Title

Intricate Interactions between the Bloom-Forming Cyanobacterium *Microcystis aeruginosa* and Foreign Genetic Elements Revealed by Diversified CRISPR Signatures

Running title

CRISPR diversity in *Microcystis aeruginosa*

Journal section

Environmental microbiology

Authors

Sotaro Kuno[1], Takashi Yoshida[1*], Takakazu Kaneko[2] and Yoshihiko Sako[1].

Affiliations

Graduate school of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan[1].

Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-Ku, Kyoto 603-8555, Japan[2].

*Corresponding author. Mailing address: Faculty of Agriculture, 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

ABSTRACT

Clustered regularly interspaced short palindromic repeats (CRISPR) confer sequence-dependent, adaptive resistance in prokaryotes against viruses and plasmids via incorporation of short sequences, called spacer, derived from foreign genetic elements. CRISPR loci are thus considered to provide records of past infections. To describe the host-parasite (i.e. cyanophages and plasmids) interactions involving the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*, we investigated CRISPR in four *M. aeruginosa* strains and in two previously sequenced genomes. The number of spacers in each locus was larger compared to the average among prokaryotes. All spacers were strain-specific except a string of 11 spacers shared in two closely-related strains, suggesting diversification of the loci. Using CRISPR repeat-based PCR, 24 CRISPR genotypes were identified in a natural cyanobacterial community. Among 995 unique spacers obtained, only 10 sequences showed similarity to *M. aeruginosa* phage Ma-LMM01. Of these, six spacers showed only silent or conservative nucleotide mutations compared to Ma-LMM01 sequences, suggesting a strategy by the cyanophage to avert CRISPR immunity dependent on nucleotide identity. These results imply host-phage interactions can be divided into combinations of *M. aeruginosa*-cyanophage rather than pandemics of population-wide infectious cyanophages. Spacer similarity

43 also showed frequent exposure of *M. aeruginosa* to small cryptic plasmids that were
44 observed only in a few strains. Thus, the diversification of CRISPR implies *M.*
45 *aeruginosa* has been challenged by diverse communities (almost entirely
46 uncharacterized) of cyanophages and plasmids.
47
48

INTRODUCTION

Bacteria and archaea have acquired a large number of defense mechanisms against viruses, plasmids, and other mobile genetic elements (32, 37). One of the defense systems, termed clustered regularly interspaced short palindromic repeats (CRISPR), was recently identified in most archaeal and nearly half of the bacterial genomes (11). The CRISPR array comprises short (21- to 48-bp) direct repeats separated by similarly-sized variable unique sequences called spacers. CRISPR loci are typically accompanied by CRISPR-associated (*cas*) genes (19, 26, 36). Although the molecular mechanisms are far from understood, CRISPR/Cas is shown to function as a heritable, acquired defense system in some bacteria and archaea (3, 5, 10, 38). In general, the spacers are derived from partial sequences (proto-spacers) from virus and plasmid genomes that invade the cell, creating individual archives (i.e. spacers) of exposure to the non-cellular parasites (3, 10, 13). Spacers are transcribed from an upstream AT-rich region (leader) as a single long precursor RNA, and then processed by specific Cas proteins into small CRISPR RNAs (crRNAs) containing a single spacer (5, 6, 33). Each crRNA forms a nucleoprotein complex to guide Cas nuclease to recognize foreign DNA (5) or RNA (20) using base-paring. Through this RNA interference (RNAi)-like manner, the CRISPR/Cas system protects bacteria and archaea from viral

67 infection and limits plasmid transfer (3, 13, 38, 39). In individual CRISPR arrays, new
68 spacer additions are polarized to the leader-end of the loci (3, 10) while simultaneously
69 spacer loss occurs (10), resulting in hypervariability within species. Therefore, the loci
70 have been used for fingerprinting of pathogenic bacteria (i.e. spoligotyping (27)) and for
71 dissecting microbial population structures (22, 49). Further, CRISPR spacers,
72 representing direct signatures of viral infection, help to understand the co-evolutionary
73 dynamics of the host-virus community in environments (2, 21).

74 The unicellular cyanobacterium *Microcystis aeruginosa* frequently forms dense
75 blooms in freshwater environments worldwide (50). Some strains of this species
76 produce potent hepatotoxins called microcystins that occasionally cause death of
77 domestic animals and humans (8). Therefore, community composition and dynamics of
78 *M. aeruginosa* populations are of great concern to water quality management (50).
79 Previous studies have shown *M. aeruginosa* to be genetically highly heterogeneous (44,
80 51); *M. aeruginosa* populations also undergo temporal changes in genotypic
81 composition (4, 52, 53). Previously we isolated a phage Ma-LMM01 infecting *M.*
82 *aeruginosa* (57), and observed the potential qualitative impact of the cyanophage and its
83 relatives on natural *M. aeruginosa* populations (54, 56). However, Ma-LMM01 has a
84 narrow host range despite the large genetic diversity of hosts, suggesting there is a

85 greater diversity of host-cyanophage combinations other than Ma-LMM01 and its host
86 strain in natural *M. aeruginosa* populations (57). Recently, a comparative genomic study
87 of bacteria and archaea revealed an abundance of diverse defense systems including a
88 CRISPR/Cas system in the *M. aeruginosa* NIES843 genome (37). Together, these
89 findings imply richness of foreign genetic elements that have coevolved with *M.*
90 *aeruginosa* through their specific interactions that probably contribute to the large
91 clonal diversity (4, 51, 52) and genomic plasticity (12, 28) of the species.

92 Here we investigated the CRISPR to understand host-parasite (i.e. cyanophages
93 and plasmids) dynamics involving *M. aeruginosa*. Our aims are 1) to estimate diversity
94 of CRISPR spacer repertoire within *M. aeruginosa* populations, and 2) to imply
95 potential impact of known cyanophages and plasmids on *M. aeruginosa*. Therefore we
96 determined CRISPR sequences in four *M. aeruginosa* strains in addition to two strains
97 whose genomes are available in the databases, and examined intra-species variability of
98 the loci. In addition, we assessed a natural *M. aeruginosa* population for diversity in the
99 leader-end CRISPR fragments obtained using CRISPR repeat-based PCR. We inferred
100 previous host-parasite interactions within the *M. aeruginosa* populations from
101 signatures in the CRISPR.

102 **MATERIALS AND METHODS**

103 ***M. aeruginosa* strains and DNA extraction.**

104 *M. aeruginosa* strains NIES87, NIES102, NIES298, and NIES1067 were
105 obtained from the National Institute for Environmental Studies (Tsukuba, Japan). The
106 strains used in this study represent different phylogenetic groups in this species (Table
107 1). In a previous report using multilocus sequence typing (MLST), *M. aeruginosa*
108 strains were largely divided into five major clades (Group A-E) (44). *M. aeruginosa*
109 NIES102 and NIES843 are closely related to each other and fall into a well-supported
110 inner clade (within Group A). Strain NIES298 and NIES1067 fall into other
111 well-supported clades (Group B and Group D, respectively) (44). Strain NIES87 is not
112 included in the major clades, but this strain is interesting for carrying two plasmids (46).
113 Phylogenetic relationships of these strains are shown in Fig. S1. The strains were
114 maintained in CB medium (29) as previously described (31). Late-exponential phase
115 cultures (2-ml aliquots) were mildly sonicated to remove gas vesicles, and then
116 centrifuged at 3,000×g for 10 min (31). DNA was extracted from the cell pellets using
117 the Nucleon Phytopure genomic DNA extraction kit (GE Healthcare, Tokyo, Japan)
118 according to the manufacturer's instructions.

119 **Identification and Sequencing of CRISPR arrays.**

120 The nucleotide sequence for the CRISPR of *M. aeruginosa* NIES843 was

obtained from the CRISPRdb database (15). To identify homologous CRISPR/Cas systems in the draft genome sequence of *M. aeruginosa* PCC7806 (12), the NIES843 CRISPR repeat sequence was searched against 116 contig sequences deposited in the EMBL database using BLASTN (1). Homologous CRISPR arrays in another four *M. aeruginosa* strains (NIES87, NIES102, NIES298, and NIES1067) were amplified using a primer MaeCRf specific to leader region together with various reverse primers. All primer sequences used in this study are shown in Table S1. The loci were amplified from strains NIES87 and NIES102 using primer pairs MaeCRf/MaeCRrtp2 and MaeCRf/MaeCRrtp3, respectively. The PCR was performed in 25μl containing 50 ng DNA, 0.2 μM each primer, 0.4 mM each dNTP, 1×LA PCR Buffer, 2.5 mM $MgCl_2$ and 1.25 U of *TaKaRa LA Taq* polymerase (Takara Bio, Otsu, Japan). The reaction conditions were: 1 min of initial denaturing at 94°C followed by 30 cycles; 94°C for 20 s and 64°C for 20 min; and a final extension at 72°C for 10 min. Because of the diversity of sequences surrounding the loci, the conventional PCR failed to amplify CRISPR from strains NIES298 and NIES1067. To obtain the NIES298 CRISPR, we searched for short genomic contigs of NIES298 (unpublished data) that matched CRISPR spacers of NIES843. Based on one of the found matches, a primer 298CRrS6 was designed and used with primer MaeCRf under the above PCR conditions. For strain

NIES1067, a primer MaeCRrGT (described below and in Fig. 1A) was used with MaeCRf under the following PCR conditions: initial 94°C for 1 min followed by 30 cycles; 98°C for 10 s and 60°C for 20 min; and a final 72°C for 10 min. PCR products were purified and then sequenced using the primer walking method. The CRISPR array was completely sequenced for strain NIES102, while those for the other three strains were partially sequenced. Therefore, amplification and sequencing was performed using successive thermal asymmetric interlaced PCR (TAIL-PCR) (34) toward the end of the arrays (primers are shown in Table S1). Reaction conditions and arbitrary primers for the TAIL-PCR were as described previously (31). A portion of the amplifications were performed using alternative PCR using MaeCRrGT and outward primers based on CRISPR spacers in the sequenced fragments.

Amplification of *M. aeruginosa* CRISPR from a natural cyanobacterial population.

M. aeruginosa CRISPR was investigated in Hirosawanoike Pond (Kyoto, Japan), a small (surface area: 14 ha) and shallow (mean depth: 1.5 m) reservoir. One liter of surface water was collected at a fixed point (35°02' N, 135°41' E) at noon on 13 Sep 2010. Cyanobacterial cells were harvested from a 50-mL aliquot of the water by mild sonication followed by centrifugation at 1,680×g for 10 min. DNA was extracted from the cell pellet using the xanthogenate method as previously described (45, 52).

157 Purified DNA was suspended in 30 µl of sterilized milliQ water.

158 Because conventional PCR amplification was not applicable to *M. aeruginosa*
159 populations, we developed a PCR strategy based on the leader region (primer
160 MaeCRf2) and repeat-spacer units (primers MaeCRrGT or its derivative MaeCRrCA)
161 (Fig. 1A); thereby amplifying the leader-side fragments of *M. aeruginosa* CRISPR
162 irrespective of their genomic contexts. The PCR was performed in 50µl containing 2.5
163 µL 1:100 dilution of the environmental DNA, 0.8 µM each primer, 0.25 mM each dNTP,
164 1×*EX Taq* Buffer and 2 U *TaKaRa EX Taq* polymerase (Takara Bio). The reaction
165 conditions were: 94°C for 4 min followed by 30 cycles; 94°C for 30 s, 62°C for 1 min
166 and 72°C for 1min; and a final extension of 72°C for 7 min. The PCR products were
167 separated using electrophoresis on 2.0% (w/v) agarose S gels (Nippon Gene, Tokyo,
168 Japan). The gels were stained with GelRed (Biotium, CA) and visualized using the Gel
169 Doc XR system (Bio-Rad Laboratories, CA). To prevent bias of preferential cloning of
170 smaller DNA fragments, small gel blocks were separately excised, and DNA was
171 purified and cloned. Thirty-nine and 50 clones were sequenced in full from PCR using
172 MaeCRrGT and MaeCRrCA, respectively, using the primer walking method (primers
173 are shown in Table S1). The CRISPRtionary program (17) was used to find identical
174 spacers in different CRISPR fragments. Then, the fragments sharing the same spacer

175 order were manually assembled into contigs. To ensure accuracy, the leader-distal
176 spacer was removed from the contigs, because the PCR used in this study was shown to
177 allow one mismatch at the 3'-end additional dinucleotide of the reverse primers.

178 **Bioinformatics analysis.**

179 CRISPR repeats and spacers were identified using the CRISPRFinder (16) with
180 manual validation. A similarity search of the unique spacer sequences was performed
181 against the NCBI nr database using BLASTN with an E-value threshold of 0.1 and the
182 word size set at 7. The best hits for bacteriophages and plasmid sequences were
183 investigated and those showing ≥80% identity over the queried spacers were considered
184 to be significant. Partial phage/plasmid sequence including the sequence match and
185 covering the spacer length was referred to as a putative proto-spacer. Sequence logos
186 were generated with the WebLogo (9) using 10-bp flanking sequences on both sides of
187 the putative proto-spacers in phage Ma-LMM01 and *M. aeruginosa* plasmids (PMA1,
188 pMA1, and pMA2).

189 Separately, spacer sequences were clustered using CD-HIT-EST web server
190 (25) where spacers showing ≥87% identity over >60% of the shorter were clustered.

191 **Nucleotide sequences**

192 The nucleotide sequences determined in this study are deposited in the

193 DDBJ/EMBL/GenBank database. The accession numbers are: AB644436 to AB644439
194 for the CRISPR arrays from cultured strains, and AB644412 to AB644435 for the
195 representative clones (i.e. the longest one) of the CRISPR contigs obtained from the
196 environmental samples. Consensus sequences for CRISPR types CT1, CT2, CT4, CT5,
197 CT7, CT8, CT19 and CT22 to CT24 (see Result section) are provided as a supplemental
198 text.

199 **RESULTS**

200 **Type I-D CRISPR/Cas system in *M. aeruginosa*.**

201 A type I-D CRISPR/Cas system homologous to that in *M. aeruginosa* NIES843
202 was identified in the genomic contig C326 of PCC7806. Makarova et al. (2011)
203 identified "divergent rare variants" of *cas8*, *cas7*, and *cas5* along with the typical type
204 I-D *cas* genes in the NIES843 genome (37); however, the three *cas* variants and the
205 downstream *cas3* were absent in the PCC7806 draft genome (Fig. 2). The rare *cas*
206 module in NIES843 may have been acquired by horizontal gene transfer (HGT), given
207 its presence in some plasmids (e.g. *Cyanothece* sp. PCC8802 plasmid pP880201) (37).
208 In PCC7806, *cas1*, presumably involved in spacer acquisition (11), was interrupted by
209 an in-frame stop codon, suggesting that spacer uptake may be no longer active in this
210 strain. The sequence between *cas2* and the first CRISPR repeat was highly conserved

211 between the strains and contained AT-rich regions including a putative promoter element
212 (5'-TTGAAG-17bp-TAYRAT-3'). Therefore the sequence was considered to be a leader
213 (26, 33). The CRISPR/Cas was located in different genomic contexts in the two strains
214 (Fig. 2).

215 **Variation in genomic position of the CRISPR arrays.**

216 The CRISPR arrays were sequenced for another four *M. aeruginosa* strains
217 (NIES87, NIES102, NIES298, and NIES1067), from *cas2* to the downstream flanking
218 sequences of the loci (Fig. 3). The flanking sequences of the loci in NIES102 and
219 NIES298 were nearly identical to those in NIES843 and PCC7806, respectively, while
220 the CRISPR locus of NIES1067 was located at another genomic position. We could not
221 determine the position of the CRISPR for NIES87 because of two insertion sequences
222 (ISs) lying at the end of the locus. CRISPR arrays in strains NIES843 and NIES87 have
223 been subjected to transposable elements including ISs, miniature inverted-repeat
224 transposable element (MITE), and other putative short sequence elements (Fig. 3).
225 These mobile elements may contribute to the variation of the CRISPR position.

226 **Sequence analysis of the CRISPR repeats.**

227 Although the CRISPR locality varied among these strains, CRISPR repeat,
228 leader, and partial *cas2* sequences were nearly identical among the strains. Thus,

homologous, comparable CRISPR arrays were obtained from six *M. aeruginosa* strains (Fig. 3). The consensus sequence of the CRISPR repeats was 37-bp long (5'-GTTCCAATTAATCTTAAACCCTATTAGGGATTGAAAC-3') and fell into cluster 5 of the previously proposed repeat-based classification (30). Strains NIES87, NIES298, and NIES1067 had repeat variants derived from single nucleotide polymorphisms (SNPs) that were situated in the hairpin loop or external bases of the predicted stem-loop RNA structure of the repeat. In general, terminal repeats tend to be degenerate at the leader-distal end of the CRISPR loci (24). Among the *M. aeruginosa* strains, an identical degenerate terminal repeat was found in strains NIES102 and NIES843 (Fig. 3).

Intra-species variability of the CRISPR spacers.

The number of spacers varied among the *M. aeruginosa* strains, from 47 (NIES102) to 174 (NIES843) (Fig. 3). These numbers are significantly larger relative to the average (27 repeats) among prokaryotes (14).

CRISPR spacer repertoire was compared in the context of phylogenetic relationships among the six strains. Between strains NIES102 and NIES843 (Fig. S1), a string of 11 spacers were shared at the leader-side of the loci, while the rest were strain-specific (Fig. 3). The other four strains, including the closest of the

NIES102/NIES843 clade (strain NIES1067), had only strain-specific spacers (Fig. 3).

Diversity of the leader-end CRISPR fragments in natural cyanobacterial populations.

To investigate the CRISPR diversity in a natural population, we used the repeat-based PCR with a pond water sample. The PCR yielded multiple bands ranging from ca. 150-bp to 3,000-bp at intervals of 70-bp that is consistent with the size of a repeat-spacer unit (Fig. 1B). Every sequence from 89 PCR clones contained a leader region and subsequent CRISPR repeats that are homologous to those of *M. aeruginosa* (with up to two SNPs), confirming specificity of the PCR. Clones containing the same spacers in the same order were assembled into contigs, resulting in 24 distinct leader-end CRISPR fragments (Table 2). We designate them as "CRISPR types (CTs)", which probably represent distinct CRISPR genotypes. Up to 32 spacers (13 on average) were obtained for each CT (Table 2). No spacers were shared between the CRISPR types, except for three pairs of CTs sharing a portion of their spacer sets (Table 2). Separately, we determined partial CRISPR sequences from eight *M. aeruginosa* strains isolated from the same pond water sample. The eight strains were identified as *M. aeruginosa* based on specific PCR targeting partial 16S-23S rDNA internal transcribed spacers (55). Two strains had a spacer set identical to CT1, another two strains were

265 CT4, and another one strain was CT3. Origins of the CTs obtained by the repeat-based
266 PCR were therefore partially confirmed. The other three strains had spacer sets different
267 from the CTs (data not shown). Considering the general trends of polarized spacer
268 addition, our data indicates there are no less than 24 coexisting genotypes that are
269 different in the recent evolutionary history of the CRISPR in the *M. aeruginosa*
270 population.

271 **CRISPR signatures of foreign DNA elements.**

272 Excluding redundant spacers, we obtained a total of 995 unique spacer
273 sequences from *M. aeruginosa* strains and the water sample. We determined the
274 sequence matches comparing the 995 spacers to sequences of phages and plasmids in
275 the NCBI nr database to show histories of host-parasite interaction. Overall, only 43
276 unique spacers (4%) had significant matches, of which 10 and 33 spacers matched
277 phage and plasmid sequences, respectively (Table S2).

278 The 10 spacers had 83-97% identity to the genome sequence of *M. aeruginosa*
279 phage Ma-LMM01. Of these, four spacers were identified in the host strain, NIES298
280 (Fig. 3). The other six spacers were identified in strain PCC7806 and CRISPR types
281 CT6, CT15, and CT19 (Fig. 3, Table 2).

282 Ten, 13, and seven unique spacers showed sequence matches to plasmids

PMA1 from *M. aeruginosa* HUB 5-2-4 (42), pMA1, and pMA2 from NIES87 (46), respectively (Fig. 3, Table 2). Of these, four spacers showed 100% identity to sequences in the corresponding plasmids. Putative proto-spacers were evenly mapped onto each plasmid sequence (Fig. 4). The other three spacers showed moderate sequence similarity across genera to *Cyanothece* sp. PCC7424 plasmid pP742402, *Streptococcus thermophilus* LMD-9 plasmid 1, and *Bacillus coagulans* plasmid pMSR0 (86, 82, and 82%, respectively).

Independent acquisition of similar spacers.

Clustering analysis of the 995 spacer sequences identified 48 distinct pairs and four triads of spacers that share nearly identical sequences (Table S3). In each pair, the similar spacers exhibited overlapping that could be merged into a contig. Further, in 19 cases the paired spacers were complementary (Table S3). Therefore, similar spacers in each pair may be derived from the same viral (or plasmid) lineages and be acquired in separate exposure incidents. The similar spacer pairs were identified in various combinations of strains and CTs; e.g. *M. aeruginosa* NIES298 shared 14 similar spacers with 5 other strains and 4 CTs.

PAMs associated with the *M. aeruginosa* CRISPR/Cas system.

We determined if the *M. aeruginosa* CRISPR/Cas system associates a specific

proto-spacer associated motif (PAM), a short nucleotide motif adjacent to the proto-spacers in target sequences (40). The WebLogo analysis of the flanking sequences identified a conserved motif of GTY immediately upstream of the putative proto-spacers (Fig. 5). No particular motif was detected when searching the downstream sequences.

DISCUSSION

CRISPR variability and inferred host-parasite interactions.

At present, we have no experimental evidence to show whether, and if so, how the *M. aeruginosa* CRISPR/Cas functions as a defense system against parasites. However, we found several spacers matching known foreign genetic elements for *M. aeruginosa* (e.g. Ma-LMM01) in the CRISPR loci. This strongly suggests the *M. aeruginosa* CRISPR/Cas had been functional at least in the past, and thereby the spacer repertoire at each locus represents a history of previous host-parasite interactions. In our dataset the spacer repertoire was unique for each *M. aeruginosa* strain (Fig. 3). This result should be interpreted with care because we cannot rule out the possibility of under-sampling given the high diversity of *M. aeruginosa*. However, considered in the context of local populations (e.g. coexisting genotypes in Hirosawanoike pond), CRISPR variability can provides insights into the interplay between hosts and parasites (especially cyanophages) involving *M. aeruginosa*.

CRISPR spacers are believed to be acquired by individuals and then selected in response to viral infection. Therefore, intra-population variability in the spacer repertoire implies extent (i.e. host range of viruses) and frequency of selection events posed on the host population (21, 22, 47, 49). For example, CRISPR sequence of *Leptospirillum* in acidophilic microbial biofilms showed that individuals in a nearly clonal population share spacers in the leader-distal half of the CRISPR arrays, suggesting population-wide selective sweep events (49). In contrast, we found 24 coexisting different CRISPR genotypes in a pond *M. aeruginosa* population, where spacers were rarely shared between the leader-end portions of the individual CRISPR types (Table 2). The lack of population-wide fixed spacers may be also suggested in another population (Lake Kasumigaura), where *M. aeruginosa* strains (NIES87, NIES298, and NIES102) isolated in Sep 1982 (Table 1) shared no single spacer (Fig. 3). These data suggests purifying selection is unlikely to be so extensive that only strains carrying specific spacers can survive. In other words, the host-phage interaction may be subdivided into diverse "susceptible combinations" each consisting of *M. aeruginosa* strains and specific cyanophage, rather than pandemics of population-wide infectious cyanophages. Sporadic distribution of the paired similar spacers among *M. aeruginosa* (Table S3) also supports the subdivided host-phage interactions.

A recent modeling study concerning CRISPR evolution showed CRISPR immunity induces allele (spacers and proto-spacers) diversification within a community of a host and a viral lineage, given that each virus carries a number of distinct, variable proto-spacers (7). Our results indicate Ma-LMM01-matching spacers could be derived from a number of different proto-spacers, and the sequences suggest proto-spacer diversification in the phage (discussed below). Further, the data from closely-related strains NIES102 and NIES843 allow inference concerning diversification of a CRISPR locus in *M. aeruginosa*. The strains NIES102 and NIES843, both isolated from Lake Kasumigaura, shared an identical degenerate terminal repeat, genomic location of CRISPR, and a small portion of their spacer repertoires (Fig. 3). This may suggest the two strains share a common ancestor at the CRISPR loci (i.e. involved in the same "susceptible combination"). However, these strains have diverged to show more than a hundred strain-specific spacers (Fig. 3), and the shared spacer block was not polarized to the leader-distal end. If apparent rate of spacer addition is constant in the two strains over time, their spacer repertoires indicate addition of 127 spacers in NIES843 during the 15 years after isolation of NIES102. Assuming host growth rate of 1 per day and bloom period of 6 months per year, spacer addition rate is approximately one spacer per 20 generations. Although this estimate is based on many assumptions, the timescale is

355 compatible with that of host genotype turnover (13 turnover events in 200 generations)
356 observed in the modeling study (7). Very few studies have examined the rate of spacer
357 addition into natural bacterial populations, and addressing this issue will help to
358 understand host-parasite co-evolutionary dynamics more clearly and deeply.

359 CRISPR array was significantly longer in *M. aeruginosa* than other organisms.
360 A modeling study where host density is kept constant predicts that larger viral diversity
361 leads to longer host CRISPR arrays (18). Therefore, the long CRISPR of *M. aeruginsa*,
362 which grows up to form dense bloom, may reflect adaptive response to highly diverse,
363 quickly diversifying cyanophages. High cyanophage diversity is also compatible with a
364 huge variety of antiviral defense systems on the genome of *M. aeruginosa* NIES843
365 (37).

366 **The ecological impact of cyanophages and cryptic plasmids**

367 CRISPR signatures of *M. aeruginosa* provided insights into the ecological
368 impact and dynamics of known cyanophage. Spacers matching cyanophage
369 Ma-LMM01 were found in *M. aeruginosa* NIES298 as well as in PCC7806 from the
370 Netherlands (Fig. 3) and three CTs from Hirosawanoike Pond (Table 2). This implies
371 the host-virus combinations involving Ma-LMM01 and possible dispersal of its related
372 phages. Conversely, all of the Ma-LMM01-matching spacers (10 spacers) were not

373 completely identical to their corresponding putative proto-spacers in Ma-LMM01,
374 suggesting high nucleotide diversity within this phage lineage. Of these, six spacers had
375 only nucleotide mutations that are translated into silent or conservative changes in
376 deduced amino acid sequences (Table 3). Although a single mutation in proto-spacers or
377 PAMs basically can abolish CRISPR-mediated immunity against phages (3, 10, 13), the
378 strict nucleotide identity required for the immunity is limited to a specific
379 seven-nucleotide region in the proto-spacers (denoted as "seed sequence") in
380 *Escherichia coli* (43). If the *M. aeruginosa* CRISPR system employs this
381 "seed"-dependent immunity, Ma-LMM01 or its related phages may evade interference
382 mediated by five of the spacers (all but CT15spc14 in Table 3), and thereby the
383 Ma-LMM01-related phage assemblage retains the conservative nucleotide
384 polymorphisms to circumvent host immunity.

385 CRISPR signatures of *M. aeruginosa* also showed repeated exposures to
386 plasmids PMA1, pMA1, and pMA2. Uniform distribution of putative proto-spacers
387 throughout these plasmids (Fig. 4) suggests the recipient CRISPR recognized the
388 plasmids themselves transferred from potential donors, rather than other plasmids
389 sharing specific components (e.g. conserved replication genes). The proposed
390 interference mechanisms of the CRISPR/Cas systems predict that spacers completely

391 matching a plasmid prevent it from establishing in the host (13, 39). In accordance, *M.*
392 *aeruginosa* NIES843 possessing a spacer with 100% identity to plasmid pMA1 (Fig. 3)
393 carries no plasmid (28). Unstable presence of the small plasmids reported in some *M.*
394 *aeruginosa* strains (42) may be attributed to such CRISPR-mediated exclusion.
395 However, strain NIES87 retains pMA2 (46), despite possessing a spacer with 100%
396 identity to the plasmid (Fig. 3). This apparent conflict may have resulted from defective
397 CRISPR interference given the proliferation of IS elements around the CRISPR array in
398 NIES87 (Fig. 3). The small plasmids carry very few genes, and thus their role in
399 ecology and evolution of *M. aerugninosa* remains unclear. Interestingly, similar small
400 cryptic plasmids were found to dominate in marine *Synechococcus* metagenomes and
401 were hypothesized to facilitate HGT and in some case phage resistance (35, 41).
402 Considered with the suggested spread of the small plasmids among *M. aeruginosa*
403 populations, they may facilitate host genetic diversity via HGT or chromosomal
404 integration.

405 **Functional characteristics of *M. aeruginosa* CRISPR.**

406 *M. aeruginosa* and several other cyanobacteria possess the recently identified
407 subtype I-D CRISPR/Cas system, which is a hybrid of type-I system and type-III
408 executive Cas module (36). Our data provides some information on the functional

409 characteristics of the *M. aeruginosa* CRISPR/Cas system. Uniform distribution of the
410 putative proto-spacers without bias toward either strand of the plasmids (Fig. 4)
411 suggests the *M. aeruginosa* CRISPR/Cas recognizes DNA rather than RNA. Combined
412 with the presence of the upstream PAM (Fig. 5), this is in accordance with the type-I
413 information processing (i.e. spacer acquisition) system (36).

414 **Conclusions.**

415 *M. aeruginosa* shows a high degree of CRISPR heterogeneity within
416 populations. Thus we infer that host-phage community can be subdivided into a number
417 of different "susceptible combinations" of *M. aeruginosa* and cyanophages. In each
418 combination, the CRISPR spacers and cyanophages may quickly diversify through
419 co-evolution. This intricate interaction is expected to shape complex host-phage
420 community. The spacer sequences imply what kinds of phages are involved in each
421 "susceptible combination" with a strain of *M. aeruginosa* (Fig. 3, Table S3).
422 Nevertheless, a significant fraction of the spacer sets in *M. aeruginosa* are of unknown
423 origin, largely because of the under-representation of the diversity of bacteriophages
424 and plasmids in the current sequence data. Thus, *M. aeruginosa* may be challenged by
425 diverse, almost entirely uncharacterized communities of cyanophages and plasmids.
426 Some of the unknown spacers would be attributed to previously observed (but not

isolated) *M. aeruginosa* cyanophages (23, 48), and a metagenomic survey will provide links between the CRISPR signatures and uncultured phage community.

Acknowledgements

This study was supported by a Grant-in-Aid for Science Research (No. 20310045). SK is a research fellow supported by JSPS for Young Scientist (No. 224469).

Figure Legends

FIGURE 1. Amplification of leader-end CRISPR fragments. (A) The reverse primer MaeCRrGT was designed to be complementary to the 5'-end 20-b of the repeat (R) but with two additional nucleotides, GT, thereby preferentially annealing to a limited number of specific spacer-repeat units. Another reverse primer, MaeCRrCA, has dinucleotide CA instead of GT. (B) PCR products from primers MaeCRrGT (left) and MaeCRrCA (right). Mar, 2-log DNA ladder; Env, DNA from Hirosawanoike Pond; 843, *M. aeruginosa* NIES843 genomic DNA.

FIGURE 2. Comparison of the CRISPR/Cas systems identified in *M. aeruginosa* NIES843 and PCC7806. ORFs and CRISPR arrays are shown as arrows and filled

445 boxes, respectively. *Cas* genes, toxin-antitoxin genes, and transposase genes are
446 indicated by gray, dotted, and striped arrows, respectively. Divergent rare *cas* variants
447 (37) are enclosed by thick lines. Gene nomenclature is in accordance with the previous
448 study (37). Pseudogenes are marked with asterisks. Conserved regions between the two
449 genomes are indicated by gray shadows.
450
451 **FIGURE 3. CRISPR arrays in six *M. aeruginosa* strains.** Strain-specific spacers are
452 indicated by circles, of which those with significant hits to sequences of phage and
453 plasmids are pattern-coded as indicated in the figure. * indicates 100% nucleotide
454 identity to phage/plasmid sequences. A string of 11 consecutive spacers shared between
455 strains NIES102 and NIES843 is indicated by rectangles with shading showing the
456 direction of the shared spacer block. The total number of the spacers is shown at the
457 right of each array. Degenerate terminal repeats are indicated by black bars. Spacer
458 duplication is indicated by open boxes. Insertion of ISs, MITE and other short sequence
459 elements, in CRISPR repeats are indicated by triangles.
460
461 **FIGURE 4. Distribution of proto-spacers on plasmids PMA1 (A), pMA1 (B) and**
462 **pMA2 (C).** Each bar indicates the position of the putative proto-spacer. Solid and

463 dashed bars indicate putative proto-spacers in clockwise and counterclockwise strands,
464 respectively. * indicates 100% nucleotide identity. Note we found additional ORFs in
465 pMA1 (ORF b) and pMA2 (ORFs a, b, e, g and h) compared with their original
466 description. ORF a and b of pMA1 may be generated from a longer replication gene by
467 a frame-shift. ORF g, h, a, b, and c may be generated by frame-shifts from a single
468 hypothetical gene whose homologues are prevalent in cyanobacteria (e. g.
469 CY0110_05002 of *Cyanothece* sp. CCY0110).
470
471 **FIGURE 5. Sequence logo of the PAM consensus.** The logo was built using the
472 WebLogo based on alignments of the flanking sequences of putative proto-spacers in
473 Ma-LMM01 and PMA1, pMA1 and pMA2. Numbers below indicate nucleotide
474 positions where -1 is just upstream of the putative proto-spacers.
475
476
477

478 **Tables**

479 TABLE 1. *M. aeruginosa* strains used in this study

480

Strain name[a]	Locality	Year	Sequence type[b]
NIES102	Lake Kasumigaura, Japan	1982	10
NIES298	Lake Kasumigaura, Japan	1982	60
NIES87	Lake Kasumigaura, Japan	1982	2
NIES1067	Chikatou Pond, Japan	1982	27
NIES843*	Lake Kasumigaura, Japan	1997	18
PCC7806*	Braakman Reservoir, The Netherlands	1972	N.D.

481 a *, Genome sequences are published in the NCBI/EMBL/DDBJ database.

482 b Sequence type were determined by Tanabe et al. (2007) (44). N.D., not determined.

483

484

485 TABLE 2. Characteristics of CRISPR types (CTs) identified in Hirosawanoike Pond.

CT	No. of clones	No. of spacers[a]		No. of unique spacers with significant hit[c]			
		Total	Shared[b]	Ma-LMM01	PMA1	pMA1	pMA2
CT1	8	16		-	4(1)	-	1
CT2	15	13	10(4-13)[A]	-	-	-	-
CT3	1	12		-	-	-	-
CT4	10	27	1(1)[B]	-	-	-	-
CT5	20	32		-	-	5	-
CT6	1	5		1	-	-	-
CT7	2	4		-	-	-	-
CT8	2	10		-	-	-	-
CT9	1	6		-	-	-	1
CT10	1	6		-	1	-	-
CT11	1	8		-	-	-	-
CT12	1	21	6(1-3, 6-8)[C]	-	-	-	-
CT13	1	15		-	-	-	-
CT14	1	7		-	-	-	-
CT15	1	28		3	-	-	-
CT16	1	11		-	-	-	-
CT17	1	13	1(1)[B]	-	-	-	-
CT18	1	13		-	-	-	-
CT19	3	17		1	-	-	-
CT20	1	17		-	-	-	-
CT21	1	16	6(1-3, 4-6)[C], 10(7-16)[A]	-	-	-	-
CT22	6	4		-	-	-	-
CT23	6	11		-	-	-	-
CT24	3	9		-	-	-	-

486 [a] Number of spacers in CT contigs.

487 [b] Spacers shared between CTs are indicated as $x(y\text{-}z)^N$. x is total number of shared

488 spacers, y-z is position of shared spacers (sequentially numbered from the leader-end

489 spacer), and N denotes a pair of strain sharing the spacers.

490 [c] -, Not found in sequenced contigs. Number of 100% match is shown in parenthesis.

491

492

TABLE 3. Spacers showing silent or conservative mutations compared to Ma-LMM01 putative proto-spacers.

spacer/phage gene[a]	nucleotide sequence[b]	predicted amino acid sequence[b]
298spc124[#]	AGTGGCGCGGCTACTTATCTCTACCAATTTTCTAC	SGAATYLYQFS
ORF40	.....**A**..**C**.....**G***G*.....**C**..	.....*V*.
7806spc34[#]	ATTTGAGGGACTAAATAATGGGATCGTATTCAAT	FEGLNNGIVFN
ORF41	.....**C***GC*...**C**.....*G*.	...*A*.....*S*
CT6spc2[#]	AATCCCCCCGTCAGGGATTCTCCCACGGGTTTCAAT	IPPSGILPRVS
ORF61	...**A**.....**A***G*.....**T**..	.....*V*.....
CT15spc12[#]	ACTCTCCTTGCGACTATAAGTATGTGGGCAAGTCT	SPCDYKYVGKS
ORF25	.....**T**.....	.....
CT15spc13[#]	TCTATCTGTTCAATACTATGCCTCTAGGAGCAGGGCAAG	YLFNTMPLGAGQ
ORF20	.....**G**.	.....
CT15spc14[#]	TTGATACAGGTGCCTTCCTAGGCTGTTATCT	DTGAFLGCY
ORF25	.....**G**.....	.....

[a] In each pair of rows, spacer and corresponding putative proto-spacer are shown in upper and lower, respectively. [#], Reverse complementary sequences are shown.

[b] Identical nucleotides and amino acids are indicated by ".". Synonymous and conservative changes are indicated in bold and italic letters, respectively. Seed region identified in the subtype I-E CRISPR of *E. coli* (43) is indicated by boxes.

REFERENCES

1. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403-410.

2. **Andersson AF, Banfield JF.** 2008. Virus population dynamics and acquired virus resistance in natural microbial communities. Science **320:**1047-1050.

3. **Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P.** 2007. CRISPR provides acquired resistance against viruses in prokaryotes. Science **315:**1709-1712.

4. **Bozarth CS, Schwartz AD, Shepardson JW, Colwell FS, Dreher TW.** 2010. Population turnover in a *Microcystis* bloom results in predominantly nontoxigenic variants late in the season. Appl. Environ. Microbiol. **76:**5207-5213.

5. **Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, Dickman MJ, Makarova KS, Koonin EV, van der Oost J.** 2008. Small CRISPR RNAs guide antiviral defense in prokaryotes. Science **321:**960-964.

6. **Carte J, Pfister NT, Compton MM, Terns RM, Terns MP.** 2010. Binding and cleavage of CRISPR RNA by Cas6. RNA **16:**2181-2188.

7. **Childs LM, Held NL, Young MJ, Whitaker RJ, Weitz JS.** 19 March 2012. Multiscale model of CRISPR-induced coevolutionary dynamics: diversification at the interface of Lamarck and Darwin. Evolution doi: 10.1111/j.1558-5646.2012.01595.x.

8. **Codd GA, Lindsay J, Young FM, Morrison LF, Metcalf JS.** 2005. Harmful cyanobacteria, p 1-23. *In* J. Huisman, H. C. P. Matthijs, and P. M. Visser (ed.), Harmful Cyanobacteria. Springer, Dordrecht, Netherlands.

9. **Crooks GE, Hon G, Chandonia J-M, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188-1190.

10. **Deveau H, Barrangou R, Garneau JE, Labonte J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S.** 2008. Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. J. Bacteriol. **190:**1390-1400.

11. **Deveau H, Garneau JE, Moineau S.** 2010. CRISPR/Cas system and its role in phage-bacteria interactions. Annu. Rev. Microbiol. **64:**475-493.

12. **Frangeul L, Quillardet P, Castets A-M, Humbert J-F, Matthijs HCP, Cortez D, Tolonen A, Zhang C-C, Gribaldo S, Kehr J-C, Zilliges Y, Ziemert N, Becker S, Talla E, Latifi A, Billault A, Lepelletier A, Dittmann E, Bouchier**

C, de Marsac NT. 2008. Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. BMC genomics **9:**274.

13. **Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S.** 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature **468:**67-71.

14. **Godde JS, Bickerton A.** 2006. The repetitive DNA elements called CRISPRs and their associated genes: evidence of horizontal transfer among prokaryotes. J. Mol. Evol. **62:**718-729.

15. **Grissa I, Vergnaud G, Pourcel C.** 2007. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics **8:**172.

16. **Grissa I, Vergnaud G, Pourcel C.** 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **35:**W52-57.

17. **Grissa I, Vergnaud G, Pourcel C.** 2008. CRISPRcompar: a website to compare clustered regularly interspaced short palindromic repeats. Nucleic Acids Res. **36:**W145-148.

18. **Haerter JO, Trusina A, Sneppen K.** 2011. Targeted bacterial immunity buffers phage diversity. J. Virol. **85:**10554-10560.

19. **Haft DH, Selengut J, Mongodin EF, Nelson KE.** 2005. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. PLoS Comput. Biol. **1:**e60.

20. **Hale CR, Zhao P, Olson S, Duff MO, Graveley BR, Wells L, Terns RM, Terns MP.** 2009. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell **139:**945-956.

21. **Held NL, Whitaker RJ.** 2009. Viral biogeography revealed by signatures in *Sulfolobus islandicus* genomes. Environ. Microbiol. **11:**457-466.

22. **Held NL, Herrera A, Cadillo-Quiroz H, Whitaker RJ.** 2010. CRISPR associated diversity within a population of *Sulfolobus islandicus*. PLoS One **5:**e12988.

23. **Honjo M, Matsui K, Ueki M, Nakamura R, Fuhrman JA, Kawabata Z.** 2006. Diversity of virus-like agents killing *Microcystis aeruginosa* in a hyper-eutrophic pond. J. Plankton Res. **28:**407-412.

24. **Horvath P, Romero DA, Coûté-Monvoisin A-C, Richards M, Deveau H, Moineau S, Boyaval P, Fremaux C, Barrangou R.** 2008. Diversity, activity,

576 and evolution of CRISPR loci in *Streptococcus thermophilus*. J. Bacteriol.
577 **190:**1401-1412.
578 25. **Huang Y, Niu B, Gao Y, Fu L, Li W.** 2010. CD-HIT Suite: a web server for
579 clustering and comparing biological sequences. Bioinformatics **26:**680-682.
580 26. **Jansen R, Embden JD, Gaastra W, Schouls LM.** 2002. Identification of genes
581 that are associated with DNA repeats in prokaryotes. Mol. Microbiol.
582 **43:**1565-1575.
583 27. **Kamerbeek J, Schouls L, Kolk A, van Agterveld M, van Soolingen D,**
584 **Kuijper S, Bunschoten A, Molhuizen H, Shaw R, Goyal M, van Embden J.**
585 1997. Simultaneous detection and strain differentiation of *Mycobacterium*
586 *tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. **35:**907-914.
587 28. **Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M,**
588 **Nakamura Y, Kasai F, Watanabe A, Kawashima K, Kishida Y, Ono A,**
589 **Shimizu Y, Takahashi C, Minami C, Fujishiro T, Kohara M, Katoh M,**
590 **Nakazaki N, Nakayama S, Yamada M, Tabata S, Watanabe MM.** 2007.
591 Complete genomic structure of the bloom-forming toxic cyanobacterium
592 *Microcystis aeruginosa* NIES-843. DNA Res. **14:**247-256.
593 29. **Kasai F, Kawachi M, Erata M, Watanabe MM.** 2004. NIES-Collection List of

Strains, 7th ed. The National Institute for Environmental Studies, Tsukuba, Japan.

30. **Kunin V, Sorek R, Hugenholtz P.** 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. Genome Biol. **8:**R61.

31. **Kuno S, Yoshida T, Kamikawa R, Hosoda N, Sako Y.** 2010. The distribution of a phage-related insertion sequence element in the cyanobacterium, *Microcystis aeruginosa*. Microbes Environ. **25:**295-301.

32. **Labrie SJ, Samson JE, Moineau S.** 2010. Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. **8:**317-327.

33. **Lillestøl RK, Redder P, Garrett RA, Brügger K.** 2006. A putative viral defence mechanism in archaeal cells. Archaea **2:**59-72.

34. **Liu YG, Whittier RF.** 1995. Thermal asymmetric interlaced PCR: automatable amplification and sequencing of insert end fragments from P1 and YAC clones for chromosome walking. Genomics **25:**674-681.

35. **Ma Y, Paulsen IT, Palenik B.** 2012. Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. Environ. Microbiol.

36. **Makarova KS, Haft DH, Barrangou R, Brouns SJJ, Charpentier E, Horvath P, Moineau S, Mojica FJM, Wolf YI, Yakunin AF, van der Oost J,**

612 **Koonin EV.** 2011. Evolution and classification of the CRISPR-Cas systems. Nat.
613 Rev. Microbiol. **9:**467-477.
614 37. **Makarova KS, Wolf YI, Snir S, Koonin EV.** 2011. Defense islands in bacterial
615 and archaeal genomes and prediction of novel defense systems. J. Bacteriol.
616 **193:**6039-6056.
617 38. **Manica A, Zebec Z, Teichmann D, Schleper C.** 2011. *In vivo* activity of
618 CRISPR-mediated virus defence in a hyperthermophilic archaeon. Mol.
619 Microbiol. **80:**481-491.
620 39. **Marraffini LA, Sontheimer EJ.** 2008. CRISPR interference limits horizontal
621 gene transfer in staphylococci by targeting DNA. Science **322:**1843-1845.
622 40. **Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C.** 2009.
623 Short motif sequences determine the targets of the prokaryotic CRISPR defence
624 system. Microbiology **155:**733-740.
625 41. **Palenik B, Ren Q, Tai V, Paulsen IT.** 2009. Coastal *Synechococcus*
626 metagenome reveals major roles for horizontal gene transfer and plasmids in
627 population diversity. Environ. Microbiol. **11:**349-359.
628 42. **Schwabe W, Weihe A, Börner T, Henning M, Kohl J-G.** 1988. Plasmids in
629 toxic and nontoxic strains of the cyanobacterium *Microcystis aeruginosa*. Curr.

Microbiol. **17:**133-137.

43. **Semenova E, Jore MM, Datsenko KA, Semenova A, Westra ER, Wanner B, van der Oost J, Brouns SJJ, Severinov K.** 2011. Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. Proc. Natl. Acad. Sci. USA **108:**10098-10103.

44. **Tanabe Y, Kasai F, Watanabe M.** 2007. Multilocus sequence typing (MLST) reveals high genetic diversity and clonal population structure of the toxic cyanobacterium *Microcystis aeruginosa*. Microbiology **153:**3695-3703.

45. **Tillett D, Neilan BA.** 2000. Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. J. Phycol. **36:**251-258.

46. **Tominaga H, Kawagishi S, Ashida H, Sawa Y, Ochiai H.** 1995. Structure and replication of cryptic plasmids, pMA1 and pMA2, from a unicellular cyanobacterium, *Microcystis aeruginosa*. Biosci. Biotechnol. Biochem. **59:**1217-1220.

47. **Touchon M, Charpentier S, Clermont O, Rocha EPC, Denamur E, Branger C.** 2011. CRISPR distribution within the *Escherichia coli* species is not suggestive of immunity-associated diversifying selection. J. Bacteriol. **193:**2460-2467.

48. **Tucker S, Pollard P.** 2005. Identification of cyanophage Ma-LBP and infection of the cyanobacterium *Microcystis aeruginosa* from an Australian subtropical lake by the virus. Appl. Environ. Microbiol. **71:**629-635.

49. **Tyson GW, Banfield JF.** 2008. Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. Environ. Microbiol. **10:**200-207.

50. **Visser PM, Ibelings BW, Mur LR, Walsby AE.** 2005. The ecophysiology of the harmful cyanobacterium *Microcystis*, p 109-142. *In* J. Huisman, H. C. P. Matthijs, and P. M. Visser (ed.), Harmful Cyanobacteria. Springer, Dordrecht, Netherlands.

51. **Wilson AE, Sarnelle O, Neilan BA, Salmon TP, Gehringer MM, Hay ME.** 2005. Genetic variation of the bloom-forming Cyanobacterium *Microcystis aeruginosa* within and among lakes: implications for harmful algal blooms. Appl. Environ. Microbiol. **71:**6126-6133.

52. **Yoshida M, Yoshida T, Takashima Y, Kondo R, Hiroishi S.** 2005. Genetic diversity of the toxic cyanobacterium *Microcystis* in Lake Mikata. Environ. Toxicol. **20:**229-234.

53. **Yoshida M, Yoshida T, Takashima Y, Hosoda N, Hiroishi S.** 2007. Dynamics

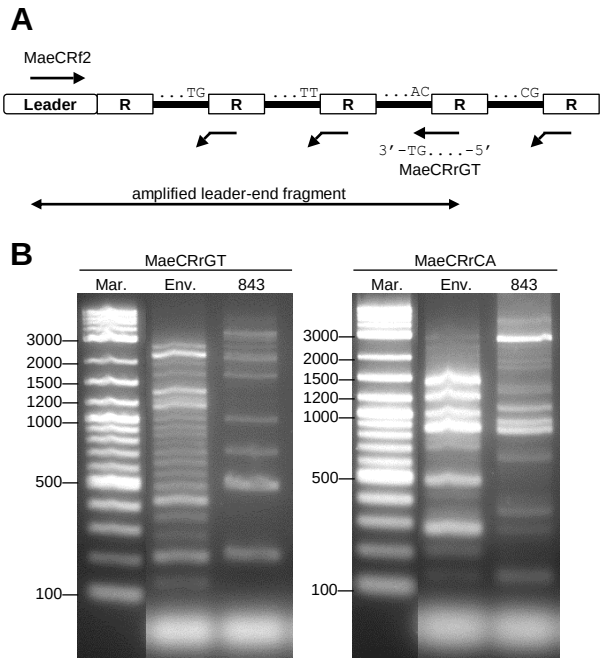
of microcystin-producing and non-microcystin-producing *Microcystis* populations is correlated with nitrate concentration in a Japanese lake. FEMS Microbiol. Lett. **266:**49-53.

54. **Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki K, Hiroishi S.** 2008. Ecological dynamics of the toxic bloom-forming cyanobacterium *Microcystis aeruginosa* and its cyanophages in freshwater. Appl. Environ. Microbiol. **74:**3269-3273.

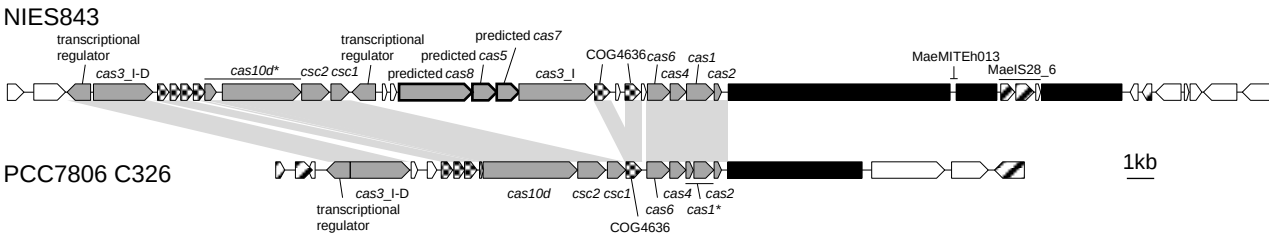
55. **Yoshida M, Yoshida T, Satomi M, Takashima Y, Hosoda N, Hiroishi S.** 2008. Intra-specific phenotypic and genotypic variation in toxic cyanobacterial *Microcystis* strains. J. Appl. Microbiol. **105:**407-415.

56. **Yoshida M, Yoshida T, Yoshida-Takashima Y, Kashima A, Hiroishi S.** 2010. Real-time PCR detection of host-mediated cyanophage gene transcripts during infection of a natural *Microcystis aeruginosa* population. Microbes Environ. **25:**211-215.

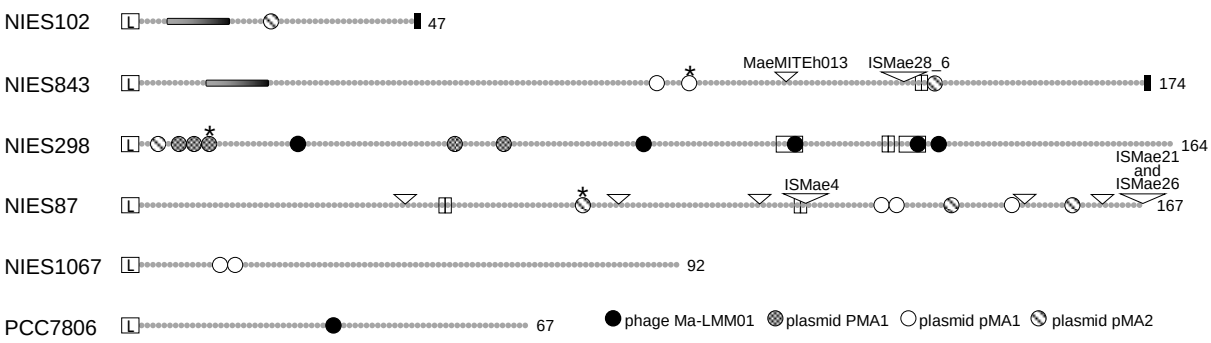
57. **Yoshida T, Takashima Y, Tomaru Y, Shirai Y, Takao Y, Hiroishi S, Nagasaki K.** 2006. Isolation and characterization of a cyanophage infecting the toxic cyanobacterium *Microcystis aeruginosa*. Appl. Environ. Microbiol. **72:**1239-1247.



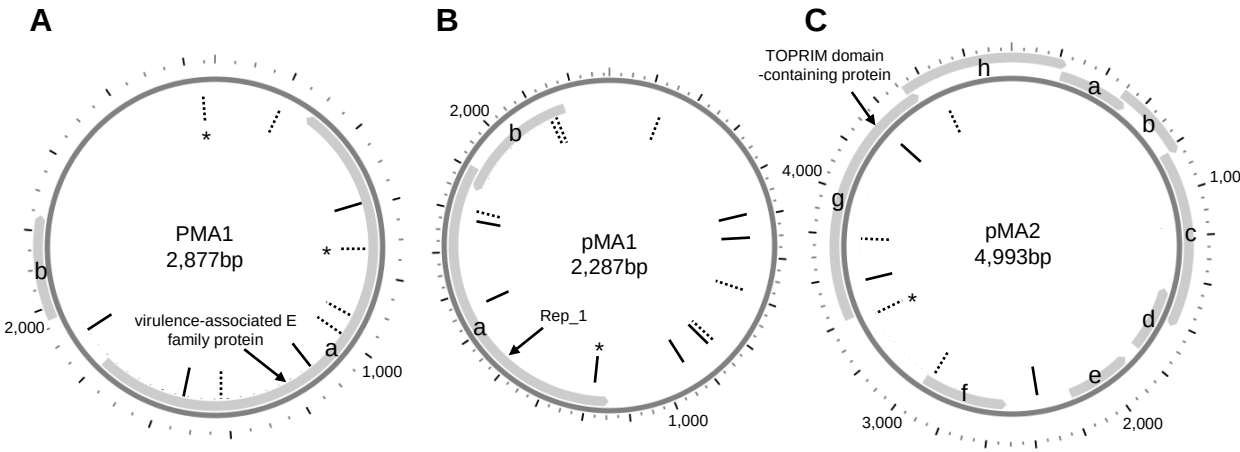
Kuno et al. Figure 1



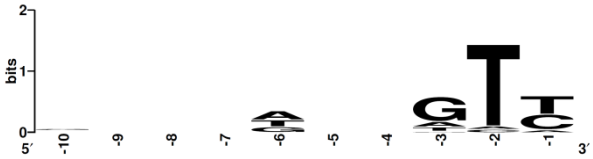
Kuno et al. Figure 2



Kuno et al. Figure 3



Kuno et al. Figure 4



Kuno et al. Figure 5