

Viruses Infecting a Freshwater Filamentous Cyanobacterium (*Nostoc* sp.) Encode a Functional CRISPR Array and a Proteobacterial DNA Polymerase B

Caroline Chénard,^{a*} Jennifer F. Wirth,^{a*} Curtis A. Suttle^{a,b,c,d}

Department of Earth Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, British Columbia, Canada^a; Departments of Botany and Microbiology & Immunology, University of British Columbia, Vancouver, British Columbia, Canada^b; Institute for the Oceans and Fisheries, University of British Columbia, Vancouver, British Columbia, Canada^c; Canadian Institute for Advanced Research, Toronto, Ontario, Canada^d

* Present address: Caroline Chénard, Singapore Centre on Environmental Life Sciences Engineering (SCELSSE), Nanyang Technological University, Singapore; Jennifer F. Wirth, Thermal Biology Institute, Montana State University, Bozeman, Montana, USA.

ABSTRACT Here we present the first genomic characterization of viruses infecting *Nostoc*, a genus of ecologically important cyanobacteria that are widespread in freshwater. Cyanophages A-1 and N-1 were isolated in the 1970s and infect *Nostoc* sp. strain PCC 7210 but remained genomically uncharacterized. Their 68,304- and 64,960-bp genomes are strikingly different from those of other sequenced cyanophages. Many putative genes that code for proteins with known functions are similar to those found in filamentous cyanobacteria, showing a long evolutionary history in their host. Cyanophage N-1 encodes a CRISPR array that is transcribed during infection and is similar to the DR5 family of CRISPRs commonly found in cyanobacteria. The presence of a host-related CRISPR array in a cyanophage suggests that the phage can transfer the CRISPR among related cyanobacteria and thereby provide resistance to infection with competing phages. Both viruses also encode a distinct DNA polymerase B that is closely related to those found in plasmids of *Cyanothece* sp. strain PCC 7424, *Nostoc* sp. strain PCC 7120, and *Anabaena variabilis* ATCC 29413. These polymerases form a distinct evolutionary group that is more closely related to DNA polymerases of proteobacteria than to those of other viruses. This suggests that the polymerase was acquired from a proteobacterium by an ancestral virus and transferred to the cyanobacterial plasmid. Many other open reading frames are similar to a prophage-like element in the genome of *Nostoc* sp. strain PCC 7524. The *Nostoc* cyanophages reveal a history of gene transfers between filamentous cyanobacteria and their viruses that have helped to forge the evolutionary trajectory of this previously unrecognized group of phages.

IMPORTANCE Filamentous cyanobacteria belonging to the genus *Nostoc* are widespread and ecologically important in freshwater, yet little is known about the genomic content of their viruses. Here we report the first genomic analysis of cyanophages infecting filamentous freshwater cyanobacteria, revealing that their gene content is unlike that of other cyanophages. In addition to sharing many gene homologues with freshwater cyanobacteria, cyanophage N-1 encodes a CRISPR array and expresses it upon infection. Also, both viruses contain a DNA polymerase B-encoding gene with high similarity to genes found in proteobacterial plasmids of filamentous cyanobacteria. The observation that phages can acquire CRISPRs from their hosts suggests that phages can also move them among hosts, thereby conferring resistance to competing phages. The presence in these cyanophages of CRISPR and DNA polymerase B sequences, as well as a suite of other host-related genes, illustrates the long and complex evolutionary history of these viruses and their hosts.

Received 21 April 2016 Accepted 12 May 2016 Published 14 June 2016

Citation Chénard C, Wirth JF, Suttle CA. 2016. Viruses infecting a freshwater filamentous cyanobacterium (*Nostoc* sp.) encode a functional CRISPR array and a proteobacterial DNA polymerase B. *mBio* 7(3):e00667-16. doi:10.1128/mBio.00667-16.

Editor Roger Hendrix, University of Pittsburgh

Copyright © 2016 Chénard et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Curtis A. Suttle, suttle@science.ubc.ca.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. External solicited reviewers: Graham Hatfull, University of Pittsburgh; Martha Clokie, University of Leicester.

Filamentous cyanobacteria of the genera *Nostoc* and *Anabaena* are abundant and active members of aquatic and terrestrial microbial communities. They occur in habitats ranging from the bottom of ice-covered polar lakes (1) and hypertrophic coastal lagoons (2) to rice paddy soils (3) and rock pool communities in karst regions (4). They fix nitrogen in nutrient-poor environments and can form symbiotic associations with a wide range of plants and fungi (4). Despite the ecological importance of these

cyanobacteria, the genomic composition of the viruses infecting them remains largely unexplored.

Bacteria and their viruses (phages) have a shared evolutionary history stretching for billions of years that has led to a myriad of adaptations for cells to avoid infection and countermeasures for phage to escape these defenses (5). One of these defense mechanisms involves the clustered regularly interspaced short palindromic repeats (CRISPRs) and the CRISPR-associated (*cas*) genes

(6, 7) (i.e., the CRISPR-Cas system). Evidence of this adaptive immune system is found in almost all archaeal genomes and in about 40% of bacterial genomes (8). The CRISPR array consists of a series of noncontiguous direct repeats (DRs) 20 to 50 bp long that are separated by variable sequences (spacers) usually derived from viruses and plasmids (6, 7, 9). A leader region that is an AT sequence of up to 550 bp directly adjoins the first regions. In some cases, *cas* genes are found upstream or downstream of CRISPR arrays. The CRISPR-Cas system provides immunity to cells by recognizing and cleaving incoming foreign genetic material with sequence similarity to the spacers (6–8, 10). Many types of CRISPRs have been recognized on the basis of the sequence similarity of the repeats (11). The distribution of closely related CRISPR/Cas systems in phylogenetically distant organisms suggests exchange by horizontal gene transfer (12). CRISPR/Cas systems are rare in plasmids and prophages; however, it has been suggested that they could mediate the exchange of CRISPRs among organisms (13–15).

Phages infecting *Nostoc* spp. were among the first cyanophages to be studied extensively (16), including host range, life cycle, and host interaction (17–19). For example, cyanophage N-1 was the first shown to reduce host photosynthesis upon infection (20). Despite the extensive work on the viruses and the effects of infection on the hosts, the viruses have remained uncharacterized genomically. In this study, we analyzed the genome sequences of viruses infecting the freshwater filamentous cyanobacterium *Nostoc* sp. strain PCC 7120. Our results demonstrate that cyanophages A-1 (21) and N-1 (17, 22) are distantly related to other genetically characterized cyanophages. Both viruses contain a distinct DNA polymerase B (Pol B) that is closely related to those found in plasmids of their cyanobacterial hosts. In addition, cyanophage N-1 has a CRISPR array that is similar to those found in cyanobacteria, suggesting that cyanophages play a role in exchanging CRISPRs among cyanobacteria.

RESULTS AND DISCUSSION

The sequencing of cyanophages A-1 and N-1, which infect *Nostoc* sp. strain PCC 7120, revealed a previously unknown lineage of myoviruses. Few of the predicted coding genes were similar to those found in other cyanophages, while the DNA Pol B sequences were similar to those found in a host plasmid. In addition, cyanophage N-1 contains a CRISPR array similar to those found in cyanobacteria, suggesting that phages may mediate the exchange of CRISPRs among cyanobacteria and confer resistance to competing phages.

Genome features. The double-stranded DNA (dsDNA) genomes of cyanophages A-1 and N-1 are 68,304 and 64,960 bp long, with G+C contents of 38.3 and 35.4%, respectively, and have circularly permuted, terminally redundant ends. Their

genome lengths are about half those of many other phages that have an obvious contractile tail. For example, coliphage T4 has a genome of 168 kb (23), while other myoviruses have genomes ranging from 161 to 231 kb in length (24–29). However, some viruses with contractile tails have much smaller genomes; the classic case is bacteriophage Mu, a temperate bacteriophage of 36 kb that reproduces by transposition. In addition, there are recently characterized (30) “dwarf” myoviruses that have genomes of <50 kb and infect a diversity of proteobacteria (i.e., *Aeromonas salmonicida*, *Vibrio cholerae*, *Bdellovibrio* spp., and *Pectobacterium carotovorum*). These viruses do not have sequence similarity to A-1 and N-1.

Bioinformatic analysis of A-1 and N-1 revealed that only about a quarter of the translated open reading frames (ORFs) had similarity to protein sequences in current databases (36 of 97 for A-1 and 33 of 91 for N-1). This is similar to other freshwater cyanophages, including recently described S-EIV1, in which 85% of the translated ORFs had no database matches (31), and S-CRM01 (26), Ma-LMM01 (29), Pf-WMP4, and Pf-WMP3 (32, 33), which range from 61 to 76% similarity. In contrast, only about 20% of the ORFs in most phage isolates do not have recognizable homologues (34), while in marine cyanophages, the percentage is typically <40% (27, 35). Clearly, there is a lack of representative genomes of freshwater cyanophages.

Translated ORFs with significant hits to proteins of known function were associated with DNA replication, DNA metabolism and repair, and structural components (Fig. 1; see Tables S1 and S2 in the supplemental material). There were also some putative genes consistent with a temperate lifestyle, such as a transposase and two phage antirepressors, in cyanophage A-1. Although no obvious repressor genes were identified, a plausible candidate is a putative LuxR-encoding gene that was found in both genomes. However, no evidence of genes encoding an integrase or partitioning proteins was found.

Although there was high similarity between the *Nostoc* cyanophages, they had little similarity to cyanophages infecting other cyanobacteria. Genomic studies of 26 T4-like phages, including 16 marine cyanophages, identified 38 T4-like core genes coding primarily for structural and DNA replication proteins (35). In contrast, only six T4-like core genes have significant similarity to ORFs in the *Nostoc* cyanophages (Table 1), and those genes are associated with replication and DNA modification. In addition, none of the 25 genes found exclusively within marine cyanomyoviruses were present in the *Nostoc* viruses.

Both genomes were analyzed for regulatory elements and motifs such as tRNA genes, promoter motifs, and transcriptional terminators. Unlike viruses infecting marine cyanobacteria, but not unusual for myoviruses with genomes smaller than 70 kb, tRNA genes were not identified in cyanophages A-1 and N-1. A recent

TABLE 1 Predicted ORFs in cyanophages A-1 and N-1 with similarity to T4-like genes

T4 core gene product	Cyanophage A-1		Cyanophage N-1	
	E value	Phage hit	E value	Phage hit
gp5 baseplate hub + tail lysozyme	4.00e ⁻⁵	Phage RB49	4.00e ⁻⁵	Cyanophage S-PM2
gp17 terminase DNA-packaging enzyme large subunit	4.00e ⁻²	Cyanophage S-PM2	4.00e ⁻²	Cyanophage S-PM2
Gp41 DNA primase-helicase	2.00e ⁻¹³	Cyanophage S-PM2	1.00e ⁻¹¹	Cyanophage S-RSM4
gp43 DNA Pol B	4.00e ⁻⁵	Phage Aeh1	9.00e ⁻⁶	Phage Aeh1
Thymidylate synthase	1.00e ⁻³⁵	Cyanophage S-CRM01	9.00e ⁻³²	Cyanophage P-SSM4

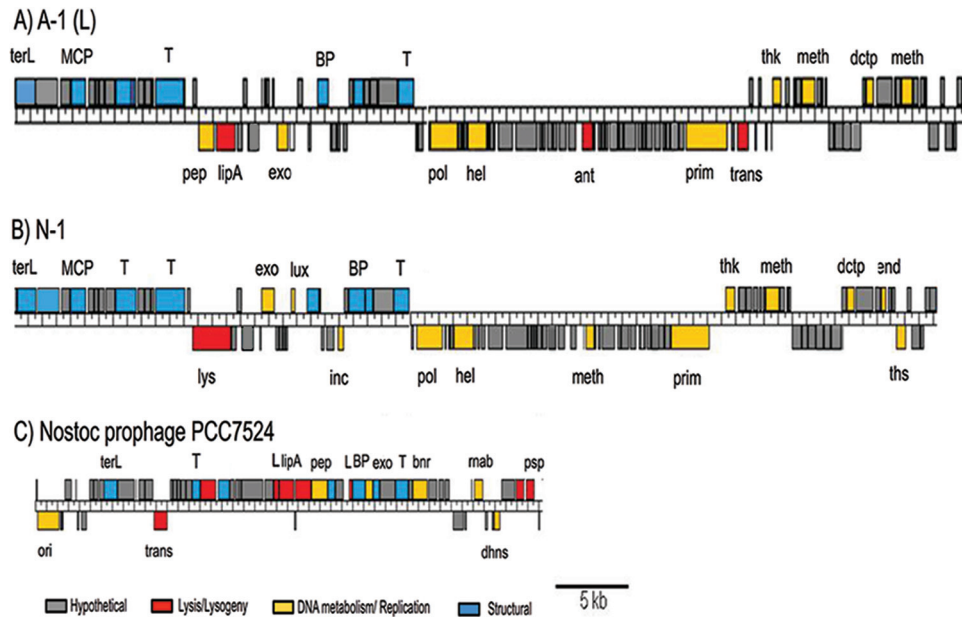


FIG 1 Comparative genomics of the two *Nostoc* cyanophages and the prophage-like element in the genome of *Nostoc* sp. strain PCC 7524. The genomes are presented as linear molecules for better comparison. A, cyanophage A-1; B, cyanophage N-1; C, prophage-like element in the genome of *Nostoc* sp. strain PCC 7524 (accession no. NC019684.1). Gene abbreviations and functions are as follows: ant, antirepressor; BP, baseplate; bnr, bnr/Asp-box repeat protein; dctp, dCTP deaminase; dhns, dihydroxynaphthoate synthase; exo, exonuclease; hel, helicase; incl, viral A inclusion factor; L, lysozyme; lipA, lipoprotein A; luX, LuxR transcription factor; lyz, lysozyme; MCP, major capsid protein; meth, methylase; ori, origin of replication; pep, peptidase; psp, phage shock protein; pol, DNA Pol B; prim, primase; T, tail; terL, terminase large subunit; thk, thymidylate kinase; ths, thymidylate synthase; trans, transposase.

study found that 7 of 10 small myoviruses did not contain tRNAs (30). In addition, the presence or absence of tRNAs might influence host range, as myoviruses that infect only *Prochlorococcus* spp. tend to code for few, if any, tRNAs, whereas those infecting *Prochlorococcus* spp. and *Synechococcus* spp. or just *Synechococcus* spp., code for more tRNAs (36). In contrast, early studies described cyanophages A-1 and N-1 as being able to infect both *Anabaena* and *Nostoc* spp. (19); however, it has been suggested that the reported host range might reflect errors in taxonomy (16, 37).

Putative promoter motifs in the *Nostoc* cyanophage genomes were identified with PHIRE (38), although the sequences differed (Fig. 2). For A-1, 43 motifs were identified that contained a con-

sensus sequence for a putative promoter motif consisting of two highly conserved regions separated by 1 bp (Fig. 2). For N-1, 21 motifs were identified that included a highly conserved region of 11 bp in the repeat motif. Consistent with functioning as promoters, the motifs were typically located between ORFs. However, the motifs were asymmetric and not consistently in the same direction as the flanking genes, which suggests that they may not function as promoters. An alternative possibility is that the motifs represent stopperators that bind a regulatory protein such as a phage repressor, as has been shown for mycobacteriophages (39).

Presence of a distinct DNA Pol B. Sequences were found in A-1 and N-1 that putatively encode a DNA Pol B that catalyzes the polymerization of deoxyribonucleotides into a DNA strand and



FIG 2 Sequence logos of the promoter motifs predicted from alignments of the 5' upstream regions (top) and putative promoter for cyanophage A-1 showing a sequence logo created from an alignment of 43 sequences (bottom). The putative promoter for cyanophage N1 shows a sequence logo created from an alignment of 21 sequences. The heights of the letters are proportional to the levels of sequence conservation of the nucleotides at the respective positions.

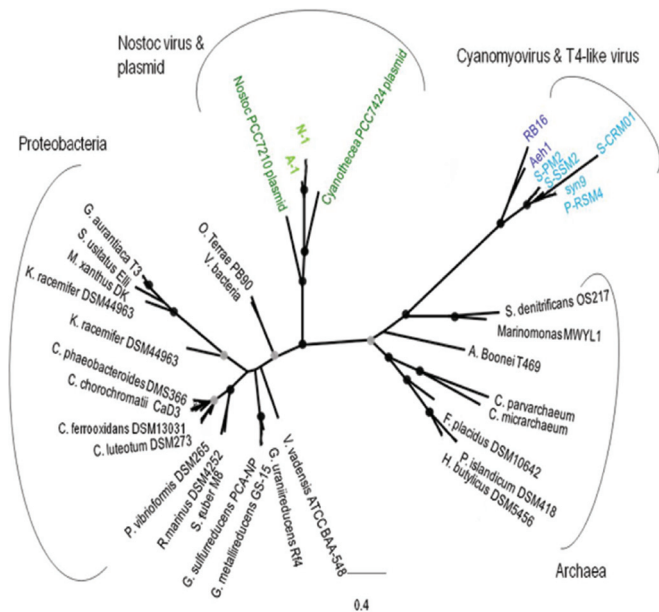


FIG 3 Unrooted ML phylogenetic tree of DNA Pol B protein sequences found in viruses, bacteria, and archaea. Bootstrap values of 90 to 100% (black circles) and 75 to 89 % (gray circles) are shown at the nodes. The sequence names are colored as follows: black, *Bacteria* and *Archaea*; light blue, marine cyanomyoviruses; dark blue, T4-like myoviruses; light green, *Nostoc* phages; dark green, cyanobacterial plasmid. The scale bar shows the number of amino acid substitutions per site.

has many viral and cellular homologues, making it a useful phylogenetic marker. The Pol B sequences in A-1 and N-1 were most closely related to plasmids in the cyanobacteria *Cyanothecae* sp. strain PCC 7424 (plasmid pP742402), *Nostoc* sp. strain PCC 7120 (plasmid pCC7120beta), and *Anabaena variabilis* ATCC 29413 (plasmid C) (Fig. 3). These viral and plasmid Pol B sequences form a distinct group that is related to proteobacterial and archaeal Pol B clades. Moreover, while Pol B is common in proteobacteria, the only representatives known in cyanobacteria are in these three plasmids.

The close phylogenetic relationship between the Pol B sequences in the phages and plasmids suggests that the gene for Pol B was transferred laterally. Moreover, the DNA Pol B in *Cyanothecae* plasmid pP742402 is adjacent to a CRISPR, a region where recombination can occur and thus be promiscuous to gene exchange. The DNA polymerases in A-1, N-1, and cyanobacterial plasmids have an ancestor in common with proteobacteria, implying that the transfer of DNA to the cyanobacterial plasmids may have been phage mediated. Recent genomic analysis of other cyanophages reveals the presence of interesting polymerases. For instance, viruses infecting the marine cyanobacterium *Acaryochloris marina* contain a Pol A that appears to be related to eukaryotic polymerases (40). In addition, a polar cyanophage (S-EIV1) recently isolated from polar inland waters contains a Pol A sequence that is phylogenetically divergent from those found in other cyanophages (31).

Phylogeny of the terminase large subunit. Further evidence of the evolutionary divergence of A-1 and N-1 from other cyanophages is provided by the gene *terL*, which encodes the terminase large subunit, a protein involved in DNA packaging in ds-DNA phages. Phylogenetic analysis reveals that the translated *terL*

sequences from A-1 and N-1 cluster separately from those in other viruses (Fig. 4) and branch most closely with those in the freshwater heterocystous cyanobacterium *Nostoc* sp. strain PCC 7524 and the marine tropical and subtropical unicellular N fixer *Crocospheera watsonii*, while marine *Synechococcus* sp. strain PCC 9605 is somewhat more distant. It has been argued that terminases are good phylogenetic markers of phage evolution and that sequences found in cyanobacteria may be remnant prophages (41). Indeed, the gene for the terminase large subunit in *C. watsonii* occurs near genes encoding a phage tail collar (NCBI reference sequence accession no. EAM53192), a transposase (NCBI reference sequence accession no. EAM53191), and a hypothetical protein (NCBI reference sequence accession no. EAM53190) that also show similarity to putative genes in the *Nostoc* cyanophages. The terminase large subunit in *Nostoc* sp. strain PCC 7524 is also part of a prophage-like element (see below). The A-1 and N-1 terminase sequences have revealed a new evolutionary group of phage terminases with similarity to prophage elements in several genera of divergent cyanobacteria, suggesting that relatives of A-1 and N-1 infect a broad range of marine and freshwater hosts.

Genetic exchange between filamentous cyanobacteria and *Nostoc* cyanophages. Some ORFs in A-1 and N-1 exhibit high similarity to genes in cyanobacteria that code for proteins with known functions (Table 2 and 3), including purine and pyrimidine metabolism. One example is dCTP deaminase, an enzyme involved in the production of dUMP, the immediate precursor of thymidine nucleotides. Phylogenetic analysis demonstrates that dCTP deaminase sequences from the *Nostoc* cyanophages are more similar to those found in cyanobacteria, whereas the dCTP deaminase homologue in the marine cyanophage S-PM2 is more closely related to other virus sequences (see Fig. S1 in the supplemental material). Thymidylate synthase and thymidylate kinase genes in A-1 and N-1 were also similar to those found in cyanobacteria. The proteins encoded by these genes likely catalyze two subsequent steps in dTTP synthesis. Thymidylate synthase is involved in the production of dTMP, while thymidylate kinase phosphorylates dTMP to dTDP. This reaction is crucial to both the *de novo* synthetic and salvage pathways for pyrimidine deoxyribonucleotides. The gene encoding thymidylate kinase is commonly found in eukaryotes and their DNA viruses and has been reported in some myoviruses with genome sizes of >200 kb (42, 43); however, to our knowledge, this is the first time a homologue of this gene in A-1 and N-1 has been reported in phages with genomes of <70 kb. Sequences with similarity to putative genes encoding DNA adenine methyltransferases and DNA cytosine methyltransferases were also found in the *Nostoc* cyanophages. In general, DNA methyltransferases mediate postreplicative methylation at a specific recognition site and protect bacterial DNA against digestion by specific restriction endonucleases, whereas unmethylated infective DNA, such as in phages, is cleaved. However, DNA methyltransferases occur in some phages and modify the viral DNA to be resistant to bacterial restriction systems. In general, phage DNA methyltransferases are similar to those of their hosts.

Putative coding sequences in A-1 and N-1, as well as in the freshwater cyanophage S-CRM01 (26) and the marine cyanophage S-PM2 (24), are similar to host-like genes encoding the rare lipoprotein A (*rlpA*). Although its function is not known, *rlpA* was strongly induced during hyperosmotic stress in *Synechocystis* sp. strain PCC 6803 (44) and is upregulated as part of the general

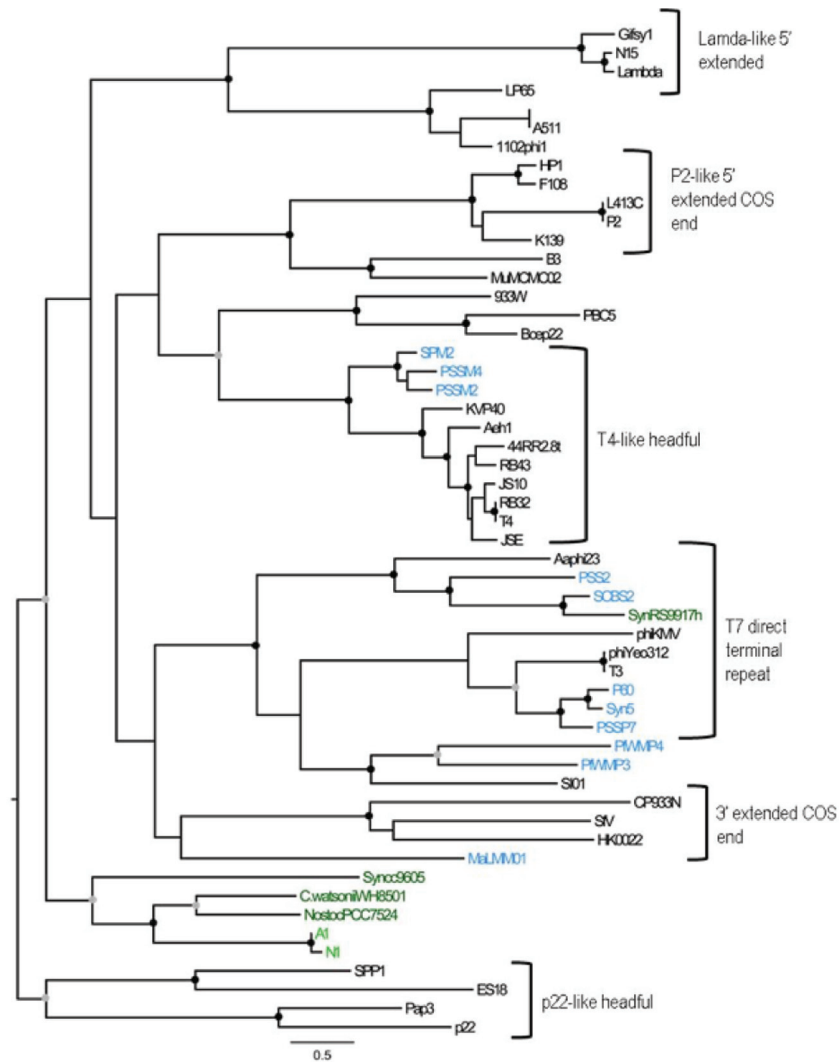


FIG 4 Phylogenetic relationship of terminase large subunit protein sequences from phages. A ML tree is shown with bootstrap values of 90 to 100% (black circles) and 75 to 89 % (gray circles) shown at the nodes. The sequence names are colored as follows: light green, *Nostoc* cyanophages; dark green, cyanobacteria; light blue, other cyanophages; black, other viruses.

stress response of *Synechococcus* sp. strain WH8102 grown under low-phosphate conditions (45).

***Nostoc* cyanophage-related genes were also found in the genome of *Nostoc* sp. strain PCC 7524.** Fourteen ORFs in cyanophages A-1 and N-1 that putatively encode structural proteins, terminases, lysozymes, and peptidases had high similarity to ORFs in *Nostoc* sp. strain PCC 7524 (NCBI reference sequence accession no. NC019684.1), leading to the reannotation of about 30 kb of sequence and the identification of a prophage-like element (Fig. 1C). Although lysogeny has been reported in natural *Synechococcus* communities (46, 47), few prophage-like elements have been detected in cyanobacterial genomes; for example, none were found in a dozen marine picocyanobacterial genomes (48, 49). However, evidence of lysogeny was recently found in the genomes of *Synechococcus elongatus* strains PCC 6301 and PCC 7942 (50). The lifestyles of cyanophages A-1 and N-1 have been reported to be lytic, but similar prophage-like elements in *Nostoc* sp. strain PCC 7524 suggest that related phages have the potential for lysogeny. Moreover, the presence of ORFs in the host with high

similarity to sequences in A-1 and N-1 indicates that genetic exchange occurs possibly via prophage integration or homologous recombination.

Cyanophage N-1 contains a CRISPR array. During genomic analysis of cyanophage N-1, a region of about 400 bp was identified (Fig. 5A) that comprises four spacers and five 37-bp-long DRs that are similar in structure to the DR5 family of CRISPRs commonly found in cyanobacteria (Table 4; Fig. 5B). Spacers in the N-1 CRISPR vary in length from 29 to 37 bp and from 24 to 54% in GC content (see Table S3 in the supplemental material) but did not have significant matches to other sequences in the NCBI nonredundant (nr) nucleotide database. An AT-rich sequence region (~25.6% G+C content) of approximately 120 bp, upstream of the CRISPR array, was considered the leader region. A leader region can be an AT-rich area that is upstream of the CRISPR array, as was observed in the genome of *Nostoc* phage N-1. Neighbor-joining analysis revealed that the DRs in cyanophage N-1 clustered among those in filamentous cyanobacteria (Fig. 5C) and were most similar to three sets of consensus DRs from

TABLE 2 Predicted ORFs in cyanophage A-1 with high similarity to cyanobacterial genes

ORF	Length (bp)	Strand ^a	Significant hit	Organism (accession no.)	E value	% Identity (no. of shared aa) ^b
1	1,359	F	Terminase large subunit	<i>Nostoc</i> sp. strain PCC 7524 (AFY48994.1)	2.28e ⁻⁵⁴	30 (129)
2	1,605	F	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY48995.1)	7.16e ⁻¹⁷	21 (113)
9	1,521	F	Tail sheath protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49006.1)	1.14e ⁻⁶⁹	38 (137)
16	1,416	R	Lysozyme-like domain, rare lipoprotein A	<i>Anabaena variabilis</i> ATCC 29413	2.3e ⁻¹⁹	57.7 (60)
19	738	R	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49010.1)	9.39e ⁻¹¹	29 (67)
23	891	R	Exonuclease RNase T and DNA polymerase	<i>Cyanobacterium aponinum</i> PCC 10605	1.63e ⁻⁰⁸	27 (48)
26	813	F	Phage-related baseplate assembly protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49015.1)	1.46e ⁻³²	34 (91)
32	738	F	Baseplate J phage tail	<i>Nostoc</i> sp. strain PCC 7524 (AFY49018.1)	3.76e ⁻³³	41 (94)
34	576	F	Phage tail protein (tail_P2_I)	<i>Nostoc</i> sp. strain PCC 7524 (AFY49020.1)	5.63e ⁻⁴⁰	59 (77)
35	1,389	F	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49021.1)	4.83e ⁻⁵⁷	37 (157)
36	1,140	F	Tail collar protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49022.1)	2.1e ⁻⁵²	50 (149)
39	2,016	R	DNA Pol B	<i>Cyanothece</i> sp. strain PCC 7425	7.00e ⁻¹¹⁵	40.4 (237)
54	939	R	Putative ant AntA/AntB antirepressor	<i>Leptolyngbya</i> sp. strain PCC 7375	1.90e ⁻²⁴	45 (51)
59	618	R	DNA N-6-adenine-methyltransferase	<i>Synechocystis</i> sp. strain PCC 7509	1.00e ⁻⁰⁶	27.3 (44)
72	1,209	F	Transposase	<i>Nostoc</i> sp. strain PCC 7120	0	388 (100)
75	417	F	Hypothetical protein	<i>Calothrix</i> sp. strain PCC 7103	2.59e ⁻⁴⁸	58 (98)
86	600	F	dCTP deaminase/dUTPase superfamily	<i>Cyanothece</i> sp. strain PCC 7425	2.00e ⁻⁷⁶	67.3 (134)
89	888	F	DNA methylase N-4/N-6 domain protein	<i>Arthrospira maxima</i>	1.00e ⁻⁶⁵	49.4 (133)
90	432	F	Endo-DNase RusA	<i>Cyanothece</i> sp. strain PCC 7425	8.00e ⁻⁰⁴	28.4 (29)
92	780	R	Thymidylate synthase complementing protein	<i>Chlorobium phaeobacteroides</i> BSI	4.38e ⁻³⁹	38.4 (86)

^a F, forward; R, reverse.^b aa, amino acids.

CRISPR arrays found in the genome of *Calothrix* sp. strain PCC 7507. Overall, DRs found within a cyanobacterial genome are not necessarily most closely related to each other. For example, some CRISPRs in *Nostoc* sp. strain PCC 7210 and *A. variabilis* ATCC 29413 have repeats that are more similar to those in N-1 than to other repeats within their own genomes (Fig. 5C).

The CRISPR array found in cyanophage N-1 is similar to the DR5 family of CRISPRs commonly found in cyanobacteria and is predicted to have the same characteristic hairpin structure (11) found in DR5 (type I-D) CRISPR repeats (data not shown) (51, 52). The CRISPR-Cas system is widespread among cyanobacteria, with 86 out of 126 sequenced genomes containing CRISPR-Cas

TABLE 3 Predicted ORFs in cyanophage N-1 with high similarity to cyanobacterial genes

ORF	Length (bp)	Strand ^a	Significant hit	Organism (accession no.)	E value	% Identity (no. of shared aa) ^b
1	1,359	F	Phage terminase, large subunit	<i>Nostoc</i> sp. strain PCC 7524	2.32e ⁻⁵⁵	31 (134)
2	1,602	F	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY48995)	3.55e ⁻¹⁸	22 (106)
6	330	F	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49001)	9.62e ⁻⁰⁶	30 (30)
9	1,521	F	Tail sheath protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49006)	3.74e ⁻⁷⁴	39 (138)
15	2,694	R	Lysozyme-like domain, rare lipoprotein A (RlpA)-like double psi beta barrel	<i>Nostoc</i> sp. strain PCC 7524 (AFY49014)	1.48e ⁻³⁷	28 (137)
18	849	R	Hypothetical protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49010)	9.82e ⁻¹⁰	27.9 (51)
19	918	F	Exonuclease RNase T and DNA Pol III	<i>Thauera</i> sp. strain MZIT	4.14e ⁻⁰⁴	28.4 (38)
29	330	F	Lysozyme	<i>Nostoc</i> sp. strain PCC 7524 (AFY49017)	2.51e ⁻⁰³	24 (32)
30	1,167	F	Baseplate J phage tail protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49018)	4.18e ⁻⁶⁸	43 (144)
31	576	F	Phage tail protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49020)	2.32E ⁺⁰⁰	26.3 (45)
32	1,395	F	Phage tail fiber protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49021)	2.97e ⁻⁶⁴	37 (172)
33	1,155	F	Tail collar protein	<i>Nostoc</i> sp. strain PCC 7524 (AFY49022)	1.62e ⁻⁴⁷	33 (137)
34	1,896	R	DNA Pol B	<i>Cyanothece</i> sp. strain PCC 7424	7.0e ⁻¹¹⁸	41.5 (243)
48	600	R	Hypothetical protein	<i>Nostoc punctiforme</i> PCC 73102	1.10e ⁻⁰⁶	21.4 (40)
52	633	R	C-5 cytosine-specific DNA methylase	<i>Nostoc punctiforme</i> PCC 73102	1.81e ⁻⁰⁵	36.2 (34)
56	240	R	ASCH domain protein	<i>Clostridium</i> phage phiMMP02	4.06e ⁻¹²	42.3 (33)
62	627	F	Thymidylate kinase	<i>Lyngbya</i> sp. strain PCC 8106	2.45e ⁻²³	36 (74)
68	1,047	F	DNA-cytosine methyltransferase	<i>Anabaena variabilis</i> ATCC 29413	1.02e ⁻⁵⁶	33.2(127)
75	579	R	Hypothetical protein	<i>Thermoanaerobacter italicus</i> Ab9	8.67e ⁻⁰³	32.7 (33)
76	507	R	Hypothetical protein	<i>Calothrix</i> sp. strain PCC 7103	1.73e ⁻⁴⁷	56 (94)
77	726	R	Hypothetical protein	<i>Calothrix</i> sp. strain PCC 7103	4.94e ⁻⁰⁶	30.4 (78)
79	585	F	dCTP deaminase	<i>Synechococcus</i> sp. strain PCC 7335	1.45e ⁻⁶⁹	65.5 (131)

^a F, forward; R, reverse.^b aa, amino acids.

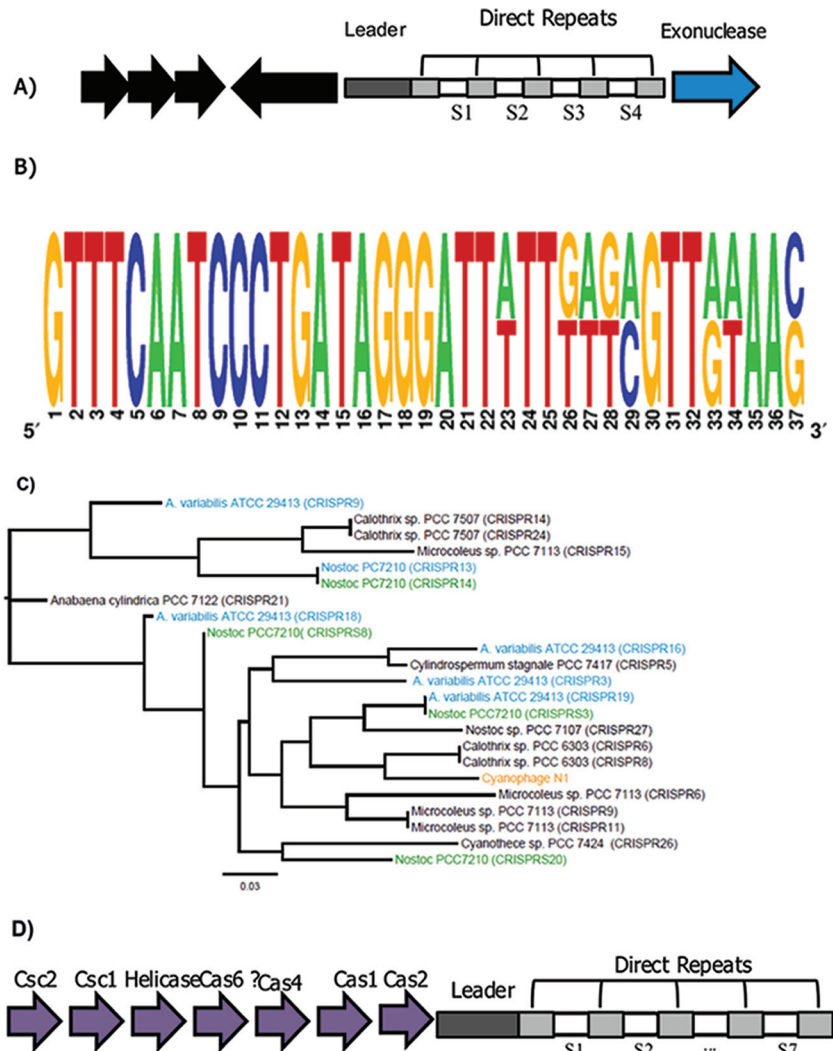


FIG 5 Characterization of the CRISPR in cyanophage N-1. (A) The CRISPR in cyanophage N-1 consists of five DRs (light gray boxes), four spacers (white boxes), and a leader sequence (dark gray box). The CRISPR array is surrounded by ORFs that putatively encode an exonuclease (blue arrows) or are hypothetical (black arrows). (B) A logo illustrating nucleotide differences in an alignment of the consensus DR from cyanophage N-1 and the consensus DR from CRISPR arrays 13 and 14 in *Nostoc* sp. strain PCC 7210. Differences are shown by placement of the letters for the nucleotides of *Nostoc* on the top and those for cyanophage N-1 on the bottom of the logo. (C) An unrooted neighbor-joining tree shows that the consensus DR in cyanophage N-1 (orange) is not most closely related to those found in its known hosts, *Nostoc* sp. strain PCC 7210 (green) and *A. variabilis* ATCC 29413 (light blue). Consensus DRs from other cyanobacteria, including those that are most closely related in *Calothrix* sp. strain PCC 6303, are shown in black. The scale bar represents 0.03 nucleotide change. (D) Schematic of CRISPR 8 in *Nostoc* sp. strain PCC 7210. The CRISPR array consists of DRs (light gray boxes), seven spacers (white boxes), and a leader sequence (dark gray box). The CRISPR array is surrounded by CAS genes (purple arrows).

systems and with multiple CRISPR arrays in many genomes (52). This includes *Nostoc* sp. strain PCC 7210 and *A. variabilis* ATCC 29413, which include 13 and 11 CRISPR arrays and 106 and 183 spacers, respectively (53). The sequence similarity of CRISPR repeats between cyanobacteria and N-1 suggests that the N-1 CRISPR was transferred from a cyanobacterium to an ancestor of N-1 during an infection, confirming that the CRISPR-Cas system has been exchanged by lateral gene transfer among microorganisms (12, 54).

Although no *cas* genes were identified in the N-1 genome, the N-1 CRISPR array was transcribed during infection (see Fig. S2 in the supplemental material). Possibly, N-1 contains unidentified genes encoding Cas proteins or host Cas proteins may be used for initiation. CRISPR loci can function without proximate *cas* genes

(11, 54), and different CRISPR loci with similar repeats in the same genome can use the same set of Cas proteins. In *Nostoc* sp. strain PCC 7210, CRISPR8 is adjacent to a *cas* operon and is similar to the repeats in the N-1 CRISPR (Fig. 5D). Nothing is known about the expression of the CRISPR-Cas system in *Nostoc*, but in another type I CRISPR-Cas system in *Escherichia coli*, Cas proteins are continuously transcribed (55). This provides a mechanism for expression of the N-1 CRISPR array, even in the absence of virus-encoded Cas proteins.

The previous isolation of phages infecting *V. cholerae* that encode a CRISPR-Cas system demonstrated that the system defeats an inhibitory chromosomal island of the bacterial host (56). CRISPR arrays were also recently identified in prophages of *Clostridium difficile* (57). The prophage-carried CRISPR ar-

TABLE 4 BLAST results for the DRs from the CRISPR array present in the genome of cyanophage N-1

Cyanobacterial stain (CRISPR ID)	E value	Start	End	DR consensus sequence
<i>Microcoleus</i> sp. strain PCC 7113 (CRISPR6)	3.0e ⁻⁰⁸	2087079	2087333	GTCTGAATCCATATAATCCCTATCAGGGATTGAAAC
<i>Microcoleus</i> sp. strain PCC 7113 (CRISPR11)	1.0e ⁻⁰⁷	3131505	3132053	GTTTAAATCCACTTAATCCCTATCAGGGATTGAAAC
<i>Microcoleus</i> sp. strain PCC 7113 (CRISPR15)	1.0e ⁻⁰⁷	3859458	3860565	GTTTCAATCCCTGATAGGGATTAAGTGGAAATTTAAAC
<i>Microcoleus</i> sp. strain PCC 7113 (CRISPR9)	1.0e ⁻⁰⁷	2916270	2917173	GTTTAAATCCACTTAATCCCTATCAGGGATTGAAAC
<i>Nostoc</i> sp. strain PCC 7107 (CRISPR27)	1e ⁻⁰⁷	5817107	5818597	GTTGCAATTTCTATTAATCCCTATCAGGGATTGAAAC
<i>A. variabilis</i> ATCC 29413 (CRISPR3)	8.0e ⁻⁰⁷	1234556	1237182	GTTTAAATTAACAAAAATCCCTATCAGGGATTGAAAC
<i>Nostoc</i> sp. strain PCC 7210 (CRISPR13)	8.0e ⁻⁰⁷	3516819	3517367	GTTTCAATCCCTGATAGGGATTTTTGTTAGTTAAAAAC
<i>Nostoc</i> sp. strain PCC 7210 (CRISPR14)	8.0e ⁻⁰⁷	3517542	3518084	GTTTCAATCCCTGATAGGGATTTTTGTTAGTTAAAAAC
<i>A. cylindrica</i> PCC 7122 (CRISPR21)	4e ⁻⁰⁷	5001036	5003639	GTTTCAATCCCTAATAGGGATTAATTTGAAATTTCAAC
<i>C. stagnale</i> PCC 7417(CRISPR5)	4e ⁻⁰⁷	1101502	1103134	GTTACAATTCACCCAAATCCCTATCAGGGATTGAAAC
<i>Calothrix</i> sp. strain PCC 6303 (CRISPR6)	4e ⁻⁰⁷	2021864	2022765	GTTTCTATAAACTAAAATCCCTATCAGGGATTGAAAC
<i>Calothrix</i> sp. strain PCC 6303 (CRISPR8)	4e ⁻⁰⁷	2038085	2039287	GTTTCTATAAACTAAAATCCCTATCAGGGATTGAAAC
<i>Calothrix</i> sp. strain PCC 7507 (CRISPR24)	4e ⁻⁰⁷	5067421	5070798	GTTTCAATCCCTGATAGGGATTTAAGTTAATTGGAAC
<i>Calothrix</i> sp. strain PCC 7507 (CRISPR14)	4e ⁻⁰⁷	3375025	3376306	GTTTCAATCCCTGATAGGGATTTAAGTTAATTGGAAC
<i>N. punctiforme</i> PCC 73102 (CRISPR16)	3.0e ⁻⁰⁷	3338172	3341197	GTTTCAATCCCTGATAGGGATTTTGATGAATTGCAAT
<i>Nostoc</i> sp. strain PCC 7210 (CRISPR8)	3.0e ⁻⁰⁷	1836813	1837723	GTTTCTATTAACACAAATCCCTATCAGGGATTGAAAC
<i>Nostoc</i> sp. strain PCC 7210 (CRISPR3)	3.0e ⁻⁰⁷	807452	807558	ATTGCAATTAATACTAAAATCCCTATCAGGGATTGAAAC
<i>A. variabilis</i> ATCC 29413 (CRISPR19)	3.0e ⁻⁰⁷	5764010	5766428	ATTGCAATTAATACTAAAATCCCTATCAGGGATTGAAAC
<i>Nostoc</i> sp. strain PCC 7210 (CRISPR20)	3.0e ⁻⁰⁷	5654133	5654384	GTTAAAAACCTCTAAAATCCCTATCAGGGATTGAAAC
<i>A. variabilis</i> ATCC 29413 (CRISPR9)	3.0e ⁻⁰⁷	2395670	2398703	GTTTCAATCCCTGATAGGGATTTAGAGGGTTTTAAC
<i>A. variabilis</i> ATCC 29413 (CRISPR18)	1.0e ⁻⁰⁶	5227213	5229282	GTTTCTATTAACACAAATCCCTATCAGGGATTGAAAG
<i>Cyanothece</i> sp. strain PCC 7424 (CRISPR26)	4.0e ⁻⁰⁶	3575124	3575742	GTTACAATTAATAATGAATCCCTATTAGGGATTGAAAC
<i>A. variabilis</i> ATCC 29413 (CRISPR16)	4.0e ⁻⁰⁶	4821250	4823752	GTTGCAACACCACATAATCCCTATTAGGGATTGAAAC

rays were generally located in the structure/morphogenesis module and were similar to CRISPR arrays in the host, *C. difficile*. The prophage-carried CRISPR arrays contained spacers that match other *C. difficile* phage sequences and hence likely confer resistance to other phages that can infect the host. In the case of the N-1 CRISPR, the origin of the spacers is unknown, which makes it impossible to identify the target of the array. However, the CRISPR array in cyanophage N-1 may act as a mechanism against coinfection and offer a fitness advantage to both the host and the virus by preventing lysis by a competing phage.

The ability of viruses to contain and likely transfer CRISPRs benefits both the host and the carrier virus by introducing new spacers that protect the host from a wider spectrum of viruses that are potential competitors of the carrier virus. Expression of the viral CRISPR may be a particular selective advantage of filamentous cyanobacteria, because molecules are believed to be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot pass through these channels, while small molecules such as CRISPR RNA (crRNA) probably can. This could allow cells adjacent to those infected by N-1 to acquire CRISPR sequences that would confer immunity to infection by other competing viruses.

The presence of a CRISPR array in cyanophage N-1 implies that phages can acquire CRISPRs and suggests that they can also transfer them among host cells. As the spacers likely originate from other phages, transferring the array among hosts has the potential to confer host immunity to competing phages.

Conclusion. Cyanophages A-1 and N-1, which infect *Nostoc* sp. strain PCC 7120, belong to a previously unrecognized evolutionary lineage of tailed phages. Most of their predicted protein-coding genes have no obvious similarity to sequences in databases, and those that do are generally most similar to genes found in filamentous cyanobacteria. Also, the closest phylogenetic relative

of the DNA polymerases in A-1 and N-1 is found in a giant plasmid in the host. Moreover, the presence of a CRISPR array in N-1 suggests that cyanophages may exchange CRISPR elements among cyanobacteria. Together, these findings indicate the important role that lateral gene transfer has played in forging the evolutionary trajectory of this previously unrecognized evolutionary lineage of phages.

MATERIALS AND METHODS

Cyanophage isolation, purification, DNA preparation, and genome sequencing. Cyanophages A-1 and N-1 (ATCC 27893-B16 and ATCC 27893-B15, respectively) are tailed phages belonging to the family *Myoviridae* that infect *Nostoc* sp. strain PCC 7120 (ATCC 27893). Cyanophages were amplified on *Nostoc* sp. strain PCC 7120 grown in 800-ml batch cultures in 1-liter Erlenmeyer flasks containing BG-11 medium (59) under constant illumination (33 μmol of photons $\text{m}^{-2} \text{s}^{-1}$, photosynthetically active radiation) at 26°C with constant shaking at 75 rpm. Exponentially growing cultures were infected with either A-1 or N-1 at a multiplicity of infection (MOI) of about 1 and left for 4 to 7 days until transparent, indicating lysis. To prevent phage binding to the filter, sodium chloride was added to the lysate at a final concentration of 0.5 M and the cultures were incubated at 4°C for 1 h before filtration through a 1.2- μm -pore-size GC50 glass fiber filter (Advantec MFS, Dublin, CA) and twice through GVWP 0.22- μm -pore-size polyvinylidene low-protein-binding filters (Millipore, Bedford, MA). Subsequently, the viral particles were concentrated by polyethylene glycol (PEG) precipitation (60). Briefly, the filtered lysate was centrifuged at 10,000 $\times g$ for 10 min in a Sorvall RC-5C centrifuge (GSA rotor, 4°C) to remove cellular debris. PEG 6000 was added to the supernatant to a 10% final concentration, and the mixture was incubated overnight at 4°C with constant shaking. The PEG solution was then centrifuged at 16,000 $\times g$ for 20 min (GSA rotor, 4°C), the supernatant was removed, and the pelleted viruses were resuspended in 200 μl of BG-11 medium.

DNA was extracted by treating the resuspended pellet with DNase I and RNase A to remove free nucleic acids with the QIAamp MinElute Virus Spin kit (Qiagen, Mississauga, ON, Canada) in accordance with the manufacturer's instructions. The DNA was sequenced by 454 GS

FLX titanium pyrosequencing at the G enome Qu ebec Innovation Centre, McGill University (Montreal, Quebec, Canada). For each phage, >36,000 reads with an average length of ~350 bp were assembled into three contiguous sequences (contigs) with the GS *de novo* Assembler (Roche) and closed into a single circular contig by PCR. The sequencing coverage was approximately 179-fold for A-1 and 250-fold for N-1.

Genome annotation. ORFs were predicted with GeneMark (61) and GLIMMER (62). To create the final predictions, the ORF calls from the two programs were combined. When ORFs predicted by both programs differed in size, the longer of the two was kept. The final set of predicted ORFs was translated and assigned putative functions by comparison with known protein sequences found in the GenBank (nr), Acclame, and Procite databases with the BLASTp program. The ORFs were considered to be homologous to a protein-encoding gene if the E value was $<10^{-4}$. Identification of tRNA genes was performed with tRNAscan-SE (63). Putative promoter motifs were identified with PHIRE (38) and the default parameters of 20-mer DNA sequences (S) with 4-bp degeneracy ($D = 4$). The motif was considered a putative promoter if it was found in the 150-bp region immediately upstream of the start codon of a predicted protein-coding gene. Sequence logos of the motifs were created with WebLogo by using the alignment of the sequences (64). Identification of rho-independent terminators was performed with FINDTERM (Softberry, Inc.). The default energy threshold was set to -16 kcal for the analysis as previously described (25).

Phylogenetic analysis. The ORFs with similarity to genes coding for dCTP deaminase, DNA Pol B, and the large terminase subunit were used for phylogenetic analysis. Inferred amino acid sequences for the dCTP deaminase and DNA Pol B were aligned in ClustalX with default parameters, while the large terminase gene (*terL*) was aligned by using the Promals web server (65, 66) with default parameters. Geneious v4.7 (67) was used to manually refine the alignment and construct neighbor-joining trees. Maximum-likelihood (ML) trees were constructed with the RAxML Web-Server rapid bootstrapping and ML search (100 replicates) (68) assuming the James-Taylor Thornton model of substitution with empirical base frequencies and estimating the proportion of invariable sites from the data.

Gene comparison with other bacteriophages. Predicted ORFs were compared to a database for T4-like phages infecting heterotrophic bacteria (10) and marine (16, 35) and freshwater (1) cyanobacteria (26). An ORF was considered shared if the E value was $<10^{-3}$.

Analysis of the CRISPR array in cyanophage N-1. During the genomic analysis of N-1, a repeat DNA region was found and identified as a CRISPR array by CRISPRFinder (69). To show that this region was not a sequencing or assembly error, the CRISPR region was checked by PCR and sequencing. Two microliters of a 0.22- μ m-filtered N-1 lysate was added to a 48- μ l PCR mixture containing Platinum *Taq* DNA polymerase assay buffer (50 mM KCl, 20 mM Tris-HCl, pH 8.4), 10 mM MgCl₂, 200 μ M deoxynucleoside triphosphate, 0.25 μ M each primer (sCRF [CA ATTGGCAAAAAGATTAGCAGC] and CR3R [GGGGAGAGGTTTGGAGAGGGGT]), and 2.0 U of Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA). Negative controls contained all of the reagents but with sterile water as the template. PCR was carried out by denaturation at 94°C for 5 min; 35 cycles of denaturation at 94°C for 30 s, annealing at 57°C for 45 s, and extension at 72°C for 1 min; and a final extension at 72°C for 10 min (70). The amplification products were subjected to electrophoresis with 1.5% agarose-0.5 \times Tris-borate-EDT buffer (45 mM Tris-borate, 1 mM EDTA [pH 8.0]) at 100 V for 60 min. Gels were stained with Gel-Green (Invitrogen) and visualized under UV illumination. The PCR amplicons were sequenced by Sanger sequencing.

Cyanophage N-1 DRs were compared with those from the CRISPR database with a BLASTN E value of $<10^{-5}$. A sequence logo of the aligned consensus DRs from N-1 with the consensus of CRISPR 13 in *Nostoc* sp. strain PCC 7210 was created with WebLogo (64). The genomes of N-1 and its host were screened for *cas* genes by using the nr database. A neighbor-

joining tree (Jukes-Cantor model) was created with the consensus DRs from the N-1 CRISPR and other cyanobacterial CRISPRs with Geneious (67) and edited with FigTree v1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree>).

RNA isolation and RT-PCR. To confirm the transcription of the N1 CRISPR loci, the presence of precursor crRNA (pre-crRNA) was analyzed by reverse transcriptase PCR (RT-PCR). Total RNA was extracted from host cells infected with N-1. Two flasks, each containing 100 ml of an exponentially growing *Nostoc* sp. strain PCC 7210 culture, were infected with N-1 lysate, and 15 ml was collected at day 5. *Nostoc* cells were pelleted by centrifugation and resuspended in BG-11 medium. RNA was extracted with TRIzol reagent in accordance with the manufacturer's protocol (Life Technologies). Briefly, 0.75 ml of TRIzol reagent was added to the resuspended pellets. The cells were lysed by being pipetted up and down several times. The samples were centrifuged at $12,000 \times g$ for 10 min at 4°C (Beckman Coulter Allegra X-22R). The supernatant was then transferred to a new microcentrifuge tube and incubated for 5 min at room temperature to permit complete dissociation of the nucleoprotein complex. A 0.2-ml volume of chloroform was then added to the tube, and it was incubated for 3 min at room temperature. The sample was then centrifuged for 15 min as described above, and the resulting aqueous phase was collected and placed into a new microcentrifuge tube. A 0.5-ml volume of 100% isopropanol was added to the aqueous phase, which was incubated at room temperature for 10 min. The sample was centrifuged for 10 min at 4°C as described above, the supernatant was removed from the tube, and the pellet was washed with 1 ml of 75% ethanol. The RNA pellet was then resuspended in 50 μ l of RNase-free water.

RT-PCR targeted the sequence between spacers 1 and 4 (~150 bp). First, an aliquot of the extracted RNA was treated with DNase I (Invitrogen) to remove DNA. The cDNA was generated with Superscript III RT (Invitrogen) by using random hexamers (50 ng/ μ l). Amplification was carried out in 25- μ l PCR mixtures containing 10 ng of a cDNA template, 1 μ M each primer, 1.5 mM MgCl₂, 0.2 mM deoxynucleoside triphosphates, and 0.5 U of Platinum *Taq* DNA polymerase (Invitrogen). The PCR cycle parameters were a single denaturation step of 95°C for 5 min; 35 cycles of 95°C for 30 s, 57°C for 1 min, and 72°C for 3 min; and a final extension step of 72°C for 10 min.

Nucleotide sequence accession numbers. The nucleotide sequences of the dsDNA genomes of cyanophages A-1 and N-1 have been deposited in the GenBank database under accession no. KU234533 and KU234532, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00667-16/-/DCSupplemental>.

Figure S1, TIF file, 1.4 MB.

Figure S2, EPS file, 2.1 MB.

Table S1, DOCX file, 0.1 MB.

Table S2, DOCX file, 0.1 MB.

Table S3, DOCX file, 0.04 MB.

ACKNOWLEDGMENTS

We are grateful to J. A. Gustavsen for helpful comments on the manuscript.

This research was funded by grants to C.A.S. from the Natural Sciences and Engineering Research Council of Canada (NSERC) and the Canadian Institute for Advanced Research, by equipment purchased through the Canadian Foundation for Innovation and the British Columbia Knowledge Development Fund, as well as scholarship funding to C.C. from the Fonds de Recherche du Qu ebec—Nature et Technologies, a UBC University graduate fellowship, and a BRITE award from the NSERC CREATE Program.

FUNDING INFORMATION

This work, including the efforts of Curtis Suttle, was funded by Canadian Foundation for Innovation. This work, including the efforts of Curtis Suttle, was funded by British Columbia Knowledge Development Fund. This work, including the efforts of Curtis Suttle, was funded by Gouvernement du Canada | Natural Sciences and Engineering Research Council of Canada (NSERC). This work, including the efforts of Curtis Suttle, was funded by Canadian Institute for Advanced Research (CIFAR). This work, including the efforts of Caroline Chénard, was funded by Fonds de Recherche du Québec - Nature et Technologies (FRQNT).

REFERENCES

- Jungblut AD, Lovejoy C, Vincent WF. 2010. Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME J* 4:191–202. <http://dx.doi.org/10.1038/ismej.2009.113>.
- Moreno J, Rodríguez H, Vargas MA, Rivas J, Guerrero MG. 1995. Nitrogen-fixing cyanobacteria as source of phycobiliprotein pigments. Composition and growth performance of ten filamentous heterocystous strains. *J Appl Phycol* 7:17–23. <http://dx.doi.org/10.1007/BF00003545>.
- Kim J-D. 2006. Screening of cyanobacteria (blue-green algae) from rice paddy soil for antifungal activity against plant pathogenic fungi. *Mycobiology* 34:138–142. <http://dx.doi.org/10.4489/MYCO.2006.34.3.138>.
- Potts M. 2000. Nostoc, p 465–504. In Whitton B, Potts M (ed), *The ecology of cyanobacteria*, Springer, Netherlands.
- Labrie SJ, Samson JE, Moineau S. 2010. Bacteriophage resistance mechanisms. *Nat Rev Microbiol* 8:317–327. <http://dx.doi.org/10.1038/nrmicro2315>.
- Deveau H, Garneau JE, Moineau S. 2010. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol* 64:475–493. <http://dx.doi.org/10.1146/annurev.micro.112408.134123>.
- Horvath P, Barrangou R. 2010. CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327:167–170. <http://dx.doi.org/10.1126/science.1179555>.
- van der Oost J, Jore MM, Westra ER, Lundgren M, Brouns SJ. 2009. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* 34:401–407. <http://dx.doi.org/10.1016/j.tibs.2009.05.002>.
- Bhaya D, Grossman AR, Steunou AS, Khuri N, Cohan FM, Hamamura N, Melendrez MC, Bateson MM, Ward DM, Heidelberg JF. 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J* 1:703–713. <http://dx.doi.org/10.1038/ismej.2007.46>.
- Bhaya D, Davison M, Barrangou R. 2011. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu Rev Genet* 45:273–297. <http://dx.doi.org/10.1146/annurev-genet-110410-132430>.
- Kunin V, Sorek R, Hugenholtz P. 2007. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* 8:R61. <http://dx.doi.org/10.1186/gb-2007-8-4-r61>.
- Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, Moineau S, Mojica FJ, Wolf YI, Yakunin AF, van der Oost J, Koonin EV. 2011. Evolution and classification of the CRISPR-Cas systems. *Nat Rev Microbiol* 9:467–477. <http://dx.doi.org/10.1038/nrmicro2577>.
- 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. <http://dx.doi.org/10.1007/s00239-005-0223-z>.
- Sebahia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdeño-Tárraga AM, Wang H, Holden MT, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabinowitz E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. 2006. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet* 38:779–786. <http://dx.doi.org/10.1038/ng1830>.
- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. 2011. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 21:1616–1625. <http://dx.doi.org/10.1101/gr.122705.111>.
- Suttle CA. 2000. Cyanophages and their role in the ecology of cyanobacteria, p 563–589. In Whitton BA, Potts M (ed), *The ecology of cyanobacteria: their diversity in time and space*. Kluwer Academic Publishers, Boston, MA.
- Adolph KW, Haselkorn R. 1971. Isolation and characterization of a virus infecting the blue-green alga *Nostoc muscorum*. *Virology* 46:200–208. [http://dx.doi.org/10.1016/0042-6822\(71\)90023-7](http://dx.doi.org/10.1016/0042-6822(71)90023-7).
- Adolph KW, Haselkorn R. 1973. Blue-green algal virus N-1: physical properties and disassembly into structural parts. *Virology* 53:427–440. [http://dx.doi.org/10.1016/0042-6822\(73\)90222-5](http://dx.doi.org/10.1016/0042-6822(73)90222-5).
- Hu N, Thiel T, Giddings TH, Wolk CP. 1981. New Anabaena and Nostoc cyanophages from sewage settling ponds. *Virology* 114:236–246. [http://dx.doi.org/10.1016/0042-6822\(81\)90269-5](http://dx.doi.org/10.1016/0042-6822(81)90269-5).
- Adolph KW, Haselkorn R. 1972. Photosynthesis and the development of the blue-green algal virus N-1. *Virology* 47:370–374. [http://dx.doi.org/10.1016/0042-6822\(72\)90272-3](http://dx.doi.org/10.1016/0042-6822(72)90272-3).
- Koz'yakov S. 1977. Cyanophages of the series A(L) specific for the blue-green alga *Anabaena variabilis*. *Exp Algal Biol Sci Res* 25:151–175.
- Adolph KW, Haselkorn R. 1972. Photosynthesis and the development of blue-green algal virus N-1. *Virology* 47:370–374. [http://dx.doi.org/10.1016/0042-6822\(72\)90272-3](http://dx.doi.org/10.1016/0042-6822(72)90272-3).
- Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Rüger W. 2003. Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67:86–156. <http://dx.doi.org/10.1128/MMBR.67.1.86-156.2003>.
- Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, Letarov A, Krisch HM. 2005. The genome of S-PM2, a “photosynthetic” T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* 187:3188–3200. <http://dx.doi.org/10.1128/JB.187.9.3188-3200.2005>.
- Weigle PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, King J, Hatfull GF, Lawrence JG, Hendrix RW. 2007. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* 9:1675–1695. <http://dx.doi.org/10.1111/j.1462-2920.2007.01285.x>.
- Dreher TW, Brown N, Bozarth CS, Schwartz AD, Riscoe E, Thrash C, Bennett SE, Tzeng S-C, Maier CS. 2011. A freshwater cyanophage whose genome indicates close relationships to photosynthetic marine cyanomyophages. *Environ Microbiol* 13:1858–1874. <http://dx.doi.org/10.1111/j.1462-2920.2011.02502.x>.
- Millard AD, Zwirgmaier K, Downey MJ, Mann NH, Scanlan DJ. 2009. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ Microbiol* 11:2370–2387. <http://dx.doi.org/10.1111/j.1462-2920.2009.01966.x>.
- Sullivan MB, Coleman ML, Weigle P, Rohwer F, Chisholm SW. 2005. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* 3:e144. <http://dx.doi.org/10.1371/journal.pbio.0030144>.
- Yoshida T, Nagasaki K, Takashima Y, Shirai Y, Tomaru Y, Takao Y, Sakamoto S, Hiroishi S, Ogata H. 2008. Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol* 190:1762–1772. <http://dx.doi.org/10.1128/JB.01534-07>.
- Comeau AM, Tremblay D, Moineau S, Rattei T, Kushkina AI, Tovkach FI, Krisch HM, Ackermann H-W. 2012. Phage morphology recapitulates phylogeny: the comparative genomics of a new group of myoviruses. *PLoS One* 7:e40102. <http://dx.doi.org/10.1371/journal.pone.0040102>.
- Chénard C, Chan AM, Vincent WF, Suttle CA. 2015. Polar freshwater cyanophage S-EIV1 represents a new widespread evolutionary lineage of phages. *ISME J* 9:2046–2058. <http://dx.doi.org/10.1038/ismej.2015.24>.
- Liu X, Kong S, Shi M, Fu L, Gao Y, An C. 2008. Genomic analysis of freshwater cyanophage Pf-WMP3 infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb Ecol* 56:671–680. <http://dx.doi.org/10.1007/s00248-008-9386-7>.
- Liu X, Shi M, Kong S, Gao Y, An C. 2007. Cyanophage Pf-WMP4, a T7-like phage infecting the freshwater cyanobacterium *Phormidium foveolarum*: complete genome sequence and DNA translocation. *Virology* 366:28–39. <http://dx.doi.org/10.1016/j.virol.2007.04.019>.
- Yin Y, Fischer D. 2008. Identification and investigation of ORFans in the viral world. *BMC Genomics* 9:24. <http://dx.doi.org/10.1186/1471-2164-9-24>.
- Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigle PR, Defrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osborne MS, Henn MR, Chisholm SW. 2010. Genomic analysis of oceanic cyanobacterial

- myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035–3056. <http://dx.doi.org/10.1111/j.1462-2920.2010.02280.x>.
36. Enav H, Béjà O, Mandel-Gutfreund Y. 2012. Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. *ISME J* 6:619–628. <http://dx.doi.org/10.1038/ismej.2011.146>.
 37. Mann NH, Clokie MRJ. 2012. Cyanophages, p 535–557. In Whitton BA (ed), *Ecology of cyanobacteria II*. Springer Verlag, Dordrecht, Netherlands.
 38. Lavigne R, Sun WD, Volckaert G. 2004. PHIRE, a deterministic approach to reveal regulatory elements in bacteriophage genomes. *Bioinformatics* 20:629–635. <http://dx.doi.org/10.1093/bioinformatics/btg456>.
 39. Brown KL, Sarkis GJ, Wadsworth C, Hatfull GF. 1997. Transcriptional silencing by the mycobacteriophage L5 repressor. *EMBO J* 16:5914–5921. <http://dx.doi.org/10.1093/emboj/16.19.5914>.
 40. Chan YW, Mohr R, Millard AD, Holmes AB, Larkum AW, Whitworth AL, Mann NH, Scanlan DJ, Hess WR, Clokie MR. 2011. Discovery of cyanophage genomes which contain mitochondrial DNA polymerase. *Mol Biol Evol* 28:2269–2274. <http://dx.doi.org/10.1093/molbev/msr041>.
 41. Sullivan MB, Krastins B, Hughes JL, Kelly L, Chase M, Sarracino D, Chisholm SW. 2009. The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial “mobilome.” *Environ Microbiol* 11:2935–2951. <http://dx.doi.org/10.1111/j.1462-2920.2009.02081.x>.
 42. Mesyanzhinov VV, Robben J, Grymonprez B, Kostyuchenko VA, Bourkaltseva MV, Sykilinda NN, Krylov VN, Volckaert G. 2002. The genome of bacteriophage phiKZ of *Pseudomonas aeruginosa*. *J Mol Biol* 317:1–9. <http://dx.doi.org/10.1006/jmbi.2001.5396>.
 43. Hertveldt K, Lavigne R, Pleteneva E, Sernova N, Kurochkina L, Korchevskii R, Robben J, Mesyanzhinov V, Krylov VN, Volckaert G. 2005. Genome comparison of *Pseudomonas aeruginosa* large phages. *J Mol Biol* 354:536–545. <http://dx.doi.org/10.1016/j.jmb.2005.08.075>.
 44. Kanesaki Y, Suzuki I, Allakhverdiev SI, Mikami K, Murata N. 2002. Salt stress and hyperosmotic stress regulate the expression of different sets of genes in *Synechocystis* sp. PCC 6803. *Biochem Biophys Res Commun* 290:339–348. <http://dx.doi.org/10.1006/bbrc.2001.6201>.
 45. Tetu SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, Paulsen IT. 2009. Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *ISME J* 3:835–849. <http://dx.doi.org/10.1038/ismej.2009.31>.
 46. McDaniel LD, delaRosa M, Paul JH. 2006. Temperate and lytic cyanophages from the Gulf of Mexico. *J Mar Biol Assoc UK* 86:517–527. <http://dx.doi.org/10.1017/S0025315406013427>.
 47. Ortmann AC, Lawrence JE, Suttle CA. 2002. Lysogeny and lytic viral production during a bloom of the cyanobacterium *Synechococcus* spp. *Microb Ecol* 43:225–231. <http://dx.doi.org/10.1007/s00248-001-1058-9>.
 48. Kettler GC, Martiny AC, Huang K, Zucker J, Coleman ML, Rodrigue S, Chen F, Lapidus A, Ferriera S, Johnson J, Steglich C, Church GM, Richardson P, Chisholm SW. 2007. Patterns and implications of gene gain and loss in the evolution of *Prochlorococcus*. *PLoS Genet* 3:e231. <http://dx.doi.org/10.1371/journal.pgen.0030231>.
 49. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, Paulsen IT, de Marsac NT, Wincker P, Dossat C, Ferriera S, Johnson J, Post AF, Hess WR, Partensky F. 2008. Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* 9:R90. <http://dx.doi.org/10.1186/gb-2008-9-5-r90>.
 50. Huang S, Wang K, Jiao N, Chen F. 2012. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol* 14:540–558. <http://dx.doi.org/10.1111/j.1462-2920.2011.02667.x>.
 51. Makarova KS, Aravind L, Wolf YI, Koonin EV. 2011. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct* 6:38. <http://dx.doi.org/10.1186/1745-6150-6-38>.
 52. Cai F, Axen SD, Kerfeld CA. 2013. Evidence for the widespread distribution of CRISPR-Cas system in the phylum Cyanobacteria. *RNA Biol* 10:687–693. <http://dx.doi.org/10.4161/rna.24571>.
 53. Agervald Å, Camsund D, Stensjö K, Lindblad P. 2012. CRISPR in the extended *hyp*-operon of the cyanobacterium *Nostoc* sp. strain PCC 7120, characteristics and putative function(s). *Int J Hydrogen Energy* 37:8828–8833. <http://dx.doi.org/10.1016/j.ijhydene.2012.01.127>.
 54. Horvath P, Couité-Monvoisin A-C, Romero DA, Boyaval P, Fremaux C, Barrangou R. 2009. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 131:62–70. <http://dx.doi.org/10.1016/j.ijfoodmicro.2008.05.030>.
 55. Pougach K, Semenova E, Bogdanova E, Datsenko KA, Djordjevic M, Wanner BL, Severinov K. 2010. Transcription, processing and function of CRISPR cassettes in *Escherichia coli*. *Mol Microbiol* 77:1367–1379. <http://dx.doi.org/10.1111/j.1365-2958.2010.07265.x>.
 56. Seed KD, Lazinski DW, Calderwood SB, Camilli A. 2013. A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature* 494:489–491. <http://dx.doi.org/10.1038/nature11927>.
 57. Hargreaves KR, Flores CO, Lawley TD, Clokie MR. 2014. Abundant and diverse clustered regularly interspaced short palindromic repeat spacers in *Clostridium difficile* strains and prophages target multiple phage types within this pathogen. *mBio* 5:e01045-13. <http://dx.doi.org/10.1128/mBio.01045-13>.
 58. Mullineaux CW, Mariscal V, Nenninger A, Khanum H, Herrero A, Flores E, Adams DG. 2008. Mechanism of intercellular molecular exchange in heterocyst-forming cyanobacteria. *EMBO J* 27:1299–1308. <http://dx.doi.org/10.1038/emboj.2008.66>.
 59. Rippka R, Desruelles J, Waterbury JB, Herdman M, Stanier R. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *Microbiology* 111:1–61.
 60. Sambrook J, Russell DW. 2001. *Molecular cloning: a laboratory manual*, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
 61. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res* 26:1107–1115.
 62. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with glimmer. *Bioinformatics* 23:673–679. <http://dx.doi.org/10.1093/bioinformatics/btm009>.
 63. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <http://dx.doi.org/10.1093/nar/25.5.0955>.
 64. Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190. <http://dx.doi.org/10.1101/gr.849004>.
 65. Pei J, Grishin NV. 2007. PROMALS: towards accurate multiple sequence alignments of distantly related proteins. *Bioinformatics* 23:802–808. <http://dx.doi.org/10.1093/bioinformatics/btm017>.
 66. Pei J, Kim B-H, Tang M, Grishin NV. 2007. PROMALS web server for accurate multiple protein sequence alignments. *Nucleic Acids Res* 35:W649–W652. <http://dx.doi.org/10.1093/nar/gkm227>.
 67. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A. 2011. *Geneious v5.4*. Biomatters Ltd., Auckland, New Zealand.
 68. Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* 57:758–771. <http://dx.doi.org/10.1080/10635150802429642>.
 69. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–W57. <http://dx.doi.org/10.1093/nar/gkm360>.
 70. Nübel U, Garcia-Pichel F, Muyzer G. 1997. PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63:3327–3332.