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,b,* Curtis A. Suttlea,b,c,d

**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 *Cyanothecales* 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.

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 cyanobacterial hosts. In addition, cyanophage N-1 has a CRISPR array similar to those found in cyanobacteria, suggesting that cyanophages play a role in exchanging CRISPRs among cyanobacteria.

**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 | E value | Phage hit      | Cyanophage N-1 | E value | Phage hit      |
| gp5 baseplate hub + tail lysozyme | 4.00e-5       | Phage RB49 | 4.00e-5       | Cyanophage S-PMP2 | 4.00e-5       | Phage RB49 |
| gp17 terminase DNA-packaging enzyme large subunit | 4.00e-2       | Phage S-PMP2 | 4.00e-2       | Cyanophage S-PMP2 | 4.00e-2       | Phage S-PMP2 |
| Gp41 DNA primase-helicase     | 2.00e-13      | Phage S-PMP2 | 1.00e-11      | Cyanophage S-RSM4 | 9.00e-6       | Phage Aeh1 |
| gp43 DNA Pol B               | 4.00e-5       | Phage Aeh1 | 9.00e-32      | Cyanophage P-SSM4 | 9.00e-32      | Phage Aeh1 |
| Thymidylate synthase         | 1.00e-35      | Cyanophage S-CRM01 | 9.00e-32      | Cyanophage P-SSM4 | 9.00e-32      | Cyanophage P-SSM4 |

May/June 2016 Volume 7 Issue 3 e00667-16
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 consensus 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 stoperators 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 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, dihydroxyanthoate 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.

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.
DNA phages. Phylogenetic analysis reveals that the translated nase large subunit, a protein involved in DNA packaging in ds-
the evolutionary divergence of A-1 and N-1 from other cya-
otic polymerases (40). In addition, a polar cyanophage (S-EIV1)
chloris marina
cyanophages reveals the presence of interesting polymerases. For
may have been phage mediated. Recent genomic analysis of other
plying that the transfer of DNA to the cyanobacterial plasmids
recombination can occur and thus be promiscuous to gene ex-
plasmid pP742402 is adjacent to a CRISPR, a region where
strain PCC 7424 (plasmid pP742402),
4
B was transferred laterally. Moreover, the DNA Pol B in
sequences in the phages and plasmids suggests that the gene for Pol
bacteria, the only representatives known in cyanobacteria are in
chaeal Pol B clades. Moreover, while Pol B is common in proteo-
form a distinct group that is related to proteobacterial and ar-
(plasmid C) (Fig. 3). These viral and plasmid Pol B sequences

has many viral and cellular homologues, making it a useful phy-
logenetick marker. The Pol B sequences in A-1 and N-1 were most
closely related to plasmids in the cyanobacteria Cyaanothece 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 ar-
cheal Pol B clades. Moreover, while Pol B is common in proteo-
bacteria, the only representatives known in cyanobacteria are in
these three plasmids.

The close phylogenetic relationship between the Pol B se-
quences in the phages and plasmids suggests that the gene for Pol
B was transferred laterally. Moreover, the DNA Pol B in Cyaano-
thece plasmid pP742402 is adjacent to a CRISPR, a region where
recombination can occur and thus be promiscuous to gene ex-
change. The DNA polymerases in A-1, N-1, and cyanobacterial
plasmids have an ancestor in common with proteobacteria, im-
plying 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 Acaryo-
chloris marina contain a Pol A that appears to be related to eukary-
otic polymerases (40). In addition, a polar cyanophage (S-EIV1)
recently isolated from polar inland waters contains a Pol A se-
fquence 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 cya-
nophages is provided by the gene terL, which encodes the termi-
nase 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 freshwa-
ter heterocystous cyanobacterium Nostoc sp. strain PCC 7524 and
the marine tropical and subtropical unicellular N fixer Crocospha-
era 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
genesis encoding a phage tail collar (NCBI reference sequence ac-
cession no. EAM53192), a transposase (NCBI reference sequence ac-
cession no. EAM53191), and a hypothetical protein (NCBI re-
ference sequence accession no. EAM53190) that also show similar-
ity to putative genes in the Nostoc cyanophages. The terminase
large subunit in Nostoc sp. strain PCC 7524 is also part of a
phage-like element (see below). The A-1 and N-1 terminase
sequences have revealed a new evolutionary group of phage ter-
minases with similarity to phage elements in several genera of
divergent cyanobacteria, suggesting that relatives of A-1and 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 pyrimi-
dine 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 supple-
mental material). Thymidylate synthase and thymidylate kinase
genies in A-1and N-1 were also similar to those found in cyano-
bacteria. The proteins encoded by these genes likely catalyze two
subsequent steps in dTTP synthesis. Thymidylate synthase is in-
volved 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 deoxy-
ribonucleotides. The gene encoding thymidylate kinase is com-
monly 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
coding DNA adenine methyltransferases and DNA cytosine
methyltransferases were also found in the Nostoc cyanophages. In
general, DNA methyltransferases mediate postreplicative methyl-
ation 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. How-
ever, DNA methyltransferase 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 cya-
nophage 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 3](http://mbio.asm.org/Downloadedfromhttp://mbio.asm.org/)

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 cyanomyo-
viruses; dark blue, T4-like myoviruses; light green, Nostoc phages; dark green, cyanobacterial plasmid. The scale bar shows the number of amino acid substi-
tutions per site.
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

---

**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.
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 systems (52).

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

<table>
<thead>
<tr>
<th>ORF length (bp)</th>
<th>Strand</th>
<th>Significant hit</th>
<th>Organism (accession no.)</th>
<th>E value</th>
<th>% Identity (no. of shared aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Terminase large subunit</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF489941.1)</td>
<td>2.28e-54</td>
<td>30 (129)</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF489951.1)</td>
<td>7.16e-37</td>
<td>21 (113)</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Tail sheath protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490006.1)</td>
<td>1.14e-09</td>
<td>38 (137)</td>
</tr>
<tr>
<td>16</td>
<td>R</td>
<td>Lysozyme-like domain, rare lipoprotein A</td>
<td><em>Anaerobacter variabilis</em> ATCC 29413</td>
<td>2.3e-19</td>
<td>57.7 (60)</td>
</tr>
<tr>
<td>738</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490101.1)</td>
<td>9.39e-11</td>
<td>29 (67)</td>
</tr>
<tr>
<td>23</td>
<td>R</td>
<td>Exonuclease RNase T and DNA polymerase</td>
<td><em>Cyanothecales pachydiscus</em> PCC 10605</td>
<td>1.63e-08</td>
<td>27 (48)</td>
</tr>
<tr>
<td>26</td>
<td>F</td>
<td>Phage-related baseplate assembly protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490151.1)</td>
<td>1.46e-32</td>
<td>34 (91)</td>
</tr>
<tr>
<td>32</td>
<td>F</td>
<td>Baseplate I phage tail</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490181.1)</td>
<td>3.76e-33</td>
<td>41 (94)</td>
</tr>
<tr>
<td>35</td>
<td>F</td>
<td>Phage tail protein (tail_P2_I)</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490201.1)</td>
<td>5.63e-40</td>
<td>59 (77)</td>
</tr>
<tr>
<td>37</td>
<td>F</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490211.1)</td>
<td>4.83e-47</td>
<td>37 (157)</td>
</tr>
<tr>
<td>48</td>
<td>F</td>
<td>Tail collar protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490221.1)</td>
<td>2.1e-52</td>
<td>50 (149)</td>
</tr>
<tr>
<td>92</td>
<td>R</td>
<td>Thymidylate synthase complementing protein</td>
<td><em>Chlorobium phaeobacteroides</em> BSI</td>
<td>4.38e-39</td>
<td>38.4 (86)</td>
</tr>
</tbody>
</table>

a F, forward; R, reverse.

<table>
<thead>
<tr>
<th>ORF length (bp)</th>
<th>Strand</th>
<th>Significant hit</th>
<th>Organism (accession no.)</th>
<th>E value</th>
<th>% Identity (no. of shared aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>F</td>
<td>Phage terminase, large subunit</td>
<td><em>Nostoc</em> sp. strain PCC 7524</td>
<td>2.32e-35</td>
<td>31 (134)</td>
</tr>
<tr>
<td>2</td>
<td>F</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF489951)</td>
<td>3.55e-18</td>
<td>22 (106)</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490011)</td>
<td>9.62e-06</td>
<td>30 (30)</td>
</tr>
<tr>
<td>9</td>
<td>F</td>
<td>Tail sheath protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490061)</td>
<td>3.74e-74</td>
<td>39 (138)</td>
</tr>
<tr>
<td>15</td>
<td>R</td>
<td>Lysozyme-like domain, rare lipoprotein A (RlpA)-like double psi beta barrel</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490141)</td>
<td>1.48e-37</td>
<td>28 (137)</td>
</tr>
<tr>
<td>18</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490101)</td>
<td>9.82e-10</td>
<td>27.9 (51)</td>
</tr>
<tr>
<td>19</td>
<td>F</td>
<td>Exonuclease RNase T and DNA Pol III</td>
<td><em>Thauera</em> sp. strain M2ZIT</td>
<td>4.14e-04</td>
<td>28.4 (38)</td>
</tr>
<tr>
<td>29</td>
<td>F</td>
<td>Lysozyme</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490171)</td>
<td>2.51e-05</td>
<td>24 (32)</td>
</tr>
<tr>
<td>30</td>
<td>F</td>
<td>Baseplate I phage tail protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490181)</td>
<td>4.18e-48</td>
<td>43 (144)</td>
</tr>
<tr>
<td>31</td>
<td>F</td>
<td>Phage tail protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490201)</td>
<td>2.32e-19</td>
<td>26.3 (45)</td>
</tr>
<tr>
<td>32</td>
<td>F</td>
<td>Phage tail fiber protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490211)</td>
<td>2.97e-44</td>
<td>37 (172)</td>
</tr>
<tr>
<td>33</td>
<td>F</td>
<td>Tail collar protein</td>
<td><em>Nostoc</em> sp. strain PCC 7524 (AF490221)</td>
<td>1.62e-47</td>
<td>33 (137)</td>
</tr>
<tr>
<td>34</td>
<td>R</td>
<td>DNA Pol B</td>
<td><em>Cyanothecales pachydiscus</em> PCC 7524</td>
<td>7.6e-18</td>
<td>41.5 (243)</td>
</tr>
<tr>
<td>48</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Nostoc</em> punctiforme PCC 731021</td>
<td>1.10e-06</td>
<td>21.4 (40)</td>
</tr>
<tr>
<td>52</td>
<td>R</td>
<td>C-5 cytosine-specific DNA methylase</td>
<td><em>Nostoc</em> punctiforme PCC 73102</td>
<td>1.81e-05</td>
<td>36.2 (34)</td>
</tr>
<tr>
<td>62</td>
<td>F</td>
<td>ASCH domain protein</td>
<td><em>Clostridium</em> phage mMPM2</td>
<td>4.06e-12</td>
<td>42.3 (33)</td>
</tr>
<tr>
<td>68</td>
<td>F</td>
<td>Thymidylate kinase</td>
<td><em>Lyngbya</em> sp. strain PCC 8016</td>
<td>2.45e-23</td>
<td>36 (74)</td>
</tr>
<tr>
<td>75</td>
<td>R</td>
<td>DNA-cytosine methyltransferase</td>
<td><em>Anaerobacter variabilis</em> ATCC 29413</td>
<td>1.02e-36</td>
<td>33.2 (127)</td>
</tr>
<tr>
<td>75</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Thermoaerobacterium</em> Ab9</td>
<td>8.67e-03</td>
<td>32.7 (33)</td>
</tr>
<tr>
<td>77</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Calothrix</em> sp. strain PCC 7103</td>
<td>1.73e-47</td>
<td>56 (94)</td>
</tr>
<tr>
<td>77</td>
<td>R</td>
<td>Hypothetical protein</td>
<td><em>Calothrix</em> sp. strain PCC 7103</td>
<td>4.94e-06</td>
<td>30.4 (78)</td>
</tr>
<tr>
<td>78</td>
<td>F</td>
<td>dCTP deaminase</td>
<td><em>Synochecoccus</em> sp. strain PCC 7335</td>
<td>1.45e-49</td>
<td>65.5 (131)</td>
</tr>
</tbody>
</table>

a F, forward; R, reverse.

b aa, amino acids.
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-

---

**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).
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 be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58). Viral particles likely cannot be exchanged through nonspecific junctions that connect the cytoplasm of adjacent cells (58).

Cyanobacterial strain (CRISPR ID) | E value | Start | End | DR consensus sequence
--- | --- | --- | --- | ---
*A. variabilis* sp. strain PCC 7113 (CRISPR6) | 3.0e⁻⁰⁶ | 2087079 | 2087333 | GTCCTGAATTCATATACCTATATCCCTATCCAGGGGTGAAAC
*Microcoleus* sp. strain PCC 7113 (CRISPR11) | 1.0e⁻⁰⁷ | 3131505 | 3132053 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Microcoleus* sp. strain PCC 7113 (CRISPR15) | 1.0e⁻⁰⁷ | 3859458 | 3860565 | GTTTCAATTCCTGATAGGGATTTTTGTTAGTTAAGAAC
*Nostoc* sp. strain PCC 7107 (CRISPR27) | 1.0e⁻⁰⁷ | 5817107 | 5818597 | GTTGCAATTCCTATATCCCTATCCAGGGGTGAAAC
A. variabilis ATCC 29413 (CRISPR3) | 8.0e⁻⁰⁷ | 1234556 | 1237182 | GTTTATAATACAAAAATCCCTATCCAGGGGTGAAAC
*Nostoc* sp. strain PCC 7210 (CRISPR13) | 3.0e⁻⁰⁷ | 3516819 | 3517367 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Nostoc* sp. strain PCC 7210 (CRISPR14) | 3.0e⁻⁰⁷ | 3517542 | 3518084 | GTTTCAATTCCTGATAGGGATTTTTGTTAGTTAAGAAC
*A. cylindrica* sp. strain PCC 7210 (CRISPR8) | 3.0e⁻⁰⁷ | 5001036 | 5003639 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Calothrix* sp. strain PCC 74714 (CRISPR20) | 3.0e⁻⁰⁷ | 1101502 | 1103134 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Calothrix* sp. strain PCC 6303 (CRISPR6) | 4.0e⁻⁰⁷ | 2021864 | 2022765 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Calothrix* sp. strain PCC 7507 (CRISPR24) | 4.0e⁻⁰⁷ | 2038058 | 2039287 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Calothrix* sp. strain PCC 7507 (CRISPR14) | 4.0e⁻⁰⁷ | 5067421 | 5070978 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*N. punctiforme* sp. strain PCC 73102 (CRISPR16) | 3.0e⁻⁰⁷ | 3338172 | 3341197 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Nostoc* sp. strain PCC 7210 (CRISPR8) | 3.0e⁻⁰⁷ | 1836813 | 1837723 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Nostoc* sp. strain PCC 7210 (CRISPR3) | 3.0e⁻⁰⁷ | 8075442 | 8075588 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*A. variabilis* ATCC 29413 (CRISPR19) | 3.0e⁻⁰⁷ | 5764010 | 5766428 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*Nostoc* sp. strain PCC 7210 (CRISPR20) | 3.0e⁻⁰⁷ | 5654133 | 5654384 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*A. variabilis* ATCC 29413 (CRISPR9) | 3.0e⁻⁰⁷ | 2395670 | 2398703 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*A. variabilis* ATCC 29413 (CRISPR18) | 1.0e⁻⁰⁶ | 5227213 | 5229822 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
Cyanobacteria (CRISPR16) | 4.0e⁻⁰⁷ | 3575124 | 3575742 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC
*A. variabilis* ATCC 29413 (CRISPR16) | 4.0e⁻⁰⁷ | 4821250 | 4823752 | GTTTAATATCCCTATATCCCTATCCAGGGGTGAAAC

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

**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 tail lysed 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 m⁻² 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, so-dium chloride was added to the lysate at a final concentration of 0.5 M and the cultures were incubated for 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 × 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 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 1 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 Génomique Québec 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^-5. 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 (Δ = 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 Pro-mals 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^-5.

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-μm-filtered N-1 lysate was added to a 48-μl PCR mixture containing Platinum Taq DNA polymerase assay buffer (50 mM KCl, 20 mM Tris-HCl, pH 8.4), 10 mM MgCl2, 200 μM deoxynucleoside triphosphate, 0.25 μM each primer (sCRF [CA ATTTGCAAAAGATTTAGCAGC] and CR3 R [GGGGAGAGTTTGGAAGGGGTT]), 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 performed in a 10-μl reaction mix containing 10 ng of a cDNA template, 1× LongAmp buffer (Invitrogen), 0.25 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 3 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 nos. KU234533 and KU234532.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org.

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 Québec—Nature et Technologies, a UBC University graduate fellowship, and a BRITE award from the NSERC CREATE Program.
REFERENCES


