
Kimura, Shigeko; Sako, Yoshihiko; Yoshida, Takashi

Applied and environmental microbiology (2013), 79(8): 2789-2795

© 2013, American Society for Microbiology.; This is not the published version. Please cite only the published version. この論文は出版社版でありません。引用の際には出版社版をご確認ご利用ください。
Title

Rapid gene diversification of *Microcystis* cyanophages revealed by long- and short-term genetic analysis of the tail sheath gene in a natural pond

Running title

Rapid diversification of *Microcystis* cyanophage

Authors

Shigeko Kimura, Yoshihiko Sako and Takashi Yoshida*

Affiliation

Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan

Corresponding author

Takashi Yoshida

Postal address: Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan. Phone: (+81) 75-753-6218. Fax: (+81) 75-7533-6226. E-mail: yoshiten@kais.kyoto-u.ac.jp

Journal section

Environmental microbiology
Abstract

Viruses influence the abundance of host populations through virus-mediated host cell lysis. Viruses contribute to the generation and maintenance of host diversity, which also results in viral diversity throughout their co-evolution. Here, to determine the phage gene diversification throughout co-evolution of host and phage in a natural environment, we investigated the genetic diversity and temporal changes in *Microcystis* cyanophage populations using a total of 810 sequences of the Ma-LMM01-type cyanophage tail sheath gene (*g*91) from 2006 to 2011 in a natural pond. The sequences obtained were highly diverse and assigned to 419 different genotypes (GT1-GT419) clustered at 100% nucleotide sequence similarity. A maximum parsimony network showed the genotypes were largely divided into three sequence groups, which were dominated by major genotypes (more than 24 sequences: GT2, GT53, and GT163 in group I; GT25 in group II; and GT1 in group III). These major genotypes co-existed and oscillated throughout the sampling periods, suggesting the *Microcystis*-cyanophage co-evolution was partly driven by a negative frequency-dependent selection. Meanwhile, the high viral genetic diversity observed was derived from a large number of the variants of each
major and moderately-frequent genotype (including 7 to 18 sequences: GT7,

GT26, GT56, GT149, and GT182 in group I; GT152 in group II) (1-2 nucleotide

substitutions). The variants almost always co-occurred with their origins. This

manner of variant emergence suggests increased contact frequency with a

host-phage population promotes rapid co-evolution in an arms race.
**Introduction**

Viruses are abundant in marine and fresh water ecosystems. They infect the hosts to replicate and, ultimately result in host cell lysis. Therefore, viruses play important roles in regulating the abundance of host populations, and catalyze the movement of nutrients and organic carbon from host cells to dissolved and particulate organic matter pools (‘viral shunt’) (20).

In addition, viruses affect host diversity largely in three ways: 1) horizontal gene transfer (5); 2) ‘arms race’: viruses promote the emergence of host defense systems against them, and subsequently viral mutations occur that enable infection of recently emergent resistant host populations, leading to rapid host-virus co-evolution and generation of their diversity (16, 17); and 3) ‘frequency-dependent selection’ (e.g. kill the winner, constant-diversity dynamics, or fluctuating selection): viruses infect host strains (genotypes or taxa) that become relatively abundant (the winner) and frequencies of host and phage strains oscillate over time, maintaining host and virus diversity (1, 18, 23).

Therefore, viruses contribute to the generation and maintenance of their host diversity, also resulting in viral diversity throughout their co-evolution.

*Microcystis aeruginosa* forms toxic cyanobacterial blooms throughout
the world. Several studies have shown that *Microcystis* populations are highly diversified and the genetic compositions of *Microcystis* populations temporally change during the development of blooms (3, 4, 19). Recently, a comparative genomic study showed the largest number of phage-defense systems in the *M. aeruginosa* NIES-843 genome included the clustered regularly interspaced short palindromic repeat (CRISPR) – CRISPR-associated (Cas) system, the restriction-modification (RM) system, and the abortive infection (Abi) system among the bacterial and archaeal genomes (15). Further, our study indicated CRISPR spacers in *M. aeruginosa* (considered to provide records of past infections by viruses) are remarkably diversified and are rarely shared between co-existing different CRISPR types in a natural *M. aeruginosa* population (12). This suggests *M. aeruginosa* is susceptible to attack by diverse viral communities and the host-phage interaction may be subdivided into diverse “susceptible combinations” of *M. aeruginosa* with its specific cyanophage (12). Therefore, the *M. aeruginosa*-phage interaction can be used to determine the co-evolution of phage and bacteria.

Ma-LMM01 is a lytic myovirus only infecting a single strain of *Microcystis aeruginosa* (NIES-298) (30). The majority of the predicted genes in its genome
have no detectable homologues in the present databases, and thus, Ma-LMM01 was assigned as a member of a new lineage in the *Myoviridae* family (6, 29). Since no strain closely related to Ma-LMM01 has been isolated, the degenerate primer set was designed based on environmental sequences of Ma-LMM01 g91 (tail sheath gene), which is phylogenetically distinct from other phages and has been used as a genetic marker of this phage (11, 25, 28). Using this primer set, clonal analysis reveals high sequence divergence that is derived from single point mutations in natural populations (11). These suggest that this gene is likely to reflect the ‘arms race’ between the phage and host populations even though there is no evidence that the gene involve in the interaction between hosts and phages. To determine the manner of the diversification of *Microcystis* cyanophage throughout co-evolution of the host and its infectious phage, we investigated the genetic diversity and its temporal change in genotypic composition of the Ma-LMM01-type phage in a natural pond by sequencing the phage tail sheath gene g91 in long- (5 years) and short- (one day) sampling periods.
Materials and Methods

Study site and sampling. Hirosawanoike Pond (35°026' N, 135°690' E) is a carp cultivation site located in central Kyoto, Japan (11). Water samples at the surface were taken from 2006 to 2011, except 2008, at a fixed point in the pond. Pond water was stored in a brown bottle and transported to the laboratory within 1 h. For phage DNA extraction, 10 mL of the pond water was filtered using a 0.2-μm polycarbonate filter (Toyo Roshi Kaisha, Ltd, Tokyo, Japan) and ultra-centrifuged at 111,000 × g for 1.5 h at 4 °C (21). The pellet was suspended in 200 μL deionized water (viral suspension) and stored at -80 °C. For host DNA extraction, 25-100 mL of the pond water was sonicated gently and harvested using centrifugation at 1,680 × g for 10 min (31). The pellet was stored at -20 °C until DNA analysis. Simultaneously, we measured water temperature and dissolved oxygen (DO) with YSI Model 55 (YSI incorporated, Yellow Springs, OH).

DNA extraction. Host DNA extraction was performed using the xanthogenate method as described previously (31). Phage DNA extraction was performed as previously described (21). To avoid contamination with dissolved DNA, viral
suspensions were treated with DNase I (SIGMA-ALDRICH, St. Louis, MO) at 37 °C for 1 h before DNA extraction. Purified DNAs were suspended in 30 µL deionized water. The amount and purity of the extracted DNA were determined using optical density comparison at 260 nm and 280 nm. Each DNA extract was used as a template for PCR to determine the g91 sequences.

**Primer design, PCR amplification, and sequencing.** We designed a degenerate primer set (g91DF1 and g91DR3). As no strain closely related to Ma-LMM01 has been isolated, the degenerated primer set was designed based on sequences obtained using products with thermal asymmetric interlaced (TAIL)-PCR products from environmental samples according to Kimura et al. (2012).

PCR amplification with primer sets g91DF1 (CTGGGTAATCAAGTTA) and g91DR3 (CGGGTGGRGTRMAYCYRCG) was performed using an i Cycler (Bio-Rad, Hercules, CA). The reaction conditions were an initial denaturation at 94 °C for 1 min, followed by 30 cycles: denaturation at 94 °C for 30 sec, annealing at 55 °C for 30 sec, and extension at 72 °C for 1 min; with a final extension at 72 °C for 10 min. The 50 µl reaction mixture contained 10×Ex Taq
Buffer (TaKaRa Bio Inc., Shiga, Japan), 200 µM dNTP mix, 0.5 µM each primer (g91DF1 and g91DR3), 1.25_U TaKaRa Ex Taq™ polymerase (TaKaRa Bio Inc., Shiga, Japan), and 1 µl of each DNA template. The PCR products (50µl) were electrophoresed in a 1.0 % (wt/vol) agarose gel in 1×TAE buffer and stained with GelRed (Biotium, Hayward, CA). The gel image was captured and analyzed with the Gel Doc XR system (Bio-Rad Laboratories, Hercules, CA). Visually confirmed bands derived from the amplicons obtained with PCR using the g91DF1-g91DR3 primer set were excised and purified using a Wizard Miniprep Purification Kit (Promega, Madison, WI) according to the manufacturer’s instructions. The purified PCR products were cloned into pTAC-1 (BioDynamics Laboratory Inc., Tokyo, Japan) and then transformed into *E. coli* DH5α-competent cells according to the manufacturer’s instructions. Positive clones containing an insert of the correct size from each clone library were verified by colony PCR. The plasmid templates were amplified using an illustra TempliPhi DNA Amplification Kit (GE Healthcare Japan Corporation, Tokyo, Japan) and isolated by PCR using the commercial primers M13 BDFw and M13 BDRev for the pTAC-1 Vector. Sequencing was performed using a 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA) with a BigDye Terminator v3.1
Cycle Sequencing Kit according to the manufacturer’s instructions (Applied Biosystems, Foster City, CA). The sequences obtained were aligned using MEGA5 (22); the primer sequences were removed from all sequences.

**Real-time PCR amplification.** To quantify abundances of total *M. aeruginosa*, a real-time PCR assay was performed using primers 188F/254R based on sequences of the phycocyanin intergenic spacer (PC-IGS) gene, as described previously (28).

**Diversity analysis.** Shannon index and Chao1 index were estimated for the obtained sequences using PAST software v2.09 (10) and EstimateS v8.2.0 (8), respectively. Coverage index (C) was calculated as $C = (1-n/N) \times 100$ (where $N$ is the number of sequences in each sample, $n$ is the number of genotypes appearing only once). Rarefaction curves were obtained for each sample using the PAST software v2.09 (10). Maximum parsimony network analysis was performed using the statistical parsimony program TCS v1.21 (7). Non-metric multi-dimensional scaling (MDS) based on Bray-Curtis similarity matrices in R 2.15.1 was used to visualize patterns in *g91* genotypic composition at different
Nucleotide sequence. The nucleotide sequences determined in this study are deposited in the DDBJ/EMBL/GenBank database. The accession numbers are AB766381 to AB767190.

Results

*M. aeruginosa* abundances and environmental parameters. We observed cyanobacterial bloom during sampling periods (early summer to autumn) in Hirosawanoike Pond. The PC-IGS copy numbers of *Microcystis aeruginosa* were almost always found between $10^5$ and $10^7$ copy numbers mL$^{-1}$ (Table 1). Water temperature was between 22.2 and 33.9 °C from July to September, and around 18 °C on October (Table 1). Dissolved oxygen concentrations were almost more than 100 %, supporting blooms occur during sampling periods (Table 1).

The temporal changes in genetic compositions in *Microcystis* cyanophage populations. In all, 810 g91 sequences (554 bp, 29-41 sequences from each
sample) were obtained (Table 1). The sequences showed significant similarities only to the corresponding region of Ma-LMM01 g91 when searched against the NCBI non-redundant protein sequence database using BLAST (data not shown). The 810 sequences were assigned to 419 different genotypes (GT1-GT419) clustered at 100% nucleotide sequence similarity. Forty three of these genotypes were found at least twice in our samples. The nucleotide sequence of the GT1 type showed 100% similarity with the same region of Microcystis cyanophage Ma-LMM01 isolated in 2003.

To determine the relationships between the phage g91 genotypes including the Ma-LMM01 g91 sequence (419 genotypes), we conducted a maximum parsimony network analysis with a 95% parsimony connection limit (Fig. 1). This network showed the genotypes were largely divided into three sequence groups: group I (339 genotypes, 642 sequences), group II (58 genotypes, 119 sequences), and group III (22 genotypes, 49 sequences). Group I was dominated by GT2, GT53, and GT163 genotypes consisting of 113, 74, and 39 sequences, respectively. Group II and group III were dominated by GT25 (47 sequences) and by GT1 (24 sequences), respectively (Fig. 1). We referred to these dominating genotypes as ‘major genotypes’. Most of the genotypes
(408/419) consisted of less than five sequences (rare genotypes). Most of them included only one sequence. We confirmed three groups were genetically distinct groups in a phylogenetic tree using the neighbor-joining method (Fig. S1). Comparing the sequences of representatives from each group (GT2, group I; GT25, group II; G1, group III), nucleotide differences between each pair were 27 (4.7 %, GT2 and GT25), 27 (4.7 %, G25 and GT1), and 33 (6.0 %, GT25 and GT1) (Fig. 1 and S2).

We also investigated the temporal changes in compositions of the g91 sequence genotype in natural cyanophage populations. Major genotypes (group I; GT2, GT53, and GT163, group II; GT25, and group III; GT1) accounted for 15 to 60 % at each sampling date (Fig. 2). Throughout the 5-year period (long-term), five major genotypes were frequently found in cyanophage populations and their compositions oscillated. For example, the GT2 genotype (group I), which included the largest number of sequences in all of the 419 genotypes, consisted of a large part of the populations between 2006 and 2010, but it greatly decreased in 2011. GT53 (group I) genotype was also found in almost all samples and dominated especially in 2011. The GT163 (group I) and GT25 (group II) genotypes were primarily observed in samples from 2009 to 2010.
GT1 genotype (group III), whose sequences were identical to Ma-LMM01, was found in 2006. This genotype was not detected between 2007 and 2010, and was observed again in 2011. The genotypic compositions in g91 sequences also changed during short-term periods (1, 3, 7, and 13-day intervals during 2009 and 2010) (Fig. 2). Further, non-metric multi-dimensional scaling (MDS) analysis showed the plot of each sampling day were scattered (data not shown), supporting that the major genotype compositions did not vary on direction, but oscillated. Rare genotypes occupied between 40 and 82 % of the populations at every sampling date (Fig. 2). Further, GT7, GT26, GT56, GT149, and GT182 genotypes in group I and the GT152 genotype in group II include 9, 18, 7, 14, 7, and 10 sequences, respectively. These genotypes (moderately-frequent genotypes; MF) accounted for between 3 and 26 % temporally during some sampling dates (Fig. 2).

Genetic diversity of *Microcystis* cyanophage. We determined the genetic diversity within the g91 gene at each sampling date (Table 1). The rarefaction curve did not reach an asymptote for any of the libraries from each sampling day (data not shown) and the sequence coverage values were low, suggesting
greater diversity of $g91$ sequences was present in the samples than shown by the sequencing (Table 1). The Chao1 index also suggested more genotypes were present in the samples than were detected in each clone library (Table 1). The Chao 1 and Shannon indices showed the high level of genetic diversity in cyanophage populations was maintained throughout the sampling periods (Table 1).

This diversity in the $g91$ gene was derived from a large number of rare genotypes, which were located at the periphery of each major and moderately-frequent genotype (1-2 nucleotide substitutions). The variants of the major genotypes almost always co-occurred with each major genotype at each sampling date (Fig. 3). For example, all the variants of the major genotype GT1 except for one clone were obtained when GT1 occurred (Fig. 3). This was observed in emergence of moderately-frequent genotypes and their variants although there were some exceptions that a few variants (ex. variants of GT152) occurred independently (Fig. 4). For example, all of GT26 variants (15/15) were only found in the sampling days when GT26 occurred. Thus, the frequency of the variants depended on the presence of the original genotypes (Fig. 3 and 4).

We found *Microcystis*-specific proto-spacer associated motifs (PAMs; GTT
of GTC) (12) on both strands of the g91 sequenced regions of each major
genotype (Fig. 5). The mutations in each genotype of the variants could be
mapped randomly within 35 bp (the average length of spacers in M. aeruginosa)
downstream from the PAM motifs (Fig. 5).

The amino acid sequences of the three major genotypes (GT2, GT53
and GT163) in group I exhibited 100% similarity to each other (Fig. 1). However,
most of the variants (191/262) harbored non-synonymous nucleotide
substitutions relative to the three major genotypes. Similarly, most of the variants
of GT25 (30/44), GT1 (17/21), and six moderately-frequent genotypes (60/81)
had non-synonymous substitutions relative to their original genotypes.

Discussion

The host-phage co-evolution that generates and maintains their diversity
has been assessed primarily through experimental studies (9, 16, 17). Here, we
observed at duration in genetic diversity and the rapid gene diversification in
natural cyanophage populations possibly throughout host-phage co-evolution.

Our data indicate five major genotypes of Ma-LMM01-type phage (group I: GT2,
GT53, and GT163; group II: GT25; and group III: GT1) co-exist in the natural
cyanophage populations throughout the sampling periods (Fig. 3). Additionally, the co-existent multiple cyanophage genotypes oscillated in the population over long- and short- periods (Fig. 3). This compositional oscillation of phage populations is a typical pattern in a frequency-dependent selection mode (9, 24).

Further, many studies show the composition of a natural _M. aeruginosa_ population (e.g., microcystin-producing and non-microcystin-producing populations, ITS types) is temporally changing during blooms (3, 4, 19, 25-27) and that the genetic shift is affected by cyanophage infection (25). Additionally, the constant-diversity dynamics model predicts that high diversity of bacterial community would be maintained steadily by phage infection against high-frequency hosts (18). These indicate that _Microcystis–cyanophage_ co-evolution is partly driven by negative frequency-dependent selection and that their genetic diversity is maintained throughout their co-evolution.

The variants of major genotypes and moderately-frequent genotypes in the _g91_ sequences almost always co-occurred with their original genotypes (Fig. 3 and 4). Several studies indicate an arms race between host and phage cannot continue because the host needs the costs of phage resistance for its growth (2, 9, 13). However, the manner for emergence of the variants where the variants
almost always co-occurred with their origins (Fig. 3 and 4) suggests increased frequency of a host-phage population promotes rapid co-evolution in the arms race mode. In general, CRISPR confers sequence-dependent, adaptive resistance in prokaryotes against viruses and plasmids via incorporation of short sequences called spacers. A single mutation in the proto-spacers can abolish CRISPR-mediated immunity against phages. Recently we have shown highly diversified CRISPR sequences in natural *Microcystis* populations. Although only a few spacers for Ma-LMM01-type phage were found (10 spacers in approx. 1,000 spacers), six of these spacers have point mutations compared to the sequence of Ma-LMM01 (proto-spacers), suggesting *Microcystis* cyanophages may evade interference mediated by the CRISPR spacers in *M. aeruginosa*. Further, distributions of PAMs and the mutations in the variants (Fig. 5) suggest that the sequenced regions are potential proto-spacers for the *Microcystis* CRISPR system. Therefore, the emerging pattern of variants from the major and moderately-frequent genotypes may be easily explained by rapid evolution of phage to avert the host CRISPR-Cas system. A simulation model predicts the CRISPR-mediated system allows a continuous arms race between host and phage (14). Combined, we infer that while the diversity of *Microcystis*
cyanophage is maintained by the negative frequency-dependent selection mode, rapid diversification occurs through co-evolution between increased host-cyanophage combinations which might contest under the CRISPR-Cas system.

Considering Ma-LMM01 has a narrow host range, the host-phage interaction may be subdivided into 'susceptible combinations'. However, we observed no succession from rare or moderately-frequent genotypes to major genotypes in this study. Three major genotypes (GT2, GT53, and GT163) dominating group I exhibited 100 % similarity at the amino acid level. In contrast, most of the rare and moderately-frequent genotypes had non-synonymous substitutions in the g91 gene compared to the three major genotypes. It is possible many of the variants trade off infection efficiency for avoidance of host defense. To understand host-phage interaction in natural environments, the Microcystis-phage system is recognized as a model system. Further studies evaluating spontaneous diversity of both of Microcystis and phage will further explain Microcystis-phage ecology, which may lead to understand other host-phage system.
Acknowledgements

This study was partially supported by a Grant-in-Aid for Scientific Research (B) (grant 20310045) and by the JSPS Research Fellowships for Young Scientists (grant 233313).

We thank Yusuke Matsui and Yoko Matsui for sampling help.

References


25. Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki...


**Figure Legends**

Figure 1. Maximum-parsimony network for the tail sheath gene g91 genotypes of the Ma-LMM01-type phage created using the TCS program version 1.21 (7). Eight hundred ten sequences generated from the samples taken during 2006 to 2011 (except 2008) from Hirosawanoike Pond were used. Circles indicate different genotypes. Cross-hatches in some connecting lines indicate mutational steps between genotypes. Each connecting line without cross-hatches that connected directly between genotypes (circles) represents a single mutational change. The sum of the number of cross-hatches, intervening genotypes, and junction nodes (smallest circles) between genotypes (circles) is the number of
nucleotide differences between them. The name of the major genotypes (GT2, GT53, GT163, GT25, and GT1) and moderately-frequent genotypes (GT7, GT26, GT56, GT149, GT182, and GT152) are indicated inside circles with the number of sequences obtained. Genotypes which exhibited 100 % similarity with each other at the amino acid level are coded as same color circles indicated in black or seven kinds of gray.

Figure 2. Changes in the proportion of the major, moderately-frequent, and rare genotypes in the tail sheath gene g91 of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond. MF: Moderately-Frequent.

Figure 3. Changes in the proportion of the five major genotypes (A) and their variants (B) from the tail sheath gene g91 of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.

Figure 4. Changes in the proportion of the moderately-frequent genotypes (A)
and their variants (B) in the tail sheath gene g91 of the Ma-LMM01-type phage
during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.

Figure 5. Distributions of specific proto-spacer associated motifs (PAMs) on the both of the sense strand (black) and on the anti-sense strand (gray) of the major genotypes of Ma-LMM01-type phage tail sheath gene and nucleotide substitutions within variants compared with the each original major genotype.

Figure S1. A neighbor-joining tree of 43 Ma-LMM01-type phage tail sheath gene genotypes that were found at least twice in samples from 2006 to 2011 (except 2008) in Hirosawanoike Pond. Bold indicates major genotypes. Bootstrap values are indicated at the nodes with more than 50 % bootstrap support.

Figure S2. Alignment of the nucleic acid sequences of major genotypes. Asterisks above the sequences indicate identical nucleotide in all major genotype sequences.
Figure 1. Maximum-parsimony network for the tail sheath gene g91 genotypes of the Ma-LMM01-type phage created using the TCS program version 1.21 (7). Eight hundred ten sequences generated from the samples taken during 2006 to 2011 (except 2008) from Hirosawanoike Pond were used. Circles indicate different genotypes. Cross-hatches in some connecting lines indicate mutational steps between genotypes. Each connecting line without cross-hatches that connected directly between genotypes (circles) represents a single mutational change. The sum of the number of cross-hatches, intervening genotypes, and junction nodes (smallest circles) between genotypes (circles) is the number of nucleotide differences between them. The name of the major genotypes (GT2, GT53, GT163, GT25, and GT1) and moderately-frequent genotypes (GT7, GT26, GT56, GT149, GT182, and GT152) are indicated inside circles with the number of sequences obtained. Genotypes which exhibited 100 % similarity with each other at the amino acid level are coded as same color circles indicated in black or seven kinds of gray.
Figure 2. Changes in the proportion of the major, moderately-frequent, and rare genotypes in the tail sheath gene $g91$ of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond. MF, Moderately-Frequent.
Figure 3. Changes in the proportion of the five major genotypes (A) and their variants (B) from the tail sheath gene g91 of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.
Figure 4. Changes in the proportion of the moderately-frequent genotypes (A) and their variants (B) in the tail sheath gene g91 of the Ma-LMM01-type phage during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.
Fig. 5. Distributions of specific proto-spacer associated motifs (PAMs) on the both of the sense strand (black) and on the anti-sense strand (gray) of the major genotypes of Ma-LMM01-type phage tail sheath gene and nucleotide substitutions within variants compared with the each original major genotype.
Table 1. *Microcystis aeruginosa* abundances and environmental parameters (water temperature and dissolved oxygen (DO)) during the sampling periods from 2006 to 2011 (except 2008) in Hirosawanoike Pond. Genetic diversity of the tail sheath gene g91 of the Ma-LMM01-type phage and the proportion of the major, moderately-frequent, and rare genotypes in Hirosawanoike Pond in each sampling day.

<table>
<thead>
<tr>
<th>Sampling date</th>
<th><em>M. aeruginosa</em> (copies mL$^{-1}$)</th>
<th>Temp. (°C)</th>
<th>DO (%)</th>
<th>No. of sequences</th>
<th>No. of genotypes</th>
<th>Coverage (%)</th>
<th>Shannon index</th>
<th>Chao 1index</th>
<th>The proportion of genotypes (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20060705</td>
<td>1.4×10$^7$ a</td>
<td>26.6 a</td>
<td>-</td>
<td>40</td>
<td>23</td>
<td>48</td>
<td>2.54</td>
<td>233 (114-508)</td>
<td>38 10 53</td>
</tr>
<tr>
<td>20060802</td>
<td>3.2×10$^6$ a</td>
<td>29.8 a</td>
<td>-</td>
<td>40</td>
<td>30</td>
<td>38</td>
<td>3.26</td>
<td>130 (62-341)</td>
<td>23 10 68</td>
</tr>
<tr>
<td>20060906</td>
<td>1.5×10$^6$ a</td>
<td>25.2 a</td>
<td>-</td>
<td>40</td>
<td>23</td>
<td>50</td>
<td>2.65</td>
<td>118 (51-344)</td>
<td>48 8 45</td>
</tr>
<tr>
<td>20070718</td>
<td>8.5×10$^4$ a</td>
<td>26.5 a</td>
<td>-</td>
<td>38</td>
<td>29</td>
<td>29</td>
<td>3.08</td>
<td>205 (85-582)</td>
<td>24 3 73</td>
</tr>
<tr>
<td>20070821</td>
<td>3.3×10$^5$ a</td>
<td>31.9 a</td>
<td>-</td>
<td>39</td>
<td>33</td>
<td>23</td>
<td>3.39</td>
<td>178 (81-467)</td>
<td>15 3 82</td>
</tr>
<tr>
<td>20070919</td>
<td>1.4×10$^6$ a</td>
<td>28.2 a</td>
<td>-</td>
<td>38</td>
<td>29</td>
<td>37</td>
<td>3.21</td>
<td>84 (48-191)</td>
<td>22 14 65</td>
</tr>
<tr>
<td>20090721</td>
<td>3.9×10$^7$ b</td>
<td>26.9 b</td>
<td>70</td>
<td>29</td>
<td>20</td>
<td>41</td>
<td>2.77</td>
<td>156 (74-363)</td>
<td>45 7 48</td>
</tr>
<tr>
<td>20090806</td>
<td>6.2×10$^6$</td>
<td>29.3</td>
<td>117</td>
<td>31</td>
<td>18</td>
<td>52</td>
<td>2.48</td>
<td>123 (57-299)</td>
<td>55 0 45</td>
</tr>
<tr>
<td>20090818</td>
<td>3.9×10$^6$</td>
<td>32.0</td>
<td>160</td>
<td>30</td>
<td>20</td>
<td>50</td>
<td>2.82</td>
<td>55 (30-147)</td>
<td>37 17 47</td>
</tr>
<tr>
<td>20090821</td>
<td>2.0×10$^6$ b</td>
<td>-</td>
<td>-</td>
<td>31</td>
<td>20</td>
<td>45</td>
<td>2.65</td>
<td>88 (39-261)</td>
<td>39 13 48</td>
</tr>
<tr>
<td>20090828</td>
<td>8.2×10$^7$</td>
<td>27.0</td>
<td>166</td>
<td>30</td>
<td>23</td>
<td>30</td>
<td>2.90</td>
<td>128 (54-373)</td>
<td>30 3 67</td>
</tr>
<tr>
<td>20090915</td>
<td>6.1×10$^6$ b</td>
<td>22.2 b</td>
<td>99</td>
<td>30</td>
<td>22</td>
<td>47</td>
<td>3.00</td>
<td>46 (29-103)</td>
<td>30 17 53</td>
</tr>
<tr>
<td>Date</td>
<td>Value</td>
<td>Frequency</td>
<td>Length</td>
<td>Value</td>
<td>Frequency</td>
<td>Value</td>
<td>Frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>-----------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20090916</td>
<td>5.2×10^6 b</td>
<td>23.1 b</td>
<td>146</td>
<td>30</td>
<td>25</td>
<td>33</td>
<td>3.17</td>
<td>57 (35-121)</td>
<td>17</td>
</tr>
<tr>
<td>20091020</td>
<td>2.5×10^6 b</td>
<td>18.8 b</td>
<td>150</td>
<td>30</td>
<td>16</td>
<td>57</td>
<td>2.38</td>
<td>94 (43-239)</td>
<td>60</td>
</tr>
<tr>
<td>20091021</td>
<td>3.5×10^6 b</td>
<td>18.2 b</td>
<td>133</td>
<td>30</td>
<td>18</td>
<td>53</td>
<td>2.58</td>
<td>64 (30-189)</td>
<td>47</td>
</tr>
<tr>
<td>20100723</td>
<td>4.4×10^3</td>
<td>33.9</td>
<td>112</td>
<td>41</td>
<td>26</td>
<td>49</td>
<td>2.99</td>
<td>96 (48-254)</td>
<td>41</td>
</tr>
<tr>
<td>20100805</td>
<td>3.4×10^5</td>
<td>-</td>
<td>-</td>
<td>39</td>
<td>24</td>
<td>46</td>
<td>2.69</td>
<td>129 (55-374)</td>
<td>46</td>
</tr>
<tr>
<td>20100814</td>
<td>2.4×10^6</td>
<td>-</td>
<td>-</td>
<td>39</td>
<td>27</td>
<td>36</td>
<td>2.96</td>
<td>327 (166-675)</td>
<td>36</td>
</tr>
<tr>
<td>20100827</td>
<td>3.6×10^5</td>
<td>32.5</td>
<td>176</td>
<td>40</td>
<td>23</td>
<td>50</td>
<td>2.68</td>
<td>213 (104-471)</td>
<td>53</td>
</tr>
<tr>
<td>20110708</td>
<td>4.0×10^6</td>
<td>27.9</td>
<td>131</td>
<td>37</td>
<td>28</td>
<td>27</td>
<td>2.99</td>
<td>379 (195-765)</td>
<td>30</td>
</tr>
<tr>
<td>20110726</td>
<td>1.9×10^7</td>
<td>-</td>
<td>-</td>
<td>38</td>
<td>27</td>
<td>37</td>
<td>3.05</td>
<td>303 (153-632)</td>
<td>11</td>
</tr>
<tr>
<td>20110804</td>
<td>5.1×10^7</td>
<td>29.6</td>
<td>174</td>
<td>35</td>
<td>17</td>
<td>57</td>
<td>2.03</td>
<td>70 (31-210)</td>
<td>54</td>
</tr>
<tr>
<td>20110909</td>
<td>6.6×10^7</td>
<td>28.2</td>
<td>&gt;200</td>
<td>38</td>
<td>20</td>
<td>53</td>
<td>2.23</td>
<td>97 (42-286)</td>
<td>50</td>
</tr>
</tbody>
</table>

^a Detected by Yoshida et al. (28). ^b Detected by Kimura et al. (11). M, Major genotype; MF, Moderately-Frequent genotype; R, Rare genotype. Dashes indicate no measured data.
Figure S1. A neighbor-joining tree of 43 Ma-LMM01-type phage tail sheath gene genotypes that were found at least twice in samples from 2006 to 2011 (except 2008) in Hirosawanoike Pond. Bold indicates major genotypes. Bootstrap values are indicated at the nodes with more than 50% bootstrap support.
Figure S2. Alignment of the nucleic acid sequences of major genotypes. Asterisks above the sequences indicate identical nucleotide in all major genotype sequences.