

1 **Title**

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

4 **Running title**

5 Rapid diversification of *Microcystis* cyanophage

6 **Authors**

7 Shigeko Kimura, Yoshihiko Sako and Takashi Yoshida*

8 **Affiliation**

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

11 **Corresponding author**

12 Takashi Yoshida

13 Postal address: Graduate School of Agriculture, Kyoto University,
14 Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan. Phone: (+81)
15 75-753-6218. Fax: (+81) 75-7533-6226. E-mail: yoshiten@kais.kyoto-u.ac.jp

16 **Journal section**

17 Environmental microbiology

18

19 **Abstract**

20 Viruses influence the abundance of host populations through
21 virus-mediated host cell lysis. Viruses contribute to the generation and
22 maintenance of host diversity, which also results in viral diversity throughout their
23 co-evolution. Here, to determine the phage gene diversification throughout
24 co-evolution of host and phage in a natural environment, we investigated the
25 genetic diversity and temporal changes in *Microcystis* cyanophage populations
26 using a total of 810 sequences of the Ma-LMM01-type cyanophage tail sheath
27 gene (*g91*) from 2006 to 2011 in a natural pond. The sequences obtained were
28 highly diverse and assigned to 419 different genotypes (GT1-GT419) clustered
29 at 100% nucleotide sequence similarity. A maximum parsimony network showed
30 the genotypes were largely divided into three sequence groups, which were
31 dominated by major genotypes (more than 24 sequences: GT2, GT53, and
32 GT163 in group I; GT25 in group II; and GT1 in group III). These major
33 genotypes co-existed and oscillated throughout the sampling periods,
34 suggesting the *Microcystis*-cyanophage co-evolution was partly driven by a
35 negative frequency-dependent selection. Meanwhile, the high viral genetic
36 diversity observed was derived from a large number of the variants of each

37 major and moderately-frequent genotype (including 7 to 18 sequences: GT7,
38 GT26, GT56, GT149, and GT182 in group I; GT152 in group II) (1-2 nucleotide
39 substitutions). The variants almost always co-occurred with their origins. This
40 manner of variant emergence suggests increased contact frequency with a
41 host-phage population promotes rapid co-evolution in an arms race.

42

43 **Introduction**

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

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

60 *Microcystis aeruginosa* forms toxic cyanobacterial blooms throughout

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

77 Ma-LMM01 is a lytic myovirus only infecting a single strain of *Microcystis*
78 *aeruginosa* (NIES-298) (30). The majority of the predicted genes in its genome

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

95

96

97 **Materials and Methods**

98 **Study site and sampling.** Hirosawanoike Pond (35°026' N, 135°690' E) is a
99 carp cultivation site located in central Kyoto, Japan (11). Water samples at the
100 surface were taken from 2006 to 2011, except 2008, at a fixed point in the pond.
101 Pond water was stored in a brown bottle and transported to the laboratory within
102 1 h. For phage DNA extraction, 10 mL of the pond water was filtered using a
103 0.2-µm polycarbonate filter (Toyo Roshi Kaisha, Ltd, Tokyo, Japan) and
104 ultra-centrifuged at $111,000 \times g$ for 1.5 h at 4 °C (21). The pellet was suspended
105 in 200 µL deionized water (viral suspension) and stored at -80 °C. For host DNA
106 extraction, 25-100 mL of the pond water was sonicated gently and harvested
107 using centrifugation at $1,680 \times g$ for 10 min (31). The pellet was stored at -20 °C
108 until DNA analysis. Simultaneously, we measured water temperature and
109 dissolved oxygen (DO) with YSI Model 55 (YSI incorporated, Yellow Springs,
110 OH).

111

112 **DNA extraction.** Host DNA extraction was performed using the xanthogenate
113 method as described previously (31). Phage DNA extraction was performed as
114 previously described (21). To avoid contamination with dissolved DNA, viral

115 suspensions were treated with DNase I (SIGMA-ALDRICH, St. Louis, MO) at 37
116 °C for 1 h before DNA extraction. Purified DNAs were suspended in 30 µL
117 deionized water. The amount and purity of the extracted DNA were determined
118 using optical density comparison at 260 nm and 280 nm. Each DNA extract was
119 used as a template for PCR to determine the *g91* sequences.

120

121 **Primer design, PCR amplification, and sequencing.** We designed a
122 degenerate primer set (g91DF1 and g91DR3). As no strain closely related to
123 Ma-LMM01 has been isolated, the degenerated primer set was designed based
124 on sequences obtained using products with thermal asymmetric interlaced
125 (TAIL)-PCR products from environmental samples according to Kimura *et al.*
126 (2012).

127 PCR amplification with primer sets g91DF1 (CTGGGGTAATCAAGTTA)
128 and g91DR3 (CGGGTGGRGTTRMAYCYRCG) was performed using an iCycler
129 (Bio-Rad, Hercules, CA). The reaction conditions were an initial denaturation at
130 94 °C for 1 min, followed by 30 cycles: denaturation at 94 °C for 30 sec,
131 annealing at 55 °C for 30 sec, and extension at 72 °C for 1 min; with a final
132 extension at 72 °C for 10 min. The 50 µl reaction mixture contained 10×*Ex Taq*

133 Buffer (TaKaRa Bio Inc., Shiga, Japan), 200 μ M dNTP mix, 0.5 μ M each primer
134 (g91DF1 and g91DR3), 1.25_U TaKaRa *Ex Taq*TM polymerase (TaKaRa Bio Inc.,
135 Shiga, Japan), and 1 μ l of each DNA template. The PCR products (50 μ l) were
136 electrophoresed in a 1.0 % (wt/vol) agarose gel in 1 $\times$ TAE buffer and stained with
137 GelRed (Biotium, Hayward, CA). The gel image was captured and analyzed with
138 the Gel Doc XR system (Bio-Rad Laboratories, Hercules, CA). Visually
139 confirmed bands derived from the amplicons obtained with PCR using the
140 g91DF1-g91DR3 primer set were excised and purified using a Wizard Miniprep
141 Purification Kit (Promega, Madison, WI) according to the manufacturer's
142 instructions. The purified PCR products were cloned into pTAC-1 (BioDynamics
143 Laboratory Inc., Tokyo, Japan) and then transformed into *E. coli*
144 DH5 α -competent cells according to the manufacturer's instructions. Positive
145 clones containing an insert of the correct size from each clone library were
146 verified by colony PCR. The plasmid templates were amplified using an illustra
147 TempliPhi DNA Amplification Kit (GE Healthcare Japan Corporation, Tokyo,
148 Japan) and isolated by PCR using the commercial primers M13 BDFw and M13
149 BDRRev for the pTAC-1 Vector. Sequencing was performed using a 3130 Genetic
150 Analyzer (Applied Biosystems, Foster City, CA) with a BigDye Terminator v3.1

151 Cycle Sequencing Kit according to the manufacturer's instructions (Applied
152 Biosystems, Foster City, CA). The sequences obtained were aligned using
153 MEGA5 (22); the primer sequences were removed from all sequences.

154

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

159

160 **Diversity analysis.** Shannon index and Chao1 index were estimated for the
161 obtained sequences using PAST software v2.09 (10) and EstimateS v8.2.0 (8),
162 respectively. Coverage index (C) was calculated as $C = (1 - n/N) \times 100$ (where N is
163 the number of sequences in each sample, n is the number of genotypes
164 appearing only once). Rarefaction curves were obtained for each sample using
165 the PAST software v2.09 (10). Maximum parsimony network analysis was
166 performed using the statistical parsimony program TCS v1.21 (7). Non-metric
167 multi-dimensional scaling (MDS) based on Bray-Curtis similarity matrices in R
168 2.15.1 was used to visualize patterns in *g91* genotypic composition at different

169 sampling days.

170

171 **Nucleotide sequence.** The nucleotide sequences determined in this study are
172 deposited in the DDBJ/EMBL/GenBank database. The accession numbers are
173 AB766381 to AB767190.

174

175 **Results**

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

184

185 **The temporal changes in genetic compositions in *Microcystis* cyanophage**
186 **populations.** In all, 810 *g91* sequences (554 bp, 29-41 sequences from each

187 sample) were obtained (Table 1). The sequences showed significant similarities
188 only to the corresponding region of Ma-LMM01 *g91* when searched against the
189 NCBI non-redundant protein sequence database using BLAST (data not shown).
190 The 810 sequences were assigned to 419 different genotypes (GT1-GT419)
191 clustered at 100 % nucleotide sequence similarity. Forty three of these
192 genotypes were found at least twice in our samples. The nucleotide sequence of
193 the GT1 type showed 100% similarity with the same region of *Microcystis*
194 cyanophage Ma-LMM01 isolated in 2003.

195 To determine the relationships between the phage *g91* genotypes
196 including the Ma-LMM01 *g91* sequence (419 genotypes), we conducted a
197 maximum parsimony network analysis with a 95% parsimony connection limit
198 (Fig. 1). This network showed the genotypes were largely divided into three
199 sequence groups: group I (339 genotypes, 642 sequences), group II (58
200 genotypes, 119 sequences), and group III (22 genotypes, 49 sequences). Group
201 I was dominated by GT2, GT53, and GT163 genotypes consisting of 113, 74,
202 and 39 sequences, respectively. Group II and group III were dominated by GT25
203 (47 sequences) and by GT1 (24 sequences), respectively (Fig. 1). We referred
204 to these dominating genotypes as 'major genotypes'. Most of the genotypes

205 (408/419) consisted of less than five sequences (rare genotypes). Most of them
206 included only one sequence. We confirmed three groups were genetically
207 distinct groups in a phylogenetic tree using the neighbor-joining method (Fig. S1).
208 Comparing the sequences of representatives from each group (GT2, group I;
209 GT25, group II; G1, group III), nucleotide differences between each pair were 27
210 (4.7 %, GT2 and GT25), 27 (4.7 %, G25 and GT1), and 33 (6.0 %, GT25 and
211 GT1) (Fig. 1 and S2).

212 We also investigated the temporal changes in compositions of the *g91*
213 sequence genotype in natural cyanophage populations. Major genotypes (group
214 I; GT2, GT53, and GT163, group II; GT25, and group III; GT1) accounted for 15
215 to 60 % at each sampling date (Fig. 2). Throughout the 5-year period (long-term),
216 five major genotypes were frequently found in cyanophage populations and their
217 compositions oscillated. For example, the GT2 genotype (group I), which
218 included the largest number of sequences in all of the 419 genotypes, consisted
219 of a large part of the populations between 2006 and 2010, but it greatly
220 decreased in 2011. GT53 (group I) genotype was also found in almost all
221 samples and dominated especially in 2011. The GT163 (group I) and GT25
222 (group II) genotypes were primarily observed in samples from 2009 to 2010.

223 GT1 genotype (group III), whose sequences were identical to Ma-LMM01, was
224 found in 2006. This genotype was not detected between 2007 and 2010, and
225 was observed again in 2011. The genotypic compositions in *g91* sequences also
226 changed during short-term periods (1, 3, 7, and 13-day intervals during 2009 and
227 2010) (Fig. 2). Further, non-metric multi-dimensional scaling (MDS) analysis
228 showed the plot of each sampling day were scattered (data not shown),
229 supporting that the major genotype compositions did not vary on direction, but
230 oscillated. Rare genotypes occupied between 40 and 82 % of the populations at
231 every sampling date (Fig. 2). Further, GT7, GT26, GT56, GT149, and GT182
232 genotypes in group I and the GT152 genotype in group II include 9, 18, 7, 14, 7,
233 and 10 sequences, respectively. These genotypes (moderately-frequent
234 genotypes; MF) accounted for between 3 and 26 % temporally during some
235 sampling dates (Fig. 2).

236

237 **Genetic diversity of *Microcystis* cyanophage.** We determined the genetic
238 diversity within the *g91* gene at each sampling date (Table 1). The rarefaction
239 curve did not reach an asymptote for any of the libraries from each sampling day
240 (data not shown) and the sequence coverage values were low, suggesting

241 greater diversity of *g91* sequences was present in the samples than shown by
242 the sequencing (Table 1). The Chao1 index also suggested more genotypes
243 were present in the samples than were detected in each clone library (Table 1).
244 The Chao 1 and Shannon indices showed the high level of genetic diversity in
245 cyanophage populations was maintained throughout the sampling periods (Table
246 1).

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

258 We found *Microcystis*-specific proto-spacer associated motifs (PAMs; GTT

259 of GTC) (12) on both strands of the *g91* sequenced regions of each major
260 genotype (Fig. 5). The mutations in each genotype of the variants could be
261 mapped randomly within 35 bp (the average length of spacers in *M. aeruginosa*)
262 downstream from the PAM motifs (Fig. 5).

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

269

270 **Discussion**

271 The host-phage co-evolution that generates and maintains their diversity
272 has been assessed primarily through experimental studies (9, 16, 17). Here, we
273 observed at duration in genetic diversity and the rapid gene diversification in
274 natural cyanophage populations possibly throughout host-phage co-evolution.
275 Our data indicate five major genotypes of Ma-LMM01-type phage (group I: GT2,
276 GT53, and GT163; group II: GT25; and group III: GT1) co-exist in the natural

277 cyanophage populations throughout the sampling periods (Fig. 3). Additionally,
278 the co-existent multiple cyanophage genotypes oscillated in the population over
279 long- and short- periods (Fig. 3). This compositional oscillation of phage
280 populations is a typical pattern in a frequency-dependent selection mode (9, 24).
281 Further, many studies show the composition of a natural *M. aeruginosa*
282 population (e.g., microcystin-producing and non-microcystin-producing
283 populations, ITS types) is temporally changing during blooms (3, 4, 19, 25-27)
284 and that the genetic shift is affected by cyanophage infection (25). Additionally,
285 the constant-diversity dynamics model predicts that high diversity of bacterial
286 community would be maintained steadily by phage infection against
287 high-frequency hosts (18). These indicate that *Microcystis*–cyanophage
288 co-evolution is partly driven by negative frequency-dependent selection and that
289 their genetic diversity is maintained throughout their co-evolution.

290 The variants of major genotypes and moderately-frequent genotypes in
291 the *g91* sequences almost always co-occurred with their original genotypes (Fig.
292 3 and 4). Several studies indicate an arms race between host and phage cannot
293 continue because the host needs the costs of phage resistance for its growth (2,
294 9, 13). However, the manner for emergence of the variants where the variants

295 almost always co-occurred with their origins (Fig. 3 and 4) suggests increased
296 frequency of a host-phage population promotes rapid co-evolution in the arms
297 race mode. In general, CRISPR confers sequence-dependent, adaptive
298 resistance in prokaryotes against viruses and plasmids via incorporation of short
299 sequences called spacers. A single mutation in the proto-spacers can abolish
300 CRISPR-mediated immunity against phages. Recently we have shown highly
301 diversified CRISPR sequences in natural *Microcystis* populations. Although only a
302 few spacers for Ma-LMM01-type phage were found (10 spacers in approx. 1,000
303 spacers), six of these spacers have point mutations compared to the sequence
304 of Ma-LMM01 (proto-spacers), suggesting *Microcystis* cyanophages may evade
305 interference mediated by the CRISPR spacers in *M. aeruginosa*. Further,
306 distributions of PAMs and the mutations in the variants (Fig. 5) suggest that the
307 sequenced regions are potential proto-spacers for the *Microcystis* CRISPR
308 system. Therefore, the emerging pattern of variants from the major and
309 moderately-frequent genotypes may be easily explained by rapid evolution of
310 phage to avert the host CRISPR-Cas system. A simulation model predicts the
311 CRISPR-mediated system allows a continuous arms race between host and
312 phage (14). Combined, we infer that while the diversity of *Microcystis*

313 cyanophage is maintained by the negative frequency-dependent selection mode,
314 rapid diversification occurs through co-evolution between increased
315 host-cyanophage combinations which might contest under the CRISPR-Cas
316 system.

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

330

331 **Acknowledgements**

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

335 We thank Yusuke Matsui and Yoko Matsui for sampling help.

336

337 **References**

- 338 1. **Avrani S, Schwartz DA, and Lindell D.** 2012. Virus-host swinging party
339 in the oceans: Incorporating biological complexity into paradigms of
340 antagonistic coexistence. *Mob Genet Elements* **2**:88-95.
- 341 2. **Bohannon BJM.** 2000. Linking genetic change to community evolution:
342 insights from studies of bacteria and bacteriophage. *Ecol Lett* **3**:362-377.
- 343 3. **Bozarth CS, Schwartz AD, Shepardson JW, Colwell FS, and Dreher**
344 **TW.** 2010. Population turnover in a *Microcystis* bloom results in
345 predominantly nontoxigenic variants late in the season. *Appl Environ*
346 *Microbiol* **76**:5207-5213.
- 347 4. **Briand E, Escoffier N, Straub C, Sabart M, Quiblier C, and Humbert**
348 **JF.** 2009. Spatiotemporal changes in the genetic diversity of a

- 349 bloom-forming *Microcystis aeruginosa* (cyanobacteria) population. ISME
350 J **3**:419-429.
- 351 5. **Brussow H, Canchaya C, and Hardt WD.** 2004. Phages and the
352 evolution of bacterial pathogens: from genomic rearrangements to
353 lysogenic conversion. Microbiol Mol Biol Rev **68**:560-602.
- 354 6. **Carstens EB.** 2010. Ratification vote on taxonomic proposals to the
355 International Committee on Taxonomy of Viruses (2009). Arch Virol
356 **155**:133-146.
- 357 7. **Clement M, Posada D, and Crandall KA.** 2000. TCS: a computer
358 program to estimate gene genealogies. Mol Ecol **9**:1657-1659.
- 359 8. **Colwell RK.** 2006. Estimates: statistical estimation of species richness
360 and shared species from samples. Version 8.
361 <http://viceroy.eeb.uconn.edu/estimates>.
- 362 9. **Hall AR, Scanlan PD, Morgan AD, and Buckling A.** 2011. Host-parasite
363 coevolutionary arms races give way to fluctuating selection. Ecol Lett
364 **14**:635-642.
- 365 10. **Hammer Ø, Harper D, and Ryan P.** 2001. PAST: Paleontological
366 statistics software package for education and data analysis. Palaeontol

- 367 Electron **4**.
- 368 11. **Kimura S, Yoshida T, Hosoda N, Honda T, Kuno S, Kamiji R,**
369 **Hashimoto R, and Sako Y.** 2012. Diurnal infection patterns and impact of
370 *Microcystis* cyanophages in a Japanese pond. Appl Environ Microbiol
371 **78**:5805-5811.
- 372 12. **Kuno S, Yoshida T, Kaneko T, and Sako Y.** 2012. Intricate interactions
373 between the bloom-forming cyanobacterium *Microcystis aeruginosa* and
374 foreign genetic elements, revealed by diversified clustered regularly
375 interspaced short palindromic repeat (CRISPR) signatures. Appl Environ
376 Microbiol **78**:5353-5360.
- 377 13. **Lennon JT, Khatana SAM, Marston MF, and Martiny JBH.** 2007. Is
378 there a cost of virus resistance in marine cyanobacteria? ISME J
379 **1**:300-312.
- 380 14. **Levin BR.** 2010. Nasty viruses, costly plasmids, population dynamics,
381 and the conditions for establishing and maintaining CRISPR-mediated
382 adaptive immunity in bacteria. PLoS Genet **6**:e1001171.
- 383 15. **Makarova KS, Wolf YI, Snir S, and Koonin EV.** 2011. Defense islands in
384 bacterial and archaeal genomes and prediction of novel defense systems.

- 385 J Bacteriol **193**:6039-6056.
- 386 16. **Marston MF, Pierciey FJ, Jr., Shepard A, Gearin G, Qi J, Yandava C,**
387 **Schuster SC, Henn MR, and Martiny JB.** 2012. Rapid diversification of
388 coevolving marine *Synechococcus* and a virus. Proc Natl Acad Sci U S A
389 **109**:4544-4549.
- 390 17. **Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson**
391 **NR, Quail M, Smith F, Walker D, Libberton B, Fenton A, Hall N, and**
392 **Brockhurst MA.** 2010. Antagonistic coevolution accelerates molecular
393 evolution. Nature **464**:275-278.
- 394 18. **Rodriguez-Valera F, Martin-Cuadrado AB, Rodriguez-Brito B, Pasic L,**
395 **Thingstad TF, Rohwer F, and Mira A.** 2009. Explaining microbial
396 population genomics through phage predation. Nat Rev Microbiol
397 **7**:828-836.
- 398 19. **Sabart M, Pobel D, Latour D, Robin J, Salencon MJ, and Humbert JF.**
399 2009. Spatiotemporal changes in the genetic diversity in French
400 bloom-forming populations of the toxic cyanobacterium, *Microcystis*
401 *aeruginosa*. Environ Microbiol Reports **1**:263-272.
- 402 20. **Suttle CA.** 2007. Marine viruses--major players in the global ecosystem.

- 403 Nat Rev Microbiol **5**:801-812.
- 404 21. **Takashima Y, Yoshida T, Yoshida M, Shirai Y, Tomaru Y, Takao Y,**
405 **Hiroishi S, and Nagasaki K.** 2007. Development and application of
406 quantitative detection of cyanophages phylogenetically related to
407 cyanophage Ma-LMM01 infecting *Microcystis aeruginosa* in fresh water.
408 Microbes Environ **22**:207-213.
- 409 22. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S.**
410 2011. MEGA5: molecular evolutionary genetics analysis using maximum
411 likelihood, evolutionary distance, and maximum parsimony methods. Mol
412 Biol Evol **28**:2731-2739.
- 413 23. **Winter C, Bouvier T, Weinbauer MG, and Thingstad TF.** 2010.
414 Trade-offs between competition and defense specialists among
415 unicellular planktonic organisms: the "killing the winner" hypothesis
416 revisited. Microbiol Mol Biol Rev **74**:42-57.
- 417 24. **Woolhouse ME, Webster JP, Domingo E, Charlesworth B, and Levin**
418 **BR.** 2002. Biological and biomedical implications of the co-evolution of
419 pathogens and their hosts. Nat Genet **32**:569-577.
- 420 25. **Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki**

- 421 **K, and Hiroishi S.** 2008. Ecological dynamics of the toxic bloom-forming
422 cyanobacterium *Microcystis aeruginosa* and its cyanophages in
423 freshwater. Appl Environ Microbiol **74**:3269-3273.
- 424 26. **Yoshida M, Yoshida T, Takashima Y, Hosoda N, and Hiroishi S.** 2007.
425 Dynamics of microcystin-producing and non-microcystin-producing
426 *Microcystis* populations is correlated with nitrate concentration in a
427 Japanese lake. FEMS Microbiol Lett **266**:49-53.
- 428 27. **Yoshida M, Yoshida T, Takashima Y, Kondo R, and Hiroishi S.** 2005.
429 Genetic diversity of the toxic cyanobacterium *Microcystis* in Lake Mikata.
430 Environ Toxicol **20**:229-234.
- 431 28. **Yoshida M, Yoshida T, Yoshida-Takashima Y, Kashima A, and**
432 **Hiroishi S.** 2010. Real-time PCR detection of host-mediated cyanophage
433 gene transcripts during infection of a natural *Microcystis aeruginosa*
434 population. Microbes Environ **25**:211-215.
- 435 29. **Yoshida T, Nagasaki K, Takashima Y, Shirai Y, Tomaru Y, Takao Y,**
436 **Sakamoto S, Hiroishi S, and Ogata H.** 2008. Ma-LMM01 infecting toxic
437 *Microcystis aeruginosa* illuminates diverse cyanophage genome
438 strategies. J Bacteriol **190**:1762-1772.

- 439 30. **Yoshida T, Takashima Y, Tomaru Y, Shirai Y, Takao Y, Hiroishi S, and**
440 **Nagasaki K.** 2006. Isolation and characterization of a cyanophage
441 infecting the toxic cyanobacterium *Microcystis aeruginosa*. Appl Environ
442 Microbiol **72**:1239-1247.
- 443 31. **Yoshida T, Yuki Y, Lei S, Chinen H, Yoshida M, Kondo R, and Hiroishi**
444 **S.** 2003. Quantitative detection of toxic strains of the cyanobacterial
445 genus *Microcystis* by competitive PCR. Microbes Environ **18**:16-23.

446

447 **Figure Legends**

448 Figure 1. Maximum-parsimony network for the tail sheath gene *g91* genotypes of
449 the Ma-LMM01-type phage created using the TCS program version 1.21 (7).
450 Eight hundred ten sequences generated from the samples taken during 2006 to
451 2011 (except 2008) from Hirosawanoike Pond were used. Circles indicate
452 different genotypes. Cross-hatches in some connecting lines indicate mutational
453 steps between genotypes. Each connecting line without cross-hatches that
454 connected directly between genotypes (circles) represents a single mutational
455 change. The sum of the number of cross-hatches, intervening genotypes, and
456 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)

475 and their variants (B) in the tail sheath gene *g91* of the Ma-LMM01-type phage
476 during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike
477 Pond.

478

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

483

484 Figure S1. A neighbor-joining tree of 43 Ma-LMM01-type phage tail sheath gene
485 genotypes that were found at least twice in samples from 2006 to 2011 (except
486 2008) in Hirosawanoike Pond. Bold indicates major genotypes. Bootstrap values
487 are indicated at the nodes with more than 50 % bootstrap support.

488

489 Figure S2. Alignment of the nucleic acid sequences of major genotypes.
490 Asterisks above the sequences indicate identical nucleotide in all major
491 genotype sequences.

492

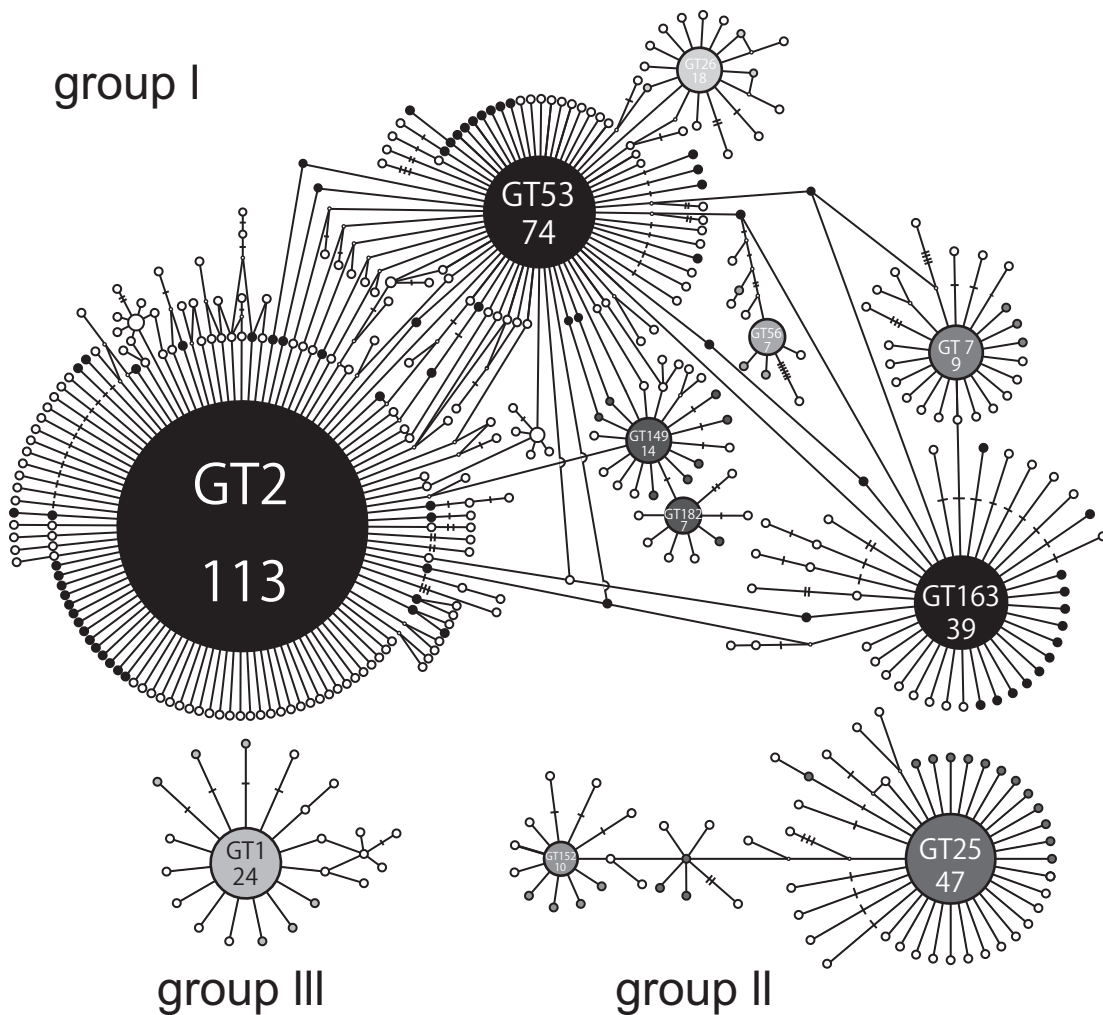


Figure 1. Maximum-parsimony network for the tail sheath gene *g91* genotypes of the Ma-LMM01-type phage created using the TCS program version 1.21 (7). Eight hundred ten sequences generated from the samples taken during 2006 to 2011 (except 2008) from Hirosawanoike Pond were used. Circles indicate different 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.

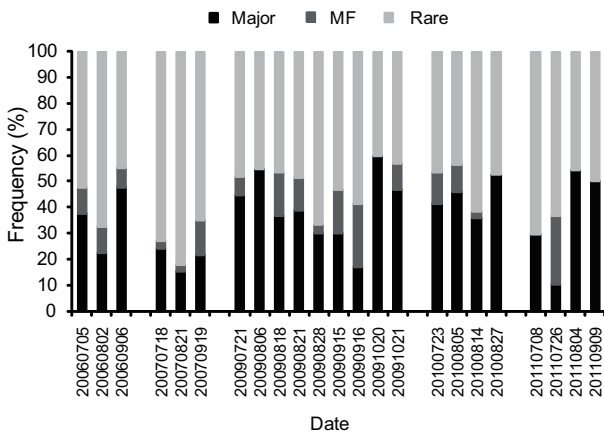


Figure 2. Changes in the proportion of the major, moderately-frequent, and rare genotypes in the tail sheath gene *g91* of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond. MF, Moderately-Frequent.

Kimura et al. Fig. 3

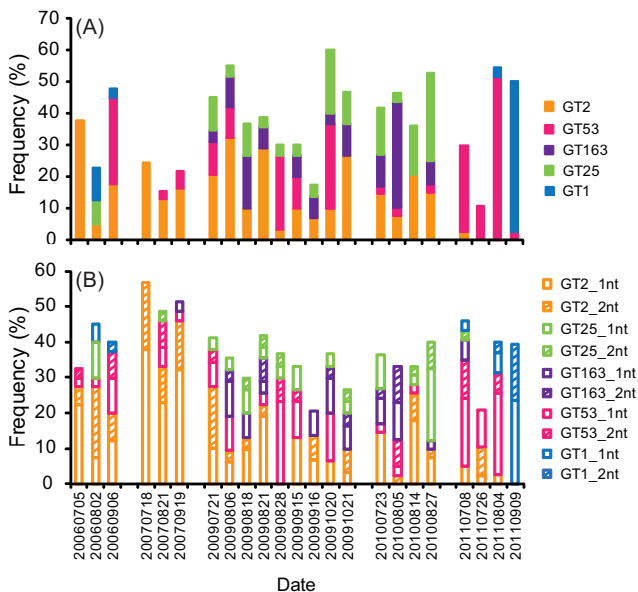


Figure 3. Changes in the proportion of the five major genotypes (A) and their variants (B) from the tail sheath gene *g91* of the Ma-LMM01-type phage populations during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.

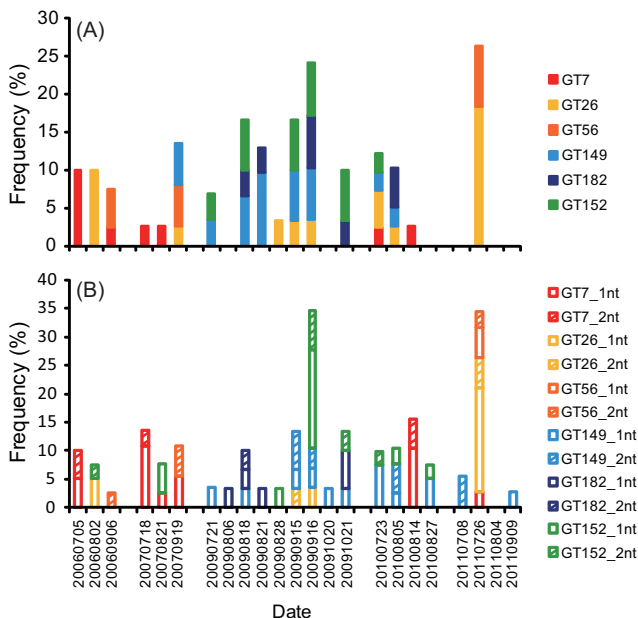


Figure 4. Changes in the proportion of the moderately-frequent genotypes (A) and their variants (B) in the tail sheath gene *g91* of the Ma-LMM01-type phage during the sampling period from 2006 to 2011 (except 2008) in Hirosawanoike Pond.

Group I



Group II



Group III



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.

Sampling date	<i>M. aeruginosa</i> (copies mL ⁻¹)	Temp. (°C)	DO (%)	No. of sequences	No. of genotypes	Coverage (%)	Shannon index	Chao 1 index	The proportion of genotypes (%)		
									M	MF	R
20060705	1.4×10 ⁷ ^a	26.6 ^a	-	40	23	48	2.54	233 (114-508)	38	10	53
20060802	3.2×10 ⁶ ^a	29.8 ^a	-	40	30	38	3.26	130 (62-341)	23	10	68
20060906	1.5×10 ⁶ ^a	25.2 ^a	-	40	23	50	2.65	118 (51-344)	48	8	45
20070718	8.5×10 ⁴ ^a	26.5 ^a	-	38	29	29	3.08	205 (85-582)	24	3	73
20070821	3.3×10 ⁵ ^a	31.9 ^a	-	39	33	23	3.39	178 (81-467)	15	3	82
20070919	1.4×10 ⁶ ^a	28.2 ^a	-	38	29	37	3.21	84 (48-191)	22	14	65
20090721	3.9×10 ⁷ ^b	26.9 ^b	70	29	20	41	2.77	156 (74-363)	45	7	48
20090806	6.2×10 ⁶	29.3	117	31	18	52	2.48	123 (57-299)	55	0	45
20090818	3.9×10 ⁶	32.0	160	30	20	50	2.82	55 (30-147)	37	17	47
20090821	2.0×10 ⁶ ^b	-	-	31	20	45	2.65	88 (39-261)	39	13	48
20090828	8.2×10 ⁷	27.0	166	30	23	30	2.90	128 (54-373)	30	3	67
20090915	6.1×10 ⁶ ^b	22.2 ^b	99	30	22	47	3.00	46 (29-103)	30	17	53

20090916	5.2×10 ⁶ ^b	23.1 ^b	146	30	25	33	3.17	57 (35-121)	17	24	59
20091020	2.5×10 ⁶ ^b	18.8 ^b	150	30	16	57	2.38	94 (43-239)	60	0	40
20091021	3.5×10 ⁶ ^b	18.2 ^b	133	30	18	53	2.58	64 (30-189)	47	10	43
20100723	4.4×10 ³	33.9	112	41	26	49	2.99	96 (48-254)	41	12	46
20100805	3.4×10 ⁵	-	-	39	24	46	2.69	129 (55-374)	46	10	44
20100814	2.4×10 ⁶	-	-	39	27	36	2.96	327 (166-675)	36	3	62
20100827	3.6×10 ⁵	32.5	176	40	23	50	2.68	213 (104-471)	53	0	48
20110708	4.0×10 ⁶	27.9	131	37	28	27	2.99	379 (195-765)	30	0	70
20110726	1.9×10 ⁷	-	-	38	27	37	3.05	303 (153-632)	11	26	63
20110804	5.1×10 ⁷	29.6	174	35	17	57	2.03	70 (31-210)	54	0	46
20110909	6.6×10 ⁷	28.2	>200	38	20	53	2.23	97 (42-286)	50	0	50

^aDetected by Yoshida *et al.* (28). ^bDetected by Kimura *et al.* (11). M, Major genotype; MF, Moderately-Frequent genotype; R, Rare genotype.
Dashes indicate no measured data.

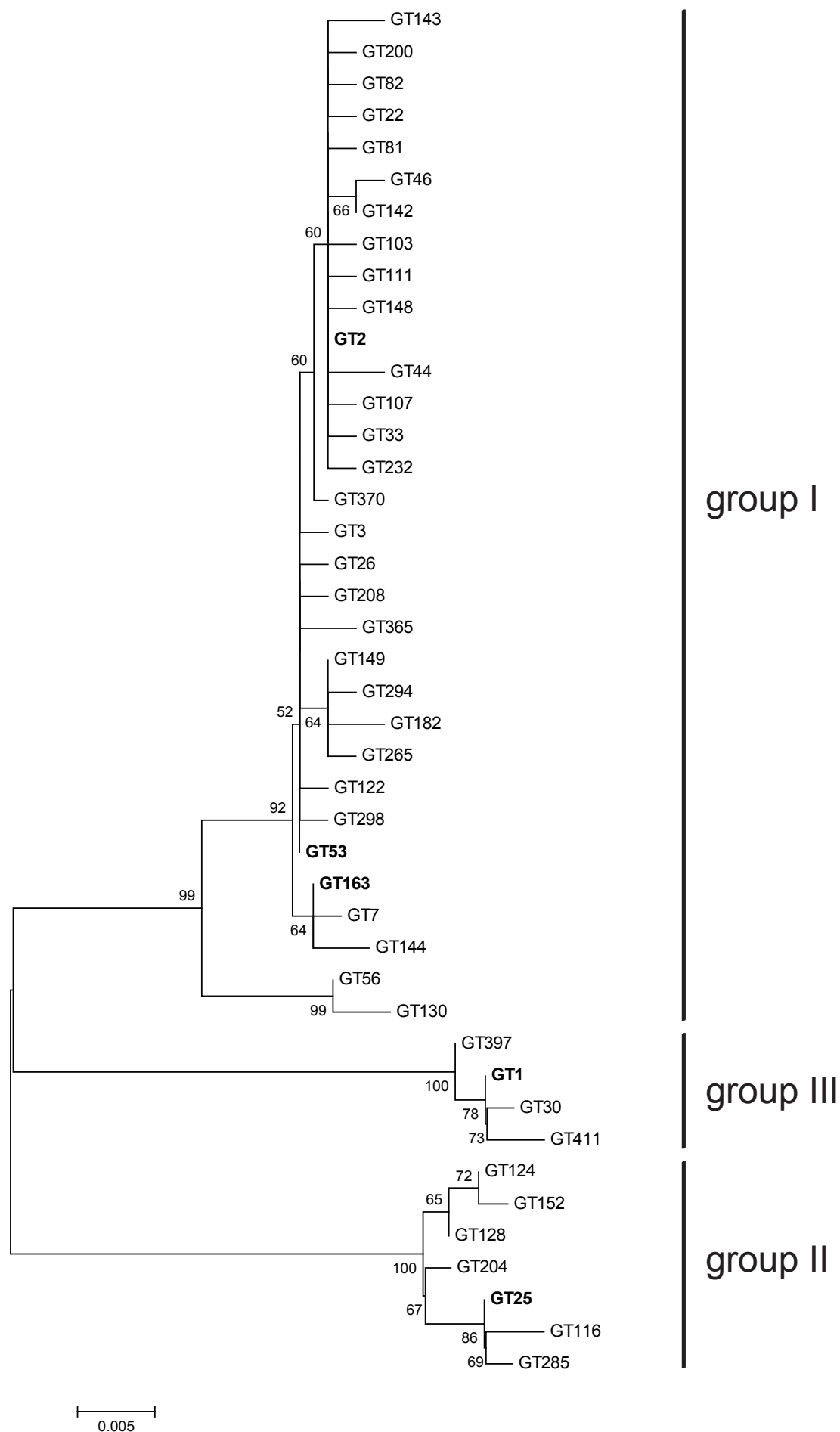


Figure S1. A neighbor-joining tree of 43 Ma-LMM01-type phage tail sheath gene genotypes that were found at least twice in samples from 2006 to 2011 (except 2008) in Hirosawanoike Pond. Bold indicates major genotypes. Bootstrap values are indicated at the nodes with more than 50 % bootstrap support.

