University of Tennessee, Knoxville
Trace: Tennessee Research and Creative Exchange

Microbiology Publications and Other Works | Microbiology

January 2009

Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells

C.J. Allender

G.R. LeCleir

J.M. Rinta-Kanto

R.L. Small

M.F. Satchwell

See next page for additional authors

Follow this and additional works at: http://trace.tennessee.edu/utk_micrpubs

Part of the Environmental Microbiology and Microbial Ecology Commons, Fresh Water Studies Commons, and the Oceanography Commons

Recommended Citation

Allender, C.J.; LeCleir, G.R.; Rinta-Kanto, J.M.; Small, R.L.; Satchwell, M.F.; Boyer, G.L.; and Wilhelm, Steven, "Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells" (2009). *Microbiology Publications and Other Works.*
http://trace.tennessee.edu/utk_micrpubs/1

This Article is brought to you for free and open access by the Microbiology at Trace: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Microbiology Publications and Other Works by an authorized administrator of Trace: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

Authors
C.J. Allender, G.R. LeCleir, J.M. Rinta-Kanto, R.L. Small, M.F. Satchwell, G.L. Boyer, and Steven Wilhelm

This article is available at Trace: Tennessee Research and Creative Exchange: http://trace.tennessee.edu/utk_micrpubs/1

AEM Accepts, published online ahead of print on 10 April 2009
Appl. Environ. Microbiol. doi:10.1128/AEM.02448-08
Copyright © 2009, American Society for Microbiology and/or the Listed Authors/Institutions. All Rights Reserved.

AEM Accepts published online ahead of print

For submission to Applied and Environmental Microbiology – Microbial Ecology

Identifying the source of unknown microcystin genes and predicting microcystin variants by comparing genes within uncultured cyanobacterial cells

Christopher J. Allender[1], Gary R. LeCleir[1], Johanna M. Rinta-Kanto[2], Randall L. Small[3], Michael F. Satchwell[4], Gregory L. Boyer[4], and Steven W. Wilhelm[1] *

[1] Department of Microbiology, University of Tennessee, Knoxville, TN 37996

[2] Depatment of Marine Sciences, University of Georgia, Athens, GA 30602

[3] Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996

[4] Department of Biochemistry, State University of New York College of Environmental Science and Forestry, Syracuse, NY, 13210

* Corresponding author: Steven W. Wilhelm, Department of Microbiology, The University of Tennessee, Knoxville, TN 37996-0845, Office: 865-974-0665, FAX: 865-974-4007, Email: Wilhelm@utk.edu

Running title: Allender et al.: Comparing microcystin gene sequences within individual cells

AEM Accepts published online ahead of print

23 **ABSTRACT**

24 While multiple phylogenetic markers have been used in the culture independent study of
25 microcystin producing cyanobacteria, in only a few instances have multiple markers been studied
26 within individual cells, and in all cases these studies have been conducted with cultured isolates.
27 Here, we isolate and evaluate large DNA fragments (> 6 kb) encompassing two genes involved
28 in microcystin biosynthesis (*mcyA2* and *mcyB1)* and use them to identify the source of gene
29 fragments found in water samples. Further investigation of these gene loci from individual
30 cyanobacterial cells allowed for improved analysis of the genetic diversity within microcystin
31 producers as well as a method to predict microcystin variants for individuals. These efforts have
32 also identified the source of the novel *mcyA* genotype previously termed "*Microcystis*-like" that
33 is pervasive in the Laurentian Great Lakes and predict the microcystin variant(s) that it produces.

AEM Accepts published online ahead of print

34 Microcystin-producing cyanobacteria are common nuisance organisms in harmful algal
35 blooms of freshwaters around the world (5). This genetically diverse group (based on 16S rRNA,
36 *mcyA, mcyD,* and *mcyE* gene sequences (7, 11, 16, 17, 23)) ranges in morphology from
37 unicellular and colonial cocci to large filamentous strands. Many species can produce a variety
38 of secondary metabolites (i.e., *Microcystis*) that can act as hepatatoxins upon ingestion by
39 animals (*e.g.,* variants of microcystin) (5, 35). Microcystin production reduces the water quality
40 in reservoirs used by human populations and fishery resources, and production of these toxins by
41 this group of cyanobacteria makes them important organisms for continued observation and
42 study (5, 35, 38). Much effort has been expended over the past fifteen years to characterize the
43 genomic and structural components of the microcystin (*mcy*) synthetase operon responsible for
44 the production of microcystins. Several complete DNA sequences of the *mcy* synthetase operon
45 are currently available in GenBank (4, 12, 31, 33).
46 Although the mechanisms of microcystin production are now better understood, recent
47 analyses of *mcyA* gene fragments from the Lakes Erie and Ontario indicated a microcystin toxin
48 producer of unknown phylogeny (8, 30). This discrepancy suggested a need for improved
49 molecular characterization of naturally occurring microcystin producers, which spurred our
50 research to identify the source of several unusual *mcyA* fragments from the cyanobacterial
51 community (8, 30). It was apparent from initial sequence data that these *mcyA* gene fragments,
52 termed "*Microcysti*s-like," were highly similar to those from *Microcystis* spp. (a colonial or
53 unicellular cocci). However, they contained a six nucleotide insert consistent with *mcyA* genes
54 from filamentous cyanobacteria (e.g., *Anabaena*, *Nostoc*, and *Planktothrix*) (30). These
55 preliminary findings suggested that these unusual *mcyA* fragments either came from: 1) a novel
56 species or strain, 2) an ancestral *Microcystis*, 3) the highly unlikely hybridization of colonial

AEM Accepts published online ahead of print

57 cocci and filamentous cyanobacteria, or 4) a chimera of cocci and filamentous PCR products. To
58 identify the source of these *mcyA* gene fragments from uncultured cyanobacteria, we used
59 culture-independent methods to amplify and isolate long regions of the *mcy* synthetase operon
60 for the simultaneous analysis of two genes, *mcyA* and *mcyB*, in one individual from a population.
61 This approach ensures that both genes are contained on the same DNA molecule, thus allowing
62 for more continuous sequence information to use in comparative phylogenetic analyses than
63 previously described. We also envisioned that this *mcy* gene combination would provide an
64 improved diagnostic tool for determining the genetic potential of naturally occurring
65 cyanobacteria to produce specific microcystin variants by comparing the phylogenetic marker in
66 *mcyA* to the predictor of amino acid incorporation (via an adenylation domain) in *mcyB1*.

67 **Sample collection and DNA extraction.** Water samples were collected from three
68 locations. Surface water samples (<1 m) collected from Lakes Erie (Aug. 2007) and Ontario
69 (Aug. 2002 and 2006) were processed onboard the C.C.G.S. *Limnos* and water samples collected
70 from Waterville Reservoir (North Carolina, USA) were chilled on ice and brought back to the lab
71 for processing (Fig. 1). Cells for DNA extraction were collected on 0.2-µm or 2.0-µm
72 polycarbonate membrane filters (Millipore) and stored at -20° C. Genomic DNA was extracted
73 (28) and stored at 4° C for immediate use or stored at -20° C.

74 **Long fragment isolation.** An initial PCR amplified long fragments (>6kb) spanning the
75 *mcyA2* and *mcyB1* gene regions. PCRs were performed using a DNA polymerase with
76 proofreading activity (Takara LA *Taq* Polymerase) in 50 µl reactions containing 200 nM of each
77 primer (Operon Biotechnologies), 400 µM of each dNTP (Fisher Bioreagents), 1X Takara LA
78 *Taq* buffer II, 0.05 U *Taq*, and 0.5-5 ng of DNA. PCR was performed under the following
79 conditions: 94° C for 30 s; 14 cycles of 94° C for 30 s, 64° C (minus 0.5° C for each cycle) for

AEM Accepts published online ahead of print

80 30 s, 68° C for 9 min; 26 cycles of 94° C for 30 s, 58° C for 30 s, 68° C for 9 min; and 72° C for
81 10 min. The following primer combinations were used: *mcyA*-CdR/*mcyB1*-MaBR, *mcyA*-
82 CdR/*mcyB1*-MaCPa1R, *mcyA*-CdR/*mcyB1*-Pa2R, and *mcyA*-CdR/*mcyB1*-AnaR (Table 2).
83 Products of the appropriate size, confirmed using agarose gel electrophoresis, were cloned (Table
84 3) into the the CopyControl™ pCC1™ Blunt Cloning-Ready Vector, (Epicentre) and
85 transformed into TransforMax™ EPI300™ electrocompetent *E.coli* cells (Epicentre). Clones
86 were induced to high-copy number, harvested and stored at -20° C or immediately processed for
87 DNA extraction (28).

88 **Sequencing and molecular data analysis.** PCR was also performed for each partial gene
89 region (*mcyA2* and *mcyB1*). PCR was performed in 25 µl reactions containing 200 nM of each
90 primer (Operon Biotechnologies) 400 µM of each dNTP (Fisher Bioreagents), 1X Green
91 GoTaq® Flexi buffer (Promega), 2 mM $MgCl_2$ (Fisher Bioreagents), 0.025 U GoTaq® Flexi
92 polymerase (Promega), and 0.5-5 ng of DNA template. The protocol was as follows: 95° C for 2
93 min; 40 cycles of 95° C for 30 s, 59° C for 30 s (or 53° C for *mcyB*), 72° C for 1 min; and 72° C
94 for 10 min. PCR products were verified using agarose gel electrophoresis and cleaned-up using
95 the Wizard® SV gel and PCR clean-up system (Promega). Sequencing was performed on a 3730
96 capillary electrophoresis DNA analyzer (Applied Biosystems) using the Big Dye® Terminator
97 version 3.1 Cycle Sequencing Mix (Applied Biosystems) at the Molecular Biology Resource
98 Facility (University of Tennessee, Knoxville, TN). Forward and reverse sequence reads were
99 obtained for each sample. Sequences were deposited in the NCBI.

100 Using phylogenetic analyses and bioinformatic prediction, the condensation domain of
101 *mcyA* and the adenylation domain of *mcyB1* were examined. Phylogenetic trees depicting
102 evolutionary relationships among cultures and clones (Table 2) were independently developed

AEM Accepts published online ahead of print

103 for each marker using Bayesian inference implemented in MrBayes v 3.1.2 (9). The best-fitting
104 nucleotide substitution models for Bayesian inference were chosen using both hierarchical
105 likelihood ratio tests and the Akaike information criterion implemented in MrModeltest v 2.3
106 (19) and PAUP* v 4.0b10 (32). The best-fitting models for *mcyA* and *mcyB1* were HKY+G and
107 GTR+G, respectively. Bayesian analysis was performed for one million generations in two
108 independent runs, saving trees every 100 generations. In both cases the average standard
109 deviation of split frequencies was less than 0.01 at the end of one million generations, indicating
110 that the independent runs had converged on a stationary tree distribution. To assess an
111 appropriate number of trees to discard as "burn-in," likelihoods were plotted against generation
112 number to determine the point at which likelihoods reached stationarity. In each case the first
113 500 (of 10,000) trees were discarded, and the remaining trees from each run were used for
114 analysis. Majority-rule consensus trees with posterior probabilities were constructed from the
115 combined pool of trees and edited in FigTree (22). For variant prediction, the deduced amino
116 acid sequence information at *mcyB1* (converted by Frameplot (10)) was used to predict what
117 amino acid would likely be incorporated into the microcystin molecule at site two (for references
118 see (1, 35)). We predicted the microcystin variant for each of the 6 kb fragments using
119 NRPSpredictor (26) (please note: predictions are not considered absolute due to the dynamic
120 environment where the synthetases occur within cells (17, 24, 34)). We also tested for
121 compatible evolutionary relationships between these two genes using a partition-homogeneity
122 test (6) in PAUP* with 1000 replicates, heuristic searches, sequence data from the six available
123 cultures in Genbank (Table 2) and our long fragments.
124 **Toxin analysis.** Filters for toxin analysis were extracted in 10 mL of 50% aqueous
125 acidified methanol, and the extracts were clarified by centrifugation, filtered through a 0.45-

AEM Accepts published online ahead of print

126 µm nylon syringe filter, and stored at -20˚ C. Microcystin concentrations were determined for
127 samples listed in Table 3 using a protein phosphatase inhibition assay (PPIA), modified from
128 (3) as described previously in (8). Positive samples were confirmed and variants determined
129 using LCMS techniques if concentrations in the voucher extract where > 0.5 µg/ml. For
130 samples with no detectable toxin levels, the method detection limits for the assay are
131 provided. These vary depending on the volume filtered and the instrument sensitivity that day.
132 Variant analysis was run on a MicroMass ZQ4000 against an RR, LR, and LF standard.
133 Unknown peaks that fell between the retention times of our most polar (microcystin-RR) and
134 non-polar (microcystin-LF) standards were considered putative microcystin variants.
135 Individual isomers were identified on the basis of their uv extinct coefficient, molecular ion
136 and retention time compared to literature values (15) and secondary standards. Percent
137 composition was determined using their UV absorbance at 232 nm and normalized to 100%.

138 **Results**

139 In total we examined fourteen 6 kb fragments (LF-1 through LF-14 in Table 3) isolated
140 from natural samples with wide ranging microcystin concentrations. Our initial success with lab
141 strains *M. aeruginosa* PCC 7806, LE3 (known microcystin-LR variant producers) and UV027 (a
142 known microcystin-RR variant producer), which were selected as a semi-diverse sampling panel
143 (unpublished data), provided a proof of concept to begin investigating natural phytoplankton
144 communities. Microcystin concentrations at sites from the Great Lakes ranged from < 0.1 to 2.2
145 µg/L; however the Waterville reservoir contained toxin concentrations in excess of 800 µg/L,
146 which was well above the levels set by the World Health Organization (1 µg/L) for safe drinking
147 and recreational water (5, 36).

AEM Accepts published online ahead of print

148 Comparing sequence information for both markers within individual 6 kb fragments
149 revealed compelling results for carrier of the toxin genes. According to our phylogenetic
150 analyses (Figs. 2A and 2B), all of our 6 kb fragments, including the previously discovered
151 "*Microcystis*-like" *mcyA* sequences, clustered with cultured isolates listed in GenBank. The
152 sequence information suggested that these unusual *mcyA* gene fragments were found in
153 organisms clustering with ancestral *Microcystis* spp. at *mcyA* (1a, 1b, and 1c in Fig. 2A) and the
154 *mcyB1*(B) genotype of *Microcystis* at *mcyB* (5 in Fig. 2B). Thus, we identified these unusual
155 *mcyA* gene fragments as coming from an ancestral *Microcystis* rather than a novel species,
156 hybrid, or a chimera. In addition to clustering with cultures known to produce the microcystin-
157 LR, LA, or YR (17), our bioinformatic prediction suggested that leucine would likely be
158 incorporated at site two of the microcystin molecule by the adenylation domain in *mcyB1*. This
159 bioinformatic result was corroborated by our toxin analysis, which indicated that 78% of the
160 microcystin content did indeed contain leucine (the "L" in microcystin-LR; See Table 3).
161 During analyses we noticed that a *mcyB1* sequence from a cultured isolate (*Microcystis*
162 sp. N-C 118-2) closely resembled two environmental *mcyB1* fragments from individuals
163 containing the unusual *mcyA* gene fragments (LF-1 and LF-2). We acquired *Microcystis* sp. N-C
164 118-2 from the Norwegian Institute for Water Research, sequenced its *mcyA* gene region, and
165 found that this was identical to some of the previously identified "*Microcystis*-like" sequences.
166 Thus, we concluded that this strain was contained within the ancestral *mcyA* genotypes (1c in
167 Fig. 2A).
168 In addition to ancestral *Microcystis*, we isolated six other 6 kb fragments from Lakes
169 Erie, Ontario, and Waterville Reservoir (LF-5 through LF-10 in Table 3). Their placement
170 among cultured isolates and predicted variants are indicated in Figs. 2A and 2B. We also

AEM Accepts published online ahead of print

171 examined our bioinformatic predictions presented in Fig. 2B with toxin analysis using LCMS
172 when available (Table 3). When sampling Sandusky Bay (Lake Erie Sta. 1163 in Fig 1A; see
173 Table 3), we isolated three 6 kb fragments that grouped together with *Planktothrix* spp. (4 and 7
174 in Figs. 2A and 2B) at both markers (*i.e., mcyA* and *mcyB*) and predicted the synthetase would
175 likely incorporate leucine at site 2 (*i.e.,* possibly corresponding to the highly toxic variants
176 microcystin-LR or LA). Our toxin analysis did indeed reveal that microcystin-LR was detected at
177 this location, though only 12%. Three other 6 kb fragments were isolated from the highly toxic
178 bloom in Waterville reservoir. These sequences clustered with *Microcystis* cultures using *mcyA*
179 and *mcyB* (2 and 6 in Figs. 2A and 2B); however, the sequence of *mcyB1* matched the *mcyB1* (C)
180 genotype (17) that is associated with the production of the microcystin-RR variant in cultures (6
181 in Fig. 2B). Our bioinformatic prediction also suggested that an arginine would likely be
182 incorporated into the microcystin molecule at site 2. Subsequent toxin analysis detected that
183 microcystin-RR was present at this location.

184 Finally, our partition-homogeneity test of phylogenetic incongruence detected a
185 significant difference (P-value = <0.01) between sequence information at *mcyA* and *mcyB1*. This
186 suggests that these two genes did not follow the same evolutionary path, which was also
187 supported by our phylogenetic analysis. The *Microcystis* cultures and clones grouped differently
188 with respect to the two gene markers, where *mcyA* demonstrated division between ancestral and
189 recently derived lineages and mcyB1 showed clusters based on predicted microcystin variants
190 (e.g. incorporation of leucine versus arginine).

191 **Discussion**

192 **Comparing *mcy* genes for individuals.** Our multi-gene approach improved the
193 molecular characterization of individual *mcy* synthetase operons from natural communities. First,

AEM Accepts published online ahead of print

194 our long fragments identified the source of the unusual *mcyA* gene fragments (i.e., "*Microcystis*-
195 like") as coming from a strain of ancestral *Microcystis* whose *mcy* biosynthesis operon had not
196 yet been described. Second, by linking the phylogenetic marker in *mcyA* to the predictor of
197 amino acid incorporation in the adenylation domain of *mcyB*, we created an improved diagnostic
198 tool for examining the genetic potential of individual cyanobacteria to produce specific
199 microcystin variants. It is important to note, however, that the mere presence of the *mcy* operon
200 does not confirm toxin production nor does the toxin prediction confirm variant presence.
201 Previous experimentation as well as our results indicated that microcystin production is
202 controlled by the interaction between genetics and environmental conditions (17, 27, 34, 35).
203 Third, our *mcyA* analysis suggested that most *Microcystis* spp. in culture represent a derived
204 lineage, while one isolate (i.e., *Microcystis* sp. N-C 118-2) appeared to be from a more ancestral
205 lineage that is commonly found in natural aquatic communities, making it perhaps a better model
206 for lab-based studies

207 The determination that the previously discovered "*Microcystis*-like" *mcyA* sequences
208 originate from a potentially highly toxic *Microcystis* spp. is not trivial, since these populations
209 have appeared in surface waters of Lake Erie (30) and Lake Ontario (8) in 2001, 2003, 2004,
210 2006, and 2007. Also, this genotype was detected from Lake Erie sediments in 2004 (29),
211 demonstrating its pervasiveness in natural communities. Regarding the isolate from Norway (i.e.,
212 *Microcystis* sp. N-C 118-2), our methods directed us from an unknown environmental sample to
213 a matching cultured isolate - the exact opposite of most conventional microbial ecology. This
214 culture will provide a useful tool for determining whether this ancestral genotype acts differently
215 than other *Microcystis*.

AEM Accepts published online ahead of print

216 The other 6 kb fragments isolated from *Planktothrix* or *Microcystis*-dominated waters
217 demonstrated the flexibility of our approach. We designed PCR primers for a wide variety of
218 microcystin producers based on sequences of full *mcy* synthetase operons for *Anabaena*,
219 *Microcystis*, and *Planktothrix* and our results indicated their probable success with at least two of
220 those genera. Since *Anabaena* was not detected in natural communities, we are not sure whether
221 our methods are limited by sampling or amplification bias or simply that this genus was not
222 present at the time of sampling. Future work is needed to verify successful isolation of 6 kb
223 fragments from natural samples of *Anabaena* sp., however our *mcyA* and *mcyB* primer sets
224 (Table 1) were successful at amplifying these targeted regions in *Anabaena* cultures (data not
225 shown).

226 Our partition-homogeneity test of phylogenetic incongruence indicated that the
227 phylogenetic signal in these two genes is incompatible. It is somewhat surprising to detect
228 different evolutionary paths for these genes, since they are located along the same operon and
229 Rantala *et al*. 2004 (23) demonstrated that several *mcy* genes (*mcyA*, *mcyD*, and *mcyE*) followed
230 similar evolutionary pathways with housekeeping genes (16S rRNA and *rpoC1*) indicating
231 evidence for vertical and not horizontal gene transfer. However, Mikalsen *et al*. 2003 (17)
232 reported evidence for recombination events between *mcyB* and *mcyC*, which may help explain
233 the differences in evolutionary pathway between the two loci at *mcyA* and *mcyB1*. Our results are
234 consistent with recombination events with one or both of these genes, but following the analysis
235 of *mcyA* from Rantala *et al*. 2004 (23), it is more likely that a recombination event took place in
236 *mcyB* than *mcyA*. Thus, our results also support the findings from Mikalsen *et al*. 2003 (17) in
237 detecting different evolutionary pathways between *mcyA* and *mcyB*.

AEM Accepts published online ahead of print

238 **Prediction of microcystin variants.** Due to our experimental design, which aimed to
239 isolate large *mcy* operon fragments from individual cells rather than test for prediction accuracy
240 in natural populations, we were only able to broadly test whether a predicted variant was present
241 in our samples. Our results did indeed confirm all predictions; however all of the samples
242 contained a mixture of variants that incorporated leucine and arginine into at site two of the
243 microcystin molecule. Our initial examination of this prediction tool in natural samples provides
244 evidence for its accuracy, but this support is scant at best.

245 **Concluding remarks.** Microbial ecologists are still far from sequencing large numbers
246 of full genomes of single cells within natural communities. While these technologies are being
247 developed, there is still a need to compare multiple genes from single cells collected from natural
248 populations. Molecular tools offer an advantage over more traditional chemical measures of
249 toxin concentration, as they may serve as a predictor of potential bloom events to come (while
250 toxin measurements suggest the event, and potentially exposures, have already occurred) While
251 our analysis comparing gene sequences from 20 different individual *mcy* operons is not
252 exhaustive, it is the most extensive to date (21). Future experimentation with larger sample sizes
253 will be required to reduce any unforeseen biases. Our methods offer an improvement over
254 traditional cloning procedures (i.e., plasmids) by providing a means for multi-gene comparisons
255 between individual *mcy* operons from distantly related cyanobacteria. Since these operons
256 encode for non-ribosomal peptide synthetases, we envision that our methods can be adapted for a
257 wide variety of secondary metabolites.

AEM Accepts published online ahead of print

258 **Acknowledgements**

259 We thank the captain and crew of the CCGS *Limnos* as well as Dr. George Bullerjahn, Dr. Andy
260 Ouellette, and Dan Walsh for assistance and discussions and Emily Rogers for supplying the
261 samples from the Waterville Reservoir. Helpful discussions with Dr. Alison Buchan were also
262 appreciated when developing these methods. This article includes research supported by the
263 National Oceanic and Atmospheric Administration Coastal Ocean Program (MERHAB-LGL,
264 #NA 160 P2788), an Oceans and Human Health grant (NA05NOS4781251), the National
265 Science Foundation (DEB-0129118), and the Helsingin Sanomat Centennial Foundation.

AEM Accepts published online ahead of print

Table 1. Primers used in this study

Primer name (designed taxon specificity and/or gene locus)	Sequence (5'-3')	Ref.
mcyB1-MaF (*Microcystis*; mcyB1)	TTTATACATCAGGTTCAACGGG	this paper
mcyB1-PaF (*Planktothrix*; mcyB1)	TCTATACATCAGGTTCCACTGG	this paper
mcyB1-AnaF (*Anabaena*)	TTTATACCTCTGGTTCTACAGG	this paper
mcyB1-MaBR (*Microcystis mcyB1* (B) genotype; mcyB1)	TATTTCCCCTAATTCAATCCG	this paper
mcyB1-MaCPa1R (*Microcystis mcyB1* (C) genotype and *Planktothrix*; mcyB1)	AATTTCTCCTAATTCAATGCG	this paper
mcyB1-Pa2R (*Planktothrix*; mcyB1)	AATTTCTCCTAATTCAATACG	this paper
mcyB1-AnaR (*Anabaena*)	TATTTCTCCTAACTCAATACG	this paper
mcyA-CdF (mcyA)	AAAATTAAAAGCCGTATCAAA	(7)
mcyA-CdR (mcyA)	AAAAGTGTTTTATTAGCGGCTCAT	(7)
mcyB.fw (mcyB1)	ATTACAGCAGAGAAAATCCAAGCA	(16)
mcyB.rev (mcyB1)	TCGCAATAGCGGGATCA	(16)
2156-F (mcyB1)	ATCACTTCAATCTAACGACT	(17)
3111-R (mcyB1)	AGTTGCTGCTGTAAGAAA	(17)

AEM Accepts published online ahead of print

Table 2. List of cultures and clones. Accession numbers are given for *mcy* gene sequences where available. [a] *grsA* was used as an out-group for *mcyB* as in (17).

Name	Strain	*mcyA* accession #	*mcyB* accession #
Anabaena circinalis	90	AY212249 (31)	AY212249 (31)
Anabaena flos-aquae	NIVA-CYA83	AJ515466 (7)	-
Anabaena lemmermannii	66A	AJ515462 (7)	-
Bacillus brevis		-	X15577 (13)[a]
Microcystis aeruginosa	K-139	AB019578 (18)	-
Microcystis aeruginosa	LE3	DQ379709 (30)	FJ411060 (this study)
Microcystis aeruginosa	N-C 228-1	-	AJ492558 (17)
Microcystis aeruginosa	N-C 31	-	AJ492552 (17)
Microcystis aeruginosa	NIES 89	AJ515459 (7)	-
Microcystis aeruginosa	NIES-298	-	AB092806 (37)
Microcystis aeruginosa	PCC 7806	AF183408 (33)	AF183408 (33)
Microcystis aeruginosa	PCC 7813	-	AY034601 (2)
Microcystis aeruginosa	PCC 7941	AJ515460 (7)	-
Microcystis aeruginosa	UV027	AF458094 (25)	AF458094 (25)
Microcystis sp.	IZANCY A5	AJ515456 (7)	-
Microcystis sp.	N-C 118-2	FJ379558 (this study)	AJ492554 (17)
Microcystis sp.	N-C 324-1	-	AJ492560 (17)
Microcystis sp.	TuM7C	AJ515458 (7)	-
Nostoc sp.	152	AJ515475 (7)	-
Nostoc sp.	IO-102-I	AY566856 (20)	-
Planktothrix agardhii	CYA 126/8	AJ441056 (4)	AJ441056 (4)
Planktothrix agardhii	NIVA-CYA 34	AJ515474 (7)	-
Planktothrix agardhii	No39	-	AJ749269 (14)
Planktothrix rubescens	CCAP1459/31	-	AJ863134 (14)
Planktothrix rubescens	NIVA-CYA 18	-	AJ749283 (14)
Planktothrix rubescens	No10	-	AJ890255 (14)
Planktothrix rubescens	No108	-	AJ749281 (14)
Planktothrix rubescens	No64	-	AJ749277 (14)
Planktothrix rubescens	No80	-	AJ749278 (14)
Planktothrix rubescens	No82	-	AJ749279 (14)
mcyA clone	01LO-LSB4	EF424297 (8)	-
mcyA clone	03LO-OS9	EF424370 (8)	-
mcyA clone	LE03-1163-G03	DQ379678 (30)	-
mcyA clone	LE04 974 C01	DQ379692 (30)	-
mcyA clone	LE04-1163-A04	DQ379674 (30)	-
mcyA clone	LE04-974-B07	DQ379691 (30)	-
mcyA clone	LE04-974-C02	DQ379683 (30)	-
mcyA clone	LE04-974-C03	DQ379704 (30)	-
mcyA clone	LE04-974-D02	DQ379684 (30)	-

AEM Accepts published online ahead of print

Table 3. Samples used to generate long fragment (LF) and short fragment (SF, 297 bp *mcyA* fragment only) products in this study. Microcystins concentrations and variants were determined by the protein phosphatase inhibition assay and LCMS, respectively. NT means not tested. ND = not detected by LCMS

ID	Sample Location	Collection Date	Accession # (*mcyA*/*mcyB1* or *mcyA* only)	Microcystin (µg/L)	Variants
LF-1	Hamilton Harbour	Aug-06	FJ379539/FJ379559	NT	ND
LF-2	Sta 971 Lake Erie	Aug-07	FJ379540/FJ379560	0.1	ND
LF-3	Sta 974 Lake Erie	Aug-07	FJ379541/FJ379561	<0.1	ND
LF-4			FJ379542/FJ379562		
LF-5	Sandusky Bay	Aug-07	FJ379543/FJ379563	2.2	63% dRR, 23% YR, 12% dLR
LF-6			FJ379544/FJ379564		
LF-7			FJ379545/FJ379565		
LF-8	Waterville NC	Oct-07	FJ379546/FJ379566	824.3	26% RR, 74% LR
LF-9			FJ379547/FJ379567		
LF-10			FJ379548/FJ379568		
LF-11	Erie (PA) Harbor	Aug-07	FJ379549/FJ379569	0.68	22% RR, 78% LR
LF-12			FJ379550/FJ379570		
LF-13			FJ379551/FJ379571		
LF-14			FJ379552/FJ379572		
SFA-1	Hamilton Harbour	Aug-06	FJ379553	NT	NT
SFA-2			FJ379554		
SFA-3	(n=6)		FJ379555		
SFA-4	(n=7)	Aug-02	FJ379556	NT	ND
SFA-5	(n=30)		FJ379557		

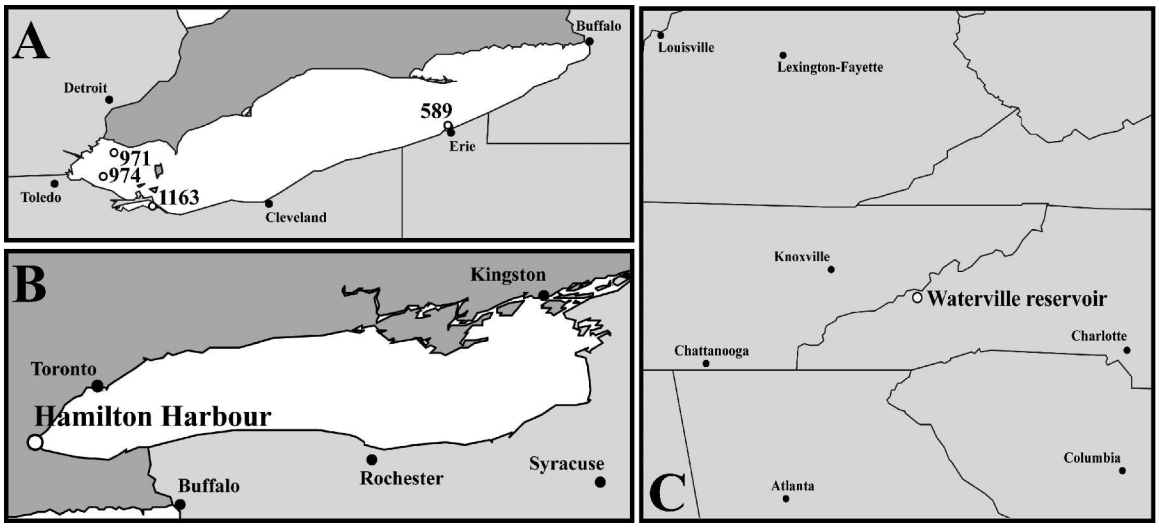
AEM Accepts published online ahead of print

Figure Legends

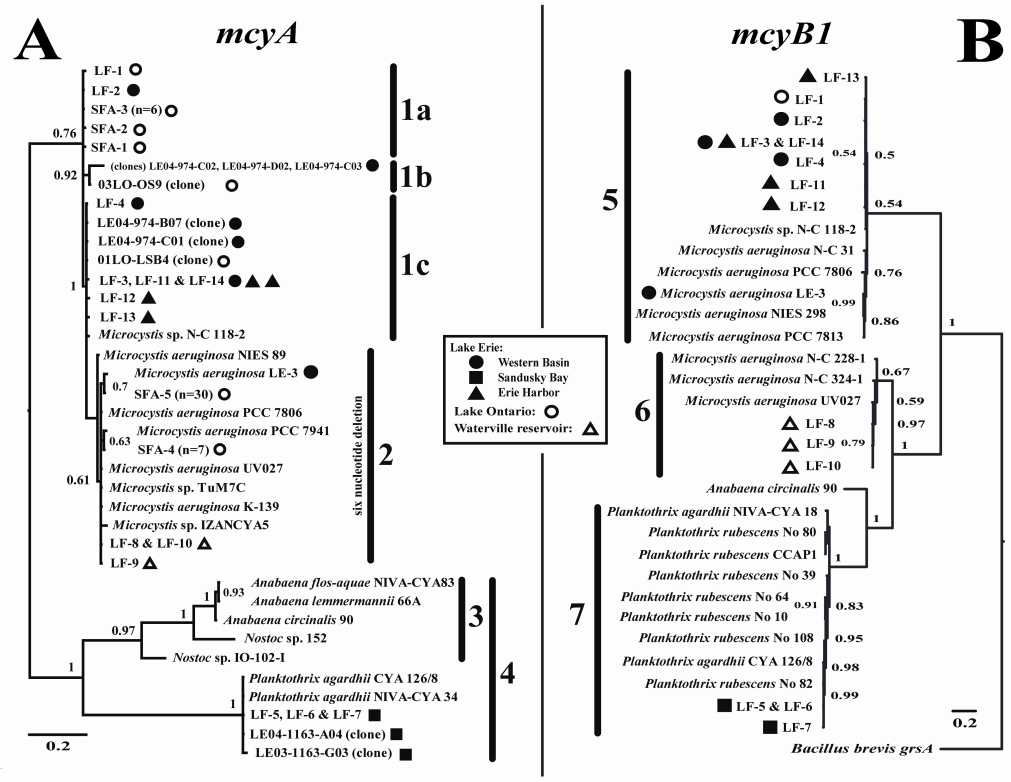
Figure 1. Maps indicating sample collection sites in A) Lake Erie, B) Lake Ontario, and C) North Carolina. Geographical coordinates (in decimal degrees): Hamilton Harbour (43.295, -79.813); Sta. 589 (42.134, -80.101); Sta. 971 (41.948, -83.054); Sta. 974 (41.730, -83.157); Sta. 1163 (41.469, -82.715); and Waterville reservoir (35.695, -83.049). Maps generated using the Planiglobe web mapping service.

Figure 2. Phylogenetic trees for the condensation domain in *mcyA* and the adenylation domain in *mcyB1*. A. Bayesian analysis for *mcyA* indicating placement of 6 kb fragments among cultures and clones. Posterior probabilities between 0.5 and 1 are listed. 1a-1c) The ancestral *Microcystis* genotypes; formerly known as the "*Microcystis*-like" sequences (8, 30) containing six nucleotides that were not present among *Microcystis* cultures, except for *Microcystis* sp. N-C 118-2. 2) *Microcystis* spp. clade including most existing cultures, which all exhibit a common six nucleotide deletion. 3) *Anabaena* and *Nostoc* spp. 4) The filamentous cyanobacteria including 6 kb fragments LF5-LF7, which cluster with *Planktothrix* cultures. B. Bayesian analysis for *mcyB1* indicating placement of 6 kb fragments among cultures with listing of posterior probabilities between 0.5 and 1. *Bacillus brevis* was used as an out-group as in (17). 5) Corresponding to *Microcystis* genotype *mcyB1* (B) known to produce microcystin-LR, LA, and YR (17). Leucine was predicted to be incorporated at site two of the resulting microcystin molecule for each member using an adenylation domain prediction tool (see text) 6) Corresponding to *Microcystis* genotype *mcyB1* (C) known to produce microcystin-RR and many others (17). All members, including 6 kb fragments LF8-LF10, were predicted to incorporate arginine into site 2 of the microcystin molecule. 7) *Planktothrix* spp. including 6 kb fragments LF5-LF7. Each member was predicted to incorporate a leucine into site two of the microcystin molecule.

AEM Accepts published online ahead of print



AEM Accepts published online ahead of print



AEM Accepts published online ahead of print

1. **Botes, D. P., A. A. Tuinman, P. L. Wessels, C. C. Viljoen, H. Kruger, D. H. Williams, S. Santikarn, R. J. Smith, and S. J. Hammond.** 1984. The structure of cyanoginosin-LA, a cyclic heptapeptide toxin from the cyanobacterium *Microcystis-aeruginosa*. Journal of the Chemical Society-Perkin Transactions 1**:**2311-2318.
2. **Botes, E.** Molecular study of selected *Microcystis aeruginosa* strains. Unpublished.
3. **Carmichael, W. W., and J. S. An.** 1999. Using an enzyme linked immunosorbent assay (ELISA) and a protein phosphatase inhibition assay (PPIA) for the detection of microcystins and nodularins. Natural Toxins **7:**377-385.
4. **Christiansen, G., J. Fastner, M. Erhard, T. Borner, and E. Dittmann.** 2003. Microcystin biosynthesis in *Planktothrix*: Genes, evolution, and manipulation. Journal of Bacteriology **185:**564-572.
5. **Dittmann, E., and C. Wiegand.** 2006. Cyanobacterial toxins - occurrence, biosynthesis and impact on human affairs. Molecular Nutrition & Food Research **50:**7-17.
6. **Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult.** 1994. Testing significance of incongruence. Cladistics-the International Journal of the Willi Hennig Society **10:**315-319.
7. **Hisbergues, M., G. Christiansen, L. Rouhiainen, K. Sivonen, and T. Borner.** 2003. PCR-based identification of microcystin-producing genotypes of different cyanobacterial genera. Archives of Microbiology **180:**402-410.
8. **Hotto, A. M., M. F. Satchwell, and G. L. Boyer.** 2007. Molecular characterization of potential microcystin-producing cyanobacteria in Lake Ontario embayments and nearshore waters. Applied and Environmental Microbiology **73:**4570-4578.

AEM Accepts published online ahead of print

9. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754-755.

10. **Ishikawa, J., and K. Hotta.** 1999. FramePlot: a new implementation of the Frame analysis for predicting protein-coding regions in bacterial DNA with a high G plus C content. FEMS Microbiology Letters **174:**251-253.

11. **Jungblut, A. D., and B. A. Neilan.** 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. Archives of Microbiology **185:**107-114.

12. **Kaneko, T., N. Nakajima, S. Okamoto, I. Suzuki, Y. Tanabe, M. Tamaoki, Y. Nakamura, F. Kasai, A. Watanabe, K. Kawashima, Y. Kishida, A. Ono, Y. Shimizu, C. Takahashi, C. Minami, T. Fujishiro, M. Kohara, M. Katoh, N. Nakazaki, S. Nakayama, M. Yamada, S. Tabatai, and M. M. Watanabe.** 2007. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Research **14:**247-256.

13. **Kratzschmar, J., M. Krause, and M. A. Marahiel.** 1989. Gramicidin-S biosynthesis operon containing the structural genes *grsA* and *grsB* has an open reading frame encoding a protein homologous to fatty-acid thioesterases. Journal of Bacteriology **171:**5422-5429.

14. **Kurmayer, R., G. Christiansen, M. Gumpenberger, and J. Fastner.** 2005. Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*. Microbiology-Sgm **151:**1525-1533.

15. **Lawton, L. A., and C. Edwards.** 2001. Purification of microcystins. Journal of Chromatography A **912:**191-209.

AEM Accepts published online ahead of print

16. **Mbedi, S., M. Welker, J. Fastner, and C. Wiedner.** 2005. Variability of the microcystin synthetase gene cluster in the genus *Planktothrix* (Oscillatoriales, Cyanobacteria). Fems Microbiology Letters **245:**299-306.

17. **Mikalsen, B., G. Boison, O. M. Skulberg, J. Fastner, W. Davies, T. M. Gabrielsen, K. Rudi, and K. S. Jakobsen.** 2003. Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. Journal of Bacteriology **185:**2774-2785.

18. **Nishizawa, T., M. Asayama, K. Fujii, K. Harada, and M. Shirai.** 1999. Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in Microcystis spp. Journal of Biochemistry **126:**520-529.

19. **Nylander, J. A. A.** 2004. MrModeltest v2.3. Program distributed by the author, Evolutionary Biology Centre, Uppsala University.

20. **Oksanen, I., J. Jokela, D. P. Fewer, M. Wahlsten, J. Rikkinen, and K. Sivonen.** 2004. Discovery of rare and highly toxic microcystins from lichen-associated cyanobacterium Nostoc sp strain IO-102-I. Applied and Environmental Microbiology **70:**5756-5763.

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

22. **Rambaut, A.** 2008. FigTree v1.1.2. Institute of Evolutionary Biology, University of Edinburgh

23. **Rantala, A., D. P. Fewer, M. Hisbergues, L. Rouhiainen, J. Vaitomaa, T. Borner, and K. Sivonen.** 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proc Natl Acad Sci U S A **101:**568-73.

AEM Accepts published online ahead of print

24. **Rapala, J., K. Sivonen, C. Lyra, and S. I. Niemela.** 1997. Variation of microcystins, cyanobacterial hepatotoxins, in Anabaena spp. as a function of growth stimuli. Applied and Environmental Microbiology **63:**2206-2212.

25. **Raps, S., D. Miller, and S. Ratner.** The peptide synthetase operon for microcystin-RR in *Microcystis aeruginosa* UV027. Unpublished.

26. **Rausch, C., T. Weber, O. Kohlbacher, W. Wohlleben, and D. H. Huson.** 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Research **33:**5799-5808.

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

28. **Rinta-Kanto, J. M., A. J. A. Ouellette, G. L. Boyer, M. R. Twiss, T. B. Bridgeman, and S. W. Wilhelm.** 2005. Quantification of toxic *Microcystis* spp. during the 2003 and 2004 blooms in western Lake Erie using quantitative real-time PCR. Environmental Science & Technology **39:**4198-4205.

29. **Rinta-Kanto, J. M., M. A. Saxton, J. M. DeBruyn, K. L. Smith, C. H. Marvin, K. A. Krieger, G. S. Sayler, G. L. Boyer, and S. W. Wilhelm.** 2008. 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.

AEM Accepts published online ahead of print

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

31. **Rouhiainen, L., T. Vakkilainen, B. L. Siemer, W. Buikema, R. Haselkorn, and K. Sivonen.** 2004. Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. Applied and Environmental Microbiology **70:**686-692.

32. **Swofford, D. L.** 1998. PAUP*: Phylogenetic Analysis Using Parsimony and Other Methods. Sinauer Associates, Sunderland, MA.

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

34. **Tonk, L., P. M. Visser, G. Christiansen, E. Dittmann, E. Snelder, C. Wiedner, L. R. Mur, and J. Huisman.** 2005. The microcystin composition of the cyanobacterium Planktothrix agardhii changes toward a more toxic variant with increasing light intensity. Applied and Environmental Microbiology **71:**5177-5181.

35. **Welker, M., and H. von Dohren.** 2006. Cyanobacterial peptides - Nature's own combinatorial biosynthesis. FEMS Microbiology Reviews **30:**530-563.

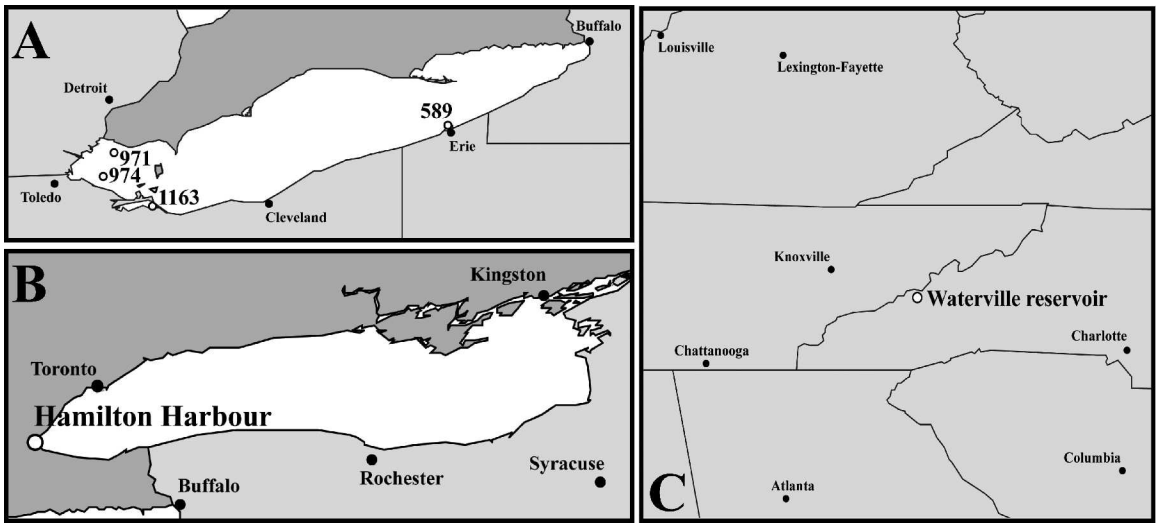
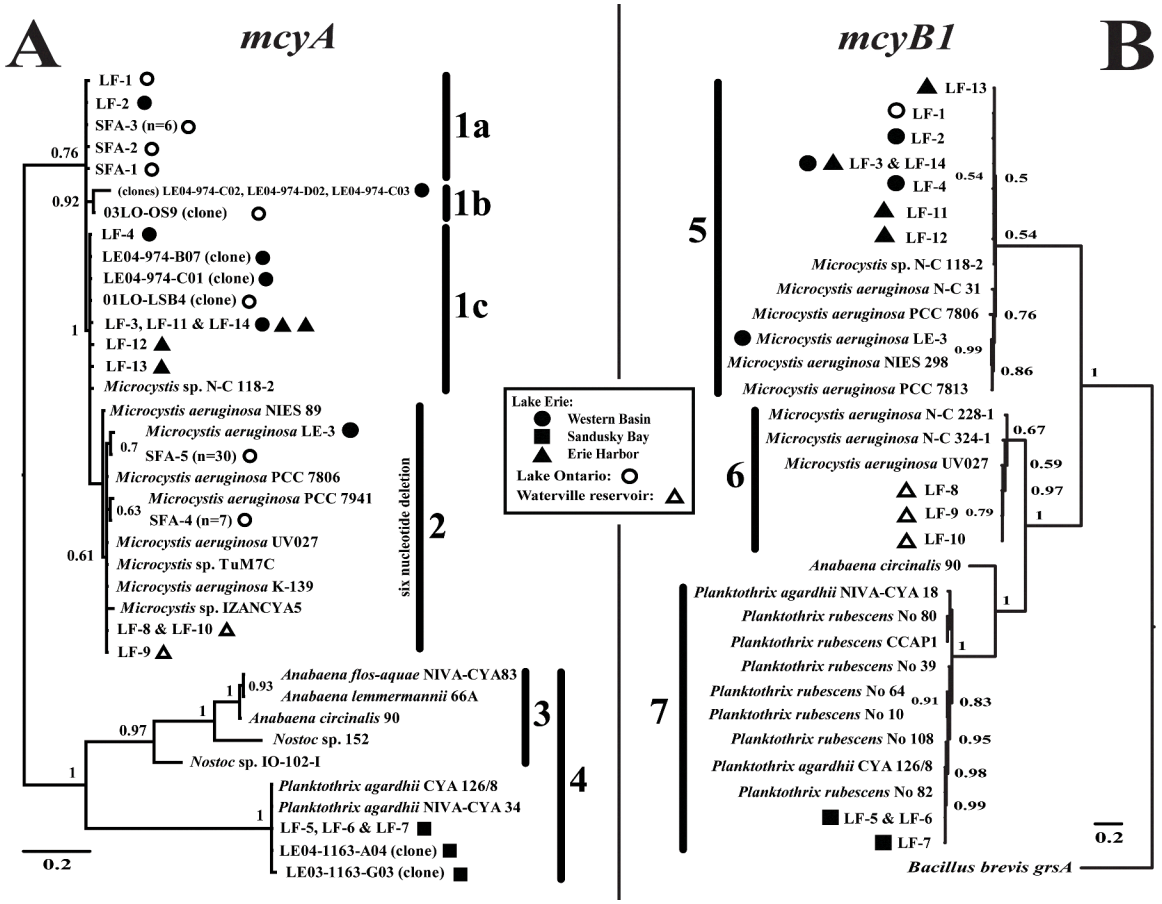
36. **WHO.** 1998. Guidelines for drinking-water quality, 2nd ed, vol. Addendum to Vol. 2, Geneva:World Health Organization.

AEM Accepts published online ahead of print

37. **Yoshida, T., Y. Yuki, S. Lei, H. Chinen, M. Yoshida, R. Kondo, and S. Hiroishi.** 2003. Quantitative detection of toxic strains of the cyanobacterial genus *Microcystis* by competitive PCR. Microbes and Environments **18:**16-23.

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

AEM Accepts published online ahead of print



AEM Accepts published online ahead of print