8-2014

SYSTEMS BIOLOGY OF *MICROCYSTIS* BLOOMS

Morgan Michelle Steffen
University of Tennessee - Knoxville, msteffe1@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Morgan Michelle Steffen entitled "SYSTEMS BIOLOGY OF *MICROCYSTIS* BLOOMS." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Microbiology.

Steven W. Wilhelm, Major Professor

We have read this dissertation and recommend its acceptance:

Alison Buchan, Erik Zinser, Loren Hauser, Jennifer DeBruyn

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

SYSTEMS BIOLOGY OF *MICROCYSTIS* BLOOMS

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Morgan Michelle Steffen
August 2014

To my parents, for cultivating my curiosity from the start.

Acknowledgements

To start, I want to thank my advisor, Dr. Steven Wilhelm, for having confidence in me, for giving me just the right amount of independence while still keeping me on the right track, and for always knowing just the right way to motivate me when I needed the push. The lessons you have taught me over the last several years are too numerous to list, but I will never forget any of them.

To my committee members, who have each contributed to this work, thank you for making this investment in my education. I especially appreciate your patience and willingness to give advice the many times I've needed it over the years.

It is impossible for me to thank everyone who has contributed to my work over the last several years, but I will certainly try. I have been lucky to have a number of enthusiastic students whose work has been imperative for moving my projects forward, so thank you to Rebecca Weir, Shafer Belisle, Mariya Campbell, Kristen Larsen, and Alex Daniels for all of your hard work. I also need to thank Dr. Matt Saxton, Dr. Greg Boyer, and Dr. Sue Watson for their contribution to my training in field work and for keeping me on my toes during all those weeks on Lake Erie. To my lab mates, both past and present, thank you for making our lab an enjoyable place to work and listening to my talks over and over. I want to extend a special thanks to Dr. Gary LeCleir for his training and willingness to always answer any question.

To my family, everything I am today I owe to you. Your support and love have been invaluable to me during this process. And finally, to my partner Louie. Meeting you has made all of the frustration, failure, and stress completely worth it. Thank you for being there at the end of every day.

Abstract

Cyanobacterial harmful algal blooms (cHABs) degrade freshwater lakes worldwide. Accumulation of nuisance biomass and production of noxious secondary metabolites can result in an expansive impact on both lake ecology and the surrounding communities. The cHAB forming organism *Microcystis aeruginosa* is known to produce the toxin microcystin, a compound nicknamed "fast death factor," which has been implicated in animal poisonings and human liver failure and cancers. *M. aeruginosa* inhabits a wide range of freshwater lakes around the world, such as Lake Erie (USA/Canada) and Lake Tai (Taihu, China), and is often a dominant member of bloom communities. Such systems are well-studied paragons of annual bloom events which have expanded both in size and duration in recent years. Decades of research on these two lakes point to eutrophication, specifically nitrogen (N) and phosphorus (P) loading and rising global temperatures as keys to the success of *Microcystis* spp. in the environment. Nevertheless, it is unknown how these environmental factors interact with microbial communities to contribute to bloom formation. To better understand these complex interactions, a series of targeted and community genomic and transcriptomic studies were performed. Due to the exploratory nature of these studies, results from each were largely in the form of new hypotheses about the *Microcystis* system. Specific findings support the established hypothesis that microbial community function is independent of community structure, at both that the DNA and mRNA levels. Novel hypotheses were also generated as a result of these exploratory studies, including that nutrient availability, specifically the availability of nitrogen and phosphorus, may drive activity of transposable elements across the genome of *Microcystis*. Taken together, this work represents a new foundation for future exploration of the success of *Microcystis* in freshwater lakes, and will lead to a series of new empirical studies to test the hypotheses presented herein. This in turn has led to the advancement of understanding of the microbial interspecies relationships of freshwater cHABs at the intersection of environmental biogeochemistry.

Table of Contents

List of Tables

CHAPTER V

List of Figures

CHAPTER I

CHAPTER II

CHAPTER III

CHAPTER I:
INTRODUCTION

PUBLICATION NOTE

This chapter is a version of a peer-reviewed article previously published in *The Journal of Great Lakes Research* 40:2 (215-225; 466-467) by Morgan M. Steffen, B. Shafer Belisle, Sue B. Watson, Gregory L. Boyer, and Steven W. Wilhelm.

My contribution to this work was the compilation of historical and current literature and data for the synopsis of *Microcystis* research in Lake Erie, as well as most of the writing.

PART I: LITERATURE REVIEW: STATUS, CAUSES AND CONTROLS OF CYANOBACTERIAL BLOOMS IN LAKE ERIE

Abstract

The Laurentian Great Lakes are among the most prominent sources of fresh water in the world. Lake Erie's infamous cyanobacterial blooms have, however, threatened the health of this valuable freshwater resource for decades. Toxic blooms dominated by the cyanobacterium *Microcystis aeruginosa* have most recently been the primary ecological concerns for the lake. These toxic blooms impact the availability of potable water, as well as public health and revenues from the tourism and fishery industries. The socioeconomic effects of these blooms have spurred research efforts to pinpoint factors that drive bloom events. Despite decades of research and mitigation efforts, these blooms have expanded both in size and duration in recent years. However, through continued joint efforts between the Canadian and United States governments, scientists, and environmental managers, identification of the factors that drive bloom events is within reach. This review provides a summary of historical and contemporary research efforts in the realm of Lake Erie's harmful cyanobacterial blooms, both in terms of experimental and management achievements and insufficiencies, as well as future directions on the horizon for the lake's research community.

Introduction

The Laurentian Great Lakes are arguably one of the most valuable natural resources in North America, if not the world. This system represents roughly 20% of the Earth's available surface freshwater, a resource that is expected to become increasingly limited in the near future (Schottler and Eisenreich 1994). Lake Erie alone provides over 7 billion dollars in revenue each year from tourism and fishery industries (USDA 2005). For the last two decades, however, Lake Erie has again been threatened (as it was in the 1960s and 1970s) by annual blooms of toxic cyanobacteria during summer months. Despite intensive research and management efforts, the duration and toxicity of blooms appear to be expanding in recent years (Stumpf *et al.* 2012).

Proliferation of undesirable plankton, whether in freshwater or marine environments, has long plagued the world's waters (Table 1.1). Among the Laurentian Great Lakes, Lake Erie is

most susceptible to recurring large-scale blooms due to the morphology of the lake, its location in a temperate climate with warm summer temperatures, and extensive anthropogenic inputs. At an average depth of 19 m, Lake Erie has a relatively short retention time (< 3 years) and consistently reaches temperatures above 25° C during summer months (Burns *et al.* 2005, NWS 2012, Stumpf et al 2012). The lake continues to receive extensive input from agricultural and industrial runoff, despite decades of international efforts to reduce nutrient loading (Waples 2008). Phosphorus (Dolan and McGunagle 2002; Dolan and Chapra 2012; Han et al. 2012) and nitrogen (Solomon *et al.* 2010) are key components in detergents, fertilizers, industrial chemicals, and common herbicides. These compounds and others are increasingly being applied within watersheds, resulting in the delivery of nutrients to the lakes through tributaries, rivers, and non-point source runoff. As in the past, constraining these two macronutrients remains a particular focus for both scientists and ecosystem stewards.

The formation of large swaths of cyanobacterial biomass across Lake Erie is not a new phenomenon. Beginning in the early twentieth century, a marked increase in phytoplankton biomass and a decline in dissolved O_2 were primarily thought to be a result of phosphorus loading *via* point sources into the system (Charlton *et al.* 1993). While largely comprised of diatoms, the Lake Erie phytoplankton community throughout the first half of the twentieth century also contained cyanobacteria from the genera *Microcystis* and *Anabaena* (Davis 1958, Nicholls *et al.* 1977). During the 1950s, studies indicated a peak in cyanobacterial biomass during the months of September and October, referred to as the "autumnal maximum." Both filamentous genera (*Anabaena* spp. – the pelagic versions are now reclassified as *Dolichospermum* spp., *Aphanizomenon* spp., *Lyngbya* spp., and planktonic *Oscillatoria* spp.- now reclassified as *Planktothrix* spp.) and the unicellular colony-forming *Microcystis* were observed to contribute to this autumnal maximum, with filamentous *Aphanizomenon* being reported as most abundant (Davis 1954). A later survey of the phytoplankton community during 1956 and 1957 reported *Microcystis* as the most abundant cyanobacterium, followed in quantity by the filamentous *Aphanizomenon flos-aquae* and *Oscillatoria* (*Planktothrix*) spp. (Davis 1962). Despite the presence of these potentially toxic organisms, the initial focus remained on reduction

Table 1.1. Selection of historical evidence of cyanobacteria blooms.

Year	Observation	Reference
~ 1000	General Zhu Ge-Ling reports illness in troops who drank from a river in the south of China that was green	(Chorus and Bartram 1999)
~ 1200	Locals aware of the toxicity of algae near *Monasterium Viridis Stagni* (Monastery of the Green Loch), located near the eutrophic, freshwater Soulseat Loch near Stranraer in south west Scotland.	(Codd 1996)
1648	Paintings by Dutch Master Artist Salomon van Ruysdael (1648) show water bodies that are visibly green with a blue sky background	(Paerl 2009)
1878	Domestic animal poisonings in Australia	(Francis 1878)
1931	Lower than normal rainfalls in Ohio cause a bloom which is washed down the river leading to a cascade of gastroenteritis that could not be attributed to any infectious disease	(Miller and Tisdale 1931)
1959	*Microcystis* blooms first reported in Lake Tai (Taihu) China. Regions that experience blooms have been reported to also have the highest incidence of cancer	(Qin *et al.* 2007)
1981	Extensive bloom in Australia associated with illness and hepatotoxicity	(Chorus and Bartram 1999)
1995	Microcystis blooms first reported on large scales in Lake Erie	(Brittain *et al.* 2000)
1996	50 deaths in Brazil due to microcystins in a dialysis unit's intake filters	(Azevedo *et al.* 2002)
2007	Blooms in *Taihu* force government to supply bottled water to *Wuxi*	(Qin et al 2010)
2011	The largest ever recorded *Microcystis* bloom in Lake Erie occurs and persists into September	(Michalak *et al.* 2013)
2012	*ABC News* and the *Food & Environmental Reporting Network* report 10 dog deaths nationwide over the past 2 years and 98 reports of illness in Wisconsin in last three year.	(ABCAvila 2012)

of overall nuisance algal biomass and restoration of dissolved O_2 levels to the hypolimnion of the lake (Charlton et al. 1993). In the mid-twentieth century, the eutrophication of Lake Erie was described by the scientific community with increasing frequency and urgency. Indeed, an article by Beeton (1961) that pointed out rapid eutrophication in Lake Erie served as a spring board for the popular press to describe the lake as suffering from "accelerated old age," to the point where the lake was declared "dead" by the popular press (Ashworth 1986, Fortner 1987). Ironically, the problem was that the lake was likely "too alive," at least microbiologically. Indeed so poor was the health of the lake at that time that it was referenced in a popular children's book:

> *You're glumping the pond where the Humming-Fish hummed! No more can they hum, for their gills are all gummed. So I'm sending them off. Oh their future is dreary. They'll walk on their fins and get woefully weary in search of some water that isn't too smeary. I hear things are just as bad up in Lake Erie (Seuss 1971).*

Spurred by popular press and an outcry among scientists alike, a scientific "call to action" resulted in a flurry of studies looking to identify the cause of the eutrophication of such an ecologically and economically important water source. Phosphorus was identified as the key nutrient contributing to eutrophication (Vollenweider 1968), which led to the implementation of the Great Lakes Water Quality Agreement of 1972 (Hasler 1969, IJC 1986, Schindler 1977). This joint effort by the Canadian and American governments effectively reduced phosphorus loading from point sources into the lake by 50% within 10 years of the peak levels observed in 1968 (Charlton *et al.* 1993, Makarewicz and Bertram 1991).

As predicted, total phytoplankton biomass (g/m^3) was observed to decline during the 1970s and 1980s, with reports suggesting an 89% total reduction in biomass in some areas of the lake (Makarewicz 1993). In fact, so effective was the remediation that the above children's literature reference to Lake Erie by Dr. Seuss was removed in 1985 (Morgan 1995). Dissolved O_2 levels did not recover, however, despite substantial efforts to limit further eutrophication of the lake by regulation of phosphate loading (Charlton 1980). However, by the early 1990s, trends in fishery populations, algal biomass decline, and sediment quality suggested a possible recovery of the system. The reasons for lack of total recovery from anoxia remain unknown but have been

tied tightly to the morphology of the lake and the orientation of the central basin. The combination of these physical characteristics results in a typically thin (~4 m) hypolimnion, which is supplied with carbon from primary producers in the much larger epilimnion (Rao *et al.* 2008, Wilhelm *et al.* 2006). Combined with carbon input from spring diatom growth as well as mid-winter blooms that do not appear to be totally consumed by bacteria during winter months, dissolved O_2 isolated in the hypolimnion in summer months is rapidly consumed (Twiss *et al.* 2012, Wilhelm et al 2006).

Historically, the major input of nutrients into Lake Erie has been from the Maumee River, which drains productive (and well-fertilized) farmland, as well as urban centers in Michigan, Indiana, and Ohio into the western basin of the lake (Dolan and Chapra 2012). In contrast, while there is no doubt that large volumes of water also pass down the Detroit River, nutrient inputs from the Detroit River have been much lower than from the Maumee River (Han et al. 2012), and there is no record of blooms occurring at the entrance of the Detroit River into the lake. After decades of reduction in P loading, Lake Erie reached its target external phosphorus loading of 11,000 MT per year in 1981. For the next decade, combined concentrations from point and nonpoint sources remained relatively static, as observed for example in the Maumee River which flows into the Western Basin of Lake Erie (Figure 1.1A). Beginning in the mid-1990s, however, a return to annual increases in soluble reactive phosphorus (SRP) loading to the lake has been reported (Joose and Baker, 2011). Measurements of SRP concentrations for at least the Maumee River contradict this observation, however, as they reflect a limited variability loading trends after 1995, with a similar overall trend from 1975 to present in this particular tributary (Figure 1C-D). Thus, even though SRP loading is widely considered a potential threat to the trophic status of Lake Erie (Bridgeman *et al.* 2012, Joosse and Baker 2011), the Maumee River tributary concentration data suggest that SRP inputs from this source may not contribute to bloom proliferation as heavily as suspected.

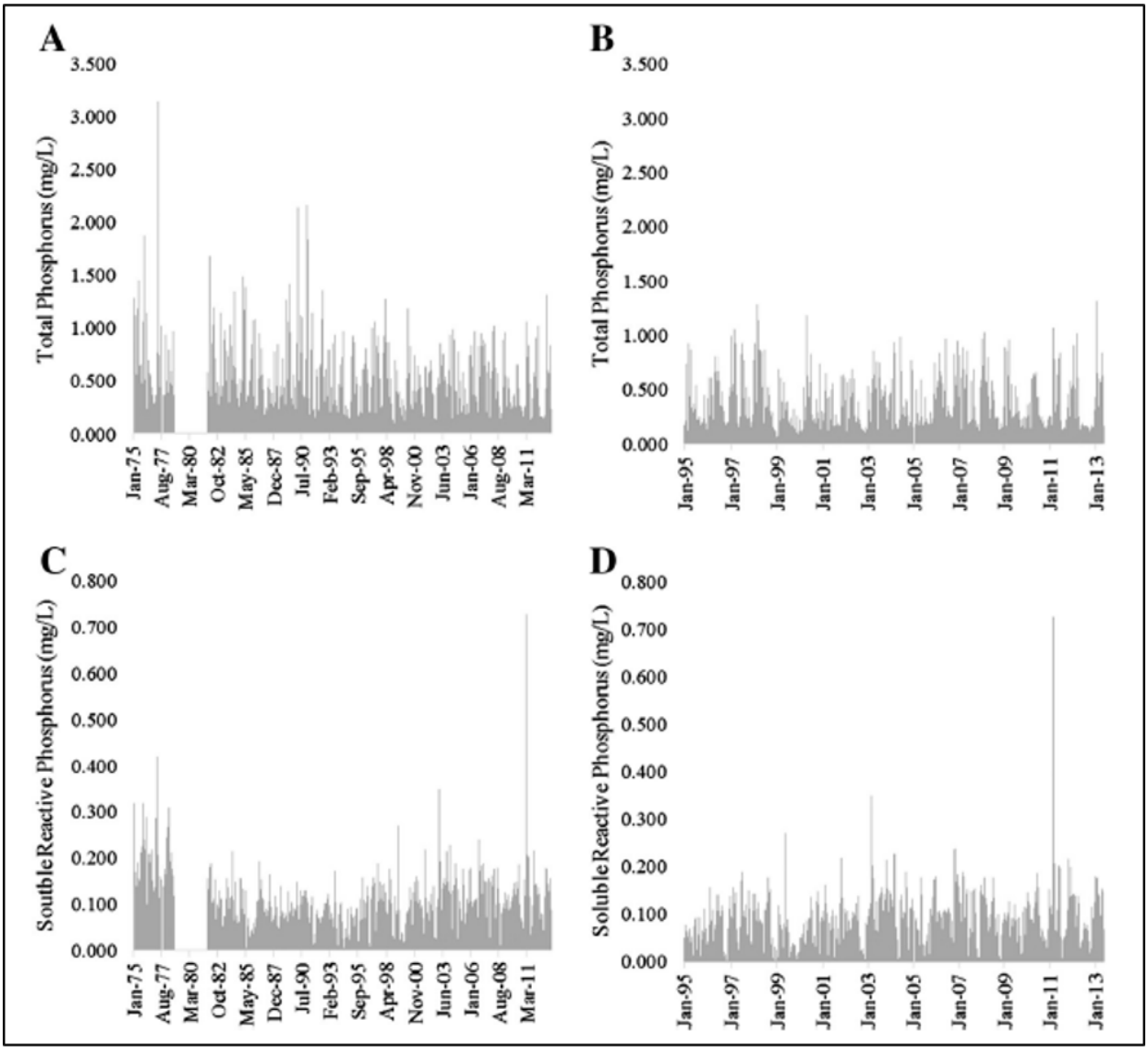

Figure 1.1. Total phosphorus (A-B) and soluble reactive phosphorus (SRP) (C-D) measurements from Maumee River tributary monitoring program taken from 1975 to 2012 (A/C). B/D depict data from 1995-present. Data available from www.heidelberg.edu/academiclife/distinctive /ncwqr, last accessed Dec 2, 2013.

Despite an apparent improvement in the overall health of the lake, cyanobacterial blooms returned to Lake Erie beginning in the mid-1990s (Conroy *et al.* 2005). The identification of toxic *Microcystis aeruginosa* in Lake Erie in 1995 marked the beginning of a second cyanobacterial-driven decline in Lake Erie water quality (Brittain *et al.* 2000). Today, annual *Microcystis* blooms occur in all basins of Lake Erie and within many of its embayments, although large scale surface blooms are generally confined to the western and west-central basins of the Lake and embayments near Sandusky (OH), Presque Isle (PA) and Rondeau Provincial Park (ON). The duration of these blooms has expanded during the last decade. The goal of this review is to underscore recent information meant to address the question, "What makes *Microcystis* bloom?" We highlight gaps in current research and potential obstacles toward moving forward in addressing toxic *Microcystis* blooms in Lake Erie.

Current Status of the Cyanobacterial Community and Cyanotoxins in Lake Erie

After initial observations of toxic *Microcystis aeruginosa* in Lake Erie in 1995 (Brittain et al. 2000), blooms of this cyanobacterium and others have formed annually in the western basin and across the lake (Conroy et al 2005, Millie *et al.* 2009, Ouellette *et al.* 2006). In the last decade, there has been a tendency for annual bloom formation to start earlier in the year and extend later in the season than in the past; the 2011 Lake Erie bloom was visible from satellites until mid-October (Michalak *et al.* 2013). The 2011 bloom has been characterized as the most extensive in Lake Erie's recent history *via* analysis of MODIS satellite imagery (Michalak et al 2013). Due to the potential ecological and public health impacts of the toxins produced in association with these blooms, scientists and government agencies have been working to identify potential causative factors for these blooms. These endeavors may contribute to eventual mitigation of annual bloom events.

Lake Erie blooms are typically dominated by the genera *Microcystis* (Figure 1.3) and to a lesser extent *Anabaena* (Figure 1.4), with other filamentous genera often existing as background populations. *Microcystis aeruginosa* is a colonial, bloom-forming cyanobacterium that has been identified in freshwater systems worldwide (Sant′Anna *et al.* 2011, van Gremberghe *et al.* 2011, Ye *et al.* 2009). Surveys of Lake Erie's flora have documented numerous *Microcystis* species, including *M. flos-aquae, M. botrys, M. novacekii, M. viridis,* and

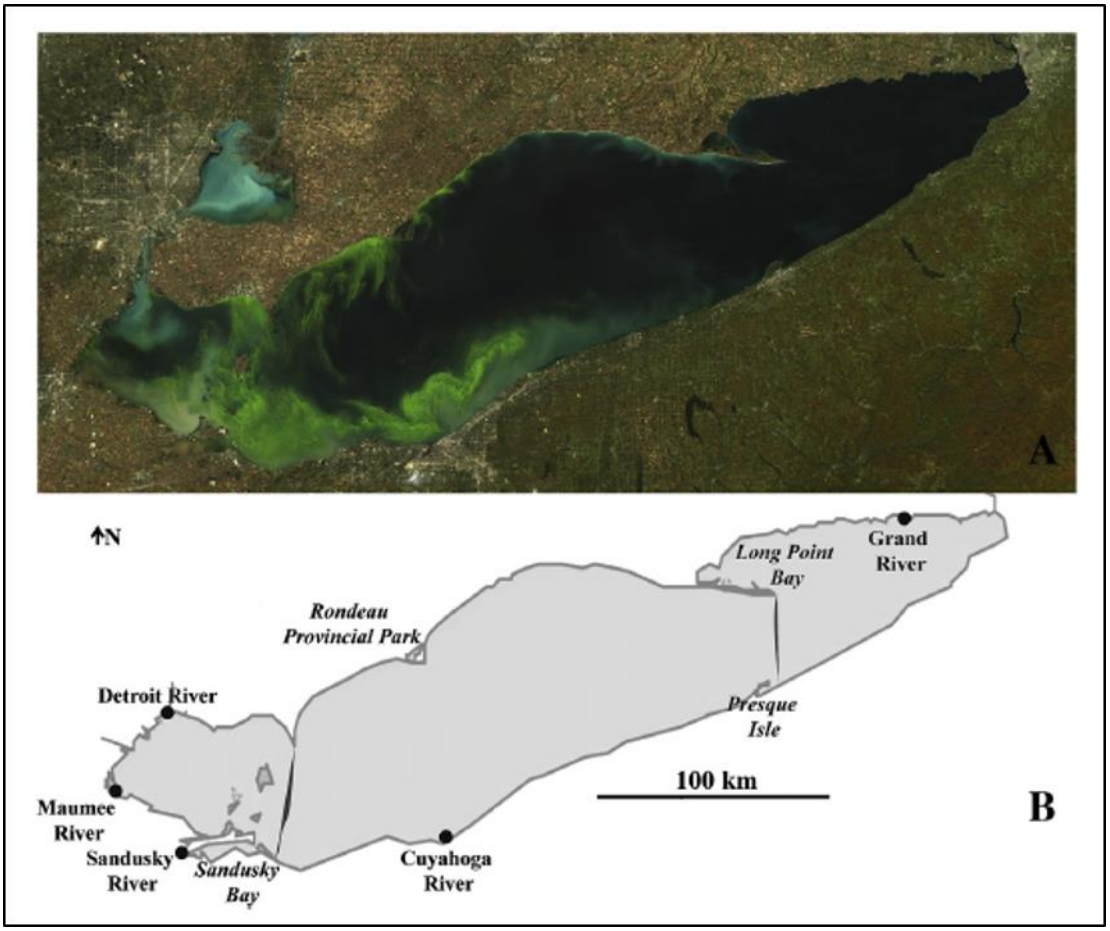

Figure 1.2. A. Simulated true-color image of *Microcystis* blooms in Lake Erie from the MODIS sensor on NASA's Aqua satellite. Image acquired October 9, 2011 and provided courtesy of Dr. Richard Stumpf (NOAA). B. Locations of embayments (in italics) and river input sites in Lake Erie discussed in this review.

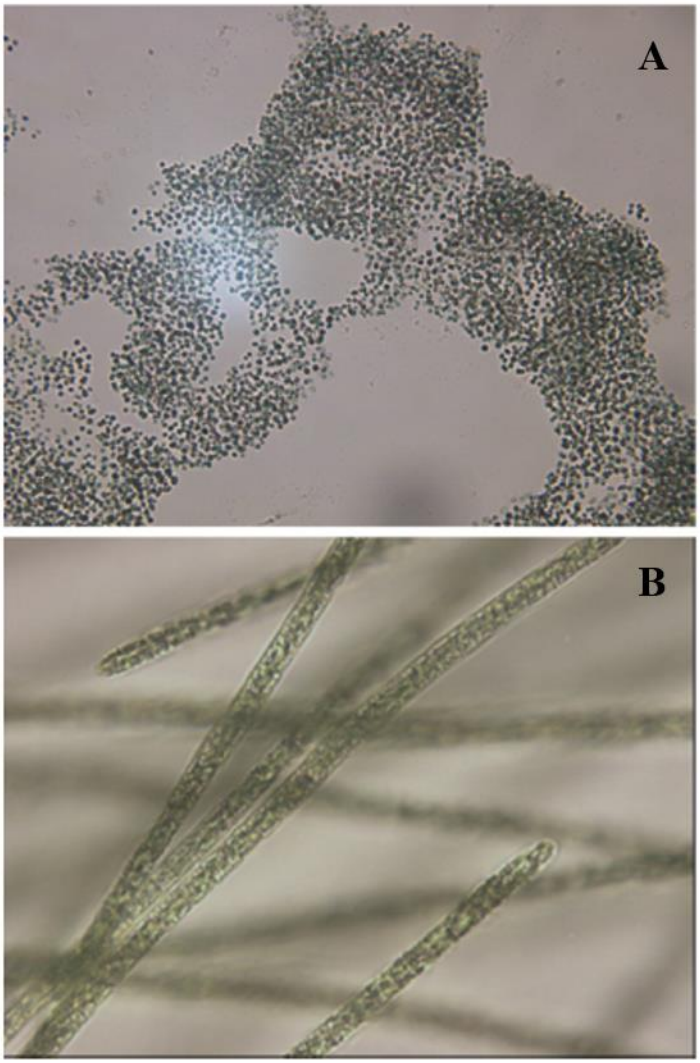

Figure 1.3. Light microscope images (viewed at 400X) of *Microcystis aeruginosa* NIES 843 (A) and *Planktothrix* spp. (B). Photo credit to SW Wilhelm and GL Boyer.

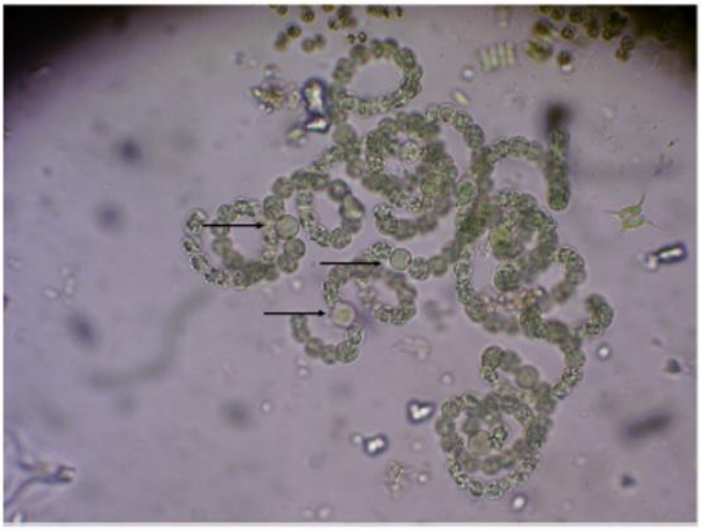

Figure 1.4. Light microscope image (viewed at 400X) of *Anabaena* (syn *Dolichospermum*) sp. collected from Lake Erie in July 2006. Heterocyst formation (arrow) is readily evident suggesting ongoing N-fixation. Image by SW Wilhelm and GL Boyer.

M. wesenbergii, although toxin production has (to date) been generally attributed to *M. aeruginosa*. Work by Dyble et al. (2008), using reconstructed phylogenies of the *mcyB* gene of the microcystin toxin operon, suggested several Lake Erie isolates were consistent with *M. botrys*, although active microcystin production by this species in Lake Erie has yet to be validated.

An important point concerning Lake Erie that is often overlooked is that the lake's cyanobacterial community is persistently dominated by small, single-celled cyanobacteria consistent with *Synechococcus* spp. and *Cyanobium* spp. (Carrick and Schelske 1997, Wilhelm et al 2006). As these cells are difficult to observe and generally not considered nuisance populations, they are commonly discounted in surveys of cyanobacterial populations. However, as Carrick and Schelske (1997) pointed out, they are important members of the community of primary producers and likely carry out a significant proportion of the total lake primary production, at least seasonally. Moreover, as cyanobacteria, they likely contribute to the pool of specific pigments used in many studies to estimate phytoplankton biomass (Vincent et al. 2004).

Other potential cyanotoxin producers, notably species of *Planktothrix* (Figure 1.3B), *Aphanizomenon*, and *Anabaena*, have also been observed with increasing frequency in the western basin of the lake (Davis *et al.* 2012, Rinta-Kanto *et al.* 2005, Saxton *et al.* 2012). While *Anabaena* and *Aphanizomenon* had been observed in Lake Erie prior to 1960 (Davis 1954), the potentially toxic *Cylindrospermopsis* was first observed in the mid-2000s (Conroy *et al.* 2007). The reasons for the recent resurgence of nitrogen-fixing filamentous cyanobacteria like *Anabaena* spp. are as yet unknown, but algal succession events favoring diazotrophs may be indicative of shifting seasonal nutrient loading patterns (Paerl and Fulton 2006). Moreover, current *Anabaena* populations in Lake Erie have demonstrated heterocyst formation consistent with the fixation of atmospheric N_2 (Figure 1.4), implying their success may be due to at least temporal nitrogen limitation in the system (see *The role of nutrients as drivers of Lake Erie cHABs* below). Along with the above populations, blooms of benthic *Lyngbya* spp. have been seen in recent years (Bridgeman and Penamon 2010). Although *Lyngbya* has been shown to be toxic in other marine and freshwater systems, to date neither chemical nor molecular screenings of Lake Erie *Lyngbya* populations have detected any of the common cyanobacterial toxins or the genetic potential to make such compounds (Boyer and Wilhelm, unpublished).

Many cyanobacteria are capable of producing a number of secondary metabolites, and the hepatotoxic microcystins are of particular concern. Microcystins - originally called *"fast-death factors"* (Bishop et al. 1959) - are cyclic heptapeptides with two variable amino acid positions that result in the existence of over 80 chemical variants (Dietrich and Hoeger 2005). Microcystins are potent inhibitors of type 1 and 2A protein phosphatases (Dawson 1998) and globally have been involved in numerous animal and, in limited cases human, intoxications and deaths (Azevedo *et al.* 2002, Jochimsen *et al.* 1998). Despite the risk, microcystin-associated fatalities have yet to be reported in the Laurentian Great Lakes. Potential tumor-promoting properties of microcystins are thought to be caused by their inhibitory effect on DNA repair mechanisms (Žegura *et al.* 2011a, Žegura *et al.* 2011b). However, the International Agency for Research on Cancer (IARC) ruled in 2006 that their existing evidence is inadequate to determine if this tumor-promoting activity actually translates into cancer in human or animal populations (Grosse *et al.* 2006). Additional secondary metabolites produced by *Microcystis* spp. have also been found to act as synthetic estrogens in zebra fish exposure experiments (Oziol and Bouaïcha 2010, Rogers *et al.* 2011). *Microcystis* is known to produce other secondary metabolites, such as aeruginosins and cyanopeptolins, which may also have an impact on consuming species (Paerl and Otten 2013, Smith *et al.* 2008). Recently, the first report of production of paralytic shellfish poison (PSP) by a strain of *Microcystis* was recorded, although to date, this is an isolated record (Sant'Anna *et al.* 2011).

When the resurgence of *Microcystis* was originally detected in Lake Erie (~ 1995), initial measurements never reached levels considered unsafe for drinking (1 µg/L) by the World Health Organization (WHO) (Table 1.2). Within 6 years, however, peak annual levels exceeded 200 µg/L during late summer (WHO 2011). Indeed, toxin concentrations appear to vary widely over the last two decades (Table 1.2). This variation may be in part due to temporal and spatial differences in sampling, as well as disparate methods used to collect and process samples. Moreover, sampling is often complicated by the ability of blooms to be transported vertically in the water column and horizontally between locations, meaning sampling often occurs outside the range of the actual bloom. Annual monitoring spearheaded by both the Canadian and US governments of multiple cyanotoxins is now standard during summer months for Lake Erie;

Table 1.2. A survey of literature data available for microcystin concentrations in the Western Basin of Lake Erie. Methods for detection vary and are available from cited sources or as noted. Concentrations are presented as maximum values reported in individual studies. *Data from the MERHAB-LGL research program using methodology similar to Rinta-Kanto et al. 2009a. ǂNOAA Center for Excellence on Great Lakes Research (GLERL, www.glerl.noaa.gov/res/Centers/HABS/western_lake_erie.html).

Date	Maximum Concentration Detected	Source
September 1995	>1 µg/L	(Brittain et al 2000)
September 1996	0.003 µg/L	(Brittain et al 2000)
September 2001	238.8 µg/L	(Murphy *et al.* 2003)
August 2002	13.9 µg/L	(Dionisio Pires *et al.* 2005b)
August 2003	45.1 µg/L	(Dionisio Pires et al 2005b)
August 2004	2.6 µg/L	(Rinta-Kanto et al 2005)
August 2005	0.2 µg/L	(Rinta-Kanto *et al.* 2009)
August 2006	2.47 µg/L	Boyer et al., unpublished*
July 2007	2.2 µg/L	Boyer et al., unpublished*
August 2008	1.12 µg/L	Boyer et al., unpublished*
August 2009	8.05 µg/L	GLERLǂ
August 2010	20.77 µg/L	GLERLǂ
July 2011	1221.2 µg/L	GLERLǂ
August 2012	201.41 µg/L	Boyer et al., unpublished*

chemical, physical, and biological data are used to track bloom progression and serve as a means of assessing water quality for the lake. Finally, sampling is also complicated by the potential of various compounds to bioaccumulate in fish populations of commercial value. Poste and colleagues (2011) demonstrated detectable levels of microcystin within fish from Lake Erie as well as other North American and Ugandan lakes, suggesting a potential route for the toxins into higher order predators, including humans. While incredibly valuable, these data (and others) require further examination and resolution of both issues of toxin-binding within tissues as well as matrix effects associated with the estimate of toxin within various tissues (Schmidt et al. 2013; Smith et al. 2010).

One problem with the chemical detection of biologically produced environmental toxins is that it is only useful after the toxins are present and exposure may have occurred (Dyble *et al.* 2008, Ouellette and Wilhelm 2003). In recent years, molecular biological approaches have not only been employed to develop predictive tools but also to better identify the members of the plankton community associated with potential toxin production. The genes encoding the nonribosomal synthesis of microcystins (*mcyA-J*) have been well characterized and as a result are often used as a means of identifying potential toxin producers in lakewater samples (Dyble et al 2008, Ouellette and Wilhelm 2003). Filamentous genera such as *Anabaena* and *Planktothrix* are also capable of producing microcystin and are often found in conjunction with *Microcystis* during bloom events. Moreover, these populations are often independently associated with toxin events, requiring that most mitigation strategies be informed of the biological suspects (Rinta-Kanto and Wilhelm 2006). As technology continues to develop, detection methods are becoming more comprehensive. Traditional and next generation sequencing techniques are currently being employed to examine the structure of bloom communities and to identify the function of bloom and sediment community members in Lake Erie and their linkage to bloom events (Bouzat *et al.* 2013, Mou *et al.* 2013, Steffen *et al.* 2012). Moreover, researchers continue to carry out controlled laboratory experiments to determine why a cell would commit ~2% of a its genomic material and a nearly an equal percentage of total dry weight to maintain this compound in hopes that an understanding of the selective advantages microcystin conveys on the cell will provide insight on better management strategies (Wilhelm and Boyer, 2011).

It is important to note that damage from cyanobacterial blooms can also occur independent of toxin production. Water column export of abundant phytoplankton biomass can lead to aquatic plant die offs, production of noxious taste and odor compounds (Watson *et al.* 2008), and regional hypoxia, as well as jeopardize socioeconomically important tourism industries by the simple "fouling" of beaches (Paerl *et al.* 1998). Moreover, the very presence of these organisms leads to a competition for nutrients, in many cases at the expense of other plankton (*e.g.,* diatoms) that are important components of fisheries-dependent food webs.

Early detection of cyanobacterial harmful algal bloom (cHAB) formation is critical to the formation of a prompt response, and available detection methods have greatly improved in sensitivity and speed in recent years. Microscopic identification of cHAB cells in environmental samples has progressed toward the usage of image cytometry devices such as the FlowCAM (Lavrentyev *et al.* 2004). Sensitivity and predictive ability of remote sensing (especially satellite) data have extended warning periods before large-scale blooms in Erie (Becker *et al.* 2009, Binding *et al.* 2012, Strong 1974). While traditional methods such as fluorescence, microscopy, and pigment analysis are still employed in the field (Saxton et al 2012), they are more frequently being supplemented and sometimes replaced by molecular biological methods that can more accurately distinguish between toxic and non-toxic populations {*e.g.,* see Ouellette and Wilhelm (2003), Ouellette et al (2006)}.

The role of nutrients as drivers of Lake Erie cHABs

Eutrophication has long been acknowledged as a driver of freshwater phytoplankton biomass. Phosphorus has historically been accepted as the key limiting nutrient in freshwater systems (Schindler 1977). Indeed, this observation was the motivation for "cleanup" efforts in Lake Erie mediated by the Great Lakes Water Quality Agreement of 1972, which substantially reduced phosphorus loading to the lake. This international effort was renewed as an amendment in 2012 with continued emphasis on restoration of Great Lakes water quality. Nutrient measurements across Lake Erie during the 1990's (Guildford et al. 2005) confirmed that the lake remains P-limited (the TN:TP whole lake average was 110, while the western basin average was 91 molar ratio) based on the criterion of TN:TP molar ratio of >50, (Guildford and Hecky 2000) being indicative of sole limitation by P. Lake Erie phosphorus amendment studies in the early

2000s also concluded that additional P stimulated algal biomass production (Wilhelm *et al.* 2003). Based on size-fractionated chl *a* measurements, however, it appears that plankton community structure across the lake did not respond in a consistent matter to P-additions (Wilhelm et al 2003). Phosphorus loading to Lake Erie remains a serious concern, with a focus on nonpoint sources emerging in recent years (Joosse and Baker 2011).

More recently, other influences have been discovered that may be shaping phytoplankton populations within this system. For example, agricultural runoff, commonly cited as a source of nutrients that drive cyanobacterial blooms, also results in seasonal presence of herbicides containing glyphosate, which has recently been shown to impact algal succession in favor of *Planktothrix* spp. over *Microcystis* spp., both by acting as a sporadic potential nutrient source (for *Planktothrix* spp.) and as an herbicidal inhibitor (to *Microcystis* spp.). Complicating matters, it is likely that heterotrophic bacterial populations associated with blooms are responsible for the breakdown of this compound, suggesting a role for both P and the microbial population in the structure of the phytoplankton community (Saxton *et al.* 2011).

There is renewed support for the hypothesis that while phosphorus may control total phytoplankton biomass in many freshwater systems, the chemical speciation of nitrogen shapes algal community structure in many lakes: for a historical review see Lewis and Wurstbaugh, (2008). This in turn has led to a call for dual nutrient management practices, although these management practices are currently under debate, in many cases due to the projected costs of N management (Conley *et al.* 2009, Lewis *et al.* 2011, Schindler 2012). Total nitrogen levels have recently been shown to be particularly important in highly P-enriched systems, although again contradictory evidence has been reported (DeBruyn *et al.* 2004, Dolman *et al.* 2012, Schindler 2012), and it is likely that these systems can experience light-limitation. This last point is of importance, as *Microcystis* spp. have the ability to regulate their own buoyancy through the production of gas vacuoles, keeping cells at the surface and within the reach of photosynthetically available radiation (Reynolds, 1973; Rinta-Kanto et al. 2009b).

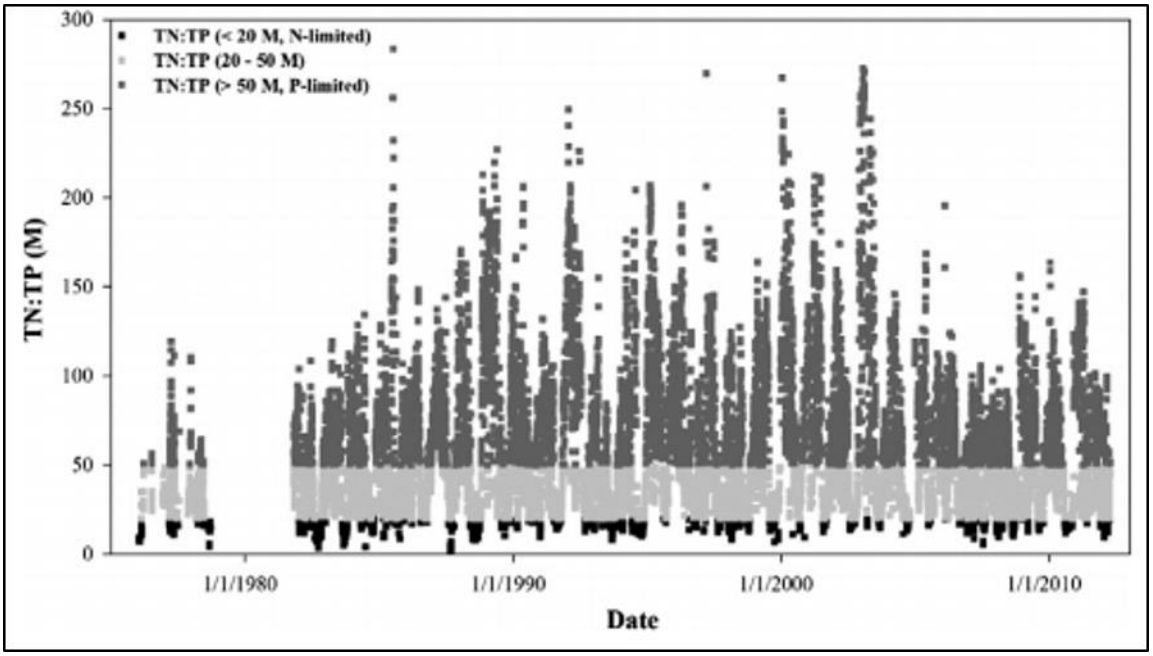

Figure 1.5. Estimated TN (TKN + NO2/NO3): TP molar ratio from the Maumee River tributary loading program between 1975 and 2012, demonstrating relative inputs of these nutrients to the western basin of Lake Erie. Ratios of TN:TP that would be consistent with N-limitation (•), balanced growth, i.e. N or P limitation may be possible (◦), and P-limitation (•), consistent with those observed populations from lakes and oceans are denoted. Data are available from www.heidelberg.edu/academiclife/distinctive/ncwqr, last access Feb. 11, 2013.

Despite historical experimental evidence suggesting phytoplankton biomass production in Lake Erie is P-limited, comprehensive nutrient surveys suggest that tributary inputs to the system can be N-depleted (Figure 1.5). From over 14,000 data points in the Maumee River, waters were N-limited 11.3% of the time and P-limited 48% of the time, using parameters established by Guildford and Hecky (2000) for lakes and oceans. Within this observation remains the caveat that a substantial portion of the suspended load is mineral-bound P, and may not be immediately available to lake phytoplankton and is instead likely rapidly exported to the sediments. Moreover, variations within the data suggest some seasonality, with TN:TP relationships being generally lower in these samples during the August to October period each year. The result is that these inputs provide for variable molar TN:TP ratios (from <25 to >100) in both the waters at the very entrance of Lake Erie as well as farther into the western basin (Chaffin *et al.* 2013). Overall, the data suggest a potential cryptic yet seasonal role for N input than historical theories dictate as well as support for season variations in limiting nutrients (Chaffin et al 2013, Hartig and Wallen 1984).

There is also mounting evidence that suggests specific chemical species of nitrogen, such as urea, may act as a driver for *Microcystis* blooms in the environment, as it is an energetically favorable nutrient source (Davis *et al.* 2010, Finlay *et al.* 2010). The global use of urea-based nitrogen fertilizers has escalated since the 1920s (Finlay 2002). In the last several decades, global urea production has increased at an average of 3.3% annually, with North America contributing 8.8% of total usage (IFA 2011). While the use of urea as a fertilizer is an effective and economical agricultural practice, urea in runoff can harm neighboring aquatic systems (Finlay et al 2010, Glibert *et al.* 2006). Reports of total urea concentration in freshwater systems vary between 1.0 and 10 μM (Berman and Chava 1999, Finlay et al 2010, Park *et al.* 1997). Urea amendments have been shown to preferentially stimulate *Microcystis* biomass over inorganic forms of nitrogen, although *in vitro* culture studies suggest inorganic N may favor growth of nontoxic strains of *Microcystis* over toxin producers (Davis et al 2010). Despite recent interest, published data on urea and its role as a potential driver of *Microcystis* blooms in Lake Erie remains limited and should receive more study.

The relationship between nutrient levels, cyanobacterial biomass and toxicity is further complicated by the impact of additional factors. Where nitrogen availability may matter most is

in the genetic regulation of toxin production. Orihel et al. (2012) found a correlation between microcystin concentration and low N:P ratio based on a statistical examination of Canadian lakes. However, actual concentrations of N:P were not taken into account when calculating the ratios and in fact, total N appeared to be the best predictor of toxin concentration (Scott *et al.* 2013). In contrast, Rinta-Kanto et al. (2009a) found significant correlations between microcystin concentrations and TP across a three year sample set from Lake Erie from 2003 to 2005. This observation was also made for samples collected across a series of North American and Ugandan lakes by Poste and colleagues (2011, 2013). Other relationships to microcystin concentrations, including SiO_2 (Rinta-Kanto et al. 2009a) and carbon depletion (Poste et al. 2013) have also been seen. This includes a series of environmental factors (temperature, light, pH/alkalinity, nutrients), which can be linked to varying degrees of toxin production (Watanabe and Oishi 1985, Wicks and Thiel 1990). At the genetic level, however, molecular evidence has clearly shown that both the ferric uptake regulator protein (FUR) and the global nitrogen regulator (NtcA) can interact with the bidirectional promoter region of the *mcy* toxin cassette (Neilan *et al.* 2013). Within individual cells, the activity of these regulators is specifically controlled by nutrient concentrations: Fe in the case of FUR, and ammonia in the case of NtcA. Although consensus on the specific regulatory mechanism has not yet been reached, it is clear that nutrient concentrations within cells can markedly shape the activity of the toxin production mechanisms within microcystin producers, implying that external nutrient availability should play a role in regulating toxin production.

In terms of the management of toxins, new focus has been placed on the removal of toxins and the use of biological controls, such as heterotrophic bacteria which can degrade cyanotoxins (Manage *et al.* 2009) and the use of barley straw to draw down biovolume of toxin producers in freshwater systems (Newman and Barrett 1993, Purcell *et al.* 2013). To our knowledge neither of these controls are currently being tested in Lake Erie, however, evidence regarding the natural biochemical pathways of degradation of microcystin is mounting (Mou et al., 2013), establishing this as a potentially valuable area of future research.

Consideration also needs to be given to exogenous factors, with the zebra mussel, *Dreissena polymorpha* being the best known example, when looking at factors that influence Lake Erie's toxic cyanobacteria. By 1996, *Dreissena* populations had markedly reduced total

phytoplankton populations in Lake Erie to 20% of the pre-invasion biovolume (Barbiero *et al.* 2006). A myriad of work suggests that filter-feeding zebra mussels may shift freshwater communities toward *Microcystis* by selective feeding (Vanderploeg *et al.* 2001), although contradictory studies suggest that *Microcystis* abundance in the Hudson River decreased in the 4 years post-invasion (Smith *et al.* 1998). This does not appear to be related to toxin formation, as both toxic and nontoxic strains of *Microcystis* are consumed at the same rate (Dionisio Pires *et al.* 2005a). Rather, large colonies of *Microcystis* have been shown to be emitted from mussels as pseudofeces while other forms of phytoplankton are digested. This may contribute to potential algal succession patterns in favor of *Microcystis*. Additional evidence suggests zebra mussels may be an important factor in phosphorus cycling in the lake (Heath *et al.* 1995). Declines in TP in Lake Michigan and Lake Huron since dreissenid invasion have been attributed to the role of mussels as a sink for P (Chapra and Dolan 2012). However, recent research on both zebra (*Dreissena polymorpha*) and quagga (*Dreissena rostriformis*) mussels suggests these bivalves can be a source of SRP in Lake Erie as well as a sink, likely due to a combination of mobilizing P from the particulate biological pool as well as enhancing P-release from sediments (Turner 2010). Finally, the presence of the mussels has led to other changes in water column chemistry: from 1991 – 1996 (around the time *Microcystis* emerged in Lake Erie) spring alkalinity levels dropped 10%, likely due to the uptake of calcium by the growing mussel populations and long-term storage of materials in the shells of dead mussels at the bottom of the lake. Decreases of this nature may increase competition for dissolved CO_{2aq} by primary producers in the systems. As a cyanobacterium, *Microcystis* can exploit bicarbonate *via* carbonic anhydrase activity, providing a selective advantage under dense bloom conditions where pH commonly increases, making CO_{2aq} even less available (Wicks and Thiel 1990; Kotak et al. 2000; Rinta-Kanto et al. 2009a; Poste et al. 2013). This competitive ability has been shown in the laboratory to be enhanced in cells capable of producing microcystin, although the exact biochemical mechanism remains unclear (Van de Waal et al. 2011).

Abiotic factors and the success of *Cyanobacteria* in Lake Erie

Microcystis and other toxin-producing cyanobacteria are thought to be most successful at temperatures above 25º C (Paerl and Huisman 2009), with several early studies demonstrating

optimal growth at temperatures of 26-28 °C for laboratory isolates (Konopka and Brock 1978, Westhuizen and Eloff 1985). The International Panel on Climate Change (IPCC) models for changes in climate predict temperature increases that would favor cyanobacterial dominance in freshwaters and may contribute to an extension of annual bloom durations (IPCC 2007). Warmer temperatures have already resulted in an earlier deterioration of ice cover on Lake Erie, with recent temperature data showing above freezing (0º C) temperatures for the entire month of March 2012 (NWS 2012). How shorter ice seasons, earlier and more extended thermal stratification, and shifts in seasonal nutrient loading will affect cyanobacterial proliferation remains an unanswered question. Shifts in community succession and nutrient cycles will likely occur and need to be considered in future studies.

In recent years, species traditionally associated with tropical climates have been invading temperate lake systems. Of note is the northward expansion of *Cylindrospermopsis raciborskii*, a diazotroph known to produce multiple neurotoxins (Briand *et al.* 2004, Stüken *et al.* 2006). While *C. raciborksii* has been found in North America for a number of years (Chapman and Schelske 1997, Hong *et al.* 2006), only recently was it identified among Lake Erie cyanobacteria (Conroy et al 2007), a sign that Lake Erie may be susceptible to further invasion of non-indigenous species as global temperatures rise. With rising temperature, there is an additional risk for invasion of toxic cyanobacteria from the lower Great Lakes to Lakes Huron and Superior. Previous work has shown *Microcystis* and *Anabaena* spp. are transported to these lakes *via* ship ballast (Doblin *et al.* 2007). However, bloom formation in Lake Superior and the central and northern basins of Lake Huron is thought to in part be constrained by temperature. Warming temperatures trends, in combination with encroaching urban centers (which will bring added nutrients), in some of the nearer shore embayments may encourage bloom events farther north, as temperatures are predicted to warm by 0.2° C per year over the next two decades (IPCC 2007).

Climate change is projected to affect not only temperature trends but also mixing and stratification patterns, precipitation, irradiance, and the frequency of storm events, all of which may potentially impact the success of cHABs in Lake Erie and beyond. Specifically, increased precipitation and resulting runoff has been correlated with recent larger-scale bloom events in Lake Erie (Koslow 2013, Michalak et al 2013). Heavier precipitation generally contributes to

elevated levels of nutrient loading, which can exacerbate levels of toxic cyanobacteria in the lake. Indeed, a recent report linking precipitation to bloom severity indicates that there may be a link between wetter years (1998, 2003, 2008, 2011) and bloom size (Koslow 2013). Conversely, drought years (1999, 2001, 2012) have been associated with below average bloom size. If weather patterns become more extreme in the future, Lake Erie bloom dynamics may continue to evolve. It is highly likely that these patterns will also affect nutrient loading in both quantity and chemical species, exacerbating the problem of anticipating harmful bloom events.

The continuing saga: current research and policy trends in Lake Erie

Lake Erie cHAB research continues to evolve as new tools are applied to the system. A new focus on the entire microbial bloom community, rather than a sole concentration on cyanobacteria, has developed due to the application of a more holistic "systems biology" approach. When applied to studying bloom ecology, this method moves beyond the study of a single organism, and provides insight on the relative contributions of biotic and abiotic driving forces behind massive bloom events. Emerging tools, such as the environmental application of genomics, transcriptomics and even metabolomics, allow researchers to examine not only what the organism of interest is doing (or potentially doing) but also now provide insight on the role of the entire planktonic community in this process. As of yet, the application of these studies to the Lake Erie systems remain limited (see Steffen et al (2012) and Mou et al (2013) for examples). As they accumulate, however, these efforts will allow for a more comprehensive approach to modeling and forecasting efforts and may eventually lead to novel management practices in the field.

Current research efforts also continue to exploit model organisms, and *in vitro* culture studies continue to prove useful for baseline knowledge of bloom-causing organisms. Design and interpretation of such studies are enhanced by the availability of detailed genomic information of model organisms (Frangeul *et al.* 2008b, Kaneko *et al.* 2007b, Rounge *et al.* 2009, Sugita *et al.* 2007). Continued advancement of research efforts is restricted by the few complete freshwater cyanobacterial genomes publicly available. When this deficit is resolved, we will have access to a plethora of new knowledge useful to understanding bloom formation and persistence, as well

as provide for the development of new tools that will allow for better diagnostics of bloom constituents and how to better mitigate bloom events.

Moving forward, application of knowledge gained from model organism studies to the study of environmental communities will provide a more comprehensive picture of bloom dynamics – especially at the biochemical level. In the recent past, valuable insight has been provided from long-term data sets on nutrients and biological abundance and activity at given locations and times. Future studies combining "-omics" methods, model organisms, and molecular approaches in the field will advance understanding of blooms and toxin production within the wider context of the entire bloom system. In addition to this redirection, a renewed focus is likely needed in hypothesis-based science. Combined experimental and descriptive approaches have proven quite powerful for interpretation, understanding, and advancement of our knowledge and policies regarding P-loading. Considering the current descriptive and limited experimental data, we should move forward with a more comprehensive experimental approach to examine issues such as the varying role of different nutrients in biomass generation and toxin production in impacted lakes.

A critical role for science within Lake Erie is to also inform both lake- and land-management stewards who represent the best interests of the general population. Efforts to improve lake quality and address trophic status are underway at local, state, national, and international levels in both Canada and the United States. Current focus on reduction of both nonpoint source loading of phosphorus and the escalating SRP levels remains a priority for national and state-level organizations (EPA 2010). And the ongoing debate within the scientific community as to the best method for nutrient management has yet to be resolved. Evidence exists to support a focus on a single nutrient (P) and a dual-nutrient management strategy (both N and P) (Lewis et al 2011, Schindler 2012). Further efforts to resolve this debate (see (Orihel *et al.* (2012), Schindler *et al.* (2008), Scott and McCarthy (2010)); Scott et al (2013) for examples) will advance both management practices and the science of freshwater cHABs. Continued sponsorship of monitoring and scientific endeavors to understand cHAB events in Lake Erie and gain insight into the forces contributing to events is also an important goal of provincial, state, and national organizations. It should remain a priority, however, to continue to engage in

experimental science along with monitoring and modeling efforts, as such efforts create context and validation.

Acknowledgements

This review is adapted from a report prepared for the International Joint Commission (IJC) on the Great Lakes by MMS and SWW. We thank the IJC as well as the National Science Foundation (IOS0841918, DEB1240870), the National Oceanic and Atmospheric Administration (NA11NOS4780021), Environment Canada and a Wallace-Dean Fellowship to MMS for support. This is contribution PCM11, funded by award number NA11NOS4780021 from the NOAA PCMHAB program. We thank Richard Stumpf (NOAA) for assistance with satellite imagery. We also thank Professors R. Michael McKay, Robert Hecky, Joe Makarewicz, Peter Richards and two anonymous reviewers for comments that greatly improved the manuscript.

PART II: APPLICATION OF SYSTEMS BIOLOGY TO AQUATIC MICROBIAL ECOLOGY

The ongoing evolution of microbial ecology.

The tools associated with the umbrella field of "molecular biology" are universally acknowledged to have revolutionized our understanding of not just microbiology, but all biological systems. The term molecular biology, first coined in 1938 (Weaver 1970), encompasses a number of paradigm shifting methodological breakthroughs, including nucleic acid sequencing (Sanger *et al.* 1977), polymerase chain reaction (PCR) (Mullis and Faloona 1987), 16S rRNA phylogenetic analysis (Sogin *et al.* 1972, Woese and Fox 1977), and next generation sequencing technologies (Bentley *et al.* 2008, Ronaghi *et al.* 1998). Widespread implementation of each of these innovations marked a shift in the study and understanding of microbial life. Currently, scientists are only beginning to grasp the power of the so-called "-omics" techniques to the study of microbial ecosystems, which will be discussed more thoroughly below.

The study of aquatic biology, however, long predates the field of molecular biology, as it is thought to originate with Aristotle (Voultsiadou and Vafidis 2007). It was not until several millennia later that the more narrow field of aquatic microbiology first arose (Oren 2002). Even later still was the application of molecular techniques to the field. In fact, some argue the deeply engrained methodologies in marine sciences prevented the implementation of modern molecular techniques longer than in other fields of biological study (McLean 2013). Today, however, molecular techniques are considered standard methodology for aquatic microbial ecology, and their application has led to a vast number of important and high-profile findings across multiple domains of life (Fischer *et al.* 2010, Spang *et al.* 2010, Venter *et al.* 2004). The development of methods to characterize microbial communities based on both 16S rRNA genes (Lozupone and Knight 2005, Muyzer *et al.* 1993) and genomic DNA (Venter et al 2004) assemblages opened new lines of research leading to a better understanding of microbial community structure in function in a variety of aquatic environments. We now know that the concept of a "microbiome," so often associated with the human body, applies across the biosphere. The microbiome, or bacterial community assemblage, of aquatic systems has been shown to play an integral role in carbon and other nutrient cycling (Azam *et al.* 1983) and pathogen defense of those systems.

This applies to both large-scale aquatic food webs and to smaller scale systems including coral reefs (Bourne *et al.* 2009), hydrothermal vents (Xie *et al.* 2011), macroalgae (Miranda *et al.* 2013), and harmful algal blooms (Mou et al 2013, Sison-Mangus *et al.* 2014).

One primary focus of aquatic microbiology, both historically and contemporarily, has been the study of harmful algal blooms (HABs). Chiefly studied due to their destructive impact on aquatic ecosystems, HAB-forming organisms are theorized to have had a long history of contributing to mass extinction events. Such theories are based on observations of the fossil record (Castle and Rodgers 2009). Cyanobacteria, which are responsible for harmful algal blooms (cHABs) in freshwater and estuarine environments, are among the most ancient extant organisms on the planet (Blank and SÁNchez-Baracaldo 2010). Due to their role in the Great Oxygenation Event ~2.3 billion years ago (Bekker *et al.* 2004), they are thought to be partly responsible for one of earth's first mass extinction events, as rising O_2 concentrations threatened anaerobic life. They continue to play a fundamental role in global oxygen cycling, as they are thought to produce 30% of the planet's atmospheric oxygen. As reviewed in the first section of this chapter, cyanobacteria can pose a potent threat to freshwater ecosystems under bloom conditions. With the expansive use of molecular and "–omics" techniques over the last decade, our knowledge of bloom forming cyanobacteria has broadened and will continue to do so with further application of hypothesis-based inquiry of such methods.

A systems biology approach.

Systems biology is broadly defined as the study of the interactions within a biological system and how those interactions lead to or influence system function. The expansion of systems biology during the "big data" era of biology represents a shift from reductionism toward a more holistic approach to understanding biology, and the historical progress of this shift is thoroughly reviewed in Trewavas (2006). Today, systems biology is often associated with "-omics" techniques, however, this holistic approach to biology is not solely a 21st century trend. Some could argue that G. Evenlyn Hutchinson, often referred to as the father of modern ecology, applied a systems biology approach to ecological study long before it was associated with the genomics era.

The toolset employed by scientists electing to take a systems biology approach often includes genomics, transcriptomics, proteomics, and/or metabolomics techniques. An important consideration when applying these techniques is that they are simply tools with which hypotheses about a system of interest can be generated or tested. In terms of their application to *Microcystis*, genomics has been the most widely and frequently used. The first genome of a *Microcystis* strain was published in 2007 (Kaneko *et al.* 2007a). Since that primary publication, several others have been published but only the genome of *M. aeruginosa* NIES 843 has been fully closed and annotated, making it the primary model for "-omics" studies of *Microcystis* species (Table 1.3). Metagenomic studies, both targeted and shotgun, have been completed for *Microcystis* communities in the environment. Toxin genes and 16S rRNA genes are often the focus of targeted studies, which are widespread in the *Microcystis* literature. More recently, shotgun metagenomic studies have provided new insight into whole bacterial communities associated with *Microcystis*. Fewer functional studies have been completed, although this represents a rapidly expanding area of interest in *Microcystis*, and more broadly, cHAB research.

Biological system selection.

In order to understand the dynamics of microorganisms in the environment, they are studied in culture under controlled laboratory conditions. Often, work on cultured isolates represents a preliminary step in the study of a microbial system; observations made in culture must be tested and validated in the field to determine ecological relevance. The value in cultured isolates lays in their potential to generate fingerprints or biomarkers in the environment for certain conditions tested in the laboratory. However, caution must be exercised, because results of environmental studies can be skewed according to laboratory-based foundational studies (Rappé 2013).

The systems studied herein include cultures of *Microcystis aeruginosa* NIES 843, as well as larger lake ecosystems in North America (Lake Erie, Grand Lake St Marys) and Asia (Lake Tai). Lake Tai, or *Taihu*, is a widely studied eutrophic shallow lake with a watershed that supports 40 million people in the Jiangsu Province of China. This lake has been plagued by annual blooms of *Microcystis* in recent years, culminating in a documented drinking water crisis in 2007 (Qin *et al.* 2010). Research on this system is largely motivated by the need to develop

Table 1.3. Summary of "-omics" studies completed on *Microcystis*.

Technique	Application	Representative References
Genomics	Isolate genomic sequencing	(Frangeul *et al.* 2008a, Humbert *et al.* 2013, Kaneko et al 2007a)
Targeted metagenomics	Specific gene environmental surveys	(Baxa *et al.* 2010, Davis *et al.* 2009, Kurmayer *et al.* 2002, Ouellette et al 2006, Rinta-Kanto et al 2005)
Whole genome shotgun (WGS) metagenomics	Community surveys of DNA to determine structure and functional potential	(Mou et al 2013, Steffen et al 2012)
Targeted transcriptomics	Measurement of mRNA transcript abundance of specific genes	(Gobler *et al.* 2007, Harke *et al.* 2012, Qian *et al.* 2012, Wood *et al.* 2011)
Transcriptomics	Analysis of isolate mRNA transcript abundance under controlled experimental conditions	(Harke and Gobler 2013, Straub *et al.* 2011)
Proteomics	2D gel analysis of protein content from cultured isolates	(Alexova *et al.* 2011, Tonietto *et al.* 2012)

more robust and effective management practices. However, Taihu may also act as a model for future bloom events as trophic status worsens for other large lake systems around the world, such as Lake Erie.

The following chapters represent a series of studies marking a comprehensive application of molecular and next-generation techniques to further our understanding of the ecologically important cHAB organism, *Microcystis aeruginosa*. These studies examine the *Microcystis* system at the cellular level both in the environment (Chapter 2) and in the lab (Chapter 4), as well as the entire *Microcystis* bloom-associated microbial community (Chapters 3 and 5). This synthesis began with the more widely applied targeted approach to understanding the cyanobacterial population dynamics in a highly toxic bloom in Grand Lakes St Marys (Ohio, USA) during summer 2010. Both the total cyanobacterial community (*via* 16S gene) and the toxin-producing subset (*via mcyA* gene) were tracked over one summer, and results indicated a shift in the toxin-producing community between July and August 2010. While an understanding of the cyanobacterial population is important from both ecological and ecosystem management standpoints, this type of study overlooks an important component of the community: the heterotrophic population. In order to better understand the total bloom-associated community structure, we undertook a more expanded approach using shotgun metagenomics to examine the microbial community from three separate samples taken from three sites around the world. This study provided a strong foundation for future genomic studies of freshwater systems, and generated new insight into the dynamics of freshwater microbial populations. However, as previously noted, DNA data can provide insight into community structure and functional potential, but further analyses are required to obtain more direct information on community function. To this end, the next study presented herein is a transcriptomic study used to examine the transcriptional response of *Microcystis aeruginosa* NIES 843 to nutrient variability. As this was a broad exploratory study using an "-omics" approach, the results are presented in the form of new hypotheses about this system. These hypotheses have both metabolic and evolutionary implications and will be empirically tested during future research endeavors. The three preceding studies resulted in the generation of numerous hypotheses, and in order to validate our observations, a final metatranscriptomic survey of the western basin of Lake Erie is reported here. Not only does this final examination of bloom communities provide environmental

validation of observations made in previous studies, but it proves to be a strong method for detection of both top-down and bottom-up ecological controls on bloom populations. Taken together, these studies provide new insight into how *Microcystis* interacts with both its abiotic and biotic environment, bringing us closer to answering the question, "What makes *Microcystis* bloom?"

LIST OF REFERENCES

Ashworth W (1986). *The late, Great Lakes: An environmental history*. Wayne State University Press.

Avila J (2012). Lake Invaders. ABC News.

Azam F, Fenchel T, Field J, Gray J, Meyer-Reil L, Thingstad F (1983). The ecological role of water-column microbes in the sea. *Marine ecology progress series Oldendorf* **10:** 257-263.

Azevedo SMFO, Carmichael WW, Jochimsen EM, Rinehart KL, Lau S, Shaw GR *et al* (2002). Human intoxication by microcystins during renal dialysis treatment in Caruaru, Brazil. *Toxicology* **181–182:** 441-446.

Barbiero RP, Rockwell DC, Warren GJ, Tuchman ML (2006). Changes in spring phytoplankton communities and nutrient dynamics in the eastern basin of Lake Erie since the invasion of *Dreissena* spp. *Can J Fish Aquat Sci* **63:** 1549-1563.

Becker RH, Sultan MI, Boyer GL, Twiss MR, Konopko E (2009). Mapping cyanobacterial blooms in the Great Lakes using MODIS. *Journal of Great Lakes Research* **35:** 447-453.

Beeton AM (1961). Environmental changes in Lake Erie. *T Am Fish Soc* **90:** 153-159.

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

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG *et al* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53-59.

Berman T, Chava S (1999). Algal growth on organic compounds as nitrogen sources. *J Plankton Res* **21:** 1423-1437.

Binding CE, Greenberg TA, Bukata RP (2012). An analysis of MODIS-derived algal and mineral turbidity in Lake Erie. *Journal of Great Lakes Research* **38:** 107-116.

Blank CE, SÁNchez-Baracaldo P (2010). Timing of morphological and ecological innovations in the cyanobacteria – a key to understanding the rise in atmospheric oxygen. *Geobiology* **8:** 1-23.

Bourne DG, Garren M, Work TM, Rosenberg E, Smith GW, Harvell CD (2009). Microbial disease and the coral holobiont. *Trends in Microbiology* **17:** 554-562.

Bouzat JL, Hoostal MJ, Looft T (2013). Spatial patterns of bacterial community composition within Lake Erie sediments. *Journal of Great Lakes Research* **39**.

Briand J-F, Leboulanger C, Humbert J-F, Bernard C, Dufour P (2004). *Cylindrospermopsis raciborskii* (*Cyanobacteria*) invasion at mid-latitudes: selection, wide physiological tolerance, or global warming? *J Phycol* **40:** 231-238.

Bridgeman TB, Penamon WA (2010). *Lyngbya wollei* in western Lake Erie. *Journal of Great Lakes Research* **36:** 167-171.

Bridgeman TB, Chaffin JD, Kane DD, Conroy JD, Panek SE, Armenio PM (2012). From river to lake: Phosphorus partitioning and algal community compositional changes in western Lake Erie. *Journal of Great Lakes Research* **38:** 90-97.

Brittain S, Mohamed ZA, Wang J, Lehmann VK, Carmichael WW, Rinehart KL (2000). Isolation and characterization of microcystins from a river nile strain of *Oscillatoria tenuis Agardh ex Gomont*. *Toxicon* **38:** 1759-1771.

Burns NM, Rockwell DC, Bertram PE, Dolan DM, Ciborowski JJH (2005). Trends in temperature, secchi depth, and dissolved oxygen depletion rates in the central basin of Lake Erie, 1983–2002. *Journal of Great Lakes Research* **31, Supplement 2:** 35-49.

Carrick HJ, Schelske CL (1997). Have we overlooked the importance of small phytoplankton in productive waters? *Limnol Oceanogr* **42:** 1613-1621.

Castle JW, Rodgers JH (2009). Hypothesis for the role of toxin-producing algae in Phanerozoic mass extinctions based on evidence from the geologic record and modern environments. *Environmental Geosciences* **16:** 1-23.

Chaffin JD, Bridgeman TB, Bade DL (2013). Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. *Adv Microbiol* **3:** 16.

Chapman AD, Schelske CL (1997). Recent appearance of *Cylindrospermopsis* (*Cyanobacteria*) in five hypereutrophic Florida lakes. *J Phycol* **33:** 191-195.

Charlton MN (1980). Oxygen depletion in Lake Erie: has there been any change? *Can J Fish Aquat Sci* **37:** 72-80.

Charlton MN, Milne JE, Booth WG, Chiocchio F (1993). Lake Erie Offshore in 1990 - restoration and resilience in the central basin. *Journal of Great Lakes Research* **19:** 291-309.

Chorus I, Bartram J (1999). Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management/edited by Ingrid Chorus and Jamie Bertram.

Codd G (1996). Harmful Algae News. *IOC of UNESCO*. United Nations Education, Scientific and Cultural Organization, Paris.

Conley DJ, Paerl HW, Howarth RW, Boesch DF, Seitzinger SP, Havens KE *et al* (2009). Controlling eutrophication: Nitrogen and phosphorus. *Science* **323:** 1014-1015.

Conroy JD, Kane DD, Dolan DM, Edwards WJ, Charlton MN, Culver DA (2005). Temporal trends in Lake Erie plankton biomass: Roles of external phosphorus loading and dreissenid mussels. *Journal of Great Lakes Research* **31:** 89-110.

Conroy JD, Quinlan EL, Kane DD, Culver DA (2007). *Cylindrospermopsis* in Lake Erie: Testing its association with other cyanobacterial genera and major limnological parameters. *Journal of Great Lakes Research* **33:** 519-535.

Davis CC (1954). A preliminary study of the plankton of the Cleveland Harbor Area, Ohio: II. The distribution and quantity of the phytoplankton. *Ecol Monogr* **24:** 322-347.

Davis CC (1958). An approach to some problems of secondary production in the western Lake Erie region. *Limnol Oceanogr* **3:** 15-28.

Davis CC (1962). The plankton of the Cleveland Harbor Area of Lake Erie in 1956-1957. *Ecol Monogr* **32:** 209-247.

Davis TW, Harke MJ, Marcoval MA, Goleski J, Orano-Dawson C, Berry DL *et al* (2010). Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Aquat Microb Ecol* **61:** 149-162.

Davis TW, Koch F, Marcoval MA, Wilhelm SW, Gobler CJ (2012). Mesozooplankton and microzooplankton grazing during cyanobacterial blooms in the western basin of Lake Erie. *Harmful Algae* **15:** 26-35.

Dawson RM (1998). The toxicology of microcystins. *Toxicon* **36:** 953-962.

DeBruyn JM, Leigh-Bell JA, McKay RML, Bourbonniere RA, Wilhelm SW (2004). Microbial distributions and the impact of phosphorus on bacterial activity in Lake Erie. *Journal of Great Lakes Research* **30:** 166-183.

Dietrich D, Hoeger S (2005). Guidance values for microcystins in water and cyanobacterial supplement products (blue-green algal supplements): a reasonable or misguided approach? *Toxicol Appl Pharm* **203:** 273-289.

Dionisio Pires LM, Bontes BM, Van Donk E, Ibelings BW (2005a). Grazing on colonial and filamentous, toxic and non-toxic cyanobacteria by the zebra mussel *Dreissena polymorpha*. *J Plankton Res* **27:** 331-339.

Dionisio Pires LM, Ibelings BW, Brehm M, Donk E (2005b). Comparing grazing on lake seston by *Dreissena* and *Daphnia*: Lessons for biomanipulation. *Microb Ecol* **50:** 242-252.

Doblin MA, Coyne KJ, Rinta-Kanto JM, Wilhelm SW, Dobbs FC (2007). Dynamics and short-term survival of toxic cyanobacteria species in ballast water from NOBOB vessels transiting the Great Lakes - implications for HAB invasions. *Harmful Algae* **6:** 519-530.

Dolman AM, Rücker J, Pick FR, Fastner J, Rohrlack T, Mischke U *et al* (2012). Cyanobacteria and cyanotoxins: The influence of nitrogen versus phosphorus. *PLoS ONE* **7:** e38757.

Dyble J, Fahnenstiel GL, Litaker RW, Millie DF, Tester PA (2008). Microcystin concentrations and genetic diversity of *Microcystis* in the lower Great Lakes. *Environmental Toxicology* **23:** 507-516.

EPA (2010). Ohio Lake Erie Phosphorus Task Force Final Report. Ohio EPA. p 109.

Finlay K, Patoine A, Donald DB, Bogard MJ, Leavitt PR (2010). Experimental evidence that pollution with urea can degrade water quality in phosphorus-rich lakes of the Northern Great Plains. *Limnol Oceanogr* **55:** 1213-1230.

Finlay MR (2002). Enriching the earth: Fritz Haber, Carl Bosch, and the transformation of world food production. *Brit J Hist Sci* **35:** 120-121.

Fischer MG, Allen MJ, Wilson WH, Suttle CA (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proceedings of the National Academy of Sciences* **107:** 19508-19513.

Fortner RWaM, V.J. (1987). *The Great Lake Erie*. The Ohio State University Research Foundation: Columbus, OH.

Francis G (1878). Poisonous Australian lake. *Nature* **18:** 11-12.

Frangeul L, Quillardet P, Castets AM, Humbert JF, Matthijs HCP, Cortez D *et al* (2008). Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *Bmc Genomics* **9**.

Glibert PM, Harrison J, Heil C, Seitzinger S (2006). Escalating worldwide use of urea - a global change contributing to coastal eutrophication. *Biogeochemistry* **77:** 441-463.

Grosse Y, Baan R, Straif K, Secretan B, El Ghissassi F, Cogliano V (2006). Carcinogenicity of nitrate, nitrite, and cyanobacterial peptide toxins. *The Lancet Oncology* **7:** 628-629.

Guildford SJ, Hecky RE (2000). Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? *Limnol Oceanogr* **45:** 1213-1223.

Hartig JH, Wallen DG (1984). Seasonal variation of nutrient limitation in western Lake Erie. *Journal of Great Lakes Research* **10:** 449-460.

Hasler AD (1969). Cultural eutrophication is reversible. *Bioscience* **19:** 425.

Heath RT, Fahnenstiel GL, Gardner WS, Cavaletto JF, Hwang SJ (1995). Ecosystem-level effects of zebra mussels (*Dreissena polymorpha*): An enclosure experiment in Saginaw Bay, Lake Huron. *Journal of Great Lakes Research* **21:** 501-516.

Hong Y, Steinman A, Biddanda B, Rediske R, Fahnenstiel G (2006). Occurrence of the toxin-producing cyanobacterium *Cylindrospermopsis raciborskii* in Mona and Muskegon Lakes, Michigan. *Journal of Great Lakes Research* **32:** 645-652.

IFA (2011). Production and International Trade: International Fertilizer Industry Association

IJC (1986). Executive summary of the US Task Force Plans for the Phosphorus Load Reduction to Lake Erie, Lake Ontario, and Saginaw Bay. International Joint Commission: Windsor, Ontario.

IPCC (2007). Climate Change 2007 Synthesis Report. In: Pachauri RLaR, A. (ed). International Panel on Climate Change Geneva, Switzerland. p 104.

Jochimsen EM, Carmichael WW, An J, Cardo DM, Cookson ST, Holmes CEM *et al* (1998). Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *New England Journal of Medicine* **338:** 873-878.

Joosse PJ, Baker DB (2011). Context for re-evaluating agricultural source phosphorus loadings to the Great Lakes. *Canadian Journal of Soil Science* **91:** 317-327.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M *et al* (2007a). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Research* **14:** 247-256.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M *et al* (2007b). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Res* **14:** 247-256.

Konopka A, Brock TD (1978). Effect of temperature on blue-green algae (*Cyanobacteria*) in Lake Mendota. *Appl Environ Microb* **36:** 572-576.

Koslow L, Benka (2013). Taken by storm: How heavy rain is worsening algal blooms in Lake Erie. National Wildlife Federation. p 23.

Lavrentyev PJ, McCarthy MJ, Klarer DM, Jochem F, Gardner WS (2004). Estuarine microbial food web patterns in a Lake Erie coastal wetland. *Microb Ecol* **48:** 567-577.

Lewis WM, Wurtsbaugh WA, Paerl HW (2011). Rationale for control of anthropogenic nitrogen and phosphorus to reduce eutrophication of inland waters. *Environmental Science & Technology* **45:** 10300-10305.

Lozupone C, Knight R (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microb* **71:** 8228-8235.

Makarewicz JC, Bertram P (1991). Evidence for the restoration of the Lake Erie ecosystem - water-quality, oxygen levels, and pelagic function appear to be improving. *Bioscience* **41:** 216-223.

Makarewicz JC (1993). Phytoplankton biomass and species composition in Lake Erie, 1970 to 1987. *Journal of Great Lakes Research* **19:** 258-274.

Manage PM, Edwards C, Singh BK, Lawton LA (2009). Isolation and identification of novel microcystin-degrading bacteria. *Appl Environ Microb* **75:** 6924-6928.

McLean TI (2013). "Eco-omics": A review of the application of genomics, transcriptomics, and proteomics for the study of the ecology of harmful algae. *Microb Ecol* **65:** 901-915.

Michalak AM, Anderson EJ, Beletsky D, Boland S, Bosch NS, Bridgeman TB *et al* (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 6448-6452.

Miller AP, Tisdale ES (1931). Epidemic of intestinal disorders in Charleston, W. Va., occurring simultaneously with unprecedented water supply conditions. *American Journal of Public Health and the Nations Health* **21:** 198-200.

Millie DF, Fahnenstiel GL, Bressie JD, Pigg RJ, Rediske RR, Klarer DM *et al* (2009). Late-summer phytoplankton in western Lake Erie (Laurentian Great Lakes): bloom distributions, toxicity, and environmental influences. *Aquat Ecol* **43:** 915-934.

Miranda LN, Hutchison K, Grossman AR, Brawley SH (2013). Diversity and abundance of the bacterial community of the red macroalga *Porphyra umbilicalis*: did bacterial farmers produce macroalgae? *PLoS ONE* **8:** e58269.

Morgan J (1995). *Dr. Seuss & Mr. Geisel : a biography / Judith & Neil Morgan*. Random House: New York.

Mou X, Lu X, Jacob J, Sun S, Heath R (2013). Metagenomic identification of bacterioplankton taxa and pathways involved in microcystin degradation in Lake Erie. *PLoS ONE* **8:** e61890.

Mullis KB, Faloona FA (1987). [21] Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. In: Ray W (ed). *Methods in Enzymology*. Academic Press. pp 335-350.

Murphy TP, Irvine K, Guo J, Davies J, Murkin H, Charlton M *et al* (2003). New microcystin concerns in the lower Great Lakes.

Muyzer G, De Waal EC, Uitterlinden AG (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microb* **59:** 695-700.

Neilan BA, Pearson LA, Muenchhoff J, Moffitt MC, Dittmann E (2013). Environmental conditions that influence toxin biosynthesis in cyanobacteria. *Environmental Microbiology* **15:** 1239-1253.

Newman JR, Barrett PRF (1993). Control of *Microcystis aeruginosa* by decomposing barley straw. *J Aquat Plant Manage* **31:** 203-206.

Nicholls KH, Standen DW, Hopkins GJ, Carney EC (1977). Declines in the near-shore phytoplankton of Lake Erie's western basin since 1971. *Journal of Great Lakes Research* **3:** 72-78.

NWS (2012). Lake Erie Water Temperatures. In: NOAA (ed): National Weather Service Forecast Office Buffalo

Oren A (2002). Halophilic Microorganisms in their Natural Environment and in Culture—an Historical Introduction. *Halophilic Microorganisms and their Environments***:** 3-16.

Orihel Diane M, Bird David F, Brylinsky M, Chen H, Donald Derek B, Huang Dorothy Y *et al* (2012). High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **69:** 1457-1462.

Ouellette AJA, Wilhelm SW (2003). Toxic cyanobacteria: the evolving molecular toolbox. *Front Ecol Environ* **1:** 359-366.

Ouellette AJA, Handy SM, Wilhelm SW (2006). Toxic *Microcystis* is widespread in Lake Erie: PCR detection of toxin genes and molecular characterization of associated cyanobacterial communities. *Microb Ecol* **51:** 154-165.

Oziol L, Bouaïcha N (2010). First evidence of estrogenic potential of the cyanobacterial heptotoxins the nodularin-R and the microcystin-LR in cultured mammalian cells. *Journal of Hazardous Materials* **174:** 610-615.

Paerl H (2009). Controlling eutrophication along the freshwater–marine continuum: Dual nutrient (N and P) reductions are essential. *Estuaries and Coasts* **32:** 593-601.

Paerl HW, Pinckney JL, Fear JM, Peierls BL (1998). Ecosystem responses to internal and watershed organic matter loading: consequences for hypoxia in the eutrophying Neuse river estuary, North Carolina, USA. *Mar Ecol-Prog Ser* **166:** 17-25.

Paerl HW, Fulton RS (2006). Ecology of Harmful Cyanobacteria
Ecology of Harmful Algae. In: Granéli E, Turner JT (eds). Springer Berlin Heidelberg. pp 95-109.

Paerl HW, Huisman J (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports* **1:** 27-37.

Paerl HW, Otten TG (2013). Harmful cyanobacterial blooms: Causes, consequences, and controls. *Microb Ecol*.

Park MG, Shim JH, Cho BC (1997). Urea decomposition activities in an ammonium enriched freshwater pond. *Aquat Microb Ecol* **13:** 303-311.

Purcell D, Parsons SA, Jefferson B, Holden S, Campbell A, Wallen A *et al* (2013). Experiences of algal bloom control using green solutions barley straw and ultrasound, an industry perspective. *Water and Environment Journal* **27:** 148-156.

Qin B, Xu P, Wu Q, Luo L, Zhang Y (2007). Environmental issues of Lake Taihu, China. In: Qin B, Liu Z, Havens K (eds). *Eutrophication of Shallow Lakes with Special Reference to Lake Taihu, China*. Springer Netherlands. pp 3-14.

Qin B, Zhu G, Gao G, Zhang Y, Li W, Paerl H *et al* (2010). A drinking water crisis in Lake Taihu, China: Linkage to climatic variability and lake management. *Environmental Management* **45:** 105-112.

Rao YR, Hawley N, Charlton MN, Schertzer WM (2008). Physical processes and hypoxia in the central basin of Lake Erie. *Limnol Oceanogr* **53:** 2007-2020.

Rappé MS (2013). Stabilizing the foundation of the house that ‘omics builds: the evolving value of cultured isolates to marine microbiology. *Current Opinion in Microbiology* **16:** 618-624.

Rinta-Kanto JM, Ouellette AJA, Boyer GL, Twiss MR, Bridgeman TB, Wilhelm SW (2005). Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. *Environmental Science & Technology* **39:** 4198-4205.

Rinta-Kanto JM, Wilhelm SW (2006). Diversity of microcystin-producing cyanobacteria in spatially isolated regions of Lake Erie. *Appl Environ Microb* **72:** 5083-5085.

Rinta-Kanto JM, Konopko EA, DeBruyn JM, Bourbonniere RA, Boyer GL, Wilhelm SW (2009). Lake Erie *Microcystis*: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. *Harmful Algae* **8:** 665-673.

Rogers ED, Henry TB, Twiner MJ, Gouffon JS, McPherson JT, Boyer GL *et al* (2011). Global gene expression profiling in larval zebrafish exposed to microcystin-LR and *Microcystis* reveals endocrine disrupting effects of cyanobacteria. *Environmental Science & Technology* **45:** 1962-1969.

Ronaghi M, Uhlén M, Nyrén P (1998). A sequencing method based on real-time pyrophosphate. *Science* **281:** 363-365.

Rounge TB, Rohrlack T, Nederbragt AJ, Kristensen T, Jakobsen KS (2009). A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. *Bmc Genomics* **10**.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74:** 5463-5467.

Sant'Anna CL, de Carvalho LR, Fiore MF, Silva-Stenico ME, Lorenzi AS, Rios FR *et al* (2011). Highly toxic *Microcystis aeruginosa* strain, isolated from São Paulo, Brazil, produces hepatotoxins and paralytic shellfish poison neurotoxins. *Neurotox Res* **19:** 389-402.

Sant′Anna C, de Carvalho L, Fiore M, Silva-Stenico M, Lorenzi A, Rios F *et al* (2011). Highly toxic *Microcystis aeruginosa* strain, isolated from São Paulo, Brazil, produces hepatotoxins and paralytic shellfish poison neurotoxins. *Neurotoxicity Research* **19:** 389-402.

Saxton MA, Morrow EA, Bourbonniere RA, Wilhelm SW (2011). Glyphosate influence on phytoplankton community structure in Lake Erie. *Journal of Great Lakes Research* **37:** 683-690.

Saxton MA, Arnold RJ, Bourbonniere RA, McKay RM, Wilhelm SW (2012). Plasticity of total and intracellular phosphorus quotas in *Microcystis aeruginosa* cultures and Lake Erie algal assemblages. *Frontiers in Microbiology* **3**.

Schindler DW (1977). Evolution of phosphorus limitation in lakes. *Science* **195:** 260-262.

Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ *et al* (2008). Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 11254-11258.

Schindler DW (2012). The dilemma of controlling cultural eutrophication of lakes. *Proceedings of the Royal Society B: Biological Sciences*.

Schottler SP, Eisenreich SJ (1994). Herbicides in the Great Lakes. *Environmental Science & Technology* **28:** 2228-2232.

Scott JT, McCarthy MJ (2010). Nitrogen fixation may not balance the nitrogen pool in lakes over timescales relevant to eutrophication management. *Limnol Oceanogr* **55:** 1265-1270.

Scott JT, McCarthy MJ, Otten TG, Steffen MM, Baker BC, Grantz EM *et al* (2013). Comment: An alternative interpretation of the relationship between TN:TP and microcystins in Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **70:** 1265-1268.

Seuss (1971). *The Lorax*. Random House: New York.

Sison-Mangus MP, Jiang S, Tran KN, Kudela RM (2014). Host-specific adaptation governs the interaction of the marine diatom, Pseudo-nitzschia and their microbiota. *ISME J* **8:** 63-76.

Smith JL, Boyer GL, Zimba PV (2008). A review of cyanobacterial odorous and bioactive metabolites: Impacts and management alternatives in aquaculture. *Aquaculture* **280:** 5-20.

Smith TE, Stevenson RJ, Caraco NF, Cole JJ (1998). Changes in phytoplankton community structure during the zebra mussel (*Dreissena polymorpha*) invasion of the Hudson River (New York). *J Plankton Res* **20:** 1567-1579.

Sogin SJ, Sogin ML, Woese CR (1972). Phylogenetic measurement in procaryotes by primary structural characterization. *J Mol Evol* **1:** 173-184.

Solomon CM, Collier JL, Berg GM, Glibert PM (2010). Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat Microb Ecol* **59:** 67-88.

Spang A, Hatzenpichler R, Brochier-Armanet C, Rattei T, Tischler P, Spieck E *et al* (2010). Distinct gene set in two different lineages of ammonia-oxidizing archaea supports the phylum Thaumarchaeota. *Trends in Microbiology* **18:** 331-340.

Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW (2012). Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS ONE* **7:** e44002.

Strong AE (1974). Remote sensing of algal blooms by aircraft and satellite in Lake Erie and Utah Lake. *Remote Sensing of Environment* **3:** 99-107.

Stüken A, Rücker J, Endrulat T, Preussel K, Hemm M, Nixdorf B *et al* (2006). Distribution of three alien cyanobacterial species (*Nostocales*) in northeast Germany: *Cylindrospermopsis raciborskii*, *Anabaena bergii* and *Aphanizomenon aphanizomenoides*. *Phycologia* **45:** 696-703.

Stumpf RP, Wynne TT, Baker DB, Fahnenstiel GL (2012). Interannual variability of cyanobacterial blooms in Lake Erie. *PLoS ONE* **7:** e42444.

Sugita C, Ogata K, Shikata M, Jikuya H, Takano J, Furumichi M *et al* (2007). Complete nucleotide sequence of the freshwater unicellular cyanobacterium *Synechococcus elongatus* PCC 6301 chromosome: Gene content and organization. *Photosynthesis Research* **93:** 55-67.

Trewavas A (2006). A Brief History of Systems Biology: "Every object that biology studies is a system of systems." Francois Jacob (1974). *The Plant Cell Online* **18:** 2420-2430.

Turner CB (2010). Influence of zebra (*Dreissena polymorpha*) and quagga (*Dreissena rostriformis*) mussel invasions on benthic nutrient and oxygen dynamics. *Canadian Journal of Fisheries and Aquatic Sciences* **67:** 1899-1908.

Twiss MR, McKay RML, Bourbonniere RA, Bullerjahn GS, Carrick HJ, Smith REH *et al* (2012). Diatoms abound in ice-covered Lake Erie: An investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. *Journal of Great Lakes Research* **38:** 18-30.

USDA UNRCS (2005). Western Lake Erie Basin Water Resources Protection Plan.

van Gremberghe I, Leliaert F, Mergeay J, Vanormelingen P, Van der Gucht K, Debeer A-E *et al* (2011). Lack of phylogeographic structure in the freshwater cyanobacterium *Microcystis aeruginosa* suggests global dispersal. *PLoS ONE* **6:** e19561.

Vanderploeg HA, Liebig JR, Carmichael WW, Agy MA, Johengen TH, Fahnenstiel GL *et al* (2001). Zebra mussel (*Dreissena polymorpha*) selective filtration promoted toxic *Microcystis* blooms in Saginaw Bay (Lake Huron) and Lake Erie. *Canadian Journal of Fisheries and Aquatic Sciences* **58:** 1208-1221.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

Vollenweider R (1968). Scientific fundamentals of the eutrophication of lakes and flowing waters, with a particular reference to phosphorus and nitrogen as factor in eutrophication. *Organ Econ Coop Devel Tech Rep OAS/CSI/6827*.

Voultsiadou E, Vafidis D (2007). Marine invertebrate diversity in Aristotle's zoology. *Contributions to Zoology* **76:** 103-120.

Waples JT, Eadie, B., Klump, J.V., Squires, M., Cotner, J., McKinley, G. (2008). The Laurentian Great Lakes. In: Hales B, Cai, W., Mitchell, G., Sabine, C.L., Schofield, O. (ed). *North American Continental Margins: A Synthesis and Planning Workshop*. U.S. Carbbon Cycle Science Program. p 110.

Watanabe MF, Oishi S (1985). Effects of environmental factors on toxicity of a cyanobacterium (*Microcystis aeruginosa*) under culture conditions. *Appl Environ Microb* **49:** 1342-1344.

Watson SB, Ridal J, Boyer GL (2008). Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **65:** 1779-1796.

Weaver W (1970). Molecular biology: origin of the term. *Science* **170:** 581-582.

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

WHO (2011). Guidelines for drinking-water quality, fourth edition. In: Organization WH (ed).

Wicks RJ, Thiel PG (1990). Environmental factors affecting the production of peptide toxins in floating scums of the cyanobacterium *Microcystis aeruginosa* in a hypertrophic African reservoir. *Environmental science & technology* **24:** 1413-1418.

Wilhelm SW, DeBruyn JM, Gillor O, Twiss MR, Livingston K, Bourbonniere RA *et al* (2003). Effect of phosphorus amendments on present day plankton communities in pelagic Lake Erie. *Aquat Microb Ecol* **32:** 275-285.

Wilhelm SW, Bullerjahn GS, Eldridge ML, Rinta-Kanto JM, Poorvin L, Bourbonniere RA (2006). Seasonal hypoxia and the genetic diversity of prokaryote populations in the central basin hypolimnion of Lake Erie: Evidence for abundant cyanobacteria and photosynthesis. *Journal of Great Lakes Research* **32:** 657-671.

Woese CR, Fox GE (1977). Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences* **74:** 5088-5090.

Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J *et al* (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME Journal* **5:** 414-426.

Ye W, Liu X, Tan J, Li D, Yang H (2009). Diversity and dynamics of microcystin—Producing cyanobacteria in China's third largest lake, Lake Taihu. *Harmful Algae* **8:** 637-644.

Žegura B, Gajski G, Štraser A, Garaj-Vrhovac V, Filipič M (2011a). Microcystin-LR induced DNA damage in human peripheral blood lymphocytes. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **726:** 116-122.

Žegura B, Štraser A, Filipič M (2011b). Genotoxicity and potential carcinogenicity of cyanobacterial toxins – a review. *Mutation Research/Reviews in Mutation Research* **727:** 16-41.

CHAPTER II:
TAXONOMIC ASSESSMENT OF A TOXIC CYANOBACTERIA SHIFT IN HYPEREUTROPHIC GRAND LAKE ST. MARYS (OHIO, USA)

PUBLICATION NOTE

This chapter is a version of a peer-reviewed article previously published in *Harmful Algae* 33 (2014) 12-18 by Morgan M. Steffen, Zhi Zhu, Robert Michael L. McKay, Steven W. Wilhelm, and George S. Bullerjahn.

My contribution to this work was the processing of field samples for the molecular component of the paper, performing data analysis, and assisting with the writing of the manuscript. It is noted in the published manuscript that MMS and ZZ contributed equally to this work.

ABSTRACT

Grand Lake St. Marys (Ohio, USA) is a hypereutrophic reservoir prone to persistent toxic cyanobacterial blooms fuelled by agricultural nutrient runoff. The pervasive and highly toxic 2010 bloom led to the collapse of the local tourism industry, with microcystin concentrations exceeding 2,000 μg mL^{-1} at some locations during the peak of the bloom. Sampling from the Celina Water Treatment Plant, chlorophyll a levels remained above 25 μg L^{-1} from June – September, reaching a maximum in excess of 100 μg L^{-1} in July. Ratios of dissolved inorganic nitrogen (DIN) to dissolved inorganic phosphorus (DIP) declined below 16 in July and August, suggesting the cyanobacterial population became N-limited during the bloom peak. Cyanobacterial biomass was dominated by *Planktothrix* spp. throughout late summer 2010, but phylogenetic analysis of *mcyA* sequences revealed the presence of toxic *Microcystis* spp. during July that coincided with the highest toxin measurements. August and September samples indicated the community shifted to yield *Planktothrix* spp. as the major toxic genus. Community shifts of this nature may impede targeted remediation efforts, and therefore a thorough understanding of the taxa involved is necessary prior to implementing strategies to limit bloom formation.

INTRODUCTION

Grand Lake St. Marys (GLSM) is Ohio's largest inland lake, having a surface area of > 5,500 ha. Developed in the early 19th century as a feeder reservoir for the Miami and Erie Canal, the lake has since attained regional importance as a center for recreation and tourism, attracting up to $150 million in annual revenue in recent years (Drake and Davenport 2011). However, with a watershed dominated by agriculture (Hoorman et al. 2008), the lake is prone to eutrophication, one outcome of which has been recurrent and persistent blooms of toxic and noxious cyanobacteria that produce the hepatotoxin, microcystin. Currently, the dominant cyanobacterium in GLSM is the filamentous genus *Planktothrix* (U.S. EPA, 1975; U.S. EPA, 2009). In July 2010, a cyanobacterial bloom in GLSM resulted in the microcystin concentrations reaching >500 µg g dry weight algal biomass^{-1} (Steffen et al., 2012; Schmidt et al., 2013). During the same week, Ohio EPA measured microcystin levels > 2,000 µg L^{-1} at several beach sites (http://www.epa.ohio.gov/portals/35/inland_lakes/glsm_microcystin_data.pdf), a value orders of magnitude higher than drinking-water (1 µg L^{-1}) or recreational exposure (20 µg L^{-1}) guidelines of the World Health Organization (WHO 2004). Toxin concentrations of this magnitude not only affect tourism industries, but also may pose a serious threat to public health. Whereas cases of acute microcystin toxicity are rare (Jochimsen et al., 1998; Pouria et al., 1998), there is increasing concern that chronic- and low-level exposure can yield an increased risk for liver cancer as supported by epidemiological studies (Ueno *et al.* 1996), consistent with microcystin's role as a tumor promoter (Yoshizawa et al., 1990).

Results of many studies support a strong relationship between phosphorus (P)-loading and cyanobacterial blooms (Heisler *et al.* 2008, Schindler 1977). Sewage, animal waste, atmospheric deposition, and groundwater inflow, as well as agricultural and other fertilizer runoff can contribute to external P-loading (Robert 1983). Release of P from sediment, especially during periods of anoxia, represents an important source of internal loading (Dennis *et al.* 1993, Robert 1983). During summer, total phosphorus (TP) concentrations reach 200 µg L^{-1} and support accumulation of phytoplankton with chlorophyll (chl) *a* concentrations reaching 250 µg L^{-1} in GLSM (Ohio EPA 2010). These levels far exceed lake management criteria characterizing a eutrophic lake as having chl *a* in excess of 25µg L^{-1}. Additionally, both the concentration and chemical speciation of N has been identified as important in promoting

blooms (Xu et al., 2010; Scott et al., 2013), complicating the processes associated with lake restoration.

To address the problem of recurrent- and persistent cyanobacterial blooms in GLSM, the state of Ohio, along with grassroots citizens groups, implemented a multi-faceted response. Some approaches addressed the cause of blooms by attempting to reduce external nutrient loads to the lake. Others, such as linear aeration and the use of AIRY-GATOR™ technology, served to aerate the lake, dispersing surface blooms and reducing internal loading of P accumulated in the sediment (OLMS 2011). Likewise, aluminum sulfate (alum) can be applied to sequester P and thus render the cyanobacterial assemblage P-deficient, retarding further growth (Welch *et al.* 1988, Welch and Schrieve 1994). Alum treatment has previously yielded successful control of cyanobacterial blooms (Soltero et al., 1981; Cooke et al., 1982; Foy 1985; Dennis et al., 1993; Welch and Schrieve 1994), albeit in lakes far smaller in surface area than GLSM. The state of Ohio funded pilot alum treatments for GLSM in fall 2010 (Tetra Tech, 2010; Nogaro et al., 2013) and spring 2011 (Tetra Tech, written communication, Experimental treatments to reduce phosphorus and algae in nearshore coves of Grand Lake St. Marys. Technical Memo, 4 pp.), followed by larger-scale alum applications during summers 2011 and 2012 (Tetra Tech, 2011; Tetra Tech, 2013). The lakewide alum treatment in 2012 reduced the internal phosphorus loading by 55 percent (Tetra Tech, 2013), but the expense of the large-scale alum application required precluded treatment during summer 2013. Further, alum application increased concentrations of dissolved aluminum and altered nutrient cycling, the results of which may yield unintended and negative ecological consequences (Nogaro et al., 2013).

Given the variability of the results to date, an informed approach to prevent and remediate toxic blooms should begin with a survey of the organisms present in the lake during the bloom season and the identification of the specific organisms responsible for toxin production (*e.g.* Rinta-Kanto and Wilhelm, 2006). A common observation is that a bloom event comprises a succession of toxic and nontoxic genotypes (Rinta-Kanto et al., 2009), and our sampling efforts during the 2010 bloom reveal that while *Planktothrix* spp. were the dominant biomass throughout, toxic *Microcystis* spp. were coincident with peak summertime toxin measurements. The aim of our study was to determine the diversity of the microcystin producers during the peak bloom months. Understanding such community shifts in concert with physical

and chemical measures will be important in assessing what factors may favor one toxic genus over another, leading to more effective management and remediation strategies in the future.

MATERIALS AND METHODS

Study site

With a surface area of 5,500 hectares, Grand Lake St. Marys (40°31'52" N, 84°29'59" W; Fig.1) is Ohio's largest inland lake. Straddling the Auglaize-Mercer County line between the towns of St. Marys and Celina, the mean depth of the lake is approximately 1.5-2 m, with approximately 1.8 m of sediment having accumulated since the lake was created. The hydraulic residence time has been reported to be approximately 1.6 years (U.S. EPA, 1975).

Samples of raw water collected were collected 2-3 times daily from the intake pipe by personnel from the City of Celina Water Treatment Plant (40º32'27" N, 84º34'41" W; Fig. 1) and analyzed for temperature, pH, dissolved oxygen and alkalinity as part of their routine water quality analysis. The water intake was at a fixed depth of 1 m. Samples for chlorophyll, DNA and nutrients were collected from April – December in acid-washed polycarbonate bottles at a site immediately adjacent to the Celina Water Treatment Plant (CWTP) intake.

Physical and chemical measurements

To determine chl *a* biomass, samples were filtered in triplicate on 0.2 µm nominal pore-sized polycarbonate (PCTE) membranes (Whatman, Ann Arbor, MI). Filters containing seston were extracted in 90% acetone for 24 h at -20°C and measured by fluorometry on a TD-700 fluorometer (Turner Designs, Sunnyvale, CA) following the method of Welschmeyer (1994). For dissolved nutrients, water samples were passed through Sterivex cartridge filters (0.22 µm; Millipore, Billerica, MA) and the filtrate collected into acid-rinsed PE bottles, which were frozen at -20°C. Dissolved nutrients were analyzed at the National Center for Water Quality Research at Heidelberg University (Tiffin, OH). Celina Utilities Water Treatment Plant personnel measured total microcystin by the ELISA method (An and Carmichael, 1994) and data from 2010 are available from a public Ohio EPA database (http://www.epa.ohio.gov/portals/35/inland_lakes/glsm_microcystin_data.pdf).

Taxonomic assessment of phytoplankton

Water samples for phytoplankton abundance and taxonomic composition were fixed by addition of glutaraldehyde to 2% (v/v). Light microscopy was used to determine phytoplankton abundance and taxonomic composition following concentration of samples in a graduated cylinder. A Palmer-Maloney nanoplankton chamber was used to count phytoplankton. For each sample, at least 300 natural units (maximum of 100 random fields) were counted (Acker 2002). Counts were converted into biovolume following Hillebrand et al. (1999). Algal taxonomy conformed to that outlined by Prescott (1962). Taxonomic keys of Patrick and Reimer (1975), Dodd (1987), as well as Wehr and Sheath (2003) were used for identification.

DNA extraction and sequencing

Bloom biomass material (100 mL) was collected directly from the lake surface adjacent to the CWTP intake (Figure 2.1) at two time points (during the July and August sample collection events) and nucleic acids extracted directly from this material. Samples were stored at -20 ºC prior to processing for nucleic acids. Total genomic DNA was extracted using the MoBio PowerWater® DNA Isolation kit (MO BIO Laboratories, Carlsbad, CA). All PCR amplifications with *Microcystis* spp. *mcyA* primers (Hisbergues et al., 2003) and sequence analyses were performed as described previously (Rinta-Kanto and Wilhelm, 2006). In brief, clone libraries were generated using the TOPO®-TA cloning kit as described by the supplier (Life Technologies, Grand Island, NY) and individual clones picked and processed for sequencing. Sequencing of amplicons was performed by the Sanger method at the University of Tennessee Molecular Biology Resource Facility. Sequences were aligned using ClustalW within Mega 5.0 using default parameters (Peterson et al., 2010). Phylogenetic trees were generated using the Neighbor-Joining method. A bootstrap test of phylogeny was performed using 5000 replications. The p-distance substitution model was chosen using amino acids as the substitution type. DNA sequences obtained in this study were deposited in GenBank under accession numbers KF871117-KF871144.

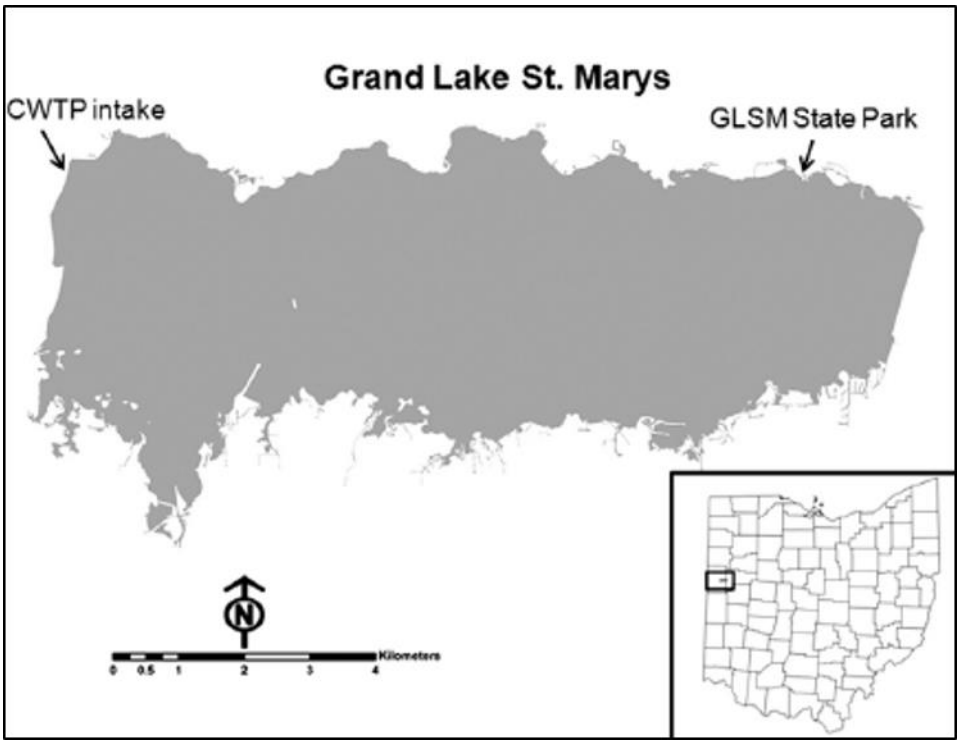

Figure 2.1. Location of the Celina water intake study site (Celina). Credit: Google Earth.

RESULTS

Physicochemical properties of Grand Lake St. Marys: 2010

The location of the CWTP water intake (Figure 2.1) served as the field site for our taxonomic analysis of GLSM toxic cyanobacteria. Typical of shallow, north temperate lakes, GLSM displays seasonality in many of its physicochemical parameters (Figure 2.2). Ice cover typically occurs on the lake during winter and is consistent with the stable low temperatures measured from December through late February. Following ice out, temperature increased through mid-late July when it attained a maximum of *ca.* 30°C, thereafter declining through late November. Alkalinity in 2010 was highest in winter, decreasing in March and reaching its lowest values in late May-June (Figure 2.2B). The pH of the lake in 2010 tracked alkalinity in an inverse manner beginning neutral in winter with pH rising to >9 in March coinciding with a large decline in alkalinity (Figure 2.2C). The pH declined to *ca.* 7.2 in late winter, increased to 9 in early spring and through the summer and fall, followed by a gradual decline to <8 by December. Dissolved oxygen (DO) in GLSM was variable, but yielded a slight downward trend through the summer, when DO dropped below 5 mg L^{-1} (Figure 2.2D). Whereas technical difficulties precluded oxygen from being measured through late spring 2010, DO was variable, reaching minima below 4 mg L^{-1} before recovering to near saturation at the end of the year.

Nutrients (N and P) concentrations were lower in the summer than in spring (Figure 2.3). Dissolved nitrogen (NH_3 and NO_3^-) became depleted by early summer (Figure 2.3A, B). By contrast, soluble reactive phosphorus (SRP) varied between 0.05 – 0.2 mg L^{-1} during the spring and summer, before declining in the fall (Figure 2.3C). Dissolved silica concentrations were lower in spring than in summer with values peaking at >5 mg L^{-1} each year (Figure 3D). Consistent with the seasonal depletion of nitrogen, the molar ratio of dissolved inorganic nitrogen (DIN): dissolved inorganic phosphorus (DIP) declined such that by early June, it was reduced to <25 and declined to *ca.* 1 from July through September (Figure 3E). These seasonal data suggest the populations were P-limited through June and after October, but N-limited during the time of the study.

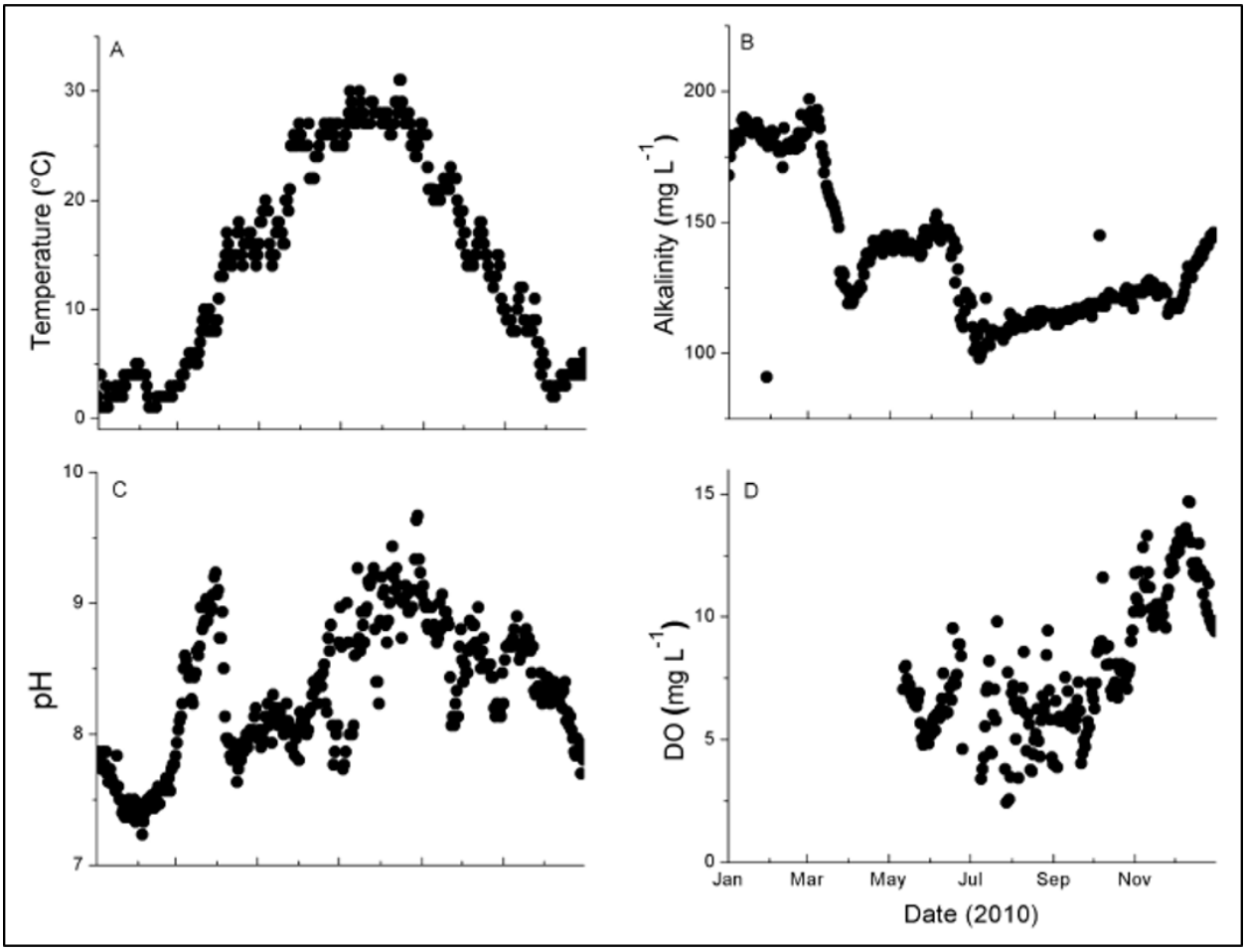

Figure 2.2. Physicochemical parameters measured from the Celina water intake in 2010. A) temperature, B) alkalinity, C) pH, and D) dissolved oxygen (DO).

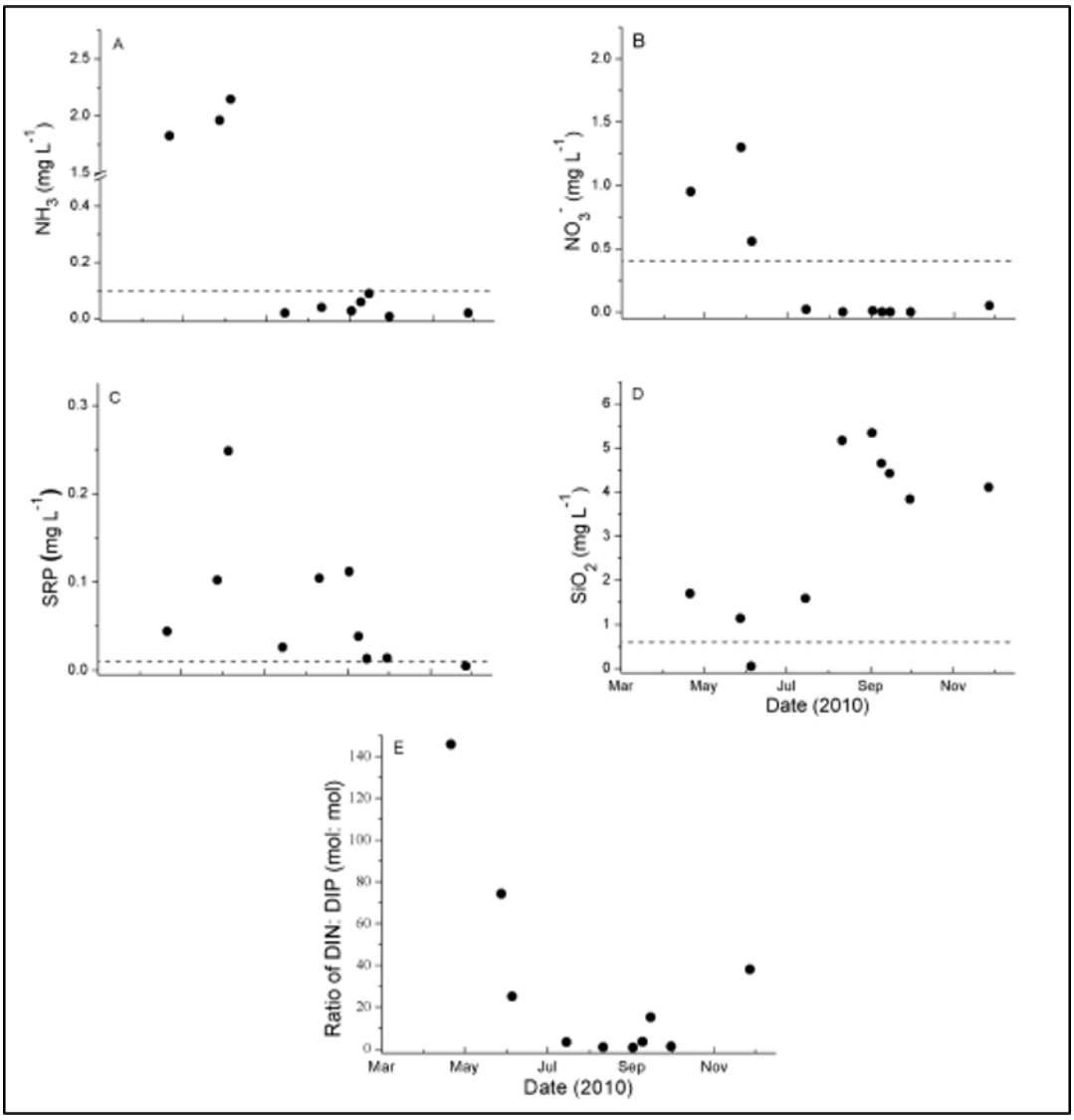

Figure 2.3. Nutrients measured at the Celina water intake channel in 2010. A) NH_3, B) NO_3^-, C) SRP and TP, D) SiO_2, E) DIN: DIP (mol: mol). Dotted line in panels A-D shows the Method Reporting Limit representing the lowest concentration of quantitative certainty for each analyte.

This maximum level roughly coincides with the July 12 maximum exceeding 2,000 µg L^{-1} at 3 beaches located at the north-east end of the lake (Grand Lake St. Marys State Park, Figure 2.1). On August 8, the day prior to a second sampling for DNA, the toxin level had declined to 2.5 µg L^{-1} (Figure 2.4, also see Table 2.1).

Phytoplankton of Grand Lake St. Marys during 2010

Seasonal trends in phytoplankton chl *a* biomass reflected the hypereutrophic status of GLSM during summer 2010, with chl *a* exceeding 25 µg L^{-1} from early June through October (Figure 2.5A). These chl values corresponded a Secchi depth of less than 35 cm throughout the summer (Tetra Tech 2011). At other times of the year, the lake ranged from meso- to eutrophic based on chl *a* biomass. Samples for phytoplankton biovolume analysis were collected at near weekly intervals from late August through September 2010. Cyanobacteria, predominantly *Planktothrix* spp., dominated the lake's phytoplankton with a biovolume in excess of 1.2×10^{11} μm^3 L^{-1}, which was one order of magnitude higher than the biovolume of other algae (Figure 2.5B). Among known microcystin producers, *Microcystis* spp. were present as a minor component of the late summer algal community, representing *ca.* 10% of the cyanobacterial biovolume (Figure 2.5C).

Phylogenetic analysis of *mcyA* sequences

PCR amplification of the *mcyA* gene of the microcystin synthetase operon was performed on DNA extracted from the algal community sampled on July 16 and August 9, 2010. The accompanying physical and chemical characteristics of the Celina site on these dates are provided in Table 1. Phylogenetic analysis of *mcyA* sequences revealed that both *Microcystis* spp. and *Planktothrix* spp. are capable of microcystin production in the lake, but the two samples are distinctly different with respect to the toxic genera present. The July sample contains largely *Microcystis* spp. *mcyA* sequences (10 of 12 clones), whereas a community shift has occurred by

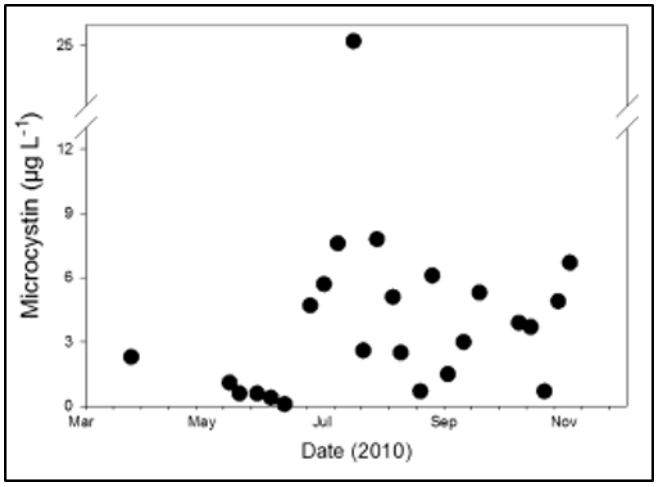

Figure 2.4. Microcystin concentrations in Grand Lake St. Marys (OH) during 2010. Toxin concentrations were determined by the Celina Utilities Water Treatment Plant for samples collected at the public water supply intake using the ELISA approach. Data provided by the Ohio EPA (http://www.epa.ohio.gov/portals/35/inland_lakes/glsm_microcystin_data.pdf). Samples with values below the detection limit (0.14 mg L^{-1}) of the microcystin ELISA assay were not included in the plot.

Table 2.1. Physical and chemical characteristics of 2010 samples obtained for phylogenetic analysis. Nutrient and chlorophyll measurements are in mgL^{-1}. Toxin levels were measured the day prior to sampling (July 15 and August 8).

Date	**Temperature (°C)**	**DO**	**Chl *a***	**NH_3**	**NO_3-**	**SiO_2**	**SRP**	**Toxin (ppb)**
July 16	27	4.5	108.7	0.019	0.019	1.59	0.0026	25.2
August 9	27	7.1	47.9	0.039	0.039	5.17	0.104	2.5

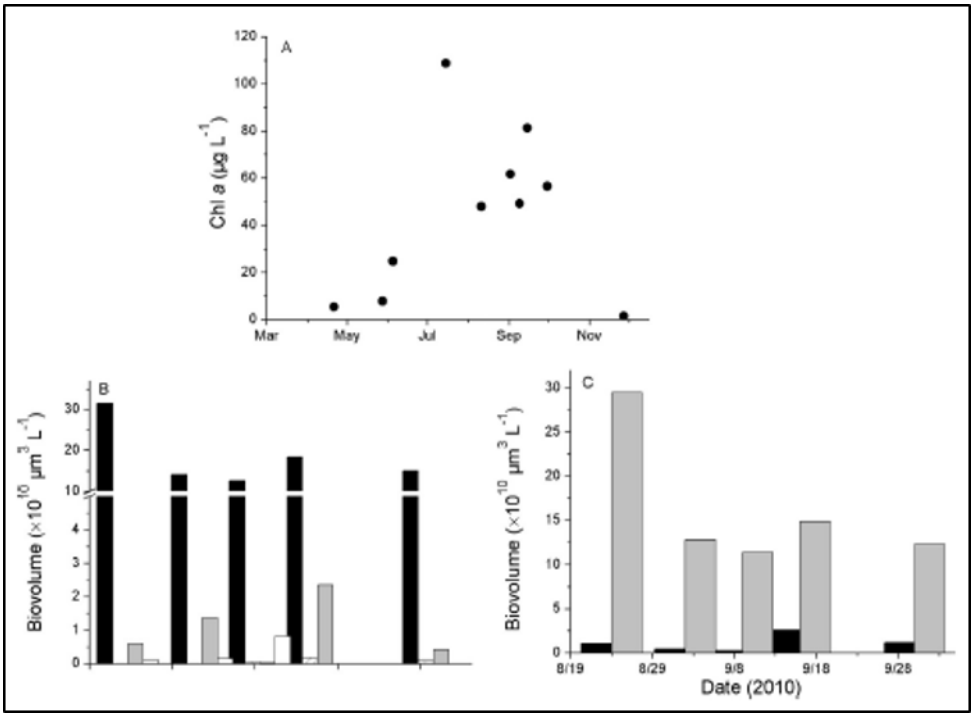

Figure 2.5. A) Phytoplankton chl *a* biomass determined from samples collected at the Celina water intake channel in 2010. B) Phytoplankton biovolume determined from samples collected at Celina. Cyanobacteria (black), green algae (striped), diatoms (grey), and cryptomonads (white). C) Cyanobacterial biovolume from same samples. *Planktothrix* spp. (grey) and *Microcystis* spp. (black).

August to yield primarily *Planktothrix* spp. sequences at the later date (16 of 16 clones, see Figure 2.6). The *Planktothrix* spp. sequences are closely related to *Planktothrix* spp. clones obtained previously from nearby sites in Ohio, including Sandusky Bay (Rinta-Kanto and Wilhelm, 2006) and the Maumee River (Kutovaya et al., 2012).

DISCUSSION

During the 2010 bloom in Grand Lake St. Marys, levels of microcystin exceeded by several orders of magnitude the World Health Organization (WHO) guidelines of 1 μg L^{-1} for drinking water and 20 μg L^{-1} for recreational waters (OLMS 2011, WHO 1998). In addition, microcystin was documented to accumulate in fish tissue in GLSM, further compromising the recreational value of the lake (Schmidt et al., 2013). Indeed, if consumption of fish from the lake is considered a route of toxin exposure, then food web structure (dictated by the presence or absence of different plankton species) may be critical in determining the risk of exposure of humans as well as birds and other animals that consume fish.

The dominant organism in GLSM, *Planktothrix* spp. (U.S. EPA, 2009), is a cyanobacterium that prefers higher temperature and water column stability compared to diatoms (Dokulil and Teubner 2000, Paerl 2008, Paerl 1996, Zhang and Prepas 1996). GLSM is a shallow and hypereutrophic lake; the long water residence time and high surface water temperatures (> 20°C) could account for the dominance of *Planktothrix* spp. This result is in accordance with Schreurs (1992) in that filamentous species dominate shallow lakes having either total phosphorus ranging from 100 to 800 μg L^{-1} or total nitrogen of 2.5-3.5 mg L^{-1}. In this case it is likely the low TN/TP ratios, high P and long water residence time are beneficial to the cyanobacterial community.

Despite their numerical dominance, the effects of the bloom event may not be due to the persistent abundance of *Planktothrix* spp. Given the finding that the genetic capability for toxin production in GLSM may at times be contributed by *Microcystis* spp. as well as *Planktothrix* spp., examination of cell counts from late summer 2010 (Fig. 2.5b) and 2011 (Zhu 2012) revealed the presence of a population of *Microcystis* spp. routinely comprising *ca.* 10% of the total cyanobacterial biovolume. Considering that the elevated July 2010 toxin levels are nearly coincident with the detection of *Microcystis* spp. *mcyA* sequences, our observations raise the

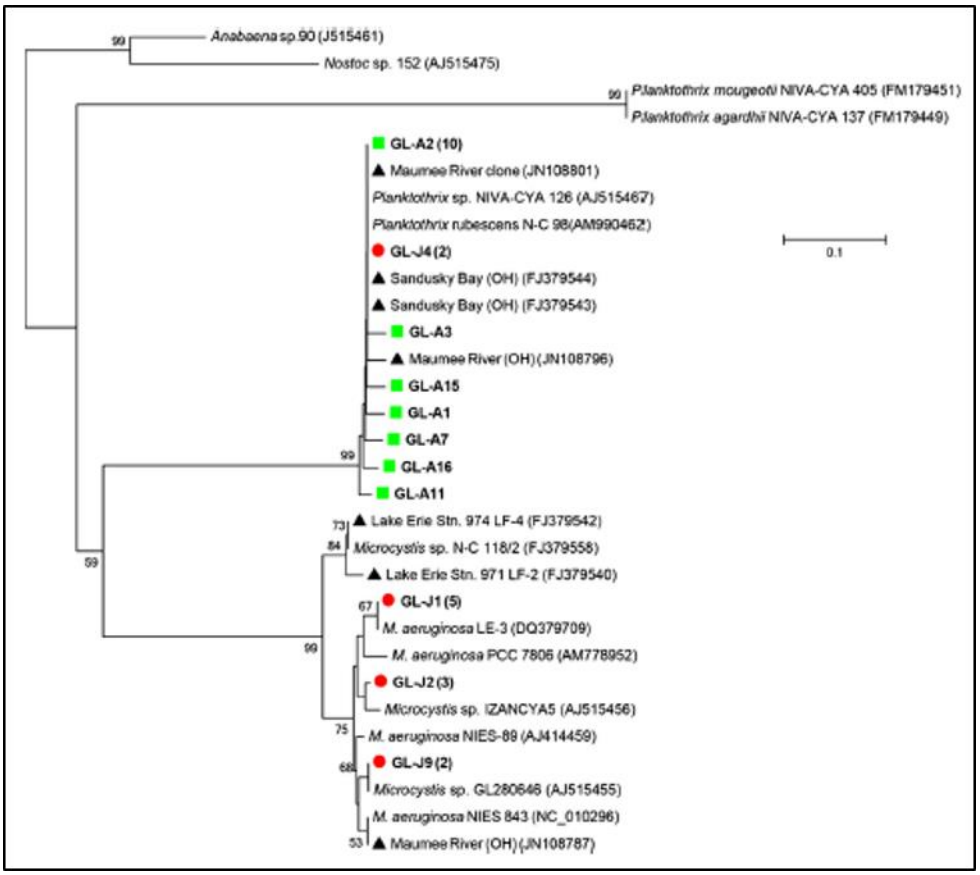

Figure 2.6. Phylogenetic analysis of *mcyA* sequences from the Celina site taken on July 16 (Julian Day 197) and August 9 (Julian Day 221). Red circles represent clones from July, green squares are clones from August, and black triangles are reference environmental clones from nearby sites in OH (Maumee River, Sandusky Bay and Lake Erie).

possibility that the *Microcystis* spp. population may have been a major producer of toxin in GLSM.

Recurrent cyanobacterial blooms like those seen in recent years likely result from the unique physical and chemical characteristics of GLSM. The temperature gradually increased to *ca.* 30℃ in the summer coincident with depletion of N (Figs 2.2A and 2.3A, B) along with the decrease of DO (Fig 2.2D) which would promote the regeneration of soluble P. Since cyanobacteria prefer high temperature and some taxa are capable of diazotrophic growth, it is reasonable to believe cyanobacteria may out-compete other algae in the lake under these conditions (Berman-Frank *et al.* 2003, Dokulil and Teubner 2000, Paerl 2008, Paerl 1996). As cyanobacteria have higher optimum growth temperatures than eukaryotic algae, it is also expected that climate warming will promote the growth of cyanobacteria over eukaryotes in the long term (Elliott *et al.* 2006, Kosten *et al.* 2011, Wagner and Adrian 2009). Furthermore, global warming increases the stratification period of lakes, so buoyant cyanobacteria can occupy the surface of the water, forming blooms that shade benthic and deeper planktonic algae, suppressing their growth (Joehnk *et al.* 2008). Considering the advantages of cyanobacteria over eukaryotic algae, the frequency and severity of cyanobacterial blooms may be enhanced in the future (O'Neil *et al.* 2012, Paerl and Huisman 2008).

The ability to manage lake systems as effective stewards is in part dependent on an understanding of the (un)desirable communities within the system. In GLSM, it appears that for at least part of 2010 there may have been contrasting values of two populations (toxin production by *Microcystis vs* hypereutrophic accumulation of *Planktothrix*). Understanding the contribution of each of these populations to the overall health of GLSM will be critical to the future management of this lake, as efforts to constrain one population may lead to the proliferation of the other. Future work examining taxonomic shifts of cyanobacteria in experimentally manipulated mesocosms will help reveal chemical (N,P) and physical factors that may favor one toxic genus over another, and/or increased toxin production by the endemic cyanobacterial assemblage. These studies will assist in the development of remediation strategies for individual bloom forming cyanobacteria in GLSM.

ACKNOWLEDGEMENTS

The authors thank T. Mike Sudman and Todd Hone (Celina Utilities Water Treatment Plant) for assistance with water sampling and access to their water quality data. Dr. Benjamin Beall and Ms. Katrina Thomas assisted in processing of samples for nutrient and phylogenetic analyses. Dr. Rex Lowe provided helpful advice on identification of cyanobacteria. We thank Kurt Thompson and Dr. Geraldine Nogaro (Annis Water Resources Institute) for providing the map template used in Fig. 1. This project was supported by funds provided by the Bowling Green State University Division of Research and Economic Development, NSF awards IOS-0841918 and DEB-1240870 to SWW and an award to GSB by the Lake Erie Protection Fund (LEPF 09-357).

LIST OF REFERENCES

Acker, F., 2002. Analysis of Soft Algae and Enumeration of Total Number of Diatoms in USGS NAWQA Program Quantitative Targeted-Habitat (RTH and DTH) Samples. *Protocols for the analysis of algal samples collected as part of the U.S. Geological Survey National Water-Quality Assessment Program.* D. F. Charles, C. Knowles and R. S. Davis. Philadelphia, PA., The Academy of Natural Sciences**,** pp. 79-85.

An, J., Carrmichael, W.W., 1994. Use of a colorimetric protein phosphatase inhibition assay and enzyme linked immunosorbent assay for the study of microcystins and nodularins. *Toxicon* **32:** 1495-1507.

Berman-Frank, I., Lundgren, P. Falkowski, P., 2003. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Research in Microbiology* **154:** 157-164.

Cooke, G. D., Heath, R.T., Kennedy, R.H., McComos, M.R., 1982. Change in lake trophic state and internal phosphorus release after aluminum sulfate application. *Journal of the American Water Resources Association* **18:** 699-705.

Davenport, T., Drake, W., 2011. Grand Lake St. Marys, Ohio - The case for source water protection: Nutrients and algae blooms. LakeLine Magazine **31(3):** 41-46.

Dennis, C. G., Eugene, W.B., Spencer, P.A., Peter, N.R., 1993. *Restoration and Management of Lakes and Reservoirs*. Boca Raton, Lewis Publishers.

Dodd, J. J., 1987. *Diatoms*. Illinois, Southern Illinois University Press.

Dokulil, M. T., Teubner, K., 2000. Cyanobacterial dominance in lakes. *Hydrobiologia* **438:** 1-12.

Elliott, J., Jones, I., Thackeray, S., 2006. Testing the sensitivity of phytoplankton communities to changes in water temperature and nutrient load, in a temperate lake. *Hydrobiologia* **559:** 401-411.

Filbrun, J.E., Conroy, J.D., Culver, J.A., 2013. Understanding seasonal phosphorus dynamics to guide effective management of shallow, hypereutrophic Grand Lake St. Marys, Ohio. *Lake Reservior Manag.* **29:** 165-178.

Foy, R. H., 1985. Phosphorus inactivation in a eutrophic lake by the direct addition of ferric aluminium sulphate: impact on iron and phosphorus. *Freshwater Biology* **15:** 613-629.

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

Hillebrand, H., Dürselen, C.-D., Kirschtel, D., Pollingher, U., Zohary, T., 1999. Biovolume calculation for pelagic and benthic microalgae. *Journal of Phycology* **35:** 403-424.

Hisbergues, M., Christiansen, G., Rouhainen, L., Sivonen, K., Börner, T., 2003. PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. Archives of Microbiology **180:** 402-410.

Hoorman, J., Hone, T., Sudman, T. Jr., Dirksen, T., Iles, J., Islam, K.R., 2008. Agricultural impacts on lake and stream water quality in Grand Lake St. Marys, Western Ohio. Water, *Air, & Soil Pollution* **193:** 309-322.

Jochimsen, E.M., Carmichael, W.W., An, J., Cardo, D.M., Cookson, S.T., Holmes, C.E.M., Antunes, M.B., de Melo Filho, D.A., Lyra, T.M., Barreto, V.S.T., Azevedo, S.M.F.O, Jarvis, W.R., 1998. Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *The New England Journal of Medicine* **338:** 873-878.

Joehnk, K. D., Huisman, J., Sharples, J., Sommeijer, B., Visser, P.M., Stroom, J.M., 2008. Summer heatwaves promote blooms of harmful cyanobacteria. *Global Change Biology* **14:** 495-512.

Kosten, S., Huszar, V.L.M., Bécares, E., Costa, L.S., Donk, E., Hansson, L.A., Jeppesen, E., Kruk, C., Lacerot, G., Mazzeo, N., De Meester, L., Moss, B., Lürling, M., Nöges T., Romo, S., Scheffer, M., 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. *Global Change Biology* **18:** 118-126.

Kutovaya, O.A., McKay, R.M.L., Beall, B.F.N., Wilhelm, S.W., Kane, D.D., Chaffin, J.D., Bridgeman, T.B., Bullerjahn, G.S., 2012. Evidence against fluvial seeding of recurrent toxic blooms of *Microcystis* spp. in Lake Erie's western basin. *Harmful Algae* **15:** 71-77.

Nogaro, G., Burgin, A.J., Schoepfer, V.A., Konkler, M.J., Bowman, K.L., Hammerschmidt, C.R., 2013. Aluminum sulfate (alum) application interactions with coupled metal and nutrient cycling in a hypereutrophic lake ecosystem. *Environmental Pollution* **176:** 267-274.

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

Ohio Lake Management Society, 2011. Management techniques at Grand Lake St. Marys. Ohio *Shorelines* **24:** 2-11.

Paerl, H., 2008. Nutrient and other environmental controls of harmful cyanobacterial blooms along the freshwater–marine continuum. In: *Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs. Advances in Experimental Medicine and Biology* **619:** 217-237.

Paerl, H. W., 1996. A comparison of cyanobacterial bloom dynamics in freshwater, estuarine and marine environments. *Phycologia* **35(6S):** 25-35.

Paerl, H. W., Huisman, J., 2008. Blooms like it hot. *Science* **320:** 57-58.

Patrick, P., Reimer, C.R., 1975. *The Diatoms of the United States*. Philadelphia, PA, Academy of Natural Sciences

Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:** 2731-2739.

Pouria, P., de Andrade, A., Barbosa, J., Cavalcanti, R.L., Barreto, V.T.S., Ward, C.J., Preiser, W., Poon, G.K., Neild, G.H., Codd, G.A., 1998. Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. *The Lancet* **352:** 21-26.

Prescott, G. W., 1962. *Algae of the Western Great Lakes area.* Wm. C. Brown Publishers, Dubuque, IA (USA).

Rinta-Kanto, J.M., Wilhelm, S.W., 2006. Diversity of microcystin-producing cyanobacteria in spatially isolated regions of Lake Erie. *Applied and Environmental Microbiology* **72:** 5083–5085.

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

Schindler, D. W., 1977. Evolution of phosphorus limitation in lakes. *Science* **195:** 260-262.

Schmidt, J. R., Shaskus, M., Estenik, J.F., Oesch, C., Khidekel, R., Boyer, G.L. 2013. Variations in the microcystin content of different fish species collected from a eutrophic lake. *Toxins* **5:** 992-1009.

Schreurs, H., 1992. *Cyanobacterial Dominance-Relations to Eutrophication and Lake Morphology.* Academisch Profschrift, Universiteit van Amsterdam.

Scott, J. T., McCarthy, M.J., Otten, T.G., Steffen, M.M., Baker, B.C., Grantz, E.M., Wilhelm, S.W., Paerl, H.W., 2013. Comment: an alternative interpretation of the relationship between TN:TP and microcystins in Canadian lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **70(8):** 1265-1268

Soltero, R. A., Nichols, D.G., Gasperino, A.F., Beckwith, M.A., 1981. Lake restoration: Medical Lake, Washington. *Journal of Freshwater Ecology* **1:** 155-165.

Steffen, M.M., Li, Z., Effler, T.C., Hauser, L.J., Boyer, G.L., Wilhelm, S.W., 2012. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLOS One* **7(8),** e44002.

Tetra Tech. 2010. Effectiveness of alum treatment demonstration projects in Grand Lake St. Marys. Technical Memo, 6 pp. December 2010. http://www.westernlakeerie.org/grand_lake_treatment_report_12_2010.pdf

Tetra Tech. 2011. 2011 Alum treatment and associated water quality monitoring for Grand Lake St. Marys. Preliminary draft report. 37 pp. September 2011. http://www.epa.ohio.gov/portals/35/hab/glsmalum2011.pdf

Tetra Tech, 2013. Preliminary assessment of effectiveness of the 2012 alum application - Grand Lake St. Marys. Columbus, Ohio, 85 pp. February 2013. http://www.lakeimprovement.com/sites/default/files/GLSM%20Alum%20Report%2002202013.pdf

Ueno, Y., Nagata, S., Tsutsumi, T., Hasegawa, A., Watanabe, M.F., Park, H.-D., Chen, G.-C., Chen, G., Yu, S.-Z. 1996. Detection of microcystins, a blue-green algal hepatotoxin, in drinking water sampled in Haimen and Fusui, endemic areas of primary liver cancer in China, by highly sensitive immunoassay. *Carcinogenesis* **17:** 1317-1321.

U.S. Environmental Protection Agency, 1975. National Eutrophication Working Paper Series Report on Grand Lake St. Marys, Auglaize and Mercer Counties, OH. EPA Region V Working Paper, No. 411. Pacific Northwest Environmental Research Laboratory.

U.S. Environmental Protection Agency, 2009. National Lakes Assessment: a collaborative survey of the nation's lakes. Office of Water and Office of Research and Development Report 841-R-09-001.

Wagner, C., Adrian R., 2009. Cyanobacteria dominance: Quantifying the effects of climate change. *Limnology and Oceanography* **54:** 2460-2468.

Wehr, J. D., Sheath, R.G., 2003. *Freshwater algae of North America: Ecology and Classification.* Academic Press, San Diego, CA.

Welch, E. B., DeGasperi, C.L., Spyridakis, D.E., Belnick, T.J., 1988. Internal phosphorus loading and alum effectiveness in shallow lakes. *Lake and Reservoir Management* **4:** 27-33.

Welch, E. B., Schrieve, G.D., 1994. Alum treatment effectiveness and longevity in shallow lakes. *Hydrobiologia* **275:** 423-431.

Welschmeyer, N. A., 1994. Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll *b* and pheopigments. *Limnology and Oceanography* **39:** 1985–1992.

Wetzel, R.G., 1983. *Limnology*. Harcourt Brace College Publishers., Fort Worth, TX (USA).

World Health Organization, 1998. Cyanobacterial toxins: Microcystin-LR in drinking-water. Geneva.

World Health Organization, 2004. Guidelines for drinking-water quality. Geneva.

Xu, H., Paerl, H.W., Qin, B., Zhu, G.H., Gao, H., 2010. Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnology and Oceanography* **55:** 420-432.

Yoshizawa, S., Matsushima, R., Watanabe, M.F., Harada, K., Ichihara, A., Carmichael, W.W., Fujiki, H., 1990. Inhibition of protein phosphatases by microcystis and nodularin associated with hepatotoxicity. *Journal of Cancer Research and Clinical Oncology* **116:** 609-614.

Zhang, Y., Prepas, E.E., 1996. Regulation of the dominance of planktonic diatoms and cyanobacteria in four eutrophic hardwater lakes by nutrients, water column stability, and temperature. *Canadian Journal of Fisheries and Aquatic Sciences* **53:** 621-633.

Zhu, Z. 2012. Monitoring the phytoplankton community response to recent geoengineering initiatives at Grand Lake St. Marys. (M.S. thesis, Bowling Green State University, 74 pp., etd.ohiolink.edu).

CHAPTER III:
COMPARATIVE METAGENOMICS OF TOXIC FRESHWATER CYANOBACTERIA BLOOM COMMUNITIES ON TWO CONTINENTS

PUBLICATION NOTE

This chapter is a version of a peer-reviewed article previously published in *PLoS ONE* 7(8): e44002 by Morgan M. Steffen, Zhou Li, Loren J. Hauser, Gregory L. Boyer, and Steven W. Wilhelm.

My contribution to this work was the preparation of samples for sequencing, data analysis, and writing most of the manuscript.

ABSTRACT

Toxic cyanobacterial blooms have persisted in freshwater systems around the world for centuries and appear to be globally increasing in frequency and severity. Toxins produced by bloom-associated cyanobacteria can have drastic impacts on the ecosystem and surrounding communities, and bloom biomass can disrupt aquatic food webs and act as a driver for hypoxia. Little is currently known regarding the genomic content of the *Microcystis* strains that form blooms or the companion heterotrophic community associated with bloom events. To address these issues, we examined the bloom-associated microbial communities in single samples from Lake Erie (North America), Lake Tai (Taihu, China), and Grand Lakes St. Marys (OH, USA) using comparative metagenomics. Together the *Cyanobacteria* and *Proteobacteria* comprised > 90% of each bloom bacterial community sample, although the dominant phylum varied between systems. Relative to the existing *Microcystis aeruginosa* NIES 843 genome, sequences from Lake Erie and Taihu revealed a number of metagenomic islands that were absent in the environmental samples. Moreover, despite variation in the phylogenetic assignments of bloom-associated organisms, the functional potential of bloom members remained relatively constant between systems. This pattern was particularly noticeable in the genomic contribution of nitrogen assimilation genes. In Taihu, the genetic elements associated with the assimilation and metabolism of nitrogen were predominantly associated with *Proteobacteria*, while these functions in the North American lakes were primarily contributed to by the *Cyanobacteria*. Our observations build on an emerging body of metagenomic surveys describing the functional potential of microbial communities as more highly conserved than that of their phylogenetic makeup within natural systems.

INTRODUCTION

The theory that microbial community structure dictates the function of that community has recently been called into question (Burke et al. 2011, Frossard et al. 2012). Previously, laboratory studies using enzyme assays and 16S rRNA gene phylogeny provided evidence for the importance of phylogenetic identity to community function (Kandeler et al. 2000, Lehman and O'Connell 2002). A shift to whole community shotgun metagenomics has, however, allowed for a more comprehensive examination of microbial functional genes present in a wealth of natural environments. This technique circumvents the need for culture-based analysis and is more representative of natural community structure and functional potential. The trend emerging from this work indicates that the function of microbes within the environment is often more important than their phylogenetic/taxonomic identity (Burke et al 2011, Oh et al. 2011).

One environment of particular concern is the freshwater systems that have in recent years been increasingly inundated by toxic cyanobacterial blooms. These blooms have been responsible for the deterioration of freshwater systems with increasing frequency and intensity and are commonly dominated by colonial cyanobacteria of the genus *Microcystis* (Carmichael et al. 2000, Chen et al. 2003, Yang et al. 2011). Toxins produced by bloom-associated cyanobacteria have been shown to adversely affect health of animals and humans. The microcystins, cyclic secondary metabolites produced by members of the genera *Anabaena, Microcystis*, and Planktothrix, are hepatotoxins that have been associated with human liver and colorectal cancers (Azevedo et al. 2002, Hernandez et al. 2009, Mankiewicz-Boczek et al. 2011). Other secondary metabolites produced by *Microcystis* have been linked to phytoestrogenic effects in fish (Rogers et al. 2011). *Microcystis aeruginosa* is commonly dominant within freshwater blooms and has been shown to impact local ecology by disruption of the food web and induction of hypoxia in large lake systems (Charlton et al. 1993, Conroy et al. 2005).

Microcystis-dominated blooms have been observed in the western basin of Lake Erie annually since the 1990s. This freshwater system is a commonly studied model of toxic bloom development and persistence (Carmichael et al 2000, Rinta-Kanto et al. 2005). Lake Tai (hereafter referred to by its Chinese name Taihu, which translates to grand or great (tai) lake (hu)) is a shallow eutrophic lake that has experienced annual cyanobacterial blooms for the last three decades. The Taihu watershed supports upwards of 40 million people and receives inputs

from both industrial and agricultural sources. During the last decade Taihu has experienced *Microcystis* bloom events on an unprecedented scale: in 2007 authorities shut down water intakes, resulting in a freshwater shortage for the 4 million residents of the city of Wuxi (Stone 2011). In contrast, Grand Lakes St. Marys (GLSM) is a smaller (54.6 km^2) inland lake in Ohio that has recently experienced recurring highly toxic cyanobacteria blooms during summer months. While GLSM also experiences high concentrations of microcystin during bloom events, this lake is dominated by the filamentous cyanobacterium *Planktothrix* spp. (RML McKay, personal communication). To this end, GLSM provides a novel and contrasting study site (relative to Taihu and Lake Erie) whereby each lake experiences eutrophic conditions leading to the presence of the toxin microcystin, but in GLSM that toxin is produced by a cyanobacterium with a different physiological ecology. Examination of these contrasting freshwater cyanobacterial blooms models may provide insight on the environmental and genetic factors that influence bloom formation worldwide.

Despite our ability to examine *Microcystis* within bloom communities, questions remain regarding the environmental triggers that facilitate bloom formation and persistence. Previous research has shown that numerous abiotic factors, including nutrient input and air temperature, are important during cyanobacterial bloom formation, (Joung et al. 2011, Rinta-Kanto et al. 2009, Wilhelm et al. 2011a) although there is a current schism regarding which nutrient inputs most exacerbate bloom formation (Schindler et al. 2008, Scott and McCarthy 2010). Recent work has demonstrated that concentrations of both nitrogen and phosphorus influence success of freshwater cyanobacteria in eutrophic environments (Finlay et al. 2010, Gobler et al. 2010, Paerl et al. 2011b, Wilhelm et al 2011a). Moreover, freshwater cyanobacteria thrive at temperatures above 25°C, and increasing mean temperatures associated with recent global climate change have been associated with the increasing frequency of these toxic bloom events (Paerl and Huisman 2009, Paerl et al. 2011a, Ye et al. 2011).

Currently, we have little information regarding bloom communities as a whole, in terms of either phylogenetic composition or functional role. Indeed recent work suggests companion heterotrophs may be essential for cyanobacteria biomass growth in response to the addition of certain nutrient sources (Saxton et al. 2011), highlighting the importance of resolving the ecology of this complex community. To gain insight into the role of bloom microbial community

members and better understand differences between geographically distinct freshwater microbial communities, we have applied comparative metagenomics to the microbial populations of three toxic freshwater blooms. This approach has enabled us to examine not only the toxin-producing organisms, but also the companion heterotrophic community and their potential role in driving these blooms. In addition, we have used environmental genomic data to identify potential differences in genomic content in bloom-associated *Microcystis* spp. and a model lab isolate (*Microcystis aeruginosa* NIES 843) to further the aim of establishing the core genome of *Microcystis.* This information provides new insight on factors that may be critical to toxic freshwater blooms and which may prove crucial to future mitigation efforts.

MATERIALS AND METHODS

Sample collection

Lake Erie was sampled during an August, 2009 cruise aboard the CCGS Limnos. Water was taken from the surface in a 20 µm net tow at Environment Canada station 589 (Erie Harbor, PA; 42° 08' N 80° 07' W). Biomass was stored at -80° C until extraction. Taihu was sampled during a large surface bloom in May, 2009 near the Taihu Laboratory for Lake Ecosystem Research (TLLR, 31°27' N, 120°12' E) using 20 µm Nytex™. Samples were stored at -20° C until extraction. The GLSM (40°31' N, 84°24' W) bloom sample was collected using a 20 µm plankton net during the July 2010 bloom event and stored at -20° C until extraction. The 20 µm size is a commonly used technique in limnology and allowed for collection of *Microcystis*, other large cyanobacteria and the microbial community associated with these cells. Nutrient and toxin measurements at each location were made as previously described (Boyer 2008, Rinta-Kanto et al 2009, Wilhelm et al 2011a). Metadata for each site are listed in Table 3.1.

DNA extraction, sequencing, assembly, and annotation

Genomic DNA was extracted from all samples using the MoBio PowerWater® DNA Isolation Kit (MoBio Laboratories, Inc., CA, USA). The GLSM sample required an additional wash step, using sterile CT medium (Ichimura 1971, Watanabe 2000) to remove organic impurities that interfered with nucleic acid purity. The quality and quantity of DNA was

Table 3.1. Bloom sample metadata.

	Collection Date	Water temp. (°C)	Chl *a* (µg/L)	Microcystin (µg/gdw)	pH
Erie	08/19/2009	25.8	11.7 ± 0.2	<0.05g/L	8.61
Taihu	05/30/2009	24.1	14.2±0.3	60.5±10.8	8.11
GLSM	07/19/2010	27.0	10.9	537±117	8.8

measured by spectrophotometric quantification in a NanoDrop 1000 (Thermo Fisher Scientific, Inc., DE, USA) and agarose gel electrophoresis. Extracted DNA was stored at -80° C prior to sequencing.

A total of 500 ng of genomic DNA per sample was used for library preparation, prepped and sequenced according to the Roche 454 GS FLX platform using Titanium chemistry (Roche/454-Life Sciences, Branford, CT, USA) at the UT/ORNL Joint Institute of Biological Sciences. We generated a total of 533 Mbp and 1.36 million reads with an average read length of 394 bp (Appendix Table 3.1). The MG-RAST v.3.0 online server quality control pipeline was used to remove reads of poor quality and short length before annotation and analysis of metagenomics data (Meyer et al. 2008). Pipeline parameters were kept at the default settings. Sequences were dereplicated and filtered by length, removing sequences more than 2.0 standard deviations from the mean sequences length. Sequences with five or more non ATCG characters were removed. Recruitment of all metagenome reads post-QC processing to the *Microcystis* aeruginosa NIES 843 genome as was performed using the SeqMan NGen® (DNASTAR, Inc., WI, USA) software package using manufacturer's suggested default parameters established for 454 datasets. Reannotation of the NIES 843 reference genome and annotation of the environmental *Microcystis* genes was performed using Prodigal (Hyatt et al. 2010).

Metagenome reads were subjected to a BLASTx comparison against the NCBI nonredundant database. Community taxonomy and function were analyzed in the MEGAN v.4.61.6 analysis toolset (Huson et al. 2007, Huson et al. 2011). To confirm reproducibility of taxonomic assignments, reads were also uploaded to the MG-RAST server for taxonomic and functional potential analysis using a minimum e-value cutoff of 1e-5, minimum percent identity of 65% and a minimum alignment length of 50 amino acids. COGs (Clusters of Orthologous Groups) were used to assign function (Tatusov et al. 2000). For MG-RAST comparative analyses, the GenBank database was used for phylogenetic analysis and RefSeq was used to annotate functional genes using the parameters described above (Pruitt et al. 2007). Phylogenetic comparisons were made using percent abundance within each library post quality control analysis. COG and functional gene comparative analyses were performed using reads normalized to the library with the greatest total number of hits to the COG database (Taihu). Sequences generated in this study are available on the MG-RAST online server

(http://metagenomics.anl.gov/) under the identification numbers 4467029.3 (Erie), 4467058.3 (Taihu), and 4467059.3 (GLSM). (Sequences will be made public upon acceptance of this manuscript).

Microcystis Genome Recruitment

Metagenomes were annotated using the RefSeq database (MG-RAST) with an e-value minimum of 1e-5, a 65% identity cutoff, and a minimum alignment length of 50. All reads annotated as *Microcystis* from each metagenome were recruited to the genome of *Microcystis aeruginosa* NIES 843 using MUMmer 3.22 (Delcher et al. 2003). The PROmer algorithm was used for alignment. Potential metagenomic islands were initially identified manually in the Lake Erie and Taihu recruitment plots as regions of 10 kb or longer with less than average recruitment as established in Rodriguez-Valera et al (Rodriguez-Valera et al. 2009). These regions were defined as MIs if they had less than 25% average similarity with the reference and coverage of the reference in the region was less than 25%. Genetic elements associated with these regions not present in the metagenomes were identified manually.

RESULTS AND DISCUSSION

The *Microcystis* Metagenome

It is widely known that not all members of the genus *Microcystis* have the genetic capability to produce microcystin – only 10-50% of cells in natural systems have these genes (Joung et al 2011, Rinta-Kanto et al 2009). It is thus likely that similar variability in content may be reflected in other genomic regions, a factor that is important to know when picking model organisms for lab study. Using SeqMan NGen® (DNAStar, Madison, WI), an initial attempt was made to generate whole genome assemblies of environmental *Microcystis* from each metagenome for genome-wide comparison. This effort was abandoned early however, as we realized there were issues with the potential recruitment of non-*Microcystis* reads to the sequenced lab isolate used as a scaffold. Instead, recruitment plots were generated in MUMmer 3.22 to identify potential genomic differences between bloom-associated *Microcystis* spp. and the sequenced lab isolate (Kurtz et al. 2004). Sequences from each metagenome assigned a best hit identity of *Microcystis* were aligned to the *Microcystis aeruginosa* NIES 843 genome

(Kaneko et al. 2007). Recruitment of GLSM sequences was limited, with only 7% total coverage of the reference genome. Out of the possible 1327 individual *Microcystis* genes identified in the GLSM metagenome, 64% of genes had at least one repeat. (Figure 3.1A). The GLSM bloom has been subsequently characterized as a predominately *Planktothrix* bloom by microscopic and molecular analysis (Steffen et al., 2014), providing a foundation for our observation of limited recruitment to the NIES 843 genome. This characterization provides a valid explanation for the limited recruitment of *Microcystis* reads from the GLSM metagenome. Taihu sequences had greater recruitment to the reference genome compared to the GLSM data set, with 51% of the reference sequence covered (Figure 3.1B). 79% of the possible 4256 recruited sequences from Taihu were present at least twice in the metagenome data set. A robust recruitment, covering of 79% of the NIES 843 reference genome, resulted from alignment of Lake Erie sequences, with 92% of the 5539 *Microcystis* sequences recruited to the NIES 843 scaffold repeated at least once in the Erie metagenome (Figure 3.1C). This difference in recruitment between Erie and Taihu is likely due to differences in abundance of *Microcystis* spp. within our microbial samples. While the method of sample collection was the same across each bloom, collection took place at different points during bloom development at each site, which may have contributed to this difference in abundance of *Microcystis* spp.

Six conserved metagenomic islands were identified in the Taihu and three in the Lake Erie recruitments (Figure 3.2A, 3.2B). All six regions initially identified as potential MIs are shown in Figure 3.2. Despite having only three regions that qualified as MIs, it is worth mentioning that the Lake Erie dataset does demonstrate reduced similarity and coverage in all of these regions. Metagenomic islands (MI's) have been defined previously as areas within a genome that are underrepresented in metagenomic datasets. It is thought that MI's represent regions of a genome that are unique between closely related strains, or components of the species pangenome (Rodriguez-Valera et al 2009). For this analysis, these were identified as regions of 10 kb or larger that had relatively poor similarity and coverage (<25%) to the reference, consistent with similar analyses performed in marine environments (Rodriguez-Valera et al 2009). For a detailed breakdown of the coverage of each MI, see Appendix Table 3.2. A number of transposases and hypothetical proteins were consistently identified as missing across all six islands: a complete list of genes within the six metagenomic islands is provided in Appendix

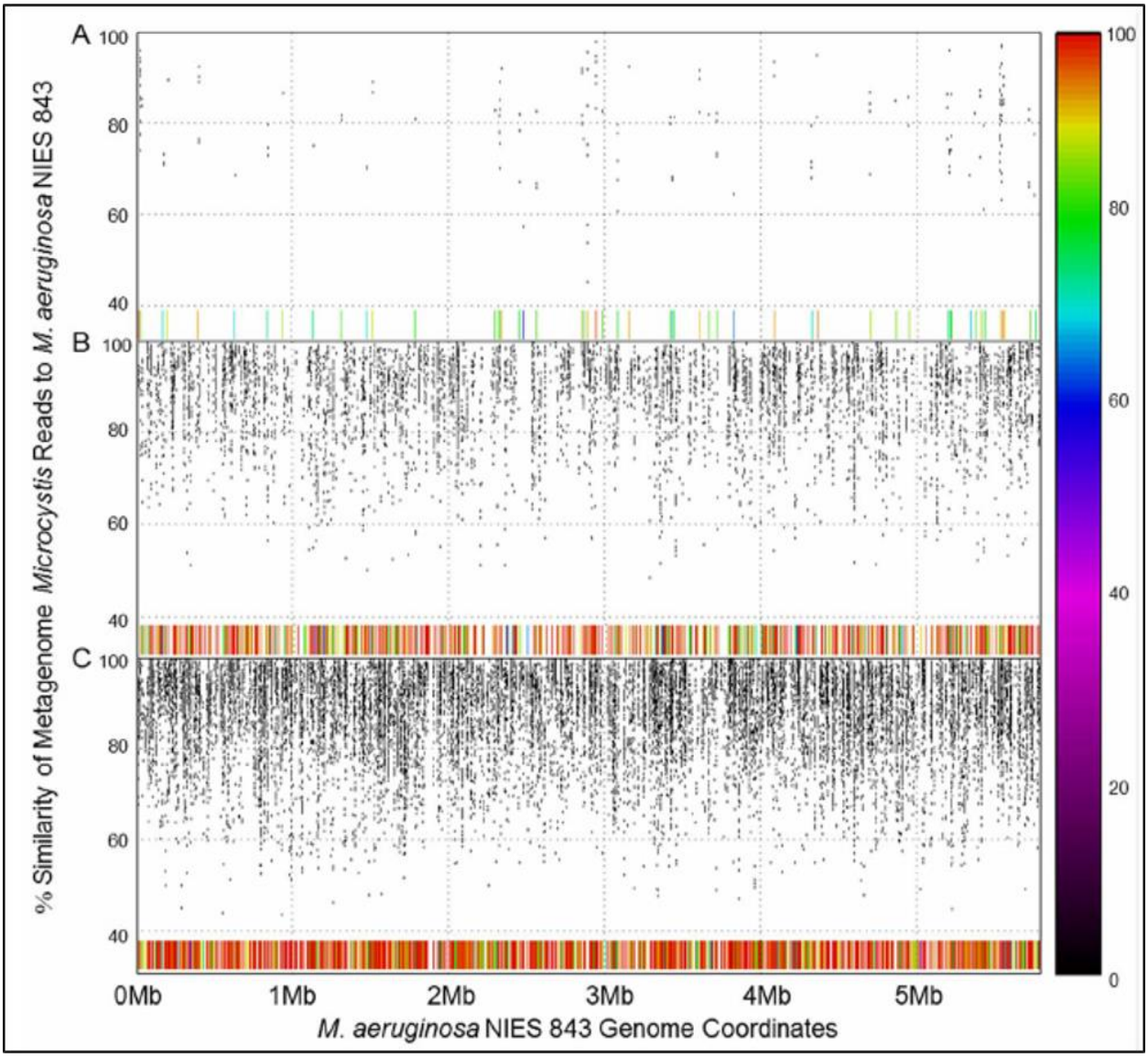

Figure 3.1. Recruitment plots of environmental *Microcystis* reads from each metagenome to the *M. aeruginosa* NIES 843 genome. A) Recruitment of GLSM *Microcystis* reads to NIES 843. B) Recruitment of Taihu *Microcystis* reads to NIES 843. C) Recruitment of Erie *Microcystis* reads to NIES 843. Position along the x-axis indicates position along the genome of *M. aeruginosa* NIES 843 from zero to 5.8 Mbp and position along the y-axis indicates percent similarity of recruited sequences. The bar along the bottom is a secondary indicator of percent similarity to the reference genome. See Appendix Table 3.2 for locations of potential MI's within each recruitment.

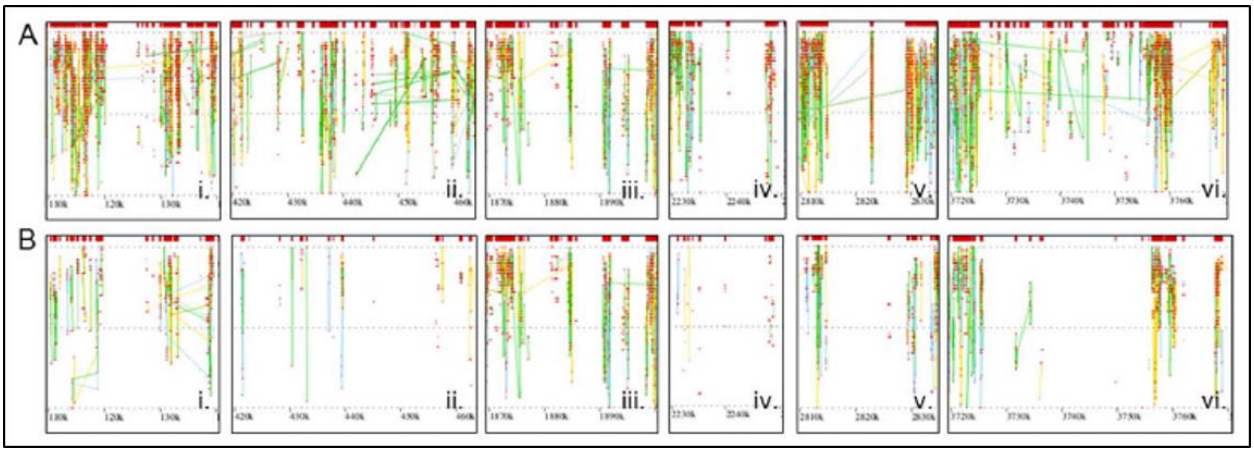

Figure 3.2. Metagenomic islands (MIs) identified in the Lake Erie and Taihu *Microcystis* recruitments. Six regions in Taihu and three in Erie qualified as MIs. MIs were defined as regions of ≥10kb in length that have an average similarity of less than 25% to the reference genome and less than 25% total coverage of the corresponding base pairs of the reference sequence. A) View of MIs of Lake Erie *Microcystis* recruitment. Boxes ii, iii, and vi were found to be MIs in Taihu, but did not satisfy our criteria in Lake Erie. B) View of MI regions in Taihu *Microcystis* recruitment.

Within the second island, two genes involved in restriction modification systems failed to recruit. Interestingly, the 5S, 16S, and 23S rRNA genes are located in the third island, although they recruited fully to the reference in both Taihu and Lake Erie recruitments. NIES 843 CRISPR-associated (Clustered Regularly Interspaced Short Palindromic Repeats) protein genes are located in the fifth metagenomic island, one of which failed to recruit from the Lake Erie metagenome. In the sixth and largest island (~ 42kb), a number of genes involved in cell surface recognition systems failed to recruit. These genes encoded for two cell surface antigen receptor proteins, a two component response regulator, and a sensor protein.

Taken together, the genes missing from each metagenomic island represent genes that may all be involved in phage recognition, consistent with metagenomic islands identified in marine and hypersaline environments (Pasic et al. 2009, Rodriguez-Valera et al 2009). Although these gaps in recruitment may be an artifact of sequencing depth, this is unlikely, as there was robust recruitment to a vast majority to the NIES 843 genome by the Erie reads and islands were conserved between the datasets. The three MIs present in Taihu that did not meet the criteria in Erie may be examples of limited recruitment due to differences in population makeup between Erie and Taihu. Alternatively, these regions may differ from the reference, but are not yet divergent enough to qualify as MIs in Erie. The future establishment of a *Microcystis* spp. core genome (as more sequences become available) will be important in the verification of these differences. One explanation for the three islands conserved in both recruitments is that they represent regions of bloom-associated *Microcystis* genomes that are too divergent from the reference to recruit.

Structure of the co-occurring community during *Microcystis* blooms

Phylogenetic composition of each bloom sample was compared at multiple taxonomic levels. A relatively conserved phylogenetic structure was observed in the cyanobacterial populations for all of the lake samples (Figure 3.3A). The orders *Chroococcales, Nostocales,* and *Oscillatoriales* were most abundant in all three systems. The oscillatorians made up a slightly higher percentage of the total cyanobacterial population within GLSM (38%) relative to Erie (24%) or Taihu (17%). At the genus level, the most common best hit matches within the *Chroococcales* were the nitrogen-fixing *Cyanothece,* as well as *Synechococcus*, with *Microcystis*

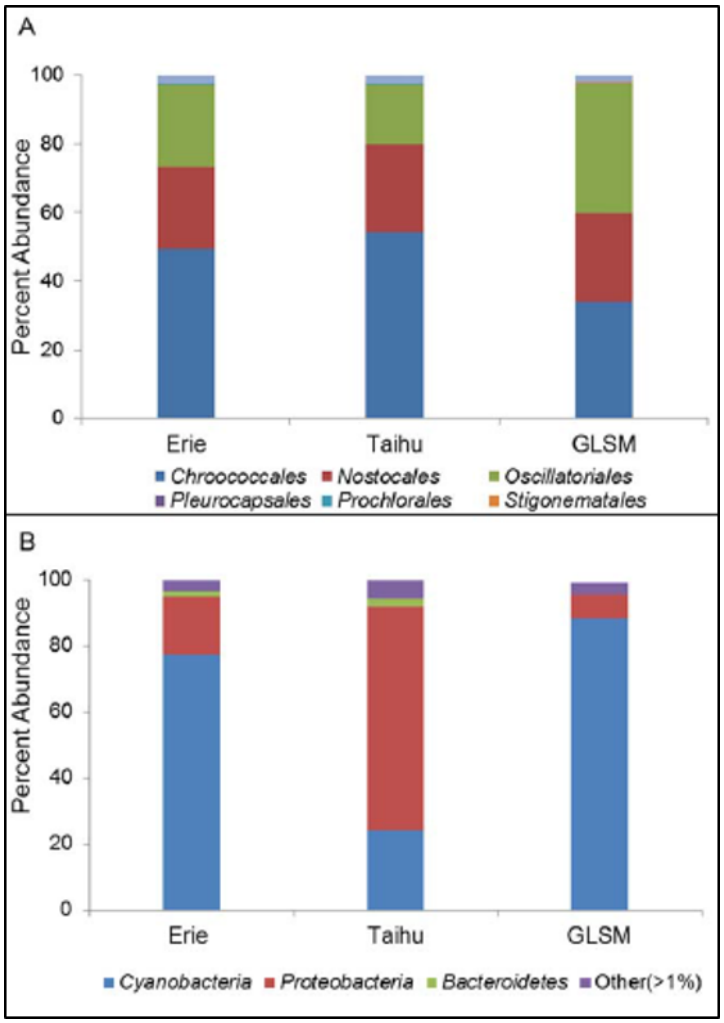

Figure 3.3. A) Percent abundance of orders belonging to the *Cyanobacteria* phylum. Abundance is a percentage of reads annotated as *Cyanobacteria* from each metagenome. B) Distribution of bacterial phyla in each metagenome. Abundance is a percentage of total bacterial reads in each metagenome. Phyla with less than 1% total abundance were grouped in the "Other" category.

spp. making up less than three percent of the total population for each bloom. Despite their relatively low abundance, *Microcystis* spp. remain important as they were the dominant toxin producers in Lake Erie and Taihu (Table 3.2) (Otten et al. 2012, Wilhelm et al. 2009). Genetic elements classifying to *Oscillatoria* were largely comprised of *Lyngbya* spp. and a group whose closest relatives are members of *Trichodesmium* at the genus level. Because *Trichodesmium* spp. are marine organisms, it is likely that a significant population of previously unclassified freshwater nitrogen fixers were present at high abundance in all three blooms. This finding is indicative of the need for more freshwater cyanobacterial genome sequences to more comprehensively study bloom events and other phenomena. As with members of *Chroococcales*, the dominant toxin producers within the oscillatorians (*Planktothrix*) are present in lower numbers than other genera (Table 3.2). At the genus level, phylogenetic resolution cannot be completely achieved, as current annotations reflect only what is currently available in the databases used for annotation, and are reported as the best hit. As databases continue to expand and more freshwater genomes become available, best hit annotations may change and give a more accurate representation of the communities. We include this information with the knowledge that as updates to databases and annotation algorithms improve, these assignments may change.

At the phylum level, microbial community composition varied across the different lakes (Figure 3.3B). *Cyanobacteria* were the most abundant members of the Erie and GLSM bloom samples (77% and 88%, respectively). The *Proteobacteria* were the second most abundant group in the North American lakes, followed by *Bacteroidetes* and several rare phyla (<1%). In Taihu, the *Proteobacteria* were the most abundant, despite higher measured levels of chl *a* (Table 3.1). *Proteobacteria* comprise 68% of the total microbial population, followed by *Cyanobacteria* at 25%, which is consistent with previous investigations of heterotrophic populations in the Taihu system (Li et al. 2011b, Wilhelm et al. 2011b). Further repetition of these datasets across multiple time points would provide more comprehensive insight into the bloom-associated community. As toxin-levels and the abundance of toxin-producers change throughout the course of a bloom, so may the associated bacterial community, as has been previously demonstrated in Taihu (Li et al. 2011a). While possibly an artifact of sequencing coverage (a re-occurring theme in all metagenomics surveys), age of the bloom at the time of sample collection or of the sample

Table 3.2. Percentage abundance of hits assigned to genera in the orders *Chroococcales* and *Oscillatoriales*. Sequences most closely identified as exclusively marine lineages (*Chrocosphera, Trichodesmium*) are denoted as "-like" and potentially are from related yet currently unclassified members of the community.

	Erie	**Taihu**	**GLSM**
Chroococcales			
Crocosphera-like	9.4	9.2	6.7
Cyanobium	0.2	0.2	0.3
Cyanothece	71.9	72.9	69.5
Microcystis	1.6	2.4	0.8
Synechococcus	14.9	13.2	21.5
Synechocystis	0.6	0.6	0.3
Unclassified	1.3	1.5	0.9
Oscillatoriales			
Arthrospira	14.7	15.6	20.6
Lyngbya	31.1	21.4	27.6
Microceleus	18.0	26.6	13.2
Planktothrix	1.0	1.3	2.6
Trichodesmium-like	34.9	34.9	35.9

collection techniques, these differences between Taihu and the North American lakes may be demonstrative of phylogenetic variability of *Microcystis*-associated bacteria between geographically distinct systems.

Functional potential of co-occurring heterotrophic bacteria

Genes from each site were assigned to broad functional categories (Clusters of Orthologous Groups, COGs) and the relative abundance of hits to each category was compared across the three metagenomes (Figure 3.4). The number of genes within several categories was closely conserved across all sites. This includes genes involved in nutrient metabolism, coenzyme transport and metabolism, secondary metabolite metabolism and transport, as well as replication, recombination, and repair.

Despite this general conservation, there were multiple categories in which genes from the co-occurring microbial community in Taihu were over- or underrepresented compared to the North American lakes. The greatest differences in abundance were in the genes involved in amino acid metabolism and signal transduction mechanisms. The largest difference was ~20% between Taihu and Erie/GLSM normalized COG distributions. Even with these differences, the functional potential of bloom community members appears relatively static between blooms, especially when compared to the striking phylogenetic variability between Taihu and the other systems.

To confirm the observed conservation of functional potential, the normalized number of hits to each COG category for each metagenome was plotted against the other (Figure 3.5). All points that fell within the calculated 95% confidence interval were assumed to be conserved in frequency of occurrence between the two blooms being compared. The only outlier in the comparison between the North American lakes was the category containing genes of unknown function (Figure 3.5A). As expected, the farthest outlier between Taihu and both Erie and GLSM was the Amino Acid Metabolism and Transport category (Figure 3.5B-C). Other conserved outliers included COGs for Signal Transduction Mechanisms, Energy Production & Conservation and Cell Wall/Membrane/Envelope Biogenesis. Taken with the original observational comparison of relative abundance of each COG category, these data suggest a conserved functional potential of the bloom-associated bacterial community members, despite

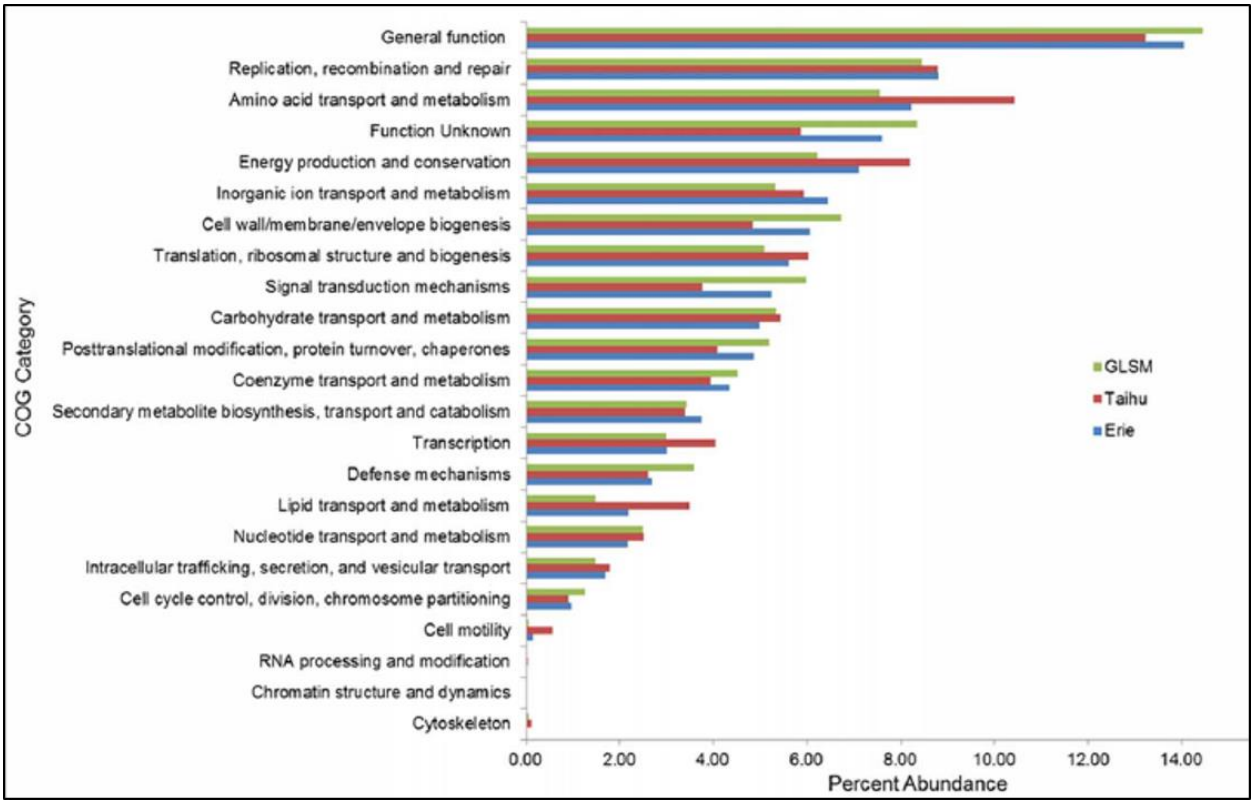

Figure 3.4. Metagenomes were annotated using the COG database. Reads were assigned to standard COG functional categories based on this annotation. Abundance is a percentage of total hits to the COG database.

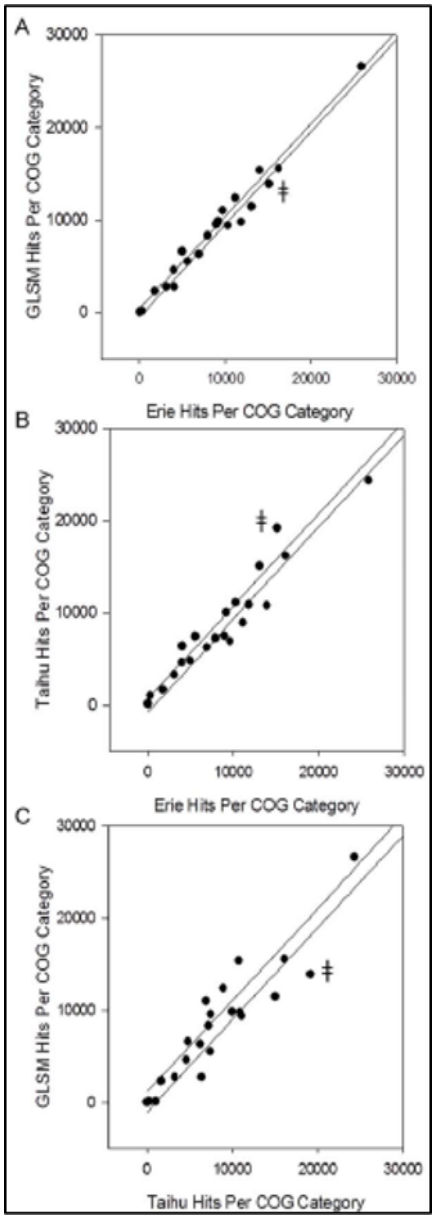

Figure 3.5. Normalized number of hits to each COG category from each metagenome plotted against the other. Data were normalized to the metagenome with the greatest number of hits to the COG database (Taihu). A) COG distributions of Erie and GLSM are compared. Data points largely fall within or close to the 95% confidence interval (CI) which is indicated by the black lines, indicating similarity. B) COG distributions of Erie and Taihu are compared. C) COG distributions of Taihu and GLSM are compared. The largest outlier in the Taihu comparisons is denoted by the (ǂ) symbol in each plot. This outlier corresponds to the Amino Acid Metabolism and Transport COG category.

variable phylogenetic makeup. This provides support to the growing body of evidence suggesting that variable phylogenetic makeup may not be reflected in the function of the environmental microbial community members (Burke et al 2011, Oh et al 2011).

Due to their potential role in the regulation of bloom formation in Taihu , we specifically took a deeper look at pathways associated with nitrogen cycling (Xu et al. 2010). Functional genes involved in nitrogen assimilation were identified and annotated at the phylum level in each dataset to highlight the genomic contribution of bloom microbes to the incorporation of nitrogen in the community. Genes examined were those involved in nitrogen fixation (*nifD, nifH, nifK*), urea metabolism (*ureA-G*), nitrate and nitrite reduction (*nar* and *nir*), and glutamine synthesis (*glnA*). A stark contrast in the contribution of genes involved in nitrogen fixation and urea metabolism emerged between Taihu and the North American lakes (Table 3.3). Cyanobacteria contribute the majority of the genes involved in these processes in Erie (78% and 58%) and GLSM (96% and 95%). In contrast, *Proteobacteria* contribute an overwhelming majority of these functional genes (86% and 74%) within the Taihu bloom community. This highlights the potential for distinct assimilation of nitrogen by different organisms in the bloom-associated through the microbial community in Taihu relative to Erie and GLSM. It should be noted that amino acids may also serve as potential nitrogen sources, and the overrepresentation of those genes involved in general amino acid metabolism in Taihu may reflect an important role for heterotrophs in nitrogen metabolism in this system (Figure 3.4). The normalized abundance of genes involved in each type of nitrogen assimilation remains relatively constant, despite their taxonomic identity (Table 3.3). These data also highlight the somewhat divergent phylogenetic makeup of each community. The overrepresentation of nitrogen assimilation genes associated with *Proteobacteria* in Taihu is largely due to numerical dominance of these organisms. To demonstrate this, the commonly used housekeeping gene *rpoB* can be used as a crude proxy for organism abundance, and has been included to allow for normalization of nitrogen assimilation genes within phyla (Table 3.3) (Case et al. 2007, Dole et al. 2010). While normalization changes the phylogenetic identity of the majority of the N assimilation genes, it again reiterates the possibility of phylogenetically distinct organisms having the potential to perform a similar functional role in bloom-associated microbial communities.

Table 3.3. The abundance of microbial nitrogen assimilation genes within metagenome. Genes from families associated with nitrogen fixation (*nifD, nif H, nifK*), urea assimilation (*ureA-G*), nitrate reduction (*nar, nir*) and ammonium utilization (*glnA*) were identified and enumerated within each metagenome using the COG database and phylogenetically classified using RefSeq. Results represent the total number of hits to functional categories (COG database) within each library and are normalized for library size.

	Erie					**Taihu**					**GLSM**				
	nifD, H & K	*ureA-G*	*nar/nir*	*glnA*	*rpoB*	*nifD, H & K*	*ureA-G*	*nar/nir*	*glnA*	*rpoB*	*nifD, H & K*	*ureA-G*	*nar/nir*	*glnA*	*rpoB*
Cyanobacteria	210	84	0	17	399	41	26	0	8	80	201	126	0	48	371
Proteobacteria	58	60	3	41	118	261	76	135	138	537	7	7	0	0	27
Actinobacteria	-	-	-	-	0	-	-	-	2	1	-	-	-	3	3
Cytophaga	-	-	-	6	1	-	-	-	-	-	-	-	-	-	-
Total Hits	268	144	3	64	517	302	102	135	148	618	208	133	0	51	401

This pattern in nitrogen assimilation gene distribution underscores the potential role of heterotrophic bacteria within toxic cyanobacterial bloom communities in some freshwater lakes. The role of heterotrophs in cyanobacterial metabolism has been previously observed in both marine and freshwater systems. Morris et al. (2008) identified heterotrophic activity as the source of *Prochlorococcus* resistance to hydrogen peroxide in the oceans, while the important microcystin-producer *Planktothrix* has recently been found to successfully use the compound glyphosate, an herbicide known to be relatively resistant to biodegradation, as a nutrient source in the presence of "helper" heterotrophs (Saxton et al 2011). Based on this metagenomic data, it appears that heterotrophs in Taihu may play an important role in nitrogen assimilation and transformation for the larger bloom microbial community.

Conclusions

This work highlights the utility of metagenomics as a tool for exploration of microbial communities. Here, we provide microbial snapshots of three separate toxic cyanobacterial blooms. Despite being single samples, these metagenomes provide a unique snapshot of the microbial community associated with toxic cyanobacterial blooms. This initial characterization is an important foundation for further study of such communities. We have attempted to include metadata that complies with the Genomic Standards Consortium, with the intention that this initial characterization of three freshwater blooms can be reexamined for future statistical comparison as the repository of freshwater sequences builds (Field et al. 2008). While we focus largely on the bacterial community associated with bloom biomass, viral sequences were also present in all data sets. Notably, sequences of the *Microcystis* phage Ma-LMM01 were detected at all three lakes. This is especially worth noting due to the importance of phage in bloom dynamics and termination (Yoshida et al. 2008). Other interesting findings include the presence of the *mlrC* gene in both Taihu and Erie. This gene is involved in microbial degradation of microcystin, and its presence warrants further inquiry into the presence of potential important microcystin degraders in these lakes (Shimizu et al. 2011).

In the current study we have used metagenomics to describe the phylogenetic makeup and functional potential of three geographically distinct cyanobacterial blooms at single time points during each bloom. This comparative approach has confirmed an increasing trend in microbial

ecology: variable bacterial phylogenetic makeup is not mirrored in the relatively constant functional potential of the bacterial community members in the environment (Burke et al 2011). This finding may reflect a need to amend current methods of describing microbial community dynamics, as the taxonomic/phylogenetic identity of community members may no longer be sufficient. Within our observations key functional genes, such as those involved in nitrogen assimilation, appear to be more informative than standard 16S rDNA gene analysis and demonstrate that within 2 similar biological events (blooms in Lake Erie and Taihu) the analogous processes are likely carried out by different members of the community. With this approach, we were able to identify potentially divergent pathways of assimilated nitrogen through the microbial communities of three different blooms. The genomic contribution of heterotrophic bacteria to nitrogen assimilation in Taihu represents a potentially critical contribution of heterotrophic bacteria in driving toxic freshwater blooms.

ACKNOWLEDGEMENTS

We thank Hans Paerl, Guangwei Zhu, Sarah Farnsley, Gary LeCleir, George Bullerjahn, R Michael McKay, Matthew Saxton and Richard Bourbonniere for assistance in sample collection and nutrient analyses. We also thank Brian Dill and Nathan VerBerkmoes for valuable insight and discussions. This work was supported by National Science Foundation grants IOS 0841918 and CBET 0826838 to SWW and a Wallace-Dean Graduate Fellowship through the University of Tennessee to MMS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

LIST OF REFERENCES

Azevedo SMFO, Carmichael WW, Jochimsen EM, Rinehart KL, Lau S, Shaw GR *et al* (2002). Human intoxication by microcystins during renal dialysis treatment in Caruaru-Brazil. *Toxicology* **181:** 441-446.

Boyer GL (2008). Chapter 7: Cyanobacterial toxins in New York and the lower great lakes ecosystems. *Adv Exp Med Biol* **619:** 153-165.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011). Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci U S A* **108:** 14288-14293.

Carmichael WW, Brittain SM, Wang J, Babcock-Jackson L, Rinehart KL, Culver DA (2000). Isolation and characterization of microcystins, cyclic heptapeptide hepatotoxins from a Lake Erie strain of *Microcystis aeruginosa*. *J Great Lakes Res* **26:** 241-249.

Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, Kjelleberg S (2007). Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* **73:** 278-288.

Charlton MN, Milne JE, Booth WG, Chiocchio F (1993). Lake Erie Offshore in 1990 - restoration and resilience in the central basin. *Journal of Great Lakes Research* **19:** 291-309.

Chen YW, Qin BQ, Teubner K, Dokulil MT (2003). Long-term dynamics of phytoplankton assemblages: Microcystis-domination in Lake Taihu, a large shallow lake in China. *J Plankton Res* **25:** 445-453.

Conroy JD, Kane DD, Dolan DM, Edwards WJ, Charlton MN, Culver DA (2005). Temporal trends in Lake Erie plankton biomass: Roles of external phosphorus loading and dreissenid mussels. *Journal of Great Lakes Research* **31:** 89-110.

Delcher AL, Salzberg SL, Phillippy AM (2003). Using MUMmer to identify similar regions in large sequence sets. *Current protocols in bioinformatics / editoral board, Andreas D Baxevanis [et al]* Chapter 10: Unit 10 13.

Dole VS, Banu LA, Fister RD, Nicklas W, Henderson KS (2010). Assessment of rpoB and 16S rRNA genes as targets for PCR-based identification of Pasteurella pneumotropica. *Comparative medicine* **60:** 427-435.

Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P *et al* (2008). The minimum information about a genome sequence (MIGS) specification. *Nature biotechnology* **26:** 541-547.

Finlay K, Patoine A, Donald DB, Bogard MJ, Leavitt PR (2010). Experimental evidence that pollution with urea can degrade water quality in phosphorus-rich lakes of the Northern Great Plains. *Limnol Oceanogr* **55:** 1213-1230.

Frossard A, Gerull L, Mutz M, Gessner MO (2012). Disconnect of microbial structure and function: enzyme activities and bacterial communities in nascent stream corridors. *Isme J* **6:** 680-691.

Gobler CJ, Davis TW, Harke MJ, Marcoval MA, Goleski J, Orano-Dawson C *et al* (2010). Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. *Aquat Microb Ecol* **61:** 149-162.

Hernandez JM, Lopez-Rodas V, Costas E (2009). Microcystins from tap water could be a risk factor for liver and colorectal cancer: A risk intensified by global change. *Med Hypotheses* **72:** 539-540.

Huson DH, Auch AF, Qi J, Schuster SC (2007). MEGAN analysis of metagenomic data. *Genome Res* **17:** 377-386.

Huson DH, Mitra S, Ruscheweyh H-J, Weber N, Schuster SC (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome Res* **21:** 1552-1560.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics* **11:** 119.

al. KNe (ed): Sexual cell division and conjugation-papilla formation in sexual reproduction of Closterium strigosum. *Proceedings of the Seventh Internationl Seaweed Symposium*.

Joung SH, Oh HM, Ko SR, Ahn CY (2011). Correlations between environmental factors and toxic and non-toxic Microcystis dynamics during bloom in Daechung Reservoir, Korea. *Harmful Algae* **10:** 188-193.

Kandeler E, Tscherko D, Bruce KD, Stemmer M, Hobbs PJ, Bardgett RD *et al* (2000). Structure and function of the soil microbial community in microhabitats of a heavy metal polluted soil. *Biol Fert Soils* **32:** 390-400.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M *et al* (2007). Complete genomic structure of the bloom-forming toxic cyanobacterium Microcystis aeruginosa NIES-843. *DNA research : an international journal for rapid publication of reports on genes and genomes* **14:** 247-256.

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C *et al* (2004). Versatile and open software for comparing large genomes. *Genome Biol* **5:** R12.

Lehman RM, O'Connell SP (2002). Comparison of extracellular enzyme activities and community composition of attached and free-living bacteria in porous medium columns. *Applied and environmental microbiology* **68:** 1569-1575.

Li H, Xing P, Chen M, Bian Y, Wu QL (2011a). Short-term bacterial community composition dynamics in response to accumulation and breakdown of Microcystis blooms. *Water Res* **45:** 1702-1710.

Li N, Zhang L, Li F, Wang Y, Zhu Y, Kang H *et al* (2011b). Metagenome of microorganisms associated with the toxic Cyanobacteria <i>Microcystis aeruginosa analyzed using the 454 sequencing platform. *Chinese Journal of Oceanology and Limnology* **29:** 505-513.

Mankiewicz-Boczek J, Palus J, Gągała I, Izydorczyk K, Jurczak T, Dziubałtowska E *et al* (2011). Effects of microcystins-containing cyanobacteria from a temperate ecosystem on human lymphocytes culture and their potential for adverse human health effects. *Harmful Algae* **10:** 356-365.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M *et al* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Morris JJ, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER (2008). Facilitation of robust growth of Prochlorococcus colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Applied and environmental microbiology* **74:** 4530-4534.

Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R *et al* (2011). Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Appl Environ Microbiol* **77:** 6000-6011.

Otten TG, Xu H, Qin B, Zhu G, Paerl HW (2012). Spatiotemporal patterns and ecophysiology of toxigenic microcystis blooms in Lake Taihu, China: implications for water quality management. *Environ Sci Technol* **46:** 3480-3488.

Paerl HW, Huisman J (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Env Microbiol Rep* **1:** 27-37.

Paerl HW, Hall NS, Calandrino ES (2011a). Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Sci Total Environ* **409:** 1739-1745.

Paerl HW, Xu H, McCarthy MJ, Zhu GW, Qin BQ, Li YP *et al* (2011b). Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Res* **45:** 1973-1983.

Pasic L, Rodriguez-Mueller B, Martin-Cuadrado AB, Mira A, Rohwer F, Rodriguez-Valera F (2009). Metagenomic islands of hyperhalophiles: the case of Salinibacter ruber. *Bmc Genomics* **10**.

Pruitt KD, Tatusova T, Maglott DR (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* **35:** D61-65.

Rinta-Kanto JM, Ouellette AJA, Boyer GL, Twiss MR, Bridgeman TB, Wilhelm SW (2005). Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. *Environmental Science & Technology* **39:** 4198-4205.

Rinta-Kanto JM, Konopko EA, DeBruyn JM, Bourbonniere RA, Boyer GL, Wilhelm SW (2009). Lake Erie Microcystis: Relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. *Harmful Algae* **8:** 665-673.

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

Rogers ED, Henry TB, Twiner MJ, Gouffon JS, McPherson JT, Boyer GL *et al* (2011). Global Gene Expression Profiling in Larval Zebrafish Exposed to Microcystin-LR and Microcystis Reveals Endocrine Disrupting Effects of Cyanobacteria. *Environmental science & technology* **45:** 1962-1969.

Saxton MA, Morrow EA, Bourbonniere RA, Wilhelm SW (2011). Glyphosate influence on phytoplankton community structure in Lake Erie. *Journal of Great Lakes Research* **37:** 683-690.

Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, Paterson MJ *et al* (2008). Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 11254-11258.

Scott JT, McCarthy MJ (2010). Nitrogen fixation may not balance the nitrogen pool in lakes over timescales relevant to eutrophication management. *Limnol Oceanogr* **55:** 1265-1270.

Shimizu K, Maseda H, Okano K, Itayama T, Kawauchi Y, Chen R *et al* (2011). How microcystin-degrading bacteria express microcystin degradation activity. *Lakes & Reservoirs: Research & Management* **16:** 169-178.

Stone R (2011). China Aims to Turn Tide Against Toxic Lake Pollution. *Science* **333:** 1210-1211.

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* **28:** 33-36.

Watanabe MM, M Kawachi, M. Hiroki, and F. Hkasai (2000). NIES collection list of strains: microalgal and protozoa. National Institute for Environmental Studies.

Wilhelm SW, Rinta-Kanto JM, Saxton MA, DeBruyn JM, Smith JL, Marvin CH *et al* (2009). The diversity and distribution of toxigenic Microcystis spp. in present day and archived pelagic and sediment samples from Lake Erie. *Harmful Algae* **8:** 385-394.

Wilhelm SW, Farnsley SE, LeCleir GR, Layton AC, Satchwell MF, DeBruyn JM *et al* (2011a). The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae* **10:** 207-215.

Wilhelm SW, Farnsley SE, LeCleir GR, Layton AC, Satchwell MF, DeBruyn JM *et al* (2011b). The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae* **10:** 207-215.

Xu H, Paerl HW, Qin B, Zhu G, Gao G (2010). Nitrogen and phosphorus inputs control phytoplankton growth in eutrophic Lake Taihu, China. *Limnol Oceanogr* **55:** 420-432.

Yang H, Ye WJ, Tan J, Liu XL, Lin SQ, Pan JL *et al* (2011). Temporal variability of cyanobacterial populations in the water and sediment samples of Lake Taihu as determined by DGGE and real-time PCR. *Harmful Algae* **10:** 472-479.

Ye C, Shen Z, Zhang T, Fan M, Lei Y, Zhang J (2011). Long-term joint effect of nutrients and temperature increase on algal growth in Lake Taihu, China. *Journal of Environmental Sciences* **23:** 222-227.

Yoshida T, Nagasaki K, Takashima Y, Shirai Y, Tomaru Y, Takao Y *et al* (2008). Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *Journal of bacteriology* **190:** 1762-1772.

Chapter III Appendix

Table 3.4. Sequence data for all samples before and after quality control processing.

	Lake Erie		**Taihu**		**GLSM**	
	Pre-QC	**Post-QC**	**Pre-QC**	**Post-QC**	**Pre-QC**	**Post-QC**
Total Reads	504,223	354,384	637,724	401,996	219,716	140,179
Total bp	204,427,100	153,986,130	240,405,900	163,139,593	88,229,075	60,656,381
Mean sequence length	405 ± 133	434 ± 104	376 ± 127	405 ± 105	401 ± 141	432 ± 113
GC Content	46 ± 10%	46 ± 10%	51 ± 13%	51 ± 13%	41 ± 8%	41 ± 8%

Table 3.5. Coordinates of each metagenomic island (MI) identified in the Lake Erie and Taihu recruitment plots. The percentage of individual base pairs with zero percent similarity (no coverage) to the *M. aeruginosa* NIES 843 reference genome and the average percent similarity across each MI are shown.

		MI 1	**MI 2**	**MI 3**	**MI 4**	**MI 5**	**MI 6**
Coordinates		120122.. 130126	421251.. 460015	1874286.. 1899881	2234993.. 2246392	2813246.. 2830250	3725469. .3768893
% with no coverage	*Erie*	75.8%	54.8%	66.9%	91.8%	80.6%	57.0%
	Taihu	88.7%	90.7%	76.7%	95.8%	85.2%	80.6%
Average % similarity with 843 reference	*Erie*	21.9%	37.3%	29.0%	7.4%	16.9%	38.8%
	Taihu	10.4%	8.4%	19.8%	3.6%	12.8%	16.6%

Table 3.6. Location and identity information for all genes located in the six Lake Erie metagenomes.

Island	Coordinates	Gene	Description
MI-1	120122..120571	MAE_01450	YP_001655159.1: transposase
	121213..121416	MAE_01470	YP_001655161.1: hypothetical protein
	121415..121675	MAE_01480	YP_001655162.1:transposase
	122392..123021	MAE_01500	YP_001655164.1:hypothetical protein
	123219..123779	MAE_01510	YP_001655165.1: hypothetical protein
	123822..124133	MAE_01520	YP_001655166.1: hypothetical protein
	124309..125967	MAE_01530	YP_001655167.1: transposase
	126308..126502	MAE_01540	YP_001655168.1: hypothetical protein
	126676..16984	MAE_01550	YP_001655169.1: hypothetical protein
	127447..128661	MAE_01560	YP_001655170.1: transposase
	129011..129658	MAE_01570	YP_001655171.1: hypothetical protein
	129735..129971	MAE_01580	YP_001655172.1: hypothetical protein
	129968..130126	MAE_01590	YP_001655173.1: hypothetical protein
MI-2	421251..421613	MAE_04740	YP_001655488.1: hypothetical protein
	421618..421785	MAE_04750	YP_001655489.1: hypothetical protein
	422089..422292	MAE_04760	YP_001655490.1: hypothetical protein
	422466..422627	MAE_04770	YP_001655491.1: hypothetical protein
	422976..423179	MAE_04780	YP_001655492.1: hypothetical protein
	423515..425173	MAE_04790	YP_001655493.1: transposase
	427533..428129	MAE_04800	YP_001655494.1: tranposase
	428328..428603	MAE_04810	YP_001655495.1: tranposase
	428570..429838	MAE_04820	YP_001655496.1: tranposase
	430411..431406	MAE_04830	YP_001655497.1: tranposase
	431565..432356	MAE_04840	YP_001655498.1: transposase
	433373..434521	MAE_04850	YP_001655499.1: transposase
	434292..434907	MAE_04860	YP_001655500.1: hypothetical protein
	435199..435573	MAE_04870	YP_001655501.1: hypothetical protein
	435914..436132	MAE_04880	YP_001655502.1:hypothetical protein
	436515..437135	MAE_04890	YP_001655503.1: hypothetical protein
	437381..437710	MAE_04900	YP_001655504.1: hypothetical protein
	437821..437982	MAE_04910	YP_001655505.1: hypothetical protein
	437985..438341	MAE_04920	YP_001655506.1:hypothetical protein
	438410..438904	MAE_04930	YP_001655507.1: SixA type phosphohistidine phosphatase
	438957..440177	MAE_04940	YP_001655508.1: hypothetical protein
	440645..442303	MAE_04950	YP_001655509.1: transposase

Table 3.6
Continued

Island	Coordinates	Gene	Description
	442341..442697	MAE_04960	YP_001655510.1: hypothetical protein
	443212..443979	MAE_04970	YP_001655511.1: hypothetical protein
	444081..445739	MAE_04980	YP_001655512.1: transposase
	446868..447023	MAE_04990	YP_001655513.1: hypothetical protein
	447147..448025	MAE_05000	YP_001655514.1: type I restriction-modification system
	448235..449293	MAE_05010	YP_001655515.1: type I restriction-modification system
	449329..449748	MAE_05020	YP_001655516.1: hypothetical protein
	449857..450027	MAE_05030	YP_001655517.1: hypothetical protein
	450156..450509	MAE_05040	YP_001655518.1: putative restriction modification system
	451244..451651	MAE_05050	YP_001655519.1: putative restriction modification system
	451737..452939	MAE_05060	YP_001655520.1: type I site-specific deoxyribonuclease
	453090..454358	MAE_05070	YP_001655521.1: transposase
	454486..454752	MAE_05080	YP_001655522.1: hypothetical protein
	454809..455483	MAE_05090	YP_001655523.1: tranposase
	455519..456031	MAE_05100	YP_001655524.1: transposase
	456171..456329	MAE_05110	YP_001655525.1: hypothetical protein
	456375..457418	MAE_05120	YP_001655526.1: trasposase
	457485..457640	MAE_05130	YP_001655527.1: type I site-specific deoxyribonuclease
	457842..459326	MAE_05140	YP_001655528.1: transposase
	459399..459551	MAE_05150	YP_001655529.1: hypothetical protein
	459602..460015	MAE_05160	YP_001655530.1: hypothetical protein
MI-3	1874286..1875272	MAE_21010	YP_001657115.1: transposase
	1876106..1876348	MAE_21020	YP_001657116.1: hypothetical protein
	1876960..1878228	MAE_21030	YP_001657117.1: transposase
	1878296..1878805	MAE_21040	YP_001657118.1: transposase
	1878992..1879162	MAE_21050	YP_001657119.1: hypothetical protein
	1879166..1879513	MAE_21060	YP_001657120.1: transposase
	1879528..1879953	MAE_21070	YP_001657121.1: hypothetical protein
	1880244..1881512	MAE_21080	YP_001657122.1: transposase
	1881671..1883329	MAE_21090	YP_001657123.1: transposase
	1883916..1884215	MAE_21100	YP_001657124.1: hypothetical protein
	1884348..1885373	MAE_21110	YP_001657125.1: hypothetical protein

Table 3.6
Continued

Island	Coordinates	Gene	Description
	1885475..1885699	MAE_21120	YP_001657126.1: hypothetical protein
	1885814..1887290	rrn16S_1	rRNA: 16S ribosomal RNA
	1887428..1887504	MAE_t009	tRNA:Ile
	1887656..1890533	rrn23S_1	rRNA: 23S ribosomal RNA
	1890592..1890709	rrn5S_1	rRNA: 5S ribosomal RNA
	1890758..1891048	MAE_21130	YP_001657127.1: hypothetical protein
	1891573..1892130	MAE_21140	YP_001657128.1: AMP-dependent synthetase and ligase
	1892134..1892340	MAE_21150	YP_001657129.1: hypothetical protein
	1892333..1893949	MAE_21160	YP_001657130.1: AMP-dependent synthetase and ligase
	1893968..1895290	MAE_21170	YP_001657131.1: class I/II aminotransferase
	1895388..1896506	MAE_21180	YP_001657132.1: glycine amidinotransferase
	1896517..1897185	MAE_21190	YP_001657133.1: hypothetical protein
	1897182..1898087	MAE_21200	YP_001657134.1: short chain dehydrogenase
	1898301..1898501	MAE_21210	YP_001657135.1: hypothetical protein
	1898655..1899881	MAE_21220	YP_001657136.1: ABC-transporter DevB family protein
MI-4	2234993..2235154	MAE_24710	YP_001657485.1: hypothetical protein
	2235494..2235691	MAE_24720	YP_001657486.1: hypothetical protein
	2235852..2236169	MAE_24730	YP_001657487.1: hypothetical protein
	2236166..2236513	MAE_24740	YP_001657488.1: hypothetical protein
	2236609..2237040	MAE_24750	YP_001657489.1: hypothetical protein
	2237041..2237427	MAE_24760	YP_001657490.1: hypothetical protein
	2237807..2237959	MAE_24770	YP_001657491.1: hypothetical protein
	2238032..2239516	MAE_24780	YP_001657492.1: transposase
	2239745..4420059	MAE_24790	YP_001657493.1: hypothetical protein
	2240067..2241335	MAE_24800	YP_001657494.1: transposase
	2241507..2243948	MAE_24810	YP_001657495.1: hypothetical protein
	2243948..2246392	MAE_24820	YP_001657496.1: hypothetical protein
MI-5	2813246..2814250	cas1	YP_001658100.1: CRISPR-associated Cas1 family protein
	2814263..2814535	cas2	YP_001658101.1: CRISPR-associated Cas2 family protein
	2814598..2814795	MAE_30880	YP_001658102.1: hypothetical protein
	2816187..2816363	MAE_30890	YP_001658103.1: hypothetical protein
	2816380..2816556	MAE_30900	YP_001658104.1: hypothetical protein

Table 3.6
Continued

Island	Coordinates	Gene	Description
	2816668..2816826	MAE_30910	YP_001658105.1: hypothetical protein
	2817526..2817696	MAE_30920	YP_001658106.1: hypothetical protein
	2818072..2818236	MAE_30930	YP_001658107.1: hypothetical protein
	2819008..2819244	MAE_30940	YP_001658108.1: hypothetical protein
	2820299..2820469	MAE_30950	YP_001658109.1: hypothetical protein
	2821343..2821627	MAE_30960	YP_001658110.1: hypothetical protein
	2821850..2822017	MAE_30970	YP_001658111.1: hypothetical protein
	2822231..2822392	MAE_30980	YP_001658112.1: hypothetical protein
	2822939..2823109	MAE_30990	YP_001658113.1: hypothetical protein
	2823272..2823433	MAE_31000	YP_001658114.1: hypothetical protein
	2823480..2823662	MAE_31010	YP_001658115.1: hypothetical protein
	2823748..2824011	MAE_31020	YP_001658116.1: hypothetical protein
	2824060..2824242	MAE_31030	YP_001658117.1: hypothetical protein
	2084735..2825247	MAE_31040	YP_001658118.1: transposase
	2825283..2825957	MAE_31050	YP_001658119.1: transposase
	2826014..2826190	MAE_31060	YP_001658120.1: hypothetical protein
	2826399..2826743	MAE_31070	YP_001658121.1: hypothetical protein
	2826940..2827149	MAE_31080	YP_001658122.1: hypothetical protein
	2828121..2828393	MAE_31090	YP_001658123.1: hypothetical protein
	2828631..2828801	MAE_31100	YP_001658124.1: hypothetical protein
	2829465..2829752	MAE_31110	YP_001658125.1: hypothetical protein
	2829903..2830250	MAE_31120	YP_001658126.1: hypothetical protein
MI-6	3725469..3726938	MAE_40650	YP_001659079.1: hypothetical protein
	3726941..3727795	MAE_40660	YP_001659080.1: hypothetical protein
	3727825..3729195	MAE_40670	YP_001659081.1: von Willebrand factor type A
	3729113..3730219	MAE_40680	YP_001659082.1: hypothetical protein
	3730231..3732540	MAE_40690	YP_001659083.1: hypothetical protein
	3732621..3735221	MAE_40700	YP_001659084.1: hypothetical protein
	3735226..3736140	MAE_40710	YP_001659085.1: hypothetical protein
	3736127..3736972	MAE_40720	YP_001659086.1: hypothetical protein
	3736992..3737765	MAE_40730	YP_001659087.1: hypothetical protein
	3738093..3738800	MAE_40740	YP_001659088.1: hypothetical protein
	3738822..3739733	MAE_40750	YP_001659089.1: hypothetical protein
	3740070..3740693	MAE_40760	YP_001659090.1: hypothetical protein
	3740745..3741206	MAE_40770	YP_001659091.1: hypothetical protein
	3741197..3741673	MAE_40780	YP_001659092.1: hypothetical protein

Table 3.6
Continued

Island	Coordinates	Gene	Description
	3741679..3742743	MAE_40790	YP_001659093.1: hypothetical protein
	3743368..3743721	MAE_40800	YP_001659094.1: hypothetical protein
	3744056..3746173	MAE_40810	YP_001659095.1: serine/threonine protein kinase
	3746285..3748108	MAE_40820	YP_001659096.1: protein serine/threonine phosphatase
	3748173..3748736	MAE_40830	YP_001659097.1: FHA domain-containing protein
	3748825..3750207	MAE_40840	YP_001659098.1: von Willebrand factor type A
	3750231..3750623	MAE_40850	YP_001659099.1: hypothetical protein
	3750643..3752817	MAE_40860	YP_001659100.1: von Willebrand factor type A
	3752923..3755109	MAE_40870	YP_001659101.1: hypothetical protein
	3755157..3755381	MAE_40880	YP_001659102.1: hypothetical protein
	3755965..3756459	MAE_40890	YP_001659103.1: hypothetical protein
	3756465..3757922	MAE_40900	YP_001659104.1: HlyD family secretion protein
	3758217..3758819	MAE_40910	YP_001659105.1: hypothetical protein
	3758903..3759883	pyrB	YP_001659106.1: aspartate carbamoyltransferase catalytic subunit
	3759907..3760188	MAE_40930	YP_001659107.1: hypothetical protein
	3760379..3760555	MAE_40940	YP_001659108.1: hypothetical protein
	3760670..3761503	MAE_40950	YP_001659109.1: surface antigen D15
	3761957..3762454	MAE_40960	YP_001659110.1: surface antigen variable number
	3762465..3763106	MAE_40970	YP_001659111.1: two-component response regulator
	3763339..3763632	MAE_40980	YP_001659112.1: hypothetical protein
	3763974..3764159	MAE_40990	YP_001659113.1: hypothetical protein
	3764204..3764371	MAE_41000	YP_001659114.1: hypothetical protein
	3764646..3765110	MAE_41010	YP_001659115.1: hemolysin-type calcium-binding protein
	3765371..3767299	MAE_41020	YP_001659116.1: sensor protein
	3767424..3768173	MAE_41030	YP_00165117.1: hypothetical protein
	3768246..3768893	MAE_41040	YP_00165118.1: hypothetical protein

CHAPTER IV:
NUTRIENTS DRIVE TRANSCRIPTIONAL CHANGES THAT MAINTAIN METABOLIC HOMEOSTASIS BUT ALTER GENOME ARCHITECTURE IN *MICROCYSTIS*

PUBLICATION NOTE

This chapter is a version of a peer-reviewed article previously published in *International Society for Microbial Ecology Journal (ISMEJ),* which is currently in press and authored by Morgan M. Steffen, Stephen P. Dearth, Brian D. Dill, Zhou Li, Kristen M. Larsen, Shawn R. Campagna, and Steven W. Wilhelm.

My contribution to this work was assistance in experimental design, sample processing, transcriptome data analysis, and writing the majority of the manuscript.

ABSTRACT

The cyanobacterium *Microcystis aeruginosa* is a globally distributed bloom-forming organism that degrades freshwater systems around the world. Factors that drive its dispersion, diversification, and success remain, however, poorly understood. To develop insight into cellular level responses to nutrient drivers of eutrophication, RNA sequencing was coupled to a comprehensive metabolomics survey of *M. aeruginosa* sp. NIES 843 grown in various nutrient-reduced conditions. Transcriptomes were generated for cultures grown in nutrient-replete (with nitrate as the nitrogen (N) source), nitrogen-reduced (with nitrate, urea, or ammonium acting as the N sources), and phosphate-reduced conditions. Extensive expression differences (up to 696 genes for urea grown cells) relative to the control treatment were observed, demonstrating that the chemical variant of nitrogen available to cells impacted transcriptional activity. Of particular note, a high number of transposase genes (up to 81) were significantly and reproducibly up-regulated relative to the control when grown on urea. Conversely, P-reduction resulted in a significant cessation in expression of transposase genes, indicating that variation in nutrient chemistry may influence transcription of transposases and may impact the highly mosaic genomic architecture of *M. aeruginosa*. Corresponding metabolomes showed comparably few differences between treatments, suggesting broad changes to gene transcription are required to maintain metabolic homeostasis under nutrient reduction. The combined observations provide novel and extensive insight into the complex cellular interactions that take place in this important bloom-forming organism during variable nutrient conditions and highlight a potential unknown molecular mechanism that may drive *Microcystis* blooms and evolution.

INTRODUCTION

Blooms of *M. aeruginosa* occur in freshwater lakes, rivers, and estuaries worldwide. Accumulation of biomass and cyanotoxins complicate both the ecological and socioeconomic health of lakes and surrounding communities (Dodds et al. 2008). For decades, P-loading has been cited as the primary cause of bloom events (Schindler 1977). This paradigm has begun to shift in recent years, however, to recognize a potential role for N concentrations and chemical forms in driving bloom formation and biological community structure (Dolman et al. 2012). While this shift remains hotly debated (Downing et al. 2001, Paerl et al. 2011, Paterson et al. 2011, Scott and McCarthy 2010), it is clear that *Microcystis* exhibits a differential response when exposed to various N-containing chemical species (Donald et al. 2011) and concentrations (Vézie et al. 2002), as well as various P concentrations both in the environment (Davis et al. 2010, Wilhelm et al. 2003) and in culture (Harke et al. 2012, Shen and Song 2007). Specifically, urea appears to stimulate rapid biomass accumulation of *Microcystis* when it enters freshwater as fertilizer runoff (Dolman et al 2012). Despite decades of research, the cellular mechanisms driving these notably different responses to N-reduction and composition versus P-reduction remain undiscovered.

In this study, we used high throughput RNA-sequencing (RNA-seq) to comparatively measure global changes in transcriptional patterns for cultures of *Microcystis aeruginosa* NIES 843 grown under different N and P nutrient regimes. Corresponding metabolic pools were measured in parallel for each culture to establish a comprehensive blueprint of the molecular response of *Microcystis* to nutrient concentration and chemical composition. These observations help to explain the complex response of *Microcystis* and toxigenic cyanobacteria to changing patterns in freshwater biogeochemistry worldwide. Moreover, they highlight a previously unknown effect of nutrient chemistry and concentration on the genomic architecture and thus the evolutionary trajectory of this globally important cyanobacterium.

MATERIALS AND METHODS

Culture Conditions

Microcystis aeruginosa NIES 843 cultures were grown in CT medium (Ichimura 1971) modified to produce five separate nutrient conditions (in triplicate) and to allow for growth across

all conditions. Control conditions were standard CT medium, which contained 1.67 mM nitrate (KNO_3 and $Ca(NO_3)_2$) and 164 µM phosphate (Na_2 β-glycerophosphate). Preliminary experiments examining growth constraints on cultures at reduced nutrient concentrations established the limits of nutrient reduction that still allowed for comparable growth rates between treatments (Appendix Table 4.1). The P-reduced treatment contained control levels of nitrogen and 82 µM phosphate (Na_2 β-glycerophosphate). The three N-reduced treatments contained 178.9 µM of the nitrogen source and 164 µM phosphate (Na_2 β-glycerophosphate). The three nitrogen variants in the nutrient-limited medium tested were nitrate (denoted N-reduced (nitrate) hereafter), ammonium (NH_4Cl), and urea. To facilitate growth in the urea treatment, this medium was also supplemented with 100 nM NiSO4, as the urease enzyme requires Ni as a co-factor and this metal is not in the original medium (Mackerras and Smith 1986). Cultures were grown under 12:12 light:dark illumination of 110 µmols photons m^{-2} s^{-1} at 26º C. Cultures were passaged three times in their respective medium before being harvested during mid-exponential growth to ensure they were physiologically equilibrated to growth conditions. Due to nutrient constraints imposed by growth on reduced N and P, the N-reduced (ammonia) and P-reduced treatments could not reach an exponential growth rate that matched the control treatment (Appendix Table 4.1). All treatments except the N-reduced (ammonia) were able to reach cell densities to match the control (Appendix Table 4.1). Chlorophyll *a* fluorescence, used as a proxy for cell biomass, was measured daily on a Turner Designs TD-700 fluorometer (Sunnyvale, CA). To allow for extraction of sufficient nucleic acid material, triplicate cultures (150 mL) were grown in 500 mL, 1% HCl-washed flasks. These samples were harvested at the estimated point of maximum growth rate (Appendix Table 4.1). Cells were harvested by centrifugation for 15 minutes at 8,500 rpm. Samples were stored at -80º C prior to RNA extraction.

RNA Isolation, Sequencing, and Analysis

Total RNA was extracted from samples collected at the beginning of the light portion of the 12:12 hour light cycle with the MoBio PowerSoil® Total RNA Isolation kit (MO BIO Laboratories, Carlsbad, CA). Yields varied between treatments, with the N-reduced (urea and ammonia) treatments yielding less total RNA from the control treatment, however, reduced yields did not impact the total amount of mRNA sequenced (Appendix Table 4.1). Residual genomic

DNA was removed using the Ambion® Turbo DNA-free™ kit (Austin, TX, USA). Total RNA was kept at -80º C prior to sequencing. Initial RNA quality assessment was performed using a NanoDrop 1000 spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA). Subsequent sample processing, including rRNA subtraction, library prep, and sequencing were performed by the HudsonAlpha Genomic Services Laboratory (Huntsville, AL, USA). RNA quality was confirmed by quantification on a Qubit® Fluorometer (Life Technologies, Grand Island, NY, USA). The Ribo-Zero Magnetic Gold™ kit was employed for rRNA subtraction from all samples (Epicentre Biotechnologies, Madison, WI, USA). Library preparation was performed according to specifications provided by Illumina®, and samples were sequenced on the Illumina® MiSeq platform (San Diego, CA, USA).

Bioinformatic Analysis

All transcriptome libraries were aligned to the *Microcystis aeruginosa* sp. NIES 843 genome (Kaneko et al. 2007) using CLC Genomics Workbench v6.0 (CLC bio, Cambridge, MA, USA). Default recruitment parameters were used, allowing only two mismatches per alignment. Expression values were calculated as reads per kilobase per million (RPKM) (Mortazavi et al. 2008) and normalized (Wagner et al., 2012; Jalan et al., 2013; Feng et al., 2013) using the "By Totals" option in CLC Genomics Workbench using default parameters After normalization, Baggerly et al.'s (2003) weighted t-type test was used for all comparisons. Statistical cutoffs were established at $P < 0.05$ and P-values were corrected for false discovery rate (FDR). Due to current controversy surrounding use of RPKM (Dillies et al. 2012; Wagner et al., 2012), normalization methods were validated using a series of established housekeeping genes (Appendix Figure 1). The genome plot was generated using BRIG (Alikhan et al. 2011). Sequences are available in the NCBI short read archive (PRJNA229846). Taxonomic identities for non-*Microcystis* reads were generated using MetaPhlAn with the very sensitive local option for read-marker similarity (Segata et al. 2012).

Metabolome and Proteome Processing and Analysis

For metabolite analysis, treatments were extracted in 40:40:20 HPLC grade methanol, acetonitrile, water with 0.1% formic acid and analysis was performed via UPLC-Exactive Plus

Orbitrap mass spectrometer (Thermo Fisher) (Lu et al. 2010). Metabolites identities were determined in silico via exact mass and retention time matching using MAVEN, which also calculated area counts for each compound (Clasquin et al. 2002, Melamud et al. 2010). For proteomic measurements, samples were analyzed by 2D-LC-MS on an LTQ-Orbitrap-Velos following protein extraction and digestion, as described previously (Gobler et al. 2011). Cell pellets of 10 mg were lysed and proteins extracted in guanidine, yielding approximately 1 mg of protein per sample. ~100 µg of protein were loaded onto Strong Cation Exchange Chromatography (SCX) for processing. Resulting MS/MS spectra were searched against an *M. aeruginosa* database. Modified normalized spectral abundance factor (NSAF) values (Abraham et al. 2011, Zybailov et al. 2006) for identified transposases were extracted and visualized as a heat map using Perseus (http://www.perseus-framework.org/).

Chlorophyll *a* and Microcystin Estimates

Microcystin concentrations were measured for each culture according to previously established methods (Hotto et al. 2008). Chlorophyll *a* (a proxy for *Microcystis* biomass) measurements were performed using the fluorescence method for normalization of toxin and metabolome data analysis on an AU-10 fluorometer (Tuner Designs, Sunnyvale, CA) (Welschmeyer 1994).

Alkaline Phosphatase Activity (APA) Measurements

APase (E.C. 3.1.3.1.) activity measurement was performed in biological triplicate across all treatments with the procedure adapted for cyanobacterial culture assays from Ivanikova et al. (2007). A TD-700 fluorometer was used to make measurements (Turner Designs, Sunnyvale, CA, USA). APase activity was normalized to chlorophyll a measured as stated above. Student's t-test was used to perform statistical analysis.

RESULTS

Growth Characteristics and Extraction Yield

Growth rates of reduced nutrient treatments were lower than the control treatment (Appendix Table 4.1). Biomass after 10 days of growth was also reduced (83.1% and 22.3%,

respectively) for the N-reduced (ammonium) and P-reduced treatments compared to the control (Appendix Table 4.1). RNA yields from all reduced nutrient treatments were lower than the yields from control samples. Despite variable yields, 500 ng of each sample was used to construct sequence libraries (Appendix Table 4.1). Microcystin concentrations were below the detection limit for all samples.

Summary of sequencing output

On average, the sequence libraries contained a mean of ~1.1 million 50-bp reads, with > 99% of sequences free of ambiguous bases (Supplementary Table 4.1). An average of 65% of sequences stringently recruited to the *M. aeruginosa* NIES 843 genome. The majority of the remaining un-recruited reads (average 30.2% of un-recruited sequences) were annotated as various other *Microcystis* spp. (Supplementary Figure 4.2). These reads likely represent the fraction that could not be aligned to the template genome due to two base pair mismatches from the published sequence and were therefore excluded from the *M. aeruginosa* NIES 843 expression analysis. Co-occurring bacteria within these non-axenic cultures were also identified and classified: all reads not mapped to the NIES 843 genome by CLC Genomics were identified using MetaPhlAn. Most abundant phyla are shown in Appendix Figure 4.2; the N-reduced (urea) treatment differed most widely from the other samples, both in most abundance of *Proteobacteria* and the number of unaligned reads (Appendix Table 4.1).

Complete coverage of the *Microcystis* genome across all treatments is shown in Appendix Figure 4.3. The greatest number of reads predictably aligned to portions of the genome containing the 5S rRNA and *ssrA* (10Sa RNA) genes.

Global transcriptional response of *M. aeruginosa* NIES 843 to nutrient variability

Transcriptomes of *M. aeruginosa* NIES 843 grown under variable nutrient conditions exhibited significantly different mean global transcription profiles. When compared to the nutrient-rich control treatment, ~11% of genes (696) were significantly up-regulated under the N-reduced (urea) treatment (Figure 4.1A), while 1.5% (Figure 4.1B), 0.3% (Figure 4.1C), and 0.02% (Figure 4.1D) of genes were up-regulated under the N-reduced (ammonium), N-reduced (nitrate), and phosphate reduced conditions, respectively. Conversely, 4.1% of genes under N-

reduced (urea) (Figure 4.1A), 2.0% of genes under N-reduced (ammonium) (Figure 4.1B), 1.4% of genes under N-reduced (nitrate) (Figure 4.1C), and 1.1% of genes under low phosphorus (Figure 4.1D) were significantly down-regulated compared to the control treatment. Drastic global transcription differences existed between the three N-reduced treatments (Figure 4.2). For example, under the N-reduced (urea) treatment, over 2,000 genes were significantly up-regulated in direct comparisons to either the N-reduced (ammonium) or N-reduced (nitrate) treatments (Figure 4.2).

Metabolic response of *Microcystis* to nutrient limitation

We examined the expression of genes involved in N and P metabolism from the *Microcystis aeruginosa* NIES 843 genome (NC_010296.1) to determine changes associated with alterations in nutrient availability and chemistry. Seven genes involved in P metabolism and transport exhibited significant differential expression from the control under at least one low nutrient treatment. Expression of alkaline phosphatase (MAE_16640) was up-regulated under three out of four low nutrient treatments, although significant up-regulation was observed only for the N-reduced (urea) treatment (Figure 4.3A). Three of four genes within the pathway (phosphate permease MAE_09270, alkaline phosphatase, and *phoU* MAE_52660) experienced significantly increased expression compared to the control under the N-reduced (urea) treatment (Figure 4.3a). Alkaline phosphatase activity under these conditions was independently verified using an enzymatic assay, with only the N-reduced (urea) treatment exhibiting increased activity (Appendix Table 4.1). Genes involved in urea metabolism largely exhibited reduced expression across all treatments when compared to the control treatment (Figure 4.3B). However, significant up-regulation occurred for two of the three structural genes (*ureA-B*) of the urease enzyme

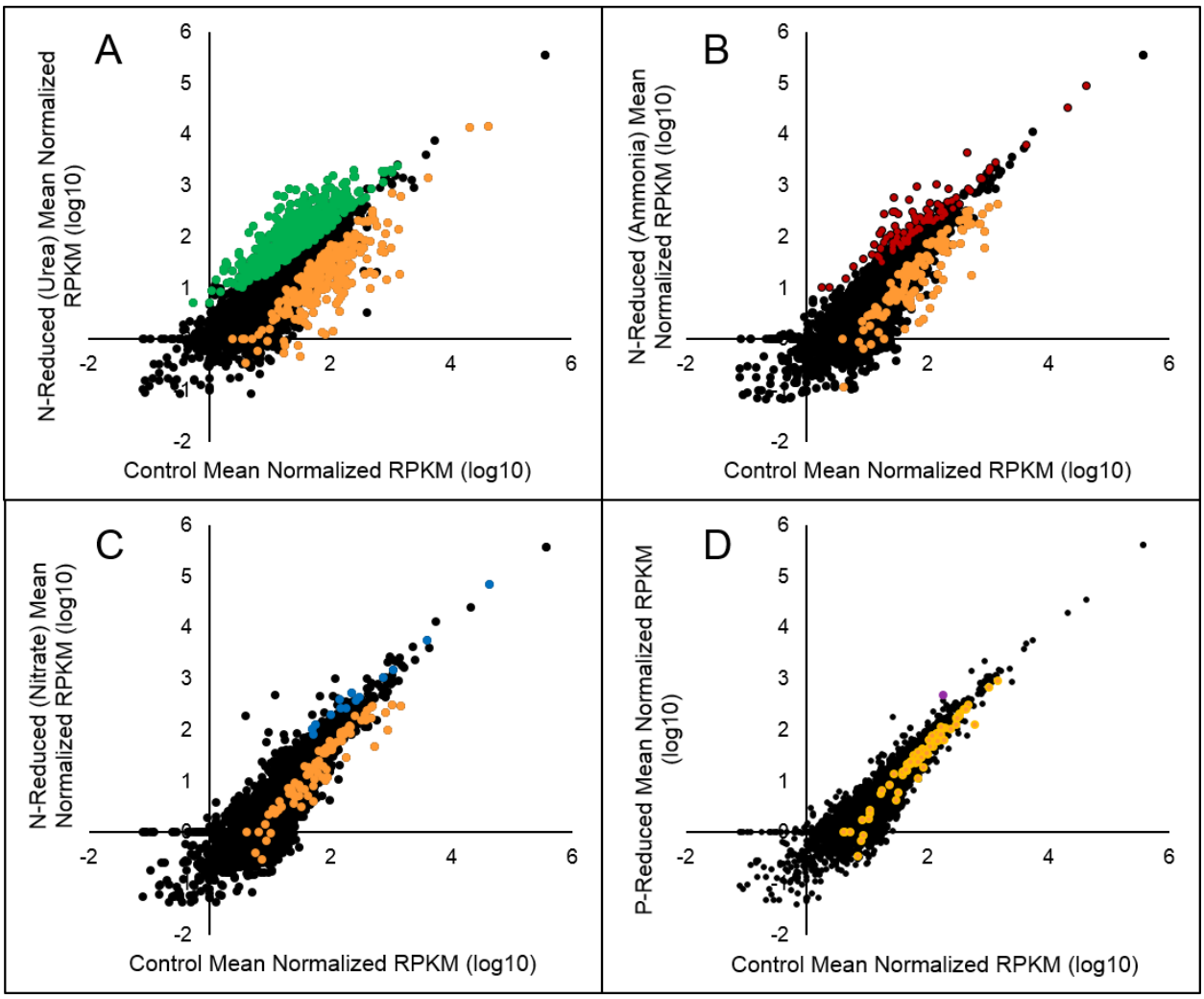

Figure 4.1. Comparison of global genomic expression between low nutrient treatments and the control treatment. Black points represent expression values of the 6364 genes of *M. aeruginosa* NIES 843. Orange points represent those genes that exhibit significantly decreased expression in the low nutrient treatment transcriptome when compared to the control at $P < 0.05$ (corrected for false discovery rate). Other colored points represent significant upregulation of expression in the low nutrient treatment compared to the control at $P < 0.05$ (corrected for false discovery rate). A) Expression of N-reduced (urea) vs. control treatments; B) Expression of N-reduced (ammonium) vs. control treatments; C) Expression of N-reduced (nitrate) vs. control treatments; D) Expression of P-reduced vs. control treatments

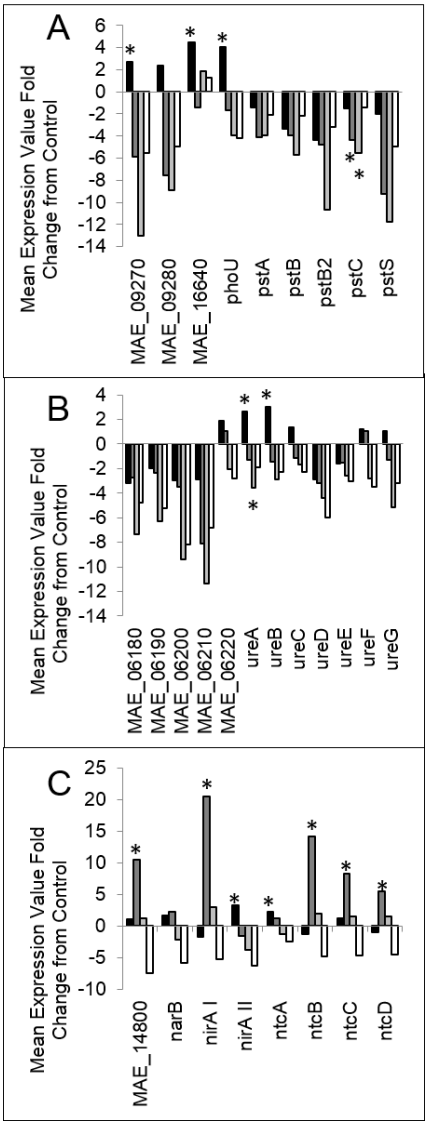

Figure 4.2. Fold change to expression in comparison to control treatment of genes involved in N and P metabolism. * indicates significant difference in expression value (RPKM) from control treatment at $P < 0.05$ (FDR corrected). A) Genes involved in phosphorus transport and metabolism; B) genes involved in urea transport and metabolism; C) genes involved in nitrate transport and metabolism. Black bars represent expression fold change of N-reduced (urea), dark grey represent N-reduced (ammonium), light grey represent N-reduced (nitrate), and white represent P-reduced treatments.

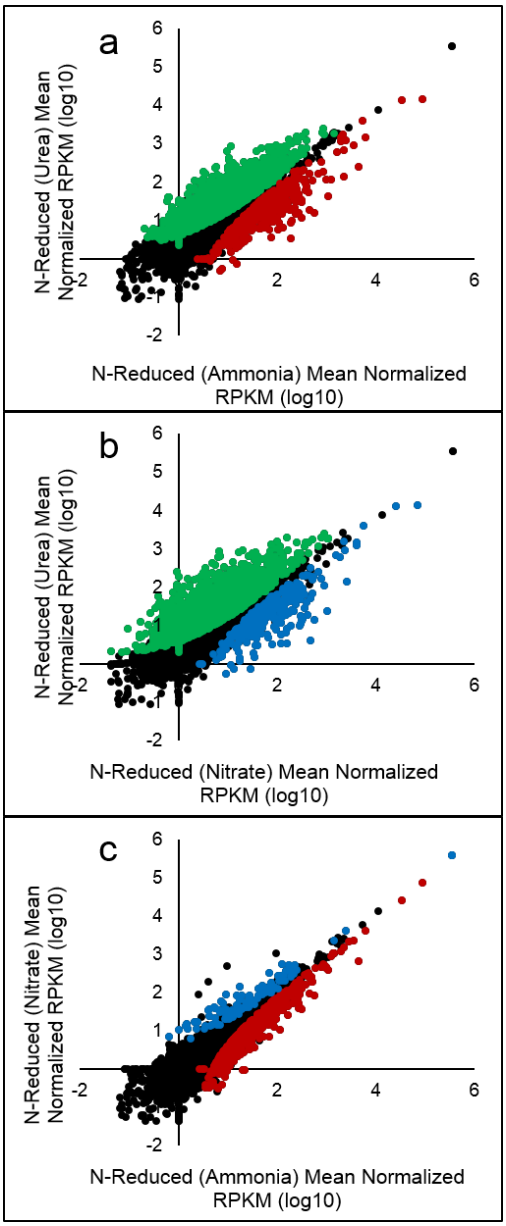

Figure 4.3. Comparison of global genomic expression between low nitrogen treatments. Black points represent expression values of the 6364 genes of *M. aeruginosa* NIES 843. Colored points represent significant upregulation of expression in the specified low nitrogen treatment compared to the second at $P < 0.05$. A) Expression of N-reduced (urea) (green) vs. N-reduced (ammonium) (red) treatments; B) Expression of N-reduced (urea) (green) vs. N-reduced (nitrate) (blue) treatments; C) Expression of N-reduced (ammonium) (red) vs. N-reduced (nitrate) (blue) treatments.

complex (Figure 4.3B) in the N-reduced (urea) treatment. A gene involved in nitrate transport (MAE_14800) was significantly up-regulated under the N-reduced (ammonium) treatment (Figure 4.3B). This transcriptional response was not emulated in the N-reduced (nitrate) treatment, however, which only exhibited significant up-regulation of *nirA*. Expression of the global nitrogen regulator gene, *ntcA* (MAE_01830) and the PII nitrogen regulator, *glnB*, were significantly different from the control under the N-reduced (urea) treatment alone (Figure 4.3B).

The water-soluble metabolome (i.e., the pool of small molecules) within cells grown under these conditions was measured in parallel with each transcriptome. After comparison to the control cultures, relatively little variation between metabolomes among the N-reduced (urea) and other low N treatments were observed (Appendix Figure 4.4). Outliers include NADP+, guanosine, and NADPH (Figure 4.4; Appendix Figure 4.4). This pattern is echoed within pathways involved in N metabolism (Figure 4.4). In comparison to the control, metabolite pools largely respond in two ways: 1) N-reduced (ammonium and nitrate) and P-reduced treatments result in a significant increase in N-acetyl-ornithine, 2-oxoglutarate, and malate) the N-reduced (nitrate) treated cells showed significant accumulation of certain metabolites (arginine, citrate, glutamate, glutamine, isocitrate) in comparison to the control treatment (Figure 4.4).

Nutrients and toxin production by *Microcystis*

The biosynthetic cassette encoding for the biosynthesis of microcystin toxin spans 55 kb of the genome and is composed of ten genes. Transcription of all ten genes was reduced in each low nutrient treatment compared to the control (Appendix Figure 4.5). Functional assessment of toxin production was attempted by measurement of toxin concentration in each sample, but levels were below detection limits (data not shown). The genes involved in synthesis of cyanopeptolin (*mcnA-G*) and aeruginosin (*aerA-N*) did not exhibit significant up-regulation under any growth treatment.

Insights into novel genetic elements of *Microcystis*

Approximately 9.6% of the *Microcystis aeruginosa* NIES 843 genome is comprised of genes encoding transposases (Kaneko et al 2007). Of these 610 genes, stark differences in transcription patterns were observed between treatments (Figure 4.5). Cells grown under N-

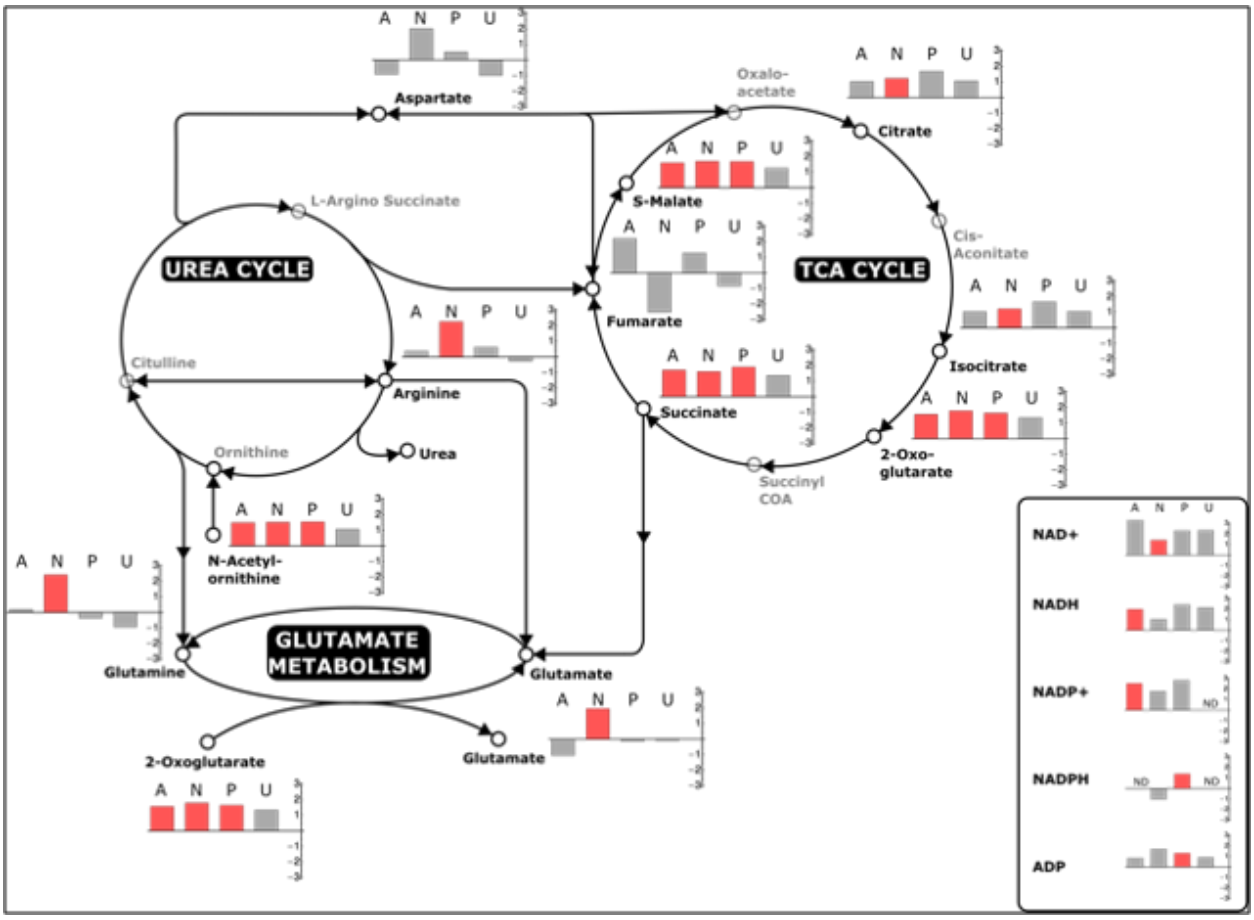

Figure 4.4. Fold changes of treatment metabolome samples compared to control samples displayed for relevant pathways. A) N-reduced (ammonium), N) N-reduced (nitrate), P) P-reduced, and U) N-reduced (urea) treatments to control. Bars represent $\log_2$ transformed fold intensity and directionality. Red indicates significant changes ($P < 0.05$).

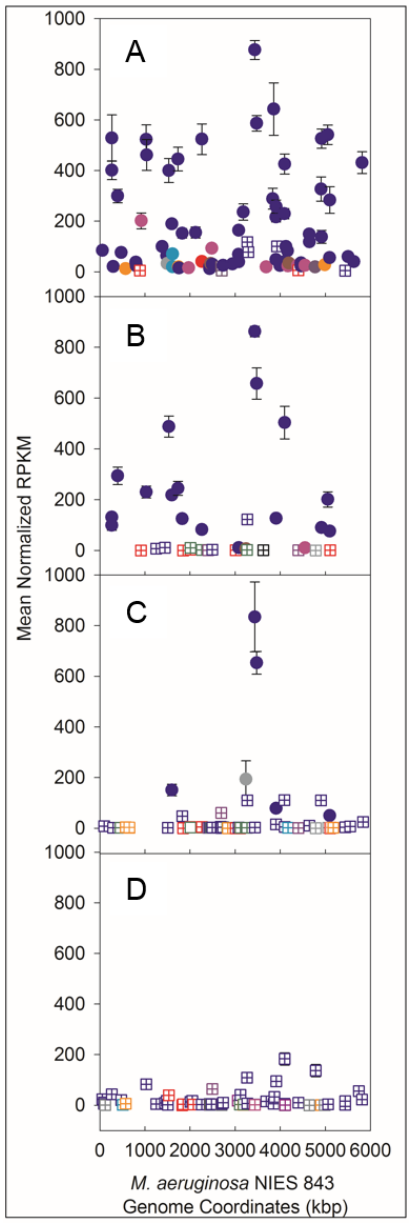

Figure 4.5. Mean normalized expression values (RPKM) of *M. aeruginosa* NIES 843 transposase genes with significantly ($P < 0.05$) different expression from control treatment. Error bars represent standard error of the mean of biological triplicates of A) N-reduced (urea); B) N-reduced (ammonium); C) N-reduced (nitrate); D) P-reduced treatments. Circles indicate increased and squares decreased expression from the control treatment. Colors represent transposase family: (●IS200/IS605), (●IS630), (●ISAzo13), (●IS1634), (●IS110), (●ISAs1), (●IS1), (●ISL3), (●IS5), (●IS701), (● IS481).

reduction with urea as the sole N source exhibited significantly increased and reproducible (mean coefficient of variation = 18.2%) transcription of 12% of these transposase genes (Figure 4.5A), while the N-reduced (ammonium) treatment showed 5.6% of transposase genes significantly up-regulated (Figure 4.5B) prior to FDR analysis, which suggested the potential for 11 and 10 false positives, respectively. Exposure to P-reduced conditions had the opposite effect on transposase gene transcription; 58 genes were significantly reduced in transcription compared to the control (Figure 4.5D). These observations were validated by proteomics, which showed a similar pattern of increased transposase expression during N-reduction (nitrate) and a quieting of this expression during P-reduction (Appendix Figure 4.6), although overlap of differential expression of specific genes from the control treatment was limited to two genes for the N-reduced and three for the P-reduced.

Another anomaly within the *M. aeruginosa* NIES 843 genome is the existence of nine genes that have been previously annotated as putative nitrogen fixation elements (Kaneko et al 2007): we consider these genes an anomaly as there is no evidence that this organism fixes nitrogen. Differential transcription of these genes was observed when comparing the mean transcriptomes of the five nutrient treatments (Table 4.1). Greatest up-regulation of transcription occurred when cells were grown on N-reduced (urea). When comparing the N-reduced (urea) culture transcriptomes to the other two N-reduced treatments, up-regulation of $\geq$ 20-times occurred for transcription of multiple genes annotated as being involved in N-fixation (Table 4.1). Transcription of a single gene (MAE_15560), encoding a NifU-like protein, was greater in N-reduced (ammonium) and N-reduced (nitrate) than in the N-reduced (urea) treatment. Compared to the P-reduced treatment, six of these genes experienced significantly increased transcription when grown in N-reduced (urea) conditions (Table 4.1).

DISCUSSION

The insights gained from decades of research attempting to identify factors that drive cyanobacterial harmful algal blooms remain limited. While factors such as temperature and nutrient availability have been implicated, their individual roles as drivers of freshwater blooms and phytoplankton speciation remain hotly debated (Paterson et al 2011, Scott and McCarthy 2010). We applied RNA-seq to replicate *Microcystis aeruginosa* cultures to study the system-

Table 4.1. Fold change different relative to the N-reduced (urea) treatment of genes annotated as having a role in N-fixation function.

Gene	Function	Control	N-Reduced (Ammonium)	N-Reduced (Nitrate)	P-Reduced
MAE_14230	Sigma factor potentially involved in expression of N_2 fixation genes	5.9	4.9	6.1	6.1
MAE_14860	Putative transcription factor for heterocyst differentiation	-	21.8	56.4	52.5
MAE_15560	NifU-like protein	-	-1.2	-1.4	-
MAE_16560	Nitrogen regulation protein NifR3-like protein	-	11.5	22.1	9.8
MAE_16580	Nitrogen regulation protein NifR3-like protein	2.7	16.5	13.0	33.6
MAE_18160	Mo-dependent nitrogenase-like	-	1.7	-	-
MAE_27840	Heterocyst glycolipid synthase	5.6	19.5	30.4	38.0
MAE_31880	Mo-dependent nitrogenase-like	2.2	2.7	4.5	9.6
MAE_38090	NifY-like protein	-	2.7	4.4	5.5
nifS	NifS cysteine desulfurase	-	-	2.3	-
sigD	Sigma factor potentially involved in expression of N_2 fixation genes and light/dark adaptation	2.8	4.3	8.4	8.3

wide molecular response of this organism to alterations in nutrient concentration and chemistry. Our results demonstrate that widespread transcriptional changes are needed to maintain metabolic homeostasis within cultured *Microcystis* in response to nutrient variability. Moreover, significant and reproducible up-regulation of transcription of transposases accompanied specific nutrient treatments, implying that anthropogenic nutrient loads may not only constrain biomass, but could potentially shape the genomic architecture and ultimately the evolutionary trajectory of *Microcystis* populations. We discuss these results in the context of the ecology of one of the world's most common freshwater harmful algal species.

Insights into physiological responses of *Microcystis* to nutrient limitation

Genome wide changes to gene transcription were observed when cultures were exposed to N-reduced as well as P-reduced conditions. Up to ~20 % of the annotated genes in nutrient-limited treatments were differentially transcribed compared to the control, suggesting expansive physiological responses to nutrient variability. Harke and Gobler (2013) demonstrated that additional strains of *Microcystis* also exhibit massive differential transcription patterns (~32%) in response to reduced N conditions. The most extensive changes in the current study were observed for cultures grown in N-reduced (urea) conditions (Figure 4.1A). Unexpectedly, the differences between the N-reduced (urea) treatment and the alternative N-reduced treatments were even more extreme (Figure 4.2). The changes in global transcription patterns suggest that the transcriptional response to N-reduction by *Microcystis* is dependent on the chemical species of nitrogen available to the cell. In contrast with N-reduced treatments, relatively limited transcriptional change was observed under growth on P-reduction relative to the controls. This limited response indicates that a more extreme reduction in P concentrations is necessary to up-regulate physiological P stress in *Microcystis* cells. Capturing this response using advanced molecular tools such as RNA-seq represents a developing area of interest in HAB research (Harke and Gober 2013, Wurch et al. 2011).

Genes involved in N metabolism exhibited anticipated responses to nutrient limitation. Specifically, genes involved in urea metabolism (*ureA-B*) were significantly up-regulated during growth on urea (Figure 4.3B). This up-regulation is supportive of the ability of *Microcystis* to use urea as its sole N source and indicates the use of the urease enzyme for the breakdown of the

compound. This observation is salient to current agricultural trends, as the application of urea fertilizers is rapidly increasingly worldwide, resulting in elevated levels of urea in agricultural runoff into aquatic systems (Glibert et al. 2006). Surprisingly, a significant up-regulation of genes involved in phosphate transport and metabolism occurred under the N-reduced (urea) treatment and was not seen during P-reduction (Figure 4.3B). The differences in alkaline phosphatase transcription and activity under N-reduced (urea) limitation were particularly unexpected, as this gene is commonly used as an environmental marker for phosphorus limitation (Berman 1970, DeBruyn et al. 2004). The lack of up-regulation under the P-reduced treatment is potentially a result of experimental conditions designed to up-regulate P stress not being limiting enough. In the case of *Microcystis*, deeper P-limiting conditions may be required for widespread of up-regulation of genes involved in phosphate metabolism, including APase (Harke and Gober 2013, Healey and Hendzel 1979). Previous studies have found that APase activity is independent of phosphate concentrations when they are above 0.2 µM in the environment (Nausch 1998), much lower than those used in this study, and may in fact be linked to cellular N:C ratios in phytoplankton (Myklestad and Sakshaug 1983, Neddermann and Nausch 2004).

Because the N-reduced (ammonia) treatment was not able to reach cell densities comparable to the other treatments, transcriptional responses unique to this treatment may be reflective of a response to differences in cell densities, rather than nitrogen conditions. Additionally, to determine whether supplementation of the N-reduced (urea) treatment with NiSO4 potentially confounded results, transcription patterns of genes involved in construction of enzymes requiring a nickel cofactor were examined. Only *hoxH* of the bidirectional hydrogenase (*hoxEFHUY*) and *ureAB* (*ureA-G*) showed significantly increased transcription in the N-reduced (urea) treatment compared to the control. Genes involved in the construction of the uptake hydrogenase (*hypA-F*), lactoylglutathionine (MAE_23250), and coenzyme F420 hydrogenase (MAE_40060) showed no difference between the urea and control treatments. Notably, the genes involved in nickel incorporation for urease (*ureE*) and hydrogenase (MAE_57880) showed no significant difference between the urea treatment and the control. Overall, these caveats highlight an important consideration often missed in single variable experiments: metabolic pathways and genetic pathways do not function in isolation, but are tightly coupled. Along with the above considerations (i.e., cell density effects on light-field, effects of growth using urea on Ni use and

transport), there are innumerable of other traits (e.g., Fe use and uptake for NO_3 assimilation, Zn use and uptake associated with alkaline phosphatase, the influence of nutrient chemistry on co-occurring / contaminating microbes, etc.) that can "confound" observations in studies with microbial isolates, but are important considerations for such culture studies. While it is likely that many of these effects are secondary to the more dramatic influences of shifts in nutrient availability, they provide a series of caveats that need be remembered in interpretation of our data.

The transcriptome and metabolome capture distinct levels of physiological response to environmental change, but do not always overlap. This is in part due to differences in turnover times for mRNA versus small metabolites (Spura et al. 2009, Steglich et al. 2010). Moreover, it is likely that many of the transcriptional responses are occurring in order to stabilize cellular metabolism due to environmental change. We therefore must exercise caution when comparing these measurements. Additionally, metabolite biosynthesis and accumulation can become convoluted, as expression of multiple pathways can result in metabolic alterations that involve a single compound.

It is, however, possible to draw certain comparisons. For example, accumulation of lysine, an essential amino acid, occurred in cells grown in N-reduction (urea) (Appendix Figure 4.4). This is also reflected in the transcript data, with a gene involved in lysine biosynthesis (*dapB*) significantly up-regulated compared to the control treatment. When looking across the pathways highlighted in Figure 4.4, relatively few changes to metabolite pools can be observed between nutrient limited treatments, especially in comparison to those occurring between the transcriptomes (Figure 4.1; Figure 4.2). These differences between the transcriptome and metabolome suggest that widespread transcriptional changes are necessary to maintain metabolic homeostasis when cells are exposed to nutrient limitation.

Interactions with co-occurring heterotrophs

Cultures used in this study were non-axenic, and therefore interactions with the co-occurring heterotrophic bacterial community have the potential to indirectly shape gene transcription response to nutrient concentration. Of note, the drastic transcriptional response observed in the N-reduced (urea) treatment coincided with the greatest number of un-recruited sequences to the NIES 843 genome (Appendix Table 4.2). Heterotrophic transformation of

organic N compounds, such as urea, to inorganic N has been shown to play an important role in the utilization of organic nitrogen by cyanobacteria in aquatic environments (Berman et al. 1999, Purvina et al. 2010). The active heterotrophic population in the cultures examined in this study may be fulfilling that role for *Microcystis*, as has been observed previously (Purvina et al 2010). The contribution to urea metabolism by the heterotrophic population may explain why genes encoding urea transporters did not exhibit significantly increased transcription in the N-reduced (urea) treatment in *Microcystis* cells, as it may have been metabolized by contaminating heterotrophs prior to incorporation by *Microcystis* cells. More broadly, the transcriptionally active heterotrophic population in the N-reduced (urea) treatment may have had other indirect impacts on the global transcriptional response of *Microcystis* (e.g., competition for Ni) and could possibly explain at least a portion of the differences between the N-reduced (urea) and the alternative treatments.

Genomic consequences of nutrient conditions

The genomic architecture of *Microcystis aeruginosa* is considered to be among the most plastic of all cyanobacteria. Comprising ~10% of the entire genome, the number of insertion sequence elements of *M. aeruginosa* outnumbers those found in the other bloom formers, including *Nostoc punctiforme* (2.0%), *Anabaena variabilis* (1.4%), and *Cylindrospermopsis raciborskii* (0.9%) (Frangeul et al. 2008, Humbert et al. 2013, Kaneko et al 2007, Lin et al. 2011). Indeed the presence of repetitive sequences and associated transposase genes have likely made successful closure of sequenced genome sequences challenging (Frangeul et al 2008). The biological causes and consequences of this unique genomic architecture, however, remain unexplained.

First reported in maize in 1950 (McClintock 1950), transposable elements have been identified in genomes across all domains of life and more recently been implicated in not only driving evolution, but also epigenetics of eukaryotic organisms (Slotkin and Martienssen 2007). To date, the factors driving transposase expression have not been well defined. The data presented herein suggest that nutrient availability and chemistry may play a role in regulating activity for 11 of the 21 major insertion sequence (ISE) and associated transposase families found in the *M. aeruginosa* genome (Kaneko et al 2007). Significantly increased expression occurred most

frequently for transposases from the family IS200/IS605, suggesting not only that nutrient limitation may regulate transposase gene expression, but that it specifically targets expression of a single family. This family of transposases is also widespread in the genomes of multiple pathogenic bacteria (Gibert et al. 1990): similar behavior of these elements among pathogens and proteomic identification of transposases in endosymbionts and an acid mine drainage community may have ramifications for the ability of such organisms to infect a host or adapt to an environment (Chao et al. 1983, Kleiner et al. 2013, Ram et al. 2005). Homologues of the IS200/605 family have been identified across both eukaryotes and eukaryote-infecting viruses, suggesting these results may have implications in all domains of life (Bao and Jurka 2013).

Somewhat surprisingly, prior transcriptome (both microarray and RNA-seq) analyses of *Microcystis* spp. transcriptomics make no reference to differential expression of transposases across light/dark and nutrient exposure treatments (Harke and Gober 2013, Straub et al. 2011). A re-examination of previously published data for *Microcystis* strain LE-3 grown under N-reduced (nitrate) conditions showed 127 transposase genes differentially expressed, with 122 showing increased expression relative to the control (Harke and Gober 2013). Other treatments in that study were unfortunately too dissimilar to compare with our work due to differences in nutrient concentrations and chemistries.

The abundance and expression of transposases within the genome of *M. aeruginosa* NIES 843 could have a substantial impact on genomic architecture of individual cells (Figure 4.6). Indeed, there are only four expansive regions (>150 kb) where no transposase genes exist (Figure 4.6 I-IV; Appendix Table 4.3). Within these regions are large numbers of hypothetical proteins, genes involved in cobalamin biosynthesis, nitrogen transport and metabolism (Figure 4.6 I-III), and DNA and RNA polymerase subunit proteins (Figure 4.6 II-III). Region II contains all genes encoding for the 5S, 16S, and 23S ribosomal subunits and ten genes involved in phosphate transport (Figure 4.6). As many of these genes are considered essential for viability, it is not surprising that they exist in genome regions free from interference by active ISEs.

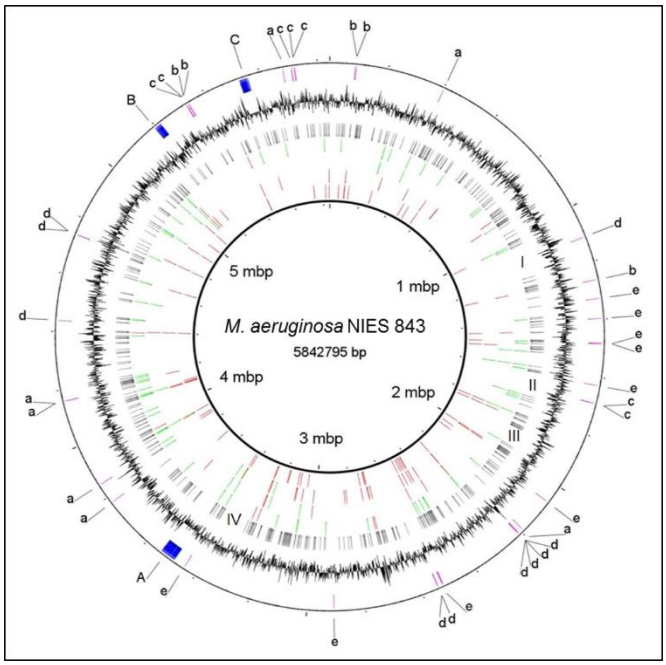

Figure 4.6. Genome map of *M. aeruginosa* NIES 843. Inner four rings show locations of significantly increased (green) or decreased (red) expression of transposase genes for the P-reduced (innermost), N-reduced (nitrate), N-reduced (ammonium), and N-reduced (urea) treatments. Location of all transposase genes in the genome in ring five and GC content is located in ring six. The outermost ring shows locations of *ureA-G* (a), *hoxF-Y* (b), *hypA-F* (c) , *cpcA1-G* (d), and putative *nif* genes (e); *mcyA-J* (A), *aerD-L* (B), *mcnA-G* (C). Gaps in transposase placement are highlighted as I-IV.

The genome of *M. aeruginosa* is also unique in that many genes, which are canonically organized in other cyanobacteria as single operons (Wang et al. 2012), are scattered throughout as individual open reading frames (Figure 4.6a-e; Appendix Table 4.3). Examples include genes involved in construction of the urease enzyme complex (*ureA-G*; Figure 4.6a), both (*hoxEFHUY*; Figure 4.6b; *hypA-F*; Figure 4.6c) hydrogenases, phycocyanin synthesis genes (*cpcA1-G*; Figure 4.6d), and even the genes annotated as being involved in nitrogen fixation (Figure 4.6e). The mosaic and highly plastic nature of the *M. aeruginosa* NIES 843 genome is suggestive of the ability to produce subpopulations of cells with alternate genomic architectures, and potentially evolutionary trajectories, in response to environmental conditions.

In contrast, genes involved in synthesis of non-ribosomal secondary metabolites appear to have maintained the operon architecture observed in other cyanobacterial genomes (Figure 4.6A-C). Regulation of synthetase genes for microcystins (Figure 4.6A), aeruginosins (Figure 4.6B), and cyanopeptolins (Figure 4.6C) (Ishida et al. 2009, Kaneko et al 2007) is likely tied to a need for juxtaposition of these gene products to each other for proper synthesis: rearrangements leading to a loss of function likely means the entire cassette would be quickly lost from the genome. In part this may explain the absence of these capabilities in some isolates: while intact, the genetic cassettes support selection for organisms producing these compounds, allowing the persistence of these capabilities within the genome. Rearrangement of these genes (which need to be proximal for product assembly) removes the positive selection that maintains the rest of the cassette. Indeed, within the presented data, the most highly expressed transposases for the N-reduced treatments were those located proximal to the microcystin cassette, which is present in only a subset of natural communities (Rinta-Kanto et al. 2009). Similar activity of transposable elements has been implicated in the loss of microcystin biosynthesis genes in *Planktothrix* strains (Christiansen et al. 2008).

Ecologically, these drivers of genomic architecture may be exacerbated by changing land use practices: urea has become the dominant nitrogenous fertilizer during the past three decades – a period where we have seen expansive proliferations of *Microcystis* spp. (Glibert et al 2006). Moreover, the potential role of nutrients as a regulatory mechanism for transposase expression may apply more broadly across microbial life given the wide distribution of the IS200/IS605 family. Indeed, a limited number of other organisms (both bacteria and archaea) have been shown

to experience increased transposase expression under various environmental conditions, including N variability (Hewson et al. 2009, Jäger et al. 2009). This suggests that anthropogenic nutrient loading may not just shape the biomass and evolutionary trajectories of single organisms, but entire microbial communities.

Our observations illustrate a possible evolutionary response which would allow for survival and dominance across the range of freshwater systems *Microcystis* inhabits. Because these data arose from an exploratory transcriptomic study, it highlights the utility of so-called "-omics" tools for hypothesis development for a system of interest. Future testing of these hypotheses are necessary to resolve how biogeochemical conditions can drive evolution (i.e., genomic rearrangements) in *Microcystis*, as increased transcription of transposase genes may not translate to enzymatic activity or genomic rearrangements (Schmitz-Esser et al. 2011). Moreover, these hypotheses require particular testing in field studies with natural populations, as variations in natural communities may shed further light on how biogeochemistry shapes not only community structure but evolutionary trajectory.

ACKNOWLEDGEMENTS

We thank Dr. GL Boyer, Dr. LJ Hauser, Dr. NC VerBerkmoes, Dr. RL Hettich, Dr. GS Bullerjahn and Dr. RML McKay for support, meaningful ideas and discussion. We also thank Shafer Belisle, Chad Effler, and Justine Schmidt for their assistance. Finally, we thank three anonymous reviewers for their constructive and helpful comments during the review process. This project was supported by grants from the National Science Foundation (IOS - 0841918, DEB – 1240870 (to SWW) and OCE-1233964 (SRC and SPD)) and a UT/ORNL Science Alliance JDRD award to SWW. MMS was supported by a Wallace-Dean fellowship from the University of Tennessee.

LIST OF REFERENCES

Abraham P, Adams R, Giannone RJ, Kalluri U, Ranjan P, Erickson B *et al* (2011). Defining the boundaries and characterizing the landscape of functional genome expression in vascular tissues of *Populus* using shotgun proteomics. *J Proteome Res* **11:** 449-460.

Alikhan N-F, Petty N, Ben Zakour N, Beatson S (2011). BLAST ring image generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12:** 402.

Baggerly KA, Deng L, Morris JS, Aldaz CM (2003). Differential expression in SAGE: accounting for normal between-library variation. *Bioinformatics* **19:** 1477-1483.

Bao W, Jurka J (2013). Homologues of bacterial TnpB_IS605 are widespread in diverse eukaryotic transposable elements. *Mobile DNA* **4:** 12.

Berman T (1970). Alkaline phosphatases and phosphorus availability in Lake Kinneret. *Limnol Oceanogr* **15:** 663-674.

Berman T, BÃ©chemin C, Maestrini SY (1999). Release of ammonium and urea from dissolved organic nitrogen in aquatic ecosystems. *Aquat Microb Ecol* **16:** 295-302.

Chao L, Vargas C, Spear BB, Cox EC (1983). Transposable elements as mutator genes in evolution. *Nature* **303:** 633-635.

Christiansen G, Molitor C, Philmus B, Kurmayer R (2008). Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Mol Biol Evol* **25:** 1695-1704.

Clasquin MF, Melamud E, Rabinowitz JD (2002). LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Current Protocols in Bioinformatics*. John Wiley & Sons, Inc.

Davis T, Harke MJ, Marcoval MA, Goleski, J *et al* (2010). Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Aquat Microb Ecol* **61:** 149-162.

DeBruyn JM, Leigh-Bell JA, McKay RML, Bourbonniere RA, Wilhelm SW (2004). Microbial distributions and the impact of phosphorus on bacterial activity in Lake Erie. *J Great Lak Res* **30:** 166-183.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N *et al* (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform* **14**: 671-683

Dodds WK, Bouska WW, Eitzmann JL, Pilger TJ, Pitts KL, Riley AJ *et al* (2008). Eutrophication of U.S. freshwaters: analysis of potential economic damages. *Environ Sci Tech* **43:** 12-19.

Dolman AM, Rücker J, Pick FR, Fastner J, Rohrlack T, Mischke U *et al* (2012). Cyanobacteria and cyanotoxins: the influence of nitrogen versus phosphorus. *PLoS ONE* **7:** e38757.

Donald DB, Bogard MJ, Finlay K, Leavitt PR (2011). Comparative effects of urea, ammonium, and nitrate on phytoplankton abundance, community composition, and toxicity in hypereutrophic freshwaters. *Limnol Oceanogr* **56:** 2161-2175.

Downing JA, Watson SB, McCauley E (2001). Predicting cyanobacteria dominance in lakes. *Can J Fish Aquat Sci* **58:** 1905-1908.

Feng, S, Euker, TP, Holly, MK, Konkel, ME, Lu, X, Wang, S (2013). Investigating the responses of *Cronobacter sakazakii* to garlic-derived organosulfur compounds: a systematic study of pathogenic-bacterium injury by use of high-throughput whole-transcriptome sequencing and conocal micro-raman spectroscopy. *Appl Environ Micro* **80:** 959-971.

Frangeul L, Quillardet P, Castets A-M, Humbert J-F, Matthijs H, Cortez D *et al* (2008). Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. *BMC Genomics* **9:** 274.

Gibert I, Barbé J, Casadesús J (1990). Distribution of insertion sequence IS200 in *Salmonella* and *Shigella*. *J Gen Microbiol* **136:** 2555-2560.

Glibert P, Harrison J, Heil C, Seitzinger S (2006). Escalating worldwide use of urea – a global change contributing to coastal eutrophication. *Biogeochemistry* **77:** 441-463.

Gobler CJ, Berry DL, Dyhrman ST, Wilhelm SW, Salamov A, Lobanov AV *et al* (2011). Niche of harmful alga *Aureococcus anophagefferens* revealed through ecogenomics. *Proc Nat Acad Sci USA* **108:** 4352-4357.

Harke M, Berry D, Ammerman J, Gobler C (2012). Molecular response of the bloom-forming cyanobacterium, *Microcystis aeruginosa*, to phosphorus limitation. *Microb Ecol* **63:** 188-198.

Harke M, Gober C (2013). Global transcriptional responses of the toxic cyanobacterium, *Microcystis aeruginosa*, to nitrogen stress, phosphorus stress, and growth on organic matter. *PLoS ONE* **8**.

Healey FP, Hendzel LL (1979). Indicators of phosphorus and nitrogen deficiency in five algae in culture. *J Fish Res Board Can* **36:** 1364-1369.

Hewson I, Poretsky RS, Beinart RA, White AE, Shi T, Bench SR *et al* (2009). In situ transcriptomic analysis of the globally important keystone N_2-fixing taxon *Crocosphaera watsonii*. *ISME J* **3:** 618-631.

Hotto AM, Satchwell MF, Berry DL, Gobler CJ, Boyer GL (2008). Spatial and temporal diversity of microcystins and microcystin-producing genotypes in Oneida Lake, NY. *Harmful Algae* **7:** 671-681.

Humbert J-F, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T *et al* (2013). A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PLoS ONE* **8:** e70747.

Ichimura T (1971). Sexual cell division and conjugation-papilla formation in sexual reproduction of Closterium stringosum. *Proceedings of the 7th international seaweed symposium, 1971*.

Ishida K, Welker M, Christiansen G, Cadel-Six S, Bouchier C, Dittmann E *et al* (2009). Plasticity and evolution of aeruginosin biosynthesis in cyanobacteria. *Appl Environ Microbiol* **75:** 2017-2026.

Ivanikova NV, McKay RML, Bullerjahn GS, Sterner RW (2007). Nitrate utilization by phytoplankton in Lake Superior is impaired by nutrient (P, Fe) availability and season light limitation - a cyanobacterial bioreporter study. *J Phycol* **43:** 475-484.

Jäger D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA (2009). Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *P Natl Acad Sci USA* **106:** 21878-21882.

Jalan, N, Kumar, D, Andrade, MO, Yu, F, Jones, JB, Graham, JH, White, FF, Setubal, JC, Wang, N (2013). Comparative genomic and transcriptome analyses of pathotypes of *Xanthomonas citri* subsp. *citri* provide insights into mechanisms of bacterial virulence and host range. *BMC Genomics* **14:** 551.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M *et al* (2007). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Research* **14:** 247-256.

Kleiner M, Young JC, Shah M, VerBerkmoes NC, Dubilier N (2013). Metaproteomics reveals abundant transposase expression in mutualistic endosymbionts. *mBio* **4**. e00223-13

Lin S, Haas S, Zemojtel T, Xiao P, Vingron M, Li R (2011). Genome-wide comparison of cyanobacterial transposable elements, potential genetic diversity indicators. *Gene* **473:** 139-149.

Lu WY, Clasquin MF, Melamud E, Amador-Noguez D, Caudy AA, Rabinowitz JD (2010). Metabolomic analysis via reversed-phase ion-pairing liquid chromatography coupled to a stand alone Orbitrap mass spectrometer. *Anal Chem* **82:** 3212-3221.

Mackerras AH, Smith GD (1986). Urease activity of the cyanobacterium *Anabaena cylindrica*. *J Gen Microbiol* **132:** 2749-2752.

McClintock B (1950). The origin and behavior of mutable loci in maize. *Proc Nat Acad Sci USA* **36:** 344-355.

Melamud E, Vastag L, Rabinowitz JD (2010). Metabolomic analysis and visualization engine for LC-MS data. *Anal Chem* **82:** 9818-9826.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Meth* **5:** 621-628.

Myklestad S, Sakshaug E (1983). Alkaline phosphatase activity of *Skeletonema costatum* populations in the Trondheimsfjord. *J Plank Res* **5:** 557-564.

Nausch M (1998). Alkaline phosphatase activities and the relationship to inorganic phosphate in the Pomeranian Bight (southern Baltic Sea). *Aquat Microb Ecol* **16:** 87-94.

Neddermann K, Nausch M (2004). Effects of organic and inorganic nitrogen compounds on the activity of bacterial alkaline phosphatase. *Aquat Ecol* **38:** 475-484.

Paerl HW, Xu H, McCarthy MJ, Zhu G, Qin B, Li Y *et al* (2011). Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Res* **45:** 1973-1983.

Paterson M, Schindler D, Hecky R, Findlay D (2011). Comment: Lake 227 shows clearly that controlling inputs of nitrogen will not reduce or prevent eutrophication of lakes. *Limnol Oceanogr* **56:** 1545-1547.

Purvina S, Bechemin C, Balode M, Verite C, Arnaud C, Maerstrini SY (2010). Release of available nitrogen from river-discharged dissolved organic matter by heterotrophic bacteria associated with the cyanobacterium *Microcystis aeruginosa*. *Estonian J Ecol* **59:** 184-196.

Ram RJ, VerBerkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC *et al* (2005). Community proteomics of a natural microbial biofilm. *Science* **308:** 1915-1920.

Rinta-Kanto JM, Konopko EA, DeBruyn JM, Bourbonniere RA, Boyer GL, Wilhelm SW (2009). Lake Erie *Microcystis*: relationship between microcystin production, dynamics of genotypes and environmental parameters in a large lake. *Harmful Algae* **8:** 665-673.

Schindler DW (1977). Evolution of phosphorus limitation in lakes. *Science* **195:** 260-262.

Schmitz-Esser S, Penz T, Spang A, Horn M (2011). A bacterial genome in transition - an exceptional enrichment of IS elements but lack of evidence for recent transposition in the symbiont Amoebophilus asiaticus. *BMC Evol Biol* **11:** 270.

Scott JT, McCarthy MJ (2010). Nitrogen fixation may not balance the nitrogen pool in lakes over timescales relevant to eutrophication management. *Limnol Oceanogr* **55:** 1265-1270.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Meth* **9:** 811-814.

Shen H, Song L (2007). Comparative studies on physiological responses to phosphorus in two phenotypes of bloom-forming *Microcystis*. *Hydrobiologia* **592:** 475-486.

Slotkin RK, Martienssen R (2007). Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8:** 272-285.

Spura J, Christian Reimer L, Wieloch P, Schreiber K, Buchinger S, Schomburg D (2009). A method for enzyme quenching in microbial metabolome analysis successfully applied to gram-positive and gram-negative bacteria and yeast. *Anal Biochem* **394:** 192-201.

Steglich C, Lindell D, Futschik M, Rector T, Steen R, Chisholm S (2010). Short RNA half-lives in the slow-growing marine cyanobacterium *Prochlorococcus*. *Genome Biol* **11:** R54.

Straub C, Quillardet P, Vergalli J, de Marsac NT, Humbert J-F (2011). A day in the life of *Microcystis aeruginosa* strain PCC 7806 as revealed by a transcriptomic analysis. *PLoS ONE* **6:** e16208.

Vézie C, Rapala J, Vaitomaa J, Seitsonen J, Sivonen K (2002). Effect of nitrogen and phosphorus on growth of toxic and nontoxic *Microcystis* strains and on intracellular microcystin concentrations. *Microb Ecol* **43:** 443-454.

Wagner, GP, Kin, K, Lynch, VJ (2012). Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory in Biosciences* **131:** 281-285.

Wang H, Sivonen K, Rouhiainen L, Fewer D, Lyra C, Rantala-Ylinen A *et al* (2012). Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. *BMC Genomics* **13:** 613.

Welschmeyer NA (1994). Fluorometric analysis of chlorophyll-a in the presence of chlorophyll-B and pheopigments. *Limnol Oceanogr* **39:** 1985-1992.

Wilhelm SW, DeBruyn JM, Gillor O, Twiss MR, Livingston K, Bourbonniere RA *et al* (2003). Effect of phosphorus amendments on present day plankton communities in pelagic Lake Erie. *Aquat Microb Ecol* **32:** 275-285.

Wurch LL, Haley ST, Orchard ED, Gobler CJ, Dyhrman ST (2011). Nutrient-regulated transcriptional responses in the brown tide-forming alga *Aureococcus anophagefferens*. *Environ Microbiol* **13:** 468-481.

Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* **5:** 2339-2347.

Chapter IV Appendix

Table 4.2. Characteristics of RNA extracted and used for sequencing. Numbers in parentheses indicate standard deviation from triplicate measurements. Chlorophyll a fluorescence (fsu) is used as a proxy for biomass and was measured from cultures grown in 25 mL volumes on a Turner Designs TD-700™ fluorometer (Sunnyvale, CA, USA). n = 3 for all means reported. Growth rates were calculated using $\mu = Ln(N_2/N_1)/(t_2-t_1)$, with N_1 = to biomass (fluorescence) at time t_1 (Guillard 1973).

Sample	Mean Concentration of Extracted RNA (ng/µL ± s.d.)	Total mRNA used for Library Preparation	Mean Biomass after 10 days growth (fsu ± s.d.)	Mean Growth Rate (µ)	Alkaline phosphatase Activity (nmol/µg chl *a*·hr)
Control	665.3 (118.7)	500 ng	479.6 (20.3)	0.55	2.218
N-Reduced (Urea)	192.0 (12.7)	500 ng	555.4 (15.3)	0.46	3.778
N-Reduced (Ammonia)	81.2 (6.3)	500 ng	107.2 (10.2)	0.32	0.923
N-Reduced (Nitrate)	308 (9.8)	500 ng	519.2 (42.6)	0.31	0.615
P-Reduced	408.7 (264.5)	500 ng	398.5 (99.0)	0.45	0.973

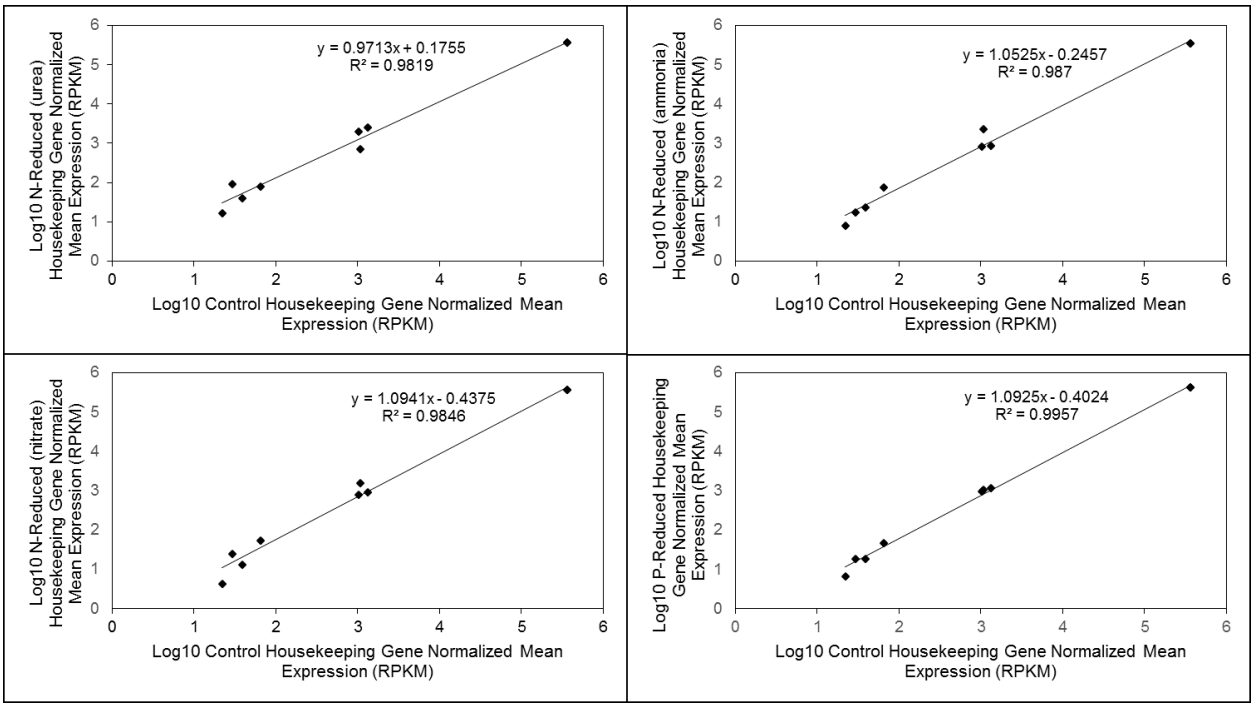

Figure 4.7. Log10 of the normalized mean expression values of genes established in the literature as common housekeeping genes (*rnpB, trpA, gltX, gyrB, rrn5S_1, rrn5S_2, recA, rrn23S_1, rrn23S_2*) (Alfonso *et al.* 2000, Christiansen *et al.* 2008, Humbert *et al.* 2013, López-López *et al.* 2005, Y *et al.* 2013) were plotted for each of the nutrient limited treatments against the corresponding values of the control treatment to validate normalization methods and calculation of expression values (Dillies *et al.* 2012).

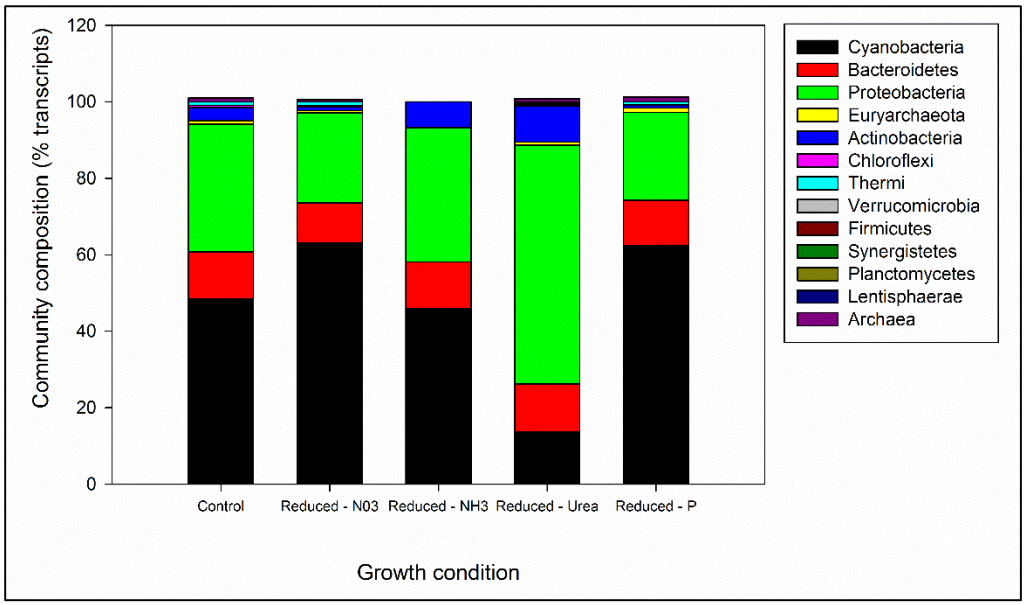

Figure 4.8. Community analysis at Kingdom and Phylum levels showing abundance across the three replicates of each treatment. Abundances are expressed in percentage of total reads that could not be recruited to the *M. aeruginosa* NIES 843 genome for comparative purposes.

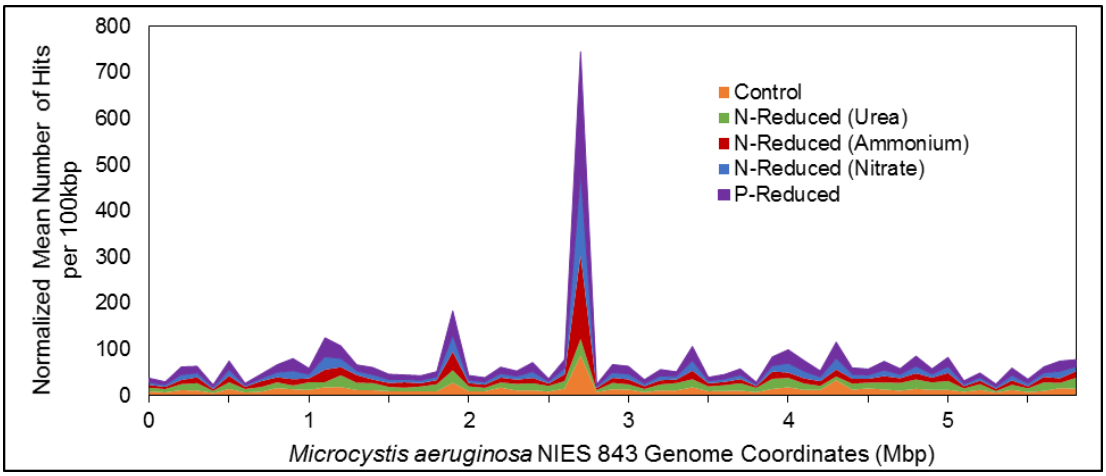

Figure 4.9. Recruitment plot (stacked) of transcriptomes to the *Microcystis aeruginosa* NIES 843 genome. Reads were binned into 100 kbp sections of the genome.

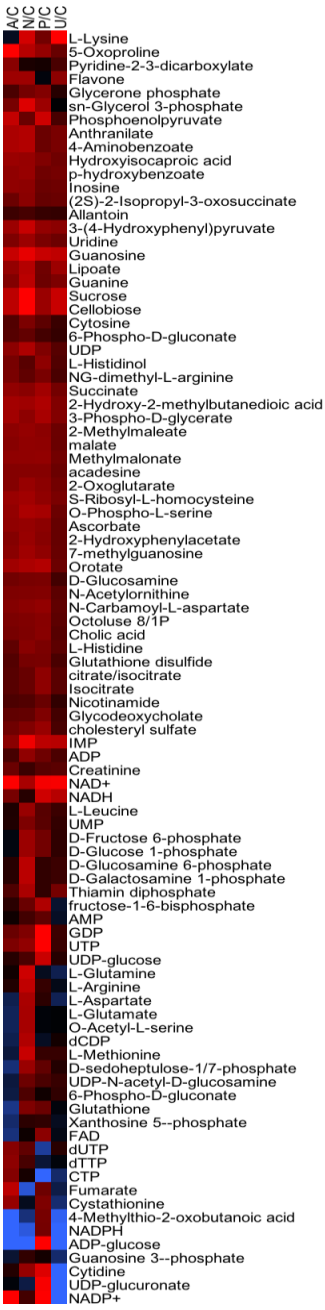

Figure 4.10. Fold changes of treatment metabolomes compared to control. A, N-limited (ammonium); N, N-limited (nitrate); P, P-limited; and U, N-limited (urea) treatments to control. Red indicates increase in metabolite concentration relative to control and blue indicates a decrease in metabolite concentration relative to control.

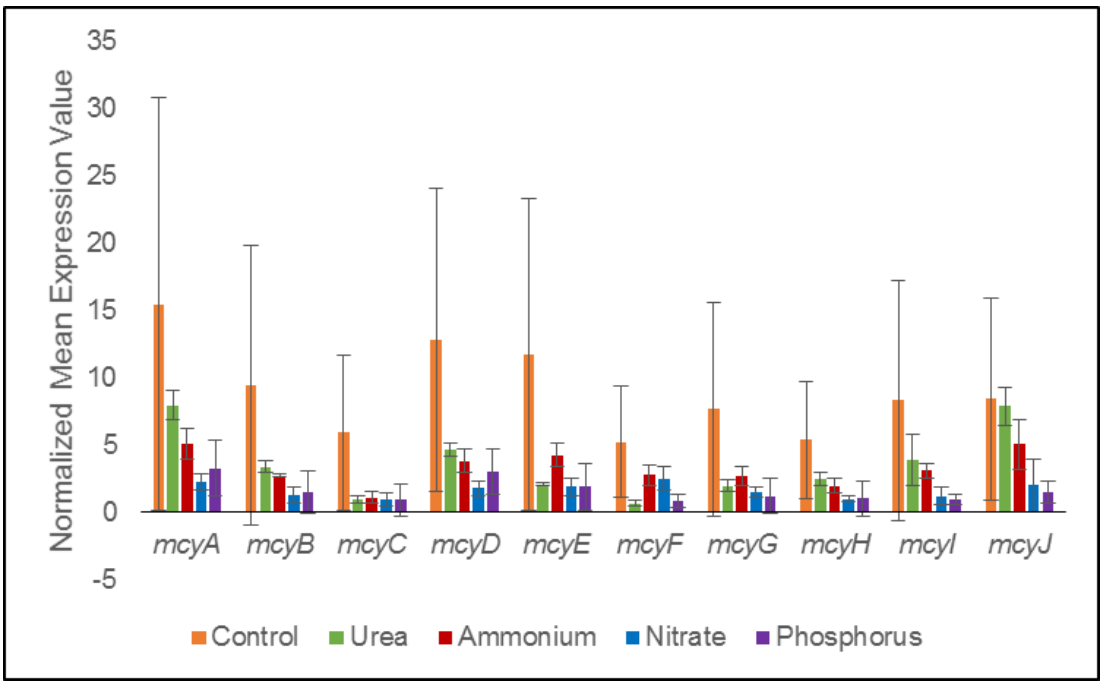

Figure 4.11. Change in expression value (RPKM) of the *mcy* toxin gene cassette.

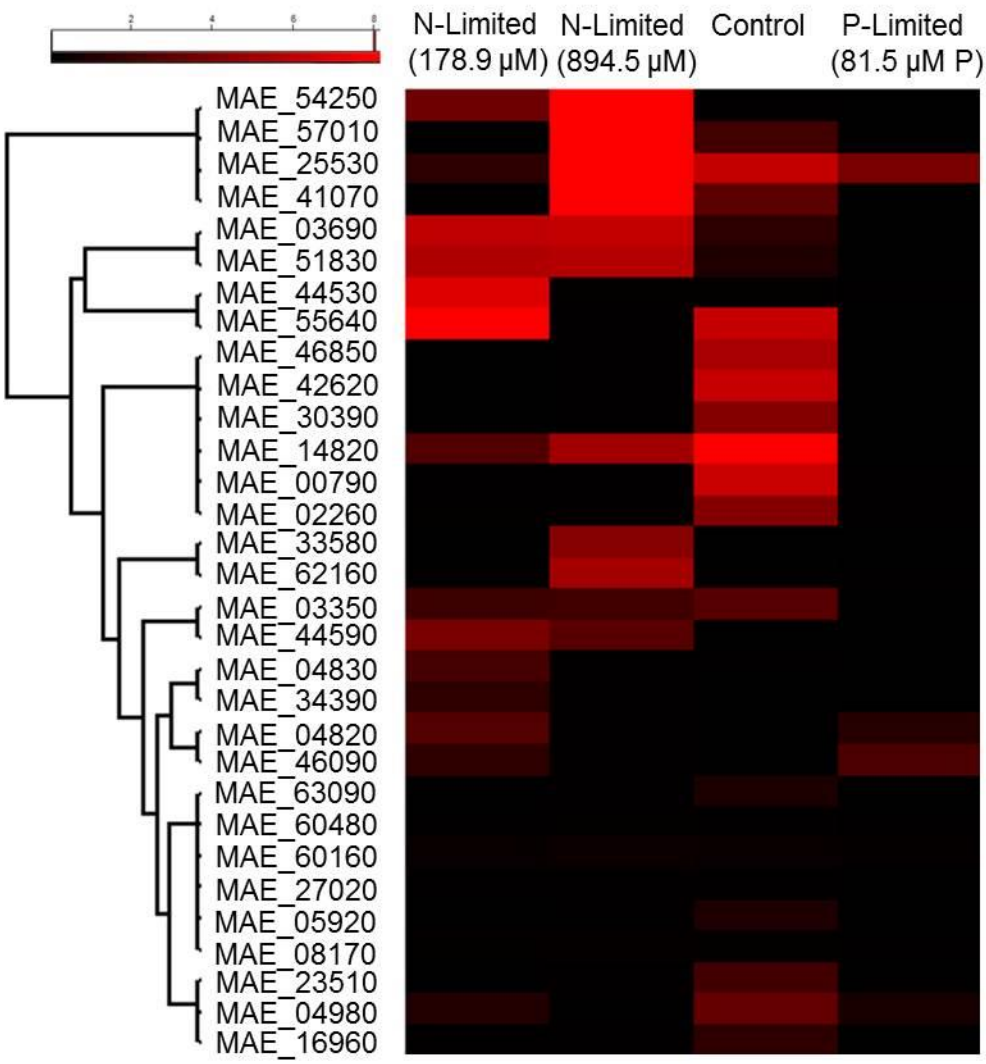

Figure 4.12. Heat map depicting protein abundances of detected transposase enzymes in *M. aeruginosa* NIES 843 cultures grown under N-reduced and P-reduced conditions. For this study, only nitrate (NO_3) was used as the N source. Modified normalized spectral abundance factor (NSAF) values from biological and technical duplicates were clustered based on Euclidean distance using average linkage.

Table 4.3. Summary of information regarding general sequencing results. All sequence libraries had a mean read length of 50 base pairs.

Sample	Reads	Reads Free of Ambiguous Bases (%)	Average GC Content (%)	Average PHRED Score	*Microcystis aeruginosa* NIES 843 (%)
Control 1	942,659	99.80	53.11	36.7	74
Control 2	1,300,284	99.79	53.21	36.7	60
Control 3	1,173,451	99.79	49.70	36.8	72
Urea 1	1,653,395	99.78	54.41	36.7	43
Urea 2	1,472,287	99.79	56.52	36.6	39
Urea 3	1,051,181	99.80	57.22	36.5	37
Ammonium 1	975,315	99.79	52.36	36.7	65
Ammonium 2	1,159,667	99.80	52.57	36.7	63
Ammonium 3	900,733	99.80	52.01	36.8	67
Nitrate 1	939,698	99.80	52.86	36.8	81
Nitrate 2	1,073,365	99.79	53.01	36.8	81
Nitrate 3	880,618	99.79	51.44	36.8	70
Phosphorus 1	914,018	99.78	51.87	36.8	73
Phosphorus 2	1,261,840	99.78	51.44	37.3	75
Phosphorus 3	1,243,944	99.79	51.74	36.8	74

Table 4.4. Coordinates of elements in the *M. aeruginosa* NIES 843 genome highlighted in Figure 4.6.

Figure 6 Element	**Description**	***M. aeruginosa* NIES 843 Genome Coordinates**
4I	Transposase gene gap I	1041344..1191289
4II	Transposase gene gap II	1620382..1731466
4III	Transposase gene gap III	1883330..1966746
4IV	Transposase gene gap IV	3309616..3427642
4a	*ureA-G*	400334..401146
		2191524..2192117
		3776565..3777041
		3849213..3849878
		4159708..4160010
		4160030..4160356
		5680106..5681815
4b	*hoxF-Y*	88744..89253
		89282..90889
		1263680..1265104
		5335966..5336514
		5337459..5338175
4c	*hypA-F*	1690174..1690530
		1690589..1691278
		5328028..5328840
		5721867..5722061
		5329833..5330903
		5709456..5711720
4d	*cpcA1-G*	1112104..1112328
		2209645..2210163
		2210230..2210718
		2210838..2211653
		2211699..2212577
		2541454..2542077
		2542713..2543525
		4438444..4439196
		4739960..4740451
		4740507..4741025
4e	Genes annotated as having putative involvement in N_2 fixation	1328194..1329135
		1397140..1397376
		1482487..1483263
		1484505..1484750
		1628877..1629560
		2065163..2066425
		2521940..2525761
		2907305..2907670
		3444826..3444990
4A	*mcyA-J*	3486436-3541027
4B	*aerD-L*	5194435-5219745
4C	*mcnA-G*	5526971-5557378

Chapter IV Appendix Reference List

Alfonso M, Perewoska I, Kirilovsky D (2000). Redox control of *psbA* gene expression in the cyanobacterium *Synechocystis* PCC 6803. Involvement of the cytochrome b(6)/f complex. *Plant Physiol* **122:** 505-515.

Christiansen G, Molitor C, Philmus B, Kurmayer R (2008). Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. *Mol Biol Evol* **25:** 1695-1704.

Dillies M-A, Rau A, Aubert J, Hennequet-Antier C, Jeanmougin M, Servant N *et al* (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform*. 14:671-683

Guillard R (1973). Division Rates. In: Stein (ed). *Handbook of Phycological Methods*. Cambridge University Press: Cambridge. pp 289-312.

Humbert J-F, Barbe V, Latifi A, Gugger M, Calteau A, Coursin T *et al* (2013). A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PLoS ONE* **8:** e70747.

López-López A, Bartual SG, Stal L, Onyshchenko O, Rodríguez-Valera F (2005). Genetic analysis of housekeeping genes reveals a deep-sea ecotype of *Alteromonas macleodii* in the Mediterranean Sea. *Environ Microbiol* **7:** 649-659.

Yannik L, Lambio A, Menzel, D, Codd, GA, Wilmotte A (2013). A cultivation-independent approach for the genetic and cyanotoxin characterization of colonial cyanobacteria. *Aquat Microb Ecol* **69:** 135-143.

CHAPTER V:
METATRANSCRIPTOMIC EVIDENCE FOR CO-OCCURRING TOP-DOWN AND BOTTOM-UP CONTROLS ON TOXIC CYANOBACTERIAL COMMUNITIES

PUBLICATION NOTE

This chapter is a version of an article in preparation for submission to *Applied and Environmental Microbiology* by Morgan M. Steffen, B. Shafer Belisle, Sue B. Watson, Gregory L. Boyer, and Steven W. Wilhelm.

My contribution to this work was the collection and processing of field samples for sequencing, performing data analysis, and most of the writing of the manuscript.

ABSTRACT

Little is known about the molecular and physiological function of co-occurring microbes within freshwater cyanobacterial harmful algal blooms (cHABs). To address this, community metatrancriptomes were examined from the western basin of Lake Erie during August 2012. Using this sequence data, we tested the hypothesis that the function of active community members is independent of community taxonomic structure. Predicted metabolic and physiological functional profiles from triplicate metatranscriptomes were determined to be ≥ 90% similar between sites. Targeted analysis of *Microcystis aeruginosa*, the historical causative agent of blooms over the past ~20 years, revealed ongoing transcription of genes involved in the acquisition of both nitrogen and phosphorus: nutrients often implicated as bottom-up drivers of eutrophication in aquatic systems. Transcription of genes involved in carbon dioxide (CO_2) concentration and metabolism also provided support for the independent hypothesis that high pH conditions and dense algal biomass result in CO_{2aq}-limiting conditions that favor cyanobacterial dominance. However, the presence of *Microcystis*-specific cyanophage sequences provided evidence of top-down virus-mediated control of cHAB populations while the activity of *Microcystis* specific transposable elements suggest the potential for ongoing changes within this organisms genome. Overall, these data provide insight into the complex series of constraints associated with *Microcystis* blooms that dominate the western basin of Lake Erie during summer months, demonstrating that multiple environmental factors work to shape cyanobacterial bloom proliferation.

INTRODUCTION

Freshwater ecosystems are considered among the most endangered in the biosphere (Dudgeon et al. 2006). One such threat to fresh waters around the world is the annual occurrence of blooms of toxic cyanobacteria. Cyanobacterial harmful algal blooms (cHABs) have been identified on every continent except Antarctica and have been occurring with increased frequency and duration in recent years (Michalak et al. 2013, Steffen et al. 2014a). The accumulation of bloom biomass has been associated with fish, avian, and mammal intoxication (Chen et al. 2009, Lindholm et al. 1999), the formation of hypoxic zones (Charlton and Milne 2004), the production of taste and odor compounds (Tabachek and Yurkowski 1976, Watson et al. 2008), and even human liver failure in extreme cases (Jochimsen et al. 1998).

The Laurentian Great Lakes are an important freshwater resource, holding ~18% of the world's potable water and ~84% of the surface waters in North America (Fuller et al. 1995). cHABs have had a persistent presence in Lake Erie and other Laurentian Great Lakes for several decades (Brittain et al. 2000, Rinta-Kanto et al. 2005, Steffen et al 2014a). Nuisance biomass and toxin production associated with cHABs in the Great Lakes have not only had a detrimental effect on ecosystem health, but also on the economic health of the surrounding communities. To date, research efforts have broadly identified nutrient inputs and temperature as contributing factors in the development of such blooms, but little focus, primarily due to methodological limitations, has been allocated to the study of the cellular physiology and metabolism of the entire bloom community. Not only do bloom-forming organisms such as *M. aeruginosa* live in competition with other phytoplankton species, they also live in concert with heterotrophic bacteria, which can attach to the mucilagenous matrix produced by *M. aeruginosa* colonies (Brunberg 1999, Shen et al. 2011). There persists a dearth of information concerning the functional role that co-occurring bacteria may play in bloom dynamics, although recent studies suggest that heterotrophs may play an important role in the metabolism of some nutrients (Jiang et al. 2007, Steffen et al. 2012).

As metagenomic surveys have provided new insight into the dynamics of microbial communities, a growing number have suggested that microbial community function is independent of the identity of the member organisms, including one study of a Lake Erie cHAB community (Burke et al. 2011, Saxton et al. 2011, Steffen et al 2012). These studies, however,

allow only for inference about community functional potential as they are based on the analysis of DNA, rather than mRNA or protein. Here we applied RNA sequencing to multiple bloom sites across Lake Erie to test the hypothesis that cHAB community function is independent of taxonomic identity. We also compared the transcript profiles of native *Microcystis* populations to a well-characterized laboratory isolate (*M. aeruginosa* NIES 843) to identify physiological differences of an important toxic cHAB organism and possible important ecological controls across the western basin of Lake Erie during the summer bloom season.

MATERIALS AND METHODS

Sample collection and survey of environmental conditions

Samples were collected in triplicate from three Environment Canada survey stations in the western basin of Lake Erie during August 2012 (Table 5.1) using the CCGS Limnos research platform. All samples were collected during daylight hours within a 24 hour period. To better concentrate the cyanobacterial community, collection was completed using a with a Nytex™ plankton net with a 20-µm pore-size. Multiple net tows at each site were homogenized then divided into 4 mL technical triplicates at each station and flash frozen in liquid nitrogen for storage and transport to the laboratory. Samples were stored at -80° C prior to processing. Chlorophyll *a* (chl *a*) concentrations for each field station sample were determined using the non-acidification method (Welschmeyer 1994, Wetzel and Likens 2000), and served as a proxy for total phytoplankton biomass. A 50 mL volume of surface water from each station was filtered through 0.2-µm nominal pore-size polycarbonate membrane filters (25 mm diam., Millipore) and stored in a 2.0 mL cryovial (Corning, NY) at -80°C, until analyzed. Chla concentrations were determined after extraction into 90% acetone overnight (24 hours) at -20°C and measured using a calibrated Turner Designs 10-AU fluorometer (Sunnyvale, CA, USA). The National Laboratory for Environmental Testing (Environment Canada) provided pH and surface temperature for each station, at the time of sampling. Toxin analyses were completed as previously described (Howard and Boyer 2007).

Phytoplankton enumeration and identification in environmental samples were performed on Lugol's Iodine preserved samples and based upon microscopic morphometrics (Auinger et al. 2008, Saxton et al. 2012). A 50 mL volume for each sample was preserved by adding 0.5 mL of

Table 5.1. Description of environmental data for collected samples.

	Station 973	**Station 882**	**Station 1163**
Latitude	41°47’30”N	41°46’00”N	41°28’10”N
Longitude	83°19’58”W	83°18’30”W	82°43’00”W
Date of Sample Collection	08/15/2012	08/15/2012	08/14/2012
pH	8.51	8.32	8.06
Microcystin Concentration	0.37 µg/L	0.11 µg/L	0.61 µg/L
Anatoxin Concentration	Not detected	Not detected	0.77 µg/L
***Microcystis* (cells mL)**	25152	12310	4027

Lugol's Iodine before storage at 4°C. Fixed samples were analyzed on a Micromaster light microscope (ThermoFisher) using a 1 mL Sedgwick-Rafter counting slide: a total of 40 fields were visualized for each sample. Due to difficulties in sometimes distinguishing cellular morphology, *Anabaena/Dolichospermum* and *Cylindrospermopsis/Raphidiopsis* populations are reported in groups.

Nucleic acid extraction, sequencing and processing

Total RNA was extracted with the MoBio RNA PowerSoil® Total RNA Isolation kit (Carlsbad, CA, USA). To remove genomic DNA contamination, samples were treated with Ambion® TURBO™ DNA Free (Grand Island, NY, USA). Samples were confirmed to be free of bacterial DNA by PCR using primers 27f and 1522r (Edwards et al. 1989). The absence of amplification was considered as the metric that the sample was DNA-free. Ribosomal RNA reduction using the Epicentre® Ribo-Zero™ rRNA removal kit (Madison, WI, USA) and sequencing on the Illumina HiSeq™ platform for 50-bp paired-end sequencing, performed by HudsonAlpha Genomic Services Laboratory (Huntsville, AL, USA). Sequences were imported into CLC Genomics Workbench version 6.5.1 (Cambridge, MA, USA) with failed reads (based on quality scores) removed. Due to the low percentage of paired end reads merging with zero mismatches, downstream contig assembly and identity analyses were performed using the 50-bp reads. To better resolve putative transcript function, reads were assembled de novo into contigs in CLC Genomics Workbench with a minimum contig length of 200-bp, with paired distances taken into account. Quantification of specific functions of interest was subsequently determined by mapping the 50-bp reads back to assembled contigs with default parameters in CLC Genomics Workbench. For specific analyses of *Microcystis* populations, RNA sequencing data was mapped to the *M. aeruginosa* NIES 843 genome, allowing 2 mismatches with default paired settings (Kaneko et al. 2007). Expression values were calculated as total counts per million (TCPM) using the CLC Genomics Workbench Transcriptomics module.

Bioinformatic and Statistical Analysis

Taxonomic identification of community members was performed using the MetaPhlAn metagenomic profiler with the "Very Sensitive Local" option (Segata et al. 2012). Functional

annotation of contigs was performed in the MG-RAST metagenomics analysis server: contig sequences are available under project ID 8691 and individual assembly IDs are listed in Table 5.2 (Glass et al. 2010, Meyer et al. 2008). Assemblies were categorized by function using the Subsystems (SEED) database with default settings (Meyer et al 2008, Overbeek et al. 2005). Contigs identified as involved in "nitrogen metabolism" and "phosphorus metabolism" functions were analyzed for Best Hit Classification using the RefSeq database with a maximum e-value cutoff of 1e-5 and minimum identity of 60%. Statistical comparisons of individual gene expression (TCPM) values was performed using a one-way ANOVA in SigmaPlot® version 12.5 (Systat Software, Inc. San Jose, CA, USA). Raw sequences will be available through the NCBI sequence read archive.

RESULTS

Sample Collection and Sequencing Output

Water samples were collected from three sites in Lake Erie during August 2012 (Table 5.1). Stations are historical Environment Canada survey sites, denoted station 973 (located in open waters of the western basin of the lake), 882 (located in the western basin at the outflow of the Maumee River) and 1163 (located within Sandusky Bay, OH). *Microcystis* (4,027 – 25, 152 cells/mL) and *Anabaena/Dolichospermum* cells (8,359 – 15,673 cells /mL) were present and abundant at all stations, although *Plantkothrix* and *Cylindrospermopsis/Raphidiopsis* populations dominated in in Sandusky Bay (station 1163) with each population at greater than 150,000 cells/mL.

RNA extraction yielded an average of 162.2 ng/µL of total RNA, although this varied between individual samples (Table 5.2). Sequence quality as measured by phred score was an average of 37.15, and the percentage of ambiguous bases in each read was an average of 0.057 (Table 5.2). Metatranscriptome libraries were sequenced in triplicate, resulting in a total of 384 million reads. Each library comprised an average of 42.7 million, 50-bp paired-end sequences. Short read paired end (50-bp) libraries assembled into contigs (>200bp) with an average of 69.9% (±10.1%) of short reads recruiting to assemblies (Table 5.2).

Table 5.2. Description of sequence data. All values represent triplicate means.

	973	**882**	**1163**
Total RNA Extracted (ng / µL)	263.3	59.68	163.7
Number of Reads	44861129	42443881	40898607
% GC Content	49.44	54.33	48.59
% Ambiguous Bases	0.057	0.055	0.058
Average Phred Score	37.45	36.59	37.42
Number of Contigs	7164	10224	11694
Number of Reads Mapped to Contigs	36769599	24622585	28585787
% Reads Mapped to *M. aeruginosa NIES 843*	3.91	4.08	0.396
Assembly MG-RAST ID	4559605.3-4559607.3	4559602.3-4559604.3	4559599.3-4559601.3

Active Lake Erie bloom community structure from metatranscriptomes

A total of 32 bacterial and archaeal phyla were detected, with bacterial cDNA representing > 98% of all sequences (Figure 5.1A). Community analysis of short read libraries showed a predominance of Cyanobacteria in the samples collected from all three sites: more than 85% of bacterial reads at stations 973 and 1163 were identified as *Cyanobacteria* (Figure 5.1A) and were similar in composition to a previous Lake Erie cyanobacterial bloom metagenome (Steffen et al 2012). The bacterial community structure at station 882 was more varied, with *Cyanobacteria* comprising ~38% of the community and *Proteobacteria* ~35%. An examination of the cyanobacterial sequences revealed differences in the transcriptionally active cyanobacterial community structure between the three sites surveyed (Figure 5.1B). At the order level, *Nostocales* (~76%) (i.e., *Anabaena, Nostoc, Cylindrospermopsis*) were dominant at station 973 and *Oscillatoriales* (~74%) (i.e., *Planktothrix, Oscillatoria, Lyngbya*) at station 1163. *Chroococcales* (~13%) (i.e., *Microcystis*) and *Oscillatoriales* (~14%) were relatively evenly represented at station 882.

Active biological functions of the Lake Erie bloom communities. In contrast to differences in community structure, functional profiles of assembled community sequences were similar across all sites (Figure 5.2). Populations at station 973 had reductions (< 4% of annotated contigs) in “photosynthesis”, “protein metabolism”, and “respiration” SEED categories and an increase in the “amino acid metabolism” and “miscellaneous functions” categories in comparison to stations 882 and 1163 (Figure 5.2). Multidimensional scaling (MDS) analysis of Bray Curtis similarity between the three sites confirms this observed functional similarity between station populations, with all samples clustering within 75% similarity to one another (Figure 5.3). Replicates clustered more closely to each other than between sites (95%), however, stations 882 and 1163 also clustered closely (90%), suggesting stronger functional similarity at those two sites (Figure 5.3), despite the differences in community structure (Figure 5.1).

To verify that functional profiles generated from assembled contigs were representative of short reads libraries, 50-bp sequences were recruited back to contigs associated with nitrogen and phosphorus transport and metabolism in order to provide quantitative insight on expression.

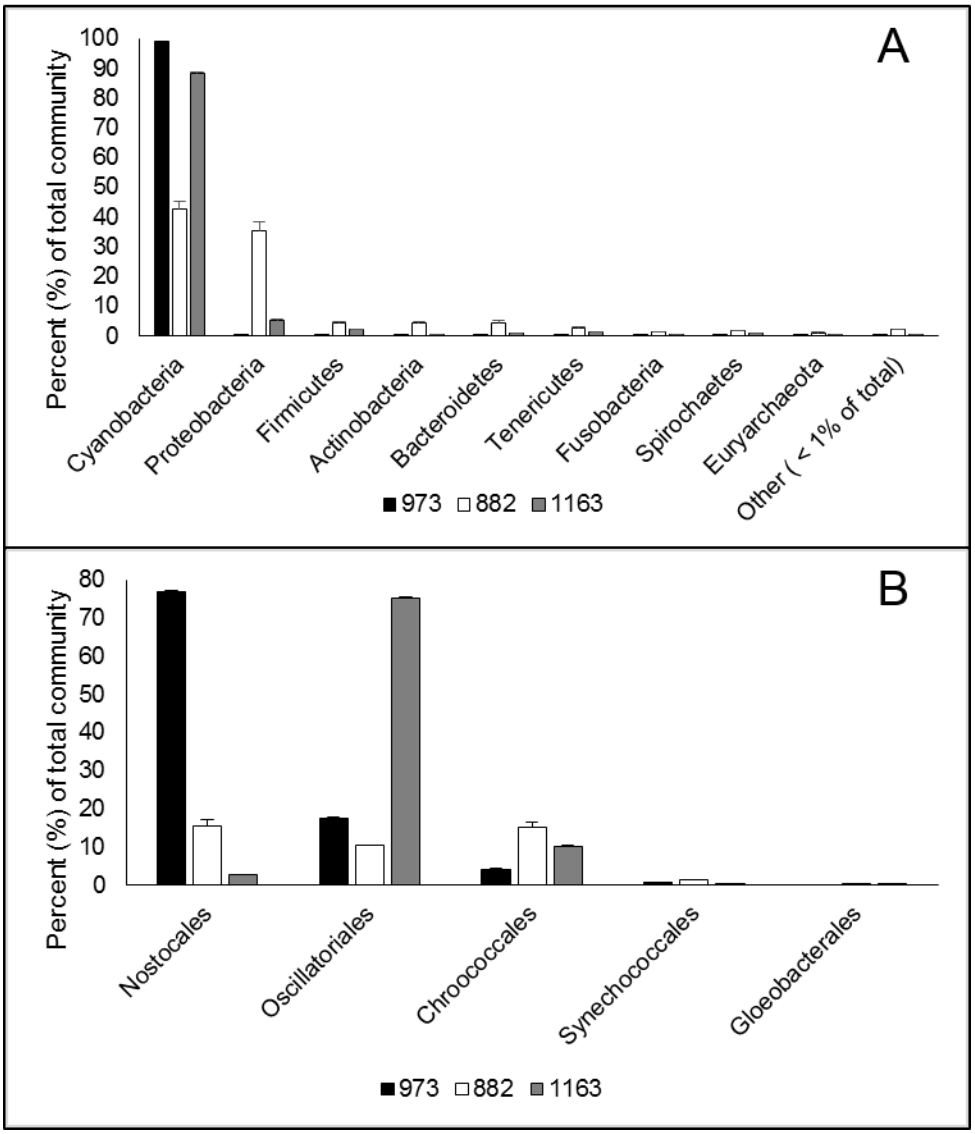

Figure 5.1. A) Metatranscriptome community structure of the particle associated (> 20 µm) bacterial population at the phylum level. B) Order-level community structure of the cyanobacterial population associated with each station. Bars represent standard deviation of triplicates.

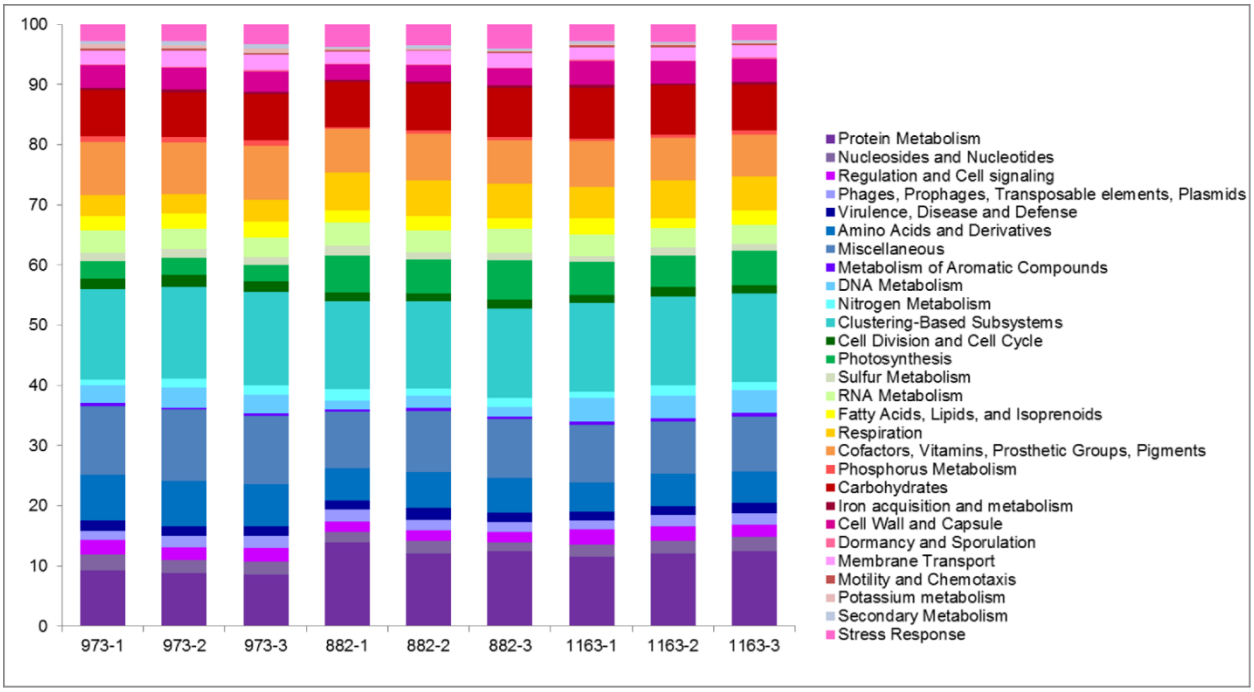

Figure 5.2. Community functional analysis of each library assembly. Functional annotation is based on the Subsystems (SEED) database performed *via* the MG-RAST server. To demonstrate reproducibility, individual replicates are shown separately.

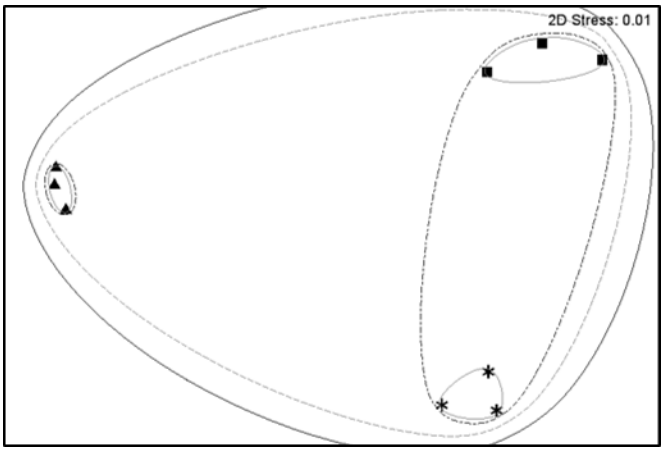

Figure 5.3. MDS analysis of a Bray-Curtis similarity matrix of functional assembly profiles from each station. Triangles represent station 973, squares represent station 882, and asterisks represent station 1163. The solid black line indicates 50% similarity, the dashed grey line is 75% similarity, the dashed black line is 90% similarity, and the solid grey line is 95% similarity between samples.

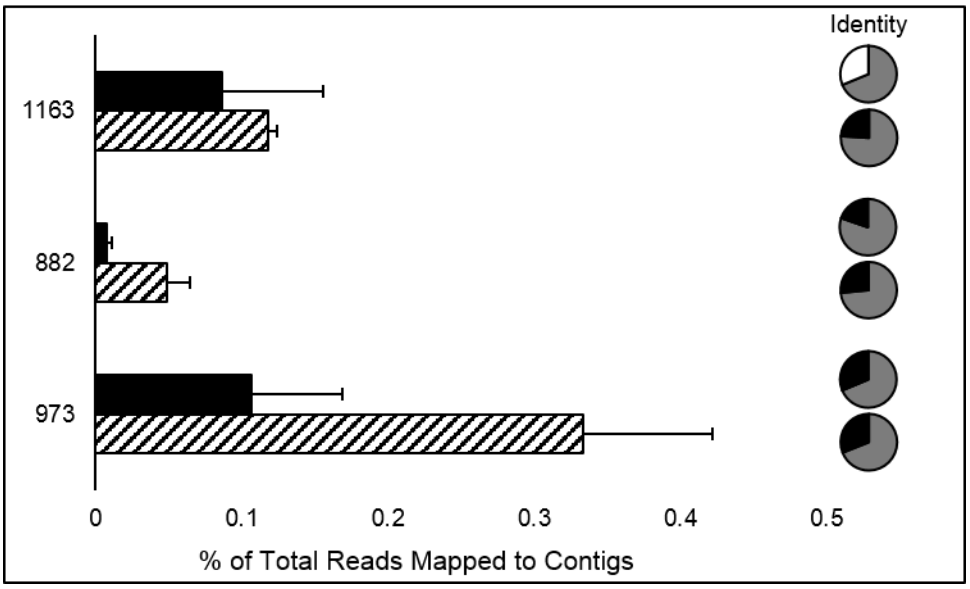

Figure 5.4. Percentage of total short read libraries mapped to the assembled contigs annotated as nitrogen or phosphorus metabolism and transport functional categories. The percentage of reads mapped is calculated from the total number of reads which mapped to the complete set of assembled contigs. Black bars represent phosphorus functional categories and striped bars represent nitrogen functional categories. Pie charts indicate the identity at the order-level of those contigs with N and P function. Black is *Chroococcales*, Grey is *Nostocales*, and white is *Osciallatoriales*.

Not all short reads recruited to the contigs, suggesting that assembly may artificially inflate the importance of certain functions compared to their abundance in unassembled sequence libraries. However, the recruitment suggests that a higher number of transcripts were dedicated to both nitrogen and phosphorus metabolism at station 973 than the other two stations (Figure 5.4). To determine which organisms were responsible for these specific functions at each station, contigs associated with nitrogen and phosphorus metabolism were identified using the Ref-Seq database on MG-RAST. Greater than 99% of all reads were identified as cyanobacterial, with members of *Chroococcales* contributing the majority of transcripts involved in N and P metabolism, followed by *Nostocales*, with members of *Oscillatoriales* only contributing transcripts involved in P metabolism and transport at station 1163 (Figure 5.4).

Activity of the *Microcystis* population

Members of the order *Chroococales*, including *Microcystis* spp, comprised between 4.1 (± 0.10) and 15.1 (± 1.47) % of the total mRNA reads from Lake Erie during our August 2012 survey (Figure 5.1B). Amongst these, 0.4 - 5.0 % of total sequence reads recruited to the genome of *M. aeruginosa* NIES 843. This limited recruitment is due to the fact that the *Chroococcales* population in Lake Erie contains multiple genera of cyanobacteria, and likely a diverse number of *Microcystis* strains (Rinta-Kanto et al. 2009). On average, 20.3% (± 1.4%) of genes within the NIES 843 genome were transcribed at station 973, with 31 of those genes only transcribed at that station across all replicates (Figure 5.5A). Transcription of 16.1% (±1.4%) and 6.7% (±0.5%) of NIES 843 genome was detected at stations 882 and 1163, respectively, with 12 genes detected exclusively at 882 and 4 at 1163. Genes encoding transposases (MAE_06490, MAE_29610, MAE_41670, MAE_45770) were among those *M. aeruginosa* NIES 843 uniquely identified at stations 882 and 973. 185 of the 610 transposase genes exhibited transcription within at least one Lake Erie sample; 29.7% of those genes were transcribed at all three sites and an additional 22.7% were transcribed at both 973 and 882. Expression patterns of environmental *Microcystis* transposases exhibited tight clustering patterns across replicates, and stations were more similar to each other than to previously published culture studies of this isolate (Figure 5.6). Those genes unique to station 973 included eight hypothetical proteins, a protein involved in iron transport (MAE_12460), and an excision element (*xisI*) (Appendix Table 5.1). Unique to 882 were a

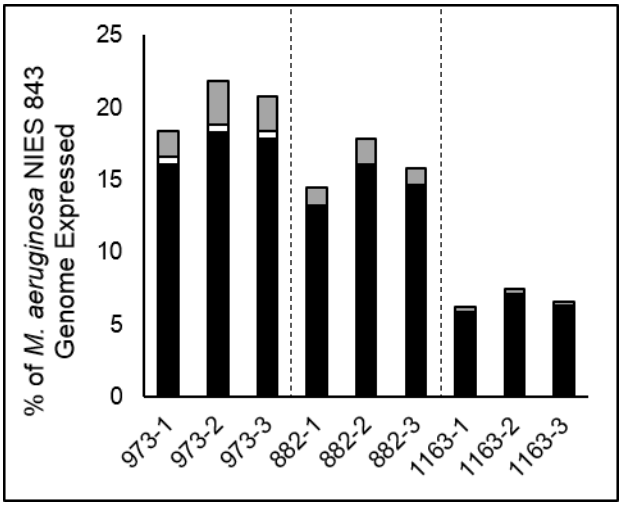

Figure 5.5. Summary of *M. aeruginosa* NIES 843 gene transcription across all samples. Black indicates the percentage of genes in the genome with calculated expression values (TCPM), grey is the percentage of genes expressed only in one replicate, and white is the percentage of genes expressed only at a single station across all replicates.

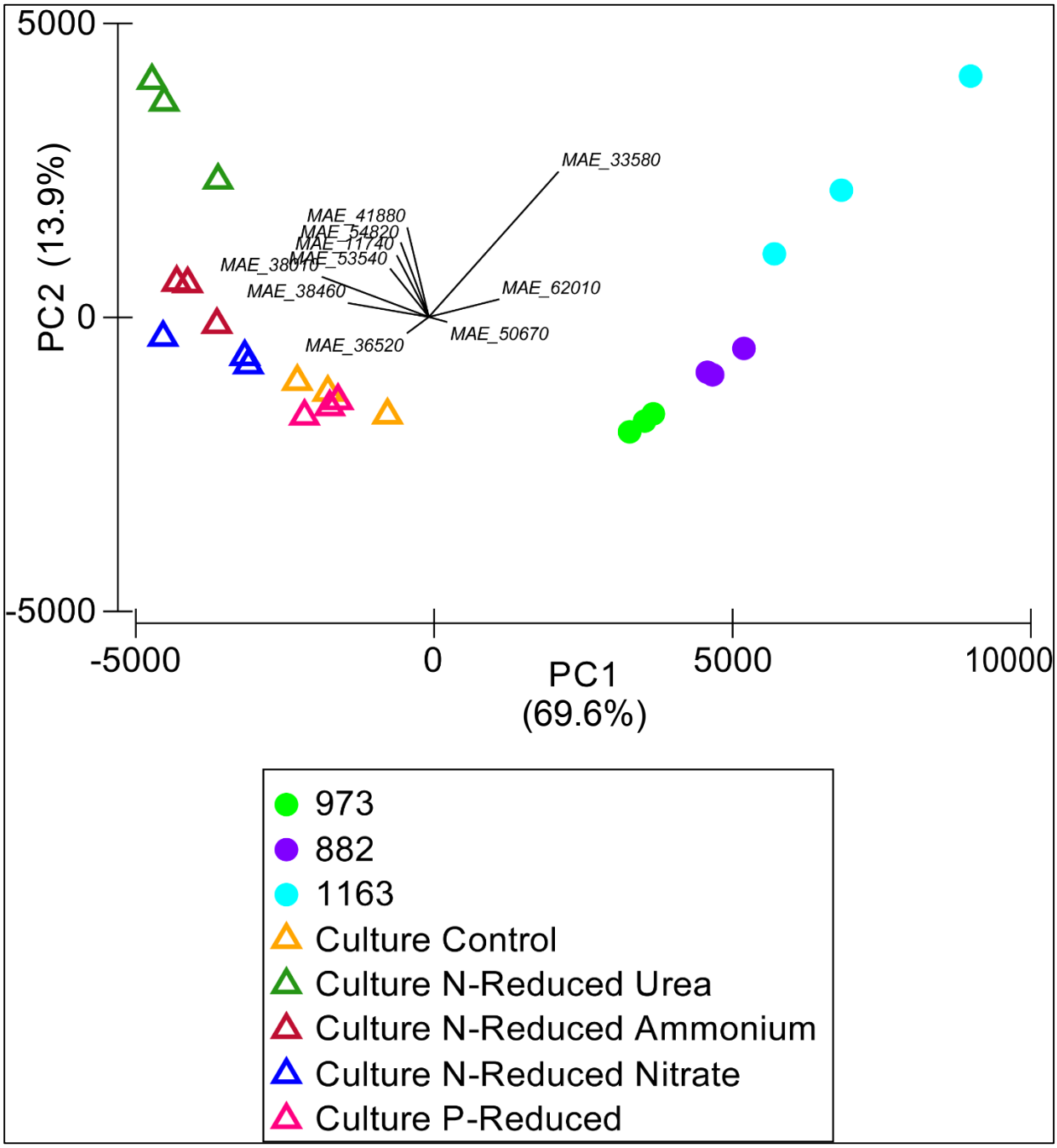

Figure 5.6. Principle component analysis (PCA) of expression patterns of 610 transposase genes in the genome of *M. aeruginosa* NIES 843. Expression values of cultured samples taken from Steffen et al. (Steffen et al 2014b).

phosphate-binding protein (MAE_32380) and an arsenite transporter (MAE_02560). *nirA* (nitrite reductase) was one of the genes unique to 1163.

Given the importance of nutrient abatement programs in managing the formation and persistence of cHABs (Paerl et al. 2011, Steffen et al 2014a), we examined the expression profiles of several key genes involved in nutrient (nitrogen and phosphorus) metabolism and transport. *glnA* (glutamate-ammonia ligase), *narB* (ferredoxin-nitrate reductase), *nirA* (ferredoxin-nitrite reductase), and ureC (urease alpha subunit) were chosen as markers for ammonium, nitrate, nitrite, and urea metabolism, respectively. While *glnA, narB,* and *ureC* were transcribed at all stations, none had significantly different transcription patterns between stations ($P < 0.05$). Across all stations, 11 additional genes involved in inorganic carbon uptake and metabolism from *M. aeruginosa* NIES 843 were detected (Figure 5.7). Station 973 exhibited significantly higher ($P < 0.05$) transcription of *ccmM* and *rbcL* than the other two stations, *ccmK2* than *882*, and *sbtA* than 1163. There was also significantly ($P < 0.05$) increased transcription of *rbcS* at 1163 over 882. These genes encode components of the CO_2 concentrating mechanism of *Microcystis* (Kaneko et al., 2007).

DISCUSSION

Application of deep sequencing technologies to the study of microbial community structure and function has resulted in the observation that variability in microbial community structure is not reflected in the function of that community (Burke et al 2011, Fan et al. 2012, Frossard et al. 2012). Here, we tested this hypothesis using RNA sequencing to identify the transcriptionally active microbial community members and the function of those genes being transcribed. Our data suggest the taxonomic identity of active Lake Erie bloom communities vary across the western basin (Figure 5.1), but the biological functions of those communities remains relatively static (Figures 5.2-5.3), and that relatively minor members of the population appear to contribute disproportionately to certain functions (Figures 5.1, 5.4). To better resolve this issue, we examined historically-cited factors that shape cyanobacterial communities (N and P cycling) as well as more recently hypothesized (CO_2 acquisition, cyanophage and transposable element activity) constraints (Van de Waal et al., 2011; Tucker and Pollard, 2005; Steffen et al., 2014b). We couch our observations within the confines of a community-

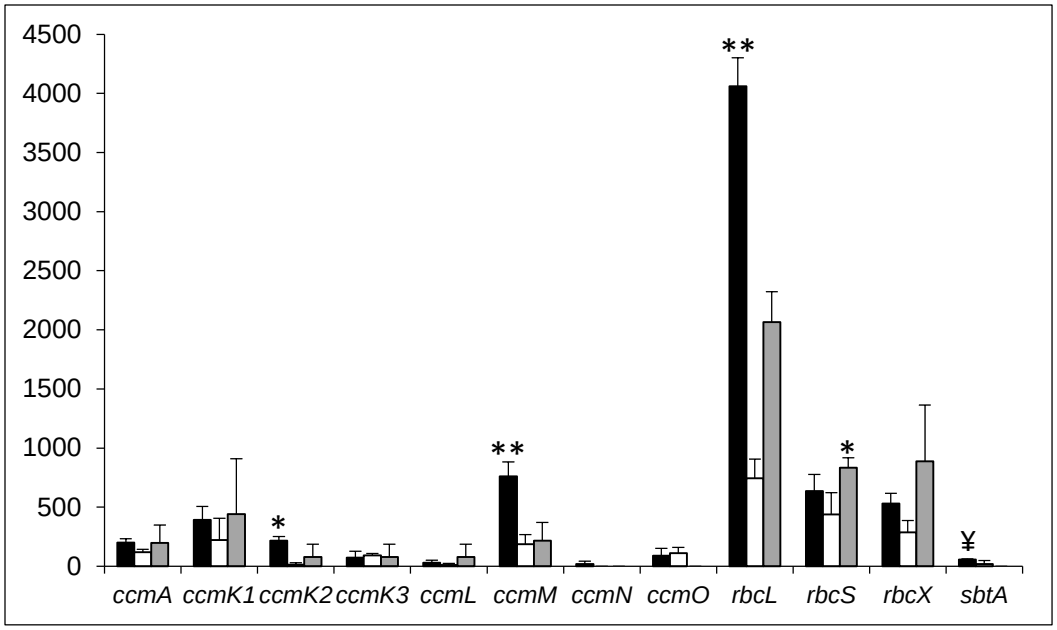

Figure 5.7. Expression values (TCPM) of *M. aerguinosa* NIES 843 genes associated with inorganic carbon uptake or metabolism function at each site. Black bars represent station 973 mean values, white bars station 882 mean values, and grey bars station 1163 mean values. Error bars indicate standard deviation. (**) indicates significantly increased expression at station 973 ($P < 0.05$) over both 882 and 1163, (*) denotes significantly increased expression over 882 at either 1163 or 973, and ¥ denotes significantly increased expression over only 1163. Error bars indicate standard deviation.

based sequencing approach to provide new insight into the complex nature of microbial populations and cyanobacterial blooms. Within the bloom associated (> 20.0 µm) samples we collected, members of *Chroococcales*, including *Microcystis* spp., appear to play an important role in nutrient cycling (Figure 5.4) as well as the perceived "health" of the system. We thus performed a targeted analysis of *Microcystis* gene transcription using the genome of *M. aeruginosa* NIES 843 as the model strain, as this is the only strain for which a fully closed genome is available (Kaneko et al 2007). The overall expression of specific genes involved in N and P metabolism as well as the identification of uniquely transcribed genes at each station provided insight into the physiology and contribution of *Microcystis* to nutrient cycling in Lake Erie. Nitrite reductase (*nirA*) mRNA was only detected at station 1163, suggesting nitrite, and thus potentially nitrate as well, played a more important role as a nitrogen source in Sandusky Bay than at other sites within the western basin (*i.e.*, it was not actively being consumed). In contrast, alkaline phosphatase (MAE_16640), a classical marker for P-limited plankton growth was detected at stations 973 and 882. However, recent observations have suggested this gene might be regulated by urea in addition to phosphate concentrations (Steffen et al. 2014b), while other studies have shown that transcription of this gene does not necessarily indicate phosphorus stress in eutrophic systems (Nausch 1998), calling into question the validity of this interpretation. Indeed concurrent microcosms efforts (Belisle, 2014) have demonstrated that primary producers were most likely N-limited during August 2012, consistent with observation from late summer in previous seasons (Chaffin et al. 2013).

Far fewer studies have been dedicated to resolving the role of CO_2 availability relative to N and P dynamics in cHAB formation and persistence. We detected transcripts for 12 genes involved in carbon concentration (Figure 5.7). Their increased transcription at station 973 suggests *Microcystis* cells were investing significant resources to acquire CO_2, a potential indicator of CO_2 limitation at this station (which also had of the highest pH of the sites surveyed, Table 5.1). Enhanced photosynthesis associated with dense algal blooms is commonly accompanied by a subsequent increase in pH, creating the potential for growth-limitation of phototrophs by decreased dissolved inorganic carbon supplies (Ibelings and Maberly 1998). In such situations, organisms with efficient carbon concentrating mechanisms are therefore thought to have a competitive advantage in bloom conditions (Shapiro 1997). The evidence presented

here supports the hypothesis that these conditions exist in the eutrophic western basin of Lake Erie, at least in *Microcystis* populations. Because little is known regarding the role of CO_2 limitation in bloom community speciation and function (Van de Waal et al. 2011), this hypothesis warrants further testing in the field. Indeed, there have been calls in the literature for further study of this phenomenon, especially as atmospheric CO_2 levels are projected to rise over the next century (O'Neil et al. 2012).

Recent observations suggest that environmental conditions may in part regulate transposase activity for not only *Microcystis*, but across all domains of life (Capy et al. 2000, Jäger et al. 2009, Steffen et al 2014b). While *Microcystis* transposase activity has not yet been surveyed in the environment, the data presented herein suggest that at least a subset of the *Microcystis* population in Lake Erie was actively transcribing transposase genes during August 2012. The four most highly expressed transposases (MAE_33580, MAE_50660, MAE_50670, and MAE_62010) are conserved across all three sites (Figure 5.6); three of these genes (MAE_33580, MAE_50660, MAE_62010) have been shown to exhibit differential transcription dependent upon nutrient conditions in a lab-based culture study (Steffen et al 2014b). Increased transposase activity is thought to play an important role in evolutionary response to variable environments (Capy et al 2000, Casacuberta and González 2013), and the high number of transcribed transposases across Lake Erie leads to the hypothesis that such a response may be associated with how *Microcystis* cells maintain bloom conditions (Steffen et al 2014b). 68% of variation between samples was explained by PC1 (which separated cultured isolates from field samples), while an additional 13% was explained by PC2. Along PC2, station 1163 aligns more closely with N-limited cultures exposed to urea as the sole N-source: deeper inquiry is necessary to determine whether transposase activity can indeed be used as an indicator of *Microcystis* nutrient status (Figure 5.6).

Microcystin concentrations were below the World Health Organization's drinking water limit of 1 µg/L, however, this toxin was detected at all three sites (Table 1). Somewhat surprisingly, no microcystin biosynthesis genes (*mcyA-J*) were detected at any site. Previous studies have shown that high microcystin concentrations (1-25 µg/L) do not always correspond to a strong transcriptional signal for *Microcystis* mcy genes (Gobler et al. 2007). Indeed, as toxicity varies within cyanobacterial populations and the toxin does not degrade quickly, it is

likely we missed the period of active toxin gene transcription. Alternatively, it is important to remember our sampling protocol focused on *Microcystis* colonies and long filaments: it is possible that there are organisms capable of producing microcystin that evaded our sampling efforts. Finally, in spite of what appears on the surface to be an impressive depth of sequencing, it is important to remember that even with ~ 384 million reads, the information only is equivalent to the genomes of ~64 individual *Microcystis* cells: to this end it remains possible, that rare transcripts were not captured.

While significant attention has been directed toward understanding bottom-up controls of cHABs (i.e., nutrients, temperature), top-down controls such as grazing and phage infection can also play an important role in bloom dynamics (Tucker and Pollard 2005). To quantify active phage infections, we recruited pooled short reads unmapped to the NIES 843 genome to the genome of the *Microcystis* phage Ma-LMM01 (Yoshida et al. 2008b). The recruitments provided from 6.4 x to 11.7 x coverage of the virus genome across all three stations, although coverage of individual genes was highly variable (Table 5.2). Historically, the tail sheath gene gp91 has been used to quantify Ma-LMM01 copies in environmental samples (Rozon and Short 2013): only afew reads recruited to this gene from our samples (Table 5.3). Gene *gp135*, which is annotated as a transposase that has only ~81% identity with host genes, recruited a higher number of reads, suggesting sampling may have captured a stage of infection prior to the transcription and manufacture of tail proteins. To assess the relative magnitude of infection at each site, the ratio of reads recruited to phage *gp135* and the *M. aeruginosa* NIES 843 *rpoB* gene was determined. This host gene is commonly used as a housekeeping gene in bacterial studies as a proxy for cell copy number (Case et al. 2007). The highest ratio of phage gene to host gene was at station 882 (0.77), implying that an active infection may have been taking place in the *Microcystis* community during our sampling efforts. This observation was made in spite of station 973 having the greatest number of *Microcystis*-associated transcripts. (Table 5.3). Negative correlations between host: phage have been observed in previous studies of Ma-LMM01 infection in *M. aeruginosa* (Yoshida et al. 2008a), and make sense given the "lethal" nature of most virus infections (Weitz and Wilhelm 2012). While additional study is required to resolve the importance of phages in the natural control of cHABs in large lake systems, our observations

Table 5.3. Summary of viral recruitment for *Microcystis* phage MA-LMM01

	Station 973	**Station 882**	**Station 1163**
Reads mapped to Ma-LMM01 Genome	37,336	2,522	4,748
Coverage of Ma-LMM01 Genome	10.9x	11.7x	6.4x
Reads mapped to *gp91*	19	8	33
Reads mapped to *gp135*	1325	1961	317
Coverage of gp135	56.4x	79x	12.8x
gp135*:*rpoB	0.13	0.77	0.07

provide preliminary evidence of transcriptionally active phages within Lake Erie bloom communities. Furthermore, these observations suggest that extraction of virus gene sequences from particle associated transcriptomes may be a useful way to better resolve the activity of DNA viruses in natural systems. Indeed, the expression of several other virus transcripts, including several hundred reads identified as an *Acinetobacter* phage AP205 (Klovins et al. 2002) at each station, highlight the broad applicability of this approach to both host and virus dynamics (Table 5.3).

Conclusions

Application of RNA-sequencing technologies to the study of cHAB-associated microbial communities provides robust evidence that microbial community function occurs independent of taxonomic identity. Despite variable taxonomies within these bacterial communities, functional profiles were 90% similar across the western basin of Lake Erie (Figures 5.1-5.3). The dominant role of Lake Erie cyanobacteria in nitrogen metabolism within the entire bloom community, hypothesized in a metagenomics study in 2012 (Steffen et al 2012), was functionally confirmed by this RNA sequence analysis (Figure 5.4). *M. aeruginosa*, an important toxin-producer in the lake, was an active member of all sites surveyed, as transcripts for genes associated with nitrogen metabolism, including urea degradation and phosphorus metabolism, were identified. Of note, transcriptional profiles indicate potential CO_2 limitation was occurring, particularly at stations with higher pH. As CO_2 levels climb toward projected levels, the role of CO_2 in driving phototroph community speciation may become increasingly important, and warrants further study (O'Neil et al 2012). Additionally, transcriptional evidence of an active infection by a *Microcystis* cyanophage were identified as a potential top-down control. As it is unlikely that infection of only a single cyanophage occurred throughout the entire population this observation highlights the potential importance of top-down in situ virus controls on cHAB dynamics. Taken together, this study provides new insight into the multiple constraints on cHAB communities across the western basin of Lake Erie during bloom conditions. This new information may be an important consideration for management practices, as a key factor in the construction of robust management strategies is a scientifically-based understanding of the function of microbial populations which make up cHAB communities.

ACKNOWLEDGEMENTS

We thank the captain and crew of the *C.C.G.S. Limnos* as well as Dr. Shawn Levy and colleagues at Hudson Alpha. We also thank Helena Pound and Katherine Perri for their assistance in the field. This project was supported by grants from the National Science Foundation (IOS - 0841918, DEB – 1240870 and CBET- 1230543) and a UT/ORNL Science Alliance JDRD award to SWW. MMS was supported by a Wallace-Dean fellowship from the University of Tennessee.

LIST OF REFERENCES

Auinger BM, Pfandl K, Boenigk J (2008). Improved methodology for identification of protists and microalgae from plankton samples preserved in Lugol's iodine solution: combining microscopic analysis with single-cell PCR. *Appl Environ Microbiol* **74:** 2505-2510.

Brittain SM, Wang J, Babcock-Jackson L, Carmichael WW, Rinehart KL, Culver DA (2000). Isolation and characterization of microcystins, cyclic heptapeptide hepatotoxins from a Lake Erie strain of *Microcystis aeruginosa*. *J Great Lakes Res* **26:** 241-249.

Brunberg A-K (1999). Contribution of bacteria in the mucilage of *Microcystis* spp. (Cyanobacteria) to benthic and pelagic bacterial production in a hypereutrophic lake. *FEMS Microbiol Ecol* **29:** 13-22.

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011). Bacterial community assembly based on functional genes rather than species. *Proc Natl Acad Sci USA* **108:** 14288-14293.

Capy P, Gasperi G, Biémont C, Bazin C (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity* **85:** 101-106.

Casacuberta E, González J (2013). The impact of transposable elements in environmental adaptation. *Mol Ecol* **22:** 1503-1517.

Case RJ, Boucher Y, Dahllöf I, Holmström C, Doolittle WF, Kjelleberg S (2007). Use of *16S rRNA* and *rpoB* genes as molecular markers for microbial ecology studies. *Applied and Environmental Microbiology* **73:** 278-288.

Chaffin JD, Bridgeman TB, Bade DL (2013). Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. *Adv Microbiol* **3:** 16.

Charlton M, Milne J (2004). Review of thirty years of change in Lake Erie water quality. *NWRI contribution*.

Chen J, Zhang D, Xie P, Wang Q, Ma Z (2009). Simultaneous determination of microcystin contaminations in various vertebrates (fish, turtle, duck and water bird) from a large eutrophic Chinese lake, Lake Taihu, with toxic *Microcystis* blooms. *Sci Total Environ* **407:** 3317-3322.

Dudgeon D, Arthington AH, Gessner MO, Kawabata Z-I, Knowler DJ, Lévêque C *et al* (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81:** 163-182.

Edwards U, Rogall T, Blöcker H, Emde M, Böttger EC (1989). Isolation and direct complete nucleotide determination of entire genes. Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Res* **17:** 7843-7853.

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS *et al* (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci USA* **109:** E1878–E1887.

Frossard A, Gerull L, Mutz M, Gessner MO (2012). Disconnect of microbial structure and function: enzyme activities and bacterial communities in nascent stream corridors. *ISME J* **6:** 680-691.

Fuller K, Shear H, Wittig J (1995). The Great Lakes: an environmental atlas and resource book.

Glass EM, Wilkening J, Wilke A, Antonopoulos D, Meyer F (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harbor Protocols* **2010:** pdb.prot5368.

Gobler CJ, Davis TW, Coyne KJ, Boyer GL (2007). Interactive influences of nutrient loading, zooplankton grazing, and microcystin synthetase gene expression on cyanobacterial bloom dynamics in a eutrophic New York lake. *Harmful Algae* **6:** 119-133.

Howard KL, Boyer GL (2007). Quantitative analysis of cyanobacterial toxins by matrix-assisted laser desorption ionization mass spectrometry. *Anal Chem* **79:** 5980-5986.

Ibelings BW, Maberly SC (1998). Photoinhibition and the availability of inorganic carbon restrict photosynthesis by surface blooms of cyanobacteria. *Limnol Oceanogr* **43:** 408-419.

Jäger D, Sharma CM, Thomsen J, Ehlers C, Vogel J, Schmitz RA (2009). Deep sequencing analysis of the *Methanosarcina mazei* Gö1 transcriptome in response to nitrogen availability. *Proc Natl Acad Sci USA* **106:** 21878-21882.

Jiang L, Yang L, Xiao L, Shi X, Gao G, Qin B (2007). Quantitative studies on phosphorus transference occuring between *Microcystis aeruginosa* and its attached bacterium (*Pseudomonas* sp.). In: Qin B, Liu Z, Havens K (eds). *Eutrophication of Shallow Lakes with Special Reference to Lake Taihu, China*. Springer Netherlands. pp 161-165.

Jochimsen EM, Carmichael WW, An J, Cardo DM, Cookson ST, Holmes CEM *et al* (1998). Liver failure and death after exposure to microcystins at a hemodialysis center in Brazil. *N Engl J Med* **338:** 873-878.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M *et al* (2007). Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Res* **14:** 247-256.

Klovins J, Overbeek GP, van den Worm SHE, Ackermann H-W, van Duin J (2002). Nucleotide sequence of a ssRNA phage from *Acinetobacter*: kinship to coliphages. *J Gen Virol* **83:** 1523-1533.

Lindholm T, Öhman P, Kurki-Helasmo K, Kincaid B, Meriluoto J (1999). Toxic algae and fish mortality in a brackish-water lake in Åland, SW Finland. *Hydrobiologia* **397:** 109-120.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass E, Kubal M *et al* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Michalak AM, Anderson EJ, Beletsky D, Boland S, Bosch NS, Bridgeman TB *et al* (2013). Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc Natl Acad Sci USA* **110:** 6448-6452.

Nausch M (1998). Alkaline phosphatase activities and the relationship to inorganic phosphate in the Pomeranian Bight (southern Baltic Sea). *Aquat Microb Ecol* **16:** 87-94.

O'Neil JM, Davis TW, Burford MA, Gobler CJ (2012). The rise of harmful cyanobacteria blooms: The potential roles of eutrophication and climate change. *Harmful Algae* **14:** 313-334.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang H-Y, Cohoon M *et al* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Paerl HW, Xu H, McCarthy MJ, Zhu G, Qin B, Li Y *et al* (2011). Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Res* **45:** 1973-1983.

Rinta-Kanto JM, Ouellette AJA, Boyer GL, Twiss MR, Bridgeman TB, Wilhelm SW (2005). Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. *Environ Sci Technol* **39:** 4198-4205.

Rinta-Kanto JM, Saxton MA, DeBruyn JM, Smith JL, Marvin CH, Krieger KA *et al* (2009). The diversity and distribution of toxigenic *Microcystis* spp. in present day and archived pelagic and sediment samples from Lake Erie. *Harmful Algae* **8:** 385-394.

Rozon RM, Short SM (2013). Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. *Freshw Biol* **58:** 2648-2663.

Saxton MA, Morrow EA, Bourbonniere RA, Wilhelm SW (2011). Glyphosate influence on phytoplankton community structure in Lake Erie. *J Great Lakes Res* **37:** 683-690.

Saxton MA, Arnold RJ, Bourbonniere RA, McKay RM, Wilhelm SW (2012). Plasticity of total and intracellular phosphorus quotas in *Microcystis aeruginosa* cultures and Lake Erie algal assemblages. *Front Microbiol* **3**.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Meth* **9:** 811-814.

Shapiro J (1997). The role of carbon dioxide in the initiation and maintenance of blue-green dominance in lakes. *Freshw Biol* **37:** 307-323.

Shen H, Niu Y, Xie P, Tao MIN, Yang XI (2011). Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshw Biol* **56:** 1065-1080.

Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW (2012). Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PLoS ONE* **7:** e44002.

Steffen MM, Belisle BS, Watson SB, Boyer GL, Wilhelm SW (2014a). Status, causes and controls of cyanobacterial blooms in Lake Erie. *J Great Lakes Res* **40:** 215-225.

Steffen MM, Dearth SP, Dill BD, Li Z, Larsen KM, Campagna SR *et al* (2014b). Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. *ISME J*.

Tabachek J-AL, Yurkowski M (1976). Isolation and identification of blue-green algae producing muddy odor metabolites, geosmin, and 2-methylisoborneol, in saline lakes in Manitoba. *Journal of the Fisheries Research Board of Canada* **33:** 25-35.

Tucker S, Pollard P (2005). Identification of cyanophage Ma-LBP and infection of the cyanobacterium *Microcystis aeruginosa* from an Australian subtropical lake by the virus. *Appl Environ Microbiol* **71:** 629-635.

Van de Waal DB, Verspagen JMH, Finke JF, Vournazou V, Immers AK, Kardinaal WEA *et al* (2011). Reversal in competitive dominance of a toxic versus non-toxic cyanobacterium in response to rising CO_2. *ISME J* **5:** 1438-1450.

Watson SB, Ridal J, Boyer GL (2008). Taste and odour and cyanobacterial toxins: impairment, prediction, and management in the Great Lakes. *Canadian Journal of Fisheries and Aquatic Sciences* **65:** 1779-1796.

Weitz JS, Wilhelm SW (2012). Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 biology reports* **4**.

Welschmeyer NA (1994). Fluorometric analysis of chlorophyll *a* in the presence of chlorophyll b and pheopigments. *Limnol Oceanogr* **39:** 1985-1992.

Wetzel RG, Likens GE (2000). *Limnological Analyses*, 3rd edn. Springer: New York.

Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki K *et al* (2008a). Ecological dynamics of the toxic bloom-forming cyanobacterium *Microcystis aeruginosa* and its cyanophages in freshwater. *Appl Environ Microbiol* **74:** 3269-3273.

Yoshida T, Nagasaki K, Takashima Y, Shirai Y, Tomaru Y, Takao Y *et al* (2008b). Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol* **190:** 1762-1772.

Chapter V Appendix

Table 5.4. Mean expression values (TCPM) of genes uniquely expressed at each station across all three replicates. S.D. denotes standard deviation.

Gene	Description	973 Mean	973 S.D.	882 Mean	882 S.D.	1163 Mean	1163 S.D.
cobB	cobyrinic acid a,c-diamide synthase	46.60	9.93	0.00	0.00	0.00	0.00
hisZ	ATP phosphoribosyltransferase regulator	70.18	43.81	0.00	0.00	0.00	0.00
MAE_11250	6-phosphogluconolactonase	57.53	22.72	0.00	0.00	0.00	0.00
MAE_12430	hypothetical protein	70.18	43.81	0.00	0.00	0.00	0.00
MAE_12460	iron(III)-transport ATP-binding protein	129.10	63.37	0.00	0.00	0.00	0.00
MAE_17230	hypothetical protein	28.73	2.71	0.00	0.00	0.00	0.00
MAE_18670	group 1 glycosyl transferase	37.70	11.67	0.00	0.00	0.00	0.00
MAE_18930	hypothetical protein	28.73	2.71	0.00	0.00	0.00	0.00
MAE_24070	ABC transporter ATP-binding protein	37.62	11.38	0.00	0.00	0.00	0.00
MAE_27680	efflux transporter MFP subunit	64.63	49.57	0.00	0.00	0.00	0.00
MAE_27810	short-chain dehydrogenase/reductase	28.73	2.71	0.00	0.00	0.00	0.00
MAE_28440	hypothetical protein	46.60	9.93	0.00	0.00	0.00	0.00
MAE_32340	hypothetical protein	37.62	11.38	0.00	0.00	0.00	0.00
MAE_32940	protein-tyrosine kinase	39.58	18.05	0.00	0.00	0.00	0.00
MAE_35820	major facilitator superfamily protein	28.73	2.71	0.00	0.00	0.00	0.00
MAE_42060	resolvase	28.73	2.71	0.00	0.00	0.00	0.00
MAE_45240	5'-methylthioadenosine phosphorylase	46.60	9.93	0.00	0.00	0.00	0.00
MAE_45770	transposase	28.73	2.71	0.00	0.00	0.00	0.00
MAE_46020	site-specific recombinase	129.10	63.37	0.00	0.00	0.00	0.00
MAE_47010	hypothetical protein	46.60	9.93	0.00	0.00	0.00	0.00
MAE_52340	hypothetical protein	55.65	36.89	0.00	0.00	0.00	0.00
MAE_58170	hypothetical protein	57.45	5.42	0.00	0.00	0.00	0.00
MAE_58980	hypothetical protein	66.43	11.23	0.00	0.00	0.00	0.00
MAE_59180	hypothetical protein	46.68	24.24	0.00	0.00	0.00	0.00
MAE_60890	putative glycosyl transferase	37.62	11.38	0.00	0.00	0.00	0.00

Table 5.4
Continued

Gene	Description	973 Mean	973 S.D.	882 Mean	882 S.D.	1163 Mean	1163 S.D.
ndhF3	NAD(P)H-quinone oxidoreductase subunit F	37.70	11.67	0.00	0.00	0.00	0.00
pstS	phosphate-binding periplasmic protein	28.73	2.71	0.00	0.00	0.00	0.00
rplV	50S ribosomal protein L22	39.58	18.05	0.00	0.00	0.00	0.00
rpsC	30S ribosomal protein S3	127.07	67.83	0.00	0.00	0.00	0.00
serS	seryl-tRNA synthetase	37.62	11.38	0.00	0.00	0.00	0.00
xisI	FdxN element excision controlling facto	39.58	18.05	0.00	0.00	0.00	0.00
accA	acetyl-CoA carboxylase carboxyltransferas	0.00	0.00	47.43	19.51	0.00	0.00
MAE_02560	Acr3 family arsenite transporter	0.00	0.00	95.16	39.99	0.00	0.00
MAE_06490	transposase	0.00	0.00	34.99	3.61	0.00	0.00
MAE_07890	HEAT repeat-containing protein	0.00	0.00	47.58	19.99	0.00	0.00
MAE_16760	hypothetical protein	0.00	0.00	59.87	36.98	0.00	0.00
MAE_17450	hypothetical protein	0.00	0.00	47.43	19.51	0.00	0.00
MAE_29610	transposase	0.00	0.00	47.43	19.51	0.00	0.00
MAE_32380	phosphate-binding protein PstS-like protein	0.00	0.00	34.99	3.61	0.00	0.00
MAE_35430	transposase	0.00	0.00	54.92	24.57	0.00	0.00
MAE_40350	hypothetical protein	0.00	0.00	34.99	3.61	0.00	0.00
MAE_41670	transposase	0.00	0.00	72.61	34.08	0.00	0.00
MAE_44440	hypothetical protein	0.00	0.00	47.58	19.99	0.00	0.00
MAE_56510	3-oxoacyl-ACP reductase	0.00	0.00	0.00	0.00	834.58	81.94
MAE_56520	non-ribosomal peptide synthetase	0.00	0.00	0.00	0.00	1431.36	606.08
MAE_56540	3-oxoacyl-ACP reductase	0.00	0.00	0.00	0.00	1422.40	1008.85
nirA	ferredoxin-nitrite reductase	0.00	0.00	0.00	0.00	292.86	54.18

CHAPTER VI:
CONCLUSION

This body of work has provided new insight into the structure and function of the microbial communities that comprise cyanobacterial harmful algal blooms in freshwater lakes around the world. This phenomenon is a widely studied and much publicized area, because it results in ecological (Paerl et al. 2001), economic (Glibert et al. 2005), and public health (Pouria et al. 1998) issues worldwide (Qin et al. 2010), often annually. A more comprehensive understanding of the microbial community dynamics of cHABs is not only compelling from an ecological perspective. This work can aid in the development of more robust and effective management practices as governments and private organizations attempt to abate the occurrence and expansion of toxic blooms.

In order to further our understanding of cHAB communities, we began with a targeted examination of the cyanobacterial population of a highly toxic bloom in Grand Lake St Marys, Ohio. We then expanded this effort by applying next generation sequencing to produce snapshots of cHAB-associated total microbial communities across two continents. Next, we narrowed our focus from entire bloom populations to the bloom forming organism *Microcystis aeruginosa,* broadening our understanding of this organism's molecular response to nutrient variability using transcriptomics. Finally, to validate a number of our hypotheses generated from the preceding studies, we surveyed the mRNA profiles of cHAB associated microbial communities across Lake Erie. In its entirety, this endeavor has generated a historical record of toxic bloom communities in freshwater lakes around the world. As our ability to analyze such data continues to develop, or if a consensus on how to handle such data is reached in the future, others may return to these publically available sequence records to generate further insight into bloom community dynamics.

While much of this work began as a series of pilot studies used to preliminarily characterize these communities, it has furthered the development of both established and novel hypotheses about the system, including:

1. Bloom microbial community function is independent of identity.
2. Nutrient dynamics drive changes to genomic architecture for *Microcystis*.
3. Activity and efficiency of carbon concentrating mechanisms during conditions of CO_{2aq} limitation and high pH contribute to bloom community speciation.

These hypothesis-based conclusions demonstrate one of the most appealing and frustrating aspects of the scientific method: there are always additional questions to be answered. Specifically, do the patterns of functional potential (from DNA) and transcription (from mRNA) translate to actual function of microbial community members? The hypothesis that variable microbial community structure is independent of function has been validated by numerous other studies (Burke et al. 2011, Oh et al. 2011). Further study is needed to clarify what factors dictate the functional role of individual organisms within a larger microbial community.

Our work demonstrated that transposase genes across the genome of *M. aeruginosa* NIES 843 are differentially transcribed during nutrient (N and P) variability. We were also able to validate this observation by accessing a proteomics dataset and mining environment transcriptomes for similar patterns. This hypothesis, however, represents an exciting area for future research, as more study is necessary to test whether this differential transcription of transposase genes results in changes to the genomic architecture of the organism.

Finally, it has been hypothesized that cyanobacteria have a competitive advantage over eukaryotic phytoplankton during high pH/low CO_{2aq} conditions due to more efficient carbon concentration mechanisms (Goldman and Shapiro 1973). What remains less clear is whether these conditions also drive speciation between cyanobacteria. While we were able to detect changes to transcription of genes involved in carbon concentration, we do not yet understand how this may impact bloom community dynamics (Paerl 1983). As atmospheric CO_2 levels continue to rise over the coming centuries, this will become an increasingly important question in cyanobacterial harmful algal bloom research. Overall, the studies presented herein not only have the potential to drive the field forward, but also better inform environmental managers and policy makers toward the critical goal of freshwater ecosystem health and conservation.

LIST OF REFERENCES

Burke C, Steinberg P, Rusch D, Kjelleberg S, Thomas T (2011). Bacterial community assembly based on functional genes rather than species. *Proceedings of the National Academy of Sciences* **108:** 14288-14293.

Glibert PM, Anderson DM, Gentien P, Graneli E, Sellner KG (2005). The global, complex phenomena of harmful algal blooms.

Goldman JC, Shapiro J (1973). Carbon dioxide and pH: effect on species succession of algae. *Science* **182:** 306-307.

Oh S, Caro-Quintero A, Tsementzi D, DeLeon-Rodriguez N, Luo C, Poretsky R *et al* (2011). Metagenomic insights into the evolution, function, and complexity of the planktonic microbial community of Lake Lanier, a temperate freshwater ecosystem. *Applied and Environmental Microbiology* **77:** 6000-6011.

Paerl HW (1983). Partitioning of CO2 Fixation in the Colonial Cyanobacterium Microcystis aeruginosa: Mechanism Promoting Formation of Surface Scums. *Applied and Environmental Microbiology* **46:** 252-259.

Paerl HW, Fulton RS, Moisander PH, Dyble J (2001). Harmful freshwater algal blooms, with an emphasis on cyanobacteria. *TheScientificWorldJOURNAL* **1:** 76-113.

Pouria S, de Andrade A, Barbosa J, Cavalcanti RL, Barreto VTS, Ward CJ *et al* (1998). Fatal microcystin intoxication in haemodialysis unit in Caruaru, Brazil. *The Lancet* **352:** 21-26.

Qin B, Zhu G, Gao G, Zhang Y, Li W, Paerl H *et al* (2010). A drinking water crisis in Lake Taihu, China: Linkage to climatic variability and lake management. *Environmental Management* **45:** 105-112.

Vita

Morgan Steffen was born in Des Moines, IA on October 19, 1987. She attended Lincoln High School and Central Academy in Des Moines and graduated in 2006. She received her Bachelor of Science degree in Biology from Iowa State University in May 2010 and began graduate school at the University of Tennessee in June 2010.