Comparative Metagenomics of Toxic Freshwater Cyanobacteria Bloom Communities on Two Continents

Morgan M. Steffen[1], Zhou Li[2], T Chad Effler[1], Loren J. Hauser[1,2,3], Gregory L. Boyer[4], Steven W. Wilhelm[1]*

1 Department of Microbiology, The University of Tennessee, Knoxville, Tennessee, United States of America, 2 Graduate School of Genome Science & Technology, The University of Tennessee, Knoxville, Tennessee, United States of America, 3 Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States of America, 4 Department of Chemistry, College of Environmental Science and Forestry, State University of New York, Syracuse, New York, United States of America

Abstract

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

Introduction

The theory that microbial community structure dictates the function of that community has recently been called into question [1,2]. Previously, laboratory studies using enzyme assays and 16S rRNA gene phylogeny provided evidence for the importance of phylogenetic identity to community function [3,4]. A shift to whole community shotgun metagenomics has, however, allowed for a more comprehensive examination of microbial functional genes present in a wealth of natural environments. This technique circumvents the need for culture-based analysis and is more representative of natural community structure and functional potential. The trend emerging from this work indicates that the function of microbes within the environment is often more highly conserved than their phylogenetic/taxonomic identity [1,5].

One environment of particular concern is the freshwater systems that have in recent years been increasingly inundated by toxic cyanobacterial blooms. These blooms have been responsible for the deterioration of freshwater systems with increasing frequency and intensity and are commonly dominated by colonial cyanobacteria of the genus *Microcystis* [6–8]. Toxins produced by bloom-associated cyanobacteria have been shown to adversely affect health of animals and humans. The microcystins, cyclic secondary metabolites produced by members of the genera *Anabaena*, *Microcystis*, and *Planktothrix*, are hepatotoxins that have been associated with human liver and colorectal cancers [9–11]. Other secondary metabolites produced by *Microcystis* have been linked to phytoestrogenic effects in fish [12]. *Microcystis aeruginosa* is commonly dominant within freshwater blooms and has been shown to impact local ecology by disruption of the food web and induction of hypoxia in large lake systems [13,14].

Microcystis-dominated blooms have been observed in the western basin of Lake Erie annually since the 1990s. This freshwater system is a commonly studied model of toxic bloom development and persistence [6,15]. Lake Tai (hereafter referred to by its Chinese name *Taihu*, which translates to grand or great (*tai*) lake (*hu*)) is a shallow eutrophic lake that has experienced annual cyanobacterial blooms for the last three decades. The Taihu watershed supports upwards of 40 million people and receives inputs from both industrial and agricultural sources. During the last decade Taihu has experienced *Microcystis* bloom events on an unprecedented scale: in 2007 authorities shut down water intakes, resulting in a freshwater shortage for the 4 million residents of the city of Wuxi [16]. In contrast, Grand Lakes St. Marys (GLSM) is a

Table 1. Bloom sample metadata.

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

smaller (54.6 km^2) inland lake in Ohio that has recently experienced recurring highly toxic cyanobacteria blooms during summer months. While GLSM also experiences high concentrations of microcystin during bloom events, this lake is dominated by the filamentous cyanobacterium *Planktothrix* spp. (RML McKay, personal communication). To this end, GLSM provides a novel and contrasting study site (relative to Taihu and Lake Erie) whereby each lake experiences eutrophic conditions leading to the presence of the toxin microcystin, but in GLSM that toxin is produced by a cyanobacterium with a different physiological ecology. Examination of these contrasting freshwater cyanobacterial blooms models may provide insight on the environmental and genetic factors that influence bloom formation worldwide.

Despite our ability to examine *Microcystis* within bloom communities, questions remain regarding the environmental triggers that facilitate bloom formation and persistence. Previous research has shown that numerous abiotic factors, including nutrient input and air temperature, are important during cyanobacterial bloom formation, [17–19] although there is a current schism regarding which nutrient inputs most exacerbate bloom formation [20,21]. Recent work has demonstrated that concentrations of both nitrogen and phosphorus influence success of freshwater cyanobacteria in eutrophic environments [19,22–24]. Moreover, freshwater cyanobacteria thrive at temperatures above 25°C, and increasing mean temperatures associated with recent global climate change have been associated with the increasing frequency of these toxic bloom events [25–27].

Currently, we have little information regarding bloom communities as a whole, in terms of either phylogenetic composition or functional role. Indeed recent work suggests companion heterotrophs may be essential for cyanobacteria biomass growth in response to the addition of certain nutrient sources [28], highlighting the importance of resolving the ecology of this complex community. To gain insight into the role of bloom microbial community members and better understand differences between geographically distinct freshwater microbial communities, we have applied comparative metagenomics to the microbial populations of three toxic freshwater blooms. This approach has enabled us to examine not only the toxin-producing organisms, but also the companion heterotrophic community and their potential role in driving these blooms. In addition, we have used environmental genomic data to identify potential differences in genomic content in bloom-associated *Microcystis* spp. and a model lab isolate (*Microcystis aeruginosa* NIES 843) to further the aim of establishing the core genome of *Microcystis*. This information provides new insight on factors that may be critical to toxic freshwater blooms and which may prove crucial to future mitigation efforts.

Methods

Sample Collection

Lake Erie was sampled during an August, 2009 cruise aboard the CCGS *Limnos*. Water was taken from the surface in a 20 µm net tow at Environment Canada station 589 (Erie Harbor, PA; 42° 08′ N 80° 07′ W). Biomass was stored at −80°C until extraction. Taihu was sampled during a large surface bloom in May, 2009 near the Taihu Laboratory for Lake Ecosystem Research (TLLR, 31°27′ N, 120°12′ E) using 20 µm Nytex^{TM}. Samples were stored at −20°C until extraction. The GLSM (40°31′ N, 84°24′ W) bloom sample was collected using a 20 µm plankton net during the July 2010 bloom event and stored at −20°C until extraction. The 20 µm size is a commonly used technique in limnology and allowed for collection of *Microcystis*, other large cyanobacteria and the microbial community associated with these cells. Nutrient and toxin measurements at each location were made as previously described [17,19,29]. Metadata for each site are listed in Table 1.

DNA Extraction, Sequencing, Assembly, and Annotation

Genomic DNA was extracted from all samples using the MoBio PowerWater® DNA Isolation Kit (MoBio Laboratories, Inc., CA, USA). The GLSM sample required an additional wash step, using sterile CT medium [30,31] to remove organic impurities that interfered with nucleic acid purity. The quality and quantity of DNA was measured by spectrophotometric quantification in a NanoDrop 1000 (Thermo Fisher Scientific, Inc., DE, USA) and agarose gel electrophoresis. Extracted DNA was stored at −80°C prior to sequencing.

A total of 500 ng of genomic DNA per sample was used for library preparationand sequenced according using a Roche 454 GS FLX platform with Titanium chemistry (Roche/454-Life Sciences, Branford, CT, USA) at the UT/ORNL Joint Institute of Biological Sciences. We generated a total of 533 Mbp and 1.36 million reads with an average read length of 394 bp (Table S1). The MG-RAST v.3.0 online server quality control pipeline was used to remove reads of poor quality and short length before annotation and analysis of metagenomics data [32]. Pipeline parameters were kept at the default settings. Sequences were dereplicated and filtered by length, removing sequences more than 2.0 standard deviations from the mean sequences length. Sequences with five or more non ATCG characters were removed. Recruitment of all metagenome reads post-QC processing to the *Microcystis aeruginosa* NIES 843 genome as was performed using the SeqMan NGen® (DNASTAR, Inc., WI, USA) software package using manufacturer's suggested default parameters established for 454 datasets. Reannotation of the NIES 843 reference genome and annotation of the environmental *Microcystis* genes was performed using Prodigal [33].

Metagenome reads were subjected to a BLASTx comparison against the NCBI nonredundant database. Community taxonomy and function were analyzed in the MEGAN v.4.61.6 analysis toolset [34,35]. To confirm reproducibility of taxonomic assignments, reads were also uploaded to the MG-RAST server for taxonomic and functional potential analysis using a minimum *e*-value cutoff of 1e-5, minimum percent identity of 65% and a minimum alignment length of 50 amino acids. COGs (Clusters of Orthologous Groups) were used to assign function [36]. For MG-
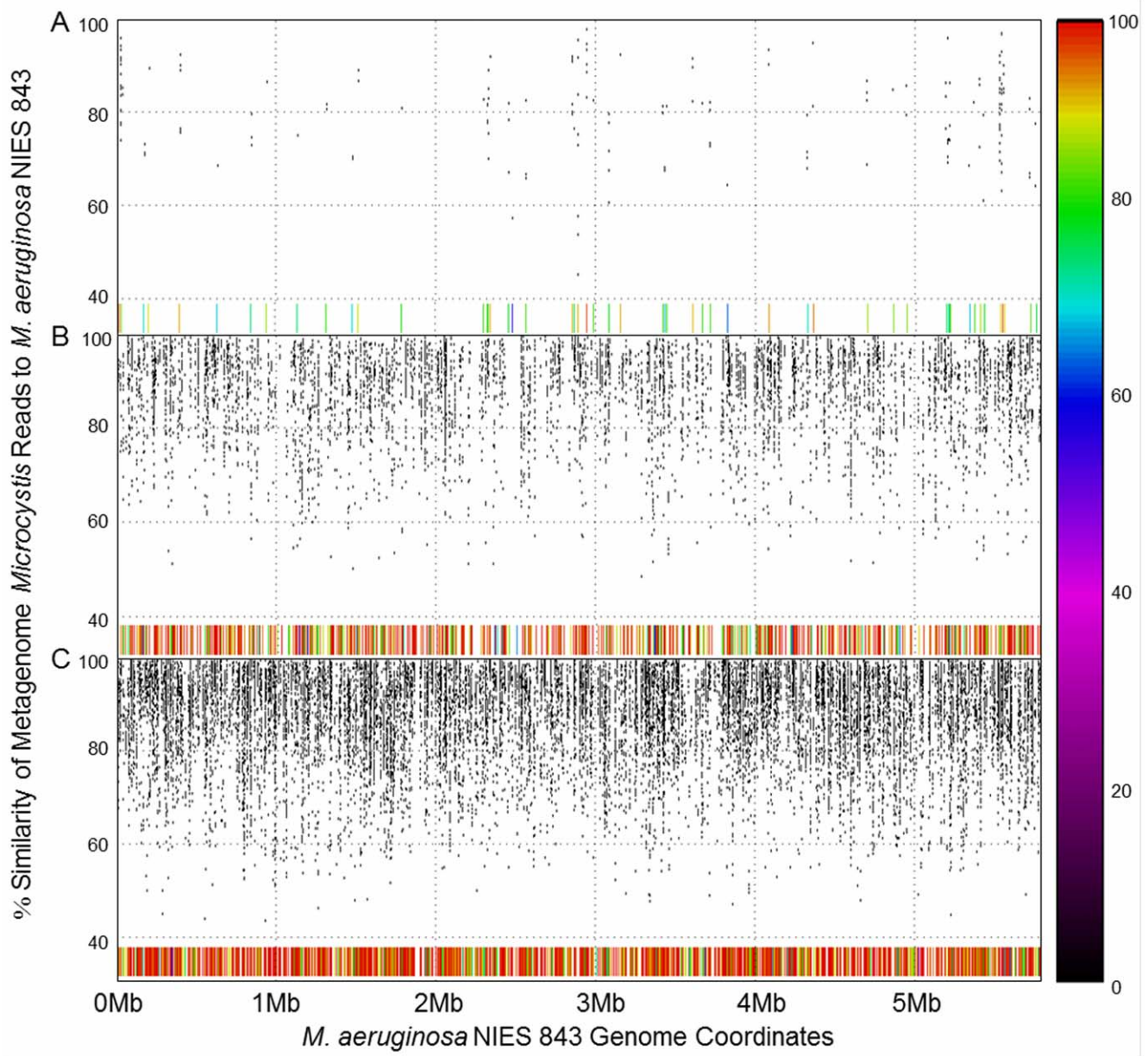
Figure 1. Recruitment plots of metagenomes to Microcystis aeruginosa NIES 843. Recruitment plots of environmental *Microcystis* reads from each metagenome to the *M. aeruginosa* NIES 843 genome. A) Recruitment of GLSM *Microcystis* reads to NIES 843. B) Recruitment of Taihu *Microcystis* reads to NIES 843. C) Recruitment of Erie *Microcystis* reads to NIES 843. Position along the x-axis indicates position along the genome of *M. aeruginosa* NIES 843 from zero to 5.8 Mbp and position along the y-axis indicates percent similarity of recruited sequences. The bar along the bottom is a secondary indicator of percent similarity to the reference genome. See Table S2 for locations of potential MI's within each recruitment. doi:10.1371/journal.pone.0044002.g001

RAST comparative analyses, the GenBank database was used for phylogenetic analysis and RefSeq was used to annotate functional genes using the parameters described above [37]. Phylogenetic comparisons were made using percent abundance within each library post quality control analysis. COG and functional gene comparative analyses were performed using reads normalized to the library with the greatest total number of hits to the COG database (Taihu). Sequences generated in this study are available on the MG-RAST online server (http://metagenomics.anl.gov/) under the identification numbers 4467029.3 (Erie), 4467058.3 (Taihu), and 4467059.3 (GLSM).

Microcystis Genome Recruitment

Metagenomes were annotated using the RefSeq database (MG-RAST) with an *e*-value minimum of 1e−5, a 65% identity cutoff, and a minimum alignment length of 50. All reads annotated as *Microcystis* from each metagenome were recruited to the genome of *Microcystis aeruginosa* NIES 843 using MUMmer 3.22 [38]. The PROmer algorithm was used for alignment. Potential metagenomic islands were initially identified manually in the Lake Erie and Taihu recruitment plots as regions of 10 kb or longer with less than average recruitment as established in Rodriguez-Valera et al [39]. These regions were defined as MIs if they had less than 25% average similarity with the reference and coverage of the reference
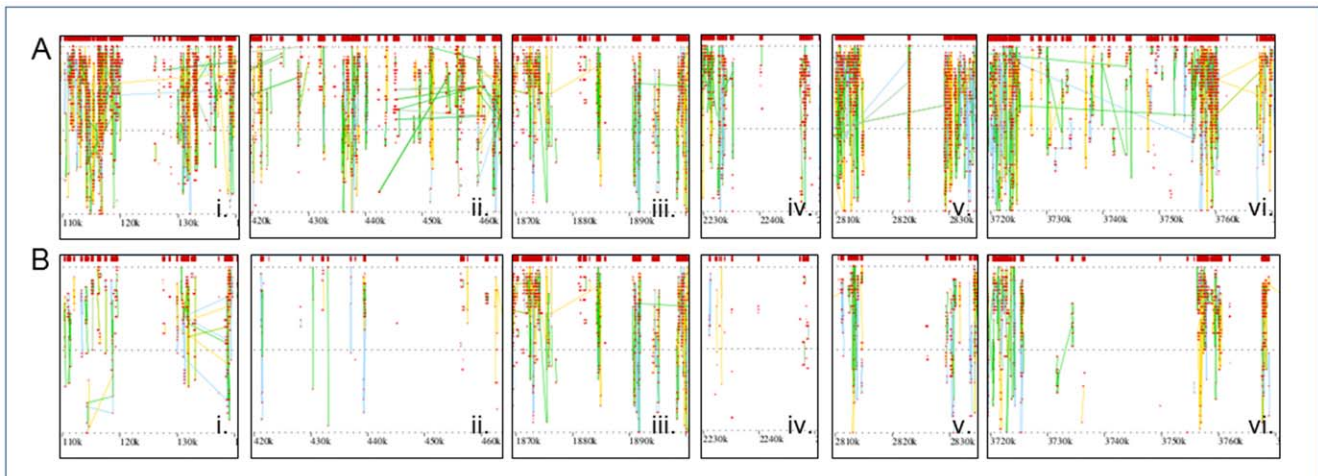
Figure 2. Metagenomic islands (MIs) of bloom-associated *Microcystis*. Metagenomic islands (MIs) identified in the Lake Erie and Taihu *Microcystis* recruitments. Six regions in Taihu and three in Erie qualified as MIs. MIs were defined as regions of ≥10 kb in length that have an average similarity of less than 25% to the reference genome and less than 25% total coverage of the corresponding base pairs of the reference sequence. A) View of MIs of Lake Erie *Microcystis* recruitment. Boxes ii, iii, and vi were found to be MIs in Taihu, but did not satisfy our criteria in Lake Erie. E) View of MI regions in Taihu *Microcystis* recruitment.
doi:10.1371/journal.pone.0044002.g002

Results and Discussion

The *Microcystis* Metagenome

It is widely known that not all members of the genus *Microcystis* have the genetic capability to produce microcystin – only 10–50% of cells in natural systems have these genes [17,18]. It is thus likely that similar variability in content may be reflected in other genomic regions, a factor that is important to know when picking model organisms for lab study. Using SeqMan NGen® (DNAStar, Madison, WI), an initial attempt was made to generate whole genome assemblies of environmental *Microcystis* from each metagenome for genome-wide comparison. This effort was abandoned early however, as we realized there were issues with the potential recruitment of non-*Microcystis* reads to the sequenced lab isolate used as a scaffold. Instead, recruitment plots were generated in MUMmer 3.22 to identify potential genomic differences between bloom-associated *Microcystis* spp. and the sequenced lab isolate [40]. Sequences from each metagenome assigned a best hit identity of *Microcystis* were aligned to the *Microcystis aeruginosa* NIES 843 genome [41]. Recruitment of GLSM sequences was limited, with only 7% total coverage of the reference genome. Out of the possible 1327 individual *Microcystis* genes identified in the GLSM metagenome, 64% of genes had at least one repeat. (Figure 1A). The GLSM bloom has been subsequently characterized as a predominately *Planktothrix* bloom by microscopic and molecular analysis (Bullerjahn and McKay, unpublished), providing a foundation for our observation of limited recruitment to the NIES 843 genome. This characterization provides a valid explanation for the limited recruitment of *Microcystis* reads from the GLSM metagenome. Taihu sequences had greater recruitment to the reference genome compared to the GLSM data set, with 51% of the reference sequence covered (Figure 1B). 79% of the possible 4256 recruited sequences from Taihu were present at least twice in the metagenome data set. A

robust recruitment, covering of 79% of the NIES 843 reference genome, resulted from alignment of Lake Erie sequences, with 92% of the 5539 *Microcystis* sequences recruited to the NIES 843 scaffold repeated at least once in the Erie metagenome (Figure 1C). This difference in recruitment between Erie and Taihu is likely due to differences in abundance of *Microcystis* spp. within our microbial samples. While the method of sample collection was the same across each bloom, collection took place at different points during bloom development at each site, which may have contributed to this difference in abundance of *Microcystis* spp.

Six conserved metagenomic islands were identified in the Taihu and three in the Lake Erie recruitments (Figure 2A, 2B). All six regions initially identified as potential MIs are shown in Figure 2. Despite having only three regions that qualified as MIs, it is worth mentioning that the Lake Erie dataset does demonstrate reduced similarity and coverage in all of these regions. Metagenomic islands (MI's) have been defined previously as areas within a genome that are underrepresented in metagenomic datasets. It is thought that MI's represent regions of a genome that are unique between closely related strains, or components of the species pangenome [39]. For this analysis, these were identified as regions of 10 kb or larger that had relatively poor similarity and coverage (<25%) to the reference, consistent with similar analyses performed in marine environments [39]. For a detailed breakdown of the coverage of each MI, see Table S2. A number of transposases and hypothetical proteins were consistently identified as missing across all six islands: a complete list of genes within the six metagenomic islands is provided in Table S3. Within the second island, two genes involved in restriction modification systems failed to recruit. Interestingly, the *5S, 16S,* and *23S* rRNA genes are located in the third island, although they recruited fully to the reference in both Taihu and Lake Erie recruitments. NIES 843 CRISPR-associated (Clustered Regularly Interspaced Short Palindromic Repeats) protein genes are located in the fifth metagenomic island, one of which failed to recruit from the Lake Erie metagenome. In the sixth and largest island (~42 kb), a number of genes involved in cell surface recognition systems failed to recruit. These genes encoded for two cell surface antigen
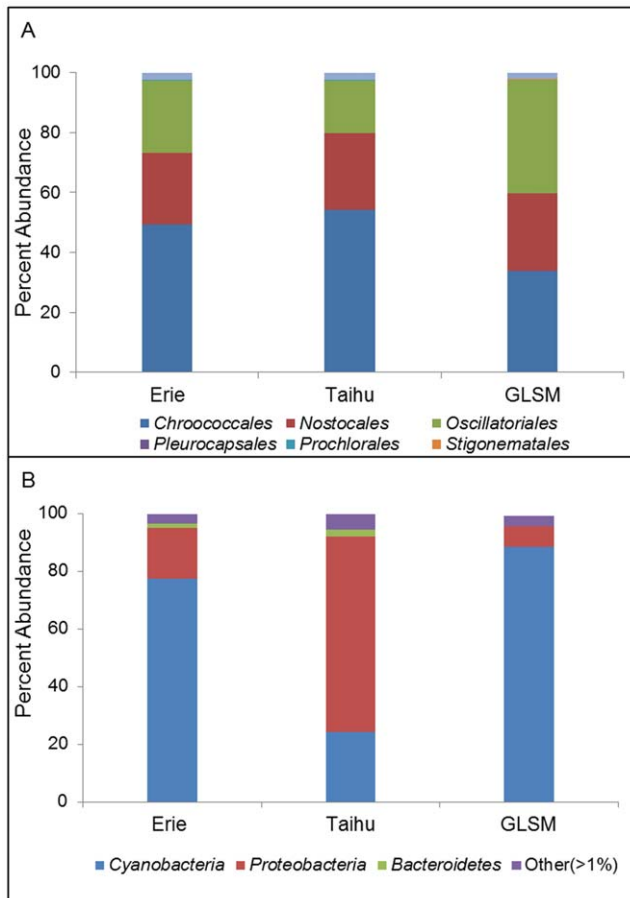
Figure 3. Identity of bloom-associated microbial community members. A) Percent abundance of orders belonging to the *Cyanobacteria* phylum. Abundance is a percentage of reads annotated as *Cyanobacteria* from each metagenome. B) Distribution of bacterial phyla in each metagenome. Abundance is a percentage of total bacterial reads in each metagenome. Phyla with less than 1% total abundance were grouped in the "Other" category.
doi:10.1371/journal.pone.0044002.g003

Table 2. Percentage abundance of hits assigned to genera in the orders *Chroococcales* and *Oscillatoriales*.

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

Sequences most closely identified as exclusively marine lineages (*Chrocosphera*, *Trichodesmium*) are denoted as "-like" and potentially are from related yet currently unclassified members of the community.
doi:10.1371/journal.pone.0044002.t002

Structure of the Co-occurring Community during *Microcystis* Blooms

Phylogenetic composition of each bloom sample was compared at multiple taxonomic levels. A relatively conserved phylogenetic structure was observed in the cyanobacterial populations for all of the lake samples (Figure 3A). The orders *Chroococcales*, *Nostocales*, and *Oscillatoriales* were most abundant in all three systems. The oscillatorians made up a slightly higher percentage of the total cyanobacterial population within GLSM (38%) relative to Erie (24%) or Taihu (17%). At the genus level, the most common best hit matches within the *Chroococcales* were the nitrogen-fixing *Cyanothece*, as well as *Synechococcus*, with *Microcystis* spp. making up less than three percent of the total population for each bloom. Despite their relatively low abundance, *Microcystis* spp. remain important as they were the dominant toxin producers in Lake Erie and Taihu (Table 2) [43,44]. Genetic elements classifying to *Oscillatoria* were largely comprised of *Lyngbya* spp. and a group whose closest relatives are members of *Trichodesmium* at the genus level. Because *Trichodesmium* spp. are marine organisms, it is likely that a significant population of previously unclassified freshwater nitrogen fixers were present at high abundance in all three blooms. This finding is indicative of the need for more freshwater cyanobacterial genome sequences to more comprehensively study bloom events and other phenomena. As with members of *Chroococcales*, the dominant toxin producers within the oscillatorians (*Planktothrix*) are present in lower numbers than other genera (Table 2). At the genus level, phylogenetic resolution cannot be completely achieved, as current annotations reflect only what is currently available in the databases used for annotation, and are reported as the best hit. As databases continue to expand and more freshwater genomes become available, best hit annotations may change and give a more accurate representation of the communities. We include this information with the knowledge that as updates to databases and annotation algorithms improve, these assignments may change.

receptor proteins, a two component response regulator, and a sensor protein.

Taken together, the genes missing from each metagenomic island represent genes that may all be involved in phage recognition, consistent with metagenomic islands identified in marine and hypersaline environments [39,42]. Although these gaps in recruitment may be an artifact of sequencing depth, this is unlikely, as there was robust recruitment to a vast majority to the NIES 843 genome by the Erie reads and islands were conserved between the datasets. The three MIs present in Taihu that did not meet the criteria in Erie may be examples of limited recruitment due to differences in population makeup between Erie and Taihu. Alternatively, these regions may differ from the reference, but are not yet divergent enough to qualify as MIs in Erie. The future establishment of a *Microcystis* spp. core genome (as more sequences become available) will be important in the verification of these differences. One explanation for the three islands conserved in both recruitments is that they represent regions of bloom-associated *Microcystis* genomes that are too divergent from the reference to recruit.
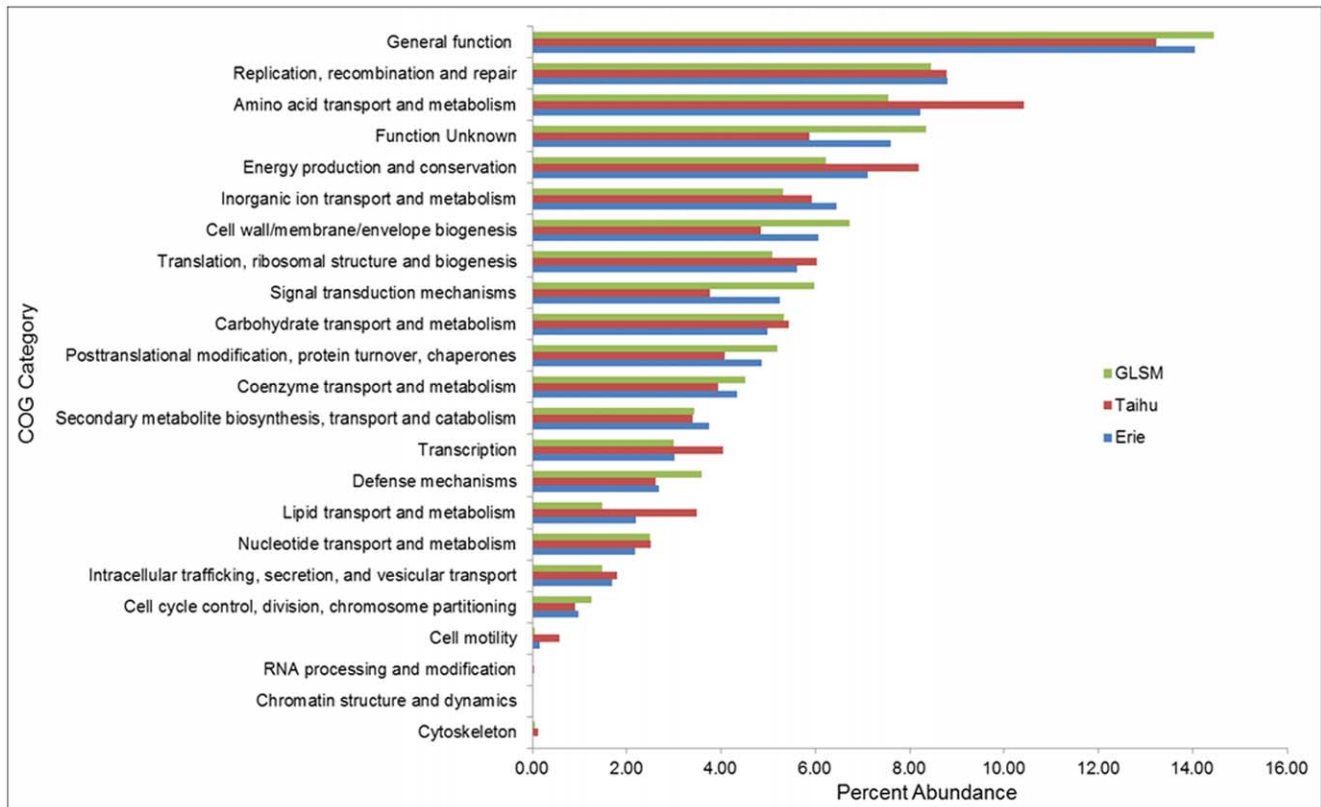
Figure 4. Abundance of metagenome reads assigned to each COG category. Metagenomes were annotated using the COG database. Reads were assigned to standard COG functional categories based on this annotation. Abundance is a percentage of total hits to the COG database. doi:10.1371/journal.pone.0044002.g004

At the phylum level, microbial community composition varied across the different lakes (Figure 3B). *Cyanobacteria* were the most abundant members of the Erie and GLSM bloom samples (77% and 88%, respectively). The *Proteobacteria* were the second most abundant group in the North American lakes, followed by *Bacteroidetes* and several rare phyla (<1%). In Taihu, the *Proteobacteria* were the most abundant, despite higher measured levels of *chl* a (Table 1). Proteobacteria comprise 68% of the total microbial population, followed by *Cyanobacteria* at 25%, which is consistent with previous investigations of heterotrophic populations in the Taihu system [45,46]. Further repetition of these datasets across multiple time points would provide more comprehensive insight into the bloom-associated community. As toxin-levels and the abundance of toxin-producers change throughout the course of a bloom, so may the associated bacterial community, as has been previously demonstrated in Taihu [47]. While possibly an artifact of sequencing coverage (a re-occurring theme in all metagenomics surveys), age of the bloom at the time of sample collection or of the sample collection techniques, these differences between Taihu and the North American lakes may be demonstrative of phylogenetic variability of *Microcystis*-associated bacteria between geographically distinct systems.

Functional Potential of Co-occurring Heterotrophic Bacteria

Genes from each site were assigned to broad functional categories (Clusters of Orthologous Groups, COGs) and the relative abundance of hits to each category was compared across the three metagenomes (Figure 4). The number of genes within several categories was closely conserved across all sites. This includes genes involved in nutrient metabolism, coenzyme transport and metabolism, secondary metabolite metabolism and transport, as well as replication, recombination, and repair.

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

To confirm the observed conservation of functional potential, the normalized number of hits to each COG category for each metagenome was plotted against the other (Figure 5). All points that fell within the calculated 95% confidence interval were assumed to be conserved in frequency of occurrence between the two blooms being compared. The only outlier in the comparison between the North American lakes was the category containing genes of unknown function (Figure 5A). As expected, the farthest outlier between Taihu and both Erie and GLSM was the Amino Acid Metabolism and Transport category (Figure 5B–C). Other conserved outliers included COGs for Signal Transduction Mechanisms, Energy Production & Conservation and Cell Wall/Membrane/Envelope Biogenesis. Taken with the original observational comparison of relative abundance of each COG category, these data suggest a conserved functional potential of the bloom-associated community, despite variable phylogenetic makeup. This
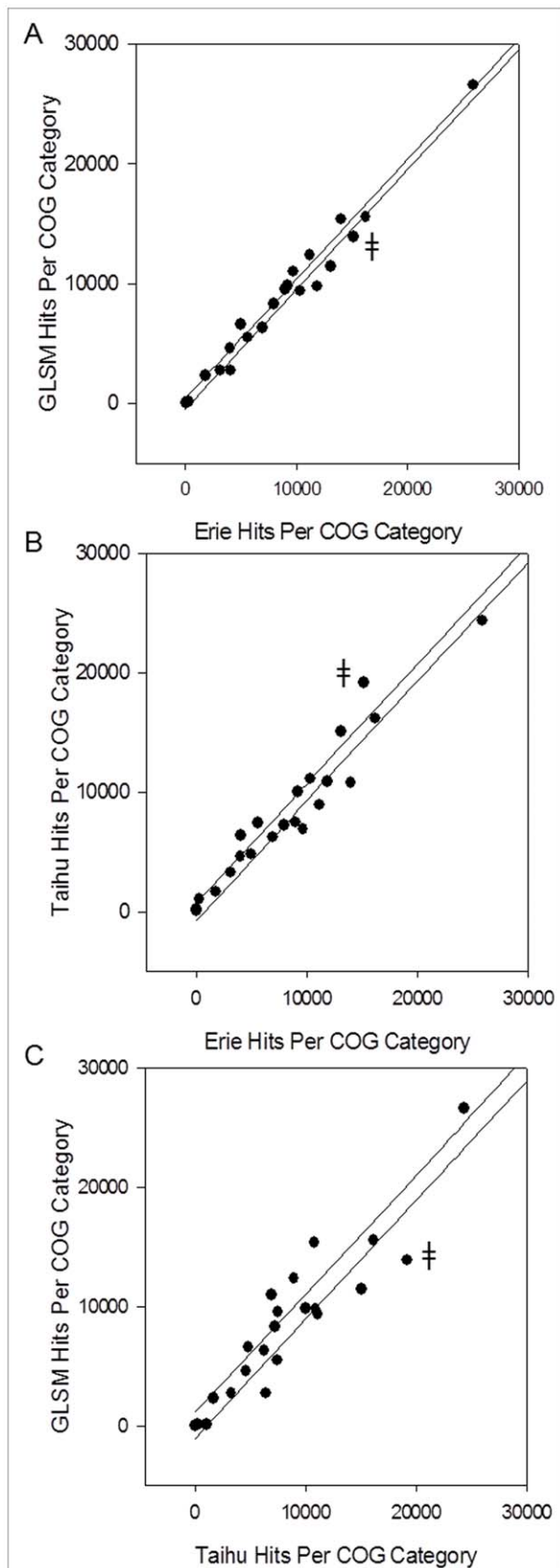
Figure 5. Comparison of number of hits to COG categories between bloom metagenomes. Normalized number of hits to each COG category from each metagenome plotted against the other. Data

were normalized to the metagenome with the greatest number of hits to the COG database (Taihu). A) COG distributions of Erie and GLSM are compared. Data points largely fall within or close to the 95% confidence interval (CI) which is indicated by the black lines, indicating similarity. B) COG distributions of Erie and Taihu are compared. C) COG distributions of Taihu and GLSM are compared. Outliers in (B) and (C) are largely conserved, indicating closer conservation of COG distribution between the North American lakes to each other than Taihu. The largest outlier in the Taihu comparisons is denoted by the (‡) symbol in each plot. This outlier corresponds to the Amino Acid Metabolism and Transport COG category.
doi:10.1371/journal.pone.0044002.g005

provides support to the growing body of evidence suggesting that variable phylogenetic makeup may not be reflected in the function of the environmental microbial community members [1,5].

Due to their potential role in the regulation of bloom formation in Taihu, we specifically took a deeper look at pathways associated with nitrogen cycling [48]. Functional genes involved in nitrogen assimilation were identified and annotated at the phylum level in each dataset to highlight the genomic contribution of bloom microbes to the incorporation of nitrogen in the community. Genes examined were those involved in nitrogen fixation (*nifD*, *nifH*, *nifK*), urea metabolism (*ureA–G*), nitrate and nitrite reduction (*nar* and *nir*), and glutamine synthesis (*glnA*). A stark contrast in the contribution of genes involved in nitrogen fixation and urea metabolism emerged between Taihu and the North American lakes (Table 3). *Cyanobacteria* contribute the majority of the genes involved in these processes in Erie (78% and 58%) and GLSM (96% and 95%). In contrast, *Proteobacteria* contribute an overwhelming majority of these functional genes (86% and 74%) within the Taihu bloom community. This highlights the potential for assimilation of nitrogen by different organisms in the bloom-associated community in Taihu relative to Erie and GLSM. It should be noted that amino acids may also serve as potential nitrogen sources, and the overrepresentation of those genes involved in general amino acid metabolism in Taihu may reflect an important role for heterotrophs in nitrogen metabolism in this system (Figure 4). The normalized abundance of genes involved in each type of nitrogen assimilation remains relatively constant, despite their taxonomic identity (Table 3). These data also highlight the somewhat divergent phylogenetic makeup of each community. The overrepresentation of nitrogen assimilation genes associated with *Proteobacteria* in Taihu is largely due to numerical dominance of these organisms. To demonstrate this, the commonly used housekeeping gene *rpoB* can be used as a crude proxy for organism abundance, and has been included to allow for normalization of nitrogen assimilation genes within phyla (Table 3) [49,50]. While normalization changes the phylogenetic identity of the majority of the N assimilation genes, it again reiterates the possibility of phylogenetically distinct organisms having the potential to perform a similar functional role in bloom-associated microbial communities.

The data further illustrate an important caveat and area for future investigation: similar biological processes (*e.g.*, those discussed above) within different organisms are likely to result in different biochemical outcomes as these similar processes are independently coupled to unique subsequent process within members of the population. To this end our data need to be considered with a reminder of the importance of looking beyond individual pathways (or even parts of pathways) when trying to address biogeochemistry and the functional level.

This pattern in nitrogen assimilation gene distribution underscores the potential role of heterotrophic bacteria within toxic cyanobacterial bloom communities in some freshwater lakes. The role of heterotrophs in cyanobacterial metabolism has been

Table 3. The abundance of microbial nitrogen assimilation genes within metagenome.

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

Genes from families associated with nitrogen fixation (*nifD, nif H, nifK*), urea assimilation (*ureA–G*), nitrate reduction (*nar, nir*) and ammonium utilization (*glnA*) were identified and enumerated within each metagenome using the COG database and phylogenetically classified using RefSeq. Results represent the total number of hits to functional categories (COG database) within each library and are normalized for library size.
doi:10.1371/journal.pone.0044002.t003

previously observed in both marine and freshwater systems. Morris et al. (2008) [51] identified heterotrophic activity as the source of *Prochlorococcus* resistance to hydrogen peroxide in the oceans, while the important microcystin-producer *Planktothrix* has recently been found to successfully use the compound glyphosate, a herbicide known to be relatively resistant to biodegradation, as a nutrient source in the presence of "helper" heterotrophs [28]. Based on this metagenomic data, it appears that heterotrophs in Taihu may play an important role in nitrogen assimilation and transformation for the larger bloom microbial community.

Conclusions

This work highlights the utility of metagenomics as a tool for exploration of microbial communities. Here, we provide microbial snapshots of three separate toxic cyanobacterial blooms. Despite being single samples, these metagenomes provide a unique snapshot of the microbial community associated with toxic cyanobacterial blooms. This initial characterization is an important foundation for further study of such communities. We have attempted to include metadata that complies with the Genomic Standards Consortium, with the intention that this initial characterization of three freshwater blooms can be reexamined for future statistical comparison as the repository of freshwater sequences builds [52]. While we focus largely on the bacterial community associated with bloom biomass, viral sequences were also present in all data sets. Notably, sequences of the *Microcystis* phage Ma-LMM01 were detected at all three lakes. This is especially worth noting due to the importance of phage in bloom dynamics and termination [53]. Other interesting findings include the presence of the *mlrC* gene in both Taihu and Erie. This gene is involved in microbial degradation of microcystin, and its presence warrants further inquiry into the presence of potential important microcystin degraders in these lakes [54].

In the current study we have used metagenomics to describe the phylogenetic makeup and functional potential of three geographically distinct cyanobacterial blooms at single time points during each bloom. This comparative approach has confirmed an increasing trend in microbial ecology: variable bacterial phylogenetic makeup is not mirrored in the relatively constant functional potential of the bacterial community in the environment [1]. This finding may reflect a need to amend current methods of describing microbial community dynamics, as the taxonomic/phylogenetic

identity of community members may no longer be sufficient. Within our observations key functional genes, such as those involved in nitrogen assimilation, appear to be more informative than standard *16S* rDNA gene analysis and demonstrate that within 2 similar biological events (blooms in Lake Erie and Taihu) the analogous processes are likely carried out by different members of the community. With this approach, we were able to identify potentially divergent pathways of assimilated nitrogen through the microbial communities of three different blooms. The genomic contribution of heterotrophic bacteria to nitrogen assimilation in Taihu represents a potentially critical contribution of heterotrophic bacteria in driving toxic freshwater blooms.

Supporting Information

Table S1 Information regarding the sequence data sets used in this study.
(DOCX)

Table S2 Coverage of *Microcystis aeruginosa* NIES 843 metagenomic islands (MIs) identified in the Erie and Taihu datasets.
(DOCX)

Table S3 All genes present in each metagenomic island (MI) of M. aeruginosa NIES 843. Genes with and without coverage are listed.
(DOCX)

Author Contributions

Conceived and designed the experiments: MMS ZL SWW. Performed the experiments: MMS ZL. Analyzed the data: MMS ZL TCE LJH GLB SWW. Contributed reagents/materials/analysis tools: LJH GLB SWW. Wrote the paper: MMS ZL LJH GLB SWW.

References

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

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

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

4. Lehman RM, O'Connell SP (2002) Comparison of extracellular enzyme activities and community composition of attached and free-living bacteria in porous medium columns. Appl Environ Microbiol 68: 1569–1575.

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

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

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

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

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

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

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

12. Rogers ED, Henry TB, Twiner MJ, Gouffon JS, McPherson JT, et al. (2011) Global gene expression profiling in larval zebrafish xxposed to microcystin-LR and Microcystis reveals endocrine disrupting fffects of cyanobacteria. Environ Sci Technol 45: 1962–1969.

13. Charlton MN, Milne JE, Booth WG, Chiocchio F (1993) Lake Erie offshore in 1990 - restoration and resilience in the central basin. J Great Lakes Res 19: 291–309.

14. Conroy JD, Kane DD, Dolan DM, Edwards WJ, Charlton MN, et al. (2005) Temporal trends in Lake Erie plankton biomass: Roles of external phosphorus loading and dreissenid mussels. J Great Lakes Res 31: 89–110.

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

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

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

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

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

20. Schindler DW, Hecky RE, Findlay DL, Stainton MP, Parker BR, et al. (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. Proc Natl Acad Sci U S A. 105: 11254–11258.

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

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

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

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

25. Ye C, Shen Z, Zhang T, Fan M, Lei Y, et al. (2011) Long-term joint effect of nutrients and temperature increase on algal growth in Lake Taihu, China. J Environ Sci 23: 222–227.

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

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

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

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

30. Ichimura T. (1971) Sexual cell division and conjugation-papilla formation in sexual reproduction of Closterium strigosum. In: Nishizawa, K ed. Proceedings of the 7th International Seaweed Symposium. 208–214. University of Tokyo Press, Tokyo.

31. Watanabe MM, M Kawachi, M Hiroki, F Hkasai (2000) NIES collection list of strains: microalgal and protozoa. National Institute for Environmental Studies. Tsukuba, Japan, 159 pp.

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

33. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, et al. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11: 119.

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

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

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

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

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

39. Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L, Thingstad TF, et al. (2009) Explaining microbial population genomics through phage predation. Nature Rev Microbiol 7: 828–836.

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

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

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

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

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

45. Li N, Zhang L, Li F, Wang Y, Zhu Y, et al. (2011) Metagenome of microorganisms associated with the toxic cyanobacteria Microcystis aeruginosa analyzed using the 454 sequencing platform. Chinese J Oceanol Limnol 29: 505–513.

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

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

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

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

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

51. Morris JJ, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER (2008) Facilitation of robust growth of Prochlorococcus colonies and dilute liquid cultures by "helper" heterotrophic bacteria. Appl Environ Microbiol 74: 4530–4534.

52. Field D, Garrity G, Gray T, Morrison N, Selengut J, et al. (2008) The minimum information about a genome sequence (MIGS) specification. Nature Biotechnol 26: 541–547.

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

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