The mosaic structure of the *mcvABC* operon in *Microcystis*

Ave Tooming-Klunderud,^{1,2} Bjørg Mikalsen,^{2†} Tom Kristensen^{1,3} and Kjetill S. Jakobsen²

Correspondence

Kjetill S. Jakobsen
kjetill.jakobsen@bio.uio.no

¹University of Oslo, Department of Molecular Biosciences, 0316 Oslo, Norway

²University of Oslo, Department of Biology, Centre for Ecological and Evolutionary Synthesis (CEES), 0316 Oslo, Norway

³University of Oslo, Microbial Evolution Research Group (MERG), 0316 Oslo, Norway

An extensive study of the *mcvABC* genes and regions flanking the *mcv* gene cluster was performed in naturally occurring *Microcystis* strains. Lack of methylation in strains producing only desmethyl⁷-microcystin was found to be associated with point mutations in substrate-binding sequence motifs of the *N*-methyltransferase (NMT) domain in *McyA*. Multiple recombination events giving rise to 'phylogenetic mosaics' were detected within the NMT-domain-encoding *mcvA* sequences and the adenylation (A) domain sequences of *mcvB* and *mcvC*. Recombination leading to exchanges between the *mcvB* and *mcvC* regions encoding A domains in modules *McyB1* and *McyC* was also detected. A previously reported replacement of the A domain in *McyB1* was found to involve the region between the conserved motifs A3 and A8/A9. In all microcystin-producing strains the *mcv* gene cluster was flanked by the genes *uma1* and *dnaN*. Clear indications of recombination, an insertion element and footprints of IS elements were found in the *dnaN*–*mcvJ* intergenic region. Among the non-microcystin producers, *uma1* and *dnaN* were linked in some, but not all strains. Most non-producing strains lacked all *mcv* genes, while one strain possessed a partially deleted *mcv* operon. Our results show that frequent horizontal gene transfer events in addition to point mutations and insertions/deletions contribute to variation in the *mcv* gene cluster.

Received 14 December 2007

Revised 16 April 2008

Accepted 17 April 2008

INTRODUCTION

Blooms of cyanobacteria are common in freshwater lakes, ponds and water reservoirs worldwide. Cyanobacteria of the genus *Microcystis* are widespread in such blooms and constitute a health risk for animals and humans through the production of hepatotoxic microcystins – a family of cyclic heptapeptides. Microcystins share a common structure of cyclo (D-Ala¹–L-X²–D-MeAsp³–L-Z⁴–Adda⁵–D-Glu⁶–Mdha⁷), where Mdha is *N*-methyldehydroalanine, D-MeAsp is 3-methylaspartic acid and Adda is 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid. X and Z indicate variable L-amino acids. At least 89 different microcystin isoforms, differing in modifications of the

peptide backbone or the type of amino acids incorporated, have been identified (Welker & von Döhren, 2006). Many strains produce several microcystin isoforms, but only one or two isoforms dominate in any single strain.

Microcystins are synthesized nonribosomally by a thio-template mechanism catalysed by microcystin synthetase (Marahiel *et al.*, 1997). Microcystin synthetase gene clusters (*mcv*) include genes for peptide synthetases, polyketide synthases, mixed peptide/polyketide synthases and tailoring enzymes, and have been characterized in detail in *Microcystis* (Nishizawa *et al.*, 1999, 2000; Tillett *et al.*, 2000), *Planktothrix* (Christiansen *et al.*, 2003) and *Anabaena* (Rouhiainen *et al.*, 2004). Insertional gene knockout experiments have demonstrated that all microcystin variants produced by a strain are synthesized by a single enzyme complex encoded by a 55 kb gene cluster (Christiansen *et al.*, 2003; Dittmann *et al.*, 1997; Nishizawa *et al.*, 1999).

The phylogenetic concurrence between a few housekeeping genes and microcystin synthetase genes from the same strains representing a wide selection of microcystin-producing genera suggests that *mcv* genes represent an

†Present address: Division of Pathology, Rikshospitalet University Hospital, 0027 Oslo, Norway.

Abbreviations: A, adenylation (domain); C, condensation (domain), ML, maximum likelihood; NJ, neighbour joining; NMT, *N*-methyltransferase (domain); SAM, S-adenosylmethionine.

The GenBank/EMBL/DDBJ accession numbers for the sequences determined in this work are shown in Table 1.

A supplementary table of primers and three supplementary figures are available with the online version of this paper.

ancestral state of many lineages (Rantala *et al.*, 2004). Thus, within lineages producing microcystin, non-producing strains would be the result of inactivation either by point mutation, or by partial or complete deletion of *mcy* genes (Christiansen *et al.*, 2006; Dittmann *et al.*, 1997; Kurmayer *et al.*, 2004; Mikalsen *et al.*, 2003).

One of the most intriguing features of microcystin is the high number of isoforms, often associated with changes in the synthetases as a result of mutations in the gene cluster (Kurmayer *et al.*, 2005; Mikalsen *et al.*, 2003). Several independent investigations on *Microcystis* and *Planktothrix* strains have concluded that the natural variation in the gene cluster within each strain is often caused by recombination. The regions with highest site variation seem to be *mcyA* (Kurmayer *et al.*, 2005; Tanabe *et al.*, 2004) and *mcyB* (Kurmayer & Gumpenberger, 2006; Mikalsen *et al.*, 2003). For example within *Planktothrix* spp., the typical N-methyltransferase (NMT)-containing adenylation (A) domain of the first module of McyA has been replaced by an A domain without NMT, leading to production of desmethyl⁷-microcystin isoforms (Kurmayer *et al.*, 2005). For *Microcystis* Mikalsen *et al.* (2003) showed that a recombination event between segments encoding two different adenylation domains has led to the presence of a second type of A domain in McyB1. Strains containing the Arg-activating A domain in McyB1 (hereafter denoted as C-like due to properties similar to the A domain of McyC) produce mainly microcystin-RR (indicating a microcystin with Arg in both the X and Z position), while strains with the Leu-activating A domain in McyB1 (hereafter denoted as B-like) produce mainly microcystin-LR.

In the previous study of Mikalsen *et al.* (2003) only a part of *mcyB* (A1) was analysed in detail. We therefore expanded the investigation using the same 17 field-collected strains to include the NMT domain of McyA and the A domain of McyC, as well as the flanking regions on both sides of the gene cluster. We employed a combination of alignments, a phylogenetic approach using split decomposition analysis and various statistical software (GENECONV, RPD and MaxChi) to address the genomic processes leading to new microcystin gene cluster variants (and thus to new *Microcystis* chemotypes, or inactivation of the cluster), the genetic basis for desmethyl⁷-microcystin variants and the recombination events involving *mcyB1* and *mcyC*.

METHODS

Bacterial strains and DNA isolation. The *Microcystis* strains used in this study (Table 1) have previously been genetically typed using RAPD and REP fingerprinting (Mikalsen *et al.* 2003). Unialgal cultures were grown at the Norwegian Institute of Water Research (NIVA) as previously described (Skulberg & Skulberg, 1990) except for strain PCC 7806, which was kindly provided by H. Utkilen (National Institute of Public Health, Norway). Genomic DNA for Southern blotting and PCR was isolated as described earlier (Mikalsen *et al.*, 2003).

Southern hybridization. Southern hybridization analyses were performed using probes derived from *dnaN*, *uma1* and *mcyA* sequences in *Microcystis aeruginosa* PCC 7806 (AF183408). Probe locations are shown in Fig. 1. Approximately 1 µg genomic DNA was digested overnight with 15 U *Hind*III (37 °C). Fragments were separated on a 1 % agarose gel and transferred to an Amersham Hybond-N membrane by Southern dry blotting overnight. Hybridization was performed at 68 °C by standard procedures (Galau *et al.*, 1986). Radiolabelled probes were generated from NMT-F/NMT-R, *dnaN*-F/*dnaN*-R and *uma1* F/*uma1*-R PCR products from strain PCC 7806 by standard MSPL (magnetic solid-phase labelling) random-primed synthesis (Espelund *et al.*, 1990) using the Random Primed Labelling kit (Roche Diagnostics).

PCR amplification and sequencing. Primers used for amplification of different regions of the *mcy* gene cluster are listed in Supplementary Table S1 (available with the online version of this paper) and their relative positions in the gene cluster are shown in Fig. 1. Primers were designed based on the publicly available *mcy* gene sequence of *M. aeruginosa* PCC 7806, accession no. AF183408 (Tillett *et al.*, 2000). The primer hypX (Nishizawa *et al.*, 2007) was used to investigate the genetic organization downstream of *uma1* in two non-toxic strains. BD Advantage 2 polymerase (BD Biosciences) was used in PCRs. Amplicons were purified using the E.Z.N.A Gel Extraction kit (Omega Biotek) and sequenced directly. DNA sequencing was performed on both strands with the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) with a capillary electrophoresis sequencer (ABI 3730). Sequences have been submitted to GenBank; accession numbers are shown in Table 1.

Phylogenetic and recombination analyses. Sequences were aligned with the publicly available *mcy* sequence of *M. aeruginosa* PCC 7806 (accession no. AF183408) using CLUSTAL W (Thompson *et al.*, 1997) and manual editing. The sequences were used to infer the phylogeny in a Bayesian framework applying the program MrBayes v3.1 (Ronquist & Huelsenbeck, 2003). Analysis was performed with the following parameters: GTR model, gamma distribution, running 2 million generations and sampling trees every 100 generation, burn-in 3000 trees. The maximum-likelihood (ML) tree was estimated for *mcyB1* and *mcyC* A domain sequences using PhyML (Guindon & Gascuel, 2003) under the GTR model, gamma distribution and with parameter values indicated by MrmodelTest (Nylander, 2004). The neighbour joining (NJ) tree was obtained under the default nucleotide substitution model using MEGA version 3.1 (Kumar *et al.*, 2004). Mrmodeltest, PhyML and MrBayes analyses were performed on the Bioportal computer platform resources (<http://www.bioportal.uio.no>). Recombination was detected by split decomposition analysis using SplitsTree version 4.8 (Huson & Bryant, 2006) with default settings (uncorrectedP method) and 1000 bootstrap replicas, and Phi test for recombination (implemented to test split decomposition analysis reliability) (Bruen *et al.*, 2006). In addition, statistical tests for detecting recombination were used: GENECONV (Padidam *et al.*, 1999), RDP and MaxChi (Martin *et al.*, 2005) analyses in the RDP version 2.608 program package. For GENECONV, G-scale values 0 and 1 were used, for detecting recent and older recombination events, respectively. Recombination was also investigated by visual analysis of informative sites (variable sites where each variant occurs in at least two strains).

RESULTS

Sequence variation in the NMT domain encoded by *mcyA*

Southern hybridization analysis with a probe derived from the NMT region of *mcyA* (Fig. 1) was used to investigate

Table 1. *Microcystis* strains investigated

<i>Microcystis</i> strain*	Geographical origin	Accession no. (PCR product, bp)					
		<i>dnaN-mcyJ</i>	<i>mcyA</i>	<i>mcyB</i>	<i>mcyC</i>	<i>mcyC-umaI</i>	<i>dnaN/ftsH/hypX-umaI</i>
N-C 31 <i>M. aeruginosa</i>	Little Rideau Lake, Canada	EF115379 (1009)	EF115388 (1491)	EU009866 (1406)	EF115396 (1063)	EF115405 (515)	–
N-C 43 <i>M. aeruginosa</i>	Wisconsin, USA (ATCC 22663)	–†	–†	–†	–†	–†	EU475874‡ (668)
N-C 57 <i>M. aeruginosa</i>	L. Frøylandsvatnet, Norway	EF115380 (1011)	EF115389 (1491)	EU009867 (1406)	EF115397 (1063)	EF115406 (515)	–
N-C 118/2 <i>M. sp.</i>	L. Gjersjøen, Norway	EF115381 (1062)	EF115390 (1491)	EU009868 (1406)	EF115398 (1060)	EF115407 (503)	–
N-C 122/2 <i>M. viridis</i>	L. Finjasjön, Sweden	–†	–†	–†	–†	–†	–
N-C 123/1 <i>M. aeruginosa</i>	L. Mälaren, Sweden	–†	–†	–†	–†	–†	EF115424§ (756)
N-C 143 <i>M. aeruginosa</i>	L. Akersvatnet, Norway	EF115382 (2165)	–¶	EU009870 (1406)	EF115399 (1063)	EF115408 (515)	–
N-C 144 <i>M. cf. flos-aquae</i>	L. Borrevatnet, Norway	–†	–†	–†	–†	–†	EF115425§ (756)
N-C 161/1 <i>M. botrys</i>	L. Mosvatnet, Norway	EF115383 (775)	EF115391 (1491)	EU009872 (1406)	EF115400 (1063)	EF115409 (515)	–
N-C 166 <i>M. aeruginosa</i>	L. Hellesjøvatnet, Norway	–†	–†	–†	–†	–†	EF115427# (1144)
N-C 169/7 <i>M. viridis</i>	L. Arresø, Denmark	EF115384 (879)	EF115392 (1491)	EU009873 (1406)	EF115401 (1063)	EF115410 (515)	–
N-C 172/5 <i>M. cf. wesenbergii</i>	L. Arresø, Denmark	–†	–†	–†	–†	–†	EF115428# (1144)
N-C 228/1 <i>M. aeruginosa</i>	L. Akersvatnet, Norway	EF115385 (2605)	EF115393 (1491)	EU009875 (1406)	EF115402 (1063)	EF115411 (515)	–
N-C 264 <i>M. botrys</i>	L. Frøylandsvatnet, Norway	EF115386 (989)	EF115394 (1491)	EU009876 (1409)	EF115403 (1063)	EF115412 (515)	–
N-C 279 <i>M. cf. ichthyoblabe</i>	L. Østensjøvatnet, Norway	–†	–†	–†	–†	–†	EF115426§ (756)
N-C 324/1 <i>M. sp.</i>	L. Tøråssjøen, Norway	EF115387 (774)	EF115395 (1491)	EU009877 (1406)	EF115404 (1063)	EF115413 (515)	–
PCC 7806 <i>M. aeruginosa</i>	Braakman Reservoir, The Netherlands	AF183408 (1009)	AF183408 (1491)	AF183408 (1406)	AF183408 (1063)	AF183408 (515)	–

*N-C, NIVA-CYA, Norwegian Institute for Water Research Cyanobacterial Culture Collection; PCC, Pasteur Culture Collection. The NIVA-CYA strains have been genetically typed using RAPD and REP fingerprinting by Mikalsen *et al.* (2003). There is no evidence for clonality among the strains.

†The strain does not possess *mcy* genes.

‡PCR product contains *hypX* and *umaI* segments.

§PCR product contains *dnaN* and *umaI* segments.

||PCR product *dnaN-mcyA* due to partial loss of *mcy* operon in this strain.

¶The strain does not possess the NMT-domain-encoding part of *mcyA*.

#PCR product contains *ftsH* and *umaI* segments.

Fig. 1. The microcystin synthetase gene cluster in *M. aeruginosa* PCC 7806. The relative positions of PCR primers used to amplify flanking regions of *mcy* gene clusters and parts of *mcyA*, *mcyB* and *mcyC* are shown. The positions of the probes used for Southern hybridization are indicated.

the presence of this domain in 17 *Microcystis* strains (see Table 1 for description of the strains). Positive hybridization signals were obtained for all nine toxic strains (Fig. 2A), including the strains producing only desmethyl⁷-microcystin isoforms (N-C 57, N-C 228/1 and N-C 264) (Mikalsen *et al.*, 2003). The same *Hind*III restriction pattern was obtained for all strains except N-C 118/2. No hybridization signal was observed for the non-toxic strain N-C 143, which possesses *mcy* genes, but is deficient in microcystin production (Mikalsen *et al.*, 2003).

Using an NMT-specific primer pair (see Fig. 1), PCR products were obtained from all toxic *Microcystis* strains, but not from strain N-C 143 (Table 1), thus confirming the Southern hybridization results. Sequencing of the NMT region (see Fig. 1) showed that strain N-C 118/2 lacked the *Hind*III restriction site present in the other hybridizing strains, thus explaining the polymorphism seen for this strain.

The amino acid residues constituting the *S*-adenosyl-methionine (SAM)-binding site of cyclosporin synthetase NMT domains (Velkov & Lawen, 2003) were identified in the deduced amino acid sequences of the NMT domains. The NMT domain sequence from *Microcystis* strain K-139 (AB019578), which exclusively produces [Dha⁷]-MC, was also included in the NMT dataset. The analysis revealed that invariant amino acid residues in the SAM-binding site (indicated in Fig. 3A) were intact in all strains producing only [Mdha⁷]-MC. In all methylation-defective strains one or more of these amino acid residues were altered (Fig. 3A). Interestingly, in strain N-C 324/1, which produces both [Mdha⁷]-MC and [Dha⁷]-MC (Mikalsen *et al.*, 2003), the invariant Gln residue in the SAM-binding site was replaced by Arg (Fig. 3A).

Apart from these functionally relevant differences, the NMT sequences showed relatively little variation (1–7 %) between strains (as might be anticipated from the almost identical Southern blots). However, an alignment of the 55 informative sites present showed a block-like structure with respect to phylogenetic affiliation. For all sequences one or more blocks showed similarity to different subsets of the strains examined, i.e. a mosaic pattern (Supplementary

Fig. S1A). Such block differences were originally demonstrated for the *rbcLX* operon in cyanobacteria; they were attributed to horizontal (lateral) gene transfer and denoted mosaic structures (Rudi *et al.*, 1998) – a term we will use in the following. The split decomposition analysis revealed a reticulate phylogeny (Fig. 3B) and the Phi test found statistically significant evidence for recombination ($P < 0.01$). All three recombination detection programs used suggested recombination events in this dataset (Table 2).

Adenylation domains in modules McyB1 and McyC

Using the *mcyB*-specific primer pair, a 1406/1409 bp segment coding for almost the entire A domain of McyB1 was amplified and sequenced from all toxin-producing strains and from strain N-C 143 (Table 1). An alignment of the sequences was used to identify the 3' breakpoint of the recombination event involving the A-domain-encoding region of *mcyB* reported by Mikalsen *et al.* (2003). The 5' breakpoint of this recombination event is located near the conserved motif A3 (Mikalsen *et al.*, 2003). Recombination detection programs suggested a 3' breakpoint near the conserved motif A8 (at position 1209 in nucleotide alignment) (Table 2) in strain N-C 31. Visual inspection of amino acid alignment revealed a putative recombination breakpoint near the conserved motif A9 in strains N-C 118/2, N-C 161/1 and PCC 7806 (Fig. 4). This breakpoint could not be detected by Mikalsen *et al.* (2003) due to the shorter McyB1 sequences studied in that work.

Aligning the four B-like nucleotide sequences revealed 3–6 % sequence variation. The 16 informative sites found in the alignment revealed a mosaic structure (Supplementary Fig. S1B). A reticulate phylogeny was obtained by split decomposition analysis (Fig. 5) and the Phi test found statistically significant evidence for recombination ($P < 0.01$). However, no recombination events received statistical support by the recombination detection programs (Table 2). The C-like McyB1 A domain nucleotide sequences showed 2–5 % sequence variation. The split decomposition analysis revealed a reticulate phylogeny

Fig. 2. Southern blot analyses of the *mcy* gene cluster and flanking region. The relative positions of primers, Southern probes and *Hind*III restriction sites (H) are shown. Genomic DNA was digested with *Hind*III and hybridized with the *dnaN* and *uma1* probes. Strains containing *mcy* genes (Mikalsen *et al.*, 2003) are indicated in bold. (A) Organization in strain PCC 7806 of the region in *mcyA* that contains the NMT domain and Southern blot analysis of the corresponding region in various strains. The A, NMT and C domains are indicated. Strains producing desmethyl⁷-microcystin isoforms are indicated by black dots. (B) Organization of the upstream region flanking the *mcy* gene cluster in strain PCC 7806 and Southern blot analysis of this region in various strains. (C) Organization of the downstream region flanking the *mcy* gene cluster in strain PCC 7806 and Southern blot analysis of this region in various strains.

Fig. 3. Region of *mcyA* encoding the NMT domain in *Microcystis* strains. Strains producing desmethyl⁷-microcystin isoforms are indicated by black dots. (A) Amino acid sequences of NMT domains. Amino acid residues constituting the SAM-binding-site according to Velkov & Lawen (2003) are indicated. Function of residues altered in non-methylating strains: Gly (G) residues form hydrogen bonds with the Met moiety of SAM. Acidic residue (D/E) forms hydrogen bonds with the adenine ribose. Gln (Q) forms hydrogen bonds with Met (Velkov & Lawen, 2003). (B) Split decomposition analysis of NMT-domain sequences constructed using SplitsTree4 at default setting, showing 1000 bootstrap replicas above 50%.

(Fig. 5) and statistically significant evidence for recombination was found by the Phi test ($P < 0.01$). Recombination was also detected by the recombination detection programs (Table 2) and suggested by the mosaic structure of informative sites (Supplementary Fig. S1C).

The A-domain-encoding region of *mcyC* was successfully amplified from all toxic strains as well as the non-toxic strain N-C 143 (Table 1). The sequences showed 0–5% sequence variation. Multiple recombination events were detected in this dataset by the mosaic structure of informative sites (Supplementary Fig. S1C); the reticulate phylogeny revealed by split decomposition (Fig. 5) was supported by the Phi test ($P < 0.01$) and all recombination detection programs (Table 2).

A comparison of *mcyB* and *mcyC* A-domain-encoding regions from the same strain revealed relatively low sequence variation (9–13%) in strains possessing the C-like A domain in McyB1. In strains with a B-like A domain in McyB1, these sequences were rather divergent, with 39–40% sequence variation. Split decomposition analysis including all *mcyB* and *mcyC* sequences revealed a reticulate phylogeny within, but not between the clusters ($P < 0.01$) (Fig. 5). The 161 informative sites found in the 1063 bp alignment of *mcyB* (C-like) and *mcyC* sequences showed a mosaic structure, indicating recombination also between these regions (Supplementary Fig. S1C). Putative recombination events involving the A domain regions of *mcyB* (C-like) and *mcyC* identified by the recombination

detection programs are listed in Table 2. The phylogenetic analyses of all A domain amino acid sequences revealed three clades: McyB1 (B-like), McyB1 (C-like) and McyC (Supplementary Fig. S2).

Regions flanking the *mcy* gene cluster

Probes derived from the genes flanking the *mcy* gene cluster in strain PCC 7806, *dnaN* and *uma1* (Fig. 1) (Tillett *et al.*, 2000), were used for Southern hybridization analysis of flanking regions. All 17 strains gave positive hybridization signals with both probes (Fig. 2B, C). For the *uma1* probe, all *mcy*-containing strains (Mikalsen *et al.*, 2003) gave a 4.5 kb *HindIII* fragment similar to that from strain PCC 7806. The *dnaN* probe displayed a far more variable restriction pattern, with few shared bands between the strains.

The primer pairs *dnaN*-*mcyJ* and *mcyC*-*uma1* (Fig. 1 and Supplementary Table S1) amplified flanking regions from all toxic strains. The *mcy*-containing non-toxic strain N-C 143 gave a PCR product with the *mcyC*-*uma1* combination but not with the *dnaN*-*mcyJ* primer pair. The region between *dnaN* and *mcyB* could however be amplified using the *dnaN*-676R (*mcyB*-located) primer pair. Sequencing revealed a deletion of the whole *mcyD–J* gene cluster and a large part of *mcyA* (everything before position 43852 in AF183408). Comparison of DNA sequences from the toxic strains showed a high degree of conservation of all coding regions. A 90 bp intergenic region between *mcyC* and

Table 2. Putative recombination events detected by RPD, GENECONV and MaxChi2 (events detected by two or more programs are listed)

Strains involved [#]	RDP*	GENECONV*		MaxChi*
		(G=1)	(G=0)	
	Fragment†, <i>P</i> -value	Fragment†, <i>P</i> -value	Fragment†, <i>P</i> -value	Fragment†, <i>P</i> -value
Putative recombination events within NMT-domain-encoding sequences of <i>mcvA</i>				
N-C 228/1	1–399, 0.001	1–440, <0.001	1–440, 0.002	15–411, <0.001
N-C 161/1				
N-C 264	406–1129, 0.006	–	–	492–1142, <0.001
N-C 169/7				
N-C 264	295–1129, 0.045	–	–	15–1128, <0.001
N-C 324				
N-C 264	–	–	1130–1490, <0.001	825–1470, <0.001
N-C 31				
N-C 264	–	517–1082, <0.001	–	517–1082, <0.001
N-C 161/1				
N-C 161/1	295–479, 0.016	–	–	295–479, 0.008
N-C 264				
Putative recombination events within A-domain-encoding sequences of <i>mcvB1</i>				
N-C 143	1–254, <0.001	1–302, <0.001	–	–
N-C 31				
N-C 264, 324/1, 169/7, 57	1–486, <0.001	169–491, 0.006	–	–
N-C 31, 118/2, 161/1, PCC 7806				
N-C 143	1209–1405, <0.001	1243–1405, <0.001	1243–1405, <0.001	1209–1404, <0.001
N-C 31				
Putative recombination events within <i>mcvB1</i> sequences encoding C-like adenylation domain‡				
N-C 143	33–953, 0.001	–	–	33–953, <0.001
N-C 57				
N-C 57 (143, 228/1)	1062–1362, 0.004	1066–1338, 0.004	–	1068–1404, <0.001
N-C 169/7				
N-C 57	33–633, 0.002	–	–	33–633, <0.001
N-C 228/1				
Putative recombination events within A-domain-encoding <i>mcvC</i> sequences				
N-C 118/2	36–169, 0.009	–	–	15–246, <0.001
N-C 264				
PCC 7806	1–203, <0.001	1–246, 0.002	–	–
N-C 228/1				
N-C 161/1	–	–	43–454, 0.004	111–417, <0.001
N-C 264				
Putative recombination events between sequences encoding A domains of McyB1 (C-like) and McyC				
McyB1 N-C 264	297–996, <0.001	262–1062, <0.001	–	249–1014, <0.001
McyC N-C 31 (57, 143, 161/1, 228/1, 264, 324/1)				
McyB1 N-C 169/7	876–1029, <0.001	–	887–1013, <0.001	707–1056, <0.001
McyC N-C 31 (57, 161/1, 169/7)				
McyB1 143	948–1062, 0.006	–	934–1055, 0.006	–
McyC N-C 161/1				

[#]Pairs of strains analysed for recombination.

*Reference to programs (see Methods).

†Relative position of the recombining regions in our alignments. The alignment of NMT-domain-encoding *mcvA* sequences covered positions 38 722–40 212 in accession no. AF183408 from strain PCC 7806. The alignments of A-domain-encoding sequences in *mcvB1* covered positions 46 789–47 834 in accession no. AF183408 from strain PCC 7806. The alignment of A-domain-encoding sequences in *mcvC* covered positions 53 181–54 242 in accession no. AF183408 from strain PCC 7806.‡No recombination events were detected within sequences encoding the B-like A domain of *McvB1*.

Fig. 4. Alignment of the A domain amino acid sequences in the first module of McyB. Conserved motifs are underlined and genetic variants (B-like and C-like) indicated to the right of the alignment.

umal was also conserved in all strains except strain N-C 118/2, which has a shorter (80 bp) region with no apparent similarity to the intergenic regions from the other strains (Fig. 6A). In contrast, the *dnaN-mcyJ* intergenic regions were highly different among the strains, with regard to

both length (from 360 to 2192 bp) and sequence. Based on the alignments, the region was divided into segments A–D (Fig. 6A and Supplementary Fig. S3) that were assigned to subgroups containing similar sequences (colour coded in Fig. 6A). The *dnaN-mcyJ* sequence from strain N-C 118/2

Fig. 5. Split decomposition analysis of gene sequences encoding A domains in modules McyB1 and McyC constructed using SplitsTree4 at default setting, showing 1000 bootstrap replicas above 50%.

was at the *dnaN* end highly similar to the corresponding sequences from strains N-C 161/1, N-C 169/7, N-C 264 and N-C 324/1, while it was most similar to the sequences from strains N-C 228/1, N-C 31, N-C 57 and PCC 7806 at the *mcyJ* end, indicating recombination.

A 1585 bp insertion containing an IS element with similarity to Tn5-like transposases from *Gloeobacter violaceus* PCC 7421 (NP923092) and bacteriophage WO (BAA89624) was found in strain N-C 228/1. This IS element (named IS*Mae7* and submitted to the ISfinder database) belongs to the IS4 family and consists of 7 bp direct repeats, 29 bp terminal inverted repeats (Fig. 6B) and a single 1407 bp ORF transcribed in the same direction as *mcyJ*. A GATC methylation site known to play a modulating role in transposition activity (Mahillon & Chandler, 1998) was found in the left inverted repeat. The transposase encoded by IS*Mae7* contains a DDE motif known to be necessary for efficient DNA transposition (Mahillon & Chandler, 1998) and has probably retained transposition activity, since no stop codons were present within the ORF. Direct repeats associated with IS*Mae7* were also found in the *dnaN*–*mcyJ* intergenic spacer in strains N-C 118/2 and N-C 143. Interestingly, another 16 bp direct repeat not associated with IS*Mae7* was present in *dnaN*–*mcyJ* intergenic sequences from several strains (Fig. 6B).

Phylogenetic analysis of the *mcy* flanking regions gave fairly similar trees, except for strain N-C 324/1, which clustered with different clades at the two flanks, and the clade including strains N-C 31, N-C 57, N-C 228/1 and PCC 7806, which showed incongruent phylogenies (Fig. 7). Split decomposition analysis indicated recombination events for

both flanks (Fig. 7), while the Phi test found statistically significant evidence for recombination only in the *dnaN*–*mcyJ* alignment ($P < 0.01$).

PCR with the *dnaN*–*uma1* primer pair gave PCR products from five of seven non-toxic strains investigated (Table 1). PCR with the *hypX* (Nishizawa *et al.*, 2007) and *uma1* primer pair gave a PCR product for one of the two remaining non-toxic strains, N-C 43 (Table 1). In three strains (N-C 123/1, N-C 144 and N-C 279), *dnaN* was found to be situated close to *uma1*. The intergenic region between these genes showed no similarity to the intergenic regions in the toxic strains, except for a 80 bp segment close to *uma1* that was similar to the region between *mcyC* and *uma1* in strain N-C 118/2 (Fig. 7A). In three other non-toxic *Microcystis* strains, the genetic organization downstream of *uma1* was different: in strains N-C 166 and N-C 172/5, an ORF encoding a protein homologous to FtsH (Kaneko *et al.*, 1995) and in strain N-C 43, a hypothetical protein (HypX) (Nishizawa *et al.*, 2007) was found (Fig. 6A). The flanking regions of strain N-C 122/2 could not be amplified using these primer pairs, indicating differences in the flanking regions of its *mcy* gene cluster, and therefore this strain is not included in Fig. 6(A).

DISCUSSION

Loss of *N*-methyltransferase activity is associated with specific point mutations in *mcyA* in *Microcystis*

Deletion of the entire NMT domain in *McyA* has been reported to cause production of desmethyl⁷-microcystin in

some strains of *Planktothrix* (Kurmayer *et al.*, 2005) and *Anabaena* (Fewer *et al.*, 2008). Our results indicate that the production of desmethyl⁷-microcystin by five of our *Microcystis* strains is not due to deletion of the NMT domain, but rather to specific point mutations altering amino acid residues in the cofactor SAM-binding site. Not all mutations were associated with total inactivation of the NMT domain, as shown for strain N-C 324/1, where replacement of invariant Gln with Arg (Fig. 3A) was associated with co-production of methylated and non-methylated microcystin isoforms (Mikalsen *et al.*, 2003). The replacement of the second conserved Gly residue (as in strain N-C 57, Fig. 3A) has previously been shown to inactivate the NMT domain of pyochelin synthetase in *Pseudomonas aeruginosa* (Patel & Walsh, 2001). The effect of the remaining amino acid residues in the SAM-binding site has not been verified by mutations. However, residues which form the SAM-binding site (Velkov & Lawen, 2003) were found to be replaced only in *Microcystis* strains producing desmethyl⁷-microcystin, and always with residues with different physicochemical properties (Fig. 3A). Although the effects of these replacements on the methyltransferase activity have not yet been demonstrated, the results suggest that the NMT domain may be inactivated by point mutations, in addition to the complete deletion of this domain described in other nonmethylating strains (Fewer *et al.*, 2008; Kurmayer *et al.*, 2005).

Inter- and intragenomic recombinations within and between domains

Previously, we have shown recombination involving the A domain of McyB1 resulting in altered substrate specificity (Mikalsen *et al.*, 2003). Here we show that this recombination included the A domain segment between the conserved motifs A3 and A8/A9 (Table 2, Fig. 4). Notably, Fewer *et al.* (2007) have described a similar recombination in *Anabaena* spp. and *Hapalosiphon hibernicus*, involving only the A domain regions of *mcyB1* and *mcyC* and not affecting the condensation (C) domains in these modules. In both the *Planktothrix* and the Nostocales (*Anabaena* and *Hapalosiphon*) case the recombinations involve larger, non-identical functional units (different A domains) and affect essentially an entire domain. For *Microcystis* the 5' recombination breakpoint (near the conserved motif A3) is the same in all strains, but the downstream breakpoint has two different locations (near conserved motifs A8 in strain N-C 31 and A9 in strains N-C 118/2, N-C 161/1 and PCC 7806), separated by about 120 bp (Fig. 4). This might indicate two independent recombination events. However, as seen from the high similarity between the conserved A8 and A9 regions of McyB1 and McyC in strains PCC 7806, K-139 and UV027 (AP183408, AB019578 and AP458094, respectively), the variable location of the 3' breakpoint may alternatively be the result of a recombination event between two different A domain sequences followed by intragenomic or intergenomic recombinations spanning the regions encoding the conserved motifs A8 and A9.

Our data also indicate several recombination events between the A domain regions of *mcyB* (C-like) and *mcyC*, covering almost the whole segment between the conserved motifs A3 and A8 (as in strain N-C 264) or shorter segments of the A domain (Supplementary Fig. S1C and Table 2) and mainly resulting in the replacement of McyB1 A domain segments with corresponding McyC segments. Recombination between these two A-domain-encoding gene regions has been shown also for cyanobacteria from other genera (Fewer *et al.*, 2007; A. Tooming-Klunderud, D. P. Fewer, T. Rohrlack, J. Jokela, L. Rouhiainen, K. Sivonen, T. Kristensen & K. S. Jakobsen, unpublished).

Our data suggest multiple recombination events involving smaller segments within all examined domains of *mcyABC* operon. These recombinations lead to a mosaic pattern of phylogenetic affinities in the alignments covering nearly the entire analysed sequence of the NMT domain and the McyB1A domain. As in other studies of the NMT domain of *Microcystis* (Tanabe *et al.*, 2004) and the McyB1A domain of *Planktothrix* (Kurmayer *et al.*, 2005; Kurmayer & Gumpenberger, 2006) many of these recombinations do not change the amino acid sequence. Within the A domain of McyC, however, such a pattern was mainly found upstream of the conserved motif A3 (Table 2). The apparent lack of recombination in the region coding for the substrate-binding part of this domain may be explained by a more restricted function of the Arg-activating McyC A domain compared to the A domain of McyB1, which will activate several amino acids (Mikalsen *et al.*, 2003).

Evidence indicates that all microcystin variants are synthesized by a single enzyme complex. Given a single *mcy* operon in each strain, the observed recombinations (Kurmayer & Gumpenberger, 2006; Tanabe *et al.*, 2004; present study) most likely represent intergenomic processes and are thus manifestations of horizontal gene transfer (HGT) between strains. The frequent HGT between closely related cyanobacterial strains was originally suggested by Rudi *et al.* (1998) and recently reinforced by whole-genome analysis of cyanobacteria (Zhaxybayeva *et al.*, 2006) and in a general bacterial context (Papke *et al.*, 2007). HGT may be the result of uptake of free DNA through natural transformation, since *Microcystis* is naturally competent (Dittmann *et al.*, 1997), or DNA shuttling catalysed by cyanophages through transduction, or possibly due to other transposable elements.

The *mcy* flanking sequences do not indicate frequent transfer of complete *mcy* gene clusters between *Microcystis* strains

In all microcystin-producing strains investigated here the same two genes (*dnaN* and *umaI*) flank the *mcy* operon, in agreement with Tillett *et al.* (2001), who previously have shown that *umaI* is located upstream of *mcyC* in toxic *Microcystis* strains. Thus, it seems likely that the genomic

Fig. 6. Regions flanking the *mcy* operon in *Microcystis* strains. (A) A schematic overview of both flanks. In strains containing *mcy* genes, the intergenic region between *dnaN* and *mcyJ* was divided into segments based on DNA sequence similarity (Supplementary Fig. S2). For each segment, sequences marked with the same colours are highly similar, while unique colours indicate different sequences. Strains producing microcystins are indicated by red diamonds. *For strains N-C 166 and N-C 172/5, the PCR product contained the *uma1*-R primer in both ends. (B) Parts of the sequences of the *dnaN*–*mcyJ* intergenic region in strains containing *mcy* genes. Inverted repeats in the IS element present in strain N-C 228/1 (IS*Mae7*) and direct repeated sequences (DR) generated on insertion are indicated. Directly repeated sequences not associated with IS*Mae7*, present in some strains, are also indicated.

location of the *mcy* gene cluster in *Microcystis* spp. is the same in all strains examined, in agreement with another recent study (Nishizawa *et al.*, 2007). These findings do not support the idea of frequent transfer of complete *mcy* gene clusters between *Microcystis* strains. The same two genes (*dnaN* and *uma1*) were also found to be neighbours in three of the non-toxic *Microcystis* strains, as also reported by Nishizawa *et al.* (2007), indicating that loss of the *mcy* gene cluster was not accompanied by further rearrangements in this genomic region. Partial loss of the *mcy* gene cluster in strain N-C 143 clearly illustrates that loss of microcystin production can be caused by deletion (Fig. 6A).

A gene encoding a putative transposase (*uma4*) has previously been identified in the *mcy* region of *Microcystis* strain PCC 7806 (Tillett *et al.*, 2000). Here we

report the presence of an IS element (IS*Mae7*) in the intergenic region between *mcyJ* and *dnaN* in strain N-C 228/1 (Fig. 6). Recently, two other genes coding for different types of transposases have been identified between *dnaN* and *mcyJ* in several *Microcystis* strains (Nishizawa *et al.*, 2007), but none of these reported transposases are similar to the one encoded by *uma4*. The additional direct repeats detected by us in this intergenic region in several strains may also indicate an additional, now lost IS element. According to these results, the intergenic region between *dnaN* and *mcyJ* seems to be a recombinatorial ‘hot spot’ while the *mcyC*–*uma1* region is more conserved.

The frequent recombinations – representing both intragenomic and intergenomic rearrangements – give rise to novel *mcy* gene cluster variants, most of which encode

Fig. 7. Phylogenetic and split-decomposition analysis of regions flanking the *mcy* gene cluster in *Microcystis* spp. Both NJ and Bayesian trees revealed the same topology (NJ tree shown). Bootstrap values (1000 replicates) and posterior probabilities are indicated at the nodes. Strain N-C 324/1, for which the two flanks cluster with different clades, is marked by a grey box. (A) Intergenic region between *dnaN* and *mcyJ*. (B) Intergenic region between *uma1* and *mcyC*.

peptide synthetases with the same biosynthetic properties, due to a constant selective pressure. Changes in selection will, however, favour some variants within a particular ecosystem. The results presented here along with some previous studies (Kurmayer *et al.*, 2005; Kurmayer & Gumpenberger, 2006; Mikalsen *et al.*, 2003; Tanabe *et al.*, 2004) suggest that the function of microcystins needs to be understood within the frame of the natural ecosystems of the bacterial strains. It seems likely that the processes promoting variation within the microcystin synthetase genes are crucial for adaptivity of a given population within a particular habitat.

ACKNOWLEDGEMENTS

We thank Randi Skulberg for providing the N-C strains and Hans Utkilen for providing the PCC 7806 strain. We also thank Trine B. Rounge and Thomas Rohrlack for fruitful discussions. This work was supported by the Research Council of Norway by a grant 157338/140 to K. S. J.

REFERENCES

- Bruen, T. C., Philippe, H. & Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665–2681.
- Christiansen, G., Fastner, J., Erhard, M., Börner, T. & Dittmann, E. (2003). Microcystin biosynthesis in *Planktothrix*: genes, evolution, and manipulation. *J Bacteriol* **185**, 564–572.
- Christiansen, G., Kurmayer, R., Liu, Q. & Börner, T. (2006). Transposons inactivate biosynthesis of the nonribosomal peptide microcystin in naturally occurring *Planktothrix* spp. *Appl Environ Microbiol* **72**, 117–123.
- Dittmann, E., Neilan, B. A., Erhard, M., von Döhren, H. & Börner, T. (1997). Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. *Mol Microbiol* **26**, 779–787.
- Espelund, M., Stacy, R. A. & Jakobsen, K. S. (1990). A simple method for generating single-stranded DNA probes labeled to high activities. *Nucleic Acids Res* **18**, 6157–6158.
- Fewer, D. P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H. & Sivonen, K. (2007). Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. *BMC Evol Biol* **7**, 183.
- Fewer, D. P., Tooming-Klunderud, A., Jokela, J., Wahlsten, M., Rouhiainen, L., Kristensen, T., Rohrlack, T., Jakobsen, K. S. & Sivonen, K. (2008). Natural occurrence of microcystin synthetase deletion mutants capable of producing microcystins in strains of the genus *Anabaena* (Cyanobacteria). *Microbiology* **154**, 1007–1014.
- Galau, G. A., Hughes, D. W. & Dure, L., III (1986). Absciscic acid induction of cloned cotton late embryogenesis-abundant (Lea) mRNAs. *Plant Mol Biol* **7**, 155–177.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267.
- Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. (1995). Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64 % to 92 % of the genome. *DNA Res* **2**, 153–166.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5**, 150–163.
- Kurmayer, R. & Gumpenberger, M. (2006). Diversity of microcystin genotypes among populations of the filamentous cyanobacteria *Planktothrix rubescens* and *Planktothrix agardhii*. *Mol Ecol* **15**, 3849–3861.
- Kurmayer, R., Christiansen, G., Fastner, J. & Börner, T. (2004). Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium *Planktothrix* spp. *Environ Microbiol* **6**, 831–841.
- Kurmayer, R., Christiansen, G., Gumpenberger, M. & Fastner, J. (2005). Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*. *Microbiology* **151**, 1525–1533.
- Mahillon, J. & Chandler, M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725–774.
- Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. (1997). Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* **97**, 2651–2674.
- Martin, D. P., Williamson, C. & Posada, D. (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260–262.
- Mikalsen, B., Boison, G., Skulberg, O. M., Fastner, J., Davies, W., Gabrielsen, T. M., Rudi, K. & Jakobsen, K. S. (2003). Natural variation in the microcystin synthetase operon *mcvABC* and impact on microcystin production in *Microcystis* strains. *J Bacteriol* **185**, 2774–2785.
- Nishizawa, T., Asayama, M., Fujii, K., Harada, K. & Shirai, M. (1999). Genetic analysis of the peptide synthetase genes for a cyclic heptapeptide microcystin in *Microcystis* spp. *J Biochem* **126**, 520–529.
- Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K., Ochi, K. & Shirai, M. (2000). Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin. *J Biochem* **127**, 779–789.
- Nishizawa, T., Nishizawa, A., Asayama, M., Harada, K. & Shirai, M. (2007). Diversity within the microcystin biosynthetic gene clusters among the genus *Microcystis*. *Microbes Environ* **22**, 380–390.
- Nylander, J. A. A. (2004). *MrModeltest v2*. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Padidam, M., Sawyer, S. & Fauquet, C. M. (1999). Possible emergence of new geminiviruses by frequent recombination. *Virology* **265**, 218–225.
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D. & Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proc Natl Acad Sci U S A* **104**, 14092–14097.
- Patel, H. M. & Walsh, C. T. (2001). *In vitro* reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry* **40**, 9023–9031.
- Rantala, A., Fewer, D. P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Börner, T. & Sivonen, K. (2004). Phylogenetic evidence for the early evolution of microcystin synthesis. *Proc Natl Acad Sci U S A* **101**, 568–573.
- Ronquist, F. & Huelsenbeck, J. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.
- Rouhiainen, L., Vakkilainen, T., Siemer, B. L., Buikema, W., Haselkorn, R. & Sivonen, K. (2004). Genes coding for hepatotoxic

- heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. *Appl Environ Microbiol* **70**, 686–692.
- Rudi, K., Skulberg, O. M. & Jakobsen, K. S. (1998). Evolution of cyanobacteria by exchange of genetic material among phylogenically related strains. *J Bacteriol* **180**, 3453–3461.
- Skulberg, R. & Skulberg, O. M. (1990). *Research with Algal Cultures*. NIVA's Culture Collection of Algae. Oslo: Norway.
- Tanabe, Y., Kaya, K. & Watanabe, M. M. (2004). Evidence for recombination in the microcystin synthetase (*mcy*) genes of toxic cyanobacteria *Microcystis* spp. *J Mol Evol* **58**, 633–641.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- 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. *Chem Biol* **7**, 753–764.
- Tillett, D., Parker, D. L. & Neilan, B. A. (2001). Detection of toxigenicity by a probe for the microcystin synthetase A gene (*mcyA*) of the cyanobacterial genus *Microcystis*: comparison of toxicities with 16S rRNA and phycocyanin operon (Phycocyanin Intergenic Spacer) phylogenies. *Appl Environ Microbiol* **67**, 2810–2818.
- Velkov, T. & Lawen, A. (2003). Mapping and molecular modeling of S-adenosyl-L-methionine binding sites in N-methyltransferase domains of the multifunctional polypeptide cyclosporin synthetase. *J Biol Chem* **278**, 1137–1148.
- Welker, M. & von Döhren, H. (2006). Cyanobacterial peptides – nature's own combinatorial biosynthesis. *FEMS Microbiol Rev* **30**, 530–563.
- Zhaxybayeva, O., Gogarten, J. P., Charlebois, R. L., Doolittle, W. F. & Papke, R. T. (2006). Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res* **16**, 1099–1108.

Edited by: K. Forchhammer