Multilocus sequence typing (MLST) reveals high genetic diversity and clonal population structure of the toxic cyanobacterium Microcystis aeruginosa

Yuuhioko Tanabe,1 Fumie Kasai1 and Makoto M. Watanabe2

1Environmental Biology Division, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki 305-8506, Japan
2Graduate School of Life and Environmental Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8572, Japan

Microcystis aeruginosa is one of the most prevalent bloom-forming cyanobacteria and has been the cause of increasing public health concern due to the production of hepatotoxins (microcystins). To investigate the genetic diversity, clonality and evolutionary genetic background with regard to the toxicity of M. aeruginosa, a multilocus sequence typing (MLST) scheme was developed, based on seven selected housekeeping loci (ftsZ, glnA, gltX, gyrB, pgi, recA and tpi). Analysis of a collection of 164 isolates from Japan and other countries identified 79 unique sequence types (STs), revealing a high level of genetic diversity (H=0.951). Although recombination between loci was indicated to be substantial by Shimodaira–Hasegawa (SH) tests, multilocus linkage disequilibrium analyses indicated that recombination between strains probably occurs at some frequency but not to the extent at which alleles are associated randomly, suggesting that the population structure of M. aeruginosa is clonal. Analysis of subsets of strains also indicated that the clonal population structure is maintained even in a local population. Phylogenetic analysis based on the concatenated sequences of seven MLST loci demonstrated that microcystin-producing genotypes are not monophyletic, providing further evidence for the gain and loss of toxicity during the intraspecific diversification of M. aeruginosa. However, toxic strains are genetically distinct from non-toxic strains in MLST allelic profiles, and it was also shown that non-toxic strains harbouring toxin genes fall into a single monophyletic clade, except for one case. These results suggest that the toxicity of M. aeruginosa is relatively stable in the short term, and therefore can be unequivocally characterized by MLST. The MLST scheme established here will be of great help for future detailed population genetic studies of M. aeruginosa.

INTRODUCTION

Microcystis aeruginosa is a unicellular cyanobacterium that produces water blooms in eutrophic freshwater environments such as ponds, lakes and reservoirs. The genus Microcystis is characterized by having gas vesicles for buoyancy, a coccoid cell shape, a tendency to form aggregates or colonies, and amorphous mucilage or a sheath (Holt et al., 1994). Traditionally, five dominant morphospecies, M. aeruginosa, Microcystis novacekii, Microcystis ichthyoblabe, Microcystis viridis and Microcystis wessenbergii are recognized within the water-bloom-forming Microcystis, primarily on the basis of colony morphology, including the arrangement of cells and sheath characteristics (Watanabe, 1996). Given that the colony characteristics of these morphospecies are known to be highly variable and that the variation sometimes exceeds species criteria, the species definition of the genus Microcystis has been called into question (Otsuka et al., 2000). Indeed, molecular and chemotaxonomic data clearly demonstrate that each of the five morphospecies is non-monophyletic and that they are genetically and biochemically highly similar to each other, leading to the recent synthesis of the five morphospecies into one species, M. aeruginosa (Otsuka et al., 2001).

Hepatotoxic compounds produced by Microcystis, called microcystins, have become widely recognized as a serious...
public health concern. Accidental drinking of water contaminated with microcystins causes severe liver damage in both humans and animals as a consequence of the inhibition of protein phosphatases 1 and 2A (PP1 and PP2A) in hepatocytes. In addition, the involvement of microcystins in tumour promotion has been suggested (reviewed by Dittmann & Wiegand, 2006). Microcystins are a family of cyclic heptapeptides that are non-ribosomally synthesized by the microcystin synthetase (mcy) complex, which is encoded by 10 genes, mcyA–mcyJ (Tillett et al., 2000). Since both toxic and non-toxic colonies coexist in nature, researchers have been focusing on developing a rapid molecular typing method to discriminate toxic strains. To date, many attempts have been made to develop a method to distinguish toxic *M. aeruginosa*; these include RAPD fingerprinting (Neilan, 1995), 16S rDNA (Lyra et al., 2001; Neilan et al., 1997; Tillett et al., 2001), 16S–23S rDNA internal transcribed spacer (ITS; Janse et al., 2004; Otsuka et al., 1999), a part of the phycocyanin operon cpcBA intergenic spacer (IGS; Neilan et al., 1995; Tillett et al., 2001), and mcy genes (Kurmayer et al., 2002; Nishizawa et al., 1999; Tillett et al., 2001). However, these analyses consistently demonstrated that genetic similarity and toxicity are not always consistent, even when employing microcystin genes themselves as markers (i.e. a single mcy genotype encompasses both toxic and non-toxic strains). This inconsistency could be overcome by the use of a higher-resolution typing method.

For the past few decades, the impact of recombination on natural genetic diversity has become a central interest in the field of population genetics of bacteria. Accumulated evidence suggests that population structures of bacteria vary greatly depending on the species or subspecies, ranging from strictly clonal to panmictic (Smith et al., 1993). Since multiclonal markers with which we can index the degree of clonality are, however, not currently available for *Microcystis*, none of the genotyping studies has addressed this issue. However, molecular evolutionary studies have suggested that recombination contributes substantially to the genetic diversity of mcy genes of *M. aeruginosa* (Mikalsen et al., 2003; Tanabe et al., 2004), highlighting the potential importance of genetic exchange to the population genetic structure of this species. At the same time, recombination is of particular concern when one attempts to reconstruct a phylogenetic history because recombination breaks up correlations between genetic distances and strain relatedness. Therefore, a molecular tool that is less affected by genetic mixing is a prerequisite for phylogenetic analysis of recombining bacteria such as *M. aeruginosa*. Unfortunately, almost all previous efforts to differentiate genotypes of *Microcystis* have used either a single gene-based approach or fingerprinting, neither of which is free from problems involved by recombination.

In this study, we developed a multilocus sequence typing (MLST) scheme for *M. aeruginosa*. MLST is a DNA sequence-based genotyping approach that indexes the genetic variation of seven or more housekeeping loci, each of which is free from vigorous selection pressure, and ideally the loci are scattered around the chromosome to reflect overall genome evolution (Maiden et al., 1998). MLST has been widely and successfully used in describing the population structure of many bacterial species, particularly pathogenic bacteria (Feil & Enright, 2004). Our MLST protocol employed seven housekeeping loci, with which we characterized the genetic variation of 164 isolates of *M. aeruginosa*. Based on the results of MLST, we performed linkage disequilibrium analyses to infer the population genetic structure of *M. aeruginosa*. We also performed phylogenetic analyses, and we discuss the potential utility of MLST for characterizing toxic genotypes. To our knowledge, this is the first report of an MLST scheme for cyanobacteria.

**METHODS**

**Strains, cultures and DNA extraction.** The 164 strains of *M. aeruginosa* used in this study are listed in Supplementary Table S1. These include strains showing colony morphology typical of the five traditional morphospecies (*M. aeruginosa*, *M. ichthyoblabe*, *M. novacekii*, *M. viridis* and *M. wesenbergii*). These strains were isolated over 25 years from Japan and East Asia. One Canadian and one British isolate were also included. Almost all strains are clones; that is, each was established from a single bacterial cell. A few strains were established from a single colony, but sequencing experiments confirmed that these strains are represented by a single genotype. All strains but PACC7941 are available at MCC-NIES (Tsukuba, Japan). Cultures were grown in 10 ml MA medium (Kasai et al., 2004) at 25 °C for 1–3 weeks under a 12 : 12 h light : dark cycle with a photon density of 15 μmol m⁻² s⁻¹. Genomic DNA was extracted and purified using a FastDNA kit (Qbiogene).

**MLST and mcy gene detection.** Seven housekeeping loci, ftsZ, **glmN**, **glmX**, **gyrB**, **pgi**, **recA** and **tpi**, were selected for MLST. All of these loci have been employed in MLST of other bacteria. Each locus is present as a single copy, but their physical positions on the genome are currently difficult to determine (T. Kaneko & S. Tabata, personal communication). In order to develop the MLST primers for **ftsZ**, **glmN**, **glmX**, **pgi** and **recA**, degenerate primers targeting segments of each candidate gene were designed and used for PCRs using the genomic DNA of *M. aeruginosa* NIES102 as a reference strain (degenerate primer sequences are shown in Supplementary Table S2). Amplicons of the expected sizes were cloned and sequenced, followed by the identification of the corresponding MLST genes with the aid of **BLAST** searching. Based on the sequences we obtained, internal primers were designed. In the case of the remaining two loci, **gyrB** and **tpi**, primers were designed using published sequences from *M. aeruginosa* (DDBJ accession numbers AB014988–89 and AY238889–91), respectively as a guide. As a result of the successful PCR amplification and following direct sequencing of several strains, we validated these primers as MLST primers (Table 1). Using ExTag DNA polymerase (Takara), each locus was PCR-amplified (in 25 μl volumes) with initial denaturation at 94 °C for 3 min, followed by 40 cycles of 94 °C for 60 s, 60–68 °C (optimal annealing temperatures for each locus are listed in Table 1) for 60 s, and 72 °C for 30 s. To detect the potential toxicity of strains, a partial segment of **mcyG** (549 bp in length) was amplified using the previously described primer pair **GF** and **GR** (Tanabe et al., 2004) under the same PCR conditions as for MLST, except that the annealing temperature was 60 °C. PCR products were purified using ExoSAP-IT (USB) and were sequenced in both
directions using a DTCS Quick Start kit and a CEQ8000 autosequencer (Beckman Coulter). Sequence data have been submitted to the DDBJ/EMBL/GenBank databases under accession numbers AB324850–AB325402. For each locus, each different allele was assigned a different arbitrary number, and a unique combination of seven allele numbers (‘allelic profile’) unambiguously defines the sequence type (ST) of a strain. Allelic profiles were used for subsequent linkage disequilibrium analysis.

Population genetic analysis. Estimates of parameters for DNA divergence, gene diversity $H = [n/(n−1)](1−\sum p^2)$ (n is the number of samples and $p_i$ is the relative frequency of the $i$th allele), nucleotide diversity ($\pi$; Nei, 1987), and a test for neutrality based on Tajima’s $D$ (equal to zero at neutral equilibrium; Tajima, 1989) were performed using DnaSP version 4.00 (Rozas et al., 2003). Infragenic recombination was investigated using RDP2 software (Martin et al., 2005) with all the methods available at the default settings. Multilocus linkage disequilibrium was assessed using the index of association ($I_A$, Smith et al., 1993) and $r_D$ (Agapow & Burt, 2001) with the program MULTILOCUS version 1.2.2 (Agapow & Burt, 2001). $r_D$ is virtually a standardized measure of $I_A$ ranging from 0 (panmixia) to 1 (absolute linkage disequilibrium). $r_D$ is free from dependence on the number of loci analysed, as encountered with $I_A$, and makes comparison among studies possible. The statistical significance of non-zero values of these indices was also inferred, based on the comparison of those values estimated from 1000 randomized datasets under a null hypothesis of panmixia.

Phylogenetic analysis. The most appropriate model of DNA sequence evolution of MLST loci was selected by the hierarchical likelihood ratio test using MODELTEST version 3.7 (Posada & Crandall, 1998). Using PAUP* version 4.0b10 (Swofford, 2002), neighbour-joining (NJ) phylogenetic trees were constructed based on the alignments generated by 1000 resamplings of the data with the same settings used for phylogenetic reconstruction. Bayesian ML phylogenetic reconstruction was performed using MrBayes version 3.1.2 (Ronquist & Huelsenbeck, 2003). Based on the DNA evolution model chosen by MODELTEST, and using the NJ tree as a starting tree, two independent runs were performed, each with four chains for 2 000 000 generations, in which trees were sampled every 100 generations. Statistical confidence for branch support was assessed using posterior probability (PP) estimated from the 50% majority consensus tree after discarding the burn-in phase of 500 000 generations. Shimodaira–Hasegawa (SH) tests (Shimodaira & Hasegawa, 1999) were performed to statistically assess phylogenetic incongruence between loci based on the ML parameter estimates, as described above.

### RESULTS

#### Genetic diversity

All sets of primers successfully recovered seven MLST loci from all 164 isolates included in this study. Neither insertions nor deletions were found within the sequences of any of the loci, and therefore sequences could be unambiguously aligned. Genetic diversity indices calculated from the MLST data are shown in Table 2. The mean overall nucleotide sequence divergence ($\pi \times 100$) was 2.3%, where the maximum was 4.3% in $pgi$, and the minimum was 1.3% in $recA$. From 164 isolates, 79 STs were found. The number of alleles ranged from 40 for $ftsZ$ to 57 for $gltX$. This allows about 550 billion different allelic profiles to be discriminated within *M. aeruginosa*. The average gene diversity over seven loci was $H=0.951$. None of the Tajima’s $D$ values for each MLST locus deviated significantly from zero ($P>0.10$).

#### Recombination and linkage disequilibrium

All of the methods included in RDP2 failed to identify any intragenic (within-locus) recombinational replacements within the seven MLST genes. On the other hand, individual gene trees indicated many discrepancies between topologies (see Supplementary Fig. S1), and the result of SH tests provided statistically significant support for phylogenetic conflicts in almost all pair-wise comparisons between loci (Table 3). Multilocus linkage disequilibrium was analysed using $I_A$ and $r_D$ based on the MLST allelic profile (Table 4). Analysis of all isolates indicated $I_A=3.608$ and $r_D=0.608$, both of which were significantly positive.

### Table 1. MLST PCR primers

<table>
<thead>
<tr>
<th>Locus</th>
<th>Gene product</th>
<th>Primers</th>
<th>Sequence (5’–3’)</th>
<th>Length (bp)</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>$ftsZ$</td>
<td>Cell division protein FtsZ</td>
<td>$ftsF$</td>
<td>GGTTTACAAGAGTCCGGTGCGATAC</td>
<td>409</td>
<td>60 °C</td>
</tr>
<tr>
<td>$ftsR$</td>
<td></td>
<td>$ftsR$</td>
<td>CCTCGATTTTTCGTTCAATTAC</td>
<td>451</td>
<td>65 °C</td>
</tr>
<tr>
<td>$glmA$</td>
<td>Glutamine synthetase</td>
<td>$glmF$</td>
<td>AACCCACCTAAACATGATGTG</td>
<td>400</td>
<td>60 °C</td>
</tr>
<tr>
<td>$glmR$</td>
<td></td>
<td>$glmR$</td>
<td>GCCACGTCTTGAGTCGATGA</td>
<td>430</td>
<td>60 °C</td>
</tr>
<tr>
<td>$gltX$</td>
<td>Glutamyl-tRNA synthetase</td>
<td>$gltF$</td>
<td>ATTCAAAGGGGCCCACAAATG</td>
<td>416</td>
<td>60 °C</td>
</tr>
<tr>
<td>$gltR$</td>
<td></td>
<td>$gltR$</td>
<td>TCCCCACGAATACATGGGTAATT</td>
<td>467</td>
<td>60 °C</td>
</tr>
<tr>
<td>$gyrB$</td>
<td>DNA gyrase subunit B</td>
<td>$gyrF$</td>
<td>GGAGCTTTGACGAGAATCGCT</td>
<td>424</td>
<td>60 °C</td>
</tr>
<tr>
<td>$gyrR$</td>
<td></td>
<td>$gyrR$</td>
<td>GGTCTGATTGTCGTCCCTCAA</td>
<td>446</td>
<td>60 °C</td>
</tr>
<tr>
<td>$pgi$</td>
<td>Glucose-6-phosphate isomerase</td>
<td>$pgiF$</td>
<td>CGCTTTGCAACTGCTCC</td>
<td>424</td>
<td>60 °C</td>
</tr>
<tr>
<td>$pgiR$</td>
<td></td>
<td>$pgiR$</td>
<td>ATGGGGAAAGCGTTAGCCA</td>
<td>446</td>
<td>60 °C</td>
</tr>
<tr>
<td>$recA$</td>
<td>Recombination protein RecA</td>
<td>$recF$</td>
<td>ATGGGGAACACGCTTGA</td>
<td>424</td>
<td>60 °C</td>
</tr>
<tr>
<td>$recR$</td>
<td></td>
<td>$recR$</td>
<td>AGGAGCCGCCACCTTATT</td>
<td>446</td>
<td>60 °C</td>
</tr>
<tr>
<td>$tpi$</td>
<td>Triosephosphate isomerase</td>
<td>$tpiF$</td>
<td>GGTCTATACGCGGGGAATCT</td>
<td>416</td>
<td>60 °C</td>
</tr>
<tr>
<td>$tpiR$</td>
<td></td>
<td>$tpiR$</td>
<td>CCGACTAAAGCCTCGTTAT</td>
<td>416</td>
<td>60 °C</td>
</tr>
</tbody>
</table>
values (P<0.001). When analysing allelic profiles of the 79 unique STs only, smaller values were obtained (I_A=1.755, r_D=0.314), but were still significantly positive (P<0.001). Subsets of local populations were also analysed individually. All of the I_A values of local populations were still significantly positive (P<0.001). Notably, all indices of local populations were larger than the values obtained by the entire dataset. Among them, those for Lake Okutama and Shirakaba indicated complete linkage disequilibrium (I_A=6.000, r_D=1.000).

Detection of toxin genes
A pair of primers targeting mcyG (Tanabe et al., 2004) was used for diagnosis of potential toxicity of a given strain. The results of PCR detection experiments are shown in Supplementary Table S1. All toxic strains were positive for mcyG, whereas all but five non-toxic strains representing three STs (ST48, ST51 and ST55) were negative for mcyG. Sequencing experiments confirmed that all of these PCR products represent authentic mcyG. Non-toxicity of these three STs was further confirmed by liquid chromatography electrospray ionization MS (LC ESI-MS) (T. Sano, personal communication).

Phylogenetic analysis
Phylogenetic analyses of individual MLST genes failed to resolve almost all of the phylogenetic relationships with confidence (Supplementary Fig. S1). This poor resolution is likely due to a few phylogenetic signals within the individual gene sequences. We therefore concatenated the seven MLST sequences, obtaining a much longer alignment of 2992 bp. Also, this procedure has an advantage in that it can buffer the consequences of local recombination that could affect the recovery of true phylogenetic relationships when a phylogeny based on one or a small number of genes is employed (Hanage et al., 2005). A hierarchical likelihood-ratio test using the MODELTEST program indicated that the DNA evolution of the concatenated data was best explained by the GTR + I + G model (GTR, the general time reversible DNA substitution model; I, invariable sites; G, rate heterogeneity among sites under a gamma distribution with four categories), in which the rate matrix was estimated as (A–C]=0.8742, R(b) [A–G]=3.0433, R(c) [A–T]=0.5537, R(d) [C–G]=1.8297, R(e) [C–T]=6.5145, R(f) [G–T]=1.0000, the gamma shape parameter is ñ=0.4669, and the proportion of invariable sites is 0.8192. An NJ midpoint-rooted tree based on the ML

Table 2. Genetic diversity indices

<table>
<thead>
<tr>
<th>Locus</th>
<th>N*</th>
<th>H†</th>
<th>π‡ (maximum sequence divergence)</th>
<th>Tajima’s D A–T</th>
<th>G–T</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftsZ</td>
<td>40</td>
<td>0.939±0.009</td>
<td>0.026 (0.061)</td>
<td>−0.472§</td>
<td></td>
</tr>
<tr>
<td>gltA</td>
<td>55</td>
<td>0.948±0.009</td>
<td>0.025 (0.051)</td>
<td>0.175§</td>
<td></td>
</tr>
<tr>
<td>gltX</td>
<td>57</td>
<td>0.969±0.005</td>
<td>0.023 (0.049)</td>
<td>−0.203§</td>
<td></td>
</tr>
<tr>
<td>gyrB</td>
<td>44</td>
<td>0.938±0.009</td>
<td>0.017 (0.034)</td>
<td>0.196§</td>
<td></td>
</tr>
<tr>
<td>pgi</td>
<td>50</td>
<td>0.956±0.006</td>
<td>0.043 (0.083)</td>
<td>1.211§</td>
<td></td>
</tr>
<tr>
<td>recA</td>
<td>49</td>
<td>0.957±0.006</td>
<td>0.013 (0.040)</td>
<td>−0.420§</td>
<td></td>
</tr>
<tr>
<td>tpi</td>
<td>47</td>
<td>0.951±0.007</td>
<td>0.019 (0.053)</td>
<td>−0.269§</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>49</td>
<td>0.951±0.004</td>
<td>0.023 (0.055)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*N, number of alleles.  
†H, gene diversity.  
‡π, nucleotide diversity.  
§Not significantly different from zero (P>0.10).

Table 3. SH test for phylogenetic incongruence

<table>
<thead>
<tr>
<th>Locus</th>
<th>ftsZ</th>
<th>gltA</th>
<th>gltX</th>
<th>gyrB</th>
<th>pgi</th>
<th>recA</th>
<th>tpi</th>
<th>Concatenated</th>
</tr>
</thead>
<tbody>
<tr>
<td>ftsZ</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.034</td>
</tr>
<tr>
<td>gltA</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.023</td>
</tr>
<tr>
<td>gltX</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>gyrB</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>pgi</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>recA</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>tpi</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>Concatenated</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Significant incongruence (P<0.001) is indicated in bold type.

Table 4. Results of multilocus linkage disequilibrium analysis

<table>
<thead>
<tr>
<th></th>
<th>I_A</th>
<th>r_D</th>
</tr>
</thead>
<tbody>
<tr>
<td>All (n=164)</td>
<td>3.608*</td>
<td>0.608*</td>
</tr>
<tr>
<td>ST (n=79)</td>
<td>1.755*</td>
<td>0.314*</td>
</tr>
<tr>
<td>Lake Kasumigaura (n=25)</td>
<td>4.227*</td>
<td>0.720*</td>
</tr>
<tr>
<td>Lake Okutama (n=15)</td>
<td>6.000*</td>
<td>1.000*</td>
</tr>
<tr>
<td>Lake Shirakaba (n=12)</td>
<td>6.000*</td>
<td>1.000*</td>
</tr>
<tr>
<td>Lake Suwa (n=14)</td>
<td>5.084*</td>
<td>0.850*</td>
</tr>
</tbody>
</table>

*Significantly positive (P<0.001).
distance inferred from the selected model and parameters is shown in Fig. 1. As expected, relationships between strains were much more resolved, as illustrated by several highly supported monophyletic clades in the tree. (Bayesian ML analysis identified a phylogenetic tree with a different topology; however, the main differences lie in internal branches between and within major clades, and the robustness of each major clade was also supported by Bayesian PP.) In this tree, most toxic strains fell into two lineages, groups A and B, and only two toxic strains (ST23 and ST57) were not included in one of the two groups. All strains included in group A were toxic, except for one of the 10 representatives of ST55, whereas group B included both toxic and non-toxic strains. Support for group A was weak (≤50 %, both in NJBP and PP), and it could be subdivided into at least three highly supported groups (>90 %, both in NJBP and PP). Support for group B was moderate to strong (52 % NJBP; 100 % PP). ST40 was, however, more or less divergent from other strains of group B, as illustrated by the long branch leading to ST40 (Fig. 1). In fact, the position of ST40 was variably placed depending on the phylogenetic method used, in some cases even located outside group B (data not shown). The two toxic STs ST23 and ST57 were located near the base of group A. However, this placement received little support, and was indeed variable depending on the phylogenetic method used. Three strongly supported (100 % NJBP; 100 % PP) monophyletic groups that exclusively contained non-toxic strains were identified (groups C, D and E). The other non-toxic strains were distributed across the tree, but in most cases with no support for their placement.

Fig. 1. NJ phylogenetic tree of 164 strains of *M. aeruginosa* based on the concatenated sequences of seven MLST loci. Statistical values for major branches (shown by thick bars) are indicated (NJBP/Bayesian PP/C190 100). The number of strains representing the same ST is indicated in parentheses.
DISCUSSION

High genetic diversity within *M. aeruginosa*

Although our estimates might not be accurate due to non-random sampling, the MLST analysis revealed that a very high level of genetic diversity is maintained within *M. aeruginosa*. All MLST alleles are highly polymorphic, each containing at least 40 alleles per locus. The average gene diversity of the entire dataset is estimated to be $H=0.951$, which is much higher than that of *Escherichia coli* ($H=0.47$; Selander & Levin, 1980) and *Bacillus subtilis* ($H=0.44$; Istock *et al.*, 1992). Direct comparison of these two values with our estimate for *M. aeruginosa* may be meaningless, since these two estimates are based on multilocus enzyme electrophoresis (MLEE), where only a genotype with a mutation that causes an electromobility shift is distinguished as a distinct allele. However, our estimate of $H$ is still higher than the reported MLST estimates for *Enterococcus faecium* ($H=0.60$; Homan *et al.*, 2002) and *Streptococcus pneumoniae* ($H=0.82$; Enright & Spratt, 1998), as well as that for a hyperthermophilic archaeon (*Halorubrum*) ($H=0.69$; Papke *et al.*, 2004). These comparisons highlight the extremely high genetic diversity of *M. aeruginosa*. Theoretically, high genetic diversity is unexpected for clonal organisms such as bacteria. This is because clonal organisms are purged of genetic diversity at all loci by recurrent selective sweeps, so-called ‘periodic selection’ (Atwood *et al.*, 1951). Then how do we explain such a high level of genetic diversity in *M. aeruginosa*?

One possibility is that *M. aeruginosa* contains several ecologically distinct populations or ‘ecotypes’. Given that ecological divergence is expected to accompany genetic divergence, it follows that if more ecotypes exist within a single species, then higher genetic diversity is maintained within it. Cohan (2002) has suggested that each ecotype falls into its own sequence-based cluster through time. In fact, a number of distinct clusters are observed in the phylogenetic tree of *M. aeruginosa* (Fig. 1), suggesting that each cluster might represent a *bona fide* cryptic ecotype. If so, it is likely that ecological divergence contributes greatly to the high genetic diversity of *M. aeruginosa*. It should be noted that STs belonging to different clusters are often isolated from a single locality, even in the same water sample. This observation implies the presence of distinct microhabitats at a very fine scale. Future investigation of ecological parameters that potentially characterize each clade is essential to test this possibility.

Recombination and clonal population structure

Our results strongly suggest that recombination is an important evolutionary force for the generation and maintenance of the genetic diversity of *M. aeruginosa*. Many phylogenetic conflicts are evident among the individual gene phylogenies (Supplementary Fig. S1). This is further supported by the results of SH tests, which indicate statistically significant discrepancies in almost all pair-wise comparisons between single-gene phylogenies (Table 3), which is most likely explained by recombination at different loci. Although recombination appears to be substantial in *M. aeruginosa*, alleles at different loci were not assorted randomly. In fact, both the total and local $I_A$ and $r_D$ values indicated a significant deviation from the null model of panmixia (Table 4), suggesting a ‘clonal’ population structure for *M. aeruginosa*. Theoretically, a clonal population structure under a high rate of recombination can be observed when several particular STs are over-represented due to non-random and limited sampling, as in this study. However, an analysis of only single representatives of 79 STs indicates slightly lower but still significant non-zero values, excluding the possibility of an ‘epidemic’ population structure (Smith *et al.*, 1993). To our knowledge, this is the second report of clonal population structure of cyanobacteria, following the Baltic Sea population of *Nodularia* revealed by an allele-specific PCR study (Barker *et al.*, 2000). The recovery of the same ST from distantly separated locations (e.g. ST26 was isolated from Japan and China) also provides circumstantial evidence for clonality. Interestingly, all local $I_A$ and $r_D$ values are higher than those for the total, as clearly exemplified by the theoretical maximum values ($I_A=6$, $r_D=1$) obtained from two lakes, Lake Okutama and Lake Shirakaba (Table 4). This result suggests that geographical isolation does not influence the strong linkage disequilibrium inferred from the entire dataset. Concurrently, this implies that a sympatric population might contain a number of genetically isolated clones, which further supports the clonal population structure.

The properties of vectors may be responsible for the clonal population structure of recombogenic bacteria. At present, vectors involved in natural genetic exchange among *Microcystis* have yet to be identified. Basically, only three are possible: plasmids, conjugating plasmids or phages. Plasmids (Takahashi *et al.*, 1996, Wallace *et al.*, 2002) and cyanophages (Tucker & Pollard, 2005; Yoshida *et al.*, 2006) have been isolated from natural populations of *Microcystis*, and hence are good candidates for being responsible for recombination in nature. *In vitro* experiments have indicated that *Microcystis* is competent but sometimes difficult to transform due to the extracellular secretion of nuclease that may cause breakdown of foreign DNA (Dittmann *et al.*, 1997; Takahashi *et al.*, 1996). This observation is in line with the clonal population structure of *M. aeruginosa*, whereby recombination appears to be more or less restricted. Although the effects of conjugation and transduction on genetic exchange in *Microcystis* have never been investigated, these two mechanisms might also play a limited role(s) in the recombination of *Microcystis*, thus contributing to the clonal population structure of *M. aeruginosa*.

Phylogenetic relationships among toxic and non-toxic *M. aeruginosa*

Phylogenetic analyses identified a number of well- (and less-well-) defined clades within *M. aeruginosa*. Notably,
most toxic strains are assigned to two clades, groups A and B. Group A is a weakly supported clade that includes three well-supported subclades. In group A, microcystin toxicity appears to be stably maintained. Group B is better supported, but, in contrast with group A, the toxicity appears to be unstable. This implies that the toxin genes in group B are frequently lost and/or gained. The two toxic STs ST23 and ST57 are exceptional, and are not located within either group A or group B. The origin of toxicity of these two STs is unknown, and awaits future mcy genealogical analysis.

There are five anomalous strains that are non-toxic but have at least a part of the toxin gene mcyG. The otherwise toxic ST55 in group A includes one non-toxic strain that is also positive for mcyG. Group B also includes four strains represented by ST48 and ST51, which are also non-toxic but positive for mcyG. The presence of these inconsistent genotypes is most likely explained by the loss of function of the mcy genes due to a recent mutation caused by transposition, or a large deletion within the toxin genes or the relevant regulatory regions, as reported for the mcy genes of Planktothrix spp. (Christiansen et al., 2006). In M. aeruginosa, transpositional inactivation has also been documented in the gas vesicle genes (gvp) for buoyancy (Mlouka et al., 2004). Indeed, sequence comparisons revealed that the mcyG of non-toxic ST55 is distinct from that of any other toxic ST55, in that it contains one base substitution relative to the sequences of toxic ST55 (Y. Tanabe, unpublished data). Although we have yet to discern whether this point mutation is directly involved in the loss of microcystins, it is likely that another large mutation (e.g. insertion of IS elements) that occurred within the essential region might be responsible. Obtaining the sequences of the entire microcystin gene regions from these non-toxic strains and their transcriptional analysis are critical to address this issue. Another possible reason for non-toxicity is that the amount of microcystin production in these STs is too low to be detected by canonical HPLC and LC ESI-MS techniques. However, we consider this unlikely, since production of trace levels of microcystins has never been documented. Interestingly, non-toxic strains that possess an mcy gene (ST48 and ST51) form a monophyletic group. The stable retention of mcy genes in this non-toxic clade through phylogenetic timescales suggests the possibility that the mcy genes of these strains might function differently from other microcystin-producing genotypes. It is possible that these two STs do not produce microcystins under laboratory culture conditions and require certain physiological factors for microcystin production. Since the toxicity of our strains was assayed from cells grown in only one standard culture medium (T. Sano, personal communication), microcystins could be produced when different media are used. In any event, these groups of non-toxic strains should be characterized by molecular genetic experiments to clarify the mechanism underlying their non-toxic nature.

In contrast to the clustering of toxic strains, non-toxic strains are widely scattered in the phylogenetic tree (Fig. 1). Rantala et al. (2004) proposed the ‘microcystin-early’ hypothesis, in which the ability to produce microcystins has been vertically inherited through the diversification of toxic cyanobacterial genera and thereafter repeatedly lost. Although relationships between toxic and non-toxic groups are not statistically supported, this pattern of distribution could be explained by later acquisition of toxin genes in the toxic clade during the evolution of M. aeruginosa. Genealogical analyses of mcy are needed to test whether the gain of microcystin genes is also an important factor in the evolution of toxic Microcystis.

Utility of MLST for research on population dynamics

Due to their ecological and environmental importance, a reliable genotyping scheme is fundamental to the monitoring and control of the occurrence of water blooms of M. aeruginosa. The MLST developed here is a convenient technique with which we can unambiguously characterize the genotype of a clone. Indeed, our MLST potentially distinguishes more than 550 billion multilocus genotypes, and thus has high resolution. Although most of the strains analysed in this study are isolates from Japan, we have successfully recovered MLST genes for isolates from other Asian regions (e.g. China, Nepal and Thailand) as well as regions of other continents (e.g. Britain and Canada) using the same sets of primers. We therefore expect that our MLST protocol will be applicable to isolates worldwide. In fact, our preliminary survey of another 100 strains of diverse origins successfully recovered all of the MLST genes without exception (data not shown). Since the results of MLST are portable and therefore easily comparable to those of other researchers, our MLST would serve as a very valuable tool to investigate the global population structure of M. aeruginosa.

Our MLST survey demonstrated that toxic and non-toxic strains are distinct in their MLST profiles. This means that our MLST is valuable in that it can be used diagnostically to discriminate toxic from non-toxic strains. As noted above, there is one exception in the toxic ST55, which contains one non-toxic strain. However, such cases appear to be scarce, and are not expected to seriously influence the diagnosis with regard to toxicity in general. If one wants to determine toxicity with confidence, it may be desirable to use chemical techniques, such as HPLC, in combination with MLST.

Conclusions

In this study, we developed an MLST scheme for M. aeruginosa, demonstrated high genetic diversity, clonal population structure and substantial recombination, and identified toxic and non-toxic clades based on a collection of diverse samples. One might think that our results are no
more than a partial snapshot and far from a full description of the global genetic diversity and toxicity of *M. aeruginosa*. Nevertheless, our MLST of *M. aeruginosa* has high discriminatory power, assuring a significant contribution to future detailed studies. Investigation of other important population genetic forces, such as gene flow, natural selection, etc., with more extensive and/or focused sampling, would provide more insights into the spatial and temporal population dynamics of *M. aeruginosa*.

ACKNOWLEDGEMENTS

We thank M. Watanabe, Department of Botany, National Science Museum, Japan, for providing the Tsukuba Algal Collection (TAC) strains used in this study. We also thank T. Sano, National Institute for Environmental Studies, Japan, for providing unpublished results of microcystin production analysis of strains used in this study, T. Kaneko and S. Tabata, Kazusa DNA Research Institute, Japan, for providing unpublished genome information for *M. aeruginosa* NIES843, and K. Kaya, Tohoku University, for helpful discussions about the toxicity of *Microcystis*.

REFERENCES


Edited by: K. Forchhammer