8-2018

MOLECULAR CHARACTERIZATION OF FACTORS CONSTRAINING THE SUCCESS AND TOXICITY OF *MICROCYSTIS* BLOOMS

Lauren Elisabeth Krausfeldt
University of Tennessee, lkrausfe@vols.utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Lauren Elisabeth Krausfeldt entitled "MOLECULAR CHARACTERIZATION OF FACTORS CONSTRAINING THE SUCCESS AND TOXICITY 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, Shawn R. Campagna, Karen G. Lloyd

Accepted for the Council:
Dixie L. Thompson

Vice Provost and Dean of the Graduate School

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

MOLECULAR CHARACTERIZATION OF FACTORS CONSTRAINING THE SUCCESS AND TOXICITY OF *MICROCYSTIS* BLOOMS

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

Lauren Elisabeth Krausfeldt
August 2018

DEDICATION

To my dad, my hero.

And my sister, forever my best friend.

ACKNOWLEDGEMENTS

I would first like the acknowledge Dr. Wilhelm for being an exceptional advisor. I am incredibly grateful for the tremendous research opportunities I have been given in graduate school. Thank you for your mentorship, support and advice in research and in life. Thank you for creating and encouraging a positive work environment. I am truly appreciative to have worked in such a collaborative atmosphere with people I call friends. To my committee members, I look up to you all with great admiration and so appreciate your guidance over the years.

I have to especially thank my undergraduate research advisor Dr. Tamara Marsh for giving me the opportunity to dive into research and always believing in me. I wouldn't be where I am today without her.

I have been lucky to have collaborated with many wonderful people – Drs. Boyer, Tang, Bullerjahn, McKay, Steffen, Bodrossy, Hector, Abigail and many more, thank you for all of your input on my work and the many valuable discussions. I also have to thank Elizabeth McPherson for not only being a tremendous teaching supervisor but also a great friend.

To my labmates, past and present – I am so grateful for our time commiserating over failed experiments, the celebration of our successes, and our talks about life and research in the lab and out. A special shout needs to go out to Gary LeCleir for all the help throughout the years!

I am so lucky to have such a wonderful support system in my family and friends. Thank you all for your love, encouragement and never-ending support even when distance sometimes keeps us apart. A special thank you goes to all of the friends I have made in Knoxville – I love you all, thank you for making Knoxville home. I have to especially thank my family for always making sure that I know how proud they are of me. And to my little family here in Knoxville – Bob, Theodore, and Rocky, I love coming home to you at the end of every day!

ABSTRACT

Harmful cyanobacterial blooms (cyanoHABs) have detrimental effects on freshwater lakes and reservoirs around the world. CyanoHABs severely reduce water quality, altering the food web and disrupting fishing and tourism industries. In addition, many bloom-forming cyanobacteria have the capacity to produce potent toxins, making the negative impacts of cyanoHABs of ecological and economic importance and a serious public health risk. *Microcystis* spp., which are the typically dominating cyanobacteria in blooms, often produce thick scums, taste and odor compounds and the hepatotoxins, microcystin, implicated in water advisories, human and animal poisonings and drinking water shutdowns globally. Primarily driven by anthropogenic nutrient loading and climate change, *Microcystis* blooms are on the rise, not only increasing in frequency but also intensity. Nitrogen (N) and phosphorus (P) are considered the most important nutrients in driving bloom formation and persistence but historically management practices have solely focused on P abatements. Large inputs of N into the environment have now been tied to the increased prevalence of toxic *Microcystis*. To better understand the role of N and other environmental factors impacting the success and toxicity of *Microcystis*, a number of molecular techniques were employed to characterize bloom community dynamics and the physiology of *Microcystis.* Results from environmental studies generated new hypotheses about the role of heterotrophic bacteria in N-cycling and microcystin degradation, while observations from laboratory studies provided novel insight into the metabolism of N by *Microcystis* and subsequent microcystin production was proposed. Together, the findings presented here can be extrapolated to the natural environment and provide greater insight into the mechanisms that contribute to cyanoHAB expansion and toxicity.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 : Introduction

Harmful cyanobacterial blooms

Cyanobacteria are one of the most abundant organisms by mass on Earth (Paerl & Otten, 2013). They are important primary producers in freshwater and marine ecosystems and play a large role in the generation of oxygen through photosynthesis. Though beneficial in several ways, they can form dense blooms that have negative consequences on ecosystem and public health (Paerl & Otten, 2013, Paerl, 2014, Bullerjahn *et al.*, 2016). Harmful cyanobacterial blooms (often termed cyanoHABS) are defined by an increase in cyanobacterial biomass dominated by one or a few species (Meriluoto *et al.*, 2017). CyanoHABs are a growing threat to freshwater resources around the world (Brooks *et al.*, 2016) and have been reported in some of the world's largest freshwater reservoirs (Carmichael, 2008, Le *et al.*, 2010, Bullerjahn *et al.*, 2016). The term harmful comes from the consequences of cyanobacterial overgrowth including the alteration of the food web, and generation of hypoxic conditions, but many also have the capacity to produce potent toxins (Paerl, 2018). Indeed, cyanoHABs have been linked to massive fish kills, wild and domesticated animal poisonings, reproductive defects in fish, water advisories and drinking water shutdowns all over the world (Oberemm *et al.*, 1999, Bullerjahn *et al.*, 2016, Carmichael & Boyer, 2016). CyanoHABs are also an economic concern because they hinder tourism and fishing industries, and these economic costs can only be presumed to be much greater during times of water restriction to domestic households and businesses. On a larger scale, the economic losses from impaired water quality for recreation alone can reach over one billion dollars from individual events (Dodds *et al.*, 2008, Steffen *et al.*, 2017). For these reasons, the alarming increase in cyanoHABs frequency and intensity globally is a major environmental and public health concern (Brooks *et al.*, 2016).

Globally, the most prevalent bloom forming cyanobacteria in freshwater systems are the non-N_2 fixing *Microcystis* and the N_2-fixing filamentous *Anabaena* and *Cylindrospermopsis*,

although other notable bloom formers include *Planktothrix, Nostoc*, and *Aphanizomenon* (Paerl & Fulton III, 2006, O'neil *et al.*, 2012). Different factors promote the dominance of these cyanobacteria, including temperature, light intensity, and nutrient concentrations (Bonilla *et al.*, 2012, O'neil *et al.*, 2012, Cirés & Ballot, 2016). One distinct commonality between these cyanobacteria is the ability to produce the aforementioned toxins (Meriluoto *et al.*, 2017). These toxins, known as cyanotoxins, include neurotoxins, hepatotoxins and cytotoxins that are highly toxic to mammals, including humans, and some cyanobacteria have the capability to produce more than one (Table 1) (Pantelic *et al.*, 2013, Meriluoto *et al.*, 2017). Mortality and morbidity cases have been reported due to cyanotoxins since the 1800s and have involved birds, livestock, dogs, fish and even humans (Chorus & Bartram, 1999, Wood, 2016). The number of cases has risen gradually over time, likely due to the increase in toxic blooms. Increased awareness about the dangers of cyanobacterial blooms has also presumably contributed the increase in reports, but it is predicted that the number of individuals affected is greatly underestimated (Wood, 2016).

Microcystins: Toxicity and risks

The hepatotoxins, microcystins (MCs), are the most commonly detected cyanotoxins around the world and are produced by several bloom-forming cyanobacteria (Table 1.1). MCs were first isolated more than half a century ago and was originally known as "*Fast Death Factor*" (Bishop *et al.*, 1959). MCs are cyclic heptapeptides, comprised of five common amino acids with two variable amino acid side chains (Figure 1). There are currently over 130 described congeners

Table 1.1 Prevalent bloom-forming cyanobacteria

List of most prevalent toxin producing bloom-forming cyanobacteria, physiological characteristics and cyanotoxins produced.

Toxin producing genera	Morphology	Cyanotoxins produced
Microcystis spp.	Coccoid/colonial	Microcystins
Anabaena spp.	Heterocystous/filamentous	Microcystins, Cylindrospermopsins, Anatoxins,
Nostoc spp.	Heterocystous/filamentous	Microcystins, Nodularins
Aphanizomenon spp.	Heterocystous/filamentous	Cylindrospermopsins, Anatoxins
Planktothrix spp.	Non-heterocystous/filamentous	Microcystins
Cylindrospermopsis spp.	Heterocystous/filamentous	Cylindrospermopsins

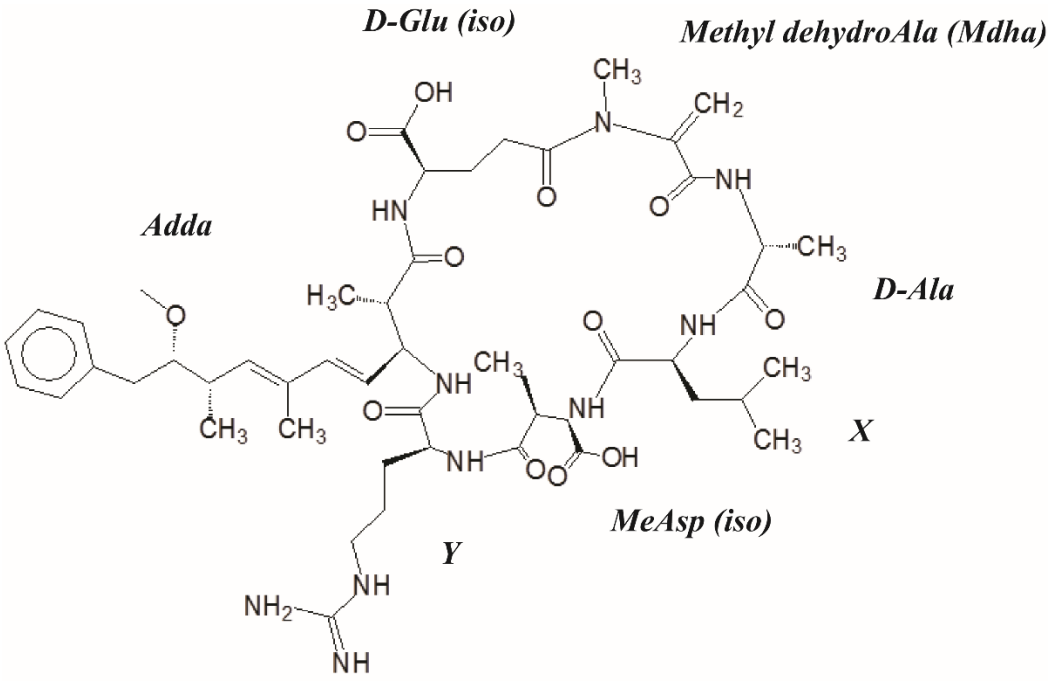

Figure 1.1 The general structure of microcystins

Microcystin-LR is shown here with leucine in the X position and arginine in the Y position.

of MCs, but the most prevalent are MC-LR and MC-RR, with leucine and arginine or two arginines in the X and Y positions (Carmichael & Boyer, 2016). While the physiological role of MCs within the cyanobacteria that produces them remains to be fully elucidated, these compounds are highly toxic to mammals, including humans, with an LD_{50} ranging from 25 - >1000 μg/kg. The cyclic structure of MCs and the -Adda group play significant roles in toxicity of MCs, but not all variants have the same level of toxicity (Zurawell *et al.*, 2005). MCs are potent protein phosphatase 1 and 2A inhibitors and long-term exposure can lead to hepatocytosis, liver and colon cancer and other gastrointestinal problems and direct skin contact can lead to rashes and blistering (Zurawell *et al.*, 2005).

Exposure to cyanotoxins usually arises from ingestion of contaminated water, food (*i.e.* fish) or through contact during recreational water activities (Pantelic *et al.*, 2013, Wood, 2016). MCs are thought to remain intracellularly within cyanobacteria until they are released during cell death, viral lysis or bloom senescence (Young *et al.*, 2005, Steffen *et al.*, 2017). The World Health Organization has deemed concentrations greater than 1 μg/L MC to be unacceptable in drinking water (Carmichael & Boyer, 2016) and 20 μg/L for recreational waters (Chorus & Bartram, 1999), while current US-EPA guidelines for finished drinking water are 0.3 μg/L for children up to the age of 6, and 1.6 μg/L per day for adults over a 10-day period (US-EPA, 2015). However, concentrations in a bloom can well-exceed this limit. For example, concentrations have reached up to 7300 μg/g of dry weight and up to 26.5 mg/L in lakes in China and New Zealand, respectively (Pham & Utsumi, 2018).

Monitoring cyanotoxins is not always a routine procedure in local jurisdictions worldwide, including in the United States (Bullerjahn *et al.*, 2016). Indeed, illnesses have been reported due to MC and other cyanotoxin exposure in humans and animals and increases in liver cancer and

colorectal cancer been linked to nearby bodies of water plagued with toxic cyanobacteria in the United States, Serbia and China (Carmichael, 2001, Hernández *et al.*, 2009, Svircev *et al.*, 2009, Carmichael & Boyer, 2016). A lack of reporting and monitoring makes it difficult to assess the real risk of MCs to the public and the highest exposure risks, however efforts are being put forth to improve upon reporting and confirming cases of cyanotoxin illnesses (Backer *et al.*, 2015, Hilborn & Beasley, 2015). The only confirmed case of severe illness and fatalities from MC exposure was in Brazil in 1996. 130 patients at a dialysis clinic were treated with contaminated water, 116 of which suffered visual impairments, vomiting and nausea, 100 developed acute liver failure, and ultimately 70 patients died (Azevedo *et al.*, 2002).

There have been several highly publicized instances where MC was an immediate risk to public health. Drinking water shutdowns have been reported worldwide due to MCs, including in the United States. Before the relatively non-toxic bloom event in 2015 on Lake Erie, the 2011 bloom event was largest on record (Michalak *et al.*, 2013, Wines, 2013). The most recent major problem occurred in Toledo, OH (USA) when a water shutdown in 2014 left 500,000 people without access to drinking water for 3 days (Bullerjahn *et al.*, 2016). Other instances have been more severe; ten million people in the city of Wuxi (China) did not have access to drinking water for 30 days in 2007 (Qin *et al.*, 2010). In June and July of 2016, a state of emergency was declared in Florida because a cyanobacterial bloom positive for MCs migrated down Lake Okeechobee waterways and accumulated on beach severely impacting tourism (Graham *et al.*, 2016, Oehrle *et al.*, 2017). Less publicized events have also occurred in the United States, such as a 2-day drinking water shutdown Carroll Township, OH in 2013, illnesses and water restrictions for the areas surrounding Utah Lake, UT in 2016 and water advisories for five states along the Ohio River in 2015 (Fitzsimmons, 2014, Brooks *et al.*, 2016, Graham *et al.*, 2016). Conventional municipal water

treatments are effective at removing or inactivating cyanobacterial cells, but these methods risk the release of MCs from the cell. MCs are not removed by flocculation processes and require secondary treatments that include the use of activated carbon, which is effective but costly (He *et al.*, 2016). Other methods such as membrane processes, UV light, and the addition of hydrogen peroxide are promising in the removal of MC, but more research is needed (He *et al.*, 2016, Kansole & Lin, 2017).

The fate of microcystins in the environment

MC is chemically stable and resistant to many abiotic factors, which has led to numerous studies questioning its fate in the environment (Harada & Tsuji, 1998). Its disappearance has been attributed to some abiotic effects including dilution, adsorption, and photodegradation (Harada & Tsuji, 1998, Corbel *et al.*, 2014). Adsorption to sediments varies depending on sediment composition and chemistry but has been estimated to be responsible for up to 20% of the disappearance of MC from the water column in environmental samples. Photodegradation also varies greatly depending on the presence of water soluble pigments, humic substances, the presence of metal ions and radical oxygen species like hydrogen peroxide (Welker & Steinberg, 2000, Wörmer *et al.*, 2010, Schmidt *et al.*, 2014, Dai *et al.*, 2017). MCs are also reported to be resistant to the activity of common peptidases (Tsuji *et al.*, 1994, Mazur & Plinski, 2001, Smith *et al.*, 2010). The cyclic nature of the compounds ring structure makes it resistant to degradation by many exopeptidases, which require a free carboxyl or amino end of the peptide chain for initial binding. At the same time, the alternating incorporation of the non-protein "R" stereoisomers of the amino acid, coupled with "iso" peptide bonds formed through the side chain of glutamate and aspartate, make the molecule resistant to degradation by traditional serine peptidases such as

trypsin and chymotrypsin (Smith *et al.*, 2010). Selective methylation of the amide nitrogen in some peptide linkages (*e.g.*, methyl DHA or methyl isoaspartate) may also increase the resistance of these linkages to hydrolysis. Despite these challenges, the current literature often refers to biodegradation as the most important route for the disappearance of MCs in nature.

Natural microbial communities from lakes, drinking water reservoirs, water treatment facilities, soils, sediments, and biofilters have been shown to degrade MCs (Li *et al.*, 2017). Typically, these studies focused on microbial communities that have been previously exposed to MCs. However, pre-exposure to the toxin in other studies was not required but shortened the lag periods and increased the degradation efficiency (Rapala *et al.*, 1994, Christoffersen *et al.*, 2002, Edwards *et al.*, 2008). Biodegradation of MC in nature occurs in a wide range of oxygen gradients, nutrient concentrations and temperatures, (Li *et al.*, 2017, and citations within), but identifying universal factors that constrain or promote degradation has been difficult, in part because of the variability between systems. In many studies, it also becomes difficult to tease apart the differences between biodegradation, adsorption to sediment particles or other materials, photodegradation and other potential factors that are not mechanistically understood to date, such as radical oxygen species or metals precipitation (Kansole & Lin, 2017). This makes discerning the total contribution to the disappearance of MCs via biodegradation in the environment unclear.

MC biodegradation is well studied in lab settings. Over fifty MC-degrading microorganisms have been reported to date (see Chapter 6). The capacity for MC biodegradation is not conserved within any one genera or species but spans across diverse groups of microorganisms. The first reported MC-degrader was MJ-PV, an Alphaproteobacteria now known as *Sphingomonas* ACM3962 (Jones *et al.*, 1994, Bourne *et al.*, 1996, Bourne *et al.*, 2001). Additional MC-degrading Alphaproteobacteria have been isolated, but microorganisms capable of

this process extend throughout the Beta- and Gammaproteobacteria, Actinobacteria, and Firmicutes. This capability has also been demonstrated in eukaryotic microorganisms, like fungi and algae. Rates of biodegradation differ between strains, ranging anywhere from hours to weeks, when MC is the sole C and N source (Dziga *et al.*, 2013). Rates and likelihood for complete removal also depend on initial concentrations of MCs, and not all MC degraders appear to completely degrade MC (Nybom *et al.*, 2012). Rates also vary between strains under differing temperatures, pH ranges and nutrient concentrations, which supports the hypothesis that MC degradation rates vary in the environment because of the differences in microbial community composition.

Bourne *et al*. (1996) characterized a genetic pathway involved in MC-LR degradation in *Sphingomonas* ACM3962, which included four enzymes: *mlrA, mlrB, mlrC, mlrD* (Bourne *et al.*, 1996, Bourne *et al.*, 2001). To broaden the understanding about the mechanisms of MC biodegradation, it is important to not only identify microorganisms (*e.g.,* at the level of the 16S rRNA gene sequence), but to develop a broader understanding of the mechanisms involved in the degradation of this compound. This may involve detecting the genes involved in the *mlrA* pathway or an alternate pathway. The gene *mlrA* has been used as a marker for the *mlrABCD* pathway because it catalyzes a rate-limiting step of MC degradation and substantially reduces toxicity. PCR amplification of *mlrA* has been used to determine if *mlrABCD* pathway is the route of MC degradation in lab isolates or in environments where MCs persist. However, standard PCR-based assays are not quantitative, and the presence of *mlrA* gene in the DNA signifies only the potential for degradation *via* this pathway, not that the pathway is active or complete (Kainz, 2000, Ouellette & Wilhelm, 2003). The presence of *mlrA* implies the capacity for degradation of MCs, but additional assays are needed to confirm activity. As of 2017, less than half of putative MC-

degrading bacteria from cultures reported in the literature had been screened for the presence of the *mlrA* gene, leaving gaps in a complete understanding of the mechanism(s) by which MC is degraded by various bacteria. Several studies suggest the *mlrABCD* pathway may not be the only mechanism for MC degradation (Nybom *et al.*, 2008, Manage *et al.*, 2009, Wu *et al.*, 2013). MC-degrading bacteria including *Arthrobacter, Brevibacterium*, *Rhodococcus*, and MC-degrading probiotic bacteria have all been *mlrA* negative; researchers have not been able to amplify the *mlrA* sequence from these cultures. This inability to amplify *mlrA* could be attributed to large differences in *mlrA* sequences between the known and unknown degraders. The primers commonly used (Saito *et al.*, 2003) are non-degenerate and based strictly on the *mlrA* sequence of *Sphingomonas* ACM3962. These primers may thus be too specific to capture true diversity of this gene. Degenerate qPCR primers have been designed targeting *mlrA* but they yield fragments that are too short to be useful for assessing diversity of this gene sequence (Hoefel *et al.*, 2009). Further complicating the understanding of this process is the lack of sequenced genomes of MC-degrading microorganisms with only one MC-degrader's genome sequenced to date, which is confirmed *mlrA*+ (Okano *et al.*, 2015).

The environmental relevance and distribution of the *mlr* cassette in the environment remains unclear. A synopsis of the literature suggests the distribution of the *mlrABCD* pathway varies not only across phylogenetic groups but also environments (see Chapter 6). The *mlrA* gene has been amplified by PCR from DNA isolated from environmental samples with MC degradation capabilities: these include biofilms, filtration sand and water samples from treatment plants used in laboratory scale studies in Japan and Australia where the microcystin degrading community was likely enriched. However, several studies have observed biodegradation that was not attributed to *mlrA* (Mou et al., 2013, Lezcano et al., 2016). With the exception of a few studies, there is a large

focus on studying *mlrA* as it is a marker gene for this pathway, yet *mlrBCD* is rarely explored in environmental samples. These observations and gaps in our knowledge of the relevance of this pathway highlight the potential problem in assuming biodegradation is attributed solely to the *mlr* pathway or that it is most important mechanism for biodegradation when the *mlrA* is amplified or detected in environmental DNA. This complicates the understanding of the fate of MCs in the environment and ability to assess the risk to public health, especially if this is truly the most important route of removal of MCs.

Microcystis and the microcystin production

Microcystis spp. are the prevalent bloom-formers and MC producers that, around the world, are often associated with detectable amounts of MCs (Figure 2). *Microcystis* is a non-N_2 fixing, buoyant and coccoid cyanobacteria that can form colonies in nature that result in thick scums and potent taste and odor compounds (Figure 3, Paerl & Fulton III, 2006, Harke *et al.*, 2016, Watson *et al.*, 2016). The proliferation of *Microcystis* is thought to have increased in the past two decades and it has now been detected on all continents with the exception of Antarctica (Harke *et al.*, 2016). While *Microcystis* is an important MC producer, not all species produce MC. The function of MCs in *Microcystis* and other cyanobacteria is not well understood but in general, MC is assumed to provide an advantage to toxic cells (Gobler *et al.*, 2016, Harke *et al.*, 2016). Toxic *Microcystis* strains have been observed to have higher growth rates and are better adapted to grow in high light conditions (Gobler *et al.*, 2016) compared to non-toxic strains. However, non-toxic strains are able to persist under nutrient limitation and even outcompete toxic strains during bloom events (Gobler *et al.*, 2016). In accordance with this, toxic *Microcystis* tend to persist early in bloom events when nutrient concentrations are presumably high, and there is shift to dominance of non-toxic strains

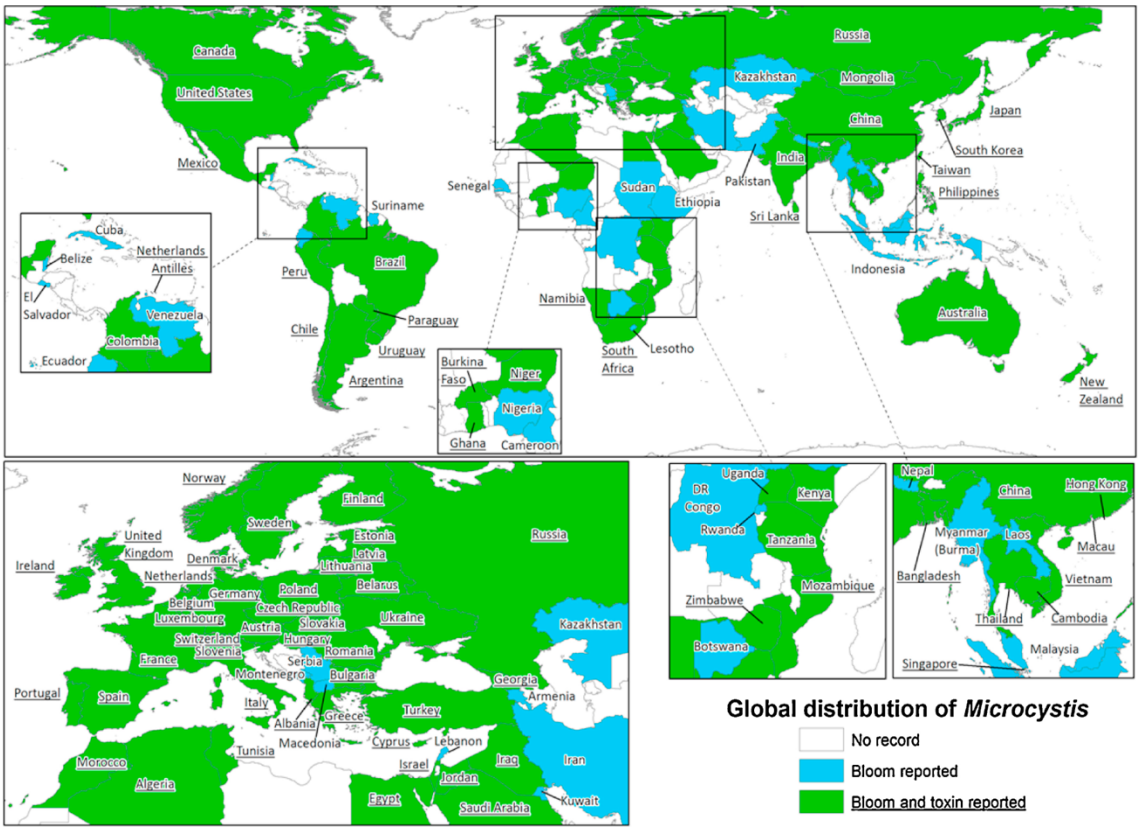

Figure 1.2 Global distribution of Microcystis and cyanotoxins reports

The global distribution of *Microcystis*. In blue are countries in which *Microcystis* blooms have been reported and in green are countries in which both *Microcystis* and microcystins have been detected (figure from Harke, *et al*. 2016).

Figure 1.3 *Microcystis* as single cells, colonies and the formation of scum

Microcystis cells under a bright light microscope (a, photo credit to Barry Rosen, USGS), colonies in nature (b, photo credit to Lauren E. Krausfeldt), and thick surface scums of *Microcystis* (c, photo credit to Lauren E. Krausfeldt).

during the peak of the blooms (Davis *et al.*, 2010). Because MCs are N rich and costly to produce (Tillett *et al.*, 2000), N availability has been argued to drive this species succession over a bloom event (Gobler *et al.*, 2016).

The physiological role of MCs and drivers of its production has been heavily debated over the past several decades. Hypothesis have included MC production is influenced by growth rate, high N, high light, temperature, grazing, and oxidative stress while roles in quorum sensing, colony formation and cell signaling have also been proposed (Harke *et al.*, 2016). Current research efforts have focused largely on the role of N because MCs are amino acid and N-rich. Generally, high MC concentrations are observed in association with high N concentrations in culture and field studies (Gobler *et al.*, 2016). However, the opposite has also been proposed and is largely supported by the observation that NtcA, the global nitrogen regulator, can bind to the promotor sequences for MC production (Ginn *et al.*, 2010). This implies that in times of high N availability, NtcA production would increase and block the transcription of MC genes. Laboratory studies have supported this observation as well (Ginn *et al.*, 2010, Kuniyoshi *et al.*, 2011, Pimentel & Giani, 2014, Peng *et al.*, 2018). The role of oxidative stress has also been of interest recently (Zilliges *et al.*, 2011, Pimentel & Giani, 2014) and may provide an explanation for the number of variables that affect MC production since cyanobacteria are subjected to a number of factors that generate reactive oxygen species (*i.e.* photosynthesis, high light, nutrient limitation).

Factors that influence the success and persistence of *Microcystis* blooms

Cyanobacteria as a group are well-adapted to handle a variety of environmental stressors, such as desiccation, low nutrient availability, and high irradiance, and are able to persist for several months in some lakes (Paerl, 2018). They also "over-winter" which likely plays an important role

in the annual recurrence of cyanoHABS globally (Preston *et al.*, 1980). Anthropogenic nutrient overloading is the primary cause of increasing cyanoHABs. Nutrient loading is a result of terrestrial derived run-off into freshwater ecosystems from surrounding areas, due to agricultural, farming and industrial practices (Paerl *et al.*, 2011, Paerl *et al.*, 2016). While still problematic, point source pollution can be more easily mitigated while non-point source is more difficult to manage (Paerl, 2018). Growing populations, increased fertilizer use, and destruction of wetlands has led to non-point source pollution contributing more than >50% of the nutrient input into freshwater lakes (Heisler *et al.*, 2008, Glibert *et al.*, 2014, Paerl, 2018). Climate change is predicted to exacerbate this problem and specifically benefit *Microcystis* spp., bringing longer periods of precipitation with subsequent rises in nutrient inputs (Paerl & Huisman, 2009). Climate change is also expected to bring longer periods of drought, and along with higher temperatures promotes the growth of cyanobacteria by increasing stratification of the water column which in turn reduces mixing and increases residence time of nutrients. Further, increased CO_2 emissions may promote the increase of blooms by enrichment of CO_2 in the surface waters needed for photosynthesis (O'neil *et al.*, 2012, Sandrini *et al.*, 2015).

The ability of *Microcystis* spp. specifically to form blooms and proliferate is multifaceted. *Microcystis* spp. have higher photosynthetic efficiencies and growth rates at high temperatures and outcompete other phytoplankton and cyanobacterial species (Paerl & Huisman, 2008, Visser *et al.*, 2016). *Microcystis* is buoyant, providing the ability to migrate vertically within the water column by the use of gas vesicles (Harke *et al.*, 2016). This is advantageous during times of nutrient limitation in the surface waters to scavenge nutrients near the sediments or evade high irradiance. In addition, *Microcystis* spp. have a variety of high affinity N and P transporters that aid in scavenging nutrients as well several storage mechanisms for N (cyanophycin) and P

(polyphosphate bodies). Over the course of a bloom event, the pH of the water often becomes more alkaline, which also promotes the growth of *Microcystis* and cyanobacteria over other phytoplankton species. Cyanobacteria have highly efficient carbon concentrating mechanisms that allow them to thrive under low dissolved CO_2 conditions present at these high pHs (Price, 2011). Many *Microcystis* spp. prefer these conditions while others prefer higher CO_2 concentrations implying this is likely species or strain specific adaptation (Sandrini *et al.*, 2014). Colony formation is also thought to promote the success of *Microcystis* spp. by becoming more buoyant, protecting cells against environmental stressors or promoting the association with heterotrophic bacteria (Xiao *et al.*, 2018). While antagonistic relationships can exist, interactions with heterotrophs are generally considered mutually beneficial (Jiang *et al.*, 2007, Gumbo *et al.*, 2008, Shen *et al.*, 2011, Kang *et al.*, 2012, Wang *et al.*, 2016). Associated heterotrophs can aid in the recycling or transfer of nutrients, like P, to *Microcystis* and has been proposed to be involved in the cycling of N in several studies (Chen *et al.*, 2012, Steffen *et al.*, 2012, Krausfeldt *et al.*, 2017).

Nitrogen and phosphorus: Has the debate come to an end?

While nutrient over-loading is well understood to drive problematic cyanoHABS, remediation strategies have primarily focused on P (Schindler *et al.*, 2008). P as the primary limiting nutrient in freshwaters has been a long-standing and stable paradigm in limnology (Wetzel, 2001, Lewis & Wurtsbaugh, 2008). Following long-term experiments conducted on the Experimental Lakes in the 1970s with N and P additions, it was concluded that P was the limiting factor driving cyanobacterial blooms, and the call for P reductions began (Schindler, 1974, Schindler *et al.*, 2008). This observation stemmed from P stimulating the growth of N_2-fixing cyanobacteria, which can produce their own N. Therefore, N was not considered a priority for

reductions although incidentally, around the same time, other researchers had noted the accumulation of more biomass from both N and P additions (Lewis & Wurtsbaugh, 2008). However, P became the main target for reduction to remediate freshwater lakes from cyanoHABs and P abatements were implemented in the United States and Europe (Makarewicz & Bertram, 1991, Conley *et al.*, 2009), a major one including the improvement of municipal waste facilities. This to some extent did help improve water quality (Makarewicz & Bertram, 1991, Jeppesen *et al.*, 2005).

In the past few decades however, cyanoHABs have expanded globally and a shift from N_2-fixing cyanobacteria to non-N_2-fixing cyanobacteria, like *Microcystis* spp., took place. Evidence of non-N_2-fixers dominating bloom communities drew the focus away from the P-only paradigm (EPA, 2015). The increased use in N fertilizers, like urea, fossil fuel combustion, legume harvesting, domestics wastes, and storm water runoff have substantially increased anthropogenic N inputs into the environment and led to great interest in the role of N in cyanoHABs and rise of *Microcystis.* Indeed, human activities have doubled inputs of N into the environment and increased atmospheric deposition on land by ~46 Tg/N/year (Schlesinger, 2009). There is mounting evidence from field and laboratory studies that increased N inputs, more so than P, increase the growth of *Microcystis* (Gobler *et al.*, 2016). While the role of N in the toxicity of blooms remains unclear, higher levels of MCs are generally associated with high exogenous N availability (Gobler *et al.*, 2016), and this alone is enough to raise concern.

It has been argued that N-only abatements have aided in mitigating eutrophic water systems (Paerl *et al.*, 2016), but seasonal influences of both N and P likely contribute to the proliferation of large blooms (Xu *et al.*, 2010). TN:TP ratios change over a bloom event with the drawn down of one nutrient of addition of "new" nutrients from run off or sediment resuspension. The ability

of *Microcystis* to be especially tolerant of low P conditions makes the addition of "new" N an even larger concern This is further exacerbated by the ability of *Microcystis* to thrive in low P conditions (Gobler *et al.*, 2016), but the same can be said for periods of low N as well. The lowest TN:TP ratios or inorganic N concentrations are observed during the peak of the bloom event, suggesting N is drawn down early, potentially stimulating the bloom, and eventually becomes limiting (Gobler *et al.*, 2016). Seasonal changes in nutrient dynamics which at times involved the co-limitation of both N and P, (Xu *et al.*, 2010, Belisle *et al.*, 2016) indicate that both nutrients are important drivers in the persistence of blooms. These observations highlight a dual management strategy for both N and P need to be reduced in order to mitigate cyanoHABs from freshwater resources.

The research presented in this dissertation addresses several aspects related to the success and toxicity of harmful cyanobacterial blooms dominated by *Microcystis* spp. on both a molecular and ecological level. We broadly investigated interactions within the community involved in nutrient transformations, with a specific focus on N. On a community level, we examined the N-cycling community associated with *Microcystis* and explored dynamic interactions within the colonial fractions in blooms. At a higher resolution, the effects of specific N sources on the metabolome of *Microcystis* and microcystin production were elucidated to extrapolate the importance of N-form on physiological responses of *Microcystis* in nature. We also aimed to broaden the understanding of the general toxicity of blooms and explored the removal of microcystins by biological degradation by the associated microbial community. Our observations provided insight into the physiological mechanisms and ecological aspects that contribute to the toxicity of cyanoHABs and proliferation of *Microcystis* globally.

The first goal of this work was to identify which biological N-cycling processes were occurring in the surface waters and how they might impact the *Microcystis* community. This was

an essential question to address because N-cycling processes can be responsible for the introduction of new N or the loss of N from the system. While the importance of these processes is clear on a global scale, the N-cycling community within cyanobacterial bloom communities has received less attention. We found evidence to suggest that biological N-cycling may be important source of new N or recycled N from the heterotrophic community to the cyanobacterial community. Because heterotrophic bacteria form tight associations with *Microcystis* colonies, our N-cycling work led to follow up experiments to investigate how colonies might be hotspots for N transformations during blooms. By taking a metatranscriptomic approach, we were able to not only screen for N-transformations, but also ask questions about other heterotrophic interactions and physiological advantages of *Microcystis* colony formation in nature. In the next set of chapters, our aim was directed towards investigating the physiological response to different N forms of *Microcystis.* We asked specific questions about the fate of nitrate, ammonium and urea and how N-species influenced microcystin production. By taking a metabolomics approach, we used stable labeled isotopes to achieve high resolution snapshots of *Microcystis*' N metabolism and potential routes of N that lead to microcystin production. On a larger ecological scale, the final chapter examines the fate of microcystin in nature by the most important mechanism of removal, biological degradation. Taken together, these studies provide insight into additional abiotic and biotic factors influencing the proliferation and success of *Microcystis* and the toxicity of blooms.

References

Azevedo S. M., Carmichael W. W., Jochimsen E. M., Rinehart K. L., Lau S., Shaw G. R. & Eaglesham G. K. 2002. Human intoxication by microcystins during renal dialysis treatment in Caruaru—Brazil. *Toxicology*. 181, 441-446.

Backer L. C., Manassaram-Baptiste D., LePrell R. & Bolton B. 2015. Cyanobacteria and algae blooms: Review of health and environmental data from the harmful algal bloom-related illness surveillance system (HABISS) 2007–2011. *Toxins*. 7, 1048-1064.

Belisle B. S., Steffen M. M., Pound H. L., Watson S. B., DeBruyn J. M., Bourbonniere R. A., Boyer G. L. & Wilhelm S. W. 2016. Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*. 42, 599-607.

Bishop C. T., Anet E. F. L. J. & Gorham P. R. 1959. Isolation and identification of the Fast-Death Factor in *Microcystis aeruginosa* NRC-1. *Can J Biochem Physiol*. 37, 453-471.

Bonilla S., Aubriot L., Soares M. C., Gonzalez-Piana M., Fabre A., Huszar V. L., Lurling M., Antoniades D., Padisak J. & Kruk C. 2012. What drives the distribution of the bloom-forming cyanobacteria *Planktothrix agardhii* and *Cylindrospermopsis raciborskii*? *FEMS Microbiol Ecol*. 79, 594-607.

Bourne D. G., Riddles P., Jones G. J., Smith W. & Blakeley R. L. 2001. Characterisation of a gene cluster involved in bacterial degradation of the cyanobacterial toxin microcystin LR. *Environ Toxicol*. 16, 523-534.

Bourne D. G., Jones G. J., Blakeley R. L., Jones A., Negri A. P. & Riddles P. 1996. Enzymatic pathway for the bacterial degradation of the cyanobacterial cyclic peptide toxin microcystin LR. *Applied and Environmental microbiology*. 62, 4086-4094.

Brooks B. W., Lazorchak J. M., Howard M. D., Johnson M. V. V., Morton S. L., Perkins D. A., Reavie E. D., Scott G. I., Smith S. A. & Steevens J. A. 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environmental Toxicology and Chemistry*. 35, 6-13.

Bullerjahn G. S., McKay R. M., Davis T. W., Baker D. B., Boyer G. L., D'Anglada L. V., Doucette G. J., Ho J. C., Irwin E. G. & Kling C. L. 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. *Harmful Algae*. 54, 223-238.

Carmichael W. 2008. A world overview—One-hundred-twenty-seven years of research on toxic cyanobacteria—Where do we go from here? *Cyanobacterial harmful algal blooms: State of the science and research needs,* 619, 105-125. Springer.

Carmichael W. W. 2001. Health effects of toxin-producing cyanobacteria:"The CyanoHABs". *Human and ecological risk assessment: An International Journal*. 7, 1393-1407.

Carmichael W. W. & Boyer G. L. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae*. 54, 194-212.

Chen X., Yang L., Xiao L., Miao A. & Xi B. 2012. Nitrogen removal by denitrification during cyanobacterial bloom in Lake Taihu. *Journal of Freshwater Ecology*. 27, 243-258.

Chorus I. & Bartram J. 1999. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management. World Health Organiziation

Christoffersen K., Lyck S. & Winding A. 2002. Microbial activity and bacterial community structure during degradation of microcystins. *Aquatic Microbial Ecology*. 27, 125-136.

Cirés S. & Ballot A. 2016. A review of the phylogeny, ecology and toxin production of bloom-forming Aphanizomenon spp. and related species within the Nostocales (cyanobacteria). *Harmful Algae*. 54, 21-43.

Conley D. J., Paerl H. W., Howarth R. W., Boesch D. F., Seitzinger S. P., Karl E., Karl E., Lancelot C., Gene E. & Gene E. 2009. Controlling eutrophication: nitrogen and phosphorus. *Science*. 123, 1014-1015.

Corbel S., Mougin C. & Bouaïcha N. 2014. Cyanobacterial toxins: modes of actions, fate in aquatic and soil ecosystems, phytotoxicity and bioaccumulation in agricultural crops. *Chemosphere*. 96, 1-15.

Dai G., Peng N., Zhong J., Yang P., Zou B., Chen H., Lou Q., Fang Y. & Zhang W. 2017. Effect of metals on microcystin abundance and environmental fate. *Environmental Pollution*. 226, 154-162.

Davis T. W., Harke M. J., Marcoval M. A., Goleski J., Orano-Dawson C., Berry D. L. & Gobler C. J. 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Aquatic Microbial Ecology*. 61, 149-162.

Dodds W. K., Bouska W. W., Eitzmann J. L., Pilger T. J., Pitts K. L., Riley A. J., Schloesser J. T. & Thornbrugh D. J. 2008. Eutrophication of US freshwaters: analysis of potential economic damages. *Environmental Science & Technology*. 43, 12-19.

Dziga D., Wasylewski M., Wladyka B., Nybom S. & Meriluoto J. 2013. Microbial degradation of microcystins. *Chemical research in toxicology*. 26, 841-852.

Edwards C., Graham D., Fowler N. & Lawton L. A. 2008. Biodegradation of microcystins and nodularin in freshwaters. *Chemosphere*. 73, 1315-1321.

EPA. 2015. Preventing Eutrophication: Scientific Support for Dual Nutrient Criteria

Fitzsimmons E. G. 2014. Tap water ban for Toledo residents. *New York Times*. 3, A12.

Ginn H., Pearson L. & Neilan B. 2010. NtcA from *Microcystis aeruginosa* PCC 7806 is autoregulatory and binds to the microcystin promoter. *Applied and environmental microbiology*. 76, 4362-4368.

Glibert P. M., Maranger R., Sobota D. J. & Bouwman L. 2014. The Haber Bosch–harmful algal bloom (HB–HAB) link. *Environmental Research Letters*. 9, 105001.

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

Graham J. L., Dubrovsky N. M. & Eberts S. M. 2016. *Cyanobacteria Harmful Algal Blooms and US Geological Survey Science Capabilities*. US Department of the Interior, US Geological Survey,

Gumbo R. J., Ross G. & Cloete E. T. 2008. Biological control of Microcystis dominated harmful algal blooms. *African Journal of Biotechnology*. 7,

Harada K.-I. & Tsuji K. 1998. Persistence and decomposition of hepatotoxic microcystins produced by cyanobacteria in natural environment. *Journal of Toxicology: Toxin Reviews*. 17, 385-403.

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

He X., Liu Y.-L., Conklin A., Westrick J., Weavers L. K., Dionysiou D. D., Lenhart J. J., Mouser P. J., Szlag D. & Walker H. W. 2016. Toxic cyanobacteria and drinking water: Impacts, detection, and treatment. *Harmful Algae*. 54, 174-193.

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. 2008. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful algae*. 8, 3-13.

Hernández J. M., 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. *Medical hypotheses*. 72, 539-540.

Hilborn E. D. & Beasley V. R. 2015. One health and cyanobacteria in freshwater systems: animal illnesses and deaths are sentinel events for human health risks. *Toxins*. 7, 1374-1395.

Hoefel D., Adriansen C. M., Bouyssou M. A., Saint C. P., Newcombe G. & Ho L. 2009. Development of an *mlrA* gene-directed TaqMan PCR assay for quantitative assessment of microcystin-degrading bacteria within water treatment plant sand filter biofilms. *Applied Environvironmental Microbiology*. 75, 5167-5169.

Jeppesen E., Søndergaard M., Jensen J. P., Havens K. E., Anneville O., Carvalho L., Coveney M. F., Deneke R., Dokulil M. T. & Foy B. 2005. Lake responses to reduced nutrient loading–an analysis of contemporary long-term data from 35 case studies. *Freshwater Biology*. 50, 1747-1771.

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.). *Hydrobiologia*. 581, 161-165.

Jones G. J., Bourne D. G., Blakeley R. L. & Doelle H. 1994. Degradation of the cyanobacterial hepatotoxin microcystin by aquatic bacteria. *Natural toxins*. 2, 228-235.

Kainz P. 2000. The PCR plateau phase–towards an understanding of its limitations. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*. 1494, 23-27.

Kang Y. H., Park C. S. & Han M. S. 2012. *Pseudomonas aeruginosa* UCBPP-PA14 a useful bacterium capable of lysing *Microcystis aeruginosa* cells and degrading microcystins. *Journal of Applied Phycology*. 24, 1517-1525.

Kansole M. M. & Lin T. F. 2017. Impacts of hydrogen peroxide and copper sulfate on the control of *Microcystis aeruginosa* and MC-LR and the inhibition of MC-LR degrading bacterium *Bacillus* sp. *Water*. 9, 255.

Krausfeldt L. E., Tang X., van de Kamp J., Gao G., Bodrossy L., Boyer G. L. & Wilhelm S. W. 2017. Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu. *FEMS Microbiology Ecology*. 93,

Kuniyoshi T. M., Gonzalez A., Lopez-Gomollon S., Valladares A., Bes M. T., Fillat M. F. & Peleato M. L. 2011. 2-oxoglutarate enhances NtcA binding activity to promoter regions of the microcystin synthesis gene cluster. *FEBS letters*. 585, 3921-3926.

Le C., Zha Y., Li Y., Sun D., Lu H. & Yin B. 2010. Eutrophication of lake waters in China: cost, causes, and control. *Environmental management*. 45, 662-668.

Lewis W. M. & Wurtsbaugh W. A. 2008. Control of lacustrine phytoplankton by nutrients: erosion of the phosphorus paradigm. *International Review of Hydrobiology*. 93, 446-465.

Lezcano M. Á., Morón-López J., Agha R., López-Heras I., Nozal L., Quesada A. & El-Shehawy R. 2016. Presence or absence of *mlr* genes and nutrient concentrations co-determine the microcystin biodegradation efficiency of a natural bacterial community. *Toxins*. 8, 318.

Li J., Li R. & Li J. 2017. Current research scenario for microcystins biodegradation–A review on fundamental knowledge, application prospects and challenges. *Science of The Total Environment*. 595, 615-632.

Makarewicz J. C. & Bertram P. 1991. Evidence for the restoration of the Lake Erie ecosystem. *Bioscience*. 41, 216-223.

Manage P. M., Edwards C., Singh B. K. & Lawton L. A. 2009. Isolation and identification of novel microcystin-degrading bacteria. *Applied Environmental Microbiology*. 75, 6924-6928.

Mazur H. & Plinski M. 2001. Stability of cyanotoxins, microcystin-LR, microcystin-RR and nodularin in seawater and BG-11 medium of different salinity. *Oceanologia*. 43, 329-339.

Meriluoto J., Spoof L. & Codd G. A. 2017. *Handbook of cyanobacterial monitoring and cyanotoxin analysis*. John Wiley & Sons,

Michalak A. M., Anderson E. J., Beletsky D., Boland S., Bosch N. S., Bridgeman T. B., Chaffin J. D., Cho K., Confesor R. & Daloğlu I. 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*. 110, 6448-6452.

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.

Nybom S., Dziga D., Heikkilä J., Kull T., Salminen S. & Meriluoto J. 2012. Characterization of microcystin-LR removal process in the presence of probiotic bacteria. *Toxicon*. 59, 171-181.

Nybom S. M., Collado M. C., Surono I. S., Salminen S. J. & Meriluoto J. A. 2008. Effect of glucose in removal of microcystin-LR by viable commercial probiotic strains and strains isolated from dadih fermented milk. *Journal of agricultural and food chemistry*. 56, 3714-3720.

O'neil J., Davis T., Burford M. & Gobler C. 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful algae*. 14, 313-334.

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

Oberemm A., Becker J., Codd G. & Steinberg C. 1999. Effects of cyanobacterial toxins and aqueous crude extracts of cyanobacteria on the development of fish and amphibians. *Environmental Toxicology: An International Journal*. 14, 77-88.

Oehrle S., Rodriguez-Matos M., Cartamil M., Zavala C. & Rein K. S. 2017. Toxin composition of the 2016 *Microcystis aeruginosa* bloom in the St. Lucie Estuary, Florida. *Toxicon*. 138, 169-172.

Okano K., Shimizu K., Maseda H., Kawauchi Y., Utsumi M., Itayama T., Zhang Z. & Sugiura N. 2015. Whole-genome sequence of the microcystin-degrading bacterium *Sphingopyxis* sp. strain C-1. *Genome announcements*. 3,

Ouellette A. J. & Wilhelm S. W. 2003. Toxic cyanobacteria: the evolving molecular toolbox. *Frontiers in Ecology and the Environment*. 1, 359-366.

Paerl H. & Fulton III R. 2006. Ecology of harmful cyanobacteria. *Ecology of harmful algae,* 189, 95-109. Springer.

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

Paerl H. W. 2018. Mitigating Toxic Planktonic Cyanobacterial Blooms in Aquatic Ecosystems Facing Increasing Anthropogenic and Climatic Pressures. *Toxins*. 10, 76.

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

Paerl H. W. & Huisman J. 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental microbiology reports*. 1, 27-37.

Paerl H. W. & Otten T. G. 2013. Harmful cyanobacterial blooms: causes, consequences, and controls. *Microbial ecology*. 65, 995-1010.

Paerl H. W., Hall N. S. & Calandrino E. S. 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment*. 409, 1739-1745.

Paerl H. W., Scott J. T., McCarthy M. J., Newell S. E., Gardner W., Havens K. E., Hoffman D. K., Wilhelm S. W. & Wurtsbaugh W. A. 2016. It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science & Technology*.

Pantelic D., Svircev Z., Simeunovic J., Vidovic M. & Trajkovic I. 2013. Cyanotoxins: characteristics, production and degradation routes in drinking water treatment with reference to the situation in Serbia. *Chemosphere*. 91, 421-441.

Peng G., Martin R., Dearth S., Sun X., Boyer G. L., Campagna S., Lin S. & Wilhelm S. W. 2018. Seasonally-relevant cool temperatures interact with N chemistry to increase microcystins produced in lab cultures of *Microcystis aeruginosa* NIES-843. *Environmental science & technology*.

Pham T.-L. & Utsumi M. 2018. An overview of the accumulation of microcystins in aquatic ecosystems. *Journal of environmental management*. 213, 520e529.

Pimentel J. S. & Giani A. 2014. Microcystin production and regulation under nutrient stress conditions in toxic *Microcystis* strains. *Applied and environmental microbiology*. 80, 5836-5843.

Preston T., Stewart W. & Reynolds C. 1980. Bloom-forming cyanobacterium *Microcystis aeruginosa* overwinters on sediment surface. *Nature*. 288, 365.

Price G. D. 2011. Inorganic carbon transporters of the cyanobacterial CO_2 concentrating mechanism. *Photosynthesis Research*. 109, 47-57.

Qin B., Zhu G., Gao G., Zhang Y., Li W., Paerl H. W. & Carmichael W. W. 2010. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *Environmental management*. 45, 105-112.

Rapala J., Lahti K., Sivonen K. & Niemelä S. 1994. Biodegradability and adsorption on lake sediments of cyanobacterial hepatotoxins and anatoxin-a. *Letters in applied microbiology*. 19, 423-428.

Saito T., Okano K., Park H.-D., Itayama T., Inamori Y., Neilan B. A., Burns B. P. & Sugiura N. 2003. Detection and sequencing of the microcystin LR-degrading gene, *mlrA*, from new bacteria isolated from Japanese lakes. *FEMS Microbiology Letters*. 229, 271-276.

Sandrini G., Jakupovic D., Matthijs H. C. & Huisman J. 2015. Strains of the harmful cyanobacterium *Microcystis aeruginosa* differ in gene expression and activity of inorganic carbon uptake systems at elevated CO_2 levels. *Applied and environmental microbiology*. 81, 7730-7739.

Sandrini G., Matthijs H. C., Verspagen J. M., Muyzer G. & Huisman J. 2014. Genetic diversity of inorganic carbon uptake systems causes variation in CO_2 response of the cyanobacterium *Microcystis*. *The ISME journal*. 8, 589.

Schindler D. W. 1974. Eutrophication and recovery in experimental lakes: implications for lake management. *Science*. 184, 897-899.

Schindler D. W., Hecky R., Findlay D., Stainton M., Parker B., Paterson M., Beaty K., Lyng M. & Kasian S. 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*. 105, 11254-11258.

Schlesinger W. H. 2009. On the fate of anthropogenic nitrogen. *Proceedings of the National Academy of Sciences*. 106, 203-208.

Schmidt J. R., Wilhelm S. W. & Boyer G. L. 2014. The fate of microcystins in the environment and challenges for monitoring. *Toxins*. 6, 3354-3387.

Shen H., Niu Y., Xie P., Tao M. & Yang X. 2011. Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshwater Biology*. 56, 1065-1080.

Smith J. L., Schulz K. L., Zimba P. V. & Boyer G. L. 2010. Possible mechanism for the foodweb transfer of covalently bound microcystins. *Ecotoxicology and Environmental Safety*. 73, 757-761.

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, e44002.

Steffen M. M., Davis T. W., McKay R. M., Bullerjahn G. S., Krausfeldt L. E., Stough J. M., Neitzey M. L., Gilbert N. E., Boyer G. L. & Johengen T. H. 2017. Ecophysiological examination of the Lake Erie *Microcystis* bloom in 2014: linkages between biology and the water supply shutdown of Toledo, Ohio. *Environmental Science & Technology*.

Svircev Z., Krstic S., Miladinov-Mikov M., Baltic V. & Vidovic M. 2009. Freshwater cyanobacterial blooms and primary liver cancer epidemiological studies in Serbia. *Journal of Environmental Science and Health Part C*. 27, 36-55.

Tillett D., Dittmann E., Erhard M., von Döhren H., Börner T. & Neilan B. A. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chemistry & biology*. 7, 753-764.

Tsuji K., Naito S., Kondo F., Ishikawa N., Watanabe M. F., Suzuki M. & Harada K.-i. 1994. Stability of microcystins from cyanobacteria: effect of light on decomposition and isomerization. *Environmental Science & Technology*. 28, 173-177.

US-EPA (2015) Drinking Water Health Advisory for the Cyanobacterial Microcystin Toxins.

Visser P. M., Verspagen J. M. H., Sandrini G., Stal L. J., Matthijs H. C. P., Davis T. W., Paerl H. W. & Huisman J. 2016. How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*. 54, 145-159.

Wang W., Shen H., Shi P., Chen J., Ni L. & Xie P. 2016. Experimental evidence for the role of heterotrophic bacteria in the formation of *Microcystis* colonies. *Journal of Applied Phycology*. 28, 1111-1123.

Watson S. B., Miller C., Arhonditsis G., Boyer G. L., Carmichael W., Charlton M. N., Confesor R., Depew D. C., Höök T. O. & Ludsin S. A. 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae*. 56, 44-66.

Welker M. & Steinberg C. 2000. Rates of humic substance photosensitized degradation of microcystin-LR in natural waters. *Environmental Science & Technology*. 34, 3415-3419.

Wetzel R. G. 2001. *Limnology: lake and river ecosystems*. Gulf Professional Publishing.

Wines M. 2013. Spring rain, then foul algae in ailing Lake Erie. *The New York Times* .

Wood R. 2016. Acute animal and human poisonings from cyanotoxin exposure—A review of the literature. *Environment international*. 91, 276-282.

Wörmer L., Huerta-Fontela M. a., Cirés S., Carrasco D. & Quesada A. 2010. Natural photodegradation of the cyanobacterial toxins microcystin and cylindrospermopsin. *Environmental science & technology*. 44, 3002-3007.

Wu X., Jiang J. & Hu J. 2013. Determination and occurrence of retinoids in a eutrophic lake (Taihu Lake, China): cyanobacteria blooms produce teratogenic retinal. *Environ Sci Technol*. 47, 807-814.

Xiao M., Li M. & Reynolds C. S. 2018. Colony formation in the cyanobacterium *Microcystis*. *Biological Reviews*.

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

Young F. M., Thomson C., Metcalf J. S., Lucocq J. M. & Codd G. A. 2005. Immunogold localisation of microcystins in cryosectioned cells of *Microcystis*. *Journal of Structural Biology*. 151, 208-214.

Zilliges Y., Kehr J.-C., Meissner S., Ishida K., Mikkat S., Hagemann M., Kaplan A., Börner T. & Dittmann E. 2011. The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of *Microcystis* under oxidative stress conditions. *PloS one*. 6, e17615.

Zurawell R. W., Chen H., Burke J. M. & Prepas E. E. 2005. Hepatotoxic cyanobacteria: a review of the biological importance of microcystins in freshwater environments. *Journal of Toxicology and Environmental Health, Part B*. 8, 1-37.

Chapter 2 : Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu

Publication note

This chapter is a version of a peer-reviewed article previously published in FEMS Microbiology Ecology by Lauren E. Krausfeldt, Xiangming Tang, Jodie van de Kamp, Guang Gao, Levente Bodrossy, Gregory L. Boyer, Steven W. Wilhelm

My contribution to this work included assistance with sample collection, RNA extraction, cDNA synthesis, data analysis and writing the majority of the paper.

Abstract

Harmful cyanobacterial blooms (cyanoHABs) are a major threat to freshwater ecosystems worldwide. Evidence suggests that both nitrogen and phosphorus are important nutrients in the development and proliferation of blooms, yet much less is known about nitrogen cycling dynamics in these systems. To assess the potential nitrogen cycling function of the cyanoHAB community, surface water samples were collected in Lake Tai (*Taihu*), China over a 5-month bloom event in 2014. The expression of six nitrogen cycling genes (*nifH*, *hzsA*, *nxrB*, *nrfA*, *amoA*, *nosZ*) was surveyed using a targeted microarray with probes designed to provide phylogenetic information. N-cycling gene expression varied spatially across *Taihu*, most notably near the mouth of the Dapu river. Expression of *nifH* was observed across the lake and attributable to both proteobacteria and cyanobacteria: proteobacteria were major contributors to *nifH* signal near shore. Other N transformations such as anaerobic ammonia oxidation and denitrification were evident in the surface waters as well. Observations in this study highlight the potential importance of heterotrophic bacteria in N-cycling associated with cyanoHABs.

Introduction

Harmful cyanobacterial blooms (cyanoHABs) are detrimental to freshwater aquatic ecosystems worldwide, and some reports have even declared cyanoHABs one of the greatest risks to freshwater aquatic ecosystems (Bullerjahn*, et al.* 2016, Council 2000, Dudgeon*, et al.* 2006, Watson*, et al.* 2016). Along with a decrease in water quality and the presence of taste and odor compounds like (*i.e.,* volatile organic compounds), cyanoHABs have been implicated in massive fish kills, the generation of hypoxia and the intoxication of various mammals (Brooks*, et al.* 2016, Watson, *et al*. 2016). The presence of cyanoHABs interferes with tourism and industry and poses a severe risk to public health and drinking water supplies (Bullerjahn*, et al.* 2016, He*, et al.* 2016). It is widely accepted that cyanoHABs are a result of eutrophication due to the domestic, agricultural and industrial introduction of nutrients (Heisler*, et al.* 2008). Traditionally, cyanoHABs have been linked strictly to phosphorus loading and many management strategies have aimed to reduce P inputs (Conley*, et al.* 2009, Cooke*, et al.* 2016). Despite these reductions, cyanoHABs continue to occur at greater frequencies and intensities globally (Harke*, et al.* 2015, Harke*, et al.* 2016, Michalak*, et al.* 2013, O'neil*, et al.* 2012)**.**

A potential role for nitrogen (N) in the promotion of freshwater cyanoHABs has recently gained increased attention (Paerl*, et al.* 2014a, Paerl*, et al.* 2016). Anthropogenic N sources, such as fertilizers, legume cultivation, and fossil fuel combustion have increased markedly in recent decades and now outweigh natural N inputs (Galloway, *et al*. 2008, Glibert, *et al*. 2015). Many lakes receive higher N loads relative to P, yet seasonal N and P limitation may be important for the initiation and duration of cyanoHABs (Belisle*, et al.* 2016, Chaffin*, et al.* 2013, Davis*, et al.* 2015, Paerl and Otten 2013, Xu*, et al.* 2010). N availability may further drive community structure, leading to the emergence of toxic algal species in some bloom communities: indeed, nitrogen

availability has been suggested to influence the production of microcystin within cells (Davis*, et al.* 2010, Harke*, et al.* 2015, Neilan*, et al.* 2013, O'neil*, et al.* 2012)**.**

Countering the idea that N-loading can drive eutrophication and species composition is the argument that N_2-fixing cyanobacteria "produce" their own nitrogen, thus they are never truly N-limited (Schindler*, et al.* 2008). While in some systems there may be merit to this idea, questions have arisen about the ability of these organisms to supplement anything more than their basic survival based N-requirements (Scott and McCarthy 2010, Willis*, et al.* 2016). Indeed, non-N_2-fixing cyanobacteria that have become the dominant bloom-formers in many lakes. *Microcystis* spp., which obtain reduced N from the aquatic environment, has been demonstrated to outcompete N_2-fixing cyanobacteria in N-limited conditions (Blomqvist*, et al.* 1994, Paerl*, et al.* 2014b). In parallel with the implications of global climate change effects on these systems (Paerl and Huisman 2008), it is becoming more apparent that both N and P are combined driving factors in bloom formation, with several reports declaring the need to reduce both N and P to combat prolonged eutrophication and cyanoHABs (Conley*, et al.* 2009, EPA 2015, Lewis Jr*, et al.* 2011, Paerl*, et al.* 2014a).

Hypereutrophic Lake Tai (or in Chinese, *Taihu*) is China's third largest freshwater lake in terms of surface area (2,338 km^2) and is plagued by cyanoHABs annually. These blooms, occurring mostly in the northern and western part of the lake, can last for more than half the year and are driven by both N and P (Qin*, et al.* 2010, Xu*, et al.* 2010). Increased anthropogenic nutrients and climate change have provided ideal conditions for cyanoHABs which are dominated by *Microcystis spp.* (Guo 2007, Paerl*, et al.* 2015, Qin*, et al.* 2007). *Taihu* is very shallow, with a mean depth of only 1.9 m and is rarely stratified due to its broad fetch and shallow depth (Xu*, et al.* 2010). There are 172 rivers and channels that flow into *Taihu*, bringing nutrients and waste

from domestic, agricultural and industrial practices (Guo 2007, Han*, et al.* 2014, Qin*, et al.* 2007). There are only a limited number of outflows, with the major ones being confined to the southeast, prolonging the residence time of nutrients within the lake (Qin*, et al.* 2007). This residence time and the shallow nature of the lake establishes a large potential for regeneration of nutrients such as N and P from the sediments (McCarthy*, et al.* 2007, Qin*, et al.* 2007). N can be released back into the atmosphere in the form of N_2 (through denitrification or anaerobic ammonia oxidation), while P, with no gaseous form to permit its escape, is recycled through the microbial food web.

There are many potentially important biological activities that may shape N-dynamics in the water column (McCarthy*, et al.* 2007, Qin*, et al.* 2007). Our objective here was to characterize the spatial and temporal variations in these processes and identify the microbes responsible for these various N transformations as they occurred within the surface dwelling bloom community. The expression of six different N-cycling functional genes, across the season and phylogenetic spectrum, were investigated using a targeted microarray approach.

Methods

Sample collection and environmental parameters

Surface water samples were collected monthly using a small boat at locations in *Taihu* from June to October 2014 (Figure 1). Samples from Station 10 in June and the Channel in October were lost during processing, and therefore are not included in this study. Samples were passed through 0.2-μm Sterivex filters (Appendix Table 1) and preserved with Ambion RNA*later* Stabilizing Solution (Invitrogen, Carlsbad, CA, USA) at the sample site. Sterivex were then stored at -80º C until further processing.

Water column depth and Secchi depth (SD) were measured using a water depth gauge

(Uwitec, Austria) and Secchi disk, respectively. Water temperature, electrical conductivity (EC), pH, dissolved oxygen (DO) and phycocyanobilin (PC) were measured *in situ* using a multiparameter water quality sonde (YSI 6600 V2, Yellow Springs Instruments Inc., USA). Total nitrogen (TN), dissolved total nitrogen (DTN), ammonium (NH_4), nitrate (NO_3), total phosphorus (TP), dissolved total phosphorus (DTP), orthophosphate (PO_4), total suspended solids (TSS), and chlorophyll *a* (chl *a*) were measured according to standard methods (Jin and Tu 1990).

RNA extraction

RNA was extracted using the Powerwater DNA Isolation kit for Sterivex (MO BIO Laboratories, Carlsbad, CA, USA) modified for RNA based on a protocol provided from the manufacturer. Modifications to the manufacturer's protocol were made to optimize the extraction of RNA preserved with RNA*later*. Briefly, RNA*later* was removed from the Sterivex cartridge with a syringe and the sample centrifuged to collect biomass that was detached from the filter. β-mercaptoethanol plus solution ST1B (included in the kit) was used to resuspend the biomass and add it back into the Sterivex to continue with the manufacturer's protocol. Other modifications to the manufacturer's protocol included vortexing the Sterivex 5 min longer than recommended and the passage of all treatments and wash steps through the binding column by centrifugation at 13000 x g for 2 min after being allowed to sit for at least 1 min. The On-Spin Column DNase I Kit (MO BIO Laboratories, Carlsbad, CA, USA) was used for DNA removal with the modification that DNase was allowed to sit for up to 30 min to increase the efficiency of DNA removal. Several wash steps were performed following DNase treatment using the provided solutions as indicated in the protocol, as well as two additional washes with 100 % ethanol to remove residual salt contamination. The resulting RNA was checked for DNA contamination fluorometrically and by

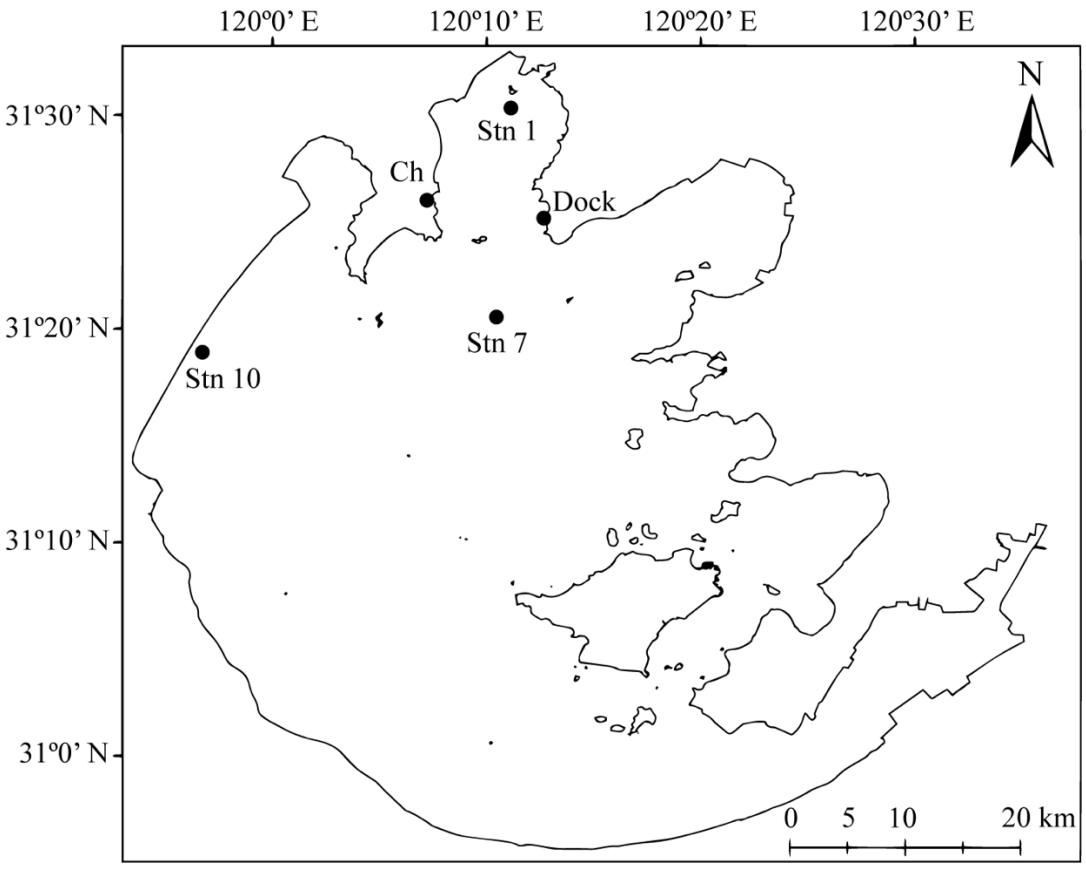

Figure 2.1 Map of Lake Taihu in China

Black circles represent stations sampled.

PCR amplification with 16S primers (27F and 1522R, Edwards*, et al.* 1989): when product was seen the DNase treatments were repeated. Conditions for 16S rDNA PCR are provided in the supplemental materials. Total RNA was converted to cDNA using SuperScriptIII First Strand Synthesis Supermix kit (Invitrogen, Carlsbad, CA, USA) and the cDNA purified using the Qiaquick PCR Purification kit (Qiagen, Hilden, Germany).

Microarray analyses

A high-throughput functional gene microarray was used to assess the community composition of key N-cycling clades. The microarray consists of 691 long oligonucleotide probes targeting 6 functional marker genes of the N cycle, on a solid substrate (glass microscope slide). The functional marker genes (archaeal and bacterial *amoA*, *hzsA*, *nifH*, *nosZ*, *nrfA* and *nxrB*) were amplified from DNA or cDNA, fluorescently labeled via T7 *in vitro* transcription, fragmented and hybridized overnight to the array. Signals were normalized to the *hyaBP* probe, targeting an externally added spike control, set to 10,000. Detailed information about the development and methods of the functional gene microarray is provided in the Apendix under Supplementary Information and Appendix Tables 2-3. Raw array data is available from the author upon request.

Multivariate Analysis

Environmental parameters were scaled and analyzed using several multivariate and statistical tests in Primer-e (V7) package (Clarke and Gorley 2006). Dissimilarities between months and stations were determined using ANOSIM at a p-value < 0.05 and driving factors of dissimilarities identified *via* SIMPER. Relationships between samples were visualized and clustered using nMDS and SIMPROF analysis. Spearman correlations and p-values between chl

a, cyanobacterial cells, and microcystin and other environmental variables were calculated in SigmaPlot v12.5 (Systat Software, San Jose, CA).

N-cycling gene expression was analyzed in Primer-e (V7). The values from positive probes were transformed (log(x+1)), and a heat map was generated with the ShadePlot function. Dissimilarities between months and locations were determined by ANOSIM (p-value < 0.05). Since almost half of the N-cycling gene signal came from probes targeting *nifH* and many were designed from *nifH* sequences from uncultured microrganisms, further phylogenetic analyses were performed. A maximum likelihood placement tree was created with the 46 positive *nifH* probes and over 650 annotated reference *nifH* sequences. Reference sequences were obtained from an extensive *nifH* database created by Gaby and Buckley (2014) and GenBank. The reference tree included sequences from *nifH* Clusters I, II and III, as well as IV and V to be certain no positive probes aligned with *nifH*-like genes, which are not involved in N-fixation. Further, only annotated sequences between 800bp and 1000bp were used in order to avoid misplacement of the probes on the reference tree. The sequences for the reference tree were aligned using MUSCLE in Mega version 7.0 (Tamura*, et al.* 2007). The reference tree was made on the PhyML server using a GTR model, and likelihood ratios and branch support were calculated using a Shimodaira-Hasegawa (SH)-like approximate likelihood ratio test (aLRT-SH-like) (Guindon*, et al.* 2010). Multiple sequence alignments were generated using HMMER to align the probes to the reference tree. The placement tree was generated using Pplacer (matsen.fhcrc.org/pplacer/) and trimmed in Mega version 7.0 for visualization purposes.

To identify relationships between samples based on *nifH* expression alone, nMDS and SIMPROF analyses (p-value < 0.05) were used. Samples were colored based on the percentage of the total *nifH* signal that came from either probes targeting cyanobacteria or proteobacteria.

Relationships between environmental parameters and N-cycling gene expression were assessed by totaling the signals from each of the N-genes achieve a total signal for each process. Spearman correlations and p-values were calculated between total signal (log(x+1) transformed and scaled in Primer-e) and environmental variable measurements (scaled in Primer-e). A p-value of < 0.10 was considered statistically significant. To investigate potential drivers of the observed differences in *nifH* expression between proteobacteria or cyanobacteria, *nifH* signal was totaled within the two phylogenetic groups and correlated to environmental variables as well. *AmoA* probes were separated into archaeal or bacterial *amoA* and totaled within their phylogenetic group. Probes targeting *nrfA, nosZ, nxrB*, and *hzsA* were not separated by phylogenetic groups because either the probes were targeting uncultured members of the community or all probes within that gene with positive hits targeted only one major group.

Cyanobacteria toxins were determined using liquid chromatography coupled with mass spectroscopy. For water samples, a known volume of water (500-1000 ml) was filtered through glass fiber filters (Whatman 934 AH) and the toxin filters extracted in 10 mL of 50 % methanol/water (v/v) containing 1 % acetic acid (v/v) using sonication (Boyer 2007). The samples were clarified by centrifugation for 10 minutes at 12,000 *g* and 0 ˚C, filtered through a 13 mm diameter nylon syringe filter (0.45 μm pore), and stored at -20 ˚C until analysis. Microcystins were determined using reverse phase liquid chromatography and screened for the molecular ion of 14 common microcystin congeners (RR, dRR, mRR, hYR, YR, LR, mLR, dLR, AR, FR, LA, LW, LF, WR and R-NOD) using a Waters ZQ4000 mass spectrometer coupled with a photodiode array spectrometer. All microcystins were quantified by mass spectroscopy against microcystin-LR and their presence confirmed using their diagnostic ADDA UV signatures. Anatoxin-a (ATX), homoanatoxin-a, cylindrospermopsin (CYL) and deoxycylindrospermopsin were determined in

the same extracts using HPLC coupled with mass selective (LCMS) or tandom mass (LC-MS/MS: Waters TQD) detection and quantified against the respective compounds. Method detection limits varied depending on the volume filtered but generally ranged from 0.1-0.3 ug MC-LR / L and were less than 0.01 µg/L for anatoxin-a, cylindrospermopsin, and their variants.

Results

Status of the lake

Water temperatures were coolest in October averaging 21.8º C ± 0.4º C and warmest in September averaging 27.5º C ± 0.5º C. The pH in all samples ranged from about 8 to 10. Lower pH values tended to be observed in June and gradually rose over the five-month period, which is typical of a cyanobacterial bloom due to the rapid utilization and depletion of CO_2 in the water column (Hudnell, 2008). Dissolved oxygen (DO) concentrations also varied greatly, with percent saturation ranging from 21.9 % to 159 % saturation. N or P mass ratios ranged from 4.12 to > 32, but were lowest in the summer and the fall months. Chl *a* concentrations spanned 2 orders of magnitude (from 17 to 16,000 µg/L). The highest concentrations of chl *a* and cyanobacterial cell numbers were observed in the Channel and the dock, and the lowest were observed at Station 10 (Appendix Table 4 and 5).

Biological parameters, including chl *a* and phycocyanin (a proxy for cyanobacterial cells), were only moderately correlated to each other (Appendix Figure 1a), suggesting that cyanobacteria did not always represent the largest component of total chl *a* concentrations. Chl *a* and cyanobacterial cells positively correlated with turbidity and negatively correlated to TN:TP. Chl *a* strongly positively correlated to both TP and TN, while cyanobacterial cells only moderately correlated to TP (Appendix Figure 1). There were statistically significant spatial differences in chl

a concentrations over the bloom event suggesting the bloom status is not the same across all stations (ANOSIM, data not shown).

Total microcystin concentrations ranged from undetectable (*ca* 0.1 μg/L) to 595 μg/L. For some samples that were not represented in microarray dataset (*ie*. the Channel in October), microcystin concentrations reached over 6,000 μg/L. The highest concentrations were observed in the Channel where chl *a* was also the highest. Total microcystin concentrations positively correlated to cyanobacterial cells abundance, chl *a*, TP and TN (Figure 2). Microcystin-LR, -RR, and -YR were the most prevalent congeners comprising approximately 85-100 % of the total microcystin in all toxic samples. A small percentage of congeners consisted of microcystin AR, FR and mLR. Microcystin-FR and mLR were only detected in the Channel, while -AR were detected in both the Channel and the dock (data not shown). Other common cyanobacterial toxins, including anatoxin-a, homoanatoxin-a, cylindrospermopsin, epi-cylindrospermopsin and deoxycylindrospermopsin were not detected (< 0.01 μg/L) in any of the water samples.

ANOSIM analysis indicated there were differences in environmental parameters between months, while cluster analysis suggested this is likely seasonal (Figure 3a, Appendix Table 6). Factors that contributed the most to these differences were related to water chemistry, such as water temperature, nutrient concentrations, pH and conductivity (SIMPER, not shown). Spatial variations based on environmental parameters were also observed in *Taihu.* These differences were most notable at Station 10 and the Channel (Figure 3b, Table S7). Major contributing factors to dissimilarities between Station 10 and other stations were nutrient concentrations, such as NH_4, TDP, and PO_4, whereas dissimilarities between the Channel and other stations were mainly driven by chl *a* and microcystin concentrations (SIMPER, not shown). Water depth varied spatially,

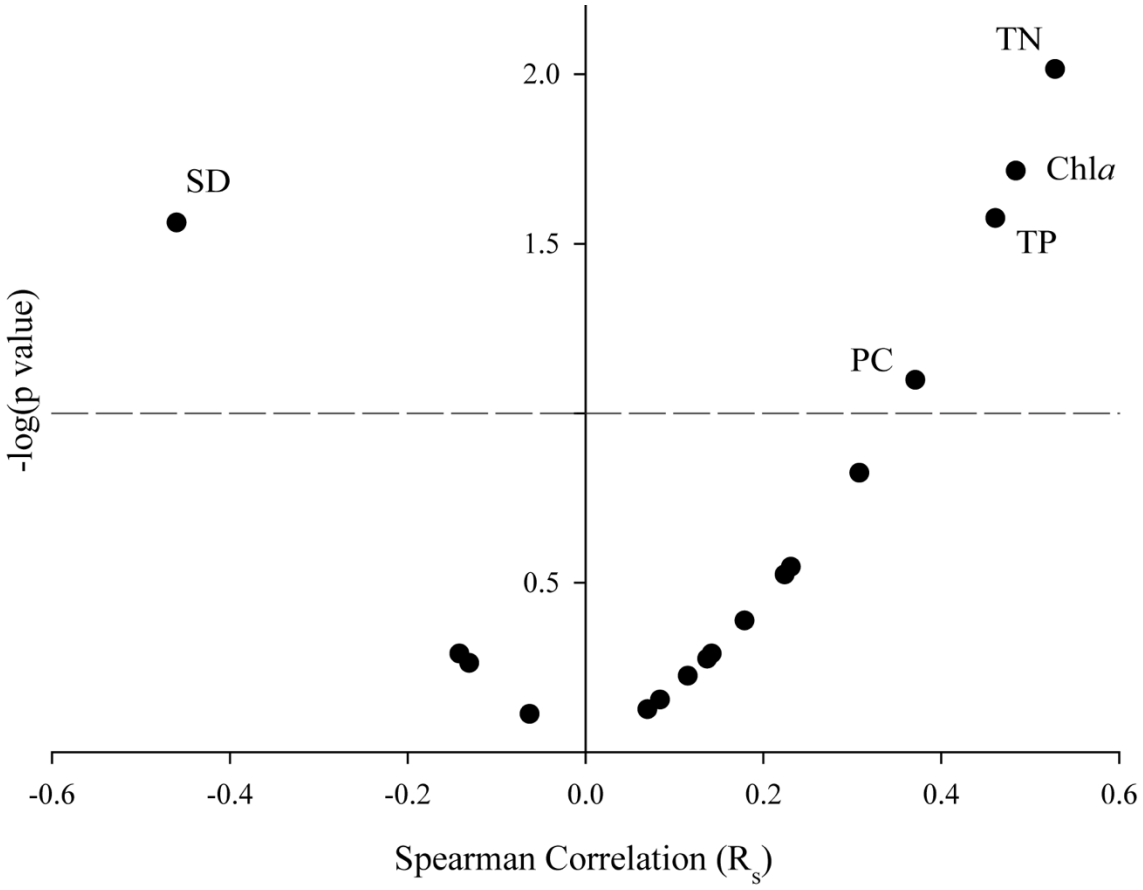

Figure 2.2 Volcano plot representing correlations to microcystin concentrations to environmental parameters

Volcano plot representing non-parametric correlations to microcystin (μg/L) in *Taihu* across the bloom event. Any parameter above the dashed line correlates to microcystin concentrations at a p value of less than 0.10.

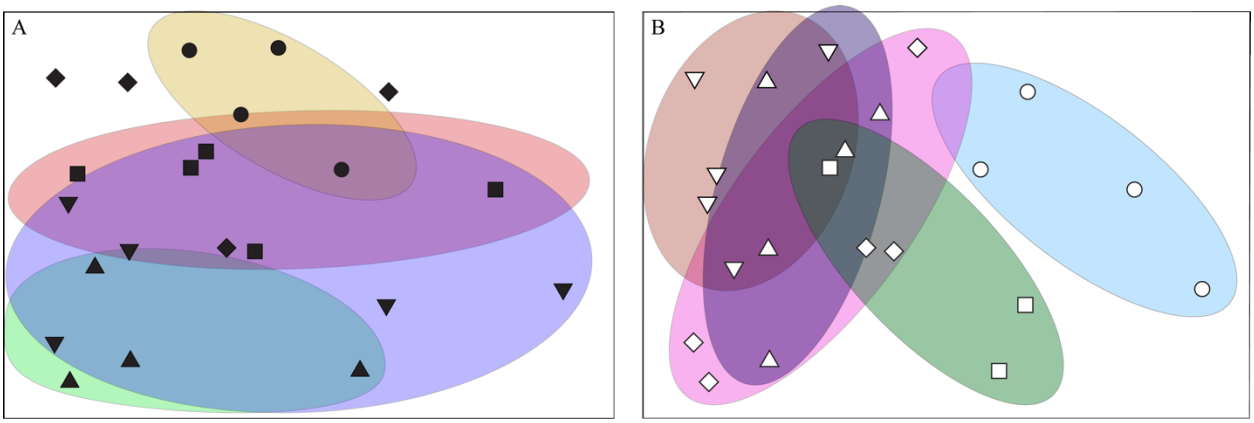

Figure 2.3 NMDS analysis plot describing relationships between samples based on environmental parameters

Circles are drawn to emphasize shift in samples over time or location with a 2D of 0.15 A). Relationships between samples change temporally. Closed upright triangles = June; closed upside down triangles =July; closed squares =August; closed diamonds = September; circles = October B). The same nMDS as in A, although demonstrating a shift in samples based on location in the lake. Open squares = Channel; open diamonds =Dock; open triangles = Stn 1; open upside down triangles= Stn 7; open circles = Stn10

although minimally, between locations (Appendix Table 4). Station 7 was typically the deepest reaching 3 m, whereas stations located near shore were as shallow as 1.2 m.

N-cycling gene expression

Across all samples, signals were observed for 90 of the 691 N-gene probes on the array (Figure 4). There were no temporal differences when considering all N-gene expression, but there were spatial differences (Table S8). The greatest dissimilarities were between Stations 7 and 10, Station 7 and the dock, and the Channel and Station 10. To look further into each process, each functional gene was analyzed separately.

Signals were observed for 45 of the 144 total probes targeting *nifH*. There were no statistical differences in *nifH* expression between months but there were statistical differences between locations (Table S8). *NifH* expression patterns at Station 10 differed from all other stations. Total *nifH* signal positively correlated with total dissolved phosphorus and phosphate while negatively correlating with chl *a* (not shown). The highest concentrations of ammonia were also observed at Station 10 in July and August which also corresponds to highest relative signal for expression of *nifH* which appears to be predominantly due to Proteobacteria (Figure 4, Table S9).

All but one positive probe aligned with sequences in *nifH* cluster I (Figure 5). Major contributors to *nifH* expression were uncultured members of the phylum Proteobacteria or Cyanobacteria (Table S9). Many probes with positive signals targeted uncultured members of these groups. Phylogenetic analysis was used to better characterize probes and suggested that many fall within Alpha-, Beta-, Gamma- and Deltaproteobacteria (Figure 5). The majority of cyanobacteria expressing *nifH* were members of the heterocystous *Nostocales* such as *Nostoc sp.*

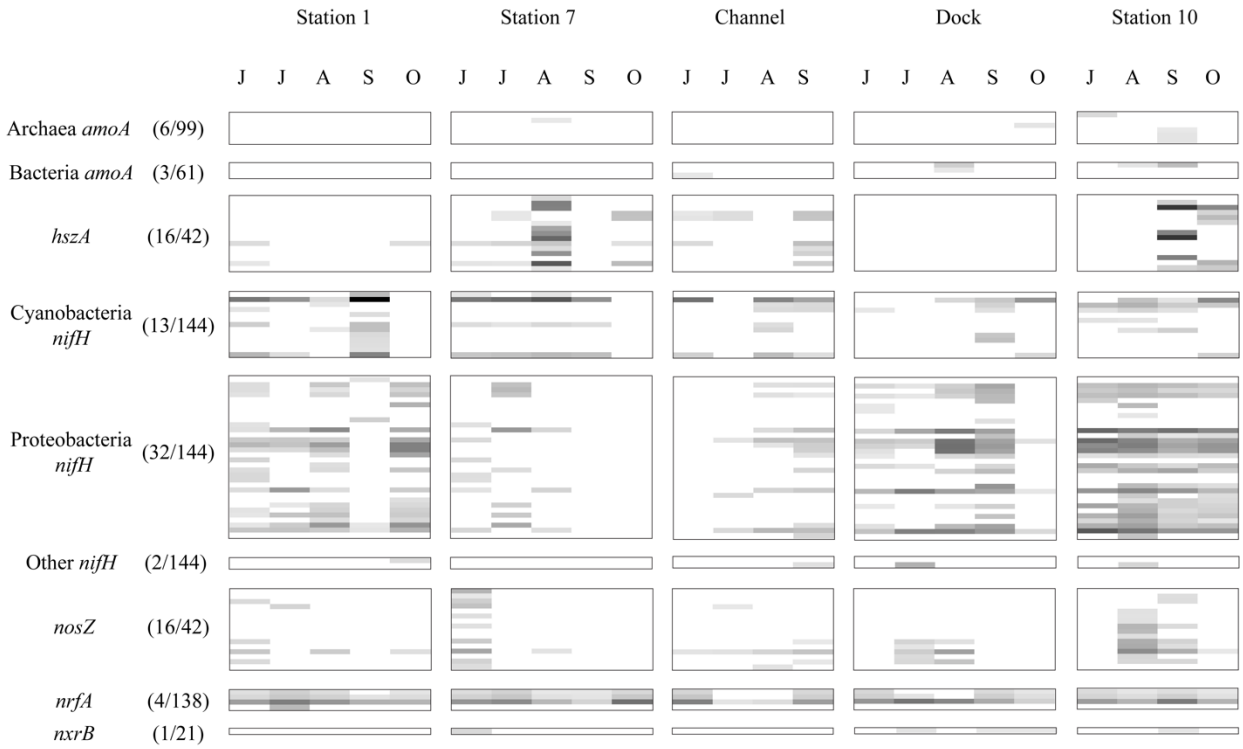

Figure 2.4 Shadeplot of N-processing gene expression for water column samples collected from *Taihu*

Only positive probes are shown. Each line represents a different probe targeting a specific functional gene. Each probe provides phylogenic information. Phylum level classification is denoted within the set of probes targeting *nifH*. Domain level classification is denoted within the set of probes targeting *amoA*.

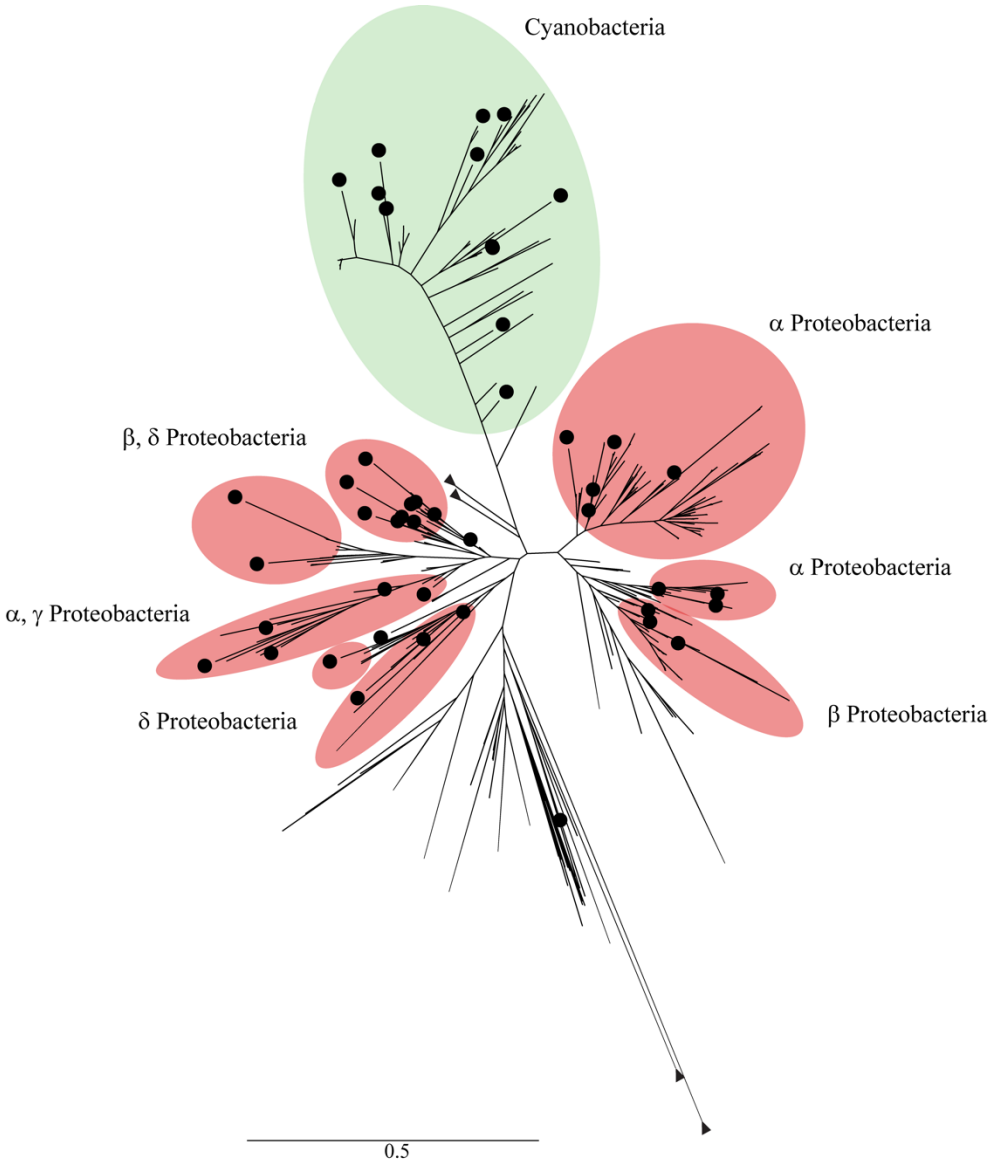

Figure 2.5 Phylogenetic placement of nifH probes with positive signals on a reference tree using annotated nifH and nifH-like sequences

▼denotes collapsed branches for visualization purposes. * indicates collapsed branches in which *nifH* Clusters IV and V align. No probes aligned with sequences within these clades. ● denotes *nifH* sequences detected in this study.

and *Mastigocladus sp.*, however there was evidence for expression from non-heterocystous *Oscillatoriales* such as *Cyanothece* as well (Appendix Figure 2). Although targeting uncultured microorganisms, many of the positive *nifH* probes aligned within closely related groups of N-fixing Proteobacteria providing some insight into phylogeny at a higher resolution (Appendix Figure 3-5). One positive probe aligned with sequences from *nifH* Cluster II, which contains iron-dependent alternative nitrogenases (Figure 5). nMDS analyses of *nifH* expression alone suggest there is clustering of samples based on who (either proteobacteria or cyanobacteria) is contributing the most to *nifH* expression (Figure 6, SIMPROF $p=<0.05$). Proteobacteria were dominantly expressing *nifH* at Station 10, as well as the dock, aside from one sample (Figure 6). In contrast, cyanobacteria dominated the *nifH* signal at Station 7, located in the middle and deepest part of the lake, with the exception of one sample in which the *nifH* signal came from both Proteobacteria and Cyanobacteria at similar levels. These results agree with the correlations to lake depth that were observed. Cyanobacterial *nifH* expression positively correlated with lake depth, suggesting cyanobacterial *nifH* expression in deeper waters, such as Station 7 (Figure 7a). However, *nifH* expression by Proteobacteria negatively correlated with lake depth, suggesting Proteobacteria were fixing N in shallower waters near shore, such as Station 10 and the dock.

Signals for 7 of the 160 probes targeting various *amoA* sequences on the array were observed which included probes specific for both bacteria and archaea (Figure 4, Table S9). The strongest signals for *amoA* were at Station 10 in September and at the dock in August. Although these appeared to be relatively low intensity, total *amoA* expression positive correlated with NH_4, TDN, TDP and PO_4 and negatively correlated to DO (Figure 7b). There were no positive hits out of a possible 21 probes for *nxrB*. Only 3 out of 138 probes specific for bacteria capable of dissimilatory nitrate reduction to ammonia (*nrfA*) were positive, though interestingly the *nrfA*

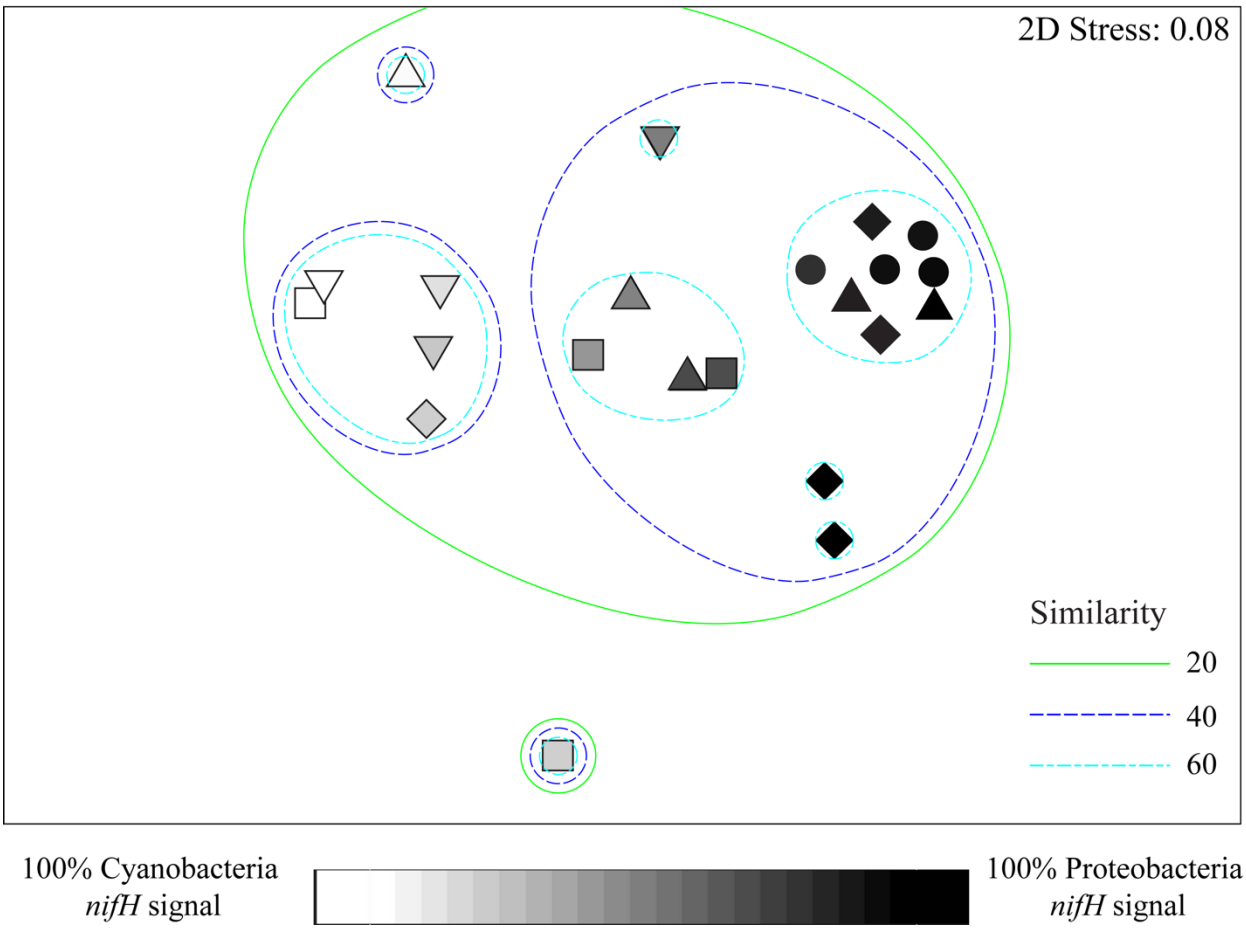

Figure 2.6 Relationships between samples based on *nifH* expression

NMDS analysis describing relationships between samples based on *nifH* expression at different locations in *Taihu.* ■ = Channel; ♦ = Dock; ▲= Stn 1; ▼= Stn 7; ● = Stn 10. 2D stress = 0.08. Shading of the symbol was dictated by percent of total *nifH* signal from either proteobacteria or cyanobacteria.

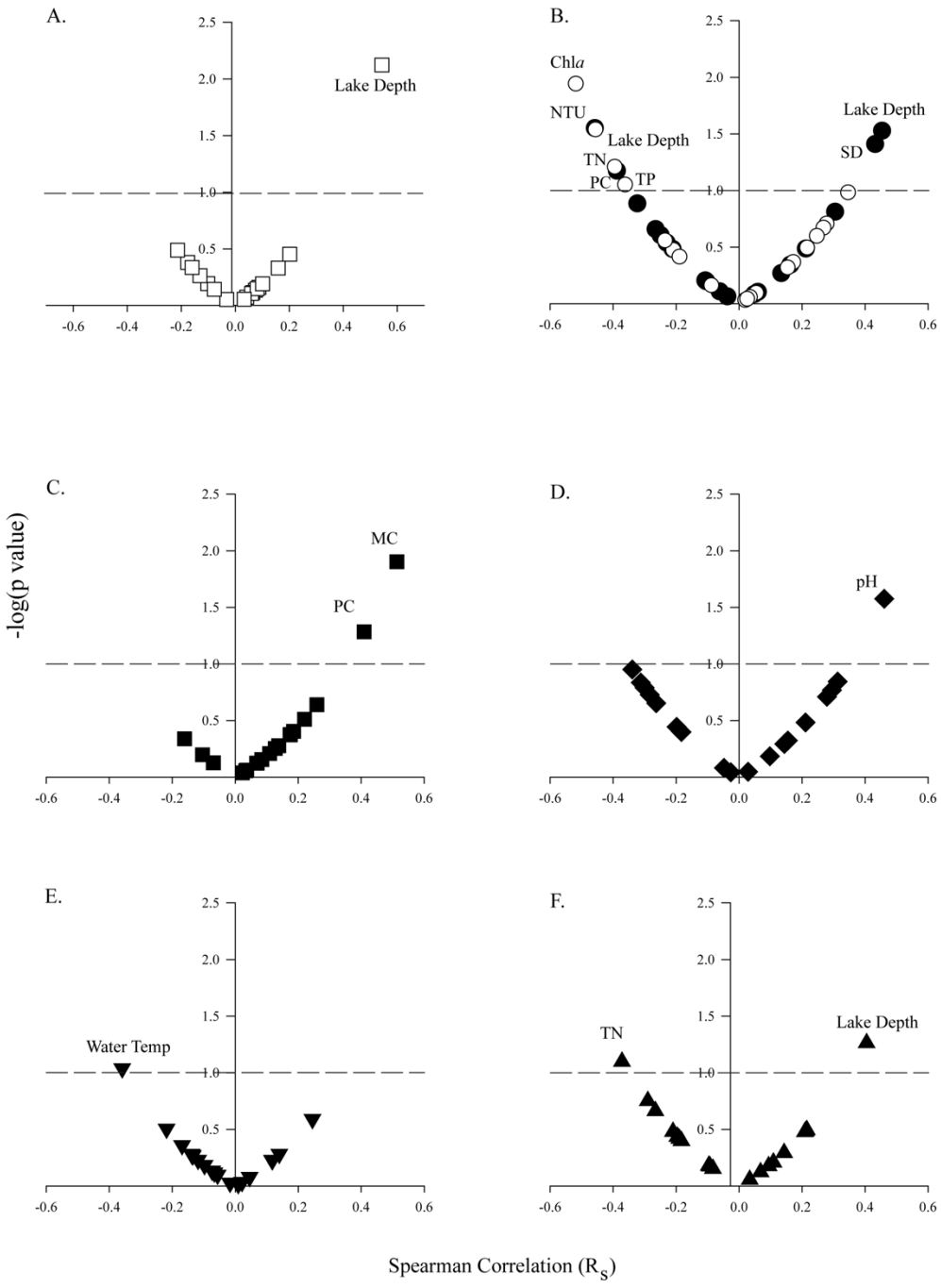

Figure 2.7 Volcano plots representing non-parametric correlations of environmental parameters to total expression of five different N-cycling genes

*nxr*B showed no statistically significant correlations to environmental parameters (not shown). In each plot the dashed line represents p = 0.10. A). *nosZ* (□) B). *nifH*; ● = Proteobacteria, ○ = Cyanobacteria C). *hzsA* (■) D). *amoA* (♦) E). *nxrB* (▼) F). *nrfA* (▲)

signal from these probes was fairly consistent across all samples (Figure 4). There were no statistical dissimilarities based on month or location in the lake (ANOSIM, not shown). All three positive probes targeted *E. coli* (Table S9). Total *nrfA* expression positively correlated to TN:TP and negatively with pH (Figure 7c).

There were 18 of 42 probes positive for Anammox bacteria (*hzsA*). Twelve probes targeted several uncultured subgroups of *Brocadia fulgida* and one targeted a clade with *Brocadia anammoxidans* and members of the nitrate-dependent anaerobic methane oxidizing (n-DAMO) clade. The remaining targeted other uncultured members of the n-DAMO clade and one positive probe targeted *Jettenia* subgroup 2 (Table S9). *HzsA* expression was sporadic throughout the samples but highest at Station 7 in August and Station 10 in September (Figure 4). Negative correlations between these transcripts and water depth were observed. (Figure 7d).

In examination of nitrous oxide reduction, 16 of the 182 probes targeting *nosZ* revealed positive signals. *NosZ* signals were most detectable in early to mid-summer in *Taihu*, prevalently at Station 10 in August and September and Station 7 in June (Figure 4). By October, all *nosZ* signals were low to undetectable. Several of the positive probes for *nosZ* targeted uncultured denitrifiers, that had originally been identified from lake sediments, agricultural soils, paddy soils or activated sludge (data not shown). Other probes suggest members *of Azospirillum, Herbaspirillum*, and *Bradyrhizobium* were expressing *nosZ* (Table S9). Signal for *nosZ* did not correlate to any of the environmental parameters measured in this study.

Discussion

The cycling of N is a complex network of transformations facilitated by groups of specialized microorganisms essential in the transport of N through various terrestrial and aquatic

ecosystems. Key processes involve the conversion of atmospheric N into the usable form of ammonia (N_2 fixation), the conversion of ammonia to nitrate (aerobic ammonia oxidation and nitrification), the reduction of nitrate to ammonia (dissimilatory nitrate reduction to ammonia) or back to N_2 gas (denitrification), and the reduction of ammonia to N_2 gas (anaerobic ammonia oxidation). With the addition of 150 Tg/yr of reactive N to the environment from fertilization and fossil fuel combustion, human activities can alter the N cycle in various systems and the dynamics of the microbial communities within them (Greaver*, et al.* 2016, Socolow 1999). N-cycling has been extensively studied in marine and soil systems, but it is not as well understood in freshwater systems, especially those plagued with cyanoHABs. The importance of excess N continues to emerge with respect to cyanoHABs and the dominance of non-N_2-fixing cyanobacteria. Understanding when and where these biological processes are active may provide insight into the requirements of the community and how different species emerge or dominate (*i.e*. blooms) under periods of N availability, limitation and processing.

To determine the N demands of the community during the bloom, TN:TP ratios were used to access nutrient limitation. Our data suggest the bloom was either P-limited (and had excess N available) or experiencing balanced growth in the early summer (Appendix Table 2; Guildford and Hecky 2000). In the late summer and fall, almost all TN:TP ratios during this time were less than 16, suggesting the bloom was potentially N-limited (and had excess P available) or experiencing balanced growth. High signals for *nifH* expression across *Taihu* indicate cells were trying to fix N_2, supporting N limitation in *Taihu* (Figure 4). This coincides with previous experiments on *Taihu* that have demonstrated that the accumulation of primary producer biomass is P-limited in the spring and N-limited or N and P co-limited in the Summer and Fall (Paerl*, et al.* 2014b, Paerl*, et al.* 2011, Xu*, et al.* 2010). Further, *nifH* signals were higher, particularly from proteobacteria, in

the late summer and fall sampling times when TN:TP ratios were the lowest, suggesting that not only is the bloom N-limited, but so are the associated bacteria. Correlations of individual TN or TP concentrations indicated that chl *a* concentrations increased but cyanobacterial cells did not when TN was high (Appendix Figure 1). However, microcystin concentrations increased with both TN and TP (Figure 2). The correlation between chl *a* or microcystin and bulk measures of TN or TP is not necessarily sufficient evidence to conclude that changing levels of TN or TP caused growth or production of toxin as biology is a major component of these measures. High TN or TP with high chl *a* or microcystin could be indicative simply of the accumulation of biomass. A positive correlation between microcystin and TN may have indicated the importance of N in the production of toxin, or even reflect that microcystin is an N-rich molecule located within the cell, and it is hypothesized to be metabolized to serve as a N source under low N conditions (Gobler*, et al.* 2016). A better understanding of the physiological role of microcystins in the cell as well as the constraints on its production would help unravel these complex relationships.

The microarray data suggested N cycling pathways were active in the surface waters of *Taihu*. Despite fluctuations in N availability, *nifH* in particular was being expressed by a diverse community of microorganisms in *Taihu* for most of the bloom event (Figure 5, Table S9). Proteobacteria dominated the *nifH* signal nearshore while cyanobacteria dominated the *nifH* signal in the middle of the lake. This observation arose from (or likely drove) the relationship between cyanobacterial and proteobacterial *nifH* expression with lake depth (Figure 6, 7b). This agrees with previous research demonstrating that N_2-fixing cyanobacteria tend to be more prevalent in the middle of the lake (McCarthy*, et al.* 2007). With the exception of one sample, when both the proteobacteria and cyanobacteria *nifH* signal were similar, the dominance of the cyanobacterial component at Station 7 agreed with these observations (Figure 6). Cluster analysis suggests

different drivers of *nifH* expression were in play for these two groups. This analysis showed no clear distinction between the different trends of *nifH* expression based on any environmental parameters that were measured in this study together or alone other than lake depth (Figure 6, Figure 7b). Since N-fixation is typically tightly regulated and very sensitive to oxygen (Dixon and Kahn 2004, Ormeño-Orrillo*, et al.* 2013), it was surprising that no negative relationships with DO were found, although relationships to phosphorus concentrations have been suggested previously since N-fixation is an expensive process (Ward*, et al.* 2007).

NifH expression primarily by proteobacteria near shore indicated strong influence from the land or sediments. Rivers and coastal run off from agricultural soils, industrial practices and domestic wastes along with sediment resuspension contribute greatly to the inflow of nutrients to *Taihu* (Duan*, et al.* 2009). Phylogenetic analysis using *nifH* probes suggest that the proteobacteria found in the surface waters of *Taihu* may have originated from these locations: several *nifH* signals were microorganisms closely related to free-living and symbiotic N_2-fixing proteobacteria, such as *Azospirillum, Rhodospirillum, Rhodobacter, Rhizobium*, *Azorhizobium, Burkholderia, Anaeromyxobacter* and *Geobacter* (Appendix Figure 3-5, Appendix Table 9). These are associated with terrestrial non-agricultural and agricultural soils as well as lakes and sediments (Ormeño-Orrillo*, et al.* 2013, Zehr*, et al.* 2003). This suggests upwelling of *Taihu* sediments and riverine inputs may also play important roles in the N-cycling microbial community in the water column as well as the addition of nutrients.

Gene expression does not always mean activity (Rocca*, et al.* 2015, Turk*, et al.* 2011), but the abundance and diversity of *nifH* expression observed in *Taihu* surface waters strongly suggests N-fixation was occurring. Moreover, given the short half-life (5 minutes in the presence of oxygen) of *nifH* transcripts (Collins*, et al.* 1986, Dixon and Kahn 2004), it is likely this process was

occurring in the lake, and transcripts were not simply washed in on resuspended materials or materials from the shoreline. Recent studies suggest that N-fixation likely does not satisfy the N requirements of a bloom (Paerl*, et al.* 2014b, Willis*, et al.* 2016) but is important for cell maintenance (Willis*, et al.* 2016). This may explain the abundance of *nifH* transcripts observed in this dataset but no positive relationship with chl *a* or cyanobacterial cells (Figure 4, Figure 7b).

Indicative of the presence of nitrates, denitrifying processes evident in *Taihu* are classical denitrification and DNRA. Higher C:N ratios favor DNRA, which reduces NO_3 to bioavailable NH_4 (Jetten 2008). Here, *nrfA* positively correlates to TN:TP (Figure 7c), and signals were consistently detected in every sample across the bloom event, consistent with the idea that hypereutrophic *Taihu* was rich in organic C. Classical denitrification is a typically anaerobic process in which NO_3 and NO_2 are reduced to N_2 gas. Signals for *nosZ* were highest in the early to mid-summer months when TN:TP ratios were high which is also when N, including nitrate, concentrations are typically high in *Taihu* (Xu, *et al.,* 2010). Sedimentary denitrification is likely important in removal of N from the lake (Chen*, et al.* 2012, McCarthy*, et al.* 2007, Zeng*, et al.* 2008), and rates have even been detected in the water column in the presence of (and perhaps even tightly associated with) cyanobacterial blooms (Chen*, et al.* 2012, McCarthy*, et al.* 2007, Zeng*, et al.* 2008)**.** Positive probes suggest the bacteria expressing *nrfA* are *E. coli*. *E. coli* is a common component of the microbial community in *Taihu* (Wilhelm et al. 2011), and to this end its potential contribution to nitrogen cycling (and likely other processes) in this *Microcystis*-dominated system should not be too surprising.

The signal from probes targeting *nosZ* are bacteria typically associated with agricultural soils, similarly to those expressing *nifH*, such as *Azospirillum, Bradyrhizobium, Herbaspirillum* and some uncultured bacteria from paddy soils and other agricultural soils (Table S9). High signals

for *nosZ* at Station 10 suggest that these denitrifying bacteria may be transported into *Taihu via* the Dapu River.

High signals were observed for probes targeting *hzsA* indicative of anaerobic ammonia oxidation (anammox) activity. In other freshwater lakes, anammox has been shown to contribute substantially to the reduction of nitrites (Schubert*, et al.* 2006, Zhu*, et al.* 2013). Anammox bacteria and activity have been detected at the sediment-water interface in *Taihu* (Wu*, et al.* 2012, Xu*, et al.* 2009) but rates have not been measured in the water column. Interestingly, anammox bacteria, members of the phylum *Planctomycetes,* have been detected in the sediments and within bloom communities in *Taihu* and colonizing *Microcystis* mucilage elsewhere (Cai*, et al.* 2013, Wilhelm*, et al.* 2011, Woodhouse*, et al.* 2016). HzsA negatively correlated to lake depth, again demonstrating influences from sediment resuspension and coastal run-off. The sporadic signals for *hzsA* and *nosZ* may also reflect competition for substrate, since both processes are dependent on NO_2. Coupling of denitrification and anammox is possible (Jensen*, et al.* 2011, Shi*, et al.* 2013), but the gene markers here are not able to address this linkage.

Frequently coupled to anammox is aerobic ammonia oxidation, the first step in complete nitrification. While ammonia oxidizing bacteria (AOB) and archaea (AOA) have been found in the sediments in *Taihu,* the signals for *amoA* expression in surface waters were very weak (Wu*, et al.* 2010)**.** In our data there was only one probe targeting *nxrB* with any signal, and the signal was very low as well. The lack of *amoA* and *nxrB* expression across most samples and low signal otherwise suggests nitrification was not occurring in the surface waters. Probes for *amoA* and *nxrB* may have also been limited as they do not target some of the more recently described nitrifiers.

It is likely that many of these active microorganisms in the surface waters originated from the lake sediments or from nearby agricultural soils and have established a niche in the water

column. Although *Taihu* is well oxygenated, suspended sediment particles, lake snow, or cyanobacterial aggregations can provide an ideal microenvironment for many of the transformations evident here, many of which are typically anaerobic processes. The potential for microbial colonization of sediment particles and cyanobacterial colonies has been established in aquatic systems including *Taihu* (Grossart and Simon 1998, McCarthy*, et al.* 2007, Schweitzer*, et al.* 2001, Tang*, et al.* 2010, Tang*, et al.* 2009). Cyanobacterial blooms can provide protection from environmental stressors to other microorganisms, and tighter associations within cyanobacterial colonies may involve the exchange or transfer of nutrients (Cai*, et al.* 2014, Jiang*, et al.* 2007, Louati*, et al.* 2016, Shen*, et al.* 2011)**.** *Microcystis* colonies in particular have been identified as "hotspots" for bacterial production and a microhabitat in which there may be nutrient, pH or oxygen gradients (Brunberg 1999, Cai*, et al.* 2014, Paerl 1996, Worm and Søndergaard 1998). The role of heterotrophic bacteria in N transformations specifically has been hypothesized when several proteobacterial N cycling genes were detected in cyanobacterial blooms largely composed of *Microcystis* in *Taihu* (Steffen*, et al.* 2012).

With the use of a targeted microarray approach, expression of important N-cycling genes in the surface waters of hypereutrophic *Taihu* is shown with their accompanying phylogenies at high resolution. Although the probe set was extensive, expression of the N cycling genes within the bloom community may even still be underestimated, as new phylogenetic information about N cycling microorganisms continues to emerge. Our data suggest that N transformations typical of sediments and soils may be occurring in the surface waters of *Taihu,* and the process itself related to the proximity of individual sites to the shoreline. In some cases, such as with *nifH*, expression appears across all stations, however the identity of the member of the community expressing these genes changes depending on location and proximity to the shore. This study also points out the

danger in treating the entire lake as a homogenous unit: even in the northern portion of *Taihu* there are clear differences between stations in both water chemistry and function of the community. As the importance of N continues to emerge with respect to cyanoHABs, it will be important to consider these possible N transformations in the surface waters as well as the role of heterotrophic bacteria and how they directly impact bloom proliferation.

Acknowledgements

This work was supported by National Natural Science Foundation of China [41471040, 41230744], the National Oceanic and Atmospheric Administration [NA11NOS4780021], the National Science Foundation [IOS1451528, DEB1240870], the *Kenneth & Blaire Mossman Endowment* to the University of Tennessee, and a CSIRO Science Leader Fellowship to LB and the Environmental Genomics project of CSIRO O&A. We gratefully acknowledge Taihu Laboratory for Lake Ecosystem Research (TLLER), Jingchen Xue and Dr. Keqiang Shao for helping with sample collection and water chemical analysis. We also thank Drs. Hans Paerl, Silvia Newell, Mark McCarthy, Guangwei Zhu, Boqiang Qin and Ferdi Hellweger for assistance and valuable discussions.

References

Belisle B. S., Steffen M. M., Pound H. L., Watson S. B., DeBruyn J. M., Bourbonniere R. A., Boyer G. L. & Wilhelm S. W. 2016. Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*. 42, 599-607.

Blomqvist P., Petterson A. & Hyenstrand P. 1994. Ammonium-nitrogen: a key regulatory factor causing dominance of non-nitrogen-fixing cyanobacteria in aquatic systems. *Archiv für Hydrobiologie*. 132, 141-164.

Boyer G. L. 2007. The occurrence of cyanobacterial toxins in New York lakes: Lessons from the MERHAB-Lower Great Lakes program. *Lake and Reservoir Management*. 23, 153-160.

Brooks B. W., Lazorchak J. M., Howard M. D., Johnson M. V. V., Morton S. L., Perkins D. A., Reavie E. D., Scott G. I., Smith S. A. & Steevens J. A. 2016. Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems? *Environmental Toxicology and Chemistry*. 35, 6-13.

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 Microbiology Ecology*. 29, 13-22.

Bullerjahn G. S., McKay R. M., Davis T. W., Baker D. B., Boyer G. L., D'Anglada L. V., Doucette G. J., Ho J. C., Irwin E. G. & Kling C. L. 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. *Harmful Algae*. 54, 223-238.

Cai H., Jiang H., Krumholz L. R. & Yang Z. 2014. Bacterial community composition of size-fractioned aggregates within the phycosphere of cyanobacterial blooms in a eutrophic freshwater lake. *PloS one*. 9, e102879.

Cai H. Y., Yan Z. S., Wang A. J., Krumholz L. R. & Jiang H. L. 2013. Analysis of the attached microbial community on mucilaginous cyanobacterial aggregates in the eutrophic Lake Taihu reveals the importance of Planctomycetes. *Microbial ecology*. 66, 73-83.

Chaffin J. D., Bridgeman T. B. & Bade D. L. 2013. Nitrogen constrains the growth of late summer cyanobacterial blooms in Lake Erie. *Advances in Microbiology*. 2013,

Chen X., Yang L., Xiao L., Miao A. & Xi B. 2012. Nitrogen removal by denitrification during cyanobacterial bloom in Lake Taihu. *Journal of Freshwater Ecology*. 27, 243-258.

Clarke K. R. & Gorley R. N. (2006) Primer v6: user manual / tutorial. Primer – E. Plymouth, UK.

Collins J., Roberts G. & Brill W. 1986. Posttranscriptional control of *Klebsiella pneumoniae nif* mRNA stability by the *nifL* product. *Journal of bacteriology*. 168, 173-178.

Conley D. J., Paerl H. W., Howarth R. W., Boesch D. F., Seitzinger S. P., Karl E., Karl E., Lancelot C., Gene E. & Gene E. 2009. Controlling eutrophication: nitrogen and phosphorus. *Science*. 123, 1014-1015.

Cooke G. D., Welch E. B., Peterson S. & Nichols S. A. 2016. *Restoration and management of lakes and reservoirs*. CRC press.

Council N. R. 2000. Clean Coastal Waters - Understanding and Reducing the Effects of Nutrient Pollution. *National Academy Press, Washington D.C.*

Davis T. W., Bullerjahn G. S., Tuttle T., McKay R. M. & Watson S. B. 2015. Effects of increasing nitrogen and phosphorus concentrations on phytoplankton community growth and toxicity during *Planktothrix* blooms in Sandusky Bay, Lake Erie. *Environmental science & technology*. 49, 7197-7207.

Davis T. W., Harke M. J., Marcoval M. A., Goleski J., Orano-Dawson C., Berry D. L. & Gobler C. J. 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of Microcystis during cyanobacterial blooms. *Aquatic Microbial Ecology*. 61, 149-162.

Dixon R. & Kahn D. 2004. Genetic regulation of biological nitrogen fixation. *Nature Reviews Microbiology*. 2, 621-631.

Duan H., Ma R., Xu X., Kong F., Zhang S., Kong W., Hao J. & Shang L. 2009. Two-decade reconstruction of algal blooms in China's Lake Taihu. *Environmental Science & Technology*. 43, 3522-3528.

Dudgeon D., Arthington A. H., Gessner M. O., Kawabata Z. I., Knowler D. J., Lévêque C., Naiman R. J., Prieur-Richard A. H., Soto D. & Stiassny M. L. 2006. Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological reviews*. 81, 163-182.

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

EPA. 2015. Preventing Eutrophication: Scientific Support for Dual Nutrient Criteria

Gaby J. C. & Buckley D. H. 2014. A comprehensive aligned nifH gene database: a multipurpose tool for studies of nitrogen-fixing bacteria. *Database*.

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

Greaver T., Clark C., Compton J., Vallano D., Talhelm A., Weaver C., Band L., Baron J., Davidson E. & Tague C. 2016. Key ecological responses to nitrogen are altered by climate change. *Nature Climate Change*. 6, 836-843.

Grossart H.-P. & Simon M. 1998. Bacterial colonization and microbial decomposition of limnetic organic aggregates (lake snow). *Aquatic Microbial Ecology*. 15, 127-140.

Guildford S. J. & Hecky R. E. 2000. Total nitrogen, total phosphorus, and nutrient limitation in lakes and oceans: Is there a common relationship? *Limnology and Oceanography*. 45, 1213-1223.

Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W. & Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systems Biology*. 59, 307-321.

Guo L. 2007. Doing battle with the green monster of Taihu Lake. *Science*. 317, 1166-1166.

Han X., Zhu G., Xu H., Wilhelm S., Qin B. & Li Z. 2014. Source analysis of urea-N in Lake Taihu during summer. *Environmental Science*. 35, 2547-2556.

Harke M. J., Davis T. W., Watson S. B. & Gobler C. J. 2015. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environmental science & technology*. 50, 604-615.

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

He X., Liu Y.-L., Conklin A., Westrick J., Weavers L. K., Dionysiou D. D., Lenhart J. J., Mouser P. J., Szlag D. & Walker H. W. 2016. Toxic cyanobacteria and drinking water: Impacts, detection, and treatment. *Harmful Algae*. 54, 174-193.

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. 2008. Eutrophication and harmful algal blooms: a scientific consensus. *Harmful algae*. 8, 3-13.

Jensen M. M., Lam P., Revsbech N. P., Nagel B., Gaye B., Jetten M. S. & Kuypers M. M. 2011. Intensive nitrogen loss over the Omani Shelf due to anammox coupled with dissimilatory nitrite reduction to ammonium. *The ISME journal*. 5, 1660-1670.

Jetten M. S. 2008. The microbial nitrogen cycle. *Environmental Microbiology*. 10, 2903-2909.

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.). *Hydrobiologia*. 581, 161-165.

Jin X. & Tu Q. 1990. The standard methods for observation and analysis in lake eutrophication. *Chinese Environmental Science Press, Beijing*. 138272,

Lewis Jr W. M., Wurtsbaugh W. A. & Paerl H. W. 2011. Rationale for control of anthropogenic nitrogen and phosphorus to reduce eutrophication of inland waters. *Environmental science & technology*. 45, 10300-10305.

Louati I., Pascault N., Debroas D., Bernard C., Humbert J.-F. & Leloup J. 2016. Correction: Structural Diversity of Bacterial Communities Associated with Bloom-Forming Freshwater Cyanobacteria Differs According to the Cyanobacterial Genus. *PloS one*. 11,

McCarthy M. J., Lavrentyev P. J., Yang L., Zhang L., Chen Y., Qin B. & Gardner W. S. 2007. Nitrogen dynamics and microbial food web structure during a summer cyanobacterial bloom in a subtropical, shallow, well-mixed, eutrophic lake (Lake Taihu, China). *Hydrobiologia*. 581, 195-207.

Michalak A. M., Anderson E. J., Beletsky D., Boland S., Bosch N. S., Bridgeman T. B., Chaffin J. D., Cho K., Confesor R. & Daloğlu I. 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*. 110, 6448-6452.

Neilan B. A., Pearson L. A., Muenchhoff J., Moffitt M. C. & Dittmann E. 2013. Environmental conditions that influence toxin biosynthesis in cyanobacteria. *Environmental microbiology*. 15, 1239-1253.

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

Ormeño-Orrillo E., Hungria M. & Martinez-Romero E. 2013. Dinitrogen-fixing prokaryotes. *The prokaryotes,* 427-451. Springer.

Paerl H. W. 1996. Microscale physiological and ecological studies of aquatic cyanobacteria: macroscale implications. *Microsc Res Tech*. 33, 47-72.

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

Paerl H. W. & Otten T. G. 2013. Harmful cyanobacterial blooms: causes, consequences, and controls. *Microbial ecology*. 65, 995-1010.

Paerl H. W., Gardner W. S., McCarthy M. J., Peierls B. L. & Wilhelm S. W. 2014. Algal blooms: Noteworthy nitrogen. *Science*. 346, 175.

Paerl H. W., Xu H., McCarthy M. J., Zhu G., Qin B., Li Y. & Gardner W. S. 2011. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Research*. 45, 1973-1983.

Paerl H. W., Xu H., Hall N. S., Rossignol K. L., Joyner A. R., Zhu G. & Qin B. 2015. Nutrient limitation dynamics examined on a multi-annual scale in Lake Taihu, China: implications for controlling eutrophication and harmful algal blooms. *Journal of Freshwater Ecology*. 30, 5-24.

Paerl H. W., Scott J. T., McCarthy M. J., Newell S. E., Gardner W., Havens K. E., Hoffman D. K., Wilhelm S. W. & Wurtsbaugh W. A. 2016. It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science & Technology*.

Paerl H. W., Xu H., Hall N. S., Zhu G., Qin B., Wu Y., Rossignol K. L., Dong L., McCarthy M. J. & Joyner A. R. 2014. Controlling Cyanobacterial Blooms in Hypertrophic Lake Taihu, China: Will Nitrogen Reductions Cause Replacement of Non-N_2 Fixing by N_2 Fixing Taxa? *PloS one*. 9, e113123.

Qin B., Xu P., Wu Q., Luo L. & Zhang Y. 2007. Environmental issues of Lake Taihu, China. *Hydrobiologia*. 581, 3-14.

Qin B., Zhu G., Gao G., Zhang Y., Li W., Paerl H. W. & Carmichael W. W. 2010. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *Environmental management*. 45, 105-112.

Rocca J. D., Hall E. K., Lennon J. T., Evans S. E., Waldrop M. P., Cotner J. B., Nemergut D. R., Graham E. B. & Wallenstein M. D. 2015. Relationships between protein-encoding gene abundance and corresponding process are commonly assumed yet rarely observed. *The ISME journal*. 9, 1693-1699.

Schindler D. W., Hecky R., Findlay D., Stainton M., Parker B., Paterson M., Beaty K., Lyng M. & Kasian S. 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*. 105, 11254-11258.

Schubert C. J., Durisch-Kaiser E., Wehrli B., Thamdrup B., Lam P. & Kuypers M. M. 2006. Anaerobic ammonium oxidation in a tropical freshwater system (Lake Tanganyika). *Environmental Microbiology*. 8, 1857-1863.

Schweitzer B., Huber I., Amann R., Ludwig W. & Simon M. 2001. α-and β-Proteobacteria control the consumption and release of amino acids on lake snow aggregates. *Applied and Environmental Microbiology*. 67, 632-645.

Scott J. T. & McCarthy M. J. 2010. Nitrogen fixation may not balance the nitrogen pool in lakes over timescales relevant to eutrophication management. *Limnology and Oceanography*. 55, 1265-1270.

Shen H., Niu Y., Xie P., Tao M. & Yang X. 2011. Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshwater Biology*. 56, 1065-1080.

Shi Y., Hu S., Lou J., Lu P., Keller J. & Yuan Z. 2013. Nitrogen removal from wastewater by coupling anammox and methane-dependent denitrification in a membrane biofilm reactor. *Environmental science & technology*. 47, 11577-11583.

Socolow R. H. 1999. Nitrogen management and the future of food: lessons from the management of energy and carbon. *Proceedings of the National Academy of Sciences*. 96, 6001-6008.

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, e44002.

Tamura K., Dudley J., Nei M. & Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology Evolution*. 24, 1596-1599.

Tang X., Gao G., Chao J., Wang X., Zhu G. & Qin B. 2010. Dynamics of organic-aggregate-associated bacterial communities and related environmental factors in Lake Taihu, a large eutrophic shallow lake in China. *Limnology and oceanography*. 55, 469-480.

Tang X., Gao G., Qin B., Zhu L., Chao J., Wang J. & Yang G. 2009. Characterization of bacterial communities associated with organic aggregates in a large, shallow, eutrophic freshwater lake (Lake Taihu, China). *Microbial ecology*. 58, 307-322.

Turk K. A., Rees A. P., Zehr J. P., Pereira N., Swift P., Shelley R., Lohan M., Woodward E. M. S. & Gilbert J. 2011. Nitrogen fixation and nitrogenase (nifH) expression in tropical waters of the eastern North Atlantic. *The ISME journal*. 5, 1201-1212.

Ward B. B., Capone D. G. & Zehr J. P. 2007. What's new in the nitrogen cycle? *Oceanography*. Washington D.C. Oceanography Society. 20, 101.

Watson S. B., Miller C., Arhonditsis G., Boyer G. L., Carmichael W., Charlton M. N., Confesor R., Depew D. C., Höök T. O. & Ludsin S. A. 2016. The re-eutrophication of Lake Erie: Harmful algal blooms and hypoxia. *Harmful Algae*. 56, 44-66.

Wilhelm S. W., Farnsley S. E., LeCleir G. R., Layton A. C., Satchwell M. F., DeBruyn J. M., Boyer G. L., Zhu G. & Paerl H. W. 2011. The relationships between nutrients, cyanobacterial toxins and the microbial community in Taihu (Lake Tai), China. *Harmful Algae*. 10, 207-215.

Willis A., Chuang A. W. & Burford M. A. 2016. Nitrogen fixation by the reluctant diazotroph *Cylindrospermopsis raciborskii* (Cyanophyceae). *Journal of Phycology*.

Woodhouse J. N., Kinsela A. S., Collins R. N., Bowling L. C., Honeyman G. L., Holliday J. K. & Neilan B. A. 2016. Microbial communities reflect temporal changes in cyanobacterial composition in a shallow ephemeral freshwater lake. *The ISME journal*. 10, 1337-1351.

Worm J. & Søndergaard M. 1998. Dynamics of heterotrophic bacteria attached to *Microcystis* spp. (Cyanobacteria). *Aquatic microbial ecology*. 14, 19-28.

Wu X., Jiang J. & Hu J. 2013. Determination and occurrence of retinoids in a eutrophic lake (Taihu Lake, China): cyanobacteria blooms produce teratogenic retinal. *Environmental Science Technology*. 47, 807-814.

Wu Y., Xiang Y., Wang J. & Wu Q. L. 2012. Molecular detection of novel anammox bacterial clusters in the sediments of the shallow freshwater Lake Taihu. *Geomicrobiology Journal*. 29, 852-859.

Wu Y., Xiang Y., Wang J., Zhong J., He J. & Wu Q. L. 2010. Heterogeneity of archaeal and bacterial ammonia-oxidizing communities in Lake Taihu, China. *Environmental microbiology reports*. 2, 569-576.

Xu H., Zhang L., Shang J., Dai J. & Fan C. 2009. Denitrification and anammox on the sediment-water interface in the Meiliang Bay of Lake Taihu. *Journal of Lake Sciences*. 6,

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

Zehr J. P., Crumbliss L. L., Church M. J., Omoregie E. O. & Jenkins B. D. 2003. Nitrogenase genes in PCR and RT-PCR reagents: implications for studies of diversity of functional genes. *Biotechniques*. 35, 996-1013.

Zeng J., Yang L. Y., Xiao L. & Du H.-W. 2008. Comparative study on transformation potential of dissolved inorganic nitrogen in different parts of Lake Taihu. *Journal of Ecology and Rural Environment*. 1, 014.

Zhu G., Wang S., Wang W., Wang Y., Zhou L., Jiang B., den Camp H. J. O., Risgaard-Petersen N., Schwark L. & Peng Y. 2013. Hotspots of anaerobic ammonium oxidation at land-freshwater interfaces. *Nature Geoscience*. 6, 103-107.

Appendix

Figures

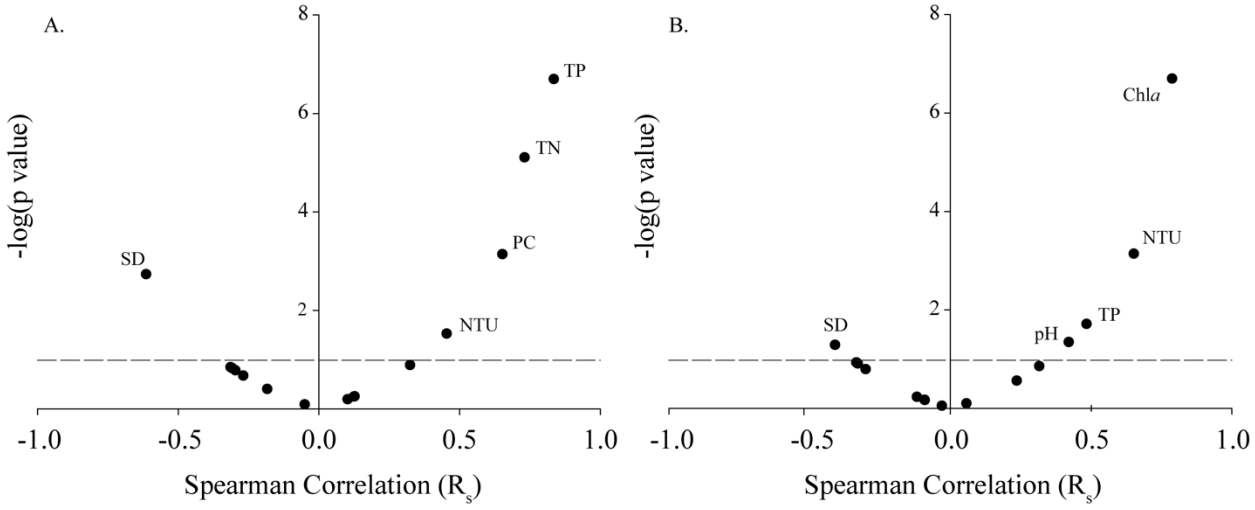

Figure A2.1. Environmental data correlations
Volcano plot representing non-parametric correlations to A). chl*a* and B). cyanobacterial cells in *Taihu* across the bloom event. Any parameter above the dashed line correlates to microcystin concentrations at a p value of less than 0.10.

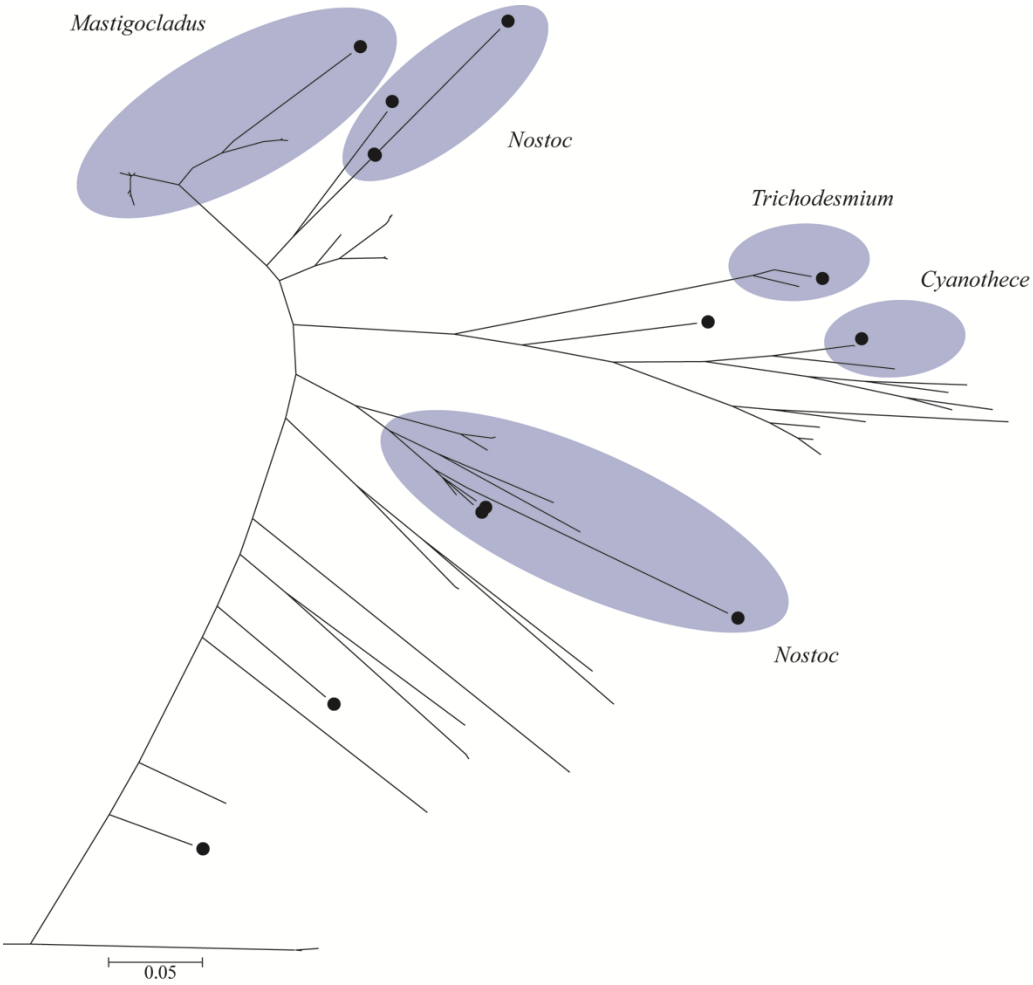

Figure A2.2. Enlarged view of Cyanobacterial branch from the *nifH* phylogenetic tree in Figure 7

Blue circles represent groups at a genus level. Black circles represent the probes with positive results. Clear phylogenetic groups could not be assigned to the probes that are not within designated groups.

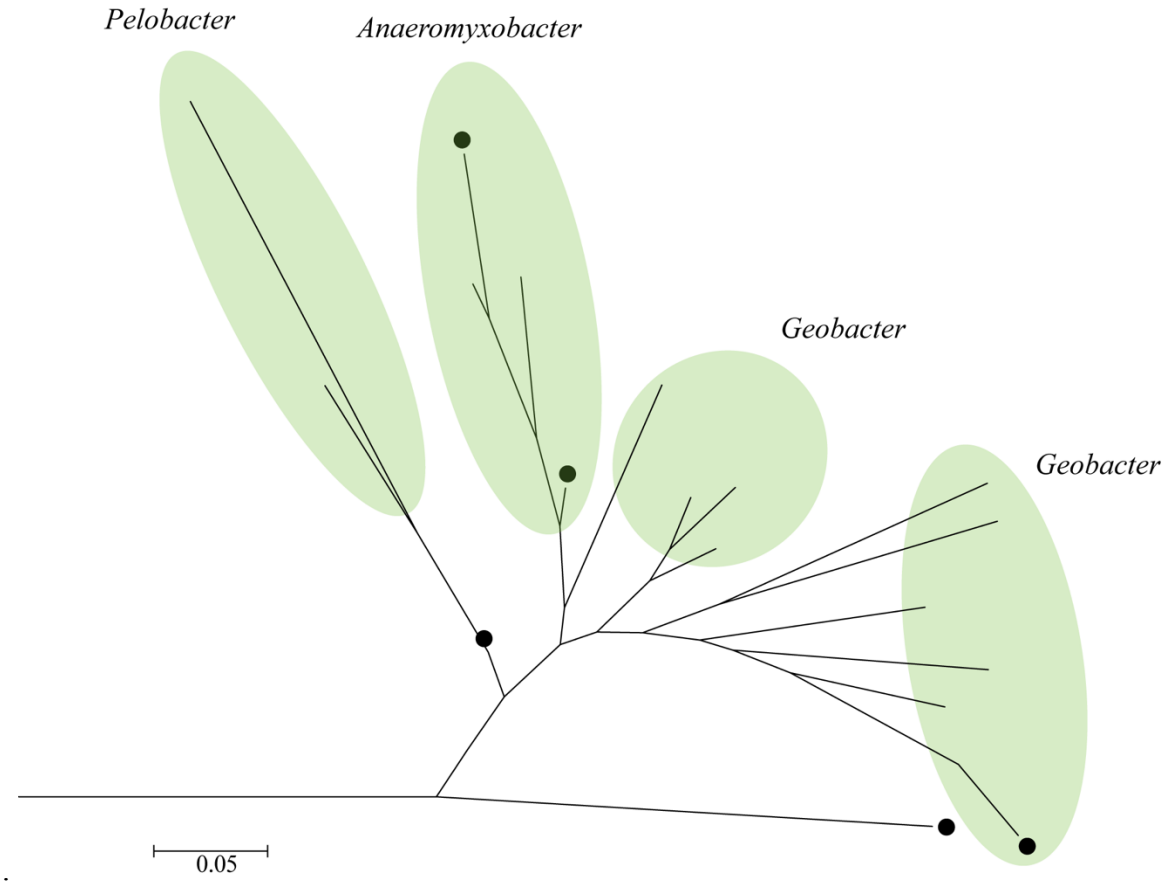

Figure A2.3. Enlarged view of Deltaproteobacteria branch from the *nifH* phylogenetic tree in Figure 7
Green circles represent groups at a genus level. Black circles represent the probes with positive results

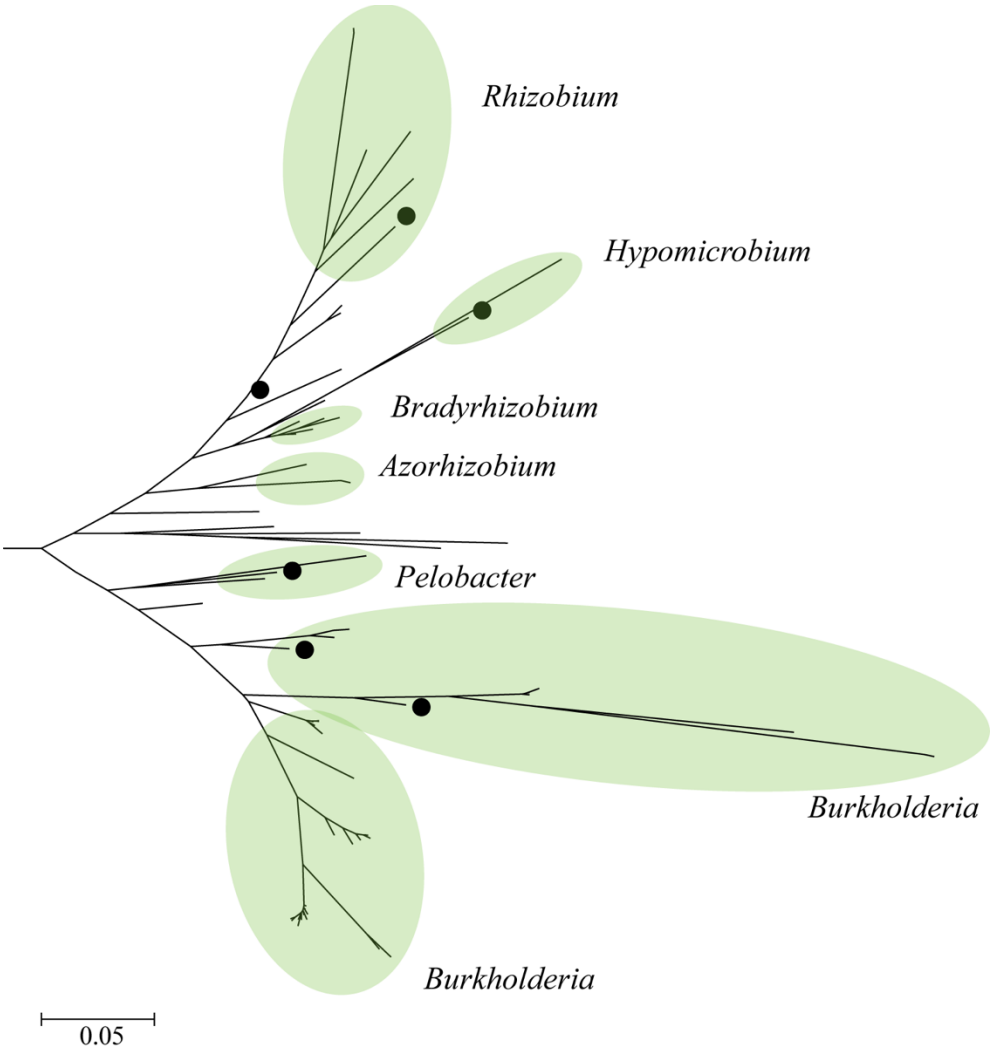

Figure A2.4. Enlarged view of an Alpha- and Betaproteobacteria branch from the *nifH* phylogenetic tree in Figure 7
Green circles represent groups at a genus level. Black circles represent the probes with positive results.

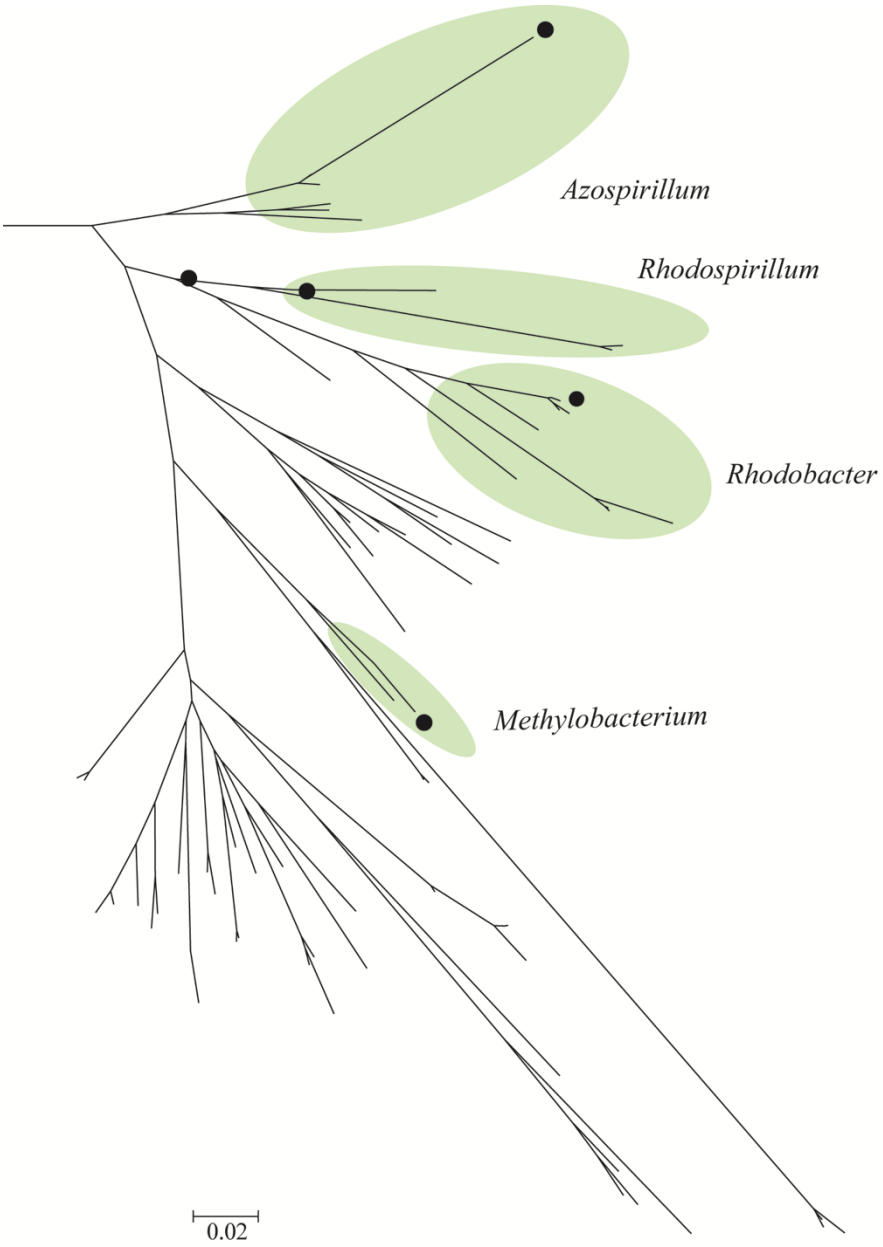

Figure A2.5. Enlarged view of an Alphaproteobacteria branch from the *nifH* phylogenetic tree in Figure 2.7
Green circles represent groups at a genus level. Black circles represent the probes with positive results.

Tables

Table A2.1. Volumes (mL) of surface water filtered through Sterivex. Volumes varied due to differences in suspended materials between samples.

	June 6, 2014	*July 3, 2014*	*August 14, 2014*	*September 9, 2014*	*October 8, 2014*
Station 1	320	n/a	120	120	180
Station 7	120	n/a	120	90	150
Station 10	220	n/a	120	120	100
Channel	120	125 +	120	20	n/a
Dock	100	n/a	120	120	120

Table A2.2. PCR primers and conditions for the N-cycle array. A and B before *amoA* distinguishes between Archaea and Bacteria, respectively.

Target	Primer	Sequence	T_{ann} / # cycles	Fragment size (bp)	Reference
A-*amoA*	Arch amoA-1F	STAATGGTCTGGCTTAGACG	63 °C - > 50 °C / 13 49 °C / 25	635	Francis *et al.* (2005)
	Arch amoA-2R-T7	TAATACGACTCACTATAGGCGGCC ATCCATCTGTATGT			
B-*amoA*	amoA-1F	GGGGTTTCTACTGGTGGT	63 °C - > 50 °C / 13 49 °C / 25	490	Rotthauwe *et al.* (1997)
	amoA-2R-T7	TAATACGACTCACTATAGCCCCTCK GSAAAGCCTTCTTC			
hzsA primary fragment	hzsA_382F	GGYGGDTGYCAGATATGGG	65 °C - > 55 °C / 10 55 °C / 25	2000	Harhangi *et al.* (2012)
	hzsA_2390R	ATRTTRTCCCAYTGYGCHCC			
hzsA final fragment	hzsA_526F	TAYTTTGAAGGDGACTGG	60 °C - > 45 °C / 30 44 °C / 30	1331	Harhangi *et al.* (2012)
	hzsA_1857R-T7	TAATACGACTCACTATAGAAABGG YGAATCATARTGGC			
nifH	nifH2_F	GYGAYCCNAARGCNGA	60 °C - > 45 °C / 30 44 °C / 30	362	Zehr & McReynolds (1989)
	nifH1_R-T7	TAATACGACTCACTATAGADNGCC ATCATYTCNCC			
nosZ	nosZ-F	CGYTGTTCMTCGACAGCCAG	63 °C - > 50 °C / 13 49 °C / 25	453	Kloos *et al.* (2001)
	nosZ 1622R-T7	TAATACGACTCACTATAGCGSACCT TSTTGCCSTYGCG			Throbäck *et al.* (2004)
nrfA	nrfA_F1	GCNTGYTGGWSNTGYAA	60 °C - > 45 °C / 30 44 °C / 30		
	nrfA_R1-T7	TAATACGACTCACTATAGTWNGGC ATRTGRCARTC		490	Mohan *et al.* (2004)
nxrB - Nitrobacter	nxrB1F	ACGTGGAGACCAAGCCGGG	63 °C - > 50 °C / 13 49 °C / 25	380	Vanparys *et al.* (2007)
	nxrB1R-T7	TAATACGACTCACTATAGCCGTGCT GTTGAYCTCGTTGA			
	T7	TAATACGACTCACTATAG			

Table A2.3. The nitrogen cycling array probe set. A and B before *amoA* distinguishes between Archaea and Bacteria, respectively.

Probe	Sequence
hyaBP60	TTTCTTCGACGTACAAAACCGCCTGAAAAAATTTGTTGAAGGCGGGCAGTTGGGGATCTT
16S-B1	CAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCC
16S-B2	GACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAAC
16S-A1	GAGAGGAGGTGCATGGCCGCCGTCAGCTCGTACCGTGAGGCGTCCTGTTAAGTCAGGCAACGAGCGAGAC
16S-A2	AATTGGCGGGGAGCACTACAACCGGAGGAGCCTGCGGTTTAATTGGACTCAACGCCGGACATCTCACCA
16S-A3	CTGAACTTAAAGGAATTGGCGGGGGAGCACCACAAGGGGTGGAACCTGCGGCTCAATTGGAGTCAACGC
BamoA-05	CATTTTTGGCCCGACCCACCTGCCGCTCGTGGCTGAAGGCGTCCTGCTCTCCCTGGCCGACTACACCGGC
BamoA-06	TGCTGTTCTACCCGGGCAACTGGCCGATCTTTGGGCCGACCCACCTGCCGCTCGTGGCTGAAGGCGTCCT
BamoA-56	GGTGGTACTTTGGCAAGGTCTACTGCACCGCCTTCTACTATGTCAAGGGTGCGCGCGGCCGAGTCAGCAT
BamoA-57	CCATGCTCATGTTCACCGTCTGGTGGTACTTTGGCAAGGTCTACTGCACCGCCTTCTACTATGTCAAGGG
BamoA-04	TGGCAAAGTCTACTGCACCGCCTTCTACTACGTAAAAGGCGCCCGTGGCCGCGTCAGCATGAAGAACGAC
BamoA-52	TACTTTGGCAAAGTCTACTGCACCGCCTTCTACTACGTAAAAGGCGCCCGTGGCCGCGTCAGCATGAAGA
BamoA-53	TAGCCGACTACACAGGCTTCCTGTATGTACGCACCGGCACGCCCGAGTACGTGCGCCTGATCGAACAAGG
BamoA-02	GGCTGACTACACCGGCTTCCTGTATGTCCGCACGGGCACCCCGGAATACGTACGGCTGATCGAACAAGGC
BamoA-03	GAATACGTACGGCTGATCGAACAAGGCTCACTACGCACCTTTGGTGGCCACACCACCGTGATTGCGGCCT
BamoA-01	CGTTCGTCTCCATGCTCATGTTCTGCGTCTGGTGGTACTTTGGCAAACTCTACTGCACCGCGTTCTACTA
BamoA-54	TCTCCGCGTTCGTCTCCATGCTCATGTTCACCGTCTGGTGGTACTTTGGCAAGGTCTACTGCACCGCCTT
BamoA-55	CCATGCTCATGTTCACCGTCTGGTGGTACTTTGGCAAGGTCTACTGCACCGCCTTCTACTATGTCAAGGG
BamoA-50	AGGCGCACTGGTCATGGACACTGTCATGCTGCTCAGCCGTAACTGGATGATCACAGCCTTGGTTGGGGGT
BamoA-51	ACACTGTCATGCTGCTCAGCCGTAACTGGATGATCACAGCCTTGGTTGGGGGTGGCGCATTCGGACTGCT
BamoA-48	ACCCATCTGCCGTTGGTAGCAGAAGGCGTCCTGTTGTCGGTTGCTGATTACACGGGCTTTCTCTATGTTC
BamoA-49	AGCTGGTTGATTACAGCACTGGTTGGTGGCGGCGCATTCGGTTTGTTGTTCTATCCGGGTAACTGGCCGA
BamoA-45	ATCGAGCAAGGATCGTTGCGTACCTTTGGTGGTCACACCACGGTAATTGCCGCGTTCTTCTCGGCCTTCG
BamoA-46	CGGTTGATCGAGCAAGGATCGTTGCGTACCTTTGGTGGTCACACCACGGTAATTGCCGCGTTCTTCTCGG
BamoA-47	ACCTGCCGCTGGTTGCAGAAGGGGTCTTGCTGTCGGTTGCTGACTACACCGGGTTCCTCTATGTTCGTAC
BamoA-43	GGTGGTGGATTCTGGGGCTTGTTCTTCTACCCAGGTAACTGGCCAATCTTTGGACCAACCCACTTGCCAC
BamoA-44	ACTGGCTGATCACAGCATTAGTTGGTGGTGGATTCTGGGGCTTGTTCTTCTACCCAGGTAACTGGCCAAT
BamoA-41	AAAATCTATTGCACCGCGTTTTACTATGTTAGAGGCGAAAGAGGCCGTATCTCGCAGAAACACGACGTCA
BamoA-42	ACTTTGGCAAAATCTATTGCACCGCGTTTTACTATGTTAGAGGCGAAAGAGGCCGTATCTCGCAGAAACA
BamoA-36	ATCAACGATGATTCCAGGTGCACTGATGCTGGATACCATCATGCTGTTGACGGGTAACTGGTTGATTACG

Table A2.3 continued.

Probe	Sequence
BamoA-37	CTGTTGGGTGGTGGTTTCTGGGGCTTGTTCTTTTACCCAGGCAACTGGCCTATTTTCGGCCCGACCCACT
BamoA-38	ATGATACCTGGCGCATTGATGCTGGACACAATCATGTTACTGACTGGTAACTGGTTGATTACTGCATTAT
BamoA-39	TGATACCTGGTGCATTGATGCTGGATACAATCATGTTACTGACTGGTAACTGGTTGATAACTGCATTATT
BamoA-40	CTGGACACAATCATGTTACTGACTGGTAACTGGTTGATTACTGCATTATTGGGTGGTGGATTCTGGGGAT
BamoA-34	TTACGGCATTGCTAGGTGGTGGTTTCTGGGGTCTGTTCTTTTACCCTGGTAACTGGCCGATTTTTGGGCC
BamoA-35	CTGGGGTCTGTTCTTTTACCCTGGTAACTGGCCGATTTTTGGGCCAACCCATTTACCGTTGGTTGTAGAA
BamoA-32	GTAGAAGGCGTATTGCTTTCAGTTGCTGACTACACTGGTTTCCTCTATGTGCGTACAGGTACACCTGAAT
BamoA-33	ATGTGCGTACAGGTACACCTGAATATGTGAGGCTGATTGAACAAGGGTCGCTACGAACCTTTGGTGGTCA
BamoA-29	ATGTTCTGCGTATGGTGGTACTTTGGCAAACTGTACTGCACCGCTTTCTACTATGTTAAAGGAGAAAGAG
BamoA-30	CTTTGGCAAACTGTACTGCACCGCTTTCTACTATGTTAAAGGAGAAAGAGGACGTATCTCGATGAAGAAC
BamoA-31	GTATCCATGCTGATGTTCTGTGTATGGTGGTACTTTGGCAAACTGTATTGCACCGCTTTCTACTATGTTA
BamoA-15	CTGTACAGCCTTTTTCTACGTTAAAGGTAAAAGAGGTCGTATCGTACATCGCAATGATGTTACCGCATTC
BamoA-16	GTATGGTGGTATCTTGGAAAAGTTTACTGTACAGCCTTTTTCTACGTTAAAGGTAAAAGAGGTCGTATCG
BamoA-17	ACCCCATCGATCATGATTCCGGGTGCATTGATGTTGGATATCACGTTGTACTTGACCCGTAACTGGTTGG
BamoA-18	ATGTTGGATATCACGTTGTACTTGACCCGTAACTGGTTGGTAACGGCGCTGATCGGAGGCGGATTCTTTG
BamoA-26	GTCGTCTGGTGGTATCTCGGTAAAATTTATTGCACAGCCTTCTACTATGTTAAAGGCAAAAGAGGACGTA
BamoA-27	GTCGTCTGGTGGTATCTCGGTAAAATTTATTGCACAGCGTTCTACTATGTTAAAGGCAGAAGAGGCCGTA
BamoA-28	AGTCGGCGGTGGATTTTTTGGTTTGTTGTTTTATCCAGGAAACTGGCCAATTTTTGGACCCACCCATCTG
BamoA-24	GGTGGATTCTTTGGTTTACTCTTCTATCCAGGCAACTGGGTAATTTTTGGACCAACTCACTTGCCAGTCG
BamoA-25	AGTTTACTGTACAGCTTTCTTCTACGTTAAAGGTAAGAGAGGCCGTATTGTGAAAAGAGACGACGTTACA
BamoA-20	ATCGATGGCTGACTATATGGGACACATGTATGTACGTACAGGTACTCCTGAATATGTTCGCTTAATCGAA
BamoA-21	TCGTAGTTGAAGGCCAATTATTATCGATGGCTGACTATATGGGACACATGTATGTACGTACAGGTACTCC
BamoA-22	AATCATGATTCCAGGCGCGTTAATGTTAGATATTACTTTATATCTAACCCGTAGCTGGCTCGTGACAGCG
BamoA-23	TGTTAGATATTACTTTATATCTAACCCGTAGCTGGCTCGTGACAGCGTTAATCGGTGGCGGGTTCTTTGG
BamoA-19	ACGGCGCTACTTGGTGGTGCATTTTTTGGTTTATTGTTCTATCCGGGCAACTGGACAATATTTGGACCTA
BamoA-13	AATCACAGCATTGCTTGGCGGCGGAGCTTTTGGTTTGCTGTTTTATCCGGGCAACTGGGTAATTTTTGGA
BamoA-14	ACAGCGTTCTACTATGTTAAAGGTCCAAGAGGCAAGATAACAGAGAAGATGGATGTTACCGCTTTTGGAG
BamoA-59	CGATGATTCCTGGGGCATTGATGATGGATACGATACTATTGTTGACCCGTAACTGGATGATTACAGCGTT
BamoA-60	GATGATGGATACGATACTATTGTTGACCCGTAACTGGATGATTACAGCGTTGCTTGGCGGCGGATGTTTT
BamoA-11	ACGATGATTCCTGGGGCATTGATGATGGATACGATACTATTGTTGACCCGTAACTGGATGATTACAGCGT
BamoA-12	AACTCACCTCCCATTAGTAGTAGAAGGGGTGTTACTCTCAGTGGCTGACTATACCGGGTTCCTATATGTA

Table A2.3 continued.

Probe	Sequence
BamoA-07	CCTGAGTATGTTCGGTTGATTGAACAAGGATCGCTACGTACCTTCGGTGGTCATACTACGGTGATTGCTG
BamoA-10	ATGTTCGTACTGGTACACCTGAGTATGTTCGGTTAATTGAACAAGGGTCACTACGTACCTTCGGTGGTCA
BamoA-58	GTTCACCGTCTGGTGGTACTTTGGCAAGGTCTACTGCACCGCCTTCTACTATGTCAAGGGTGCGCGCGGC
BamoA-08	CGGCCTTTGTATCCATGTTAATGTTCGTTGTCTGGTGGTACCTGGGAAGGTTCTACTGCACTGCGTTCTA
BamoA-09	GCGGAGCCTTTGGGCTACTGTTCTATCCGGGTAACTGGCCGATCTTCGGACCGACTCACCTCCCGTTAGT
AamoA-06	AACAGTCCTGTTGCTTTAGGTGCTGGTGCTGGTTCGGTATTGGGAGTAACGTTTACAGCACTTGGTTGTA
AamoA-07	AAGTTCTATAACAGTCCTGTTGCTTTAGGTGCTGGTGCTGGTTCGGTATTGGGAGTAACGTTTACAGCAC
AamoA-08	CAGGTAGGAAAGTTCTATAACAGTCCTGTTGCTTTAGGTGCTGGTGCTGGTTCGGTATTGGGAGTAACGT
AamoA-09	ACTTCGTACACGGTATTCTCAATATCGCAAACGTTGATGCTTGTAGTAGGTGCAACATATTACCTTACAT
AamoA-10	ACTGGGCTTGGACTTCGTACACGGTATTCTCAATATCGCAAACGTTGATGCTTGTAGTAGGTGCAACATA
AamoA-12	ATATCCAAGACCAACGCTACCACCATACATGGTACCAATTGAACCGCAGGTCGGAAAGTTCTATAACAGT
AamoA-13	TGTTGACAGTCGCAGACCCACTAGAAACAGCGTTCAAATATCCAAGACCAACGCTACCACCATACATGGT
AamoA-11	GTAGGAATGTCATTGCCACTGTTTAACATGGTTAACTTGATTACGGTCGCAGACCCACTAGAAACGGCAT
AamoA-14	ACATCTTCTACACTGACTGGGCCTGGACTTCATACACTGTATTCTCAATTTCACAAACATTGATGCTTAC
AamoA-15	AACCCCAAGTAGGCAAGTTCTATAACAGCCCTGTAGCATTAGGTGCTGGAGCTGGTGCAGTATTGTCGGT
AamoA-01	ACACTGTTAACAATTAATGCAGGAGACTACATCTTCTACACTGACTGGGCTTGGACTTCGTACACGGTAT
AamoA-02	GACTACATCTTCTACACTGACTGGGCTTGGACTTCGTACACGGTATTTTCAATATCGCAAACGTTGATGC
AamoA-04	AGTTAACTCAACACTGTTAACAATTAATGCAGGAGACTACATTTTCTATACTGACTGGGCTTGGACTTCG
AamoA-05	ACTACATTTTCTATACTGACTGGGCTTGGACTTCGTTCACGGTATTTTCAATATCGCAAACGTTGATGCT
AamoA-03	TTCAACATGGTAAACCTGATAACAGTAGCAGACCCACTAGAAACGGCATTCAAATATCCAAGACCAACAT
AamoA-16	ACACGTGGGTGGCCAAAGGTGCATGGTTTGCACTGGGTTATCCATACGACTTCATCGTAGTGCCAGTATG
AamoA-17	CCCTGTTATCAATCAATGCAGGTGACTACATCTTCTACTCAGACTGGGCCTGGACATCATTTGTAGTATT
AamoA-19	CTTGGTATACTGGGCCACCAAGAAGAACAAACACAGTCTGATACTGTTCGGAGGTGTGTTGGTAGGCATG
AamoA-18	AATGTTACTTGACTTGGTATACTGGGCCACCAAGAAGAACAAACACAGTCTGATACTGTTCGGAGGTGTG
AamoA-20	CAATCAACGCGGGTGACTACATCTTCTACACAGACTGGGCTTGGACATCGTTTGTCATATTCTCAATATC
AamoA-21	ATCTTCTACACAGACTGGGCTTGGACATCGTTTGTCATATTCTCAATATCACAAACGCTGATGCTCTCAG
AamoA-22	GTGGAATGTCATTGCCGTTGTTTAACATGGTCAACCTCATCACTGTGGCTGATCCATTAGAGACTGCATT
AamoA-23	TTAACATGGTCAACCTCATCACTGTGGCTGATCCATTAGAGACTGCATTCAAATATCCAAGACCAACGCT
AamoA-24	AGAACAAGCACTCACTGATTCTATTCGGTGGTGTGTTGTGTGGAATGTCATTGCCATTGTTCAACATGGT
AamoA-25	ACTCACTGATTCTATTCGGTGGTGTGTTGTGTGGAATGTCATTGCCATTGTTCAACATGGTCAACCTAAT

Table A2.3 continued.

Probe	Sequence
AamoA-26	AGTTCTATAACAGTCCAGTTGCACTCGGTGCAGGAGCAGGTGCTGTATTGTCAGTTACCATGTGCGCACT
AamoA-27	ACACTGACTGGGCCTGGACTTCGTTTGTCGTATTCTCAATATCGCAAACGTTGATGTTGGTAGTCGGCGC
AamoA-28	CATCTTCTACACTGACTGGGCCTGGACTTCGTTTGTCGTATTCTCAATATCGCAAACGTTGATGTTGGTA
AamoA-29	CAGATTGGGCTTGGACATCCTTTGTAGTATTTTCCATCTCACAAACTTTGATGCTTTGTGTTGGTGCAA
AamoA-30	CATCCTTTGTAGTATTTTCCATCTCACAAACTTTGATGCTTTGTGTTGGTGCAACATATTACCTAACATT
AamoA-31	ACATGGTAAACCTGATTACAGTAGCGGATCCATTGGAGACTGCATTCAAGTATCCAAGACCTACGTTGCC
AamoA-32	ATGGTCTGGCTTAGACGTTGTACTCACTATTTATTCATAATAGTGGTCGCGGTAAACGGAACTCTGCTTA
AamoA-33	ACGTTGTACTCACTATTTATTCATAATAGTGGTCGCGGTAAACGGAACTCTGCTTACAATTAATGCCGGT
AamoA-34	AACATTTACAGGCGTTCCAGGGACAGCTACTTACTACGCGTTAATTATTACTGTGTATACTTGGGTAGCA
AamoA-35	CAGTAAACTCAACATTACTTACCATTAATGCTGGAGACTACATTTTCTACACCGATTGGGCCTGGACATC
AamoA-36	GACAAAAAAGAACAAGCACTCTCTGATTCTATTCGGCGGTGTATTGGTAGGAATGTCGTTACCTTTGTTC
AamoA-37	TGGACATCATACACCGTATTTTCAATTTCACAAACTTTGATGTTAACGGTAGGTGCAGTATACTACCTTA
AamoA-38	TCAATATCGCAAACGTTGATGCTAGCTGTTGGAGCTTCGTATTATCTTACATTTACTGGGGTTCCTGGAA
AamoA-39	TGGAGTTTTAGTCGGAATGTCGTTACCGCTATTCAACATGGTAAATTTGATTACGGTAGCAGATCCACT
AamoA-40	ATGCTTACGGTAGGAGCAGTATACTATCTAACATTTACAGGTGTTCCAGGAACCGCAACATACTACGCAT
AamoA-41	CTAACGTTTACAGGTGTTCCAGGAACTGCAACATATTATGCACTAATCATGACGGTGTATACATGGGTAG
AamoA-42	CTTTATTGTTACACCAGTATGGATCCCATCGGCGATATTGTTAGATTTGGCGTATTGGGCTACAAAGAAG
AamoA-43	TGTTCGGCGGAGTCTTAGTTGGAATGTCATTACCACTATTCAACATGGTCAATTTGATCACAGTGGCTGA
AamoA-44	GCGTACTGGGCAACAAAGAAGAATAAGCACTCTCTGATACTGTTCGGCGGAGTCTTAGTTGGAATGTCAT
AamoA-48	TTAATTCAGGAGACTATATCTTCTATACTGACTGGGCTTGGACATCATTCGTAGTTTTCTCTATCGCCAA
AamoA-45	TACTGGTGTACCTGGTACTGCAACGTACTATGGTCTGATTATGCAAGTCTATACTTGGGTTGCAAAAGTT
AamoA-46	ATCGTATTACGGTCTGATTATGCAGGTCTATACATGGGTTGCAAAAGTTGCATGGTTTGCACTTGGCTAT
AamoA-47	CCGATAGAACCGCAGGTAGGTAAGTTCTACAATAGCCCAGTAGCTCTGGGAGCGGGAGCTGGATCTGTAC
AamoA-49	GAAAATTCTATAATAGCCCAGTCGCGCTGGGTGCGGGTGCTGGTGCAGTATTAGCCGTTACGATGGCAGC
AamoA-50	TCGACCTTGTTGACAATCAACGCAGGAGACTACATCTTCTATACTGACTGGATGTGGACATCTTATGTGA
AamoA-51	TGTCAATGTCGTTGTTCAACATGATCAATCTAATCACAATAGATGATCCTCTTGAGACTGCTTTCAAATA
AamoA-52	GCTTGCTGTCGGTGCCGCGTACTACCTTACATTCACTGGAGTTCCAGGAACTGCAACTTACTATGCACTG
AamoA-53	TATTCTCAGTATCACAGACACTCATGCTTGCTGTCGGTGCCGCGTACTACCTTACATTCACTGGAGTTCC
AamoA-56	GTTGTGGGAGCAATATACTACATGCTCTTTACTGGAGTCCCTGGAACGGCGACATATTATGCTACCATCA
AamoA-57	TTTCGCAATCTACGATGTTAGTTGTGGGAGCAATATACTACATGCTCTTTACTGGAGTCCCTGGAACGGC
AamoA-58	TTACATATTCTACACGGACTGGGCCTGGACATCATTTGTAGTATTTTCCATCTCACAATCTACTATGCTT

Table A2.3 continued.

Probe	Sequence
AamoA-59	GGCCTGGACATCATTTGTAGTATTTTCCATCTCACAATCTACTATGCTTGTTGTTGGGGCGATTTACTAC
AamoA-60	CCCTCGGCAATGTTGTTGGATCTTACATATTGGGCTACAAGGCGCAATAAACACGCCGCCATTATCATTG
AamoA-64	CCTCTCGAGGTCGCGTTCAAGTATCCAAGACCTACACTTCCTCCCTATATGACACCTATAGAGCCTCAGG
AamoA-61	GTAGGGGCTATATACTATATGCTCTTTACAGGTGTTCCAGGAACCGCTACATACTATGCAACCATTATGA
AamoA-62	AATCTACACATGGGTTGCAAAAGGTGCATGGTTTGCATTAGGATATCCATATGACTTCGTAACAACACCA
AamoA-63	GTTATTAGATCTTACATACTGGGCTACAAGACGTAACAAGCATGCTGCAATAATTATTGGTGGAACTTTG
AamoA-65	ATACCTTCAGCCATGTTGTTGGACCTTACGTATTGGGCTACGCGGCGCAATAAGCACATGCTAATACTTG
AamoA-66	GGGCTTGGACATCCTTTGTAATATTTTCGATTTCACAATCGACTATGCTTGCAGTAGGAGCTGTCTATTA
AamoA-67	AGGTACAGCGACATATTACGCAACCATTATGACAATTTATACATGGGTTGCAAAAGGAGCTTGGTTTGCA
AamoA-68	AGGTCGGTAAGTTTTACAACAGTCCTGTTGCCTTAGGAGCAGGTGCAGGGGCCGTGTTGACTGTGCCTAT
AamoA-69	TTCGTCGTGTTCTCTATATCGCAGTCAACGATGCTTGTTGTGGGTGCCATCTATTACATGTTGTTTACGG
AamoA-70	AACGATGCTTGTTGTGGGTGCCATCTATTACATGTTGTTTACGGGAGTGCCAGGGACAGCGACCTACTAT
AamoA-71	ACTATAATGACGATTTATACATGGGTTGCAAAAGGGGCTTGGTTTGCTCTTGGATATCCATACGACTTCA
AamoA-72	AAAGGGCGCCTGGTTTGCATTAGGATACCCATACGACTTCGTTACAGTACCGGTTTGGATACCTTCAGCA
AamoA-73	ACGATATATACGTGGGTTGCAAAGGGCGCCTGGTTTGCATTAGGATACCCATACGACTTCGTTACAGTAC
AamoA-74	AATATACTATATGCTATTTACAGGAGTGCCGGGTACGGCTACTTATTATGCAACCATTATGACGATATAT
AamoA-75	TACACTGATTGGGCATGGACTTCATTCGTCGTATTCTCTATCTCTCAATCGACAATGCTTGTGGTGGGGG
AamoA-76	GCACCTTACTTACTATCAATGCAGGAGACTACATTTTCTACACTGACTGGGCTTGGACATCATTCGTAGT
AamoA-77	CACCTTACTTACTATCAATGCGGGAGACTACATTTTCTACACTGACTGGGCGTGGACATCCTTTGTAGTA
AamoA-78	TAGTCGTTGCTGTTAACAGTACATTATTGACTATCAATGCAGGTGACTATATCTTCTATACTGATTGGGC
AamoA-79	ACAACGCACTATCTATTCATAGTAGTCGTTGCTGTTAACAGTACATTATTGACTATCAACGCTGGTGACT
AamoA-80	ACCAGTTTGGATACCTTCGGCTATGTTATTGGACCTTACTTACTGGGCAACAAGACGAAACAAACATGCT
AamoA-81	ATTTTCGATTTCTCAATCCACAATGCTTGTGGTAGGAGCTATATACTATATGTTGTTTACAGGTGTCCCT
AamoA-82	AGTAGGAGCTATATACTATATGTTATTTACGGGTGTTCCTGGAACGGCTACGTATTATGCAACTATTATG
AamoA-83	ACAATGCTTGTAGTAGGAGCTATATACTATATGTTATTTACGGGTGTTCCTGGAACGGCTACGTATTATG
AamoA-84	AATTATCATTGGAGGCACTTTGGTAGGATTATCATTACCAATTTTCAATATGATAAACCTCTTGCTTGTA
AamoA-90	CTACACAGACTGGATGTGGACATCATTTGTCGTATTTTCGATCTCTCAATCTACAATGCTCGCAGTGGGA
AamoA-87	GCTTAGACGTACAACGCACTACCTGTTCATAGTAGTGGTTGCTGTCAACAGCACACTGCTAACCATCAAC
AamoA-88	AAATCTATTGCTGGTACAGGATCCGCTTGCAATGGCGTTCAAGTATCCTAGACCGACACTGCCTCCGTAT
AamoA-89	ACCACAGGTGGGGAAGTTCTATAACAGCCCCGTGGCCTTGGGAGCAGGCATTGGCGCCGTGATATCGGTT
AamoA-85	ATAGTAGTCGTTGCTGTCAACAGCACGCTGCTAACCATCAACGCAGGAGACTACATCTTCTACACAGACT

Table A2.3 continued.

Probe	Sequence
AamoA-86	GGCACGGCCACATACTACGCCACAATCATGACCATCTACACATGGGTTGCAAAGGGTGCATGGTTCGCAC
AamoA-94	TACAGATTGGGCATGGACATCATTTGTTGTATTCTCAGTATCCCAATCAACAATGCTTGTAGTGGGAGCA
AamoA-97	ACATGGGTAGCCAAAGGAGCTTGGTTTGCGCTAGGATATCCAATGGACTTCATCACCGTACCTGTTTGGA
AamoA-99	TGGCGCTAGGATCGGGAGCTGGAGCTGTGCTAAGTGTTCCGATAGCTGCACTGGGTGCGAAACTCAATAC
AamoA-95	GTTCTACAATAGTCCCGTGGCGCTAGGATCGGGAGCTGGAGCTGTGCTAAGTGTTCCGATAGCTGCACTG
AamoA-96	TGCTTGTAGTTGGTGCAATCTATTACATGCTATTCACCGGAGTACCAGGGACTGCAACATATTATGCAAC
AamoA-91	GTGTACCAGGGACTGCAACATATTATGCAACAATCATGACTATCTATACATGGGTAGCCAAAGGAGCTTG
AamoA-92	TCTATTACATGCTATTCACCGGTGTACCAGGGACTGCAACATATTATGCAACAATCATGACTATCTATAC
AamoA-93	TATTGGGCAACAAGAAGGAATAAACATGCTGCCATTATAATTGGCGGAACATTGGTTGGACTTTCATTGC
AamoA-98	AATCATGACTATCTATACATGGGTAGCCAAAGGAGCTTGGTTTGCGCTAGGATATCCAATGGACTTCATC
AamoA-54	ATGGTCTGGCTTAGACGTACAACCCACTATCTGTTCATTGTTGTAGTTGCTGTCAATAGCACGCTGCTGA
AamoA-55	ATGGTCTGGCTTAGACGYACNACRCACTANCTRTTCATNGTNGTAGTTGCTGTNAAYAGYACNYTNYTNA
hzsA-001	CGGGGACAGGAGATTGGTGTATGTAGAGTCGCCGTATATGAACTGGGGAGTAGGTCAGTTGGCATCGGTA
hzsA-002	CGGTAAGCTGGGATGCCCCGTATAACAAGACCTATGAGAGGTTGACGAAGGATGATGGTGGATTGTACCG
hzsA-003	AACTGGGGAGTAGGTCAGTTGGCATCGGTAAGCTGGGATGCCCCGTATAACAAGACCTATGAGAGGTTGA
hzsA-035	AACCCCGGATGGGAATATCCTGTTCTCCAGCACCCAGGCCAATGGGAGCAGGGCAGGTGGCAAGGGCAGG
hzsA-036	AAAGGTGGAAGGATATCCACATTCCTGGGGCACATGGATTTGTTTCGATACGACGCTGACCGATCTTCCG
hzsA-010	AGGCCGATCTACGGGAACTGCGATGATGAGATTGGTGGTACCAGTGGCAAGTCGCAGGCCAAGATCACGT
hzsA-011	CTTATCCGAGGCCGATCTACGGGAACTGCGATGATGAGATTGGTGGTACCAGTGGCAAGTCGCAGGCCAA
hzsA-039	ACAACTGGGATGGTGCTTATCCGAGGCCTATCTATGGGAACTGCGATGAGGAGATTGGTGGGGCAAATGG
hzsA-040	ACCGGTCACGTGCTGTCTGGTATAGAGTTTCCTAGTATAAATACATCAATAGACCAGATTACCTACAACA
hzsA-032	AAACTGCGATGATGAGATTGGCGGTGCATGTGGCAAGTCACAGGCCAAGATCACTTTTGGGGAAAGAAGG
hzsA-033	AATGATCCTGAGTGGAACGATCATCAACCAGCCCCGGTCTATATCAAGTACAGGCCGAGGTGGATCAATA
hzsA-012	AACAAGACCTATGAGAGGTTGACCAAGGACGAGGGTGGGTTATACAGGAGTCCCTATCCGCTTCCCGATG
hzsA-013	ACATGAACTGGGGAGTAGGTCAGTTGGCGGCGGTAAGTTGGGATGCCCCCTATAACAAGACCTATGAGAG
hzsA-014	AGCATAAATACATCAATAGACCAGATTACGTACAACATCAGTTCGAACTTTGATCCTGCGCTTACGCCGG
hzsA-015	ACGATAGGATGTTAGTATCGTATGCGGAGAGGGGAGATTTTGGTATTTACTGGTTTGATTGCAAGAACGG
hzsA-016	ACTTCCGGACGATAGGATGTTAGTATCGTATGCGGAGAGGGGAGATTTTGGTATTTACTGGTTTGATTGC
hzsA-037	GTGCGTATCCAAGACCGATCTATGGAAACTGTGATGACGAAATTGGCGGCGCATCATGCAAGTCACAGGC
hzsA-038	GGATAACTGGGATGGTGCGTATCCAAGACCGATCTATGGAAACTGTGATGACGAAATTGGCGGCGCATCA

Table A2.3 continued.

Probe	Sequence
hzsA-041	AAATTGAGCAAGGGTGAAGGTGGTTTATACAGGAGTCCATATCCACTTCCAGACGATAGGATGTTGGTAT
hzsA-042	AAATTGGCGGTGCATCTGGCAAGTCACAGGCCAAGGTCACCTTTGGTGACAGGAAGTTAGTATACATAGA
hzsA-004	CAATGTGAGTTCGAACTTTGATCCTGCGCTTACGCCGGATGGAAACATCCTGTTTTCAAGCGTGCAGGCC
hzsA-005	AGGTGAAGGTGGAGGGTTATCCACACTCATGGGGTACGTGGATTTGTTTCGATACGACATTAACGGATCT
hzsA-006	GCGCTTACGCCGGATGGAAACATCCTGTTTTCAAGCGTGCAGGCCAATGGGAGCAGGGCAGGTGGCAAAG
hzsA-007	AAACATCCTGTTTTCCAGCGTCCAGGCCAATGGTAGCAGGGCAGAAGGCAAAGGCAGGGTGTTGCTGGAA
hzsA-008	AAAGCTGGTATATGTAGAGTCACCCTATATGAACTGGGGTGTTGGTCAATTGGCAGCAGTCAGTTGGGAT
hzsA-009	AACAAGACCTACGAGAGGCTGACGAAGGATGATGGTGGATTGTACAGGAGCCCATACCCACTTCCAGATG
hzsA-034	ATAGAGTTTCCGAACTTAAATACGTCAGTTGATCAGATTACGTATAATGTGAGTTCGAACTTTGATCCAG
hzsA-017	AGGTCCGTATCCAAGCCAGAGGGCAAAGGCAACGAAGCCAGGGGATGTAAAGGCAGTGAGAATCGTAGAA
hzsA-018	CTCCTTACTACATTCAGAATTTGGATGAGAGGGGTATGGCAGTGCAGACGGCTCTGATGTGGGCATATTT
hzsA-019	ACATTCAGAATTTGGATGAGAGGGGTATGGCAGTGCAGACGGCTCTGATGTGGGCATATTTGAGGCCATA
hzsA-020	AGGATGTTAGTGTCCTATGCAGCAAGAGGTGACTTTGGTATTTATTGGTTTGACTGTAAGAATGGGAGAG
hzsA-021	GTCCTATGCAGCAAGAGGTGACTTTGGTATTTATTGGTTTGACTGTAAGAATGGGAGAGCCGGTGAGGTA
hzsA-023	AATCTGTTCAGATTAGATACGCAGGGTGGTCATGTAACCGAAAGAATCACCGGCCACGTTCTCTCCGGTA
hzsA-022	AATCTGTTCAGGTTGGATACGCAGGGCGGTCATGTAACCGAAAGAATTACGGGACACGTTCTTTCCGGTA
hzsA-024	AGGGCGGTCATGTAACCGAAAGAATTACGGGACACGTTCTTTCCGGTATAGAGTTTCCGAGTTTAAACAC
hzsA-025	AAACTGTGATGGGGAAATCGGTGGCACCAGCGGAAGGTCTCAGGCGAAGATCACCTTTGGAGACAGGAAG
hzsA-026	AAAGAGGGGATTTTGGCATATACTGGTTTAATTTTAGCAAATGTGCTGCAGGGGACAAGGTATATGATGA
hzsA-027	AAGATCATACAGAGAATCACCGGACACGTGCTATCTGGCATAGAGTTTCCGCACCTCAATACCACGATAG
hzsA-030	AATGGCAGCACATGCCTGTTAGTTAACAACTGGACTGGCGCTTACCCAAGACATATCTACGGTAACGAGG
hzsA-031	ATGAATTTGGAAGACCATTCGCTTTTGACTTATACAGGCTTGATCCGCAGGGTGGAAAGTCAATGGATCG
hzsA-028	CAGTTCTGGCAGATGAGCCGGATTCTAAGAGATATCTTCAGGGTGCTGGTGCACACCTTCTTGGTGGCGC
hzsA-029	TGGTGTCTTCTGCTGAAAGACAGGATTTTGGTATTTATTACTTCTGCGCTGATAAGGGCACTGTTTCAGA
nifH-003	GGAGACGTTGTTTGTGGTGGTTTCGCTATGCCTATCCGTGAAGGAAAAGCACAAGAAATCTACATCGTTA
nifH-008b	GTATTAGGTGACGTTGTATGTGGTGGTTTTGCTATGCCTATCCGTGAAGGTAAAGCACAAGAAATCTACA
nifH-007	CATCAACTTCCTGGAAGAAGAAGGCGCTTACGAAGATCTAGATTTCGTTTCCTACGACGTATTGGGTGAC
nifH-005	GCTGTGCCGGTCGGGGCATCATCACCGCCATCAACTTCCTGGAAGAAGAAGGCGCTTACGAAGACCTCGA
nifH-004	ACCGCCATCAACTTCCTGGAAGAAGAAGGCGCTTACGAAGACCTCGATTTCGTTTCCTACGACGTATTGG
nifH-006	ATCATCACCGCCATCAACTTCCTAGAAGAAGAAGGCGCTTACGAAGACCTCGATTTCGTCTCCTACGACG

Table A2.3 continued.

Probe	Sequence
nifH-560	GCCGGGTGTAGGCTGTGCCGGTCGTGGCATCATCACCGCCATCAACTTCCTTGAAGAAGA
nifH-552	CTGGCTATAACAACGTGCGCTGCGTTGAGTCTGGCGGTCCCGTACCTGGTGTTGGCTGTG
nifH-537	ACTTCACTTAGCTGCTGAACGCGGTGCAGTGGAAGACTTAGAACTTGATGAAGTAGTACT
nifH-551	GTAGGTTGTGCAGGTCGTGGTATTATCACAGCTATCAACTTCCTCGGAGAAGAAGGTGCT
nifH-547	TGAGTCCGGCGGTCCAGAACCAGGAGTAGGCTGCGCCGGTCGCGGTATCATCACCGCCAT
nifH-548	GTCCGGCGGTCCAGAACCAGGAGTAGGCTGCGCCGGTCGCGGTATCATCACCGCCATCAA
nifH-579	GTTTTACATCTAGCAACAGAAAGAGGAACTGTAGAGGATATCGAACTTGATGAAGTAGTA
nifH-580	GCTATCAACTTGATTGAAGGAGAAGGTGCTTACGAAAATCTAGATTTCGTATCTTATGAC
nifH-563	TACCCGTTTGATGCTACACACTAAAGCACAAACTACCATTCTTCACTTAGCAGCAGAACG
nifH-564	TTTGATGCTACACACTAAAGCACAAACTACCATTCTTCACTTAGCAGCAGAACGGGGAAC
nifH-008	CCTATCCGTGAAGGTAAAGCACAAGAAATCTACATCGTTACCTCTGGTGAAATGATGGCGATGTACGCTG
nifH-535	AACTTTATACACTCACTGTCTAACAAGCGAGAATCAACTATGCGTCAGATTGCATTTTAC
nifH-561	TCCCGCTTAATGCTTCACTGTAAAGCGCAAACCACCATTCTGCACTTAGCTGCTGAAAGA
nifH-542	CTGCACAGCAAAGCCCAGACCACGACTTTGCACCTGGCCGCCGAGCGTGGCGCAGTCGAA
nifH-570	CCTCCGTGCTGCAACTAGCCGCCGAACTCGGTGCCGTTGAAGATGTCGAACTTGAGCAAG
nifH-055	AAAATGGCGCGTACGAAGACCTGGATTTCGTCTCCTACGACGTATTAGGTGACGTTGTGTGCGGTGGTTT
nifH-056	AAATCGACGAAGTATTGCTCACAGGCTTCAACGGTGTCAAGCGCGTTGAATCTGGCGGTCCAGAACCAGG
nifH-549	TACCGTATTGCACTTGGCTGCTGAGAAAGGTGCTGTAGAAGACCTGGAATTGGAAGAAGT
nifH-516	TTCGTATTCTACGATGTATTGGGTGACGTGGTATGTGGTGGATTCGCCATGCCCATTCGT
nifH-573	GTATTAGGTGACGTTGTGTGCGGTGGTTTCGCAATGCCTATCCGTGAAGGCAAGGCACAA
nifH-574	TTAGGTGACGTTGTGTGCGGTGGTTTCGCAATGCCTATCCGTGAAGGCAAGGCACAAGAA
nifH-569	GTGTGCTGGTCGTGGTGTTATCACTGCCATTAATTTCCTCGAAGAGGAAGGCGCGTATGA
nifH-556	TGCACTCCAAAGCACAAACCACAATTCTCAGCTTAGCTGCAGAACGGGGTGCAGTAGAAG
nifH-557	ACCTAGCTGCCGAGCGGGGCGCTGTCGAAGACTTGGAACTCGATGAAGTGCTGCTGACCG
nifH-559	TCTACTCGCCTGATGCTGCACAGCAAGGCTCAGACTACGATTCTGCACTTGGCCGCCGAG
nifH-565	CCTGAGCCTGGTGTTGGTTGCGCAGGTCGCGGCATCATCACAGCCATCAACTTCCTCGAA
nifH-541	CTGATGCTGCACAGCAAAGCCCAGACCACGACTTTGCACCTGGCCGCCGAGCGTGGCGCA
nifH-578	TGCACAGCGAGGCTCAAACCACCATCCTGCACCTGGCTGCCGAGCGCGGTGCAGTAGAAG
nifH-558	TACGATTCTTCACTTAGCTGCTGAGCGTGGCGCGGTTGAAGATCTAGAGCTAGAGGAAGT
nifH-555	TGGCAGCCGAGCGCGGCGCAGTCGAAGATCTGGAACTCGAAGAAGTACTACTCGAAGGCT
nifH-571	GCAGTTGAAGATGTGGAACTTGAAGAAGTTCTACTCACTGGCTACAAAGGCGTTAAGTGC

Table A2.3 continued.

Probe	Sequence
nifH-567	CAGTCGAAGATCTAGAGCTCGACGAAGTACTACTCACTGGCTACAAAGATGTGAAGTGCG
nifH-582	CTACACACTAAAGCTCAGACTACTATTCCATCCCTTGCAGCAGAAAGAGGTGCAGTAGAA
nifH-553	GACTACAATTCTACACTTAGCCGCAGAACAGGGTTCAGTGGAAGACATCGAACTGGAAGA
nifH-554	TCGAACTTGAAGAAGTATTACTCACTGGCTACGACAATGTACGTTGCGTTGAGTCTGGCG
nifH-522	AGAACCAGGTGTTGGTTGTGCTGGTCGCGGTGTAATCACTGCTATCAACTTCCTTGAGGA
nifH-572	GCTCAAACAACCGTTCTTCACCTCGCCGCAGAACGTGGTGCAGTAGAAGATATCGAACTC
nifH-550	CTCACTCGTTTAATGCTTCACTCTAAGGCTCAAACCACTGTGCTGCACTTAGCAGCAGAA
nifH-575	AAAGCTCAAACCACCGTGCTGCACTTAGCAGCAGAAAGAGGTACAGTAGAAGACTTAGAA
nifH-576	GCACTTAGCAGCAGAGAGAGGCGCAGTAGAAGACTTAGAGCTCGAAGAGGTAATGTTACC
nifH-543	CGTTGTTTGCGGTGGCTTTGCAATGCCAATTCGAGAAGGCAAAGCACAAGAGATCTACAT
nifH-009	AGGTGACGTTGTTTGCGGTGGATTTGCAATGCCTATCCGTGAAAACAAAGCTCAAGAAGTCTACATCGTA
nifH-010	GTATTATTACTGCTATCAACTTCCTAGAAGAAGAAGGAGCTTATACAGATCTAGATTTCGTAAGCTATGA
nifH-011	AAGGAGCTTATGCAGATCTAGATTTCGTAAGCTATGACGTACTAGGTGACGTTGTTTGCGGTGGATTTGC
nifH-538	CCGTAAAGTTGATCGTGAAACTGAGTTAATTGAGAACTTAGCAGCTCGTCTCAACACTCA
nifH-012	ACCCGGTTGATTCTGCACTGTAAAGCCCAAACCACCGTGTTGCACTTGGCTGCGGAACGGGGTGCTGTAG
nifH-041	CGGCGACATCAAGTGCGTTGAGTCCGGTGGTCCTGAGCCCGGTGTCGGCTGCGCCGGTCGCGGCGTTATC
nifH-545	TCTTATGACGTACTTGGTGACGTTGTATGTGGTGGTTTTGCTATGCCTATCCGCGAAGGA
nifH-546	GAGATTGACTTAATCGAAGAGTTAGCTGAACGCTTGAACACCCAAATGATCCACTTTGTA
nifH-014	TTGTCCAGAAGGGCTTCGGTGACATTCTGAACGTGGAATGTGGCGGTCCAGAGCCTGGTCCAGAGCCTGG
nifH-021	AGCTGGAGGATGTTGTCCAGAAGGGCTTCGGTGACATTCTGAACGTGGAATGTGGCGGAGCTGGAGGATG
nifH-577	TTCTGGTGAAATGATGGCAATGTACGCTGCTAACAACATTGCAAGAGGTATTCTTAAATA
nifH-566	GCAGTTGAAGACCTCGAACTCGGAGAAGTGCTATTAGATGGCTACAACGGTGTCCGTTGC
nifH-013	AACTTCCTGGAAGAGCAAGGCGCTTATGACGGCATGGACTTCATCTCCTACGACGTACTGGGTGACGTTG
nifH-016	ATTATCACCTCCATCAACTTCCTGGAGGAGCAAGGCGCTTATGAGGATATGGACTTCGTTTCCTATGACG
nifH-015	AAGAAGAAGGCGCTTATGAAGGGAAAGATTTTATTTCTTATGACGTATTAGGAGACGTTGTTTGTGGTGG
nifH-539	ACTAAGCTTGGCTGCTGAGCGCGGTGCAGTAGAAGATTTAGAACTTGAAGAAGTAATGCT
nifH-562	CTTCACTTGGCTGCTGAAAGAGGATCTGTTGAAGATTTAGAACTTCATGAAGTAATGCTC
nifH-536	AGATTTAGAACTCGAAGAAGTAATGCTCGCAGGCTTCCGTGGTGTGAAGTGCGTTGAGTC
nifH-043	GGCGACATCAGGTGCGTTGAGTCCGGTGGTCCTGAGCCTGGTGTCGGCTGCGCCGGTCGCGGCGTTATCA
nifH-042	GGCGACATCAGGTGCGTTGAGTCCGGTGGTCCTGAGCCCGGTGTCGGCTGCGCCGGTCGCGGCGTCATCA
nifH-050	ATCAACTTCCTGGAAGAGAACGGCGCCTATGACGACGTCGACTACGTCTCCTACGACGTGCTGGGCGACG

Table A2.3 continued

Probe	Sequence
nifH-051	GCCGGCCGCGGCGTCATCACCGCCATCAACTTCCTGGAAGAGAACGGCGCCTACGACGACGTGGACTACG
nifH-052	CTGATCCTGAACTCCAAGGCGCAGGACACCGTGCTGCACCTGGCCGCCGACATGGGCTCGGTCGAGGATC
nifH-053	ATCAACTTCCTGGAAGAGAACGGCGCCTATGACGATGTGGACTATGTGTCCTATGACGTTCTGGGCGACG
nifH-054	GGGGTTGGCTGCGCTGGTCGCGGTGTGATCACCTCGATCAACTTCCTGGAAGAGAACGGCGCCTATGACG
nifH-046	TACGTGTCCTACGACGTGCTGGGCGACGTGGTCTGCGGCGGCTTCGCGATGCCGATCCGCGAGAACAAGG
nifH-504	CTGGCCGCCGACATGGGCTCGGTCGAGGATCTGGAGCTGGAGGACGTCCTCAAGGTCGGC
nifH-047	TACGTCTCCTACGACGTGCTCGGCGACGTCGTCTGCGGCGGCTTCGCCATGCCGATCCGCGAGAACAAGG
nifH-048	TACGTCTCCTACGACGTGCTCGGCGACGTCGTCTGCGGCGGCTTCGCGATGCCGATCCGCGAGAACAAGG
nifH-049	ACGACGTCGACTACGTCTCCTACGACGTGCTGGGCGACGTGGTCTGCGGCGGCTTCGCCATGCCGATCCG
nifH-032	AAGTGCGTTGAATCAGGCGGTCCTGAGCCCGGTGTCGGCTGTGCCGGCCGCGGTGTTATTACCGCTATCA
nifH-062	CGGCTGCGCCGGTCGCGGTGTTATCACCGCCATCAACTTCCTGGAAGAGGAAGGCGCTTACGACGAAGAT
nifH-531	GCGTCGAGGACCTCGAACTCGAAGACGTGATGAAGATTGGCTACAAGGACATCCGTTGCG
nifH-532	CGAGGACCTCGAACTCGAAGACGTGATGAAGATTGGCTACAAGGACATCCGTTGCGTGGA
nifH-534	CTCGACCTCGCCGAGGCCCTGGCCAAGCGCCTGAACTCGCAGCTGATCCACTTCGTGCCG
nifH-517	GATATCGACTACGCGTCCTACGACGTTCTCGGCGACGTCGTCTGCGGCGGCTTCGCGATG
nifH-530	CAAGGTAGCGTCGAGGACCTCGAACTCGAAGACGTGATGAAGATTGGCTACAAGGACATC
nifH-060	AAGACCTGGACTTCGTGTTCTACGATGTGCTGGGTGACGTTGTCTGCGGTGGTTTTGCCATGCCGATTCG
nifH-533	CGCGACAACATCGTGCAGCACGCGGAACTCCGCCGCATGACCGTGATCGAGTATGCCCCG
nifH-501	TCAGGACACCATCCTGTCGCTGGCCGCTGAAGCCGGTTCGGCGGAAGACCTCGAGATCGA
nifH-037	ATGCCGATCCGCGAGAACAAGGCCGAGGAGATCTACATCGTGGTCTCCGGCGAAATGATGGCGATAAGGG
nifH-057	AGCCAGGGGTTGGTTGTGCCGGTCGTGGTGTTATCACGGCGATTAACTTCCTTGAGGAAGAAGGCGCTTA
nifH-515	TGGTGTTATCACCGCTATCAACTTCCTCGAAGAGGAAGGCGCCTACGAAGACGATCTCGA
nifH-521	GGTGTAATCACTGCAATCAACTTCCTGGAAGAGGAAGGTGCATATGAAGATGACCTGGAC
nifH-044	AAGGCGCAGGACACCATCATGCAGATGGCGGCTGACGCGGGCTCGGTTGAGGATCTGGAGCTTGAGGATG
nifH-045	CGACCCGTCTGATTCTGCACTCCAAGGCGCAGGACACCATCATGCAGATGGCGGCTGACGCGGGTTCGGT
nifH-039	AACTGGAAGATGTATTAAAAGTCGGTTACGGCGGTGTGCGCTGTGTTGAGTCGGGCGGTCCTGAGCCAGG
nifH-040	AAGACTTAGAACTGGAAGATGTATTAAAAGTCGGTTACGGCGGTGTGCGCTGTGTTGAGTCGGGCGGTCC
nifH-526	GCACAGGCTGGTACGGTAGAGGATCTTGAGCTTGAAGAAGTGCTTAAGGTTGGCTTTGGC
nifH-527	TGAAGAAGTGCTTAAGGTTGGCTTTGGCGATATTAAGTGCGTGGAGTCAGGTGGTCCTGA
nifH-001	AGAAGTGCTTAAGGTTGGCTTTGGCGACATTAAGTGCGTGGAGTCAGGTGGTCCTGAGCCTGGAGTTGGC
nifH-002	AGAGGATCTTGAGCTTGAAGAAGTGCTTAAGGTTGGCTTTGGCGACATTAAGTGCGTGGAGTCAGGTGGT

Table A2.3 continued.

Probe	Sequence
nifH-505	CATGGAAATGGCTGCCCAGGCAGGTACGGTTGAGGATTTAGAGCTTGATGACGTACTTAA
nifH-525	ATCTCGAGCTCGATGACGTGCTTAAGGTTGGTTATGGTAACATTAAGTGTGTAGAAGCAG
nifH-513	TTGATCCTTCACTCCAAAGCTCAGAACACCATTATGGAAATGGCTGCTGAAGCTGGCACC
nifH-514	CAAAGCTCAGAACACCATTATGGAAATGGCTGCTGAAGCTGGCACCGTTGAAGATCTGGA
nifH-520	GAAATGGCTGCTGAAGCCGGCACCGTGGAAGATCTGGAACTGGAAGATGTATTAAAAGCC
nifH-519	TGCGATCCGAAAGCGGACTCCACTCGTCTGATTCTTCACTCCAAAGCACAGAACACCATC
nifH-523	TGCTATCAACTTCCTTGAGGAAGAAGGTGCGTACGAAGACGATCTTGACTTCGTATTCTA
nifH-038	ACACTAAGATGCAAAACACCATCATGGAAATGGCAGCGGAAGCCGGTACTGTTGAAGACATCGAATTAGA
nifH-529	TGGTTGCGCTGGTCGCGGTGTAATAACAGCGATTAACTTCCTTGAAGAGGAAGGTGCATA
nifH-020	TACGACGACGACCTCGACTTCGTGTTCTACGACGTGCTGGGCGACGTGGTCTGCGGCGGTTTCGCCATGC
nifH-019	CGTACGACGACGAACTCGACTTCGTGTTCTACGACGTGCTGGGCGACGTGGTGTGCGGCGGTTTCGCGAT
nifH-568	GTTGCGCTGAATCAGGTGGTCCCGAGCCCGGTGTTGGGTGTGCTGGTCGTGGTGTTATCA
nifH-058	AAAACACAATCATGCAGATGGCACCTGATGCCGGTTCTGTAGAACATCTGGAACTCGAAGATGTACTGAA
nifH-059	AAAGCAGATGCTACACGTCTAATTCTTCATTCAAAGGCTCAGGAGACTATTATGCATCTGGCCGCCGATG
nifH-507	ATGGCGGCCGAGAAGGGCTCCGTTGAAGATCTGGAACTCGAAGATGTGCTGCAAATCGGT
nifH-061	AAAAGCACAAAACACCATTATGGAAATGGCGGCCGAAGCGGGCACCGTTGAAGATCTGGAATTGGAAGAT
nifH-508	ACCGTGGAGGATCTGGAGCTCGAAGACGTGCTCAAGACCGGCTTCGGCGACATCAAGTGC
nifH-017	AAACTACTGTTATGCATCTGGCTGCTGAGGCCGGTACCGTAGAAGACCTGGAGCTGGAAGATGTATTGTC
nifH-018	ACCGTAGAAGACCTGGAGCTGGAAGATGTATTGTCTGTCGGTTACGGCGACGTTAAATGTGTTGAGTCTG
nifH-506	CGCAGACCACGGTGATGCACCTGGCCGCTGAAGCCGGCTCGGTGGAAGATCTCGAACTCG
nifH-036	ACGTCGTCTGTGGCGGATTCGCTATGCCGATCCGTGAGAACAAGGCGGAAGAGATTTACATCGTTGTCTC
nifH-031	ACCCGTCTGATCCTTCATGCCAAAGCGCAGGAAACGGTGATGGACAAGGTCCGGGAACTCGGCACCGTCG
nifH-035	AGGGTTGGCTACGGCGGCATCAAGTGCGTTGAATCAGGCGGTCCTGAGCCCGGTGTCGGCTGTGCCGGCC
nifH-064	ACCCGCCTGATCCTTCATGCCAAAGCCCAGGACACGGTTATGGACAAGGTCCGTGAACTCGGCACCGTTG
nifH-065	ACCCGTCTGATCCTCCATGCCAAAGCCCAGGAAACGGTTATGGACAAGGTCCGGGAACTCGGCACCGTCG
nifH-518	TGATCCTGCACGCCAAGGCTCAGGATACCGTTATGGACAAGGTCCGTGAGCTCGGTACCG
nifH-034	AAGAAGAAGGCGCTTACACCCCCGACCTCGACTTTGTTTTTTATGATGTTCTTGGTGACGTCGTCTGTGG
nifH-033	AAACGGTCATGGACAAGGTCCGGGAGCTGGGCACCGTTGAGGATCTGGAACTGGATGATGTGCTCCGGAT
nifH-528	TTAGAGTTAGAAGATGTCATGAAGCGCGGTTATGGCGAAGTGATGTGCGTTGAATCTGGT
nifH-063	GCGGCGTCATCACCGCCATCAACTTCCTCGAGGAGAACGGCGCCTACACCCCCGATCTCGACTTCGTCTT
nifH-030	ATCACCGCCATCAACTTCCTGGAAGAGAACGGCGCCTACACCCCCGATCTCGACTTCGTCTTCTACGACG

Table A2.3 continued.

Probe	Sequence
nifH-029	CCGGACCTGGACTTCGTATTCTACGACGTTCTCGGCGACGTCGTCTGCGGCGGGTTTGCCATGCCGATCC
nifH-022	AAAAAGGCGCAGGCGACAGTCATGGACATGGCAAGGGAGATGGGAAGTGTCGAAGACCTTGAGATAGGGG
nifH-023	AAAACGGGGCCTATGATGCTGACACCGATTTCGTCTTTTACGATGTGCTCGGCGACGTCGTCTGCGGCGG
nifH-024	AGGCTGCGCAGGTCGCGGCGTTATCACTGCCATCAACTTTCTTGAAGAGAACGGCGCGTACGGCGATGAT
nifH-524	TGAGGATCTCGAGCTAGAAGATGTTCTAAAGGTTGGATACGGCGATATTAAGTGTGTAGA
nifH-502	TCCTGGTTTCGGCGGTTCATTGTGCACCGAATCTGGTGGCCCAGAGCCTGGCGTTGGCTG
nifH-503	ATCAACATGCTTGAGCAGCTCGGTGCCTACGAGACCCACGACCTCGACTACACCTTCTAC
nifH-509	TCATCTTGCACTCCAAGGCTCAAAACACCATCATGGTGATGGCTGCTCAGGCAGGCACGG
nifH-510	GGCGCGGTATCATCACCTCCATCGGTCTGCTGGAACGCCTTGGCGCTTACACCGAAGACC
nifH-511	CTGATCCTGGGCGGCAAGCCGCAGGAGTCGCTGATGGAAGTGCTGCGCGAGGAAGGCGCC
nifH-512	GTGAAGATGTTGATATTGAAGACATTGTAAAGGCAGGGTACAGTAATGTACGTTGCGTTG
nifH-540	TTGGCCGCCGAGCGTGGTGCTGTAGAAGATCTCGAACTCGATGAGGTCCTCCTCACTGGC
nifH-544	CATGTGCCGGCGGAACTGGGCGCGGTGGAAGATGTGGAACTGGATCAGGTGCTGAAACCG
nifH-581	CATCAATATGCTAGAGCAGCTCGGCGCCTACGAAGAGGACGAAGAACTCGACTACGTGTT
nrfA-01	AAAAACAAAGCGGTTAAATTCCCGTGGGATGACGGCATGAAAGTCGAAAATATGGAGCAGTATTACGACA
nrfA-02	ACTTCGACGGCAAAAACAAAGCGGTTAAATTCCCGTGGGATGACGGCATGAAAGTCGAAAATATGGAGCA
nrfA-03	GCTGGAGTTGTAAAAGCCCGGATGTGGCGCGTCTGATCCAGAAAGACGGCGAAGATGGCTACTTCCACGG
nrfA-04	AACAGTCTGCCCGTGAAGATGCGCTGGCGGAAGACCCTCGTCTGGTGATCCTGTGGGCGGGCTATCCCTT
nrfA-05	AACGATATTCGCCACGCGCAGTGGCGTTGGGATCTGGCGATTGCGTCACACGGTATTCATATGCATGCCC
nrfA-06	AAAGCCGCATGGGATGCGGGGGCGACAGACGCGGAAATGAAACCGATTCTGAACGATATTCGCCACGCGC
nrfA-94	AAACAGGATTTCCTGAACACGGTGGTGCCGCAGTGGGATGAGCAAGCGCGTAAAGCTGGACGACTGAACT
nrfA-95	AAACCTGGAGCATTGGTATTCACGGTAAAAACAACGTGACCTGTATCGACTGCCATATGCCGAAAGTGAA
nrfA-13	AAAAAATGCTGAAGACGGTCCAATGCCTATGGCATGTTGGAGCTGTAAAAGCCCTGACGTACCTCGCGTG
nrfA-10	AAAGAGTTCATGGTTAACCTAGATAACGAGCGTAAAGCTAAGGTTAAGCAGATCCAACTTCAAGCTGAAG
nrfA-11	AAGCTGGCGCAACTGAAGCTGAAATGGCACCAATCCTAATGGATATCCGTCATGGTCAATGGCGTTGGGA
nrfA-12	AAAAAATCCTTTGCACTGAGTGCATTGGTCGCAGCAAGCCTAATGGCTTCTGGTGCGATGGCAAGTGATA
nrfA-08	AAAAAAGAAGGCAGAAAAGGTTTCGTTAAATTCCCTTGGGATATGGGCACAACCGTTGAGCAAATGGAAA
nrfA-09	CTGTCACATGCCTAAAGTGACTAACGACAAAGGTCGTAAGTACACAGATCATAAAGTGGGTAACCCATTC
nrfA-14	AAAAAGGCACCCCTAAACTGCGCCTGTCTCGTCCTTTTGCTTCCCGCGCCATGGAAGCCATTGGTACCCC
nrfA-07	AAAACTGAACCTCGCAATGAGGTTTATAAAGATAAATTTAAAAATCAATACAATAGCTGGCATGATACCG

Table A2.3 continued.

Probe	Sequence
nrfA-120	CGACCTTATGTTGAGCGTGCATTCGAAGCGATTGGCAAGAAGTTTGATGAGCAGAGCCGCCTCGATCAAC
nrfA-121	AGATGTGGCTCGTGTTATCGAAGAGCGTGGCGAAGATGGTTACTTCGAAGGCAAATGGGCACGCCTAGGC
nrfA-23	AAAAAATGAACGCAGATAAGAAACACTTCCTAGAGACTGTCGTTCCGGACTGGGATAAAGCCGCGGCCGA
nrfA-122	AACGGTGAACCAGCGCTGGCAATAACTCGCCCTTATGTAGAACGCGCTTTCGATGCCATCGGCAAAAAGT
nrfA-123	AAAGATTGGACACACGCAGTATCGAAGGCACCAATGTTGAAAGCACAGCACCCAGGTTACGAAACATGGC
nrfA-15	AAAAAATGAACGCGGATAAAAAGCACTTCTTAGACACCGTTGTGCCAGATTGGGATAAAGCAGCGGCCGA
nrfA-18	AAACAAGCGTCTGTGTGTGCTCAGTGCCATGTCGAATACTACTTTACAGGACCAACGAAAGCGGTGAAAT
nrfA-24	AAAAAGAAATTGGCGACCCTCGTAACGACCAGTTCGAGCAAAACCACCCAGATCAATACCATTCATGGAG
nrfA-25	AAAAACAAATCGTTGCGGCTCACTTTGAAGCGGGTGCAGCGTGGAAAGCTGGCGCTACTGAAGAAGAAAT
nrfA-26	AAAAACAAAGTGGCGTGTGTCGACTGTCATATGCCAAAAGTGACCAAAGAAGATGGCACTGTCTACACCG
nrfA-16	AAAAAGAAATGGCACCAATCCTACAAGACATTCGCCACGCACAATGGCGTTGGGATTATGCGATTGCTTC
nrfA-17	AAAAAGCAATGGACCCGTCGTTCAGCCGCTGCAATAGCAATGGTGACTACTTTATTATTAAGTAGCCACA
nrfA-19	AAAGACTGGACTCACAAAGTATCTAAAGCGCCTATGTTAAAAGCGCAGCATCCTGGTTATGAAACATGGC
nrfA-20	AAGGATTACAACAAGCCGCGCGGCCACTTCTATGCGCTGACCGACGTGCGGGAAACCCTGCGCACCGGCG
nrfA-21	GCCAGAACAACGTCGCCTGTGTCGACTGTCACATGCCCAAGGTGCAAAACGAGCAGGGCAAGGTCTACAC
nrfA-22	ATCCGTCACGCCCAATGGCGCTGGGACTTCTCCATCGCTTCCCACGGGGTGCAGATGCACGCCCCGGAAG
nrfA-27	AAAAACTAGGAGTGGTAGTAGAATCAGCAAACCATAAATTTGCTGAAAAATATCCGCTTCAATATAACTC
nrfA-28	AAAAAGGTGAACGTGGTTATTTTGATCCTAAATGGGCGAAATACGGTGCGGAAATCGTAAACTCAATCGG
nrfA-29	AAACACGGCGTGACAACACCGGTAGCAATTCCTGATATCTCTACTAAAGAAAAAGCACAAAAAGCGATTG
nrfA-31	AAAAAGCACAAAAAGCGATTGGTTTAGATATTCCTAAAGAACAAGCTGCAAAAGATGAATTCTTACGTAC
nrfA-32	AAAAATACTACGATGAAATTGAATTTGCTGACTGGACTCACGCATTATCAAAAGCACCAATGTTAAAAAC
nrfA-30	AAAAAACAAGTAAAAGATGCAATGATCAAATTAGAAGATCAATTAGTTAAGGCGCACTTTGAAGCGAAAG
nrfA-34	AAAATGATGGCCCGCAACCAATGGCTTGTTGGACATGTAAAGGCCCAGATGTTCCTCGCTTAATCGAAGA
nrfA-33	AAAAGTAGAAACTAAAGATGGTAAAGTCTATACCGATCACAAAATCGGCAATCCGTTTGATAACTTTGAG
nrfA-36	AAAGCATGGAAAGTAATGGGTATTGATATCGAGAAAGAACGTAAAGCGAAAAAAGAATTCTTAGAAACTG
nrfA-38	AAAAACGGCGTGACCTGTATCGATTGCCACATGCCGAAAGTGCAAGGGGCGGATGGTAAAGTTTATACCG
nrfA-35	AAACTGTCACGATCAAAGCAAAGAAAAATTACGCGATATCGTCACTTCACGTAAAAAAGAAGTGAAAGAT
nrfA-37	AACTCAGCGGTAGCTGATATGGTTTATAAGCCATTAGAACAACCTGTTGAACCAGTAAAACCAGATTTGA
nrfA-39	TATGAAACCTGGATGCTCGGTGTGCATGGTAAAAATGGAGTGACTTGTATTGACTGCCATATGCCAAAAGTA CA
nrfA-40	AAAAGCCGAATTCCTCAAAACTGTTGTTCCTCAATGGGATAAAGAAGCCCGTGAAAAAGGCTTACTTCCT

Table A2.3 continued.

Probe	Sequence
nrfA-41	AAGAGGCTTCCGGGACAGAAGGGGAGTTGTCTACAGTGCAAGGGTTCGTATGTTTATGATGTGTACTTCA
nrfA-42	TTGAGTAACAGACAGAAAATCATCGCGGGTGTTGCCGGCGTCTGTGCCCTGTTCTTTGGATTCGTGGCCGTT CG
nrfA-43	AACGAGATGCGCGCCTACGTTTGCGGACAGTGCCATACGGAGTATTACTTCACGGCTGAGGACGGCCGCG
nrfA-44	AAAAAAACAGATCGGCTCCTGTCTGACCTGCAAATCGGCGGAAGTGCCCGGTATGATCGCCAAGATGGGC
nrfA-45	AAAAAAGAAATTTTTATGAATGGGTACATCCTTTAAGCGGCACTAAAATGCTGAAATCCCGCCATCCGGA
nrfA-126	AAAAACAACGAGACGTCCGTCATGACCGAGTACAAGGGTTCGGTGCCGTTCCACAAGAACGACGACGTGA
nrfA-128	AACAAGCTCTCCCGCAACGAAAAGCGCACCCTGGTCTGCGCCCAGTGCCATGTGGAATACTACTTCACCC
nrfA-129	AACAAGCTCTCGCGCAACGAAAAGCGCACCCTTGTCTGCGCCCAGTGCCATGTGGAATACTACTTCACCC
nrfA-127	AAAAACCTCTGGCTGGGCTATCCGTTCATGTATGAATATAATGAAACGCGCGGCCACACCTATGCCATTG
nrfA-124	AACGAATTCCGCGACAAGATCGACATGAAGGATCACACCATCGGGTGTGCCACCTGTCACGACCCCCAGA
nrfA-125	AACAAGAAGCCGGTGTTCCCGTGGGCCGAGGGTTTCGACCCTGCCGACATGTATCGCTACTACGACAAGC
nrfA-46	AAAAAGCAACATTTGATCAATTGTTGATAGCTCAGGAGCTTTCTGTAAAAGCCCATGAAGCAGTACGGTT
nrfA-49	AAAAGAAATTAAAATCATGGCAAGGATGGCTGTTGTTCGGAGGTTCAATGGTCGTTGTATTTGTTTTGGG
nrfA-52	TTGTATTTGTTTTGGGATTATGCGTTTCTGCGTTGATGGAACGGAGGGCGGAAGTTGCCAGCATTTTCAA
nrfA-53	AACGAGAGACATGCGGAGGTCACTAGCGTAATGAATAATAAGAAGACGGAGATCACAGGAATCGAGGCTC
nrfA-54	AAAGTGTTAGCGAAATTAGGTTATACGGATGATGTTCCGATGCCGGACTTCTCTACGAAAGAAAAGGCTC
nrfA-50	AAAAAGCACAGGAATACATCGGTCTGGATATGGAGAAGGAGCGTAAGGCGAAAGACAAATTCCTGAAAAC
nrfA-51	AAAAGATATAGACAAAGCGACTCCACAGGAGATGCGTTCGCTGGTATGTGCGCAATGTCATGTAGAGTAC
nrfA-47	AAAAACAAATGGAGCAGTCTGGGCAGTGACATAGTGAATCCGATAGGGTGTGCAGACTGCCATGATCCGG
nrfA-48	AAAAACGTTTACGAACGTCAGCGGAAAGCAAATGAAGTGCGCAATCAGTTGGAAAACGAACTGGCCAAAG
nrfA-56	AAAAACTATAATACGCCCCGAGGGCACCGTCATGCCATTGAAGACATGCGCAAGATTCTGCGCACGGGCA
nrfA-58	AAAAGAAAGTGTACGACTTCGCGGTCAAGGTGAACCATGAATTGGCTTTAGCCCACCTCGAAGCAGAGTT
nrfA-57	AAACAAAGCCAGTCATCAAGAAATGCGTTCGTTGGTTTGTGCACAGTGCCATACCGAGTATTACTTTGAA
nrfA-55	AAAGGAACGGATAAATTCTATGCTGCCAAGTGGAGCGACTGGGGACCAGAGGTCATGAACACCATCGGTT
nrfA-59	AAAGATGCACGTCTGCAGATAGCCCGTGTCGTAGCACGTCACGGATTTACAGGTCAGATTCCTCTGCCCG
nrfA-132	AAAAAGACAATGGGAACTATCTTCACTTCCCACAAGAGAAGGGACTTACATGCGAAGCAGCTGAAGAATA
nrfA-133	AAAAATACTGCGTACCGGAAGCCCGGGAGTAGACGGTCAAGCAGACATACAGCCCGGCACTTGCTGGACT
nrfA-131	AAAAACATGTTTAAAGAAGACATTGTTGCACAGAACGAGAAGTTTGCAGGCGACTTTCCACGCGAATACG
nrfA-130	AAAAACAACGGTTCGTCAACAACATAGTACCCAAGTGGATAGACGACGCCCGCAAGAACCACCGATTCGT
nrfA-134	AAAACAATGGACCTCAAGCCCGCTCGTCCTGCACTGTACGAGGCCTGGCAACGCGTAGGTAAAGACGTTA

Table A2.3 continued.

Probe	Sequence
nrfA-136	AAAACCATGAATCTTCGTATTACACGTCCGGCTCTGGTCGAAGCGTTTCAGCGTAGAGGCAAAGATATCA
nrfA-135	GTAAAAGTACGGACGTTCCGCGCGTGATGAACGAACGCGGCGTGGCTGAATTCTATAAGGACAAATGGGA
nfrA-96	AAAAGTCCCGATGTACCCAGGTTAATGAATGAGCTGGGAGTTGCAGAATTTTATGGCGGATCCTGGGAAT
nfrA-97	AAAGCGCCAATGCTGAAGGCACAGCATCCGGGCTATGAATTGTATATGACAGGGATACATGCCGAAAGGG
nfrA-98	GTGAACCCGATTGGCTGTGCCGATTGCCACGATTCGAAAACCATGAACCTGCACATTTCGCGCCCCGGAT
nfrA-99	CTGAAGTAGTGAACCCGATTGGTTGTGCCGATTGCCACGATTCGAAAACCATGAACCTGCACATTTCGCG
nrfA-64	AAAAAAAGGAATTTAAACAAAATGTGCTGCCTCAATGGTTAAAAGCAGCAAAAGAACGTGAAGCGAAAAT
nfrA-100	AAAGAAATTCTGGAAGGTATCAGGCACGCCCAATGGAGATGGGACTTTGTAGTGGCATCACATGGGGGAG
nrfA-60	AACGCCTCCTTAGCGATGGTCTTATCTATGCTTATCAAGCTGAAAATAATTTAGATGTGCTTAAAGAAAA
nrfA-61	AAACCTACTTTGGCACTTATCCGCAAATCACAATGGCGTTGGGATATGGTACACTCTTCACACGGTGCCG
nrfA-62	AAGTCTCAATGGCGTTGGGATATGGTACACTCCTCACATGGAGCGGCTTTCCATGCTCCTATAGAGTCAG
nrfA-63	AAAGTGAGTCGTGCTCCTATTATCAAGGCACAACACCCCGATTATGAGCTTTCTCAGCTCGGTATCCACG
nrfA-66	AAAAAATATCTTACCTTCCCTTGGGATGAGGGAAAAGCAGTGGAGGATATGGAGAAATATTATGATGAAT
nrfA-137	AAAAAATGGACTGACCTGGGCTCTGAAATAATTAATCCAATCGGATGTCAGGATTGCCATGATCCTAAAA
nrfA-138	ACGCAATTTTCCAGACAAGAGATGCGTTCACTTGTTTGTGCACAATGCCACGTTGAATATTACTTTAAAG
nrfA-65	AAAAAACTATTGCAGCAGCACATATTGAGGCAAAAATTGCGTGGGATAATGGTGCAACACAAGAGCAGAT
nrfA-69	AAAGAGGGCATTTTTGGGTTCAAGTAGATCAAATGAAAACAGCAAGAAACAACAAAGATTTTCTTAATGC
nrfA-70	ATCCAAATTTTGAAAAATGGGGTAAGGTTTTTCCAGAGCAACTAAAAATGTATTTAACGGTTGAAAAAGA
nrfA-76	AAAAAACTTACGATGAATTTGACACTTGGAGAGATGGTAACAAGCCTACCGAAATCGAAGTAGCGGGGAT
nrfA-74	AAAAAACAAAATGAAGGTACAGGAGGCATAGCCTCTAAGGAATTTGTAGAGCTTAGCGATAATAACCCTA
nrfA-75	AAAAAACGTTCCAGCAGTAAATGGTATGGAAGAAAACTCTGTTGAACACAGCGGTCCTCACGGTGGTAAA
nrfA-77	AAAAAAGGAACTCCATTTAGAATTGAAATGTTCGATGATCATTATGAAGCGGTACGCGGTGTATTTGATA
nrfA-82	AAAAACGGTATAACCGGCTTCAAAACATCTAATCTCGGTTTTGAAGATATGCAAAAACTAAATCCGGGCG
nrfA-84	AAAAATCCAACACCCAGAGTACGAGCTATACAGCGGCGGCGTACATGCTGCAAACGGCGTAAGCTGCGCT
nrfA-71	AAAAAAAGTTACTAACCATAACATCACTTCACCTTTAGTTGATATTAACTCAGCTTGTAAAACTTGCCAT
nrfA-79	AAAAACAAGGCATTAAAGCTTCAAGAAGCGAGATGAGAACTTTAGTATGCTCTCAATGCCATGTTGAGTA
nrfA-72	AAAAAAATCGCGGCTAAACACGGCGAGAGTATCCATGGGTCGCATCTAGGCAGCACCTGTGCGGACTGCC
nrfA-78	AAAAACAAAGTGTTGTATGTAGCTACTTTCTTAGTGGCTGTCGTACTCGGTGCGGCTATGTTCGCGCTTT
nrfA-80	AAAAACCGCGTAAGTTTTATACAAAACCGCCACGCTTATGAACTTAGAGGTTGCGAAAACGCTCTGCTTT
nrfA-73	AAAAAAATGGAGCATAAAAGCGACGAAAATCCAAACGGAACCGAGCTAATAGAGACTCCTTTCGGCGGTT
nrfA-81	AAAAACGCTTCTTTTCCATAGAAAAGCTCATATTCGCTGGGACTTCAGCTTTAGCGAAAACAGCTACGGA

Table A2.3 continued.

Probe	Sequence
nrfA-83	AAAAAGGTACCTGTGTAAACTGTCACACCGGACACCTTATGGCTATGATGGTAGATACGGACTACAAACA
nrfA-89	AAAAAATTTACCGGAACCACCCTCATCCTCACCGGCATCATTGTGGTGCTGAGCGCCCTGCTGATTGGCG
nrfA-93	AAACACGCCGACTCCGGCGCAAACATGATCAAGATGCAGCACCCCGATTATGAGATGTACACCGCCAACA
nrfA-92	AAAATTACGGCCCGACAACTTACGGGGGCTCTGAACCGTACAGTAAGCTCGAACGCTATCCGGCGATGGT
nrfA-67	AAAACGGCCACGTTGACTTTGTCAATGCAAAAACCGGTGCAGACATTGTGAAAGCACAGCATCCGGAATT
nrfA-68	AAAGAAAGCCGTCAGCCGGATGACAAAGATCCACGGACTGAAGTAGCTAAGAGCCGCATCGAAGAAGACC
nfrA-101	AACCAGGAGATGCGCACCTACGTGTGCGCCCAGTGCCACGTCGAGTACTACTTCGCCGGTGAGGGCAAGA
nrfA-91	AAAACCCTCACGTTCCCCTGGAAGAACGGCCTGACCGTCTACGACGAGATGAACTACTACGACGAGGTCG
nfrA-102	AAAAAGACCGAAGAGCCCACCGTGGCGGGTAAGAGCCGGTACAAGAAGGGGTATGACGACGACGGCGTTT
nfrA-103	AAGCAGATCGACAAGGCCCTCTATGACCAAGCCAGGGATCACTACGAAGAGGCGTTCTACCGGGTAGTCT
nrfA-90	AAAAACGACATGAAGGCCTGCATCCAGTGCCACACCGAAAGCCCGGACTGGCTCAAGCAGCAGGTCATCA
nfrA-104	AAAAAGCTCACCCGCCAGGAAATGCGCTCCGTGGTTTGCGCTCAGTGCCACGTGACGTACAACATCCCCA
nfrA-105	CCGGAAGTGCCTGGTCTGATCGCTCAAATGGGAGACAGCTATTATACGGCTTCTTTTGATTCCTTAAAGG
nrfA-86	AAAAAGATGGGGTAGATGTAAACTCTATGGATTTTGATGAATTTGTAGATTCAAAGCCAGATATGTCTAC
nfrA-106	AAATCACCTGATGTACCCCGGCTGATTGACAAGATGGGTGAGCTTGACTATTTCACCGGCAAATGGGCCC
nfrA-107	ATGACAGAAGTTTTACGGACTGGACAAATAAAGTCAGCAAGGCGCCAATGCTTAAAGCGCAGCATCCCGG
nfrA-108	AACATTCCCATGGGATAACGGTTTCAGCGCTGAGGCCATGGAAAAGTACTACGACGCCATTAGCTTTGTA
nfrA-109	AGTGCCATTCCGAATACTACTTTAAGAAGACCGCATGGACCGACAAAAAAGGTAAGGAACAAACCGCCGG
nfrA-110	AAAACTGAGTGGACCGATCCCAATGGAGACAAACAGACGGCCGCCGTAGTTACCTATCCCTGGGACAACG
nfrA-111	AAACTGACCGTTACCAGGGAGTACCTGAAGCGGGGTCTCGCTGCTGAAGGCAGCCTCAAGTATGCCGATG
nrfA-112	CACGTAGAGTATTATTTCAAAAAAACCGAATGGAAAGATGCCAAAGGTGTGGATAAAACAGCAATGGTGG
nrfA-113	AAAAACACCCATGCTAAAGGCGCAACATCCTGATTATGAGTTTTGGAAAACAGGTATTCATGGTCAAAAA
nrfA-114	AAAAAGACCAAAGAGAGCGACAATATTACCGATATGCTCAAAGAGAAGCCCGCCCTTGTGGTGGCTTGGG
nrfA-115	AAAACAGCCCACAGCCAAGGCATCGAGGGCAAGGCAATGAGCGAGGAGTGGGCGAGATACTATCCAAGAC
nrfA-116	AGACCTTATCTCAAACGCGGCCTCGAAGCGAGCGGCAGAAAACTGGAAGATCTCACTTTTCAGGATATGC
nrfA-117	CTCTGGCCTGCGCCCAGTGCCACTCTGAGTATTACTTCAAAAAGACCCCTTACACCGACGAAGCAGGCAA
nrfA-88	AAAAAAAAGCGACAAAATCACCGATCTTCTGAGAGAAAAGCCTCAACTTGCGATCATCTGGGCCGGCTAT
nrfA-85	AAAAGAATTTTAGTTTTGGCCTGTATCGTTGTAGCGATACTTGCGATGATTTTTATCGGGTTTGACATAA
nrfA-87	AAAAAAAGAGGAGCAGCGTATCGCGATAAATACACCTCGCGGAGTGCCGCAGATGCCTAGTAAAAGTTCC
nrfA-118	AAGACTGCCGAGTTGACCGTCGGCAGGCCCTCACTGCAAAGTGGGCTCGAAAGTGCCGGTCTGGACATGA
nrfA-119	GGGCTCGAAAGTGCCGGTCTGGACATGAAGAAAGTCAGTGAAACAGAGATGCGCTCGCTCGTTTGCGCCC

Table A2.3 continued.

Probe	Sequence
nosZ-012	ACGCCACCATCGTGCACCGTTCCAAGATCAATCCGATCTCGATCTGGGACCGCGCCGATCCGATGTTTGC
nosZ-013	CCGAGCTTCGCCGAACCCCACGACGCCACCATCGTGCACCGTTCCAAGATCAATCCGATCTCGATCTGGG
nosZ-014	GTCTACATGACCTCGACCGCGCCGGCGTTCGGCCTCGAGCAGTTCCAGGTCAAGCAGGGCGATGAAGTCA
nosZ-015	ACGGCGTCAACATGGAAGTCGCGCCGATGGCGACGGCGTCGGTAACGTTCGCCGCCGATAAGGCCGGCGT
nosZ-019	ATCAGCCCGGCCACAACCACTCCTCGATGGGTCAGACCAAGGAGGCGGATGGAAAGTGGCTGATCTCGCT
nosZ-020	CCTCGATGGGTCAGACCAAGGAGGCGGATGGAAAGTGGCTGATCTCGCTCAACAAGTTCTCGAAGGACCG
nosZ-021	AGATCAATCCGATTTCGATCTGGGATCGCGCCGATCCGTTCTTTGCCGAAGCGGTCAAACAGGCCAAGGC
nosZ-022	ATGGAAGTTGCGCCGATGGCGACGGCCTCGGTATCGTTCTCGGCCGATAAGGCCGGCGTCTACTGGTACT
nosZ-023	TGGAGCAGTTCCAGATCAAGCAGGGCGATCAGGTCACCGTCTACGTCACCAACATCGATGCGGTGGAAGA
nosZ-005	GTCAAGCAGGGCGACGAGGTTACGGTCTACGTCACCAATATCGACGAAGTCGAAGACCTCACGCACGGCT
nosZ-006	AACCCAACCTTTGCGGAGCCACATGACGCCACGATCGTGCATGCCTCCAAGATCAACCCGGTCAGCCTGT
nosZ-007	AACGATCAGCTGATCGACATTTCGGGCGACCAGATGGTTCTGGTGCACGATAATCCGACCTTCGCCGAGC
nosZ-008	AAAAAGTAGACCCGATCCGCCAGAAGCTTGATGTCCAATACCAGCCCGGACACAACCACACGTCCATGGG
nosZ-009	ATGGGTCAGACCAAAGACGCGGACGGAAAGTGGCTCATTTCGCTGAACAAGTTCTCGAAGGATCGCTATC
nosZ-010	AAAGTGGCTCATTTCGCTGAACAAGTTCTCGAAGGATCGCTATCTTAACGTCGGCCCCCTGAAGCCGGAG
nosZ-011	ACTGTCTACGTCACCAACATCGACGAGGTCGAGGACCTGACGCACGGCTTCGCCATCGTCAACTACGGCA
nosZ-024	ACCGAGTTCACCGTCCAGCAGGGCGATGAGGTGACGGTGTATGTCACCAATATCGACGAGATCGAAGATC
nosZ-025	AACTACGGCATCAACTTCGAGGTTGCGCCGCAGGCCACCGCCTCGGCCACCTTCACCGCCGACAAGCCGG
nosZ-026	ACCAGCTGATCGATATATCGGGCGACGAGATGGTGCTGGTGCACGACAATCCGACCTTCGCCGAGCCGCA
nosZ-027	AACAAGGTGCGCGTTTACATGACGTCCTCGGCGCCGGCCTTCGGGCTGGAGAGCTTCAGCGTCAAGCAGG
nosZ-028	AACATGGAGGTGGCCCCCCAAGCCACGGCATCGGTGACGTTCAAGGCAAGCAAGCCCGGCGTCTACTGGT
nosZ-029	GAGCTTCACCGTGAAGCAGGGCGACGAGGTGACGGTGTATGTCACCAACATCGACGATGTCGAGGACCTG
nosZ-030	GAGGTGACGGTCTATGTCACCAACATCGACGACGTCGAGGACCTGACCCACGGCTTCACCATCGTCAATT
nosZ-038	TCGACAGCCAGGTCTGCAAGTGGAGCATCGATCTCGCCAAGCGCGCCTTCAAGGGCGAGAAGGTCAATCC
nosZ-033	AAGATCAACCCGTTGCACGTGTGGAAGCGCGATGACCCGTTCTTCGCCGACGCGGTGCAGCAGGCCAAGG
nosZ-035	AGACCAAGGAGGCGGACGGCAAATGGCTGATCTCGCTTAACAAGTTCTCGAAGGACCGCTTCCTCAACGT
nosZ-031	ACAAGTTCTCGAAAGACCGTTTCCTCAACGTTGGTCCGCTCAAGCCCGAGAACGACCAACTCATCGACAT
nosZ-034	GGTGCTCGTGCACGACGGCCCGAGCTTCGCGGAGCCGCATGACGCCACCATCGTGCATCGCTCGAAGATC
nosZ-032	ACGGCTTCTGCATCAACAACTACGGCATCAACATGGAGATCGGTCCTCAGCAGACATCCTCCGTCACCTT
nosZ-111	GTGCACTATCAGCCAGGCCACAACCACTCGTCCATGGGCCAGACCAAGGAGGCCGATGGCAAATGGCTCA
nosZ-110	AAACAGGGGGATGAGGTCACTGTCTTCGTCACCAATATTGACGAGGTTGAGGACCTGACCCACGGCTTCT

Table A2.3 continued

Probe	Sequence
nosZ-113	ACGATCAGCTCATCGATATTTCGGGCGACGAGATGGTACTCGTGCACGATGGTCCGAGCTTCGCCGAACC
nosZ-109	AACGAATGGTGCTCGTTCACGACGGCCCGAGCTTTGCAGAGCCCCATGACGCCACCATCGTCCACCGCTC
nosZ-112	CTACATGATCTCCACGGCGCCGCAGTTTGGGCTCGAGGAGTTCCAGGTGAAGCAGGGCGACGAGGTGACG
nosZ-036	AACAAGGTTCGCGTCTACATGACCTCCACCGCCCCGGCCTTCGGCCTCGAAAGCTTTCAGGTGAAGCAGG
nosZ-037	AACCTCCTCGGTCACGTTCAAGGCCGATAAGCCGGGCGTCTACTGGTATTACTGCACATGGTTCTGCCAT
nosZ-016	TGGTATTACTGCTCGTGGTTCTGCCACGCCATGCATATGGAAATGCAGGGCCGTATGCTCGTCGAACCGA
nosZ-017	ATTACTGCTCGTGGTTCTGCCACGCCATGCATATGGAAATGCAGGGCCGTATGCTCGTCGAACCGAAAAA
nosZ-018	ATTTCGCTGAACAAGTTCTCCAAGGACCGTTACCTCAATGTCGGTCCGCTCAAGCCTGAGAACGACCAGC
nosZ-039	ACGCACGGTTTCTGCATCGTCAACTACGGCGTCAACATGGAGATCGGCCCGCAGCAGACCTCGTCGGTGA
nosZ-040	ACTACGGCGTCAACATGGAGATCGGCCCGCAGCAGACCTCGTCGGTGACCTTCATCGCCGACAAGCCGGG
nosZ-043	AAACCAAGGAGGCCGACGGCAAGTGGGCCGTCGCGTTGTGCAAGTTCTCCAAGGACCGCTTCCTGAACGT
nosZ-044	AAATGGAACATTGAAGAGGCCATCCGCGCCTACAAGGGCGAGACGGTGAACCCGATCAAGCAGAAAATCG
nosZ-041	AACGCCACGCAGGCCGAAACCAAGGAGGCTGACGGCAAGTGGTGCGTCGCGCTGAACAAGTTCTCCAAGG
nosZ-042	AACAAGGGCGACGAGGTCACAATCTTCGTAACCAATGTCGACAGCGTCGTCGATCTCACTCACGGCTTTA
nosZ-045	CCTCCATGGGCCAGACCAAGGAGGCTGACGGCCAGTGGCTGTTGTCCTTGAACAAGTTCTCCAAGGACCG
nosZ-045B	CAGCCTGGAGCAGTTTACCGTCAAGCAGGGCGACGAGGTGACGGTCTACGTCACCAACATCGACGACGTC
nosZ-046	CCGACGGCGTGGTGCTGGAGGACGCGGCCGACGTGATCCGCGACGGCAACAAGGTGCGCGTCTACATGTA
nosZ-047	CTGTCGCTGAACAAGTTCTCCAAGGACCGGTTCCTCAACGTCGGGCCGCTGAAGCCCGAGTGCGACCAGC
nosZ-001	CCGGCGACGAGATGAAGCTGGTGCATGACGGCCCCACCTTCGCCGAGCCGCACGACACGCTGATCGTCCA
nosZ-002	ATGGGCGAGACCAAGGAGGCCGACGGCAAGTGGCTGGTGTCGCTGAACAAGTTCTCCAAGGACCGCTTCC
nosZ-048	CTGTCGCTGAACAAGTTCTCCAAGGACCGCTTCCTGAACGTGGGGCCGCTGAAGCCCGAGTGCGACCAGC
nosZ-049	AAAGACCGGTTCCTCAATGTCGGCCCGCTGAAACCTGAAAACGAGCAGCTGATCGACATTTCAACGGACG
nosZ-050	ACGGCGTGGTTCTGGAGGACGGTTCCGACACGGTGATCCGCGACGGCAACAAGGTGCGTGTCTACATGTG
nosZ-159	AAGCTGGTGCACGATAGCCCGACCTTCGCCGAGCCGCACGACTGCATCATCGTGCGCGCCGACATCGTCA
nosZ-051	AAGGTCAATCCGCTGCGCGTGTGGAAGCGCGACGATCCGAGTTGGGAGGACGCGCGCCGGTGGGCGGCAA
nosZ-052	AAACCGGACAACGATCAGCTGATCGACATTTCCGGTGATGAGATGAAGCTGGCGCACGACGGGCCCGTCT
nosZ-156	AAAAGGACGGGGTCAAGCTCGAGGAAGCGCAGCATGTCATTCGCGACGGCAACAAGGTCCGCGTCTACAT
nosZ-157	AAACTCGTTCATGATGGCCCAAGTTTCGCCGAGCCGCATGACTGCCTCCTCGTTCACCGCTCCAAGGTGA
nosZ-155	ACGACATCGAGAACTTGACACACGGCTTCACCATCGTGCGCTACGGCATTGCGATGGAGATCAGCCCGCA
nosZ-157B	AACATGGACGACATCGAGAACTTGACGCACGGCTTCACCATCGTCCGCTACGGCATTGCGATGGAGATCA
nosZ-158	AAAAAGGATGGCGTCGATCTGGACGACGCGGCGAAGGTCGTTCGAGACGGCAAAAAGGTGCGCGTTTACA

Table A2.3 continued.

Probe	Sequence
nosZ-144	ATCTTCGGCCTGGGCACCTTCAAGGTGAAGAAGGGCGACGAGGTGACCGTGGTGGTGACCAACATCGACG
nosZ-143	AAGAAGGTCAACTACCTGCGCCAGAAGATCGACGTGGCCTACCAGCCCGGGCACTGCCACACCTCGATGG
nosZ-145	AAGAAGGTCAACTACCTGCGCCAGAAGCTCGACGTGGCCTACCAGCCCGGTCACTGCCACACCTCGATGG
nosZ-136	AAACCATGGAAATCGCATGGCAAGTCATGGTGGACGGCAACTTGGACAACGGCGATGCCGACTACCAAGG
nosZ-137	AAACCATGGACATCGCATGGCAAGTATTGGTGGACGGCAACTTGGACAACGGCGATGCCGACTACCAAGG
nosZ-138	AAACCGGAATGCGACCAGTTGATCGACATCTCCGGCGACGAAATGCGCCTGGTACACGACAACCCGACCT
nosZ-146	ACGATCCGTTCTTCGCCGAAACCGTTGCCCAGGCCAAAAAGCACGGCATCGCCCTGGAGAGCGACAGCAA
nosZ-053	ACAAGGTGGAAGACCTGACCCACGGTTGCGCGATCCCGACCTACAACATCAACTTCATCGTCAATCCGCA
nosZ-054	TGGAAGATCTGACCCACGGTTGCGCGATCCCGATGTACAACATCAACTTCATCGTCAATCCGCAGGAAAC
nosZ-056	CGATGGGCGAAACCAAGGAAGCCGACGGCAAGTACTACAACTCGGGCAACAAGTTCTCGAAGGATCGCTT
nosZ-058	AACCGCACGACGGCATCATCGTGCGCCGCGACCGCGTGAAAACGCGCCAGGTGCACACCATGACCGACTT
nosZ-076	AACATCAATTTCATTGTCAACCCGCAGGAGACCAAATCGGTCACCTTCAAGGCTGACAAGGCGGGCGTGT
nosZ-057	CAAGATCGACGTGCACTATCAGCCCGGCCACGGCTACTCGTCGATGGGCGAAACCAAGGAAGCCGACGGC
nosZ-059	AAACCAAATCGGTCACGTTCAAGGCCGACAAGCTGGGCGCGTACTGGATCTATTGCACCCACTTCTGCCA
nosZ-060	AACAAGGTGAATGTGCGCCTGATGTCGCAGGCGCCGGCCTACAGCATGCCGATCATCAAAGTGAAGAAAG
nosZ-055	CCGACGGCAAGTACTACAACTCGGGCAACAAGTTCTCGAAGGATCGCTTCCTGCCGGTCGGCCCGCTGCA
nosZ-061	AATTCCGGAAACAAGTTCTCGAAGGACCGCTTCCTCCCGGTGGGGCCGCTGCATGTCGAGACCGAGCAGC
nosZ-063	GTGAAACTGGGCGACGAAGTCACCATTACCCTGACCAACCACGACAAGGTGGAAGACCTGACCCACGGCT
nosZ-062	AACGCGTACACGACGCTCTTCCTCGACAGCCAGATCGTGAAGTGGAGCATCGACAAGGCGATCGCGCAGT
nosZ-071	AGCCTGCGGGAGTTCAAGGTCAAGAAGGGTGACGAAGTGACCATCATCCTGACCAACCACGACAAGGTGG
nosZ-072	AGCCTGCGGGAGTTCAAGGTCAAGAAGGGCGACGAGGTGACCATCATCCTGACCAACCACGACAAGGTGG
nosZ-073	AAGGTCAAGAAGGGCGATGAAGTGACCATCATCCTGACCAACCACGACAAGGTGGAAGATCTGACCCACG
nosZ-068	AGATGATCGACATCAGCGGCGACAAGATGAAGCTGGTGGCCGACCACTCGGTGCTGTCCGAGCCGCACGA
nosZ-069	CGCAGCTGATCGACATCAGCGGAGACAAGATGAAGCTGGTGGCCGACCACGCCGTCTATTCCGAGCCGCA
nosZ-070	GCAAGTTCTTCATCTCGGACAACAAGTTCTCGAAGGATCGCTTCCTGCCGGTGGGGCCGCTGCACCCGGA
nosZ-067	AAGATGAAGCTGGTGGCCGACCACTCGGTGCTGTCCGAGCCGCACGACTCCATCGTCATCCGCCGCGACA
nosZ-066	CAGCGGCGACAAGATGATGCTGGTCGCCGACCACTCGGTGACGTCCGAGCCGCACGATTCGATCGTCATC
nosZ-074	AAGGTGGTGCTCGACCGCATTGACGTCCACTATCAGCCGGGTCACAGCTACGCCTCGATGGGTGAAACCA
nosZ-075	TGCGTGAATTCAAGGTCAAGCGCGGTGACGAAGTGACCATCATTCTGACCAACCTGGACAAAGTGGAAGA
nosZ-077	GCGACAAGATGGTGCTGCTGGCGGACCACCCGGTGCGCGGCGAGCCGCACGACTTCATCATCTTCAAGCG
nosZ-078	CTCGCTGGCGCCGGCCTTCAGCCTGCGCGAGTTCAAGGTGAAGAAGGGCGACGAGGTCACGCTGATCCTG

Table A2.3 continued.

Probe	Sequence
nosZ-079	AGTGGCTGTGCGTGGGCTGCAAGTTCTCCAAGGACCGCTTCCTGCCGGTCGGCCCGCTGCACGCCGAAAC
nosZ-080	CAGCCGGGCCACATCAACGCCTCGCAATCGGAAACGAAAGCGGCCGACGGCAAGTGGCTGGCGGTGGGCT
nosZ-081	TCGCGATCCCGAAGTACAACGTCAACTTCATCGTCAATCCGCAGGAAACCGCGTCGGTGACCTTCAAGGC
nosZ-082	AACGCGAAGGTCGTGCTCGATCGCATCGACTGCCATTACCAACCCGGGCACCTGAACGCCAGCATGGCCG
nosZ-083	AAGATGAAGCTGCTCGCCGACCATCCGGCGCATCCGGAACCGCACGATTTCATCATCGTGAAGCGCGAAA
nosZ-085	AAATCATCGACATCTCCGGCGAGAAGATGAAGCTGCTCGCGTCGTTCCCGACGCCGCCCGAGCCGCACGA
nosZ-086	GGAAATCATCGACATCTCCGGCGAGAAGATGAAGCTGCTCGCGTCGTTCCCGACGCCGCCCGAGCCGCAC
nosZ-087	TCCGACTACATCGTCGACCGCATCAACGTGCACTACAACATCGGCCATCTGCAGGCGGTGGGCGGCGACC
nosZ-088	CCAAGACGATCACGTTCGATGCGGGCAAGCCGGGCGTCTACTGGTACTACTGCACGAACTTCTGCCACGC
nosZ-089	TCGAGACGATTCGCGACATGATTCATGGCTTCGCGCTGCCCGACCACAACCTGAACATCGCGCTGGCGCC
nosZ-090	ACCCGGCCGGCGACTACCTGATCGCGCTCAACAAGCTGTCGAAGGACATGTATGTCCCGGTTGGGCCCGA
nosZ-091	AAATCATCGACATCTCCGGCGAGAAGATGAAGCTGCTCGCGTCGTTCCCGACGCCGCCCGAGCCGCACGA
nosZ-094	AAATCATCGACATCTCGGGCGAGAAGATGAAGATGCTCGCGTCGTTTCCGACGCCGCCCGAGCCGCACGA
nosZ-092	AAGCCGCTCGTTCGCCAGACCTACACGCCGGCGGCGGACGCGGTTGAAGCCGGAAAGGAGCGAGTGGTGC
nosZ-093	AACATCGGCCATCTGCAGGCGGTGGGCGGCGACCATACGCAACCGGCCGGCGACTACCTGATCGCGCTCA
nosZ-093B	AACGTGCACTACAACATCGGGCACCTGCAGGCGGTCGGCGGCGACTCGACGCATCCGGCCGGCGACTACC
nosZ-096	AACCAGGAAATCATCGACATCTCAGGCGAGAAGATGAAGATGCTCGCGTCGTTCCCCACGCCGCCCGAGC
nosZ-097	AACCAGGAGATCATCGACATCTCGGGCGAGAAGATGAAGATGCTCGCGTCGTTCCCCACGCCGCCCGAGC
nosZ-095	CGACTACATCGTCGACCGCATCAACGTGCACTACAACATCGGTCACCTCCAGGCGGTCGGCGGCGACTCG
nosZ-098	AACATCGCGCTCGGCCCCGGCTACACCAAGACCATGACGTTCGACGCTGGTAAACCGGGCGTCTACTGGT
nosZ-100	AAAGATCAGTACCTCGGCGTCGGCCCCGACCTGCCCGAGAACCAGGAAATCATCGACATCTCCGGCGAGA
nosZ-101	AACATCGGCCACCTGCAGGCGGTCGGCGGCGATCACATGCATCCCGCGGGCGATTACCTGCTCGCGCTCA
nosZ-104	ATCGACATCAGCGGCGACAAGATGAAGTTGCTGCTGGATTTCCCGACGATCGGCGAGCCGCACTACGCGC
nosZ-103	AAATACATCATCGGCAGCGGCAAGCTGCAGTCGATCACCACCGCCTTCAATATGGAGAAGATCCTCACCG
nosZ-105	AGCCGATCGAGGTCTACCCGAAGGAGGAGAACAAGAACCCGCACGCGATCTGGGACGTCAAGGACGCGGG
nosZ-154	TCGAGCCGGGCAAGACGGTGACGGTGAAGTTCAAGGCCGACAAGGAAGGCGTCTATCCGTACTACTGCAC
nosZ-152	ACCGCTTCGTGCCGGTGGGCCCGCTGCACCCGCAGAACCACCAGCTGATCGACATCAGCAACGACAAGAT
nosZ-147	AAAACGGCTACGGCTATTCGGAAGAAACGAAGCCGATGCTGGAGACCACCTTCGGCAACATCCCGTGGGA
nosZ-107	AAACAGAAACTCTCGTTTGGAAACCTACCAAAGTAGGGGTATGGCCATTCTACTGTACAGACTTCTGTTC
nosZ-106	CCCTTCACCAAGAAATGCAAGGATACATTCGTGTATCTGCTGCTAACGCCAACACTCCGCTTAAGTGGAG
nosZ-153	CGAGAAGATCCAGGCCGCCATCAAGGCCGGCAAGTTCGAATCCAAGGATCCGTACGGCATTCCGGTCATC

Table A2.3 continued.

Probe	Sequence
nosZ-004	GAGAAGATCCAGGCTGCCATCAAGGCCGGCAAGTTCGAATCGAAGGATCCGTACGGCATTCCGGTGATCG
nosZ-133	GGGCGGTGATCTTCGACATCAAGGCGATCGAGGCCGCGGTCAAGGGCGGCAAGGCGGAGATGATCAACGG
nosZ-134	TTTGACGGCAAGGGCAACGCCTTTACCACGCTGTTCCTGGATAGCCAGGCGGTGAAATGGAATATCCAGA
nosZ-135	TTTGACGGCAAGGGCAACGCCTTTACCACGCTGTTCCTGGACAGCCAGGCGGTGAAGTGGAATATCCAGA
nosZ-124	ACCACCCGCACATGTCGTTCACCGATGGCACCTATGACGGTCGCTATCTGTTCATGAACGACAAGGCCAA
nosZ-139	AACGCCTATACGACCCTGTTCATCGACAGCCAGGTCTGCAAATGGAACATCGAGGACGCCAAGCGCGCCT
nosZ-140	AACGCCTATACGACCCTGTTCATCGACAGCCAGATTTGCAAATGGAACATCGAGGACGCCAAGCGCGCCT
nosZ-141	AACGGCCCGCACGGCATGAACACCGCGCCGGACGGCATCCATATCGTCGCCAACGGCAAGCTGTCGCCGA
nosZ-142	CCCGCACGGCATGAACACCGCGCCGGACGGCATCCATATCGTCGCCAACGGCAAGCTGTCGCCGACCGTG
nosZ-162	AGCATCGTGTGCCGATCCCAAATGACGGCAGCATTCTGGATGAGCCGGAAAAGTATCACGCGATTTTTAC
nosZ-163	CGCGGTGGCACCTGGCAAAACGGTGACCTGCATCATCCGCATATGTCGTTTACAGATGGAACCTATGACG
nosZ-160	GCTCGAAGTACACGCGCTACATACCGGTGCCCAACAGCCCGCACGGCTGCAACACCGCGCCGGACGGCAT
nosZ-161	GCTCGAAGTACACGCGCTACATCCCGGTGGCCAACAGCCCGCACGGCTGCAACACCGCGCCGGACGGCAT
nosZ-163B	ACACCATGGAAGTCGCCTGGCAGGTTATGGTGGACGGCAACCTGGACAACTGCGATGCCGACTACCAGGG
nosZ-165	ACACGATGGAAGTCGCCTGGCAGGTGATGGTGGACGGCAACCTTGATAACTGCGATGCCGACTATCAGGG
nosZ-164	ACGTCACCAAGGTCCCGGACGTGTTCGCCGACAAGATCGAACCGCGCGGCTGCATCGTCGCCGAGCCGGA
nosZ-166	AGCCGCGGCGGCATCTTCCTGAACGGCGACCTGCATCATCCGCATATGTCGTTCACCGACGGCACCTACG
nosZ-167	AAAACATATAGTCATCAACGGCAAGCTGTCACCGACAGTTTCCGTCATTGACGTCACCAAACTGCCGGAC
nosZ-168	AAAGCACATCGTCATCAACGGCAAGCTTTCGCCGACGGTATCGGTGATCGATGTCACCAAGCTGCCCGAC
nosZ-172	AAAAAATATGTATCCATTTTCTCGGCGATTGATGGAGACACCATGAAGGTCGCCTGGCAGGTGATCGTCG
nosZ-173	AAAAATTTCCGCGTACCGGTTATGTTTTCGCAAATGGTGAGCACCGGATTCCCATTCCCAACGACGGAAT
nosZ-171	ACGGCAAGGGCAATGCCTATACAACGCTGTTCCTGGACAGCCAGGTGGCCAAGTGGAACATCGACAAGGC
nosZ-169	AGCAGATGACCGCCAACGAGCAGGACTGGGCGGTCATCTTCGACCTCAAGGCGATCGAAGCGGCCGTCGC
nosZ-170	ACGGCCAAAAGGGCTCGCCCCTCACCCGCTACGTGCCGATCTCCAACAGCCCCCACGGCTGCAACACGGC
nosZ-176	AACCTGGCAAGTCGTCGTCGACGGCAACCTCGACAACGTCGACGCCGACTATCAGGGCAAGTATGCGATT
nosZ-174	AAGCTGCCGGACGTGTTTGCTGACAAGATCAAGCCGCGCGACGCAATCGTTGCCGAGCCGGAACTGGGTC
nosZ-175	AACACACGCGTTGCCCGGATCCGCTGCGACATCATGAAGACCGACAAGATCATATCGATCCCGAATGCAT
nosZ-177	AACGGCAAGCTTTCACCCACAGTGACGGTCATAGACGTTCGCAAACTGCCGGACGTCTTTGCCGACAAGA
nosZ-178	AATGTGGACGCGGACTACCAAGGTAAGTACGCCATCTCCAGCTGCTATAACTCGGAGGAAGGCGTCACGC
nosZ-179	GGCGAGGAAGGCGTCACTCTCGCCGAGATGACCGCCAACGAGCAGGACTGGGCGGTGGTGTTCGACATCG
nosZ-180	ATACGGCCTTCGACGGCAAGGGGAATTGCTATACGACCCTCTTCCTCGATAGCCAGATGTGCAAGTGGAA

Table A2.3 continued.

Probe	Sequence
nosZ-181	AACCTGGCGGAGATGACGTCGAGCGAGCAGGACTGGGCGGTGATCTTCGATATCAAGGCGATCGAAGCGG
nosZ-122	ACCTCGACAACACCGATTGTGACTATCAGGGCAACTATGCCTTCTCGACCTCGTACAACTCGGAAATGGG
nosZ-123	GACGTGATGAAATGCGACAAGATCATCGAGATCCCGAACGCGCATGACATCCACGGCCTGCGTCCGCAGA
nosZ-119	GGCAGGTGATGGTGTCGGGCAACCTCGACAACACCGACTGCGACTATCAGGGCAAATACGCCTTCTCGAC
nosZ-120	ACCTCGACAACACCGACTGCGACTATCAGGGCAAATACGCCTTCTCGACCTCCTACAACTCGGAAATGGG
nosZ-121	TGCACCACCCGCACATGTCGTTCACCGACGGTACCTATGACGGCCGTTTCCTGTTCATGAACGACAAGGC
nosZ-131	GTGATCAACGGCAAGCTGTCGCCGACCGTCTCGGTGATCGACGTCACCAAGGTCGACGCCCTGTTCGAGG
nosZ-132	AGATCCCGAACGCCCATGACATCCACGGCATGCGCCCGCAGAAATATCCGCGCACCGGCTATGTCTTCGC
nosZ-125	AAACACCGATCGTCAATGACGGCTCGATCCTGGATGATCCCAGCCAATACGTGAACTTCTTCACCGCGCT
nosZ-126	AAATCGAAGCGGGTATCGCCGCCGGTGACTTCACCGAGATGAACGGCGTGAAGGTGATCGACGGTCGCAA
nosZ-127	GGGCTGACCAACGAGTCGCTCAAGATCCTGACCGAAGGCCTGCTGCCGGAAACCAAAGAATTTCTCGCCA
nosZ-128	CCGCGGTCCTGCATCGTGGCGGAACCGGAACTGGGCCTTGGCCCGCTTCACACTGCCTTCGATGGCAAGG
nosZ-129	AAAAGGGCATCGCCGAGGGCGACTATCAGGAGCTGCAGGGCGTCAAGGTCATCGACGGGCGCAAGGGCAA
nosZ-130	AAATACGCTTTCTCGACCAGCTATAACTCAGAAATGGGCACTAACCTTGCCGAAATGACCGAATCTGAGC
nosZ-003	ACCACCCGCACATCTCGATGACCGACGGCAAGTACGACGGCAAGTACCTGTTCATCAACGACAAGGCCAA
nosZ-114	ACCACCCGCACATCTCCATGACCGACGGCAAGTACGACGGCAAGTACCTGTTCATCAACGACAAGGCCAA
nosZ-115	ACCTGCTACAACTCCGAGAAGGCCTACGACCTGGGCGGCATGATGCGCAACGAGCGCGACTGGGTGGTGG
nosZ-116	AACGGCGACTGCCACCACCCGCACATCTCCATGACCGACGGCCGCTACGACGGCAAGTACCTGTTCATCA
nosZ-117	AAGTACCTGTTCATCAACGACAAGGCCAACACCCGGGTGGCCCGCATCCGTCTCGACATCATGAAGACGG
nosZ-118	CCTGGGGCCGCTGCACACCACCTACGATGGCCGCGGCAACGCCTACACCACGCTGTTCATCGACAGCCAG
nosZ-064	AAAGTCGTCGTCGACCGGATCGACGTGCACTACCAGCCGGGCCACGGCTTCACGTCGATGGGCGAGACGA
nosZ-065	AACATCGACTGCATGTCCGGCTGGGGCATCACGAACGAATCGCGCAAGATCATCGGCACCAAGGCCAACG
nosZ-084	AACGTCCAGGGCTTCCACGGCATCTTTCCGGACAAGCGCGACCCGGTCGACCCGAAGATCAACTACACGA
nosZ-149	AGAAGATGAAGCTGCTGTACGACGCGTTCACCGAGCCGGAGCCGCACTACGCGCAGATCATCAAGGCCGA
nosZ-148	ACGACGCGTTCACCGAGCCGGAGCCGCACTACGCGCAGATCATCAAGGCCGACAAGCTCCACCCGATCGA
nosZ-150	AACACTGAGGAGGCGACCGGCCGTCTCGAGGTGACCGCGAGCCAGAAGGATCGTGACTACCTCGTGTTCG
nosZ-151	AAGTACCTCGTCTCGATGAACAAGATGTCGAAGGGCCAGCACCTGAACATCGGTCCCTCCGAGCCCGAAT
nxrB-001	TACCCGACACGTTGGGAAGACCAGACCAAGTATCGCGGCGGCTGGGTGGTTGACGGTCAGAGGCAGAAGA
nxrB-002	CCAGACCAAGTATCGCGGCGGCTGGGTGGTTGACGGTCAGAGGCAGAAGAGCCTTCGGCTGCGGCTGCAG
nxrB-003	CCGAACTGGGACGACGACCTCGGCGGTTCGCAGGTTTACGCCAACAACGACCCGAACTTCGACGGCGCCT

Table A2.3 continued.

Probe	Sequence
nxrB-007	ACGATCGAGGCGGGCCCGAACTGGGACGACGACCTCGGCGGTTCGCAGGTCTATGCGAACAACGACCCGA
nxrB-004	ACCAGAACCTGATCAATGCGCCGCTGGCTGACGAGCAGCCGACCGCGCGGGCGATCTCGATGGTGACCGG
nxrB-006	TCGAGCCGTGGACCTACGATTACCAGAACCTGATCAATGCGCCGCTGGCTGACGAGCAGCCGACCGCGCG
nxrB-005	CGATGGTGACCGGCAAGTACATGGACACGATCGAGGCGGGCCCGAACTGGGACGACGACCTCGGCGGTTC
nxrB-010	ACGAGGATCTCTTCAACGCGCCTGAAGGCGCAGACCAGCCAACCGCCCGACCAGTCTCGATGGTCACGAA
nxrB-012	AACGCGCCTGAAGGCGCAGACCAGCCAACCGCCCGACCAGTCTCGATGGTCACGAACCGGCCGATCGATA
nxrB-013	AACGGCACGACCGATCTCGATGGTCACGAACAGGCCGATCGATATCAATGCCGGGCCGAACTGGGATGAC
nxrB-014	CCAACGATCCCAATTTCGCAGCCCTCACTCCTGAGCAGCGCGAGGAGATATTCGAAGTAGAACGGCTGGT
nxrB-008	AGGCTCGCCGATCTATGCGGAGAACGATCCCAACCTGTCAGCGCTTTCGCCAGAACAGCGCGAACAGTTG
nxrB-009	CGGAGAACGATCCCAACCTGTCAGCGCTTTCGCCAGAACAGCGCGAACAGTTGTTCGCGACCGAGCGGTT
nxrB-011	CAGCCAAGCCAATCTCGATGATTACGGGCGAATACGTGAACATCGAGAGCGGGCCGAACTGGGATGACGA
nxrB-015	CGATGATCGATGGTGAGAAGATCGATGTGCAGGCCGGCCCGAACTGGGACGACGACTTGGGGGGATCGCC
nxrB-016	CTCACCGAGCAGCAAAAGTTGCAATTGAGTTCCGTCGAGCGGCTCGTCTTCTTCTACTTGCCGCGCATCT
nxrB-020	CGGTCTCGATGGTGACAGGTGACCTGATCCAGATCCAGTCCGGTCCGAACTGGGATGACGATCTTGGCGG
nxrB-021	GGTTCCGGTCTCGATGGTGACAGGTGACCTGATCCAGATCCAGTCCGGTCCAAACTGGGACGACGATCTT
nxrB-017	ACGAGCACTTGTTCACGGCGCCGCGGGAAGCGGATCAGCCAACCGCGCGGCCGATTTCCTTGATCAGCGG
nxrB-018	AGTCGGGGCCGAACTGGGACGACGACCTCGGCGGCTCGCCGGTCTATGCCGAGCGCGATCCGAATCTCGA
nxrB-019	ACTACGAGCACTTGTTCACGGCGGCGCAGGAAGCGGATCAGCCGACCGCCCGGCCGATTTCCCTGATCAG

Table A2.4. Environmental parameters for *Taihu* in 2014. EC = electric conductivity; TDS = total dissolved solids; NTU = turbidity; DO = dissolved oxygen; SD = Secchi depth

Month	Station	Water Temp (ºC)	Lake Depth (m)	EC (µS/cm)	TDS (g/L)	Salinity (‰)	pH	NTU	DO (%)	DO (mg/L)	SD (cm)
June	1	24.80	2.0	670	0.437	0.33	8.19	13.3	113.9	9.43	40
	7	25.32	2.6	608	0.392	0.29	8.31	21.8	113.5	9.31	30
	Channel	25.31	2.8	632	0.408	0.30	7.94	18.1	54.4	4.46	30
	Dock	27.87	1.2	712	0.439	0.33	8.77	26.4	128.7	10.08	15
July	1	26.03	2.4	620	0.359	0.29	8.75	20.8	106.7	8.64	30
	7	27.08	3.0	608	0.380	0.28	9.14	12.9	134.4	10.68	30
	10	26.14	1.8	602	0.383	0.28	8.93	33.4	47.5	3.84	30
	Channel	25.14	1.8	507	0.329	0.24	9.04	55.5	67.1	5.53	0
	Dock	27.26	1.5	665	0.415	0.31	8.90	143.8	98.3	7.79	0
Aug	1	26.70	2.5	543	0.342	0.25	9.57	57.9	82.6	6.60	35
	7	26.43	3.1	598	0.378	0.28	9.38	28.8	98.0	7.88	40
	10	26.68	2.2	517	0.326	0.24	9.20	45.0	36.9	2.95	32
	Channel	26.65	3.1	552	0.348	0.26	9.38	71.3	58.0	4.64	25
	Dock	26.47	1.5	557	0.352	0.26	8.98	51.5	29.1	2.33	40
Sept	1	27.37	2.6	542	0.337	0.25	10.03	28.2	126.2	9.98	40
	7	27.12	3.1	482	0.301	0.22	9.94	25.0	159.0	12.63	30
	10	27.26	2.1	424	0.264	0.19	9.93	27.4	71.2	5.64	35
	Channel	28.25	2.7	515	0.315	0.23	9.76	881.9	67.7	5.27	0
	Dock	27.71	1.6	544	0.336	0.25	8.51	24.1	74.8	5.84	30
Oct	1	21.29	2.4	492	0.344	0.26	9.38	34.2	68.6	6.07	30
	7	22.20	3.1	467	0.321	0.24	9.77	47.0	32.8	2.85	25
	10	21.70	2.0	479	0.322	0.25	9.21	85.9	21.9	1.93	10
	Dock	21.91	1.5	473	0.327	0.24	10.05	48.5	32.1	2.81	25

Table A2.5. Environmental parameters for *Taihu* 2014 continued. TN = total nitrogen; TDN = total dissolved nitrogen; TP = total phosphorus; TDP = total dissolved phosphorus; PC = phycocyanin; Chl *a* = chlorophyll a; MC = microcystin

Month	Station	TN (mg/L)	TDN (mg/L)	NH_4 (mg/L)	TP (mg/L)	TDP (mg/L)	PO_4 (mg/L)	TN:TP	PC (cells/mL)	Chl *a* (µg/L)	Total MC (µg/L)
June	1	2.97	2.50	0.077	0.092	0.028	0.005	32.17	2831	40.18	0.00
	7	2.38	1.79	0.076	0.077	0.010	0.012	30.83	2125	27.34	2.64
	Channel	10.77	3.89	0.827	0.728	0.093	0.069	14.80	1374	368.28	2.64
	Dock	17.23	1.83	0.076	1.176	0.029	0.006	14.65	11107	959.76	69.39
July	1	1.78	1.40	0.084	0.0602	0.0200	0.006	29.59	2482	18.75	0.45
	7	2.51	1.60	0.025	0.0761	0.0133	0.003	32.91	2139	48.79	0.85
	10	4.66	4.20	1.672	0.2273	0.1283	0.101	20.48	1425	15.25	45
	Channel	9.80	4.92	0.339	0.4995	0.0717	0.057	19.62	31956	487.69	2.96
	Dock	4.21	1.17	0.119	0.2333	0.0167	0.003	18.05	62212	119.33	2.24
Aug	1	1.38	0.71	0.094	0.162	0.080	0.050	8.52	22185	74.58	2.71
	7	1.10	0.76	0.050	0.067	0.017	0.001	16.43	4534	20.93	0.40
	10	3.73	2.97	1.155	0.343	0.138	0.112	10.89	2159	21.11	0.00
	Channel	21.52	1.29	0.283	2.132	0.061	0.036	10.09	36136	3257.60	111.57
	Dock	2.40	1.35	0.652	0.203	0.037	0.019	11.79	18561	71.61	2.53
Sept	1	1.33	0.52	0.070	0.1704	0.0746	0.060	7.79	10618	57.14	0.00
	7	39.33	0.67	0.162	4.2445	0.0452	0.027	9.27	15009	1204.72	75.04
	10	2.99	2.70	0.864	0.1687	0.0891	0.071	17.72	4357	27.83	0.00
	Channel	48.17	1.31	0.363	11.6860	0.0343	0.059	4.12	239256	16,182	592.69
	Dock	1.26	0.66	0.241	0.1483	0.0691	0.057	8.46	4022	61.66	0.47
Oct	1	1.37	0.77	0.097	0.141	0.064	0.054	9.70	4989	32.6	0.19
	7	20.34	0.62	0.131	1.905	0.037	0.026	10.68	13945	1507.6	0.62
	10	3.87	2.10	0.382	0.374	0.073	0.058	10.37	15407	97.5	0.24
	Dock	12.66	0.64	0.123	1.199	0.056	0.044	10.56	4267	1037.9	0.00

Table A2.6. Statistically significant dissimilarities (ANOSIM) between samples based on environmental parameters within a given month.

	R Statistics	*Significance Level (%)*
June and August	0.344	0.048
June and September	0.388	0.016
June and October	0.823	0.029
July and October	0.35	0.056
August and October	0.456	0.008
September and October	0.469	0.016

Table A2.7. Statistically significant dissimilarities (ANOSIM) between samples based on environmental parameters within a given month.

	R Statistic	*Significance Level (%)*
1 and Channel	0.375	0.016
1 and 10	0.681	0.008
7 and Channel	0.406	0.016
7 and Dock	0.436	0.024
7 and 10	0.913	0.008
Channel and Dock	0.356	0.016
Dock and 10	0.556	0.008

Table A2.8. Positive probes and their targeted specificity, including the targeted clade, sequences within that clade and the origin of sequences when available. The annotation of the probe is also provided from the best BLASTn hit. All probes were annotated as targeted gene of interest. Clades represented as 1X (ie. 1J) are referring to clades within the phylum Proteobacteria. The order of the positive probes in Figure 4 corresponds to the order in this table.

Probe	Targeted clade, sequences within that clade and origin of sequence	Best BLAST Hit Description	% Identity	E-value
AamoA-11	Two major subclades of corals #1 Np3 and Np7	Uncultured Thaumarchaeote clone MT1-A19	99	1.00E-25
AamoA-44	Uncultured bacteria within Sargasso Sea lineage	Uncultured Thaumarchaeote clone 5mT1C	100	3.00E-27
AamoA-91	Uncultured soil fosmid 54d9 lineage, broad specificity probe	Uncultured archaeon clone AOA-OUT9	100	3.00E-27
AamoA-92	Uncultured soil fosmid 54d9 lineage, broad specificity probe	Uncultured archaeon clone AOA - OUT9	100	3.00E-27
BamoA-06	*Nitrosospira* Nsp 65 clade	Uncultured bacterium clone AOB 93	100	3.00E-27
BamoA-02	Uncultured Ammonia oxidizing bacteria II, *Nitrosospira* NSP12, NSP5 cluster	Uncultured bacterium clone AOB 95	100	3.00E-27
BamoA-08	Uncultured ammonia oxidizing bacteria I, broad probe	Uncultured bacterium clone EB23 C10	100	3.00E-27
hzsA-001	*Brocadia fulgida* subgroup 1	Candidatus *Brocadia fulgida*	100	3.00E-27
hzsA-003	*Brocadia fulgida* subgroup 1	Candidatus *Brocadia fulgida*	100	3.00E-27
hzsA-035	*Brocadia fulgida* subgroup 5	Uncultured anaerobic ammonia oxidizing bacterium clone BL3	94	8.00E-18
hzsA-036	*Brocadia fulgida* subgroup 5	Anaerobic ammonium-oxidizing bacteria enrichment culture clone DC1-HZ-9	93	2.00E-14
hzsA-010	*Brocadia fulgida* subgroup 2	Uncultured anaerobic ammonium-oxidizing bacterium clone BL5	100	3.00E-27
hzsA-011	*Brocadia fulgida* subgroup 2	Uncultured anaerobic ammonium-oxidizing bacterium clone BL5	100	3.00E-27
hzsA-039	*Brocadia fulgida* subgroup 6	Uncultured anaerobic ammonium-oxidizing bacterium clone HSL5	100	3.00E-27
hzsA-032	*Brocadia fulgida* subgroup 8	Uncultured anaerobic ammonium-oxidizing bacterium clone L6	100	3.00E-27
hzsA-012	*Brocadia fulgida* subgroup 6 + 8 + 3	Anammox bacterium enrichment culture clone BS23	100	3.00E-27
hzsA-013	*Brocadia fulgida* subgroup 6 + 8 + 3	Anammox bacterium enrichment culture clone BS23	100	3.00E-27
hzsA-015	*Brocadia fulgida* subgroup 4 + 7 + 9	Uncultured anaerobic ammonium-oxidizing bacterium clone BL1	100	3.00E-27
hzsA-016	*Brocadia fulgida* subgroup 4 + 7 + 9	Uncultured anaerobic ammonium-oxidizing bacterium clone BL1	100	3.00E-27
hzsA-038	*Brocadia fulgida* subgroup 4 + 7	Anammox bacterium enrichment culture clone cc2A	100	3.00E-27
hzsA-042	*Brocadia fulgida* subgroup 9	n/a		3.00E-27

Table A2.8 continued.

Probe	Targeted clade, sequences within that clade and origin of sequence	Best BLAST Hit Description	% Identity	E-value
hzsA-004	*Brocadia anammoxidans*, n-DAMO clade	Anammox bacterium enrichment culture clone BS18	100	3.00E-27
hzsA-005	n-DAMO clade	Anammox bacterium enrichment culture clone BS10	100	3.00E-27
hzsA-006	n-DAMO clade	Anammox bacterium enrichment culture clone BS6	100	3.00E-27
hzsA-022	*Jettenia* subgroup 2	Candidatus *Jettenia asiatica*	100	3.00E-27
nifH-003	1B; Cosmopolitan marine clade #1; *Cyanothece* related sequences	Uncultured bacterium	100	3.00E-27
nifH-008B	1B: Cosmopolitan marine clade #1; *Cyanothece* related sequences, *Nostoc, Anabaena*	*Dolichopermum circinale* FSS-124	100	3.00E-27
nifH-005	1B: *Acaryochloris* and *Synechococcus*	Uncultured bacterium clone nif1004424U95		3.00E-27
nifH-560	1B: Uncultured salt pond microbal mat, marine sediment, alpine glacier, mangrove sediment bacteria	Uncultured microorganism clone 20126A20	100	1.00E-21
nifH-551	1B; Marine sponge isolate	Uncultured microorganism clone 20107A04	100	1.00E-21
nifH-548	1B; Intertidal microbial mat cyanobacteria	Uncultured microorganism clone 20125A18	100	1.00E-21
nifH-008	1B; *Anabaena, Nostoc, Cylindrospermopsis* clade	*Nostoc piscinale* CENA21 genome	100	3.00E-27
nifH-549	1B; *Cylindrospermopsis*	Uncultured bacterium	100	1.00E-21
nifH-573	1B; *Anabaena, Cyanothece*	Uncultured bacterium clone N328	100	1.00E-21
nifH-574	1B; *Anabaena* + *Cyanothece*	Uncultured bacterium clone K92.nifH.67	100	1.00E-21
nifH-550	1B: Sequences sequence related to *Anabaena*	Uncultured marine bacterium clone HT43A2 t7rose	100	1.00E-21
nifH-009	1B; *Trichodesmium*	Uncultured bacterium clone M1470A02	100	3.00E-27
nifH-545	1B; Uncultured Great Barrier Reef	Uncultured microorganism clone 20115A40	10	1.00E-21
nifH-516	1G; Sequences related to *Azotobacteri vinelandii/chroococcum*	Uncultured marine bacterium clone ETSP2 50254A74	100	1.00E-21
nifH-041	1G; Uncultured bacteria from sandy intertidal beach, rhizosphere	Uncultured bacterium clone II-5, 102b	97	6.00E-24
nifH-043	1G, 1F; Uncultured bacteria from sandy intertidal beach; 1O; Uncultured from rhizosphere	Uncultured bacterium clone III-1.73b	97	6.00E-24
nifH-042	1G, 1F; uncultured from sandy intertidal beach	Uncultured bacterium clone II-5.102b	100	3.00E-27
nifH-057	1O-5 - SPOT#1	Uncultured microorganism clone GoMnifH 06	100	3.00E-27

Table A2.8 continued.

Probe	Targeted clade, sequences within that clade and origin of sequence	Best BLAST Hit Description	% Identity	E-value
nifH-045	1O: *Thiorodovibrio*	Uncultured bacterium clone III-1.86b	100	3.00E-27
nifH-520	1G + 1O: Mangrove rhizosphere clade #1, uncultured bacteria from rhizosphere, and South Pacific Gyre	Uncultured marine bacterium clone ETSP DNA 44879A55	100	1.00E-21
nifH-020	1P: uncultured bacteria from rhizosphere and sandy intertidal beach	Uncultured bacterium clone MDE elv 20e10	97	6.00E-24
nifH-568	1G: Uncultured marine bacteria from ALOHA station	Uncultured bacterium clone u_HY39A108	100	1.00E-21
nifH-050	1J; *Rhodopseudomonas, Rhodobacter, Rhodospirillum, Magnetospirillum* also including *Rhodovolum* and *Phaeospirillum* and uncultured microbes from various agricultural soils, roots of *Oryza sativa*, saline-alkaline soil	*Rhodobacer sphaeroides* strain MBTLJ-13	100	3.00E-27
nifH-051	1J; Uncultured bacteria from various agricultural soils, roots of *Oryza sativa*, saline-alkaline soil	Uncultured bacterium clone ND184	100	3.00E-27
nifH-054	1J: Uncultured marine surface clade #1	Uncultured marine bacterium clone A0-0M-17	100	3.00E-27
nifH-046	1J, 1K; Very broad probe. Uncultured from marine, agricultural soil, saline-alkaline soils	Uncultured nitrogen-fixing bacterium clone SG5A12	100	3.00E-27
nifH-047	1J, 1K; Very broad probe. Uncultured from marine, agricultural soil, saline-alkaline soils	Uncultured bacterium	100	3.00E-27
nifH-048	1J, 1K; Very broad probe. Uncultured from marine, agricultural soil, saline-alkaline soils	Uncultured bacterium clone N 224IT 021	100	3.00E-27
nifH-517	Span a number of the 1X clades - VERY patchy	Uncultured marine bacterium clone ETSP DNA 43976A3	100	1.00E-21
nifH-530	1K: Norwegian Fjord clade #1, uncultured from soil	Uncultured marine bacterium clone 12BF	100	1.00E-21
nifH-060	1O: Mangrove rhizosphere clade #1	Uncultured marine bacterium clone 42030A08	97	6.00E-24
nifH-032	1A; Uncultured bacteria from sandy intertidal beach	Uncultured bacterium clone I-5.96b	100	3.00E-27
nifH-531	1K; Norwegian Fjord clade #1, uncultured *Bradyrhizobium, Rhizobium*, and uncultured bacteria from rhizosphere and agricultural soils	Uncultured marine bacterium clone 12BF	100	1.00E-21

Table A2.8 continued.

Probe	Targeted clade, sequences within that clade and origin of sequence	Best BLAST Hit Description	% Identity	E-value
nifH-532	1K: Uncultured *Pseudomonas*, *Bradyrhizobium*, *Burkholderia,* from agricultural soil, seawater, rood nodules of legumes	Uncultured marine bacterium clone 12BF	100	1.00E-21
nifH-501	1K: sequences from the Baltic Sea 3m Landsort deep St. BY31	Uncultured bacterium clone 46402A13	100	1.00E-21
nifH-019	1P: Uncultured *Siderooxydans, Azoarcus,* bacteria from mangrove rhizosphere, agricultural soils, root tissue of various plants	Uncultured soil bacterium clone T15	100	3.00E-27
nifH-506	1P: Chesapeake Bay phylotype - *Dechloromonas aromatica*	Uncultured bacterium clone M1647A18	100	1.00E-21
nifH-035	1A - Uncultured bacteria from sandy intertidal mat	Uncultured bacterium clone I-5.92b	100	3.00E-27
nifH-062	1P; Uncultured bacteria from South Pacific Gyre #2	Uncultured bacterium clone HH1840	100	3.00E-27
nifH-037	1A, 1J, 1K, 1G: Uncultured from alkaline soil bacterium	Uncultured bacterium clone nifH19	100	3.00E-27
nifH-063	1A: patchy coverage	Uncultured bacterium clone P2-8	100	3.00E-27
nifH-030	1A	Uncultured microorganism clone P03_20	100	3.00E-27
nifH-029	1A: Uncultured bacteria from various soils and rhizosphere, freshwater	Uncultured bacterium clone_u CB907H89	99	1.00E-25
nifH-507	1G: one sequence, unknown	Uncultured bacterium clone 46426A2	100	1.00E-21
nifH-023	1C; Western channel observatory clade #1	Uncultured marine bacterium clone inside15	100	3.00E-27
nifH-510	3A; Uncultured bacteria from marine sponge	Uncultured bacterium clone 46413A13	100	1.00E-21
nosZ-019	*Bradyrhizobium*	Uncultured bacterium clone J25	100	3.00E-27
nosZ-043	Uncultured lake sediment clade #1; paddy soil	Uncultured bacterium clone DMSb2-2514	100	3.00E-27
nosZ-001	*Azospirillum* #1; Uncultured *Azospirillum* from paddy soil	*Azospirillum* sp. TSA36t	100	3.00E-27
nosZ-002	*Azospirillum* #1, *Achromobacter* and Coastal marine sediment clade #1 , uncultured members from paddy soil and other agricultural soils	Uncultured bacterium isolate DGGE gel band K	100	3.00E-27
nosZ-048	Salt marsh clade #1; Uncultured soil bacteria	Uncultured bacterium clone IGAZ29	100	3.00E-27
nosZ-159	Coastal Marine Sediment clade #1; Uncultured marine sediment and forest soil	Uncultured denitrifying bacterium clone86	100	3.00E-27

Table A2.8 continued.

Probe	Targeted clade, sequences within that clade and origin of sequence	Best BLAST Hit Description	% Identity	E-value
nosZ-057	Agricultural Soil clade #2	Uncultured bacterium clone ZA8	100	3.00E-27
nosZ-055	*Herbaspirillum* #1, #2 and rice soil clade #1	*Herbaspirillum* sp TSO45-3	100	3.00E-27
nosZ-061	Paddy soil clade #1	Uncultured bacterium clone ISA00348	100	3.00E-27
nosZ-068	Lake sediment clade #1 and *Aromatoleum* related	Uncultured bacterium clone 5_26	100	3.00E-27
nosZ-070	*Azospirillum largimobile;* paddy soil and agricultural soil	Uncultured bacterium clone OGF_Z8	100	3.00E-27
nosZ-077	Agricultural soil clade #4	Uncultured bacterium clone Z30O42(3)	100	3.00E-27
nosZ-079	Activated sludge clade #1	Uncultured denitrifying bacterium isolate DGGE gel band Z27	100	3.00E-27
nosZ-080	Lake sediment clade #3	Uncultured bacterium clone DMZb2-2817	100	3.00E-27
nosZ-152	Activated sludge clade #1	Uncultured bacterium clone MWWT94	100	3.00E-27
nosZ-064	*Burkholderia*	*Burkholderia pseudomallei* strain VB976100 chromosome 1	100	3.00E-27
nrfA-01	*E.coli*	*Escherichia coli* strain NCTC 13441	100	3.00E-27
nrfA-02	*E.coli*	*Escherichia coli* strain NCTC 13441	100	3.00E-27
nrfA-03	*E.coli*	*Escherichia coli* strain NCTC 13441	100	3.00E-27

Supplementary Information

Microarray development

Seven functional marker genes were selected, corresponding to 6 key steps in the microbial N cycle. Long oligonucleotide (70 - mer) probes were designed to cover broad clades of the functional genes to the maximum extent possible. To design probes, all available data in GenBank was gathered, including pulling out specific genes from genomes, that were identifiable by their gene name. This was followed by phylogenetic analysis in ARB. Neighbor Joining phylogenetic trees were used to identify clades and probes were designed to target larger clades. Where possible, hierarchically nested and multiple probes were designed (Loy et al. 2002; http://aem.asm.org/content/68/10/5064.abstract). Selected 60 - mer *nifH* probes were also included from the MicroTools microarray (Shilova *et al.* 2014). The final NC array contains 691 probes spotted in 4 replicates (Table 2). These include a normalisation control targeting an externally spiked PCR product (*E.coli hyaB*); 5 negative controls with broad coverage of bacterial and archaeal 16s rDNA; 159 *amoA* probes; 42 *hzsA* probes; 144 *nifH* probes; 182 *nosZ* probes; 138 *nrfA* probes; and 21 *nxrB* probes. Based on literature consensus from a number of long oligonucleotide microarray works (Deng *et al.* 2008; Vilchez-Vargas 2013; Hughes *et al.* 2001; Shilova *et al.* 2014), we expected probes with up to 5 weighted mismatches to give clear positive signals and probes with over 10 weighted mismatches to be negative; while probes with 5 - 10 weighted mismatches were expected to give varying results depending on their exact sequence. Validation with seven environmental clones from five different functional markers confirmed these overall specificity expectations (Figure 9). Out of 27 probe-target pairs with < 5 wMM, no false negative hybridisations were detected. Out of 4747 probe-target pairs with > 10 wMM, only 4 false positive hybridisations were detected (with 10.2, 10.9, 12.0 and > 15 wMM, respectively).

Due to the intrinsically complicated issue of exact specificity predictions for long oligonucleotide arrays, results were interpreted with extreme care, considering nested and multiple probes to increase confidence. Mismatch weights for basepairs were as follows (probe-target bases): CC, CT, CA 1.5; GA 1.0; GT 0.6; AG 0.5; TG 0.2; all other mismatched basepairs 1.1. Mismatches were further weighted according to their position using the algorithm:

$$\exp\left[\left(- \ln (10) / 3\right) \times \left((2 \times \text{Pos-Length}) / \text{Length}\right)^2 \right]$$

Microarray printing

Oligonucleotides for immobilization were custom synthesized (Integrated DNA Technologies, Coralville, Iowa, USA). A 384 well flat bottom plate was prepared with 30 µl of 50 µM oligonucleotide solutions in ArrayIt spotting solution (ArrayIt Corp., Sunnyvale, CA, USA). Samples were spotted with a NanoPrint spotter, using a single 946MP3 pin (ArrayIt Corp., Sunnyvale, CA, USA) at 70 % relative humidity at room temperature onto Corning® GAPS™ II Coated Slides (Corning Incorporated, New York, USA). Before spotting, slides were incubated for 4 h in the spotter at 70 % relative humidity. Arrays were always spotted in triplicate to enable a statistical correction for errors (SI 7). Spotted slides were given a UV dose of 65 000 µJ in a Spectrolinker UV Crosslinker (Spectroline, New York, USA) to assist binding of the probes to the slide surface then stored at room temperature and at low humidity in the dark prior to use. Slides were processed on the day of hybridization and all solutions were prepared fresh immediately prior to processing. Slides were rinsed in 5 × SSC, 0.1 % lauroylsarcosine wash solution followed by 15 min incubation in blocking solution (3.6 g succinic-anhydride in 200 mL methyl-pyrrolidinone and 8.96 mL 1 M sodium borate buffer pH 8.0) at room temperature in the dark with vigorous agitation to remove the unbound DNA. Slides were rinsed five times with Milli-Q water (dH_2O),

incubated 2 min in 100 °C dH_2O (without agitation) and incubated 1 min in - 20 °C absolute (100 %) ethanol (without agitation). Slides were then dried individually with an oil-free air gun.

Microarray target preparation and hybridisation

Amplification of partial fragments of selected N-cycle functional marker genes was achieved via single PCR using PCR primers and annealing temperatures and cycle numbers shown in Table 1. The *hzsA* fragment was amplified via a nested protocol (Harhangi *et al.* 2012). PCR amplifications were carried out in 96 well plates, with 25 µL volumes, and contained 1 x GoTaq mix (Promega), 40 nM of forward primer, 0.1 µL of 50 ng / µL molecular grade BSA (Promega) and 10 ng environmental DNA. Amplification reactions for *nifH*, *nrfA* and *nxrB* contained 40 nM of the corresponding reverse primers with a T7 RNA polymerase promoter site at the 5' end. Amplification reactions for *hzsA*, *AOA*, *AOB* and *nosZ* contained 10 nM of the corresponding reverse primers with a T7 promoter tag at the 5' end plus 30 nM of the T7 promoter as a second reverse primer. Following amplification PCR products were pooled and purified using AMPure-XP PCR purification (Beckman Coulter, USA). Purified, pooled PCR products (150 ng) were labelled via *in vitro* transcription as described elsewhere (Stralis-Pavese *et al.* 2011). Briefly, PCR products were transcribed into Cy3 labelled RNA via T7 RNA polymerase and fragmented via chemical fragmentation to ~ 50 nt fragments. A rotary hybridization oven and conventional hybridization tubes were preheated for at least 2 h at 55 °C. HybriWell (Grace BioLabs) adhesive hybridization chambers (custom made, containing 3 chambers per slide, 100 µL each in volume, order number 46170) were applied onto the microarray slides. Slides were preheated to 55 °C on a dry heating block. For each hybridization, 24 µL DEPC-treated water, 40 µL formamide, 30 µL 20 x SSC, 1 µL 10 % SDS, 2 µL 50 x Denhardt's reagent (Sigma) and 1 µL salmon sperm DNA (10 mg / mL) were added to a 1.5 mL Eppendorf tube and incubated at 55 °C for 1 - 15 min.

Two µL of unpurified denatured PCR target kept on ice was added to the pre-warmed hybridization buffer immediately prior to loading into HybriWells. Chambers were sealed with seal spots (Grace BioLabs) and slides were placed into preheated, conventional hybridization tubes and incubated overnight at 55 °C in a rotary hybridization oven at lowest rotation setting (approximately 10 rpm). Following hybridization, HybriWell chambers were individually removed and slides were immersed immediately into 2 x SSC, 0.1 % (w / v) SDS pre-warmed to 55 °C. Washing was performed in four successive steps for 5 min each : 2 × SSC + 0.1 % SDS at 55 °C; 0.2 x SSC (two times) and 0.1 × SSC at room temperature. Slides were dried individually with an oil-free air gun. Slides were stored at room temperature in the dark and scanned the same day.

Microarray analysis

Hybridized slides were scanned at 10 µm resolution with a GenePix 4000B laser scanner (Molecular Devices, Sunnyvale, CA, USA) at wavelengths of 532 nm and 635 nm for Cy3 and Cy5, respectively. Fluorescent images were captured as multi-layer tiff images and analyzed with the GenePix Pro 6.0 software (Molecular Devices). Hybridization signal for each probe was expressed as the F635 / B635 ratio where F635 is the median signal and B635 is the median background signal for the spot. Signals were corrected for background by deducting the signal of a negative control probe (targeting archaeal 16S rRNA genes), and then normalized to a positive control probe, hyaBP60, targeting the *hyaB* gene of *E. coli*, used as a spike added to the Klenow labeling reaction. Final results were expressed as average percentage of the hyaBP60 control signals.

16S PCR conditions PCR condition for checking genomic DNA contamination in RNA samples were as follows: 94° C for 2 min, 30 cycles of 94° C for 1 min, 60° C for 1 min 72° C for 1min, and a final elongation step of 72° C for 10 min

Chapter 3 : Cyanobacterial colonies are microniches that are functionally distinct from the free-living community

Publication note

This chapter serves as a pilot study for future work on Lake Taihu, China by Lauren E. Krausfeldt and Steven W. Wilhelm.

My contributions to this work included the sample collection, RNA extraction, bioinformatic analysis and writing most of the manuscript.

Abstract

Colony formation has been proposed to be advantageous to *Microcystis* spp. for several reasons, including to deter grazers, increase buoyancy, and provide protection from environmental stressors. Further, *Microcystis* colonies are able to form tight associations with heterotrophic bacteria and these interactions may help promote the success of *Microcystis*. However, the underlying mechanism for colony formation is not well understood because *Microcystis* in culture tends to lose its ability to form colonies. Here, metatranscriptomes were generated from fractionated samples representing "colonies" and "free-living" communities from Lake Taihu, a hypereutrophic lake in China, to assess how *Microcystis* and the co-occurring phycosphere were different in colonies relative to when these cyanobacteria are "free-living". Data support the general observation that colony formation by *Microcystis* is a stress response. Many genes involved in responses to high temperature, nutrient limitation and phage infection, were upregulated in the free-living *Microcystis* compared to those within the colonies. Differences in the expression of transporters for N and P nutrients were also observed in the phycosphere, dominated by Proteobacteria, suggesting there is a functional niche separation between the colonial and free-living community members. This study suggests colonies are a hotspot for nutrient transformations and points to colony formation as another advantageous mechanism for the success of *Microcystis* in freshwater ecosystems around the world.

Introduction

The strategies that *Microcystis* employs to outcompete other phytoplankton and cyanobacteria in freshwater ecosystems are multifaceted. *Microcystis* can adapt to a wide range of conditions that may be unfavorable to other bloom-formers. For example, dissolved CO_2 concentrations are drawn down during blooms and the pH increases to conditions that may not favor growth of other phytoplankton (Shapiro, 1997, Yamamoto & Nakahara, 2005). *Microcystis* spp. and other cyanobacteria are particularly well equipped to handle these low CO_2 conditions because of their efficient carbon concentrating mechanisms (Sandrini *et al.*, 2014, Beardall & Raven, 2017). *Microcystis* is also able to handle periods of low inorganic phosphorus (P) and nitrogen (N) availability because they possess high affinity transporters and an ability to assimilate organic N and P sources (Harke *et al.*, 2012, Belisle *et al.*, 2016). Further, *Microcystis* can store high amounts of these nutrients (P in the form of polyphosphate bodies, and N as cyanophycin) (Kromkamp, 1987). Finally, *Microcystis* spp. produce gas vesicles that allow for vertical movement through the water column. Buoyancy provides *Microcystis* the ability to maintain a position in the water column with optimal light availability or avoid times of high irradiation (Ibelings *et al.*, 1991), but the added benefit of scavenging nutrients throughout the water column (*i.e.,* CO_2 near the surface or P or N near the sediments) (Bormans *et al.*, 1999).

Another potential factor in the success of *Microcystis* is its ability form colonies. Colony formation has been suggested to provide protection from chemical stressors, high-light conditions, grazing by predators, as well as to enhance photosynthetic activity and buoyancy (Xiao *et al.*, 2018). Drivers of colony formation are not clear and the literature investigating the effect of predation, temperature, nutrients and other factors, suggests control on colony formation is more complex than regulation by one factor. Moreover, variables that induce colony formation may even

differ between species (Xiao *et al.*, 2018). Understanding colony formation is further complicated by the fact that *Microcystis* tends to lose its ability to form colonies in laboratory culture, despite its common persistence as colonies in nature (Reynolds *et al.*, 1981, Yang *et al.*, 2008)

Another potential benefit *Microcystis* gains from colony formation is the tight association with heterotrophic bacteria within the exopolysaccharide mucilage surrounding the colony (Shen *et al.*, 2011, Xiao *et al.*, 2018). The mucilage associated with *Microcystis* colonies harbors diverse groups of bacteria in the phyla Proteobacteria, Actinobacteria, Bacteriodetes, and others (Tang *et al.*, 2010, Shi *et al.*, 2012, Steffen *et al.*, 2012, Penn *et al.*, 2014, Chen *et al.*, 2018). These associated bacteria, often referred to as the phycosphere, carry with them various functional capabilities and may be involved in nutrient recycling within the colony. Many studies have characterized the functional potential of heterotrophic bacteria in a bloom using molecular techniques (*i.e.,* metagenomes, metatranscriptomes) (Shia *et al.*, 2010, Steffen *et al.*, 2012, Parveen *et al.*, 2013, Penn *et al.*, 2014, Chen *et al.*, 2018, Shi *et al.*, 2018). Roles for these bacteria in P exchange, N-cycling and organic matter degradation have been proposed (Jiang *et al.*, 2007, Steffen *et al.*, 2012, Krausfeldt *et al.*, 2017, Shi *et al.*, 2018). Further, there are distinct differences in community composition between *Microcystis* colonies compared to the free-living microbial community within blooms (Parveen *et al.*, 2013, Shi *et al.*, 2018). Certain heterotrophic bacteria have even been suggested to stimulate colony formation (Shen *et al.*, 2011, Wang *et al.*, 2016) implying the members of the phycosphere likely benefit from this relationship as well.

There have been several studies that have characterized the community, functional potential and gene expression within *Microcystis* colonies, however to date no studies have compared gene expression within the colony and phycosphere to that of the free-living community. The goal of this study was to broaden the understanding of colony formation by *Microcystis*, the

advantages of being within the colony and assess activity of the co-occurring bacteria that may benefit *Microcystis*. To do this, a comparison of physiological differences between the two communities was performed using metatranscriptomes generated from size-fractionated samples collected during a *Microcystis* bloom in hypereutrophic Lake Tai (*Taihu*), China. The data presented here is hypothesis generating and serves as pilot study to plan for future field studies examining the role of colony formation, associated bacteria, and viruses in the success of *Microcystis*

Methods

Sample collection and RNA extraction

Surface water samples were collected from a dock that stretches approximately 100 yards into Meilang Bay near the TLLER field station on *Taihu* on 06/23/2016 at 13:30 and 06/24/2016 at 5:00 and 13:30. To enrich for colonies, water was filtered through a 28 μm mesh while the < 28 μm filtrate was captured as the representative "free-living" community. Duplicate samples were collected at each time point and both fractions were filtered through Sterivex, preserved with RNA*later* and stored at -80°C until processing. RNA extraction and DNA removal were performed using the MoBio (now Qiagen) DNA isolation kit for Sterivex™, modified for RNA, as previously described in Krausfeldt *et al*. (2017) on one set of replicates. The smaller fractionated samples (< 28μm) were difficult to extract due to low biomass and the presence of sediments.

RNA yields were very low, so a different protocol using a Phenol:Chloroform protocol developed by the Moran lab and modified (manuscript in preparation, Lauren Quigley) on the second set of replicates in attempt to achieve higher yields and is described as follows. RNA*later* was removed from Sterivex™ as previously described (Krausfeldt *et al.*, 2017) and the bottom of

the cartridge was cut with a pipe cutter to open the Sterivex™. The filter was removed carefully using a razor and added to a 2 mL screw cap bead tube. 750 μL of CTAB extraction buffer and 750 μL of phenol:chloroform:isoamyl alcohol (25:24:1, pH 6.7) was added to the bead tube. Tubes were inverted several times and vortexed at maximum speed for 10 min. Tubes were then centrifuged at 4 °C at 5000 rpm for 10 min. The upper aqueous phase was removed and added to a fresh tube and mixed well with the addition of chloroform:isoamyl alcohol (24:1). Tubes were centrifuged again (4 °C at 5000 rpm for 10 min). The upper aqueous phase was removed and transferred to a solution of 0.5 M $MgCl_2$ (final concentration 2.5mM) and 0.1 volumes of 3 M Na-Acetate. Finally, 0.7 volumes of -20 °C isopropanol was added last. Tubes were incubated for at least 1 hour or overnight at -80 °C. Following incubation, tubes were centrifuged for 45 min (4 °C at 10000 rpm) and the supernatant was carefully discarded. The pellet was washed with 400uL of 70% -20 °C RNase-free ethanol and centrifuged again at 4 °C at 10000 rpm for 10 min. This was repeated 2-3 times. The pellet was then air died at 37 °C for 2- 5 min and resuspended in 50 uL of RNase free water. DNase treatment was performed using the Turbo DNase kit (Invitrogen). Concentrations and DNA checks were performed as previously described. While the protocol for RNA extraction yielded slightly higher RNA concentrations in the “free” samples (Table 1), the RNA extraction from the kit had higher quality scores as assessed by the Bioanalyzer (Figure 1).

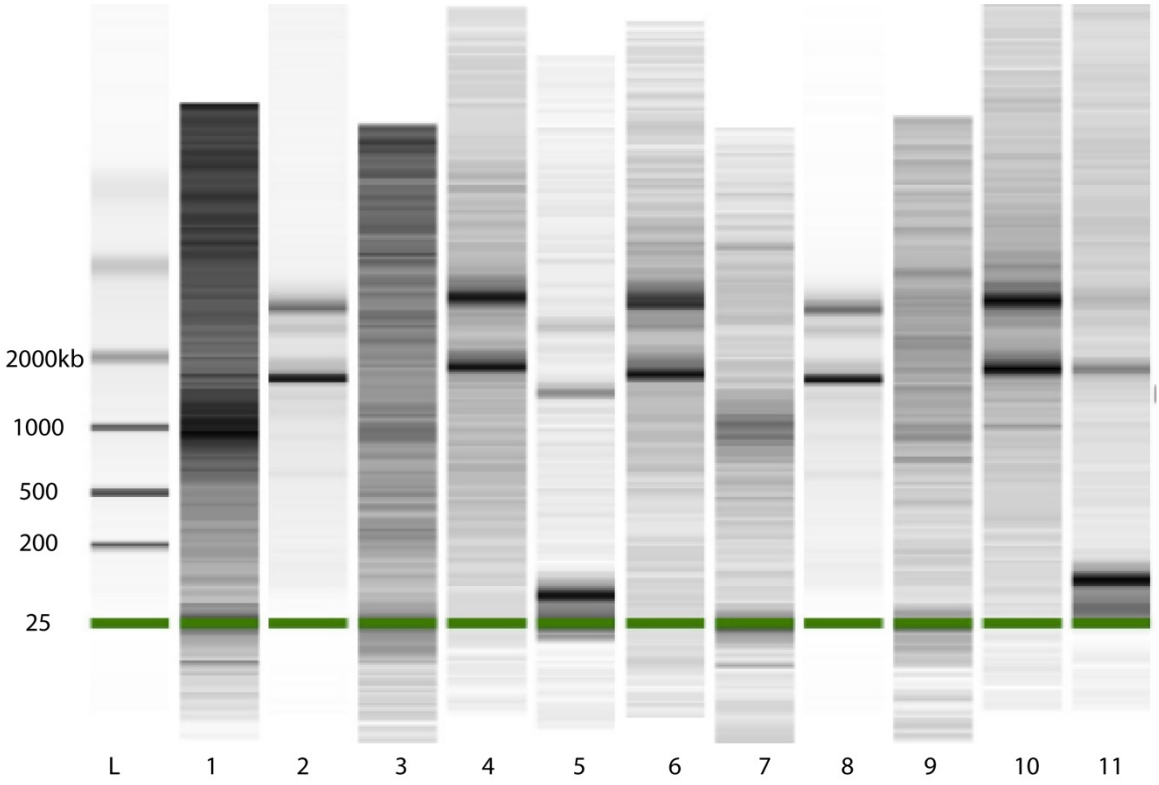

Figure 3.1 Bioanalyzer results for samples with low RNA yields extracted from Sterivex kit versus phenol:chloroform method

Only "free" samples were run through the bioanalyzer to determine which method provided the best quality of low quantity RNA. Odd numbered lanes are samples extracted by phenol:chloroform. Even numbered lanes are samples extraction by the kit. "L" is the ladder.

Table 3.1 RNA yields from extractions performed with Sterviex kit versus phenol:chloroform method

Time	Sample	Total RNA (μg) kit	Total RNA (μg) P:C
6/23/16 13:30	Colony	14.19	37.98
6/23/16 13:30	"Free"	0.26	0.24
6/24/16 5:00	Colony	20.01	21.26
6/24/16 5:00	"Free"	0.27	0.56
6/24/16 13:30	Colony	20.75	52.96
6/24/16 13:30	"Free"	0.73	0.47

To stay consistent with extraction efficiencies across methods, only one set of replicates was used for sequencing efforts.

Sequencing and analysis

Purified RNA was sent to Hudson Alpha for subsequent library prep, ribosomal RNA reduction, and sequencing on the Illumina NextSeq platform. Despite the use of an rRNA reduction kit (Epidemiology Ribo-Zero Gold rRNA Removal Kit), 80-90% of the total RNA reads were still rRNA and were removed *in silico* using SortMeRNA (http://bioinfo.lifl.fr/RNA/sortmerna/). Reads were trimmed, filtered for quality and assembled in CLC Genomics Workbench using default parameters. For annotation of the resulting contigs, data was uploaded to MGRAST. The quality filtered mRNA reads were also uploaded to the MGRAST server. Functional and taxonomic annotation for both contigs and reads was performed using the SEED database and Refseq, respectively. Upon completion, functional annotations of reads and accompanying counts were downloaded from MGRAST, normalized by library size as a proxy for community gene expression and analyzed using Primer-E v7.

Analysis of *Microcystis* gene expression from both colonial and free samples was performed in CLC Genomics Workbench. Normalized gene expression values (TPM) were calculated by recruitment of mRNA reads to the genome of *Microcystis aeruginosa* NIES 843 using default parameters (a similarity fraction and length fraction of 0.8). Differential gene expression for *Microcystis* was determined using CLC Genomics Workbench and the *Microcystis aeruginosa* NIES 843 for a reference genome. For validation of the gene expression metrics used in this study, TPM values for *Microcystis* were averaged for both colony samples and free samples and plotted against each other. This was also performed using averaged values from reads/library

size for the same housekeeping genes specific to *Microcystis.* Differential gene expression for the *Microcystis* phage were performed using the same metrics in CLC and using the *Microcystis* phage LMM-01 genome as the reference. Differentially expressed genes reported for both *Microcystis* and phage had a p value of < 0.05 after correction for false discovery rate (FDR). Because normalization of reads in CLC is based on total number of reads (TPM), reads that mapped to the phage genome based on the parameters described above were extracted and also normalized to the host housekeeping gene, *Microcystis rpoB*. Total reads that mapped to the phage genome and reads that mapped to the genes encoding *Microcystis* phage tail sheath protein (*gp091*) and capsid protein (*gp086*) in each sample were normalized by *Microcystis* specific *rpoB* (determined using the read mappings described above).

Results

Size - fractionated samples were collected from *Taihu* during a bloom event to compare activity of the microbial community within *Microcystis* colonies ("colonies") versus the free-living ("free") portion of the community. RNA yields for the samples pre-filtered through 28 μm mesh were low (Table 1), but RNA sequencing was successful for two of the free-living samples and all three colony-associated samples (Table 2). These were collected at different times, so to determine if these samples could be considered as biological replicates for statistical power, the samples were clustered based on transcript abundances normalized by library size as a proxy for community gene expression (Figure 2a). Samples from the colony and the free clustered together based on this metric. Because *Microcystis* gene expression was of particular interest, this was also confirmed by evaluating similarity as TPM values for *Microcystis* genes (Figure 2b). Based on these results, the

Table 3.2 Final mRNA reads counts after *in silico* rRNA reduction

Time	Sample	Total reads	*Microcystis* reads	Non-*Microcystis* reads
6/23/16 13:30	Colony 1	5,458,466	2,018,704	3,439,762
6/24/2016 13:30	Colony 2	4,457,222	1,440,763	3,016,459
6/24/2016 5:00	Colony 3	5,054,116	1,788,337	3,265,779
6/23/2016 13:30	Free 1	1,916,187	89,330	1,826,857
6/24/2016 13:30	Free 2	3,485,440	274,662	3,210,778

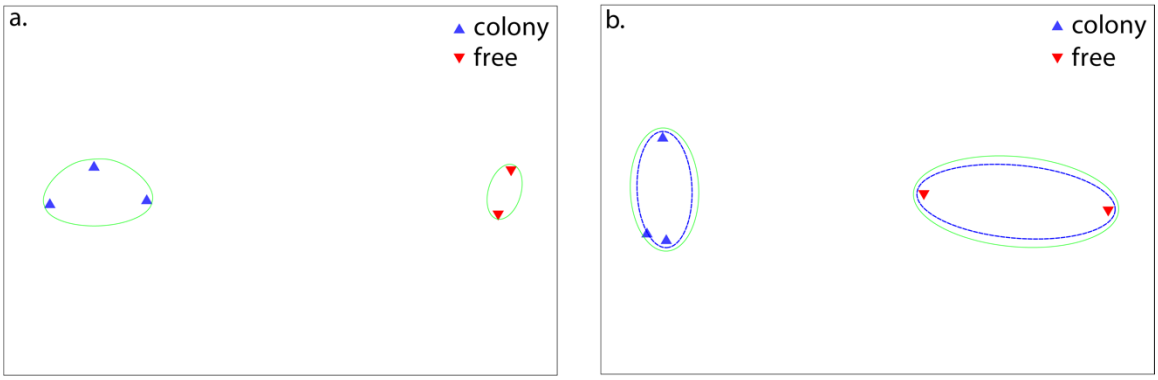

Figure 3.2 NMDS showing relationships between samples based on gene expression

Genes expression was determined by all annotated read abundances within the microbial community (a) and TPM calculated from read recruitments to the *Microcystis aeruginosa* genome (b). 2D stress = 0; green circles = 85% similarity

colony samples and the free samples were grouped to be used as replicates for statistical power and to support observations in the gene expression of the community. To validate normalization techniques and expression values, the relationships between housekeeping genes from *Microcystis*, one of the dominant taxa in the community, were determined between the free and colony samples. Both metrics for gene expression and normalization displayed a linear relationship between the colony and free with an R^2 value of >0.91 (Figure 3).

Phylogenetically, cyanobacteria produced the most transcripts in all samples (Figure 4a), making up 92%-95% of the total transcripts in the colonies and 42%-45% in the free samples. On average, 76% of the cyanobacteria transcripts originated from the Order Chroococcales (data not shown). In the colony and the free, transcripts assigned to *Microcystis* spp. and *Cyanothece* spp. were the two most abundant cyanobacterial genera, respectively (Figure 4b). Because it is difficult to obtain accurate genus level information from short transcripts, annotations were analyzed at the phylum level. Functional annotation of the cyanobacterial transcripts suggested that the cyanobacterial members of the community were actively scavenging both N and P (Figure 5 and 6). Most of the nitrogen metabolism transcripts were associated with ammonia, nitrate and urea transport. Urea transport transcripts from cyanobacteria were comparable across the colony and free samples, but there was a slightly higher relative abundance of nitrate and ammonium transport transcripts in the colony. The majority of the P metabolism transcripts were involved in the transport and scavenging of phosphate (*i.e*. *PstSCBA*, *PhoUHBR*) (Figure 6). Notably, there were low transcript abundances for cyanobacterial alkaline phosphatases in all samples, and polyphosphate utilization machinery was expressed only in the free group.

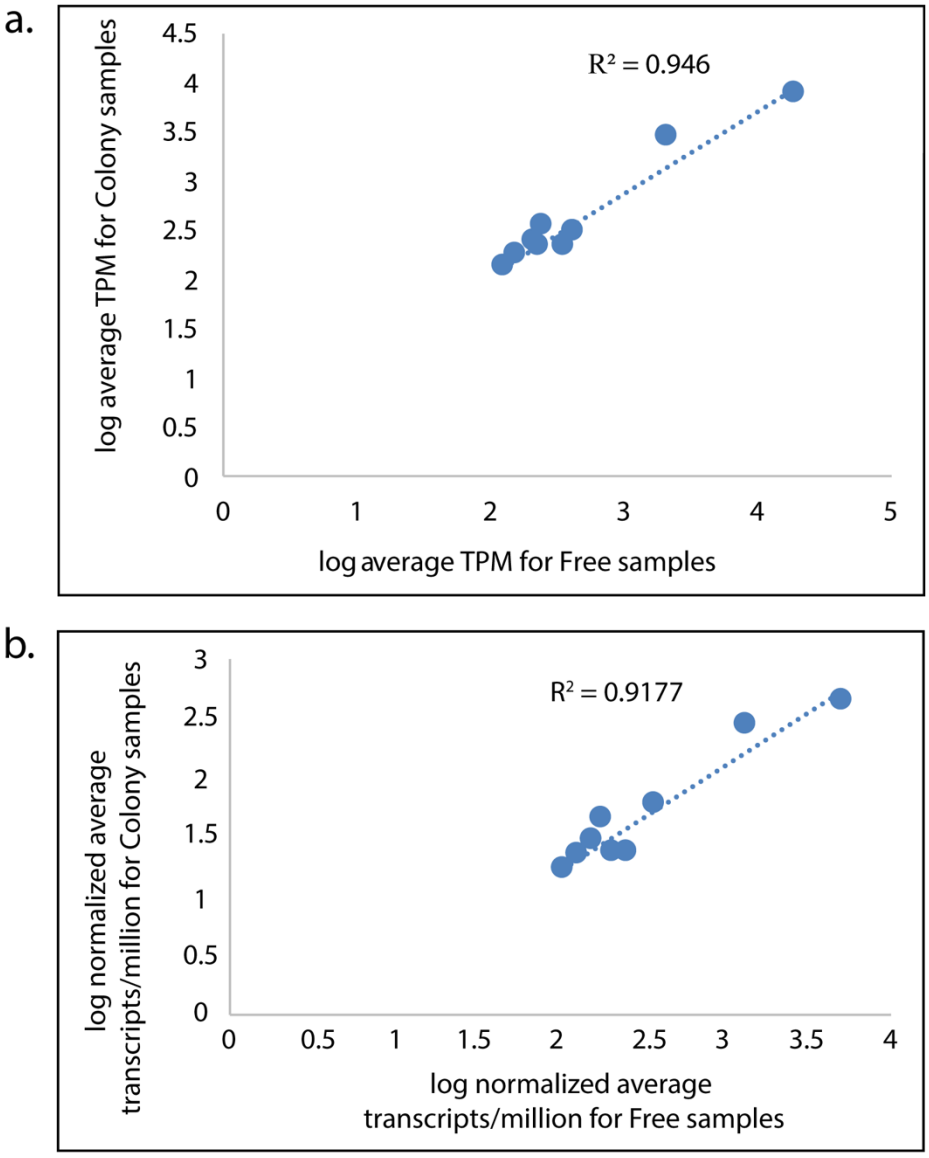

Figure 3.3 Relationship between average gene expression values (log_{10} normalized) of housekeeping genes for *Microcystis* spp. in colony and free samples

Gene expression was determined by a.) TPM and b.) transcripts per million total transcripts for nine housekeeping genes from the dominant taxa within the community,

***Microcystis* (*gltX*, *gyrB*, *recA*, *rpoB*, *tuf*, *ftsA*, *glnA*, *pgi*, and *rnpB*) for colony and the free.**

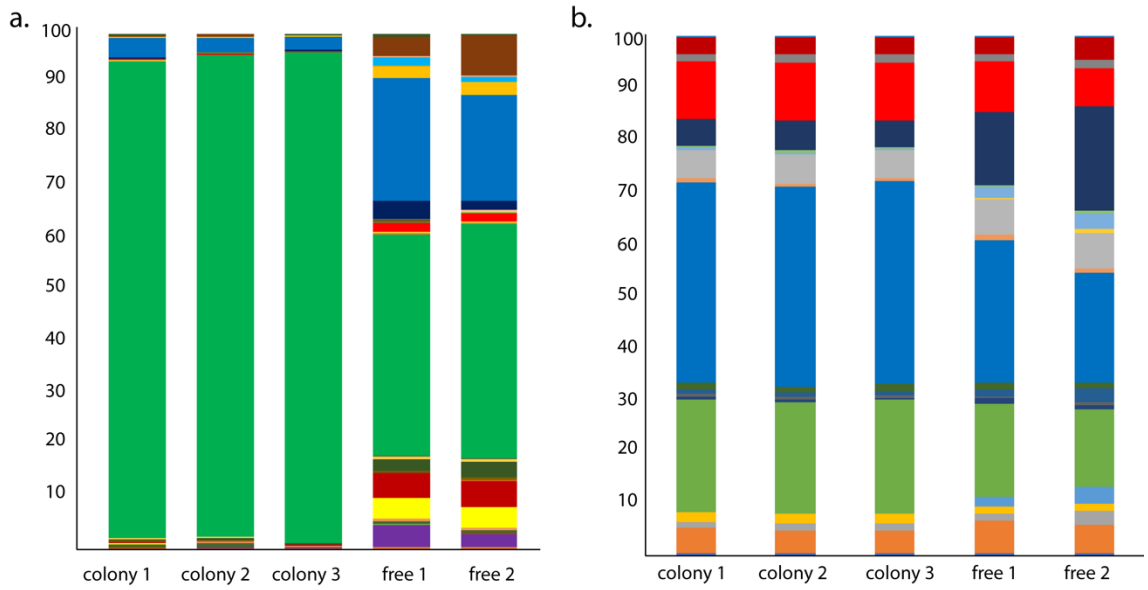

Figure 3.4 Percent abundance of reads for different phyla (a) and cyanobacterial genera (b) in colony and free samples

a). green = Cyanobacteria; blue = Proteobacteria; brown = unclassified from Eukaryota; red = Bacteriodetes; purple = Actinobacteria; and yellow = Bacillariophyta; b). genera of Cyanobacteria (blue = *Microcystis*; green = *Cyanothece*; dark blue = *Synechococcus;* red = S*ynechocystis*; gray = *Nostoc*; and orange = *Anabaena)*

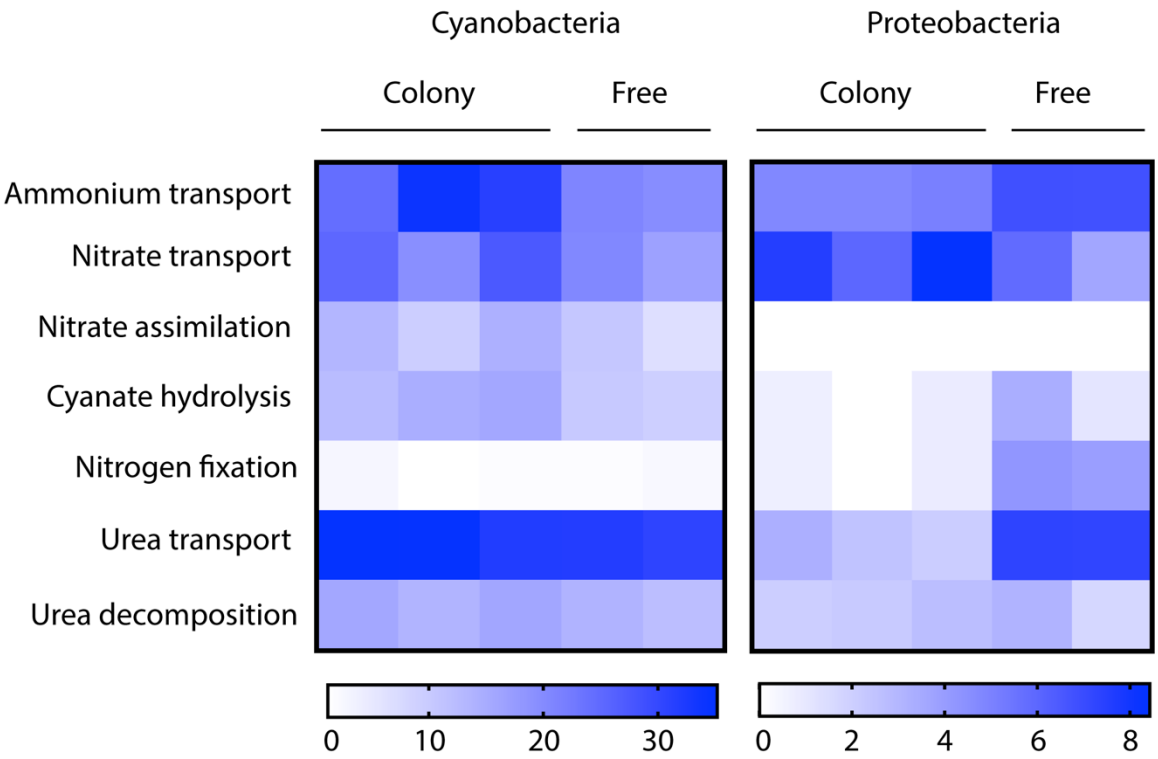

Figure 3.5 Nitrogen metabolism transcripts normalized per 1,000,000 transcripts (sqare root transformed) for Cyanobacteria and Proteobacteria

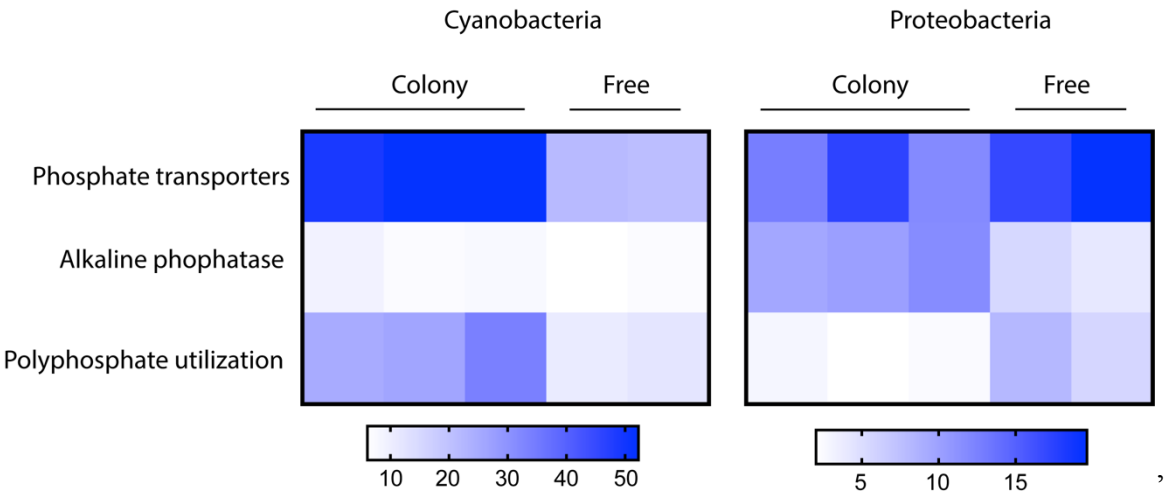

Figure 3.6 Phosphorus metabolism transcripts normalized per 1,000,000 transcripts (square root transformed) for Cyanobacteria and Proteobacteria

Proteobacteria transcripts were phylogenetically the next most represented in all samples next to those from Cyanobacteria, however these made up a greater proportion in the free samples (20%-23%) compared to the colonies (2.5%) (Figure 4b). Functional characterization of the Proteobacteria transcripts suggest that the Proteobacteria were actively scavenging both N and P in all samples as well (Figure 5 and 6). The largest number of N metabolism transcripts were for ammonium, nitrate and urea transport, similar to cyanobacteria. In contrast to the cyanobacteria, urea transport transcripts were higher in the free compared to the colony. Proteobacteria nitrate ABC transporters showed a similar trend to that of cyanobacteria. Very few N fixation transcripts were detected and those that were detected were in the free samples. Urea consumption transcripts (*i.e*. those encoding for the urease gene) were present but not in very high abundance compared with urea transport transcripts. P metabolism reads specific to Proteobacteria were also similar to cyanobacteria and involved in phosphate scavenging. While the abundance of transcripts involved in phosphate transport seem to be consistent across all samples, there was a higher proportion of alkaline phosphatases reads in the colonies, and higher proportion of transcripts for polyphosphate utilization in the free.

Analysis of the assembled Proteobacteria reads (contigs) supported the observations from individual reads that Proteobacteria in the colony and the free samples were actively involved in phosphorus metabolism (Table 3). Contigs annotated as ABC phosphate transporter, *PstS*, were detected in all three colony samples and one free sample (day 2). In the free samples, there were more diverse groups expressing these genes (*Pelagibacter*-like*, Herminiimonas, Methylobacillus, Polynucleobacter* and *Rhodospirillum*), whereas in the colony four of the five contigs came from members of the family Nitrosomonadaceae (*Nitrosomonas* and *Nitrosospira)* and one from

Table 3.3 Proteobacterial contig annotations for genes involved in nitrogen metabolism, phosphorus metabolism and methane oxidation

	Colony 1	Colony 2	Colony 3	Free 1	Free 2
Nitrogen metabolism					
Ammonium transporter	0	0	0	0	2
Nitrate ABC transporter, nitrate-binding protein	2	0	0	0	0
Nitrate ABC transporter, permease protein	0	0	0	0	1
Urea ABC transporter, urea binding protein	0	0	0	0	3
Urea carboxylase-related ABC transporter	0	0	1	0	0
Urease beta subunit (EC 3.5.1.5)	1	0	0	0	0
Phosphorus metabolism					
Alkaline phosphatase (EC 3.1.3.1)	1	3	3	1	0
Phosphate ABC transporter PstS (TC 3.A.1.7.1)	2	1	2	0	6
Phosphate transport system permease protein PstA (TC 3.A.1.7.1)	0	1	1	0	0
Phosphate transport system permease protein PstC (TC 3.A.1.7.1)	0	0	0	0	1
Phosphate transport system regulatory protein PhoU	0	0	0	0	1
Probable low-affinity inorganic phosphate transporter	0	0	0	0	1
Methane oxidation					
Particulate methane monooxygenase A-subunit (EC 1.14.13.25)	0	0	0	2	5
Particulate methane monooxygenase B-subunit (EC 1.14.13.25)	1	0	0	5	23
Particular methane monooxygenase C-subunit (EC 1.14.13.25)	1	0	0	5	12

*Geobacte*r. No clear trends were observed in N metabolism when analyzing the Proteobacteria contigs. Only a few contigs were identified to be involved in nitrate and ammonia utilization, but these were sporadic across samples. Interestingly, a total of 52 contigs were annotated as the three subunits of methane monooxygenase (Table 3). Only two of these were present in the colony and the rest were detected in the free samples. All but one of these genes was taxonomically classified as *Methylococcus* spp.

Microcystis-specific gene expression between the colony and the free samples was investigated using the genome sequence from model strain *Microcystis aeruginosa* NIES843. Seventeen genes were significantly differentially expressed between the two groups. Fourteen were more highly expressed in the free-living *Microcystis*, and only three were more highly expressed in the colony (Table 4). Two of the highly expressed genes in the colony were transposases, MAE_RS18410 and MAE_RS24045. The COG category of the first transposase was COG0675 related to InsQ, an IS609 transposase. The latter transposase remained uncharacterized. The third gene upregulated in the colony was a cytochrome b6f complex subunit 4 (MAE_RS14485). The remaining differentially expressed genes were up in the free *Microcystis* population compared to the colony. Two additional transposases were upregulated in the free *Microcystis* (MAE_RS04410 and MAE_RS18410), both of which were classified as IS5 transposases. MAE_RS20125 had the greatest decrease in fold change of all genes differentially expressed (502-fold). One PEP-CTERM sorting domain-containing protein (MAE_RS04990) was identified to be less expressed in the colony compared to the free. Genes involved in metabolic processes such as photosynthesis, glycolysis and amino acid metabolism were also downregulated in the colony compare to the free *Microcystis*; *psbA* (MAE_RS04475) pyruvate kinase

Table 3.4 Differentially expressed *Microcystis* genes in the colony compared to the free

Gene ID	Annotation	COG category	Fold change in colony	FDR p-value
MAE_RS01810	hypothetical protein	n/a	-23.13	0
MAE_RS21190	group II intron reverse transcriptase/maturase	COG 5-methylcytosine-specific restriction endonuclease McrA (Defense mechanism)	-28.9	0
MAE_RS13335	hypothetical protein	n/a	-7.06	8.92E-09
MAE_RS04990	PEP-CTERM sorting domain-containing protein	n/a	-8.38	1.68E-08
MAE_RS04475	photosystem II protein D1	*psbA*	-6.18	0.0000325
MAE_RS01800	pyruvate kinase	n/a	-3.69	0.000649
MAE_RS01460	histidine kinase	BaeS signal transduction, responds to envelope stress	-4.85	0.00238
MAE_RS20125	IS5 family transposase ISmae6	n/a	-503.98	0.00534
MAE_RS18410	transposase	IS609 transposase	10.27	0.00536
MAE_RS04410	pseudo	IS5 IS1182 transposase	-2.79	0.00948
MAE_RS24045	transposase	n/a	4.28	0.01
MAE_RS04510	polyamine ABC transporter substrate binding protein	PotD; Spermidine/putrescine-binding periplasmic protein [Amino acid transport and metabolism	-2.82	0.03
MAE_RS04505	zinc metalloprotease	Spo IVFB site-2 protease regulates intramembrane proteolysis	-3.35	0.03
MAE_RS13905	FAD-dependent oxidoreductase	n/a	-2.68	0.03
MAE_RS05040	heatshock protein	IbpA; Molecular chaperone IbpA, HSP20 family [Posttranslational modification, protein turnover	-3.27	0.04
MAE_RS14485	cytochrome b6f complex subunit 4	n/a	2.69	0.04
MAE_RS21960	hypothetical protein	*yyaL*, contains thioredoxin and six hairpin glycosidase-like domains	-3.64	0.04

(MAE_RS01800) and a polyamine ABC transporter (MAE_RS04510), respectively. The remaining genes that were down in the colony compared to the free were either hypothetical or involved in stress response. In particular, two genes were involved in responding and regulating membrane stress and one was a group II intron reverse transcriptase and maturase.

The free and colonial *Microcystis* population was also likely experiencing an active phage infection. Analysis of total *Microcystis* phage gene expression using TPM suggested that four genes in the phage were more highly expressed in the colony samples (Figure 7). However, when the number of read were normalized by *Microcystis* specific *rpoB* used as a *Microcystis* housekeeping gene in previous phage studies, the opposite trend was observed (Figure 8). A phage infection in both the free and colonial *Microcystis* populations was supported by analysis of the contigs that detected phage contigs identified specifically as *Microcystis* phage as well as other unclassified viruses from the same family (Myoviridae) (data not shown).

Discussion

As *Microcystis* spp. continue to proliferate in freshwater ecosystems around the world, it has become increasingly important to understand the factors that contribute to its success. Colony formation is considered an important factor, yet the specific drivers of colony formation are not well defined (Xiao *et al.*, 2018). In the laboratory, studying this is notoriously difficult as most species remain only in the single celled state, but some studies have been able to induce colony formation (Xiao *et al.*, 2018). In general, lab and field studies have demonstrated colony formation seems to be a response to a variety of stressors (nutrient, temperature, grazers), though heterotrophic bacteria have also been shown to induce mucilage production and colony formation

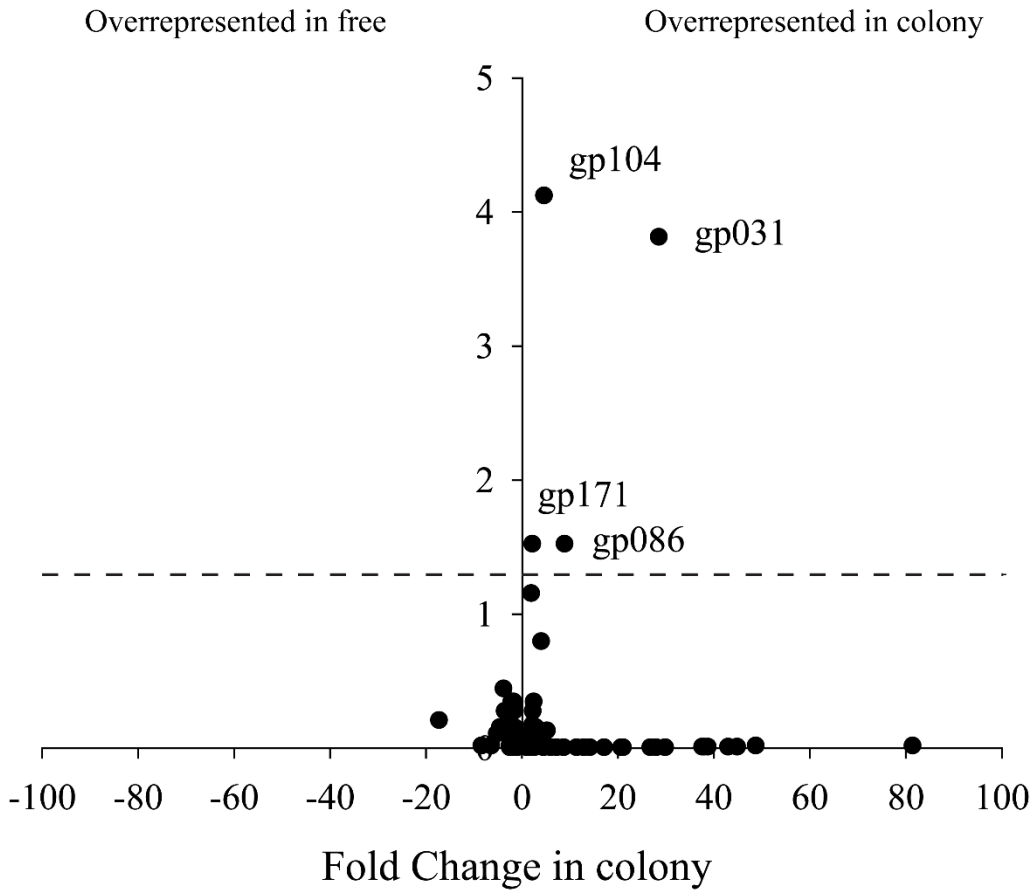

Figure 3.7 Differential expression of genes in *Microcystis* phage based on TPM

Y axis is -log(p value) and the dotted line and above represents a p value of <0.05.

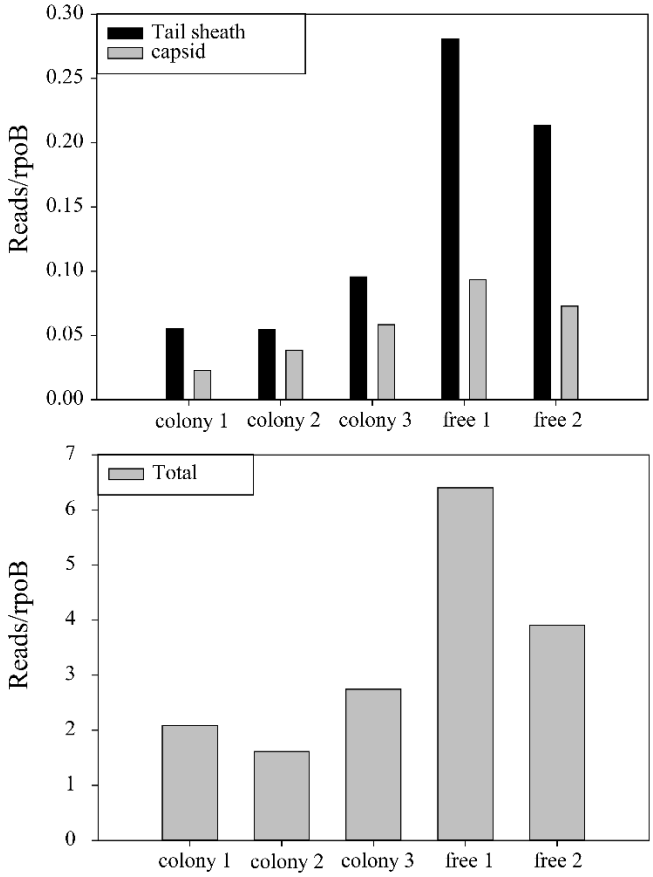

Figure 3.8 Total *Microcystis* phage transcripts, tail sheath and capsid transcripts normalized by *Microcystis rpoB*

(Shen *et al.*, 2011, Wang *et al.*, 2016). Expression analyses examining colonies and free-living *Microcystis* community from the field during a bloom provided the ability to capture a snapshot of how the microbial community in the colony is responding to its environment differently than the community in the single celled state. This provided information about the real-time advantages to *Microcystis* as well as the associated bacteria of being in a colony.

Gene expression analysis of colonial Microcystis *and free-living* Microcystis

Seventeen *Microcystis* genes were differentially expressed in the colony compared to the free-living population. *Microcystis* colonies are associated with exopolysaccharide (EPS) production and even though the composition of the EPS can vary between species and growth conditions, EPS concentrations will typically increase with colony size (Xiao *et al.*, 2018). Surprisingly, no genes were involved in EPS production were differential expressed between the colony and free samples, except for one with a PEP-CTERM sorting domain. Proteins with this domain are typically associated with the production of EPS in microorganisms that aggregate (Haft *et al.*, 2006) but was more highly expressed in the free than the colony. Another gene encoding a hypothetical protein containing a glycosidase domain had lower expression in the colony, suggesting EPS repression in the free-living *Microcystis*. There are two proposed mechanisms for colony formation that may be stimulated by different factors and both of these mechanisms involve the production of EPS. In "cell division" driven colony formation, EPS is secreted to envelop the daughter cells, and in "cell adhesion" driven formation cells aggregate as a result of the secretion of EPS (Xiao *et al.*, 2018). The apparent lack of differentially expressed genes for EPS in the colonial *Microcystis* may indicate the formation of colonies was driven by cell division rather than EPS production or that these were already well-established colonies.

Genes involved in photosynthesis were differentially expressed between the colony and the free. *PsbA* was underrepresented in colony versus the free samples was observed, which encodes a key subunit of photosystem II, protein D1. Interestingly, there was also a 2.5-fold increase in cytochrome b6f complex subunit 4 in the colony compared to the free. These genes were also differentially expressed when the sample taken at 500 was removed from the analysis, supporting that this observation was indeed inherent to the physiology within the colony and the observations were not skewed due to the absence of light or low light at one time point (data not shown). Cytochrome b6f complex is involved in the transfer of electrons between PSI and PSII (Bryant, 2006). The potential decrease in PSII activity and increase in cytochrome b6f may suggest cyclic electron flow is being favored over linear electron flow during photosynthesis (Munekage *et al.*, 2004). This ultimately yields higher production of ATP rather than the replenishment of reducing power in the form of NADPH. Favoring cyclic electron flow in *Microcystis* and other photosynthetic microorganisms has been reported as a protective response to a variety of stressors, such as exposure to heavy metals, low and high light, high salt concentrations, low dissolved CO_2, and high temperatures (Herbert *et al.*, 1995, Jeanjean *et al.*, 1998, Zhou *et al.*, 2006, Wang *et al.*, 2013).

Exposure of free-living *Microcystis* to multiple stressors is evident in the sequencing data. Higher expression of MAE_RS05040, which encodes a heat shock protein, in the free *Microcystis* community suggests these cells are responding to temperature or oxidative stress (Zheng *et al.*, 2001, Seo *et al.*, 2006). Higher expression of a pyruvate kinase (MAE_RS01800), may be indicative that carbon flow is directed towards glycolytic pathways rather than storage, which has been observed in low CO_2 conditions in *Synechococcus* (Schwarz *et al.*, 2011), consistent with the conditions that occur in Taihu during summer months. Histidine kinases are important in signal

transduction in bacteria and could be a response to a number of stressors, like P-starvation (Santos-Beneit, 2015). In *Microcystis* specifically, histidine kinases have been shown to be upregulated in response to grazing activity (Harke *et al.*, 2017). Differential expression of transposases has also been observed as a response to stress, such as during nutrient limitation (or growth on certain nutrient sources) and predatory grazing (Steffen *et al.*, 2014, Harke *et al.*, 2017). The high proportions of transcripts for N and P transporters by the cyanobacterial community and Proteobacteria suggests the community was likely nutrient-limited by N, P or both N and P.

Transposase activity, indicative of genome rearrangement, has also been proposed as a response to phage infection (Harke *et al.*, 2015). Indeed, *Microcystis* phage activity was detected in these samples. Some phage genes were more highly expressed in the colony when phage reads were normalized by metrics driven by library size. This trend shifted when expression was normalized to *Microcystis* specific *rpoB,* a host housekeeping gene, and instead suggested that the free-living *Microcystis* carried a larger burden of phage per cell compared to the colonial *Microcystis* and may be more sensitive to phage infection. It has been previously proposed that colony formation by cyanobacteria can protect against viral infection (Sulcius *et al.*, 2014) while other have shown colonies or particle-associated bacteria are more prone to viral infection (Bettarel *et al.*, 2016). Here, while phage activity was detected in *Microcystis* colonies, the free-living *Microcystis* carried a higher burden of phage. This could explain the differential expression of *Microcystis* genes involved in membrane stress and suggest the colonies may be more resistant to viral infection or more effective at fighting off the virus.

Gene expression by the co-occurring Proteobacterial community

It has been previously noted that the *Microcystis* phycosphere community differs in composition from the free-living portion of the community (Shi *et al.*, 2018), and here we show their transcriptional activity is different as well. Living within the colony may promote a different metabolism than when cells are free-living. Proteobacteria were the most active group of heterotrophic bacteria in both colonies and the free-living community. Proteobacteria are often found in association with *Microcystis* blooms and were suggested to be involved in nutrient transfer or exchange numerous times. Evidence from several studies has suggests their specific role in nitrogen transformations on both DNA and RNA level such as denitrification processes, nitrogen fixation, urea degradation and nitrification (Chen *et al.*, 2012, Steffen *et al.*, 2012, Steffen *et al.*, 2015, Belisle *et al.*, 2016, Krausfeldt *et al.*, 2017). These processes have been measured during blooms as well (McCarthy *et al.*, 2007, Belisle *et al.*, 2016, Hampel *et al.*, 2018). Here, we show that the nitrogen metabolism of the associated Proteobacteria within the colony may differ from those in the free-living bacteria. Transcripts for ammonium transporters are more represented in the free samples, the same as previously described for urea transporter transcripts. This is likely due to competition with *Microcystis* spp. and other cyanobacteria for NH_4, their preferable N source (Blomqvist *et al.*, 1994, Glibert *et al.*, 2016). Interestingly, there were also differences in P metabolism by Proteobacteria across samples. In the colony, there was a higher proportion of reads annotated as alkaline phosphatases. This is typically a marker for P limitation, but expression of alkaline phosphatases has also been demonstrated to be stimulated in response to low urea concentrations (Steffen *et al.*, 2014). Differences in expression of genes for utilization and transport of N and P between the Cyanobacteria and Proteobacteria with the fractionated samples suggests these communities employ different strategies for nutrient scavenging. For example,

Proteobacteria in the free-living fraction are more actively scavenging urea and ammonium and fixing N_2. These observations were not fully supported by the assembled data, but this is likely due to the low number of total mRNA reads as a result of the contaminating rRNA sequences. The sequencing depth may not have been sufficient enough to assemble most effectively.

A high number of Proteobacterial contigs were annotated as particulate class methane monooxygenases which indicated that methanotrophs were active in the free-living fraction of the bloom. Methane is an important greenhouse gas, and studies have estimated lakes account for 6-16% of the methane released into the atmosphere (Bastviken *et al.*, 2004). Because *Taihu* is rich in organic matter (Wang *et al.*, 2018), it is not surprising that there is evidence for release of methane production. Indeed, periodic fluxes of methane from *Taihu* sediments have been detected previously (Wang *et al.*, 2006). The activity of methanotrophs *via* methane monooxygenases in the surface waters may help to alleviate the emission of methane from *Taihu*. While it may be advantageous for some bacteria to use organic carbon from the colony, others may gain a higher benefit outside the colony utilization C1 carbons, further supporting there are different metabolic strategies between free-living bacteria and bacteria within colonies.

Conclusions

In the present study, differences in gene expression between cyanobacterial colonies and the free-living community were examined. Based on differential gene expression, *Microcystis* within the colony is less responsive to stress on a transcript level. This suggested there is a benefit to being within the colony to evade stressors the free-living *Microcystis* was responding to, such as nutrient limitation, heat stress, grazing and possibly phage infection. Evidence favoring cyclic electron flow supports this and would likely be an important mechanism for maintaining

photosynthetic efficiency or redirecting energy to scavenge nutrients. This was evident in the cyanobacterial community, dominated by *Microcystis*, who were more actively transcribing nitrate and phosphate transporters. Insights into the genetic mechanism for EPS production, which is important for colony formation, still remains unclear with no clear trends in expression of EPS genes. Interestingly, Proteobacteria within the phycosphere and as a part of the free-living community employed different strategies for nutrient scavenging. This is likely an adaptive strategy of competition with *Microcystis* and other cyanobacteria despite to benefit from the protection within the colony. These observations support that colony formation is a stress response to potentially a variety of factors also serving as a microniche for bacteria that perform different functions from the free-living community.

Acknowledgements

We would like to thank Xiangming Tang and colleagues at the TLLER field station for hosting our sampling trip to Lake Taihu. This work was made possible due to grant NA11NOS4780021 from NOS National Center for Coastal Ocean Science (NCCOS) under their Harmful Algal Bloom Prevention Control and Mitigation program to SWW and GLB, and National Science Foundation awards (DEB-1240870; IOS-1451528) to SWW. The authors also acknowledge the generous support of the *Kenneth & Blaire Mossman* endowment to the University of Tennessee.

References

Bastviken D., Cole J., Pace M. & Tranvik L. 2004. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global biogeochemical cycles*. 18,

Beardall J. & Raven J. A. 2017. Cyanobacteria vs green algae: which group has the edge? *Journal of experimental botany*. 68, 3697-3699.

Belisle B. S., Steffen M. M., Pound H. L., Watson S. B., DeBruyn J. M., Bourbonniere R. A., Boyer G. L. & Wilhelm S. W. 2016. Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*. 42, 599-607.

Bettarel Y., Motegi C., Weinbauer M. G. & Mari X. 2016. Colonization and release processes of viruses and prokaryotes on artificial marine macroaggregates. *FEMS microbiology letters*. 363,

Blomqvist P., Petterson A. & Hyenstrand P. 1994. Ammonium-nitrogen: a key regulatory factor causing dominance of non-nitrogen-fixing cyanobacteria in aquatic systems. *Archiv für Hydrobiologie*. 132, 141-164.

Bormans M., Sherman B. & Webster I. 1999. Is buoyancy regulation in cyanobacteria an adaptation to exploit separation of light and nutrients? *Marine and Freshwater Research*. 50, 897-906.

Bryant D. A. 2006. *The molecular biology of cyanobacteria*. Springer Science & Business Media,

Chen X., Yang L., Xiao L., Miao A. & Xi B. 2012. Nitrogen removal by denitrification during cyanobacterial bloom in Lake Taihu. *Journal of Freshwater Ecology*. 27, 243-258.

Chen X., Huang Y., Chen G., Li P., Shen Y. & Davis T. W. 2018. The secretion of organics by living Microcystis under the dark/anoxic condition and its enhancing effect on nitrate removal. *Chemosphere*.

Chen Z., Zhang J., Li R., Tian F., Shen Y., Xie X., Ge Q. & Lu Z. 2018. Metatranscriptomics analysis of cyanobacterial aggregates during cyanobacterial bloom period in Lake Taihu, China. *Environmental Science and Pollution Research*. 25, 4811-4825.

Glibert P. M., Wilkerson F. P., Dugdale R. C., Raven J. A., Dupont C. L., Leavitt P. R., Parker A. E., Burkholder J. M. & Kana T. M. 2016. Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions. *Limnology and Oceanography*. 61, 165-197.

Haft D. H., Paulsen I. T., Ward N. & Selengut J. D. 2006. Exopolysaccharide-associated protein sorting in environmental organisms: the PEP-CTERM/EpsH system. Application of a novel phylogenetic profiling heuristic. *BMC biology*. 4, 29.

Hampel J. J., McCarthy M. J., Gardner W. S., Zhang L., Xu H., Zhu G. & Newell S. E. 2018. Nitrification and ammonium dynamics in Taihu Lake, China: seasonal competition for ammonium between nitrifiers and cyanobacteria. *Biogeosciences*. 15, 733.

Harke M. J., Berry D. L., Ammerman J. W. & Gobler C. J. 2012. Molecular response of the bloom-forming cyanobacterium, *Microcystis aeruginosa,* to phosphorus limitation. *Microbial ecology*. 63, 188-198.

Harke M. J., Davis T. W., Watson S. B. & Gobler C. J. 2015. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environmental science & technology*. 50, 604-615.

Harke M. J., Jankowiak J. G., Morrell B. K. & Gobler C. J. 2017. Transcriptomic responses in the bloom-forming cyanobacterium *Microcystis* induced during exposure to zooplankton. *Applied and environmental microbiology*. 83, e02832-02816.

Herbert S. K., Martin R. E. & Fork D. C. 1995. Light adaptation of cyclic electron transport through photosystem I in the cyanobacterium *Synechococcus* sp. PCC 7942. *Photosynthesis research*. 46, 277-285.

Ibelings B. W., Mur L. R., Kinsman R. & Walsby A. 1991. *Microcystis* changes its buoyancy in response to the average irradiance in the surface mixed layer. *Archiv für Hydrobiologie*. 120, 385-401.

Jeanjean R., Bédu S., Havaux M., Matthijs H. C. & Joset F. 1998. Salt-induced photosystem I cyclic electron transfer restores growth on low inorganic carbon in a type 1 NAD (P) H dehydrogenase deficient mutant of *Synechocystis* PCC6803. *FEMS microbiology letters*. 167, 131-137.

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.). *Hydrobiologia*. 581, 161-165.

Krausfeldt L. E., Tang X., van de Kamp J., Gao G., Bodrossy L., Boyer G. L. & Wilhelm S. W. 2017. Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu. *FEMS Microbiology Ecology*. 93,

Kromkamp J. 1987. Formation and functional significance of storage products in cyanobacteria. *New Zealand Journal of Marine and Freshwater Research*. 21, 457-465.

McCarthy M. J., Lavrentyev P. J., Yang L., Zhang L., Chen Y., Qin B. & Gardner W. S. 2007. Nitrogen dynamics and microbial food web structure during a summer cyanobacterial bloom in a subtropical, shallow, well-mixed, eutrophic lake (Lake Taihu, China). *Hydrobiologia*. 581, 195-207.

Munekage Y., Hashimoto M., Miyake C., Tomizawa K.-I., Endo T., Tasaka M. & Shikanai T. 2004. Cyclic electron flow around photosystem I is essential for photosynthesis. *Nature*. 429, 579.

Parveen B., Ravet V., Djediat C., Mary I., Quiblier C., Debroas D. & Humbert J. F. 2013. Bacterial communities associated with *Microcystis* colonies differ from free-living communities living in the same ecosystem. *Environmental microbiology reports*. 5, 716-724.

Penn K., Wang J., Fernando S. C. & Thompson J. R. 2014. Secondary metabolite gene expression and interplay of bacterial functions in a tropical freshwater cyanobacterial bloom. *The ISME journal*. 8, 1866.

Reynolds C. S., Jaworski G., Cmiech H. & Leedale G. 1981. On the annual cycle of the blue-green alga *Microcystis aeruginosa* Kütz. emend. Elenkin. *Philosophical Transactions of the Royal Society B*. 293, 419-477.

Sandrini G., Matthijs H. C., Verspagen J. M., Muyzer G. & Huisman J. 2014. Genetic diversity of inorganic carbon uptake systems causes variation in CO_2 response of the cyanobacterium *Microcystis. The ISME journal*. 8, 589.

Santos-Beneit F. 2015. The *Pho* regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology*. 6, 402.

Schwarz D., Nodop A., Hüge J., Purfürst S., Forchhammer K., Michel K.-P., Bauwe H., Kopka J. & Hagemann M. 2011. Focus Issue on Plastid Biology: Metabolic and Transcriptomic Phenotyping of Inorganic Carbon Acclimation in the Cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Physiology*. 155, 1640.

Seo J. S., Lee Y.-M., Park H. G. & Lee J.-S. 2006. The intertidal copepod *Tigriopus japonicus* small heat shock protein 20 gene (Hsp20) enhances thermotolerance of transformed *Escherichia coli*. *Biochemical and biophysical research communications*. 340, 901-908.

Shapiro J. 1997. The role of carbon dioxide in the initiation and maintenance of blue-green dominance in lakes. *Freshwater biology*. 37, 307-323.

Shen H., Niu Y., Xie P., Tao M. & Yang X. 2011. Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshwater Biology*. 56, 1065-1080.

Shi L., Cai Y., Kong F. & Yu Y. 2012. Specific association between bacteria and buoyant *Microcystis* colonies compared with other bulk bacterial communities in the eutrophic Lake Taihu, China. *Environmental microbiology reports*. 4, 669-678.

Shi L., Huang Y., Zhang M., Shi X., Cai Y., Gao S., Tang X., Chen F., Lu Y. & Kong F. 2018. Large buoyant particles dominated by cyanobacterial colonies harbor distinct bacterial communities from small suspended particles and free-living bacteria in the water column. *MicrobiologyOpen*.

Shia L., Cai Y., Wang X., Li P., Yu Y. & Kong F. 2010. Community structure of bacteria associated with *Microcystis* colonies from cyanobacterial blooms. *Journal of Freshwater Ecology*. 25, 193-203.

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, e44002.

Steffen M. M., Belisle B. S., Watson S. B., Boyer G. L., Bourbonniere R. A. & Wilhelm S. W. 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Applied and environmental microbiology*. 81, 3268-3276.

Steffen M. M., Dearth S. P., Dill B. D., Li Z., Larsen K. M., Campagna S. R. & Wilhelm S. W. 2014. Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. *The ISME journal*. 8, 2080-2092.

Sulcius S., Staniulis J., Paskauskas R., Olenina I., Salyte A., Ivanauskaite A. & Griniene E. 2014. Absence of evidence for viral infection in colony-embedded cyanobacterial isolates from the Curonian Lagoon. *Oceanologia*. 56,

Tang X., Gao G., Chao J., Wang X., Zhu G. & Qin B. 2010. Dynamics of organic-aggregate-associated bacterial communities and related environmental factors in Lake Taihu, a large eutrophic shallow lake in China. *Limnology and oceanography*. 55, 469-480.

Wang H., Lu J., Wang W., Yang L. & Yin C. 2006. Methane fluxes from the littoral zone of hypereutrophic Taihu Lake, China. *Journal of Geophysical Research: Atmospheres*. 111,

Wang S., Chen F., Mu S., Zhang D., Pan X. & Lee D. J. 2013. Simultaneous analysis of photosystem responses of *Microcystis aeruginoga* under chromium stress. *Ecotoxicology and environmental safety*. 88, 163-168.

Wang W., Shen H., Shi P., Chen J., Ni L. & Xie P. 2016. Experimental evidence for the role of heterotrophic bacteria in the formation of *Microcystis* colonies. *Journal of Applied Phycology*. 28, 1111-1123.

Wang W., Wang S., Jiang X., Zheng B., Zhao L., Zhang B. & Chen J. 2018. Differences in fluorescence characteristics and bioavailability of water-soluble organic matter (WSOM) in sediments and suspended solids in Lihu Lake, China. *Environmental Science and Pollution Research*. 1-15.

Xiao M., Li M. & Reynolds C. S. 2018. Colony formation in the cyanobacterium Microcystis. *Biological Reviews*.

Yamamoto Y. & Nakahara H. 2005. Competitive dominance of the cyanobacterium Microcystis aeruginosa in nutrient-rich culture conditions with special reference to dissolved inorganic carbon uptake. *Phycological Research*. 53, 201-208.

Yang Z., Kong F., Shi X., Zhang M., Xing P. & Cao H. 2008. Changes in the morphology and polysaccharide content of *Microcystis aeruginosa* (cyanobacteria) during flagellate grazing. *Journal of Phycology*. 44, 716-720.

Zheng M., Wang X., Templeton L. J., Smulski D. R., LaRossa R. A. & Storz G. 2001. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *Journal of bacteriology*. 183, 4562-4570.

Zhou W., Juneau P. & Qiu B. 2006. Growth and photosynthetic responses of the bloom-forming cyanobacterium *Microcystis aeruginosa* to elevated levels of cadmium. *Chemosphere*. 65, 1738-1746.

Chapter 4 : Assimilation of liberated CO_2 during *Microcystis*' growth on urea

Publication Note

This chapter is a version of a manuscript in preparation by Lauren E. Krausfeldt, Marcus Pryor, Abigail T. Farmer, Hector Castro Gonzalez, Shawn R. Campagna, and Steven W. Wilhelm.

My contributions included assistance in experimental design, the majority of the culture work, assistance with extraction of metabolites, most of the data analysis and writing most of the manuscript.

Abstract

The use of urea as nitrogenous (N) fertilizer has increased substantially over the past decade, now comprising greater than 50% of the total N fertilization use globally. Urea is detected in freshwater systems in areas of high agricultural practice and contributes to eutrophication and cyanobacterial blooms. While the use of urea as a N source by cyanobacteria has been well-established, its potential role as a carbon (C) source has received little attention. Here we use ^{13}C-urea and demonstrate that that a model toxic cyanobacterium, *Microcystis aeruginosa* NIES843, assimilates CO_2 from hydrolyzed urea into central carbon metabolism and amino acid biosynthesis pathways. This ability to incorporate CO_2 from the breakdown of urea would be advantageous to cyanobacteria during dense algal blooms, when high rates of photosynthesis draw down dissolved CO_2 concentrations and increases pH *in situ*. In parallel with this work, we demonstrate in culture that *M. aeruginosa* is more successfully at high pH, especially when urea is the sole N source. Our results imply that CO_2 assimilation from urea degradation may provide an intriguing growth advantage.

Introduction

Fertilizer consumption has increased drastically over the last several decades around the world, including within the United States (Glibert et al., 2014, Cao et al., 2017). The use of urea (CH_4N_2O) as a N fertilizer has increased 100-fold in the last four decades, and now makes up > 50% of the total N fertilizer consumption worldwide (Glibert et al., 2006). Residual urea in soils not utilized by plants is generally thought to be consumed by the microbial community via the enzymes urease, oxidized to nitrate for denitrification, or volatilized to NH_4 and lost to the atmosphere (Glibert et al., 2014). Various agricultural practices such as the application of urease inhibitors to soils, no-tillage land management and the timing of fertilizer application before rainfall or prior to snow melt increases the likelihood that urea will be exported to nearby aquatic systems (Glibert et al., 2006). Indeed, urea does contribute to the eutrophication of inland and coastal waters (Finlay et al., 2010, Glibert et al., 2014). Urea can reach concentrations as high as 150 μM and represent over 50% of the total dissolved N pool, especially locally to heavy agricultural areas (Bogard et al., 2012, Chaffin & Bridgeman, 2013, Glibert et al., 2014).

The role of urea as a driver of harmful cyanobacterial blooms has received increased attention recently as many bloom-forming cyanobacteria, such as *Microcystis* spp., possess the enzyme urease and can utilize urea as a N source. It has even been argued that the increase in urea use has corresponded to a rise in reports and the duration of blooms in high fertilization areas (Glibert *et al.*, 2014, Paerl *et al.*, 2016). These correlations are well-supported by the field studies demonstrating the ability of cyanobacteria to readily consume urea as a N source, and even in some cases a preference for urea over oxidized forms of N such as nitrate (Davis *et al.*, 2010, Chaffin & Bridgeman, 2013, Belisle *et al.*, 2016). Lab studies confirm these observations, and even suggest a specific role for urea in toxicity and genomic evolution (Steffen *et al.*, 2014, Peng *et al.*, 2018).

While the importance of urea as a N source in cyanobacterial blooms is becoming more well-established, urea as a carbon source has not been extensively studied. This may be significant during dense blooms when the dissolved CO_2 in the water is drawn down well below the CO_2 concentrations in the atmosphere. CO_2-limitation can manifest during the course of a bloom and is observable by proxy of increasing pH (Paerl & Huisman, 2009, Visser *et al.*, 2016, Ji *et al.*, 2017). Many times, the pH of lakes experiencing cyanobacterial blooms increases over 9 and as high as 10 (Paerl *et al.*, 2011, Krausfeldt *et al.*, 2017). When the pH reaches these levels, the inorganic carbon pool is primarily composed of bicarbonate and carbonate (Figure 1, Wetzel, 2001). Cyanobacteria are equipped to handle low CO_2 concentrations by use of highly efficient carbon concentrating mechanisms (Price, 2011, Visser *et al.*, 2016). Yet free CO_2 is still preferable over bicarbonate due to the energetics associated with its assimilation (Sandrini *et al.*, 2015). Since the breakdown of urea by urease occurs intracellularly, the CO_2 from urea hydrolysis should be easily accessible for carbon fixation (Finlay *et al.*, 2010, Sandrini *et al.*, 2014). This may provide an additional advantage during dense blooms when dissolved CO_2 is limited in extreme alkaline conditions. It can be argued that heterotrophic bacteria are breaking down urea (Belisle *et al.*, 2016) as urease is found across phylogenetic diverse microorganisms (Koper *et al.*, 2004) and urease (+) heterotrophs are found in association with cyanobacterial blooms (Steffen *et al.*, 2012). However, growth experiments in the field and gene expression studies would suggest cyanobacteria are major contributors to urea degradation during bloom events (see Chapter 2, Davis *et al.*, 2010, Chaffin & Bridgeman, 2013, Steffen *et al.*, 2015, Chaffin *et al.*, 2018).

The goal of this study was to determine if *M. aeruginosa* gained a fitness advantage from growing on urea at high pH because of its potential to serve as a supplemental carbon source. We first examined the relationship between pH and cyanobacteria in Lake Erie, which experiences

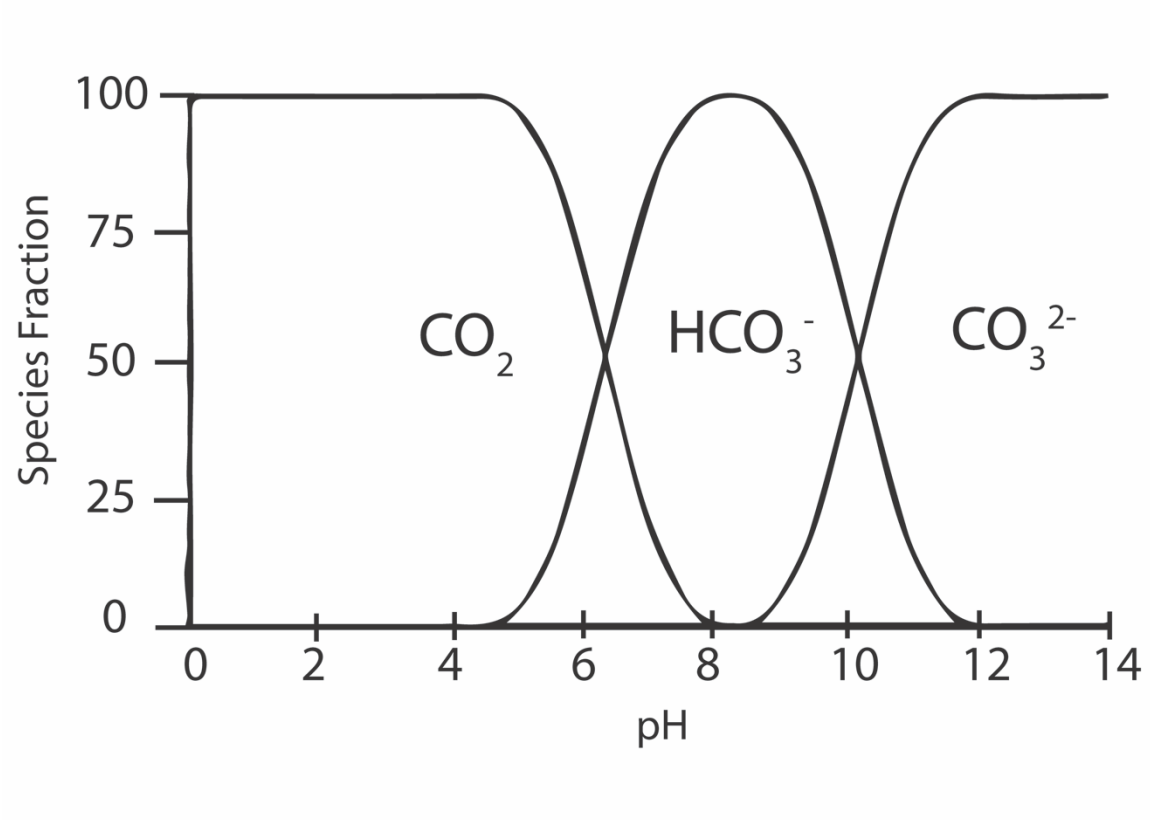

Figure 4.1 Inorganic carbon speciation in freshwater systems in relation to pH (adapted from Wetzel, 2001)

harmful cyanobacterial blooms annually. We then performed buffered growth experiments with nitrate, ammonium and urea as the sole N source to establish the effects of growth on different N sources across a range of environmentally relevant pH conditions. These were followed by experiments to test the hypothesis that the CO_2 from urea hydrolysis can be consumed by cyanobacteria using ^{13}C. Experimental data was contrasted against available environmental data to determine how these factors contributed to the persistence of cyanobacteria.

Methods

Field data analysis

Real-time phycocyanin (RFU) and pH data from monitoring buoys in the western basin of Lake Erie (Figure 2) was accessed through the NOAA Great Lakes Environmental Research Laboratory website (https://www.glerl.noaa.gov/res/HABs_and_Hypoxia/rtMonSQL.php, March 22, 2018). Available data was collected at 15-min intervals from July 2015 to September 2015, which captured before, during and after one of the biggest blooms on record in Lake Erie (Ho & Michalak, 2017). Data was downloaded and phycocyanin and pH were averaged over each calendar day for analysis to account for diel variations.

Effect of pH on Microcystis aeruginosa maintained on different N sources

Axenic *Microcystis aeruginosa* NIES843 was grown on modified CT medium (Steffen *et al.*, 2014) with different N sources (nitrate provided as KNO_3/$CaNO_3$*$4H_2O$; ammonium provided as NH_4Cl, or urea) at 26°C and on a diel cycle (12:12) with ~50-60 μmol photon m^{-2} s^{-1} of light. Cultures were checked for contamination routinely by microscopy and growth in purity tubes of rich and minimal media (LB and CT + 0.05% glucose, 0.05% acetate, 0.05% pyruvate and 0.05%

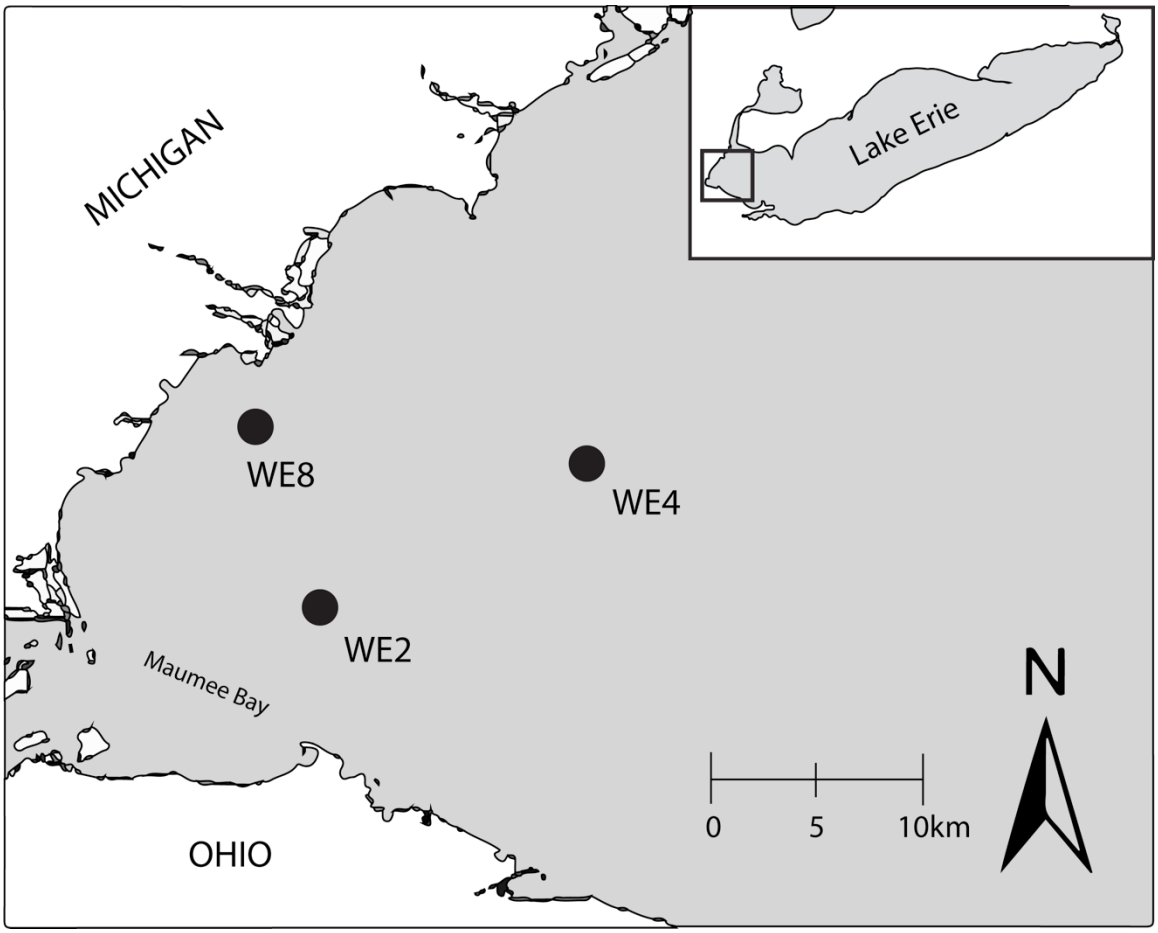

Figure 4.2 The locations of real-time monitoring buoys in Lake Erie (black circles)

lactate, respectively) incubated in the dark. N was provided at 0.585 μM. Cultures were maintained at 4 different pH (7.7, 8.2, 8.7 and 9.2) in the TAPS buffered CT medium. Prior to starting each growth curve, cells were collected on a 1.0 μm nominal pore-size polycarbonate filter and resuspended in its respective media. Cells were then added to 50 mL glass culture tubes at a final volume of 25 mL at a starting concentration of ~75,000 cell/mL. Chlorophyll *a* autofluorescence was measured as a proxy for cell density at approximately the same time every day using a Turner Designs TD-700 fluorometer (Sunnyvale, CA, USA).

Measuring cellular metabolites and ^{13}C incorporation from urea

Axenic *M. aeruginosa* NIES 843, acclimated to experimental culture conditions using ^{12}C-urea, was transferred to media containing ^{13}C-urea as the sole N source in CT medium to a concentration of 0.585 μM as described above at three different pH values. For each condition pH was empirically determined at the onset of inoculation (after autoclaving media, which can alter pH) and resulted in pH of 7.5 ± 2, 8.4 ± 2 and 9.5 ± 2 for our experiments. On day 7 (late log growth phase, Table 1), cells were filtered onto a 1.0 μm nominal pore size polycarbonate filter and immediately extracted for metabolites. Metabolites were extracted as previously described (Peng *et al.*, 2018) and kept at -20°C until analysis by methods in Peng *et al*. (2018).

Data analysis

Growth rates of *M. aeruginosa* on nitrate, ammonium, and urea at different pH values were determined by using exponential regression to fit natural log transformed fluorescence data during log-linear growth. Statistical differences between growth rates within each treatment were determined using One-way ANOVA and corrected for multiple comparisons using a Tukey test.

Statistical differences in growth on different N-forms across pH treatments were determined using a Two-way ANOVA and corrected for multiple comparison using the Tukey test as well.

Total metabolite pools were first examined using multivariate analyses in Primer-e v7 (Clarke and Gorley, 2006). Abundances of each metabolite from individual replicates were log transformed and clustered by nMDS analysis. For visualization, abundances of each metabolite (^{12}C +^{13}C peak areas normalized by cell number) were averaged within each treatment and log transformed for heatmaps generated in Graphpad Prism. Statistical analyses between total abundances of metabolites were determined by a One-way ANOVA (as described above). Percent incorporation was determined from ^{13}C : ^{12}C ratio for individual metabolites. Statistical differences in percent incorporation were determined as described above.

Results

During algal bloom season in the Western basin of Lake Erie, real-time data was collected from buoys as a part of the GLERL's efforts to monitor algal blooms and hypoxia. We analyzed data collected every 15 min from three stations (WE2, WE4 and WE8, Figure 2) during the summer months to examine relationships with pH at station WE2 and WE 4 for the period of July through September of 2015 and WE8 in August (data was not collected at this station outside of August). pH and phycocyanin were variable across the summer months, both tended to be lower earlier in the summer (July) and gradually increased throughout August and September (Figure 3). Phycocyanin increased linearly with pH with the strongest correlation being at station WE2 located at the mouth of Maumee Bay (R^2=0.87 at WE2; R^2=0.75 at WE4; R^2=0.54 at WE8; Figure 3). The average daily pH ranged from ~8.2 to 9.5, and peak phycocyanin values were observed at average pH values of 9.2 and higher.

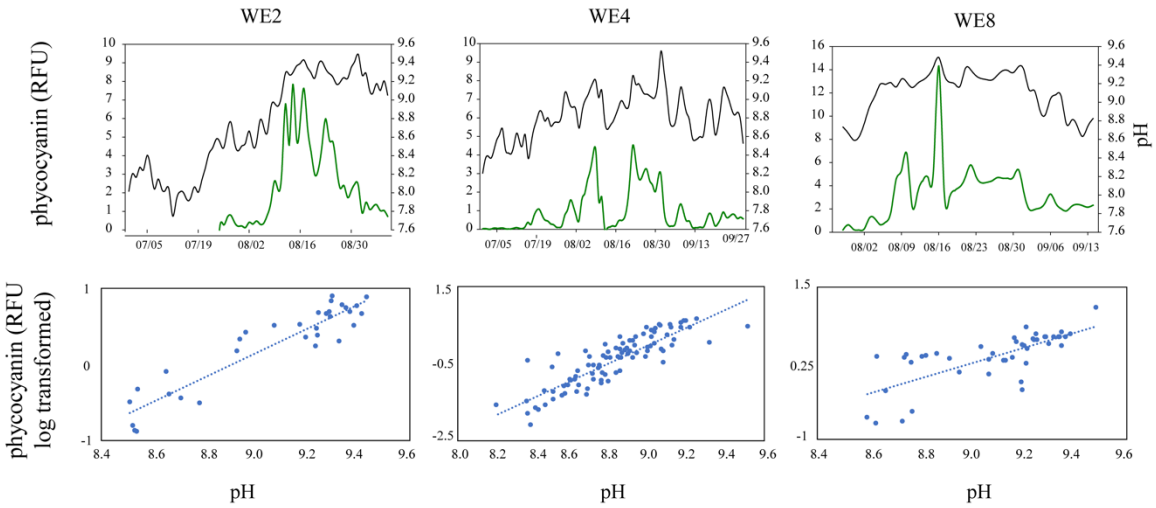

Figure 4.3 Relationships between pH and phycocyanin (a proxy for cyanobacterial biomass) at three monitoring buoys in the Western basin of Lake Erie

The green line represents changes in phycocyanin and the black line represents changes in pH.

Differences in growth rates of *M. aeruginosa* were driven by both the chemistry of available N and the pH of the growth medium (Two-way ANOVA, $p < 0.0001$) (Figure 4). The growth rate of *M. aeruginosa* grown on urea significantly increased (One-way ANOVA, $p < 0.05$) with increasing pH until 8.7 and 9.2 (Figure 4a, d). Growth on ammonium was optimal at pH of 8.2 (One-way ANOVA, $p < 0.0001$; Figure 4c, f)) while growth rates on nitrate were not affected by pH (One-way ANOVA, $p < 0.232$; Figure 4b, e). However, the most biomass accumulation was observed at the highest pH values regardless of N chemistry, with urea and nitrate reaching an FSU greater than 600 and average FSU on ammonium reached greater than 400. Cultures maintained at lower pH, particularly at 7.7, resulted in an accumulation of less biomass on all N chemistries.

To determine if *M. aeruginosa* incorporated C during the breakdown of urea, cells were grown on ^{13}C-urea until mid-log phase, harvested, and examined by LCMS to determine if the label was incorporated into the metabolites. Three different pH values were used (final pH after autoclaving: 7.5, 8.4 and 9.5) to test the hypothesis that *M. aeruginosa* incorporated more C per cell from urea when dissolved inorganic carbon in the media was low. The pH range was environmentally relevant and captured values observed across a typical bloom event. Thirty total metabolites were confidently identified from a database of 262 metabolites (Figure 5), and metabolomic profiles differed between treatments (Figure 6). Glycerol-3-phosphate was responsible for over 50% of the dissimilarity between the treatments (SIMPER) and this metabolite was more abundant at pH 7.7 relative to pH 8.4 and 9.5 (Figure 5). Other metabolites involved in central C metabolism and amino acid biosynthesis were also more abundant at pH 7.7 compared to 8.4 and 9.5 as well (Figure 5): notably, arginine and citrulline, which intermediates of the urea cycle.

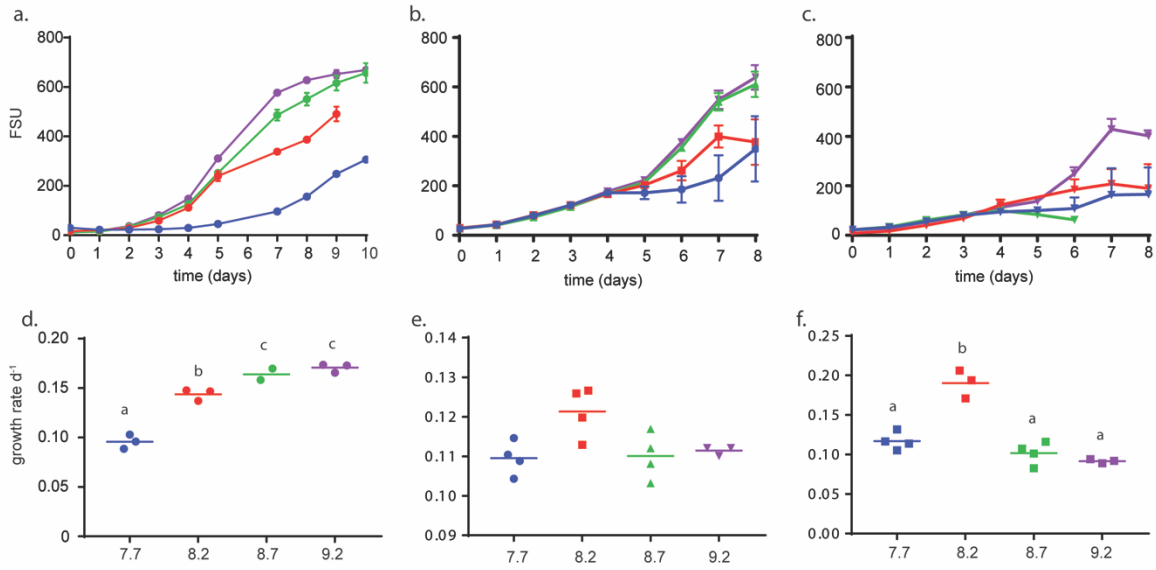

Figure 4.4 Growth dynamics on urea (a, d), nitrate (b, e), and ammonium (c, f) as the sole N source at varying pH values (7.7 in blue, 8.2 in red, 8.7 in green, 9.2 in purple)

Growth rates calculated by ln(fluorescence units) in exponential phase and are presented below each growth curve. Statistical differences in growth rates are represented as letters (a, b, c). Different lettering indicates statistically different growth rates within that N treatment. FSU = fluorescence units.

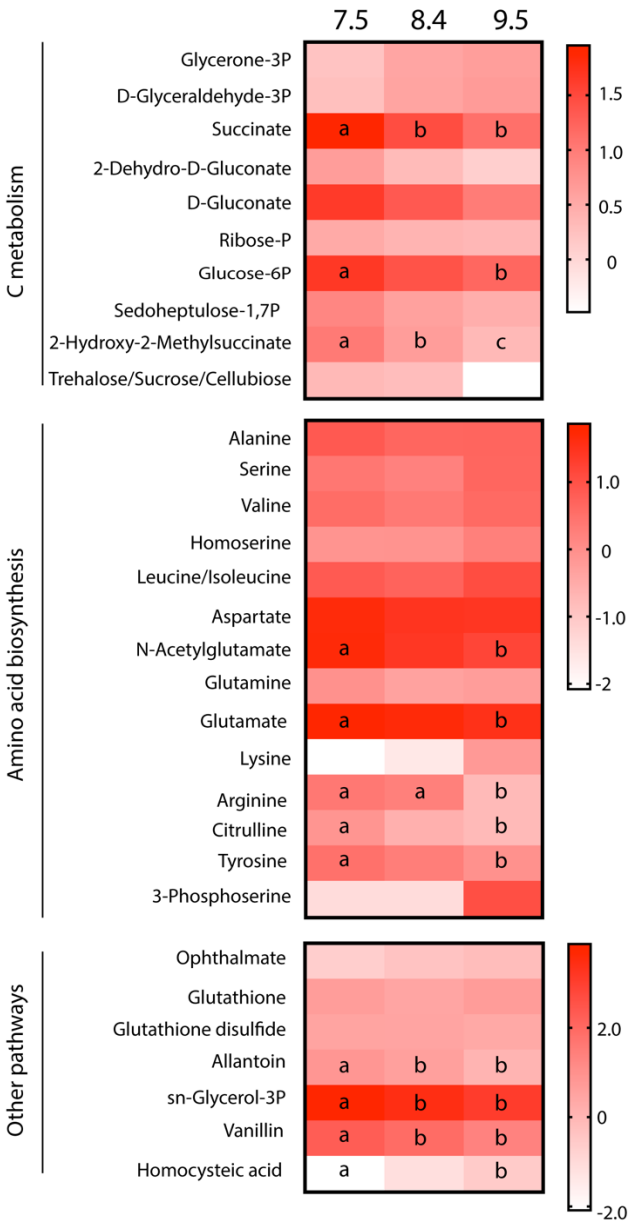

Figure 4.5 Heatmap of average metabolite abundances in *M. aeruginosa* cells growing at pH values of 7.5, 8.4, and 9.5

Heatmap of average metabolite abundances n *M. aeruginosa* cells growing at pH values of 7.5, 8.4, and 9.5 (columns left to right) normalized by cell number. Letters a and b are indicative of statistically different metabolites abundances per cell in that metabolite (one-way ANOVA, p value <0.05). If there were no statistically significant relationships between any treatment, no letter is present. Abundances of each metabolite per cell were log transformed for visualization purposes.

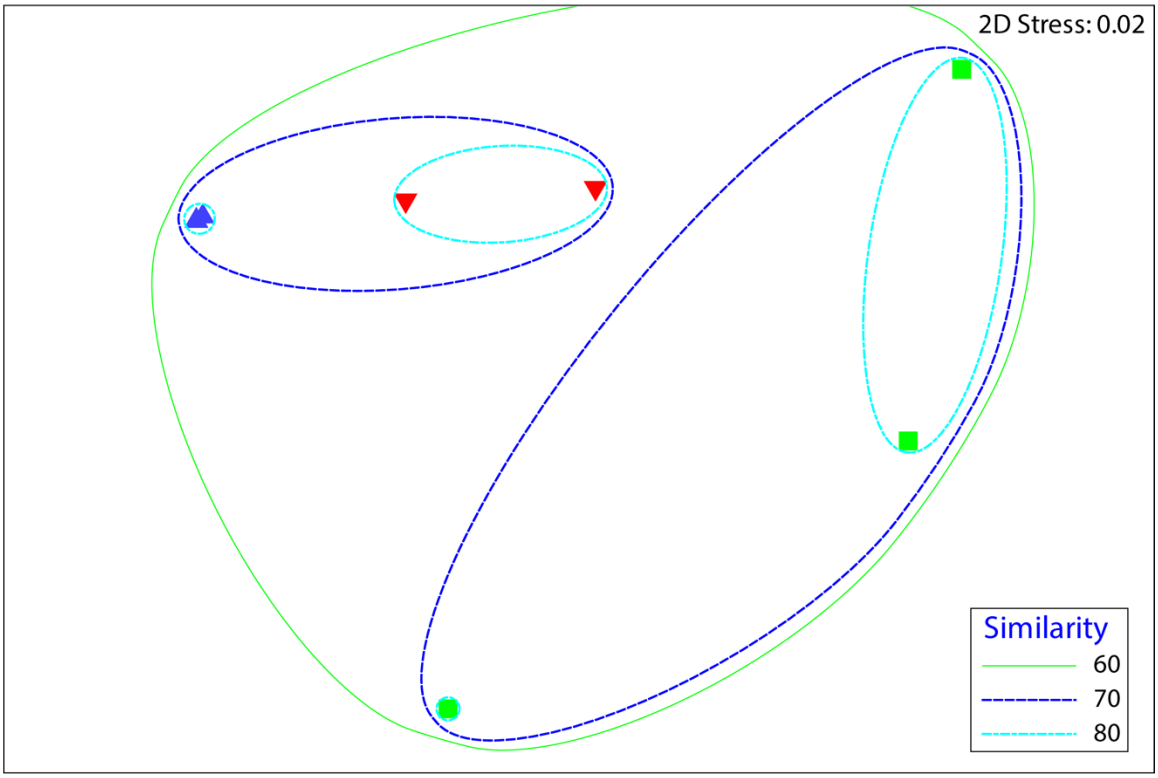

Figure 4.6 NMDS describing relationships between metabolites from *M. aeruginosa* NIES843 when growing at different pH

Total abundances for metabolites were normalized by cell number, log(x+1) transformed and clustered using Bray-Curtis similarity. 2D stress = 0.02; blue = pH 7.5; red = pH 8.4; green = pH 9.5

Glucose-6P and succinate were the only central C cycling metabolites with differences in relative abundances across pH values. Relative abundances of succinate differed across all treatments and were highest at a pH of 7.5 and lowest at 9.5. Abundances of glucose-6P were also highest at a pH of 7.5, but only statistically differed from cells grown at 9.5. While glutamine was not statistically different across treatments of varying pH, glutamate was statistically more abundant at the lowest pH. Glutamate:glutamine was not statistically significant across cultures of varying pH ($p = 0.4393$).

Of the 30 metabolites detected, ^{13}C incorporation was observed in 23 of them. Most metabolites with ^{13}C incorporation were involved in amino acid biosynthesis and central C metabolism (Figure 7, 8). Specifically, several metabolites with ^{13}C signatures are key intermediates in the Calvin cycle, glycolysis and the pentose phosphate pathway (Figure 7). ^{13}C was incorporated to metabolites at all pH values, but several metabolites from both central C metabolism and amino acid biosynthesis pathways were more enriched in ^{13}C at higher pH values: valine, glutamate, glutamine, aspartate, glucose 6-P, and sedoheptulose-1,7-BP (Figure 7). Arginine was the only urea cycle intermediate that incorporated C from urea hydrolysis. Other metabolites with high percentages of incorporation included glutathione (GSH), glutathione disulfide (GSSG) and ophthalmate, while others had very low but detectable ^{13}C signatures, such as vanillin, sn-glycerol-3P and allantoin (Figure 9). ^{13}C incorporation into GSH was not different across varying pH, but GSH had little or undetectable incorporation at pH of 7.5 and 8.4, and an average of approximately 65% at pH of 9.5.

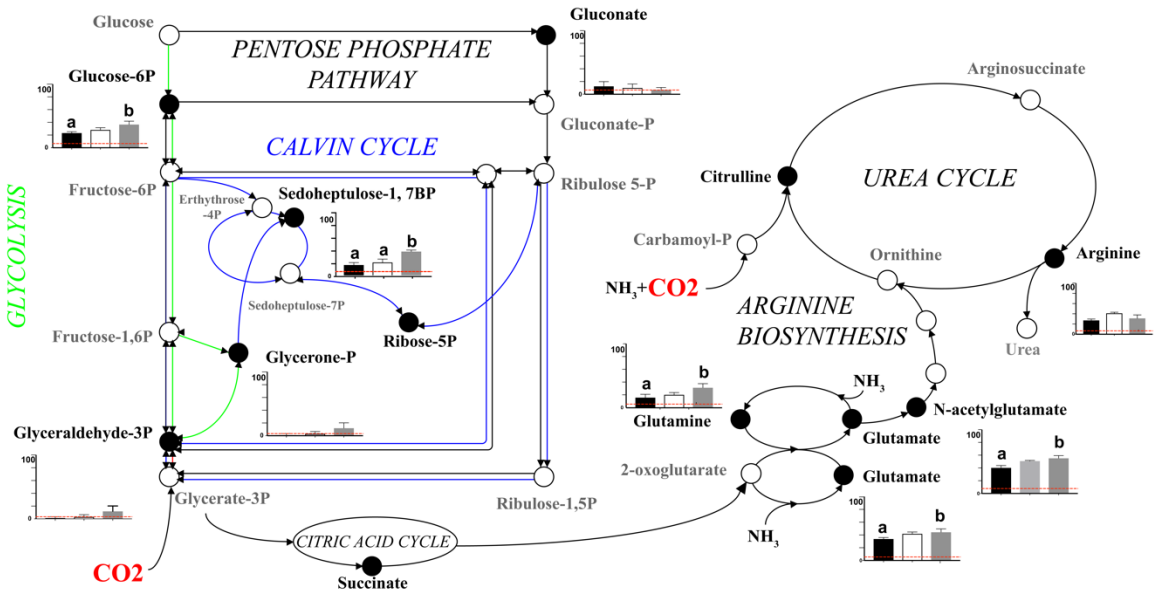

Figure 4.7 Metabolites detected across all pH treatments and pathways as possible entry points of C released from urea

Central carbon metabolism pathways in which labeled and unlabeled metabolites were detected (for clarity, glycolysis is in green and the Calvin cycle is in blue) are shown on the left. The arginine biosynthesis and urea cycle are shown on the right. Closed circles represent metabolites detected in this study and open circles are intermediates that were not detectable. Percent incorporation across treatments are shown in bar graphs between metabolites in which ^{13}C was detected (bar graphs are ordered by pH: 7.5, 8.4, 9.5 from left to right). If no incorporation graph is shown corresponding to a detected metabolite, this indicates no ^{13}C-labeling was detected in that metabolite at the time of sampling. Letters a and b are indicative of statistically different percentages of ^{13}C-incorporation for that metabolite (one-way ANOVA, p value <0.05). If there were no statistically significant relationships between any treatment, no letter is present.

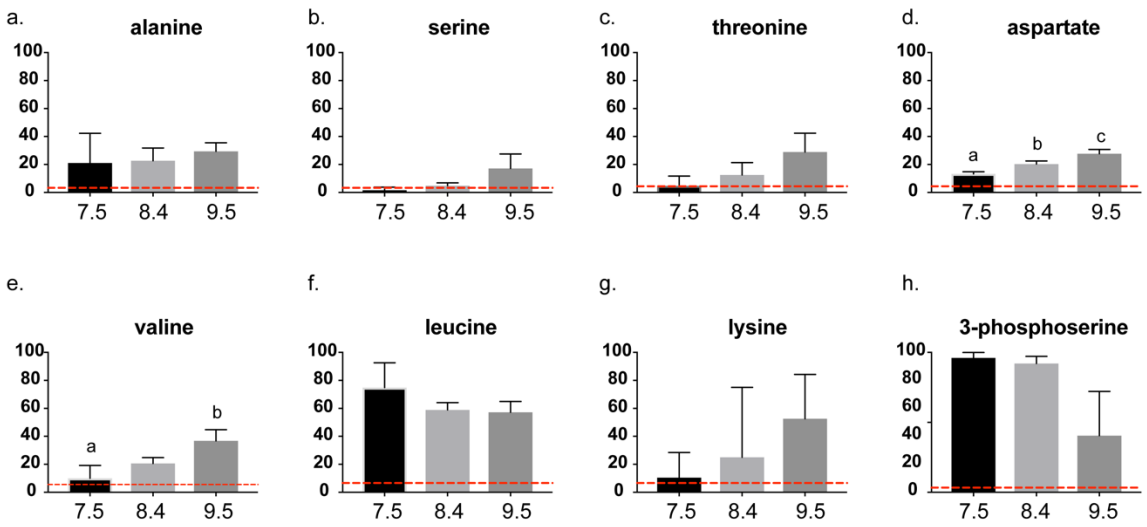

Figure 4.8 Percent ^{13}C incorporation from urea in amino acids and intermediates in amino acid biosynthesis pathways

Letters represent statistically significant differences between treatments and red dotted line is the baseline for naturally occurring ^{13}C for that metabolite.

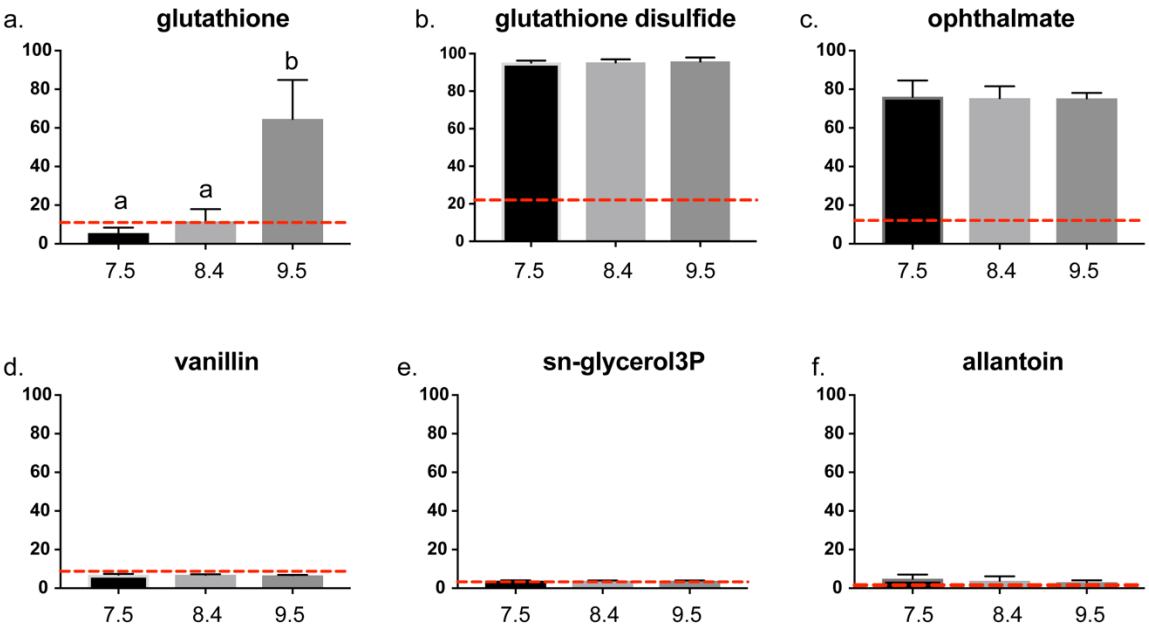

Figure 4.9 Percent ^{13}C incorporation in urea in other metabolites detected across pH treatments

Letters represent statistically significant differences between metabolites and red dotted line is the baseline for naturally occurring ^{13}C for that metabolite.

Discussion

During algal blooms, an increase in pH is commonly observed and is often attributed to the drawdown of CO_2 in the water column by the photosynthetic community. This was clearly demonstrated in the western basin of Lake Erie during the 2015 *Microcystis* bloom (Figure 2). The relationship between cyanobacterial abundance and pH was strongly and positively correlated, implying that the pH increased concomitantly with cyanobacterial abundances. This agrees with the general observation that there is an overall rise in pH over bloom events potentially due to the drawn down of dissolved CO_2 (O'neil *et al.*, 2012). Moreover, these observations and those from lab cultures are consistent with the hypothesis that increasing pH during bloom events triggers a positive feedback loop: algal blooms (including cyanobacteria) lead to increases in pH which lead to more prolific cyanobacterial growth.

Cultures of *M. aeruginosa* in the lab, grown in buffered medium, grew well at high pH and accumulated more biomass (Figure 3). Further, growth rates on different forms of N were affected by pH. In agreement with previous observations that cyanobacteria, including *Microcystis* spp. prefer ammonium as a N source (Kappers, 1980, Glibert *et al.*, 2016). *M. aeruginosa* had the highest growth rates on ammonium but only at a pH of 8.2. When grown at pH 9.2 with ammonium as a N source, biomass plateaued sooner and accumulated less. This same pattern was also observed in cells grown on nitrate. In contrast, when *M. aeruginosa* was maintained on urea as the sole N source, growth rates and biomass accumulation increased with pH. These observations imply that exogenous N inputs, especially those rich in urea, and the gradual alkalinity changes during bloom or late bloom season, play an important role in the ability of *Microcystis* to proliferate for such long periods of time in nature.

During the 2015 bloom, the average daily pH reached ~9.5 at stations WE2 and WE4 on Aug 31 and Sept 1, respectively and at WE8 on Aug 16. These peaks also corresponded to the highest peak in cyanobacterial biomass. If just the effect of pH is considered, this would mean that over 95% of the dissolved inorganic C pool is in the form of bicarbonate (HCO_3) and carbonate (CO_3) (Wetzel, 2001). This suggests that the *Microcystis* species in this bloom was more competitive during the low dissolved CO_2 concentrations. This also offers another explanation for the competitive advantage of *Microcystis* species over other phytoplankton. For example, winter and spring diatom blooms often occur in Lake Erie and are major contributors to productivity and hypoxia that follows bloom events (Bullerjahn *et al.*, 2016, Reavie *et al.*, 2016). Cyanobacteria are generally thought to outcompete these taxa and succeed in high pH/low dissolved CO_2 concentrations, though this may be species specific and consequence of other factors such as buoyancy during stratifying conditions (O'neil *et al.*, 2012, Ji *et al.*, 2017). However, these observations suggest that cyanobacteria are better competitors in these conditions.

Cyanobacteria are able to be successful in low dissolved CO_2 because of effective CO_2 concentrating mechanisms (CCMs) that allow them to transport HCO_3 into the cell when external CO_2 concentrations are low. There are five different inorganic C uptake systems in cyanobacteria: the BCT1 system, an ATP dependent complex with a high affinity for HCO_3; SbtA, a Na^+ dependent protein with a high affinity for HCO_3; BicA, also a Na^+ dependent protein with a low affinity for HCO_3; $NDH1_3$, which is constitutive and converts CO_2 that diffused into the cell to HCO_3 requiring NADPH; and $NDH1_4$, which is inducible and also converts CO_2 to HCO_3 via NADPH (Price, 2011). CCMs not only vary between different cyanobacteria (Badger *et al.*, 2005) but even with different strains of *Microcystis* spp. (Sandrini *et al.*, 2014). All species of *Microcystis* have the NDH1 complexes for converting CO_2 to HCO_3 in the cytoplasm, but vary in the presence

of *bicA*, *sbtA* and BCT1 systems. Even among *Microcystis* strains that possess the same uptake systems, their function and response to inorganic C varies as well (Sandrini *et al.*, 2015). These observations along with the environmental and growth data presented in this study may explain the fluctuation of different genotypes throughout a summer bloom that is usually attributed to nutrients like N and P (Xu *et al.*, 2010, Deng *et al.*, 2014). Our model strain, *M. aeruginosa* NIES843, contains the HCO_3 uptake systems SbtA, which requires sufficient sodium concentrations, and BCT1, which requires 1 ATP per HCO_3 to be functional (Sandrini *et al.*, 2014).

Of particular interest in the current study is the role of urea during times of low dissolved CO_2. *Microcystis* can not only utilize urea as a N-source, but there is also potential for it to be used as an inorganic C source. This has been suggested to be the case in the literatur but has yet to be experimentally confirmed (Finlay *et al.*, 2010, Glibert *et al.*, 2016). *M. aeruginosa* hydrolyses urea by means of the urease enzyme (Figure 9). The mechanism of urea degradation by urease has long been debated and those proposed are shown in Figure 10. The first reaction involves the catalyzed hydrolysis of urea to NH_3 and carbamate. Carbamate is considered unstable and spontaneously dissociates into another molecular of NH_3 and CO_2 (Zimmer, 2000, Mikkelsen *et al.*, 2010). Another mechanism of urea degradation, referred to as the elimination method, involves the uncatalyzed degradation of urea by urease and yields NH_3 and cyanate (Krajewska, 2009). The elimination method was considered very slow and not competitive with hydrolysis until recently (Barrios & Lippard, 2001, Krajewska, 2009). While this remains controversial as the elimination method has not been confirmed experimentally and only by *in silico* models, the end products would still be the same as hydrolytic decomposition of urea (Krajewska, 2009). Cyanate will either eventually dissociate spontaneously to $2NH_3$ and CO_2, or by action of the enzyme cyanase, hydrolyzes cyanate to produce the same end products (Palatinszky *et al.*, 2015). *M. aeruginosa* has

a copy of the gene encoding cyanase (MAE_RS04540), suggesting it carries the ability do this. Regardless of the mechanism, this would suggest that ultimately the point of entry of C from urea would likely be *via* C fixation (*i.e.* the Calvin Cycle). If this is true, then growing on urea may be cost effective to the cell also under C-limitation, because it is freely given rather than having to be transported into the cell using CCMs.

Another possible entry point of C is through phosphorylation *via* ATP of carbamate to yield carbamoyl-P by the enzymes carbamate kinase or carbamoyl phosphate synthetases (Kim & Raushel, 2004). Carbamoyl-P feeds into the arginine biosynthesis pathway and subsequent purine metabolism pathways (see Figure 7). *M. aeruginosa* NIES843 carries genes for the large and small subunits for the latter protein (MAE_RS21880, MAE_RS12410), suggesting this mechanism for C incorporation is also possible.

The hypothesis that *Microcystis* can assimilate C from urea breakdown is supported by the detection of ^{13}C in the majority of the detected metabolites. The detection of ^{13}C in several metabolites that are part of the Calvin cycle and other central C metabolism pathways (glycolysis and the pentose phosphate pathway) strongly suggest that the entry of C from urea is through C fixation. Unlabeled Carbamoyl-P was not detectable, however could have been missed due to the growth stage in which the cultures were processed. Because of the tight coupling between the N and C cycle, ^{13}C incorporation into glutamate, glutamate and N-acetylglutamate, along with detectable citrulline is indicative that C from urea entered via central carbon metabolism rather than this mechanism. The number of labeled amino acids and intermediates in their biosynthesis

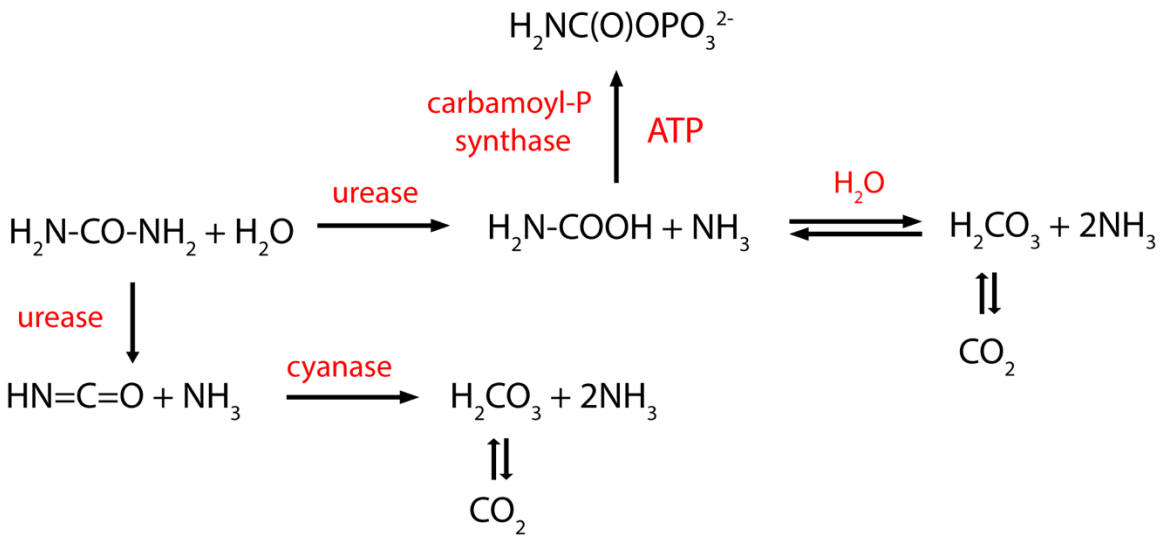

Figure 4.10 Proposed pathways of urea degradation and CO_2 assimilation by *M. aeruginosa* NIES843

suggest the C is likely being recycled through amino acid biosynthesis pathways and was likely the route for incorporation into arginine, which is an important metabolite for production of proteins, the N-rich storage molecule cyanophycin and several congeners of microcystin (Tillett *et al.*, 2000, Llácer *et al.*, 2008). Whether or not urea was degraded to carbamate or cyanate was not clear from our data, however it has been reported previously that the addition of urea in water collected from a *Microcystis* bloom, *Microcystis* has increased expression of cynS (cyanase) (Harke *et al.*, 2015) which could indicate cyanate may be an important intermediate in urea degradation by *Microcystis*. Cyanate hydrolase gene expression was also observed in our study on Lake Taihu in 2016 (see Chapter 3).

^{13}C from urea was incorporated into biomass in all cultures regardless of pH, but several metabolites had higher percentage of incorporation at higher pH values. It is important to note that this observation could simply be a factor relating to the total pool of a metabolite in each condition at the time of sampling. For example, N-acetylglutamate is higher in abundance at pH of 7.5 and lower at pH of 9.5 (Figure 5), and there is a greater percentage of ^{13}C incorporation at 9.5 (Figure 6). Incorporation percentages could be misleading because the same total C could have been incorporated, but the abundances of individual metabolites differed. This is the case for glutamate and glucose-6P as well. To address these, we examined metabolites with equal abundances per cell across all pH treatments to determine if there was higher incorporation of labeled –C at higher pH. We observed enrichment of ^{13}C for several metabolites (valine, glutamine, sedoheptulose-1,7-BP, aspartate and glutathione) at the higher pH, implying increased incorporation into metabolites.

Our data provide evidence that C from urea degradation is consumed by *M. aeruginosa* and potentially at higher rates in more alkaline conditions. Metabolically, growth on urea as the

sole N source yields 2 moles of N and 1 of C. If a Redfield ratio of C:N of 7:1 is assumed to be a conserved quota for *Microcystis* cells (Redfield, 1958), it can be concluded that each mole of urea can support 7.6% of a cell's C requirements. Since BCT1 requires 1 mole of ATP per HCO_3 molecule (Price, 2011), each urea-derived C molecule saves 1 ATP, suggesting this mechanism for obtaining inorganic C should be energetically favorable. This phenomenon may provide an additional advantage to *Microcystis* during times of low CO_2 in the environment. However, C is incorporated at all pH values, suggesting these data suggest that urea acts as a supplemental C source regardless of dissolved CO_2.

Another metabolite with more ^{13}C incorporation at the highest pH was GSH. At a pH of 9.5, approximately 60% of the total pool of C from GSH was labeled, whereas at the lower pH ^{13}C incorporation from urea was very little or completely absent. GSH abundances however did not differ between treatments which highlights the importance of this metabolite in *Microcystis* physiology regardless of pH. Although GSH is likely involved in several metabolic processes within the cell, in phototrophs it is an important antioxidant (Smirnova & Oktyabrsky, 2005, Cameron & Pakrasi, 2010, Noctor *et al.*, 2012, Cassier-Chauvat & Chauvat, 2014, Narainsamy *et al.*, 2016). In cyanobacteria, the ratio and concentrations of GSH and its oxidized form GSSG help to balance to redox state and buffering capacity of the cell (Smirnova & Oktyabrsky, 2005). One of the benefits of CCMs and subsequent increased CO_2 is the contribution to reducing photoinhibitory effects under low CO_2 conditions by inhibiting photorespiration (Tabita, 1994). A higher percent of incorporation of ^{13}C into GSH may be indicative that the C from urea is valuable to cells at a higher pH by reducing the cost to maintain a cellular redox balance.

In addition to the differences observed in ^{13}C incorporation at varying pH, total metabolite pools differed between treatments. There are a variety of mechanisms that facultative and obligate

alkaliphilic microorganisms possess to maintain a neutral internal pH, including increasing the production of polyamines and acidic amino acids, regulation of Na^+ transport, and managing a unique lipid composition to regulate fluidity of the membrane (Krulwich, 2006). Interestingly, all statistically significant differences in metabolites in *M. aeruginosa* cells grown at different pH values were higher in abundance at a lower pH, except for one. While the explanation of this general trend is unclear, the differences in glycerol-3P, an intermediate in glycerolipid metabolism, between low and high pH suggest may a role in regulation of lipid composition or concentration to adapt to these conditions. Homocysteic acid is the oxidized form of homocysteine (Jakubowski, 2013). This metabolite is mostly studied in humans for its role in various diseases as a glutamate analog, and it has been shown to be toxic to *E. coli* (Essenberg, 1984) by inhibiting its ability to use glutamate as a sole carbon source. In human cell lines, homocysteic acids causes oxidative stress (Arzumanyan *et al.*, 2008) and can be the result of photooxidation of methionine (Bern *et al.*, 2010). The higher abundances of homocysteic acid at high a pH could be suggestive that the redox state of *M. aeruginosa* is more unstable when grown at a high pH and may explain the higher incorporation of ^{13}C in glutathione.

Conclusions

The use of urea as a N fertilizer in agriculture has increased rapidly over the past several decades. Its role in the eutrophication is important as it is commonly found in freshwater systems and can serve a N source to many bloom-forming cyanobacteria, like *Microcystis* spp. Here, we show using a model toxic cyanobacterium that growth rates and biomass accumulation are higher when urea is source of N at high pH values commonly observed during blooms. We also demonstrated that urea can be used a source of C by *Microcystis aeruginosa* NIES843. Metabolites

in the Calvin cycle and other central carbon pathways had ^{13}C signatures supporting that the CO_2 was being fixed and incorporated into amino acids and other metabolic pathways but assimilation of C in the form of the intermediate carbamate cannot be ruled out. There is likely preferential incorporation of C at higher pH values, when dissolved CO_2 in the water is low or absent, which may aid in managing the redox state and contribute to the success of *Microcystis* for prolonged periods of time. In addition to being an important N source, especially at high pH, we propose that urea can be an energetically favorable C source that helps support metabolic homeostasis and photosynthetic efficiency of *Microcystis* during dense blooms. While this physiological mechanism may not fully explain how *Microcystis* manages to dominate planktonic communities during late summer months in temperate regions, it does add an intriguing and relevant piece to the puzzle.

Acknowledgements

This work was made possible due to grant NA11NOS4780021 from NOS National Center for Coastal Ocean Science (NCCOS) under their Harmful Algal Bloom Prevention Control and Mitigation program to SWW and GLB, and National Science Foundation awards (DEB-1240870; IOS-1451528) to SWW. The authors also acknowledge the generous support of the *Kenneth & Blaire Mossman* endowment to the University of Tennessee.

References

Arzumanyan E., Makhro A., Tyulina O. & Boldyrev A. (2008) Carnosine protects erythrocytes from the oxidative stress caused by homocysteic acid. 418, 44-46. Springer.

Badger M. R., Price G. D., Long B. M. & Woodger F. J. 2005. The environmental plasticity and ecological genomics of the cyanobacterial CO_2 concentrating mechanism. *Journal of experimental botany*. 57, 249-265.

Barrios A. M. & Lippard S. J. 2001. Decomposition of alkyl-substituted urea molecules at a hydroxide-bridged dinickel center. *Inorganic chemistry*. 40, 1250-1255.

Belisle B. S., Steffen M. M., Pound H. L., Watson S. B., DeBruyn J. M., Bourbonniere R. A., Boyer G. L. & Wilhelm S. W. 2016. Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*. 42, 599-607.

Bern M., Saladino J. & Sharp J. S. 2010. Conversion of methionine into homocysteic acid in heavily oxidized proteomics samples. *Rapid Communications in Mass Spectrometry*. 24, 768-772.

Bogard M. J., Donald D. B., Finlay K. & Leavitt P. R. 2012. Distribution and regulation of urea in lakes of central North America. *Freshwater Biology*. 57, 1277-1292.

Bullerjahn G., Phuntumart V., Wilhelm S. & McKay R. 2016. Adaptations to photoautotrophy associated with seasonal ice cover in a large lake revealed by metatranscriptome analysis of a winter diatom bloom. *Journal of Great Lakes Research*. 42, 1007-1015.

Cameron J. C. & Pakrasi H. B. 2010. Essential role of glutathione in acclimation to environmental and redox perturbations in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant physiology*. 154, 1672-1685.

Cao P., Lu C. & Yu Z. 2018. Historical nitrogen fertilizer use in agricultural ecosystem of the continental United States during 1850–2015: Application rate, Timing, and fertilizer types. *Earth System Science Data.* 10.2, 969

Cassier-Chauvat C. & Chauvat F. 2014. Responses to oxidative and heavy metal stresses in cyanobacteria: recent advances. *International journal of molecular sciences*. 16, 871-886.

Chaffin J. D. & Bridgeman T. B. 2013. Organic and inorganic nitrogen utilization by nitrogen-stressed cyanobacteria during bloom conditions. *Journal of Applied Phycology*. 26, 299-309.

Chaffin J. D., Davis T. W., Smith D. J., Baer M. M. & Dick G. J. 2018. Interactions between nitrogen form, loading rate, and light intensity on *Microcystis* and *Planktothrix* growth and microcystin production. *Harmful algae*. 73, 84-97.

Davis T. W., Harke M. J., Marcoval M. A., Goleski J., Orano-Dawson C., Berry D. L. & Gobler C. J. 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Aquatic Microbial Ecology*. 61, 149-162.

Deng J., Qin B., Paerl H. W., Zhang Y., Wu P., Ma J. & Chen Y. 2014. Effects of nutrients, temperature and their interactions on spring phytoplankton community succession in Lake Taihu, China. *PloS one*. 9, e113960.

Essenberg R. C. 1984. Use of homocysteic acid for selecting mutants at the *gltS* locus of *Escherichia coli* K12. *Microbiology*. 130, 1311-1314.

Finlay K., Patoine A., Donald D. B., Bogard M. J. & Leavitt P. R. 2010. Experimental evidence that pollution with urea can degrade water quality in phosphorus-rich lakes of the Northern Great Plains. *Limnology and Oceanography*. 55, 1213-1230.

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

Glibert P. M., Maranger R., Sobota D. J. & Bouwman L. 2014. The Haber Bosch–harmful algal bloom (HB–HAB) link. *Environmental Research Letters*. 9, 105001.

Glibert P. M., Wilkerson F. P., Dugdale R. C., Raven J. A., Dupont C. L., Leavitt P. R., Parker A. E., Burkholder J. M. & Kana T. M. 2016. Pluses and minuses of ammonium and nitrate uptake and assimilation by phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions. *Limnology and Oceanography*. 61, 165-197.

Harke M. J., Davis T. W., Watson S. B. & Gobler C. J. 2015. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environmental science & technology*. 50, 604-615.

Ho J. C. & Michalak A. M. 2017. Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *Journal of Great Lakes Research*. 43, 221-228.

Jakubowski H. 2013. An overview of homocysteine metabolism. *Homocysteine in Protein Structure/Function and Human Disease,* 7-18. Springer.

Ji X., Verspagen J. M. H., Stomp M. & Huisman J. 2017. Competition between cyanobacteria and green algae at low versus elevated CO_2: who will win, and why? *Journal of experimental botany*. 68, 3815-3828.

Kappers F. I. 1980. The cyanobacterium *Microcystis aeruginosa* Kg. and the nitrogen cycle of the hypertrophic Lake Brielle (The Netherlands). *Hypertrophic ecosystems,* 37-43. Springer.

Kim J. & Raushel F. M. 2004. Access to the carbamate tunnel of carbamoyl phosphate synthetase. *Archives of Biochemistry and Biophysics*. 425, 33-41.

Koper T. E., El-Sheikh A. F., Norton J. M. & Klotz M. G. 2004. Urease-encoding genes in ammonia-oxidizing bacteria. *Applied and Environmental Microbiology*. 70, 2342-2348.

Krajewska B. 2009. Ureases I. Functional, catalytic and kinetic properties: A review. *Journal of Molecular Catalysis B: Enzymatic*. 59, 9-21.

Krausfeldt L. E., Tang X., van de Kamp J., Gao G., Bodrossy L., Boyer G. L. & Wilhelm S. W. 2017. Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu. *FEMS Microbiology Ecology*. 93,

Krulwich T. A. 2006. Alkaliphilic prokaryotes. *The Prokaryotes: Volume 2: Ecophysiology and Biochemistry*. 283-308.

Llácer J. L., Fita I. & Rubio V. 2008. Arginine and nitrogen storage. *Current opinion in structural biology*. 18, 673-681.

Mikkelsen M., Jørgensen M. & Krebs F. C. 2010. The teraton challenge. A review of fixation and transformation of carbon dioxide. *Energy & Environmental Science*. 3, 43-81.

Narainsamy K., Farci S., Braun E., Junot C., Cassier-Chauvat C. & Chauvat F. 2016. Oxidative-stress detoxification and signalling in cyanobacteria: the crucial glutathione synthesis pathway supports the production of ergothioneine and ophthalmate. *Molecular microbiology*. 100, 15-24.

Noctor G., Mhamdi A., Chaouch S., Han Y., Neukermans J., Marquez-Garcia B., Queval G. & Foyer C. H. 2012. Glutathione in plants: an integrated overview. *Plant, Cell & Environment*. 35, 454-484.

O'neil J., Davis T., Burford M. & Gobler C. 2012. The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful algae*. 14, 313-334.

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

Paerl H. W. & Huisman J. 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental microbiology reports*. 1, 27-37.

Paerl H. W., Hall N. S. & Calandrino E. S. 2011. Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Science of the Total Environment*. 409, 1739-1745.

Paerl H. W., Scott J. T., McCarthy M. J., Newell S. E., Gardner W., Havens K. E., Hoffman D. K., Wilhelm S. W. & Wurtsbaugh W. A. 2016. It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science & Technology*.

Palatinszky M., Herbold C., Jehmlich N., Pogoda M., Han P., von Bergen M., Lagkouvardos I., Karst S. M., Galushko A. & Koch H. 2015. Cyanate as an energy source for nitrifiers. *Nature*. 524, 105-108.

Peng G., Martin R., Dearth S., Sun X., Boyer G. L., Campagna S., Lin S. & Wilhelm S. W. 2018. Seasonally-relevant cool temperatures interact with N chemistry to increase microcystins produced in lab cultures of *Microcystis aeruginosa* NIES-843. *Environmental science & technology*.

Price G. D. 2011. Inorganic carbon transporters of the cyanobacterial CO 2 concentrating mechanism. *Photosynthesis Research*. 109, 47-57.

Reavie E. D., Cai M., Twiss M. R., Carrick H. J., Davis T. W., Johengen T. H., Gossiaux D., Smith D. E., Palladino D. & Burtner A. 2016. Winter–spring diatom production in Lake Erie is an important driver of summer hypoxia. *Journal of Great Lakes Research*. 42, 608-618.

Redfield A. C. 1958. The biological control of chemical factors in the environment. *American scientist*. 46, 230A-221.

Sandrini G., Jakupovic D., Matthijs H. C. & Huisman J. 2015. Strains of the harmful cyanobacterium *Microcystis aeruginosa* differ in gene expression and activity of inorganic carbon uptake systems at elevated CO_2 levels. *Applied and environmental microbiology*. 81, 7730-7739.

Sandrini G., Matthijs H. C., Verspagen J. M., Muyzer G. & Huisman J. 2014. Genetic diversity of inorganic carbon uptake systems causes variation in CO_2 response of the cyanobacterium *Microcystis*. *The ISME journal*. 8, 589.

Smirnova G. & Oktyabrsky O. 2005. Glutathione in bacteria. *Biochemistry (Moscow)*. 70, 1199-1211.

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, e44002.

Steffen M. M., Belisle B. S., Watson S. B., Boyer G. L., Bourbonniere R. A. & Wilhelm S. W. 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Applied and environmental microbiology*. 81, 3268-3276.

Steffen M. M., Dearth S. P., Dill B. D., Li Z., Larsen K. M., Campagna S. R. & Wilhelm S. W. 2014. Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. *The ISME journal*. 8, 2080-2092.

Tabita F. R. 1994. The biochemistry and molecular regulation of carbon dioxide metabolism in cyanobacteria. *The molecular biology of cyanobacteria,* 437-467. Springer.

Tillett D., Dittmann E., Erhard M., von Döhren H., Börner T. & Neilan B. A. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chemistry & biology*. 7, 753-764.

Visser P. M., Verspagen J. M. H., Sandrini G., Stal L. J., Matthijs H. C. P., Davis T. W., Paerl H. W. & Huisman J. 2016. How rising CO_2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae*. 54, 145-159.

Wetzel R. G. 2001. *Limnology: lake and river ecosystems*. gulf professional publishing,

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

Zimmer M. 2000. Molecular mechanics evaluation of the proposed mechanisms for the degradation of urea by urease. *Journal of Biomolecular Structure and Dynamics*. 17, 787-7

Chapter 5 : Toxicity and physiology as related to nitrogen chemistry as revealed by tracing ^{15}N through the metabolome of *Microcystis aeruginosa*

Publication note

This is chapter is a version of a manuscript in preparation by Lauren E. Krausfeldt, Abigail T. Farmer, Hector F. Castro Gonzalez, Shawn R. Campagna, and Steven W. Wilhelm.

My contribution to this work was assistance with experimental design, culture work, metabolite extraction, data analysis and writing the majority of the manuscript.

Abstract

Large nitrogen (N) inputs into the environment from anthropogenic activities and the proliferation of non-N_2-fixing cyanobacteria has led to increased attention on N as a driver of toxic cyanobacterial blooms. Several studies have shown an importance of N in the accumulation of biomass, succession of toxic to non-toxic species and toxin production. Cyanobacteria encounter many forms of exogenous N in the environment in either the organic (*i.e.* urea, amino acids) or inorganic (*i.e.* nitrate and ammonium) form, but despite the general importance of N, the specific influence of these forms on the success and toxicity of cyanobacteria still remains unclear. In this study, axenic cultures of the model toxic cyanobacteria *M. aeruginosa* were grown on either ^{15}N-nitrate, ammonium and urea as the sole N source in order to examine the specific effects of N form on the metabolism of cells growing exponentially. Core metabolic intermediates in ATP synthesis, nucleotide metabolism, peptidoglycan synthesis and glutathione metabolism were incorporated rapidly. Several metabolites in the arginine biosynthesis pathway and urea cycle were identified in cells growing on urea whereas cells grown on nitrate had very few and none were detected in cells grown on ammonium. Further, cells grown on urea incorporated N into microcystin and quotas were 100 to 1000x times higher than in cells grown on ammonium and nitrate, respectively. Our results suggest that while there are core metabolites that N feeds into across all treatments, urea is an important nutrient in driving intracellular MC concentrations. Overall, N chemistry influences *Microcystis* cell physiology and toxicity, and these data supports that increased N inputs from anthropogenic activities will indeed influence toxicity of blooms.

Introduction

Nitrogen (N) has recently emerged as an important factor constraining the success of toxic cyanobacteria in fresh waters (Paerl *et al.*, 2014). For several decades, the focus has been on phosphorus (P) as a driving nutrient in blooms because of the dominance of N_2-fixing cyanobacteria in blooms in the 1960s and 1970s (Schindler *et al.*, 2008, Conley *et al.*, 2009). N-fixing cyanobacteria do not depend on exogenous N sources, because they can directly convert dissolved N_2 gas to ammonia. Thus, reductions in N were considered to be unnecessary (Schindler *et al.*, 2008). However, studies have shown that N_2-fixation does not support bloom N demands (Scott & McCarthy, 2010, Willis *et al.*, 2016). In addition, in the past two decades the rise of non-N_2-fixing cyanobacteria has turned the focus to the dual role of N and P in cyanobacterial blooms (Paerl *et al.*, 2011, Paerl *et al.*, 2016). Non-N_2 fixing *Microcystis* spp. are often dominant bloom formers in lakes plagued by toxic blooms and depend on exogenous forms of N, like nitrate, ammonium, and urea (Harke *et al.*, 2016). Several studies have shown that both N and P are important in the formation and proliferation of *Microcystis* blooms (Paerl *et al.*, 2016). N chemistry and concentration have been suggested to be a contributing driver to the dominance of toxic or non-toxic *Microcystis* species and their succession throughout a bloom event (Davis *et al.*, 2010).

N availability has also been implicated as a key factor driving toxicity of cyanobacterial blooms (Gobler *et al.*, 2016). Microcystins (MCs) are potent hepatotoxins produced by several strains of cyanobacteria, including *Microcystis* spp. (Carmichael & Boyer, 2016). Because MCs are N-rich, N-availability to some extent must be important in microcystin production. However, whether or not MCs are produced in response to N concentrations specifically remains debated in the scientific literature (Gobler *et al.*, 2016). This uncertainty is likely in large part due to an incomplete understanding of the general function of MCs in cyanobacterial cells, including why

and when they are produced (Bullerjahn *et al.*, 2016). MC concentrations have been observed to increase in response to both increased (Harke & Gobler, 2015, Gobler *et al.*, 2016), and decreased (Ginn *et al.*, 2010, Harke *et al.*, 2015, Gobler *et al.*, 2016, Peng *et al.*, 2018) N-availability in cultures and natural systems.

Until recently, the majority of studies examining the effects of N on the growth and toxicity of *Microcystis* blooms have been performed using nitrate as the sole N source. This is likely due to difficulty in maintaining *Microcystis* cultures on ammonium, despite this being the preferred N source (Kappers, 1980). While nitrate is still a relevant source of N in freshwater systems, the switch to urea as a N fertilizer instead of those that are nitrate-based in agricultural practices over the past few decades has led to investigating the importance of urea as a contributing factor to toxic cyanobacterial blooms (Paerl *et al.*, 2016). Indeed, *Microcystis* grows well on both urea and nitrate and can produce MCs when grown on both forms in the lab and in the field (Steffen *et al.*, 2015, Belisle *et al.*, 2016, Peng *et al.*, 2018). Indeed, while urea is found in fresh waters directly influenced by agricultural practices (Glibert *et al.*, 2006), it is commonly ignored in surveys for N-loading where nitrate remains the focus. While the general implications linking N-loading to the rise of toxic cyanobacteria has become well-established, the effect of these different chemical forms of nitrogen on the physiology of *Microcystis* and implications for toxin production, are less characterized.

The goal of this study was to investigate the effect of N chemistry on the metabolism of the model toxic cyanobacteria *Microcystis aeruginosa* NIES843. Using ^{15}N stable isotopes of ammonium, urea and nitrate, N was traced through cell metabolites to characterize the physiological responses to different N chemistries and determine which biosynthetic pathways were important in the fate of N through the cell and into MC.

Methods

Growth on urea, NO3 and NH4

Axenic *Microcystis aeruginosa* NIES 843 was grown at 26°C on a diel cycle (12 h light/12 h dark) at 50-60 μmol photons m^{-2} s^{-1} of light in CT medium at pH = 8.2 ± 0.05 (Steffen *et al.*, 2014) with the same molar concentration of N in three different forms: urea, NH_4Cl, and a nitrate combination of KNO_3/$CaNO_3$*$4H_20$). Urea and NH_4Cl are toxic or significantly reduced the growth of *M. aeruginosa* at 2.34 mM N (the concentration of N in standard CT), so cultures were maintained at N concentrations of 0.585 mM. Cultures were routinely checked for contamination via microscopy after staining with SYBR green. Cultures were acclimated to experimental media for at least three transfers prior to the onset of experimentation. To initiate experiments, *Microcystis* cultures were concentrated on a 1.0 μm nominal pore-size polycarbonate filter to remove residual media and resuspended in fresh media prior to inoculations. Growth curves were performed at a starting concentration of ~75,000 cells/mL. Cell counts were determined by flow cytometry (Guava easyCyteHT, Millipore). Chlorophyll *a* autofluorescence (FSU) was measured daily using a Turner Designs TD-700 fluorometer (Sunnyvale, CA, USA). Both cell counts and fluorescence were collected daily during the logarithmic phase of growth.

15N incorporation experiment

For the ^{15}N-incorporation experiment, acclimated cells were grown on ^{14}N in 10 x 50mL culture tubes with 25mL of media until they reached mid-log phase (~300,000 - 400,000 cell/mL). Cells were concentrated on 1.0 μm polycarbonate filters and resuspended in 50 mL culture tubes with 30 mL of fresh ^{14}N media ~24 hours prior to the start of the experiment to keep cells in log

phase. Cultures were checked for contamination *via* microscopy using SYBR green. Cells were counted on a Guava flow cytometer and aliquots transferred to ^{15}N media (made with >98% ^{15}N saturated NH_4Cl, -urea, or $KNO_3/Ca(NO_3)_2$), with a starting concentration of ~1.15 x 10^5 cells/mL to reach a final volume of 25 mL in 50 mL glass culture tubes. For each nitrogen source twenty-five individual tubes were prepared and harvested at 6 h, 12 h, 24 h, 36 h, and 48 h (x5 for each time point with the exception of a few samples that were lost in processing) after inoculation. Five control tubes were prepped with 25 mL of ^{14}N media per treatment as well and inoculated at the same starting concentration to serve as time 0. Fluorescence and cell counts were collected as described previously for all tubes at time 0 to ensure growth was being observed when sampling occurred. At each time point, five tubes were randomly selected, fluorescence (FSU) was measured and 2 x 200 µL were collected for replicate cell counts. The remaining culture in each tube were harvested by filtration on a 1.0 µm nominal pore-size polycarbonate filter. Filters were flash frozen in 2 mL cryovials using liquid N and stored at -80°C until extraction of metabolites. Extraction of metabolites and metabolomics were performed as described in Peng *et al.* (2018).

Analysis

Relative metabolite abundances were calculated by adding the ^{14}N and ^{15}N peak areas and normalizing by cell number. Global metabolite trends ($^{14}N+^{15}N$ for each metabolite detected) were assessed using NMDS analyses to determine variability amongst replicates in Primer-E v7 (Clarke & Gorley, 2006). Pathways for metabolites with and without ^{15}N signatures were determined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa & Goto, 2000). Incorporation percent was calculated by dividing ^{15}N peak areas by sum of $^{15}N+^{14}N$ peak areas for each metabolite.

Results

Growth dynamics on different N sources

Axenic *M. aeruginosa* NIES843 grown on nitrate (NO_3), ammonium (NH_3), and urea as the sole N sources had an average doubling time of 1.26, 1.88, and 1.14 days respectively (Figure 1). Differences in doublings were not statistically significant ($p = 0.09$), however cells grown on urea and NO_3 accumulated more biomass overtime (Figure 1). For the ^{15}N experiment, cells were harvested and added to ^{15}N media during log phase and continued to grow exponentially over the 48 h (Figure 2a, b). A lag phase in the urea treatment was observed in the direct cell count data during the first 24 h, although fluorescence increased over this period. Over the course of 48 h, for cells grown on urea, fluorescence did increase, and cultures reached the same cell concentration as the other treatments, maintaining a maximum doubling time similar to those described above. Because of the potential variability that comes with the destructive sampling method utilized here, metabolite abundances were compared between replicates in each treatment by NMDS analysis (Figure 3a-c). Replicates for each time point per treatment clustered together with the exception of a few outliers in the NO_3-grown cells (Figure 3a) and these were removed from the analysis.

15N incorporation into metabolites

^{15}N incorporation was detected in all samples. All metabolites detected with ^{15}N across treatments were labeled within 6 h and remained labeled for the full 48 h. Several metabolites involved in amino sugar and nucleotide metabolism (AMP, UDP-glucose, UDP-N-

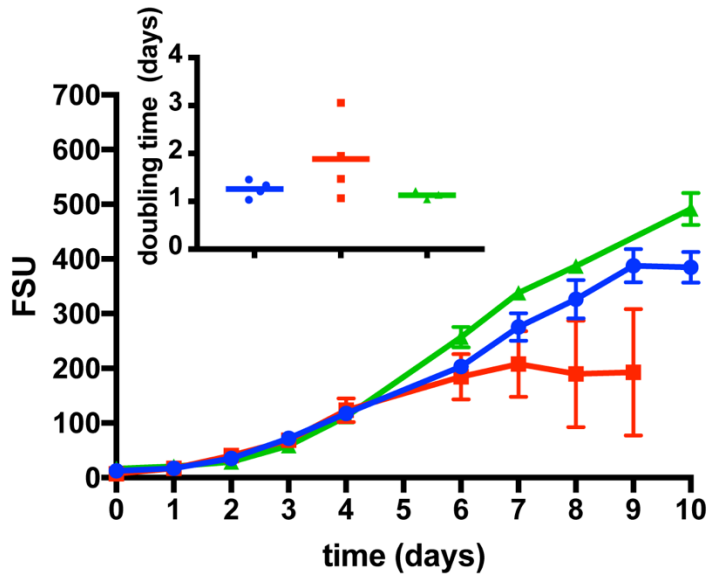

Figure 5.1 Growth of *M. aeruginosa* on nitrate, ammonium, and urea

Growth dynamics of *M. aeruginosa* by proxy of fluorescence on different N-forms (NO_3 = blue; NH_3 = red, urea = green). Doubling times were calculated based on cell concentration during exponential phase. FSU = fluorescence

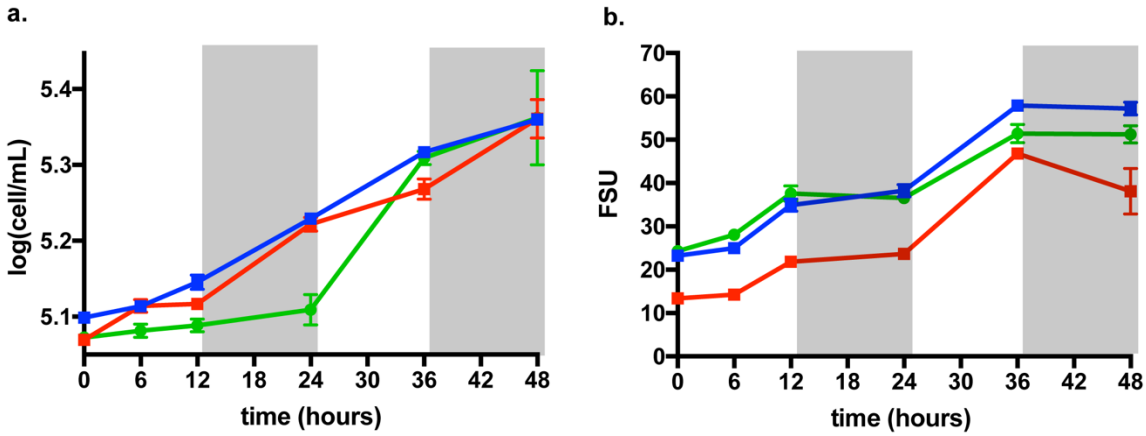

Figure 5.2 Growth of *M. aeruginosa* before and after ^{15}N additions

Growth *of M. aeruginosa* before (time 0) and after ^{15}N additions (NO_3 = blue; NH_3 = red, urea = green) by cell number (a) and fluorescence (b). Shading represents the dark cycle. FSU = fluorescence

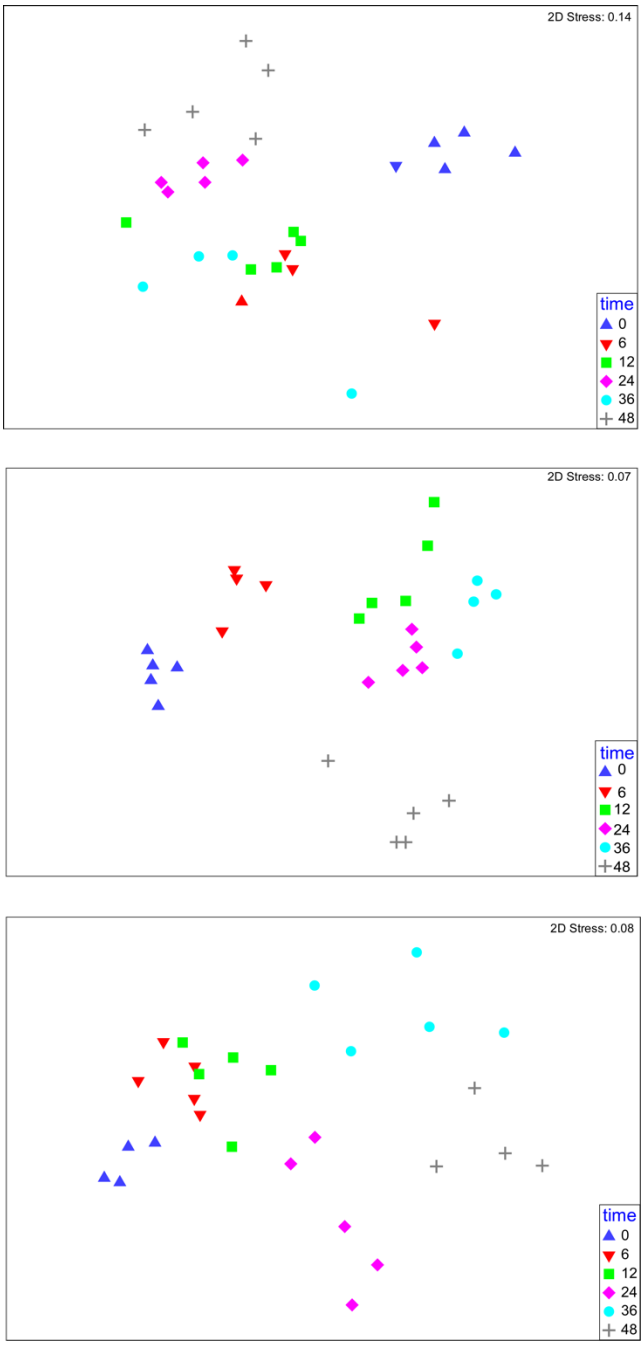

Figure 5.3 NMDS analysis of metabolite abundances across replicates in N treatment

Analysis of metabolite abundances for nitrate (a), urea (b), and ammonium (c) over time nmds. Abundances were log(x+1) transformed and clustered using Bray Curtis Similarity.
highest number of different labeled metabolites that were detected occurred in cells grown on urea.

Acetylglucosamine) and glutathione disulfide were 80-100% labeled by 6 h in the urea and NH_3 treatments and remained at 100% ^{15}N over 48 h (Figure 4a-d). This same trend was observed for these metabolites in cells grown on NO_3 but after 12 hours. Glutamine was 100% labeled by 6 h in the urea and NH_3 treatments, while glutamate plateaued at 80% ^{15}N (Figure 4 e, f). For cells grown on NO_3, different incorporation trends were observed. It took 48 h for glutamine and glutamate to reach maximum incorporation, which was 100% for glutamine and 80% for glutamate (same as urea and NH_3). This occurred over the second dark cycle and corresponded to an increase in the relative abundances of ^{15}N and total glutamate (Figure 5a). Over the course of the prior light cycle, there was a significant decrease in the percentage of labeled glutamate in both metabolites (Figure 4 e, f). The drop in the percentage of labeled glutamate and glutamine pools in cells grown on NO_3 corresponded to a drawn down in the relative abundance of total glutamate. However, the concentration of glutamine remained the same (Figure 5a). Trends in the relative abundance of total glutamine were similar for all N treatments, but glutamate concentrations were highly variable and reflected in the observed glutamate:glutamine ratios (Figure 5). For cells grown on NO_3 and urea, glutamate:glutamine ratios were generally higher at the end of the dark cycle and decreased throughout the day, while ratios in NH_3 remained low (~1) and consistent ratio for the duration of the experiment.

One readily detected pathway was for arginine biosynthesis (Figure 6). ^{15}N from urea was incorporated into several key metabolites in this pathway and the intersecting pathway, commonly referred as the urea cycle. ^{15}N from NO_3 was incorporated into some metabolites necessary for arginine biosynthesis, including citrulline, aspartate, and N-acetylglutamate (Figure 7). Arginine, although detectable, demonstrated no incorporation of ^{15}N from NO_3 (Figure 6, 7). ^{15}N

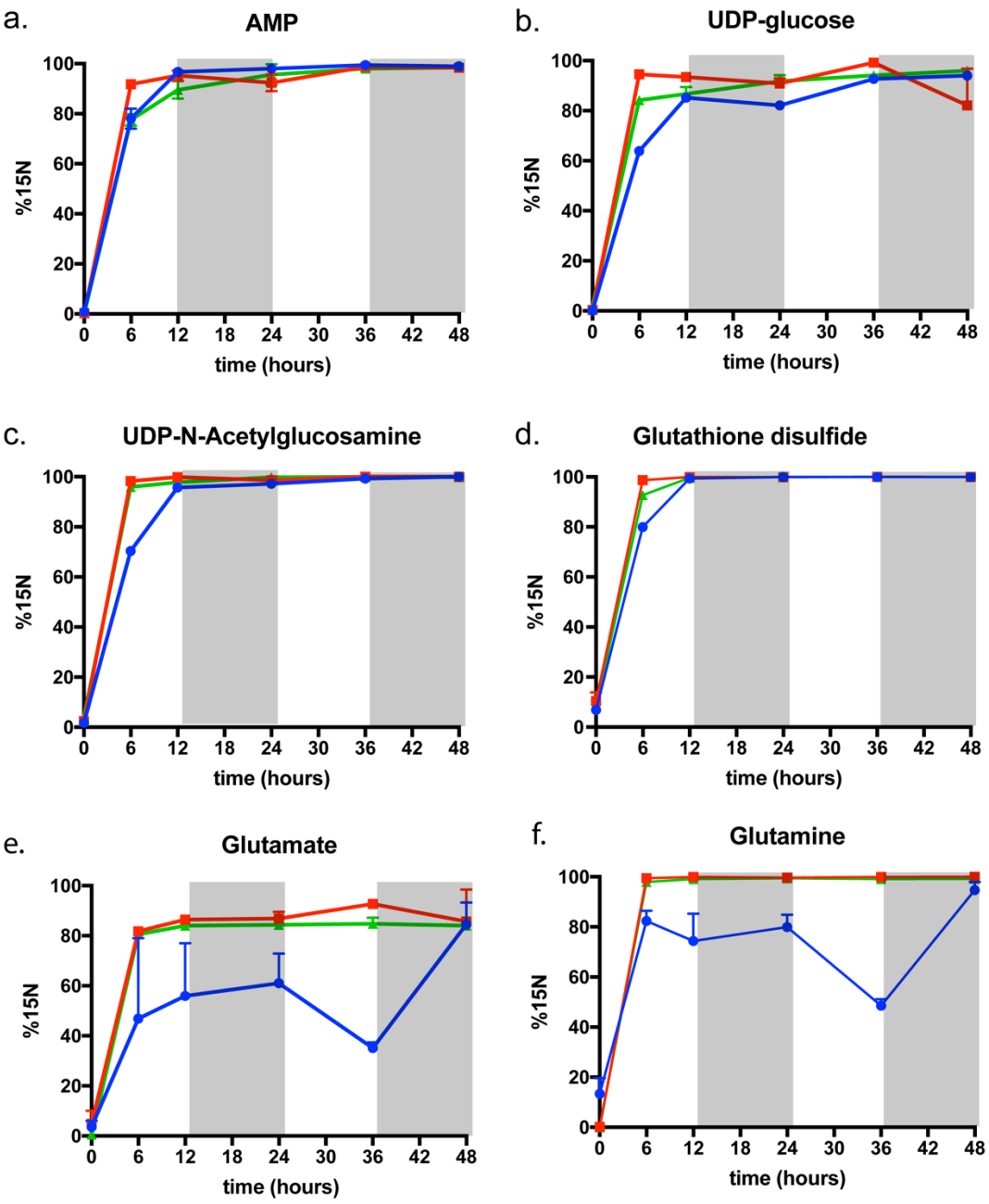

Figure 5.4 Incorporation of ^{15}N into core metabolites

Metabolites detected across all N-forms (NO_3 = blue; NH_3 = red, urea = green) by the end of first light cycle. Shading represents the dark cycle.

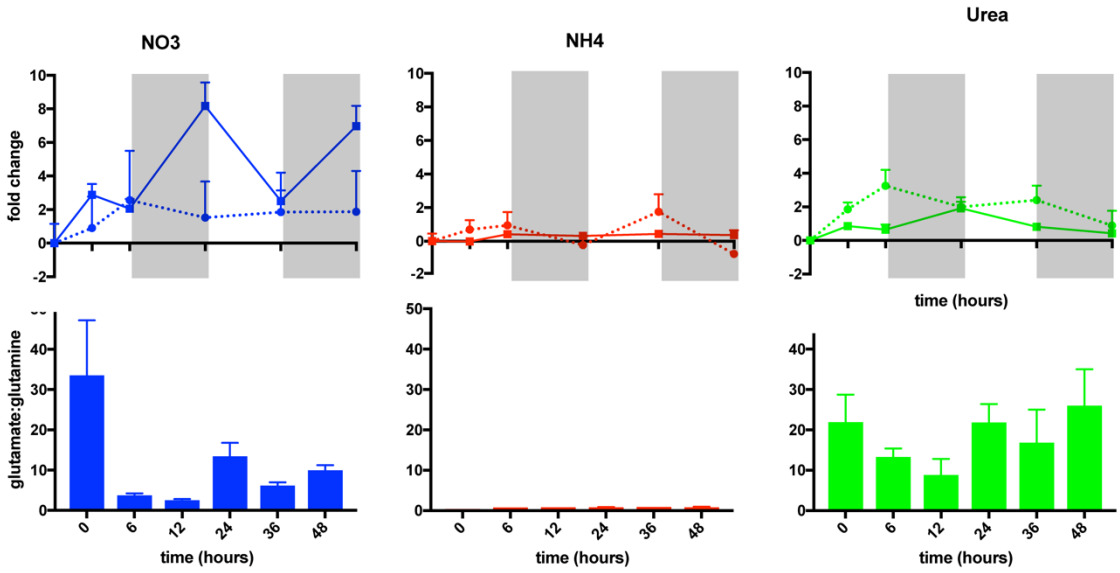

Figure 5.5 Glutamate and glutamine dynamics

Glutamate and glutamine dynamics in cells grown on N-forms (NO_3 = blue; NH_3 = red, urea = green). Top row is fold change over time compared to time 0 of glutamate (solid line) and glutamine (dotted line). Shading represents the dark cycle. Higher ratios are indicative of N-limitation and lower ratios are suggest cells are N-replete.

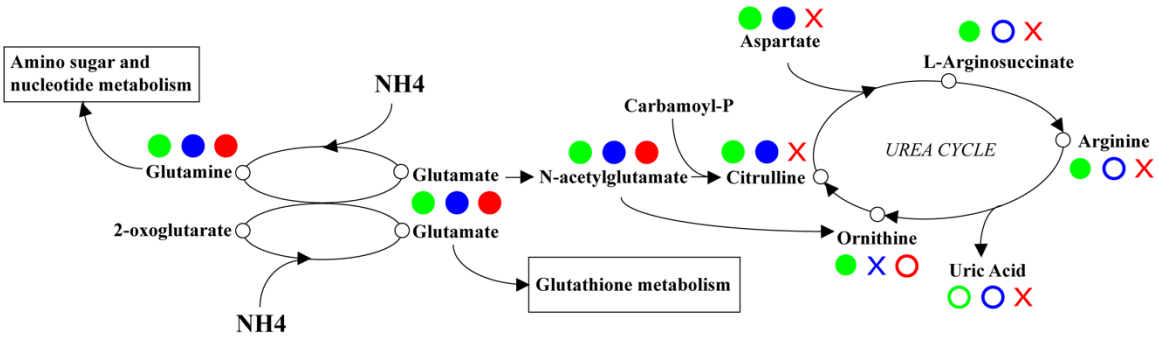

Figure 5.6 Fate of N in metabolic pathways of *M. aeruginosa*

Fate of N in *M. aeruginosa* through the GS-GOGAT cycle and arginine biosynthesis pathway. Metabolites detected are shown as circles. Closed circles indicate the metabolite had ^{15}N incorporated, open circles are indicative of detection and no ^{15}N signature, and an X represents the lack of detection of the metabolite. NO_3 = blue, NH_3 = red and urea =green.

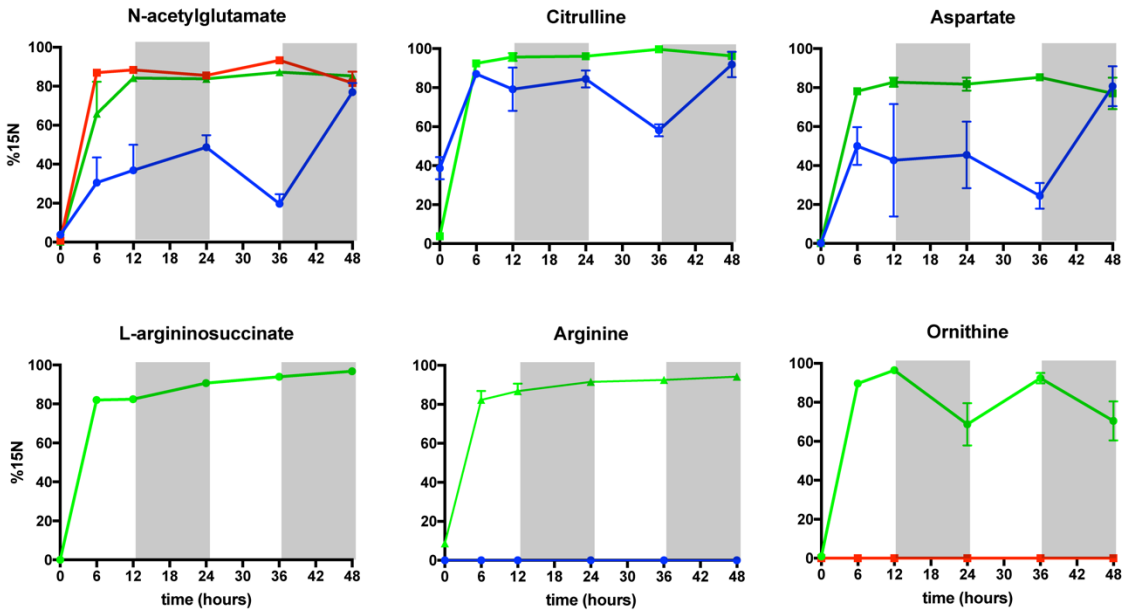

Figure 5.7 Incorporation into arginine biosynthesis pathways

Incorporation of ^{15}N into arginine biosynthesis and urea cycle metabolites. Absence of a treatment indicates this metabolite was not detected and corresponds to Figure 6. Shading represents the dark cycle. NO_3 = blue, NH_3 = red and urea =green.

incorporated into arginine biosynthesis metabolites in a trend similar to glutamate and glutamine from NO_3 treatment (Figure 4, 7). ^{15}N from NH_3 was not detected in any of the metabolites in these pathways except for N-acetylglutamate, however unlabeled ornithine was detectable (Figure 6, 7). Decrease in abundances of these metabolites after the addition of ^{15}N was observed in cells grown on NO_3 except for citrulline, which increased 30-fold during the light cycle and decreased during the night cycle. Fluctuations in these metabolites were observed in cells grown on urea over the time course of this study (Figure 8).

Effect of N form on microcystin and 15N incorporation

Cells grown on urea contained 100 to 1000-times more microcystin-LR (MC-LR) per cell than cells grown on NO_3 and NH_3, respectively (Figure 9a). MC-LR per cell reached concentrations of 0.5 fg / cell. MC-LR decreased significantly over time in both urea and NO_3 treatments (Figure 9b, c). In NH_3, MCLR quotas were comparatively very low and did not change significantly over the course of the experiment (Figure 9d). MCLR in urea decreased significantly approximately 2-fold after 24 h and MCLR in cells grown on NO_3 decreased, less dramatically, after 12 h. ^{15}N from urea was incorporated into all amino acid precursors necessary for MCLR production as well as other congener production (Figure 10), though only MCLR was investigated in this study. Only alanine and glutamate had ^{15}N signatures in NH_3 (data not shown), and aspartate and glutamate had ^{15}N signatures in NO_3 (Figure 4, 7). Although a very low percentage, ^{15}N was incorporated into MCLR in cells grown on urea and a general increase in $\%^{15}N$ was observed overtime (Figure 10b). Incorporation of ^{15}N into MCLR in cells grown on NO_3 and NH_3 was not evident based the lack of observable mass changes in MCLR in this experiment.

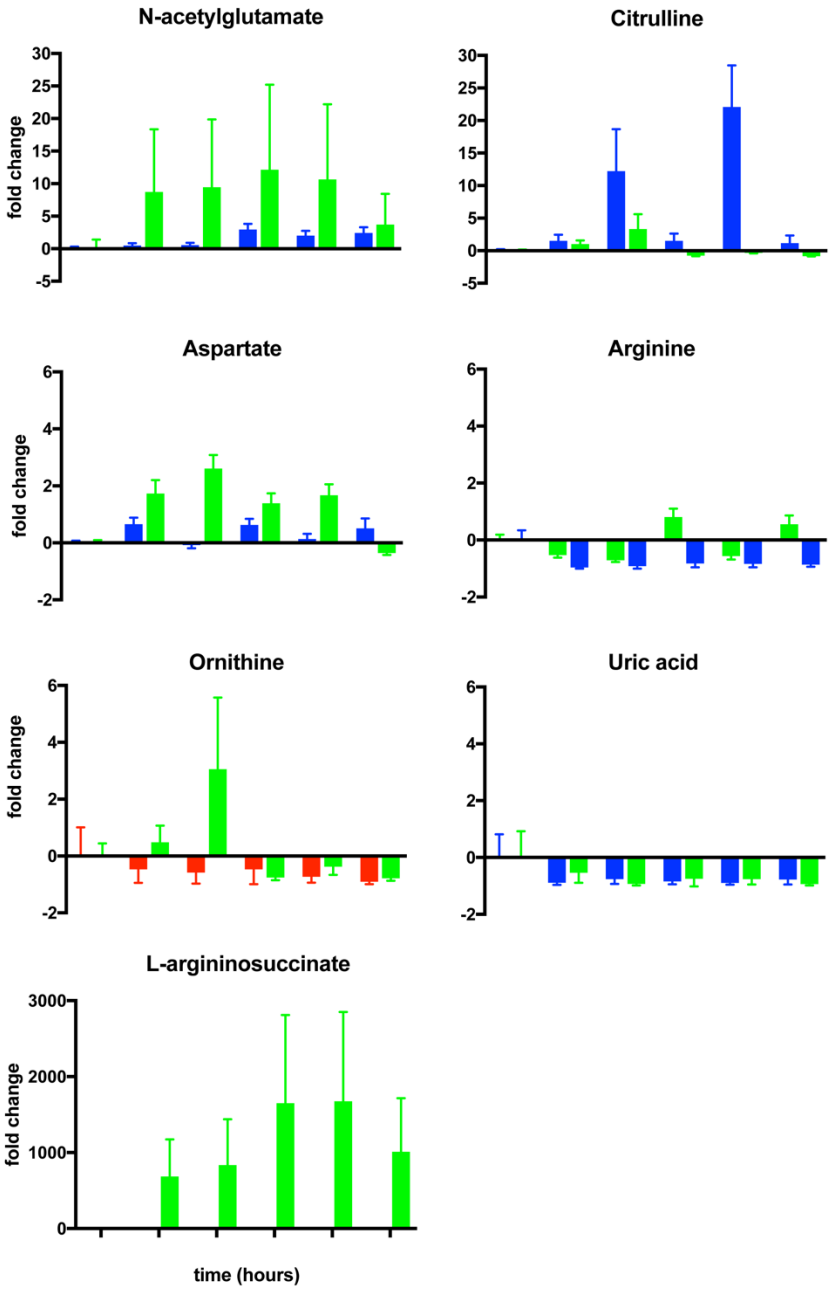

Figure 5.8 Fold changes in arginine biosynthesis metabolites over time

Fold changes over time compared to time 0 of abundances (^{15}N + ^{14}N) of metabolites within arginine biosynthesis pathway and urea cycle. NO_3 = blue, NH_3 = red and urea =green.

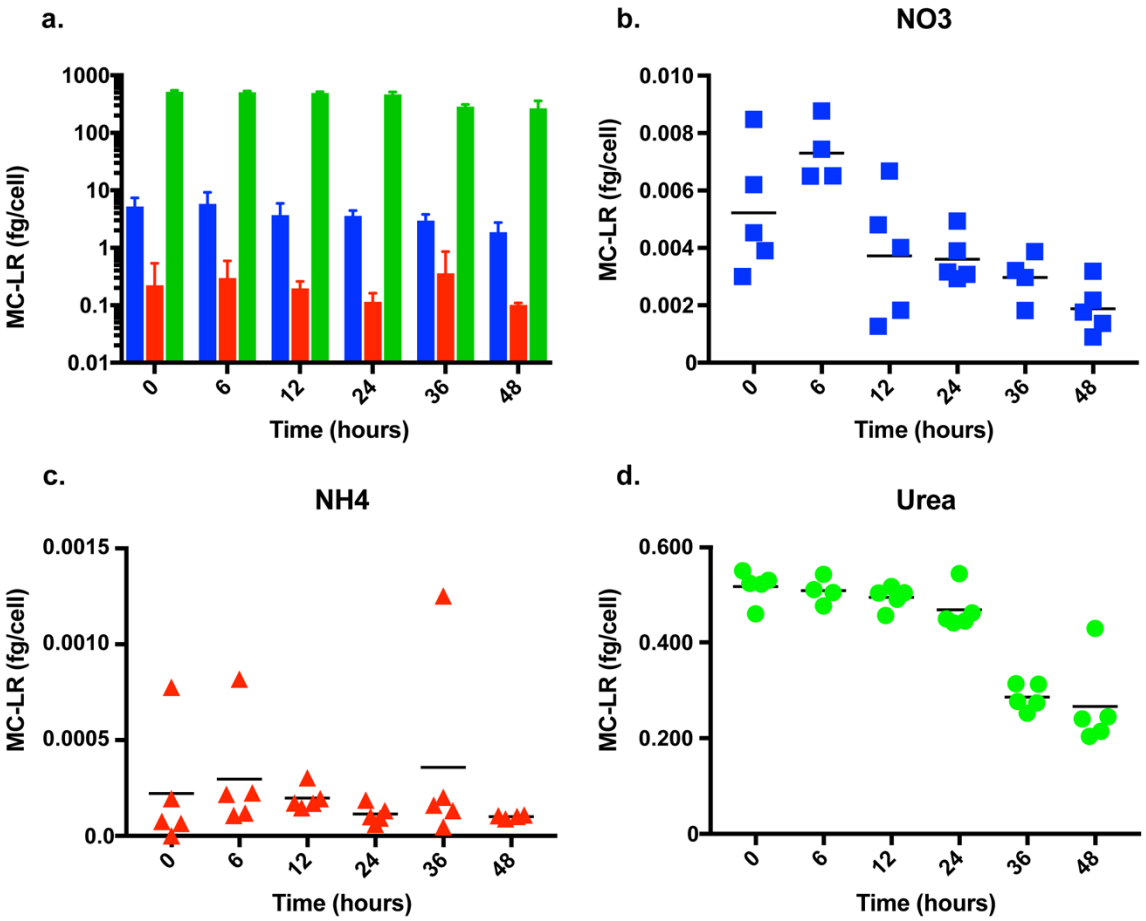

Figure 5.9 Microcystin concentrations over time on different N forms

Intracellular Microcystin-LR concentrations over time. NO_3 = blue, NH_3 = red and urea =green.

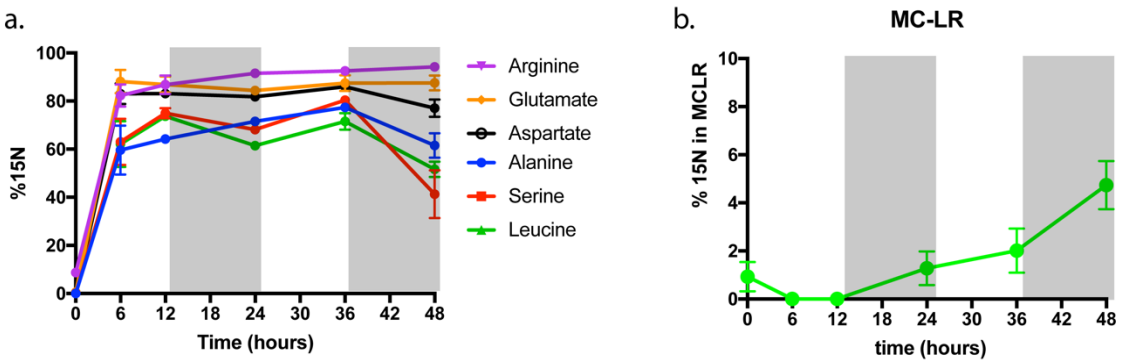

Figure 5.10 ^{15}N incorporation into microcystin precursors and microcystin in cells grown on urea

Metabolites for MC production (a) and MC-LR (b). Shading represents the dark cycle.

Discussion

In this study, ^{15}N-urea, $^{15}NO_3$ and $^{15}NH_3$ were traced through the metabolome of toxic *M. aeruginosa* to examine the effects of N form on the physiology and production of toxin in a common bloom-forming cyanobacteria. ^{15}N was incorporated into metabolites of cells grown on all treatments by six hours, indicating that *M. aeruginosa* is capable of utilizing all forms of exogenous N provided. Despite the apparent lag phase observed in cells grown on urea after being transferred to fresh media, ^{15}N from urea was rapidly incorporated as new ^{15}N incorporated from the byproduct of urea degradation, NH_4. The observed lag phase is common in these axenic cultures when transferred to new media and may be reflective of the toxicity of urea at higher concentrations. There were several metabolites (AMP, UDP-glucose, UDP-N-Acetylglucosamine) that labeled in all treatments by 6 h in NH_3 and urea, and by 12 h in NO_3 consistent with these core metabolites being rapidly turned over. This is not surprising as these metabolites are key intermediates in the production of ATP, nucleotides, and peptidoglycan. Glutathione disulfide (GSSG) was another metabolite completely incorporated with new N in early time points. GSSG is the oxidized form of glutathione (GSH), and the production of these metabolites are essential for maintaining the redox state of the cell (Smirnova & Oktyabrsky, 2005, Cameron & Pakrasi, 2010). GSH combines with reactive oxygen species and form GSSG to neutralize them, which is especially important in photosynthetic organisms that produce oxygen as a byproduct. Rapid incorporation of N and maintenance of this metabolite highlights the importance of this system to maintain redox balance in *M. aeruginosa*, as is the case for other cyanobacteria (Masip *et al.*, 2006, Cameron & Pakrasi, 2010, Narainsamy *et al.*, 2016).

The main entry point for N assimilation from all forms of N is through the GS-GOGAT cycle (Muro-Pastor *et al.*, 2005). N can enter in the form of NH_3, combining with 2-oxogluturate

to generate 2 molecules of glutamate or NH_3 can combine with glutamate to generate glutamine. NO_3 and urea are thought to both need to be converted to NH_3 prior to assimilation requiring an additional step prior to assimilation. NO_3 is converted to NH_3 *via* nitrate reductase, and urea hydrolysis by the enzyme urease produces 2 NH_3 and CO_2 (Glibert *et al.*, 2016).

Glutamate:glutamine ratios are useful as an indicator of the N-availability within the cell, with increased ratios acting as an indicator of N-limitation (Flynn *et al.*, 1989, Ankrah *et al.*, 2014). Glutamate:glutamine ratios in cells grown on urea and NO_3 were similar in that they decreased during the light cycle and increased during the dark cycle. This corresponded to trends in relative abundances of glutamate in cells grown on NO_3, where glutamate was drawn down in the light and accumulated in the dark. For cells grown on urea, the trend in this ratio overtime was instead driven by glutamine abundance, which were drawn down during the dark cycle. Cells grown on NO_3 and urea also had consistently higher (20-30x) glutamate:glutamine ratios than those grown on NH_3, indicating that NH_3 satisfies and maintains the N demand of the cells during exponential growth. Interestingly, despite the differences in glutamate and glutamine dynamics, ^{15}N from urea and NH_3 saturated these metabolites at similar rates, whereas ^{15}N from NO_3 saturated the glutamate and glutamine pools differently. This may be reflective of the differences in initial abundances of glutamate and glutamine, or indicative of differences in turnover rates. The fluctuations in ^{14}N-glutamate and ^{14}N-glutamine in NO_3 grown cells suggests these cells may be recycling previously stored ^{14}N rather than bringing new ^{15}N into the cell, whereas cells grown on urea and NH_3 maintained ^{15}N label in these metabolites throughout the experiment, indicating the utilization of exogenous N over intracellular N stores. These observations may reflect differences in the regulation of uptake mechanisms and utilization of different N species. Together, these observations suggest that physiologically, cells responded differently to N chemistry.

Growth on urea led the production of several metabolites within the arginine biosynthesis pathway and the urea cycle, and they were all labeled within 6 h. In animals, this pathway is a route for removal of excess N in the form of uric acid (Cunin *et al.*, 1986), but the absence of labeled uric acid indicates this is not the case for *M. aeruginosa*. Instead, this pathway was likely being used to produce arginine. In addition to synthesizing arginine for proteins, cyanobacteria can synthesize arginine and aspartate to produce cyanophycin, a N-rich storage molecule, when N is replete (Shively, 2006, Llácer *et al.*, 2008). ^{15}N incorporation in aspartate and arginine suggests that the cells growing on urea may have been shuttling new N into N reserves. Recently a connection between N assimilation and arginine biosynthesis was observed in a *Synechocystis,* another non-N_2 fixing freshwater cyanobacterium (Zhang *et al.*, 2018). This pathway, which they called the ornithine arginine cycle (OAC), was determined to be an important for cells to adapt to fluctuations in N availability. When cells were starved for nitrate and transferred to N-replete media, N was rapidly (within minutes) incorporated into metabolites in the OAC, such as L-argininosuccinate, citrulline, arginine and ornithine. It was determined that *Synechocystis* and many other cyanobacteria, including the one used in this study, have an arginine dihydrolase, *ArgZ*, that degrades arginine to ornithine and NH_3, offering an additional benefit for the production arginine in cells.

In this study, coupling between the GS-GOGAT cycle and the OAC is evident in *M. aeruginosa* as well for cells grown on NO_3 in particular. ^{15}N incorporation trends between glutamate, glutamine, citrulline and aspartate were similar in cells grown on NO_3. However, there was no evidence of incorporation into other metabolites in the OAC, such as arginine or the intermediate formed by aspartate and citrulline, L-argininosuccinate (Figure 6, 7). Unlabeled arginine was detectable (Figure 6, 8), suggesting this may have served as N storage molecule and

partly explain the increase in ^{14}N glutamate. New arginine was probably not synthesized in these conditions and this is supported by the unchanging arginine abundances in cells grown on NO_3. Notably, ^{15}N from NH_3 was not incorporated into any metabolites in the OAC and only unlabeled ornithine was detectable. Conversely, ^{15}N was incorporated in all OAC metabolites in cells grown on urea. As previously mentioned, this pathway has been shown to be activated as a response to shifts from N-limitation to N-replete conditions, and this could be indicative of differences in the N status of the cells growing the different forms of N prior to ^{15}N additions (mid-log). This may explain apparent lag phase in urea treatments that was not observed in the fluorescence data. Instead, fluorescence increased over the course of the light period as expected, suggesting the rapid incorporation of N into metabolites of the OAC may be involved in maintaining N-rich accessory pigments and photosynthetic efficiency. Cells grown on urea may have been more N-limited than those grown on NO_3 and NH_3, although glutamate:glutamine ratios do not reflect these differences compared to NO_3. Glutamate:glutamine ratios for NO_3 and urea were similarly indicative of N-limitation at time 0. Of course, it is possible that incorporation into OAC metabolites was missed in cells grown on NO_3 or NH_3 because of rapid turnover and not captured at these time points, especially considering the 30-fold decrease in citrulline over the dark cycle in cells grown on NO_3. For example, ^{15}N incorporation into the OAC metabolites in *Synechocystis* was observed within 120 min. However, data presented here suggests that newly assimilated N into this pathway from NO_3 was halted at aspartate and citrulline. The conversion of aspartate and citrulline to L-arginosuccinate requires 1 ATP, which indicates that cells growing on nitrate are not expending energy towards making storage molecules like arginine or arginine-rich cyanophycin.

Another difference between growth using NO_3 compared to growth using urea is the intermediate in urea degradation, carbamate. It is generally assumed that carbamate spontaneously

dissociates to $2NH_3$ and CO_2, and the resulting NH_3 enters the GS-GOGAT cycle (Krajewska, 2009, Glibert *et al.*, 2016). However, *M. aeruginosa* contains the gene carbamoyl synthase which can phosphorylate carbamate to form the metabolite carbamoyl phosphate, and this metabolite can feed directly into the arginine biosynthesis pathway. While there was no evidence of ^{15}N-labeled carbamate or carbamoyl phosphate detected, this could possibly explain the number of labeled OAC metabolites in cells grown on urea and not the other N forms. Further, while it's not clear from data whether there is a role for carbon availability in regulating N incorporation into the arginine biosynthesis pathway, a pathway where N and C metabolism are tightly linked (Tandeau de Marsac *et al.*, 2001, Flores & Herrero, 2005). Indeed, the C from urea breakdown can be incorporated into metabolites and shuttled through the OAC (see previous chapter). The contribution of the "free" C from urea may alleviate some of the C demand necessary for production of arginine and intermediates.

Arginine is also an important precursor for the synthesis of the most commonly detected MC congeners in nature, such as MC-LR and MC-RR (Tillett *et al.*, 2000, Meriluoto *et al.*, 2017). In addition to arginine, ^{15}N incorporation into all of the other amino acid precursors necessary for MC-LR production (glutamate, alanine, aspartate, serine and leucine) was detected in cells grown on urea, whereas ^{15}N from NO_3 and NH_4 was only detected in a subset of these amino acids. This observation was consistent with the large differences in MC-LR concentrations that we measured. *M. aeruginosa* produced 100-times more MC-LR when grown on urea compared to cells grown on NO_3, and three orders of magnitude more than cells grown on NH_3. Curiously ^{15}N was only incorporated at low percentage into MC-LR in cells grown on urea, whereas there was no evidence for ^{15}N incorporation into MC-LR in cells grown on NO_3 or NH_3. The underlying function for of MCs in cyanobacteria is still not clear and has been attributed to many factors not related to N *per*

se (Harke *et al.*, 2016), though being a N-rich molecule, microcystin biosynthesis requires N. In the literature, studies have suggested that high N availability equals high MCs, and there is good evidence for this (Gobler *et al.*, 2016). However, there have also been several studies suggesting an inverse relationship. For example, recently it was determined that N form or concentration had very little effect on MC quota in comparison to the very large changes driven by low temperatures (Peng *et al.*, 2018). Here, only N chemistry was considered under the assumption that N supply, to some extent, has to play some role in the synthesis of MCs. Our data suggest that N in the form of urea is more important in MC production because of the substantially higher MC quota as well as MC-LR and precursor labeling. Despite the apparent lack of incorporation of ^{15}N from NO_3 into MC-LR and precursors, these cells were still quite toxic, albeit less than those on urea. This suggests that N form alone is not the driving factor in MC-LR production and that *M. aeruginosa* can produce MC on all three N chemistries. The production of MCs in response to oxidative stress has been hypothesized several times (Zilliges *et al.*, 2011). The possibility that urea hydrolysis causes higher levels of reactive oxygen species cannot be ruled out. Higher concentrations of urea than used in this study cause severe defects in growth rate (data not shown) and even with the concentrations used in this study, a short lag phase was observed in the first 24 h when MCs were simultaneously the highest. As mentioned previously fluorescence continued to increase during this time. The lag phase was followed by rapid growth and decreased MC quotas, suggesting MCs might be involved in maintaining photosynthetic efficiency either as an antioxidant, N source or by another unknown mechanisms.

It is also possible that MC production may have little to do with the available form in the surrounding environment but the availability of intracellular NH_3 specifically. *NtcA,* an important transcriptional regulator of N transcribed at higher levels when N is low (Flores & Herrero, 2005),

can weakly bind to the bidirectional promoter between *mcyA* and *mcyD* genes and repress synthesis of MC (Ginn *et al.*, 2010). MC has therefore been proposed to be inversely related to intracellular N availability rather than high exogenous N concentrations. Here, we present the cells with the same amount of exogenous N and when considering the glutamate:glutamine ratios, the relationship to MC concentrations in this study support the high MC/low N observation. The significant decrease in MC-LR over time in NO_3 and urea, could also suggest a potential function of MCs as a N storage molecule. This hypothesis has been proposed previously (Gobler *et al.*, 2016) but has received little attention, likely due to the expense of making MCs (Tillett *et al.*, 2000). Biological degradation of extracellular MCs by other bacteria has been extensively studied (Li *et al.*, 2017), however the intracellular degradation of MCs and mechanisms for changes over time has not yet been elucidated.

Generally, in natural environments high MC concentrations are associated with high exogenous N availability (Gobler *et al.*, 2016). N amendment experiments in the field have demonstrated this as well, however the effects of N in either the form of NH_3 or NO_3 have been conflicting. Some laboratory studies have supported the positive correlation with N and MC by also showing that MC concentrations are associated with N-dependent growth rates. N can also alter the congener composition, with high N availability supporting congeners with high C: N ratios, such as MC-RR. MC are rich in N and so it is logical that MC production is in some way tied to N availability and culture work has confirmed this by repression of the genes in *mcy* cassette for MC production in N-deplete conditions (Harke and Gobler, 2013; Gobler *et al*. 2016). As mentioned above, other studies have observed that MC production is repressed in the presence of high N-availability or depressed under N-limitation, and this is likely due to regulation by *NtcA* (Ginn *et al.*, 2010, Pimentel & Giani, 2014) and our work supports this.

Recently, experiments using the same strain of *M. aeruginosa,* medium and molar concentrations of N as this study supported the mechanism of high MC/low N-availability. However, they also examined the effect of concentration (exogenous N availability) and found that higher concentrations of NO_3 and urea reduced MC quota per cell. Further, while there were slightly higher concentrations of MCs in cells grown on $NO_{3,\ but}$ the difference from urea was not supported statistically. It was concluded that N form had no effect on MC concentration or congener formation and instead, high MC concentrations and a shift in congener composition was attributed to low temperatures (~18 °C). A key difference that may explain conflicting observations on the effect of N form between these two studies is that *Microcystis* was harvested in late log phase where they were presumably N limited. In the current study, *Microcystis* was harvested during exponential growth and a fluctuation of MC quotas was observed. These fluctuations present an important consideration when extrapolating the role of nutrients on MC concentrations when sampling at just one time point. Another intriguing difference between the two studies is the *Microcystis* in Peng *et al*. (2018) was non-axenic. The associated heterotrophic bacteria may be influencing exogenous N availability, subsequent intracellular N availability and MC quotas and presents an additional variable for consideration. While using non-axenic cultures are perhaps a more environmentally relevant scenario, a key stepping stone in teasing apart the factors that influence MC production is to characterize them on *Microcystis* alone to reduce these confounding but very intriguing factors.

Conclusions

In this study, the fate of N in different forms was traced through the metabolome of *M. aeruginosa* NIES843 using stable labeled isotopes (^{15}N) of nitrate, ammonium and urea. Core

metabolites from essential pathways, such as peptidoglycan synthesis, nucleotide metabolism and glutathione metabolism, demonstrated rapid incorporation of ^{15}N regardless of N form, indicating all forms of N were assimilated and that these pathways were a preferred route for N once in the cell. Cells grown on ammonium and nitrate were not expending energy to synthesize arginine, a potential storage product, and key precursor for cyanophycin and microcystin production, whereas this pathway was active in cells grown on urea based on detection of ^{15}N. This pathway was likely being used by cells grown on urea used as a storage and N remobilization pathway. This has previously been observed in other cyanobacteria as a potential response to sudden N availability. The urea grown cells in our study were also substantially more toxic than those grown on nitrate and ammonium, and ^{15}N incorporation into microcystin only occurred in urea-grown cells. The direct link between the arginine biosynthesis pathways and microcystin production is not completely understood, however arginine is an important amino acid for several commonly detected microcystin congeners in nature (*i.e.* MC-RR, MC-LR). These data presented here support that microcystins are produced differentially in response to available N-form and urea may promote more microcystin production. However, observed trends in toxicity of the cells may be related to indirect factors, such as photosystem health or intracellular rather than extracellular N-availability. These data suggest that N chemistry is important when considering the relationship between N and microcystin and generalizations about N made using only NO_3 is not sufficient. Overall, the call for management practices that reduce anthropogenic N inputs, especially those high in urea, into freshwaters is supported here.

Acknowledgements

This work was made possible due to grant NA11NOS4780021 from NOS National Center for Coastal Ocean Science (NCCOS) under their Harmful Algal Bloom Prevention Control and Mitigation program to SWW and GLB, and National Science Foundation awards (DEB-1240870; IOS-1451528) to SWW. The authors also acknowledge the generous support of the *Kenneth & Blaire Mossman* endowment to the University of Tennessee.

References

Ankrah N. Y. D., May A. L., Middleton J. L., Jones D. R., Hadden M. K., Gooding J. R., LeCleir G. R., Wilhelm S. W., Campagna S. R. & Buchan A. 2014. Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *The ISME journal*. 8, 1089.

Belisle B. S., Steffen M. M., Pound H. L., Watson S. B., DeBruyn J. M., Bourbonniere R. A., Boyer G. L. & Wilhelm S. W. 2016. Urea in Lake Erie: Organic nutrient sources as potentially important drivers of phytoplankton biomass. *Journal of Great Lakes Research*. 42, 599-607.

Bullerjahn G. S., McKay R. M., Davis T. W., Baker D. B., Boyer G. L., D'Anglada L. V., Doucette G. J., Ho J. C., Irwin E. G. & Kling C. L. 2016. Global solutions to regional problems: Collecting global expertise to address the problem of harmful cyanobacterial blooms. A Lake Erie case study. *Harmful Algae*. 54, 223-238.

Cameron J. C. & Pakrasi H. B. 2010. Essential role of glutathione in acclimation to environmental and redox perturbations in the cyanobacterium *Synechocystis* sp. PCC 6803. *Plant physiology*. 154, 1672-1685.

Carmichael W. W. & Boyer G. L. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae*. 54, 194-212.

Clarke K. & Gorley R. 2006. PRIMER v6: user manual/tutorial (Plymouth routines in multivariate ecological research). *Plymouth: Primer-E Ltd*.

Conley D. J., Paerl H. W., Howarth R. W., Boesch D. F., Seitzinger S. P., Karl E., Karl E., Lancelot C., Gene E. & Gene E. 2009. Controlling eutrophication: nitrogen and phosphorus. *Science*. 123, 1014-1015.

Cunin R., Glansdorff N., Pierard A. & Stalon V. 1986. Biosynthesis and metabolism of arginine in bacteria. *Microbiological reviews*. 50, 314.

Davis T. W., Harke M. J., Marcoval M. A., Goleski J., Orano-Dawson C., Berry D. L. & Gobler C. J. 2010. Effects of nitrogenous compounds and phosphorus on the growth of toxic and non-toxic strains of *Microcystis* during cyanobacterial blooms. *Aquatic Microbial Ecology*. 61, 149-162.

Flores E. & Herrero A. (2005) Nitrogen assimilation and nitrogen control in cyanobacteria. Portland Press Limited.

Flynn K. J., Dickson D. M. & Al-Amoudi O. A. 1989. The ratio of glutamine: glutamate in microalgae: a biomarker for N-status suitable for use at natural cell densities. *Journal of plankton research*. 11, 165-170.

Ginn H., Pearson L. & Neilan B. 2010. NtcA from *Microcystis aeruginosa* PCC 7806 is autoregulatory and binds to the microcystin promoter. *Applied and environmental microbiology*. 76, 4362-4368.

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

Glibert P. M., Wilkerson F. P., Dugdale R. C., Raven J. A., Dupont C. L., Leavitt P. R., Parker A. E., Burkholder J. M. & Kana T. M. 2016. Pluses and minuses of ammonium and nitrate uptake and assimilation by

phytoplankton and implications for productivity and community composition, with emphasis on nitrogen-enriched conditions. *Limnology and Oceanography*. 61, 165-197.

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

Harke M. J. & Gobler C. J. 2015. Daily transcriptome changes reveal the role of nitrogen in controlling microcystin synthesis and nutrient transport in the toxic cyanobacterium, *Microcystis aeruginosa. BMC genomics*. 16, 1068.

Harke M. J., Davis T. W., Watson S. B. & Gobler C. J. 2015. Nutrient-controlled niche differentiation of western Lake Erie cyanobacterial populations revealed via metatranscriptomic surveys. *Environmental science & technology*. 50, 604-615.

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

Kanehisa M. & Goto S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 28, 27-30.

Kappers F. I. 1980. The cyanobacterium *Microcystis aeruginosa* Kg. and the nitrogen cycle of the hypertrophic Lake Brielle (The Netherlands). *Hypertrophic ecosystems,* Vol. p.^pp. 37-43. Springer.

Krajewska B. 2009. Ureases I. Functional, catalytic and kinetic properties: A review. *Journal of Molecular Catalysis B: Enzymatic*. 59, 9-21.

Li J., Li R. & Li J. 2017. Current research scenario for microcystins biodegradation–A review on fundamental knowledge, application prospects and challenges. *Science of The Total Environment*. 595, 615-632.

Llácer J. L., Fita I. & Rubio V. 2008. Arginine and nitrogen storage. *Current opinion in structural biology*. 18, 673-681.

Masip L., Veeravalli K. & Georgiou G. 2006. The many faces of glutathione in bacteria. *Antioxidants & redox signaling*. 8, 753-762.

Meriluoto J., Spoof L. & Codd G. A. 2017. *Handbook of cyanobacterial monitoring and cyanotoxin analysis*. John Wiley & Sons,

Muro-Pastor M. I., Reyes J. C. & Florencio F. J. 2005. Ammonium assimilation in cyanobacteria. *Photosynthesis research*. 83, 135-150.

Narainsamy K., Farci S., Braun E., Junot C., Cassier-Chauvat C. & Chauvat F. 2016. Oxidative-stress detoxification and signalling in cyanobacteria: the crucial glutathione synthesis pathway supports the production of ergothioneine and ophthalmate. *Molecular microbiology*. 100, 15-24.

Paerl H. W., Gardner W. S., McCarthy M. J., Peierls B. L. & Wilhelm S. W. 2014. Algal blooms: Noteworthy nitrogen. *Science*. 346, 175.

Paerl H. W., Xu H., McCarthy M. J., Zhu G., Qin B., Li Y. & Gardner W. S. 2011. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): the need for a dual nutrient (N & P) management strategy. *Water Research*. 45, 1973-1983.

Paerl H. W., Scott J. T., McCarthy M. J., Newell S. E., Gardner W., Havens K. E., Hoffman D. K., Wilhelm S. W. & Wurtsbaugh W. A. 2016. It takes two to tango: When and where dual nutrient (N & P) reductions are needed to protect lakes and downstream ecosystems. *Environmental Science & Technology*.

Peng G., Martin R. M., Dearth S. P., Sun X., Boyer G. L., Campagna S. R., Lin S. & Wilhelm S. W. 2018. Seasonally Relevant Cool Temperatures Interact with N Chemistry to Increase Microcystins Produced in Lab Cultures of *Microcystis aeruginosa* NIES-843. *Environmental science & technology*. 52, 4127-4136.

Peng G., Martin R., Dearth S., Sun X., Boyer G. L., Campagna S., Lin S. & Wilhelm S. W. 2018. Seasonally-relevant cool temperatures interact with N chemistry to increase microcystins produced in lab cultures of *Microcystis aeruginosa* NIES-843. *Environmental science & technology*.

Pimentel J. S. & Giani A. 2014. Microcystin production and regulation under nutrient stress conditions in toxic *Microcystis* strains. *Applied and environmental microbiology*. 80, 5836-5843.

Schindler D. W., Hecky R., Findlay D., Stainton M., Parker B., Paterson M., Beaty K., Lyng M. & Kasian S. 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*. 105, 11254-11258.

Scott J. T. & McCarthy M. J. 2010. Nitrogen fixation may not balance the nitrogen pool in lakes over timescales relevant to eutrophication management. *Limnology and Oceanography*. 55, 1265-1270.

Shively J. M. 2006. *Inclusions in prokaryotes*. Springer Science & Business Media,

Smirnova G. & Oktyabrsky O. 2005. Glutathione in bacteria. *Biochemistry (Moscow)*. 70, 1199-1211.

Steffen M. M., Belisle B. S., Watson S. B., Boyer G. L., Bourbonniere R. A. & Wilhelm S. W. 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Applied and environmental microbiology*. 81, 3268-3276.

Steffen M. M., Dearth S. P., Dill B. D., Li Z., Larsen K. M., Campagna S. R. & Wilhelm S. W. 2014. Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. *The ISME journal*. 8, 2080-2092.

Tandeau de Marsac N. T., Lee H., Hisbergues M., Castets A. & Bédu S. 2001. Control of nitrogen and carbon metabolism in cyanobacteria. *Journal of Applied Phycology*. 13, 287-292.

Tillett D., Dittmann E., Erhard M., von Döhren H., Börner T. & Neilan B. A. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. *Chemistry & biology*. 7, 753-764.

Willis A., Chuang A. W. & Burford M. A. 2016. Nitrogen fixation by the reluctant diazotroph *Cylindrospermopsis raciborskii* (Cyanophyceae). *Journal of Phycology*.

Zhang H., Liu Y., Nie X., Liu L., Hua Q., Zhao G.-P. & Yang C. 2018. The cyanobacterial ornithine–ammonia cycle involves an arginine dihydrolase. *Nature chemical biology*. 1.

Zilliges Y., Kehr J.-C., Meissner S., Ishida K., Mikkat S., Hagemann M., Kaplan A., Börner T. & Dittmann E. 2011. The cyanobacterial hepatotoxin microcystin binds to proteins and increases the fitness of *Microcystis* under oxidative stress conditions. *PloS one*. 6, e17615.

Chapter 6 : Metatranscriptome analysis suggests the *mlr* pathway is not a route for microcystin biodegradation in Lake Erie and Lake Taihu

Publication note

This chapter is a version of a manuscript in review for the Journal of Great Lakes Research by Lauren E. Krausfeldt, Morgan M. Steffen, Robert M. McKay, George S. Bullerjahn, Gregory L. Boyer and Steven W. Wilhelm

My contribution to this work was assisting in RNA extraction, data analysis and writing the majority of the paper.

Abstract

Microcystins are potent hepatotoxins that are frequently detected in freshwater lakes plagued by toxic cyanobacteria. Microbial biodegradation is considered to be the most important avenue for removal of microcystin from aquatic environments. The biochemical pathway most commonly reported to drive the breakdown of microcystin is encoded by the *mlr*ABCD (*mlr*) cassette and is reported to be present within diverse groups of bacteria. However, not all microcystin-degrading bacterial isolates have these genes, and the ecological significance of this pathway remains unclear. Metatranscriptomes generated from toxic *Microcystis* blooms were used to assess the activity of this pathway in environmental samples. A total of sixty-seven samples were collected from Lake Erie, USA/Canada and Lake Tai (*Taihu*), China, and screened for the presence of *mlr* gene transcripts. Analysis of assembled reads resulted in no detection of the standard marker gene for microcystin degradation, *mlr*A. Further, only 157 of the collective 2.8 billion reads mapped to any part of the *mlr* operon, suggesting expression of the other genes in this pathway were also not confidently identifiable. These observations were made despite the presence of 3.2 million transcripts that mapped to the toxin biosynthesis *mcy* operon associated with microcystin production. Transcripts for genes within the CAAX protease and bacteriocin processing family (those most closely related proteins to *mlrA*) were sparsely detected in assembled reads from Lake Erie and not highly expressed. Glutathione S-transferases and alkaline proteases have been previously hypothesized to be alternative pathways for microcystin degradation. High expression of these genes was detected across space and time in both lakes, suggesting that these genes may individually or collectively be more important. This study suggests that the *mlr* pathway is not environmentally relevant during toxic *Microcystis* blooms and supports the ongoing hypothesis that other pathways for microcystin biodegradation exist.

Introduction

Microcystins (MCs) are the one of the most frequently detected cyanotoxins within freshwater harmful cyanobacterial blooms (cyanoHABs). With cyanoHABS on the rise around the world, understanding the fate of MCs in the environment has assumed greater importance as they pose both ecological and public health risks. MCs are potent protein phosphatase 1 and 2A inhibitors as well as potential tumor promoters. Consumption of MCs can result in acute hepatocytosis, cancer and various gastrointestinal problems (Carmichael, 2001, Zurawell *et al.*, 2005, Carmichael & Boyer, 2016). Conventional municipal water treatments are effective at removing or inactivating cyanobacterial cells, but the methods risk the release of MCs from the cell. MCs are not removed by flocculation processes and require secondary treatments that include the use of activated carbon, which is effective but costly (He *et al.*, 2016). Indeed, liver cancer and colorectal cancer has been linked to nearby bodies of water plagued with toxic cyanobacteria in the United States, Serbia and China (Carmichael, 2001, Hernández *et al.*, 2009, Svircev *et al.*, 2009). Cases of mortality and morbidity due to cyanotoxins have been reported since the 1800s involving birds, livestock, dogs, fish and even humans (Chorus & Bartram, 1999, Wood, 2016). The number of cases has risen gradually over time and it is predicted that the number of individuals affected is greatly underestimated (Wood, 2016).

MCs are chemically stable and resistant to many abiotic factors, which has led to numerous studies questioning their fate in the environment (Harada & Tsuji, 1998). Whereas their disappearance has been attributed to some abiotic effects including dilution, adsorption, and photodegradation (Harada & Tsuji, 1998, Corbel *et al.*, 2014), they are generally considered to be resistant to abiotic factors. MCs are also resistant to the activity of common peptidases (Tsuji *et al.*, 1994, Mazur & Plinski, 2001, Smith *et al.*, 2010) due to their cyclic structure and the alternating

incorporation of the non-protein "R" stereoisomers of the amino acid, coupled with "iso" peptide bonds formed through the side chains of glutamate and aspartate. Despite these challenges, microorganisms that span multiple phyla and even domains have been isolated that are reported to degrade MCs (Appendix Table 1, Dziga *et al.*, 2013, Li *et al.*, 2017). Further, experiments assessing the biodegradation of MCs which employ natural microbial communities from lakes, drinking water reservoirs, water treatment facilities, soils, sediments, and biofilters, have demonstrated that degradation capacity is seemingly ubiquitous across a wide range of environments and occurs under a variety of conditions (Appendix Table 2, Li *et al.*, 2014, Lezcano *et al.*, 2016, Wang *et al.*, 2016, Dai *et al.*, 2017). For these reasons, the current literature often refers to biodegradation as the most important route for the disappearance of MCs in nature.

The reported genetic pathway involved in MC-LR degradation involves four enzymes encoded by *mlrA, mlrB, mlrC, mlrD* (Bourne *et al.*, 1996, Bourne *et al.*, 2001). MlrA, termed microcystinase, is a metalloprotease linearizing the MC-LR at the Arg-ADDA bond. It is commonly used as marker gene and considered to catalyze the most important step in biodegradation of MCs (Dziga *et al.*, 2012, Dexter *et al.*, 2018, Lezcano *et al.*, 2018). MlrB is a serine protease that drives the hydrolysis of the linearized MC-LR at the Ala-Leu bond, forming a tetrapeptide. The tetrapeptide is then subject to further degradation by a second metalloprotease, MlrC. It was recently demonstrated that MlrC can also cleave the linearized MC at the -Adda group (Dziga *et al.*, 2016). MlrD is a putative transporter, but whether MC degradation occurs intracellularly or extracellularly remains inconclusive (Bourne *et al.*, 2006).

This pathway has been extensively studied in lab settings, but the environmental relevance and distribution of the *mlr* cassette in the environment remains unclear. The *mlrA* gene has been amplified by PCR from DNA isolated from microbial communities with MC degradation

capabilities from freshwater samples, but is more commonly detected in enriched samples such as biofilms, filtration sand and water samples from treatment plants (Ho *et al.*, 2006, Ho *et al.*, 2007, Hoefel *et al.*, 2009, Ho *et al.*, 2010, Jimbo *et al.*, 2010, Li *et al.*, 2011, Li *et al.*, 2012, Kansole & Lin, 2016, Lezcano *et al.*, 2016, Wang *et al.*, 2016). Still, there are very few representative *mlrA* sequences currently available. Most often, *mlrA* is detected by PCR in environmental samples, which is not quantitative. Other genes in the cassette are not as widely studied in the environment and sequences of proteins with confirmed function are lacking. Since the methods used to screen for *mlrA* in native microbial communities have been based on DNA sequences (Hoefel *et al.*, 2009, Mou *et al.*, 2013, Lezcano *et al.*, 2018), only the potential of the pathway to be a route for MC biodegradation has thus far been considered. The activity of all the genes in the *mlr* pathway, either in the form of transcribed RNA or expressed proteins, has not been measured in natural samples.

Our goal in this study was to determine the significance of this pathway for biodegradation of toxins by means of examining the relative expression of the genes in the *mlr* pathway in the natural environment. Knowledge of MC biodegradation in the natural environment on a molecular level has potential use in predicting the capacity of the native microbial community to remove MCs from that environment. For this work we used relevant environmental samples to investigate the potential for this pathway to be an important mechanism for removal of MCs *via* biodegradation during cyanoHABs. For comparison, we screened for the presence of transcribed genes involved in other hypothesized (and less studied) mechanisms for MC degradation, including glutathione S-transferases (GSTs) and alkaline proteases (Takenaka & Watanabe, 1997, Mou *et al.*, 2013). This was done using shotgun metatranscriptomes from five surveys in eutrophic freshwater lakes with a combined total of 2.8 billion transcript reads. The use of metatranscriptomes relative to PCR-based or shotgun genomic assays allowed for a more accurate

representation of the active function of the microbial community. These metatranscriptomes provide insight on the proportional representation of a gene's activity within a community and provide the ability to detect more divergent sequences of the genes of interest.

Methods

Phylogenetic analysis of mlrA sequences

Protein sequences arising from the *mlrA* gene were downloaded from NCBI's GenBank and accession numbers are displayed in Figure 1. These sequences were aligned after eight iterations using MUSCLE and a maximum likelihood tree was created in Mega v7.0 (Tamura *et al.*, 2007).

Sample collection and sequencing

A total of five surveys, encompassing 67 metatranscriptomes, were used to screen for the expression of genes involved in MC biodegradation. Twenty-six of the metatranscriptomes were generated from surface waters in the western basin of Lake Erie (USA/Canada) with fourteen from a bloom event in early August 2014 (Figure 2, Steffen *et al.*, 2017), seven tracking a bloom over a diurnal cycle in late August 2014 (Davenport, 2016) and five from a bloom event in July 2013 (Steffen *et al.*, 2017). The Lake Erie dataset from early August of 2014 by Steffen et al. (2017) was of special interest because samples were taken at and around the City of Toledo's water intake where MC concentrations in the source water were above 10 μg/L and exceeded 2.5 μg/L in the finished water, resulting in a 3-day water ban affecting nearly 500,000 people (Carmichael & Boyer, 2016, Steffen *et al.*, 2017).

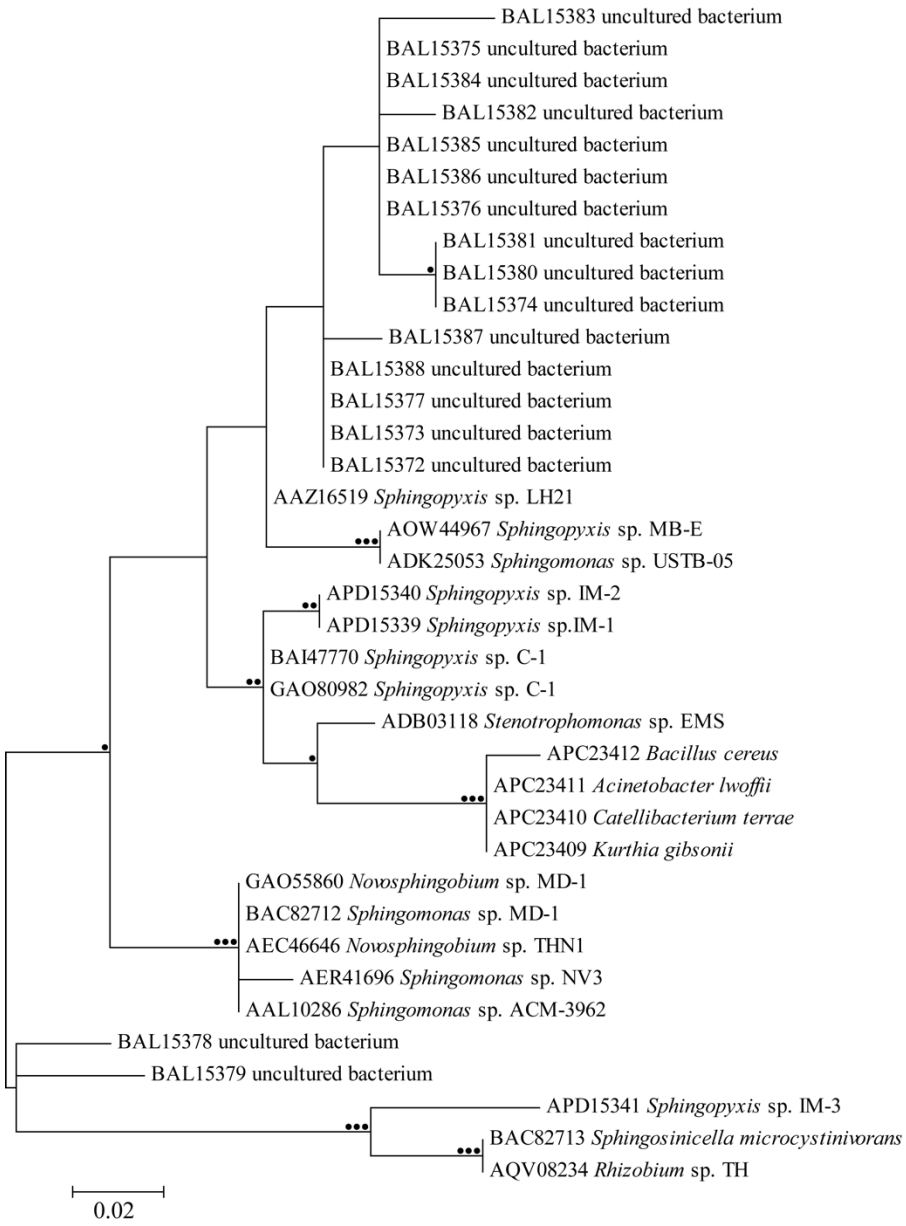

Figure 6.1 Phylogenetic tree of publicly available MlrA sequences

Maximum likelihood tree describing the phylogenetic distribution of MlrA protein sequences. Tree was bootstrapped 1000 times. Bootstrap values are represented by circles at nodes. ●●● >90, ●● >70, ● >50

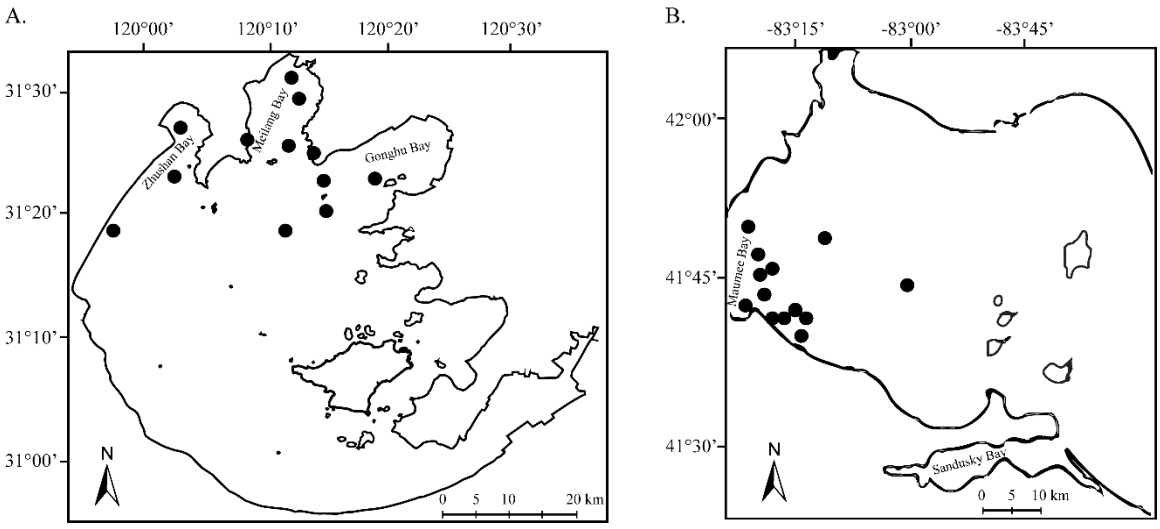

Figure 6.2 Study site and sample locations

Map of *Taihu* (A) and Lake Erie (B). Stations sampled in metatranscriptomes are depicted as closed circles. Coordinates for each sample locations can be found in Appendix Table 3 and 4.

Forty-one metatranscriptomes were generated from surface water collected in the northern part of Lake Tai (*Taihu*), China in 2013 and 2014 during bloom events. The *Taihu* 2014 metatranscriptomes were composed of thirty-five samples collected across nine locations at monthly intervals from June to October, capturing an entire bloom event (Stough *et al.*, 2017). Six metatranscriptomes were generated from samples collected in August 2013. Sample collection, RNA extraction and sequencing for the *Taihu* 2013 metatranscriptomes were performed using the same approaches as previously described for published studies on Lake Erie (Steffen *et al.*, 2015). Sampling information and environmental parameters collected at each location is available in Appendix Table 3 and 2. The *Taihu* 2013 metatranscriptomes are publicly available on the MG-RAST server (ID: mgm4768721 mgm4768720, mgm4768724, mgm4768726, mgm4768728, mgm47687130) and the Lake Erie 2014 diel metatranscriptomes are available from the SRA database (SRP128942, SRP128954, SRP128945, SRP117911, SRP117914, SRA117922, SRA117915). More information on location and dates of sampling for all metatranscriptomes can be found in Appendix Table 3 and 4.

Read recruitment to the mlr cassette

Expression of *mlr* pathway was first evaluated by recruiting reads from Lake Erie and *Taihu* metatranscriptomes to the *mlrA-D* from *Sphingomonas* sp. ACM3962 in CLC Genomics Workbench (Qiagen, Hilden, Germany). Reads were trimmed and filtered for quality using CLC Genomics Workbench default parameters. Recruitments were performed using a similarity fraction of 0.8 and a length fraction of 0.5 to encompass known and potential unknown diversity of these genes. For comparisons, reads were also recruited to the *mcy* operon (*mcyA-J*) from the model

species *Microcystis aeruginosa* NIES 843 with the more stringent similarity fraction and length fraction of 0.9 but remains inclusive of other toxin producing species (Steffen *et al.*, 2017).

Screening for microcystinase, GST, and alkaline proteases in assembled data using BLASTx

Reads were assembled from individual samples into contiguous sequences ("contigs") in CLC Genomics Workbench using the default parameters. All contigs from the metatranscriptomes were screened using BLASTx against the currently available MlrA amino acid sequences in NCBI's GenBank. Contigs with positive hits to *mlrA* and an E-value of <0.00001 were collected and translated to amino acid sequences using the appropriate reading frame identified by BLASTx. For comparison, this method of detection was also applied to a metagenome from a study on Lake Erie where MC biodegradation and *mlrA*-like genes have previously been reported (Mou *et al.*, 2013).

Putative GSTs were identified in the metatranscriptomes by first screening contigs using BLASTx against >50,000 full length amino acid sequences annotated as bacterial GSTs from UNIPROT (Search: "glutathione S-transferase"). Alkaline protease sequences from *Pseudomonas aeruginosa* isolates were used to screen contigs in a similar manner (Accession numbers: AAX84042, AFI809901, CAA45858, AAC38255, AAR20883, WP_079455220, WP_078455733, WP_078465318). Contigs with positive hits to either GST or alkaline protease with an E-value of $<1x10^{-11}$ and at least 100 high similarity positions were collected for further analysis. Reads from each sample with positive contigs were mapped to their respective putative GST or alkaline protease contigs.

Reads were mapped to contigs of interest at a 0.97 DNA sequence identity and length fraction. This stringent parameter was used so as to only recruit very specific reads to each contig.

Mapped reads per sample were totaled and normalized by library size. Shadeplots were constructed in Primer-e (v7) after being fourth root transformed for visualization.

Phylogenetic analysis of mlrA-like contigs

Amino acid reference sequences were downloaded from NCBI and UNIPROT to build a reference tree for phylogenetic analysis of *mlr*A-like contigs to determine if these sequences were likely *mlr*A. Alignments of reference sequences were performed using MUSCLE in Mega v7.0 (Tamura *et al.*, 2007). A maximum likelihood tree was made using the PhyML server using the LG model and likelihood ratios and branch support were calculated using a Shimodaira-Hasegawa (SH)-like approximate likelihood ratio test (aLRT-SH-like, Guindon *et al.*, 2010). Contigs were incorporated into the tree using pplacer (matsen.fhcrc.org/pplacer/) as in Krausfeldt *et al.* (2017).

Results and Discussion

Reads from the metatranscriptomes were first recruited to all four genes in the *mlr* pathway from *Sphingomonas* sp. ACM3962 with loose parameters to account for diversity. Only 157 reads out of the total over 3.5 billion reads surveyed mapped to genes in the *mlr* pathway; 138 from Lake Erie and 19 from *Taihu* (Figure 3a, b). These reads mapped to *mlrA*, *mlrB* and *mlrC* and none mapped to *mlrD*. Notably, mapped reads did not provide full coverage of any of the individual genes in the *mlrABCD* cassette. This is important because it indicated that the reads that did map were potentially non-specific. Together this suggested that none of the genes in the *mlr* pathway

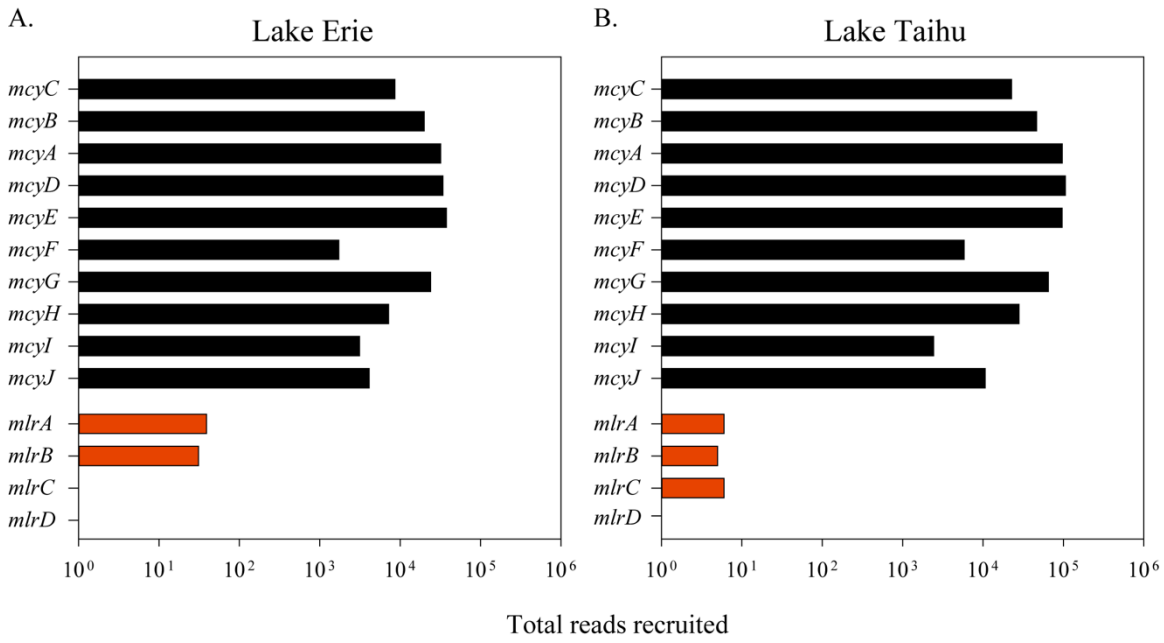

Figure 6.3 Read mappings to the *mlr* and *mcy* cassette

Reads mapped on a log scale to the *mlr* cassette and *mcy* operon from metatranscriptomes from Lake Erie (A) and *Taihu* (B).

were expressed concomitantly or alone in these samples. In comparison, over three million reads mapped to the *mcy* operon, the genes for MC production (Figure 3a, b), indicating the presence of active toxin-producing cyanobacterial species and likely the presence of MCs in the water column. The extremely low level of *mlr* gene expression was surprising because MC biodegradation has been observed by the natural microbial community residing in both Lake Erie and *Taihu* previously, and notably, several bacterial strains reported to degrade MCs (with and without *mlrA*) have been isolated from *Taihu* (Hu *et al.*, 2009, Chen *et al.*, 2010, Jiang *et al.*, 2011, Yang *et al.*, 2014, Yang *et al.*, 2014, Zhu *et al.*, 2016). In addition, *mlrA* is hypothesized to be constitutively expressed (Dziga *et al.*, 2012), which would imply if $mlrA^+$ microbes were present and active, these genes should be detectable in metatranscriptomes.

One limitation to PCR-based approaches for detecting *mlr* genes in isolates or environmental samples is that the diverse sequences may be missed due to mismatch of primers. Currently, all of the publicly available *mlrA* sequences share 87-100 percent identities on both the nucleotide and amino acid level (Figure 1). Repeated use of non-degenerate primers based on highly identical sequences may explain why there are still few

representative *mlrA* sequences available and consequently, why many MC degraders remain "negative" for a microcystinase homolog. Shotgun metagenomes and metatranscriptomes provide the ability to expand the knowledge about the diversity of genes not yet discovered. Reads from the metatranscriptomes in this study were assembled into contigs and screened for the marker gene *mlrA* using BLASTx. No contigs from any sample fell within the *mlrA* clade (Figure 4). Only five contigs, all from Lake Erie, were similar to *mlrA* sequences; the highest identity was only 40% and all of them had an E-value of 0.00001 (Figure 4, Appendix Figure 1), suggesting low

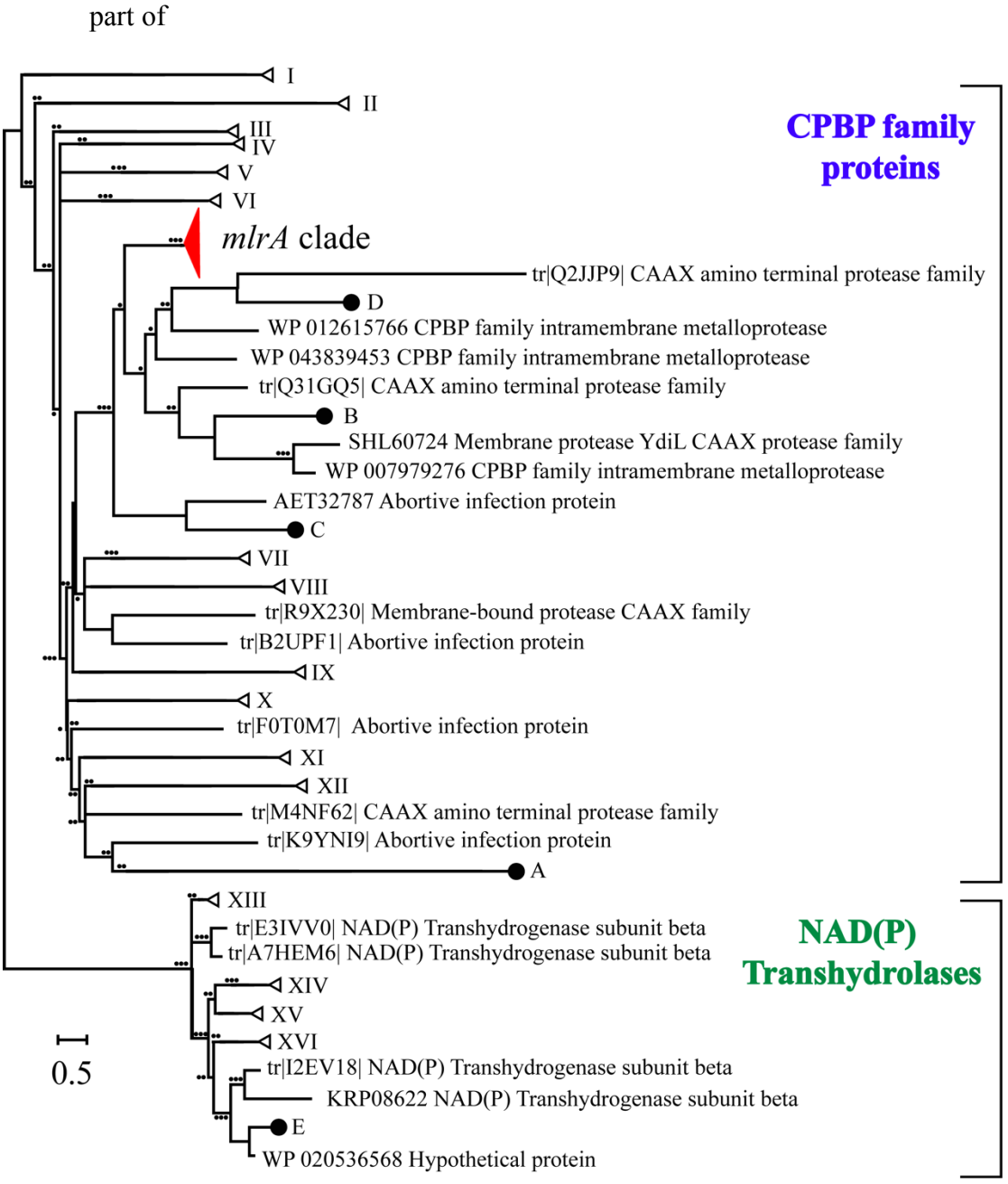

Figure 6.4 Phylogenetic analysis of putative *mlrA* contigs from Lake Erie and Taihu

Maximum likelihood phylogenetic tree with *mlrA*-like contigs aligned using pplacer indicated by large closed circles. The letters correspond to which sample they were found and this information is in Appendix Figure 1. The *mlrA* clade is collapsed for visualization purposes in red. Roman numerals represent collapsed branches and sequences within those branches are listed in the Appendix Table 5-20. Smaller closed circles at nodes represent likelihood ratios. ●●● >90, ●● >70, ● >50

probability that these sequences encoded *mlrA* function. Phylogenetic analyses of four of these contigs suggest they were part of the same protein family as *mlrA,* the CPBP family (Pei *et al*., 2011). The fifth contig aligned within a clade of putative NAD(P) transhydrolases and very distantly related to *mlrA*.

For comparison, this technique was also used on 454-generated metagenomic data from Mou *et al*. (2013) from mesocosm experiments using water from Lake Erie. Four metagenomes were generated: two that followed treatment with MCs and two which were controls to examine changes in community structure and function upon MC additions. One 500 bp sequence from one MC treatment shared 95.63% identity with *mlrA* from *Sphingopyxis* sp. C1 with an E-value of 0 using BLASTx (Appendix Figure 2). No other sequences were similar to *mlrA* using the previously described cutoffs listed above, but several sequences with a higher E-value than 0.00001 shared identity with *mlrA*. It is suspected that these sequences are also very distantly related to *mlrA* but perhaps the same protein family. These observations align with those from Mou *et al*. (2013) and suggest the *mlr* pathway was not a dominant mechanism for removal of MCs in Lake Erie. This demonstrates the ability of the technique described here to detect these genes of interest in sequence data.

The function of proteins within the CPBP family in prokaryotes remains largely unclear. These sequences are typically annotated as CAAX proteases or abortive infection proteins, and the literature suggests they may confer immunity to antimicrobial peptides secreted by the self or other bacteria (Pei *et al.*, 2011, Barrett *et al.*, 2012). A role for these proteins in MC degradation has been hypothesized previously (Kansole & Lin, 2016), but this was based on presence/absence and remains to be experimentally confirmed. To determine if these putative genes were highly expressed and potential candidates for MC removal, reads from the samples in which the contigs

were found were mapped to the respective contigs. Reads mapped accounted for a low percentage of the total number of reads in each individual sample (Figure 5, Appendix Table 4). For comparison, reads were to *mcyD* from *Microcystis aeruginosa* NIES 843 (Figure 5, Appendix Table 4). These results confirm the prevalence of toxic cyanobacteria across time and space in both lakes at the time of sampling, and despite the high proportion of reads mapping to *mcyD*, *mlrA* or *mlrA*-like CPBP protein sequences are not nearly as prevalent or abundant in the samples.

The existence of alternative pathways for MC degradation has been suggested numerous times in the literature when *mlr* genes have not been detected. No other also mapped mechanisms have been fully elucidated to date, but a few have been hypothesized. Mou *et al*. (2013) observed genes involved in xenobiotic metabolism, such as glutathione S-transferases (GSTs), were in higher abundance when MCs were degraded in Lake Erie. GST has also been implicated in the detoxification of MCs in some higher plants, invertebrates and vertebrates by first conjugating to the MCs to be targeted for removal by further processes (Campos & Vasconcelos, 2010). Takenaka *et al*. (1997) provides experimental evidence for the role of alkaline proteases from *Pseudomonas aeruginosa* in the degradation of MCs. Consequently, the metatranscriptomes presented in this study were screened for the presence of these genes as potential routes for removal of MCs. Despite using stricter cutoffs, contigs were identified across all studies in Lake Erie and *Taihu* as putative GSTs and putative alkaline proteases. Reads from samples with putative GST or alkaline proteases were mapped to the respective contigs to compare relative abundance of reads from these genes to *mlrA*-like genes (Figure 5, Appendix Table 4). These results suggest that GSTs and alkaline

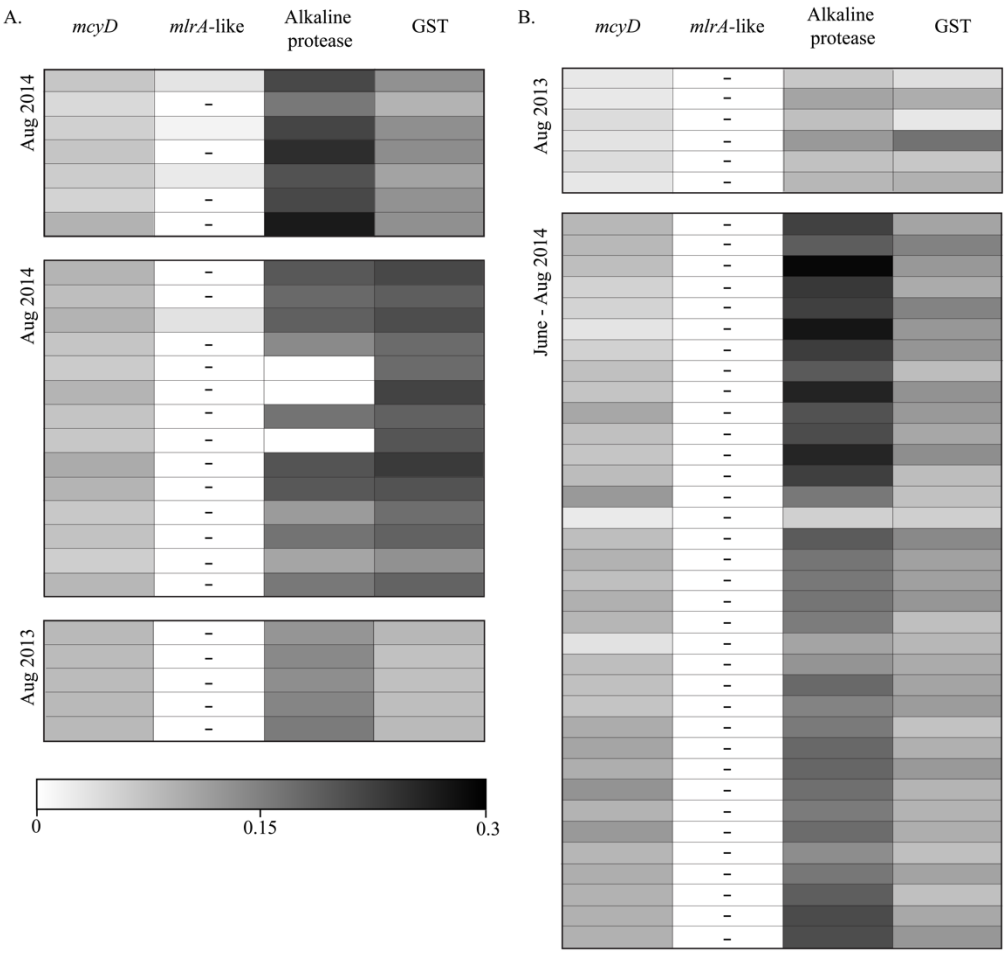

Figure 6.5 Glutathione S-transferases, alkaline proteases and *mlrA* reads in Lake Erie and Taihu

Shadeplots representing a qualitative comparison of reads that mapped to putative *mlrA*, GST, alkaline protease contigs and *mcyD* from *M. aeruginosa* NIES 843, across all samples. The reads that mapped to any putative gene of interest were totaled within each sample and normalized by total library size. For visualization purposes, these numbers were fourth root transformed. A dashed line within a box depicts a value of 0. Library sizes and values used for the shadeplot prior to transformation can be found in Appendix Table 4. Samples in the shadeplot also appear in the same order as they do in the Appendix Table 4 for reference.

proteases are actively being transcribed throughout both Lake Erie and *Taihu* across space and time and may play a larger role in biodegradation of MCs than the *mlr* pathway.

The metatranscriptomes used in this study were generated from environmentally relevant samples in which toxic cyanobacteria were present at varying densities. Surprisingly, our analyses indicated *mlr* pathway was not actively being transcribed in the surface waters of these lakes despite the presence of toxic cyanobacteria. There is growing evidence that MC biodegradation appears to be ubiquitous across different environmental samples worldwide and MC degraders will even form tight associates with toxic cyanobacteria (Maruyama *et al.*, 2003). Many studies in which *mlrA* has been detected have involved enriching or selecting for MC degraders. Lezcano *et al*. (2018) observed increased abundances of *mlrA* while the bloom was on the decline and proposed that the MC-degrading community may not be able to outcompete the cyanobacteria during times of large blooms. Despite consecutive daily and monthly sampling in this study, the samples were collected during on-going bloom events in the water column and this may explain the complete lack of detection of *mlr* genes in these samples. This work also suggests different MC degradation pathways may be occurring in the water column from those that are occurring in sediments or enriched samples. Many of the microcystin-degrading bacteria were initially isolated from enriched biofilms (Yang *et al.*, 2014) or sand filters (Ho *et al.*, 2006, Lee *et al.*, 2006, Hoefel *et al.*, 2009, Eleuterio & Batista, 2010) from water treatment systems, which may represent a very different microbial community than exists in the water column. The impact of these sediment processes on biological degradation in the water column will be highly dependent on lake geomorphology and climate events. Indeed, it is also possible that MC biodegradation was not occurring during times in which these samples were collected. If this is true, this highlights the crucial need for routine detection and implementation of appropriate water treatments, as

biodegradation of MCs in surface waters may not be as universally important as previously thought. However, evidence for the activity of other proposed mechanisms, such as putative GSTs and alkaline proteases, across almost all of the samples suggests these potential pathways and probably several more may be more relevant in *Taihu* and Erie during toxic bloom events as mechanisms for MC biodegradation.

Conclusions

With the prevalence and duration of cyanobacterial harmful algal blooms on the rise worldwide, it is important to understand the fate of the toxins produced during these blooms. The capacity for MC biodegradation has been demonstrated in microbial communities from a variety of environments all over the world suggesting the ability to biodegrade MCs is ubiquitous. The *mlr* pathway is currently the only characterized mechanism for MC biodegradation. However, there are many gaps in the general understanding of the diversity and ubiquity of this pathway as the route of MC biodegradation, especially in the natural environment. The expression of the genes in the *mlr* pathway in environmental samples had not been investigated previously with or without the addition of MCs.

In the present study, an extensive and quantitative survey of five studies yielding eighty metatranscriptomes from two eutrophic freshwater great lakes suggested that *mlr* genes were not being expressed in either of these systems. These samples were taken during MC-producing bloom events where cyanobacteria were actively transcribing *mcy* genes for MC production, yet genes in the *mlr* pathway was not detected. Despite the detection of a small number of *mlrA*-like genes sequences, these do not appear to be highly expressed or ubiquitous across samples examined in this multi-survey study. These data confirm that the *mlr* pathway was not the dominant mechanism

in these systems for the degradation of MCs. This is in accordance with previous work and supports the on-going hypothesis that other pathways are involved in this process. Evidence for expression of putative GSTs and alkaline proteases was observed across almost all of the samples suggesting these may more relevant route for biodegradation in the environment and warrant further investigation.

Acknowledgements

This work was made possible due to grant NA11NOS4780021 from NOS National Center for Coastal Ocean Science (NCCOS) under their Harmful Algal Bloom Prevention Control and Mitigation program to SWW and GLB, and National Science Foundation awards (DEB-1240870; IOS-1451528) to SWW. We thank Dr. Xiaozhen Mou for providing access to metagenomes. The sequencing work conducted by the U.S. DOE Joint Genome Institute (RMLM), a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. DOE under Contract No. DE-AC02-05CH11231. The authors also acknowledge the generous support of the *Kenneth & Blaire Mossman* endowment to the University of Tennessee.

References

Barrett A. J., Woessner J. F. & Rawlings N. D. 2012. *Handbook of proteolytic enzymes*. Elsevier.

Bourne D. G., Blakeley R. L., Riddles P. & Jones G. J. 2006. Biodegradation of the cyanobacterial toxin microcystin LR in natural water and biologically active slow sand filters. *Water Res*. 40, 1294-1302.

Bourne D. G., Riddles P., Jones G. J., Smith W. & Blakeley R. L. 2001. Characterisation of a gene cluster involved in bacterial degradation of the cyanobacterial toxin microcystin LR. *Environmental Toxicology*. 16, 523-534.

Bourne D. G., Jones G. J., Blakeley R. L., Jones A., Negri A. P. & Riddles P. 1996. Enzymatic pathway for the bacterial degradation of the cyanobacterial cyclic peptide toxin microcystin LR. *Applied and Environmental microbiology*. 62, 4086-4094.

Campos A. & Vasconcelos V. 2010. Molecular mechanisms of microcystin toxicity in animal cells. *International journal of molecular sciences*. 11, 268-287.

Carmichael W. W. 2001. Health effects of toxin-producing cyanobacteria:"The CyanoHABs". *Human and ecological risk assessment: An International Journal*. 7, 1393-1407.

Carmichael W. W. & Boyer G. L. 2016. Health impacts from cyanobacteria harmful algae blooms: Implications for the North American Great Lakes. *Harmful Algae*. 54, 194-212.

Chen J., Hu L. B., Zhou W., Yan S. H., Yang J. D., Xue Y. F. & Shi Z. Q. 2010. Degradation of microcystin-LR and RR by a *Stenotrophomonas* sp. strain EMS isolated from Lake Taihu, China. *Int J Mol Sci*. 11, 896-911.

Chorus I. & Bartram J. 1999. Toxic cyanobacteria in water: a guide to their public health consequences, monitoring and management.

Corbel S., Mougin C. & Bouaïcha N. 2014. Cyanobacterial toxins: modes of actions, fate in aquatic and soil ecosystems, phytotoxicity and bioaccumulation in agricultural crops. *Chemosphere*. 96, 1-15.

Dai G., Peng N., Zhong J., Yang P., Zou B., Chen H., Lou Q., Fang Y. & Zhang W. 2017. Effect of metals on microcystin abundance and environmental fate. *Environmental Pollution*. 226, 154-162.

Davenport E. J. 2016. Diel regulation of metabolic functions of a western Lake Erie *Microcystis* bloom informed by metatranscriptomic analysis. Thesis, Bowling Green State University.

Dexter J., Dziga D., Lv J., Zhu J., Strzalka W., Maksylewicz A., Maroszek M., Marek S. & Fu P. 2018. Heterologous expression of *mlrA* in a photoautotrophic host–Engineering cyanobacteria to degrade microcystins. *Environmental Pollution*.

Dziga D., Wladyka B., Zielinska G., Meriluoto J. & Wasylewski M. 2012. Heterologous expression and characterisation of microcystinase. *Toxicon*. 59, 578-586.

Dziga D., Wasylewski M., Wladyka B., Nybom S. & Meriluoto J. 2013. Microbial degradation of microcystins. *Chemical research in toxicology*. 26, 841-852.

Dziga D., Zielinska G., Wladyka B., Bochenska O., Maksylewicz A., Strzalka W. & Meriluoto J. 2016. Characterization of Enzymatic Activity of MlrB and MlrC Proteins Involved in Bacterial Degradation of Cyanotoxins Microcystins. *Toxins*. 8, 76.

Eleuterio L. & Batista J. R. 2010. Biodegradation studies and sequencing of microcystin-LR degrading bacteria isolated from a drinking water biofilter and a fresh water lake. *Toxicon*. 55, 1434-1442.

Guindon S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W. & Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59, 307-321.

Harada K.-I. & Tsuji K. 1998. Persistence and decomposition of hepatotoxic microcystins produced by cyanobacteria in natural environment. *Journal of Toxicology: Toxin Reviews*. 17, 385-403.

He X., Liu Y.-L., Conklin A., Westrick J., Weavers L. K., Dionysiou D. D., Lenhart J. J., Mouser P. J., Szlag D. & Walker H. W. 2016. Toxic cyanobacteria and drinking water: Impacts, detection, and treatment. *Harmful Algae*. 54, 174-193.

Hernández J. M., 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. *Medical hypotheses*. 72, 539-540.

Ho L., Gaudieux A. L., Fanok S., Newcombe G. & Humpage A. R. 2007. Bacterial degradation of microcystin toxins in drinking water eliminates their toxicity. *Toxicon*. 50, 438-441.

Ho L., Meyn T., Keegan A., Hoefel D., Brookes J., Saint C. P. & Newcombe G. 2006. Bacterial degradation of microcystin toxins within a biologically active sand filter. *Water Research*. 40, 768-774.

Ho L., Hoefel D., Palazot S., Sawade E., Newcombe G., Saint C. P. & Brookes J. D. 2010. Investigations into the biodegradation of microcystin-LR in wastewaters. *J Hazard Mater*. 180, 628-633.

Hoefel D., Adriansen C. M., Bouyssou M. A., Saint C. P., Newcombe G. & Ho L. 2009. Development of an *mlrA* gene-directed TaqMan PCR assay for quantitative assessment of microcystin-degrading bacteria within water treatment plant sand filter biofilms. *Appl Environ Microbiol*. 75, 5167-5169.

Hu L. B., Yang J. D., Zhou W., Yin Y. F., Chen J. & Shi Z. Q. 2009. Isolation of a *Methylobacillus* sp. that degrades microcystin toxins associated with cyanobacteria. *New biotechnology*. 26, 205-211.

Jiang Y., Shao J., Wu X., Xu Y. & Li R. 2011. Active and silent members in the mlr gene cluster of a microcystin-degrading bacterium isolated from Lake Taihu, China. *FEMS Microbiol Lett*. 322, 108-114.

Jimbo Y., Okano K., Shimizu K., Maseda H., Fujimoto N., Utsumi M. & Sugiura N. 2010. Quantification of microcystin-degrading bacteria in a biofilm from a practical biological treatment facility by real-time PCR. *Journal of Water and Environment Technology*. 8, 193-201.

Kansole M. M. & Lin T.-F. 2016. Microcystin-LR Biodegradation by *Bacillus* sp.: Reaction Rates and Possible Genes Involved in the Degradation. *Water*. 8, 508.

Krausfeldt L. E., Tang X., van de Kamp J., Gao G., Bodrossy L., Boyer G. L. & Wilhelm S. W. 2017. Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu. *FEMS Microbiology Ecology*. 93,

Lee Y.-J., Jung J.-M., Jang M.-H., Ha K. & Joo G.-J. 2006. Degradation of microcystins by adsorbed bacteria on a granular active carbon (GAC) filter during the water treatment process. *Journal of Environmental Biology*. 37, 317-322.

Lezcano M. Á., Quesada A. & El-Shehawy R. 2018. Seasonal dynamics of microcystin-degrading bacteria and toxic cyanobacterial blooms: Interaction and influence of abiotic factors. *Harmful algae*. 71, 19-28.

Lezcano M. Á., Morón-López J., Agha R., López-Heras I., Nozal L., Quesada A. & El-Shehawy R. 2016. Presence or absence of *mlr* genes and nutrient concentrations co-determine the microcystin biodegradation efficiency of a natural bacterial community. *Toxins*. 8, 318.

Li J., Li R. & Li J. 2017. Current research scenario for microcystins biodegradation–A review on fundamental knowledge, application prospects and challenges. *Science of The Total Environment*. 595, 615-632.

Li J., Peng L., Li J. & Qiao Y. 2014. Divergent responses of functional gene expression to various nutrient conditions during microcystin-LR biodegradation by *Novosphingobium* sp. THN1 strain. *Bioresource technology*. 156, 335-341.

Li J., Shimizu K., Utsumi M., Nakamoto T., Sakharkar M. K., Zhang Z. & Sugiura N. 2011. Dynamics of the functional gene copy number and overall bacterial community during microcystin-LR degradation by a biological treatment facility in a drinking water treatment plant. *Journal of bioscience and bioengineering*. 111, 695-701.

Li J., Shimizu K., Maseda H., Lu Z., Utsumi M., Zhang Z. & Sugiura N. 2012. Investigations into the biodegradation of microcystin-LR mediated by the biofilm in wintertime from a biological treatment facility in a drinking-water treatment plant. *Bioresource technology*. 106, 27-35.

Maruyama T., Kato K., Yokoyama A., Tanaka T., Hiraishi A. & Park H.-D. 2003. Dynamics of microcystin-degrading bacteria in mucilage of *Microcystis. Microbial Ecology*. 46, 279-288.

Mazur H. & Plinski M. 2001. Stability of cyanotoxins, microcystin-LR, microcystin-RR and nodularin in seawater and BG-11 medium of different salinity. *Oceanologia*. 43, 329-339.

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.

Pei J., Mitchell D. A., Dixon J. E. & Grishin N. V. 2011. Expansion of type II CAAX proteases reveals evolutionary origin of γ-secretase subunit APH-1. *Journal of molecular biology*. 410, 18-26.

Smith J. L., Schulz K. L., Zimba P. V. & Boyer G. L. 2010. Possible mechanism for the foodweb transfer of covalently bound microcystins. *Ecotoxicology and Environmental Safety*. 73, 757-761.

Steffen M. M., Belisle B. S., Watson S. B., Boyer G. L., Bourbonniere R. A. & Wilhelm S. W. 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Applied and environmental microbiology*. 81, 3268-3276.

Steffen M. M., Davis T. W., McKay R. M., Bullerjahn G. S., Krausfeldt L. E., Stough J. M., Neitzey M. L., Gilbert N. E., Boyer G. L. & Johengen T. H. 2017. Ecophysiological examination of the Lake Erie Microcystis bloom in 2014: linkages between biology and the water supply shutdown of Toledo, Ohio. *Environmental Science & Technology*.

Stough J. M., Tang X., Krausfeldt L. E., Steffen M. M., Gao G., Boyer G. L. & Wilhelm S. W. 2017. Molecular prediction of lytic vs lysogenic states for *Microcystis* phage: Metatranscriptomic evidence of lysogeny during large bloom events. *PloS one*. 12, e0184146.

Svircev Z., Krstic S., Miladinov-Mikov M., Baltic V. & Vidovic M. 2009. Freshwater cyanobacterial blooms and primary liver cancer epidemiological studies in Serbia. *Journal of Environmental Science and Health Part C*. 27, 36-55.

Takenaka S. & Watanabe M. F. 1997. Microcystin LR degradation by *Pseudomonas aeruginosa* alkaline protease. *Chemosphere*. 34, 749-757.

Tamura K., Dudley J., Nei M. & Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 24, 1596-1599.

Tsuji K., Naito S., Kondo F., Ishikawa N., Watanabe M. F., Suzuki M. & Harada K.-i. 1994. Stability of microcystins from cyanobacteria: effect of light on decomposition and isomerization. *Environmental Science & Technology*. 28, 173-177.

Wang X., Utsumi M., Gao Y., Li Q., Tian X., Shimizu K. & Sugiura N. 2016. Influences of metal ions on microcystin-LR degradation capacity and dynamics in microbial distribution of biofilm collected from water treatment plant nearby Kasumigaura Lake. *Chemosphere*. 147, 230-238.

Wood R. 2016. Acute animal and human poisonings from cyanotoxin exposure—A review of the literature. *Environment international*. 91, 276-282.

Yang F., Zhou Y., Yin L., Zhu G., Liang G. & Pu Y. 2014. Microcystin-degrading activity of an indigenous bacterial strain *Stenotrophomonas acidaminiphila* MC-LTH2 isolated from Lake Taihu. *PloS one*. 9, e86216.

Yang F., Zhou Y., Sun R., Wei H., Li Y., Yin L. & Pu Y. 2014. Biodegradation of microcystin-LR and-RR by a novel microcystin-degrading bacterium isolated from Lake Taihu. *Biodegradation*. 25, 447-457.

Zhu X., Shen Y., Chen X., Hu Y. O., Xiang H., Tao J. & Ling Y. 2016. Biodegradation mechanism of microcystin-LR by a novel isolate of *Rhizobium* sp. TH and the evolutionary origin of the mlrA gene. *International Biodeterioration & Biodegradation*. 115, 17-25.

Zurawell R. W., Chen H., Burke J. M. & Prepas E. E. 2005. Hepatotoxic cyanobacteria: a review of the biological importance of microcystins in freshwater environments. *Journal of Toxicology and Environmental Health, Part B*. 8, 1-37.

Appendix

Table A6.1. Microcystin degrading bacteria, their origin and the microcystin congeners that were degraded are listed below. The detection method for microcystin and whether or not *mlrA* or degradation productions were detected is also indicated.

Isolate	Site	Congener	MC Detection Method	*mlrA*	Deg. roduct	Reference
Sphingomonas sp. ACM3962	Murrumbidgee River, New South Wales	LR	HPLC	yes[a]	yes	(Jones *et al.*, 1994, Bourne *et al.*, 1996, Bourne *et al.*, 2001),
Pseudomonas aeruginosa	Ohi water purification facility, Japan	LR	HPLC	-	yes	(Takenaka & Watanabe, 1997)
Sphingosinicella microcystinivorans strain Y2 (AB084247)	Lake Suwa, Japan	LR, 6(Z)-Adda LR, RR, YR	HPLC	yes[b]	-	(Park *et al.*, 2001, Maruyama *et al.*, 2003, Saito *et al.*, 2003)
Sphingomonas sp. MD-1 (AB110635)	Lake Kasumigaura, Japan	LR, RR, YR	HPLC	yes[b]	-	(Saito *et al.*, 2003)
Sphingomonas sp. Strain 7CY (AB076083)	Lake Suwa, Japan	LR, LY, LW, LF. RR	HPLC	-	yes	(Ishii *et al.*, 2004)
Poterioochromonas (AY699607)	Culture of *Microcystis*	LR, RR	ELISA	-	-	(Ou *et al.*, 2005)
Paucibacter toxivorans sp.	Lake Tuusulanjarvi, Finland	-	-	-	-	(Rapala *et al.*, 2005)
Cultivated microbes from GAC filter	-	LR, RR	LCMS	-	-	(Lee *et al.*, 2006)
Sphingomonas sp. Strain B9 (AB159609)	Lake Sagami and Tsukui, Japan	LR, RR	HPLC	-	yes	(Tsuji *et al.*, 2006)
Sphingomonas sp. CBA4 (AY920497)	San Roque reservoir, Argentina	RR	hplc	-	yes	(Valeria *et al.*, 2006)
Sphingopyxis LH21 (DQ112242)	sand filter	LR, LA	HPLC	yes[b]	-	(Ho *et al.*, 2007)
Bifidobacteria and lactobacilli strains	Culture collection and dadih	LR	HPLC, ELISA, PPIA	no[c]	no	(Nybom *et al.*, 2007, Nybom *et al.*, 2008, Nybom *et al.*, 2012)
Bulkholderia sp. (DQ459360)	Patos Lagoon, Brazil	LR	ELISA, HPLC	-	-	(Lemes *et al.*, 2008)
Poterioochromonas ZX1 (EF165114)	-	LR	HPLC	-	-	(Zhang *et al.*, 2008)
Sphingopyxis sp. Strain C-1 (AB161684)	Hongfeng Lake, China	LR	HPLC	yes[b]	yes	(Okano *et al.*, 2009)
Methylobacillus sp. strain J10 (FJ418599)	Lake Taihu sludge, China	LR, RR	HPLC	-	-	(Hu *et al.*, 2009)
Arthrobacter sp., *Brevibacterium* sp., *Rhodococcus* sp.	Loch Rescobie, United Kingdom	LR, RR, LF, LW, LY	Biolog MT2, HPLC	no	yes	(Manage *et al.*, 2009, Lawton *et al.*, 2011)
Sphinogpyxis sp. USTB-05 (EF607053)	Lake Dianchi, China	RR	HPLC	yes[d]	yes	(Junfeng *et al.*, 2010, Zhang *et al.*, 2010) (Yan *et al.*, 2012, Yan *et al.*, 2012)
Morganella morganii (LAAFP-C25216), *Pseudomonas* sp. (*Sphingomonas* C25358 and Lake Mead C25459)	Lake Mead and Los Angeles Aqueduct Filtration Plant	LR	HPLC	-	-	(Eleuterio & Batista, 2010)
Stenotrophomonas sp EMS (FJ712028)	Lake Taihu sludge, China	LR, RR	HPLC	yes[b]	yes	(Chen *et al.*, 2010)

[a] : Fosmid library; [b]: Amplification, Saito et al. 2003 [c]: Amplification, Wang; [d]: Genomic searches [e]: Amplification, Jiang 2014; [f]: Amplification, Hoefel 2009; [g]: Amplification, Ho 2007

Table A6.1 continued.

Isolate	Site	Congener	MC Detection Method	*mlrA*	Deg. roduct	Reference
Bacillus sp. AMRI-03 (GU294753)	Tendaha Lake, Saudia Arabia	RR	ELISA, PPIA	yes[b]	-	(Alamri, 2010)
Ralstonia solanacearum	Lake Dianchi, China	LR	HPLC, LCMS	-	yes	(Zhang *et al.*, 2011)
Microbacterium sp. strain DC8 and *Rhizobium gallicum* strain DC7 (AY972457)	Lake Okeechobee	LR	HPLC	-	-	(Ramani *et al.*, 2012)
Novosphingobium sp. THN1 (HQ664117)	Lake Taihu, China	LR	HPLC	yes[e]	-	(Jiang *et al.*, 2011)
Trichaptum abietinum 1302BG	bamboo forest, China	LR	HPLC	-	-	(Jia *et al.*, 2012)
Bacillus flexus strain SSZ01 (GU112451)	Tendaha Lake, Saudi Arabia	RR	ELISA, PPIA	yes[b]	-	(Alamri, 2012)
Bacillus sp. strain EMB (FJ526332)	Soil from heap of algae, China	LR, RR	HPLC	yes[b]	no	(Hu *et al.*, 2012)
Pseudomonas aeruginosa PA14 and *Pseudomonas putida* KCCM10464	Culture Collection	-	ELISA	-	-	(Kang *et al.*, 2012)
Sphingopyxis sp. TT25 (JQ398614)	Myponga Reservoir	LR, RR, YR, LA	HPLC	yes[f]	-	(Ho *et al.*, 2012)
Novosphingobium sp. KKU03	Nong Kin Moo, Khon Kaen, Thailand	LR, [Dha^7] LR	HPLC	-	-	(Somdee *et al.*, 2013)
KKU-12	Kaen Nakhon Lake, Thailand	-	HPLC	-	-	(Phujomjai & Somdee, 2013)
Sphingmonas sp. NV3	Lake Rotoiti, New Zealand	[Dha^7] LR	HPLC	yes[b]	yes	(Somdee *et al.*, 2013)
Ochrobactrum sp. FDT5	activated sludge	LR	HPLC	-	-	(Jing *et al.*, 2014)
Portulaca oleracea cv.	-	LR	PPIA	-	-	(Isobe *et al.*, 2014)
Bordetella petrii (KC7348821.1)	Lake Taihu sludge	LR, RR	HPLC	yes[b]	yes	(Yang *et al.*, 2014)
Stenotrophomonas acidaminiphila MC LTH2 (KF305533)	Lake Taihu sludge	LR, RR	HPLC	no[b]	yes	(Yang *et al.*, 2014)
Trichoderma citronoviride strain kkuf-0955	Saudi eutrophic lake	-	ELISA	-	-	(Mohamed *et al.*, 2014)
Pseudomonas sp. WC-5, WC-4	Lake Taihu sediment, China	LR, RR	UPLC-ESI-MS/MS	-	-	(Li & Pan, 2014)
Bacillus nanhaiencis strain JZ-2013 (KF841622)	Lake Chaohu China	LR	ELISA	-	-	(Zhang *et al.*, 2015)
Arthobacter sp. strain T11	Lake Taoranting sediment, China	LR, RR	LCMS	-	-	(Qu *et al.*, 2015)
Aeromonas veronii w-s-03 (JF490063.1)	Sulejow Reservoir, Poland	LR	HPLC	no[b]	-	(Mankiewicz-Boczek *et al.*, 2015)
Psuedomonas aeruginosa DMXS	Palos Lagoon sediment, Brazil	LR	HPLC	-	-	(Lemes *et al.*, 2015)
Steroidobacter flavus	forest soil, China	LR	Biolog MT2	-	-	(Gong *et al.*, 2016)
Rhizobium sp. TH (KM365438)	Lake Taihu sediment	LR	HPLC	yes[b]	yes	(Zhu *et al.*, 2016)
Novosphingobium sp. KKU25s	Bueng nong Khot reservoir, Thailand	[Dha^7] LR	HPLC	yes[b]	yes	(Phujomjai *et al.*, 2016)
anaerobic amino acid degrading bacteria Ala-1 (DSM 12261)	DSM, Germany	LR	HPLC	no[b]	yes	(Bao & Wu, 2016)

[a] : Fosmid library; [b]: Amplification, Saito et al. 2003 [c]: Amplification, Wang; [d]: Genomic searches [e]: Amplification, Jiang 2014; [f]: Amplification, Hoefel 2009; [g]: Amplification, Ho 2007

Table A6.1 continued.

Isolate	Site	Congener	MC Detection Method	*mlrA*	Deg. roduct	Reference
Sphingopyxis sp IM-1 (KX085478), IM-2 (KX085479), IM-3 (KX085480)	Pond water in Alberche's River, Spain	LR, YR	HPLC, LCMS	yes[g]	-	(Lezcano *et al.*, 2016)
Paucibacter toxinivorans strain IM-4 (KX085481)		LR	HPLC	no[g]		
Bacillus sp.	Hulupi Lake, Taiwan	LR	ELISA	no	-	(Kansole & Lin, 2016)

[a] : Fosmid library; [b]: Amplification, Saito et al. 2003 [c]: Amplification, Wang; [d]: Genomic searches [e]: Amplification, Jiang 2014; [f]: Amplification, Hoefel 2009; [g]: Amplification, Ho 2007

Table A6.2. Environmental samples where MC biodegradation has been observed.

Environmental Sample	Site	Congeners	MC Detection Method	*mlrA*	Deg products	Reference
Surface water	Lake Centenary, Australia	LR*	HPLC, PPIA	-	-	(Jones & Orr, 1994)
Sediments and Water	Lake Tuusulanjarvi, Lake Vesijarvi, Lake Enojarvi, Lake Paijanna, Finland	LR, RR	HPLC	-	-	(Rapala *et al.*, 1994)
Activated sludge	WWTP, Alberta	LR	HPLC, PPI (with P^{32} labeled phosphatase	-	-	(Lam *et al.*, 1995)
Reservoir water, reservoir water with bed sediment	Grafham water	LR	HPLC		no	(Cousins *et al.*, 1996)
Lake surface water	Lake Tuusulanjarvi, Finland	LR	HPLC	-	-	(Lahti *et al.*, 1997)
Soil	Lake Plains and McLaren Flat, Australia	-	HPLC	-	-	(Miller & Fallowfield, 2001)
Water	Frederiksborg Slotsso	LR, RR	ELISA	-	-	(Christoffersen *et al.*, 2002)
Surface water	Frederiksborg Slotsso	LR	HPLC, ELISA, ^{14}C MC	-	-	(Hyenstrand *et al.*, 2003)
Groundwater/sediment	Molgarde moraine sediment, Denmark	Mostly LR*	^{14}C MC and PPIA	-	-	(Holst *et al.*, 2003)
Subsurface sediment and porewater	Water recharge facility and areas adjacent to Lake Arreso, Denmark					
Phototrophic biofilms	Svratka River, Czech Republic	YR, LR	HPLC	-	-	(Babica *et al.*, 2005)
Biofilm	Honeycomb tube settled in biological treatment facility, Japan	-	HPLC			(Saito *et al.*, 2003)
Estuary and marine water	Patos lagoon, Brazil	LR	PPIA	-	yes	(Lemes *et al.*, 2008)
Water samples	Loch Rescobie, Loch Balgavies, Loch Leven, Forfar Loch, River Cowie, River Carron, Scotland	LR	LCMS/MS	-	-	(Edwards *et al.*, 2008)
	Loch Rescrobe, River Carron, and Forfar Loch			-	yes	
Surface water, overlying water, sediments	Lake Taihu China	LR, RR, Dha^7 LR	HPLC and ELISA			(Chen *et al.*, 2008)
Soil	Leek cropland, China	RR	HPLC			(Bibo *et al.*, 2008)
Anthracite media and lake water	Drinking water filter and Lake Mead CA, USA	LR	HPLC	-	-	(Eleuterio & Batista, 2010)
water, aquifer sediment, filter sand	Lake Wannsee, Shore of Lake Wannsee, technical-scale slow sand and bank filtration, Germany	LR, RR, YR	ELISA	-	-	(Grützmacher *et al.*, 2009)
Sediment	Lake Dianchi, China	LR	HPLC	no[g]	yes	(Chen *et al.*, 2010)
biofilm	WTP near Lake Kasumigaura, Japan	-	-	yes	-	(Jimbo *et al.*, 2010)
Biofilm	WTP near Lake Kasumigaura, Japan	LR	HPLC	yes[f]	-	(Li *et al.*, 2011)
Biofilm	WTP near Lake Kasumigaura, Japan	LR	HPLC	yes[f]	-	(Li *et al.*, 2011)

[a] : Fosmid library; [b]: Amplification, Saito et al. 2003 [c]: Amplification, Wang; [d]: Genomic searches [e]: Amplification, Jiang 2014; [f]: Amplification, Hoefel 2009; [g]: Amplification, Ho 2007

Table A6.2 continued.

Environmental Sample	Site	Congeners	MC Detection Method	*mlrA*	Deg products	Reference
Anaerobically digested dairy cattle manure slurry added to water	Biogas plant and Lake Taihu, China	-*	ELISA	-	-	(Yuan *et al.*, 2011)
Water	Myponga Reservoir, Austrailia	LR, RR, YR, LA	HPLC	-	-	(Ho *et al.*, 2012)
Winter Biofilm	WTP near Lake Kasumigaura, Japan	LR	HPLC	yes[f]	-	(Li *et al.*, 2012)
Lake Water	Lake Erie, USA	LR	ELISA	yes[d]	-	(Mou *et al.*, 2013)
Sediment	Lake Yangebup, Australia	LR	HPLC	-	-	(Song *et al.*, 2014)
Rumen microorganisms	fistulated Swedish red cow	LR, RR, YR	LCMS	-	-	(Manubolu *et al.*, 2014)
Drinking water sludge	Quehua Drinking WTP, China	not clear, LR	ELISA	-	-	(Ma *et al.*, 2014)
Lake Water and Sediments	Lake Taihu, China	LR	HPLC	-	-	(Li *et al.*, 2016)
Biofilm	WTP near Lake Kasumigaura , Japan	LR	HPLC	yes[f]	-	(Wang *et al.*, 2016)
Water	Pond in Alberche's River, Spain	LR, YR	HPLC, LCMS	yes	-	(Lezcano *et al.*, 2016)
Water	Hulupi Lake, Taiwan	LR	ELISA	yes[f]	-	(Kansole & Lin, 2016)
Water and sediment	Poyang Lake, China	LR	HPLC	-	-	(Dai *et al.*, 2017)
Water	18 European lakes	dmLR	HPLC	no	yes	(Dziga *et al.*, 2017)
Biofilter enrichment studies						
Storage pond water	Experimental field of the German Federal Environmental Agengy	-	ELISA, HPLC	-	-	(Grützmacher *et al.*, 2002)
Irrigation Drainage Water	Murrumbidgee River (Australia)	LR	HPLC	yes*	no*	(Bourne *et al.*, 2006)
Sand	Morgan WTP, Australia	LR, LA	HPLC	yes[b]	-	(Ho *et al.*, 2006)
Water	Myponga Reservoir	LR	HPLC	yes[g]	-	(Ho *et al.*, 2007)
Sand	Morgan and Happy WTP, Australia				-	
Water	Myponga Reservoir	LR	HPLC	yes[f]	-	(Hoefel *et al.*, 2009)
TTE and ASTE	Bolivar waste water treatment plant (Austrailia)	LR	HPLC, PP2A	yes[f]	-	(Ho *et al.*, 2010)

[a] : Fosmid library; [b]: Amplification, Saito et al. 2003 [c]: Amplification, Wang; [d]: Genomic searches [e]: Amplification, Jiang 2014; [f]: Amplification, Hoefel 2009; [g]: Amplification, Ho 2007

Table A6.3. Environmental data collected for *Taihu* 2013 metatranscriptomes collected on 8/8/2013. Stn = Taihu station number; EC = electric conductivity; TDS = total dissolved solids; NTU = turbidity; DO = dissolved oxygen

Sample ID	Latitude/ Longitude	Stn	Sampling Time	Water temp (°C)	EC (µs/cm)	TDS (g/L)	Sal (‰)	pH	NTU	DO (%)	DO (mg/ L)
mgm4768724	31.40/120.03	3	13:06	33.5	693	0.387	0.28	8.8	39.7	124	8.83
mgm47687228	31.48/120.19	6	13:28	33.74	692	0.386	0.28	8.8	74.4	118.4	8.37
mgm4768720	31.45/120.12	Ch	13:46	33.74	636	0.354	0.26	8.7	73.8	132.7	9.41
mgm4768721	31.35/120.33	17 A	14:25	34.21	570	0.315	0.23	9.4	159.2	213.2	14.96
mgm4768726	31.42/120.22	31	15:13	35.07	714	0.39	0.28	9	23.2	144.1	9.74
mgm47687230	31.39/120.30	13	15:36	35.01	714	0.391	0.29	8.9	51.3	119.3	8.45

Table A6.4. Information for published metatranscriptomes screened for microcystin degradation genes. Values represent the number of transcripts that mapped to putative MC degradation genes or the *mcyD* from *M. aeruginosa* normalized per 1,000,000 transcripts for each library.

Study site	Sample collection date	Latitude/ Longitude	Total reads	Putative GST	Putative alkaline protease	*mlrA*-like	*mcyD*
Lake Taihu	8/8/2013	31.40/120.03	45444090	1.94	17.46	0	0.55
	8/8/2013	31.48/120.19	50900990	86.10	127.00	0	0.57
	8/8/2013	31.45/120.12	43007896	0.65	31.50	0	2.85
	8/8/2013	31.35/120.33	65726452	752.00	206.00	0	1.10
	8/8/2013	31.42/120.22	65825864	18.50	27.90	0	2.98
	8/8/2013	31.39/120.30	48868494	70.80	51.10	0	0.61
Lake Taihu	6/7/2014	31.51/120.19	25765226	130.00	2440.00	0	49.90
	6/7/2014	31.42/120.22	25916472	460.00	1290.00	0	48.70
	6/7/2014	31.39/120.23	24642268	210.00	7130.00	0	34.10
	6/7/2014	31.44/120.19	25361148	100.00	2930.00	0	7.60
	6/7/2014	31.45/120.12	29455482	440.00	2500.00	0	7.20
	6/7/2014	31.34/120.18	14132414	220.00	5680.00	0	1.00
	6/7/2014	31.31/119.95	18754882	240.00	2670.00	0	8.40
	7/3/2014	31.51/120.19	15217108	40.00	1400.00	0	30.40
	7/3/2014	31.42/120.22	22913270	260.00	4450.00	0	22.30
	7/3/2014	31.39/120.23	24230822	210.00	1680.00	0	112.20
	7/3/2014	31.44/120.19	27860460	110.00	1940.00	0	29.50
	7/3/2014	31.45/120.12	24019620	310.00	4280.00	0	21.20
	7/3/2014	31.45/120.03	27331358	40.00	2590.00	0	38.80
	7/3/2014	31.40/120.03	29839878	30.00	630.00	0	203.00
	7/3/2014	31.31/119.95	36080570	10.00	10.00	0	0.40
	8/14/2014	31.51/120.19	26393394	360.00	1380.00	0	33.30
	8/14/2014	31.42/120.22	25006404	150.00	680.00	0	67.20
	8/14/2014	31.44/120.19	23556748	150.00	620.00	0	32.00
	8/14/2014	31.45/120.12	21064130	230.00	690.00	0	74.20
	8/14/2014	31.40/120.03	22781612	30.00	560.00	0	54.10
	8/14/2014	31.31/119.95	25866280	50.00	130.00	0	1.40
	9/9/2014	31.51/120.19	26900236	90.00	250.00	0	33.40
	9/9/2014	31.39/120.23	24182112	130.00	950.00	0	30.10
	9/9/2014	31.44/120.19	26357046	180.00	450.00	0	23.30
	9/9/2014	31.45/120.12	25226958	30.00	590.00	0	89.40
	9/9/2014	31.45/120.03	25766550	70.00	970.00	0	121.50
	9/9/2014	31.40/120.19	24550814	200.00	1030.00	0	81.70
	9/9/2014	31.31/119.95	27341606	60.00	800.00	0	259.90
	10/8/2014	31.51/120.19	36054800	60.00	560.00	0	62.50
	10/8/2014	31.42/120.22	25758458	80.00	870.00	0	203.00
	10/8/2014	31.39/120.23	24526958	30.00	320.00	0	53.80

Table A6.4 continued.

Study site	Sample collection date	Latitude/ Longitude	Total reads	Putative GST	Putative alkaline protease	*mlrA*-like	*mcyD*
	10/8/2014	31.44/120.19	32543446	130.00	650.00	0	79.50
	10/8/2014	31.45/120.03	27724526	30.00	1270.00	0	65.10
	10/8/2014	31.40/120.19	43244796	110.00	1970.00	0	71.10
	10/8/2014	31.31/119.95	16735292	220.00	1910.00	0	69.20
Lake Erie	7/2013	41.78/83.33	54356308	50.00	237.00	0	50.00
	7/2013	41.78/83.33	61014100	30.00	368.00	0	40.00
	7/2013	41.78/83.33	80920356	30.00	310.00	0	40.00
	7/2013	41.76/-83.30	75038180	30.00	420.00	0	50.00
	7/2013	41.76/-83.30	104554048	40.00	572.00	0	50.00
Lake Erie	8/4/2014	41.70/-83.26	20300164	2090.00	1467.00	0	60.00
	8/4/2014	41.83/-83.36	22024928	1270.00	966.00	0	30.00
	8/4/2014	41.68/-83.23	22458988	1900.00	1190.00	1.291	60.00
	8/4/2014	41.83/-83.19	24016810	910.00	356.00	0	20.00
	8/4/2014	41.76/-83.33	22882652	910.00	0.00	0	10.00
	8/4/2014	41.74/-83.13	23578916	2370.00	0.00	0.721	60.00
	8/4/2014	41.77/-83.33	37413128	1180.00	717.00	0	20.00
	8/4/2014	41.71/-83.38	34135828	1580.00	0.00	0	20.00
	8/4/2014	41.72/-83.33	20657236	2840.00	1597.00	0	90.00
	8/4/2014	41.43/-83.20	25596208	1670.00	1455.00	0	60.00
	8/4/2014	41.43/-83.20	15586146	840.00	192.00	0	20.00
	8/4/2014	41.83/-83.19	27690326	1150.00	731.00	0	30.00
	8/4/2014	41.68/-83.23	19494314	270.00	126.00	0	10.00
	8/4/2014	41.70/-83.26	28095358	1120.00	632.00	0	50.00
Lake Erie	8/26/2014	41.70/-83.25	122728418	280.00	2093.00	1.002	20.00
	8/27/2014	41.70/-83.25	133304579	60.00	624.00	0	0.00
	8/27/2014	41.70/-83.25	125015545	300.00	2243.00	0.04	10.00
	8/27/2014	41.69/-83.25	131515291	320.00	3665.00	0	20.00
	8/27/2014	41.69/-83.25	120490984	130.00	1674.00	0.3652	10.00
	8/28/2014	41.70/-83.27	130524006	270.00	2138.00	0	10.00
	8/28/2014	41.70/-83.29	149931720	280.00	5152.00	0	70.00

Table A6.5. Amino acid reference sequences collapsed in Clade I of Figure 5. All sequences were downloaded from UniProt. An asterisk denotes sequences downloaded from NCBI's Genbank.

Entry code	Protein name
WP_033489078.1*	alpha/beta hydrolase [*Bifidobacterium pseudolongum*]
WP_068567594.1*	hypothetical protein. partial [*Oleiphilus* sp. HI0050]
tr|K9WBR7|	CAAX amino terminal protease family OS=*Microcoleus* sp. PCC 7113
tr|D5BHI3|	CAAX amino terminal protease *Zunongwangia profunda* strain DSM 18752
tr|A0A0S1TDB4|	CAAX protease *Nonlabens* sp. MIC269
tr|A0A0S1SD85|	CAAX protease *Sediminicola* sp. YIK13
tr|A0A0D6B8G4|	Caax amino terminal protease family protein *Rhodovulum sulfidophilum*
tr|A8LKN6|	Abortive infection protein *Dinoroseobacter shibae* strain DSM 16493
tr|K9SJS7|	Abortive infection protein *Pseudanabaena* sp. PCC 7367
tr|D3E8G0|	Abortive infection protein *Geobacillus* sp. strain Y412MC10
tr|A0A0S3TQ91|	Abortive infection protein *Fischerella* sp. NIES-3754

Table A6.6. Amino acid reference sequences collapsed in Clade II of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|A0A0A0WPN2|	CAAX protease self-immunity family protein *Bacillus weihenstephanensis*
tr|A0A0K2HBS1|	CAAX protease *Geobacillus stearothermophilus* 10
tr|A0A077J9C1|	CAAX protease *Bacillus* sp. X1
tr|A0A0B4RD33|	CAAX amino terminal protease family protein *Planococcus* sp. PAMC 21323
tr|L0KUH7|	CAAX amino terminal protease family *Methanomethylovorans hollandica* strain DSM 15978
tr|G2MQ64|	Abortive infection protein halophilic archaeon DL31
tr|D6DFN5|	CAAX amino terminal protease family *Clostridium cf. saccharolyticum* K10
tr|A0A0F8W958|	CAAX amino terminal protease self-immunity *Lokiarchaeum* sp. GC14_75
tr|F0LTH6|	CAAX amino terminal protease family protein *Vibrio furnissii* strain DSM 14383 / NCTC 11218
tr|F2LMZ6|	Abortive infection protein *Burkholderia gladioli* strain BSR3
tr|E3HK26|	CAAX amino terminal protease family protein 1 *Achromobacter xylosoxidans* strain A8
tr|A0A075PFS4|	CAAX protease *Pseudomonas fluorescens*
tr|A0A0G3II96|	CAAX protease *Pandoraea oxalativorans*
tr|A0A0E1B1C1|	CAAX amino protease *Pseudomonas aeruginosa* MTB-1
tr|A0A0D5WSU2|	CAAX protease *Klebsiella michiganensis*
tr|A0A0H3B5J1|	*Yersinia pseudotuberculosis* serotype O:3 strain YPIII
tr|A0A0F7HC32|	CAAX protease *Serratia fonticola*
tr|C8U3T9|	CAAX amino terminal protease family protein *Escherichia coli* O103:H2 strain 12009
tr|A0A089WSZ2|	CAAX protease *Cedecea neteri*
tr|A0A0H3L170|	CAAX amino Terminal protease family protein *Pantoea ananatis* strain AJ13355
tr|A4VNN3|	CAAX amino terminal protease family protein *Pseudomonas stutzeri* strain A1501
tr|W0V3T6|	CAAX amino terminal protease family protein *Janthinobacterium agaricidamnosum* NBRC 102515
tr|A0A097EQA8|	CAAX protease *Francisella* sp. FSC1006
tr|D4HBT4|	CAAX amino terminal protease family protein *Propionibacterium acnes* strain SK137
tr|E4U6Z1|	Abortive infection protein *Oceanithermus profundus* strain DSM 14977

Table A6.7. Amino acid reference sequences collapsed in Clade III of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|A8MEK1|	Abortive infection protein *Alkaliphilus oremlandii* strain OhILAs
tr|A0A0M3RJN2|	CAAX protease *Actinomyces* sp. oral taxon 414
tr|F2NKE7|	Abortive infection protein *Marinithermus hydrothermalis* strain DSM 14884
tr|A7H9T1|	Abortive infection protein *Anaeromyxobacter* sp. strain Fw109-5
tr|A0A0K1PAV6|	CAAX amino terminal protease family protein *Vulgatibacter incomptus*
tr|A0A0H4V8W5|	Abortive infection protein *Erythrobacter atlanticus*
tr|A0A0G3GTF8|	CAAX protease self-immunity *Corynebacterium mustelae*
tr|Q7B3U8|	CAAX amino protease *Escherichia coli*
tr|I7CJS1|	Abortive infection protein *Natrinema* sp. strain J7-2
tr|A0A0B5GL82|	CAAX protease *Haloarcula* sp. CBA1115
tr|W0JTI8|	Abortive phage infection protein *Halostagnicola larsenii* XH-48
tr|A0A085UF37|	CAAX amino protease family protein *Staphylococcus agnetis*
tr|B3EFA3|	Abortive infection protein *Chlorobium limicola* strain DSM 245
tr|B4SCD9|	Abortive infection protein *Pelodictyon phaeoclathratiforme* strain DSM 5477

Table A6.8. Amino acid reference sequences collapsed in Clade IV of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|F8CQK7|	CAAX amino terminal protease family protein *Myxococcus fulvus* strain ATCC BAA-855
tr|A0A0A1DS92|	Abortive infection protein *Nocardioides simplex*
tr|K4L176|	CAAX amino terminal protease family *Dehalobacter* sp. CF
tr|A0A0X1KJN7|	CAAX amino terminal protease *Thermococcus guaymasensis* DSM 11113

Table A6.9. Amino acid reference sequences collapsed in Clade V of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|Q1GZP0|	Abortive infection protein *Methylobacillus flagellatus* strain KT / ATCC 51484
tr|F0JCC6|	CAAX prenyl protease-related protein *Desulfovibrio desulfuricans* ND132

Table A6.10. Amino acid reference sequences collapsed in Clade VI of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|Q01Q98|	Abortive infection protein *Solibacter usitatus* strain Ellin6076
tr|B8J575|	Abortive infection protein *Anaeromyxobacter dehalogenans* strain 2CP-1 / ATCC BAA-258

Table A6.11. Amino acid reference sequences collapsed in Clade VII of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|F0NZ71|	*Abortive infection protein Weeksella virosa strain ATCC 43766*
tr|A0A0G3M4Q6|	CAAX protease *Chryseobacterium gallinarum*
tr|U3Q4F6|	CAAX Amino Terminal Protease Family Protein *Blattabacterium* sp./*Nauphoeta cinerea*
tr|M4Z7H3|	Putative CAAX family protease *Bradyrhizobium oligotrophicum* S58
tr|D9TPI2|	Abortive infection protein *Thermoanaerobacterium thermosaccharolyticum* strain ATCC 7956
tr|A0A0A7FX98|	CAAX protease self-immunity family protein *Clostridium baratii* str. Sullivan
tr|A2U1D0|	CAAX amino terminal protease family *Polaribacter* sp. MED152
tr|A0A0U1QXW4|	CAAX amino terminal protease family protein *Yersinia pseudotuberculosis* serotype O:1b (train IP 31758
tr|A0A0N1QWF4|	Caax amino protease family *Salmonella schwarzengrund* strain CVM19633

Table A6.12. Amino acid reference sequences collapsed in Clade VIII of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|A0A0D3LFF2|	Abortive infection bacteriophage resistance protein Abi superfamily *Flammeovirgaceae* bacterium 311
tr|K4IDE8|	Abortive infection bacteriophage resistance protein Abi superfamily *Psychroflexus torquis* strain ATCC 700755
tr|F6IG19|	Amino terminal protease *Novosphingobium* sp. PP1Y
tr|A0A0B8QV89|	CAAX amino protease *Listeria monocytogenes*
tr|C6PTE5|	Abortive infection protein *Clostridium carboxidivorans* P7
tr|G8QUV3|	CAAX amino terminal protease family *Sphaerochaeta pleomorpha* strain ATCC BAA-1885
tr|Q73N67|	CAAX amino terminal protease family protein *Treponema denticola* strain ATCC 35405
tr|A1R023|	CAAX amino terminal protease family *Borrelia turicatae* strain 91E135
tr|Q834J9|	Abortive infection protein *Enterococcus faecalis* strain ATCC 700802
tr|D6ZLM4|	CAAX amino terminal protease family protein *Streptococcus pneumoniae* serotype A19 strain TCH8431
tr|F4FTG9|	Abortive infection protein *Lactobacillus buchneri* strain NRRL B-30929
tr|G8PA89|	CAAX amino terminal protease self-immunity family protein *Pediococcus claussenii* strain ATCC BAA-344
tr|F4FWD9|	Abortive infection protein *Lactobacillus buchneri* strain NRRL B-30929
tr|B7HRK6|	Caax amino protease family *Bacillus cereus* strain AH187
tr|A0A0K2J9F4|	CAAX protease *Leptotrichia* sp. oral taxon 212
tr|K7VUH9|	CAAX amino terminal protease family protein *Lactococcus* lactis subsp. cremoris UC509.9
tr|D2SEX6|	Abortive infection protein *Geodermatophilus obscurus* strain ATCC 25078
tr|A0R2P6|	Abortive infection protein *Mycobacterium smegmatis* strain ATCC 700084
tr|A0A0K2GXN5|	Abortive infection protein *Corynebacterium lactis* RW2-5
tr|A0A0H3A1E6|	Caax amino protease family protein *Mycobacterium avium* strain 104
tr|A0A0M3QCP8|	Caax amino protease *Sphingopyxis* sp. strain 113P3
tr|G8UP72|	CAAX amino terminal protease family protein *Tannerella forsythia* strain ATCC 43037
tr|A0A0H3CB38|	CAAX amino terminal protease family *Caulobacter crescentus* strain NA1000
tr|B1IGW2|	CAAX amino terminal protease family protein *Clostridium botulinum* strain Okra / Type B1
tr|K0N4D2|	CAAX amino terminal protease family protein *Lactobacillus casei* W56
tr|S4ZH90|	CAAX amino terminal protease family *Lactobacillus casei* LOCK919
tr|A0A0K1S4B6|	Abortive infection protein *Microcystis panniformis* FACHB-1757
tr|K8E5Y6|	CAAX amino terminal protease family protein *Carnobacterium maltaromaticum* LMA28
tr|A0A0K2LFP2|	CAAX protease *Lactobacill*us heilongjiangensis
tr|A0A0G3XVQ5|	CAAX amino terminal protease self-immunity *Geobacillus* sp. 12AMOR1
tr|A0A0S3KA66|	CAAX protease *Enterococcus silesiacus*

Table A6.13. Amino acid reference sequences collapsed in Clade IX of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|Q81V64|	CAAX amino terminal protease family protein *Bacillus anthracis*
tr|A0A0U2XT44|	CAAX protease *Planococcus kocurii*
tr|W0U3N9|	CAAX amino terminal protease family protein *Ruminococcus bicirculans*
tr|H6LCK2|	Abortive infection protein *Acetobacterium woodii* strain ATCC 29683
tr|E6U5Y2|	Abortive infection protein *Ethanoligenens harbinense* strain DSM 18485
tr|I1YUK4|	CAAX protease self-immunity *Prevotella* intermedia strain 17
tr|B2A6P2|	Abortive infection protein *Natranaerobius* thermophilus strain ATCC BAA-1301
tr|K9R1C0|	CAAX amino terminal protease family *Nostoc* sp. strain ATCC 29411
tr|B9MT38|	CAAX amino terminal protease family protein *Populus trichocarpa*

Table A6.14. Amino acid reference sequences collapsed in Clade X of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|F8J5I3|	Abortive infection protein *Hyphomicrobium* sp. strain MC1
tr|A0A023X0T0|	CAAX protease self-immunity *Rubrobacter radiotolerans*
tr|A0A0D5ABX3|	Caax amino protease family protein *Rhodococcus* sp. B7740
tr|Q00UJ8|	CAAX amino terminal protease *Ostreococcus tauri*
tr|G2LDC3|	CAAX amino terminal protease family *Chloracidobacterium thermophilum* strain B
tr|F0RLC4|	Abortive infection protein *Deinococcus proteolyticus* strain ATCC 35074
tr|A0A089NZX3|	Abortive infection protein *Methylobacterium oryzae* CBMB20
tr|U3TN86|	CAAX amino terminal protease *Streptococcus dysgalactiae* subsp. *equisimilis* 167
tr|C9ABR6|	CAAX amino terminal protease *Enterococcus casseliflavus* EC20
tr|A1R9Q7|	Putative CAAX amino terminal protease family protein *Arthrobacter aurescens* strain TC1
tr|C5F8W6|	Abortive infection protein *Lactobacillus paracasei* subsp. paracasei
tr|A0A0K1DZM5|	CAAX protease self-immunity *Pyrobaculum* sp. WP30
tr|E6WNX0|	Abortive infection protein *Pseudoxanthomonas suwonensis* strain 11-1
tr|G7VCA5|	Abortive infection protein *Pyrobaculum ferrireducens*

Table 15. Amino acid reference sequences collapsed in Clade XI of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|K4KLX0|	Abortive infection protein *Simiduia agarivorans* strain DSM 21679
tr|D5X7L5|	Abortive infection protein *Thermincola potens* strain JR
tr|A0A076JJZ6|	CAAX amino terminal protease *Bifidobacterium* coryneforme
tr|A0A0E3YYJ1|	Abortive infection protein *Verrucomicrobia* bacterium IMCC26134

Table A6.16. Amino acid reference sequences collapsed in Clade XII of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|C1F1F6|	Protease. CAAX amino terminal family *Acidobacterium capsulatum* strain ATCC 51196
tr|E8V6K5|	Abortive infection protein *Terriglobus saanensis* strain ATCC BAA-1853
tr|A0A0E3V4X5|	Uncharacterized protein *Spirosoma radiotolerans*
tr|G0IVD9|	Abortive infection protein *Cyclobacterium marinum* strain ATCC 25205
tr|A0A0J8YB95|	Alpha/beta hydrolase bacteria symbiont BFo1 of *Frankliniella occidentalis*
tr|A0A148N4D0|	Alpha/beta hydrolase *Methylothermaceae* bacteria B4
tr|A0A148N686|	Alpha/beta hydrolase *Methylothermaceae* bacteria B42
tr|A4F611|	Abortive infection protein *Saccharopolyspora erythraea* strain ATCC 11635
tr|J7ITK2|	CAAX amino terminal protease family *Desulfosporosinus meridiei* strain ATCC BAA-275
tr|A0A023NMQ4|	Abortive infection protein *Dyella jiangningensis*
tr|A0A0K2J9F4|	CAAX protease *Leptotrichia* sp. oral taxon 212
tr|R4LSV1|	Abortive infection protein *Actinoplanes s*p. N902-109
tr|D9TBI2|	Abortive infection protein *Micromonospora aurantiaca* strain ATCC 27029

Table A6.17. Amino acid reference sequences collapsed in Clade XIII of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|A0A0T7BYY1|	NAD(P) transhydrogenase subunit beta *Calothrix* sp. 336
tr|K9R611|	NAD(P) transhydrogenase subunit beta *Rivularia* sp. PCC 7116
tr|A0A0S3PJY4|	NAD(P) transhydrogenase subunit beta *Nostoc* sp. NIES-3756
tr|K7W4I3|	NAD(P) transhydrogenase subunit beta *Anabaena* sp. 90
tr|B1WRZ8|	NAD(P) transhydrogenase subunit beta *Cyanothece* sp. strain ATCC 51142
tr|B0JL21|	NAD(P) transhydrogenase subunit beta *Microcystis aeruginosa* strain NIES-843

Table A6.18. Amino acid reference sequences collapsed in Clade XIV of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|A0A0U3ELT2|	NAD(P) transhydrogenase subunit beta *Paucibacter* sp. KCTC 42545
tr|F6ALX0|	NAD(P) transhydrogenase subunit beta *Delftia* sp. strain Cs1-4
tr|F8FVK9|	NAD(P) transhydrogenase subunit beta *Pseudomonas putida* strain S16
tr|A0A0E3MMH3|	NAD(P) transhydrogenase subunit beta *Pseudomonas fluorescens*
tr|A0A072ZSP2|	NAD(P) transhydrogenase subunit beta *Pseudomonas aeruginosa*
tr|A0A0C4WIX5|	NAD(P) transhydrogenase subunit beta *Azotobacter chroococcum* NCIMB 8003
tr|A4VFS8|	NAD(P) transhydrogenase subunit beta *Pseudomonas stutzeri* strain A1501
tr|M9V1X5|	NAD(P) transhydrogenase subunit beta *Comamonas* sp. 7D-2
tr|C5CJ56|	NAD(P) transhydrogenase subunit beta *Variovorax paradoxus* strain S110
tr|Q122W3|	NAD(P) transhydrogenase subunit beta *Polaromonas* sp. strain JS666
tr|A0A0H4WIE8|	NAD(P) transhydrogenase subunit beta *Bordetella hinzii*
tr|E3HEK2|	NAD(P) transhydrogenase subunit beta *Achromobacter xylosoxidans* strain A8
tr|A0A0D5V7C5|	NAD(P) transhydrogenase subunit beta *Paraburkholderia fungorum*
tr|B1JVW7|	NAD(P) transhydrogenase subunit beta *Burkholderia cenocepacia* strain MC0-3
tr|V5UGV1|	NAD(P) transhydrogenase subunit beta *Pandoraea pnomenusa*
tr|A0A0C5KBH9|	NAD(P) transhydrogenase subunit beta *Pandoraea vervacti*
tr|V5VI87|	NAD(P) transhydrogenase subunit beta *Acinetobacter baumannii*
tr|A4SHQ6|	NAD(P) transhydrogenase subunit beta *Aeromonas salmonicida* strain A449
tr|G0AZ73|	NAD(P) transhydrogenase subunit beta *Shewanella baltica* BA175
tr|B6ERP5|	NAD(P) transhydrogenase subunit beta *Aliivibrio salmonicida* strain LFI1238
tr|C6C545|	NAD(P) transhydrogenase subunit beta *Dickeya dadantii* strain Ech703
tr|A0A089UWL9|	NAD(P) transhydrogenase subunit beta *Cedecea neteri*
tr|A0A0H3FCN7|	NAD(P) transhydrogenase subunit beta *Rahnella* sp. strain Y9602
tr|A0A0M5L6H0|	NAD(P) transhydrogenase subunit beta *Proteus vulgaris*
tr|A4W9Z8|	NAD(P) transhydrogenase subunit beta *Enterobacter* sp. strain 638
tr|A0A0K2PE55|	NAD(P) transhydrogenase subunit beta *Cronobacter muytjensii* ATCC 51329
tr|A8AGX9|	NAD(P) transhydrogenase subunit beta *Citrobacter koseri* strain ATCC BAA-895
tr|W6J265|	NAD(P) transhydrogenase subunit beta *Kosakonia sacchari* SP1
tr|X2KE40|	NAD(P) transhydrogenase subunit beta *Salmonella enterica* subsp. enterica serovar Tennessee str. TXSC_TXSC08-19
sp|P0AB70|	NAD(P) transhydrogenase subunit beta *Shigella flexneri*
sp|P0AB67|	NAD(P) transhydrogenase subunit beta *Escherichia coli* strain K12
tr|W9B7K9|	NAD(P) transhydrogenase subunit beta *Klebsiella pneumoniae*
tr|W0LF76|	NAD(P) transhydrogenase subunit beta *Chania* multitudinisentens RB-25
tr|A0A076LQF6|	NAD(P) transhydrogenase subunit beta *Edwardsiella anguillarum* ET080813
tr|E3D6D2|	NAD(P) transhydrogenase subunit beta *Neisseria meningitidis* serogroup B strain alpha710
tr|W0PBG8|	NAD(P) transhydrogenase subunit beta *Advenella mimigardefordensis* DPN7
tr|A5UEJ8|	NAD(P) transhydrogenase subunit beta *Haemophilus influenzae* strain PittGG
tr|A3N0K1|	NAD(P) transhydrogenase subunit beta *Actinobacillus pleuropneumoniae* serotype 5b strain L20
tr|A0A0U3N4G8|	NAD(P) transhydrogenase subunit beta *Streptomyces* sp. CdTB01
tr|Q0SD62|	NAD(P) transhydrogenase subunit beta *Rhodococcus* jostii strain RHA1
tr|I7EN34|	NAD(P) transhydrogenase subunit beta *Phaeobacter gallaeciensis* strain 2.10
tr|A0A126V4K0|	NAD(P) transhydrogenase subunit beta *Halocynthiibacter arcticus*
tr|W8IJQ4|	NAD(P) transhydrogenase subunit beta *Ensifer adhaerens* OV14
tr|A0A0S2EP23|	NAD(P) transhydrogenase subunit beta *Aureimonas* sp. AU20
tr|A0A0K1U8T9|	NAD(P) transhydrogenase subunit beta *Marinobacter* sp. CP1
tr|K0IA78|	NAD(P) transhydrogenase subunit beta *Acidovorax* sp. KKS102
tr|C7RPC5|	NAD(P) transhydrogenase subunit beta *Accumulibacter phosphatis* strain UW-1
tr|A0A0U3ATY5|	NAD(P) transhydrogenase subunit beta *Herbaspirillum rubrisubalbicans* M1
tr|A0A0D6AUA4|	NAD(P) transhydrogenase subunit beta *Geminocystis* sp. NIES-3709
tr|B0CDX5|	NAD(P) transhydrogenase subunit beta *Acaryochloris marina* strain MBIC 11017
tr|A0A0K1E8K8|	NAD(P) transhydrogenase subunit beta *Chondromyces crocatus*
tr|D3RQH7|	NAD(P) transhydrogenase subunit beta *Allochromatium vinosum* strain ATCC 17899
tr|H6QKQ0|	NAD(P) transhydrogenase subunit beta *Rickettsia massiliae* str. AZT80
tr|Q5WY32|	NAD(P) transhydrogenase subunit beta *Legionella pneumophila* strain Lens

Table A6.19. Amino acid reference sequences collapsed in Clade XV of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein code
tr|A0A060DHD0|	NAD(P) transhydrogenase subunit beta *Azospirillum brasilense*
sp|Q2RSB4|	NAD(P) transhydrogenase subunit beta *Rhodospirillum rubrum* strain ATCC 11170
tr|A0A0G3XHT7|	NAD(P) transhydrogenase subunit beta *Croceicoccus naphthovorans*
tr|A0A0G3X717|	PntB. NAD(P) transhydrogenase beta subunit *Altererythrobacter marensis*
tr|A0A097EHV7|	NAD(P) transhydrogenase subunit beta *Sphingomonas taxi*
tr|E3HZ90|	NAD(P) transhydrogenase subunit beta *Rhodomicrobium vannielii* strain ATCC 17100
tr|I3XD95|	NAD(P) transhydrogenase subunit beta *Sinorhizobium fredii* USDA 257
tr|A0A0B4X6S5|	NAD(P) transhydrogenase subunit beta *Rhizobium gallicum* bv. gallicum R602
tr|I0G0Y4|	NAD(P) transhydrogenase subunit beta *Bradyrhizobium* sp. S23321
tr|B2IFK1|	NAD(P) transhydrogenase subunit beta *Beijerinckia indica* subsp. indica strain ATCC 9039
tr|A9MCQ1|	NAD(P) transhydrogenase subunit beta *Brucella* canis strain ATCC 23365
tr|L0KFI3|	NAD(P) transhydrogenase subunit beta *Mesorhizobium australicum* strain HAMBI 3006
tr|D8JTA7|	NAD(P) transhydrogenase subunit beta *Hyphomicrobium denitrificans* strain ATCC 51888
tr|A0A089NU63|	NAD(P) transhydrogenase subunit beta *Methylobacterium oryzae* CBMB20
tr|A0A0P0P295|	NAD(P) transhydrogenase subunit beta *Caulobacter henricii*
tr|E6UBX7|	NAD(P) transhydrogenase subunit beta *Ruminococcus albus* strain ATCC 27210
tr|A0A0E0YIL3|	NAD(P) transhydrogenase subunit beta *Alteromonas macleodii* strain English Channel 673
tr|A0A0B4XNA3|	NAD(P) transhydrogenase subunit beta *Alcanivorax pacificus* W11-5
tr|A0A076H2Z7|	NAD(P) transhydrogenase subunit beta *Synechococcus* sp. KORDI-100
tr|A4VFS8|	NAD(P) transhydrogenase subunit beta *Pseudomonas stutzeri* strain A1501

Table A6.20. Amino acid reference sequences collapsed in Clade XVI of Figure 5. All sequences were downloaded from UniProt.

Entry code	Protein name
tr|C4FEV3|	NAD(P) transhydrogenase subunit beta *Bifidobacterium angulatum* DSM 20098
tr|J9W3M7|	NAD(P) transhydrogenase subunit beta *Lactobacillus buchneri* CD034
tr|Q1LN61|	NAD(P) transhydrogenase subunit beta *Cupriavidus metallidurans* strain ATCC 43123
tr|H0PYA8|	NAD(P) transhydrogenase subunit beta *Azoarcus s*p. KH32C
tr|Q47BC7|	NAD(P) transhydrogenase subunit beta *Dechloromonas aromatica* strain RCB
tr|A0A0G3G195|	NAD(P) transhydrogenase subunit beta *Thioalkalivibrio versutus*
tr|A0A0S2FV81|	NAD(P) transhydrogenase subunit beta *Lysobacter capsici*
tr|B9K3Q9|	NAD(P) transhydrogenase subunit beta *Agrobacterium vitis* strain S4 / ATCC BAA-846
tr|S6CXN5|	NAD(P) transhydrogenase subunit beta *Acetobacter pasteurianus* 386B
tr|A0A109QFJ9|	NAD(P) transhydrogenase subunit beta *Thermus parvatiensis*
tr|A0A0K2RL61|	NAD(P) transhydrogenase subunit beta *Arthrobacter* sp. Hiyo8
tr|A0LWI2|	NAD(P) transhydrogenase subunit beta *Acidothermus cellulolyticus* strain ATCC 43068
tr|J9SI02|	NAD(P) transhydrogenase subunit beta *Gordonia* sp. KTR9
tr|R4V124|	NAD(P) transhydrogenase subunit beta *Mycobacterium abscessus* subsp. bolletii
tr|W5TA36|	NAD(P) transhydrogenase subunit beta *Nocardia nova* SH22a

Chapter 7 : Conclusions

The global increase in the frequency of harmful cyanobacterial blooms poses a serious threat to freshwater ecosystems. The impacts on this precious supply of water have severe ecological and economic impacts from both the accumulation of biomass and production of cyanotoxins that is responsible for massive fish kills, decreased diversity of plant and animal life, beach closures and drinking water shutdowns. Further, cyanotoxins as a risk to public health is becoming more substantial with the increasing number of water advisories and reported number of human and animal illnesses attributed to cyanobacterial blooms. Blooms have been largely attributed to increased anthropogenic activities like agriculture, farming and other industrial practices to support the growing population worldwide, as well as the many consequences of climate change, including warmer waters and increased precipitation. In addition, prevalent bloom-forming cyanobacteria are especially well adapted to not only succeed in these "perfect storm" conditions but withstand fluctuating environmental conditions and stressors. While these generalized statements are true, the factors that contribute to the formation and proliferation of cyanobacterial blooms are actually quite complex and multifaceted. There are still many gaps in our understanding of the biological mechanisms behind their success. This collection of work has addressed several aspects pertaining to the success of the toxic cyanobacteria *Microcystis*, while also examining pertinent interactions within the associated microbial communities that impact nutrient availability and toxin production during *Microcystis* dominated blooms. At the ecosystem level, we addressed questions about nitrogen (N) cycling in bloom communities, colony formation, and degradation of microcystins. On the laboratory scale, we asked specific questions about the influence of N on the physiology of *Microcystis* and the pathways relevant to microcystin production.

By collecting surface water samples from hypereutrophic Lake Taihu in China over a large spatial and temporal range, we were able to provide evidence through the use of a targeted gene expression microarray that N-cycling, primarily by the heterotrophic community, was occurring in surface waters and not restricted to sediments as previously thought. These processes, many of which are typically anaerobic, are usually only examined in the sediments or at the sediment-water interface in lakes. Spatially the function of the community was not different but who was responsible for these process differed between sites, especially near the mouth of a major river. Evidence for these processes in the surface waters suggests they may play a role in N-availability to the cyanobacterial community either by generating new N or contributing to the loss of N. This work was exploratory and hypothesis generating. Hypotheses generated from this work included:

1. Sediment resuspension or run off from terrestrial soils influence the N-cycling community and function in surface waters of shallow lakes.
2. Sediment particles and *Microcystis* colonies may be hotspots for N-cycling activity.
3. N-fixing heterotrophic bacteria in the surface waters are a source of new N to freshwater ecosystems.

This work was expanded upon by generating metatranscriptomes of fractionated samples from Lake Taihu to broadly investigate what nutrient transformations were occurring within the cyanobacterial colonies. We found that Proteobacteria were the most abundant phylum associated with colonies and the strategies of nutrient scavenging by these bacteria were different outside of the colony than within, supporting the hypothesis that cyanobacterial colonies are microniches and functionally different from the community outside of the colony. This is an important

consideration for experimental design and for sampling in the field. Differences in the *Microcystis* population were also observed between the colonies and the free-living fraction, and the observed differences implied that *Microcystis* forms colonies in response to a variety of stressful conditions. Hypotheses generated from this work included:

1. *Microcystis* forms colonies as a general stress response.
2. Colony formation protects against lytic phage infection.

To gain a better understanding of the effect of N form on *Microcystis* specifically, we also performed metabolomics experiments on *Microcystis aeruginosa* cultures growing on ^{15}N-nitrate, ^{15}N-ammonium, ^{15}N-urea and ^{13}C-urea. We were able to track N through the metabolome of *M. aeruginosa* and identify different physiological responses to varying N form. We found that cultures grown on urea were using "new" N to make arginine as well as more microcystin and its precursor amino acids, whereas cells grown on nitrate and ammonium were not. Arginine is not only an important precursor for microcystin production but may also be a storage molecule itself, and our results support that the urea cycle is being utilized by *M. aeruginosa* a way to store and remobilize N. We also observed that incorporated of ^{13}C from urea into *M. aeruginosa* which implies urea can be utilized as a N and C source and potentially be a more energetically favorable mechanisms of carbon acquisition. C from urea was also incorporated into microcystin precursors suggesting that the C and N from urea can be used to make microcystin. Urea was observed to be a valuable nutrient at high pH values commonly observed during dense blooms. We propose urea as a N and C source plays a significant role in the proliferation and success of *Microcystis*-dominated cyanobacterial blooms.

Finally, the fate of microcystins in the environment by the mechanism of biological degradation was investigated. Biodegradation is considered to be the most important route for removal of MCs in nature. We surveyed several metatranscriptomes generated from cyanobacterial bloom samples for evidence of this process by means of the only elucidated pathway to date, the *mlr* pathway. Our inability to detect the expression of genes or genes of similarity involved in this process indicated it was not the major mechanism for removal in the environment. We proposed the importance of other mechanisms from this data that have received little attention, however this work was also hypothesis generating. The hypothesis generated are important considerations and possible explanation for our observations.

1. Glutathione S-transferases and alkaline proteases are an environmentally relevant microcystin biodegradation pathway.
2. Microcystin biodegradation is due to the non-specific activity of peptidases in the nature.
3. Microcystin biodegradation is occurring in the sediments.
4. Microcystin biodegradation is not occurring.

Microcystin biodegradation remains an important area of research and future work should aim to determine the relevance of this process on an environmentally relevant scale. Characterization of new pathways in *mlr* negative bacteria and biodegradation field studies coupled with metatranscriptomics or peptidase activity assays over time would move this field forward.

In summary, we have utilized a series of molecular and culture-based techniques to ask pertinent and ecologically relevant questions about interactions within bloom communities to

address how individual factors influence *Microcystis*. The observations and conclusions described above have provided a broad array of new insights into the success of *Microcystis* on an ecological and physiological scale. This work generated several new hypotheses and presented novel ideas for future work to continue to tease apart the complex mechanisms that support the proliferation of cyanobacterial blooms around the world.

Vita

Lauren Elisabeth Krausfeldt was born in Arlington Heights, IL on February 10th, 1989. She attended Bartlett High School in the northwestern suburbs of Chicago and graduated in 2007. She received her Bachelor of Science degree from Elmhurst College in February of 2012 where she majored in Biology and a received a minor in Chemistry. While at Elmhurst College, Lauren discovered her love of research and microbiology while working for Dr. Tamara Marsh on a project investigating the effects of agricultural fertilizers on N-fixing bacteria. She also devoted her time to serving the community as a volunteer by leading and participating in several charitable and service organizations on campus. She started in the Department of Microbiology at the University of Tennessee in August of 2012 as a Master's student. Lauren joined the Wilhelm lab in January 2013 where she first started working on the biodegradation of microcystin and shortly after decided to switch to pursue her PhD. Since then, she has been involved in numerous projects involving harmful cyanobacterial blooms and has presented her work at several conferences in the United States and in China. She has also traveled internationally for field work; China to collect samples on Lake Taihu and Canada to participate in a research cruise on Lake Erie. Lauren is very grateful for her experiences at Elmhurst College and the University of Tennessee and is looking forward to what the future holds for her career.