canada Centre for Inland Waters, I would like to thank Dr. Arthur Zastepa for allowing me to use his research facilities. Big thanks to Marianne Racine and Anqi Liang for teaching me how to analyze cyanotoxins and process so many samples. It was a pleasure working with all of you.

Lastly, I am extremely grateful to so many friends and family, without whom this journey would not have been possible. Thank you to my incredible roommates, particularly Michelle

Gros. The hardships of the pandemic were made easier with your company. Most importantly, thank you to my parents for your continuous encouragement and weekly calls. Thank you for your patience throughout this journey and your enthusiasm along each step of the way. I hope I'll have made you proud.

ABSTRACT

The proliferation of cyanobacteria and harmful algal blooms have long been considered a major threat to water quality, human and ecosystem health, and freshwater biodiversity. Over recent decades, increases in the frequency and magnitude of blooms have been documented in many lakes worldwide, which in turn has galvanized efforts to identify the underlying drivers of cyanobacteria biomass and their community structure. Despite the attention on cyanobacteria as a target metric of water quality, there has been no systematic sampling program or complementary examination of the factors influencing cyanobacterial abundance and their toxins in lakes at a national scale in Canada. My PhD thesis utilizes a large-scale lake sampling program involving hundreds of lakes to provide a broad view of contemporary dynamics of cyanobacteria across Canada. Specifically, it aims to (1) quantify the spatial patterns of cyanobacteria biomass and identify their predictors; (2) assess the congruency between traditional and modern methods of generating community composition; and (3) quantify the abundance of cyanotoxins (microcystins) and their predictors. In my first chapter, I showed that cyanobacteria biomass was significantly higher in central Canada (within the Prairies and Boreal Plains ecozones), where several bloom-forming and toxin-producing taxa dominated (i.e., *Aphanizomenon*, *Microcystis* and *Dolichospermum*). Through a series of empirical modelling approaches, I showed that despite considering an exhaustive range of variables, the results corroborate earlier studies identifying nutrients, particularly phosphorus, as the best predictor of cyanobacteria biomass and community composition. In this chapter, I was also able to shed light on an often overlooked component of cyanobacteria predictive models; biotic variables. Specifically, I found daphnid and copepod biomass were positively related to cyanobacteria

biomass, which could be due to a variety of factors including predation release of zooplankton, alternative food sources for zooplankton and/or evolutionary adaptations. Overall, given that a large matrix of potential predictors were considered, and yet the best models explained only ~50% of the variation of cyanobacteria, I believe that we have potentially reached a maximum in predictive power in analysing snapshot samples on morphologically identified cyanobacteria. In my second chapter, I conducted a comparative analysis using two different methods of taxonomic assignment of cyanobacteria communities: traditional light microscopy and contemporary DNA metabarcoding. Using a 370 lake-dataset, I quantified the congruency between cyanobacteria communities generated by each taxonomic method and found a moderate but significant agreement between methods. The two methods were most congruent in the nutrient-rich lake subset, and at coarser levels of taxonomic assignment. Overall, we demonstrated that microscopy and DNA metabarcoding do not necessarily yield identical cyanobacteria communities but give complementary information. Finally, in my third chapter, I investigated the concentration, distribution and predictors of total microcystins and characterized their specific congener profiles across Canadian lakes. Microcystin concentrations were generally low relative to regulatory guidelines, but high values were observed in the Prairies and Boreal Plains lakes. A key finding of this chapter was that a variety of less-commonly found microcystin congeners were present across Canadian lakes, but the two most abundant were MC-LR and MC-LA. Overall, the relative abundance of congener composition was only moderately correlated to environmental variables, suggesting other factors regulate their production. Overall, my PhD thesis provided key data to evaluate cyanobacteria abundances and composition across a large-scale landscape of Canadian lakes. I have demonstrated the overwhelming importance of phosphorus and other nutrients as the leading predictors of cyanobacteria and their toxins. The

large geographic extent and standardized quality of my dataset make the model results and comparative analyses more generalizable to researchers and managers of temperate, boreal and subarctic watersheds. Finally, my thesis highlighted the strengths and weaknesses of snapshot sampling, of which is also relevant to lake management decisions.

RÉSUMÉ

La prolifération des cyanobactéries et des efflorescences d'algues nuisibles est depuis longtemps considérée comme une menace pour la qualité de l'eau, la santé humaine et des écosystèmes et la biodiversité des eaux douces. Au cours des dernières décennies, l'augmentation de la fréquence et magnitude des efflorescences a été documentée dans plusieurs lacs mondiaux, ce qui a galvanisé les efforts visant à identifier les facteurs qui promouvaient la biomasse des cyanobactéries et de leur structure communautaire. Malgré l'attention aux cyanobactéries comme mesure cible de la qualité de l'eau, il n'y a pas eu de programme d'échantillonnage systématique ou d'examen complémentaire des facteurs influençant l'abondance des cyanobactéries et de leurs toxines dans les lacs à l'échelle nationale au Canada. Ma thèse de doctorat utilise un programme d'échantillonnage de lacs à grande échelle avec des centaines de lacs pour fournir un portrait global des dynamiques contemporaine des cyanobactéries à travers le Canada. Spécifiquement, elle vise à (1) quantifier les modèles spatiaux de la biomasse des cyanobactéries et à identifier leurs prédicteurs; (2) évaluer la congruence entre les méthodes traditionnelles et modernes pour générer la composition des communautés; et (3) quantifier l'abondance des cyanotoxines (microcystines) et leurs prédicteurs. Dans mon premier chapitre, j'ai démontré que la biomasse des cyanobactéries était significativement plus élevée dans le centre du Canada (dans les écozones des Prairies et des Plaines boréales), ou plusieurs taxons formant des efflorescences et produisant des toxines dominaient (c'est-à-dire *Aphanizomenon*, *Microcystis* et *Dolichospermum*). Avec une série d'approches de modélisation empirique, j'ai démontré que malgré la prise en compte d'une gamme exhaustive de variables, les résultats corroborent avec les études antérieures identifiant

les nutriments, en particulier le phosphore, comme le meilleur prédicteur de la biomasse et de la composition des communautés de cyanobactéries. Dans ce chapitre, j'ai également été en mesure d'éclairer sur une composante souvent négligée des modèles de prédiction des cyanobactéries: les variables biotiques. Plus précisément, j'ai constaté que la biomasse des daphnies et des copépodes était positivement liée à la biomasse des cyanobactéries, ce qui pourrait être dû à une variété de facteurs, notamment la libération de la prédation de zooplancton, les sources de nourriture alternatives pour le zooplancton et/ou les adaptations évolutives. Dans l'ensemble, étant donné qu'une grande matrice de prédicteurs potentiels a été prise en compte et que les meilleurs modèles ont expliqué ~50% de la variation des cyanobactéries, je pense que nous avons potentiellement maximisé le pouvoir prédictif dans l'analyse des échantillons ponctuels sur les cyanobactéries morphologiquement identifiées. Dans mon deuxième chapitre, j'ai mené une analyse comparative en utilisant deux méthodes différentes d'assignation taxonomique des communautés de cyanobactéries: la microscopie optique traditionnelle et le métabarcodage de l'ADN. En utilisant un ensemble de données sur 370 lacs, j'ai quantifié la congruence entre les communautés de cyanobactéries générées par chaque méthode taxonomique et l'accord modéré mais significatif entre les méthodes. Les deux méthodes étaient plus congruentes dans le sous-ensemble des lacs riches en nutriments, et à des niveaux plus grossiers d'assignation taxonomique. Dans l'ensemble, nous avons démontré que la microscopie et le métabarcodage de l'ADN ne donnent pas nécessairement des communautés de cyanobactéries identiques mais fournissent des informations complémentaires. Enfin, dans mon troisième chapitre, j'ai étudié la concentration, la distribution et les prédicteurs des microcystines totales et caractérisé les profils de leurs congénères spécifiques dans les lacs canadiens. Les concentrations de microcystines étaient généralement faibles par rapport aux directives réglementaires, mais plus élevées dans les

lacs des Prairies et des Plaines boréales. L'une des principales conclusions de ce chapitre est qu'une variété de congénères de microcystines moins courante étaient présente dans les lacs canadiens, mais que les deux plus abondants étaient MC-LR et MC-LA. En général, l'abondance relative des congénères n'était que modérément corrélée aux variables environnementales, ce qui suggère que d'autres facteurs régulent leur production. En général, ma thèse de doctorat a fourni des données essentielles pour évaluer l'abondance et la composition des cyanobactéries à une plus grande échelle spatiale dans les lacs canadiens. J'ai démontré l'importance écrasante du phosphore et d'autres nutriments comme principaux prédicteurs des cyanobactéries et de leurs toxines. La grande étendue géographique et la qualité standardisée de mon ensemble de données rendent les résultats du modèle et les analyses comparatives plus généralisables aux chercheurs et aux gestionnaires des bassins versants tempérés, boréaux et subarctiques. Enfin, ma thèse a mis en évidence les forces et les faiblesses de l'échantillonnage ponctuel, qui est également pertinent dans les processus décisionnels en gestion des lacs.

PREFACE

Thesis format and style

This is a manuscript-based thesis in accordance with McGill University's regulations. It is comprised of three manuscripts numbered Chapter I, Chapter II and Chapter III, all of which have been published in a peer-reviewed academic journal. The thesis begins with a general introduction, which provides background information, context and the objectives of the thesis. Each chapter is linked by a connecting statement and the general conclusions highlight the most significant findings of the thesis. Supplementary materials are included as appendices for each of the three chapters.

The three manuscripts that comprise the body of the thesis are as follows:

MacKeigan, P. W., Z. E. Taranu, F. R. Pick, B. E. Beisner, and I. Gregory-Eaves. 2023. Both biotic and abiotic predictors explain significant variation in cyanobacteria biomass across lakes from temperate to subarctic zones. Limnol. Oceanogr. 1–16. doi:10.1002/lno.12352

MacKeigan, P. W., R. E. Garner, M. È. Monchamp, and others. 2022. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. Harmful Algae **113**. doi:10.1016/j.hal.2022.102187

MacKeigan, P. W., A. Zastepa, Z. E. Taranu, J. A. Westrick, A. Liang, F. R. Pick, B. E. Beisner, and I. Gregory-Eaves. 2023. Microcystin concentrations and congener composition in relation to

environmental variables across 440 north-temperate and boreal lakes. Sci. Total Environ. **884**. doi:10.1016/j.scitotenv.2023.163811

To make the formatting style consistent throughout this thesis, references and the numbering of figures and tables have been modified to follow the formatting guidelines from the journal *Limnology and Oceanography*. The use of first-person plural refers to all co-authors included in each chapter. First person singular is used in all other sections of the thesis. All tables, figures, and appendices related to each chapter can be found at the end of each respective chapter.

Contribution of Authors

Each chapter was conceptualized and written in close collaboration with my supervisors: Irene Gregory-Eaves and Beatrix Beisner. I contributed to formulating the overarching research goals and developed the hypotheses. The data used in each chapter was generated from samples collected by the NSERC Canadian LakePulse Network (LakePulse Network), of which I participated in the field work. I led the formal analyses including much of the laboratory work, performed all statistical analyses and wrote the original drafts of all manuscripts. Co-authors contributed ideas to each chapter, provided material resources and critically reviewed and edited each manuscript. Much of the funding was acquired by Irene Gregory-Eaves and Beatrix Beisner and the rest of the LakePulse Network.

Chapter I

The original project idea was conceptualized by Irene Gregory-Eaves and Beatrix Beisner. I was responsible for curating all of the metadata, creating the statistical framework, data analyses and writing of the original manuscript. I, along with the LakePulse sampling teams, conducted all of the field work to generate the dataset. The curated zooplankton data was provided by Cindy Paquette. Zofia Taranu and Frances Pick, along with all co-authors, contributed to reviewing and editing the manuscript.

Chapter II

The framework for chapter II was developed by Irene Gregory-Eaves and I. The extraction of DNA was performed by Vera Onana, Rebecca Garner and Susanne Kraemer in David Walsh's laboratory at Concordia University. Library preparation and 16S rRNA sequencing was performed in Jesse Shapiro's laboratory at the University of Montreal by Vera Onana, Susanne Kraemer and Naíla Barbosa da Costa. All phytoplankton counts were performed by Michael Agbeti. I performed the bioinformatic and statistical analyses, in consultation with Rebecca Garner and Marie-Ève Monchamp. All participating authors were involved in reviewing and editing the manuscript.

Chapter III

The framework for chapter III was developed by Irene Gregory-Eaves and my supervisors. I prepared all of the cyanotoxin extracts and conducted measurements of total microcystin concentrations in Arthur Zastepa's laboratory at the Canadian Centre for Inland Waters in Burlington, Ontario. Anqi Liang taught me how to perform the ELISA method and performed the analysis on a few remaining samples that needed to be re-run. I selected and

prepared the samples to for microcystin congener analysis, which was performed in Judy Westrick's laboratory at Wayne State University. All participating authors were involved in reviewing and editing the manuscript.

Statement of Originality

As a consequence of ongoing eutrophication and climate warming, cyanobacteria remain a persistent threat to many freshwater lakes around the world. The research projects that comprise this thesis are an original and exhaustive in their evaluation of the structure of cyanobacteria communities as well as of the predictors of cyanobacteria biomass and their toxins across Canadian lakes. Grounding the analyses in a large-scale lake sampling program with standardized sampling, my work provides a comprehensive consideration of both abiotic and biotic variables (the latter of which are often ignored). Across these chapters, the strength of this body of work lies in: 1) the geographic scope; 2) the standardization of variable collection (e.g., a single taxonomist completed all of the phytoplankton counts); 3) the wide range of predictors considered for both the cyanobacteria and toxin response models; and 4) the comparative analysis of traditional vs. emerging genetic approaches.

Chapter I

To date, cyanobacteria response models from large-scale datasets have provided support for increasing nutrients and climate change related effects as the leading drivers of cyanobacteria biomass (Huisman et al. 2018). However, large-scale analyses are often restricted to just a few predictors, and as a result, predictive power is generally low. Furthermore, several biotic processes have not been widely considered, and as such it was not clear when I started my thesis

if these predictors could improve models substantially. In chapter I, we quantified the composition of cyanobacteria communities from 640 lakes across Canada and used a series of empirical modelling techniques to identify the best predictors of cyanobacteria biomass and the biomasses of several key toxin and bloom-forming taxa. Despite the inclusion of an exhaustive suite of approximately 50 biotic and abiotic predictors, we found that cyanobacteria biomass was overwhelmingly explained by nutrients, primarily total phosphorus. While some top predictors varied by cyanobacteria genus, all models included total phosphorus. An additional key finding was that although biotic predictors of cyanobacteria have been overlooked in many large-scale studies, we found strong, positive relationships between cyanobacteria biomass and the biomass of daphnids and copepods. Although contrary to the traditional viewpoint, there is evidence that zooplankton groups may be indirectly promoting cyanobacteria growth through several mechanisms. Overall, this chapter is relevant for lake and watershed managers as our models emphasize the importance of regulating nutrients, especially phosphorus for controlling cyanobacteria biomass. Additionally, we identified limited regional variation within models suggesting the cyanobacteria response models are consistent across temperate to subarctic ecosystems, and therefore can provide foundational information to lakes across the country.

Chapter II

The conventional approach for cyanobacteria identification and enumeration is morphology-based taxonomy with light microscopy. In recent decades, the development of high-throughput sequencing technologies has led to the greater use of DNA metabarcoding, which can increase sample processing time, reduce costs and circumvent biases associated with microscopy. In view of the fact that both methods are used to quantify the composition of

cyanobacteria communities today, it is important to evaluate how well traditional and genetic methods compare. Considering certain taxa are capable of forming blooms and producing toxins, comparisons are particularly useful to those who are making decisions which method to use. Chapter II addresses this gap by analyzing hundreds of lakes from the LakePulse dataset, where DNA and phytoplankton samples were taken in parallel. We observed moderate congruence between DNA metabarcoding and microscopy, with the highest level of concordance occurring at coarser levels of taxonomic identification (i.e., Order) and when only lakes of higher nutrient state were considered. Since cyanobacteria bloom in eutrophic and hypereutrophic lakes, my work highlights that DNA metabarcoding could be a meaningful way to monitor eutrophication. However, DNA metabarcoding across all lakes tend to give complementary rather than identical data, as this genetic approach is more sensitive to the detection pico-sized taxa. Meanwhile, microscopy can account for taxa that do not as yet have representation in DNA reference databases. This chapter also reviews many of the biological and technical issues that may influence the congruency between methods, of which provides guiding information to managers.

Chapter III

A major concern arising from cyanobacterial blooms is the ability of some species to produce an array of toxins that have water quality and health implications for humans and lake biota. Hepatotoxin microcystins are among the most commonly cyanotoxins found in lakes. As a result of their toxicity and potential increase in lakes worldwide, there has been a widespread effort to incorporate them into monitoring programs and identify their drivers. To date, field studies have produced varied results when attempting to determine which environmental conditions are linked to microcystin production. This could be due to the uncertainty about

microcystins' physiological functions or a lack of standardized sampling. In chapter III, I conducted a broad-scale field study of microcystins in 440 lakes across Canada. We first quantified the concentration of total microcystins, then modeled the occurrence and concentration of microcystin using wide set of biotic and abiotic predictors. We found that the percentage of Canadian lakes with detectable microcystins low and similar to other temperate regions. We also showed that the best predictors of microcystins were variables related directly to cyanobacteria biomass (i.e., phosphorus and the biomass of particular genera). Perhaps the most novel addition of chapter III is microcystin congener analyses, which is one of the largest datasets available. We targeted 14 microcystin congeners using a tandem liquid chromatography-mass spectrometry technique, across a 190-lake subset. Among the key findings were that the two most toxic forms, MC-LR and -LA were the most widely detected, but a variety of less commonly found congeners were also present across Canadian lakes. The relative composition of congeners was only moderately correlated with environmental variables. Overall, this chapter provides key data and models that are insightful to researchers as well as lake managers.

References

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

GENERAL INTRODUCTION

Cyanobacteria and harmful algal blooms

Cyanobacteria (formerly referred to as blue-green algae) are a diverse group of photosynthetic prokaryotes. They are among the oldest known oxygen-producing organisms on Earth, with fossil records dating 3 to 3.5 billion years ago (Schopf 2012; Whitton 2012). Their photosynthetic activity is thought to have significantly altered the Earth's atmosphere, resulting in the Great Oxygenation Event some 2.4 billion years ago (Garcia-Pichel et al. 2019; Chorus and Welker 2021). Cyanobacteria are omnipresent and represent one of the most abundant organisms on Earth with an estimated thousand million metric tons (10^{15}; ~1 million Petagrams) of wet biomass (Garcia-Pichel et al. 2003; Bonilla and Pick 2017). Due to their long evolutionary history, they have evolved to inhabit a wide range of environmental conditions, spanning tropical and polar regions. They can occur in terrestrial habitats such as deserts and soil, but are especially capable of dominating phytoplankton assemblages in aquatic ecosystems (Whitton 2012; Bonilla and Pick 2017). The cyanobacteria phylum is highly diverse taxonomically, morphologically and in their levels of biological organization. There are an estimated 6,000-8,000 species, many of which have yet to be discovered (Guiry 2012; Guiry and Guiry 2020). Taxa vary substantially in size, from picocyanobacteria (~0.6 μm) (detection limit of light microscopy) to taxa that form large colonies visible to the naked eye (~5 mm) (Bonilla and Pick 2017). Cyanobacteria can also exist in the pelagic or littoral zones of lakes either as single cells (unicellular), colonies or filaments. Traditionally, their taxonomic classification was based on morphological characteristics such as cell size, shape and the presence of specialized cells (Rippka et al. 1979). More recently, the advancement of molecular tools has identified new

species and continues to revise previous taxonomic organization (Komárek 2016). In addition to being taxonomically and morphologically diverse, they possess a number of physiological traits and tolerances that make them highly competitive over eukaryotic phytoplankton in aquatic ecosystems (Mantzouki et al. 2016).

Although cyanobacteria are a natural component of phytoplankton communities in all lakes, under certain conditions, they can form dense aggregations commonly referred to as blooms. Throughout the water column, it is possible for them to form blooms at least temporarily in lakes of all trophic levels (Paerl and Paul 2012). Over the course of their evolutionary history, cyanobacteria have developed an extensive set of traits that enable them to form blooms and dominate phytoplankton assemblages (Carey et al. 2012; Mantzouki et al. 2016; Huisman et al. 2018). Foremost, several taxa have the ability to fix atmospheric nitrogen (diazotrophy), providing them with an additional source of inorganic nitrogen when availability is low. They are the only phytoplankton capable of directly accessing this pool of nutrients. Nitrogen fixation is typically carried out inside specialized cells called heterocysts and performed by several common bloom-forming cyanobacteria, including *Dolichospermum*, *Aphanizomenon* and *Nodularia* among other. Heterocysts are characterized by thick cell walls to limit the diffusion of oxygen, which would otherwise inactivate the enzyme complex responsible for nitrogen fixation (Muro-Pastor and Hess 2012; Huisman et al. 2018). However, because this process is energetically costly and requires high light intensities, its use is limited in turbid waters and is typically restricted to conditions wherein alternative nitrogen sources are depleted (Chorus and Welker 2021).

Many cyanobacteria are also capable of regulating their buoyancy in the water column (Walsby 1994). Through the synthesis of hollow protein structures referred to as gas vesicles,

buoyant cells can regulate their carbohydrate ballast to migrate between several meters to a few millimetres over the course of a day (Molot et al. 2014). This provides a competitive advantage as it not only allows them to adjust their position in the water column and gain access to nutrients and light, but also to shade non-buoyant phytoplankton (Huisman et al. 2018).

Lastly, cyanobacteria possess complex carbon concentrating mechanisms that allow them to utilize different forms of inorganic carbon (Pick and Lean 1987; Visser et al. 2016). These include three uptake systems for bicarbonate and two for CO_2 (Visser et al. 2016; Huisman et al. 2018). During dense blooms, cyanobacteria can deplete surface waters of dissolved CO_2, which raises pH and shifts the predominant form of inorganic carbon (Huisman et al. 2018). These uptake mechanisms allow them to respond to changes in carbon availability and maintain photosynthetic rates. Additional traits that may facilitate growth include their large storage capacity for nutrients, accessory pigments for efficient light harvesting, and resting stages in the form of akinetes for overwintering or during periods of environmental stress (Mantzouki et al. 2016; Bonilla and Pick 2017).

Blooms of cyanobacteria have been of concern to lake managers and relevant stakeholders for decades as they are associated with a multitude of factors that risk human, ecosystem and economic health. The synthesis of various toxins, herein referred to as cyanotoxins, is often considered the most severe effect that cyanobacteria can have, particularly to human and animal health (Merel et al. 2013; Carmichael and Boyer 2016). Cyanotoxins are produced by many genera and exhibit a range of toxicities including hepatotoxins that can damage the liver (e.g., microcystins), neurotoxins that target the nervous system (e.g., anatoxin-a, saxitoxins) and cytotoxins that effect cellular structure in multiple organs (e.g., cylindrospermopsin) (Chorus and Welker 2021). Human intoxication may occur from a number

of exposure routes including drinking water or during recreation activities (Lévesque et al. 2014; Chorus and Welker 2021). Moreover, their harmful effects are not limited to humans as pet, livestock and other wildlife exposures from interactions with impacted waters have resulted in poisonings and animal deaths (Backer et al. 2013; Merel et al. 2013; Ash and Patterson 2022). Aside from cyanotoxins, cyanobacteria can also produce taste and odor compounds that present a risk to lakes used for drinking water and recreation. The most common of these metabolites are geosmin and 2-methylisoborneol (MIB), which produce strong earth and musty odors at relatively low concentrations (Jüttner and Watson 2007; Chorus and Welker 2021). Although these metabolites are not considered a risk to human health, they can lower public trust and increase the cost of water treatment (Chorus and Welker et al. 2021).

Cyanobacteria can alter the physical and chemical characteristics of lakes, generally associated with cultural eutrophication, which in turn affect resident biotic communities. During high growth periods, increased photosynthetic rates can deplete CO_2 concentrations, increasing the pH of surface waters. Additionally, aerobic decomposition of cyanobacteria biomass consumes oxygen, causing hypoxia and anoxia in bottom waters of stratified water bodies. These conditions can result in mortalities of fish and invertebrates, and further degrade water quality through internal nutrient loading (Paerl and Otten 2013; Molot et al. 2014). Altogether, cyanobacteria blooms impose a substantial financial burden on lake managers, communities and governments in the form of costs related to water treatment, monitoring, and losses in recreational use and property value (Dodds et al. 2008; Smith et al. 2019). For example, it is estimated that the per annum costs of freshwater algal blooms across the United States total $4.6 billion (Kudela et al. 2015; Ho et al. 2019). If bloom frequency and magnitude increase in the future, these costs will only go up.

Despite ongoing efforts to mitigate cyanobacteria development in many lakes worldwide, numerous studies predict global expansion and the rise in cyanobacteria bloom frequency and magnitude (O'Neil et al. 2012; Huisman et al. 2018). However, reports on the occurrence and negative effects of blooms are not a new phenomenon (Huisman et al. 2018). For example, Francis (1878) was the first scientific documentation of toxic cyanobacteria resulting in the mass mortality of livestock (Huisman et al. 2018). Additionally, some have argued that the perceived increase in cyanobacteria stems from heightened awareness and increased reporting (Ewing et al. 2020; Hallegraeff et al. 2021). While this may have an effect, several paleolimnological and satellite-imagery based studies report quantitative data that suggests cyanobacteria are increasing in many parts of the world in recent decades. For instance, by analyzing three decades of high-resolution satellite imagery data from large lakes around the world, Ho et al. (2019) found that peak phytoplankton bloom intensity increased over time in > 65% of the lakes considered. An analogous trend was observed by Taranu et al. (2015), who conducted an analysis of pigments from more than 100 sediment cores taken across north temperate lakes and found that cyanobacteria trends increased at a significantly greater rate than other phytoplankton groups, especially since ~1945. Overall, cyanobacteria pose a persistent threat to the health of fresh waters today. For that reason, their continues to be a need to understand the factors that promote their dominance.

Environmental factors that promote cyanobacteria

For decades, the direct drivers of cyanobacteria biomass have been extensively studied in lakes worldwide. Empirical models from regional and large-scale datasets have identified a variety of physical, chemical and biological factors that act to promote their biomass as well as

their relative dominance (Paerl and Paul 2012; Huisman et al. 2018). Historically, nutrients, which have long been recognized as the most important variables for regulating cyanobacteria biomass (O'Neil et al. 2012; Pick 2016).

The problem of cultural eutrophication persists today in many freshwater ecosystems (MacDonald et al. 2011; Huisman et al. 2018). Increasing nutrient inputs stem from a multitude of anthropogenic sources including non-point sources (i.e., urban and agricultural land), atmospheric deposition, and legacy nutrients from internal loading (Carpenter et al. 1998; Smith and Schindler 2009; Orihel et al. 2017). Cyanobacteria can become an increasingly dominant component of phytoplankton communities as a result of these increases (Watson et al. 1997).

Both total phosphorus (TP) and total nitrogen (TN) have been shown to have a significant predictive power in large-scale datasets (Downing et al. 2001; Dolman et al. 2012; Beaulieu et al. 2013), including in Canadian lakes (Beaulieu et al. 2014). In particular, TP is heavily cited as the most important predictor and key limiting nutrient across lakes (Downing et al. 2001; Giani et al. 2005; Carvalho et al. 2013). For example, using 99 large temperate lakes, Downing et al. (2001) found the risk of cyanobacteria becoming dominant increases above 20-30 µg/L of TP, and plateaus at ~100 µg/L. A similar relationship was observed in a 102-German lake study from Dolman et al. (2012), who found a sigmoidal relationship between cyanobacteria biomass and TP at ~50 µg/L. Further support for TP, particularly from anthropogenic sources comes from experimental studies (e.g., Schindler 1977) and paleolimnological evidence (Taranu et al. 2015). In large datasets, TN has also been identified as a strong predictor (Kosten et al. 2012). In a modelling analysis of over 1,100 lakes across the continental United States, Beaulieu et al. (2013) identified TN as the best predictor of cyanobacteria biomass, potentially as a result of intrinsically higher nitrogen demands and nitrogen fixation indirectly increasing TN. The role of

the N:P ratio has also been considered in predictive models and nutrient management (Downing et al. 2001). Its importance stems from resource theory, whereby nitrogen fixing cyanobacteria become dominant to offset imbalances when nitrogen is limiting (Smith 1983; Pick 2016). However, low N:P ratio may simply be the result of increasing eutrophication from phosphorus enrichment, which in turn may be directly driving the increase in biomass. Overall, nutrients, particularly TP and TN are an essential prerequisite from biomass growth (Chorus et al. 2021).

Although increasing biomass is largely attributed to nutrients, increased temperatures and climate warming have also been identified as key factors promoting cyanobacteria growth (O'Neil et al. 2012; Paerl and Paul 2012; Huisman et al. 2018). There is evidence that lakes worldwide have increased in temperature by 0.34°C/decade between 1985 and 2009 (O'Reilly et al. 2015). This can directly and indirectly promote the dominance of cyanobacteria, giving them a competitive advantage under warming conditions (Huisman et al. 2018). First, increasing temperature can increase cyanobacteria through enhanced growth rates. Several taxa, particularly *Microcystis*, have presumed higher growth optima relative to eukaryotic algae (Robarts and Zohary 1987). Temperature can also enhance the strength of thermal stratification and limit mixing, benefitting cyanobacteria that regulate their buoyancy (Wagner and Adrian 2009). Further, shorter ice cover during winter and an elongated growing season can both benefit cyanobacteria growth (Huisman et al. 2018). Indirectly, temperature can regulate the supply of nutrients through enhanced mineralization and anoxia-mediated phosphorus release during periods of increased stratification (Kosten et al. 2012; Paerl and Barnard 2020). Large-scale datasets also highlight the predictive strength of temperature (Kosten et al. 2012; Beaulieu et al. 2013). For example, across a 143 lake latitudinal transect, Kosten et al. (2012) identified temperature as the best predictor, followed by TN and TP. Globally, temperatures are projected

to increase 2°C by the end of the 21st century (IPCC 2021). Although nutrients are the prerequisite, this increase in temperature may act synergistically to promote cyanobacteria in many lakes worldwide (Taranu et al. 2012; Richardson et al. 2018).

Despite the overwhelming support for nutrients and temperature, several additional environmental factors have been identified as important secondary drivers of increasing cyanobacteria biomass. Many of which are indirectly related to nutrients and temperature. For instance, hydrological changes via intense precipitation events can trigger nutrient resuspension and increase inputs from terrestrial sources (Paerl and Paul 2012; Errat et al. 2022). Additionally, increased salinity through droughts and anthropogenic ion sources (e.g., salting of roads) have been positively linked to cyanobacteria (Amorim et al. 2020). Several taxa show a tolerance to saline conditions, including common bloom-forming genera such as *Dolichospermu*m and *Microcystis* (Merel et al. 2013). Lastly, in many large-scale analyses, cyanobacteria have been associated with high alkalinity, pH and dissolved inorganic carbon (DIC) (Carvalho et al. 2011; Huisman et al. 2018; Richardson et al. 2018). Although cyanobacteria directly influence pH, they can maintain high growth rates under these conditions by using different forms of inorganic carbon (Paerl and Paul 2012; Huisman et al. 2018).

Previous empirical models from large spatial scale datasets have identified many conditions that favor the dominance of cyanobacteria. However, models are regularly restricted to a limited number of candidate predictors, and explanatory power is often low. Furthermore, since cyanobacteria are a diverse group with different ecological niches, the one-size-fits-all model may not be appropriate for all cyanobacteria. The relative importance of key drivers has been shown to differ between taxa and functional groups (Rigosi et al. 2014; Shan et al. 2019; Vuorio et al. 2020). However, a comprehensive understanding of what promotes specific genera

is still lacking. Finally, although the issue of cyanobacterial blooms persists and appears to be worsening in many lakes worldwide, including in Canada (Pick 2016), there has not been a systematic sampling program that incorporates an expanded set of predictors across a broad range of lake types.

Interactions between cyanobacteria and zooplankton

In laboratory and field studies, conflicting results have been observed regarding how zooplankton communities respond to increasing cyanobacteria biomass (Ger et al. 2014; Ger et al. 2016). With a traditional view of eutrophic lakes, increased nutrient availability causes the phytoplankton community to become dominated by bloom-forming and toxic cyanobacteria, which are largely considered inedible to most zooplankton grazers. This inedibility distorts the transfer of energy and carbon to higher trophic levels and selects for zooplankton that can coexist with, rather than control, cyanobacteria growth (Ger et al. 2014). Here, there are three attributes that make cyanobacteria inedible to zooplankton grazers: 1) filamentous and colonial morphology resulting in mechanical interference, 2) the production of toxins and other secondary metabolites that have lethal and sub-lethal effects, and 3) their poor nutritional quality due to a lack of polyunsaturated fatty acids (PUFAs) and sterols used in zooplankton growth and reproduction (Fulton and Paerl 1987; Demott et al. 1991; Ger et al. 2014; Ger et al. 2016). These attributes can result in a decrease of larger-bodied generalist zooplankton, an increase in selective feeders and/or a shift towards grazing on more edible phytoplankton (Ger et al. 2014). Under these conditions, total zooplankton biomass may decrease due to a selection for smaller-bodied taxa with more selective feeding, such as *Bosmina* (Ghadouani et al. 2006).

In contrast, several field studies have found that cyanobacterial biomass tends to be associated with increased zooplankton abundances (Chislock et al. 2013; Cremona et al. 2018; Briland et al. 2020). For example, in a time-series analysis of a shallow eutrophic lake, Cremona et al. (2018) assessed the influence of 28 biotic and abiotic predictors of cyanobacteria biomass and found a strong, positive relationship with copepod and Cladoceran abundance. The positive association between cyanobacteria and zooplankton may stem from several tolerance traits, primarily, selective feeding and a physiological adaptation to toxins (Ger et al. 2016). Copepods show a high degree of selective feeding, opting for alternative prey when available. Many taxa use chemosensory signals such as toxins and other metabolites, or cell size as detection cues to avoid consumption (Ger et al. 2011; Ger et al. 2014; Agasild et al. 2019). Large, generalist taxa such as *Daphnia* are unable to do this and as a result, may experience a reduction in feeding rates with higher colonial cyanobacteria abundances (Ghadouani et al. 2004). Selective feeding strategies by copepods may be the mechanism behind the observed linear increase in copepod biomass with cyanobacteria detected in multiple recent field studies (Shan et al. 2019; Amorim et al. 2020).

Long-term exposure to cyanobacteria blooms in historically eutrophic lakes has been shown to select for more tolerant genotypes to toxin exposure (Sarnelle and Wilson 2005; Chislock et al. 2013; Ger et al. 2016). *Daphnia* in particular has been shown to develop tolerant genotypes, whereby clones isolated from high nutrient lakes are more likely to be tolerant and show less inhibition from exposure to toxic cyanobacteria than clones from low nutrient lakes (Sarnelle and Wilson 2005). For example, Chislock et al. (2013) found that *Daphnia* clones exposed to over 80 years of eutrophication were not only able to graze upon cyanobacteria, but also supress a *Microcystis* bloom with high microcystin concentrations. Further evidence of rapid

evolution comes from Hairston et al. (2001), who found that hatched *Daphnia* genotypes from diapaused eggs preserved in lake sediments showed more tolerance to toxic cyanobacteria when retrieved from historical periods of intense eutrophication. In addition to tolerant genotypes, zooplankton can also maintain growth with cyanobacteria through phenotypic changes (Ghadouani and Pinel-Alloul 2002), food supplementation (Briland et al. 2020) and trophic upgrading (Bec et al. 2006). To date, many of these biotic processes have not been considered in cyanobacteria predictive models, particularly large-scale studies.

Cyanotoxins and microcystins

One of the major public health concerns from modifying the frequency and magnitude of cyanobacteria blooms in many lakes worldwide is that numerous taxa are known to produce a variety of toxins and other bioactive metabolites that can be harmful (Carmichael and Boyer 2016; Chorus and Welker 2021). There exist several classes of cyanotoxins that are typically organized according to chemical structure and mode of toxicity. Structurally, they can be classified as cyclic peptides, alkaloids, amino acids and lipopolysaccharides (Chorus and Welker 2021). These chemicals can have adverse health effects on vertebrates by acting as hepatotoxins, neurotoxins, cytotoxins, dermatoxins and irritants (Chorus and Welker 2021). Different toxins can be produced within the same genus, and the same toxin can be produced across different genera (Bonilla and Pick 2017). They exhibit a range of toxicities and reports regarding human and animal poisonings are documented from lakes worldwide (Bláha et al. 2009; Buratti et al. 2017).

Among the many types of cyanotoxins, the microcystins are found globally, including in Canada (Kotak and Zurawell 2007; Loftin et al. 2016; Mantzouki et al. 2018). Microcystins are a

diverse group of toxins with producing strains found in all orders and by many common bloom-forming taxa such as *Microcystis*, *Dolichospermum* and *Planktothrix* (Huisman et al. 2018; Chorus and Welker 2021). The list of toxin producers is expanding as we learn more, and now includes taxa from planktonic and benthic habitats as well as pico-cyanobacteria such as *Aphanocapsa* (Bernard et al. 2017; Chorus and Welker 2021). Structurally, microcystins are cyclic heptapeptides comprised of a seven amino acid chain. To date, there are 275 known congeners, which differ principally at two positions with variable amino acids (Bouaïcha et al. 2019). Congener names reflect the variation at these positions, for example, MC-LR contains a Leucine (L) and an Arginine (R) at the second and fourth position of the seven amino acid chain respectively. Amongst congeners, there is considerable variation in toxicity and levels of persistence in the environment (Chernoff et al. 2020; Chorus and Welker 2021). In the environment, microcystins succumb to several fates and accumulate in different pools within ecosystems (Shingai and Wilkinson 2022) including: 1) bioaccumulation in biotic communities (Kozlowsky-Suzuki et al. 2012; Flores et al. 2018), and 2) storage in sediment (Zastepa et al. 2015). Furthermore, microcystins are subject to several degradational processes including photodegradation and biodegradation by bacteria and fungi (Chorus and Welker 2021).

Microcystins operate primarily as a liver toxin and are associated with numerous adverse acute and chronic health effects to humans and wildlife (Carmichael and Boyer 2016). They act as potent inhibitors of protein phosphatase type 1 and type 2a in the cytoplasm of liver cells. This causes an increase in the phosphorylation of proteins, resulting in cell death (Chorus and Welker 2021). Exposure to low concentrations can cause headaches, nausea, vomiting, and skin irritation (Carmichael and Boyer 2016). Meanwhile high exposure is associated with liver damage, cancer (Zhang et al. 2015), and although exceedingly rare, has been implicated in the deaths of humans

(Pouria et al. 1998), pets (Backer et al. 2013) and wildlife (Miller et al. 2010). Humans can be exposed via several pathways including oral ingestion through drinking or recreational activities, via food irrigated with contaminated water and supplements, or exposed through the inhalation of aerosols (Carmichael and Boyer 2016; Plaas and Paerl 2021). As a result of their toxic effects, many governmental bodies have established drinking water and recreational exposure guideline limits. For example, the World Health Organization (WHO) and Health Canada have drinking water guidelines of 1 µg/L and 1.5 µg/L of total microcystins, respectively.

Microcystins are produced intracellularly within the thylakoid but are released to the environment upon cell death. They are synthesized through two multi-enzyme complexes referred to as non-ribosomal peptide synthetases (NRPS) and polyketide synthase (PKS). These complexes are encoded by a large gene cluster that spans 55 kb and is comprised 10 genes (mcyA-J), organized in two operons (Chorus and Welker 2021). There has been numerous laboratory and field based studies that have attempted to identify the factors that regulate their synthesis (Kaebernick and Neilan 2001; Dai et al. 2016). Factors identified to date include nutrients, pH, salinity, light and temperature (Kaebernick and Neilan 2001; Neilan et al. 2013). From large-scale studies, microcystins are positively related to increasing total cyanobacteria biomass (Wood et al. 2012), and the biomass of specific taxa such as *Planktothrix* (Dolman et al. 2012) and *Microcystis* (Giani et al. 2005). Several studies have provided support for the role of temperature, whereby microcystin production is increased when temperatures exceed 20°C (Mowe et al. 2015; Walls et al. 2018). Higher concentrations of microcystins may be the result increased cyanobacteria growth rates at higher temperatures, or function to protect cells from oxidative stress when photosynthetic rates are high (Dziallas and Grossart 2011; Dai et al. 2016; Omidi et al. 2018). Additionally, large-scale analyses reinforce the link between microcystins

and nutrients, particularly TN (Yuan and Pollard 2014; Taranu et al. 2017; Buley et al. 2021). Microcystins are nitrogen rich compounds (14% by mass) and have been shown to increase their synthesis with there is greater nitrogen availability (Davis et al. 2009; Gobler et al. 2016). Overall, microcystins have shown contradictory responses to environmental drivers, perhaps due to their elusive physiological function (if any) (Bonilla and Pick 2017; Omidi et al. 2018).

As microcystin congeners display a range of toxicities and levels of persistence, there has been an increasing effort to understand the environmental factors that promote specific variants. Congeners can be produced by several genera and multiple variants can be produced simultaneously (Puddick et al. 2014). Despite this, most research has focused on the MC-LR congener, which was previously considered the most toxic (Diez-Quijada et al. 2019). However, broad-scale surveys have highlighted the prominence of other congeners in different regions, including MC-LA in North American lakes (Taranu et al. 2019), and MC-YR across European waterbodies (Mantzouki et al. 2018). To date, the composition of microcystin congeners has been shown to be shaped by temperature related variables (Mantzouki et al. 2018), weather and nutrients (Taranu et al. 2019) and nitrogen species (Monchamp et al. 2014). For example, the dominant congener has been shown to shift with increasing nitrogen availability, mainly from MC-LR to MC-RR; the latter of which is a more nitrogen rich compound (Van de Waal 2014). Despite the growing interest, congener specific data are missing from many regions, including most Canadian lakes. Furthermore, explanatory power for modelling congener profiles remains low.

Incorporating DNA based methods to cyanobacteria management

Accurately identifying biological assemblages is a fundamental component of ecology and essential to ecosystem management (Pawlowski et al. 2018). This is particularly true of cyanobacterial communities, as they can contain toxin-producing and other nuisance taxa that are of interest to many stakeholders (Huisman et al. 2018). Traditionally, the taxonomic classification of cyanobacteria was inferred from morphological characteristics identified by microscopy (Reynolds 2006; Chorus and Welker 2021). As previously stated, these include features like the size and shape of cells, and the presence and position of specialized cells such as heterocysts. Although identifications based on microscopy continue to be used today, this method is associated with several technical biases and limitations that may hamper the classification of cyanobacteria communities. For example, microscopy is considered a time-consuming and laborious process and there are observer biases amongst taxonomists (Lee et al. 2014; Bailet et al. 2020). Cryptic taxa and pico-cyanobacteria (0.2-2 μm) are typically overlooked (Li et al. 2019; Esenkulova et al. 2020). Finally, phenotypic changes between different environments and culture conditions add to the difficulty of deciphering cyanobacterial communities (Komárek 2006; Li et al. 2019).

With advances in molecular genetic methods, such as high-throughput sequencing, DNA-based approaches have become more widely used for quantifying and characterizing aquatic biodiversity (Hering et al. 2018; Pawlowski et al. 2018). Methods that utilize DNA in biomonitoring often rely on metabarcoding, whereby DNA is extracted from field samples and a marker gene that is taxonomically revealing for a particular group of organisms is amplified, sequenced and annotated against a reference database (Hébert et al. 2003). This approach allows researchers and managers to overcome many of the biases related to microscopy as they can detect cryptic, pico-sized and rare taxa (Pawlowski et al. 2018). For cyanobacteria, the most

widely used marker gene for analyzing their taxonomy has been the 16S rRNA gene. This gene is a structural component that encodes ribosomal RNA of the small ribosome subunit. Sequencing of the 16S rRNA gene can provide taxonomic information as it has both highly conserved regions, whereby primers can be developed to target a broad group of bacteria, such as Cyanobacteria. The gene also has nine hypervariable regions that are sensitive to genetic changes, which can be used to distinguish species (Chorus and Welker 2021). Detailed taxonomic analyses of the 16S rRNA gene have revealed that their evolutionary relationships do not necessarily correspond to morphological changes (Lane et al. 1985; Komárek 2006; Komárek 2016). Phylogenetic reconstructions have reorganized some of their taxonomy, with new names being given to many taxa in recent years (Komárek 2016).

Molecular genetic methods, particularly DNA metabarcoding, are becoming more cost-effective and faster in processing time, making these approaches increasingly useful for monitoring aquatic communities. They have been used to investigate cyanobacteria communities in lakes and drinking water reservoirs worldwide (Gao et al. 2018; Casero et al. 2019) and have been incorporated in several paleolimnological assessments to understand how their species composition has changed through time (e.g., Monchamp et al. 2016; Tse et al. 2018; Pilon et al. 2019). With the growing use of molecular methods, and the continuing use of microscopy, knowledge regarding the congruency and variability between methods is essential for researchers and managers deciding which approach is most suitable to their study. The benchmarking of molecular methods against morphological-based taxonomy has been performed on a variety of aquatic organisms including fish (Hänfling et al. 2016), macroinvertebrates (Cowart et al. 2015; Elbrecht et al. 2017; Leese et al. 2021) and phytoplankton (Eiler et al. 2013; Abad et al. 2016; Li et al. 2019; Bailet et al. 2020), which show varying degrees of congruency (Pawlowski et al.

2021). A recent meta-analysis by Keck et al. (2022) found that across 215 comparative datasets, DNA metabarcoding was more congruent with traditional taxonomic methods in larger organisms such as fish. Meanwhile with smaller organisms, such as planktonic ones, there were differences in the communities generated with each method. These comparative analyses produce vital taxonomic information regarding your community of interest. Despite this, there has not been an assessment of cyanobacteria communities generated by DNA metabarcoding and microscopy from the same samples across a broad range of lake types.

Large-scale sampling of Canadian lakes

Canadian lakes comprise a significant portion of the landscape and provide many ecosystem services. Lakes occupy 10% of the country's territory and Canada is home to 37% of Earth's total lake area (Minns et al. 2008; Messager et al. 2016; Huot et al. 2019). With over a million lakes greater than 10 hectares, Canada is faced with the difficult task of monitoring and managing the multitude of stressors that may be influencing their structure and function (Pick 2016; Huot et al. 2019). Despite this, there was no standardized sampling program of lakes across Canada until quite recently. The LakePulse Network was established to target this gap and set about to sample over 650 lakes following a standardized sampling program (Huot et al. 2019). The network's overarching goal is to assess the health status of Canadian lakes, identify their most pressing stressors and understand how they have altered aquatic communities and ecosystem services (Huot et al. 2019). Cyanobacteria and their blooms are well-described indicators of lake health and are a target metric of water quality for managers across the country. Their documented increases media report and in empirical studies across many lakes serve as

important indicators of eutrophication and climate warming; two of the most pressing issues lakes face today (Reid et al. 2018).

Thesis objectives

To address many of the limitations and research gaps that have been identified in this review, my thesis aims to generate a broad overview of cyanobacteria patterns and predictors across Canadian lakes. One of the central shortcomings towards improving our understanding of their distribution has been a lack of a standardized sampling protocol applied across Canada. My PhD thesis was part of the LakePulse Network and was designed to consider a broader range of biotic and abiotic predictors to provide a contemporary and historical understanding of the drivers of cyanobacteria and their toxins across a wide range of Canadian lakes. The objectives of the first chapter were to quantify the biomass of cyanobacteria and identify their community composition across a ~640 lake dataset. Using a very large range of predictors, this chapter utilizes a series of empirical modelling techniques to identify the best predictors of total cyanobacteria biomass. To account for the widescale applicability of models, this chapter assesses how the relationship between cyanobacteria and its predictors vary by region. The last objective of this chapter was to identify the predictors of their community composition, and since taxa occupy different ecological niches, assess the predictors of certain genera of interest.

With the advancement of molecular techniques, DNA metabarcoding has become a widely used method for measuring diversity and identifying the community composition of many aquatic groups, including cyanobacteria. With the ongoing use of both light microscopy and DNA metabarcoding, there remain questions regarding their comparability. Using a subset of lakes in Chapter I, the second chapter of my thesis quantifies the congruency between

cyanobacteria communities generated by traditional light microscopy and DNA metabarcoding. Furthermore, using the 16S rRNA sequencing data generated for this chapter, the second objective was to assess the regional variation of *Microcystis* genotypes.

Factors related to the production of particular cyanotoxins were also unclear when I began this project on 2017. The objective of the third chapter was to quantify the concentration of total microcystins (i.e., the most commonly found toxin) across a 440-lake subset. This chapter aims to identify their key predictors using the same expanded range of standardized biotic and abiotic variables and a similar modelling method from chapter I. The second objective of this chapter was to quantify the abundance of microcystin congeners and identify the environmental variables the lead to their respective dominances. To do so, I measured the concentration of 14 microcystin congeners from a 190-lake subset and performed correlational analyses that consider environmental and cyanobacterial community composition as predictors.

References

Abad, D., A. Albaina, M. Aguirre, A. Laza-Martínez, I. Uriarte, A. Iriarte, F. Villate, and A. Estonba. 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar. Biol. **163**: 149–150. doi:10.1007/s00227-016-2920-0

Agasild, H., K. Panksep, I. Tõnno, and others. 2019. Role of potentially toxic cyanobacteria in crustacean zooplankton diet in a eutrophic lake. Harmful Algae **89**. doi:10.1016/j.hal.2019.101688

Amorim, C. A., Ê. W. Dantas, and A. do N. Moura. 2020. Modeling cyanobacterial blooms in tropical reservoirs: The role of physicochemical variables and trophic interactions. Sci. Total Environ. **744**. doi:10.1016/j.scitotenv.2020.140659

Ash, A. K., and S. Patterson. 2022. Reporting of Freshwater Cyanobacterial Poisoning in Terrestrial Wildlife: A Systematic Map. Animals **12**. doi:10.3390/ani12182423

Backer, L. C., J. H. Landsberg, M. Miller, K. Keel, and T. K. Taylor. 2013. Canine cyanotoxin poisonings in the United States (1920s-2012): Review of suspected and confirmed cases from three data sources. Toxins (Basel). **5**: 1597–1628. doi:10.3390/toxins5091597

Bailet, B., L. Apothéloz-Perret-Gentil, A. Baričević, and others. 2020. Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Sci. Total Environ. **745**. doi:10.1016/j.scitotenv.2020.140948

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. **58**: 1736–1746. doi:10.4319/lo.2013.58.5.1736

Beaulieu, M., F. Pick, M. Palmer, S. Watson, J. Winter, R. Zurawell, and I. Gregory-Eaves.

2014. Comparing predictive cyanobacterial models from temperate regions. Can. J. Fish. Aquat. Sci. **71**: 1830–1839. doi:10.1139/cjfas-2014-0168

Bec, A., D. Martin-Creuzburg, and E. Von Elert. 2006. Trophic upgrading of autotrophic picoplankton by the heterotrophic nanoflagellate Paraphysomonas sp. Limnol. Oceanogr. **51**: 1699–1707. doi:10.4319/lo.2006.51.4.1699

Bernard, C., A. Ballot, S. Thomazeau, S. Malufi, A. Furey, J. Mankiewicz-Boczek, B. Pawlik-Skowronska, C. Capelli, N. Salmaso. 2017. Appendix 2: cyanobacteria associated with the production of cyanotoxins. J. Meriluoto, L. Spoof, G.A. Codd (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis, Wiley, Chichester UK (2017), 500-525. doi:10.1002/9781119068761.app2

Bláha, L., P. Babica, and B. Maršálek. 2009. Toxins produced in cyanobacterial water blooms - toxicity and risks. Interdiscip. Toxicol. **2**: 36–41. doi:10.2478/v10102-009-0006-2

Bonilla, S., and F.R. Pick. 2017. Freshwater bloom-forming cyanobacteria and anthropogenic change. Limnol. Oceanogr e-Lectures. **7.2**: 1-62. doi:10.1002/loe2.10006

Bouaïcha, N., C. O. Miles, D. G. Beach, Z. Labidi, A. Djabri, N. Y. Benayache, and T. Nguyen-quang. 2019. Structural Diversity, Characterization and Toxicology of Microcystins. Toxins (Basel). **11**: 714. doi:10.3390/toxins11120714

Briland, R. D., J. P. Stone, M. Manubolu, J. Lee, and S. A. Ludsin. 2020. Cyanobacterial blooms modify food web structure and interactions in western Lake Erie. Harmful Algae **92**. doi:10.1016/j.hal.2019.03.004

Buley, R. P., H. E. Correia, A. Abebe, T. B. Issa, and A. E. Wilson. 2021. Predicting microcystin occurrence in freshwater lakes and reservoirs: assessing environmental variables. Inl. Waters **11**: 430–444. doi:10.1080/20442041.2021.1938491

Buratti, F. M., M. Manganelli, S. Vichi, M. Stefanelli, S. Scardala, E. Testai, and E. Funari. 2017. Cyanotoxins: producing organisms, occurrence, toxicity, mechanism of action and human health toxicological risk evaluation. Arch. Toxicol. **91**: 1049–1130. doi:10.1007/s00204-016-1913-6

Carey, C. C., B. W. Ibelings, E. P. Hoffmann, D. P. Hamilton, and J. D. Brookes. 2012. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. Water Res. **46**: 1394–1407. doi:10.1016/j.watres.2011.12.016

Carmichael, W. W., and G. L. Boyer. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. Harmful Algae **54**: 194–212. doi:10.1016/j.hal.2016.02.002

Carpenter, S. R., N. F. Caraco, D. L. Correll, R. W. Howarth, A. N. Sharpley, and V. H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. Ecol. Appl. **8**: 559–568. doi:10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Carvalho, L., C. Mcdonald, C. de Hoyos, and others. 2013. Sustaining recreational quality of European lakes: Minimizing the health risks from algal blooms through phosphorus control. J. Appl. Ecol. **50**: 315–323. doi:10.1111/1365-2664.12059

Carvalho, L., C. A. Miller, E. M. Scott, G. A. Codd, P. S. Davies, and A. N. Tyler. 2011. Cyanobacterial blooms: Statistical models describing risk factors for national-scale lake assessment and lake management. Sci. Total Environ. **409**: 5353–5358. doi:10.1016/j.scitotenv.2011.09.030

Casero, M. C., D. Velázquez, M. Medina-Cobo, A. Quesada, and S. Cirés. 2019. Unmasking the identity of toxigenic cyanobacteria driving a multi-toxin bloom by high-throughput sequencing of cyanotoxins genes and 16S rRNA metabarcoding. Sci. Total Environ. **665**:

367–378. doi:10.1016/j.scitotenv.2019.02.083

Chernoff, N., D. Hill, J. Lang, J. Schmid, T. Le, A. Farthing, and H. Huang. 2020. The comparative toxicity of 10 microcystin congeners administered orally to mice: Clinical effects and organ toxicity. Toxins (Basel). **12**: 1–18. doi:10.3390/toxins12060403

Chislock, M. F., O. Sarnelle, L. M. Jernigan, and A. E. Wilson. 2013. Do high concentrations of microcystin prevent Daphnia control of phytoplankton? Water Res. **47**: 1961–1970. doi:10.1016/j.watres.2012.12.038

Chorus, I., J. Fastner, and M. Welker. 2021. Cyanobacteria and cyanotoxins in a changing environment: Concepts, controversies, challenges. Water **13**: 1–41. doi:10.3390/w13182463

Chorus, I., and M. Welker. 2021. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management, 2nd ed. CRC Press.

Cowart, D. A., M. Pinheiro, O. Mouchel, M. Maguer, J. Grall, J. Miné, and S. Arnaud-Haond. 2015. Metabarcoding is powerful yet still blind: A comparative analysis of morphological and molecular surveys of seagrass communities. PLoS One **10**: 1–26. doi:10.1371/journal.pone.0117562

Cremona, F., L. Tuvikene, J. Haberman, P. Nõges, and T. Nõges. 2018. Factors controlling the three-decade long rise in cyanobacteria biomass in a eutrophic shallow lake. Sci. Total Environ. **621**: 352–359. doi:10.1016/j.scitotenv.2017.11.250

Dai, R., P. Wang, P. Jia, Y. Zhang, X. Chu, and Y. Wang. 2016. A review on factors affecting microcystins production by algae in aquatic environments. World J. Microbiol. Biotechnol. **32**: 1–7. doi:10.1007/s11274-015-2003-2

Davis, T. W., D. L. Berry, G. L. Boyer, and C. J. Gobler. 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during

cyanobacteria blooms. Harmful Algae **8**: 715–725. doi:10.1016/j.hal.2009.02.004

DeMott, W. R., Q. -X Zhang, and W. W. Carmichael. 1991. Effects of toxic cyanobacteria and purified toxins on the survival and feeding of a copepod and three species of Daphnia. Limnol. Oceanogr. **36**: 1346–1357. doi:10.4319/lo.1991.36.7.1346

Díez-Quijada, L., A. I. Prieto, R. Guzmán-Guillén, A. Jos, and A. M. Cameán. 2019. Occurrence and toxicity of microcystin congeners other than MC-LR and MC-RR: A review. Food Chem. Toxicol. **125**: 106–132. doi:10.1016/j.fct.2018.12.042

Dodds, W. K., W. W. Bouska, J. L. Eitzmann, T. J. Pilger, K. L. Pitts, A. J. Riley, J. T. Schloesser, and D. J. Thornbrugh. 2009. Eutrophication of U. S. freshwaters: Analysis of potential economic damages. Environ. Sci. Technol. **43**: 12–19. doi:10.1021/es801217q

Dolman, A. M., J. Rücker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. PLoS One **7**. doi:10.1371/journal.pone.0038757

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/cjfas-58-10-1905

Dziallas, C., and H. P. Grossart. 2011. Increasing oxygen radicals and water temperature select for toxic microcystis sp. PLoS One **6**. doi:10.1371/journal.pone.0025569

Eiler, A., S. Drakare, S. Bertilsson, and others. 2013. Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. PLoS One **8**: 1–10. doi:10.1371/journal.pone.0053516

Elbrecht, V., E. E. Vamos, K. Meissner, J. Aroviita, and F. Leese. 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods Ecol. Evol. **8**: 1265–1275. doi:10.1111/2041-210X.12789

Erratt, K. J., I. F. Creed, and C. G. Trick. 2022. Harmonizing science and management options to reduce risks of cyanobacteria. Harmful Algae **116**. doi:10.1016/j.hal.2022.102264

Esenkulova, S., B. J. G. Sutherland, A. Tabata, N. Haigh, C. M. Pearce, and K. M. Miller. 2020. Operational taxonomic unit comparing metabarcoding and morphological approaches to identify phytoplankton taxa associated with harmful algal blooms. Facets **5**: 784–811. doi:10.1139/FACETS-2020-0025

Ewing, H. A., K. C. Weathers, K. L. Cottingham, and others. 2020. "New" cyanobacterial blooms are not new: two centuries of lake production are related to ice cover and land use. Ecosphere **11**. doi:10.1002/ecs2.3170

Flores, N. M., T. R. Miller, and J. D. Stockwell. 2018. A global analysis of the relationship between concentrations of microcystins in water and fish. Front. Mar. Sci. **5**: 1–14. doi:10.3389/fmars.2018.00030

Francis, G. 1878. Poisonous Australian lake. Nature. **18**: 11-12. doi:10.1038/018011d0

Fulton, R. S., and H. W. Paerl. 1987. Toxic and inhibitory effects of the blue-green alga Microcystis aeruginosa on herbivorous zooplankton. J. Plankton Res. **9**: 837–855. doi:doi.org/10.1093/plankt/9.5.837

Gao, W., Z. Chen, Y. Li, Y. Pan, J. Zhu, S. Guo, L. Hu, and J. Huang. 2018. Bioassessment of a drinking water reservoir using plankton: High throughput sequencing vs. traditional morphological method. Water (Switzerland) **10**. doi:10.3390/w10010082

Garcia-Pichel, F., J. Belnap, S. Neuer, and F. Schanz. 2003. Estimates of global cyanobacterial biomass and its distribution. Arch. Hydrobiol. Suppl. Algol. Stud. **109**: 213–227. doi:10.1127/1864-1318/2003/0109-0213

Garcia-Pichel, F., J. Lombard, T. Soule, S. Dunaj, S. H. Wu, and M. F. Wojciechowski. 2019.

Timing the evolutionary advent of cyanobacteria and the later great oxidation event using gene phylogenies of a sunscreen. MBio **10**. doi:10.1128/mBio.00561-19

Ger, K. A., L. A. Hansson, and M. Lürling. 2014. Understanding cyanobacteria-zooplankton interactions in a more eutrophic world. Freshw. Biol. **59**: 1783–1798. doi:10.1111/fwb.12393

Ger, K. A., R. Panosso, and M. Lurling. 2011. Consequences of acclimation to Microcystis on the selective feeding behavior of the calanoid copepod Eudiaptomus gracilis. Limnol. Oceanogr. **56**: 2103–2114. doi:10.4319/lo.2011.56.6.2103

Ger, K. A., P. Urrutia-Cordero, P. C. Frost, L. A. Hansson, O. Sarnelle, A. E. Wilson, and M. Lürling. 2016. The interaction between cyanobacteria and zooplankton in a more eutrophic world. Harmful Algae **54**: 128–144. doi:10.1016/j.hal.2015.12.005

Ghadouani, A., and B. Pinel-Alloul. 2002. Phenotypic plasticity in Daphnia pulicaria as an adaptation to high biomass of colonial and filamentous cyanobacteria: Experimental evidence. J. Plankton Res. **24**: 1047–1056. doi:10.1093/plankt/24.10.1047

Ghadouani, A., B. Pinel-Alloul, K. Plath, G. A. Codd, and W. Lampert. 2004. Effects of Microcystis aeruginosa and purified microcystin-LR on the feeding behavior of Daphnia pulicaria. Limnol. Oceanogr. **49**: 666–679. doi:10.4319/lo.2004.49.3.0666

Ghadouani, A., B. Pinel-Alloul, and E. E. Prepas. 2006. Could increased cyanobacterial biomass following forest harvesting cause a reduction in zooplankton body size structure? Can. J. Fish. Aquat. Sci. **63**: 2308–2317. doi:10.1139/F06-117

Giani, A., D. F. Bird, Y. T. Prairie, and J. F. Lawrence. 2005. Empirical study of cyanobacterial toxicity along a trophic gradient of lakes. Can. J. Fish. Aquat. Sci. **62**: 2100–2109. doi:10.1139/f05-124

Gobler, C. J., J. A. M. Burkholder, T. W. Davis, M. J. Harke, T. Johengen, C. A. Stow, and D. B. Van de Waal. 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae **54**: 87–97. doi:10.1016/j.hal.2016.01.010

Guiry, M. D. 2012. How many species of algae are there? J. Phycol. **48**: 1057–1063. doi:10.1111/j.1529-8817.2012.01222.x

Guiry, M.D., and G.M. Guiry. 2020. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway

Hairston, N. G., C. L. Holtmeier, W. Lampert, and others. 2001. Natural selection for grazer resistance to toxic cyanobacteria: Evolution of phenotypic plasticity? Evolution (N. Y). **55**: 2203–2214. doi:10.1111/j.0014-3820.2001.tb00736.x

Hallegraeff, G. M., D. M. Anderson, C. Belin, and others. 2021. Perceived global increase in algal blooms is attributable to intensified monitoring and emerging bloom impacts. Commun. Earth Environ. **2**. doi:10.1038/s43247-021-00178-8

Hänfling, B., L. L. Handley, D. S. Read, and others. 2016. Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. Mol. Ecol. **25**: 3101–3119. doi:10.1111/mec.13660

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. **270**: 313–321. doi:10.1098/rspb.2002.2218

Hering, D., A. Borja, J. I. Jones, and others. 2018. Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Res. **138**: 192–205. doi:10.1016/j.watres.2018.03.003

Ho, J. C., A. M. Michalak, and N. Pahlevan. 2019. Widespread global increase in intense lake

phytoplankton blooms since the 1980s. Nature **574**: 667–670. doi:10.1038/s41586-019-1648-7

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. **695**: 133668. doi:10.1016/j.scitotenv.2019.133668

IPCC, 2021. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press

Jüttner, F., and S. B. Watson. 2007. Biochemical and ecological control of geosmin and 2-methylisoborneol in source waters. Appl. Environ. Microbiol. **73**: 4395–4406. doi:10.1128/AEM.02250-06

Kaebernick, M., and B. A. Neilan. 2001. Ecological and molecular investigations of cyanotoxin production. FEMS Microbiol. Ecol. **35**: 1–9. doi:10.1016/S0168-6496(00)00093-3

Keck, F., R. C. Blackman, R. Bossart, and others. 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for biological community assessment. Mol. Ecol. **31**: 1820–1835. doi:10.1111/mec.16364

Komarek, J. 2006. Cyanobacterial Taxonomy: Current Problems and Prospects for the Integration of Traditional and Molecular Approaches. Algae **21**: 349–375. doi:10.4490/algae.2006.21.4.349

Komárek, J. 2016. Review of the cyanobacterial genera implying planktic species after recent

taxonomic revisions according to polyphasic methods: State as of 2014. Hydrobiologia **764**: 259–270. doi:10.1007/s10750-015-2242-0

Kosten, S., V. L. M. Huszar, E. Bécares, and others. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Glob. Chang. Biol. **18**: 118–126. doi:10.1111/j.1365-2486.2011.02488.x

Kotak, B. G., and R. W. Zurawell. 2007. Cyanobacterial toxins in Canadian freshwaters: A review. Lake Reserv. Manag. **23**: 109–122. doi:10.1080/07438140709353915

Kozlowsky-Suzuki, B., A. E. Wilson, and A. da Silva Ferrão-Filho. 2012. Biomagnification or biodilution of microcystins in aquatic foodwebs? Meta-analyses of laboratory and field studies. Harmful Algae **18**: 47–55. doi:10.1016/j.hal.2012.04.002

Kudela, R. M., and others. 2015. Harmful algal blooms. A scientific summary for policy makers. IOC/UNESCO

Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. U. S. A. **82**: 6955–6959. doi:10.1073/pnas.82.20.6955

Lee, E., U. M. Ryan, P. Monis, G. B. McGregor, A. Bath, C. Gordon, and A. Paparini. 2014. Polyphasic identification of cyanobacterial isolates from Australia. Water Res. **59**: 248–261. doi:10.1016/j.watres.2014.04.023

Leese, F., M. Sander, D. Buchner, V. Elbrecht, P. Haase, and V. M. A. Zizka. 2021. Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environ. DNA **3**: 261–276. doi:10.1002/edn3.177

Lévesque, B., M. C. Gervais, P. Chevalier, and others. 2014. Prospective study of acute health effects in relation to exposure to cyanobacteria. Sci. Total Environ. **466**–**467**: 397–403.

doi:10.1016/j.scitotenv.2013.07.045

Li, X. C., S. Huo, J. Zhang, C. Ma, Z. Xiao, H. Zhang, B. Xi, and X. Xia. 2019. Metabarcoding reveals a more complex cyanobacterial community than morphological identification. Ecol. Indic. **107**: 105653. doi:10.1016/j.ecolind.2019.105653

Loftin, K. A., J. L. Graham, E. D. Hilborn, S. C. Lehmann, M. T. Meyer, J. E. Dietze, and C. B. Griffith. 2016. Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. Harmful Algae **56**: 77–90. doi:10.1016/j.hal.2016.04.001

MacDonald, G. K., E. M. Bennett, P. A. Potter, and N. Ramankutty. 2011. Agronomic phosphorus imbalances across the world's croplands. Proc. Natl. Acad. Sci. U. S. A. **108**: 3086–3091. doi:10.1073/pnas.1010808108

Mantzouki, E., M. Lürling, J. Fastner, and others. 2018. Temperature effects explain continental scale distribution of cyanobacterial toxins. Toxins (Basel). **10**: 1–24. doi:10.3390/toxins10040156

Mantzouki, E., P. M. Visser, M. Bormans, and B. W. Ibelings. 2016. Understanding the key ecological traits of cyanobacteria as a basis for their management and control in changing lakes. Aquat. Ecol. **50**: 333–350. doi:10.1007/s10452-015-9526-3

Merel, S., D. Walker, R. Chicana, S. Snyder, E. Baurès, and O. Thomas. 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. Environ. Int. **59**: 303–327. doi:10.1016/j.envint.2013.06.013

Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. **7**: 1–11. doi:10.1038/ncomms13603

Miller, M. A., R. M. Kudela, A. Mekebri, and others. 2010. Evidence for a novel marine harmful algal bloom: Cyanotoxin (microcystin) transfer from land to sea otters. PLoS One **5**: 1–11. doi:10.1371/journal.pone.0012576

Minns, C. K., J. E. Moore, B. J. Shuter, and N. E. Mandrak. 2008. A preliminary national analysis of some key characteristics of Canadian lakes. Can. J. Fish. Aquat. Sci. **65**: 1763–1778. doi:10.1139/F08-110

Molot, L. A., S. B. Watson, I. F. Creed, and others. 2014. A novel model for cyanobacteria bloom formation: The critical role of anoxia and ferrous iron. Freshw. Biol. **59**: 1323–1340. doi:10.1111/fwb.12334

Monchamp, M. E., F. R. Pick, B. E. Beisner, and R. Maranger. 2014. Nitrogen forms influence microcystin concentration and composition via changes in cyanobacterial community structure. PLoS One **9**. doi:10.1371/journal.pone.0085573

Monchamp, M. E., J. C. Walser, F. Pomati, and P. Spaak. 2016. Sedimentary DNA reveals cyanobacterial community diversity over 200 years in two perialpine lakes. Appl. Environ. Microbiol. **82**: 6472–6482. doi:10.1128/AEM.02174-16

Mowe, M. A. D., C. Porojan, F. Abbas, S. M. Mitrovic, R. P. Lim, A. Furey, and D. C. J. Yeo. 2015. Rising temperatures may increase growth rates and microcystin production in tropical Microcystis species. Harmful Algae **50**: 88–98. doi:10.1016/j.hal.2015.10.011

Muro-Pastor, A. M., and W. R. Hess. 2012. Heterocyst differentiation: From single mutants to global approaches. Trends Microbiol. **20**: 548–557. doi:10.1016/j.tim.2012.07.005

Neilan, B. A., L. A. Pearson, J. Muenchhoff, M. C. Moffitt, and E. Dittmann. 2013. Environmental conditions that influence toxin biosynthesis in cyanobacteria. Environ. Microbiol. **15**: 1239–1253. doi:10.1111/j.1462-2920.2012.02729.x

O'Neil, J. M., T. W. Davis, M. A. Burford, and C. J. Gobler. 2012. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. Harmful Algae **14**: 313–334. doi:10.1016/j.hal.2011.10.027

O'Reilly, C. M., R. J. Rowley, P. Schneider, J. D. Lenters, P. B. Mcintyre, and B. M. Kraemer. 2015. Rapid and highly variable warming of lake surface waters around the globe. Geophysical Research Letters, 42: 1-9. Geophys. Res. Lett. 1–9. doi:10.1002/2015GL066235.Received

Omidi, A., M. Esterhuizen-Londt, and S. Pflugmacher. 2018. Still challenging: the ecological function of the cyanobacterial toxin microcystin–What we know so far. Toxin Rev. **37**: 87–105. doi:10.1080/15569543.2017.1326059

Orihel, D. M., H. M. Baulch, N. J. Casson, R. L. North, C. T. Parsons, D. C. M. Seckar, and J. J. Venkiteswaran. 2017. Internal phosphorus loading in canadian fresh waters: A critical review and data analysis. Can. J. Fish. Aquat. Sci. **74**: 2005–2029. doi:10.1139/cjfas-2016-0500

Paerl, H. W., and M. A. Barnard. 2020. Mitigating the global expansion of harmful cyanobacterial blooms: Moving targets in a human- and climatically-altered world. Harmful Algae **96**: 101845. doi:10.1016/j.hal.2020.101845

Paerl, H. W., and T. G. Otten. 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microb. Ecol. **65**: 995–1010. doi:10.1007/s00248-012-0159-y

Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. **46**: 1349–1363. doi:10.1016/j.watres.2011.08.002

Pawlowski, J., M. Kelly-Quinn, F. Altermatt, and others. 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic

ecosystems. Sci. Total Environ. **637**–**638**: 1295–1310. doi:10.1016/j.scitotenv.2018.05.002

Pick, F. R. 2016. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria. Can. J. Fish. Aquat. Sci. **73**: 1149–1158. doi:10.1139/cjfas-2015-0470

Pick, F. R., and D. R. S. Lean. 1987. The role of macronutrients (C, N, P) in controlling cyanobacterial dominance in temperate lakes. New Zeal. J. Mar. Freshw. Res. **21**: 425–434. doi:10.1080/00288330.1987.9516238

Pilon, S., A. Zastepa, Z. E. Taranu, I. Gregory-Eaves, M. Racine, J. M. Blais, A. J. Poulain, and F. R. Pick. 2019. Contrasting histories of microcystin-producing cyanobacteria in two temperate lakes as inferred from quantitative sediment DNA analyses. Lake Reserv. Manag. **35**: 102–117. doi:10.1080/10402381.2018.1549625

Plaas, H. E., and H. W. Paerl. 2021. Toxic Cyanobacteria: A Growing Threat to Water and Air Quality. Environ. Sci. Technol. **55**: 44–64. doi:10.1021/acs.est.0c06653

Pouria, S., A. De Andrade, J. Barbosa, and others. 1998. Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. Lancet **352**: 21–26. doi:10.1016/S0140-6736(97)12285-1

Puddick, J., M. R. Prinsep, S. A. Wood, S. A. F. Kaufononga, S. C. Cary, and D. P. Hamilton. 2014. High levels of structural diversity observed in microcystins from microcystis CAWBG11 and characterization of six new microcystin congeners. Mar. Drugs **12**: 5372–5395. doi:10.3390/md12115372

Reid, A. J., A. K. Carlson, I. F. Creed, and others. 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. **94**: 849–873. doi:10.1111/brv.12480

Reynolds, C.S. 2006. The Ecology of Phytoplankton. Cambridge: Cambridge University Press (Ecology, Biodiversity and Conservation). doi:10.1017/CBO9780511542145

Richardson, J., C. Miller, S. C. Maberly, and others. 2018. Effects of multiple stressors on cyanobacteria abundance vary with lake type. Glob. Chang. Biol. **24**: 5044–5055. doi:10.1111/gcb.14396

Rigosi, A., C. C. Carey, B. W. Ibelings, and J. D. Brookes. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. Limnol. Oceanogr. **59**: 99–114. doi:10.4319/lo.2014.59.1.0099

Rippka, R., J. Deruelles, and J. B. Waterbury. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. J. Gen. Microbiol. **111**: 1–61. doi:10.1099/00221287-111-1-1

Robarts, R. D., and T. Zohary. 1987. Temperature effects on photosynthetic capacity, respiration, and growth rates of bloom-forming cyanobacteria. New Zeal. J. Mar. Freshw. Res. **21**: 391–399. doi:10.1080/00288330.1987.9516235

Sarnelle, O., and A. E. Wilson. 2005. Local adaptation of Daphnia pulicaria to toxic cyanobacteria. Limnol. Oceanogr. **50**: 1565–1570. doi:10.4319/lo.2005.50.5.1565

Schindler, D. W. 1977. Evolution of phosphorus limitation in lakes. Science (80-.). **195**: 260–262. doi:10.1126/science.195.4275.260

Schopf, J. W. 2012. The fossil record of cyanobacteria. Ecology of cyanobacteria II: their diversity in space and time. 15-36.

Shan, K., L. Song, W. Chen, and others. 2019. Analysis of environmental drivers influencing interspecific variations and associations among bloom-forming cyanobacteria in large, shallow eutrophic lakes. Harmful Algae **84**: 84–94. doi:10.1016/j.hal.2019.02.002

Smith, R. B., B. Bass, D. Sawyer, D. Depew, and S. B. Watson. 2019. Estimating the economic costs of algal blooms in the Canadian Lake Erie Basin. Harmful Algae **87**: 101624. doi:10.1016/j.hal.2019.101624

Smith, V. H. 1983. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. Science (80-.). **221**: 669–671. doi:10.1126/science.221.4611.669

Smith, V. H., and D. W. Schindler. 2009. Eutrophication science: where do we go from here? Trends Ecol. Evol. **24**: 201–207. doi:10.1016/j.tree.2008.11.009

Taranu, Z. E., I. Gregory-Eaves, P. R. Leavitt, and others. 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. Ecol. Lett. **18**: 375–384. doi:10.1111/ele.12420

Taranu, Z. E., I. Gregory-Eaves, R. J. Steele, M. Beaulieu, and P. Legendre. 2017. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework for modelling an important health risk factor. Glob. Ecol. Biogeogr. **26**: 625–637. doi:10.1111/geb.12569

Taranu, Z. E., F. R. Pick, I. F. Creed, A. Zastepa, and S. B. Watson. 2019. Meteorological and Nutrient Conditions Influence Microcystin Congeners in Freshwaters. Toxins (Basel). **11**: 1–21. doi:10.3390/toxins11110620

Taranu, Z. E., R. W. Zurawell, F. Pick, and I. Gregory-Eaves. 2012. Predicting cyanobacterial dynamics in the face of global change: The importance of scale and environmental context. Glob. Chang. Biol. **18**: 3477–3490. doi:10.1111/gcb.12015

Tse, T. J., L. E. Doig, S. Tang, and others. 2018. Combining High-Throughput Sequencing of sedaDNA and Traditional Paleolimnological Techniques to Infer Historical Trends in Cyanobacterial Communities. Environ. Sci. Technol. **52**: 6842–6853.

doi:10.1021/acs.est.7b06386

Visser, P. M., J. M. H. Verspagen, G. Sandrini, L. J. Stal, H. C. P. Matthijs, T. W. Davis, H. W. Paerl, and J. Huisman. 2016. How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. Harmful Algae **54**: 145–159. doi:10.1016/j.hal.2015.12.006

Vuorio, K., M. Järvinen, and N. Kotamäki. 2020. Phosphorus thresholds for bloom-forming cyanobacterial taxa in boreal lakes. Hydrobiologia **847**: 4389–4400. doi:10.1007/s10750-019-04161-5

Van de Waal, D. B., V. H. Smith, S. A. J. Declerck, E. C. M. Stam, and J. J. Elser. 2014. Stoichiometric regulation of phytoplankton toxins. Ecol. Lett. **17**: 736–742. doi:10.1111/ele.12280

Wagner, C., and R. Adrian. 2009. Cyanobacteria dominance: Quantifying the effects of climate change. Limnol. Oceanogr. **54**: 2460–2468. doi:10.4319/lo.2009.54.6_part_2.2460

Walls, J. T., K. H. Wyatt, J. C. Doll, E. M. Rubenstein, and A. R. Rober. 2018. Hot and toxic: Temperature regulates microcystin release from cyanobacteria. Sci. Total Environ. **610–611**: 786–795. doi:10.1016/j.scitotenv.2017.08.149

Walsby, A. E. 1994. Gas vesicles. Microbiol Rev. Mar. **58(1)**: 94-144. doi: 10.1128/mr.58.1.94-144.1994

Watson, S. B., E. McCauley, and J. A. Downing. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol. Oceanogr. **42**: 487–495. doi:10.4319/lo.1997.42.3.0487

Whitton, B. A. 2012. Ecology of cyanobacteria: Their diversity in space and time. Springer. doi:10.1007/0-306-46855-7

Wood, S. A., D. R. Dietrich, S. Craig Cary, and D. P. Hamilton. 2012. Increasing microcystis cell density enhances microcystin synthesis: A mesocosm studys. Inl. Waters **2**: 17–22. doi:10.5268/IW-2.1.424

Yuan, L. L., A. I. Pollard, S. Pather, J. L. Oliver, and L. D'Anglada. 2014. Managing microcystin: Identifying national-scale thresholds for total nitrogen and chlorophyll a. Freshw. Biol. **59**: 1970–1981. doi:10.1111/fwb.12400

Zastepa, A., F. R. Pick, J. M. Blais, and A. Saleem. 2015. Analysis of intracellular and extracellular microcystin variants in sediments and pore waters by accelerated solvent extraction and high performance liquid chromatography-tandem mass spectrometry. Anal. Chim. Acta **872**: 26–34. doi:10.1016/j.aca.2015.02.056

Zhang, F., J. Lee, S. Liang, and C. Shum. 2015. Cyanobacteria blooms and non-alcoholic liver disease: Evidence from a county level ecological study in the United States. Environ. Heal. A Glob. Access Sci. Source **14**: 1–11. doi:10.1186/s12940-015-0026-7

CHAPTER I

Both biotic and abiotic predictors explain significant variation in cyanobacteria biomass across lakes from temperate to subarctic zones

Paul W. MacKeigan[1,2], Zofia E. Taranu[3], Frances R. Pick[4], Beatrix E. Beisner[2,5] and Irene Gregory-Eaves[1,2]

[1]Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada;

[2] Interuniversity Research Group in Limnology (GRIL), Quebec, Canada;

[3]Aquatic Contaminants Research Division, Environment and Climate Change Canada, Montreal, Quebec, H2Y 2E7, Canada;

[4]Department of Biology, University of Ottawa, Ottawa, Ontario, K1N 9A7, Canada;

[5]Department of Biological Sciences, University of Quebec at Montreal, Montreal, Quebec, H2X 1Y4, Canada

Manuscript published in *Limnology and Oceanography* in April 2023

Abstract

The development of cyanobacteria blooms is of increasing concern in many lakes worldwide, and as a result, modeling their predictors is vital for understanding where and why they occur. In this study, we developed and analyzed a 640-lake data set that spans Canada and twelve ecozones to identify the drivers of cyanobacteria biomass and of several key toxin- and bloom-forming genera (*Microcystis*, *Aphanizomenon* and *Dolichospermum*). The database consisted of an exhaustive list of potential predictors (n=55), including water chemistry, land-use and zooplankton variables. We applied a series of empirical modeling approaches to identify significant predictors and thresholds (generalized linear and additive models, mixed effect regression trees), all while accounting for ecozone variability. Across all modeling approaches, and ecozones total phosphorus (TP) was identified as the most important predictor of total cyanobacterial and focal genera biomass. In addition, cyanobacteria across Canada showed significant associations with increasing dissolved organic and inorganic carbon, and several ions. Despite the widely held notion that cyanobacteria are often toxic and/or a poor food source for zooplankton, we found a positive relationship between cyanobacteria and zooplankton, particularly with daphnid and copepod biomass. Localized top-down forces and evolutionary adaptations resulting from long-term exposure in eutrophic lakes are among the possible explanations for this observed positive association. By considering a suite of complementary modeling approaches, we found that non-linear models provided greater predictive power and the random ecozone effect was minor due to the overarching importance of local abiotic and biotic factors.

Introduction

The ecological functioning of many lakes worldwide is being challenged by numerous human-induced stressors such as eutrophication and climate change (Reid et al. 2018). These stressors often result in conditions favorable for excessive phytoplankton growth, with many lakes becoming dominated by cyanobacteria and other harmful algae (O'Neil et al. 2012; Huisman et al. 2018). Blooms of cyanobacteria are visible colorations of the water that are of particular concern to lake managers and relevant stakeholders as they present a risk to human and animal health through the production of toxins along with taste and odor compounds (Merel et al. 2013; Pick 2016). Additionally, blooms indirectly lower oxygen concentrations in bottom waters that can kill fish and benthic invertebrates, and lead to internal nutrient loading (Paerl and Otten 2013; Molot et al. 2014; Huisman et al. 2018).

Despite efforts to mitigate blooms, several studies have reported their global exacerbation. Ho et al. (2019) analyzed three decades of high-resolution satellite imagery data from large lakes (surface area > 100 km^2) around the world and found that peak summertime phytoplankton bloom intensity increased in over two-thirds of their lakes. Moreover, an analysis of cyanobacterial pigments from over 100 north-temperate lake sediment cores found that cyanobacteria abundance has increased substantially relative to other phytoplankton groups (e.g. diatoms), particularly since ~1945 (Taranu et al. 2015). Due to their potentially negative impacts on aquatic ecosystems, and their projected increase, cyanobacteria remain a primary pressing and emerging threat to freshwater biodiversity (Reid et al. 2018).

Several lines of evidence have identified numerous physical, chemical and biological predictors that act to promote high cyanobacteria biomass or dominance, most relating directly or indirectly to eutrophication and climate warming (O'Neil et al. 2012; Paerl and Paul 2012;

Huisman et al. 2018) (Table S1). Large-scale statistical models based on at least 50 lakes have been useful in generalizing the key drivers of cyanobacteria across broad regions and lake types (Table S1). Generally, macronutrients (i.e., total phosphorus (TP), total nitrogen (TN)) and temperature have emerged as the best predictors of cyanobacterial biomass (Downing et al. 2001; Kosten et al. 2012; Beaulieu et al. 2013), with these factors potentially acting synergistically to further promote biomass (Taranu et al. 2012) (Table S1).

Although valuable, previous large-scale empirical models are commonly restricted to few predictor variables (Table S1). Many likewise omit integral components of the lake-landscape ecosystem such as watershed land-use, and biological community metrics. The latter represents an important shortcoming given that different zooplankton groups have been shown to be negatively or positively correlated with cyanobacteria biomass (e.g., Cremona et al. 2018). Compared to eukaryotic algae, cyanobacteria are generally considered a poor food source for key zooplankton grazers (e.g., *Daphnia*) because of their potential toxicity, poor nutritional value (low amounts of polyunsaturated fatty acids and sterols), and aggregatory morphology typical of numerous taxa (Von Elert et al. 2003; Ger et al. 2016a). Smaller-bodied, specialist zooplankton grazers have been reported to compete more successfully than larger-bodied generalists such as *Daphnia* under bloom conditions (Ghadouani et al. 2006; Ger et al. 2016a; Fig. S12 A-B). Smaller zooplankton may enable positive feedback for cyanobacteria biomass by selectively feeding on other phytoplankton, thereby minimizing the presence of competitors for cyanobacteria (Leitão et al. 2018). Despite the interest in large-scale predictive cyanobacteria models, there has been no systematic evaluation of the relative importance of abiotic and biotic factors explaining variation in cyanobacterial biomass across a broad suite of lake types. Furthermore, there is an insufficient understanding of which conditions best explain community

composition and regional variability in response models, as well as the predictors of some potentially harmful genera.

In this study, we develop large-scale cyanobacteria predictive models by quantifying the importance of abiotic (physiographic, water quality, land use, and climate) and biotic (zooplankton) variables across 640 Canadian lakes using standardized sampling and analytical protocols (Huot et al. 2019; NSERC Canadian Lake Pulse Network 2021). Our study presents an opportunity to study a wide range of potential cyanobacteria predictors across a landscape covering over 400,000 km^2. We hypothesized that increasing nutrients (TP and TN) would be overarching positive drivers of cyanobacteria biomass, but that other factors would act both at local and regional scales to modulate these relationships, and thus, account for much of the remaining unexplained variation. Primarily, we expected that variability in water temperature and residence time would account for substantial residual variation given that cyanobacteria have higher temperature growth optima relative to other phytoplankton, can regulate their buoyancy during thermal stratification of the water column and cannot establish blooms in fast flushing waters (Huisman et al. 2018). We also hypothesized that lakes from the same ecozone would show greater similarity in their cyanobacteria responses than those from other regions, as ecozones are defined by their geology, climate and vegetation. Thus, we expected that ecozone identity would be a significant factor in cyanobacteria biomass prediction.

Numerous studies have also shown that due to physiological and life history variation, the importance of biomass drivers would differ between cyanobacteria taxa and functional groups (Rigosi et al. 2014; Shan et al. 2019; Amorim et al. 2020). We constructed empirical models for three cyanobacteria taxa that are common bloom-forming genera in temperate lakes: *Microcystis*, *Aphanizomenon* and *Dolichospermum* (the latter two also capable of fixing atmospheric

nitrogen). Given that *Microcystis* has a higher maximum growth rate at warmer temperatures (Paerl and Huisman 2009; Paerl et al. 2011), and inspired by Shan et al (2019), we hypothesized that *Microcystis* biomass would be related mainly to increasing water temperature, followed by increasing TP as well as cladoceran and copepod biomass. In contrast, we hypothesized that the biomass of *Dolichospermum* and *Aphanizomenon* would be most strongly related to increasing nutrients (Rigosi et al. 2014). Lastly, we were also interested in exploring which variables explained differences in cyanobacteria community composition.

Materials and methods

Lake selection

The LakePulse Network represents the first pan-Canadian survey aimed at investigating the health and functioning of Canadian lakes within the geographic range of 52° to 118° West and 41° to 60° North. A total of 664 lakes were sampled once across three summers (2017, 2018, 2019) following a standardized protocol for over a hundred variables to create a nationwide dataset (NSERC Canadian Lake Pulse Network, 2021). Lakes were selected following a stratified random sampling design using ecozones, lake sizes (0.1-1km^2, 1-10km^2, 10-100km^2) and watershed human impact index categories (low, medium, high) as stratification groups (Huot et al. 2019. Lakes had a minimum depth of 1 m and were located within 1 km of a road for accessibility purposes. Lakes were sampled across twelve Canadian ecozones (regions with unique geological, climatic and ecological features; Ecological Stratification Working Group, 1996).

Sample collection

In total, 55 variables grouped as physiographic, water quality, land use, zooplankton, climate data were collected and used as potential predictors for cyanobacteria biomass (Table S2). Sampling occurred during the period of maximum thermal stratification (between July and early September) to limit seasonal variability (Huot et al. 2019). For each lake, the sampling station was located at the deepest point ("Index site") using a depth sounder (see details in NSERC Canadian Lake Pulse Network, 2021). Briefly, field teams used an acid washed, integrated tube sampler to collect surface water for water quality and phytoplankton samples. At the Index site, water was collected across a depth equal to twice the Secchi disk depth (euphotic zone) or 2 meters (tube length), whichever was shallower.

To characterize the crustacean zooplankton community, an integrated water-column sample was collected using a 100 μm mesh Wisconsin net at the same index site. In the field, zooplankton samples were narcotized and preserved in ethanol. Samples were identified and biomass estimated by BSA Environmental Services (Ohio, U.S.A.; details in Paquette et al. 2021).

A multiparameter meter (RBRmaestro) was deployed to sample the entire water column for temperature, dissolved oxygen, and specific conductivity. Lake physiographic variables were determined using HydroLAKES v.1.0 (Messager et al. 2016). The percentage of different land use types (agriculture, forestry, mines, natural landscape, pasture urban and water) were characterized for each watershed, as described by Huot et al. (2019). Climate variables were recorded over thirty days leading up to the sampling date, and were accessed from ERA5-Land hourly data (Muñoz Sabater 2019). Climatic variables were then averaged over a 7-day and 30-day period to provide a weekly and monthly average, respectively.

Phytoplankton counting and identification

From the Index site, a 120 mL subsample of surface water (euphotic zone) was fixed and preserved in acid Lugol's iodine solution. Phytoplankton samples were collected and analyzed for 96% of lakes of the original LakePulse nationwide dataset (n = 640). Identifications were conducted on 2-10 mL aliquots following Utermöhl's sedimentation method (Lund et al. 1958) using a Zeiss Axiovert 40 CFL inverted microscope at 250 X and 500 X magnifications and a minimum of 400 units enumerated. Biomass was calculated from recorded abundance and specific volume estimates based on geometric solids(Rott 1981; Hillebrand et al. 1999). Identifications done by a single taxonomist and were based on the following keys: Komárek and Anagnostidis (1998, 2005) and Komárek (2013).

Data analysis

All statistical analyses were conducted in R v. 4.1.0 (R Core Team 2021). To identify predictors of total cyanobacteria biomass, we used several modeling approaches to evaluate the predictive strength of a pool of 55 potential explanatory variables. Response variables were either total cyanobacteria or genus-specific biomass (mg/L). A summary of our statistical workflow is provided in Fig. S1.

To select the most parsimonious set of predictors for downstream analyses, we followed Feld et al. (2016). Briefly, unrecorded values (NAs) for each variable were first replaced by medians of the respective ecozones (see Table S2). Next, variables were transformed with either logarithmic, square root, or logit transformations to ensure that variables were normally distributed with a small constant (X_i+1) being added to zero values including the biomasses of each zooplankton group and the biotic response variables of cyanobacteria, *Microcystis*, *Aphanizomenon* and *Dolichospermum* biomasses (Table S2). Predictors were evaluated for collinearity using Variance Inflation Factors (VIFs) (*vifstep* function in the *usdm* package; Naimi

2015), setting a VIF threshold of 10 (Borcard et al. 2018). After discarding variables due to collinearity, Random Forest (RF) was used to reduce the variable list and select the strongest potential predictors for further analyses. RF regression trees were fit for total cyanobacteria, *Microcystis*, *Aphanizomenon* and *Dolichospermum* biomass using the *rfsrc* function in the randomForestSRC package (Breiman 2001; Ishwaran and Kogalur, 2016) by bootstrapping data subsets and ranking predictors by importance as single and interactive terms in the model (Feld et al. 2016). We retained the top 20 variables ranked by each RF for subsequent analyses.

To first examine linear relationships and quantify inter-ecozone correlations, we constructed a linear mixed-effect model (LMMs) using the nlme package (Pinheiro et al. 2016). We developed a LMM with cyanobacterial biomass as a response variable, by applying a "beyond optimal" model (all 20 fixed effects and ecozone as a random effect), followed by a stepwise selection of of the top 20 potential predictors using BIC values and $p < 0.05$ as the inclusion criterion. To meet model assumptions, lakes without cyanobacteria detected were removed, leaving 602 sites. Since the relationships between cyanobacteria biomass and the set of selected environmental predictors are in some cases non-linear, we then applied generalized additive mixed models (GAMMs) using the *gam* function in the mgcv package (method = REML, select = TRUE; Wood 2021) and followed a similar variable selection approach as done with the LMM (ecozone as a random effect). Given that we also wanted to identify the best predictors of the biomass of the three focal genera from this same lake set, we applied modeling approaches appropriate for these data (i.e., zero-altered hurdle model to address the inflated number of zeros in the responses). To do so, the presence-absence of each genus was first modeled with a binomial generalized linear mixed model (GLMM) (part 1 of hurdle model), then GAMMs were constructed only on lakes that had detected biomass (part 2 of hurdle model)

(Zuur and Ieno 2016). Lastly, we used linear mixed-effect regression trees (LMRTs) for the biomass of all cyanobacteria and the focal genera to identify predictor thresholds and interactions among drivers (*glmertree* function and package; Fokkema et al. 2018). Ecozone was included as a random effect in each model to determine whether relationships (intercepts and/or slopes) varied across the nation.

Finally, we used a multivariate redundancy analysis (RDA) (*rda* function in the vegan package; Oksanen et al. 2019) to examine the ensemble of predictors in relation to the cyanobacteria community composition. This analysis also helped evaluate whether the focal species dominated the community across ecozones. The community biomass matrix was Hellinger transformed and the environmental variables were normalized and scaled. Variation partitioning (*rdacca.hp* package (Lai et al. 2022)) was used to quantify and explain the unique and shared portions of variation explained by cyanobacteria drivers more commonly-tested in broad-scale studies (water quality) versus those typically omitted (land use and zooplankton) (Table S2).

Results

Distribution of cyanobacteria across Canada

Total cyanobacteria biomass (mg/L) was measured across all 640 lakes and varied considerably (Fig. 1; Table S3; median biomass = 0.16 mg/L, max = 903 mg/L). On average, cyanobacteria was 36% of total phytoplankton biomass, reaching their greatest biomasses in the Prairies and Boreal Plains ecozones (Fig. 1; representing ~62% and 58% of the phytoplankton biomass in these central Canadian ecozones). This pattern echoes trends in trophic state, with 82% of lakes in the Prairies and Boreal Plains being eutrophic or hypereutrophic, compared to

just 19% of lakes in the Atlantic Highlands and Atlantic Maritimes (trophic state based on TP thresholds in Wetzel (2001)).

In total, 31 cyanobacteria genera were identified with microscopy. *Aphanocapsa* occurred in 63% of lakes (Fig. 2; Table S3), followed by *Dolichospermum* (47%), *Microcystis* (34%), *Limnothrix* (34%) and *Aphanizomenon* (27%). Communities in central Canada were dominated by large colonial bloom-forming taxa, primarily *Microcystis*, *Aphanizomenon* and *Planktothrix* (Fig. S2 A,B; Fig. S3). On average, bloom-forming taxa represent 60% of biomass across all lakes but was >80% in central Canadian ecozones (Fig. S4). In Eastern Canada, *Aphanocapsa* was the dominant (Fig. 2). *Dolichospermum* exhibited less regional specificity; representing on average 15% to 25% of cyanobacteria biomass across all ecozones (Fig. 2; Fig. S2 C).

Models of total cyanobacteria biomass

After reducing the potential predictor list from 55 to 20 with the RF analysis (Fig. S1; Table S2), we identified the most parsimonious set of predictors for total cyanobacterial biomass as TP, DOC, TN, Ca^{2+}, DIC, cyclopoid copepod biomass, Na^{+}, daphnid biomass, % pastureland and residence time using LMM analysis (Table 1). Although the inclusion of ecozone as a random intercept was significant, this random effect only accounted for 3% of the residual variance (difference between marginal R^2 and conditional R^2; Table 1). Thus, baseline cyanobacteria biomass did not vary substantially among ecozones. For the fixed effect slopes, the linear modeling approach showed that total cyanobacteria biomass was positively related to the above predictors, with the exception of DIC and Na^{+} which were negatively related (Table 1).

To account for potential non-linear trends and improve model fit, we also developed GAMMs by considering the top 20 RF-selected predictors. After removing non-significant terms, the final model explained 55% of the variation and incorporated TP, DOC, DIC, Na^{+},

daphnid biomass, color, residence time as fixed factors, with ecozone as a random effect (Table 1; Fig. 3). In this model, hump-shaped relationships were visible in the partial fits of TP, DOC, DIC and color (Fig. 3; Fig. S6). Many predictors between the linear and non-linear models were similar, but the GAMM identified nonlinearities, explained more variation and captured the effect of color which had a pronounced unimodal fit that would have been modeled as a flat line in the LMM. Additionally, although the relationship between cyanobacteria and nitrogen was significant in the LMM and consistent across ecozones (Fig. S10), TP was the stronger predictor across both model types (Table 1), with the highest percentage of cyanobacteria biomass occurring at low TN:TP (Fig. S11).

LMRTs were then used to identify key thresholds in predictor variables and their interactions. Similar to the LMM and GAMM results, we found TP to be the most important variable but the LMRT identified key thresholds, including elevated cyanobacteria biomasses in lakes with TP above 41 µg/L (node 1; threshold = 41 µg/L and node 2; threshold = 17 µg/L; Fig. 4A). Cyanobacteria biomass was also greater in sites with longer residence times (node 6; threshold = 227 days) and higher Ca^{2+} (node 9; threshold = 8.6 mg/L); these results echo the LMM and GAMM findings. Meanwhile Ca^{2+} was significant in the LMM but not the GAMM. In all three model types, ecozone accounted for a small percentage of the variance and moderate intra-ecozone correlation, suggesting that cyanobacteria biomass was only slightly more similar within than among ecozones (Table 1; Fig. S5A).

Genus-specific cyanobacteria models

We applied a two-part hurdle approach with the focal cyanobacterial genus data because of the presence of many zeros in the 602 lake set. The presence of *Microcystis* (binomial GLMM; part 1) increased with the biomass of small cladoceran and warmer water-column

temperatures, while the biomass of *Microcystis* (GAMM; part 2) continued to increase with TP, cyclopoid copepod biomass, water-column temperature and % pastureland. Ecozone significantly explained the presence of *Microcystis* (part 1 of hurdle model; Fig. S5B), but once present, the continued increase in Microcystis was driven by local lake factors (part 2 of hurdle model) (Table 2). Threshold increases in *Microcystis* biomass occurred when TP increased beyond 51 µg/L (node 2) and again beyond 70 µg/L (node 1). Interestingly, the first TP node (70 µg/L) was higher than that for total cyanobacteria biomass (41 µg/L). Other *Microcystis* threshold responses recapitulate the hurdle model, with increasing cyclopoid copepod biomass (node 9; threshold = 14 µg/L), water-column temperature (node 3; threshold= 17°C), and soluble reactive phosphorus (SRP) (node 4; threshold = 12 µg/L) emerging as significant (Fig. 4B).

Of the three focal genera, the fixed effects best predicted *Aphanizomenon* presence (Table 2). On the other hand, the generalized models (part 2) for *Aphanizomenon* and *Microcystis* biomasses were comparable (Table 2). TP was the most important variable in modeling the presence and biomass of *Aphanizomenon*. The binomial model (part 1) identified increasing TP, daphnid biomass, depth, color, dissolved oxygen and decreasing Na^+ as key drivers of *Aphanizomenon* presence, while The GAMM for *Aphanizomenon* identified TP, daphnid biomass, altitude, color, %forestry and Na^+ as drivers of the continued rise in *Aphanizomenon* biomass. The most parsimonious LMRT for *Aphanizomenon* identified two initial splits for TP (node 1; threshold = 47 µg/L, node 2; threshold = 21 µg/L). A third split was identified, with greater amounts of daphnid biomass being associated with higher *Aphanizomenon* biomass (node 5; threshold = 6.8 µg/L) (Fig. 4C).

The presence and biomass of *Dolichospermum* was the least well predicted of the three genera (Table 2). *Dolichospermum* presence and biomass increased in deeper lakes.

Dolichospermum presence was driven by DOC, daphnid biomass, dissolved oxygen, precipitation and K^+. *Dolichospermum* biomass was additionally driven by TP, DIC and copepod biomass. The LMRT identified DIC (node 1; threshold = 2.1 µg/L), DOC (node 3; threshold = 4.4 µg/L), depth (node 5; threshold = 4.4 m) and then TP, but at much lower threshold than the other two focal genera and total cyanobacteria (node 7; threshold = 18 µg/L) (Fig. 4D). Like the LMRTs for *Microcystis* and *Aphanizomenon*, the random effect accounted for little additional residual variance (Fig. S5 B-D). Overall, TP was the only predictor significant across the biomass of all three genera; a relationship of which was consistent across ecozones (Fig. S7-9). However, genera were also associated with greater biomasses of zooplankton groups, which may be driven by several mechanisms (Fig. S12C) including food supplementation (Fig. S13).

Factors associated with variation in cyanobacteria community structure

To identify the factors that explain variation in the cyanobacteria overall community, a redundancy analysis (RDA) paired with variation partitioning was performed (Fig. 5; Table 3). Forward selection of the RDA identified TP, depth, daphnid biomass, color, DIC, Na^+, SO_4^{2-}, DO, % urban land and 30-day sum of heat degree days as the most important explanatory variables, and resemble those identified from modeling biomasses (Fig. 5). However, the amount of variation explained for the full community was lower compared to analyses on individual biomasses, with the RDA and variation partitioning explaining 13% of the total variation (Fig. 5; Table 3). The cyanobacteria community was primarily distributed along a trophic gradient (RDA axis 1). The constrained ordination biplot showed that *Aphanizomenon* and *Microcystis* were associated with higher TP concentrations, primarily sites in the Boreal Plains and Prairies. These lakes were also those with higher daphnid biomass, DIC, color and ions. In contrast, the picocyanobacteria (i.e., *Aphanocapsa*, *Chroococcus* and *Merismopedia*) were associated with

lower TP and sites primarily from Eastern Canada. Along the second RDA axes, *Dolichospermum* was associated with deeper systems with higher profundal DO and % urban lands; echoing findings from the *Dolichospermum*-specific models. Finally, variance partitioning showed that predictors grouped under water quality accounted for the greatest unique proportion of community composition variation, followed by the zooplankton groups and physiographic variables (Table 3).

Discussion

Based on the analysis of 640 lakes across Canada, we conducted a rigorous statistical modeling of total cyanobacteria and genera-specific biomasses. In agreement with much of the earlier literature, total phosphorus (TP) emerged as the most important explanatory variable of total cyanobacteria biomass, the biomass of select bloom-forming genera, as well as community composition. However, we also identified several other significant water quality predictors (including TN, DOC, DIC, Ca^{2+} and Na^{+}), physiographic (water residence time and depth) and zooplankton variables. The zooplankton results are perhaps the most interesting because while the traditional view is to expect daphnid biomass to be negatively associated with cyanobacteria due to their potential toxicity, colonial morphology and poor food quality, we found instead that total cyanobacteria biomass was positively related to zooplankton biomass, especially daphnid and cyclopoid copepod biomass. Lastly, we examined spatial variation in baseline values within models and found that the within ecozone correlation in biomass value was moderate.

Predictors of cyanobacteria community composition

Water quality variables explained most of the variation community composition, despite the inclusion of several variable categories (Table 3). Most taxa were organized along the first

RDA axis, which reflected primarily a trophic state gradient. High TP sites were associated with *Microcystis* and *Aphanizomenon* biomass, with many from the Boreal Plains and Prairies. These sites and taxa were also associated with higher daphnid biomass, DIC, ions (Na^+ and SO_4^{2-}) and color. At the opposite end of the RDA axis, small cyanobacteria such as *Aphanocapsa* were associated with lower nutrient lakes in eastern Canada. No other environmental variables were associated with these taxa, highlighting the importance of TP in determining community composition. Support for the predominant importance of trophic state explaining variation in cyanobacteria community composition extends to lakes worldwide (Beaulieu et al. 2014; Wood et al. 2017).

Along the second RDA axis, *Dolichospermum* clustered separately from the other dominant taxa being associated with deeper lakes with high profundal dissolved oxygen. Our results generally agree with the literature which reports this taxon as abundant in deeper systems and capable of blooming in low nutrient lakes (Mantzouki et al. 2016; Salmaso et al. 2015). The importance of depth further supports our univariate results, as depth was the only variable significant in modeling both *Dolichospermum* presence and biomass. Depth was also included in its genus-specific LMRT, with increasing biomass in lakes deeper than 4.4m (Fig. 4D).

Nutrients as the top predictors of cyanobacteria biomass across Canada

Across Canadian lakes, we identified TP as the top predictor of total cyanobacteria biomass (Table 1) in both linear and non-linear modeling methods, irrespective of region (Fig. S6-9). Our findings, notably the non-linear responses, echo earlier work such as Downing et al. (2001) who identified an increasing risk of cyanobacteria dominance above 20-30 µg/L of TP, plateauing at approximately 100 µg/L. More recently, Carvalho et al. (2013) noted a similar initial linear response in cyanobacteria biomass from TP of 20 to 100 µg/L, followed by an

asymptote just above 100 µg/L. Our non-linear modeling and LMRT identified a threshold at approximately 40 µg/L, above which our highest cyanobacteria biomasses occurred; a slightly greater inflection point (50 µg/L) was reported by Dolman et al. (2012). In many of these large-scale analyses, the relation between cyanobacteria and nitrogen remains linear, and as a result, can explain more variation than TP (Downing et al. 2001; Dolman et al. 2012). In the case of our Canadian lake set, TN was a significant predictor of cyanobacteria biomass only in the linear model (Table 1; Fig. S10), whereas TP explained more variation across all the statistical models. Several studies have also pointed to a modest relationship between cyanobacteria dominance and the TN:TP ratio, especially when both nutrients are in sufficient supply (Downing et al. 2001; Chorus et al. 2021). Our univariate models showed weak support for the role of the N:P in predicting cyanobacteria biomass, although we did observe a similar threshold as described in Smith (1983) when examining cyanobacteria dominance, where below N:P of 29:1 there was generally higher and much more variable percent cyanobacteria (Fig. S11).

Our models for the biomasses of *Microcystis*, *Aphanizomenon* and *Dolichospermum* also identified TP as the top predictor across model types, although with different thresholds (Table 2; Fig. 4B-C). *Microcystis* had the highest TP threshold at approximately 70 µg/L whereas the TP threshold for *Aphanizomenon* was 47 µg/L and *Dolichospermum* was the lowest at 18 µg/L. Similarly, Vuorio et al. (2020) found that the highest TP threshold was observed for *Microcystis* at 50 µg/L.

Positive relationship with zooplankton biomass

Interactions between zooplankton and cyanobacteria have been of interest in laboratory and field studies for decades, but contradictory results have emerged across studies (Haney 1987; Ger et al. 2014). As a result of various cyanobacterial grazing defenses, total zooplankton

biomass could decrease with increasing cyanobacteria biomass. The mechanical interference and poor nutritional value of many cyanobacteria (Ger et al. 2014; 2016a), along with their production of toxic metabolites (DeMott et al. 1991), are all factors causing potential negative effects on zooplankton and may reinforce cyanobacteria dominance. However, opposite relationship and mechanisms have also been put forward and align with our observations. In all our statistical cyanobacteria models, the biomass of either one or several zooplankton groups were selected as significant positive predictors, with specific groups being selected to separate low from high cyanobacteria biomass sites. For example, increasing cyclopoid biomass was selected in the hurdle (part 2) and LMRT for *Microcystis*, which tends to increase when cyclopoid biomass is above 13 µg/L of cyclopoid copepods (Table 2; Fig. 4B). Similarly, daphnid biomass was selected as a significant predictor of *Aphanizomenon* biomass (part 2 of hurdle model) and was a threshold predictor of *Daphnia* (Fig. 4C).

Zooplankton adaptations and physical factors can explain the positive relationships between cyanobacteria and zooplankton biomass (Fig. S12C). First, zooplankton commonly exposed to cyanobacterial blooms can supplement their diet with other more edible algal groups and heterotrophic flagellates that co-exist with cyanobacteria (Briland et al. 2020). In turn, the consumption of more palatable algae may indirectly promote cyanobacteria through a reduction of competitive pressure Cremona et al. (2018). We found a slight positive correlation between the biomass of cyanobacteria and all other algal groups (Fig. S13), and as such, zooplankton in the cyanobacteria-dominated lakes may have access to more palatable food sources. Additionally, cyanobacteria may indirectly fuel the microbial loop and zooplankton growth through trophic upgrading (Bec et al., 2006), whereby heterotrophic nanoflagellates that ingest cyanobacteria can upgrade the fatty acid quality, thus increasing food quality for zooplankton.

An active avoidance of harmful cells can also enable zooplankton to coexist with high cyanobacteria biomass (Tillmanns et al. 2011; Ger et al. 2016a). Copepods in particular have shown a high degree of selectivity, making use of chemosensory signals (toxins and other secondary metabolites) as detection cues (Ger et al. 2014; Ger et al. 2016b), even distinguishing between potentially toxic and non-toxic *Microcystis* strains (Ger et al. 2011). Prey selectivity in copepods (Ger et al. 2011; Leitão et al. 2018) may explain the importance of cyclopoid copepod in the models for *Microcystis*. After TP, cyclopoid copepod biomass was a significant predictor of increasing *Microcystis* biomass (Table 2), which was highest in lakes with cyclopoid copepod biomass greater than 13 µg/L (Fig. S5B).

Selection for tolerant *Daphnia* genotypes and phenotypes may further contribute to the observed positive relationship between *Daphnia* and cyanobacteria biomass, as numerous sites have been eutrophic for decades or more. Increasing *Daphnia* biomass was selected in both linear and non-linear models for total cyanobacteria biomass, and in the *Aphanizomenon* Dolichospermum hurdle models (Table 2). Daphnid biomass was higher compared to other crustacean zooplankton in the most eutrophic ecozones (Paquette et al. 2021). Daphnids have been shown to evolve a tolerance to toxic cyanobacteria (*Microcystis*) in less than 10 generations (Gustafsson and Hansson 2004; Tillmanns et al. 2011). Furthermore, their tolerance and detoxification strategy may be strengthened by a longer exposure history to blooms (Sarnelle and Wilson 2005). *Daphnia pulicaria*, in particular, has been shown to have high tolerance to nutrient enrichment by exhibiting rapid clonal evolution, metabolic variation and local adaptation (Chislock et al. 2019; Ghadouani and Pinel-Alloul 2002), allowing it to increase growth rates and fecundity in hypereutrophic lakes (Moody et al. 2022). In our study, 80% of all *Daphnia* biomass was *D. pulicaria*.

Lastly, the positive relationship between cyanobacteria and zooplankton groups may also be the result of reduced fish predation and its cascading effects on the food web. The highest biomasses of cyanobacteria and zooplankton were recorded in the Prairies and Boreal Plains. Many of these shallow, eutrophic lakes may be fishless, or experience regular winter fish kills (Jackson et al. 2007; Balayla et al. 2010), thus resulting in stronger zooplankton control over phytoplankton.

Weak effect of water temperature

Contrary to our hypothesis, water temperature did not emerge as a significant predictor of total cyanobacteria biomass across Canadian lakes. While some large-scale analyses have found temperature as a strong predictor (Beaulieu et al. 2013), these studies captured a much larger temperature gradient across sites than ours. Our study sampling occurred over a shorter time window and a more limited latitudinal range (range in temperature: Kosten et al. (2012) = 22ºC; Beaulieu et al. (2013) = 27ºC, present study = 18°C). It is also possible that the indirect effects of temperature were captured by lake depth and altitude, which did emerge as significant variables in the genus specific models for *Dolichospermum* and *Aphanizomenon* (Table 2). Only in the models for *Microcystis* did we detect a direct temperature signal with increased water column temperature predicting presence as well as increased biomass (Table 2). The LMRT identified an average water-column temperature threshold of 17°C, above which *Microcystis* biomass increased (Fig. 4B). *Microcystis* is commonly considered to have a higher temperature optima relative to other cyanobacteria (Paerl and Otten 2013).

Support for additional water quality predictors

Several additional chemical and physical variables were identified as second tier predictors. DIC and numerous major ions (including Ca^{2+} and Na^{+}) exhibited hump-shaped

relationships with total cyanobacteria biomass. DIC was selected in both linear and non-linear models for total cyanobacteria and was the top distinguishing variable between low and high *Dolichospermum* biomass (Table 2; Fig. 4D). Here, DIC exhibited a slight negative response in the linear model, but with the GAMM it became clear that the relationship is non-linear. Cyanobacteria possess carbon concentrating mechanisms and can utilize different forms of inorganic carbon depending on availability (Talling 1976; Pick and Lean 1987). With this, their photosynthesis is less impaired, particularly in high pH environments. For ions, Ca^{2+} was selected in the linear model and as the second most important splitting variable of low and high biomass sites after TP in the LMRT (Fig. 4A). Meanwhile, Na^{+} was included in the final linear and non-linear models for cyanobacteria biomass. Like DIC, Na^{+} exhibited a slight negative relationship with cyanobacteria in the linear model, but GAMMs identified a non-linear relationship. Conditions likely become uninhabitable for cyanobacteria at the highest concentrations of DIC and Na^{+}, which led to the observed negative relationship in the linear model. Combining this analysis with the GAMM captured cyanobacteria's full response along the extensive gradient. Calcium, Na^{+} and overall salinity have been positively and negatively linked to cyanobacteria in a number of studies, including a temporal analysis of ten Brazilian reservoirs, where Ca^{2+} and Na^{+} had the most positive effects on non-heterocystous filamentous cyanobacteria and picocyanobacteria (Amorim et al. 2020). Several freshwater genera are tolerant of increasing salinity, including *Dolichospermum* and *Microcystis* (Merel et al. 2013). The highest cyanobacteria biomasses were recorded in the Prairies and Boreal Plains, regions in which many lakes are high in salinity and ion content. However, increasing major ion concentrations are unlikely to directly promote cyanobacteria across Canadian lakes; rather, tolerance to salinity could provide cyanobacteria with a competitive advantage up to certain

thresholds where conditions become too harsh. Several of these same variables, including DIC, Na^+ and K^+, were also significant predictors of composition, as well as *Aphanizomenon* and *Dolichospermum* biomass. These relationships were again negative in the linear models, suggesting the genera may initially benefit from increased ions, but are not present at the highest concentrations of DIC and ions. Overall, it is difficult to decipher whether the association we observed between DIC and ions with cyanobacteria is mechanistic or just correlative.

Our DOC findings resemble the results from a previous mesocosm study, where brown, DOC-rich waters had a humped-shaped relationship with cyanobacteria (Feuchtmayr et al. 2019). DOC was included in our final models for both total cyanobacteria and *Dolichospermum* biomass (Table 1, 2). DOC was also selected after DIC in the genus specific LMRT for *Dolichospermum*. Across all of our lakes, DOC was moderately related to color ($r^2 = 0.42$). We also observed that DOC was positively related to nutrient concentrations (TP~DOC, $r^2 = 0.45$), as has been reported in a continental study of lakes in the US (Stetler et al. 2021). Overall, our findings were concordant with DOC and nutrients being positively associated with algal production metrics, especially in low to intermediate DOC sites. However, this relationship was disappeared in high DOC and colored lakes due to light limitation.

Among physical variables, we found longer water residence time promoted increasing cyanobacteria biomass (Table 1; Fig. 4A), as observed in other regions (Giani et al. 2020). Several cyanobacteria have slower reproductive and growth rates compared to many smaller phytoplankton (Paerl and Otten 2013). As a result, cyanobacteria may directly benefit from decreased flushing rates from longer residence times.

Reflections on statistical approaches

Across the statistical analyses included in this study, there was consistency in the importance of TP as a predictor. The use of multiple analyses provided additional insights. For example, the LMRT identified the unique importance of temperature for *Microcystis*, daphnid biomass for *Aphanizomenon* and depth for *Dolichospermum*. We chose to start with linear models as this work builds on decades of literature set in this approach, but as expected, the GAMMs demonstrated some important nonlinearities and increase proportion of variance explained with a smaller number of predictors. Differences in predictors between linear and non-linear approaches were apparent but most of these variables were relatively weak unique predictors. We suggest that when sampling across a broad environmental gradient, one should adopt non-linear models as they can provide stronger explanatory power and may uncover the true relationship between the predictor and response (e.g., DOC). In contrast, if the study is local or along a more constrained gradient, the relationships may come out as linear. If interested in key thresholds for setting management guidelines and water quality advisories, LMRTs are effective in detecting what tipping points, and identify how they may vary depending on the study organism. In contrast, GAMMs can help better understand the ecological questions.

Despite the efforts to standardize sampling and taxonomy, as well as considering an exhaustive suite of predictor variables across a large gradient of lakes, the amount of variation explained overall remained modest. We may have approached the predictive limit of cyanobacteria models when based on snap-shot sampling (one-time in summer). Further improvements are likely to be gained through repeated sampling and growing season averages to capture seasonal variability.

Conclusions

Our pan-Canadian study considered a wider range of biotic and abiotic predictor variables than most to date. Overall, we corroborated that nutrients, notably TP, as the most important driver of total cyanobacteria biomass and community composition. In addition, analyses identified a positive relationship between total cyanobacteria and all zooplankton groups, the latter of which may indirectly promoting cyanobacterial growth through selective grazing and microbial trophic upgrading of cyanobacteria. Several additional drivers of cyanobacteria biomass were retained in this study of north-temperate to subarctic lakes, and the linear, non-linear and threshold models used herein can serve as a key baseline for both fundamental (GAMMs) and applied (LMRT) research.

Acknowledgements

This project was made possible thanks to the NSERC LakePulse network (grant awarded to IGE & BEB, and numerous others). We thank the LakePulse Director, Yannick Huot, and all the field and lab teams for their incredible work. We also thank Dr. Michael Agbeti for his dedicated work on the phytoplankton taxonomy and biomass estimates. Additional funding was provided by NSERC in the form a scholarship to PM and the Canada Research Chair program (awarded to IGE). Feedback from members of the Gregory-Eaves and Beisner labs were helpful in shaping this manuscript.

References

Amorim, C. A., Ê. W. Dantas, and A. do N. Moura. 2020. Modeling cyanobacterial blooms in tropical reservoirs: The role of physicochemical variables and trophic interactions. Sci. Total Environ. **744**. doi:10.1016/j.scitotenv.2020.140659

Balayla, D., T. L. Lauridsen, M. Søndergaard, and E. Jeppesen. 2010. Larger zooplankton in Danish lakes after cold winters: Are winter fish kills of importance? Hydrobiologia **646**: 159–172. doi:10.1007/s10750-010-0164-4

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. **58**: 1736–1746. doi:10.4319/lo.2013.58.5.1736

Beaulieu, M., F. Pick, M. Palmer, S. Watson, J. Winter, R. Zurawell, and I. Gregory-Eaves. 2014. Comparing predictive cyanobacterial models from temperate regions. Can. J. Fish. Aquat. Sci. **71**: 1830–1839. doi:10.1139/cjfas-2014-0168

Bec, A., D. Martin-Creuzburg, and E. von Elert. 2006. Trophic upgrading of autotrophic picoplankton by the heterotrophic nanoflagellate Paraphysomonas sp. Limnol. Oceanogr. **51**: 1699–1707. doi:10.4319/lo.2006.51.4.1699

Borcard, D., F. Gillet, and P. Legendre. 2018. *In* R. Gentleman, K. Hornik, and G. Parmigiani [eds.], Numerical ecology with R, 2nd ed. Springer Nature.

Breiman, L. 2001. Random forests. Mach. Learn. **45**: 5–32. doi:10.1023/A:1010933404324

Briland, R. D., J. P. Stone, M. Manubolu, J. Lee, and S. A. Ludsin. 2020. Cyanobacterial blooms modify food web structure and interactions in western Lake Erie. Harmful Algae **92**. doi:10.1016/j.hal.2019.03.004

Carvalho, L., C. Mcdonald, C. de Hoyos, and others. 2013. Sustaining recreational quality of

European lakes: Minimizing the health risks from algal blooms through phosphorus control. J. Appl. Ecol. **50**: 315–323. doi:10.1111/1365-2664.12059

Chislock, M. F., R. R. B. Kaul, K. A. Durham, O. Sarnelle, and A. E. Wilson. 2019. Eutrophication mediates rapid clonal evolution in *Daphnia pulicaria*. Freshw. Biol. **64**: 1275–1283. doi:10.1111/fwb.13303

Chorus, I., J. Fastner, and M. Welker. 2021. Cyanobacteria and cyanotoxins in a changing environment: Concepts, controversies, challenges. Water **13**: 1–41. doi:10.3390/w13182463

Cremona, F., L. Tuvikene, J. Haberman, P. Nõges, and T. Nõges. 2018. Factors controlling the three-decade long rise in cyanobacteria biomass in a eutrophic shallow lake. Sci. Total Environ. **621**: 352–359. doi:10.1016/j.scitotenv.2017.11.250

DeMott, W. R., Q. -X Zhang, and W. W. Carmichael. 1991. Effects of toxic cyanobacteria and purified toxins on the survival and feeding of a copepod and three species of *Daphnia*. Limnol. Oceanogr. **36**: 1346–1357. doi:10.4319/lo.1991.36.7.1346

Dolman, A. M., J. Rücker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. PLoS One **7**. doi:10.1371/journal.pone.0038757

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/cjfas-58-10-1905

Ecological Stratification Working Group (Canada), Center for Land, Biological Resources Research (Canada). State of the Environment Directorate. 1996. A national ecological framework for Canada. Centre for Land and Biological Resources Research; Hull, Quebec: State of the Environment Directorate.

Von Elert, E., D. Martin-Creuzburg, and J. R. Le Coz. 2003. Absence of sterols constrains

carbon transfer between cyanobacteria and a freshwater herbivore (*Daphnia galeata*). Proc. R. Soc. B Biol. Sci. **270**: 1209–1214. doi:10.1098/rspb.2003.2357

Feld, C. K., P. Segurado, and C. Gutiérrez-Cánovas. 2016. Analysing the impact of multiple stressors in aquatic biomonitoring data: A 'cookbook' with applications in R. Sci. Total Environ. **573**: 1320–1339. doi:10.1016/j.scitotenv.2016.06.243

Feuchtmayr, H., T. G. Pottinger, A. Moore, M. M. De Ville, L. Caillouet, H. T. Carter, M. G. Pereira, and S. C. Maberly. 2019. Effects of brownification and warming on algal blooms, metabolism and higher trophic levels in productive shallow lake mesocosms. Sci. Total Environ. **678**: 227–238. doi:10.1016/j.scitotenv.2019.04.105

Fokkema, M., N. Smits, A. Zeileis, T. Hothorn, and H. Kelderman. 2018. Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees. Behav. Res. Methods **50**: 2016–2034. doi:10.3758/s13428-017-0971-x

Ger, K. A., R. Panosso, and M. Lurling. 2011. Consequences of acclimation to *Microcystis* on the selective feeding behavior of the calanoid copepod *Eudiaptomus gracilis*. Limnol. Oceanogr. **56**: 2103–2114. doi:10.4319/lo.2011.56.6.2103

Ger, K. A., L. A. Hansson, and M. Lürling. 2014. Understanding cyanobacteria-zooplankton interactions in a more eutrophic world. Freshw. Biol. **59**: 1783–1798. doi:10.1111/fwb.12393

Ger, K. A., P. Urrutia-Cordero, P. C. Frost, L. A. Hansson, O. Sarnelle, A. E. Wilson, and M. Lürling. 2016a. The interaction between cyanobacteria and zooplankton in a more eutrophic world. Harmful Algae **54**: 128–144. doi:10.1016/j.hal.2015.12.005

Ger, K. A., E. Leitao, R. Panosso, and K. Havens. 2016b. Potential mechanisms for the tropical copepod *Notodiaptomus* to tolerate *Microcystis* toxicity. J. Plankton Res. **38**: 843–854.

doi:10.1093/plankt/fbw036

Ghadouani, A., and B. Pinel-Alloul. 2002. Phenotypic plasticity in *Daphnia pulicaria* as an adaptation to high biomass of colonial and filamentous cyanobacteria: experimental evidence. J. Plankton Res. **24**: 1047–1056. doi:10.1093/plankt/24.10.1047

Ghadouani, A., B. Pinel-Alloul, and E. E. Prepas. 2006. Could increased cyanobacterial biomass following forest harvesting cause a reduction in zooplankton body size structure? Can. J. Fish. Aquat. Sci. **63**: 2308–2317. doi:10.1139/F06-117

Giani, A., Z. E. Taranu, G. von Rückert, and I. Gregory-Eaves. 2020. Comparing key drivers of cyanobacteria biomass in temperate and tropical systems. Harmful Algae **97**. doi:10.1016/j.hal.2020.101859

Gustafsson, S., and L. A. Hansson. 2004. Development of tolerance against toxic cyanobacteria in *Daphnia*. Aquat. Ecol. **38**: 37–44. doi:10.1023/B:AECO.0000020985.47348.5e

Haney, J. F. 1987. Field studies on zooplankton-cyanobacteria interactions. New Zeal. J. Mar. Freshw. Res. **21**: 467–475. doi:10.1080/00288330.1987.9516242

Hillebrand, H., C. D. Dürselen, D. Kirschtel, U. Pollingher, and T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. J. Phycol. **35**: 403–424. doi:10.1046/j.1529-8817.1999.3520403.x

Ho, J. C., A. M. Michalak, and N. Pahlevan. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature **574**: 667–670. doi:10.1038/s41586-019-1648-7

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. **695**: 133668. doi:10.1016/j.scitotenv.2019.133668

Ishwaran, H., and U.B. Kogalur. 2016. rf-src: Random Forests for Survival, Regression and Classification. R package version 2.2.

Jackson, L. J., T. L. Lauridsen, M. Søndergaard, and E. Jeppesen. 2007. A comparison of shallow Danish and Canadian lakes and implications of climate change. Freshw. Biol. **52**: 1782–1792. doi:10.1111/j.1365-2427.2007.01809.x

Komárek, J. 2013. Cyanoprokaryota 3. Teil: Heterocytous genera. In: Büdel, B., Gärtner, G., Krienitz, L., Schagerl, M. (eds.), Süswasserflora von Mitteleuropa/ Freshwater flora of Central Europe. Springer Spektrum Berlin, Heidelberg.

Komárek, J., and K. Anagnostidis. 1998. Cyanoprokaryota 1.Teil: Chroococcales. In: Ettl, H., Gärtner, G., Heynig, H., Mollenhauer, D. (eds.), Süsswasserflora von Mitteleuropa 19/1. Jena-Stuttgart-Lübeck-Ulm, Gustav Fischer.

Komárek, J., and K. Anagnostidis. 2005. Cyanoprokaryota 2. Teil: Oscillatoriales. In: Büdel, B., Krienitz, L., Gärtner, G., Schagerl, M. (eds.), Süsswasserflora von Mitteleuropa 19/2. Elsevier/Spektrum, Heidelberg.

Kosten, S., V. L. M. Huszar, E. Bécares, and others. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Glob. Chang. Biol. **18**: 118–126. doi:10.1111/j.1365-2486.2011.02488.x

Lai, J., Y. Zou, J. Zhang, and P. R. Peres-Neto. 2022. Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package. Methods Ecol. Evol. 1–7. doi:10.1111/2041-210x.13800

Leitão, E., K. A. Ger, and R. Panosso. 2018. Selective grazing by a tropical copepod (*Notodiaptomus iheringi*) facilitates microcystis dominance. Front. Microbiol. **9**: 1–11. doi:10.3389/fmicb.2018.00301

Lund, J. W. G., C. Kipling, and E. D. Le Cren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia **11**: 143–170. doi:10.1007/BF00007865

Mantzouki, E., P. M. Visser, M. Bormans, and B. W. Ibelings. 2016. Understanding the key ecological traits of cyanobacteria as a basis for their management and control in changing lakes. Aquat. Ecol. **50**: 333-350. doi:10.1007/s10452-015-9526-3

Merel, S., D. Walker, R. Chicana, S. Snyder, E. Baurès, and O. Thomas. 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. Environ. Int. **59**: 303–327. doi:10.1016/j.envint.2013.06.013

Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. **7**: 1–11. doi:10.1038/ncomms13603

Molot, L. A., S. B. Watson, I. F. Creed, and others. 2014. A novel model for cyanobacteria bloom formation: The critical role of anoxia and ferrous iron. Freshw. Biol. **59**: 1323–1340. doi:10.1111/fwb.12334

Moody, E. K., T. J. Butts, R. Fleck, P. D. Jeyasingh, G. M. Wilkinson. 2022. Eutrophication-driven eco-evolutionary dynamics indicated by differences in stoichiometry traits among populations of *Daphnia pulicaria*. Freshw. Biol **67**: 353-364. doi:10.1111/FWB.13845

Muñoz Sabater, J. 2019. ERA5-Land hourly data from 1981 to present. Copernicus Clim. Chang. Serv. Clim. Data Store. doi:10.24381/cds.e2161bac

Naimi, B. 2015. Usdm: Uncertainty Analysis for Species Distribution Models. R package version 1.1-15.

NSERC Canadian Lake Pulse Network. 2021. In M.-P. Varin, M.-L. Beaulieu, and Y. Huot [eds.], NSERC Canadian Lake Pulse Network field manual 2017 - 2018 - 2019 surveys. Université de Sherbrooke. doi:10.17118/11143/18662

Oksanen, J. and others. 2019. vegan: The community ecology package. R package version 2.5-7.

O'Neil, J. M., T. W. Davis, M. A. Burford, and C. J. Gobler. 2012. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. Harmful Algae **14**: 313–334. doi:10.1016/j.hal.2011.10.027

Paerl, H. W., and J. Huisman. 2009. Climate change: A catalyst for global expansion of harmful cyanobacterial blooms. Environ. Microbiol. Rep. **1**: 27–37. doi:10.1111/j.1758-2229.2008.00004.x

Paerl, H. W., N. S. Hall, and E. S. Calandrino. 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. Sci. Total Environ. **409**: 1739–1745. doi:10.1016/j.scitotenv.2011. 02.001

Paerl, H. W., and T. G. Otten. 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microb. Ecol. **65**: 995–1010. doi:10.1007/s00248-012-0159-y

Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. **46**: 1349–1363. doi:10.1016/j.watres.2011.08.002

Paquette, C., I. Gregory-Eaves, and B. E. Beisner. 2021. Multi-scale biodiversity analyses identify the importance of continental watersheds in shaping lake zooplankton biogeography. J. Biogeogr. **48**: 2298–2311. doi:10.1111/jbi.14153

Pick, F. R. 2016. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria.

Can. J. Fish. Aquat. Sci. **73**: 1149–1158. doi:10.1139/cjfas-2015-0470

Pick, F. R., and D. R. S. Lean. 1987. The role of macronutrients (C, N, P) in controlling cyanobacterial dominance in temperate lakes. New Zeal. J. Mar. Freshw. Res. **21**: 425–434. doi:10.1080/00288330.1987.9516238

Pinheiro, J., D. Bates, S. DebRoy, D. Sarkar, R Core Team. 2016. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-128.

R Core Team. 2021. R: A language and environment for statistical computing.

Reid, A. J., A. K. Carlson, I. F. Creed, and others. 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. **94**: 849–873. doi:10.1111/brv.12480

Rigosi, A., C. C. Carey, B. W. Ibelings, and J. D. Brookes. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. Limnol. Oceanogr. **59**: 99–114. doi:10.4319/lo.2014.59.1.0099

Rott, E. 1981. Some results from phytoplankton counting intercalibrations. Schweizerische Zeitschrift für Hydrol. **43**: 34–62. doi:10.1007/BF02502471

Salmaso, N., C. Capelli, S. Shams, and L. Cerasino. 2015. Expansion of bloom-forming Dolichospermum lemmermannii (Nostocales, Cyanobacteria) to the deep lakes south of the Alps: Colonization patterns, driving forces and implications for water use. Harmful Algae **50**: 76–87. doi:10.1016/j.hal.2015.09.008

Sarnelle, O., and A. E. Wilson. 2005. Local adaptation of *Daphnia pulicaria* to toxic cyanobacteria. Limnol. Oceanogr. **50**: 1565–1570. doi:10.4319/lo.2005.50.5.1565

Shan, K., L. Song, W. Chen, and others. 2019. Analysis of environmental drivers influencing interspecific variations and associations among bloom-forming cyanobacteria in large,

shallow eutrophic lakes. Harmful Algae **84**: 84–94. doi:10.1016/j.hal.2019.02.002

Smith, V. H. 1983. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. Science. **221**: 669–671. doi:10.1126/science.221.4611.669

Stetler, J. T., L. B. Knoll, C. T. Driscoll, and K. C. Rose. 2021. Lake browning generates a spatiotemporal mismatch between dissolved organic carbon and limiting nutrients. Limnol. Oceanogr. Lett. **6**: 182–191. doi:10.1002/lol2.10194

Talling, J. F. 1976. The Depletion of Carbon Dioxide from Lake Water by Phytoplankton. J. Ecol. **64**: 79-121. doi:10.2307/2258685

Taranu, Z. E., I. Gregory-Eaves, P. R. Leavitt, and others. 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. Ecol. Lett. **18**: 375–384. doi:10.1111/ele.12420

Taranu, Z. E., R. W. Zurawell, F. Pick, and I. Gregory-Eaves. 2012. Predicting cyanobacterial dynamics in the face of global change: The importance of scale and environmental context. Glob. Chang. Biol. **18**: 3477–3490. doi:10.1111/gcb.12015

Tillmanns, A. R., S. K. Burton, and F. R. Pick. 2011. Daphnia Pre-Exposed to Toxic Microcystis Exhibit Feeding Selectivity. Int. Rev. Hydrobiol. **96**: 20-28. doi:10.1002/iroh.201011298

Vuorio, K., M. Järvinen, and N. Kotamäki. 2020. Phosphorus thresholds for bloom-forming cyanobacterial taxa in boreal lakes. Hydrobiologia **847**: 4389–4400. doi:10.1007/s10750-019-04161-5

Wetzel, R. G. 2001. Limnology: lake and river ecosystems. Gulf professional publishing. doi:10.1016/C2009-0-02112-6

Wood, S. N. 2021. mgcv: Mixed GAM computation vehicle with automatic smoothness estimation. R package version 1.8-38.

Wood, S. A., M. Y. Maier, J. Puddick, X. Pochon, A. Zaiko, D. R. Dietrich, and D. P. Hamilton. 2017. Trophic state and geographic gradients influence planktonic cyanobacterial diversity and distribution in New Zealand lakes. FEMS Microbiol. **93**: fiw234. doi:10.1093/femsec/fiw234

Zuur, A. F., and E. N. Ieno. 2016. Beginner's Guide to Zero-Inflated Models with R. Highl. Stat. Beginner's Guide.

Tables Chapter I

Table 1. Best linear mixed model (LMM) and generalized additive mixed model (GAMM) for cyanobacteria biomass across lakes where cyanobacteria were identified (n=602). Ecozone was tested as a random effect intercept. The final model predictors for each response variables are listed with significance level denoted as ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$. The marginal R^2 (marg r^2) represents the amount of variation explained by the fixed effects. The conditional R^2 (cond r^2) describes the amount of variation explained by both fixed and random effects. Standard deviation of the random intercept and residual are given (σ). For the GAMM, results include the R^2 and deviance explained for the full model as well as the p-values and estimated degrees of freedom (edf) for each predictor.

Linear mixed effect model (LMM)					
Significant linear predictors	**Model coefficients (SE)**	**Marg r^2**	**Cond r^2**	**σ-intercept**	**σ-residual**
logTP	0.51(0.07)***	0.44	0.47	0.21	0.99
logDOC	0.15(0.07)*				
logTN	0.23(0.08)**				
logCa^{2+}	0.29(0.08)***				
logDIC	-0.27(0.09)**				
logCYCL	0.11(0.05)*				
logNa^{+}	-0.29(0.06)***				
logDAPH	0.19(0.05)***				
logitPasture	0.16(0.05)**				
logRES_TIME	0.14(0.05)**				

Generalized additive mixed model (GAMM)					
Significant additive predictors	**edf**	***p*-value**	***F*-value**	**Dev.expl**	**R^2 adj**
s(logTP)	5.43	<0.001	24.2	57.4%	0.55
s(logDOC)	2.80	<0.001	3.40		
s(logDIC)	3.32	<0.001	2.55		
s(logNA^{+})	3.51	<0.001	2.33		
s(logDAPH)	0.95	<0.001	2.69		
s(logCOLOR)	3.06	<0.001	3.10		
s(logRES_TIME	1.82	<0.001	1.41		
s(ecozone)	5.74	0.006	1.28		

Table 2. Genera-specific hurdle models showing the best binomial generalized linear mixed model (GLMM; part 1) and generalized additive mixed models (GAMM; part 2) for three select genera. The GLMM is based on the presence-absence of each genus across all lakes (*n*=640). The GAMM models are based on biomass from lakes where that genus was present: *Microcystis* (MCYST, *n*=220), *Aphanizomenon* (APHZ, *n*=175), *Dolichospermum* (DOLI, *n*=303). Ecozone was tested as a random effect in each model and included where significant. The final model predictors for each response variables are listed with the significance level denoted as ***$p <$ 0.001; **$p < 0.01$; *$p < 0.05$. The marginal R^2 (marg r^2) represents the amount of variation explained by the fixed effects. The conditional R^2 (cond r^2) describes the amount of variation explained by both fixed and random effects. Results for the GAMM include the R^2 and deviance explained (dev.expl%). N.S = not significant.

	Binomial GLMM		GAMM	
Taxon	Predictors	Intercepts (SE)	Predictors	edf (F-values)
MCYST	logSMCLAD	0.31(0.11)**	logTP	2.78(7.41)***
	TEMP_WATERCOL	0.56(0.13)***	logCYCL	1.80(1.70)***
			DO	2.03(0.87)*
			TEMP_WATERCOL	1.42(0.58)*
			logitPASTURE	0.90(0.97)**
	Marg r^2= 0.08	**Cond r^2= 0.50**	**R^2 adj= 0.57**	**Dev.expl (%) = 58.5**
APHZ	logTP	1.02(0.17)***	logTP	2.79(4.21)***
	logDAPH	0.33(0.12)**	logDAPH	0.91(1.09)***
	logDEPTH	0.39(0.15)**	sqrtALTITUDE	0.95(0.92)**
	logCOLOUR	0.30(0.14)*	logCOLOUR	2.44(1.47)***
	DO	0.28(0.10)**	logitFORESTRY	1.74(0.59)*
	$logNa^+$	-0.31(0.14)*	$logNa^+$	2.42(0.87)*
	Marg r^2= 0.22	**Cond r^2= 0.28**	**R^2 adj= 0.55**	**Dev.expl (%) = 58.1**
DOLI	logDOC	0.63(0.14)***	logTP	4.65(9.61)***
	logDEPTH	0.44(0.11)***	logDIC	2.71(1.74)***
	logDAPH	0.34(0.09)***	logCYCL	0.79(0.42)*
	DO	0.35(0.09)***	logDEPTH	1.62(0.59)*
	sqrtPCPN7	0.31(0.11)**	logCALA	1.49(0.61)*
	$logK^+$	-0.35(0.12)**		
	sqrtPCPN30	-0.23(0.11)*		

Marg r^2= 0.12 **Cond r^2= N.S.** **R^2 adj= 0.34** **Dev.expl (%) = 36.1**

Table 3. Partitioning of variance in the structure of the full cyanobacteria community according to physiography, water quality, land-use, zooplankton and climate variable matrices. Partitioning of cyanobacteria biomass and the biomass of focal genera are also included. Total variation explained by the full model is represented by adjusted R^2. The values represent the unique contributions of each predictor group towards the explained variation (shared variation not shown). See Table S2 for the list of variables in each category.

	Variance explained	**Physiography**	**Water quality**	**Land-use**	**Zooplankton**	**Climate**
Cyanobacteria community	0.13	0.014	0.042	0.000	0.010	0.007
Total cyanobacteria	0.48	0.019	0.138	0.002	0.019	0.002
Microcystis	0.44	0.006	0.098	0.005	0.023	0.010
Aphanizomenon	0.34	0.024	0.103	0.000	0.016	0.001
Dolichospermum	0.16	0.032	0.063	0.000	0.009	0.014

Figures Chapter I

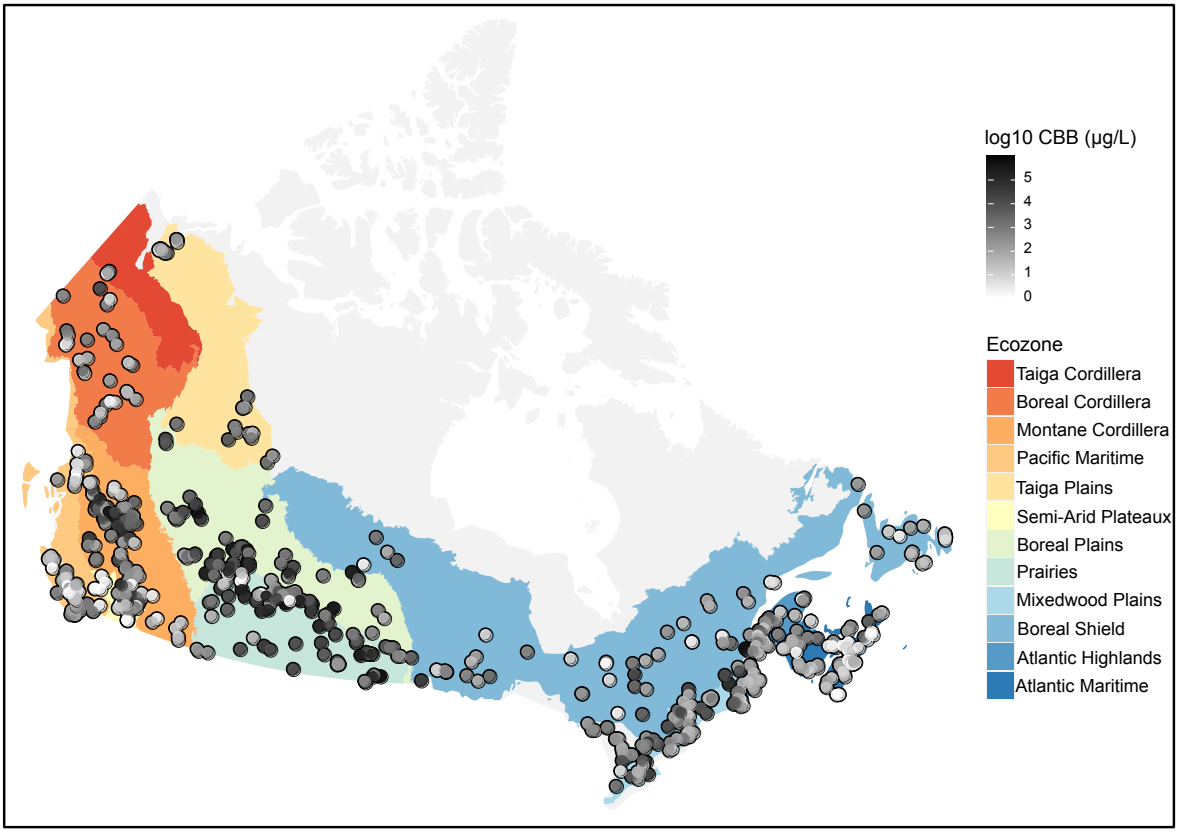

Figure 1. Map of log transformed total cyanobacteria biomass across study sites (n=640). Ecozones are color coded.

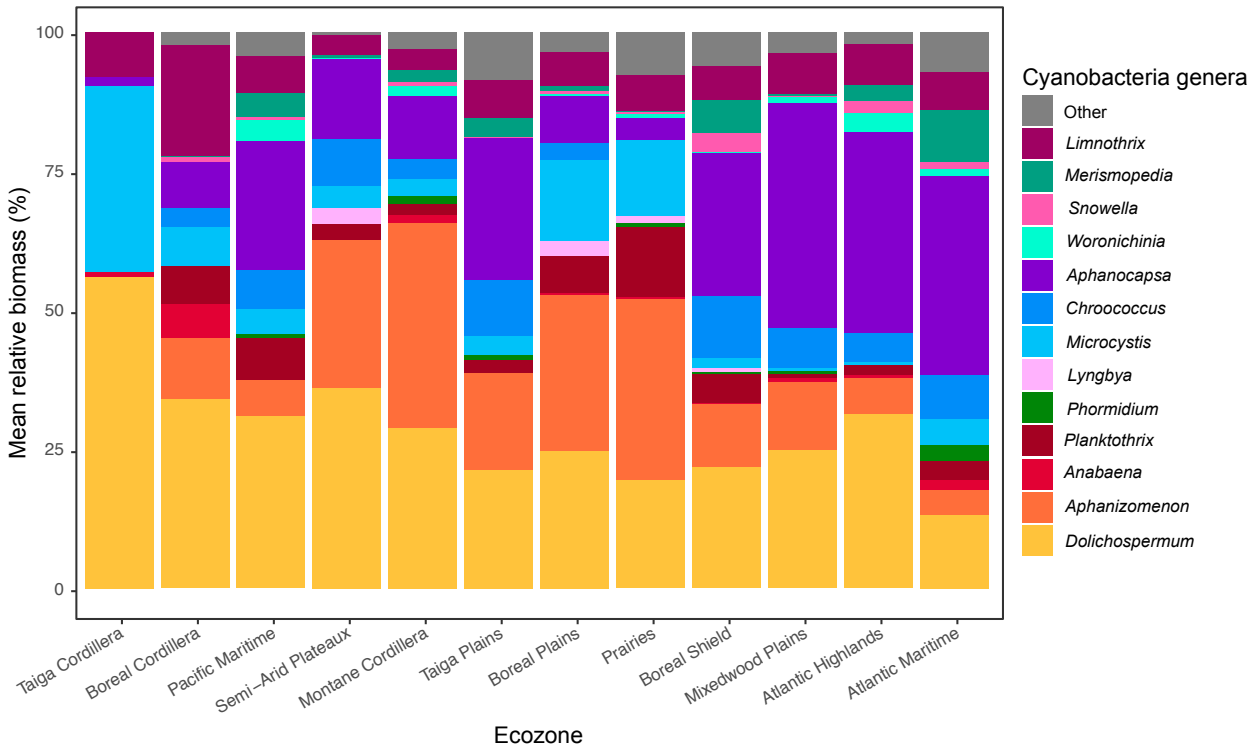

Figure 2. Cyanobacterial characterization of lakes by ecozone (west to east) represented by mean relative biomass of core genera (%). The number of lakes varies by ecozone: Taiga Cordillera (n=3), Boreal Cordillera (n=30), Pacific Maritime (n=67), Semi-Arid Plateaux (n=36), Montane Cordillera (n=69), Taiga Plains (n=25), Boreal Plains (n=71), Prairies (n=68), Boreal Shield (n=89), Mixedwood Plains (n=56), Atlantic Highlands (n=59), Atlantic Maritime (n=67).

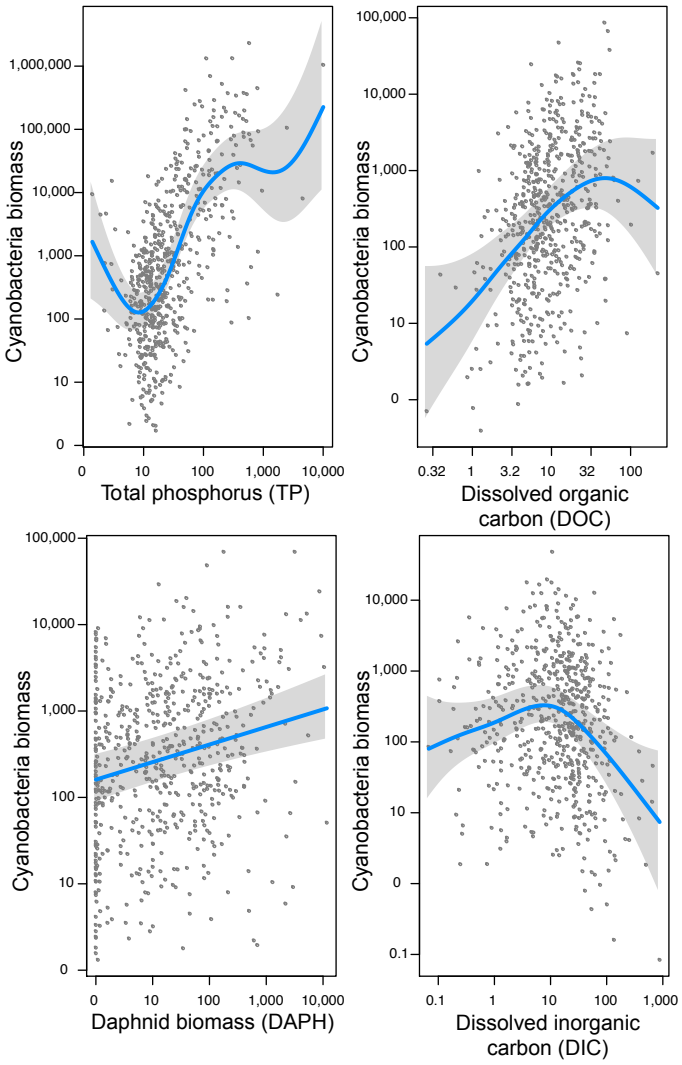

Figure 3. Fitted GAMM values (blue lines) for log cyanobacteria biomass (µg/L) versus select explanatory variables in the final model (full list of predictors shown in Table 1). The 95% confidence intervals are displayed (grey bands) on each plot.

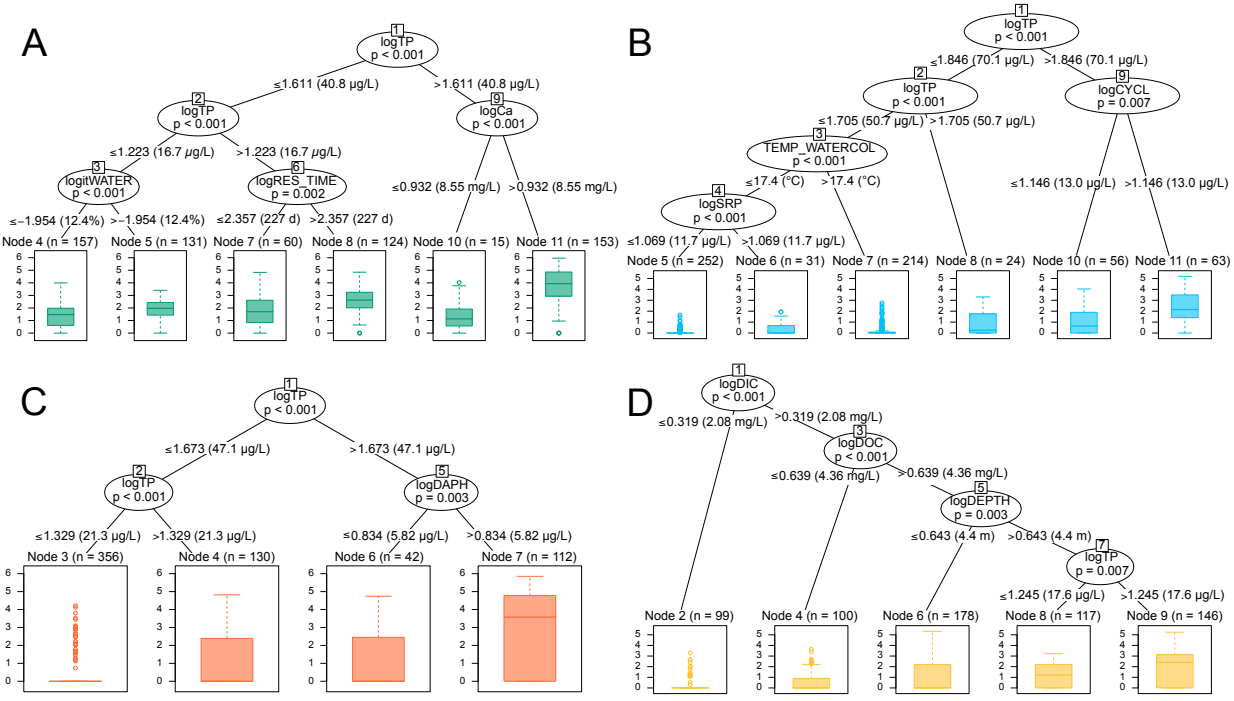

Figure 4. Linear mixed-effect regression trees for total cyanobacteria biomass (µg/L) (A), *Microcystis* biomass (B), *Aphanizomenon* biomass (C) and *Dolichospermum* biomass (D). Ecozone was tested as a random effect in each model and results for this are shown in supplemental Fig S5.

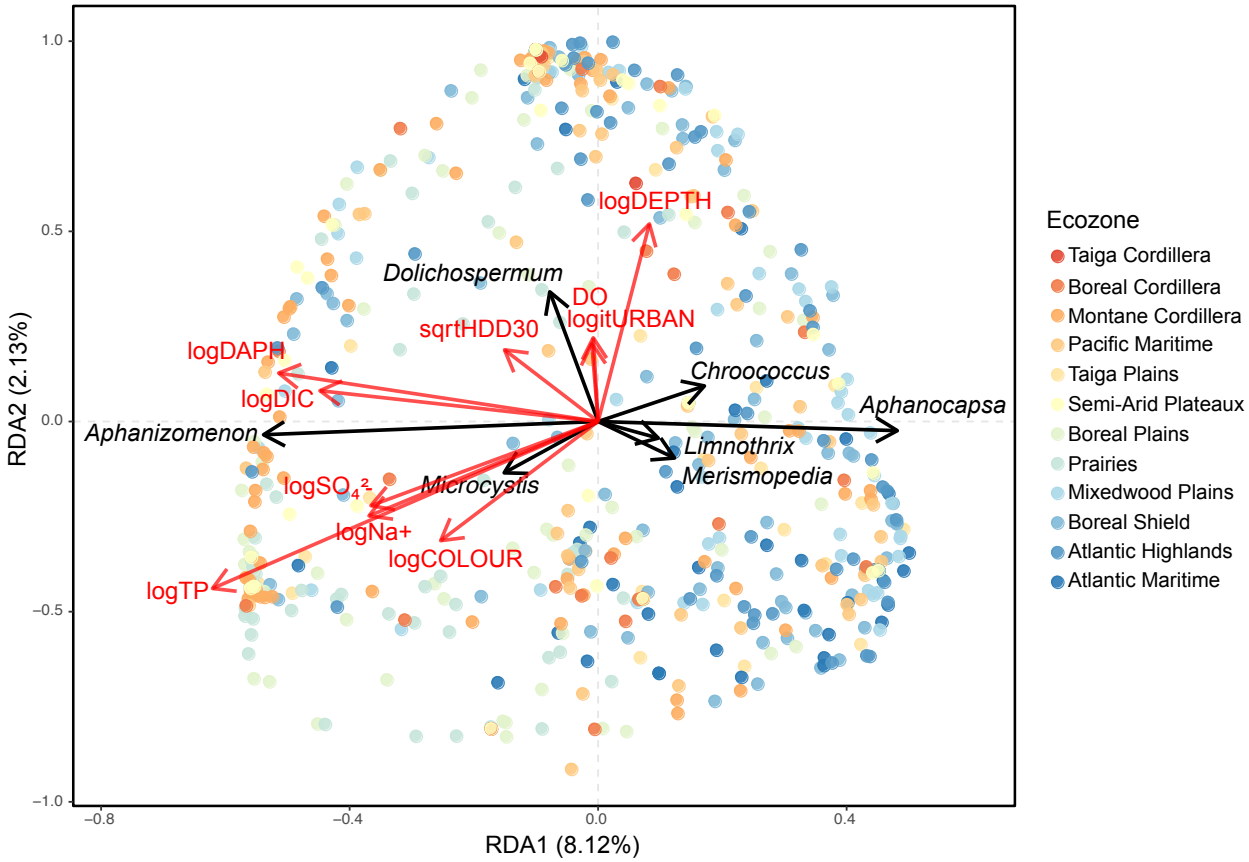

Figure 5. Redundancy analysis of the relationship between genera biomass with environmental variables. Key cyanobacteria genera are represented by black arrows, and environmental variables are shown in red. Sites are colored according to ecozone. Cyanobacteria genera with low loadings in the RDA (axis 1 or 2 scores less than 0.1) are not shown.

Appendices Chapter I

Table S1. List of cyanobacteria predictive modelling publications with large-scale (≥50 lake) datasets. This table expands on similar versions by Beaulieu et al. (2014) and Giani et al. (2020).

Author	Region	No. of sites	Predictors tested	Top predictors
Watson et al. 1997	North temperate	91	Total phosphorus	Total phosphorus
Downing et al. 2001	Worldwide	99	Total phosphorus, total nitrogen, TN:TP ratio	Total phosphorus, total nitrogen
Carvalho et al. 2011	United Kingdom	134	Area, altitude, mean depth, alkalinity, water, colour, retention time, total nitrogen, total phosphorus, chlorophyll	Water colour, alkalinity, retention time, total phosphorus
Dolman et al. 2012	Germany	102	Total phosphorus, total nitrogen	Total phosphorus and total nitrogen
Kosten et al. 2012	Europe and South America	143	Water temperature, total phosphorus, total nitrogen, pH, conductivity, depth, secchi depth, area, latitude	Water temperature, total nitrogen, total phosphorus
Beaulieu et al. 2013	United states	1147	Water temperature (surface and water column), total phosphorus, total nitrogen, TN:TP ratio, pH, conductivity, depth, area, latitude	Total nitrogen, water temperature and total phosphorus
Carvalho et al. 2013	Europe	1506	Total phosphorus, alkalinity	Total phosphorus, alkalinity
Beaulieu et al. 2014	Canada	149	Total phosphorus, total nitrogen, total Kjeldahl nitrogen, pH, TN:TP ratio,	Total phosphorus, total nitrogen

			conductivity, strength of stratification, water temperature, nitrite, nitrate, ammonium	
Doubek et al. 2015	United States	236	Land use (pasture, crop, agriculture, developed), water temperature, depth, surface area, total nitrogen, total phosphorus, TN:TP ratio	Anthropogenic land use (predicts cyanobacterial dominance), total phosphorus and total nitrogen (predicts cyanobacterial biovolume)
Mowe et al. 2015	Worldwide	186	Total phosphorus, total nitrogen, TN:TP ratio, precipitation, water temperature	Total phosphorus, total nitrogen, precipitation
Rigosi et al. 2014	United States	1076	Total phosphorus, total nitrogen, water temperature	Water temperature, total phosphorus
Filstrup et al. 2016	United States	137	Total phosphorus, total nitrogen, TN:TP ratio, light availability	Total phosphorus, total nitrogen, TN:TP ratio, light availability
Ghaffar et al. 2016	United states	116	Total phosphorus, total nitrogen, TN:TP ratio	Total phosphorus, total nitrogen
Chapra et al. 2017	United states	310	Inorganic phosphorus, organic phosphorus, organic nitrogen, Nitrate, Nitrite, ammonium, dissolved oxygen, water temperature	Inorganic phosphorus, inorganic nitrogen, water temperature, hydrological variables
Richardson et al. 2018	Europe	494	Total phosphorus, air temperature, retention times	Total phosphorus, air temperature, retention times (varied by lake type)
Ho et al. 2019	Worldwide	71	Temperature, precipitation, historic fertilizer use	Temperature, precipitation, historic fertilizer use

Ho and Michalek 2020	United States	1260	Seasonal air temperature (spring, summer and annual), water temperature (surface, water column, bottom), seasonal precipitation (spring, summer, annual), total nitrogen, total phosphorus, fertilizer, area, depth, latitude, longitude, day of the year, stratification	Total nitrogen, total phosphorus, spring air temperature, depth
Rose et al. 2019	United States	60	Drainage area, depth, surface area, Secchi depth, chlorophyll-a, total phosphorus, inorganic phosphorus, total nitrogen, TN:TP ratio, developed land (open space, low-high intensity), barren land, deciduous forest, evergreen forest, mixed forest, shrub, grassland, pasture, cultivated crops, woody wetlands, emergent herbaceous wetlands	Chlorophyll-a, inorganic phosphorus, total phosphorus, mixed forest land, developed-open space land, developed- low intensity land, drainage area
Liu et al. 2020	United States	998	Diatom inferred phosphorus, total nitrogen, total phosphorus, TN:TP ratio, conductivity, depth, pH, surface water temperature, average temperature in upper 2 m	Surface temperature, pH, diatom inferred phosphorus, total nitrogen, total phosphorus, TN:TP ratio, conductivity, depth, average temperature (relative importance of each varies by deep vs. shallow lakes and natural vs. man-made)
Mellios et al. 2020	Europe	822	Latitude, elevation, surface area, mean depth, maximum depth, chlorophyll-a, total nitrogen, total phosphorus, TN:TP ratio, mean air temperature, maximum air temperature	Chlorophyll-a, total nitrogen. Mean depth and TN:TP ratio (when the data is divided int shallow and deep lakes)

Watson, S. B., E. McCauley, and J. A. Downing. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol. Oceanogr. **42**: 487–495. doi:10.4319/lo.1997.42.3.0487

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/cjfas-58-10-1905

Carvalho, L., C. A. Miller, E. M. Scott, G. A. Codd, P. S. Davies, and A. N. Tyler. 2011. Cyanobacterial blooms: Statistical models describing risk factors for national-scale lake assessment and lake management. Sci. Total Environ. **409**: 5353–5358.

Dolman, A. M., J. Rücker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. PLoS One **7**. doi:10.1371/journal.pone.0038757

Kosten, S., V. L. M. Huszar, E. Bécares, and others. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Glob. Chang. Biol. **18**: 118–126. doi:10.1111/j.1365-2486.2011.02488.x

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. **58**: 1736–1746. doi:10.4319/lo.2013.58.5.1736

Carvalho, L., C. Mcdonald, C. de Hoyos, and others. 2013. Sustaining recreational quality of European lakes: Minimizing the health risks from algal blooms through phosphorus control. J. Appl. Ecol. **50**: 315–323. doi:10.1111/1365-2664.12059

Beaulieu, M., F. Pick, M. Palmer, S. Watson, J. Winter, R. Zurawell, and I. Gregory-Eaves. 2014. Comparing predictive cyanobacterial models from temperate regions. Can. J. Fish. Aquat. Sci. **71**: 1830–1839. doi:10.1139/cjfas-2014-0168

Doubek, J. P., C. C. Carey, and B. J. Cardinale. 2015. Anthropogenic land use is associated with N-fixing cyanobacterial dominance in lakes across the continental United States. Aquat. Sci. **77**: 681–694. doi:10.1007/s00027-015-0411-x

Mowe, M. A. D., S. M. Mitrovic, R. P. Lim, A. Furey, and D. C. J. Yeo. 2015. Tropical cyanobacterial blooms: A review of prevalence, problem taxa, toxins and influencing environmental factors. J. Limnol. **74**: 205–224. doi:10.4081/jlimnol.2014.1005

Rigosi, A., C. C. Carey, B. W. Ibelings, and J. D. Brookes. 2014. The interaction between climate warming and eutrophication to promote cyanobacteria is dependent on trophic state and varies among taxa. Limnol. Oceanogr. **59**: 99–114. doi:10.4319/lo.2014.59.1.0099

Filstrup, C. T., A. J. Heathcote, D. L. Kendall, and J. A. Downing. 2016. Phytoplankton taxonomic compositional shifts across nutrient and light gradients in temperate lakes. Inl. Waters **6**: 234–249. doi:10.5268/IW-6.2.939

Ghaffar, S., R. J. Stevenson, and Z. Khan. 2016. Cyanobacteria Dominance in Lakes and Evaluation of Its Predictors: A study of southern appalachians Ecoregion, USA. MATEC Web Conf. **60**: 1–5. doi:10.1051/matecconf/20166002001

Chapra, S. C., B. Boehlert, C. Fant, and others. 2017. Climate Change Impacts on Harmful Algal Blooms in U.S. Freshwaters: A Screening-Level Assessment. Environ. Sci. Technol. **51**: 8933–8943. doi:10.1021/acs.est.7b01498

Richardson, J., C. Miller, S. C. Maberly, and others. 2018. Effects of multiple stressors on cyanobacteria abundance vary with lake type. Glob. Chang. Biol. **24**: 5044–5055. doi:10.1111/gcb.14396

Ho, J. C., A. M. Michalak, and N. Pahlevan. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s.

Nature **574**: 667–670. doi:10.1038/s41586-019-1648-7

Ho, J. C., and A. M. Michalak. 2020. Exploring temperature and precipitation impacts on harmful algal blooms across continental U.S. lakes. Limnol. Oceanogr. **65**: 992–1009. doi:10.1002/lno.11365

Rose, V. J., W. M. Forney, R. A. Norton, and J. A. Harrison. 2019. Catchment characteristics, water quality, and cyanobacterial blooms in Washington and Oregon Lakes. Lake Reserv. Manag. **35**: 51–63. doi:10.1080/10402381.2018.1518940

Liu, B., S. Chen, H. Liu, and Y. Guan. 2020. Modeling cyanobacteria biomass by surface sediment diatoms in lakes: problems and suggestions. Ecol. Modell. **430**. doi:10.1016/j.ecolmodel.2020.109056

Mellios, N., S. J. Moe, and C. Laspidou. 2020. Machine learning approaches for predicting health risk of cyanobacterial blooms in Northern European Lakes. Water **12**. doi:10.3390/W12041191

Table S2. List of environmental variables and descriptive summary of the predictors from 640 lakes used in the current study. Whether they were selected for after assessing collinearity using VIFs is indicated by VIF selected (VIF <10). RF selected refers to which variables were selected for each response variable by the random forests (Tot=total cyanobacteria biomass, M=*Microcystis* biomass, A=*Aphanizomenon* biomass, D=*Dolichospermum* biomass). Variables are organized into five broader categories: physiography, water quality, land-use, zooplankton and climate.

Variables	Units	Abbreviation	Minimum	Maximum	Median	% NA	VIF Selected	RF Selected
Physiography								
Lake area	km^2	AREA	0.0009	99.03	0.6922	0	No	
Lake depth	m	DEPTH	0.25	151	7.7	0	Yes	Tot, A, D
Circularity	-	CIRCULARITY	0.0079	0.9269	0.3092	0	Yes	M
Discharge	m^3/sec	DISCHARGE	0	9218	0.133	4.4	Yes	
Watershed area	km^2	WATERSHED	0.2080	37,459	14.51	0	Yes	A
Altitude	m	ALTITUDE	2	1555	341	0	Yes	Tot, M, A, D
Residence time	days	RES_TIME	0.1	270,759	402.5	6.1	Yes	Tot, M
Shoreline length	m	SHORE_LENGTH	155	218,835	5177	0	No	
Slope 100m	°	SLOPE100	0.09	25.1	2.955	4.4	Yes	D
Human population	$People/km^2$	POPULATION	0	335,522	31.5	0	Yes	A
Water quality								
Conductivity	mS/cm	CONDUCTIVITY	0.0062	40.77	0.1682	10.9	No	
Specific conductivity	uS/cm	SPEC_CONDUCTIVITY	7.369	37,392	185.8	10.9	No	
Temperature-euphotic	°C	TEMP_EUPHOTIC	11.80	29.73	20.32	6.7	Yes	
Temperature- water column	°C	TEMP_WATERCOL	6.1	30	17.4	6.7	Yes	M
Average Brunt-Väisälä	s^{-1}	BRUNT_VAISALA	0	0.0075	0.0008	11.1	Yes	

Dissolved oxygen	mg/L	DO	1.85	15.34	8.936	19.2	Yes	M, A, D
Colour	mg/L Pt	COLOUR	0	369.5	20.59	3.3	Yes	Tot, A
Dissolved organic carbon	mg/L	DOC	0.12	220.2	8.498	3.0	Yes	Tot, M, A, D
Dissolved inorganic carbon	mg/L	DIC	0.0675	879	16.13	1.6	Yes	Tot, A, D
Total phosphorus	µg/L	TP	1.39	10,053	18.87	1.4	Yes	Tot, M, A, D
Total nitrogen	mg/L	TN	0.025	4.358	0.248	8.3	Yes	Tot, M, A, D
Soluble reactive phosphorus	µg/L	SRP	0.5	4,066	4.961	0.31	Yes	Tot, M, A
Total nitrogen: Total phosphorus ratio	-	TNTP	0.5446	114.9	10.92	9.5	No	
Calcium	mg/L	CA	0.005	535.5	19.17	0.16	Yes	Tot, A, D
Potassium	mg/L	K	0.005	279.2	0.9871	0.16	Yes	Tot, M, A, D
Sodium	mg/L	NA	0.01	14,804	5.622	0.16	Yes	Tot, A, D
Chloride	mg/L	CL	0.015	12,358	5.59	1.3	Yes	Tot, M
Magnesium	mg/L	MG	0.005	2,249	4.918	0.16	No	
Sulfate	mg/L	SO4	0.02	16,971	4.75	1.1	Yes	Tot, M, D
Land use								
Agriculture	%	AGRI	0	0.8623	0	0	Yes	Tot, M, A
Forestry	%	FORESTRY	0	0.5086	0.0025	0	Yes	A
Mines	%	MINES	0	0.2322	0	0	Yes	
Natural landscape	%	NATLAND	0.0213	0.9839	0.7162	0	No	
Pasture	%	PASTURE	0	0.4771	0	0	Yes	Tot, M
Urban	%	URBAN	0	0.9264	0.0196	0	Yes	
Water	%	WATER	0.0007	0.6675	0.1029	0	Yes	Tot, M
Zooplankton								
Total zooplankton	µg/L	ZOOP	0.3133	13,848	75.73	5.6	Yes	Tot, M, A, D

Cyclopoid	µg/L	CYCL	0	2,643	7.638	5.6	Yes	Tot, M, A, D
Calanoid	µg/L	CALA	0	13,403	15.82	5.6	Yes	D
Cladoceran	µg/L	CLAD	0	11,665	25.88	5.6	No	
*Daphnid	µg/L	DAPH	0	11,663	12.61	5.6	Yes	Tot, M, A, D
Chydorid	µg/L	CHYD	0	214.03	0	5.6	No	
*Bosminid	µg/L	BOSM	0	2,561	0.2787	5.6	No	
Small cladocerans	µg/L	SMCLAD	0	2,571	2.339	5.6	Yes	M
Climate								
7-day average air temperature	°C	AIRTEMP7	6.5959	26.25	17.42	0	Yes	D
30-day average air temperature	°C	AIRTEMP30	10.09	23.06	17.43	0	No	
7-day total precipitation	m	PCPN7	0.0002	1.141	0.1886	0	Yes	D
30-day total precipitation	m	PCPN30	0.0340	2.933	0.9941	0	Yes	D
7-day net solar radiation	J/m^2	SOLAR7	2,615,096	8,046,217	5,321,336	0	Yes	
30-day net solar radiation	J/m^2	SOLAR30	4,052,577	7,724,438	5,648,340	0	Yes	
7-day average wind speed	m/s	WIND7	0.5323	5.757	2.183	0	No	
30-day average wind speed	m/s	WIND30	0.6651	5.5684	2.264	0	Yes	M, A, D
7-day total heat degree days	days	HDD7	0	79.19	8.549	0	No	
30-day total heat degree days	days	HDD30	0	239.1	39.82	0	Yes	

Water quality variables expect Brunt-Väisälä were integrated from the euphotic, defined as twice the Secchi depth, up to a maximum of 2 m below the surface. *Daphnids include: *Daphnia*, *Ceriodaphnia*, *Simocephalus*, *Scapholeberis*. *Bosminids include: *Bosmina*, *Eubosmina*.

Table S3. Biomass (µg/L) summary statistics for each cyanobacteria genus detected across all study lakes (n=640), and the number of lakes they were detected in. Value for total cyanobacteria biomass is also reported.

Taxon	Maximum (µg/L)	Mean (µg/L)	Median (µg/L)	Lakes (n)
Anabaena	5,447	19.09	0	27
Anabaenopsis	42,400	75.38	0	2
Aphanizomenon	707,768	12,621	0	175
Aphanocapsa + *Synechocystis*	6,295	75.62	4.277	402
Aphanothece	117.6	0.75	0	24
Arthrospira	160.4	0.2506	0	1
Chroococcus	8,468	26.95	0	157
Chrysosporum	1,467	2.292	0	1
Coelosphaerium	72.17	0.1868	0	3
Cuspidothrix	2,998	8.136	0	6
Dolichospermum	212,350	2,231	0	303
Geitlerinema	8,115	14.01	0	17
Gloeotrichia	1,069	2.993	0	3
Gomphosphaeria	461.9	2.005	0	10
Limnothrix	14,750	96.84	0	218
Lyngbya	144,343	273.8	0	6
Merismopedia	164.2	3.27	0	119
Microcystis	155,776	1,053	0	220
Nodularia	30,282	47.51	0	3
Oscillatoria	779.5	1.957	0	4
Phormidium	157.17	0.626	0	18
Planktolyngbya	2,900	15.96	0	24
Planktothrix	258,085	984	0	79
Pseudanabaena	10,852	21.49	0	70
Rhabdoderma	164.6	0.6376	0	14
Rhabdogloea	1.804	0.003	0	2
Romeria	588.1	1.541	0	5
Snowella	134.7	1.638	0	38
Spirulina	240.6	1.068	0	4
Synechococcus	8.554	0.0297	0	21
Woronichinia	3,897	20.64	0	84
Total cyanobacteria	**920,516**	**17,604**	**158.2**	**602**

Aphanocapsa and *Synechocystis* were counted together. The minimum for each genus was 0 µg/L.

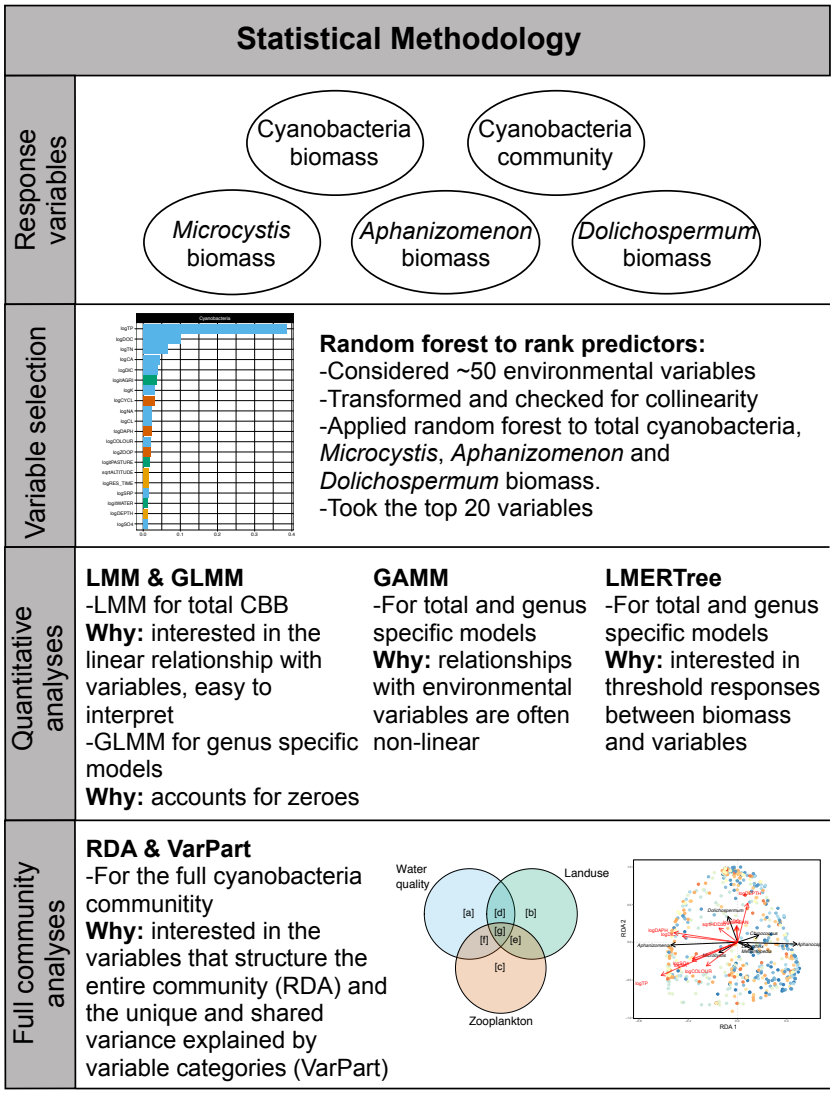

Figure S1. Statistical workflow schematic. Each analysis used in this study is listed along with its purpose.

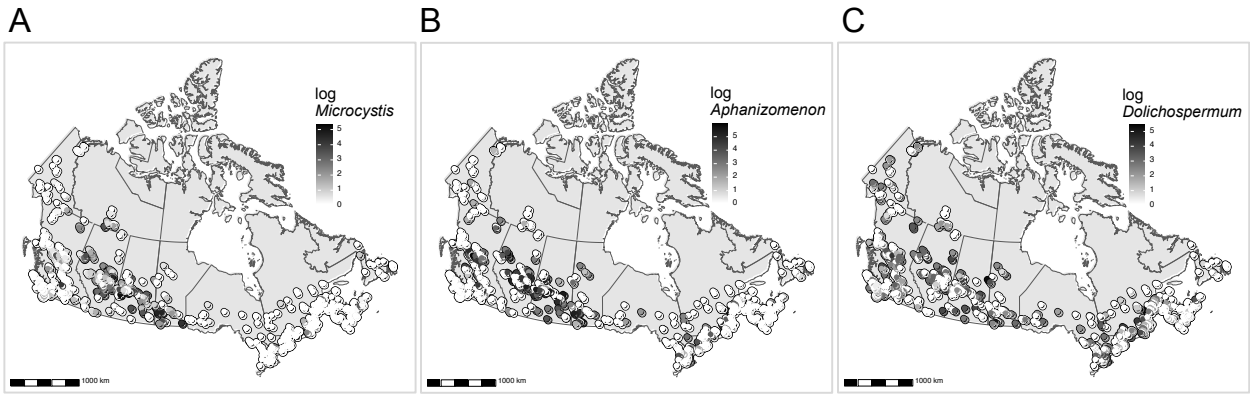

Figure S2. Cyanobacteria biomass (µg/L) or selected genera of interest *Microcystis* (A), *Aphanizomenon* (B) and *Dolichospermum* (C).

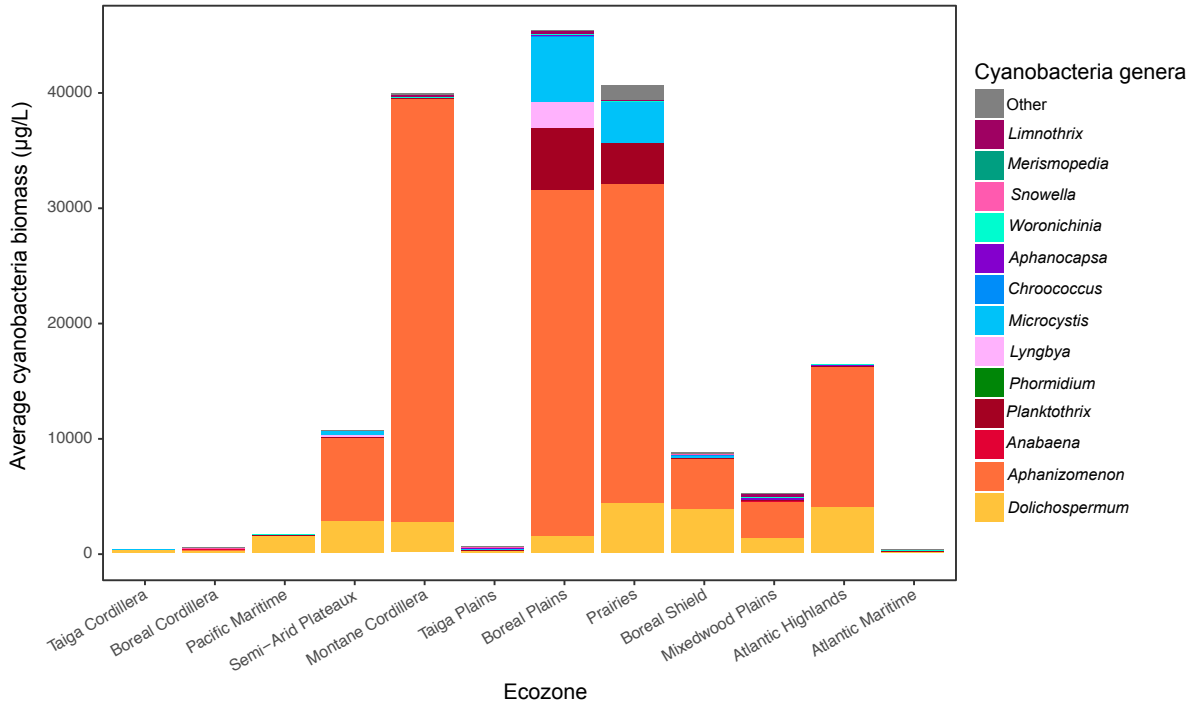

Figure S3. Mean total biomass (µg/L) of cyanobacteria genera by ecozone.

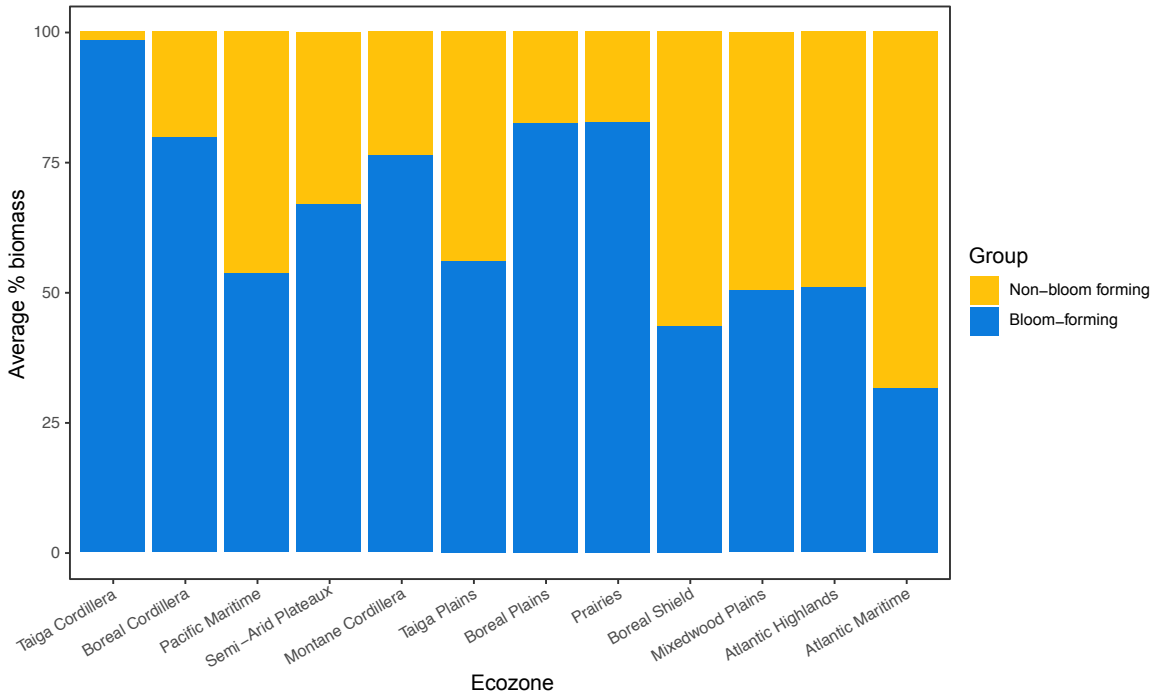

Figure S4. Bar plot of average % bloom-forming biomass by ecozone. Bloom forming taxa included *Chrysosporum, Dolichospermum, Anabaena, Anabaenopsis, Aphanizomenon, Cuspidothrix, Arthrospira, Gloeotrichia, Gomphosphaeria, Limnothrix, Oscillatoria, Microcystis, Nodularia, Planktothrix, Spirulina,* and *Woronichinia*.

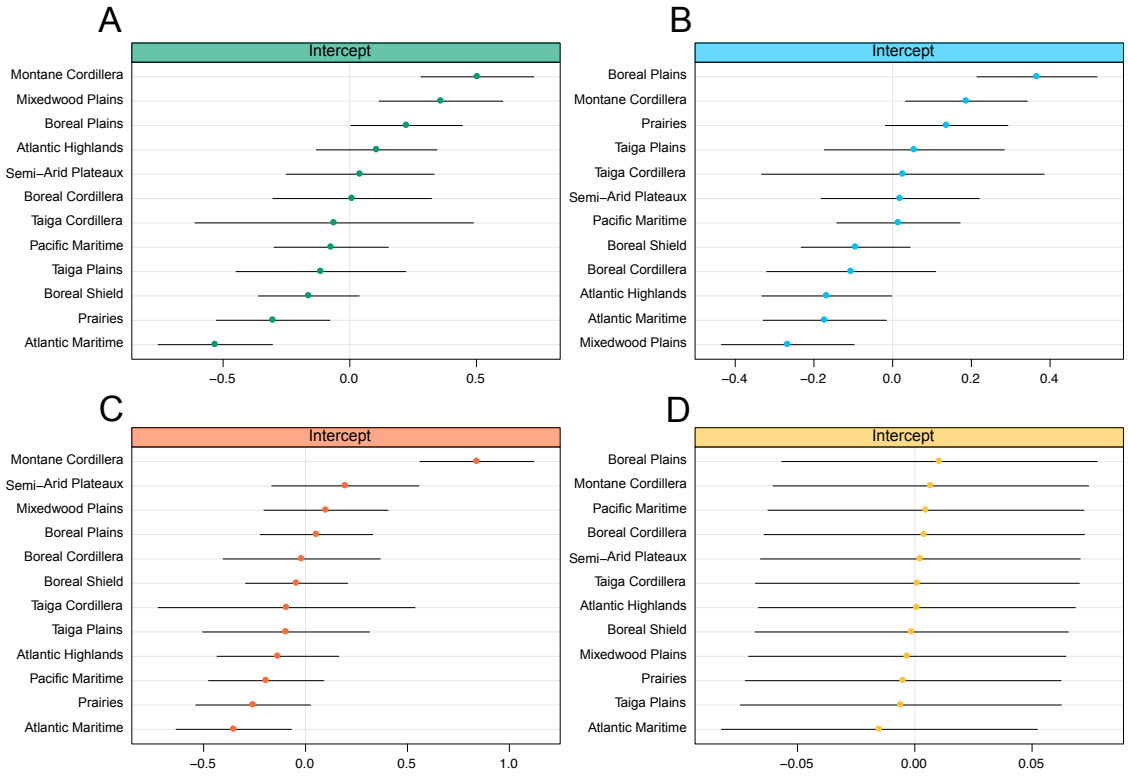

Figure S5. Random effect plots (ecozone) from the linear mixed-effect regression trees analysis for total cyanobacteria biomass (A), *Microcystis* (B), *Aphanizomenon* (C) and *Dolichospermum* (D). Random effect intercepts (mean cyanobacteria biomass) are indicated for each ecozone.

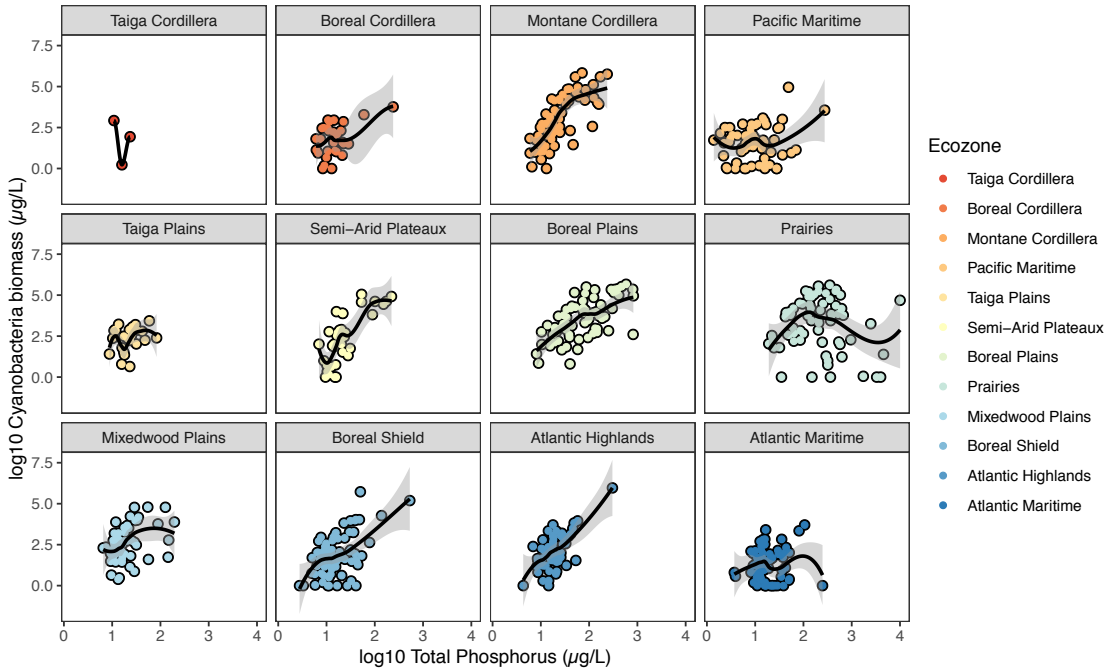

Figure S6. Relationship between cyanobacteria biomass and total phosphorus faceted by ecozone. Trends are depicted using a loess curve.

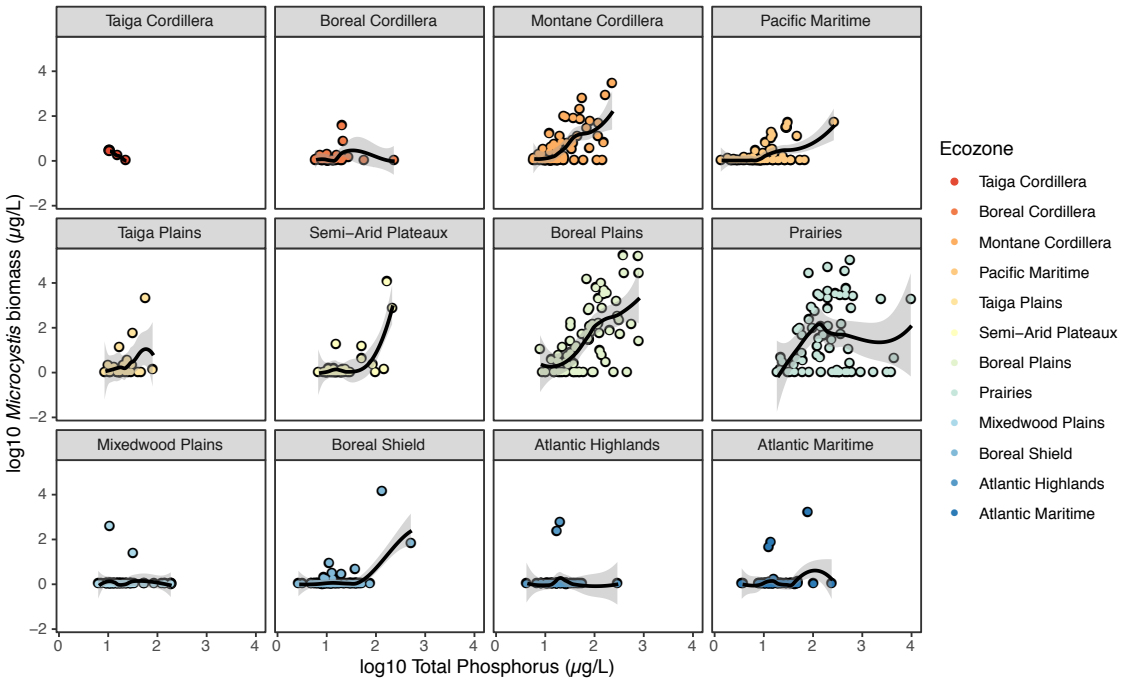

Figure S7. Relationship between Microcystis biomass and total phosphorus faceted by ecozone. Trends are depicted using a loess curve.

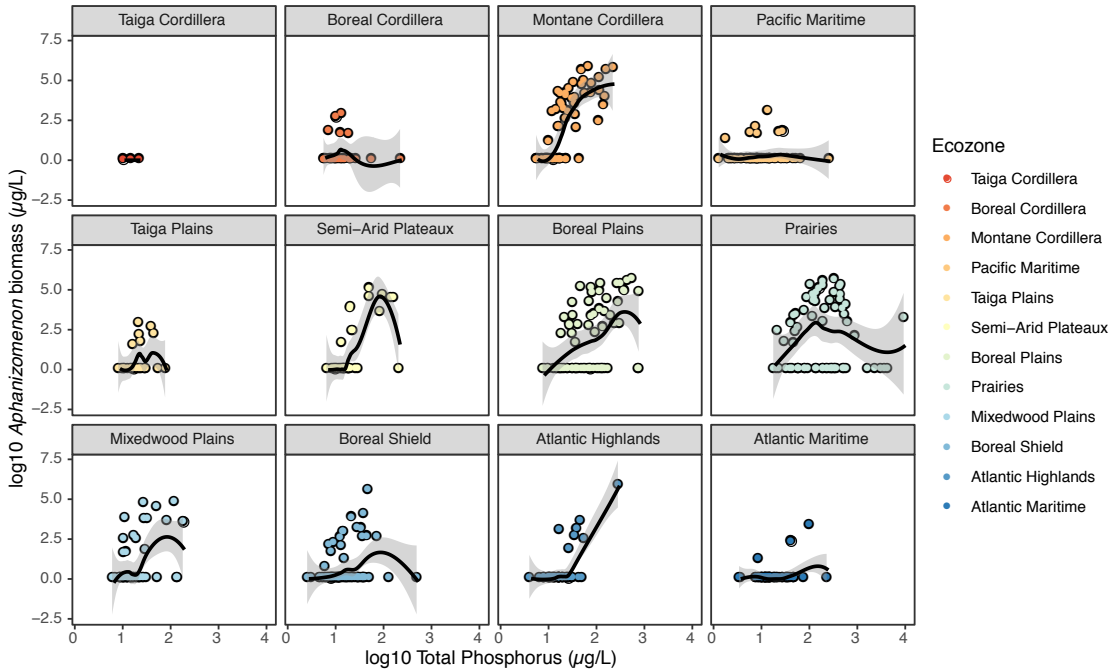

Figure S8. Relationship between Aphanizomenon biomass and total phosphorus faceted by ecozone. Trends are depicted using a loess curve.

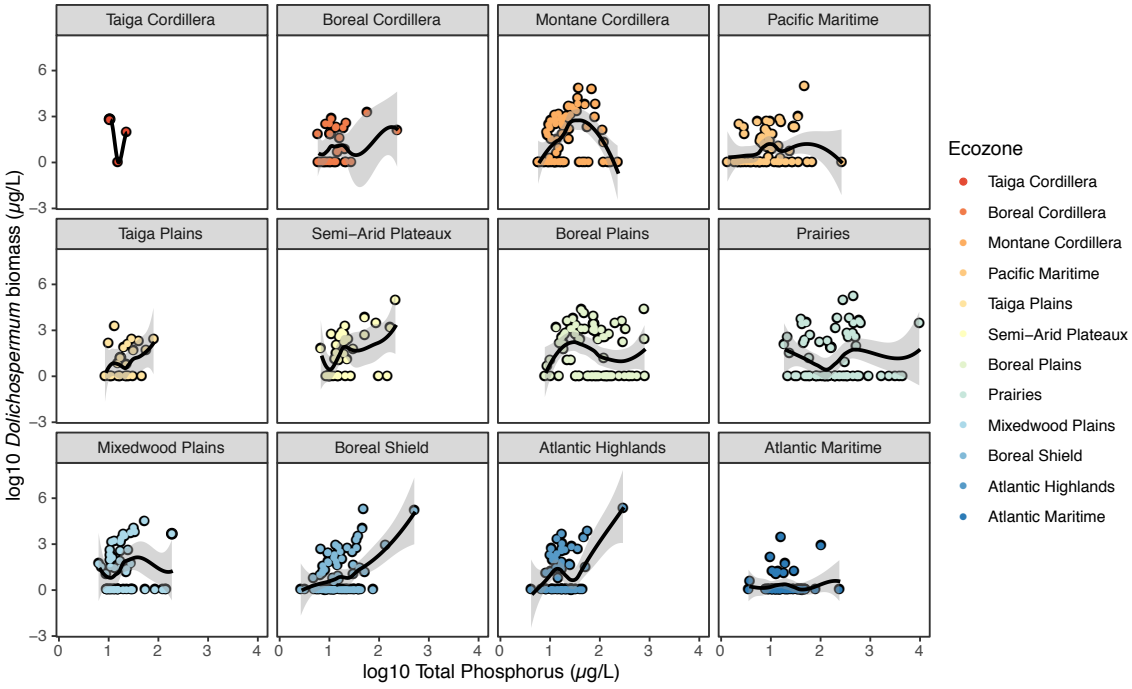

Figure S9. Relationship between Dolichospermum biomass and total phosphorus faceted by ecozone. Trends are depicted using a loess curve.

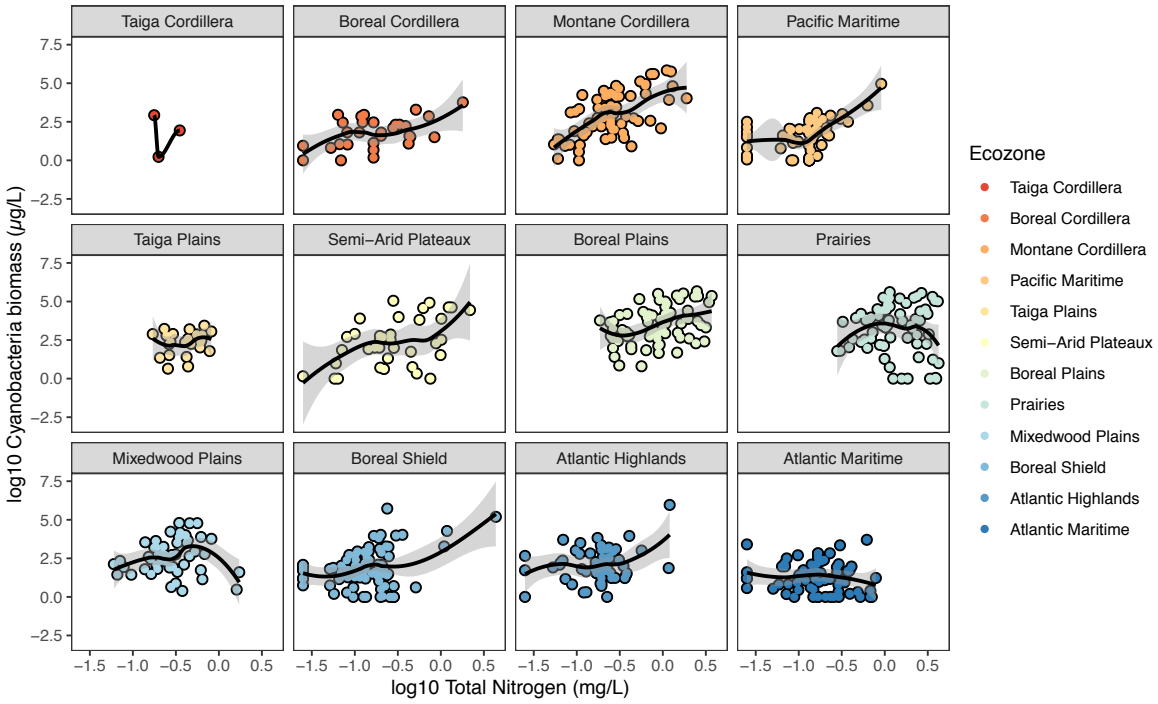

Figure S10. Relationship between cyanobacteria biomass and total nitrogen faceted by ecozone. Trends are depicted using a loess curve.

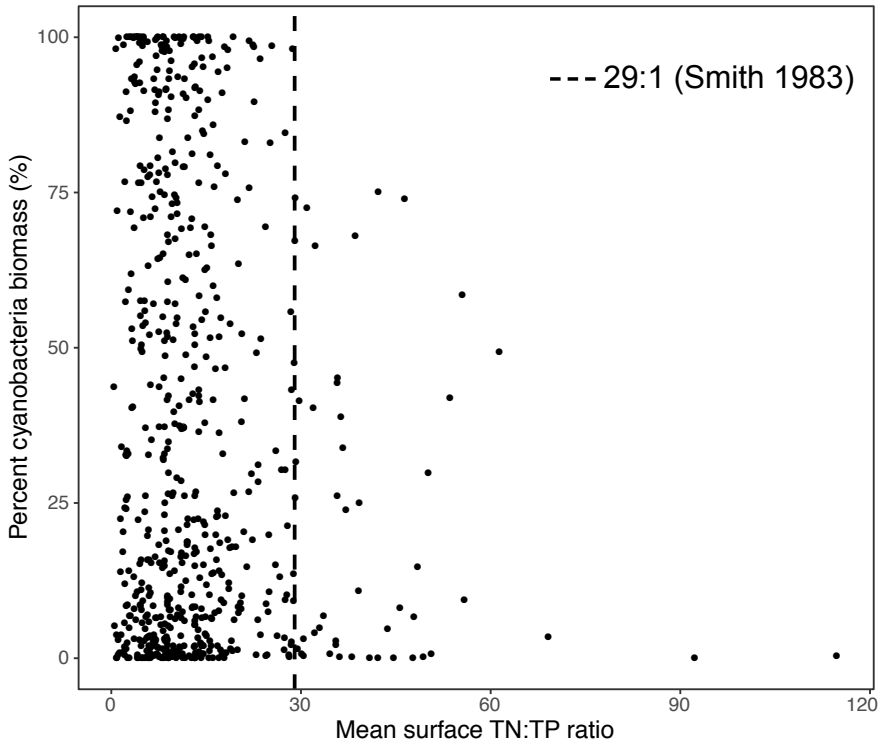

Figure S11. Percent cyanobacteria biomass by the ratio of TN:TP (µg/L). The dashed line represents the 29:1 ratio, identified by Smith (1983) as the threshold at which cyanobacteria are favored below.

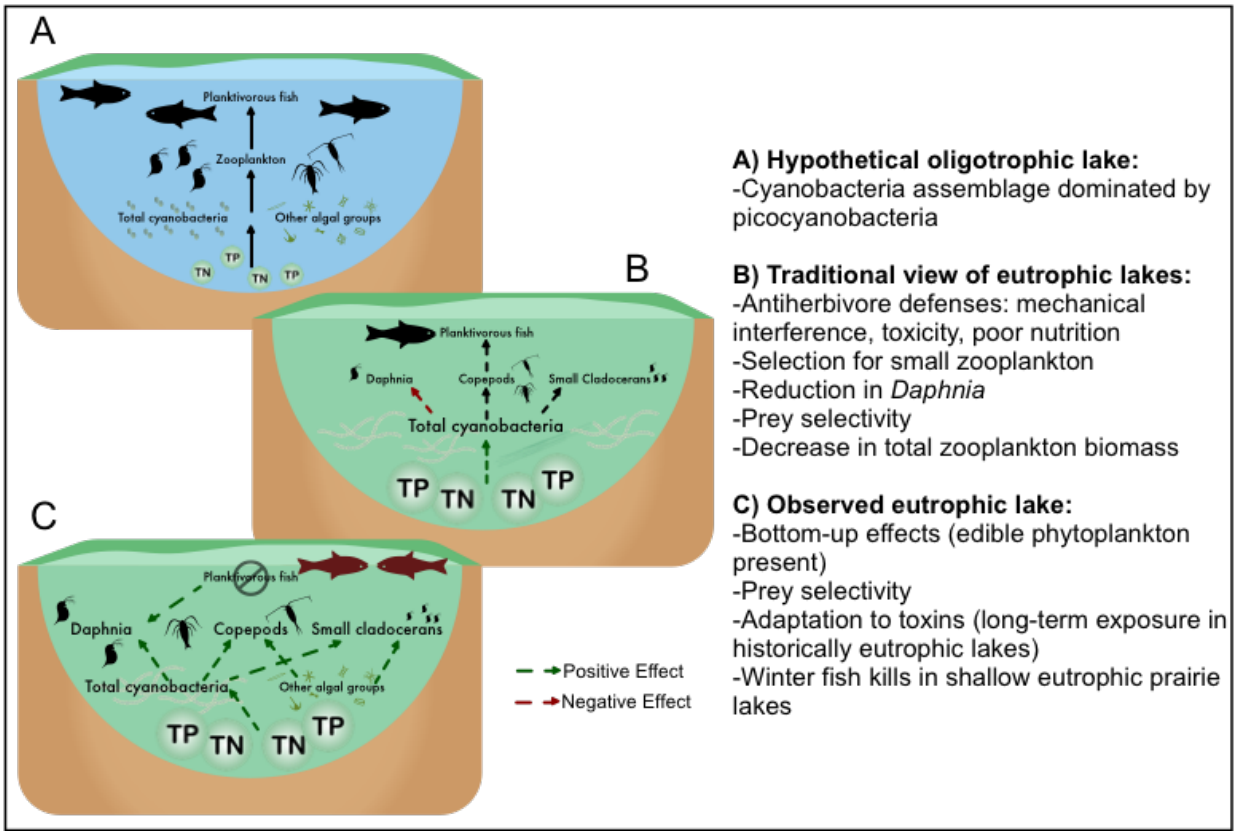

Figure S12. Depicted relationships between total cyanobacteria and zooplankton groups. (A) Hypothesized relationships between zooplankton and cyanobacteria in oligotrophic to mesotrophic lakes. (B) Traditional relationships between cyanobacteria and zooplankton in eutrophic to hypereutrophic lakes. Here, cyanobacteria proliferate with increased nutrients, limiting the amount of energy transfer to higher trophic levels due to inedibility. Smaller zooplankton are selected, therefore total zooplankton biomass decreases. (C) Observed relationships between cyanobacteria and zooplankton groups in our lakes. Higher cyanobacteria biomass correlated with greater biomasses of all zooplankton groups.

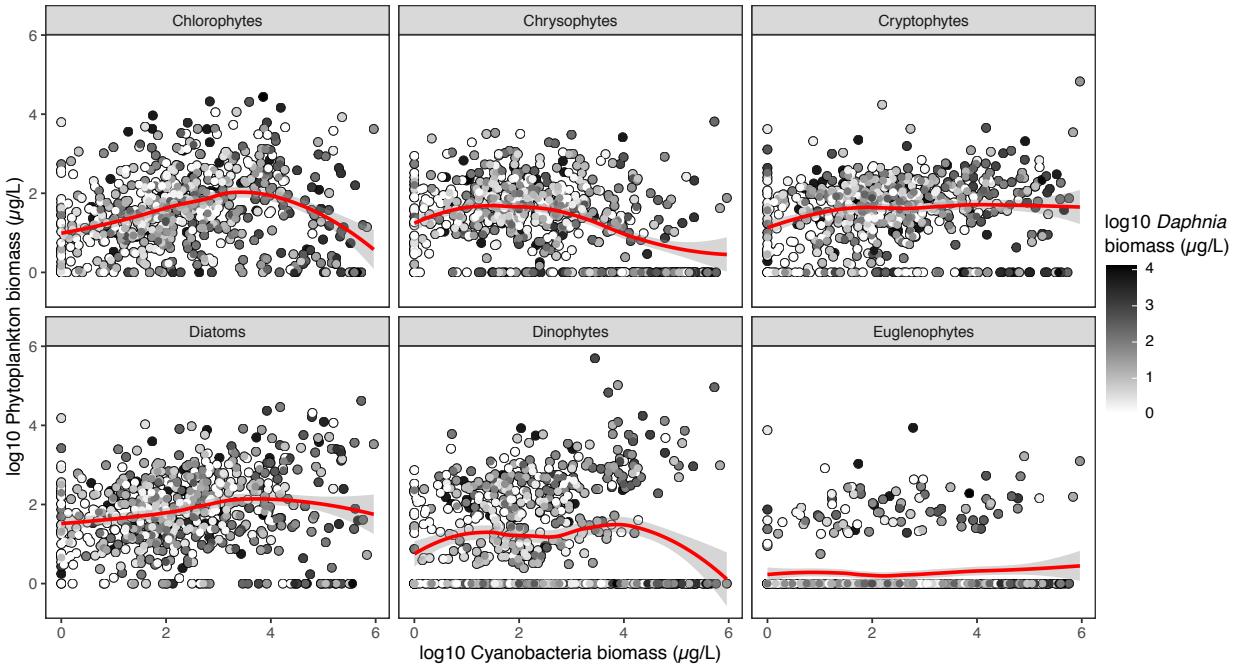

Figure S13. Relationships between the biomass of other phytoplankton groups (µg/L) and cyanobacteria biomass. The red line represents the LOESS trend, and the points are colored by *Daphnia* biomass (µg/L).

Connecting statement between Chapters I and II

In the first chapter of my thesis, we quantified the biomass and community composition of cyanobacteria across a large set of Canadian lakes. We found significantly higher biomass within the Prairies and Boreal Plains ecozones, where *Microcystis*, *Aphanizomenon* and *Dolichospermum* were the dominant genera. However, *Aphanocapsa*, a group of picocyanobacteria, was the most prevalent taxon across all lakes. While considering a broad suite of predictor variables that have been documented to have a hypothesized relationship with cyanobacteria, we developed empirical models for total cyanobacteria biomass and that of three key genera of interest. Our results demonstrated the importance of total phosphorus as the leading predictor of the biomass of total cyanobacteria, as well as of several key genera and overall cyanobacterial community composition. We observed different phosphorus thresholds for individual genera considered. The empirical models developed in chapter I also highlighted the predictive power and positive correlation between cyanobacteria and the biomass of zooplankton variables (i.e., *Daphnia* and copepod biomass). Lastly, we observed a limited effect of ecozone across models, suggesting the fixed effects alone account for most of the regional variation within models.

Up to the present, including in chapter I, monitoring programs have mostly relied on taxonomists to identify phytoplankton assemblages using morphological characteristics by light microscopy. However, this traditional method is often considered time-consuming, is unable to distinguish cryptic taxa and requires a significant amount of expertise by the taxonomist, of whom also have individual biases. Consistently and accurately characterizing the phytoplankton assemblage are foundational to many aquatic bioassessment programs, particularly for the

cyanobacteria, as several bloom-forming and toxin-producing taxa may pose a number of ecosystem and human health risks. In recent years, the advancement and cost-effectiveness of next-generation sequencing technologies has allowed for this method to become more efficient and may alleviate some of biases associated with light microscopy. With the growing use of molecular methods, and some continued use of light microscopy, the comparability of methods has been brought into question. Researchers and lake managers are faced with the decision of which method employ if budgets are limited. To address these issues, the second chapter of my thesis assessed the congruency between cyanobacteria communities generated by light microscopy and DNA metabarcoding. Using a subset of 379 lakes from chapter I, I developed a statistical approach to quantify the comparability of traditional, morphologically-based data as well as metabarcoding data were. Lastly, this chapter reviewed several of the technical and biological factors that may have resulted in community differences between methods.

CHAPTER II

Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes

Paul W. MacKeigan[1,2], Rebecca E. Garner[2,3], Marie-Ève Monchamp[1,2], David A. Walsh[2,3], Vera E. Onana[2,3], Susanne A. Kraemer[2,3], Frances R. Pick[4], Beatrix E. Beisner[2,5], Michael D. Agbeti[6], Naíla Barbosa da Costa[2,7], B. Jesse Shapiro[8], and Irene Gregory-Eaves[1,2]

[1]Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada;

[2]Interuniversity Research Group in Limnology (GRIL), Quebec, Canada;

[3]Department of Biology, Concordia University, Montreal, Quebec, H4B 1R6, Canada;

[4]Department of Biology, University of Ottawa, Ottawa, Ontario, K1N 9A7, Canada;

[5]Department of Biological Sciences, University of Quebec at Montreal, Montreal, Quebec, H2X 1Y4, Canada;

[6]Bio-Limno Research & Consulting, Halifax, Nova Scotia, B3N 3J2, Canada;

[7]Department of Biological Sciences, University of Montreal, Montreal, Quebec, H2V 0B3 Canada;

[8]Department of Microbiology and Immunology, McGill University, Montreal, H3A 2B4 Canada

Manuscript published in *Harmful Algae* in February 2022

Abstract

Accurately identifying the species present in an ecosystem is vital to lake managers and successful bioassessment programs. This is particularly important when monitoring cyanobacteria, as numerous taxa produce toxins and can have major negative impacts on aquatic ecosystems. Increasingly, DNA-based techniques such as metabarcoding are being used for measuring aquatic biodiversity, as they could accelerate processing time, decrease costs and reduce some of the biases associated with traditional light microscopy. Despite the continuing use of traditional microscopy and the growing use of DNA metabarcoding to identify cyanobacteria assemblages, methodological comparisons between the two approaches have rarely been reported from a wide suite of lake types. Here, we compare planktonic cyanobacteria assemblages generated by inverted light microscopy and DNA metabarcoding from a 379-lake dataset spanning a longitudinal and trophic gradient. We found moderate levels of congruence between methods at the broadest taxonomic levels (i.e., Order, RV=0.40, $p < 0.0001$). This comparison revealed distinct cyanobacteria communities from lakes of different trophic states, with *Microcystis, Aphanizomenon* and *Dolichospermum* dominating with both methods in eutrophic and hypereutrophic sites. This finding supports the use of either method when monitoring eutrophication in lake surface waters. The biggest difference between the two methods was the detection of picocyanobacteria, which are typically underestimated by light microscopy. This reveals that the communities generated by each method currently are complementary as opposed to identical and promotes a combined-method strategy when monitoring a range of trophic systems. For example, microscopy can provide measures of cyanobacteria biomass, which are critical data in managing lakes. Going forward, we believe that molecular genetic methods will be increasingly adopted as reference databases are routinely

updated with more representative sequences and will improve as cyanobacteria taxonomy is resolved with the increase in available genetic information.

Introduction

Photosynthetic cyanobacteria are ubiquitous in inland waters and may form at least temporary blooms in lakes of all trophic states (Paerl and Paul 2012). The reported incidence of cyanobacteria has risen considerably in recent decades (Winter et al. 2011; Ho et al. 2019) and they have become relatively dominant over other widespread phytoplankton groups in many north temperate lakes in the last century (Taranu et al. 2015). Their proliferation can have major negative impacts on aquatic ecosystems. Toxins produced by cyanobacteria have adverse health effects on aquatic animals, domestic pets and humans (Azevedo et al. 2002; Miller et al. 2010; Pick 2016). Moreover, the decomposition of blooms may deplete oxygen, leading to bottom water anoxia that can cause fish kills and promote further declines in water quality through internal loading (Paerl and Otten 2013; Huisman et al. 2018). Owing to the plethora of health and economic consequences, key priorities of lake managers are to monitor cyanobacteria, to identify their environmental drivers, and to mitigate their harmful effects (Downing et al. 2001; Paerl et al. 2011; O'Neil et al. 2012). These priorities rest on the ability to correctly determine cyanobacteria community structure and identify potentially troublesome taxa.

Correctly identifying and quantifying the species present in an ecosystem is essential to bioassessment programs (McElroy et al. 2020). Conventional assessments are performed by trained taxonomists who identify taxa via morphological and structural traits (Reynolds 2006; Gao et al. 2018; Vuorio et al. 2020) which include the organization, shape and size of cells as well as specialized cells (Castenholz 2015; Li et al. 2019a). Light microscopy is still widely used

today to identify and count algal cells (Utermöhl 1958; Karlson et al. 2010; Li et al. 2019b). Whilst routine, it is considered a time-consuming process that requires expertise honed through experience and interpretation of taxonomic literature, which can vary amongst analysts leading to observer bias (Gao et al. 2018; Rimet et al. 2018; Bailet et al. 2020). The availability of taxonomic expertise is also of concern, with the supply of newly trained taxonomists struggling to meet the demand (Esenkulova et al. 2020; Santi et al. 2021). Despite the widespread use of this method, considerable technical issues impose limitations (described in Komárek 2006; Kormas et al. 2011; Abad et al. 2016; Li et al. 2019a; Esenkulova et al. 2020). For instance, the detection of picocyanobacteria (cells 0.2–2 μm in diameter) using standard light microscopy is limited; often a completely different form of microscopy is used to examine this group (i.e., fluorescence). Identification based on morphology can be confounded by the presence of cryptic species (e.g., Engene et al. 2018; Esenkulova et al. 2020; Li et al. 2020; Vuorio et al. 2020), by cyanobacteria species that lack distinguishing features or by the clustering of filamentous, bloom-forming species that make taxonomic distinction more challenging (Li et al. 2019b). Lastly, phenotypic variation from changing environmental and culturing conditions can lead to diverging identifications between taxonomists, including of reference strains (Komárek and Anagnostidis 1989; Komárek 2006; Lee et al. 2014; Li et al. 2019a).

Molecular genetic techniques such as DNA metabarcoding, where a short taxonomically informative DNA region is amplified, sequenced and annotated against a reference sequence library (Taberlet et al. 2012), are increasingly being used for assessing aquatic species biodiversity in the 21st century (Clark et al. 2018). Molecular genetic techniques may overcome limitations of microscopy as they can yield faster and high throughput sample processing, detect rare, cryptic and small-sized species, and are generally considered more cost-effective (Eiler et

al. 2013; Abad et al. 2016; Pawlowski et al. 2018; Santi et al. 2021). The taxonomy of cyanobacteria is continuously going through reorganization with the development of such molecular tools (Komárek 2016a, 2016b). Metabarcoding frequently targets the 16S rRNA gene or the internal transcribed spacer region to identify cyanobacteria taxa and elucidate patterns of their biodiversity (Komárek 2006; Woodhouse et al. 2016). These have provided a more comprehensive community profile of cyanobacteria blooms in several ecosystems (Parulekar et al. 2017; Batista et al. 2018; Oliveira et al. 2019), including drinking water reservoirs (Gao et al. 2018; Casero et al. 2019). Furthermore, the genotypic analysis could provide a new level of ecological information as they can provide insight into cryptic species or ecotypes (Callahan et al. 2017). However, there are several limitations and biases introduced at each step of the metabarcoding procedure beginning with water sampling, DNA extraction method (Vasselon et al. 2017), marker gene selection (Kermarrec et al. 2013), PCR amplification, choice of bioinformatic pipeline used to process high-throughput sequencing data (Bailet et al. 2020) and clustering level of sequences (Tapolczai et al. 2019). Additionally, molecular approaches cannot discern between life-stages (Costa et al. 2016), and incomplete taxonomic coverage or erroneous taxonomic annotations in DNA reference libraries limit and confound the number of taxa that can be reliably identified to species (Komárek and Anagnostidis 1989; Eiler et al. 2013; Apothéloz-Perret-Gentil et al. 2017; Pawlowski et al. 2018).

With growing recognition of the potential to enhance biodiversity studies with molecular genetic techniques, some studies have focused on how the taxonomic composition of communities compares between conventional morphological identification and that inferred from DNA metabarcoding (Pawlowski et al. 2018; Keck et al. 2021). Comparative studies have been performed on a variety of phytoplankton communities (Kormas et al. 2011; Costa et al. 2016;

Albrecht et al. 2017; Batista et al. 2018; Rimet et al. 2018; Casero et al. 2019; Li et al. 2019a; Huo et al. 2020) and have shown varying levels of congruence between the relative sequence abundance and/or the occurrence of taxa obtained by metabarcoding and the counts or biomass of organisms identified with microscopy. A recent meta-analysis comparing methods from 215 aquatic community datasets revealed that metabarcoding of microorganisms gave complementary, as opposed to identical, community data generated by microscopy (Keck et al. 2021). Numerous explanations have been provided to explain differences between communities obtained with each method, including incomplete reference libraries (Abad et al. 2016), marker gene copy number variation (Kembel et al. 2012), primer bias (Elbrecht and Leese 2015; Elbrecht et al. 2017) and microscopy-based constraints such as observer bias (Gao et al. 2018).

Despite the relevance to lake managers and basic science programs, there is a paucity of work performed across a wide suite of lakes comparing metabarcoding and traditional microscopic identifications of cyanobacteria from exactly the same samples (but see Li et al. 2019a). As a part of the NSERC Canadian Lake Pulse Network (Huot et al. 2019), we sampled over 300 lakes across Canada and assessed the correlation of cyanobacteria community data generated with each method. We quantified the congruence between the microscopic and metabarcoding datasets, considering several key issues including the presence of picoplankton, sequencing depth and lake type differences. Furthermore, we conducted an exploratory analysis of genotypic diversity from one cosmopolitan, bloom-forming cyanobacteria genus, *Microcystis*, to determine whether there is any apparent ecological structure to their distribution across lakes.

A global synthesis of *Microcystis* molecular analyses revealed its variable spatial distribution and genome diversity (Harke et al. 2016). Here, we hypothesized that regional patterns of *Microcystis* genotypes would emerge, with strains adapting to specific lake types (i.e.,

trophic state). Overall, our findings provide new insights into cyanobacterial community structure across Canada and highlight some of the limitations of each taxonomic identification approach that must be considered in future applications.

Materials and Methods

Lake selection

Three hundred and seventy-nine lakes spanning seven Canadian ecozones were sampled following a standardized protocol (NSERC Canadian Lake Pulse Network 2021) across eastern and central Canada (Huot et al. 2019; Fig. 1). A full description of lake selection within the context of the NSERC Canadian Lake Pulse Network is available in Huot et al. (2019). Briefly, sampled lakes were selected from a subset of natural Canadian lakes located within 1 km of a road and with a maximum depth of at least 1 m. Lakes were selected using a random sampling design that was stratified across ecozones, lake sizes (0.1-1km^2, 1-10km^2, 10-100km^2) and watershed human impact index categories. Ecozones (from east to west: Atlantic Highlands, Atlantic Maritime, Boreal Shield, Mixedwood Plains, Boreal Plains, Prairies and Semi-Arid Plateaux) are regions with distinct geological, climatic and ecological features, and range from approximately 52° to 118° West and 41° to 60° North (Ecological Stratification Working Group 1996). Maps of sites were constructed using ArcGIS 10.5.1© (ESRI 2016) with the NAD 83 Canada Atlas Lambert coordinate reference system. Ecozone shapefiles were sourced from the Canada Council of Ecological Areas (Wiken et al. 1996).

Sample collection

Lakes were sampled in 2017 or 2018 during the season of maximum thermal stratification (between July and early September) to minimize seasonal variability (Huot et al.

2019). Sampling equipment was soaked in concentrated HCl, then triple rinsed with lake water from the sampled lake before use to avoid cross-contamination. The sampling station was located at the deepest point of each lake (as measured on site using a depth sounder). Surface water was sampled using an integrated tube sampler inserted up to the shallower of the following two possible depths: (a) twice the Secchi disk depth (i.e., the approximate limit of the euphotic zone) or (b) two meters (the tube length). Carboys of sampled lake water were shielded from direct sunlight inside icepack-chilled coolers until they were processed, within a few hours. For microscopic analysis, a 120 mL subsample of surface water was fixed and preserved in Lugol's iodine. For 16S rRNA gene analysis, up to 500 mL of lake water was first pre-filtered through an acid-washed 100 µm nylon mesh (Nitex), then vacuum-filtered through a Durapore 0.22 µm membrane (Sigma-Aldrich, St. Louis, USA) using a Gast Pressure pump (Fisher Scientific, Quebec, Canada) until the filter clogged. The filters were immediately frozen at −80°C and kept frozen until analysis. Please refer to the NSERC Canadian Lake Pulse Network Field Manual for the detailed sample collection protocols (NSERC Canadian Lake Pulse Network 2021).

Morphological identification

Microscopic identification of the phytoplankton samples were all conducted by the same taxonomist (M.A.) on 2–10 mL aliquots following Utermöhl's sedimentation method (Lund et al., 1958) using a Zeiss Axiovert 40 CFL inverted microscope at 250 X and 500 X magnifications. A minimum of 400 units were counted along 2–10 transects of the chamber per sample. Counting units were single-celled individuals, filaments or colonies, depending on the organization of the algae. Weight biomass was calculated from recorded abundance and specific volume estimates based on geometric solids (Rott 1981; Hillebrand et al. 1999). Identifications

were based primarily on the following texts: Komárek et al. (1998, 2005) and Komárek et al. (2013).

DNA extraction, sequencing & bioinformatic processing

As described in Kraemer et al. (2020), DNA was extracted from the filters with DNeasy PowerWater DNA isolation kits (Qiagen., Hilden, Germany) following the manufacturer's protocol with the following additional steps: after bead beating and centrifugation (step 7 of the detailed protocol), 1 µL of RNase A was added, followed by a 30-min incubation at 37 °C. A 250 bp fragment of the 16S rRNA gene V4 region was amplified using the universal bacterial primers U515_F (5'-GTGCCAGCMGCCGCGGTAA-3') and E806_R (5'-GGACTACHVGGGTWTCTAAT-3') (Caporaso et al. 2011). Each PCR contained a 25 µL total volume with the following components: 5 µL Phusion High Fidelity Buffer (5X), 0.5 µL dNTPs (10 mM), 1.8 µL of each primer (5 µM), 0.25 µL Phusion polymerase, 13.65 µL ultrapure nucleic acid-free water and 2 µL of genomic DNA. PCR conditions were an initial denaturation at 98 °C for 30 s, followed by 22 cycles of 98 °C for 20 s, 54 °C for 35 s, 72 °C for 30 s, and a final elongation at 72 °C for 1 min. Sequencing of 250 bp paired-end fragments was performed on an Illumina MiSeq platform in six runs in B. Jesse Shapiro's laboratory at Université de Montréal.

Reads were processed using the DADA2 package in R (Callahan et al. 2016). They were trimmed and filtered, amplicon sequence variants (ASVs) were inferred, paired-end reads were merged, and chimeras were removed. Taxonomy was assigned to ASVs against the SILVA database v.132 (Quast et al. 2012) at a minimum bootstrap confidence level of 80%. We enriched the SILVA database with additional 16S rRNA gene sequences of unrepresented cyanobacteria including *Aphanocapsa* and *Planktolyngbya*. These additional sequences were pulled in June of

2017 from NCBI and Ramos et al. (2017), a curated database of cyanobacteria strains. We chose to conduct our analyses on ASV and genus level groupings. To confirm taxonomic assignments, sequences of the most abundant ASVs were crosschecked using NCBI BLAST. We extracted all ASVs assigned to cyanobacteria to obtain a site by ASV matrix. For all subsequent analyses, we removed ASVs assigned to chloroplasts, non-photosynthetic cyanobacterial groups and sequences which could not be assigned below phylum level. Since we knew *a priori* that the identification of cyanobacteria based on light microscopy is difficult for taxa without easily distinguishable features, genus level or complex groupings were constructed to account for this bias (Table S1). Groupings were first proposed by the lead author (P.M.) and then finalized by our taxonomist (M.A.) by reviewing the list of taxa identified by light microscopy and clustering closely related cyanobacteria that are easily mistaken for one another. The taxonomist confirmed these groupings without knowledge of the metabarcoding results. All taxon names from molecular genetic and microscopy data were harmonized according to the taxonomy used in the continuously updated Algaebase (Guiry and Guiry 2020).

Statistical analyses

All statistical analyses were performed in R v.4.0.2 (R Core Team 2020). The composition of the cyanobacteria assemblages obtained with microscopy and metabarcoding were compared across sample matrices. For this, we first performed Principal Component Analyses (PCA) using the vegan R library (Oksanen et al. 2019) on each site by species matrix considering each dataset at genus or complex level. Prior to performing PCAs, matrices were Hellinger-transformed using the *decostand()* function, which converts the data into the square root of relative abundances (Legendre and Gallagher 2001). For the data based on microscopic identification, we used the calculated density (cell/mL) and biomass (µg/L) of each group as the

units of observation. The sequence abundance of each ASV was used to represent the metabarcoding identification. PCAs were performed separately on each sample matrix and taxonomic method, and the first three axes of sites scores were extracted using the *scores()* function and visually compared. Sites were categorized by trophic state according to total phosphorus concentration thresholds by Wetzel (2001): oligotrophic (total phosphorus 0 - 10 µg/L), mesotrophic (10 - 30 µg/L), eutrophic (30 - 100 µg/L), and hypereutrophic (>100 µg/L). We applied an RV coefficient to quantify similarity between two competing matrices using the first three PCA axes (which captured the largest fraction of variation). The RV coefficient is a generalization of the Pearson correlation that correlates two matrices of quantitative data with corresponding rows (sites), homologous to an R^2 (Legendre and Legendre 2012). RV coefficients between sample matrices were computed with the *coeffRV()* and their statistical significance was calculated using *RV.rtest()*, from the library FactoMineR (Lê et al. 2008) and ade4 (Dray and Dufour 2007), respectively. To address inherent differences or challenges associated with each taxonomic method, several iterations of the community matrices were compared including removing picocyanobacteria (cells < 2µm) and lakes with fewer than 500 sequences assigned to cyanobacteria. In order to reduce noise in both datasets and eliminate artefacts potentially introduced by sequencing errors, we removed low abundant genera (i.e., below 2% relative abundance in any single lake) and did not consider unassigned sequences at the genus level when computing RV coefficients between matrices (Pawlowski et al. 2018). The 2% relative abundance rule removed 5 genera (*Aphanothece*, *Arthrospira*, *Limnothrix*, *Oscillatoria* and *Spirulina*) from the metabarcoding dataset that were identified in the microscopy dataset, however this had no effect on the strength of correlation between methods.

Results

Cyanobacteria community structure revealed by light microscopy

Across the 379 lakes, a total of 31 cyanobacteria genera were identified using morphological characteristics, with an average of 12 genera detected in each lake. The cyanobacteria represented between 0.28% to 100% of the total phytoplankton cell abundance, ranging from 4 cells/ mL to 761,679 cells/mL, and with a median of 340 cells/mL. The highest cell densities were recorded in central Canada within the Prairies and Boreal Plains ecozones (Fig. 1). The most frequent genera by relative abundance of cell densities across all sites were *Microcystis* (53.8%), *Aphanizomenon* (16.7%), *Aphanocapsa* (8.8%), *Dolichospermum* (6.7%), and *Woronichinia* (5.8%). When taxa that are more difficult to distinguish were grouped into complexes, we observed that the relative abundances of cell densities were dominated by the *Synechococcus* complex, *Microcystis, Aphanizomenon* complex, *Dolichospermum* complex and *Lyngbya* complex (Table 1; Fig. 2A).

Cyanobacteria community structure revealed by 16S rRNA gene metabarcoding

After quality filtering and preprocessing, 16S rRNA gene metabarcoding using universal bacterial primers yielded 41,786,071 sequences amongst 41,289 ASVs (total across all samples). Of these, 4,103 were identified as singletons or doubletons and were subsequently removed from the dataset. The number of sequences was highly variable among lakes ranging from 14,334 to 576,574 with a median of 95,337 sequences per lake. After removing ASVs that were assigned to other phyla, chloroplasts or non-photosynthetic cyanobacteria (Monchamp et al. 2019), a total of 1,016 ASVs remained, comprising 3,675,227 sequences, with a median of 4,203 cyanobacteria sequences per lake.

Most of the 1,016 cyanobacteria ASVs were numerically rare. We were able to assign 56% of the ASVs to genus level, which comprise 98.4% of all cyanobacteria sequences. Following the data filtration steps described above, we identified 57 cyanobacteria genera, including 174 ASVs assigned to *Cyanobium*, 63 to *Aphanizomenon*, 38 to *Pseudanabaena*, 34 to *Dolichospermum*, and 10 to *Microcystis*. The top five genera with the highest sequence abundances represented just over 80% of cyanobacteria sequences: *Cyanobium* (38.7%), *Planktothrix* (17.4%), *Aphanizomenon* (11.6%), *Microcystis* (7.6%), and *Dolichospermum* (5.8%). When the molecular dataset was grouped into the complexes defined as hard to distinguish taxa by light microscopy, we found that *Microcystis*, the *Synechococcus* complex (which includes *Cyanobium*), and the *Aphanizomenon* complex had the greatest relative sequence abundances (Table 1; Fig. 2A).

Coherence between light microscopy and 16S rRNA gene metabarcoding

Direct comparisons between molecular and microscopic datasets were made using DNA sequence counts and cell density (cells/mL) or biomass (µg/L). Although all of the main comparisons were done using cell density, biomass was checked as it is a commonly used metric to analyze phytoplankton community structure. Based on the screened datasets, there were 30 and 33 genera from the microscopy and metabarcoding datasets, respectively. Of these, we found that 19 genera were shared across analytical platforms, with 11 unique to microscopy and 14 unique to metabarcoding (Fig. S3). When direct comparisons were made using genus complexes, we detected 17 genera/complexes with microscopy and 18 with metabarcoding with high overlap between groups (Fig. 2B).

The proportion of the dominant genera/complexes identified by each method differed in magnitude but followed a similar pattern along a trophic gradient (Fig. 3). In the metabarcoding

approach, the *Synechococcus* complex represented 78% of sequences in oligotrophic lakes but was reduced to 29% in hypereutrophic sites. Similarly, the *Synechococcus* complex decreased in relative abundance from 45% in oligotrophic lakes to just 13% in hypereutrophic lakes in the microscopy dataset. *Microcystis* and the *Aphanizomenon* complexes became more dominant in the cyanobacteria assemblages of eutrophic and hypereutrophic sites in both datasets. A notable difference between methods was the evenness of assemblages across trophic states. Microscopy-generated communities consistently contained at least several genera and complexes with over 10% relative abundance, whereas metabarcoding results showed that sites tended to be dominated by one or two complexes in each trophic state (Fig. S4).

Microscopy- and molecular-generated communities showed comparable dominance patterns in terms of average relative abundance (Fig. 2A, Table 1) but diverged in frequency of occurrence (Fig. 2B). For example, the *Synechococcus* complex was dominant in both the metabarcoding and microscopy datasets (i.e., mean relative abundance of 61% across all sites in the former, compared to 36% in the latter). The divergence in mean relative abundances and frequency of occurrence results are expected for the *Synechococcus* complex as the single-celled picocyanobacteria will be underestimated with Utermöhl microscopic analyses (Fig. S1 and S2). In contrast, average relative abundances for the *Planktothrix* complex and *Microcystis* were similar between microscopy and metabarcoding. However, subtle differences were noted: microscopy detected on average more *Merismopedia* (5.04% *vs*. 0.05% with metabarcoding), *Chroococcus* complex (5.04% *vs*. 0.29%), *Lyngbya* complex (9.40% *vs*. 1.14%) and *Dolichospermum* complex (11.89% *vs*. 2.65%). These differences were consistent even when we considered the microscopy data in terms of biomass. In addition, microscopic analyses distinguished fewer generalist taxa, with just one complex (i.e., the *Synechococcus* complex)

detected in more than 200 lakes (Fig. 2B). In contrast, the metabarcoding dataset had five complexes that were present in over 300 lakes (Fig. 2B), including the *Synechococcus* complex which was detected in every lake in this study.

Analysis of the variation in cyanobacteria community composition by ordination showed generally similar patterns between taxonomic platforms. The PCA biplot of the community identified by microscopy showed a clear trophic gradient, with oligotrophic and mesotrophic lakes largely distinguished from hypereutrophic lakes. *Microcystis*, *Aphanizomenon* complex and *Planktothrix* complex were characteristic of eutrophic sites, whereas *Synechococcus* complex was characteristic of oligotrophic and mesotrophic sites (Fig. 4). The first two axes of the PCA carried out on cyanobacteria composition based on microscopy explained 44% of the variance. A similar pattern of variation was observed for the metabarcoding generated community, with the same complexes split along a trophic gradient. Here, the first two axes of the PCA explained 56% of the variance (Fig. 5).

Using RV coefficients, we calculated the level of correspondence between sample matrices and taxonomic methods using the site scores of PCA axes 1, 2 and 3. In particular, we calculated the correlation between methods across all sites as well as just the eutrophic to hypereutrophic lakes, where cyanobacteria bloom monitoring tends to be more concerted. We also considered a range of taxonomic levels as well as a minimum threshold of sequences retained. Finally, a separate analysis was performed with picocyanobacteria removed, in recognition that these will be underestimated with the Utermöhl microscopy approach. Overall, we found the strongest correlations between sample matrices and taxonomic methods when broader taxonomic levels were considered (i.e., order level in sites with TP >30 µg/L: RV = 0.40, $p < 0.0001$), or when we examined only eutrophic-hypereutrophic communities where an

adequate number of sequences (>500 sequences) was analyzed and picocyanobacteria were removed (RV = 0.34, $p < 0.0001$; Fig. 6).

Distribution of Microcystis ASVs

In total, 10 ASVs were assigned to *Microcystis*, and collectively occurred in 351 of the 379 lakes. Individual ASVs ranged widely in occurrence and relative abundance. For example, the most common *Microcystis* ASV, "ASV0004," was found in 309 lakes and varied between 0.18% and 100% relative abundance of all *Microcystis* sequences. In contrast, ASV0053 occurred in 186 predominantly oligotrophic and mesotrophic lakes and represented on average 5% of *Microcystis* sequences. Using PCA to visualize the variation in *Microcystis* assemblages across our sites, we found a clear trophic gradient (the first two axes of this ordination explained 69.3% of the variation). Frequently occurring *Microcystis* ASVs were also oriented along a trophic state gradient, with ASVs 0004 and 0016 associated with more eutrophic to hypereutrophic sites. Conversely, ASV 0053 was more characteristic of mesotrophic and oligotrophic lakes (Fig. 7).

Discussion

We found a broad congruence between cyanobacteria assemblages identified using morphological and molecular genetic techniques in 379 lakes ranging in size and trophic state. Both methods captured the expected shift in community composition along a trophic gradient, with a decrease in picocyanobacteria relative abundance and an increase in common bloom formers such as *Microcystis* and *Aphanizomenon* in eutrophic to hypereutrophic sites. Though the molecular genetic approach detected more rare and benthic taxa than microscopy, most of these taxa accounted for a small percentage of genetic sequences and rarely reached 2% relative

abundance in any lake. The greatest correspondence between methods was recorded when broader levels of taxonomic assignment were applied (i.e., Family and Order), and when the analysis was restricted to eutrophic and hypereutrophic lakes. Our findings are relevant for lake managers as they can assume with confidence that cyanobacteria community data from both methods are relatively comparable in lakes prone to bloom events. Because they are relatively fast and cost-effective, metabarcoding techniques may be employed even more widely in the future as reference libraries continue to improve.

Cyanobacterial community structure across Canadian lakes

The structure of cyanobacteria communities at this landscape-scale showed a relationship with lake trophic state that was expected based on previous research (Fig. 3; Watson et al. 1997). Across the analytical platforms, we detected similar trends in the dominant taxon complexes along a trophic gradient, but at the genus-level resolution we detected some differences in dominant taxa (Figs. 3 and S4). For example, approximately 50% of the cyanobacteria community based on cell counts from oligotrophic and mesotrophic lakes was represented by the *Synechococcus* complex, mainly species of *Aphanocapsa*. From these same sites, amplicon sequencing also identified the *Synechococcus* complex as the dominant group, representing over 75% of the cyanobacteria community; however, *Cyanobium* demonstrated a clear predominance in these lakes (Figs. 3 and S4). Interestingly, based on fluorescence microscopy taxonomy, *Cyanobium* cells are typically called *Synechococcus* or *Synechocystis*. As we transitioned into nutrient-rich sites, *Microcystis*, *Aphanizomenon* complex, and *Planktothrix* complex were characteristic of eutrophic and hypereutrophic lakes in both datasets (Figs. 4 and 5). However, the relative abundances of these complexes did not always concord. Although we identified

statistically significant clusters corresponding to lake trophic status using both survey methods, several taxonomic differences were observed and are discussed in detail below (Figs. 4 and 5).

Congruence between the microscopic and metabarcoding datasets

With the growing use of DNA sequencing techniques to assess biodiversity in aquatic ecosystems, there is a need to quantify the congruence between these methods with more traditional ways of identifying community structure (Pawlowski et al. 2021, 2018). Phytoplankton surveys from around the world have incorporated and compared microscopy and molecular genetic analyses, and have recorded a range of correlations including some strong relationships between sequences and cell counts or biomass (Eiler et al. 2013; Stoeck et al. 2014; Xiao et al. 2014; Abad et al. 2016; Costa et al. 2016; Rimet et al. 2018; Li et al. 2019a; Esenkulova et al. 2020; Huo et al. 2020; Vuorio et al. 2020). However, there remain few studies specifically quantifying the congruency in cyanobacteria assemblages between survey methods. This is a necessary step towards the benchmarking of molecular analyses and a gap this study addresses.

Highest congruency in eutrophic and hypereutrophic lakes

Within the phytoplankton literature, inconsistencies between the relative abundances of different dominant taxa identified by each method have also been reported (Abad et al. 2016; Albrecht et al. 2017; Esenkulova et al. 2020). Some authors stress the complementary role that molecular based approaches serve to traditional microscopy (Esenkulova et al. 2020; Keck et al. 2021). A clear strength of metabarcoding is the identification of cryptic species, as well as pico-sized taxa which are often impossible to differentiate with microscopy (Costa et al. 2016; Vuorio et al. 2020). Across our 379 lakes, the microscopy and 16S rRNA gene sequence analyses were in broad agreement for the genus complexes, particularly in eutrophic and hypereutrophic sites,

despite reducing the number of genera prior to constructing the complexes. When considering how the genus complexes compared across study sites, eight of the top ten complexes emerged from both microscopy and metabarcoding, but in a different order of relative abundance (Table 1). The most significant difference occurred in the Synechococcus complex, which is comprised largely of unicellular and colonial picocyanobacteria, that represented an average of 9% of the cell counts but was the overwhelmingly dominant complex of the metabarcoding data, representing 41% of sequences. This five-fold difference in relative abundances highlights the sensitivity of molecular based analyses to capture a larger portion of the picocyanobacteria assemblage. Although the most abundant genus complexes were the same, they were identified by metabarcoding in more samples, suggesting that metabarcoding more effectively identified a broader diversity of cyanobacteria (Figs. 2A and 2B). Our findings are consistent with numerous other cyanobacteria metabarcoding studies of water and sediment core samples, where picocyanobacterial taxa are often dominant and overall richness tends to be greater than with microscopic counts (Monchamp et al. 2016; Li et al. 2019a; Vuorio et al. 2020).

While assessing the correlation between methods at different taxonomic levels, the highest correlation was recorded at the broadest taxonomic grouping (i.e., Order level) (Fig. 6 and S1). At this level, microscopy and metabarcoding were approximately 40% in concordance within the eutrophic sites. The strength of correlation gradually declined at Family and Genus level, with the lowest observed correlation occurring when picocyanobacteria were removed from both datasets to account for the detection limits of microscopy. However, removing the picocyanobacterial fraction resulted in the deletion of over 75% of the cyanobacteria sequences, leaving a fraction of the original data to compare. To account for this, we screened the data to only retain sites with a minimum threshold of sequences and found that the correlation increased

again. Establishing a minimum threshold of sequences to make adequate comparisons might be something other studies should consider in the future. Overall, we found that when comparing high sequence sites with the picocyanobacteria bias eliminated, correlations between methods reached over 30% at the genus level. In all instances of method comparisons, correlations were systematically higher when just eutrophic lakes were considered versus the entire lake dataset. This trend is best explained by the fact that the oligotrophic and mesotrophic sites are represented by greater richness of rare taxa; often taxa that cannot be identified microscopically and/or ones that are more often missing from the reference databases (Vuorio et al. 2020). Eutrophic and hypereutrophic lakes were characterized by greater proportions of *Microcystis, Aphanizomenon* complex and *Dolichospermum* complex: groups that were consistently detected by both methods.

Technical issues that may influence congruency between methods

Differences observed between methods may stem from numerous biases that make perfect congruency impossible (Pawlowski et al. 2021, 2018). We briefly discuss some of the biological and technical sources of bias below, but see (Pawlowski et al. 2018) for an exhaustive discussion. To start, the sample volumes settled (2–10 mL for microscopy versus up to 500 mL for metabarcoding), and underlying units used for microscopy (individual cells) and those used for metabarcoding (ASV sequences) are quite different, making direct comparison imperfect (Pawlowski et al. 2018; Vuorio et al. 2020). As discussed previously, we evaluated the influence of sequencing depth by conducting analyses with a minimum sequence threshold. For technical factors, several studies comparing taxonomic approaches have found DNA extraction efficiency and primer amplification bias to be the leading cause of discrepancies (Elbrecht and Leese 2015; Elbrecht et al. 2017). Applying a primer set designed specifically to amplify cyanobacteria

(Nübel et al. 1997) or regions of the 16S rRNA gene more variable than V4 could enhance the assessment of freshwater cyanobacteria diversity, particularly picocyanobacteria (Huber et al. 2019). Furthermore, sequencing a short (~250 bp) gene fragment with universal primers may have limited the taxonomic resolution, yet we were able to assign 56% of ASVs to genus level after extensive data curation. If the available representative sequences had missing or incomplete sections on the V4 region, then this might explain why some taxa were identified by microscopy but were not assigned via metabarcoding, despite having sequence representation in the database (i.e., *Aphanocapsa, Aphanothece* and *Limnothrix*). However, it is worth recognizing that the fraction of unassigned ASVs from the total pool of cyanobacteria reads was very small (1.6%).

The most striking difference between microscopy and metabarcoding was the proportion of picocyanobacteria detected by each method (Fig. S2). For cells below the detection limit of light microscopy and Utermöhl (i.e., approximately 1 μm in diameter), it can be difficult to even distinguish a heterotrophic bacterium from a photosynthetic one. In addition, pico-sized cells that may not sink within the Utermöhl counting chamber are often missed or underrepresented by microscopy (Albrecht et al. 2017; Batista et al. 2018). Furthermore, picocyanobacteria can lack distinctive morphological characteristics commonly used for visual identification. Yet, sufficiently counting the picocyanobacteria is important given their ubiquity in temperate lakes, their importance in biogeochemical cycles (Callieri 2008) and the capability of some taxa/strains to produce toxins (Jakubowska and Szeląg-Wasielewska 2015). In many oligotrophic to mesotrophic lakes, picocyanobacteria can reach concentrations as high as 10^5 cells/mL in summer in contrast to other cyanobacteria and eukaryotes at 10^3 cells/mL: a potential 100 times difference in abundance/sequences (Pick and Agbeti 1991). Microscopy overwhelmingly identified *Aphanocapsa*, whereas metabarcoding yielded a dominance of *Cyanobium*. Similar

mismatches between microscopy and metabarcoding have been reported previously. For example, in study of a coastal lagoon, another pico-sized genus, *Aphanothece,* was very abundant in the microscopic counts, whereas metabarcoding yielded no sequences of *Aphanothec*e, but rather a dominance of *Cyanobium* (Albrecht et al. 2017). Moreover, in a comparative analysis of over 50 lakes and reservoirs, (Li et al. 2019a) observed no *Cyanobium* based on microscopy, despite the presence of *Cyanobium* in most samples detected by metabarcoding. In each study, inconsistencies were ascribed to similar cell shapes between genera, thus making morphological distinction challenging. In addition to picocyanobacteria, the Utermöhl method may also underestimate the abundance of buoyant, colonial cyanobacteria such *Microcystis*, whose colonies can float at the top of the chamber. Although our microscopy dataset was still able to detect large abundances of *Microcystis*, this effect may lead to minor discrepancies and should be taken into consideration.

Biological issues that may influence congruency between methods

Among the biological biases, copy number variation is known to decrease the quantitative value of metabarcoding data (Kembel et al. 2012; Schirrmeister et al. 2012). Heterocystous cyanobacteria have been shown to contain up to five 16S rRNA gene copies, whereas genera in the Synechococcales and Chroococcales orders generally have only one or two copies (Kembel et al. 2012; Schirrmeister et al. 2012). The high copy number of the 16S rRNA gene among heterocystous cyanobacteria can lead to their overestimation in PCR-based methods. Copy number has also been reported to vary as a function of cell size in other phytoplankton groups, such as diatoms (Godhe et al. 2008). To explore whether cell size could influence the coherence between methods, we also ran RV coefficients with cell biomass data (as

opposed to density) but found very similar results (Fig. S5). As such, we think that with cyanobacteria, the variation in cell numbers is more important than the variation in cell sizes.

Incomplete reference databases

Lastly, the most cited reason for incongruence of taxonomic assemblages obtained from microscopy and metabarcoding analyses are incomplete reference databases (Xiao et al. 2014; Zimmermann et al. 2015; Vasselon et al. 2017; Pawlowski et al. 2018). Although we constructed genus complexes to account for this bias, some of the most abundant taxa observed by microscopy either lacked corresponding sequences in the enriched SILVA database, or were restricted to a small number of partial sequences (i.e., *Aphanocapsa, Aphanothece, Chroococcus, Merismopedia* and *Limnothrix*). To improve identification, more reference sequences are needed for rare taxa and the many small (0.2 - 2 μm) colonial cyanobacteria that were detected in a higher number of samples by microscopy. In addition to missing taxa from curated reference databases, sequences deposited with incorrect taxonomy can lead to erroneous designations downstream (Komárek and Anagnostidis 1989; Komárek 2010; Lee et al. 2014). Strain and sequence names will need to be updated as the taxonomy of cyanobacteria is changing as we generate more molecular information. For example, several species previously named *Aphanizomenon* are now *Cuspidothrix*; a change that may not be corrected in sequence databases. Further phylogenetic analyses may help validate the taxonomic assignment of sequence data.

ASV level diversity reveals structured Microcystis profiles

A potential strength of the metabarcoding approach is its ability to detect cryptic taxa and identify whether these taxa occupy distinct ecological niches. We explored these issues by examining the distinction of key ASVs within *Microcystis*, which was a dominant genus detected

in both methods (Table 1). Previous work has identified different strains and genotypes of *Microcystis*, demonstrating that multiple genotypes can coexist in the same environment, while identifying some distinction in their ecological niches (Berry et al. 2017; Otten et al. 2017; Guan et al. 2018; Tromas et al. 2018; Chun et al. 2020; Cook et al. 2020; Jankowiak and Gobler 2020; Smith et al. 2021). For example, the succession of *Microcystis* genotypes have been shown to vary inter-annually (Ninio et al. 2020) and seasonally, with dominant genotypes differing between summer and autumn blooms in the Daechung Reservoir (Korea), and during different phases of blooms (Chun et al. 2020). A network analysis conducted on a microbial time series from the same reservoir by (Chun et al. 2020) revealed that other members of the community (bacterial and eukaryotic) were associated with certain *Microcystis* genotypes in summer but associated with different genotypes in the fall. Spatial and temporal variation in *Microcystis* genotypes have also been correlated to environmental drivers including temperature (Ninio et al. 2020) and phosphorus gradients (Berry et al. 2017). Based on a PCA of M*icrocystis* ASVs across our sites, we found a clear distribution of taxa along a trophic gradient (Fig. 7). Although a single *Microcystis* ASV comprised the majority of sequences assigned to *Microcystis*, ASVs varied across different trophic states. This trend is relevant as both spatial and temporal changes in the dominant *Microcystis* genotype have corresponded to shifts in bloom toxicity (Gobler et al. 2016; Berry et al. 2017; Chun et al. 2020; Ninio et al. 2020). Important ASV level variation has been reported in other cyanobacteria genera as well. For example, Costa et al. (2016) found that during a bloom of *Anabaenopsis elenkinii*, one genotype was always the most abundant. The authors attributed the presence of other, less abundant genotypes as an adaptive strategy to maintain the population with changing ecological conditions. Access to genotype level

information supports the use of molecular genetic techniques in lake management programs. We encourage more work in this relatively emerging area of study.

Conclusions

Our large microscopy and metabarcoding datasets from 379 lakes revealed both a clear separation of the cyanobacteria communities from lakes of different trophic state and moderate levels of congruence between datasets, particularly in eutrophic to hypereutrophic lakes where picocyanobacteria are less dominant. We believe this work shows that metabarcoding could become more widely applied in water quality status assessments, providing the means to efficiently monitor eutrophication, one of the main environmental problems in surface freshwaters (Eiler et al. 2013). However, for the molecular work to achieve its full potential, more work is needed in developing and curating reference databases. These findings highlight some of the challenges in the taxonomic classification of cyanobacteria.

Acknowledgements

We would like to thank all members of the LakePulse Network who helped sample the many lakes in this study, and especially Yannick Huot who coordinated the network. Funding for this project came from the Natural Sciences and Engineering Research Council of Canada (NSERC) Canadian Lake Pulse Network (funding to IGE, BEB and DAW), Genome Canada/Genome Quebec (funding to BJS) as well as NSERC and Fonds de recherche du Québec en nature et technologies (FRQNT) scholarships to PWM, a FRQNT Fellowship to REG, a postdoctoral fellowship from the Groupe de recherche interuniversitaire en limnologie (GRIL), McGill University and Liber Ero Funding to MEM, and Canada Research Chair awards to IGE

and DAW. We also appreciate feedback from the Gregory-Eaves lab on an earlier draft of the manuscript.

References

Abad, D., A. Albaina, M. Aguirre, A. Laza-Martínez, I. Uriarte, A. Iriarte, F. Villate, and A. Estonba. 2016. Is metabarcoding suitable for estuarine plankton monitoring? A comparative study with microscopy. Mar. Biol. **163**: 149–150. doi:10.1007/s00227-016-2920-0

Albrecht, M., T. Pröschold, and R. Schumann. 2017. Identification of Cyanobacteria in a eutrophic coastal lagoon on the Southern Baltic Coast. Front. Microbiol. **8**: 1–16. doi:10.3389/fmicb.2017.00923

Apothéloz-Perret-Gentil, L., A. Cordonier, F. Straub, J. Iseli, P. Esling, and J. Pawlowski. 2017. Taxonomy-free molecular diatom index for high-throughput eDNA biomonitoring. Mol. Ecol. Resour. **17**: 1231–1242. doi:10.1111/1755-0998.12668

Azevedo, S. M. F. O., W. W. Carmichael, E. M. Jochimsen, K. L. Rinehart, S. Lau, G. R. Shaw, and G. K. Eaglesham. 2002. Human intoxication by microcystins during renal dialysis treatment in Caruaru - Brazil. Toxicology **181**–**182**: 441–446. doi:10.1016/S0300-483X(02)00491-2

Bailet, B., L. Apothéloz-Perret-Gentil, A. Baričević, and others. 2020. Diatom DNA metabarcoding for ecological assessment: Comparison among bioinformatics pipelines used in six European countries reveals the need for standardization. Sci. Total Environ. **745**. doi:10.1016/j.scitotenv.2020.140948

Batista, A. M. M., C. C. Figueredo, and A. Giani. 2018. Variability in a permanent cyanobacterial bloom: species-specific responses to environmental drivers. FEMS Microbiol. Ecol. **94**: 1–11. doi:10.1093/femsec/fiy197

Berry, M. A., J. D. White, T. W. Davis, S. Jain, T. H. Johengen, G. J. Dick, O. Sarnelle, and V. J. Denef. 2017. Are oligotypes meaningful ecological and phylogenetic units? A case study of

Microcystis in Freshwater lakes. Front. Microbiol. **8**: 1–7. doi:10.3389/fmicb.2017.00365

Callahan, B. J., P. J. McMurdie, and S. P. Holmes. 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. **11**: 2639–2643. doi:10.1038/ismej.2017.119

Callahan, B. J., P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, and S. P. Holmes. 2016. Dada2: high-resolution sample inference from Illumina amplicon data. Nat. Methods **13**: 581–583. doi:10.1007/978-1-4614-9209-2_118-1

Callieri, C. 2008. Picophytoplankton in Freshwater Ecosystems: The Importance of Small-Sized Phototrophs. Freshw. Rev. **1**: 1–28. doi:10.1608/frj-1.1.1

Caporaso, J. G., C. L. Lauber, W. A. Walters, D. Berg-Lyons, C. A. Lozupone, P. J. Turnbaugh, N. Fierer, and R. Knight. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. **108**: 4516–4522. doi:10.1073/pnas.1000080107

Casero, M. C., D. Velázquez, M. Medina-Cobo, A. Quesada, and S. Cirés. 2019. Unmasking the identity of toxigenic cyanobacteria driving a multi-toxin bloom by high-throughput sequencing of cyanotoxins genes and 16S rRNA metabarcoding. Sci. Total Environ. **665**: 367–378. doi:10.1016/j.scitotenv.2019.02.083

Castenholz, R. W. 2015. General Characteristics of the Cyanobacteria . Bergey's Man. Syst. Archaea Bact. 1–23. doi:10.1002/9781118960608.cbm00019

Chun, S. J., Y. Cui, J. J. Lee, I. C. Choi, H. M. Oh, and C. Y. Ahn. 2020. Network analysis reveals succession of Microcystis genotypes accompanying distinctive microbial modules with recurrent patterns. Water Res. **170**. doi:10.1016/j.watres.2019.115326

Clark, D. R., R. M. W. Ferguson, D. N. Harris, and others. 2018. Streams of data from drops of

water: 21st century molecular microbial ecology. WIREs Water **5**: 1–10. doi:10.1002/wat2.1280

Cook, K. V., C. Li, H. Cai, and others. 2020. The global Microcystis interactome. Limnol. Oceanogr. **65**: S194–S207. doi:10.1002/lno.11361

Costa, N. B., M. A. Kolman, A. Giani, and P. Moisander. 2016. Cyanobacteria diversity in alkaline saline lakes in the Brazilian Pantanal wetland: A polyphasic approach. J. Plankton Res. **38**: 1389–1403. doi:10.1093/plankt/fbw066

Cunha de Oliveira, E., A. Da Cunha, N. Da Silva, and others. 2019. Morphological and molecular characterization of cyanobacterial isolates from the mouth of the Amazon River. Phytotaxa **387**: 269–288. doi:https://doi.org/10.11646/phytotaxa.387.4.1

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/cjfas-58-10-1905

Dray, S., and A. B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. **22**: 1–20. doi:10.18637/jss.v022.i04

Ecological Stratification Working Group (Canada), Center for Land, Biological Resources Research (Canada) and Canada. 1996. State of the Environment Directorate. A national ecological framework for Canada. Centre for Land and Biological Resources Research; Hull, Quebec: State of the Environment Directorate.

Eiler, A., S. Drakare, S. Bertilsson, and others. 2013. Unveiling Distribution Patterns of Freshwater Phytoplankton by a Next Generation Sequencing Based Approach. PLoS One **8**: 1–10. doi:10.1371/journal.pone.0053516

Elbrecht, V., and F. Leese. 2015. Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative

metabarcoding protocol. PLoS One **10**: 1–16. doi:10.1371/journal.pone.0130324

Elbrecht, V., E. E. Vamos, K. Meissner, J. Aroviita, and F. Leese. 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. Methods Ecol. Evol. **8**: 1265–1275. doi:10.1111/2041-210X.12789

Engene, N., A. Tronholm, and V. J. Paul. 2018. Uncovering cryptic diversity of Lyngbya: the new tropical marine cyanobacterial genus Dapis (Oscillatoriales). J. Phycol. **54**: 435–446. doi:10.1111/jpy.12752

Esenkulova, S., B. J. G. Sutherland, A. Tabata, N. Haigh, C. M. Pearce, and K. M. Miller. 2020. Operational taxonomic unit comparing metabarcoding and morphological approaches to identify phytoplankton taxa associated with harmful algal blooms. Facets **5**: 784–811. doi:10.1139/FACETS-2020-0025

ESRI (Environmental Systems Research Institute). 2016. ArcGIS desktop: Release 10.5. Redlands, CA.

Gao, W., Z. Chen, Y. Li, Y. Pan, J. Zhu, S. Guo, L. Hu, and J. Huang. 2018. Bioassessment of a drinking water reservoir using plankton: High throughput sequencing vs. traditional morphological method. Water (Switzerland) **10**. doi:10.3390/w10010082

Gobler, C. J., J. A. M. Burkholder, T. W. Davis, M. J. Harke, T. Johengen, C. A. Stow, and D. B. Van de Waal. 2016. The dual role of nitrogen supply in controlling the growth and toxicity of cyanobacterial blooms. Harmful Algae **54**: 87–97. doi:10.1016/j.hal.2016.01.010

Godhe, A., M. E. Asplund, K. Härnström, V. Saravanan, A. Tyagi, and I. Karunasagar. 2008. Quantification of diatom and dinoflagellate biomasses in coastal marine seawater samples by real-time PCR. Appl. Environ. Microbiol. **74**: 7174–7182. doi:10.1128/AEM.01298-08

Guan, D. X., X. Wang, H. Xu, L. Chen, P. Li, and L. Q. Ma. 2018. Temporal and spatial

distribution of Microcystis biomass and genotype in bloom areas of Lake Taihu. Chemosphere **209**: 730–738. doi:10.1016/j.chemosphere.2018.06.141

Guiry, M.D., and G.M. Guiry. 2020. AlgaeBase. World-wide electronic publication, National University of Ireland, Galway.

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

Hillebrand, H., C. D. Dürselen, D. Kirschtel, U. Pollingher, T. Zohary. 1999. Biovolume calculation for pelagic and benthic microalgae. J. Phycol. **35**: 403–424. doi: 10.1046/j.1529-8817.1999.3520403.x

Ho, J. C., A. M. Michalak, and N. Pahlevan. 2019. Widespread global increase in intense lake phytoplankton blooms since the 1980s. Nature **574**: 667–670. doi:10.1038/s41586-019-1648-7

Huber, P., F. M. Cornejo-Castillo, I. Ferrera, and others. 2019. Primer design for an accurate view of picocyanobacterial community structure by using high-throughput sequencing. Appl. Environ. Microbiol. **85**: 1–17. doi:10.1128/AEM.02659-18

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser. 2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

Huo, S., X. Li, B. Xi, H. Zhang, C. Ma, and Z. He. 2020. Combining morphological and metabarcoding approaches reveals the freshwater eukaryotic phytoplankton community. Environ. Sci. Eur. **32**. doi:10.1186/s12302-020-00321-w

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network:

A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. **695**: 133668. doi:10.1016/j.scitotenv.2019.133668

Jakubowska, N., and E. Szeląg-Wasielewska. 2015. Toxic picoplanktonic cyanobacteria - Review. Mar. Drugs **13**: 1497–1518. doi:10.3390/md13031497

Jankowiak, J. G., and C. J. Gobler. 2020. The Composition and Function of Microbiomes Within Microcystis Colonies Are Significantly Different Than Native Bacterial Assemblages in Two North American Lakes. Front. Microbiol. **11**: 1–26. doi:10.3389/fmicb.2020.01016

Karlson, B., C. Cusack, and E. Bresnan, (eds). 2010. Microscopic and molecular methods for quantitative phytoplankton analysis. Paris, France, UNESCO, Intergovernmental Oceanographic Commission Manuals and Guides. **55**. doi: 10.25607/OBP-1371

Keck, F., R. C. Blackman, R. Bossart, and others. 2021. Meta-analysis shows both congruence and complementarity of DNA metabarcoding to traditional methods for biological community assessment. bioRxiv 2021.06.29.450286.

Kembel, S. W., M. Wu, J. A. Eisen, and J. L. Green. 2012. Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. PLoS Comput. Biol. **8**: 16–18. doi:10.1371/journal.pcbi.1002743

Kermarrec, L., A. Franc, F. Rimet, P. Chaumeil, J. F. Humbert, and A. Bouchez. 2013. Next-generation sequencing to inventory taxonomic diversity in eukaryotic communities: A test for freshwater diatoms. Mol. Ecol. Resour. **13**: 607–619. doi:10.1111/1755-0998.12105

Komarek, J. 2006. Cyanobacterial Taxonomy: Current Problems and Prospects for the Integration of Traditional and Molecular Approaches. Algae **21**: 349–375. doi:10.4490/algae.2006.21.4.349

Komárek, J. 2010. Recent changes (2008) in cyanobacteria taxonomy based on a combination of

molecular background with phenotype and ecological consequences (genus and species concept). Hydrobiologia **639**: 245–259. doi:10.1007/s10750-009-0031-3

Komárek, J. 2013. Cyanoprokaryota 3. Teil/3rd part: Heterocytous genera. In: Büdel, B., G.rtner, G., Krienitz, L., Schagerl, M. (Eds.), Süswasserflora von Mitteleuropa/ Freshwater flora of Central Europe. Springer Spektrum Berlin, Heidelberg, 1–400.

Komárek, J. 2016a. A polyphasic approach for the taxonomy of cyanobacteria: principles and applications. Eur. J. Phycol. **51**: 346–353. doi:10.1080/09670262.2016.1163738

Komárek, J. 2016b. Review of the cyanobacterial genera implying planktic species after recent taxonomic revisions according to polyphasic methods: State as of 2014. Hydrobiologia **764**: 259–270. doi:10.1007/s10750-015-2242-0

Komárek, J., K. Anagnostidis. 1989. Modern approach to the classification system of Cyanophytes 4-Nostocales. Archiv für Hydrobiologie. Supplementband. Monographische Beiträge. **82**: 247-345

Komárek, J., K. Anagnostidis. 1998. Cyanoprokaryota 1.Teil: Chroococcales. In: Ettl, H., G.rtner, G., Heynig, H., Mollenhauer, D. (Eds.), Süsswasserflora von Mitteleuropa 19/1. Jena-Stuttgart-Lübeck-Ulm, Gustav Fischer. 1-200.

Komárek, J., and K. Anagnostidis. 2005. Cyanoprokaryota. 2. Oscillatoriales. In: Büdel, B., Krienitz, L., G.rtner, G., Schagerl, M. (Eds.), Süsswasserflora von Mitteleuropa 19/2. Elsevier/Spektrum, Heidelberg, 1-200.

Kormas, K. A., S. Gkelis, E. Vardaka, and M. Moustaka-Gouni. 2011. Morphological and molecular analysis of bloom-forming Cyanobacteria in two eutrophic, shallow Mediterranean lakes. Limnologica **41**: 167–173. doi:10.1016/j.limno.2010.10.003

Kraemer, S. A., N. Barbosa da Costa, B. J. Shapiro, M. Fradette, Y. Huot, and D. A. Walsh.

2020. A large-scale assessment of lakes reveals a pervasive signal of land use on bacterial communities. ISME J. **14**: 3011–3023. doi:10.1038/s41396-020-0733-0

Lê, S., J. Josse, and F. Husson. 2008. FactoMineR: an R package for multivariate analysis. Journal of statistical software. **25**: 1-18.

Lee, E., U. M. Ryan, P. Monis, G. B. McGregor, A. Bath, C. Gordon, and A. Paparini. 2014. Polyphasic identification of cyanobacterial isolates from Australia. Water Res. **59**: 248–261. doi:10.1016/j.watres.2014.04.023

Legendre, P., and L. Legendre. 2012. Numerical ecology. Elsevier.

Legendre, P., and E. D. Gallagher. 2001. Ecologically meaningful transformations for ordination of species data. Oecologia **129**: 271–280. doi:10.1007/s004420100716

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

Li, X. C., S. Huo, J. Zhang, C. Ma, Z. Xiao, H. Zhang, B. Xi, and X. Xia. 2019b. Metabarcoding reveals a more complex cyanobacterial community than morphological identification. Ecol. Indic. **107**: 105653. doi:10.1016/j.ecolind.2019.105653

Li, X., S. Huo, and B. Xi. 2020. Updating the resolution for 16S rRNA OTUs clustering reveals the cryptic cyanobacterial genus and species. Ecol. Indic. **117**: 106695. doi:10.1016/j.ecolind.2020.106695

Lund, J. W. G., C. Kipling, and E. D. Le Cren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia **11**: 143–170. doi:10.1007/BF00007865

McElroy, M. E., T. L. Dressler, G. C. Titcomb, and others. 2020. Calibrating Environmental

DNA Metabarcoding to Conventional Surveys for Measuring Fish Species Richness. Front. Ecol. Evol. **8**: 1–12. doi:10.3389/fevo.2020.00276

Miller, M. A., R. M. Kudela, A. Mekebri, and others. 2010. Evidence for a novel marine harmful algal bloom: Cyanotoxin (microcystin) transfer from land to sea otters. PLoS One **5**: 1–11. doi:10.1371/journal.pone.0012576

Monchamp, M. E., J. C. Walser, F. Pomati, and P. Spaak. 2016. Sedimentary DNA reveals cyanobacterial community diversity over 200 years in two perialpine lakes. Appl. Environ. Microbiol. **82**: 6472–6482. doi:10.1128/AEM.02174-16

Ninio, S., A. Lupu, Y. Viner-Mozzini, T. Zohary, and A. Sukenik. 2020. Multiannual variations in Microcystis bloom episodes – Temperature drives shift in species composition. Harmful Algae **92**: 101710. doi:10.1016/j.hal.2019.101710

NSERC Canadian Lake Pulse Network. 2021. NSERC Canadian Lake Pulse Network field manual 2017 – 2018 - 2019 surveys (M-P. Varin, M. Beaulieu, & Y. Huot, Eds.). Université de Sherbrooke. doi: 10.17118/11143/18662.

Nübel, U., F. Garcia-Pichel, and G. Muyzer. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. Appl. Environ. Microbiol. **63**: 3327–3332. doi:10.1128/aem.63.8.3327-3332.1997

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt. 2019. Package 'vegan'. community ecology package. R package Version 2, 5–6

O'Neil, J. M., T. W. Davis, M. A. Burford, and C. J. Gobler. 2012. The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. Harmful Algae **14**: 313–334. doi:10.1016/j.hal.2011.10.027

Otten, T. G., H. W. Paerl, T. W. Dreher, W. J. Kimmerer, and A. E. Parker. 2017. The molecular

ecology of Microcystis sp. blooms in the San Francisco Estuary. Environ. Microbiol. **19**: 3619–3637. doi:10.1111/1462-2920.13860

Paerl, H. W., N. S. Hall, and E. S. Calandrino. 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. Sci. Total Environ. **409**: 1739–1745. doi:10.1016/j.scitotenv.2011.02.001

Paerl, H. W., and T. G. Otten. 2013. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. Microb. Ecol. **65**: 995–1010. doi:10.1007/s00248-012-0159-y

Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. **46**: 1349–1363. doi:10.1016/j.watres.2011.08.002

Parulekar, N., P. Kolekar, A. Jenkins, and others. 2017. Characterization of bacterial community associated with phytoplankton bloom in a eutrophic lake in South Norway using 16S rRNA gene amplicon sequence analysis. PLoS One **12**: 1–22. doi:10.1371/journal.pone.0173408

Pawlowski, J., A. Bonin, F. Boyer, T. Cordier, and P. Taberlet. 2021. Environmental DNA for biomonitoring. Mol. Ecol. **30**: 2931–2936. doi:10.1111/mec.16023

Pawlowski, J., M. Kelly-Quinn, F. Altermatt, and others. 2018. The future of biotic indices in the ecogenomic era: Integrating (e)DNA metabarcoding in biological assessment of aquatic ecosystems. Sci. Total Environ. **637**–**638**: 1295–1310. doi:10.1016/j.scitotenv.2018.05.002

Pick, F. R. 2016. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria. Can. J. Fish. Aquat. Sci. **73**: 1149–1158. doi:10.1139/cjfas-2015-0470

Pick, F. R., and M. Agbeti. 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. doi:10.1002/iroh.19910760409

Quast, C., E. Pruesse, P. Yilmaz, J. Gerken, T. Schweer, P. Yarza, J. Peplies, and F. O. Glöckner.

2012. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. Nucleic Acids Res. **41**: 590–596. doi:10.1093/nar/gks1219

R Core Team. 2020. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ramos, V., J. Morais, and V. M. Vasconcelos. 2017. A curated database of cyanobacterial strains relevant for modern taxonomy and phylogenetic studies. Sci. Data **4**: 1–8. doi:10.1038/sdata.2017.54

Reynolds, C.S. 2006. The Ecology of Phytoplankton. Cambridge: Cambridge University Press (Ecology, Biodiversity and Conservation). doi: 10.1017/CBO9780511542145

Rimet, F., V. Vasselon, B. A.-Keszte, and A. Bouchez. 2018. Do we similarly assess diversity with microscopy and high-throughput sequencing? Case of microalgae in lakes. Org. Divers. Evol. **18**: 51–62. doi:10.1007/s13127-018-0359-5

Rott, E. 1981. Some results from phytoplankton counting intercalibrations. Schweizerische Zeitschrift für Hydrol. **43**: 34–62. doi:10.1007/BF02502471

Santi, I., P. Kasapidis, I. Karakassis, and P. Pitta. 2021. A comparison of DNA metabarcoding and microscopy methodologies for the study of aquatic microbial eukaryotes. Diversity **13**: 1–12. doi:10.3390/d13050180

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

Smith, D. J., J. Y. Tan, M. A. Powers, X. N. Lin, T. W. Davis, and G. J. Dick. 2021. Individual Microcystis colonies harbour distinct bacterial communities that differ by Microcystis oligotype and with time. Environ. Microbiol. **23**: 3020–3036. doi:10.1111/1462-2920.15514

Stoeck, T., H. W. Breiner, S. Filker, V. Ostermaier, B. Kammerlander, and B. Sonntag. 2014. A morphogenetic survey on ciliate plankton from a mountain lake pinpoints the necessity of lineage-specific barcode markers in microbial ecology. Environ. Microbiol. **16**: 430–444. doi:10.1111/1462-2920.12194

Taberlet, P., E. Coissac, F. Pompanon, C. Brochmann, E. Willerslev. 2012. Towards next-generation biodiversity assessment using DNA metabarcoding. Mol. Ecol. **21**: 2045-2050. doi: 10.1111/j.1365-294X.2012.05470.x

Tapolczai, K., F. Keck, A. Bouchez, F. Rimet, M. Kahlert, and V. Vasselon. 2019. Diatom DNA Metabarcoding for Biomonitoring: Strategies to Avoid Major Taxonomical and Bioinformatical Biases Limiting Molecular Indices Capacities. Front. Ecol. Evol. **7**: 1–15. doi:10.3389/fevo.2019.00409

Taranu, Z. E., I. Gregory-Eaves, P. R. Leavitt, and others. 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. Ecol. Lett. **18**: 375–384. doi:10.1111/ele.12420

Tromas, N., Z. E. Taranu, B. D. Martin, A. Willis, N. Fortin, C. W. Greer, and B. J. Shapiro. 2018. Niche separation increases with genetic distance among bloom-forming cyanobacteria. Front. Microbiol. **9**: 1–12. doi:10.3389/fmicb.2018.00438

Utermöhl, H. 1958. Zur Vervollkommnung der quantitativen Phytoplankton-Methodik. SIL Commun. 1953-1996 **9**: 1–38. doi:10.1080/05384680.1958.11904091

Vasselon, V., I. Domaizon, F. Rimet, M. Kahlert, and A. Bouchez. 2017. Application of high-throughput sequencing (HTS) metabarcoding to diatom biomonitoring: Do DNA extraction methods matter? Freshw. Sci. **36**: 162–177. doi:10.1086/690649

Vuorio, K., A. Mäki, P. Salmi, S. L. Aalto, and M. Tiirola. 2020. Consistency of Targeted

Metatranscriptomics and Morphological Characterization of Phytoplankton Communities. Front. Microbiol. **11**. doi:10.3389/fmicb.2020.00096

Watson, S. B., E. McCauley, and J. A. Downing. 1997. Patterns in phytoplankton taxonomic composition across temperate lakes of differing nutrient status. Limnol. Oceanogr. **42**: 487–495. doi:10.4319/lo.1997.42.3.0487

Wetzel, R. G. 2001. Limnology: lake and river ecosystems. Gulf professional publishing. doi: 10.1016/C2009-0-02112-6.

Wiken, E. B., D. Gauthier, I. Marshall, K. Lawton, and H. Hirvonen. 1996. A perspective on Canada's ecosystems: An overview of the terrestrial and marine ecozones. Canadian Council on Ecological Areas.

Winter, J. G., A. M. Desellas, R. Fletcher, L. Heintsch, A. Morley, L. Nakamoto, and K. Utsumi. 2011. Algal blooms in Ontario, Canada: Increases in reports since 1994. Lake Reserv. Manag. **27**: 107–114. doi:10.1080/07438141.2011.557765

Woodhouse, J. N., A. S. Kinsela, R. N. Collins, L. C. Bowling, G. L. Honeyman, J. K. Holliday, and B. A. Neilan. 2016. Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. ISME J. **10**: 1337–1351. doi:10.1038/ismej.2015.218

Xiao, X., H. Sogge, K. Lagesen, A. Tooming-Klunderud, K. S. Jakobsen, and T. Rohrlack. 2014. Use of high throughput sequencing and light microscopy show contrasting results in a study of phytoplankton occurrence in a freshwater environment. PLoS One **9**. doi:10.1371/journal.pone.0106510

Zimmermann, J., G. Glöckner, R. Jahn, N. Enke, and B. Gemeinholzer. 2015. Metabarcoding vs. morphological identification to assess diatom diversity in environmental studies. Mol. Ecol. Resour. **15**: 526–542. doi:10.1111/1755-0998.12336

Tables Chapter II

Table 1. Comparison of the ten most common taxonomic complexes and ungrouped genera detected with microscopy (biomass and density) and DNA metabarcoding (reads) across all 379 lakes.

Microscopy biomass	(%)	Microscopy density	(%)	16S rRNA sequences	(%)
Aphanizomenon C.	64.2	*Microcystis*	53.8	*Synechococcus C.*	41.2
Dolichospermum C.	12.9	*Aphanizomenon C.*	16.8	*Planktothrix C.*	17.4
Microcystis	9.38	*Synechococcus C.*	8.88	*Aphanizomenon C.*	11.6
Planktothrix C.	8.54	*Dolichospermum C.*	6.95	*Snowella C.*	7.88
Lyngbya C.	3.33	*Snowella C.*	5.81	*Microcystis*	7.60
Synechococcus C.	0.48	*Lyngbya C.*	3.91	*Dolichospermum C.*	5.80
Nodularia	0.44	*Planktothrix C.*	1.70	*Pseudanabaena C.*	4.05
Chroococcus C.	0.22	*Pseudanabaena C.*	1.10	*Lyngbya C.*	1.98
Pseudanabaena C.	0.19	*Romeria*	0.45	*Nodularia*	0.37
Snowella C.	0.14	*Chroococcus C.*	0.24	*Rhabdogloea C.*	0.26

C. indicates genus complex.

Figures Chapter II

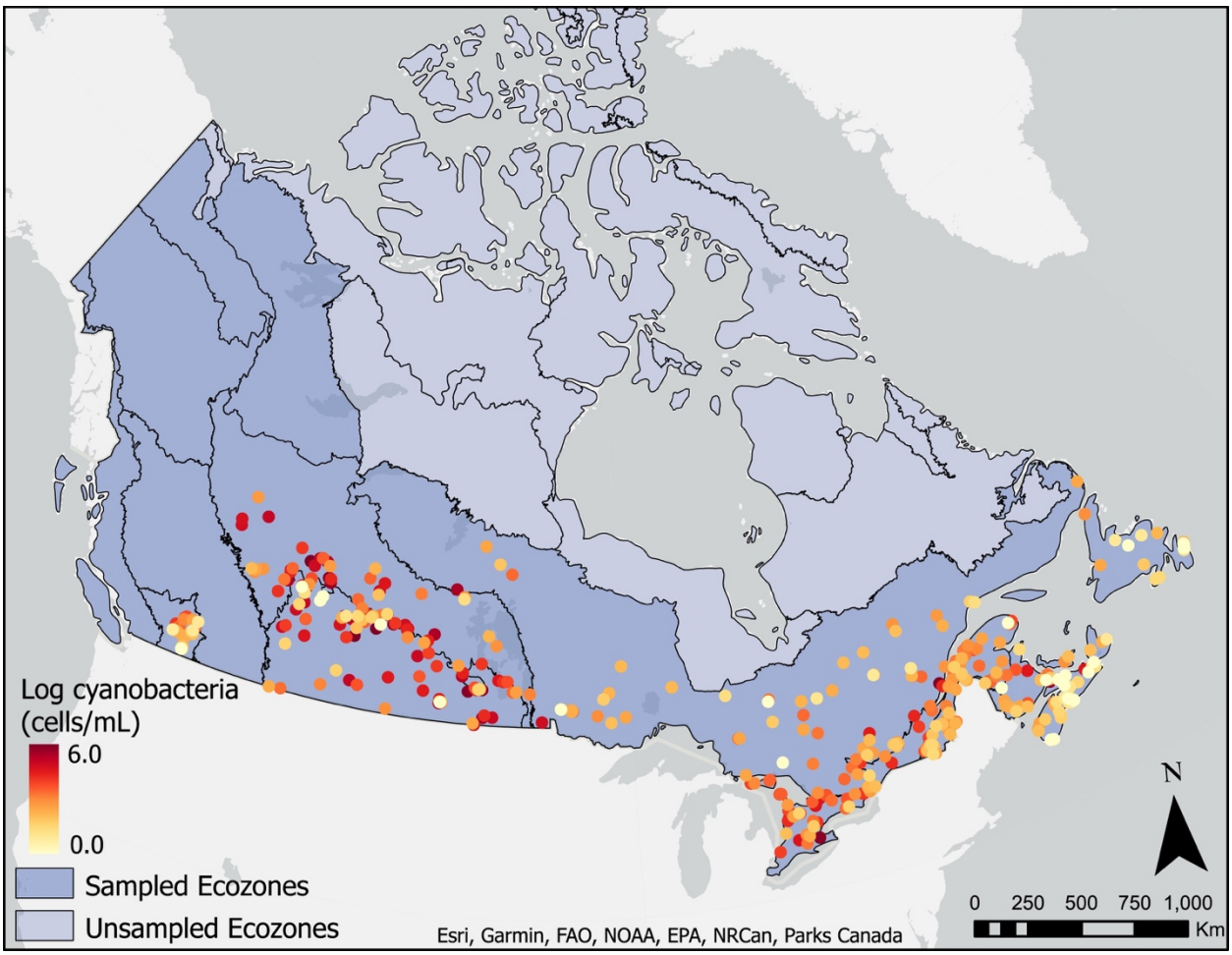

Figure 1. Distribution of log transformed cyanobacteria density (cells/mL) across 379 sites.

Water samples for microscopy and DNA metabarcoding analyses were collected at each site.

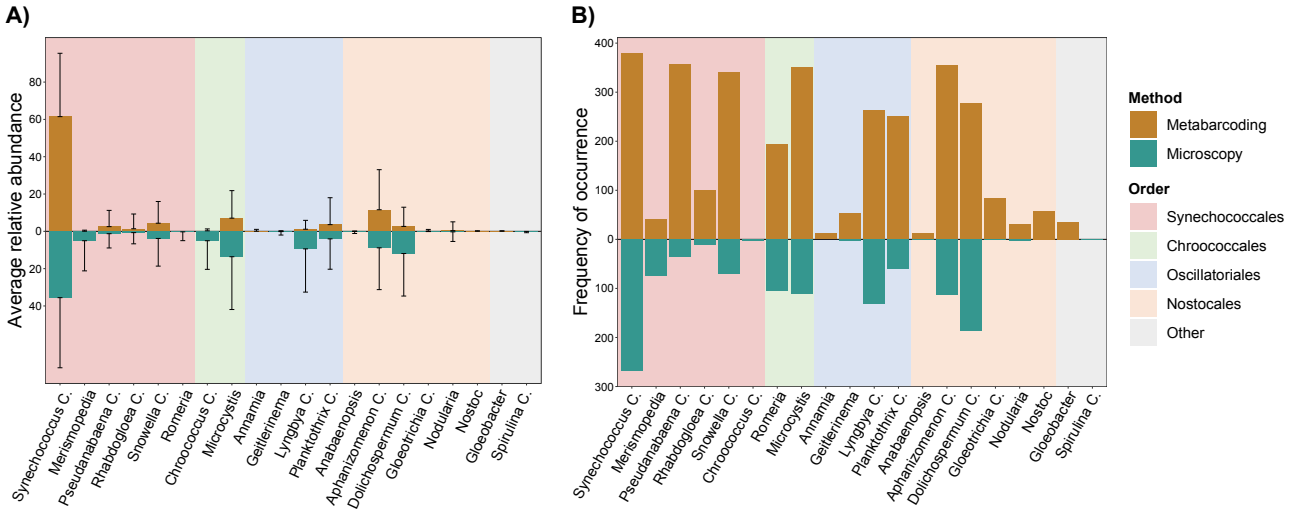

Figure 2. Average relative abundance (A) and frequency of occurrence (i.e., the number of lakes in which each taxon was observed) (B) for complexes and ungrouped genera detected by DNA metabarcoding and microscopy. Standard error bars are included for the relative abundances of each group.

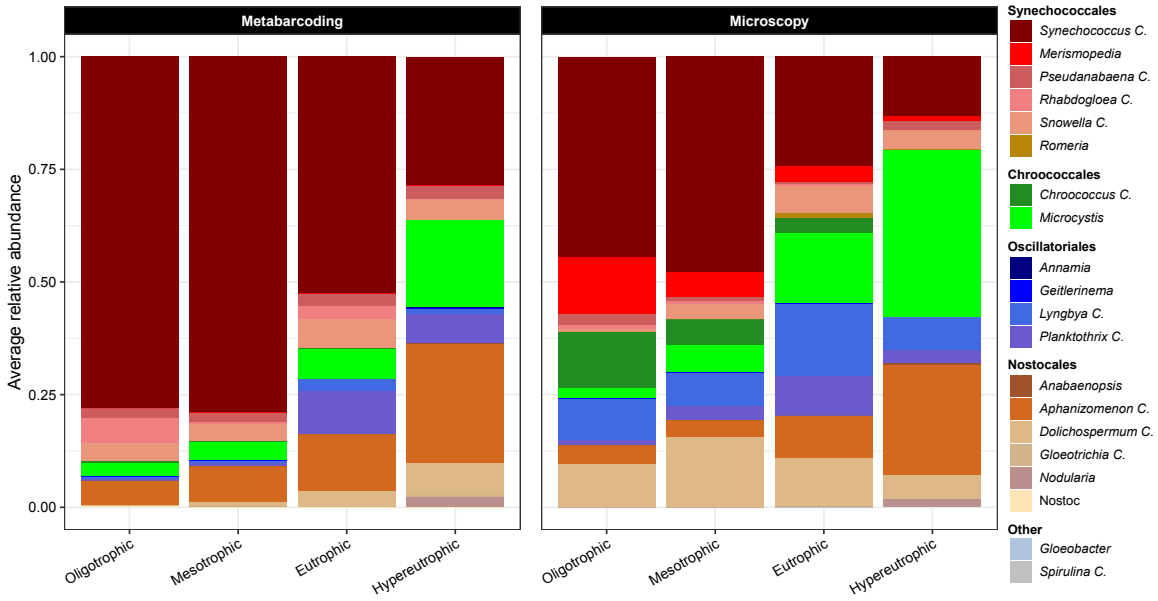

Figure 3. Average relative abundance of cyanobacteria complexes and ungrouped genera across trophic states. Number of reads were used for DNA metabarcoding (left panel), and cell abundance was used for microscopy (right panel).

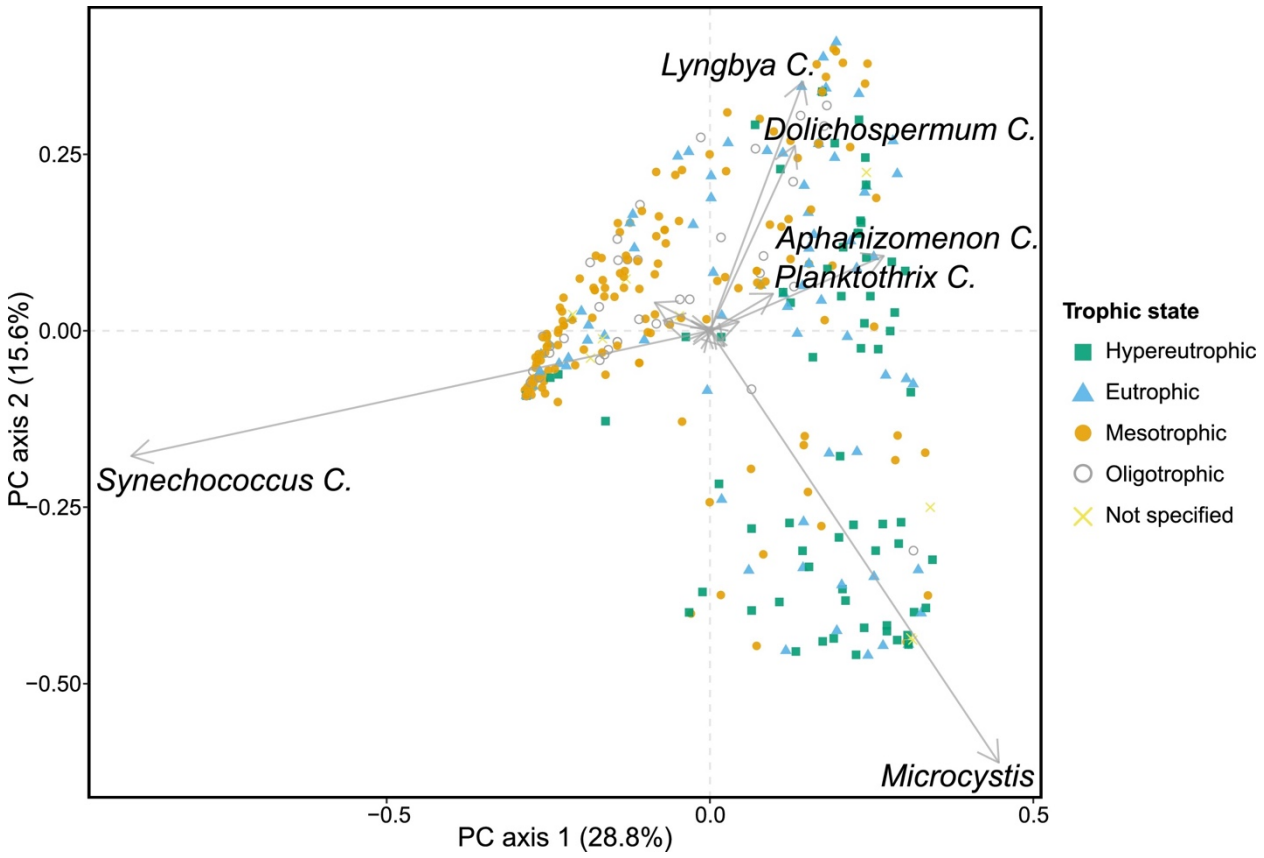

Figure 4. PCA biplot of the cyanobacteria community identified by microscopy (cells/mL). Genus complexes and ungrouped genera abundances were Hellinger transformed prior to the ordination.

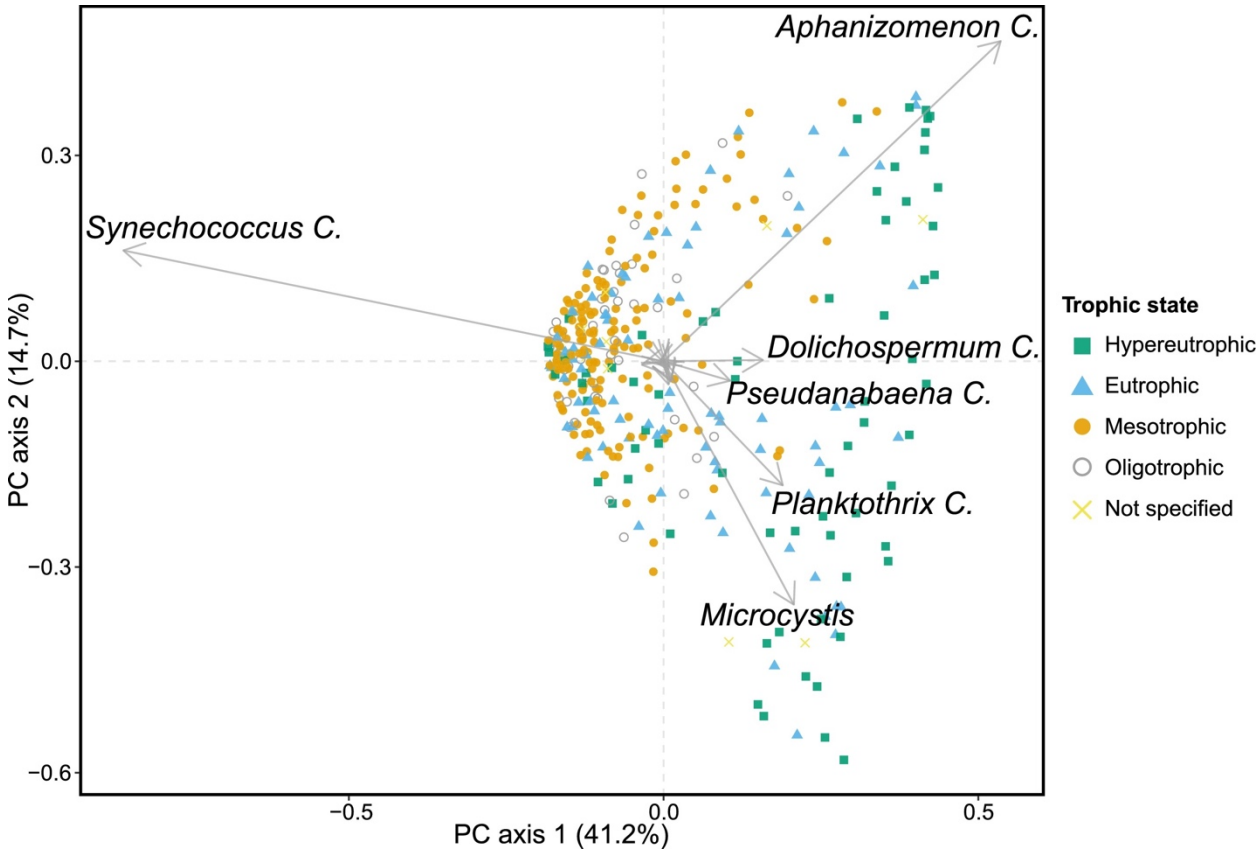

Figure 5. PCA biplot of the cyanobacteria community identified by DNA metabarcoding (reads). Genus complexes and ungrouped genera reads were Hellinger transformed prior to the ordination.

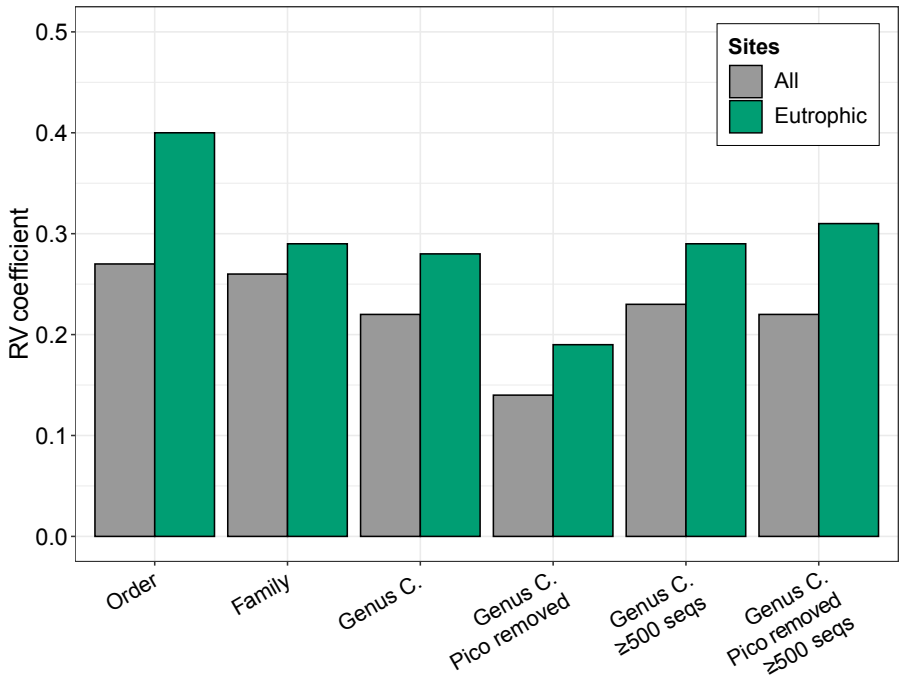

Figure 6. RV coefficients between microscopy (density- cells/mL) and DNA metabarcoding (reads) approaches on PCA site scores. Grey bars represent matrix comparisons using all 379 sites, and green bars represent comparisons between just eutrophic sites. ≥500 refers to comparisons done only on sites with at least 500 sequences. All RV coefficients were significant ($p < 0.0001$).

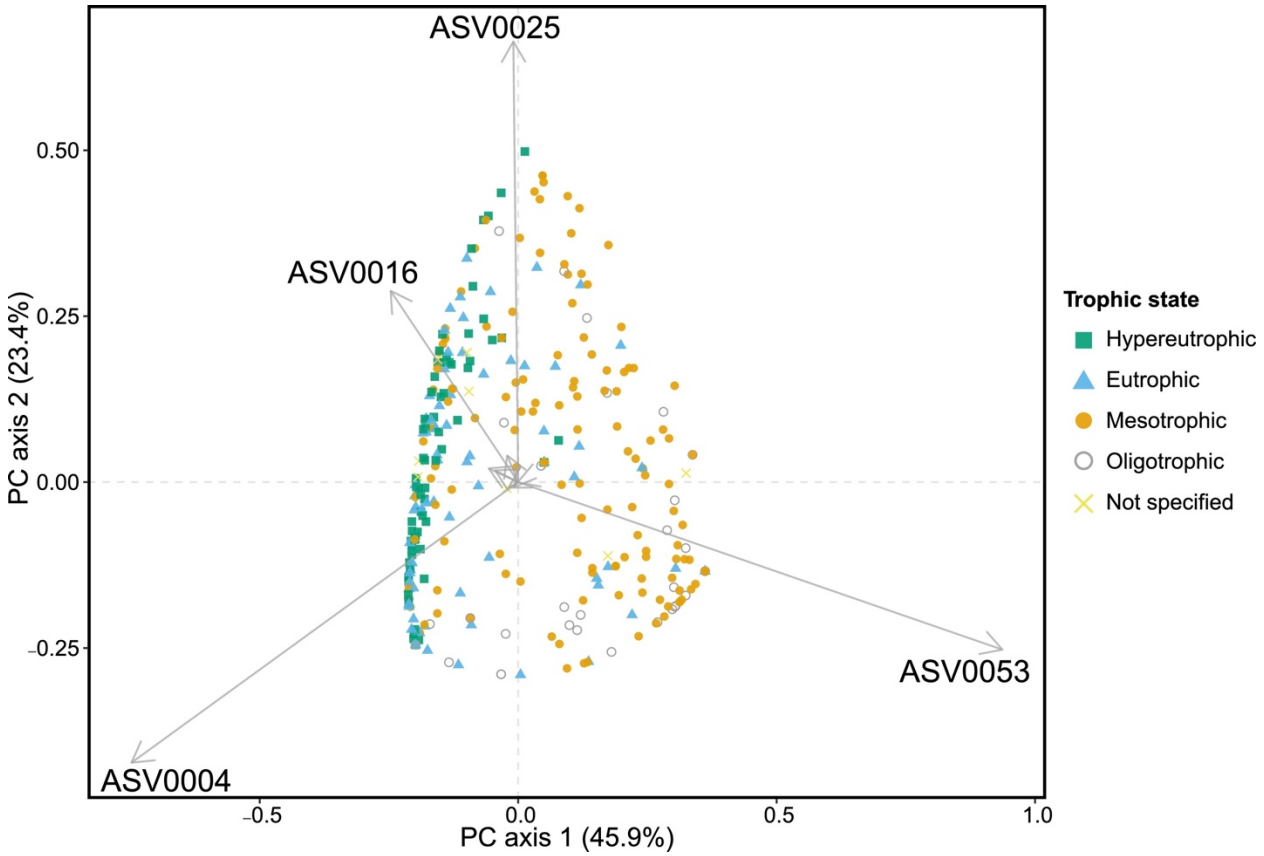

Figure 7. PCA biplot of the ASVs assigned to *Microcystis* using DNA metabarcoding (reads). Genus complexes and ungrouped genera reads were Hellinger transformed prior to the ordination.

Appendices Chapter II

Table S1. List of genera included in each complex.

Genus or complex name	
Synechococcus C.	*Cyanobium, Synechococcus, Synechocystis, Cyanothece, Aphanocapsa, Aphanothece*
Merismopedia	*Merismopedia*
Pseudanabaena C.	*Pseudanabaena, Prochlorothrix*
Rhabdogloea C.	*Rhabdogloea, Rhabdoderma*
Snowella C.	*Snowella, Gomphosphaeria, Woronichinia, Coelosphaerium*
Romeria	*Romeria*
Chroococcus C.	*Chroococcus, Limnococcus*
Microcystis	*Microcystis*
Annamia	*Annamia*
Geitlerinema	*Geitlerinema*
Lyngbya C.	*Lyngbya, Leptolyngbya, Limnolyngbya, Limnothrix, Nodosilinea, Planktolyngbya*
Planktothrix C.	*Planktothrix, Oscillatoria, Planktothricoides, Phormidium, Tychonema*
Anabaenopsis	*Anabaenopsis*
Aphanizomenon C.	*Aphanizomenon, Cuspidothrix*
Dolichospermum C.	*Dolichospermum, Anabaena, Chrysosporum*
Gloeotrichia C.	*Gloeotrichia, Calothrix*
Nodularia	*Nodularia*
Nostoc	*Nostoc*
Gloeobacter	*Gloeobacter*
Spirulina C.	*Arthrospira, Spirulina*

C. indicates genus complex. Only genera that reached a 2% relative abundance in one lake from either dataset is listed.

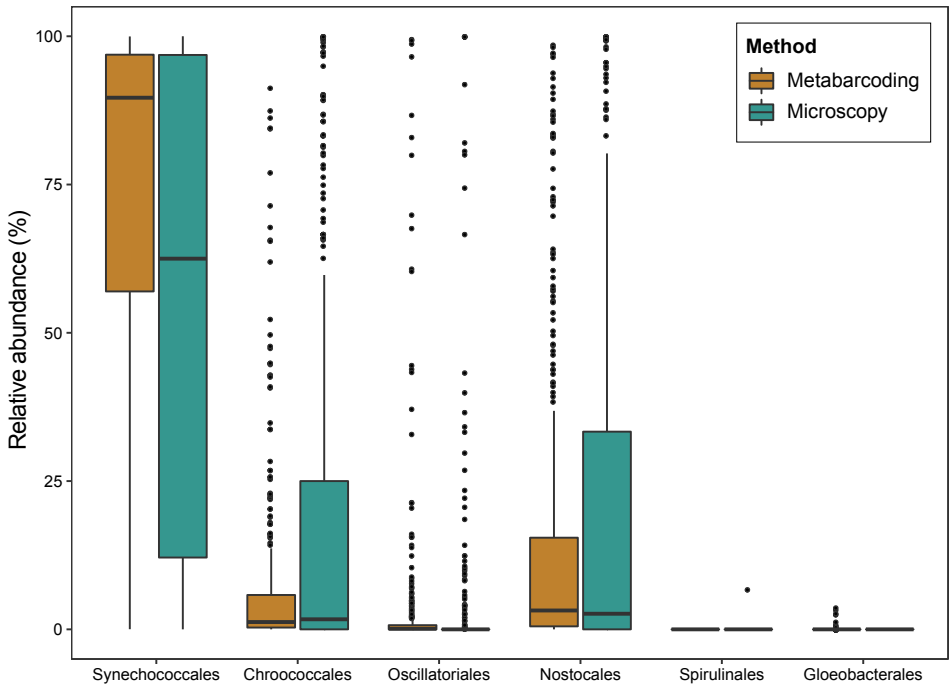

Figure S1. Relative abundances of cyanobacteria grouped at Order level analyzed by DNA metabarcoding (reads) and microscopy (cells/mL). Boxplots includes the median, standard deviation, and 25 and 75% percentiles over the whole set of 379 lake samples.

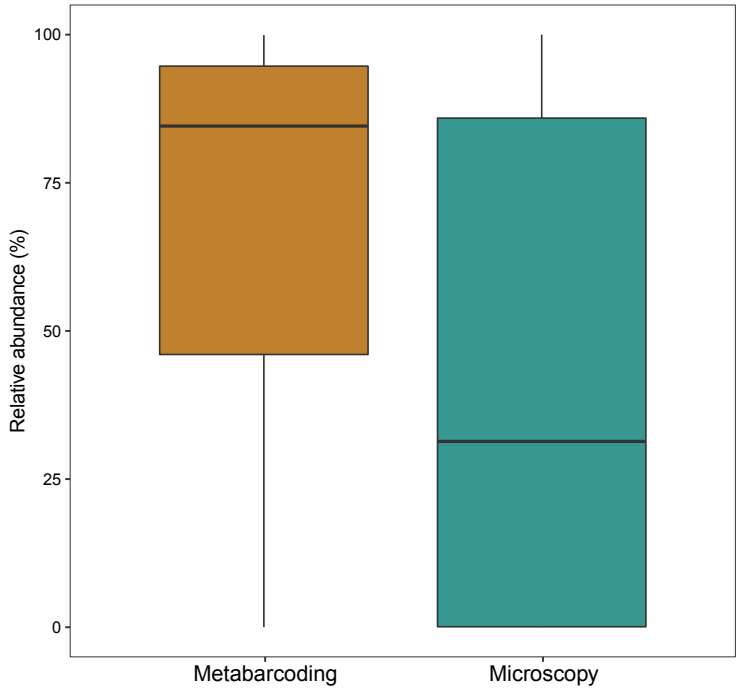

Figure S2. Relative abundances of picocyanobacteria assigned by DNA metabarcoding (reads) and microscopy (cells/mL). Boxplots includes the median, standard deviation, and 25 and 75% percentiles over the whole set of 379 lake samples.

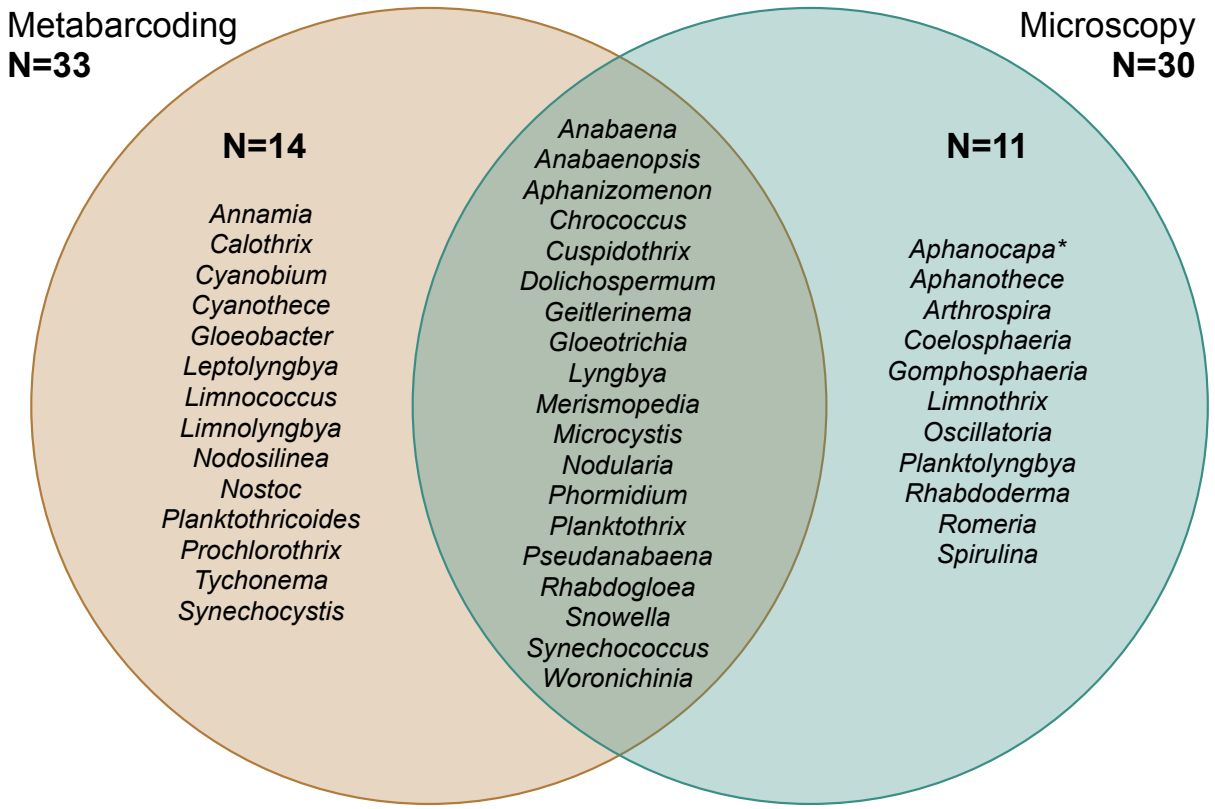

Figure S3. Venn diagram showing which genera were uniquely detected and co-detected between methods. Only genera that reached 2% relative abundance in at least one lake are included. The asterisk(*) indicates that *Synechocystis* was counted with *Aphanocapsa*; it was not identified as its own genus in the microscopy dataset.

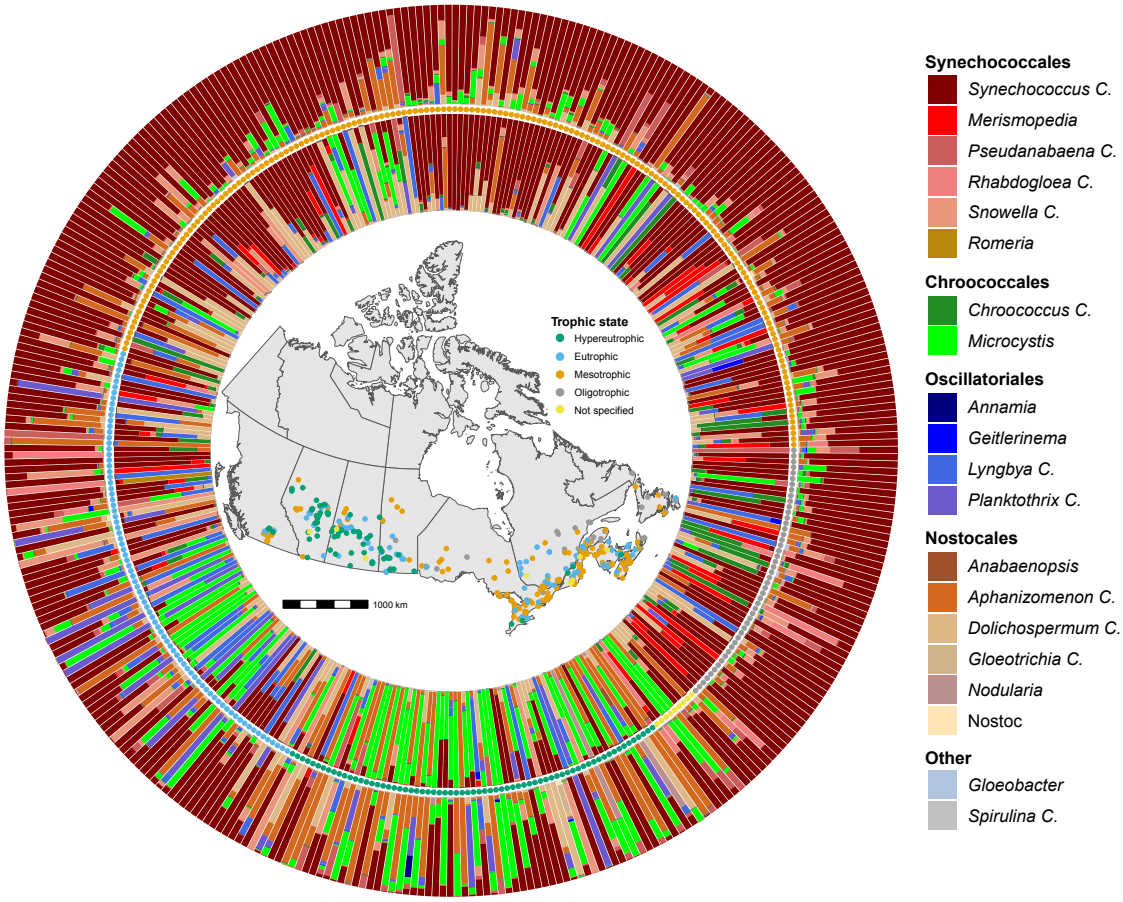

Figure S4. Taxonomic composition of cyanobacteria assemblages from each of the sampled 379 lakes. Each bar represents the relative abundance of cyanobacteria genus complexes and ungrouped genera. The inner ring is comprised of each community generated by microscopy (cells/mL). The outer ring is comprised of each community generated by DNA metabarcoding (reads). Lakes are featured on the centered map and colored according to trophic state.

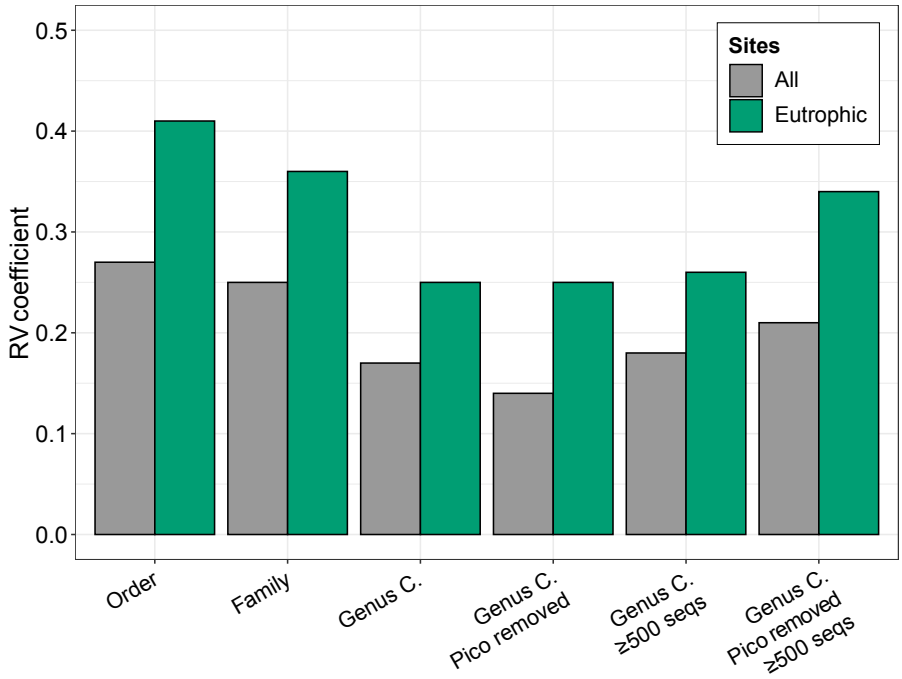

Figure S5. RV coefficients between microscopy (biomass- µg/L) and DNA metabarcoding (reads) approaches on PCA site scores. Grey bars represent matrix comparisons using all 379 sites, and green bars represent comparisons between just eutrophic sites. ≥500 refers to comparisons done only between sites with at least 500 sequences.

Connecting statement between Chapters II and III

The comparison developed in chapter II identified a moderate level of congruence between microscopy and DNA metabarcoding. In particular, the cyanobacteria communities generated by each method were more comparable at coarser taxonomic levels (i.e., Order), with correlation coefficients peaking at approximately 40%. Both methods also became more similar along a trophic gradient, with greater congruency apparent when comparing lakes of higher nutrient state. This finding is especially relevant to researchers and managers monitoring eutrophication and resulting blooms, as it shows both DNA metabarcoding and microscopy may be applied to accurately identify community structure under the most eutrophic conditions. Where the methods differed more significantly was under oligotrophic and mesotrophic conditions. In these lakes, the composition of cyanobacteria generated from DNA metabarcoding was dominated by picocyanobacteria; a size fraction of cells that can be difficult to isolate and morphologically distinguish using inverted microscopy of Lugol's preserved samples. This highlighted the complementary rather than identical information DNA can provide to morphological methods. It is possible that congruency among approaches will improve in the future if: a) fluorescence microscopy is used as an additional method to more fully capture the picocyanobacterial community and b) reference databases of cyanobacterial sequences become more fully populated.

The datasets generated in chapter I and II showcased how cyanobacteria communities vary across Canadian lakes, and how the two methods of taxonomic assignment agree more closely in lakes of high nutrient state. In these eutrophic lakes, both microscopy and DNA metabarcoding showed increased relative abundances of several potential toxin-producing

cyanobacteria, particularly, *Microcystis*, *Aphanizomenon*, *Dolichospermum* and *Planktothrix*. Many of these taxa represented the majority of cyanobacteria biomass and were the dominant genera in various lakes across Canada. The third chapter in my thesis examined another key issue relevant for lake management: predicting the occurrence and concentration of cyanotoxins. Specifically, this chapter investigated the geographic distribution and predictors of microcystins, one of the most widespread group of cyanotoxins. Using a subset of 440-lakes, I quantified the concentration of total microcystins and developed empirical models that consider an exhaustive range of biotic and abiotic predictors. Chapter III also quantified the abundance of microcystin congeners and identified the variables that correlated to their respective concentrations. Given that congeners vary considerably in toxicity and persistence, and that there is very limited information on their abundances across temperate to subarctic lakes there is a need to gain a greater understanding of the factors that may influence their production.

CHAPTER III

Microcystin concentrations and congener composition in relation to environmental variables across 440 north-temperate and boreal lakes

Paul W. MacKeigan[1,2], Arthur Zastepa[3], Zofia E. Taranu[4], Judy A. Westrick[5], Anqi Liang[3], Frances R. Pick[6], Beatrix E. Beisner[2,7] and Irene Gregory-Eaves[1,2]

[1]Department of Biology, McGill University, Montreal, Quebec, H3A 1B1, Canada;

[2]Interuniversity Research Group in Limnology (GRIL), Quebec, Canada;

[3]Environment and Climate Change Canada, Canada Centre for Inland Waters, Burlington, Ontario, L7S 1A1, Canada;

[4]Aquatic Contaminants Research Division, Environment and Climate Change Canada, Montreal, Quebec, H2Y 2E7, Canada;

[5]Department of Chemistry, Wayne State University, Detroit, Michigan, 48202, United States;

[6]Department of Biology, University of Ottawa, Ottawa, Ontario, K1N 9A7, Canada;

[7]Department of Biological Sciences, University of Quebec at Montreal, Montreal, Quebec, H2X 1Y4, Canada

Manuscript published in *Science of the Total Environment* in April 2023

Abstract

Understanding the environmental conditions and taxa that promote the occurrence of cyanobacterial toxins is imperative for effective management of lake ecosystems. Herein, we modeled total microcystin presence and concentrations with a broad suite of environmental predictors and cyanobacteria community data collected across 440 Canadian lakes using standardized methods. We also conducted a focused analysis targeting 14 microcystin congeners across 190 lakes, to examine how abiotic and biotic factors influence their relative proportions. Microcystins were detected in 30% of lakes, with the highest total concentrations occurring in the most eutrophic lakes located in ecozones of central Canada. The two most commonly detected congeners were MC-LR (61% of lakes) and MC-LA (37% of lakes), while 11 others were detected more sporadically across waterbodies. Congener diversity peaked in central Canada where cyanobacteria biomass was highest. Using a zero-altered hurdle model, the probability of detecting microcystin was best explained by increasing *Microcystis* biomass, *Daphnia* and cyclopoid biomass, soluble reactive phosphorus, pH and wind. Microcystin concentrations increased with the biomass of *Microcystis* and other less dominant cyanobacteria taxa, as well as total phosphorus, cyclopoid copepod biomass, dissolved inorganic carbon and water temperature. Collectively, these models accounted for 34% and 70% of the variability, respectively. Based on a multiple factor analysis of microcystin congeners, cyanobacteria community data, environmental and zooplankton data, we found that the relative abundance of most congeners varied according to trophic state and were related to a combination of cyanobacteria genera biomasses and environmental variables.

Introduction

The severity of cyanobacterial blooms is projected to increase in some waterbodies worldwide throughout the 21st century (Chapra et al. 2017; Kakouei et al. 2021). Increases in the frequency and duration of blooms in inland waters have been linked to a variety of environmental drivers including, but not restricted to, eutrophication (Pick and Lean 1987; Schindler et al. 2008) and increasing temperature (Kosten et al. 2012; Paerl and Paul 2012). A major concern is that many cyanobacteria can produce an array of toxins and other bioactive metabolites (collectively referred to here as cyanotoxins) that can be harmful to humans and wildlife.

Microcystins (MCs) are the most prevalent and commonly measured form of cyanotoxins, in Canada and worldwide (Kotak and Zurawell 2007; Orihel et al. 2012). MCs are cyclic heptapeptides with more than 275 congeners identified to date (Bouaïcha et al. 2019; Spoof and Catherine 2017). They vary in structure and toxicity, and are produced by a growing list of planktonic and benthic cyanobacteria genera distributed globally (Chernoff et al. 2020; Chorus and Welker 2021). Toxigenic strains arise in several of the common bloom-forming genera such as *Microcystis*, *Dolichospermum* and *Planktothrix*, as well as some picocyanobacteria (e.g., *Aphanocapsa* and *Synechococcus)* (Bernard et al. 2017; Chorus and Welker 2021). Although MCs are produced and stored within cells, they can leak into surrounding water following senescence and cell lysis (McKindles et al. 2020). Within and outside cells, MCs are relatively stable, resistant to chemical hydrolysis and oxidation at neutral pH, and can persist in the water weeks after the disappearance of a bloom (Zastepa et al. 2014; U.S. EPA 2019). Within the environment, MCs can bioaccumulate in food webs (Flores et al. 2018; Kozlowsky-Suzuki et al. 2012) and are stored in lake sediments (Zastepa et al. 2015).

Human health implications arise via several exposure routes including dermal contact or ingestion from contaminated drinking and recreational waters, inhalation of atmospheric aerosols (Plaas and Paerl 2021), or consumption of contaminated food such as fish and algal dietary supplements (Jia et al. 2014; Miller and Russell 2017; Roy-Lachapelle et al. 2017). To reduce human exposure, recreational and drinking water regulatory (1.0 μg/L) and advisory concentrations (recreational thresholds at 2-4, 5-20, >20 μg/L) have been created by the World Health Organization (WHO) as well as many countries and jurisdictions.

The distribution of toxigenic strains and occurrence of MCs are associated with eutrophic conditions but are also influenced by numerous physical, chemical and biological factors (reviewed in Dai et al. 2016). Several large-scale surveys and meta-analyses have identified significant correlations between MCs and total phosphorus (TP) (Kotak et al. 2000; Orihel et al. 2012; Scott et al. 2013), dissolved inorganic and total nitrogen (TN) (Buley et al. 2021; Dolman et al. 2012; Giani et al. 2005), temperature (Mowe et al. 2015), pH (Buley et al. 2021), dissolved organic carbon (DOC) (Beaver et al. 2014), and watershed morphology (Hayes and Vanni 2018). Others have also found positive relationships between MCs and zooplankton, including cladocerans and copepods, likely due to the presence of other edible algae or a tolerance by some zooplankton to toxic cyanobacteria (Wang et al. 2022). Despite the ubiquity of MCs in large-scale datasets (Loftin et al. 2016), and the interest in modeling MCs as a target metric of water quality, uncertainties remain regarding the relative importance of MC predictors in field studies. Previous studies using large-scale datasets have often been limited to few predictors and have showed mixed results. These discrepancies may be the result of blooms being highly dynamic, and a lack of standardized sampling design and analytic techniques among studies (Buley et al. 2021; Tillmanns and Pick 2011). Herein, we have developed a large dataset that includes

measurements of both total microcystins and specific congeners that vary in their toxicities. We considered a broader range of standardized predictors than previous studies to advance our understanding of the patterns and drivers at a continental scale.

While total MC concentrations have been included in many local and several national-scale lake management programs (e.g., National Lakes Assessment (NLA) and the European Multi-Lake Sampling programs), analyses of the distribution and drivers of MC congeners remain overlooked at large scales. Moreover, although the number of identified congeners has increased in recent years, attention has focused primarily on MC-LR and -RR (Díez-Quijada et al. 2019). Thus far, congener composition has been associated with temperature (Mantzouki et al., 2018), nitrogen forms (Monchamp et al. 2014), as well as climate and nutrient conditions (Taranu et al. 2019). Such environmental predictors have also been related to changes in cyanobacteria community structure, but in much smaller datasets (Monchamp et al. 2014). Total MCs are commonly measured using enzyme-linked immunosorbent-based assays (ELISA) which targets the ADDA-moiety specific to microcystins and their congeners, although this technique is unable to distinguish congeners. Given the growing list of identified congeners, their range in toxicities and degradation rates, the ELISA approach may mis-represent the potential risk posed by MCs in a particular waterbody (Chernoff et al. 2020; Zastepa et al. 2014).

Based on the results from other large-scale surveys, we hypothesized that cyanobacteria biomass, nutrient concentrations (primarily TN), and temperature would be the strongest predictors of total microcystin concentration (Taranu et al. 2017; Yuan et al. 2014). We also hypothesized that congener relative abundances would be strongly correlated to cyanobacteria community composition and nutrient variables, mainly TN (Monchamp et al. 2014; Taranu et al. 2019). Overall, this study is an essential resource for understanding the conditions that promote

cyanotoxin production across many lakes and ecozones in an era of accelerated environmental change.

Materials and methods

Lake selection

As part of the NSERC Canadian LakePulse Network (Huot et al. 2019), we focused on 440 lakes sampled once across two summers (2018, 2019) following a standardized protocol (NSERC Canadian Lake Pulse Network, 2021). Lakes were selected from twelve ecozones (regions defined by unique climate, geology and vegetation; Wiken et al. 1996) following a stratified random block design, using three lake sizes (0.1-1km^2, 1-10km^2, 10-100km^2) and three watershed human impact categories (low, medium, high) as the stratification groups (Huot et al. 2019). Additionally, lakes were supposed to have a minimum depth of 1 m and be located within 1 km of road access. To minimize seasonal effects, sampling occurred between the end of June to the beginning of September, a time which corresponds to the period of maximal thermal lake stratification.

Sample collection and processing

From each lake, a suite of physiographic, water quality, land use, plankton and climate variables were collected and categorized into thematic groups (Table S1). Full descriptions of methods are found in Huot et al. (2019) and the LakePulse field manual (NSERC Canadian Lake Pulse Network, 2021). Briefly, physiographic variables such as lake depth, watershed size, slope and residence time were obtained from HydroLAKES v. 1.0 (Messager et al. 2016). The percentages of different land use categories (agriculture, forestry, mines, natural landscape, pasture, urban and water) were estimated for the watershed of each lake (Huot et al. 2019).

Climate variables were calculated as averages or sums of 7 and 30 days before each sampling. These data were accessed from ERA5-Land hourly data (Muñoz Sabater 2019).

The deepest point of each lake was selected to collect surface water for water quality, plankton and microcystin analyses. Here, an integrated sample representing the euphotic zone, defined as twice the Secchi disk depth to a maximum of 2-m, was taken using and acid-washed integrated tube sampler. Water samples were used to measure a variety of water quality parameters. In brief, TP was analyzed using a standard protocol at the Université du Québec à Montréal (UQAM) (Wetzel and Likens 2000) using molybdenum-blue method following potassium persulfate digestion. Similarly, water samples for SRP were filtered through 0.45 µm filters in the field and then followed the TP analytical protocol but without the first oxidation step. TN was analyzed using an OI Analytic Flow Solution 3100, following a potassium persulfate digestion, coupled with a cadmium reactor. Ions were filtered in the field through 0.45 µm filters and measured by the Biogeochemical Analytical Service Laboratory at the University of Alberta following two protocols. Anions were determined by ion chromatography using a Dionex DX-600 and followed the US EPA Method 300.1. Cations were measured using an inductively coupled argon plasma optical emission spectrometer following the US EPA Method 200.7. DIC and DOC were also filtered in the field through 0.45 µm filters and later analyzed with an OI Analytical Aurora 1030W TOC Analyzer using a persulfate oxidation method at UQAM.

In addition to this integrated surface-water sample, full water-column profiles for temperature, pH, dissolved oxygen, and specific conductivity were measured using a multiparameter RBR logger. Using a 100 µm mesh Wisconsin net, an integrated water-column zooplankton sample was collected. Zooplankton samples were narcotized with CO_2 and

preserved in ethanol (70% final concentration) immediately after collection. Samples were identified and biomass was estimated by BSA Environmental Services laboratory (Ohio, U.S.A.; further details on zooplankton sampling provided in Paquette et al. (2021)).

Phytoplankton counting and identification

Phytoplankton samples were collected from the sampling site of each lake. A 120 mL subsample of surface water was fixed and preserved in acidic Lugol's iodine solution. Samples were identified and biomass was calculated by a sole taxonomist following Utermöhl's sedimentation method (Lund et al. 1958) using an inverted microscope. Taxonomic names were provided to species level. Further details regarding the phytoplankton identification protocol are available in MacKeigan et al. (2022).

Total microcystin and congener sample processing

For all 440 lakes included in this analysis, a 250 mL subsample of surface water was stored in a clean amber plastic bottle, immediately frozen and then shipped for laboratory processing. Total MCs were determined by enzyme-linked immunosorbent assay (ELISA). First, samples were lysed through three freeze-thaw cycles, filtered through 0.2 µm syringe filters, then analyzed on ELISA kits acquired from Abraxis, LLC (duplicates were carried out on 10% of lakes). The ELISA detection limit was 0.1 µg MC-LR/L.

Based on the ELISA results, a subset of 190 lakes were selected to measure the MC congener profiles. This subset included all lakes which had a detectable amount of MC from ELISA or had high cyanobacteria biomasses. We also randomly selected an additional 10% of lakes from the remaining 250 without detectable MC concentrations. The full analytical protocol for measuring the congeners is described in Zastepa et al. (2023). In brief, fourteen MC congeners targeted: MC-LR, MC-RR, MC-YR, MC-LA, MC-LY, MC-LW, MC-LF, MC-WR,

MC-HtyR, MC-HilR, D-Asp3-MC-RR, D-Asp3-MC-LR, Dha7-MC-LR, and Leu1-MC-LR. Microcystin standards were purchased from Enzo Life Sciences (Farmingdale, NY, USA), GreenWater Laboratories (Palatka, FL) and National Research Council Canada (Ottawa, Ontario, Canada). Metabolites were first extracted using a four-step process. Microcystins were measured using a Thermo Scientific TSQ Altis™ triple quadrupole mass spectrometer (Waltham, MA, USA) with a TriPlus™ RSH EQuan 850 system. Congeners were separated in under 6 minutes. These measurements were conducted at Wayne State University (Detroit, MI, USA).

Data analysis

A series of modeling approaches were used to identify the most parsimonious set of predictors of total microcystins and MC congener composition. In all modeling approaches, we considered ~50 potential predictors including physiographic, water chemistry, land use, zooplankton, climate and cyanobacteria taxonomic data. Our response variables were the concentration of total MC or the full MC congener matrix. These analyses were performed using 435 out of the 440 lakes in the dataset, as five lakes did not have phytoplankton count data.

All statistical analyses were conducted in R version 4.1.0 (R Core Team, 2021). For modeling total MC concentrations, we followed the data preparation and exploratory analyses outlined in Feld et al. (2016) to reduce the number of predictors to a robust subset for downstream quantitative analyses. First, missing values (NA) were replaced by medians from their respective ecozones. Variables were then $\log_{10}$, square root, or logit (used for land use percentage data; Fox and Weisberg 2011) transformed to normalize their distribution. For predictors with zero values, including cyanobacteria (total and individual genera) and zooplankton (Cyclopoid, Calanoid, Cladoceran, Daphnid, Chydorid, Bosminid, and small cladocerans), a small constant (X+1) was added to offset their distribution. Since we were

interested in potential direct relationships between MCs and the three most common genera in terms of biomass (i.e., *Microcystis*, *Dolichospermum* and *Aphanizomenon*), we included the common genera as individual predictors and considered the remaining cyanobacteria biomass (CBB_{other}) as a fourth biomass variable. Collinearity among predictor variables was then assessed using Variance Inflation Factors (VIFs) (*vifstep* function in the *usdm* package; Naimi 2015), setting a VIF threshold of 10 or higher to indicate high collinearity (Borcard et al. 2018). After removing collinear variables, random forest (RF) models were constructed to rank the importance of the remaining 45 candidate predictors. RFs were fitted using the *cforest* function using the *party* package (Hothorn et al. 2021). While considering the potential mechanistic links between MC concentrations and highly ranked predictors, the top 20 predictors identified by the RF were selected for ensuing quantitative modeling analyses (Feld et al. 2016).

To account for the high number of non-detects in the MC data (70% of the dataset), we employed a zero-altered hurdle model approach, whereby the presence-absence of total MC is modeled first in a binomial generalized linear model (GLM), then the concentrations from lakes with a detectable amount of MC (30% of lakes) is modeled in a GLM fit with a gamma distribution (Zuur and Ieno, 2016). This technique was first used to model MC concentrations by Taranu et al. (2017) using data from the NLA in the USA, which contained a similar distribution of MC detects (present in 32% of lakes). This model type was justified as ecologically relevant for this distribution of data as the predictors determining MC presence may differ from those predicting its increase. It also allows the retention of a large number of observations and reduces biased estimates of standard errors potentially introduced when ignoring the high number of non-detects (Taranu et al. 2017; Zuur and Ieno 2016). For each component of the two-part model, the most parsimonious predictors were fitted following a stepwise selection of variables using

Bayesian Information Criterion (BIC) values and $p < 0.05$ as the inclusion criterion. Only the top 20 transformed environmental predictors identified in the RF analysis were considered in each part of the hurdle model.

For the multivariate MC congener data (n=190), we examined which environmental and cyanobacterial variables were best correlated with composition using a multiple factor analysis (MFA). The MFA is a correlative analysis that computes a principal component analysis (PCA) for each matrix. The PCAs are then centered and weighted according to the first eigenvalue and presented in one ordination plot to visualize relationships among three or more multivariate data matrices (Borcard et al. 2018). The MFAs were performed using the *FactomineR* package in R (Lê et al. 2008). The matrices considered were the MC congener composition, environmental variables, zooplankton groups and the cyanobacteria community assemblage. Prior to running the MFA, congener concentrations were converted to relative abundances, environmental and zooplankton variables were transformed, centered and scaled, and the cyanobacteria community biomass matrix was Hellinger transformed using the *vegan* package (Oksanen et al. 2019). To provide a bridge to the multivariate hurdle model on total MCs, a second MFA was run using only the lakes which had at least one congener detected (n=123). Lastly, to quantify and explain the unique and shared portions of variation explained among predictor groups, we ran a variation partitioning analysis using the *varpart()* function of the *vegan* package. This was performed on both 190- and 123- lake datasets.

Results

Cyanobacteria community structure

We captured wide gradients in cyanobacteria biomass and water chemistry across the full collection of 440 lakes (Table S1 and S2). Cyanobacteria taxa were detected in 92% of lakes, but across lakes we detected considerable variation in biomass (median = 0.23 mg/L, maximum = 681.5 mg/L) (Fig. S1A and B). On average, cyanobacteria represented 40% of total phytoplankton biomass across all sites. In total, 31 genera were identified by microscopy. *Aphanocapsa* was the most pervasive genus (occurring in 52% of lakes), followed by *Dolichospermum* (48%), *Microcystis* (48%), *Limnothrix* (35%) and *Aphanizomenon* (32%) (Table S2). Cyanobacterial biomass was highest in the Prairies and Boreal Plains ecozones of central Canada where bloom-forming taxa predominated (e.g., *Aphanizomenon*, *Dolichospermum*, *Microcystis* and *Planktothrix*). Lakes in the Prairies and Boreal Plains also had elevated TP concentrations (medians of 204 µg/L and 76 µg/L, respectively) relative to the national median (20 µg/L TP).

Modelling total microcystins across Canada

MCs were detected in 30% (n=130) of lakes sampled across Canada and in all 10 ecozones considered (Fig. 1A). The highest concentration measured by ELISA was 31.6 µg/L but often concentrations remained low (i.e., below drinking water guidelines). Only 10% of samples had concentrations above the WHO drinking water guideline (1 µg/L), and 8.6% exceeded the Canadian drinking water guideline (1.5 µg/L) (Health Canada 2022). Similar to cyanobacteria biomass, MC concentrations varied by ecozone, with 67% of all detections occurring in the Prairies and Boreal Plains (Fig. 1B). This distribution is also associated with trophic state, with detectable concentrations (≥ 0.1 µg/L) often occurring in lakes characterized as eutrophic or hypereutrophic (i.e., ≥30 µg/L total phosphorus, in Wetzel (2001)).

Based on VIF analysis, we removed 11 collinear variables from the list of potential predictors, leaving 45 candidate environmental and cyanobacterial predictor variables. RF analyses on this set of predictors identified *Microcystis* biomass, TP, and *Daphnia* biomass as the three top ranking predictors of total MC concentrations. Additional important predictors included a range of biotic and abiotic variables including TN, soluble reactive phosphorus (SRP), cyclopoid copepod biomass, cyanobacteria biomass, temperature, and percents agriculture and pasture within the watershed.

Using the top 20 variables identified in the RF of the full lake dataset (n=435), the first part of the hurdle model showed that the probability of MC presence depended on a combination of biotic and abiotic variables; including increasing *Microcystis* biomass, *Daphnia* and cyclopoid copepod biomass, SRP, pH and 30-day average wind speed. These variables collectively explained 34% of MC presence (Table 1). In sites where MC was detected (n=130), increasing MC concentrations were best explained again by increasing *Microcystis* biomass, but also TP, CBB_{other}, cyclopoid copepod biomass, dissolved inorganic carbon (DIC), and euphotic zone temperature. The best-fit model also included a negative relationship with *Dolichospermum* biomass and watershed slope (within a 100 m of the lake). This second part of the hurdle model explained 70% of the variation in MC concentration (Table 1).

Distributions of microcystin congeners across Canada

Based on our congener analyses, we found that 65% of target lakes (123 out of 190) had at least one MC congener, and the lake-specific sum of concentrations varied from 0 to 22.5 µg/L (Table S3). The most widely distributed congener was MC-LR, occurring in 61% of lakes (Fig. 2C). This was followed by MC-LA (37%), MC-HtyR (25%), MC-YR (21%) and D-Asp-3-MC-LR (21%) (Table S3; Fig. 2C). As many as ten congeners were present in a given lake, with

MC-LR and MC-LA typically representing the dominant forms. Specifically, MC-LR was the dominant (most abundant) in 42% of the 190-lake subset, while MC-LA was the dominant in 15% of the subset (Fig. 2A and B). However, most of the targeted congeners were scarcely found, and one of the 14 congeners was not detected in any lake (i.e., MC-LW; Fig. 2C).

The diversity of MC congeners (calculated using the Shannon index) was highest in central Canadian ecozones, where total microcystin and cyanobacteria biomass values were greatest (Fig. 3). In lakes with at least one congener detected, 79% (97 out of 123) contained multiple variants, and 21% had more than five. Overall, MC congener diversity showed a positive with relationship with total cyanobacteria biomass, but interestingly, no relation with cyanobacteria genus diversity (Fig. S2A and B).

The MFA for the 190-lake dataset highlighted that the composition of congeners was correlated with specific biotic (cyanobacteria and zooplankton communities) and abiotic (environmental) factors (Fig. 4A and B). The first and second dimensions cumulatively explained 31.5% of the variance and showed that congeners were strongly correlated to one another (clustered together in the MFA ordination space). Their distributions were associated with lake trophic state. Almost all congeners showed higher relative abundances in eutrophic and hypereutrophic lakes and were collectively correlated with elevated nutrients (TN and TP), water temperature, DOC, DIC, ions, zooplankton and agricultural land-use. Only MC-YR was situated opposite to eutrophic conditions, but this effect was marginal. Several congeners were also correlated with specific cyanobacteria genera. For example, MC-RR, -LA, -WR and Leu1-MC-LR clustered with *Planktothrix* and *Microcystis*. MC-LR and Dha7-MC-LR were associated with *Aphanizomenon* (i.e., a non-MC producer) and MC-YR was associated with *Dolichospermum* biomass. Finally, we also noted congruence between MC congeners and zooplankton variables.

For instance, the cluster of MC-RR and -LA was categorized by lakes with elevated cyclopoid copepod and small cladoceran biomass (as well as increased air and water temperature, wind speed and ion concentrations (chloride, sulfate and sodium)). Meanwhile, higher relative abundances of MC-LR were characteristic of lakes with greater *Daphnia* biomass (as well as chlorophyll-a, color and moderately with TP and SRP).

To identify the proportion of variation explained uniquely and jointly by our three predictor groups, we ran a variation partitioning analysis. Although environmental variables explained a slightly more unique proportion of the variation, overall explanatory power was low (<10%), and the highest proportion of variation was explained jointly by environmental variables and the cyanobacteria community (Fig. 5).

In the subset of lakes where at least one congener was detected (n=123), the first and second MFA dimensions explained a comparable proportion of variation (27.1%), and overall, displayed similar relationships between congeners and the predictor matrices as the 190-lake dataset. However, we noted a better separation of the MC congeners (Fig. S4A and B). In particular, the dominant congeners opposed one another in ordination space, with MC-LR being correlated again to *Aphanizomenon*, *Daphnia* and color, whereas MC-LA was more strongly related to *Aphanocapsa*, *Microcystis* and water temperature.

Discussion

In this study, we developed a national-scale dataset of 440 lakes to examine the distributions and drivers of microcystins. Based on this extensive dataset, we found that a combination of elevated cyanobacteria biomass (i.e., *Microcystis*), nutrients (TP and SRP), zooplankton biomass (*Daphnia* and cyclopoid copepod biomass), temperature, DIC and pH were

the top predictors of total microcystins and were correlated with several congeners. We also detected a strong geographical signal, with the greatest concentration of most MC congeners detected in lakes from the Prairies and Boreal Plains; the most eutrophic ecozones. The two most dominant and widespread congeners were MC-LR and MC-LA. However, several less studied congeners were widely detected across Canadian lakes. Overall, our study highlights that the risk of detecting MCs varies substantially according to geography, and despite the inclusion of dozens of predictor variables, most of the significant drivers (i.e., nutrients, temperature and zooplankton) either directly or indirectly promote cyanobacteria biomass. For that reason, management efforts targeting nutrient controls are well suited for mitigating both cyanobacteria and their microcystins.

Predictors of total microcystin concentrations

MCs are highly variable in space and time, and although extensive laboratory and field studies have identified biotic and abiotic environmental conditions associated with increasing MC occurrence (reviewed in Dai et al. 2016; Neilan et al. 2013; Rastogi et al. 2014), the factors that influence their production are variable from strain to strain (Dai et al. 2016; Huisman et al. 2018). As a result, taxa that may produce MC are referred to as potentially toxic taxa. Furthermore, the factors that explain variation can differ between spatial and temporal scales (Tillmanns and Pick 2011). Overall, reliably predicting MC concentrations can be challenging, but we have aimed to maximize variance explained by incorporating a broader range of potential predictors (biotic and abiotic) and standardizing the way in which the data were collected across a national sampling campaign.

Cyanobacteria biomass

Foremost, we observed a positive relationship between MC concentrations and the biomass of several genera including *Microcystis*, *Dolichospermum*, *Aphanocapsa* and *Aphanizomenon* (Fig. S5; Fig. S6). Whereas the occurrence of potential MC producers (Table S2) does not always translate into occurrence of MCs (Hollister and Kreakie 2016; Merel et al. 2013; Sinang et al. 2013), MC concentrations are often positively correlated to the abundance of potential toxin-producing taxa in large-scale datasets (Beaver et al. 2014; Pick 2016; Shan et al. 2020). For example, Dolman et al. (2012) found that MC concentrations were positively associated with *Planktothrix* biomass in a 102-lake study of German sites. In North American lakes, the substantial variation in MC concentration has been explained by *Microcystis* and *Dolichospermum* biomass, even before the inclusion of abiotic variables (Giani et al. 2005; Monchamp et al. 2014; Rolland et al. 2005). Across our dataset, *Microcystis* biomass was the top predictor of total MC and was retained in both parts of the hurdle model (Table 1). As a key producer of MCs, this positive association was expected. Some authors have even provided evidence of an association between enhanced MC synthesis with increasing cell densities of *Microcystis* (Wood et al. 2012, 2021). Upregulation of MC synthesis may be related to cell-to-cell signaling, or as a response to oxidative stress and photodamage (Omidi et al. 2018; Wood et al. 2021).

Knowledge of which taxa are capable of producing MC is continuously expanding (Chorus and Welker 2021; Erratt et al. 2022). Several studies have found taxa not reported to be producers of MCs are correlated to MC concentrations, such as *Aphanizomenon* (Chen et al. 2007; Zastepa et al. 2017a). Likewise, we found *Aphanizomenon* biomass to be positively associated with MCs (Fig. S6). Based on the available data, we cannot discern whether *Aphanizomenon* strains were in fact producing MCs, and it is possible that the positive affiliation

is derived from known MC producers co-existing within *Aphanizomenon* blooms (Zastepa et al. 2017a). Among the more recent discoveries of potential toxin-producers are several picocyanobacteria (Jakubowska and Szeląg-Wasielewska 2015). Most notably, the picocyanobacteria *Aphanocapsa* (present in 52% of our lakes) was also significantly, positively associated with MC concentrations (Fig. S6).

Nutrients

Both nitrogen and phosphorus are essential for phytoplankton metabolism, and their concentrations have been widely identified as important predictors of both cyanobacteria biomass and MC production at continental scales (Dolman et al. 2012; Yuan and Pollard 2017) and laboratory studies (Wagner et al. 2021, 2019). After *Microcystis* biomass, phosphorus (TP and SRP) was the most important predictor of MC concentrations. Although TN was not selected in the final model, the highest MC concentrations occurred when both TN and TP were elevated (50 µg/L and 300 µg/L, respectively) and the ratio of TN:TP was low (<23:1) (Fig. S7A and B). The relationship with TN was low but increased linearly above ~300 µg/L, whereas the relationship with TP was unimodal (increasing up to 800 µg/L and decreasing thereafter: Fig. S8 A and B). In the final model, we identified SRP as the nutrient most strongly related to the probability of detecting MCs, while TP was strongly correlated with increasing MC concentrations (Table 1). SRP has been recognized as a main driver of MCs in lakes, primarily as an available nutrient source for cell growth (Lee et al. 2015; Wang et al. 2022). Previous studies have found strong correlations between MC concentrations and TP, invoking that phosphorus increases the growth rate of cells and facilitates MC production (Harke and Gobler 2013; Rinta-Kanto et al. 2009).

Other physical-chemical predictors

Both increasing pH and DIC were associated with MC presence and concentrations, respectively. Correlations between MC and DIC have also been reported in previous studies. However, as with TP, the mechanism is most likely indirect through cyanobacterial growth (Buley et al. 2021; Tao et al. 2012). Cyanobacteria possess complex carbon concentration mechanisms, allowing them to utilize different forms of inorganic carbon depending on what is available (Pick and Lean 1987; Talling 1976), and thus help them maintain elevated photosynthetic rates and possibly MC production rates across different environments. Increased photosynthetic rates deplete dissolved carbon dioxide from the water and increase pH. Furthermore, MC may be produced to allow cyanobacterial cells to adapt to different environmental conditions, including low and high DIC (Omidi et al. 2018), though more work is needed in this area.

Despite the modest temperature gradient (since we restricted our sampling period to the warmest months), average euphotic zone temperature emerged as a significant predictor in modeling MC concentrations (Table 1). Elevated temperatures can promote MC production in a number of ways (Dai et al. 2016; Omidi et al. 2018). First, temperature stimulates cyanobacteria growth, notably *Microcystis,* which have higher optimum growth temperatures relative to other taxa (Paerl et al. 2011; Paerl and Paul 2012), although MC production appears to have a lower temperature optimum than *Microcystis* growth (Martin et al. 2020; Peng et al. 2018). In national scale models, temperature (along with nutrients) has been one of the leading predictors of cyanobacteria biomass (Beaulieu et al. 2013; Kosten et al. 2012). Temperature may also increase MC cellular quota and regulate release (Mowe et al. 2015; Walls et al. 2018). Toxic to non-toxic biomass ratios have also been observed to rise with temperature (Davis et al. 2009; Dziallas and Grossart 2011). Davis et al. (2009) showed that an increase of 4°C yielded significantly higher

growth rates of toxic *Microcystis* strains over non-toxic genotypes. Peak MC production and release appears tightly coupled with optimum growth rates and photosynthesis, with several previous studies showing the highest concentrations of MCs between 20°C and 25°C; a similar range to what we observed (Kelly et al. 2019; Tao et al. 2012; Walls et al. 2018). One hypothesis as to why MC cellular quotas increase with temperature is that MC protects against oxidative stress that occurs at high temperatures when photosynthetic rates are elevated (Dziallas and Grossart 2011; Omidi et al. 2018). However, lower temperatures may also result in oxidative stress (Martin et al. 2020).

Zooplankton

The strong, positive correlations between MCs and the biomass of multiple zooplankton groups, including *Daphnia* and cyclopoid copepods, were intriguing results from our analyses. Increasing *Daphnia* biomass was a predictor of MC presence, while increasing cyclopoid copepod biomass was a significant predictor of both the presence-absence (part 1 of hurdle model) and the continued rise (part 2 of hurdle model) in MCs (Table 1). Previously, cladoceran and copepod biomass were also found to be positively correlated with intracellular and extracellular MCs in a large Chinese lake; this pattern was attributed to zooplankton having increased prey selectivity, tolerance to cyanobacteria and/or experience predator release concurrent with blooms (Ger et al. 2016; Wang et al. 2022). Copepods exhibit selectivity when cyanobacteria are dominant, with some populations exhibiting increased selective avoidance quickly after exposure to *Microcystis* cells (Ger et al. 2011; Leitão et al. 2018). Copepods may even use microcystins as detection cues to distinguish between toxic and non-toxin *Microcystis* strains (Agasild et al. 2019; Ger et al. 2016).

Daphnia have been shown to evolve tolerances to MCs across short time frames (<10 generations) allowing them to coexist and even suppress toxic blooms (Ger et al. 2016; Gustafsson and Hansson 2004; Jiang et al. 2016). Tolerance to MCs is strengthened by previous exposure to toxic cyanobacteria (Tillmanns et al. 2011). In numerous cases, *Daphnia* genotypes isolated from high nutrient lakes were less inhibited by MCs than those from low nutrient lakes (Chislock et al. 2013; Wilson et al. 2006). In the present study, the majority (66%) of MCs were detected in the Prairies and Boreal Plains, where numerous lakes have been eutrophic for decades and possibly centuries. It is conceivable that zooplankton in these lakes have been exposed to toxic blooms for generations, allowing ample time to develop tolerances (Zastepa et al. 2017b). The positive relationships between MC and zooplankton groups may also stem from alternative mechanisms: 1) predation release due to anoxia-driven fish kills; and 2) increased food quality and quantity fueled by the microbial loop (Bec et al. 2006; Wang et al. 2022). Overall, zooplankton biomass is not likely a direct driver of MC production, but these toxins may not be such a deterrent and could provide an important link in food web transfer of MCs (Sotton et al. 2014).

Microcystin congener distribution and drivers

There is a long history of MC-LR being the most targeted and widespread found MC congener (Pick 2016). However, several recent analyses have highlighted the frequency of additional, often overlooked congeners (Díez-Quijada et al. 2019; Turner et al. 2018). A broad survey of European lakes recently identified MC-YR as the most prevalent congener across a 137-lake dataset (Mantzouki et al. 2018). In contrast, MC-LA is encountered more frequently in North American lakes (Pick 2016; Taranu et al. 2019) and is largely absent from European sites. Across our 190-lake Canadian dataset, we targeted 14 congeners and found MC-LR and -LA to

be the most common (Fig. 2A and C), with MC-LA reaching the highest concentrations. Our MC-LA results are interesting as it is as toxic as MC-LR (Chernoff et al. 2020), exhibits greater persistence in surface water (Zastepa et al. 2014) and has been linked to wildlife fatalities (Miller et al. 2010). The other congeners varied in concentrations and frequencies of detection from 0.5% of lakes (D-Asp3-MC-RR) to 25% (MC-HtyR). Thus, several congeners may be more common than previously thought in Canadian lakes, and correlated to different environmental variables and cyanobacterial taxa (Díez-Quijada et al. 2019; Pick 2016).

We observed that congener richness and diversity was elevated in the Prairies and Boreal Plains, and both increased with cyanobacteria biomass and MC concentrations (Fig. 3; Fig. S3). Although direct comparisons to other studies are challenging because the availability of standards has changed over time, increased MC richness and diversity with elevated cyanobacteria biomass has been documented in other regions (Bouhaddada et al. 2016; Mantzouki et al. 2018). In our dataset, 51% of lakes had at least two congeners. Congeners can be produced by several cyanobacteria genera, each of which can synthesize multiple MCs simultaneously (Puddick et al. 2014). However, congener profiles and their diversity most likely reflect strain level differences within species, which have shown considerable variation in the number of congeners produced by each strain within a bloom. Congener diversity decreased in lakes outside of the Prairies and Boreal Plains; deeper, low-nutrient and strongly stratified waters are more characteristic of lakes outside of this region. Comparably, Mantzouki et al.'s (2018) broad-scale European synthesis noted reduced diversity with water column stability. Deeper, lower nutrient sites may represent more selective conditions for cyanobacteria communities, promoting a single taxon to dominate the assemblage, and correspondingly leading to a decrease

in potential MC producers. Lower-nutrient conditions also limit biomass growth, which we also found to be positively related to increasing congener diversity.

Among field and laboratory analyses, congener composition has been shown to vary in relation to several environmental conditions including nitrogen availability (Van De Waal et al. 2009), light (Tonk et al. 2005), temperature (Mantzouki et al. 2018) and weather conditions (Taranu et al. 2019). MC-LR has been associated with strong winds, higher temperatures and elevated nutrients, whereas MC-LA has been associated with intermediate winds, wetter and low nutrient conditions (Taranu et al. 2019). Across our 190-lake dataset, the detection and relative abundance of congeners was first dependent upon nutrient status (Fig. 4). Considering the subset of lakes with at least one congener detected (n=123), the relative abundance of congeners was structured by several biotic and abiotic variables (Fig. S3). In particular, MC-LR was positively correlated with higher *Daphnia* biomass, color and phosphorus, whereas MC-LA was more closely correlated to increased urban land and euphotic temperature and *Dolichospermum* biomass (Fig. S3). Interestingly, the occurrence of several less routinely monitored congeners were not positively correlated to a particular taxon but were most associated with eutrophic lake conditions. For instance, MC-HilR and -HtyR correlated positively to TP, SPR, TN, DOC, DIC and agricultural land, all of which were elevated with congener diversity and total microcystin concentration (Fig. S3).

Although environmental factors explain some variation in MC congener composition, these relationships are likely mediated by the cyanobacteria community, which in turn are directly affected by environmental gradients (Monchamp et al. 2014; Taranu et al. 2019). In an analysis of congener profiles across 70 English lakes, Turner et al. (2018) found weak relationships between environmental variables and congeners directly, suggesting their profiles

are linked to cyanobacterial species successional patterns. On a relative scale, we noted MC-LR appears to be more closely related to *Aphanizomenon* biomass; an unidentified producer of MCs, whereas MC-LA was related to *Planktothrix, Microcystis* and *Dolichospermum.* Ultimately, due to strain level differences within genera, it is difficult to fully disentangle the unique and shared effects of environmental variables and cyanobacteria genera on MC congener composition. Future research in metagenomics could help resolve some of the strain level differences among lakes and supplement our understanding from the classical taxonomy.

From a management perspective, changes in land use can result in greater nutrient loads to lakes, which in turn influence MC production. In previous multi-lake analyses, elevated MC concentrations were largely reported in agriculturally dominated regions with low percentages of forested land cover (Beaver et al. 2014). Furthermore, agricultural practices and the subsequent nutrient loading contributed to the record-breaking cyanobacteria bloom in Lake Erie in 2011 (Michalak et al. 2013). This connection also has health implications as crops from agricultural fields irrigated with MC-contaminated water can bioaccumulate MCs and create a new exposure route for consumers (Melaram et al. 2022; Miller and Russell 2017). Overall, we identified the positive influence of several environmental predictors including chemical, zooplankton and climate variables. However, mitigation of MC exposure should center on nutrient control.

Conclusions

Across the 440 temperate-boreal lakes examined herein, total MCs were detected in relatively low concentrations, with only 10% of sites exhibiting MC concentrations above WHO drinking water guidelines. Lakes with elevated levels of MC were largely concentrated in the Prairies and Boreal Plains ecozones of central Canada, in which lakes are characterized by

eutrophic and ion-rich waters. We identified that high nutrients (mainly soluble reactive and total phosphorus) and *Microcystis* biomass were the most important predictors of MC, whereas alkalinity (DIC and pH), wind, temperature, and zooplankton biomass were identified as explanatory variables of secondary importance. Congener analyses provided further insights into the toxicity of MC, and the drivers associated with MC production. A variety of less commonly measured congeners were present across Canadian lakes, however MC-LR and -LA remained the most abundant and therefore should be considered in monitoring programs. Overall, congener relative concentrations are positively, but only moderately related to many of the same environmental predictors that control cyanobacteria biomass, including nutrients and ion concentrations. Our analysis serves as an essential resource for evaluating the current incidence of MCs and congeners in Canadian lakes, and for estimating MC occurrence under future scenarios of environmental change.

Acknowledgements

We would like to thank the NSERC LakePulse network (grant awarded to IGE and BEB, among others). We thank all the field and laboratory teams for sample collection and preparation, as well as the director of LakePulse; Yannick Huot. We thank Marianne Racine for assisting with the microcystin laboratory work at ECCC. Additional funding for this project was provided by an NSERC scholarship to PM and a Canada Research Chair to IGE.

References

Agasild, H., K. Panksep, I. Tõnno, and others. 2019. Role of potentially toxic cyanobacteria in crustacean zooplankton diet in a eutrophic lake. Harmful Algae **89**. doi:10.1016/j.hal.2019.101688

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. **58**: 1736–1746. doi:10.4319/lo.2013.58.5.1736

Beaver, J.R., E. Manis, K. Loftin, J. Graham, A. Pollard, R. Mitchell. 2014. Land use patterns, ecoregion, and microcystin relationships in U.S. lakes and reservoirs: A preliminary evaluation. Harmful Algae **36**: 57–62. doi:10.1016/j.hal.2014.03.005

Bec, A., D. Martin-Creuzburg, and E. Von Elert. 2006. Trophic upgrading of autotrophic picoplankton by the heterotrophic nanoflagellate Paraphysomonas sp. Limnol. Oceanogr. **51**: 1699–1707. doi:10.4319/lo.2006.51.4.1699

Bernard, C., A. Ballot, S. Thomazeau, S. Malufi, A. Furey, J. Mankiewicz-Boczek, B. Pawlik-Skowronska, C. Capelli, N. Salmaso. 2017. Appendix 2: cyanobacteria associated with the production of cyanotoxins. J. Meriluoto, L. Spoof, G.A. Codd (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis, Wiley, Chichester UK (2017), 500-525. doi:10.1002/9781119068761.app2

Borcard, D., F. Gillet, and P. Legendre. 2018. *In* R. Gentleman, K. Hornik, and G. Parmigiani [eds.], Numerical ecology with R, 2nd ed. Springer Nature.

Bouaïcha, N., C. O. Miles, D. G. Beach, Z. Labidi, A. Djabri, N. Y. Benayache, and T. Nguyen-quang. 2019. Structural Diversity, Characterization and Toxicology of Microcystins. Toxins (Basel). **11**: 714. doi:10.3390/toxins11120714

Bouhaddada, R., S. Nélieu, H. Nasri, G. Delarue, and N. Bouaïcha. 2016. High diversity of microcystins in a Microcystis bloom from an Algerian lake. Environ. Pollut. **216**: 836–844. doi:10.1016/j.envpol.2016.06.055

Buley, R. P., H. E. Correia, A. Abebe, T. B. Issa, and A. E. Wilson. 2021. Predicting microcystin occurrence in freshwater lakes and reservoirs: assessing environmental variables. Inl. Waters **11**: 430–444. doi:10.1080/20442041.2021.1938491

Chapra, S. C., B. Boehlert, C. Fant, and others. 2017. Climate Change Impacts on Harmful Algal Blooms in U.S. Freshwaters: A Screening-Level Assessment. Environ. Sci. Technol. **51**: 8933–8943. doi:10.1021/acs.est.7b01498

Chen, H., J. Burke, W. Dinsmore, E. Prepas, P. Fedorak. 2007. First assessment of cyanobacterial blooms and microcystin-LR in the Canadian portion of Lake of the Woods. Lake Reserv. Manag. **23**: 169–178. doi:10.1080/07438140709353920

Chernoff, N., D. Hill, J. Lang, J. Schmid, T. Le, A. Farthing, and H. Huang. 2020. The comparative toxicity of 10 microcystin congeners administered orally to mice: Clinical effects and organ toxicity. Toxins (Basel). **12**: 1–18. doi:10.3390/toxins12060403

Chislock, M. F., O. Sarnelle, L. M. Jernigan, and A. E. Wilson. 2013. Do high concentrations of microcystin prevent Daphnia control of phytoplankton? Water Res. **47**: 1961–1970. doi:10.1016/j.watres.2012.12.038

Chorus, I., and M. Welker. 2021. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management, 2nd ed. CRC Press.

Dai, R., P. Wang, P. Jia, Y. Zhang, X. Chu, and Y. Wang. 2016. A review on factors affecting microcystins production by algae in aquatic environments. World J. Microbiol. Biotechnol. **32**: 1–7. doi:10.1007/s11274-015-2003-2

Davis, T. W., D. L. Berry, G. L. Boyer, and C. J. Gobler. 2009. The effects of temperature and nutrients on the growth and dynamics of toxic and non-toxic strains of Microcystis during cyanobacteria blooms. Harmful Algae **8**: 715–725. doi:10.1016/j.hal.2009.02.004

Díez-Quijada, L., A. I. Prieto, R. Guzmán-Guillén, A. Jos, and A. M. Cameán. 2019. Occurrence and toxicity of microcystin congeners other than MC-LR and MC-RR: A review. Food Chem. Toxicol. **125**: 106–132. doi:10.1016/j.fct.2018.12.042

Dolman, A. M., J. Rücker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. PLoS One **7**. doi:10.1371/journal.pone.0038757

Dziallas, C., and H. P. Grossart. 2011. Increasing oxygen radicals and water temperature select for toxic microcystis sp. PLoS One **6**. doi:10.1371/journal.pone.0025569

Erratt, K. J., I. F. Creed, and C. G. Trick. 2022. Harmonizing science and management options to reduce risks of cyanobacteria. Harmful Algae **116**. doi:10.1016/j.hal.2022.102264

Feld, C. K., P. Segurado, and C. Gutiérrez-Cánovas. 2016. Analysing the impact of multiple stressors in aquatic biomonitoring data: A 'cookbook' with applications in R. Sci. Total Environ. **573**: 1320–1339. doi:10.1016/j.scitotenv.2016.06.243

Fox, J., S. Weisberg. 2011. An R Companion to Applied Regression. second ed. Sage. URL, Thousand Oaks. Available at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion

Flores, N. M., T. R. Miller, and J. D. Stockwell. 2018. A global analysis of the relationship between concentrations of microcystins in water and fish. Front. Mar. Sci. **5**: 1–14. doi:10.3389/fmars.2018.00030

Ger, K. A., R. Panosso, and M. Lurling. 2011. Consequences of acclimation to Microcystis on the selective feeding behavior of the calanoid copepod Eudiaptomus gracilis. Limnol.

Oceanogr. **56**: 2103–2114. doi:10.4319/lo.2011.56.6.2103

Ger, K. A., P. Urrutia-Cordero, P. C. Frost, L. A. Hansson, O. Sarnelle, A. E. Wilson, and M. Lürling. 2016. The interaction between cyanobacteria and zooplankton in a more eutrophic world. Harmful Algae **54**: 128–144. doi:10.1016/j.hal.2015.12.005

Giani, A., D. F. Bird, Y. T. Prairie, and J. F. Lawrence. 2005. Empirical study of cyanobacterial toxicity along a trophic gradient of lakes. Can. J. Fish. Aquat. Sci. **62**: 2100–2109. doi:10.1139/f05-124

Gustafsson, S., and L. A. Hansson. 2004. Development of tolerance against toxic cyanobacteria in Daphnia. Aquat. Ecol. **38**: 37–44. doi:10.1023/B:AECO.0000020985.47348.5e

Harke, M. J., and C. J. Gobler. 2013. Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. PLoS One **8**. doi:10.1371/journal.pone.0069834

Hayes, N. M., and M. J. Vanni. 2018. Microcystin concentrations can be predicted with phytoplankton biomass and watershed morphology. Inl. Waters **8**: 273–283. doi:10.1080/20442041.2018.1446408

Health Canada 2022. Guidelines for Canadian Recreational Water Quality: Cyanobacteria and their Toxins

Hollister, J. W., and B. J. Kreakie. 2016. Associations between chlorophyll a and various microcystin-LR health advisory concentrations. F1000Research **5**: 151. doi:10.12688/f1000research.7955.1

Hothorn, T., K. Hornik, C. Strobl, A. Zeileis. 2021. party: A laboratory for Recursive Partytioning. R Package version 1.3-8

Huisman, J., G. A. Codd, H. W. Paerl, B. W. Ibelings, J. M. H. Verspagen, and P. M. Visser.

2018. Cyanobacterial blooms. Nat. Rev. Microbiol. **16**: 471–483. doi:10.1038/s41579-018-0040-1

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. **695**: 133668. doi:10.1016/j.scitotenv.2019.133668

Jakubowska, N., and E. Szeląg-Wasielewska. 2015. Toxic picoplanktonic cyanobacteria - Review. Mar. Drugs **13**: 1497–1518. doi:10.3390/md13031497

Jia, J., W. Luo, Y. Lu, and J. P. Giesy. 2014. Bioaccumulation of microcystins (MCs) in four fish species from Lake Taihu, China: Assessment of risks to humans. Sci. Total Environ. **487**: 224–232. doi:10.1016/j.scitotenv.2014.04.037

Jiang, X., H. Gao, L. Zhang, H. Liang, and X. Zhu. 2016. Rapid evolution of tolerance to toxic Microcystis in two cladoceran grazers. Sci. Rep. **6**: 1–7. doi:10.1038/srep25319

Kakouei, K., B. M. Kraemer, O. Anneville, and others. 2021. Phytoplankton and cyanobacteria abundances in mid-21st century lakes depend strongly on future land use and climate projections. Glob. Chang. Biol. **27**: 6409–6422. doi:10.1111/gcb.15866

Kelly, N. E., A. Javed, Y. Shimoda, A. Zastepa, S. Watson, S. Mugalingam, and G. B. Arhonditsis. 2019. A Bayesian risk assessment framework for microcystin violations of drinking water and recreational standards in the Bay of Quinte, Lake Ontario, Canada. Water Res. **162**: 288–301. doi:10.1016/j.watres.2019.06.005

Kosten, S., V. L. M. Huszar, E. Bécares, and others. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. Glob. Chang. Biol. **18**: 118–126. doi:10.1111/j.1365-2486.2011.02488.x

Kotak, B. G., A. K. Y. Lam, E. E. Prepas, and S. E. Hrudey. 2000. Role of chemical and physical

variables in regulating microcystin-LR concentration in phytoplankton of eutrophic lakes. Can. J. Fish. Aquat. Sci. **57**: 1584–1593. doi:10.1139/f00-091

Kotak, B. G., and R. W. Zurawell. 2007. Cyanobacterial toxins in Canadian freshwaters: A review. Lake Reserv. Manag. **23**: 109–122. doi:10.1080/07438140709353915

Kozlowsky-Suzuki, B., A. E. Wilson, and A. da Silva Ferrão-Filho. 2012. Biomagnification or biodilution of microcystins in aquatic foodwebs? Meta-analyses of laboratory and field studies. Harmful Algae **18**: 47–55. doi:10.1016/j.hal.2012.04.002

Lê, S., S. Josse, F. Husson. 2008. FactoMineR: a package for multivariate analysis. J. Stat. Softw. **25**: 1-18. doi:10.18637/jss.v025.i01

Lee, T. A., G. Rollwagen-Bollens, S. M. Bollens, and J. J. Faber-Hammond. 2015. Environmental influence on cyanobacteria abundance and microcystin toxin production in a shallow temperate lake. Ecotoxicol. Environ. Saf. **114**: 318–325. doi:10.1016/j.ecoenv.2014.05.004

Leitão, E., K. A. Ger, and R. Panosso. 2018. Selective grazing by a tropical copepod (Notodiaptomus iheringi) facilitates microcystis dominance. Front. Microbiol. **9**: 1–11. doi:10.3389/fmicb.2018.00301

Loftin, K. A., J. L. Graham, E. D. Hilborn, S. C. Lehmann, M. T. Meyer, J. E. Dietze, and C. B. Griffith. 2016. Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. Harmful Algae **56**: 77–90. doi:10.1016/j.hal.2016.04.001

Lund, J. W. G., C. Kipling, and E. D. Le Cren. 1958. The inverted microscope method of estimating algal numbers and the statistical basis of estimations by counting. Hydrobiologia **11**: 143–170. doi:10.1007/BF00007865

MacKeigan, P. W., R. E. Garner, M. È. Monchamp, and others. 2022. Comparing microscopy and DNA metabarcoding techniques for identifying cyanobacteria assemblages across hundreds of lakes. Harmful Algae **113**. doi:10.1016/j.hal.2022.102187

Mantzouki, E., M. Lürling, J. Fastner, and others. 2018. Temperature effects explain continental scale distribution of cyanobacterial toxins. Toxins (Basel). **10**: 1–24. doi:10.3390/toxins10040156

Martin, R., M. Moniruzzaman, G. Stark, E. Gann, D. Derminio, B. Wei, F. Hellweger, A. Pinto, G. Boyer, S. Wilhelm. 2020. Episodic Decrease in Temperature Increases mcy Gene Transcription and Cellular Microcystin in Continuous Cultures of Microcystis aeruginosa PCC 7806. Front. Microbiol. **11**: doi:10.3389/fmicb.2020.601864

McKindles, K., M. Manes, J. DeMarco, A. McClure, R. McKay, T. Davis, G. Bullerjahn. 2020. Dissolved microcystin release coincident with lysis of a Microcystis-dominated bloom in western Lake Erie attributed to a novel cyanophage. Appl. Environ. Microbiol. doi:10.1128/AEM.01397-20

Melaram, R., A. R. Newton, and J. Chafin. 2022. Microcystin Contamination and Toxicity: Implications for Agriculture and Public Health. Toxins (Basel). **14**. doi:10.3390/toxins14050350

Merel, S., D. Walker, R. Chicana, S. Snyder, E. Baurès, and O. Thomas. 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. Environ. Int. **59**: 303–327. doi:10.1016/j.envint.2013.06.013

Messager, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt. 2016. Estimating the volume and age of water stored in global lakes using a geo-statistical approach. Nat. Commun. **7**: 1–11. doi:10.1038/ncomms13603

Michalak, A. M., E. J. Anderson, D. Beletsky, and others. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. Proc. Natl. Acad. Sci. U. S. A. **110**: 6448–6452. doi:10.1073/pnas.1216006110

Miller, A., and C. Russell. 2017. Food crops irrigated with cyanobacteria-contaminated water: an emerging public health issue in Canada. Environ. Heal. Rev. **60**: 58–63. doi:10.5864/d2017-021

Miller, M. A., R. M. Kudela, A. Mekebri, and others. 2010. Evidence for a novel marine harmful algal bloom: Cyanotoxin (microcystin) transfer from land to sea otters. PLoS One **5**: 1–11. doi:10.1371/journal.pone.0012576

Monchamp, M. E., F. R. Pick, B. E. Beisner, and R. Maranger. 2014. Nitrogen forms influence microcystin concentration and composition via changes in cyanobacterial community structure. PLoS One **9**. doi:10.1371/journal.pone.0085573

Mowe, M. A. D., C. Porojan, F. Abbas, S. M. Mitrovic, R. P. Lim, A. Furey, and D. C. J. Yeo. 2015. Rising temperatures may increase growth rates and microcystin production in tropical Microcystis species. Harmful Algae **50**: 88–98. doi:10.1016/j.hal.2015.10.011

Muñoz Sabater, J. 2019. ERA5-Land hourly data from 1981 to present. Copernicus Clim. Chang. Serv. Clim. Data Store. doi:10.24381/cds.e2161bac

Naimi, B. 2015. Usdm: Uncertainty Analysis for Species Distribution Models. R Package version 1.1-15

Neilan, B. A., L. A. Pearson, J. Muenchhoff, M. C. Moffitt, and E. Dittmann. 2013. Environmental conditions that influence toxin biosynthesis in cyanobacteria. Environ. Microbiol. **15**: 1239–1253. doi:10.1111/j.1462-2920.2012.02729.x

NSERC Canadian LakePulse Network. 2021 NSERC Canadian LakePulse Network

M.-P. Varin, M.-L. Beaulieu, Y. Huot (Eds.). NSERC Canadian Lake Pulse Network Field Manual 2017–2018–2019 Surveys, Université de Sherbrooke (2021).

Oksanen, A. J., F. G. Blanchet, M. Friendly, and others. 2019. Vegan. Encycl. Food Agric. Ethics 2395–2396. doi:10.1007/978-94-024-1179-9_301576

Omidi, A., M. Esterhuizen-Londt, and S. Pflugmacher. 2018. Still challenging: the ecological function of the cyanobacterial toxin microcystin–What we know so far. Toxin Rev. **37**: 87–105. doi:10.1080/15569543.2017.1326059

Orihel, D. M., D. F. Bird, M. Brylinsky, and others. 2012. High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. Can. J. Fish. Aquat. Sci. **69**: 1457–1462. doi:10.1139/F2012-088

Paerl, H. W., N. S. Hall, and E. S. Calandrino. 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. Sci. Total Environ. **409**: 1739–1745. doi:10.1016/j.scitotenv.2011.02.001

Paerl, H. W., and V. J. Paul. 2012. Climate change: Links to global expansion of harmful cyanobacteria. Water Res. **46**: 1349–1363. doi:10.1016/j.watres.2011.08.002

Paquette, C., I. Gregory-Eaves, and B. E. Beisner. 2021. Multi-scale biodiversity analyses identify the importance of continental watersheds in shaping lake zooplankton biogeography. J. Biogeogr. **48**: 2298–2311. doi:10.1111/jbi.14153

Peng, G., R. Martin, S. Dearth, X. Sun, G. Boyer, S. Campagna, S. Lin, S. Wilhelm. 2018. Seasonally Relevant Cool Temperatures Interact with N Chemistry to Increase Microcystins Produced in Lab Cultures of Microcystis aeruginosa NIES-843. Environ. Sci. Technol. **52**: 4127–4136. doi:0.1021/acs.est.7b06532

Pick, F. R. 2016. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria.

Can. J. Fish. Aquat. Sci. **73**: 1149–1158. doi:10.1139/cjfas-2015-0470

Pick, F. R., and D. R. S. Lean. 1987. The role of macronutrients (C, N, P) in controlling cyanobacterial dominance in temperate lakes. New Zeal. J. Mar. Freshw. Res. **21**: 425–434. doi:10.1080/00288330.1987.9516238

Plaas, H. E., and H. W. Paerl. 2021. Toxic Cyanobacteria: A Growing Threat to Water and Air Quality. Environ. Sci. Technol. **55**: 44–64. doi:10.1021/acs.est.0c06653

Puddick, J., M. R. Prinsep, S. A. Wood, S. A. F. Kaufononga, S. C. Cary, and D. P. Hamilton. 2014. High levels of structural diversity observed in microcystins from microcystis CAWBG11 and characterization of six new microcystin congeners. Mar. Drugs **12**: 5372–5395. doi:10.3390/md12115372

Rastogi, R. P., R. P. Sinha, and A. Incharoensakdi. 2014. The cyanotoxin-microcystins: Current overview. Rev. Environ. Sci. Biotechnol. **13**: 215–249. doi:10.1007/s11157-014-9334-6

Rinta-Kanto, J. M., E. A. Konopko, J. M. DeBruyn, R. A. Bourbonniere, G. L. Boyer, and S. W. Wilhelm. 2009. Lake Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. Harmful Algae **8**: 665–673. doi:10.1016/j.hal.2008.12.004

Rolland, A., D. F. Bird, and A. Giani. 2005. Seasonal changes in composition of the cyanobacterial community and the occurrence of hepatotoxic blooms in the eastern townships, Québec, Canada. J. Plankton Res. **27**: 683–694. doi:10.1093/plankt/fbi042

Roy-Lachapelle, A., M. Solliec, M. F. Bouchard, and S. Sauvé. 2017. Detection of cyanotoxins in algae dietary supplements. Toxins (Basel). **9**: 1–17. doi:10.3390/toxins9030076

Schindler, D. W., R. E. Hecky, D. L. Findlay, and others. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment.

Proc. Natl. Acad. Sci. U. S. A. **105**: 11254–11258. doi:10.1073/pnas.0805108105

Scott, J., M. McCarthy, T. Otten, M. Steffen, B. Baker, E. Grantz, S. Wilhelm, H. Paerl. 2013. Comment: An alternative interpretation of the relationship between TN:TP and microcystins in Canadian lakes. Can. J. Fish. Aquat. Sci. **70**: 1265–1268. doi:10.1139/cjfas-2012-0490

Shan, K., X. Wang, H. Yang, B. Zhou, L. Song, and M. Shang. 2020. Use statistical machine learning to detect nutrient thresholds in Microcystis blooms and microcystin management. Harmful Algae **94**: 101807. doi:10.1016/j.hal.2020.101807

Sinang, S. C., E. S. Reichwaldt, and A. Ghadouani. 2013. Spatial and temporal variability in the relationship between cyanobacterial biomass and microcystins. Environ. Monit. Assess. **185**: 6379–6395. doi:10.1007/s10661-012-3031-0

Sotton, B., J. Guillard, O. Anneville, M. Maréchal, O. Savichtcheva, and I. Domaizon. 2014. Trophic transfer of microcystins through the lake pelagic food web: Evidence for the role of zooplankton as a vector in fish contamination. Sci. Total Environ. **466**–**467**: 152–163. doi:10.1016/j.scitotenv.2013.07.020

Spoof, L., and A. Catherine. 2017. Appendix 3: Tables of Microcystins and Nodularins. Handb. Cyanobacterial Monit. Cyanotoxin Anal. 526–537. doi:10.1002/9781119068761.app3

Talling, J. F. 1976. The Depletion of Carbon Dioxide from Lake Water by Phytoplankton. J. Ecol. **64**: 79. doi:10.2307/2258685

Tao, M., P. Xie, J. Chen, and others. 2012. Use of a generalized additive model to investigate key abiotic factors affecting microcystin cellular quotas in heavy bloom areas of lake Taihu. PLoS One **7**. doi:10.1371/journal.pone.0032020

Taranu, Z. E., I. Gregory-Eaves, R. J. Steele, M. Beaulieu, and P. Legendre. 2017. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework

for modelling an important health risk factor. Glob. Ecol. Biogeogr. **26**: 625–637. doi:10.1111/geb.12569

Taranu, Z. E., F. R. Pick, I. F. Creed, A. Zastepa, and S. B. Watson. 2019. Meteorological and Nutrient Conditions Influence Microcystin Congeners in Freshwaters. Toxins (Basel). **11**: 1–21. doi:10.3390/toxins11110620

Team, R.C. 2021. R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna.

Tillmanns, A. R., S. K. Burton, and F. R. Pick. 2011. Daphnia Pre-Exposed to Toxic Microcystis Exhibit Feeding Selectivity. Int. Rev. Hydrobiol. **96**: 20–28. doi:10.1002/iroh.201011298

Tillmanns, A. R., and F. R. Pick. 2011. The effect of sampling scales on the interpretation of environmental drivers of the cyanotoxin microcystin. Lake Reserv. Manag. **27**: 183–193. doi:10.1080/07438141.2011.587095

Tonk, L., P. M. Visser, G. Christiansen, E. Dittmann, E. O. F. M. Snelder, C. Wiedner, L. R. Mur, and J. Huisman. 2005. The microcystin composition of the cyanobacterium Planktothrix agardhii changes toward a more toxic variant with increasing light intensity. Appl. Environ. Microbiol. **71**: 5177–5181. doi:10.1128/AEM.71.9.5177-5181.2005

Turner, A. D., M. Dhanji-Rapkova, A. O'Neill, L. Coates, A. Lewis, and K. Lewis. 2018. Analysis of microcystins in cyanobacterial blooms from freshwater bodies in England. Toxins (Basel). **10**. doi:10.3390/toxins10010039

US EPA. 2019. Recommended human health recreational ambient water quality criteria or swimming advisories for microcystins and cylindrospermopsins (EPA 822-R-19-001).

Van De Waal, D. B., J. M. H. Verspagen, M. Lürling, E. Van Donk, P. M. Visser, and J. Huisman. 2009. The ecological stoichiometry of toxins produced by harmful cyanobacteria:

An experimental test of the carbon-nutrient balance hypothesis. Ecol. Lett. **12**: 1326–1335. doi:10.1111/j.1461-0248.2009.01383.x

Wagner, N. D., F. S. Osburn, J. Wang, R. B. Taylor, A. R. Boedecker, C. K. Chambliss, B. W. Brooks, and J. T. Scott. 2019. Biological stoichiometry regulates toxin production in microcystis aeruginosa (UTEX 2385). Toxins (Basel). **11**. doi:10.3390/toxins11100601

Wagner, N. D., E. Quach, S. Buscho, and others. 2021. Nitrogen form, concentration, and micronutrient availability affect microcystin production in cyanobacterial blooms. Harmful Algae **103**: 102002. doi:10.1016/j.hal.2021.102002

Walls, J. T., K. H. Wyatt, J. C. Doll, E. M. Rubenstein, and A. R. Rober. 2018. Hot and toxic: Temperature regulates microcystin release from cyanobacteria. Sci. Total Environ. **610**–**611**: 786–795. doi:10.1016/j.scitotenv.2017.08.149

Wang, X., L. Wang, M. Shang, L. Song, and K. Shan. 2022. Revealing Physiochemical Factors and Zooplankton Influencing Microcystis Bloom Toxicity in a Large-Shallow Lake Using Bayesian Machine Learning. Toxins (Basel). **14**. doi:10.3390/toxins14080530

Wetzel, R.G. 2001. Limnology: Lake and River Ecosystems. (3rd edn), Academic Press, San Diego. doi:10.1016/C2009-0-02112-6

Wetzel, R.G., and G.E. Likens. 2000. Limnological Analyses. 3rd Edition, Springer, New York. doi:10.1007/978-1-4757-3250-4

Wiken, E.B., D. Gauthier, I. Marshall, K. Lawton, H. Hirvonen. 1996. A perspective on Canada's ecosystems: an overview of the terrestrial and marine ecozones. In Canadian Council on Ecological Areas occasional papers. Canadian Council on Ecological Areas, Ottawa, Canada.

Wilson, A. E., O. Sarnelle, and A. R. Tillmanns. 2006. Effects of cyanobacterial toxicity and

morphology on the population growth of freshwater zooplankton: Meta-analyses of laboratory experiments. Limnol. Oceanogr. **51**: 1915–1924. doi:10.4319/lo.2006.51.4.1915

Wood, S. A., D. R. Dietrich, S. Craig Cary, and D. P. Hamilton. 2012. Increasing microcystis cell density enhances microcystin synthesis: A mesocosm studys. Inl. Waters **2**: 17–22. doi:10.5268/IW-2.1.424

Wood, S. A., J. Puddick, I. Hawes, K. Steiner, D. R. Dietrich, and D. P. Hamilton. 2021. Variability in microcystin quotas during a Microcystis bloom in a eutrophic lake. PLoS One **16**: 1–13. doi:10.1371/journal.pone.0254967

Yuan, L.L., A. Pollard, S. Pather, J. Oliver, L. D'Anglada. 2014. Managing microcystin: Identifying national-scale thresholds for total nitrogen and chlorophyll a. Freshw. Biol. **59**: 1970–1981. doi:10.1111/fwb.12400

Yuan, L. L., and A. I. Pollard. 2017. Using National-Scale Data to Develop Nutrient-Microcystin Relationships That Guide Management Decisions. Environ. Sci. Technol. **51**: 6972–6980. doi:10.1021/acs.est.7b01410

Zastepa, A., F. R. Pick, and J. M. Blais. 2014. Fate and Persistence of Particulate and Dissolved Microcystin-LA from Microcystis Blooms. Hum. Ecol. Risk Assess. **20**: 1670–1686. doi:10.1080/10807039.2013.854138

Zastepa, A., F. R. Pick, J. M. Blais, and A. Saleem. 2015. Analysis of intracellular and extracellular microcystin variants in sediments and pore waters by accelerated solvent extraction and high performance liquid chromatography-tandem mass spectrometry. Anal. Chim. Acta **872**: 26–34. doi:10.1016/j.aca.2015.02.056

Zastepa, A., S. B. Watson, H. Kling, and B. Kotak. 2017a. Spatial and temporal patterns in microcystin toxins in Lake of the Woods surface waters. Lake Reserv. Manag. **33**: 433–443.

doi:10.1080/10402381.2017.1384415

Zastepa, A., Z. Taranu, L. Kimpe, J. Blais, I. Gregory-Eaves, R. Zurawell, F. Pick. 2017b. Reconstructing a long-term record of microcystins from the analysis of lake sediments. Sci. Total Environ. **579**: 893–901. doi:10.1016/j.scitotenv.2016.10.211

Zastepa, A., J. Westrick, A. Liang, J. Birbeck, E. Furr, L. Watson, J. Stockdill, B. Ramakrishna, S. Crevecoeur. 2023. Broad screening of toxic and bioactive metabolites in cyanobacterial and harmful algal blooms in Lake of the Woods (Canada and USA), 2016–2019. J. Great Lakes Res. **49**: 134–146. doi:10.1016/j.jglr.2022.12.006

Zuur, A. F., and E. N. Ieno. 2016. Beginner's Guide to Zero-Inflated Models with R. Highl. Stat. Beginner's Guid. Ser. 414.

Tables Chapter III

Table 1. Summary statistics for part 1 and 2 of the zero-altered hurdle model.

Part 1: microcystin presence/absence binomial GLM (n=435)

Intercept	Micro	DAPH	SRP	CYCL	pH	Wind30	Pseudo-r^2
-1.17	0.69	0.37	0.49	0.33	0.54	0.35	0.34
(0.15)	(0.17)	(0.15)	(0.17)	(0.15)	(0.19)	(0.15)	

Part 2: detection-limit truncated gamma GLM (n=130)

Intercept	Micro	TP	CBB_{other}	CYCL	DIC	Temp_euph	SLOPE	Dolicho	Pseudo-r^2
-0.53	0.40	0.17	0.10	0.12	0.14	0.11	-0.16	-0.13	0.70
(0.04)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	

Regression coefficients reported with their standard error in parenthesis. Micro=$\log_{10}$(*Microcystis* biomass); DAPH=$\log_{10}$(*Daphnia* biomass); SRP=$\log_{10}$(soluble reactive phosphorus); CYCL=$\log_{10}$(cyclopoid copepod biomass); Wind30=square-root(average 30-day wind speed); TP=$\log_{10}$(total phosphorus); CBB_{other}=$\log_{10}$(cyanobacteria biomass excluding *Microcystis*, *Aphanizomenon* and *Dolichospermum*); DIC=$\log_{10}$(dissolved inorganic carbon); Temp_euph=euphotic zone temperature; SLOPE=$\log_{10}$(watershed slope within 100m of lake); Dolicho=$\log_{10}$(*Dolichospermum* biomass).

Figures Chapter III

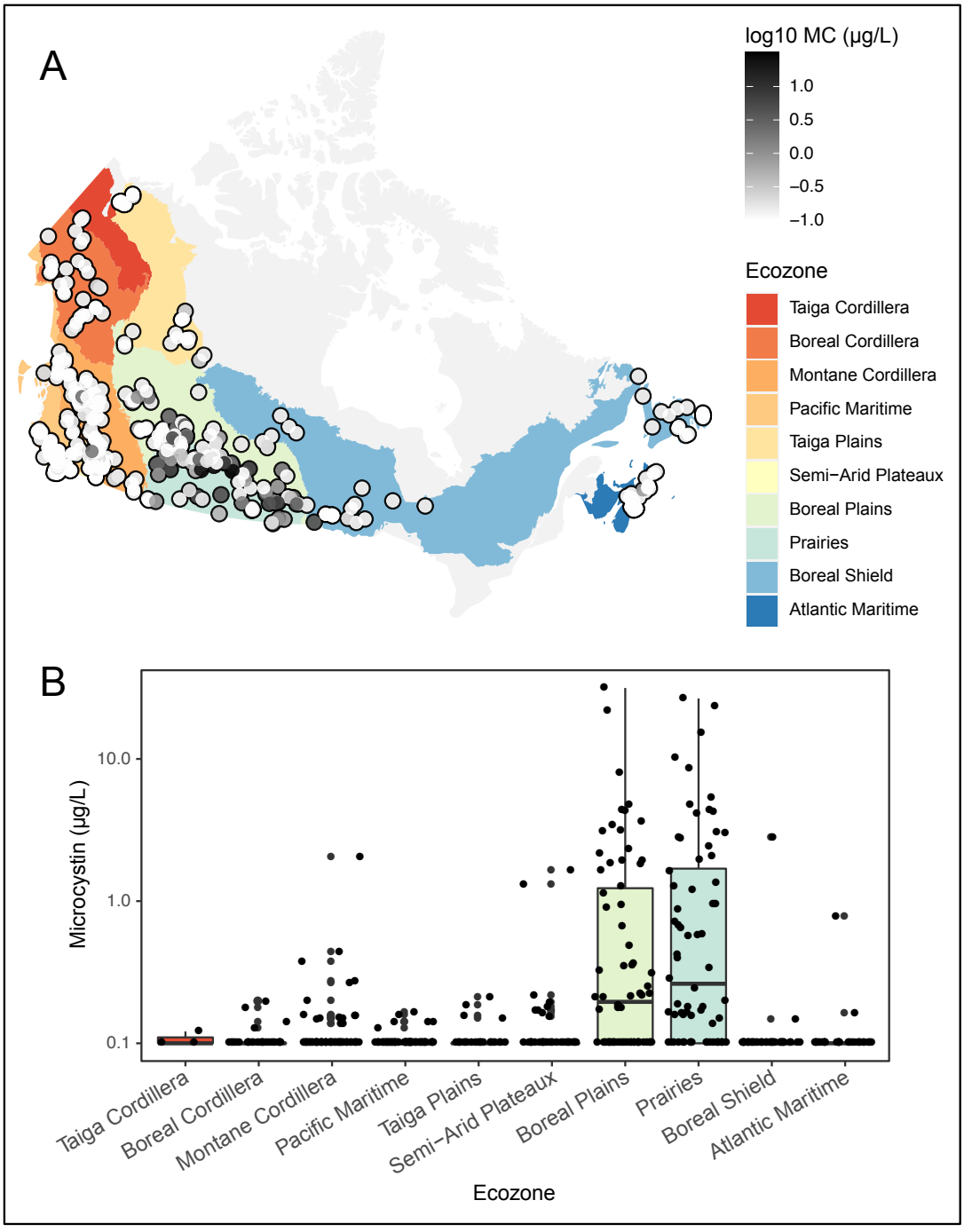

Figure 1. A) Map of log-transformed total microcystin concentrations (µg/L) across the 440-lake dataset. Ecozones are highlighted in color. B) Boxplot of log-transformed total microcystins (µg/L) by ecozone.

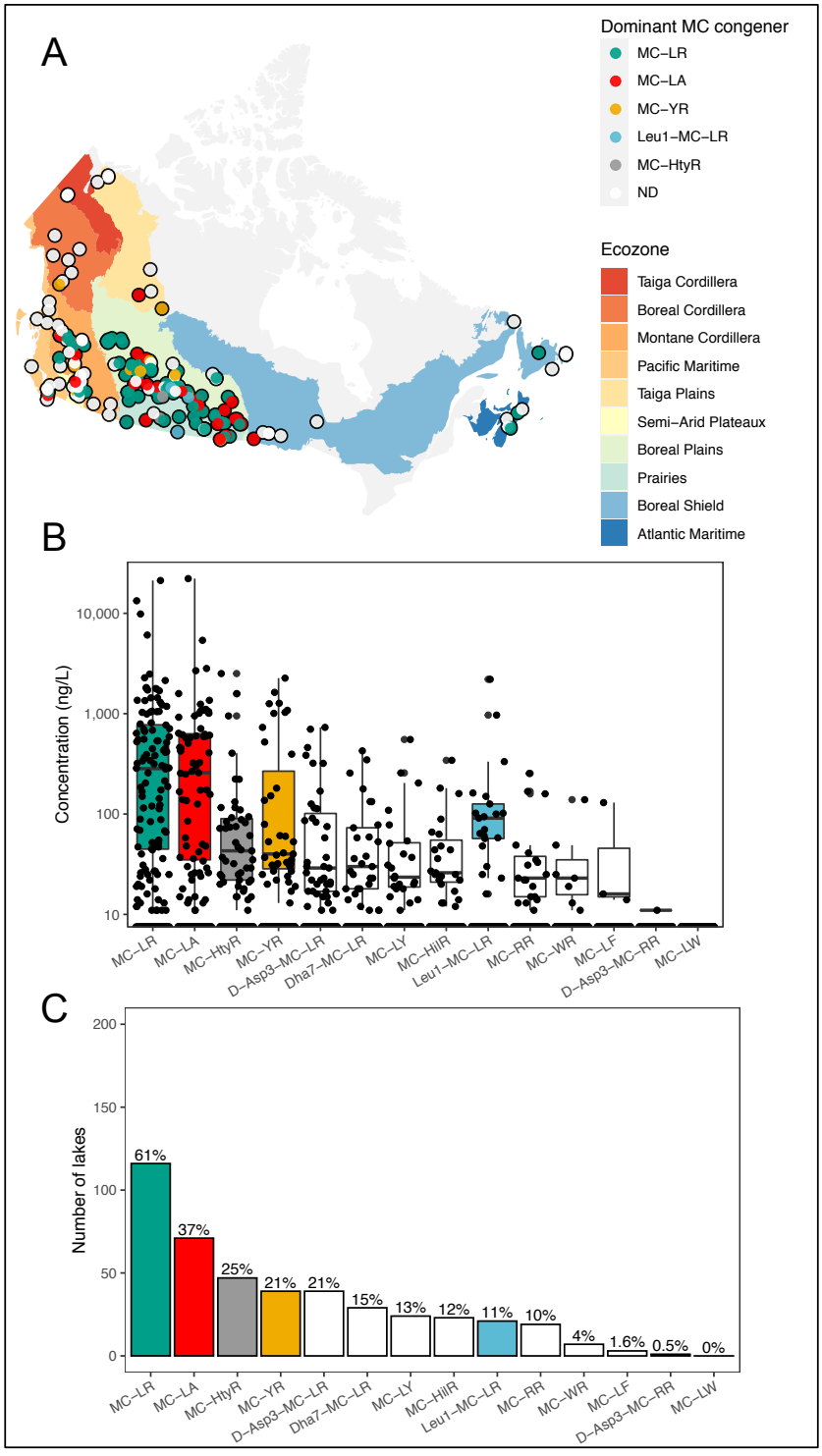

Figure 2. Map of the dominant microcystin congener present in each lake sampled in the congener analysis (n=190). Note: ND = sites with no MC congener detected. B) Boxplot of congener concentrations (ng/L). MC-LW was not detected in any of the study lakes. C)

Frequency of occurrence of microcystin congeners based on the 190 subset considered for congener analyses.

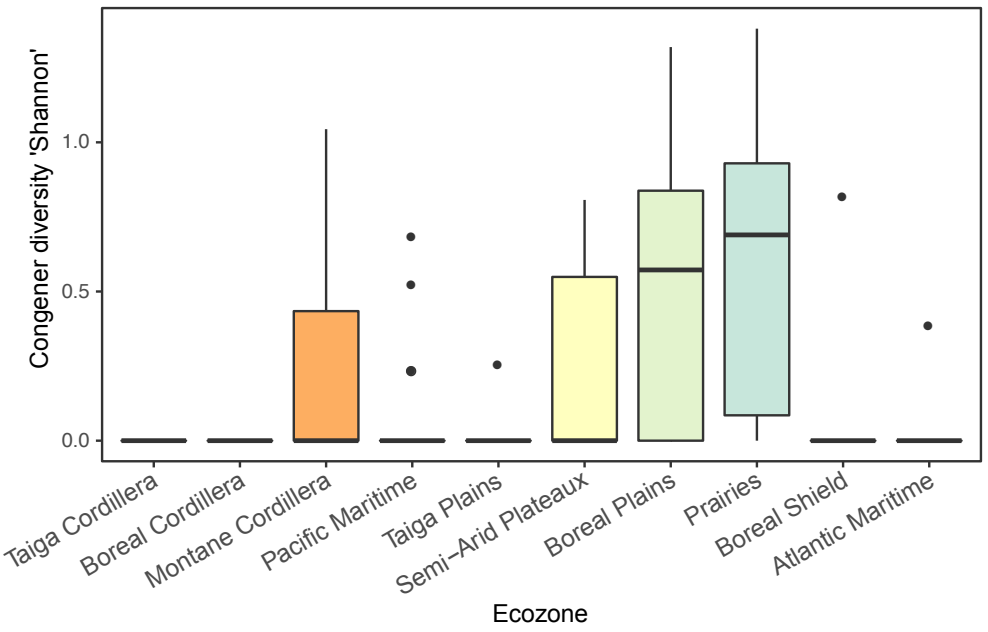

Figure 3. Boxplot of microcystin congener diversity (calculated using the Shannon diversity index) by ecozone.

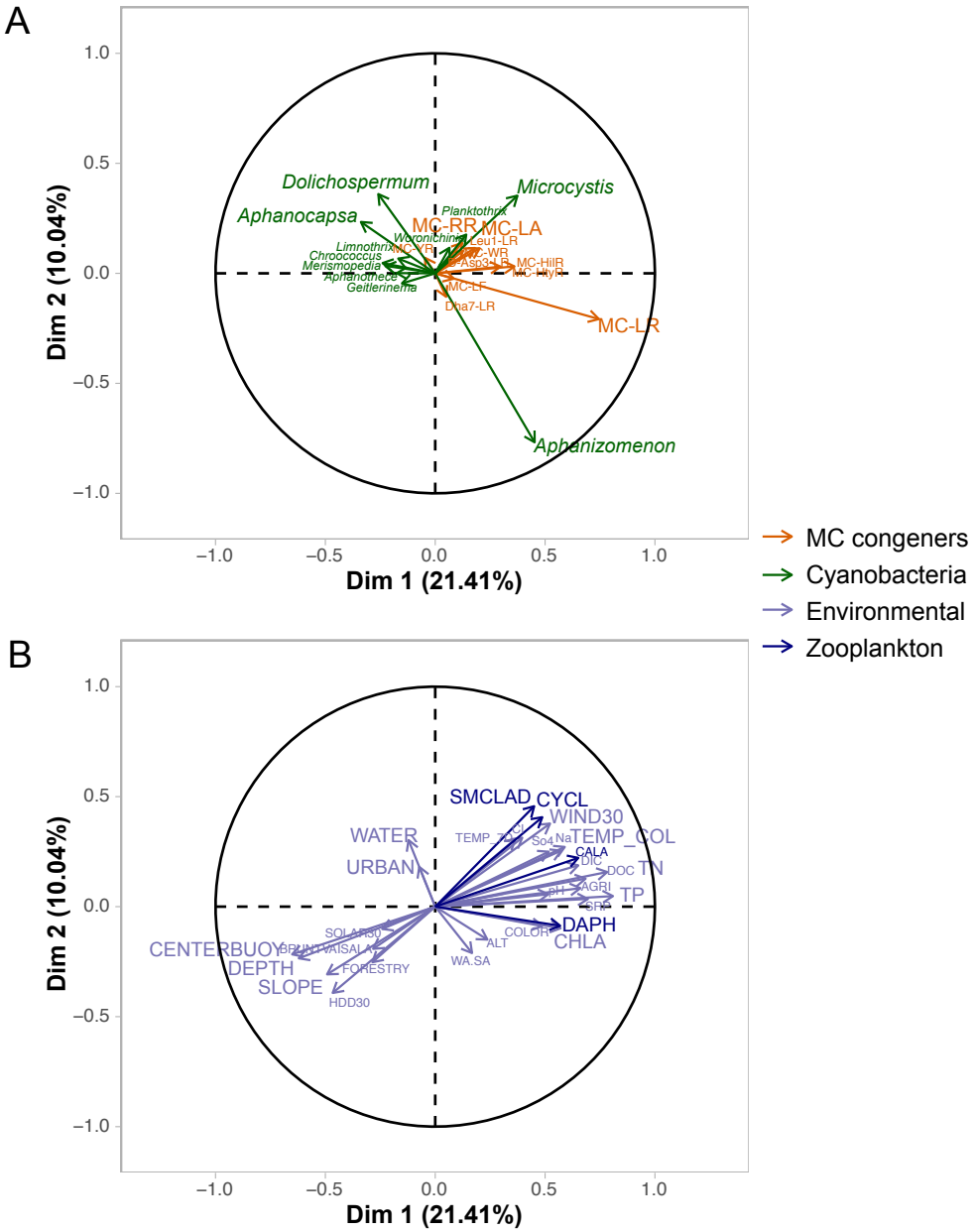

Figure 4. Multiple factorial analysis (MFA) of microcystin congeners, cyanobacteria community and environmental variables on the 190-lake dataset. The MFA was divided into two plots of A) microcystin congeners with the cyanobacteria assemblage and B) zooplankton assemblage and environmental variables.

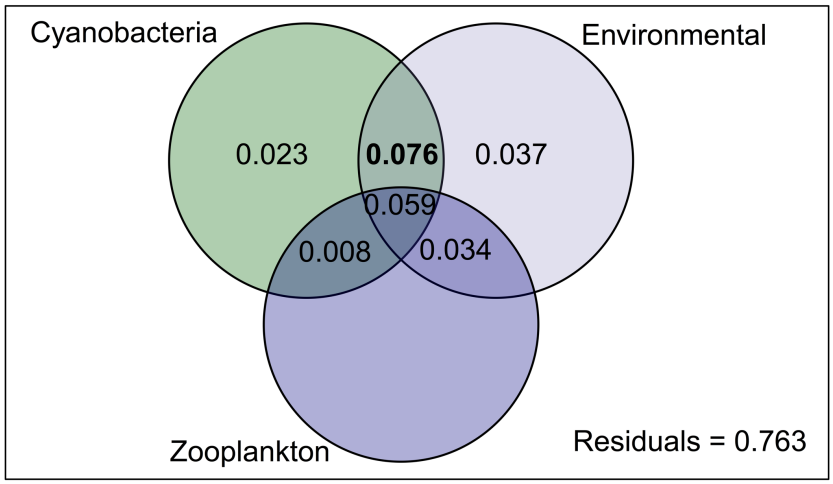

Figure 5. Variation partitioning analysis of microcystin congener relative abundances on the 190-lake dataset, constrained by cyanobacterial, environmental and zooplankton variables.

Appendices Chapter III

Table S1. Summary statistics of the environmental predictors used in the current study. Variables are grouped into five broader categories: physiography, water quality, land use, plankton and climate.

Variables	Units	Minimum	Maximum	Median
Physiography				
Lake area	km^2	0.0009	94.82	0.6736
Watershed area	km^2	0.208	37,460	14.34
Watershed area:surface area	-	1.545	11,632	16.19
Lake depth	m	0.25	151	8.6
Circularity	-	0.0079	0.9269	0.3156
Altitude	m	2	1555	568
Residence time	days	0.1	270,759	572.6
Shoreline length	m	155	218,835	5,044
Slope 100m	°	0.09	25.1	3.415
Human population	People/km^2	0	335,522	18
Water quality				
Specific conductivity	uS/cm	13.11	37,392	221.4
Temperature- euphotic	°C	11.80	29.73	19.67
Temperature- water column	°C	6.1	30	16.6
Average Brunt-Väisälä	s^{-1}	0	0.0075	0.0008
Centre buoyancy	-	0.5676	20.80	4.289
Colour	mg/L Pt	0.092	396.5	20.59
Dissolved organic carbon	mg/L	0.12	220.2	10.60
Dissolved inorganic carbon	mg/L	0.105	879	19.76
pH	-	5.51	10.28	8.42
Chlorophyll-*a*	µg/L	0.0534	381.5	2.592
Total phosphorus	µg/L	1.39	10,053	20.41
Total nitrogen	mg/L	0.025	4.358	0.3109
Soluble reactive phosphorus	µg/L	0.5	4,066	6.842
Total nitrogen: Total phosphorus ratio	-	0.5446	92.39	11.44
Calcium	mg/L	0.005	535.5	20.63
Potassium	mg/L	0.005	279.2	1.601
Sodium	mg/L	0.01	14,805	7.867
Chloride	mg/L	0.015	12,358	7.245
Magnesium	mg/L	0.005	2,249	7.444
Sulfate	mg/L	0.02	16,971	8.448
Land use				
Agriculture	%	0	86.2	0
Forestry	%	0	50.9	0.18
Mines	%	0	17.8	0

Natural landscape	%	2.13	98.4	71.9
Pasture	%	0	36.3	0
Urban	%	0	92.6	1.93
Water	%	0.07	66.8	9.9
Zooplankton				
Cyclopoid	µg/L	0	2,643	6.905
Calanoid	µg/L	0	13,403	19.17
Cladoceran	µg/L	0	11,665	29.76
Daphnid	µg/L	0	11,663	17.88
Small cladocerans	µg/L	0	2,367	1.319
Climate				
7-day average air temperature	°C	6.5959	26.25	17.07
30-day average air temperature	°C	10.09	23.06	16.89
7-day total precipitation	m	0.0002	1.141	0.1560
30-day total precipitation	m	0.0340	2.673	0.9031
7-day net solar radiation	J/m^2	2,821,571	8,046,217	5,376,189
30-day net solar radiation	J/m^2	4,260,946	7,724,438	5,774,651
7-day average wind speed	m/s	0.5323	5.30	2.052
30-day average wind speed	m/s	0.6651	5.225	2.135
7-day total heat degree days	days	0	79.19	10.39
30-day total heat degree days	days	0	239.1	57.02

Table S2. Biomass summary statistics (µg/L) for each cyanobacteria genus detected across study lakes with phytoplankton count data (n=435).

Taxon	Maximum (µg/L)	Mean (µg/L)	Median (µg/L)	% Detected in (n) lakes
Anabaena	**4,253**	**12.71**	**0**	**3.22**
Anabaenopsis	**42,400**	**110.9**	**0**	**0.46**
Aphanizomenon	626,764	15,084	0	32.2
Aphanocapsa+ Synechocystis	**6,294**	**63.77**	**0.855**	**52.6**
Aphanothece	117.6	0.684	0	2.99
Arthrospira	**160.4**	**0.369**	**0**	**0.23**
Chroococcus	2,376	13.56	0	18.9
Chrysosporum	1,467	3.372	0	0.23
Coelosphaerium	72.17	0.275	0	0.69
Cuspidothrix	1,876	4.312	0	0.23
Dolichospermum	**165,502**	**2,122**	**0**	**48.5**
Geitlerinema	8,115	20.61	0	3.91
Gloeotrichia	1,069	4.404	0	0.69
Gomphosphaeria	461.9	2.805	0	1.61
Limnothrix	14,750	105.9	0	35.2
Lyngbya	144,343	401.4	0	1.15
Merismopedia	**164.2**	**3.631**	**0**	**15.4**
Microcystis	**155,776**	**1,546**	**0**	**48.3**
Nodularia	30,282	69.90	0	0.69
Oscillatoria	**779.5**	**1.792**	**0**	**0.23**
Phormidium	**157.2**	**0.774**	**0**	**3.68**
Planktolyngbya	1,943	14.68	0	3.68
Planktothrix	**258,085**	**1,420**	**0**	**14.9**
Pseudanabaena	**10,852**	**31.17**	**0**	**13.1**
Rhabdoderma	93.8	0.293	0	1.15
Rhabdogloea	1.804	0.004	0	0.23
Romeria	237.4	0.915	0	0.92
Snowella	134.7	1.598	0	5.75
Spirulina	172.0	1.018	0	0.69
Synechococcus	**8.554**	**0.035**	**0**	**2.99**
Woronichinia	3,586	16.38	0	13.8
Total cyanobacteria	**681,497**	**21,060**	**232.0**	**92.9**
Total potential MC-producing	**258,146**	**5,314**	**86.07**	**87.1**

Aphanocapsa and *Synechocystis* were counted together. The minimum for each genus was 0 µg/L. Potential MC producing genera are bolded, based on list in (Chorus and Welker, 2021).

Table S3. Microcystin congener summary statistics (ng/L) for each measured variant across the subset of lakes chosen for congener analysis (n=190 lakes). Total microcystin concentrations based on ELISA (µg/L) for the full dataset (n=440 lakes).

Microcystin congener	Minimum (ng/L)	Maximum (ng/L)	Mean (ng/L)	%Detection (n=190)
MC-LR	0	21,220	550	61
MC-LA	0	22,150	309	37
MC-HtyR	0	2,510	41.4	25
MC-YR	0	2,260	66.1	21
D-Asp3-MC-LR	0	730	22.8	21
Dha7-MC-LR	0	426	11.5	15
MC-LY	0	553	8.83	13
MC-HilR	0	343	7.02	12
Leu1-MC-LR	0	2,200	25.7	11
MC-RR	0	254	5.06	10
MC-WR	0	139	1.47	3.7
MC-LF	0	130	0.842	1.6
D-Asp3-MC-RR	0	11	0.058	0.53
MC-LW	ND	ND	ND	ND
	Minimum (µg/L)	**Maximum (µg/L)**	**Mean (µg/L)**	**%Detection (n=440)**
MC-total (ELISA)	0	31.6	0.61	30

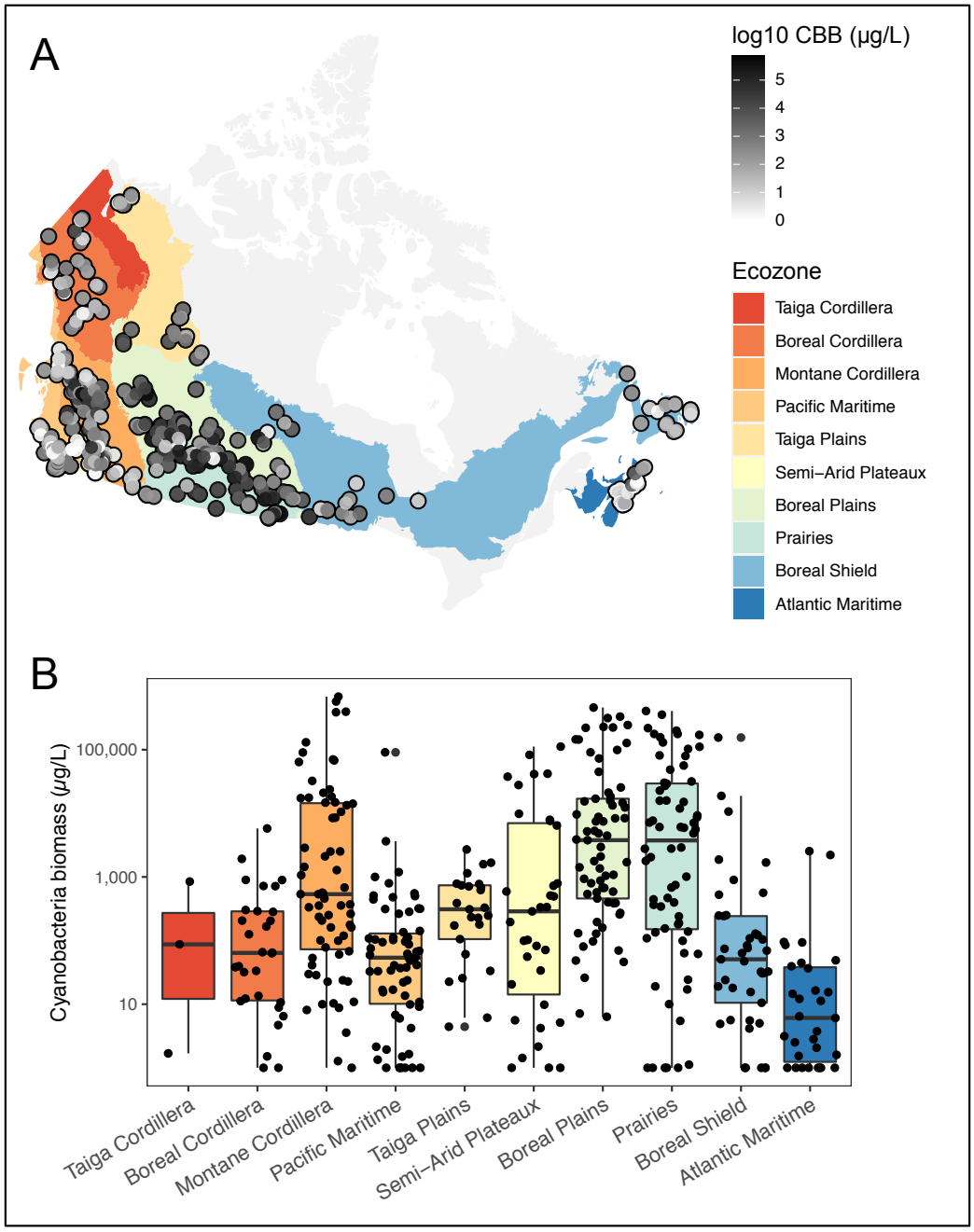

Figure S1. A) Map of log-transformed total cyanobacteria biomass (CBB) across the 440-lake dataset and B) boxplot of CBB by ecozone.

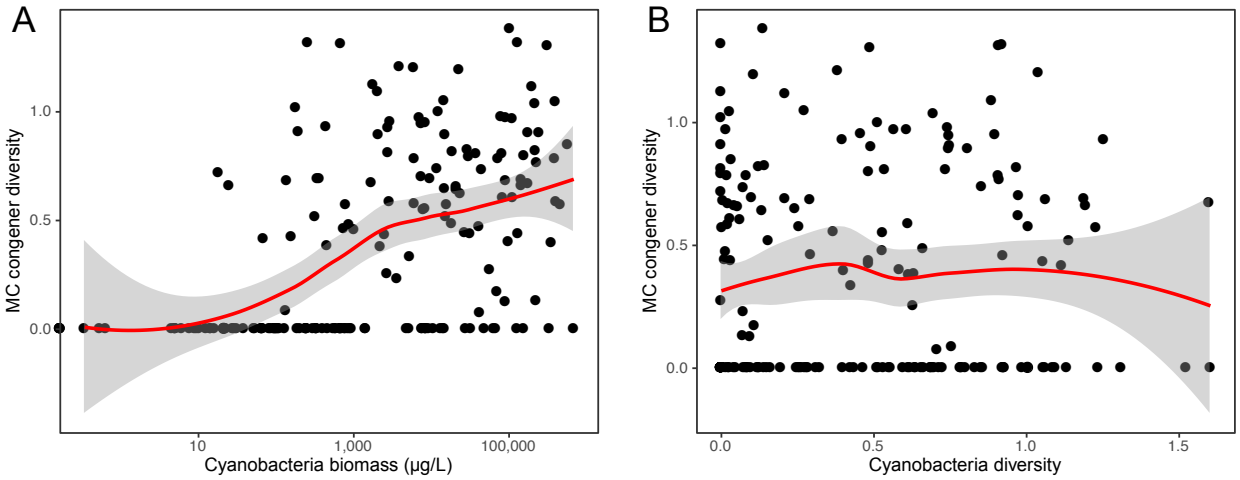

Figure S2. Relationship between microcystin congener diversity (calculated using the Shannon index) and total cyanobacteria biomass (µg/L) (A) and cyanobacteria Shannon diversity (B). Trends fit using a LOESS curve.

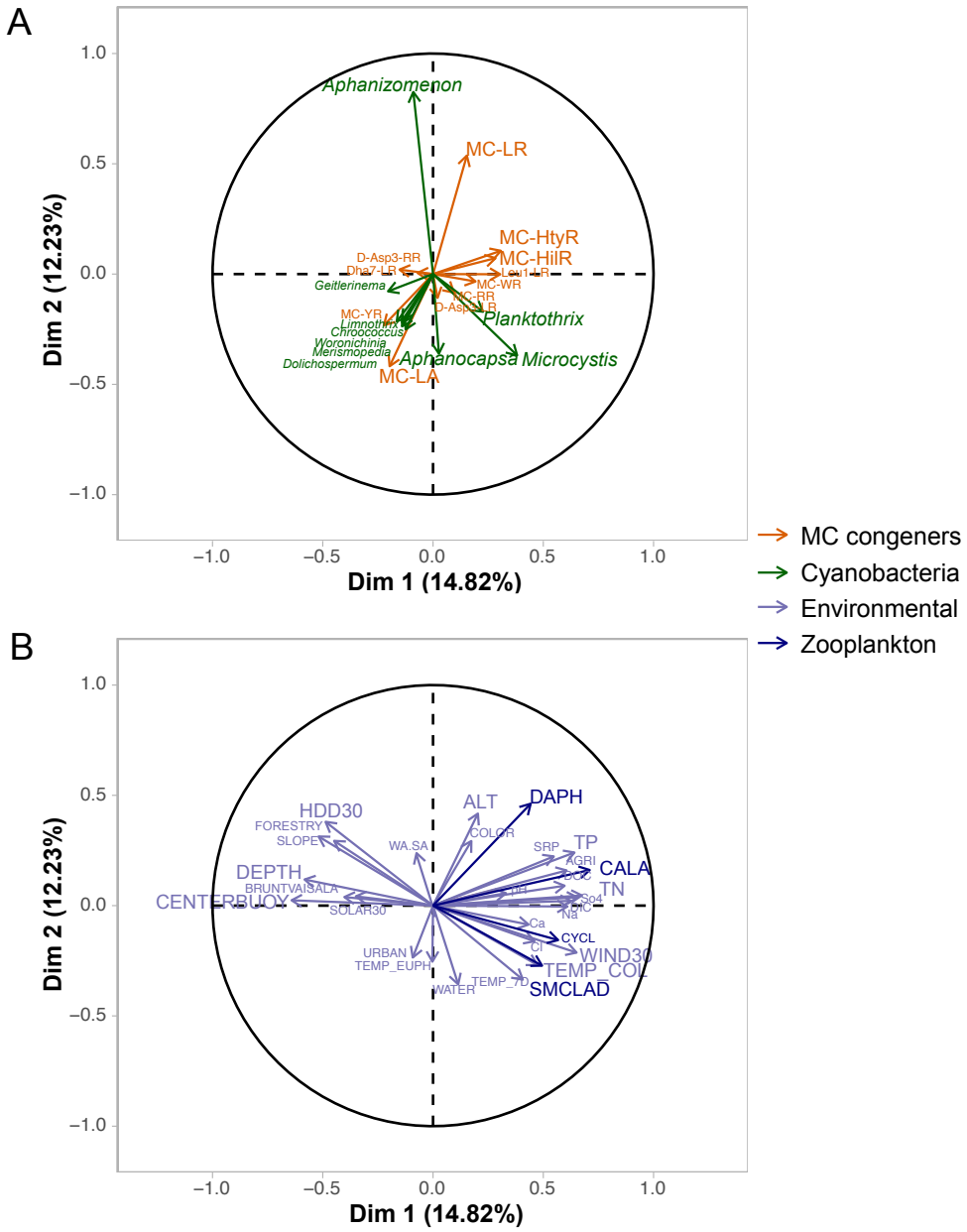

Figure S3. Multiple factorial analysis (MFA) of microcystin congeners, cyanobacteria and zooplankton community, and environmental variables for lakes where at least one congener was detected (n=123). The MFA was divided into two plots of A) microcystin congeners with the cyanobacteria assemblage and B) zooplankton assemblage and environmental variables.

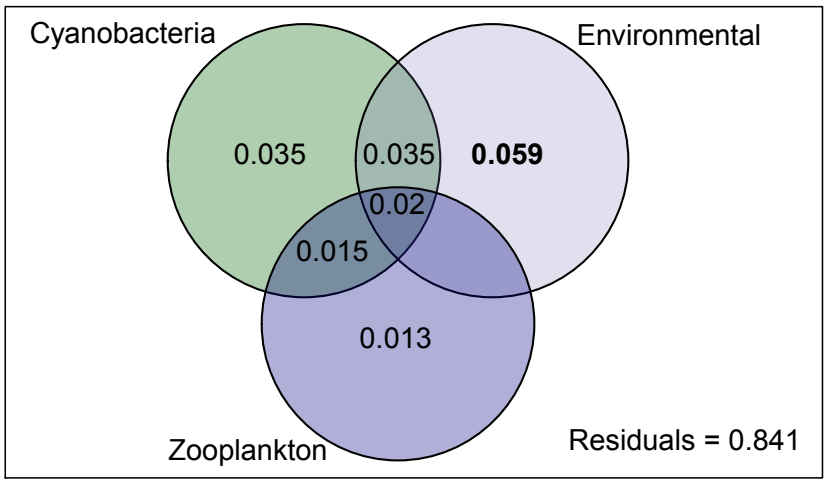

Figure S4. Variation partitioning analysis of microcystin congener relative abundances on the 123-lake dataset (sites with at least congener detected), constrained by cyanobacterial, environmental and zooplankton variables.

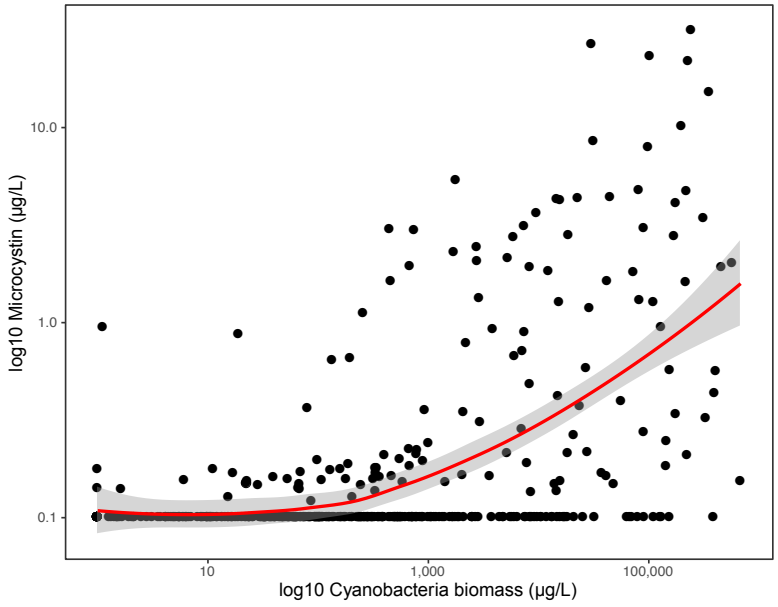

Figure S5. Relationship between total microcystin concentration (µg/L) and total cyanobacteria biomass. Trend fit using a LOESS curve.

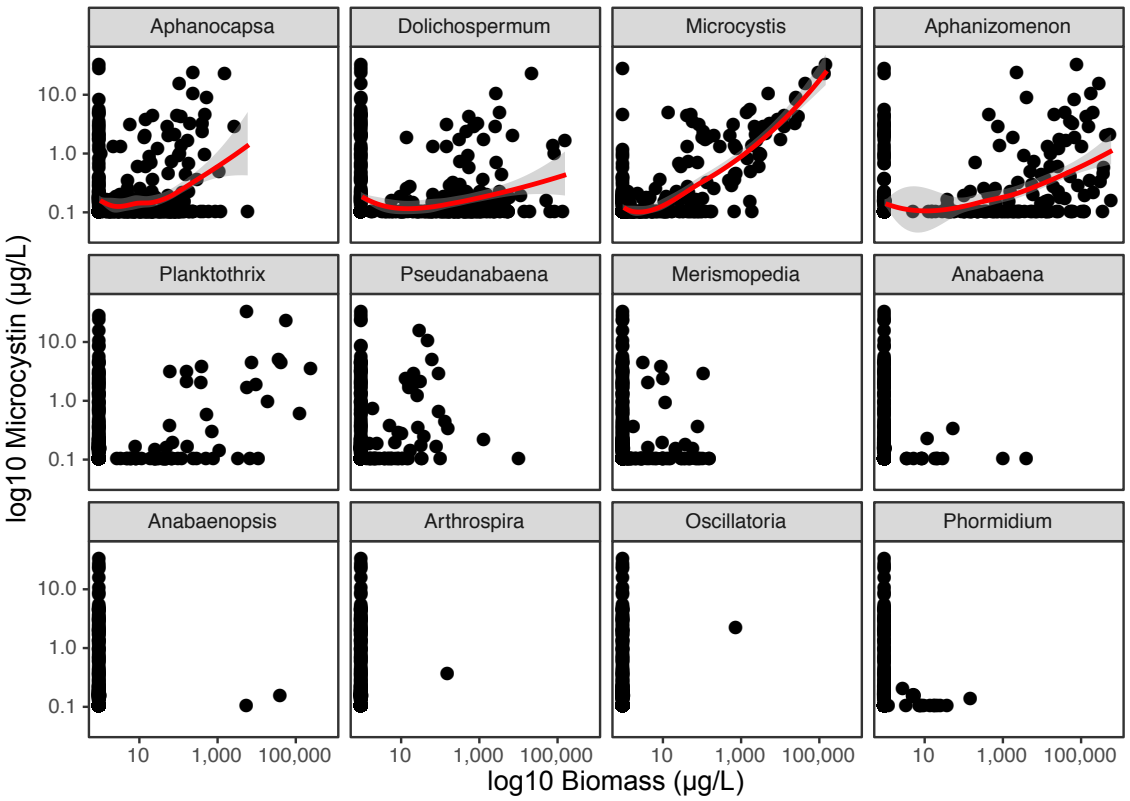

Figure S6. Relationship between total microcystin concentrations (µg/L) and the biomass of several cyanobacteria faceted by genus. Trends fit with a LOESS curve.

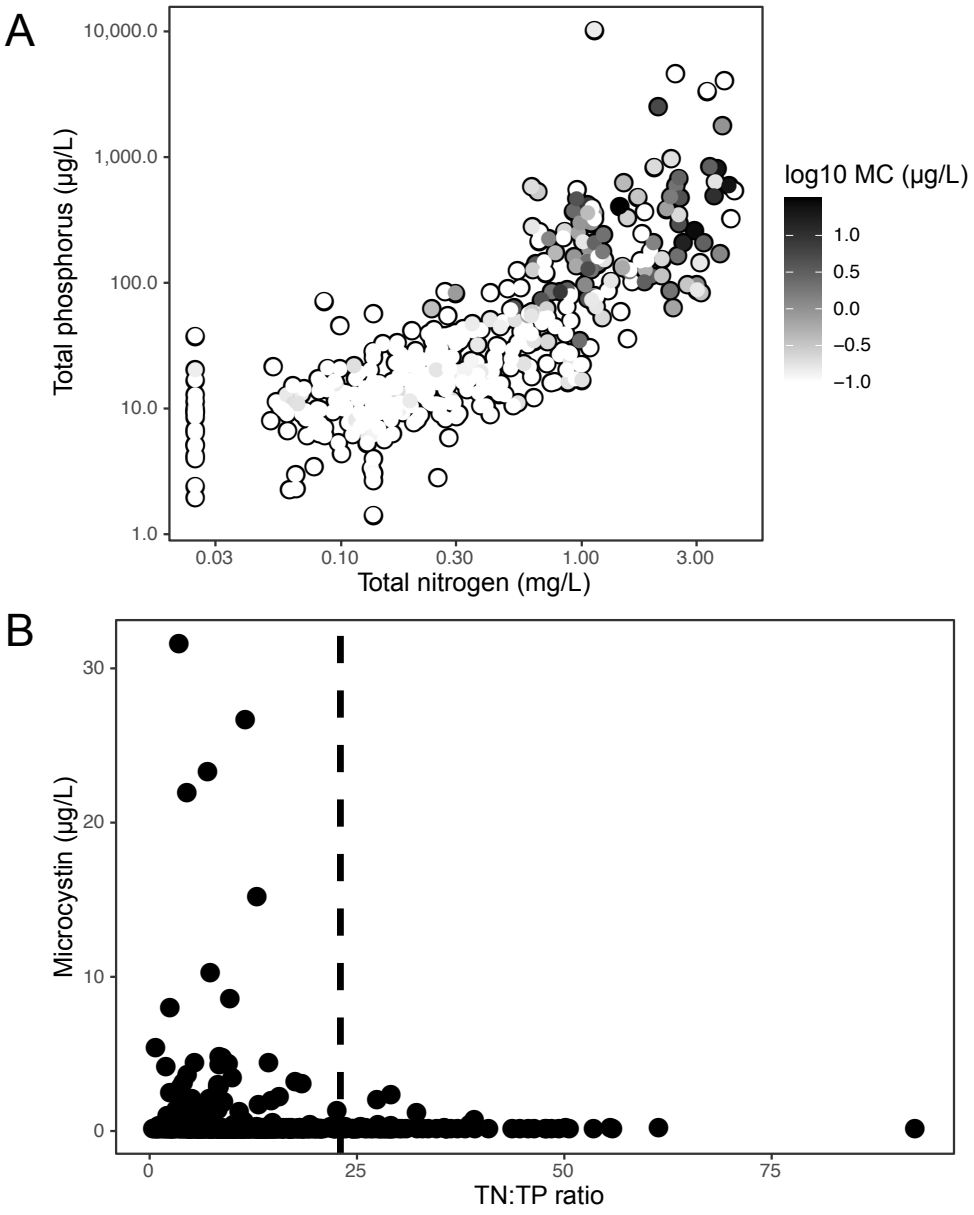

Figure S7. Relationship between total phosphorus (µg/L) and total nitrogen (mg/L), with points colored by total microcystin concentration (µg/L) (A), and the relationship between microcystin concentration and the molar ratio of total nitrogen to total phosphorus (B). The black dashed line indicates the 23:1 ratio.

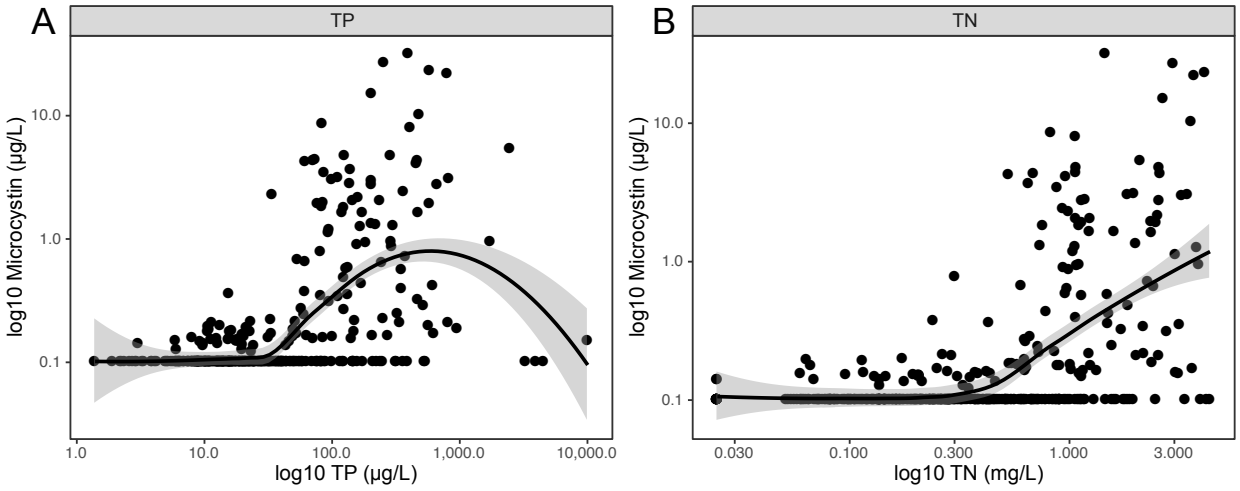

Figure S8. Relationship between total microcystin concentrations (µg/L) and total phosphorus (TP- µg/L) (A) and total nitrogen (TN- mg/L) (B). Trends fit with a LOESS curve.

COMPREHENSIVE SCHOLARLY DISCUSSION AND CONCLUSIONS

Lakes are of paramount importance to the health of humans and wildlife, providing enumerable ecosystem services (e.g., water for drinking, agriculture and recreation). Among the many recognized hazards that threaten lake biodiversity and their security as a water supply, the excessive growth of cyanobacteria remains one of the most consequential stressors (Reid et al. 2018; Chorus and Welker 2021). Although cyanobacteria have been a natural component of aquatic ecosystems for some 2 billion years, there is mounting evidence that nutrient enrichment associated with agricultural and urban development are exacerbating their dominance in many lakes worldwide, including in Canada (Taranu et al. 2015; Pick 2016). Due to their negative ecological impacts and toxicity to humans and animals, there is a global effort to improve our understanding of cyanobacterial dynamics, particularly in the context of global change (Burford et al. 2020). This served as the primary source of motivation for this thesis. The main research goal was to provide a contemporary understanding of the distribution and predictors of cyanobacteria and their toxins across Canadian lakes. With the first pan-Canadian standardized lake sampling program, I sought to address three core objectives in three chapters. In the first chapter (Chapter I), I identified the best predictors of cyanobacteria biomass sampled across Canada, from an initial set of over 50 biotic and abiotic potential predictor variables. Second (Chapter II), I turned to the methods for quantifying the composition of cyanobacterial communities and assessed the congruency between traditional microscopy and modern DNA metabarcoding. In the last chapter (Chapter III), I quantified the concentration of total microcystins and many of the individual congeners; with these datasets I identified the most parsimonious set of predictors. Collectively, the empirical models and methodological

comparison developed in this thesis provide insights that can be used directly by lake managers in both the monitoring and mitigation of cyanobacteria. For example, I recently participated in an online symposium with lake managers from the state of California who were interested in the findings published in chapter 2. Together, the chapters in this thesis generated original contributions towards advancing our understanding of cyanobacteria growth and the current state of cyanobacterial patterns in Canada. However, many of the findings could inform decision making about temperate to subarctic lakes globally.

Significance of findings and original contributions

In Chapter I, my data visualization and modeling provides a quantitative portrait of total cyanobacteria biomass and its composition from 640 lakes across Canada. This work revealed regional hotspots, with the highest concentrations of cyanobacteria occurring centrally, within the Prairies and Boreal Plains ecozones. Lakes in these regions were dominated by common bloom-forming and potential toxin-producing taxa including *Microcystis*, *Aphanizomenon* and *Dolichospermum*. Additional areas with elevated biomass include the Montane Cordillera in western Canada and the Mixedwood Plains in the southeast part of the country. Our empirical models corroborate earlier work and show the overwhelming significance of nutrients, particularly TP, as the leading predictors of cyanobacteria biomass. This finding extends on our current knowledge of cyanobacterial drivers, as several pervious large-scale models also emphasize the importance of nutrients (Downing et al. 2001; Beaulieu et al. 2013; Carvalho et al. 2013). However, these models are often had access to only a select number of potential predictors and explained low amounts of variation. Despite the inclusion of a broader range of biotic and abiotic predictors, and developing a dataset entirely enumerated by a single and

highly-trained taxonomist, TP was still the most significant, exhibiting a non-linear log-log relationship with total biomass (back-transformed, this relationship increased until ~100 µg/L of TP, then tapered). The overwhelming importance of nutrients highlights that lake management strategies should continue to focus on nutrient reduction strategies when controlling cyanobacteria. Our findings also identified thresholds between biomass and significant predictors, which in the case of TP, demonstrated an abrupt increase cyanobacteria biomass above 40 µg/L of TP. By expanding the range of possible predictors of cyanobacteria, a more novel finding was the significance of zooplankton biomass. Contrary to the traditional viewpoint, we detected positive relationships between cyanobacteria and multiple zooplankton groups from small, specialized grazers like copepods, to large generalists like *Daphnia*. Although a range of possible mechanisms could be enabling these positive associations, zooplankton interactions should definitely be considered in future studies. Along with having constructed models for total cyanobacteria, we accounted for distinct ecological niches within the cyanobacteria community and identified the predictors of specific genera of interest. We observed that some of the top predictors vary by genus, which emphasizes the need to take into account specific variables depending on the dominant genera in a given lake. Interestingly, TP was the only variable significant in all genus-specific models, albeit at different critical thresholds. Collectively, our analyses presented in Chapter I identified the relative influence of physical, chemical and biotic predictors on cyanobacterial biomasses. Due to the standardization of variable collection and limited regional variation within models, these findings offer widescale applicability and may be used to estimate cyanobacteria biomass across temperate regions.

In Chapter II, I focused on how two methods for quantifying the composition of cyanobacterial communities compare: traditional microscopy and modern DNA metabarcoding.

Our understanding of the environmental variables that promote the proliferation of cyanobacteria and the ability to construct models for specific genera (Chapter I) relies upon accurately characterizing their assemblages. However, acceptance of molecular methods requires benchmarking against the traditional morphological-based taxonomic approach (Pawlowski et al. 2021). With the ongoing use of both methods, there is a necessity to assess the level of congruency between the communities generated by each approach. I used a subset of 379 lakes from Chapter I, as an opportunity to compare cyanobacteria communities from samples taken in parallel for both microscopy and DNA metabarcoding. This comparison represents an original contribution as there has yet to be a comparative analysis across a wide suite of lake types. Furthermore, lake managers can use these findings directly when deciding which taxonomic method to use. Our findings in Chapter II demonstrated a broad level of congruence between methods, with a clear separation of similarity among platforms among lakes of different trophic states. Mainly, microscopy and DNA metabarcoding generated more similar assemblages in eutrophic and hypereutrophic lakes. This has management implications as it provides support for the use of DNA to study and track blooms, of which are characteristic of eutrophic surface waters. I also observed the highest level of congruence between methods at coarser levels of levels of taxonomic, beginning at the Order level, and decreased when comparing genus level assignments. The biggest difference between methods was the detection of picocyanobacteria, which were underestimated by microscopy. It remains to be seen if the similarity could be further improved up on the future when picoplankton counts are performed using fluorescence microscopy as the form of light microscopy used is known to underestimate picoplankton (which themselves are numerically important in terms of individual cells, but often not in terms of biomass). Furthermore, future work could investigate whether amplicon length varies by taxa,

which could produce a consistent difference between metabarcoding and microscopic results. To date, comparative studies of different aquatic communities have suggested that DNA gives complementary as opposed to identical results (Keck et al. 2022). In this sense, light microscopy is able to provide quantitative measures of cyanobacteria biomass, upon which management guideline are based, and identify taxa that lack representation in reference sequence databases. Meanwhile, DNA metabarcoding is able is distinguish cryptic taxa and account for picocyanobacteria, which are more dominant taxa in oligotrophic and mesotrophic lakes. The main findings of this chapter shed light on discrepancies that remain between methods such as the incomplete state of reference databases, and the technical and biological limitations that managers should consider when deciding which method to use. Furthermore, our analysis utilized a key strength of the DNA metabarcoding approach which is the detection of genotype level variation within a particular genus. We observed a separation of *Microcystis* ASVs along a trophic gradient, highlighting distinct ecological niches within this genus. Shifts in dominant genotypes has been studied in previous analyses, including *Microcystis*, but never at this scale. Earlier work has associated changes in genotypes to shifts in bloom toxicity (Chun et al. 2020; Ninio et al. 2020). Our work provides additional support for incorporating molecular approaches into routine monitoring and tool for lake managers to target their analyses on toxigenic genotypes in the environment.

In Chapter III, I shifted focus onto one of the main ecological and public health concerns from the proliferation of cyanobacteria: their production of toxins. Using the same pan-Canadian standardized lake sampling program, I conducted a spatial analysis that targeted microcystins; a potent liver toxin and the most widely found class of cyanotoxin in lakes. I found that most elevated microcystin concentrations were located in the Prairies and Boreal Plains ecozones, but

in general total concentrations were low across Canada relative to regulatory guidelines. Given that Chapter I revealed that these regions contain the highest biomass of cyanobacteria, including of several toxin producers, the pattern in toxin concentrations was expected. I used a similar statistical approach as in Chapter I, but using an expanded number of potential predictors, including a wide range of biotic and biotic variables, including zooplankton and the biomass of known microcystin producers. Our findings demonstrated once again the overwhelming importance of nutrients, mainly TP and soluble reactive phosphorus (SRP), and *Microcystis* biomass as the best predictors. To account for the high number microcystin non-detects, we utilized a statistical framework to first model the presence-absence of microcystins, followed by modelling of concentrations only from just the lakes that had detectable microcystins (30%). In both cases, the best predictors were nutrients and *Microcystis* biomass. Like Chapter I, this chapter also identified positive correlations between zooplankton variables and microcystins. The positive influence is not likely a result of microcystin representing a defense strategy for cyanobacteria that produce the toxin, but rather, we suspect that co-existing zooplankton may have developed tolerance and avoidance traits to withstand exposure to such toxins. One of the key contributions to knowledge from this chapter is that the best predictors of microcystins identified by empirical models are all directly associated with the promotion of cyanobacteria biomass. This finding has important management implications as it suggests managers can focus their efforts on controlling cyanobacterial biomass alone, as opposed to separate efforts that address deal specifically with toxin concentrations.

This third chapter also addressed new questions regarding the distribution of microcystin congeners across Canadian lakes. Congeners exhibit a range of toxicities, yet to date, most studies focus exclusively on the distribution of MC-LR. Using a newly developed method

consisting of liquid chromatography and mass spectrometry, this complementary analysis conducted the largest spatial evaluation of microcystin congeners and evaluated their predictors. Despite the known occurrence of MC-LR in every Canadian province, our findings revealed the presence of 14 congeners across Canadian lakes, most of which were detected in relatively few lakes. The two most commonly found were MC-LR and -LA, considered to be the two most toxic. The relatively high diversity of congeners across the country suggests that targeting them may provide a more accurate assessment of the level of toxicity present in a given system. Our findings also demonstrated that the composition of congeners was only moderately related to a broad suite of environmental variables and cyanobacterial community data. Our current understanding of congener occurrence in relation to environmental conditions is incomplete and highlights the need of future investigators to consider alternative predictors or examine the conditions that promote the production of particular congeners at the strain level.

Collectively, one of the most pertinent contributions from this thesis is the geographic scale at which cyanobacteria and their toxins were analyzed, and the resulting benefits to management associated with this approach. There has been a growing effort to combine science and management needs to reduce the risks associated with the proliferation of cyanobacteria (Errat et al. 2022). Large-scale sampling plays an essential role in these efforts and have recently become more common. For example, the European Multi-Lake Survey (EMLS) sampled over 360 lakes across 26 countries and has generated valuable information regarding the distribution of cyanotoxins and their drivers (Mantzouki et al. 2018b; Mantzouki and Ibelings 2018). The National Lakes Assessment (NLA) has sampled over 1,000 water bodies across the United States multiple times since 2007 and was used to build some of the earliest large-scale cyanobacteria and microcystin predictor models (Beaulieu et al. 2013; Taranu et al. 2017; Pollard et al. 2018).

Since cyanobacteria are a global phenomenon, there is a need to collect data at wide spatial scales and study their responses to environmental gradients across many regions and lake types (Mantzouki et al. 2018a; Mantzouki and Ibelings 2018). We adopted this framework and provided the first standardized sampling program of cyanobacteria and their toxins across Canadian lakes. There were numerous benefits associated with this national scale integration of data, including the standardization of sampling, which reduced collection biases and produced synchronic data in highly comparable datasets (Mantzouki et al. 2018a; Pérez-Jvostov et al. 2019). The centralization of several analyses and the quality control of data promoted efficient data sharing and increased accessibility to many users. Lastly, the standardization of sample collection is highly significant to the development of robust models that can be applied to larger scales. This is particularly relevant in Canada, as although water guidelines are standardized nationally, cyanobacteria management strategies vary drastically by province (Rashidi et al. 2021). The empirical model from this thesis, exhibited limited regional variation and utilized variables that were all collected using standardized protocols, including a single taxonomist. This promotes their use in management programs across the country.

A second key contribution is the consistency across empirical models, in emphasizing the role of phosphorus as the best predictor of cyanobacteria and microcystins. Across chapters, statistical models were constructed for total cyanobacteria, specific bloom-forming genera, their community composition as well as the presence and concentrations of microcystin. Despite considering an expanded range of biotic and abiotic variables, phosphorus was still the leading predictor across model types, including over nitrogen variables. Traditionally, management efforts to mitigate cyanobacteria blooms have centered around nutrient enrichment, with a particular emphasis on regulating phosphorus inputs (Paerl and Barnard 2020; Chorus and

Welker 2021). Support for phosphorus as the key nutrient for reverting eutrophication stems from field studies conducting in the Experimental Lakes Area originating in the 1970s (Schindler et al. 2008). Here, an experimental lake (Lake 227) was fertilized with constant annual inputs of phosphorus and decreasing inputs of nitrogen over the course of several decades. Despite halting nitrogen fertilization, cyanobacteria abundance remained elevate due to their ability to fix atmospheric nitrogen, thus compensate for the reduced loading. This led to conclusion that nitrogen management is futile and efforts should be focused on curtailing phosphorus inputs. In recent decades, this has challenged and debated extensively in the literature, with several researchers arguing for dual nutrient management. This is based on several authors noting that nitrogen limitation is common in many shallow water bodies, and that the fixation of nitrogen by some cyanobacteria is unable to meet ecosystem demands (Conley et al. 2009; Scott and McCarthy 2011; Paerl et al. 2016). Furthermore, there has been both heavy phosphorus and nitrogen pollution from a variety of human activities, including increased nitrogen loading to many lakes worldwide (Bogard et al. 2020; Tanvir et al. 2021). Large-scale models have provided important data and insights to this ongoing debate. For example, Beaulieu et al. (2013) found nitrogen to be a better predictor of cyanobacteria biomass in over 1,000 water bodies across the continental U.S. and argued that this could be due to intrinsically higher nitrogen demands in cyanobacteria. Others have noted the significance of phosphorus, including as the best predictor of microcystin concentrations (Giani et al. 2005; Wagner et al. 2019). However, cyanobacteria exhibit a non-linear increase with phosphorus, whereas its relationship with nitrogen increases linearly (Downing et al. 2001; Dolman et al. 2012). Across Canadian lakes, our models provide strong support for controlling phosphorus as the limiting nutrient to cyanobacteria. This does not exclude the importance of nitrogen, in fact, nitrogen alone still

explained a significant amount of variation in univariate biomass models. However, the consistency in response models highlights the close association between the cyanobacteria community, microcystins and phosphorus across Canada.

Current limitations and future considerations

Overall, this thesis provides a broad overview of the distribution and predictors of cyanobacteria and their toxins in Canadian lakes. Together, the chapters provide data in regions where it was previously missing, and the standardization of sampling allow the results of our empirical models to be widely applied. The research advances our understanding of cyanobacteria in temperate regions and contributes relevant information towards emerging management issues. Nonetheless, there are still considerable gaps in knowledge and questions this thesis was not able to consider directly. Furthermore, we generated a plethora of different pieces of data that can be used in the future studies to address many of these gaps.

Although we identified the many benefits of snap-shot sampling, there are limitations to the multi-lake approach that may have restricted explanatory power in our models. For example, time-series analyses on individual, or few select lakes are able to capture seasonal shifts in cyanobacteria community composition and identify idiosyncratic trends specific to each lake. As a result, these analyses may explain more variation than do large-scale studies (Cremona et al. 2018). In a future statistical analysis, it would be ecologically informative to decipher the direct and indirect roles of the drivers of cyanobacteria biomass using the predictors we identified as most significant. Others have made attempts to clarify the directionality of predictors using structural equation models (Beaulieu et al. 2013; Shan et al. 2019; Amorim et al. 2020). This approach may provide further evidence to lake managers regarding which variables to focus on

regulating; those being the ones that have a direct effect on cyanobacteria biomass. This may also help identify whether the positive correlation between zooplankton and cyanobacteria biomass we observed was direct or if it was moderated by other variables. For example, Shan et al. (2019) noted a positive relationship between zooplankton biomass and the biomass of multiple bloom-forming genera, but that this was partially mediated through increasing water temperature. Overall, there is no single monitoring strategy that is perfect, but the development of improved models with increased predictive power may be bolstered with yearly, repeated sampling of lakes (Mantzouki et al. 2018a).

Another limitation from our sampling approach is the spatial coverage in the water column. Most research on cyanobacteria blooms focus on near-surface samples, as was the case in the LakePulse sampling protocol. However, some cyanobacteria taxa are capable of occupying the metalimnion and benthic regions of the water column. The omission of subsurface taxa is an emerging management issue within the cyanobacteria community, as over 20 benthic species are known to produce an array of cyanotoxins, including but limited to microcystins, anatoxins and saxitoxins (Chorus and Welker 2021; Errat et al. 2022). The death of dogs and livestock has been associated with exposure to toxins in benthic mats, even in clear water and mesotrophic lakes (Fastner et al. 2018; Chorus and Welker 2021). By not accounting for these, we are potentially underestimating cyanobacteria risk in a given water body. Furthermore, there is evidence that toxic benthic taxa appear to be increasing in many lakes, largely as a result warming temperatures and longer drought periods (Chorus and Welker 2021). Given the ecological and health related issues associated with benthic cyanobacteria, there is a need to include them in monitoring efforts and to investigate their drivers. Across our dataset, the microscopic identification of cyanobacteria was typically based on the top 2m of the water column, thus

missing many benthic taxa. However, as noted in Chapter II, DNA metabarcoding is a much more sensitive method and was able to pick up on this component of the cyanobacteria community. In fact, many of the taxa that were only identified by DNA metabarcoding were benthic taxa, including *Calothrix*, *Leptolyngbya*, *Limnolyngbya* and *Tychonema*. This provides further support for the inclusion of molecular methods in routine monitoring and highlights its complementary role to traditional microscopy. In addition to benthic taxa, cyanobacteria within the metalimnion are often omitted. An interesting future study would be to analyze BBE Fluoroprobe profiles that were taken across the full water column in over 100 lakes from our dataset. Based on spectral fluorescence, Fluoroprobe profiles quantify the concentration of several algal groups, including cyanobacteria, and identify how they distribute throughout the water column. This would reveal how common metalimnetic blooms are across lakes and capture this component of cyanobacteria assemblages. This future study could also investigate the environmental variables that promote the formation and biomass of metalimnetic blooms. The combination of DNA metabarcoding and fluoroprobe profiles would provide a much more completed understanding of the full cyanobacteria community across Canadian lakes and complement the photic zone water samples that were taken across all LakePulse lakes.

In light of the findings from Chapter II, which showed that DNA metabarcoding could become more widely applied to monitor eutrophication, advances in molecular tools offer unprecedented opportunities to study cyanobacteria communities and their dynamics (Burford et al. 2020; Wells et al. 2020). The LakePulse Network collected DNA and RNA samples from most of its lakes, of which can be leveraged to access gene regulatory networks of key traits and genotype level information, including niche requirements and toxicity. Cyanobacteria exhibit variability in strain level responses to environmental conditions, altering the genetic composition

of communities (Harke et al. 2016; Chun et al. 2020). Future studies should incorporate these data to account for the acclimation and adaptation of different strains and identify the environmental conditions that select for specific genotypes.

A key limitation that may be addressed using molecular data collected by the LakePulse Network, is to investigate further the environmental conditions that regulate the production of toxins. To date, several environmental factors have been shown to increase or decrease the transcription of genes responsible in toxin production, including light, iron and nitrogen concentrations (Kaebernick and Neilan 2001; Harke and Gobler 2013), although results have not been consistent among studies. Chapter III revealed the relatively high diversity of microcystin congeners across Canadian lakes, however, their occurrences were only moderately related to environmental variables and the full cyanobacteria community as predictors. Correlational analyses alone may be constrained in identifying the complex factors that lead to toxin production, as the presence of toxigenic taxa is a pre-requisite for production but does not guarantee toxicity (Buley et al. 2022). Within a bloom of a single species, there may toxigenic and non-toxigenic strains present, which cannot be distinguished by morphological identification (Pick 2016). To verify the presence of and expression by toxigenic strains, a follow-up study could complement the identification of potential toxin producers by amplifying gene fragments responsible for toxin production. Identifying the presence and distribution of toxigenic genotypes could lead to a better understanding of the environmental variables that regulate their occurrences and the respective congeners that they are producing. Nonetheless. the initial correlational analyses and predictive models we developed for microcystins and their congeners represent important knowledge that can be rolled out immediately in terms of policy actions. We now know that across Canadian lakes, phosphorus and the biomass of potential producers (i.e.,

Microcystis) were the top predictors of total microcystin concentrations. Knowledge regarding the regulation of microcystins by environmental factors would be essential to future management efforts and may elucidate the physiological function (if any) toxins possess (Boopathi and Ki 2014).

Although microcystins are the most widespread group of cyanotoxins in lakes worldwide, cyanobacteria are known to produce hundreds of additional secondary metabolites. Several of which exhibit similar and even greater toxicities. Yet currently only a small fraction are frequently monitored (Janssen et al. 2019). Large-scale analyses in freshwater that expand their list of targeted cyanobacteria metabolites often reveal more than just microcystins (Loftin et al. 2016; Mantzouki et al. 2018b). As such, their exclusion, like with missing benthic taxa, may be underestimating the health risk from toxins in a given lake (Merel et al. 2013; Janssen et al. 2019; Errat et al. 2022). To account for this, I plan on collaborating with project partners from Environment and Climate Change Canada to provide more representative profiles of the broader suite of cyanotoxins present across Canada, using the same subset of LakePulse sites that were analyzed for microcystin congeners. Using a similar quantification method to the microcystins, we have sampled for 15 additional alkaloid and cyanopeptides, which will be related to environmental and cyanobacteria community data in a future modelling exercise.

A final aspect of this thesis that should be considered in a future study is central to the goals of the LakePulse Network in understanding the health status of Canadian lakes (Huot et al. 2019). LakePulse defines “health status” as a lakes departure from its natural state due to anthropogenic stressors. These alter the lakes ability to provide ecosystem services that it formerly could (Huot et al. 2019). To address this, lakes could place in their historical context by analyzing shifts in bioindicators, geochemical trends and different communities of interest

throughout sediment cores. The LakePulse Network collected full sediment cores from over 100 lakes across Canada, of which have been dated. In the context of cyanobacteria, it is often reported that cyanobacteria blooms are increasing, however, there is generally a paucity of long-term monitoring records to place contemporary phytoplankton assemblages in their historical context. Through the rapid development of high-throughput sequencing technologies, an increasingly common method of reconstructing past communities in paleolimnological analyses are DNA-based approaches (Domaizon et al. 2017; Capo et al. 2021). Sedimentary DNA has been shown to record community compositional trends that match historical phytoplankton counts from the water column (Monchamp et al. 2016). And as observed in Chapter II, DNA can be applied to monitor blooms and lakes that have undergone eutrophication. The advancement of sedimentary DNA techniques and their successful application offers a unique opportunity to explore historical cyanobacteria occurrences as well as community composition, including from lakes or regions where there is little long-term data available. This approach is increasingly being used to track how anthropogenic stressors, primarily climate change and eutrophication, have historically influenced cyanobacteria abundance and community structure in lakes worldwide (e.g., Pal et al. 2015; Pilon et al. 2019; Yan et al. 2019; Picard et al. 2022). For example, Pal et al. (2015) used quantitative PCR and sedimentary pigments to analyze shifts in cyanobacteria abundance over the past 150 years in five lakes located within and outside a protected area. The abundance of cyanobacterial gene copies (16S rRNA gene) had significantly increased in all lakes, particularly in the last 30 years. The authors concluded that this was most likely in response to regional climate warming and landuse changes for lakes outside the protected area. I plan on adding to the contemporary assessment of cyanobacteria captured in this thesis by examining their temporal dynamics from a selection of lakes and making use of traditional

paleolimnological proxies with cyanobacteria communities reconstructed from sedimentary DNA. This approach also offers an opportunity to build off findings from Chapter I, to see how historical zooplankton populations respond to changes in cyanobacteria abundance and community composition. Using Cladoceran subfossils preserved in lake sediments, we can observe whether increases in cyanobacteria with *Daphnia* found in water column, match historical patterns between these two communities.

References

Amorim, C. A., Ê. W. Dantas, and A. do N. Moura. 2020. Modeling cyanobacterial blooms in tropical reservoirs: The role of physicochemical variables and trophic interactions. Sci. Total Environ. **744**. doi:10.1016/j.scitotenv.2020.140659

Beaulieu, M., F. Pick, and I. Gregory-Eaves. 2013. Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set. Limnol. Oceanogr. **58**: 1736–1746. doi:10.4319/lo.2013.58.5.1736

Bogard, M. J., R. J. Vogt, N. M. Hayes, and P. R. Leavitt. 2020. Unabated Nitrogen Pollution Favors Growth of Toxic Cyanobacteria over Chlorophytes in Most Hypereutrophic Lakes. Environ. Sci. Technol. **54**: 3219–3227. doi:10.1021/acs.est.9b06299

Boopathi, T., and J. S. Ki. 2014. Impact of environmental factors on the regulation of cyanotoxin production. Toxins (Basel). **6**: 1951–1978. doi:10.3390/toxins6071951

Buley, R. P., M. F. Gladfelter, E. G. Fernandez-Figueroa, and A. E. Wilson. 2022. Can correlational analyses help determine the drivers of microcystin occurrence in freshwater ecosystems? A meta-analysis of microcystin and associated water quality parameters. Environ. Monit. Assess. **194**. doi:10.1007/s10661-022-10114-8

Burford, M. A., C. C. Carey, D. P. Hamilton, J. Huisman, H. W. Paerl, S. A. Wood, and A. Wulff. 2020. Perspective: Advancing the research agenda for improving understanding of cyanobacteria in a future of global change. Harmful Algae **91**. doi:10.1016/j.hal.2019.04.004

Capo, E., C. Giguet-Covex, A. Rouillard, and others. 2021. Lake sedimentary dna research on past terrestrial and aquatic biodiversity: Overview and recommendations. Quaternary **4**: 1–61. doi:10.3390/quat4010006

Carvalho, L., C. Mcdonald, C. de Hoyos, and others. 2013. Sustaining recreational quality of European lakes: Minimizing the health risks from algal blooms through phosphorus control. J. Appl. Ecol. **50**: 315–323. doi:10.1111/1365-2664.12059

Chorus, I., and M. Welker. 2021. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management, 2nd ed. CRC Press.

Chun, S. J., Y. Cui, J. J. Lee, I. C. Choi, H. M. Oh, and C. Y. Ahn. 2020. Network analysis reveals succession of Microcystis genotypes accompanying distinctive microbial modules with recurrent patterns. Water Res. **170**. doi:10.1016/j.watres.2019.115326

Conley, D. J., H. W. Paerl, R. W. Howarth, D. F. Boesch, S. P. Seitzinger, K. E. Havens, C. Lancelot, and G. E. Likens. 2009. Controlling Eutrophication: Nitrogen and Phosphorus. Science (80-.). **323**: 1014–1015. doi:10.1126/science.1167755

Cremona, F., L. Tuvikene, J. Haberman, P. Nõges, and T. Nõges. 2018. Factors controlling the three-decade long rise in cyanobacteria biomass in a eutrophic shallow lake. Sci. Total Environ. **621**: 352–359. doi:10.1016/j.scitotenv.2017.11.250

Dolman, A. M., J. Rücker, F. R. Pick, J. Fastner, T. Rohrlack, U. Mischke, and C. Wiedner. 2012. Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. PLoS One **7**. doi:10.1371/journal.pone.0038757

Domaizon, I., A. Winegardner, E. Capo, J. Gauthier, and I. Gregory-Eaves. 2017. DNA-based methods in paleolimnology: new opportunities for investigating long-term dynamics of lacustrine biodiversity. J. Paleolimnol. **58**: 1–21. doi:10.1007/s10933-017-9958-y

Downing, J. A., S. B. Watson, and E. McCauley. 2001. Predicting Cyanobacteria dominance in lakes. Can. J. Fish. Aquat. Sci. **58**: 1905–1908. doi:10.1139/cjfas-58-10-1905

Erratt, K. J., I. F. Creed, and C. G. Trick. 2022. Harmonizing science and management options to

reduce risks of cyanobacteria. Harmful Algae **116**. doi:10.1016/j.hal.2022.102264

Fastner, J., C. Beulker, B. Geiser, and others. 2018. Fatal neurotoxicosis in dogs associated with tychoplanktic, anatoxin-a producing tychonema sp. In mesotrophic lake tegel, Berlin. Toxins (Basel). **10**: 1–11. doi:10.3390/toxins10020060

Giani, A., D. F. Bird, Y. T. Prairie, and J. F. Lawrence. 2005. Empirical study of cyanobacterial toxicity along a trophic gradient of lakes. Can. J. Fish. Aquat. Sci. **62**: 2100–2109. doi:10.1139/f05-124

Harke, M. J., and C. J. Gobler. 2013. Global Transcriptional Responses of the Toxic Cyanobacterium, Microcystis aeruginosa, to Nitrogen Stress, Phosphorus Stress, and Growth on Organic Matter. PLoS One **8**. doi:10.1371/journal.pone.0069834

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

Huot, Y., C. A. Brown, G. Potvin, and others. 2019. The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. Sci. Total Environ. **695**: 133668. doi:10.1016/j.scitotenv.2019.133668

Janssen, E. M. L. 2019. Cyanobacterial peptides beyond microcystins – A review on co-occurrence, toxicity, and challenges for risk assessment. Water Res. **151**: 488–499. doi:10.1016/j.watres.2018.12.048

Kaebernick, M., and B. A. Neilan. 2001. Ecological and molecular investigations of cyanotoxin production. FEMS Microbiol. Ecol. **35**: 1–9. doi:10.1016/S0168-6496(00)00093-3

Keck, F., R. C. Blackman, R. Bossart, and others. 2022. Meta-analysis shows both congruence and complementarity of DNA and eDNA metabarcoding to traditional methods for

biological community assessment. Mol. Ecol. **31**: 1820–1835. doi:10.1111/mec.16364

Loftin, K. A., J. L. Graham, E. D. Hilborn, S. C. Lehmann, M. T. Meyer, J. E. Dietze, and C. B. Griffith. 2016. Cyanotoxins in inland lakes of the United States: Occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. Harmful Algae **56**: 77–90. doi:10.1016/j.hal.2016.04.001

Mantzouki, E., M. Beklioglu, J. D. Brookes, and others. 2018a. Snapshot surveys for lake monitoring, more than a shot in the dark. Front. Ecol. Evol. **6**: 1–5. doi:10.3389/fevo.2018.00201

Mantzouki, E., and B. W. Ibelings. 2018. The Principle and Value of the European Multi Lake Survey. Limnol. Oceanogr. Bull. **27**: 82–86. doi:10.1002/lob.10259

Mantzouki, E., M. Lürling, J. Fastner, and others. 2018b. Temperature effects explain continental scale distribution of cyanobacterial toxins. Toxins (Basel). **10**: 1–24. doi:10.3390/toxins10040156

Merel, S., D. Walker, R. Chicana, S. Snyder, E. Baurès, and O. Thomas. 2013. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. Environ. Int. **59**: 303–327. doi:10.1016/j.envint.2013.06.013

Monchamp, M. E., J. C. Walser, F. Pomati, and P. Spaak. 2016. Sedimentary DNA reveals cyanobacterial community diversity over 200 years in two perialpine lakes. Appl. Environ. Microbiol. **82**: 6472–6482. doi:10.1128/AEM.02174-16

Ninio, S., A. Lupu, Y. Viner-Mozzini, T. Zohary, and A. Sukenik. 2020. Multiannual variations in Microcystis bloom episodes – Temperature drives shift in species composition. Harmful Algae **92**: 101710. doi:10.1016/j.hal.2019.101710

Paerl, H. W., and M. A. Barnard. 2020. Mitigating the global expansion of harmful

cyanobacterial blooms: Moving targets in a human- and climatically-altered world. Harmful Algae **96**: 101845. doi:10.1016/j.hal.2020.101845

Paerl, H. W., J. T. Scott, M. J. McCarthy, and others. 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. doi:10.1021/acs.est.6b02575

Pal, S., I. Gregory-Eaves, and F. R. Pick. 2015. Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. J. Paleolimnol. **54**: 87–101. doi:10.1007/s10933-015-9839-1

Pawlowski, J., A. Bonin, F. Boyer, T. Cordier, and P. Taberlet. 2021. Environmental DNA for biomonitoring. Mol. Ecol. **30**: 2931–2936. doi:10.1111/mec.16023

Pérez-Jvostov, F., W. J. Sutherland, R. D. H. Barrett, and others. 2020. Horizon scan of conservation issues for inland waters in Canada. Can. J. Fish. Aquat. Sci. **77**: 869–881. doi:10.1139/cjfas-2019-0105

Picard, M., X. Pochon, J. Atalah, and others. 2022. Using metabarcoding and droplet digital PCR to investigate drivers of historical shifts in cyanobacteria from six contrasting lakes. Sci. Rep. **12**: 1–17. doi:10.1038/s41598-022-14216-8

Pick, F. R. 2016. Blooming algae: A Canadian perspective on the rise of toxic cyanobacteria. Can. J. Fish. Aquat. Sci. **73**: 1149–1158. doi:10.1139/cjfas-2015-0470

Pilon, S., A. Zastepa, Z. E. Taranu, I. Gregory-Eaves, M. Racine, J. M. Blais, A. J. Poulain, and F. R. Pick. 2019. Contrasting histories of microcystin-producing cyanobacteria in two temperate lakes as inferred from quantitative sediment DNA analyses. Lake Reserv. Manag. **35**: 102–117. doi:10.1080/10402381.2018.1549625

Rashidi, H., H. Baulch, A. Gill, L. Bharadwaj, and L. Bradford. 2021. Monitoring, Managing,

and Communicating Risk of Harmful Algal Blooms (HABs) in Recreational Resources across Canada. Environ. Health Insights **15**. doi:10.1177/11786302211014401

Reid, A. J., A. K. Carlson, I. F. Creed, and others. 2018. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol. Rev. **94**: 849–873. doi:10.1111/brv.12480

Schindler, D. W., R. E. Hecky, D. L. Findlay, and others. 2008. Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. Proc. Natl. Acad. Sci. U. S. A. **105**: 11254–11258. doi:10.1073/pnas.0805108105

Scott, T. J., and M. J. McCarthy. 2011. Response to comment: nitrogen fixation has not offset declines in the Lake 227 nitrogen pool and shows that nitrogen control deserves consideration in aquatic ecosystems. Limnol. Oceanogr. **56**: 1548-1550. doi: 10.4319/lo.2011.56.4.1548

Shan, K., L. Song, W. Chen, and others. 2019. Analysis of environmental drivers influencing interspecific variations and associations among bloom-forming cyanobacteria in large, shallow eutrophic lakes. Harmful Algae **84**: 84–94. doi:10.1016/j.hal.2019.02.002

Tanvir, R. U., Z. Hu, Y. Zhang, and J. Lu. 2021. Cyanobacterial community succession and associated cyanotoxin production in hypereutrophic and eutrophic freshwaters. Environ. Pollut. **290**. doi:10.1016/j.envpol.2021.118056

Taranu, Z. E., I. Gregory-Eaves, P. R. Leavitt, and others. 2015. Acceleration of cyanobacterial dominance in north temperate-subarctic lakes during the Anthropocene. Ecol. Lett. **18**: 375–384. doi:10.1111/ele.12420

Taranu, Z. E., I. Gregory-Eaves, R. J. Steele, M. Beaulieu, and P. Legendre. 2017. Predicting microcystin concentrations in lakes and reservoirs at a continental scale: A new framework

for modelling an important health risk factor. Glob. Ecol. Biogeogr. **26**: 625–637. doi:10.1111/geb.12569

Wagner, N. D., F. S. Osburn, J. Wang, R. B. Taylor, A. R. Boedecker, C. K. Chambliss, B. W. Brooks, and J. T. Scott. 2019. Biological stoichiometry regulates toxin production in microcystis aeruginosa (UTEX 2385). Toxins (Basel). **11**. doi:10.3390/toxins11100601

Wells, M. L., B. Karlson, A. Wulff, and others. 2020. Future HAB science: Directions and challenges in a changing climate. Harmful Algae **91**. doi:10.1016/j.hal.2019.101632

Yan, D., H. Xu, M. Yang, and others. 2019. Responses of cyanobacteria to climate and human activities at Lake Chenghai over the past 100 years. Ecol. Indic. **104**: 755–763. doi:10.1016/j.ecolind.2019.03.019