Intricate interactions between the bloom-forming cyanobacterium Microcystis aeruginosa and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat (CRISPR) signatures.

Kuno, Sotaro; Yoshida, Takashi; Kaneko, Takakazu; Sako, Yoshihiko

Applied and environmental microbiology (2012), 78(15): 5353-5360

© 2012, 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はありません。引用の際には出版社版をご確認ご利用ください。
Title

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 Kuno¹, Takashi Yoshida¹*, Takakazu Kaneko² and Yoshihiko Sako¹.

Affiliations

Graduate school of Agriculture, Kyoto University, Kitashirakawa-Oiwa, Sakyo-ku, Kyoto 606-8502, Japan¹.

Faculty of Life Sciences, Kyoto Sangyo University, Motoyama, Kamigamo, Kita-Ku, Kyoto 603-8555, Japan².
*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
also showed frequent exposure of *M. aeruginosa* to small cryptic plasmids that were observed only in a few strains. Thus, the diversification of CRISPR implies *M. aeruginosa* has been challenged by diverse communities (almost entirely uncharacterized) of cyanophages and plasmids.
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
infection and limits plasmid transfer (3, 13, 38, 39). In individual CRISPR arrays, new spacer additions are polarized to the leader-end of the loci (3, 10) while simultaneously spacer loss occurs (10), resulting in hypervariability within species. Therefore, the loci have been used for fingerprinting of pathogenic bacteria (i.e. spoligotyping (27)) and for dissecting microbial population structures (22, 49). Further, CRISPR spacers, representing direct signatures of viral infection, help to understand the co-evolutionary dynamics of the host-virus community in environments (2, 21).

The unicellular cyanobacterium *Microcystis aeruginosa* frequently forms dense blooms in freshwater environments worldwide (50). Some strains of this species produce potent hepatotoxins called microcystins that occasionally cause death of domestic animals and humans (8). Therefore, community composition and dynamics of *M. aeruginosa* populations are of great concern to water quality management (50). Previous studies have shown *M. aeruginosa* to be genetically highly heterogeneous (44, 51); *M. aeruginosa* populations also undergo temporal changes in genotypic composition (4, 52, 53). Previously we isolated a phage Ma-LMM01 infecting *M. aeruginosa* (57), and observed the potential qualitative impact of the cyanophage and its relatives on natural *M. aeruginosa* populations (54, 56). However, Ma-LMM01 has a narrow host range despite the large genetic diversity of hosts, suggesting there is a
greater diversity of host-cyanophage combinations other than Ma-LMM01 and its host strain in natural *M. aeruginosa* populations (57). Recently, a comparative genomic study of bacteria and archaea revealed an abundance of diverse defense systems including a CRISPR/Cas system in the *M. aeruginosa* NIES843 genome (37). Together, these findings imply richness of foreign genetic elements that have coevolved with *M. aeruginosa* through their specific interactions that probably contribute to the large clonal diversity (4, 51, 52) and genomic plasticity (12, 28) of the species.

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

**MATERIALS AND METHODS**
M. aeruginosa strains and DNA extraction.

M. aeruginosa strains NIES87, NIES102, NIES298, and NIES1067 were obtained from the National Institute for Environmental Studies (Tsukuba, Japan). The strains used in this study represent different phylogenetic groups in this species (Table 1). In a previous report using multilocus sequence typing (MLST), M. aeruginosa strains were largely divided into five major clades (Group A-E) (44). M. aeruginosa NIES102 and NIES843 are closely related to each other and fall into a well-supported inner clade (within Group A). Strain NIES298 and NIES1067 fall into other well-supported clades (Group B and Group D, respectively) (44). Strain NIES87 is not included in the major clades, but this strain is interesting for carrying two plasmids (46).

Phylogenetic relationships of these strains are shown in Fig. S1. The strains were maintained in CB medium (29) as previously described (31). Late-exponential phase cultures (2-ml aliquots) were mildly sonicated to remove gas vesicles, and then centrifuged at 3,000×g for 10 min (31). DNA was extracted from the cell pellets using the Nucleon Phytopure genomic DNA extraction kit (GE Healthcare, Tokyo, Japan) according to the manufacturer’s instructions.

Identification and Sequencing of CRISPR arrays.

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 MgCl₂ 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).
Purified DNA was suspended in 30 μl of sterilized milliQ water. Because conventional PCR amplification was not applicable to *M. aeruginosa* populations, we developed a PCR strategy based on the leader region (primer MaeCRf2) and repeat-spacer units (primers MaeCRrGT or its derivative MaeCRrCA) (Fig. 1A); thereby amplifying the leader-side fragments of *M. aeruginosa* CRISPR irrespective of their genomic contexts. The PCR was performed in 50μl containing 2.5 μL 1:100 dilution of the environmental DNA, 0.8 μM each primer, 0.25 mM each dNTP, 1×EX Taq Buffer and 2 U TaKaRa EX Taq polymerase (Takara Bio). The reaction conditions were: 94°C for 4 min followed by 30 cycles; 94°C for 30 s, 62°C for 1 min and 72°C for 1 min; and a final extension of 72°C for 7 min. The PCR products were separated using electrophoresis on 2.0% (w/v) agarose S gels (Nippon Gene, Tokyo, Japan). The gels were stained with GelRed (Biotium, CA) and visualized using the Gel Doc XR system (Bio-Rad Laboratories, CA). To prevent bias of preferential cloning of smaller DNA fragments, small gel blocks were separately excised, and DNA was purified and cloned. Thirty-nine and 50 clones were sequenced in full from PCR using MaeCRrGT and MaeCRrCA, respectively, using the primer walking method (primers are shown in Table S1). The CRISPRtionary program (17) was used to find identical spacers in different CRISPR fragments. Then, the fragments sharing the same spacer
order were manually assembled into contigs. To ensure accuracy, the leader-distal
spacer was removed from the contigs, because the PCR used in this study was shown to
allow one mismatch at the 3’-end additional dinucleotide of the reverse primers.

**Bioinformatics analysis.**

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

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

**Nucleotide sequences**

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

RESULTS

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

A type I-D CRISPR/Cas system homologous to that in M. aeruginosa NIES843 was identified in the genomic contig C326 of PCC7806. Makarova et al. (2011) identified “divergent rare variants” of cas8, cas7, and cas5 along with the typical type I-D cas genes in the NIES843 genome (37); however, the three cas variants and the downstream cas3 were absent in the PCC7806 draft genome (Fig. 2). The rare cas module in NIES843 may have been acquired by horizontal gene transfer (HGT), given its presence in some plasmids (e.g. Cyanothece sp. PCC8802 plasmid pP880201) (37).

In PCC7806, cas1, presumably involved in spacer acquisition (11), was interrupted by an in-frame stop codon, suggesting that spacer uptake may be no longer active in this strain. The sequence between cas2 and the first CRISPR repeat was highly conserved
between the strains and contained AT-rich regions including a putative promoter element (5’-TTGAAG-17bp-TAYRAT-3’). Therefore the sequence was considered to be a leader (26, 33). The CRISPR/Cas was located in different genomic contexts in the two strains (Fig. 2).

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

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

**Sequence analysis of the CRISPR repeats.**

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

**CRISPR signatures of foreign DNA elements.**

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

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

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

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

The ecological impact of cyanophages and cryptic plasmids

CRISPR signatures of *M. aeruginosa* provided insights into the ecological impact and dynamics of known cyanophage. Spacers matching cyanophage Ma-LMM01 were found in *M. aeruginosa* NIES298 as well as in PCC7806 from the Netherlands (Fig. 3) and three CTs from Hirosawanoike Pond (Table 2). This implies the host-virus combinations involving Ma-LMM01 and possible dispersal of its related phages. Conversely, all of the Ma-LMM01-matching spacers (10 spacers) were not
completely identical to their corresponding putative proto-spacers in Ma-LMM01, suggesting high nucleotide diversity within this phage lineage. Of these, six spacers had only nucleotide mutations that are translated into silent or conservative changes in deduced amino acid sequences (Table 3). Although a single mutation in proto-spacers or PAMs basically can abolish CRISPR-mediated immunity against phages (3, 10, 13), the strict nucleotide identity required for the immunity is limited to a specific seven-nucleotide region in the proto-spacers (denoted as “seed sequence”) in *Escherichia coli* (43). If the *M. aeruginosa* CRISPR system employs this “seed”-dependent immunity, Ma-LMM01 or its related phages may evade interference mediated by five of the spacers (all but CT15spc14 in Table 3), and thereby the Ma-LMM01-related phage assemblage retains the conservative nucleotide polymorphisms to circumvent host immunity.

CRISPR signatures of *M. aeruginosa* also showed repeated exposures to plasmids PMA1, pMA1, and pMA2. Uniform distribution of putative proto-spacers throughout these plasmids (Fig. 4) suggests the recipient CRISPR recognized the plasmids themselves transferred from potential donors, rather than other plasmids sharing specific components (e.g. conserved replication genes). The proposed interference mechanisms of the CRISPR/Cas systems predict that spacers completely
matching a plasmid prevent it from establishing in the host (13, 39). In accordance, *M. aeruginosa* NIES843 possessing a spacer with 100% identity to plasmid pMA1 (Fig. 3) carries no plasmid (28). Unstable presence of the small plasmids reported in some *M. aeruginosa* strains (42) may be attributed to such CRISPR-mediated exclusion. However, strain NIES87 retains pMA2 (46), despite possessing a spacer with 100% identity to the plasmid (Fig. 3). This apparent conflict may have resulted from defective CRISPR interference given the proliferation of IS elements around the CRISPR array in NIES87 (Fig. 3). The small plasmids carry very few genes, and thus their role in ecology and evolution of *M. aeruginosa* remains unclear. Interestingly, similar small cryptic plasmids were found to dominate in marine *Synechococcus* metagenomes and were hypothesized to facilitate HGT and in some case phage resistance (35, 41). Considered with the suggested spread of the small plasmids among *M. aeruginosa* populations, they may facilitate host genetic diversity via HGT or chromosomal integration.

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

*M. aeruginosa* and several other cyanobacteria possess the recently identified subtype I-D CRISPR/Cas system, which is a hybrid of type-I system and type-III executive Cas module (36). Our data provides some information on the functional
characteristics of the \textit{M. aeruginosa} CRISPR/Cas system. Uniform distribution of the
putative proto-spacers without bias toward either strand of the plasmids (Fig. 4)
suggests the \textit{M. aeruginosa} CRISPR/Cas recognizes DNA rather than RNA. Combined
with the presence of the upstream PAM (Fig. 5), this is in accordance with the type-I
information processing (i.e. spacer acquisition) system (36).

\textbf{Conclusions.}

\textit{M. aeruginosa} shows a high degree of CRISPR heterogeneity within
populations. Thus we infer that host-phage community can be subdivided into a number
of different “susceptible combinations” of \textit{M. aeruginosa} and cyanophages. In each
combination, the CRISPR spacers and cyanophages may quickly diversify through
coevolution. This intricate interaction is expected to shape complex host-phage
community. The spacer sequences imply what kinds of phages are involved in each
“susceptible combination” with a strain of \textit{M. aeruginosa} (Fig. 3, Table S3).

Nevertheless, a significant fraction of the spacer sets in \textit{M. aeruginosa} are of unknown
origin, largely because of the under-representation of the diversity of bacteriophages
and plasmids in the current sequence data. Thus, \textit{M. aeruginosa} may be challenged by
diverse, almost entirely uncharacterized communities of cyanophages and plasmids.

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
boxes, respectively. Cas genes, toxin-antitoxin genes, and transposase genes are indicated by gray, dotted, and striped arrows, respectively. Divergent rare cas variants (37) are enclosed by thick lines. Gene nomenclature is in accordance with the previous study (37). Pseudogenes are marked with asterisks. Conserved regions between the two genomes are indicated by gray shadows.

**FIGURE 3. CRISPR arrays in six M. aeruginosa strains.** Strain-specific spacers are indicated by circles, of which those with significant hits to sequences of phage and plasmids are pattern-coded as indicated in the figure. * indicates 100% nucleotide identity to phage/plasmid sequences. A string of 11 consecutive spacers shared between strains NIES102 and NIES843 is indicated by rectangles with shading showing the direction of the shared spacer block. The total number of the spacers is shown at the right of each array. Degenerate terminal repeats are indicated by black bars. Spacer duplication is indicated by open boxes. Insertion of ISs, MITE and other short sequence elements, in CRISPR repeats are indicated by triangles.

**FIGURE 4. Distribution of proto-spacers on plasmids PMA1 (A), pMA1 (B) and pMA2 (C).** Each bar indicates the position of the putative proto-spacer. Solid and
dashed bars indicate putative proto-spacers in clockwise and counterclockwise strands, respectively. * indicates 100% nucleotide identity. Note we found additional ORFs in pMA1 (ORF b) and pMA2 (ORFs a, b, e, g and h) compared with their original description. ORF a and b of pMA1 may be generated from a longer replication gene by a frame-shift. ORF g, h, a, b, and c may be generated by frame-shifts from a single hypothetical gene whose homologues are prevalent in cyanobacteria (e.g. CY0110_05002 of Cyanothece sp. CCY0110).

FIGURE 5. Sequence logo of the PAM consensus. The logo was built using the WebLogo based on alignments of the flanking sequences of putative proto-spacers in Ma-LMM01 and PMA1, pMA1 and pMA2. Numbers below indicate nucleotide positions where -1 is just upstream of the putative proto-spacers.
<table>
<thead>
<tr>
<th>Strain name</th>
<th>Locality</th>
<th>Year</th>
<th>Sequence type</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIES102</td>
<td>Lake Kasumigaura, Japan</td>
<td>1982</td>
<td>10</td>
</tr>
<tr>
<td>NIES298</td>
<td>Lake Kasumigaura, Japan</td>
<td>1982</td>
<td>60</td>
</tr>
<tr>
<td>NIES87</td>
<td>Lake Kasumigaura, Japan</td>
<td>1982</td>
<td>2</td>
</tr>
<tr>
<td>NIES1067</td>
<td>Chikatou Pond, Japan</td>
<td>1982</td>
<td>27</td>
</tr>
<tr>
<td>NIES843*</td>
<td>Lake Kasumigaura, Japan</td>
<td>1997</td>
<td>18</td>
</tr>
<tr>
<td>PCC7806*</td>
<td>Braakman Reservoir, The Netherlands</td>
<td>1972</td>
<td>N.D.</td>
</tr>
</tbody>
</table>

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

Sequence type were determined by Tanabe et al. (2007) (44). N.D., not determined.
TABLE 2. Characteristics of CRISPR types (CTs) identified in Hirosawanoike Pond.

<table>
<thead>
<tr>
<th>CT</th>
<th>No. of clones</th>
<th>No. of spacers$^a$</th>
<th>No. of unique spacers with significant hit$^c$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Shared$^b$</td>
</tr>
<tr>
<td>CT1</td>
<td>8</td>
<td>16</td>
<td>-</td>
</tr>
<tr>
<td>CT2</td>
<td>15</td>
<td>13</td>
<td>10(4-13)$^A$</td>
</tr>
<tr>
<td>CT3</td>
<td>1</td>
<td>12</td>
<td>-</td>
</tr>
<tr>
<td>CT4</td>
<td>10</td>
<td>27</td>
<td>1(1)$^B$</td>
</tr>
<tr>
<td>CT5</td>
<td>20</td>
<td>32</td>
<td>-</td>
</tr>
<tr>
<td>CT6</td>
<td>1</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td>CT7</td>
<td>2</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>CT8</td>
<td>2</td>
<td>10</td>
<td>-</td>
</tr>
<tr>
<td>CT9</td>
<td>1</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>CT10</td>
<td>1</td>
<td>6</td>
<td>-</td>
</tr>
<tr>
<td>CT11</td>
<td>1</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>CT12</td>
<td>1</td>
<td>21</td>
<td>6(1-3, 6-8)$^C$</td>
</tr>
<tr>
<td>CT13</td>
<td>1</td>
<td>15</td>
<td>-</td>
</tr>
<tr>
<td>CT14</td>
<td>1</td>
<td>7</td>
<td>-</td>
</tr>
<tr>
<td>CT15</td>
<td>1</td>
<td>28</td>
<td>3</td>
</tr>
<tr>
<td>CT16</td>
<td>1</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>CT17</td>
<td>1</td>
<td>13</td>
<td>1(1)$^B$</td>
</tr>
<tr>
<td>CT18</td>
<td>1</td>
<td>13</td>
<td>-</td>
</tr>
<tr>
<td>CT19</td>
<td>3</td>
<td>17</td>
<td>1</td>
</tr>
<tr>
<td>CT20</td>
<td>1</td>
<td>17</td>
<td>-</td>
</tr>
<tr>
<td>CT21</td>
<td>1</td>
<td>16</td>
<td>6(1-3, 4-6)$^C$, 10(7-16)$^A$</td>
</tr>
<tr>
<td>CT22</td>
<td>6</td>
<td>4</td>
<td>-</td>
</tr>
<tr>
<td>CT23</td>
<td>6</td>
<td>11</td>
<td>-</td>
</tr>
<tr>
<td>CT24</td>
<td>3</td>
<td>9</td>
<td>-</td>
</tr>
</tbody>
</table>

$^a$ Number of spacers in CT contigs.

$^b$ Spacers shared between CTs are indicated as x(y-z)$^N$. x is total number of shared spacers, y-z is position of shared spacers (sequentially numbered from the leader-end spacer), and N denotes a pair of strain sharing the spacers.

$^c$ -, Not found in sequenced contigs. Number of 100% match is shown in parenthesis.
TABLE 3. Spacers showing silent or conservative mutations compared to Ma-LMM01 putative proto-spacers.

<table>
<thead>
<tr>
<th>spacer/phage gene&lt;sup&gt;a&lt;/sup&gt;</th>
<th>nucleotide sequence&lt;sup&gt;b&lt;/sup&gt;</th>
<th>predicted amino acid sequence&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>298spc124&lt;sup&gt;4&lt;/sup&gt;</td>
<td>AGTGGCCGGGCTACTTATCTTCTACCAATCTAC</td>
<td>SGAATLYQFS</td>
</tr>
<tr>
<td>ORF40</td>
<td>..............A...C.................G..C.. ..............V.</td>
<td></td>
</tr>
<tr>
<td>7806spc34&lt;sup&gt;4&lt;/sup&gt;</td>
<td>TTTGAGGGACTAAATAATGGGATCGAAAATCAAT</td>
<td>FEGLNNINGVFN</td>
</tr>
<tr>
<td>ORF41</td>
<td>..............C...C.................G....G..............A.....S</td>
<td></td>
</tr>
<tr>
<td>CT6spc2&lt;sup&gt;4&lt;/sup&gt;</td>
<td>AATCCCCCGTCAGGGATTTCTCCACGGGTCAAT</td>
<td>IPPSGILPRVS</td>
</tr>
<tr>
<td>ORF61</td>
<td>..............A...G.................G...T..............V..</td>
<td></td>
</tr>
<tr>
<td>CT15spc12&lt;sup&gt;8&lt;/sup&gt;</td>
<td>ACTCTCCTTTGCCACATAGATAGTGCTAGTCTAC</td>
<td>SPCDKYVGKS</td>
</tr>
<tr>
<td>ORF25</td>
<td>..............T................. ..............</td>
<td></td>
</tr>
<tr>
<td>CT15spc13&lt;sup&gt;8&lt;/sup&gt;</td>
<td>TCTATCTTCTTAATAGCTCTATGGCTAGTAGGGAATAAGCAACTAAG</td>
<td>YLFNTMPLGAGQ</td>
</tr>
<tr>
<td>ORF20</td>
<td>..............T................. ..............</td>
<td></td>
</tr>
<tr>
<td>CT15spc14&lt;sup&gt;8&lt;/sup&gt;</td>
<td>TGATACAGTGCTCTTAGGCTAGTCTAC</td>
<td>DTGAFGLGCY</td>
</tr>
<tr>
<td>ORF25</td>
<td>..............G................. ..............</td>
<td></td>
</tr>
</tbody>
</table>

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

<sup>b</sup> 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 <i>E. coli</i> (43) is indicated by boxes.
REFERENCES


29. **Kasai F, Kawachi M, Erata M, Watanabe MM.** 2004. NIES-Collection List of...


Microbiol. 17:133-137.


40


of microcystin-producing and non-microcystin-producing *Microcystis*


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