Chapter 1

General introduction

Toxic cyanobacterial blooms have become a frequent occurrence in eutrophic freshwater environment throughout the world (1). Recent increase of the frequency, intensity, and duration of cyanobacterial blooms are likely to be related to eutrophication (2), the elevated CO_2 levels (3, 4), and global warming (5, 6). The bloom-forming cyanobacterial genera include *Aphanizomenon*, *Cylindrospermopsis*, *Dolichospermum*, *Microcystis*, *Nodularia*, *Planktothrix* and *Trichodesmium* (7). Of these, *Microcystis aeruginosa* is one of the most pervasive and notorious bloom-forming cyanobacterium (1).

M. aeruginosa spends winter period in the bottoms and rises to the water surface where it can accumulate to form blooms and scums in the summer in the temperate region (8). This cyanobacterium possesses gas vesicles which are hollow protein structures filled with gas, providing buoyancy to the cells (9, 10). The vertical migration by their buoyancy is important for accessing nutrient and optimizing the utilization of light energy (1). A part of *Microcystis* strains can produce the potent hepatotoxin microcystin which is originally identified as Fast-Death Factor (1). The cyclic heptapeptides, microcystins are comprised of several unusual amino acids including 3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4(E),6(E)-dienoic acid (ADDA) in position five, two conventional D-amino acids in positions one and six, D-erythro-β-methylaspartic acid in position three, and N-methyldehydroalanine in position seven (11, 12). To date, over 100 different congeners of microcystin showing various toxicities (from non-toxic; LD_{50} > 1200 μg/kg to highly toxic; LD_{50} > 50 μg/kg) have been characterized, mostly due to differences in L-amino acids in positions two and four (13, 14). Such microcystins form an irreversible

covalent bond with protein phosphatase (15, 16), especially in hepatocytes, and thereby can lead to subsequent cell structure damage, liver disease and nephrotoxicity (17–19). Indeed, numerous fatalities and severe poisonings of livestock, pets and wild life caused by microcystin-containing *Microcystis* blooms have been reported (20). Likewise, microcystin toxicity poses serious problems for human who use impaired water resources for drinking water supplies (21), recreational activities (22), fisheries (23), and dialysis treatment (24). These accidents have led to the World Health Organization (WHO) to propose a drinking water guideline of 1 μg/L for microcystins (25). Thus, it is global issues for maintaining safe water supplies to control the toxic *Microcystis* blooms.

Traditionally, chemical and physical environmental factors have been the focus for controlling *Microcystis* blooms as well as other cyanobacterial blooms (1). Unlike to the other major bloom-forming cyanobacteria except for *Planktothrix*, *M. aeruginosa* is incapable of supplying nitrogen requirements by N_2 fixation during nutrient limitation (1). This physiological feature has led to the hypothesis that the inflow of nitrogen into the freshwater environments plays a key role in the growth and proliferation of *M. aeruginosa* (26) as well as phosphorus. Indeed, it has been shown that the world-wide *Microcystis* proliferation appears to link to the increase in both phosphorus and nitrogen inflow generated by various human activities (26–30). Especially, the low ratio of nitrogen to phosphorus is thought to favor *Microcystis* blooms (and also increase the microcystin concentration (31)) as well as other cyanobacteria (32). However, the universal optimum ratio is not determined and some contradictory results have been reported (33–36). Therefore, the effects of the ratio of nitrogen and phosphorus for *Microcystis* blooms are still under debate. Physical factors such as irradiance, temperature, turbulence, vertical mixing and hydrologic flushing have been also only implicated as the potential control of

Microcystis blooms under certain circumstances (1, 29, 37). Thus, the impacts of chemical and physical environmental factors on *Microcystis* bloom formation and termination have not been fully understood. Furthermore, grazing by zooplankton or bivalves on *M. aeruginosa* have been studied as a biological environmental factor. Although grazers generally exert top-down ecological controls in aquatic ecosystems (38, 39), *Microcystis* colony formation can function as grazing deterrents (40). Indeed, *Microcystis* large colonies are poorly grazed, particularly by small crustacean zooplankton (40). In addition, it was reported that a certain *Microcystis* strain respond to small flagellated zooplankton that could not consume the colonies, and thereby transformed from unicellular to colonial morphology (41). Thus, *M. aeruginosa* are thought to be an inadequate food source for zooplankton (40).

Viruses infecting microorganisms are ubiquitous and the highly abundant in aquatic ecosystems (42). Such viruses inject their genome into host cells, redirect host metabolisms for their reproduction, and finally lysed host cells to release their progeny into the environments (43). Therefore, viral infections are thought to affect host mortality, the composition of microorganism communities, and the biogeochemical cycles (42, 44, 45). It was reported that a lytic agent that formed plaques on *Microcystis* lawns (46) or virus-like particles that inhibited *Microcystis* growth about two decades ago (47). In addition, an increase of cyanovirus titers accompanied by a large decrease in the *Microcystis* abundance was also observed in the natural environment (48). These findings suggest that viruses infecting this cyanobacterium could be also one of the most important factor to affect its bloom dynamics and termination in the environment. Under such circumstances, Yoshida *et al*. (2006) successfully isolated cyanovirus Ma-LMM01 that specifically infects a toxic bloom-forming cyanobacterium *M. aeruginosa*. Ma-LMM01

is a new lineage of the *Myoviridae* family that has an icosahedral head (86 nm in diameter) and a tail complex consisting of a central tube (9 nm in width) and a sheath (24 nm in width, 209 nm in length) that contracted to 90 nm in length (49, 50). The latent period and burst size of Ma-LMM01 were estimated to be 6-12 h and 50 to 120 infectious units/cell, respectively (50). To date, Takashima *et al*. (2007) developed a real-time PCR method for quantification of the abundance of Ma-LMM01, and thereby revealed that the relatively high ratio *M. aeruginosa* to Ma-LMM01 numbers in the summer (52). Comparing the expression levels of *gp091* (tail sheath gene) in the environment what in the culture experiment, furthermore, Ma-LMM01 infection likely to occur in 0.01-2.9 cell/mL of the natural *Microcystis* cell population (53). These findings indicate that Ma-LMM01 might have the ability only to infect a small percentage of *Microcystis* population in the natural environment (52, 53). Meanwhile, several studies have revealed that Ma-LMM01 exists in *Microcystis* blooms throughout the world (54–56). In accordance with these, cyanovirus MaMV-DC, which infects *M. aeruginosa* and shows high average nucleotide identity (86.1%) with Ma-LMM01, was also isolated from Lake Dianchi, China (57, 58).

Current population studies of *M. aeruginosa* have also implicated the interactions between this cyanobacterium and its infecting viruses in the environment. CRISPR-Cas systems, which is one of antiviral defense systems and composed of short direct repeats separated by unique sequences (spacers), incorporate foreign DNA fragments such as viruses into the leader-end of the CRISPR loci as a new spacer that provides a sequence memory of the invasion of exogenous genetic elements like viruses (59). Based on such spacer arrangement (CRISPR array), *Microcystis* bloom is consisted of the diverse populations that possess different CRISPR arrays (60–63). These

observations indicate that the diverse combinations of *M. aeruginosa* and its viruses exist in the natural environments. Coincide with this, *M. aeruginosa* has the highest number of putative antiviral defense genes (corresponding to 29% protein-coding genes) among all prokaryote or archeal species as of 2011 (64). Such genomic features of *M. aeruginosa* also suggest that infection profiles of *Microcystis* viruses will provide better understanding of the host-virus interactions as well as their impact on *Microcystis* blooms (60).

Despite the potential importance of cyanoviruses in *Microcystis* blooms, little is currently known about whole host transcriptional responses to viral infection, and the infection program of even the only isolated strain Ma-LMM01 to escape the highly abundant host defense systems. Especially, Ma-LMM01 lacks almost all of the T4 core genes needed for appropriating host metabolic machinery, replicating the viral genome, and building viral particles and contains none of the auxiliary metabolic genes except for *nblA* usually carried by marine T4-like cyanoviruses (49). These genomic features indicate that Ma-LMM01 may employ an infection program which is different from that of other marine cyanoviruses. Considering that Ma-LMM01-matching spacers are present in very low in natural *Microcystis* populations (10/995 spacers) (61), furthermore, numerous uncharacterized cyanoviruses exist that can affect *Microcystis* bloom dynamics and termination process. To date, however, no comprehensive studies have been done to investigate for the existence of other *Microcystis* viruses, or whole transcriptional dynamics of both *M. aeruginosa* and its viruses in the environment. In my doctoral thesis, to reveal the infection processes of *Microcystis* viruses coping with the highly abundant host antiviral defense system, I first investigated Ma-LMM01 infection profiles in the culture experiment using transcriptome analysis (in Chapter 2). Then, I revealed

the novel *Microcystis* viruses and their transcriptional patterns in the environment using cross-omics analysis (in Chapter 3).

Chapter 2

Transcriptome analysis of a bloom-forming cyanobacterium *Microcystis aeruginosa* during Ma-LMM01 phage infection

Abstract

Microcystis aeruginosa forms massive blooms in eutrophic freshwaters, where it is constantly exposed to lytic cyanophages. Unlike other marine cyanobacteria, *M. aeruginosa* possess remarkably abundant and diverse potential antiviral defense genes. Interestingly, T4-like cyanophage Ma-LMM01, which is the sole cultured lytic cyanophage infecting *M. aeruginosa*, lacks the host-derived genes involved in maintaining host photosynthesis and directing host metabolism that are abundant in other marine cyanophages. Based on genomic comparisons with closely related cyanobacteria and their phages, Ma-LMM01 is predicted to employ a novel infection program that differs from that of other marine cyanophages. Here, I used RNA-seq technology and *in silico* analysis to examine transcriptional dynamics during Ma-LMM01 infection to reveal host transcriptional responses to phage infection, and to elucidate the infection program used by Ma-LMM01 to avoid the highly abundant host defense systems. Phage-derived reads increased only slightly at 1 h post-infection, but significantly increased from 16% of total cellular reads at 3 h post-infection to 33% of all reads by 6 h post-infection. Strikingly, almost none of the host genes (0.17%) showed a significant change in expression during infection. However, like other lytic dsDNA phages, including marine cyanophages, phage gene dynamics revealed three expression classes: early (host-takeover), middle (replication), and late (virion morphogenesis). The early genes were concentrated in a single ~5.8-kb window spanning 10 open reading frames (gp054–

gp063) on the phage genome. None of the early genes showed homology to the early genes of other T4-like phages, including known marine cyanophages. Bacterial RNA polymerase (σ^{70}) recognition sequences were also found in the upstream region of middle and late genes, whereas phage-specific motifs were not found. These findings suggest that unlike other known T4-like phages, Ma-LMM01 achieves three sequential gene expression patterns with no change in host promoter activity. This type of infection that does not cause significant change in host transcriptional levels may be advantageous in allowing Ma-LMM01 to escape host defense systems while maintaining host photosynthesis.

Introduction

Viruses are extremely abundant in aquatic environments, with global estimates reaching 10^{30} virus-like particles (42). Viruses are thought to play important roles in regulating the abundance, clonal diversity, and composition of bacterial populations (45), and thus have the potential to affect biogeochemical cycles through the process of host cell lysis (42, 45). Therefore, it is essential to elucidate viral infection mechanisms to better understand the impact of viruses on host populations and biogeochemical cycles.

In general, infection dynamics of T4-like phages show that following infection, host genomic DNA is degraded and there is an almost complete shift to phage transcription, leading to the shutdown of host metabolism (43, 65). The phage transcriptional program generally follows the three temporal expression classes of early, middle, and late genes, which correspond to host takeover, replication, and virion morphogenesis, respectively (43, 66). In T4 phage, this expression program is regulated by the sequential modification of the host RNA polymerase and associated σ factor,

leading to consecutive changes in affinity for different promoter sequences. The expression of early genes relies on the primary host σ^{70} factor which recognizes early T4 promoters that resemble the major *Escherichia coli* promoters and is stronger than any bacterial promoters (65). The internal head protein Alt increases affinity for the early T4 promoters and supports preferential transcription from early T4 promoter by ADP-ribosylation of one of the two α subunits of the host RNA polymerase (67, 68). Middle-gene promoters have a distinctive motif sequence that is again recognized by the host σ^{70} factor, aided by phage-encoded proteins AsiA and MotA. Anti-σ factor AsiA forms the heterodimers with host σ^{70} factor, and activates the transcription from middle T4 promoters. Transcriptional activator MotA binds to the MotA box sequence and recruits the host RNA polymerase to middle T4 promoters (69, 70). Phage-encoded proteins endoribonuclease RegB (71) and ADP-ribosyltransferase ModA (72) also contribute to switch from the early transcription to middle transcription. In contrast to transcription from early and middle-gene promoters, recognition of late-gene promoters requires a phage-encoded σ factor, gp55. Co-activator gp33 and DNA-loaded sliding clamp gp45 also involve in efficient transcription from late-gene promoters (43).

T4-like cyanophages infecting marine cyanobacterial genera *Synechococcus* and *Prochlorococcus* contain homologs of the T4 replication and virion structural genes that are shared among T4-like phages (T4 core genes) (73). According to their genomic features, transcriptome analyses for marine T4-like cyanophages clearly indicate the three temporal classes of early, middle, and late genes as seen in T4 phage (74–77). In addition, marine T4-like cyanophages possess a number of auxiliary metabolic genes (AMGs) that are derived from hosts and are involved in processes such as photosynthesis, carbon metabolism, and phosphorus utilization (78, 79). Such AMGs are thought to provide

support during key steps in host metabolism that are relevant to phage, thereby boosting and redirecting host metabolism after the shutoff of host metabolism caused by phage infection (77, 79). Indeed, pentose phosphate pathway are augmented to provide a more direct mechanism for NADPH production, where NADPH is not destined for carbon fixation, but rather both ATP and NADPH are available for nucleotide synthesis for viral genome replication during cyanophage P-HM2 infection (77). In this way, T4-like cyanophages maintain host photosynthesis activity and redirect carbon flux from the Calvin cycle to the pentose phosphate pathway, although they lack T4-like middle-gene promoters and the *motA* and/or *asiA* genes (75, 77).

Toxic bloom-forming cyanobacterium *Microcystis aeruginosa*, along with its phages, provides an excellent model to study the co-evolution of viruses and their hosts (60) because it contains the largest number of defense genes out of all studied bacteria and archaea (64), and is frequently exposed to phage infection (61, 80). Cyanophage Ma-LMM01, which is known to only infect *M. aeruginosa* strain NIES-298 among tested strains, is a member of the *Myoviridae* family (50) and is phylogenetically distinct from other known marine T4-like cyanophages (49). Coincidentally, Ma-LMM01 lacks almost all of the T4 core genes involved in appropriating host metabolic machinery, replicating the viral genome during infection, and building viral particles (49). In particular, it does not contain any homologs of phage-encoded σ factor gp55 and transcription factor gp33, both of which are required for late-gene transcription in T4-like phage (Roucourt and Lavigne, 2009). In addition, Ma-LMM01 contains none of the AMGs usually carried by marine cyanophages. These findings indicate that Ma-LMM01 may employ an infection program that differs from that of other marine cyanophages. Ma-LMM01 does possess a homolog of *nblA*, which plays a central role in the degradation of phycobilisomes (49).

The phage-encoded NblA is predicted to be involved in maintaining host photosynthesis (49, 81, 82). Furthermore, Honda *et al*. (2014) previously reported that there was no difference in the level of host *psbA* transcription during Ma-LMM01 infection and that levels of genes involved in the Calvin cycle and pentose phosphate pathway also did not change, or were slightly decreased. Together, these findings suggested that Ma-LMM01 maintains host photosynthesis activity and carbon metabolism by protecting photosystem II using phage-encoded *nblA*. However, little is currently known about whole host transcriptional responses to phage infection, and the infection program employed by Ma-LMM01 to avoid the highly abundant host defense systems during infection.

In this study, I investigated the infection process and transcriptional program of Ma-LMM01 during infection of its sole host, *M. aeruginosa* NIES-298, and assessed host transcriptional responses to infection using RNA sequencing (RNA-seq) analysis.

Material and Methods

Bacterial and phage culture conditions and experimental design

M. aeruginosa NIES-298 was obtained from the Natural Institute for Environmental Studies (NIES, Japan; http://www.nies.go.jp). *M. aeruginosa* was cultured in CB medium (83) under a 12/12-h light/dark photocycle (light intensity: 21 μmol photons/m^2/s) at 30°C with 0.5% CO_2 (v/v) aeration.

To prepare the phage lysate, a 1-L *M. aeruginosa* culture (9.14×10^5 cells/mL) was infected with Ma-LMM01 at a multiplicity of infection of 0.02 and then incubated as above for 3 days. The resultant lysate was filtered through a sterile 3.0-μm polycarbonate membrane filter (Millipore, Billerica, MA, USA) and stored as an original lysate at 4°C (84).

2. Transcriptome analysis during Ma-LMM01 infection

A 9-L volume of *M. aeruginosa* culture in exponential-phase was prepared as described above and then divided between six different flasks containing 1.5 L of medium. I performed the one-step growth experiment for Ma-LMM01 to obtain simultaneously infected cells without multiple infection for RNA-seq analysis according to the previous study (50). Using this infection experiment, at least two different temporal classes (early and late genes) have been observed in expression of phage genes (81). In brief, cell division is well synchronized compared with untreated cells when *Microcystis* cells arrested once by 36 h darkness are transferred to continuous illumination (the block-released method) (81, 85). Thereby the variation of infection stage was minimized at each time point. For the phage infected cultures, 250 mL of original phage lysate were added to each of three synchronized cultures after the light was turned on. In parallel with infection experiment, infectious phage concentration was determined using the most probable number (MPN) method (3.59×10^6 infectious units/mL) (50, 84), resulting in a multiplicity of infection (MOI) of 0.62–0.89. From the wide range of upper value and lower value within the confidence limits (95%) in MPN method (86), the MOIs varied. However, almost complete lysis was observed at 24 h after infection. In addition, growth experiments of Ma-LMM01 should be carried out with MOIs less than 1 because an MOIs greater than 2 results in a small decrease in the number of host cells (50). For the control cultures, an equivalent volume of CB medium was added to a further three flasks in place of the phage lysate. To determine the number of phage particles and host cells, samples were collected from the flasks at different time points during the lytic cycle (0, 0.5, 1, 2, 3, 4, 5, 6, 8, 10, 12, and 24 h after phage addition). To enumerate phage particles, samples were passed through a 3.0-μm PTFE membrane filter and then immediately fixed in 20% glutaraldehyde at a final concentration of 1% and stored at 4°C

until analysis. To enumerate host cells, cells were immediately fixed in 20% glutaraldehyde at a final concentration of 1% and stored at 4°C until analysis. Densities of the host cells and phage particles were measured using epifluorescence microscopy (Nicon ECLIPSE E800; Nicon, Tokyo, Japan) with SYBR Gold staining (Molecular Probes, Eugene, OR, USA). As described in the previous studies, the estimated latent period of Ma-LMM01 is 6-12 h (50) and host transcriptional profiles do not show remarkable change between 6 h and 8 h post-infection (81). For RNA extraction, therefore, 100-mL aliquots of the infected and control cultures were collected at 0, 1, 3, and 6 h post-infection, and cells were collected on 3.0-μm PTFE membrane filters. The cells were then resuspended in 5 mL of stop solution (phenol:ethanol, 5:95 v/v) and stored at −80°C (85). At each time point, these procedures were complete within 20 min (87).

Sequencing and analysis of the *M. aeruginosa* NIES-298 genome

Genomic DNA extraction from *M. aeruginosa* cells was performed using a combination of the potassium xanthogenate-sodium dodecyl sulfate and phenol/chloroform/isoamyl alcohol procedures, as described previously (88, 89). The extracted DNA was sheared using a Covaris M220 focused-ultrasonicator (Covaris, Woburn, MA, USA) to an average size of 300 bp. A mate pair library was then prepared using a Nextera Mate Pair Library Prep Kit (Illumina, San Diego, CA, USA) according to the manufacturer's instructions. The mate pair library was sequenced using a MiSeq Reagent Kit v3 (2 × 150-bp read length; Illumina) and the Illumina MiSeq platform, and assembled using SPAdes ver.3.7.0 (90). Open reading frames (ORFs) were predicted using GenemarkS (91), and predicted ORFs were annotated by blastp analysis against the National Center for Biotechnology Information (NCBI) non-redundant database (nr) (E-value thresholds of $< 1e^{-3}$).

RNA-seq library preparation for Illumina sequencing

Total RNA was extracted from 2 mL of the stored cell suspension as described previously (85). The total RNA concentration was measured using a Qubit Fluorometer (Life Technologies, Paisley, UK) according to the manufacturer's instructions. RNA integrity was also verified by gel electrophoresis. DNA was removed using TURBO DNase (Ambion, Austin, TX, USA). Genomic DNA depletion was checked to eliminate the effects of DNA contamination on the following RNA-seq analysis using RT-PCR assay and gel electrophoresis with DNA-depleted RNA samples as non-reverse transcribed control (data not shown). For depletion of ribosomal RNA, a Ribo-Zero rRNA removal kit (Bacteria) (Epicentre, Madison, WI, USA) was used according to the kit instructions, and rRNA depletion was verified by Agilent 2100 bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). The rRNA-depleted RNA was then purified using Agencourt RNAClean XP beads (Beckman Coulter Genomics, Danvers, MA, USA) according to the manufacturer's instructions. The purified RNA was converted to double stranded cDNA using a PrimeScript Double Stranded cDNA Synthesis Kit (TaKaRa Bio, Otsu, Japan), and cDNA libraries (not strand-specific) were prepared using a Nextera XT DNA sample preparation Kit (Illumina). RNA-seq libraries were sequenced using a MiSeq Reagent Kit v3 (2 × 75 bp read length; Illumina) and the Illumina MiSeq platform.

Mapping and counting RNA-seq reads

Reads from each library were aligned separately to the merged reference genome (Ma-LMM01 plus *M. aeruginosa* NIES-298) using bowtie2 (92) with option "--score-min L,0,-0.6". Host 16S and 23S rRNA reads were removed manually from the total reads prior to read mapping. An average of 1 million reads were recovered from each cDNA library at 0, 1, 3, and 6 h post-infection (**Table 2-1**). Rarefaction curves and chao1 indices

for each host and viral reads were separately constructed using PAST ver.3.17 (93). Reads from the whole transcriptome library were counted for each gene. Host and viral transcript counts were each normalized as FPKM (fragments per kilobases of exon per million mapped reads). The reads mapping to viral genome at each time point were visualized independently with Integrative Genomics Viewer (94).

Table 2-1. Summary of sequencing data generated in this study.

(A) In control samples

	0h	1h	3h	6h
paired reads	1,524,163	1,901,677	1,501,410	1,572,588
Q30 filtered	1,446,831	1,788,926	1,428,977	1,469,486
rRNA removed	1,446,826	1,788,922	1,428,972	1,469,481
Mapped reads	1,010,716	1,342,197	1,000,001	949,935

(B) In infected samples

	0h	1h	3h	6h
paired reads	1,747,642	1,468,898	1,399,842	1,615,758
Q30 filtered	1,647,597	1,361,217	1,336,867	1,515,406
rRNA removed	1,647,592	1,361,211	1,336,862	1,515,399
Mapped reads	1,041,952	878,383	930,510	1,109,726

Identification of differentially expressed genes

Differentially expressed host transcripts were identified using the R package DESeq (95) with blind method for estimateDispersions and DESeq2 (96) in Bioconductor (97). This analysis was conducted for the total host reads, considering the global depletion of all host transcripts relative to the total transcript population due to the influx of viral transcripts. Transcript abundances were analyzed separately at each time point, comparing the infected and uninfected treatments. An adjusted P-value (P-value with a

multiple-test correction) < 0.05 indicated a significant difference. In DESeq2, the dispersion estimation procedure replaces the different methods from the DESeq, and treats the samples as replicates for the purpose of dispersion estimation. Due to the differences of dispersion estimation procedures, DESeq analyses with blind method for estimateDispersions detected differentially expressed genes (DEGs), while DESeq2 analyses could not detect any DEGs. I also investigated whether DEGs with unknown function showed similarity to any of the defense islands in the genome of *M. aeruginosa* NIES-843 (GenBank accession no. NC_010296) using a blastn search with default parameters.

Clustering of phage gene expression

Cluster analysis of Ma-LMM01 gene expression patterns was performed for the normalized and $\log_2$-transformed transcript levels of the phage genes derived from the RNA-seq data. Hierarchical clustering was performed using Euclidean distance and average linkage metrics as implemented in the R package “stats”. The Jaccard coefficient was used to assess the quality and stability of the number of clusters obtained from hierarchical clustering (75). The Jaccard coefficient provides a measure for the similarity of two different sets of clusters, and ranges from zero (dissimilar) to one (similar). For statistical evaluation of the clustering stability, random subsets of the samples containing 70% of the genes were repeatedly (1000 replicates) selected and clustered, and then the Jaccard coefficient was calculated. This procedure was performed with varying numbers of clusters (k = 2 to 7), and the distribution of the Jaccard coefficients obtained was displayed as a histogram. The benhur function in the R/Bioconductor package “clusterStab” was used to carry out the clustering and calculation. The dendrogram of Ma-LMM01 expression clusters was plotted using the dendrogram function (hclust; R

package stats). The heat map analyses for viral gene expression profile were conducted using heatmap.2 function in the R/Bioconductor package “gplots”.

Computational identification of promoter motifs

The upstream regions of all phage genes (300 bp) were collected from the Ma-LMM01 genome. The primary sigma factor recognition sequences were predicted using BPROM (http://softberry.com) with default parameters. The promoter sequences were aligned separately for each motif (−10 box and −35 box) using ClustalW (98). Logos were prepared using weblogo (99).

Verification of RNA-seq results

RNA-seq results were verified using quantitative reverse-transcription polymerase chain reaction (qRT-PCR) analysis of host (*sigA*, *rnpB*) and phage genes (*gp005*, *gp054*, *gp062*, *gp087*, *gp091* and *gp134*) during Ma-LMM01 infection. Total RNA derived from the same samples was reverse transcribed with random hexamers using a SuperScript III First-Strand Synthesis System (Invitrogen, Carlsbad, CA, USA) as per the manufacturer’s instructions. cDNA copies were quantified using SYBR Premix Ex Taq (Tli RNaseH Plus; TaKaRa Bio) with 5 pmol of each of the forward and reverse primers as described previously (84). I designed novel primer sets for *gp054* (ATGCCGAACTAAGAAGCCCACGG and CACTTGCTTCACTCGCTGCTCG), *gp062* (GGTGAACCCATCGTGAATGTGCCA and AAGATTTGGGCAACGGCATCACC), *gp087* (GGGATCCGCTAGCGCAGCTG and AGGCGCACGCCAGAAGGAAC) and *gp134* (ATGCTCCTCCTGGTGGTC and ATAGTAATCCTCGCCGTCC). Transcript levels for each gene were normalized to host *rnpB* transcript levels (85).

Public data

The genome sequence of *M. aeruginosa* NIES-298 was deposited in the DNA Data Bank of Japan (DDBJ) Mass Submission System (MSS) under the accession numbers BEIU01000001–BEIU01000088. The mRNA expression data were deposited in the DDBJ Sequence Read Archive (DRA) under accession numbers DRR101368–DRR101375.

Results

Phage infection and transcriptome dynamics

I first investigated the infection process and transcriptome dynamics of the Ma-LMM01 phage during infection. Phage particles were released from the infected cells within 8–12 h of infection (**Figure 2-1A**), and the number of phage particles increased from 1.38×10^8 particles/mL at 0 h to 3.34×10^8 particles/mL at 24 h post-infection (**Figure 2-1A**). This infection profile was consistent with previously reported results **(**Yoshida et al., 2006). In the control culture, the *M. aeruginosa* cell density increased from 6.36×10^5 cells/mL at 0 h to 8.11×10^5 cells/mL at 24 h post-infection (**Figure 2-1A**). In contrast, in the infected culture, *Microcystis* cells were lysed at the point of phage particle release (**Figure 2-1A**), with a corresponding decrease in *M. aeruginosa* cell density from 7.78×10^5 cells/mL at 0 h to 1.61×10^5 cells/mL at 24 h post-infection (**Figure 2-1A**). Considering that Ma-LMM01 infection only occurs in a light cycle (100) and the latent period of this phage is 6-12 h (50), the decrease in host cell number between 0 h and 24 h post-infection represented lysis dynamics of infected host cells without multiple-infection. Therefore, I calculated that >79% of host cells were finally infected by Ma-LMM01 in the infection experiment. Also, Ma-LMM01 infection was thought to occur within 1 h post-infection according to transcriptome dynamics as described below.

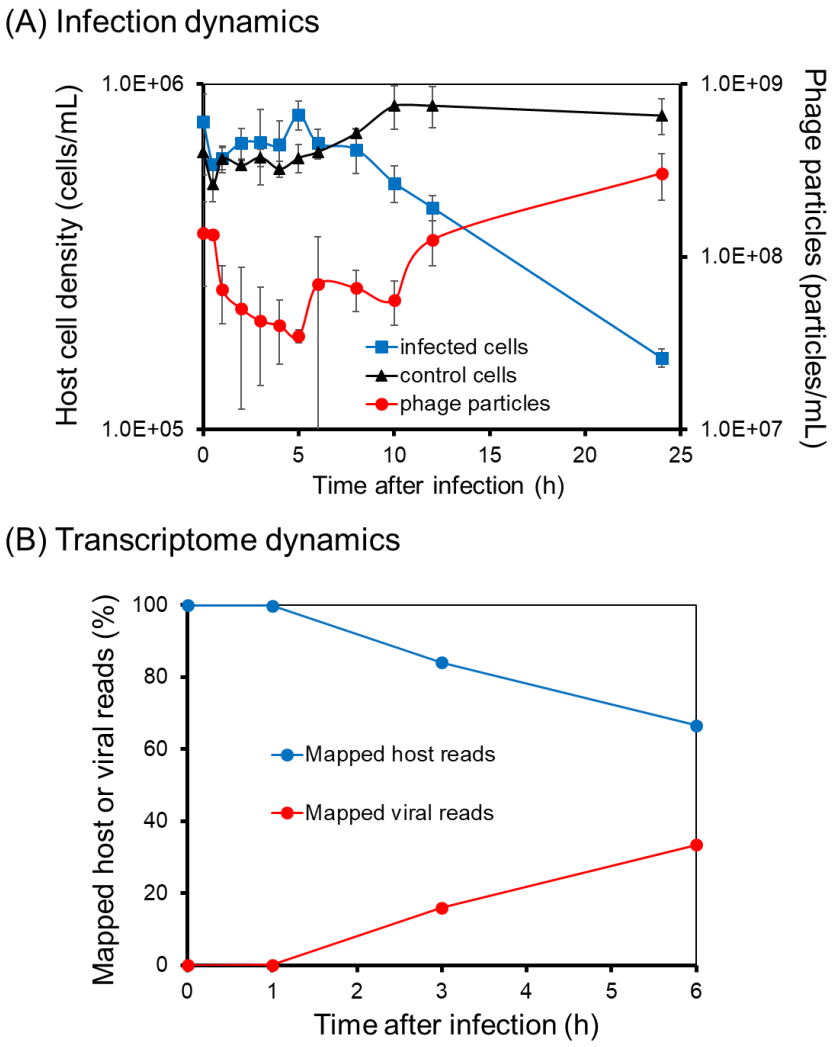


Figure 2-1. Infection dynamics and transcriptome dynamics of *Microcystis aeruginosa* NIES-298 by myovirus Ma-LMM01. Host cell density and phage particle number were determined by direct count using microscopy with SYBR Gold (A). Ratios of phage and host mRNA at different time points following infection were determined from RNA-seq reads that mapped to the phage and host genomes, respectively (B).

Rarefaction analyses for the number of host ORFs clearly demonstrated that each sequencing data was exhaustive to describe the transcriptional profile (**Figure 2-2**). Rarefaction curves generated from viral reads at 0 h and 1 h post-infection did not reach

an asymptote, and then finally reached at 3 h and 6 h post-infection with the progress of viral infection (**Figure 2-2**). The chao1 indices for each library supported these results (**Figure 2-2**). At 1 h post-infection, phage transcripts inside the infected cell accounted for 0.13% of total cellular transcription, but by 3 h and 6 h post-infection, phage transcripts constituted 16% and 33% of total cellular transcription, respectively (**Figure 2-1B**). Therefore, even at 6 h post-infection, host transcripts still accounted for 67% of cellular transcription. Also, qRT-PCR analyses normalized to host *rnpB* transcript levels (**Figure 2-3**) showed that *sigA* transcription did not change until 8 h post-infection and that *gp091* transcription increased gradually and reached peak levels within 6-8 h post-infection, indicating that the transcript profiles of host and phage genes were well represented by the RNA-seq data (see below) (**Figure 2-4**).

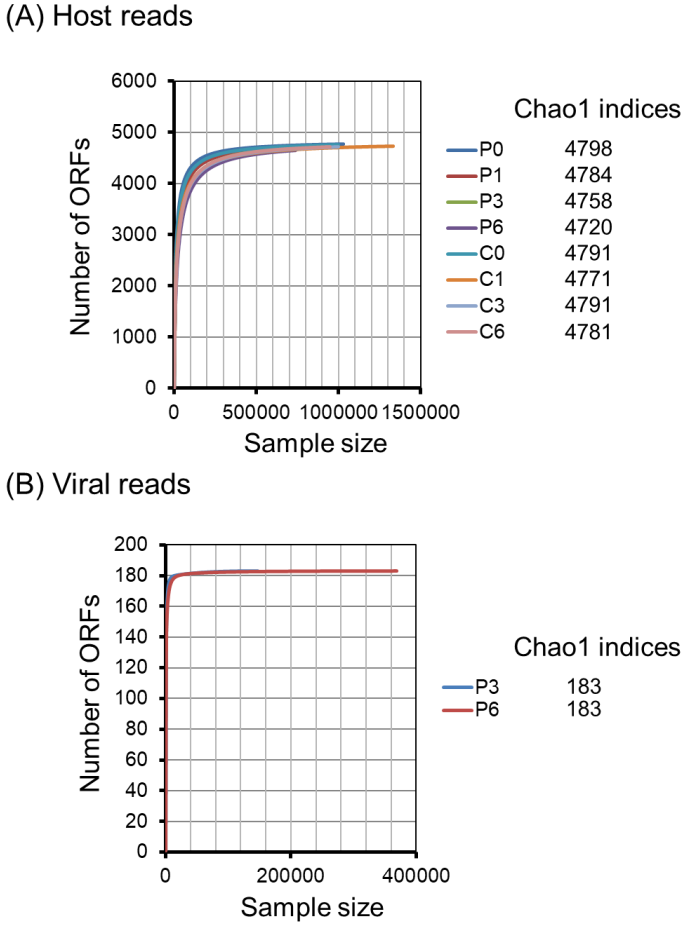


Figure 2-2. Rarefaction curves and chao1 indices for each library. Rarefaction curves of host (A) and viral (B) reads were generated from library at each time point. The chao1 indices were also used to evaluate each library.

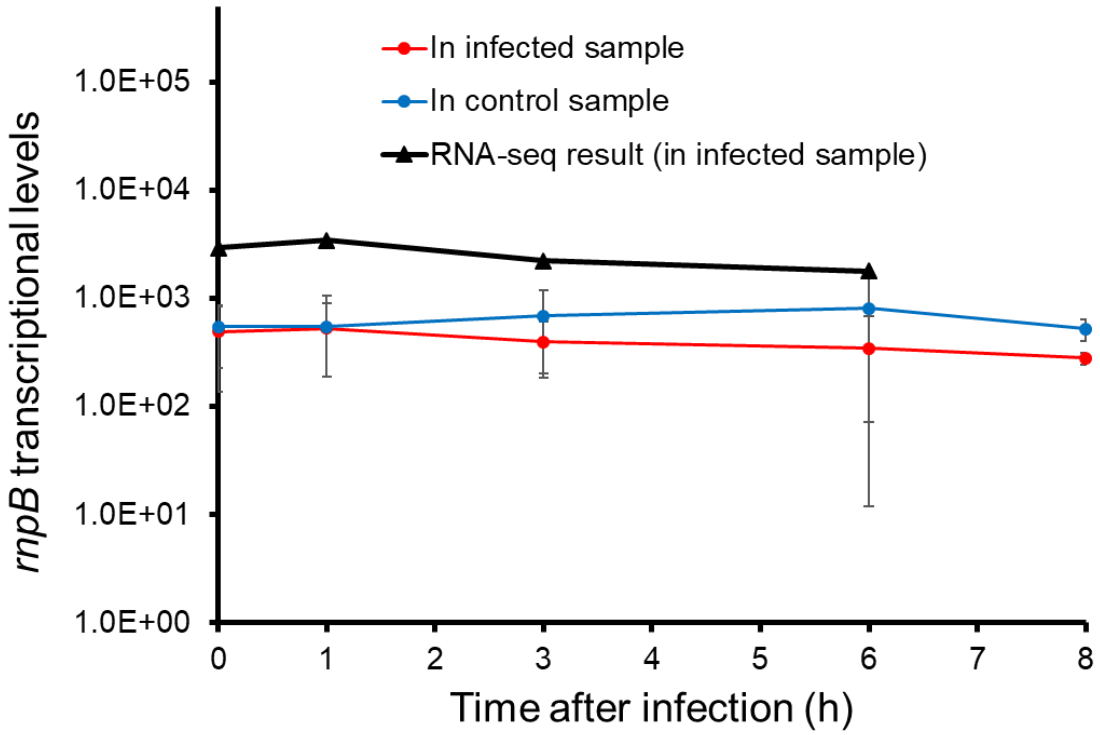


Figure 2-3. *rnpB* transcript levels during infection. Transcript levels of *rnpB* from *Microcystis aeruginosa* NIES-298 during the 8-h latent period of infection by the Ma-LMM01 phage as determined by quantitative real-time polymerase chain reaction analysis. The copy number of *rnpB* at each time point was normalized per nanogram of total RNA. Results corresponding to the infected culture are shown in red, while the uninfected control culture results are shown in blue. Three technical replicates were carried out for each biological replicate. This result was verified by transcriptional dynamics from RNA-seq data that total reads at each time point were mapped to *rnpB* sequence (shown in black). Transcript count for *rnpB* was normalized as FPKM.

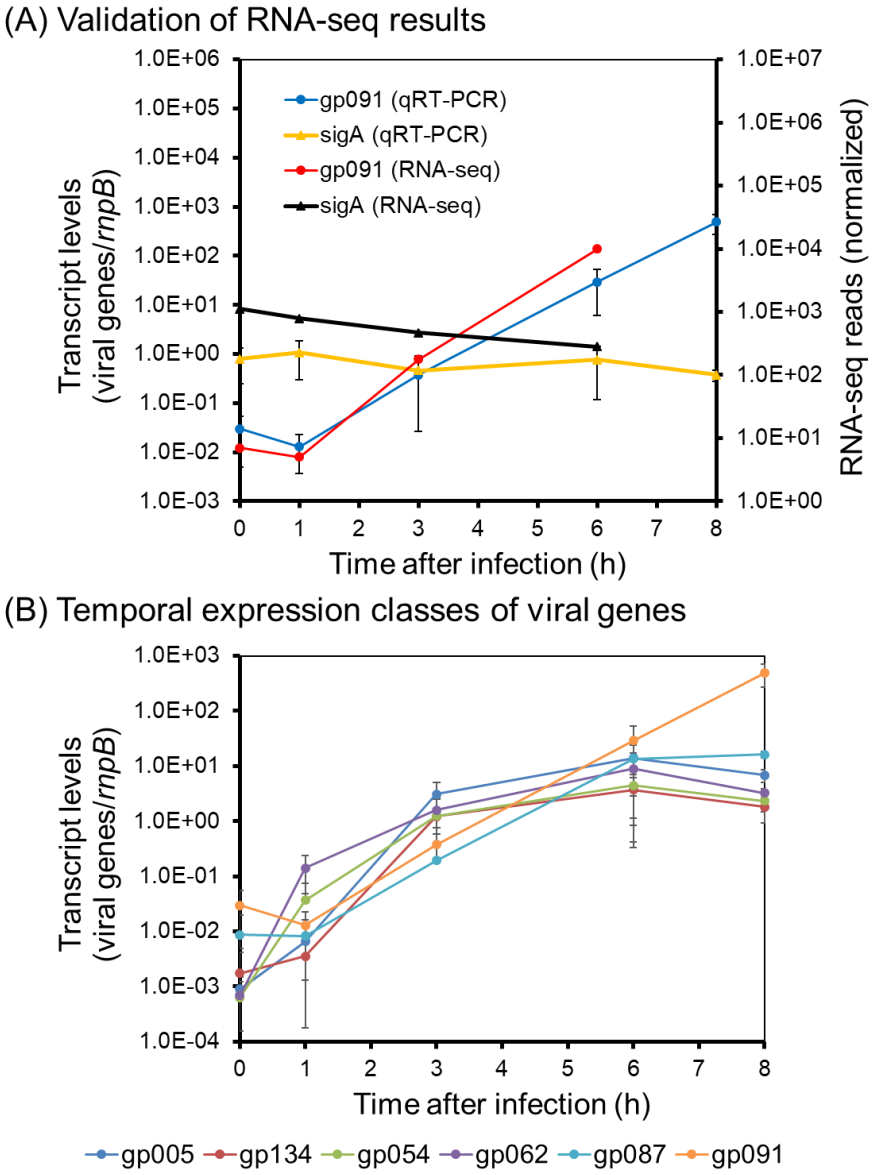


Figure 2-4. Quantitative real-time polymerase chain reaction (qRT-PCR) validation of RNA-seq results. The RNA-seq results were verified using qRT-PCR analysis of host *sigA* and viral *gp091* gene expression during infection (A). Temporal expression classes were also verified with putative early (*gp054*, *gp062*), middle (*gp087*, *gp134*), and late (*gp087*, *gp091*) genes respectively (B). Transcript levels for each gene were normalized to host *rnpB* transcript levels.

Host transcriptional responses to phage infection

Because a complete switch to phage transcription did not occur by 6 h post-infection, I investigated host transcriptional responses to phage infection in the infected cells. I generated a 4.92-Mb draft genome sequence for *M. aeruginosa* in 88 contigs (≥500 bp), containing a predicted 4749 ORFs. Strikingly, very few (0.17%) of the host genes showed significant changes in expression during infection (**Figure 2-5**). However, 8 differentially expressed genes (DEGs) were identified during Ma-LMM01 infection although an immediate response was not observed (**Table 2-2**). Of these, three genes coding for hypothetical proteins were up-regulated after both 3 and 6 h of Ma-LMM01 infection (**Figure 2-5**, **Table 2-3**). Also, the type I-D CRISPR-associated protein Cas10d/Csc3 gene, membrane protein gene, and three heat shock genes (coding for co-chaperone GroES, molecular chaperone GroEL, and heat-shock protein) were upregulated after 6 h of Ma-LMM01 infection (**Figure 2-5**, **Table 2-3**).

Table 2-2. Summary of the protein-coding host response genes.

	NIES-298 Total in genome		
	1 h	3 h	6 h
Up-regulated genes	0	3	8
Down-regulated genes	0	0	0
Unchanged genes	4749	4746	4741
Total genes		4749	

Table 2-3. Host genes with a significant change in transcript levels during phage infection. Log_2 fold changes (FC) in gene expression of infected versus control cultures with time (h) following infection. Positive values indicate an increase in transcript levels during infection, with significant differences in Log_2 FC ($P<0.05$) shown in red.

NIES-298 ID	Region in NIES-843	Predicted function	Expression pattern	Log_2 FPKM ratio at 0 h	Log_2 FPKM ratio at 1 h	Log_2 FPKM ratio at 3 h	Log_2 FPKM ratio at 6 h
NIES298_45380	N.D.	hypothetical protein	M, L	1.204	1.000	1.785	1.764
NIES298_25170	N.D.	hypothetical protein	M, L	0.835	1.060	1.307	1.827
NIES298_43020	N.D.	hypothetical protein	M, L	0.984	1.043	1.240	1.195
NIES298_45610	-	heat-shock protein	L	1.145	1.211	1.102	1.174
NIES298_19470	-	type I-D CRISPR-associated protein Cas10d/Csc3	L	1.012	0.952	1.012	1.304
NIES298_08400	-	co-chaperone GroES	L	1.074	0.991	0.978	1.140
NIES298_32000	-	membrane protein	L	1.083	1.063	1.060	1.141
NIES298_08390	-	molecular chaperone GroEL	L	1.065	1.006	0.972	1.162

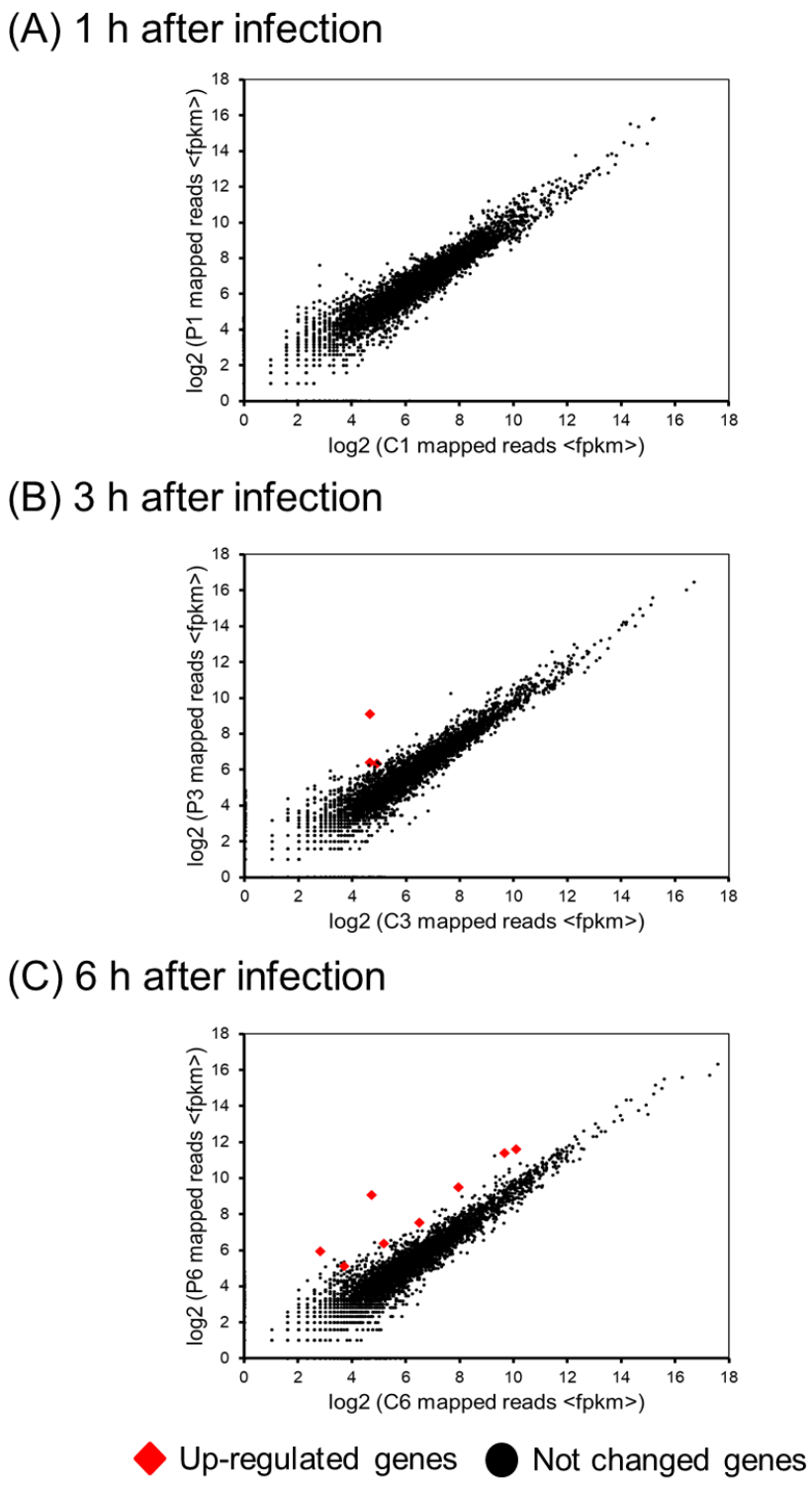


Figure 2-5. Impact of phage infection on the bacterial transcriptome. Scatter plot of the *Microcystis aeruginosa* transcriptome following phage infection (P1-P6) compared with the uninfected control (C1-C6). Each dot represents an open reading frame, with up-regulated genes shown in red and unchanged genes in black.

Amongst the DEGs with annotated function, only the type I-D CRISPR-associated protein Cas10d/Csc3 gene were associated with host defense. However, bacterial genes of unknown function in genomic islands are often differentially expressed in response to viral infection (75, 101), as well as in response to environmental stressors (102–105). Therefore, I next explored where the DEGs with unknown function were located in the *M. aeruginosa* genome. This analysis revealed that three DEGs (identified at 3 h and 6 h post-infection) showed no similarity to genes found in the *M. aeruginosa* NIES-843 defense islands (64). Therefore, three hypothetical protein genes might respond to not phage infection as host defense systems but various stresses for viral production.

Phage temporal expression patterns

Although there was very little change in host gene expression during the course of phage infection, the phage temporal expression classes were apparent when the genes were clustered according to their expression patterns (**Figure 2-6, Figure 2-7**). I assessed the quality and gene composition of the clusters obtained by hierarchical clustering analysis, using the Jaccard coefficient as a stability measure. The most stable solutions were obtained when the phage gene expression profiles were divided into two clusters, with a high frequency of high Jaccard coefficients for this number of clusters (**Figure 2-8**). Cluster 1 was composed of five genes concentrated in a single ~3.5-kb window in the Ma-LMM01 genome (*gp037–gp041*; **Figure 2-7**). Cluster 2 comprised 177 genes and could be further divided into five subclasses displaying a variety of expression patterns (B1–B5; **Figure 2-7**). For example, the expression of subclass B1 genes, which constitute 3.5% of the genome, increased drastically at 1 h post-infection, and then remained high across the remaining sampling points (**Figure 2-7**, **Table 2-4**).

These early genes, spanning 10 ORFs (Gp054–Gp063), were concentrated in a single ~5.8-kb window in the phage genome (**Figure 2-9**). However, none of the early genes appeared to have homologs in the databases, and showed no homology to early genes of T4 phage and other cyanophages.

All genes in subclass B5 were highly expressed at 3 h post-infection, and expression levels remained high throughout the infection process (**Figure 2-7**). I identified 162 genes in this middle phase of gene expression, including those involved in DNA replication, recombination/repair, and nucleotide metabolism, as well as genes coding for lysozyme and phage DNA terminase. These included DNA primase (*gp134*), DNA polymerase I (*gp178*), holB-like ATPase (*gp169*), and 3′-5′ exonuclease (*gp180*), all of which are required for phage DNA replication. In addition, the B5 subclass contained genes coding for a T4 RNA-DNA helicase UvsW homolog (*gp166*), RecA-like recombinase UvsX (*gp008*), ATP-dependent RecD-like helicase (*gp160*), and uracil-DNA glycosylase (*gp173*). These four genes are putatively associated with DNA recombination and repair functions. Further middle genes included the α (*gp006*) and β (*gp002*) subunits of ribonucleotide reductase, a flavin-dependent thymidylate synthase ThyX homolog (*gp020*), and dUTPase (*gp181*), all of which are involved in nucleotide metabolism. In addition, a phage-encoded lysozyme and terminase, which are required for cell lysis and DNA packaging, belonged to the middle cluster. The expression patterns of subclasses B2 and B3 were also similar to that of subclass B5 (**Figure 2-7**, **Table 2-4**).

The expression of all genes in subclass B4 increased gradually during infection, and reached peak levels at 6 h post-infection (**Figure 2-7**). Overall, I identified five late genes, including those coding for two major head proteins (*gp086* and *gp087*) and a phage tail sheath protein (*gp091*) (Yoshida et al., 2008; **Table 2-4**). This result is consistent with

our current understanding of the construction of T4-like phage particles, because phage structural genes tend to be transcribed later in the infection process (65, 66, 75).

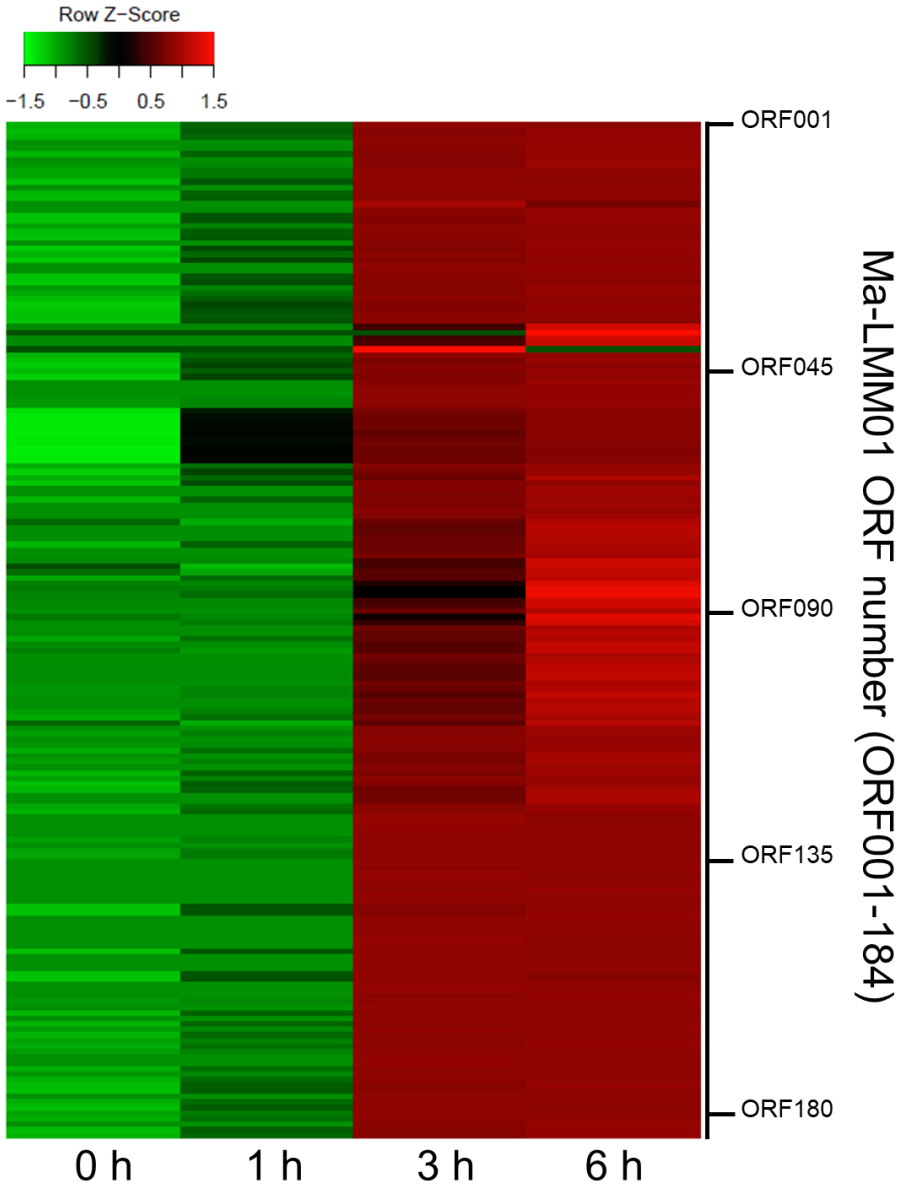


Figure 2-6. Heat map of Ma-LMM01 genes during infection. Genes are listed in order on the Ma-LMM01 genome. The color gradient indicates gene transcripts unchanged (green) or enriched (red).

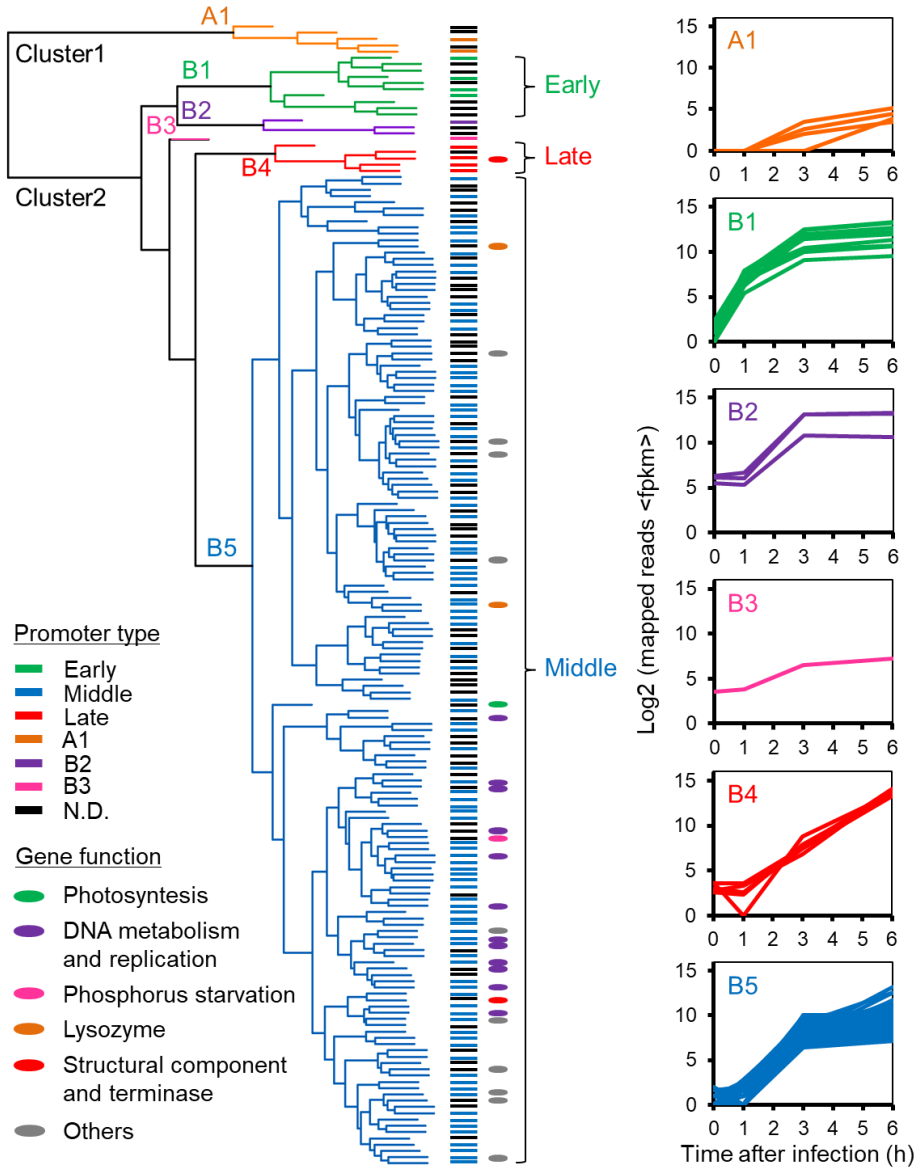


Figure 2-7. Temporal expression pattern of Ma-LMM01 phage genes. Clustering analysis of phage genes by their expression patterns is presented in the dendrogram, with the A1, B1, B2, B3, B4, and B5 expression subclasses shown in orange, green, purple, pink, red, and blue, respectively. Gene names at the dendrogram tips are colored according to the promoter class driving their expression (see legend, bottom left). The colors of the ovals adjacent to the genes denote the major classes of gene functions (see legend, bottom left). Graphs at the right of the subclasses show expression profiles of the individual genes in that subclasses as a function of time after infection. Subclass designation at the top left corner of each graph is as in Table 2-4.

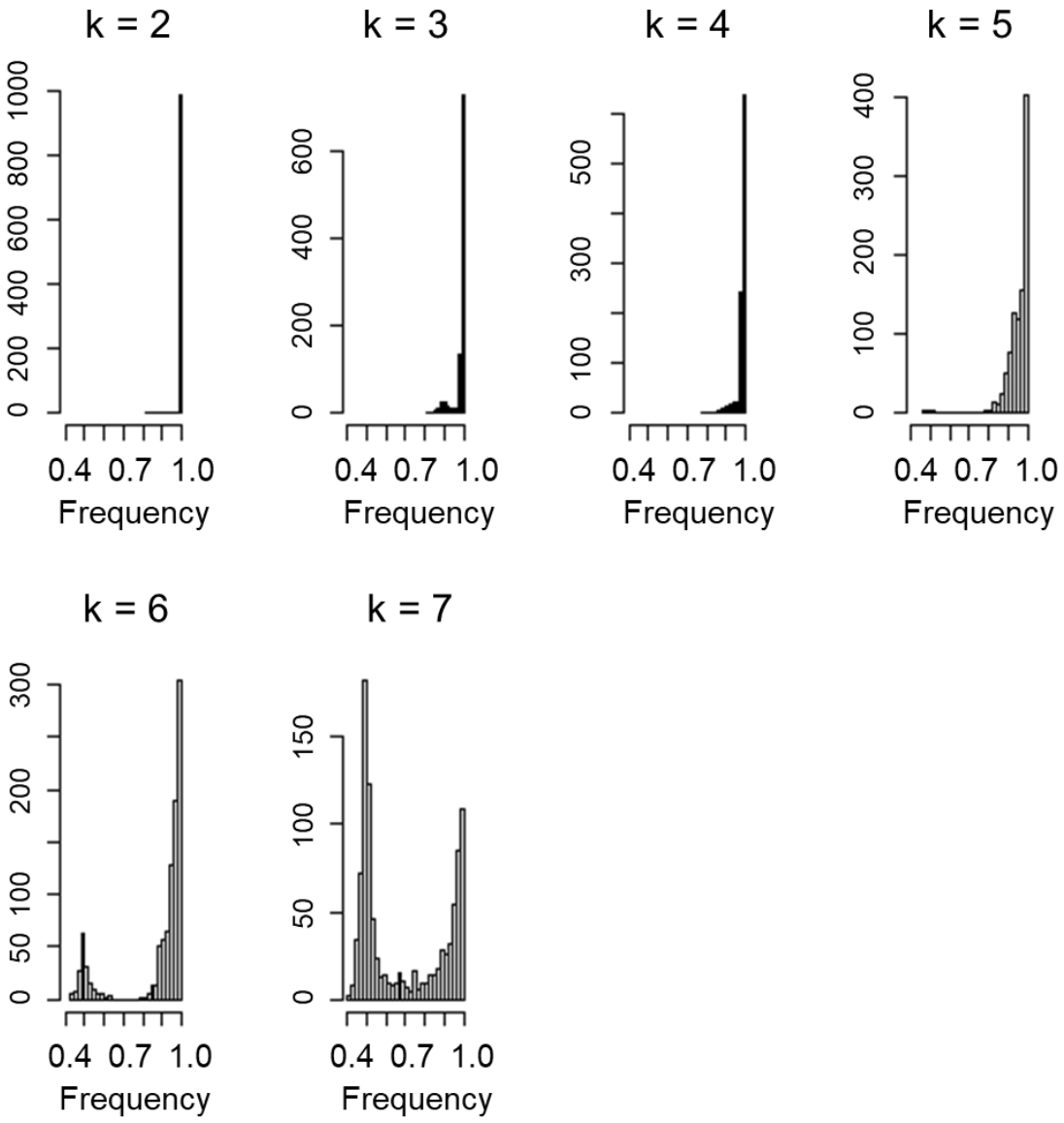


Figure 2-8. Jaccard similarity coefficients. The Jaccard coefficient was used as a stability measure to determine the most stable number of clusters obtained by hierarchical clustering analysis of the phage genes. Individual histograms show the distribution of the Jaccard coefficients for 1000 subsets of genes when 70% of the genes are randomly selected for each subset. Each histogram displays the distribution of Jaccard coefficients for a different number of clusters, k (k = 2 to k = 7). Clustering was considered stable when the majority of Jaccard values were close or equal to one.

Table 2-4. Summary of viral gene expression patterns, promoters, and expression data.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp001	9	2417	rIIA-like protein	Cluster 2	B5	-	-	-	0	2.321928	10.47675	10.38478
MaLMM01_gp002	2449	3495	ribonucleoside-diphosphate reductase beta subunit (nrdB)	Cluster 2	B5	-	-	-	0	3.169925	11.22641	11.8607
MaLMM01_gp003	3500	4294	hypothetical protein	Cluster 2	B5	M	GCCTATGCTAAGTTTAC	GTGACC	1	3.321928	11.75739	11.92815
MaLMM01_gp004	4294	4854	hypothetical protein	Cluster 2	B5	M	AGATAATCTATACGCGG	TTGAAG	0	0	12.3208	12.76466
MaLMM01_gp005	4899	5138	phycobilisome degradation protein NblA	Cluster 2	B5	-	-	-	2.321928	2.584963	12.32305	13.01681
MaLMM01_gp006	5113	7341	ribonucleoside-diphosphate reductase alpha subunit (nrdA)	Cluster 2	B5	M	AATTATACTCCTACTAA	TTGGTA	0	2.807355	11.75071	12.48809
MaLMM01_gp007	7346	7780	hypothetical protein	Cluster 2	B5	-	-	-	0	0	11.14911	11.94727
MaLMM01_gp008	7777	8841	recA recombinase	Cluster 2	B5	M	ACATACAATGATACGTA	TTTACT	1	1.584963	10.63027	11.56843
MaLMM01_gp009	8925	9413	prophage antirepressor	Cluster 2	B5	M	AGTTAATAGTTAATGGT	TTGATA	1.584963	3	11.18921	11.31911
MaLMM01_gp010	9535	10056	hypothetical protein	Cluster 2	B5	M	TGGTACACAACAGAGTA	TTGACG	0	1.584963	10.60641	10.50084
MaLMM01_gp011	10044	10169	hypothetical protein	Cluster 2	B5	M	CTCTAAAGTTACTAACT	ATGACA	0	3.321928	10.88722	10.9211
MaLMM01_gp012	10265	10663	hypothetical protein	Cluster 2	B5	-	-	-	0	0	10.90011	10.98513
MaLMM01_gp013	10636	11073	hypothetical protein	Cluster 2	B5	M	TGATCTACTACGAGATC	TTTAAT	0	2.584963	11.08015	11.18982
MaLMM01_gp014	11045	13135	hypothetical protein	Cluster 2	B5	M	TTGTATAGTGATGAACA	TTCAAT	0	2.807355	10.66622	10.78218
MaLMM01_gp015	13338	13463	hypothetical protein	Cluster 2	B5	M	TGGTAATCTCTTCCATA	TTAAAA	0	0	9.751544	8.271463
MaLMM01_gp016	13627	14169	hypothetical protein	Cluster 2	B5	M	CGCTATTATTAATCTAC	GTGTTA	2.584963	2.807355	11.25326	11.54593
MaLMM01_gp017	14166	14525	hypothetical protein	Cluster 2	B5	M	ATATATAGTGATTGGCC	CTCACG	0	3.321928	10.95783	11.53236
MaLMM01_gp018	14530	14736	hypothetical protein	Cluster 2	B5	M	GGTTATACTATCGGTGT	TAGCTA	0	3.584963	10.84078	11.51274
MaLMM01_gp019	14713	15720	hypothetical protein	Cluster 2	B5	-	-	-	1.584963	2.584963	10.71425	10.97154
MaLMM01_gp020	15717	17192	putative thymidylate synthase	Cluster 2	B5	M	ATGTTTAATGGGGCATC	TTATCA	0	3	11.03755	11.06743
MaLMM01_gp021	17250	17978	hypothetical protein	Cluster 2	B5	M	ACCTACAAAGTCCTCGT	TTGACT	0	3.169925	10.80977	11.11114
MaLMM01_gp022	17964	18128	hypothetical protein	Cluster 2	B5	M	CGGTAGCGTAACCCACA	TTATTT	0	0	9.834471	10.3696
MaLMM01_gp023	18360	18740	hypothetical protein	Cluster 2	B5	M	TTGTAGAATAATACTGT	TCGCAA	0	4.247928	11.01611	11.69349
MaLMM01_gp024	18772	19584	prophage antirepressor	Cluster 2	B5	M	GACTATAATGCAGAACC	TTCTCT	0	2.321928	10.56129	10.80655
MaLMM01_gp025	19926	21146	serine/threonine protein phosphatase	Cluster 2	B5	-	-	-	0	3.906891	10.18982	10.6865
MaLMM01_gp026	21359	21562	hypothetical protein	Cluster 2	B5	M	TGATACAATGGGCTACA	TTCACG	0	0	10.84078	10.73978
MaLMM01_gp027	21566	21688	hypothetical protein	Cluster 2	B5	-	-	-	0	0	10.46761	10.80494
MaLMM01_gp028	21685	22179	hypothetical protein	Cluster 2	B5	M	TGGTATAACTAAAGATG	TCGCAG	0	3.321928	10.783	11.05325
MaLMM01_gp029	22169	22759	hypothetical protein	Cluster 2	B5	-	-	-	0	3.70044	10.50779	10.65642
MaLMM01_gp030	22746	24266	hypothetical protein	Cluster 2	B5	M	AGGTATGATTATGCTAA	TTCATA	1.584963	2.321928	10.39124	10.98085

Table 2-4. continued.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp031	24416	25645	transposase	Cluster 2	B5	M	TGCTATAATCATAGCGT	TTGACA	1	2.584963	9.400879	9.918863
MaLMM01_gp032	25710	26129	transposase	Cluster 2	B5	M	CGCTATGATTATAGCAA	TTGACT	0	2.584963	9.85331	9.906891
MaLMM01_gp033	26239	26418	hypothetical protein	Cluster 2	B5	-	-	-	0	3.807355	9.98014	10.34207
MaLMM01_gp034	26351	26599	hypothetical protein	Cluster 2	B5	M	TTCTATCATATATTCAC	ATGATG	0	3.321928	10.18239	10.79035
MaLMM01_gp035	26606	26938	hypothetical protein	Cluster 2	B5	M	ACCTATACTCGAGACGG	TTAGTA	0	3	10.41574	10.60733
MaLMM01_gp036	26928	27611	hypothetical protein	Cluster 2	B5	M	GTTTATGCTTTCTATAG	TTAAAA	0	3.169925	9.995767	9.894818
MaLMM01_gp037	28025	28765	hypothetical protein	Cluster 1	A1	A1	CTATATACTGGTCGGGG	ATGAAA	0	0	2	3.459432
MaLMM01_gp038	28829	28954	hypothetical protein	Cluster 1	A1	-	-	-	0	0	0	3.906891
MaLMM01_gp039	29163	29633	hypothetical protein	Cluster 1	A1	-	-	-	0	0	2.584963	4.459432
MaLMM01_gp040	29780	30211	hypothetical protein	Cluster 1	A1	A1	TGGTAAAATATGACAAT	TGGATT	0	0	3.459432	5.169925
MaLMM01_gp041	30742	31509	hypothetical protein	Cluster 1	A1	-	-	-	0	0	3	0
MaLMM01_gp042	31515	32003	hypothetical protein	Cluster 2	B5	-	-	-	0	1.584963	7.169925	7.285402
MaLMM01_gp043	32096	32293	hypothetical protein	Cluster 2	B5	M	AGTTAAATTTGATTTAA	TAGATG	0	2.807355	8.672425	9.598053
MaLMM01_gp044	32400	32642	hypothetical protein	Cluster 2	B5	M	TATTATAATTATTTGTG	TTGACT	0	3.906891	9.83289	10.06743
MaLMM01_gp045	32748	33689	hypothetical protein	Cluster 2	B5	M	TTGTATGCTATTATTAG	TTGACT	0	2.584963	9.649256	10.14338
MaLMM01_gp046	33824	34039	hypothetical protein	Cluster 2	B5	M	ATTCATAATTAATAGGC	TCGATT	0	4.087463	10.00983	10.57365
MaLMM01_gp047	34602	34763	hypothetical protein	Cluster 2	B5	M	TGTTAACCTAGATGTAA	TTATAG	0	0	9.861087	10.66622
MaLMM01_gp048	35477	35755	hypothetical protein	Cluster 2	B5	M	GTTTAAGATTAATTGGA	TTTCAA	0	0	9.047124	9.087463
MaLMM01_gp049	35776	35913	hypothetical protein	-	-	-	-	-	0	3.169925	9.134426	8.879583
MaLMM01_gp050	35879	36091	hypothetical protein	-	-	-	-	-	0	3.584963	8.797662	8.714246
MaLMM01_gp051	36090	36773	hypothetical protein	Cluster 2	B2	-	-	-	6.129283	6	13.18658	13.29892
MaLMM01_gp052	36781	37011	hypothetical protein	Cluster 2	B2	-	-	-	6.321928	6.643856	13.15466	13.21189
MaLMM01_gp053	37004	37411	hypothetical protein	Cluster 2	B5	M	AGAGAGAATCTAATAGA	TTTCAA	2.584963	3.169925	10.79442	11.13699
MaLMM01_gp054	37513	38043	hypothetical protein	Cluster 2	B1	E	AAGCATCATCCCATCAA	TAGTAG	0	6.169925	11.7065	12.18921
MaLMM01_gp055	38096	38272	hypothetical protein	Cluster 2	B1	-	-	-	0	7.276124	12.4525	13.25798
MaLMM01_gp056	38265	38516	hypothetical protein	Cluster 2	B1	E	TCCCATACTAATAGTCG	TAGATA	0	7.707359	12.51988	13.32629
MaLMM01_gp057	38595	39641	hypothetical protein	Cluster 2	B1	-	-	-	2	7	10.29347	10.73978
MaLMM01_gp058	39647	39889	hypothetical protein	Cluster 2	B1	-	-	-	2.321928	7.693487	10.46352	11.3072
MaLMM01_gp059	39909	40505	hypothetical protein	Cluster 2	B1	-	-	-	1.584963	6.72792	9.971544	10.59898
MaLMM01_gp060	40662	41066	hypothetical protein	Cluster 2	B1	E	TTTTAAAGTAAAAGAAC	TTTATT	0	5.426265	9.082149	9.564149

Table 2-4. continued.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp061	41222	41686	hypothetical protein	Cluster 2	B1	-	-	-	0	7.636625	11.57554	12.18735
MaLMM01_gp062	41679	42509	hypothetical protein	Cluster 2	B1	-	-	-	0	7.257388	11.38748	11.9647
MaLMM01_gp063	42567	43226	hypothetical protein	Cluster 2	B1	E	GGATATAACCTCTAACC	TAGCCA	1	7.942515	11.94691	12.63844
MaLMM01_gp064	43811	48832	hypothetical protein	Cluster 2	B5	-	-	-	0	1.584963	7.499846	8.049849
MaLMM01_gp065	48893	49672	hypothetical protein	Cluster 2	B5	-	-	-	0	2.807355	6.629357	7.257388
MaLMM01_gp066	50144	50659	hypothetical protein	Cluster 2	B5	M	ATAAATAATAAATAAGG	TCCACT	0	1.584963	6.768184	8.357552
MaLMM01_gp067	50915	51319	hypothetical protein	Cluster 2	B5	M	GTGTATTATAGGGTGTA	TTGCTG	0	2.807355	7.876517	8.169925
MaLMM01_gp068	51615	52256	hypothetical protein	Cluster 2	B5	M	GCCTATAATACACCAAA	TATCCA	0	0	6.375039	7.129283
MaLMM01_gp069	52314	53069	chitinase	Cluster 2	B5	-	-	-	0	0	6.629357	7.643856
MaLMM01_gp070	53039	53476	hypothetical protein	Cluster 2	B5	M	TGATATAAACGCATTCA	ATGAAA	0	2	8.290019	8.918863
MaLMM01_gp071	53941	54459	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.164907	9.333155
MaLMM01_gp072	54512	55231	hypothetical protein	Cluster 2	B5	M	ATCTAGAATAGCCTCAG	GTGAAT	0	0	9.317413	9.702173
MaLMM01_gp073	55259	55621	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.290019	10.48985
MaLMM01_gp074	55590	55871	hypothetical protein	Cluster 2	B5	M	GGGTATCCTTGTAACAG	TTGATG	2	0	8.72792	10.80332
MaLMM01_gp075	55882	56049	hypothetical protein	Cluster 2	B5	M	TCATATGATGATTAATT	TTGAGT	0	0	8.005625	10.79035
MaLMM01_gp076	56049	56951	hypothetical protein	Cluster 2	B5	M	CGTTATAATAACAGGTA	TTATAG	0	0	8.011227	10.44605
MaLMM01_gp077	56939	57187	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.84549	9.960002
MaLMM01_gp078	57184	57447	hypothetical protein	Cluster 2	B5	M	GATTAGAATAAGTAAGA	CTGATA	0	2.321928	8.388017	10.18982
MaLMM01_gp079	57441	57716	hypothetical protein	Cluster 2	B5	M	TAGTAGAATTTAATGTT	ATGCTA	0	0	8.357552	10.30492
MaLMM01_gp080	57700	68388	hypothetical protein	Cluster 2	B5	M	TATTATAATGCTTGATC	TTCAGG	0	0	7.321928	8.48784
MaLMM01_gp081	68445	68807	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.894818	13.24079
MaLMM01_gp082	68990	69238	hypothetical protein	Cluster 2	B4	L	AGTTATTCTGTAAGAAT	TTCCAC	3.70044	0	8.813781	13.55051
MaLMM01_gp083	69235	70575	hypothetical protein	Cluster 2	B5	-	-	-	2	0	9.116344	12.57175
MaLMM01_gp084	70568	70990	hypothetical protein	Cluster 2	B5	M	GGGTATCAATAGTCTCG	TTGAAT	0	2	8.383704	11.42364
MaLMM01_gp085	71028	72380	hypothetical protein	Cluster 2	B4	-	-	-	2.584963	2.321928	7.864186	13.43723
MaLMM01_gp086	72407	73720	major head proteins	Cluster 2	B4	L	AGCTATTAGATATTGCT	ATGTAA	3.584963	3.584963	7.5157	14.06432
MaLMM01_gp087	73749	74855	major head proteins	Cluster 2	B4	L	GTTTATAATTTCTCACC	ATCCAA	2.807355	3.321928	6.83289	13.98344
MaLMM01_gp088	74913	75581	hypothetical protein	Cluster 2	B5	M	TCTGATGCTCTCACTAT	TTGAAG	0	0	6.714246	11.76777
MaLMM01_gp089	75594	77471	hypothetical protein	Cluster 2	B5	M	AGTTAGAAAAACTGACA	TTTAGA	1	1	7.629357	11.70044
MaLMM01_gp090	77468	78166	hypothetical protein	Cluster 2	B5	M	CGCCATACTAGACTCTA	TTAAAT	0	0	7.857981	10.05257
MaLMM01_gp091	78218	80542	putative phage tail sheath protein	Cluster 2	B4	L	CTTGAGTCTATCTAATA	TTTACT	3	2.584963	7.475733	13.26927

Table 2-4. continued.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp092	80547	81257	hypothetical protein	Cluster 2	B5	-	-	-	1	1.584963	7.507795	12.54424
MaLMM01_gp093	81316	83187	hypothetical protein	Cluster 2	B5	M	ACTCAGGATTTCCTGAG	TTCTCT	0	0	8.159871	10.0634
MaLMM01_gp094	83187	84752	hypothetical protein	Cluster 2	B5	-	-	-	1	1	8.434628	10.8933
MaLMM01_gp095	84752	85945	lysozyme/metalloendopeptidase	Cluster 2	B5	M	TTTTACAATCCCTGAAG	TTTCGA	0	1.584963	8.011227	10.16742
MaLMM01_gp096	85950	86462	hypothetical protein	Cluster 2	B5	-	-	-	1.584963	1.584963	7.658211	10.36085
MaLMM01_gp097	86452	87546	hypothetical protein	Cluster 2	B5	M	CTTTATGATGCTAGCTA	CTGATA	1	0	7.577429	10.51865
MaLMM01_gp098	87543	91076	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.016808	9.733015
MaLMM01_gp099	91207	91560	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.761551	9.881114
MaLMM01_gp100	91563	91940	hypothetical protein	Cluster 2	B5	M	GCTTATATTAGAGGCAA	TTAACT	0	0	7.686501	10.45943
MaLMM01_gp101	92000	92980	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.366322	10.30492
MaLMM01_gp102	92973	95762	hypothetical protein	Cluster 2	B5	-	-	-	1	1	7.209453	9.791163
MaLMM01_gp103	95771	96586	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.467606	9.057992
MaLMM01_gp104	96683	97483	hypothetical protein	Cluster 2	B3	-	-	-	3.584963	3.807355	6.507795	7.257388
MaLMM01_gp105	97622	99355	hypothetical protein	Cluster 2	B5	M	TATTATAATATATAGTG	TTAAAT	1.584963	2	7.73471	10.24317
MaLMM01_gp106	99480	100493	putative lysine/ornithine N-monooxygenase	Cluster 2	B5	-	-	-	0	0	8.164907	10.42626
MaLMM01_gp107	100493	101050	hypothetical protein	Cluster 2	B5	M	TCGTCTAATAAAGAATG	TTTCCA	0	0	7.83289	10.5411
MaLMM01_gp108	101054	105133	hypothetical protein	Cluster 2	B5	M	ATGTAGTATTACAAACA	CTGCAT	0	1	8.049849	10.41363
MaLMM01_gp109	105120	107642	hypothetical protein	Cluster 2	B5	M	AGTCATACTAGATGCAG	TTCTAA	0	1.584963	8.303781	9.787903
MaLMM01_gp110	107627	107962	hypothetical protein	Cluster 2	B5	M	CATTAGATTACAGATAA	TGGCCA	2	0	8.174926	10.66534
MaLMM01_gp111	107955	109535	hypothetical protein	Cluster 2	B5	M	CAGTATATTCATATACC	TAGAGA	0	0	8.299208	8.945444
MaLMM01_gp112	109519	111477	hypothetical protein	Cluster 2	B5	-	-	-	0	1	8.581201	9.364135
MaLMM01_gp113	111492	112640	hypothetical protein	Cluster 2	B5	M	TTATATAAAAGTTCTGA	TTGCTA	1	1	8.921841	9.243174
MaLMM01_gp114	112618	113463	hypothetical protein	Cluster 2	B5	M	ATCTATAAGAGAATGGA	GTGACT	1.584963	2	9.558421	10.00282
MaLMM01_gp115	113457	113825	hypothetical protein	Cluster 2	B5	M	AGATAGTATTTTATACA	TAGATA	0	2	9.571753	10.51767
MaLMM01_gp116	113827	114297	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.100662	10.61471
MaLMM01_gp117	114328	115116	hypothetical protein	Cluster 2	B5	-	-	-	1	2	9.154818	10.58214
MaLMM01_gp118	115106	116884	terminase	Cluster 2	B5	-	-	-	1	1	9.355351	10.02791
MaLMM01_gp119	117001	117636	hypothetical protein	Cluster 2	B5	-	-	-	0	2.321928	8.894818	10.0348
MaLMM01_gp120	117618	126314	hypothetical protein	Cluster 2	B5	M	TAGTAGAATACGATACA	TTACCT	0	1	8.784635	9.019591
MaLMM01_gp121	126316	126504	hypothetical protein	Cluster 2	B5	M	CTTTATAATAGCCTCAA	TATACT	0	2.807355	9.756556	10.47371
MaLMM01_gp122	126494	126721	hypothetical protein	Cluster 2	B5	M	AAGGAGAATCCAGCTAT	TTGTAA	0	2.584963	9.960002	11.79035

Table 2-4. continued.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp123	126721	127431	hypothetical protein	Cluster 2	B5	M	GGGTAATGTCTATACAG	ATCCCT	0	0	9.60548	11.8815
MaLMM01_gp124	127412	127858	hypothetical protein	Cluster 2	B5	M	CGATAAGATTAATGGGT	TTGACA	0	0	9.651052	11.68956
MaLMM01_gp125	127908	129050	hypothetical protein	Cluster 2	B5	M	TGCTAATTTATTCAGGA	TTGCTG	0	1.584963	9.769838	10.20823
MaLMM01_gp126	129022	129567	hypothetical protein	Cluster 2	B5	-	-	-	0	2.321928	10.15861	10.61287
MaLMM01_gp127	129616	129819	hypothetical protein	Cluster 2	B5	M	CAGTATGCTCCTACTAG	TTGATG	0	0	9.455327	9.041659
MaLMM01_gp128	129804	130721	putative Fe-S oxidoreductase	Cluster 2	B5	-	-	-	0	0	10.01402	9.658211
MaLMM01_gp129	130726	131064	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.375039	9.276124
MaLMM01_gp130	131113	131610	hypothetical protein	Cluster 2	B5	M	GGGTAGTCTAGTATCAG	CTGTCG	0	0	9.957102	9.994353
MaLMM01_gp131	131837	132637	hypothetical protein	Cluster 2	B5	M	GCGTATTATCAAGAAAG	TTGCGA	0	1	10.35315	10.32193
MaLMM01_gp132	132642	132983	hypothetical protein	Cluster 2	B5	M	GGGTATACTAGATCATA	ATCACG	0	0	10.04166	10.03892
MaLMM01_gp133	132973	133473	hypothetical protein	Cluster 2	B5	M	CCGTATACTATTCGTAG	CTGATA	0	1.584963	10.99789	11.06542
MaLMM01_gp134	133470	134663	putative DNA primase	Cluster 2	B5	M	AGGTTTAATAGAACGGC	TTGCTC	1	2.321928	10.85487	10.90162
MaLMM01_gp135	134712	135884	transposase	Cluster 2	B5	-	-	-	0	0	9.124121	9.167418
MaLMM01_gp136	135862	136470	putative site-specific integrase-resolvase	Cluster 2	B5	-	-	-	0	0	8.854868	9.087463
MaLMM01_gp137	136752	136943	hypothetical protein	Cluster 2	B5	M	TGGTATAATAATTGGAT	TTTCTA	0	0	9.243174	8.933691
MaLMM01_gp138	136930	137118	hypothetical protein	Cluster 2	B5	M	CTCTATGCTTATTCCTG	TCTACA	0	0	8.442943	8.23362
MaLMM01_gp139	137257	137517	hypothetical protein	Cluster 2	B5	M	GTGTATAATAAAATGAA	TTGATT	0	0	8.787903	8.8009
MaLMM01_gp140	137532	137687	hypothetical protein	Cluster 2	B5	M	TTTTATTATTATGACAA	TTGTAA	0	0	9.036174	9.301496
MaLMM01_gp141	137674	137883	hypothetical protein	Cluster 2	B5	M	CGTAATAATATGAATAT	TTGTAA	0	0	8.060696	8.011227
MaLMM01_gp142	137834	138049	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.111136	8.982994
MaLMM01_gp143	138050	138238	hypothetical protein	Cluster 2	B5	-	-	-	0	2.807355	8.936638	9.126704
MaLMM01_gp144	138225	138464	hypothetical protein	Cluster 2	B5	-	-	-	0	2.584963	8.124121	8.483816
MaLMM01_gp145	138433	138813	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.071462	8.189825
MaLMM01_gp146	138825	139067	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.149747	8.124121
MaLMM01_gp147	139054	139260	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.262095	8.413628
MaLMM01_gp148	139368	139523	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8	7.930737
MaLMM01_gp149	139547	139750	hypothetical protein	Cluster 2	B5	-	-	-	0	0	7.651052	7.285402
MaLMM01_gp150	139737	140057	hypothetical protein	Cluster 2	B5	M	CTTTATAATTCTGGAGG	TTTATA	0	0	7.554589	7.426265
MaLMM01_gp151	140096	140329	hypothetical protein	Cluster 2	B5	-	-	-	0	2.584963	8.661778	8.442943
MaLMM01_gp152	140298	140471	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.818582	9.651052
MaLMM01_gp153	140458	140622	hypothetical protein	Cluster 2	B5	M	GTTTATTATAGATTAGA	TTTACT	0	0	9.41996	9.285402

Table 2-4. continued.

Locus tag	from	to	Product (Reference)	Cluster	Sub cluster	Promoter type	-10 position motif	-35 position motif	Log2 FPKM at 0 h	Log2 FPKM at 1 h	Log2 FPKM at 3 h	Log2 FPKM at 6 h
MaLMM01_gp154	140609	140818	hypothetical protein	Cluster 2	B5	M	TTGTATTATACATCCTA	TTTATT	0	0	9.126704	9.169925
MaLMM01_gp155	140805	141017	hypothetical protein	Cluster 2	B5	-	-	-	0	2.584963	8.797662	8.577429
MaLMM01_gp156	141004	141189	hypothetical protein	Cluster 2	B5	M	CTTTATAATGCTGGGGG	TTTATA	0	2.807355	8.643856	8.366322
MaLMM01_gp157	141155	141397	hypothetical protein	Cluster 2	B5	-	-	-	0	0	8.861087	8.848623
MaLMM01_gp158	141947	142273	hypothetical protein	Cluster 2	B5	M	CGCTATTGTTAGTTTAA	TTGTTA	0	0	8.661778	8.361944
MaLMM01_gp159	142399	142680	hypothetical protein	Cluster 2	B5	M	TTCTATAATCAGTTAGG	TTGTTA	0	0	9.820179	10.22038
MaLMM01_gp160	142680	143828	putative helicase	Cluster 2	B5	M	GTTGATTATCTATCTCG	TTGAAA	1	1.584963	9.941048	10.19353
MaLMM01_gp161	143800	144255	hypothetical protein	Cluster 2	B5	-	-	-	0	0	9.491853	9.463524
MaLMM01_gp162	144239	144850	hypothetical protein	Cluster 2	B5	-	-	-	0	2.321928	9.507795	9.590587
MaLMM01_gp163	145289	145915	hypothetical protein	Cluster 2	B5	M	TGTTAAAGTTAATCAAT	TTGACA	0	0	7.78136	7.70044
MaLMM01_gp164	146035	146298	hypothetical protein	Cluster 2	B5	M	GCAAAAAATAACTAATT	TTTCCA	0	2.321928	10.25503	10.53041
MaLMM01_gp165	146295	146894	hypothetical protein	Cluster 2	B5	M	TGATAACGTTACTGTAT	TAGCCA	1.584963	2.321928	11.63662	11.72494
MaLMM01_gp166	146894	148345	superfamily II DNA/RNA helicase	Cluster 2	B5	-	-	-	0	2.584963	12.0372	12.17524
MaLMM01_gp167	148342	149568	hypothetical protein	Cluster 2	B5	-	-	-	1	2.321928	11.78382	12.12089
MaLMM01_gp168	149617	150303	hypothetical protein	Cluster 2	B5	M	GGATATCATTAATAGAA	TAGCAG	1	3	12.06676	12.38613
MaLMM01_gp169	150291	151394	holB-like ATPase involved in DNA replication	Cluster 2	B5	-	-	-	0	1	11.80373	11.93701
MaLMM01_gp170	151375	151545	hypothetical protein	Cluster 2	B5	-	-	-	0	0	11.59199	11.45635
MaLMM01_gp171	151605	152753	hypothetical protein	Cluster 2	B2	-	-	-	5.459432	5.357552	10.80896	10.6591
MaLMM01_gp172	153022	153285	hypothetical protein	Cluster 2	B5	M	GGATATAAATATCATCA	ATGATG	0	2.321928	11.65687	11.8392
MaLMM01_gp173	153269	154024	uracil-DNA glycosylase	Cluster 2	B5	-	-	-	1	1.584963	11.6425	11.65777
MaLMM01_gp174	154027	155163	hypothetical protein	Cluster 2	B5	-	-	-	0	2.321928	11.63072	11.58965
MaLMM01_gp175	155187	155558	hypothetical protein	Cluster 2	B5	-	-	-	0	3.321928	12.01959	12.61494
MaLMM01_gp176	155656	156759	hypothetical protein	Cluster 2	B5	-	-	-	0	3.321928	11.57601	11.7719
MaLMM01_gp177	156716	157282	hypothetical protein	Cluster 2	B5	M	ATATACCATGCCGGGCT	TTAAGG	0	0	11.27088	11.27903
MaLMM01_gp178	157270	158232	DNA polymerase	Cluster 2	B5	M	GCGTATACTAGCCAACC	TTAGTA	0	2.321928	11.2246	11.11699
MaLMM01_gp179	158210	158365	hypothetical protein	Cluster 2	B5	M	TAGTAATGTTGAGCTGA	TTCAAT	0	3	11.14275	11.10852
MaLMM01_gp180	158352	159635	3'-5' exonuclease	Cluster 2	B5	M	TGGGAGACTTATCTAAG	TTCACA	1	2.321928	11.41521	11.43775
MaLMM01_gp181	159632	160306	dUTPase	Cluster 2	B5	M	CGTTCTATTAGTCCAGA	TTGCTG	1	3	11.224	11.18302
MaLMM01_gp182	160333	160542	hypothetical protein	Cluster 2	B5	M	AGGTATAATACGCTTCG	TTGTGG	2.584963	2.584963	11.3072	11.91551
MaLMM01_gp183	160545	161216	phoH-like phosphate starvation-inducible protein	Cluster 2	B5	-	-	-	0	3	11.14466	11.59945
MaLMM01_gp184	161274	162089	hypothetical protein	Cluster 2	B5	M	TACTATAATGCCCTTAA	TCTAAA	1	3.459432	12.17305	12.73471

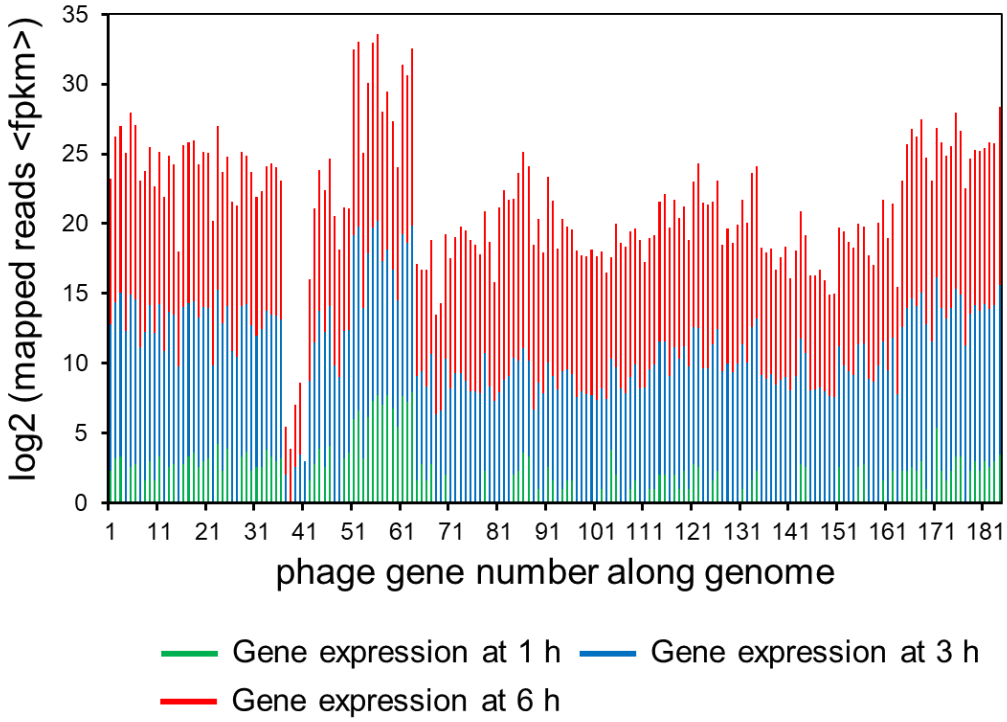


Figure 2-9. Genome organization of phage early, middle, and late genes. Log_2 fold changes in expression of phage genes during infection of *Microcystis aeruginosa* NIES-298 at 1 h (green), 3 h (blue), and 6 h (red) post-infection.

Viral regulation of the transcriptional patterns

To understand how the phage expression patterns are regulated with no change in host transcriptional levels, I examined the upstream regions of all phage genes. Grouping of promoters according to the timing of gene expression revealed similar motif signatures that are likely to be responsible for directing the expression of the early, middle, and late genes (**Figure 2-10**). A canonical cyanobacterial σ^{70} (SigA) recognition-like sequence was found upstream of the early genes, and comprised two palindromic 6-bp motifs, separated by 16–18 bp, located 6–8 bp upstream of the transcriptional start site (**Figure 2-10A**). This is consistent with current knowledge on phage early transcription,

as early phage gene expression is usually regulated by the host core transcriptional machinery, and hence, phage early promoters are expected to resemble host σ^{70} promoters (43, 65). Similarly, the promoters of the Ma-LMM01 middle genes were characterized by the SigA recognition site, suggesting that these genes were also transcribed using the core host transcriptional machinery (**Figure 2-10B**). Finally, although late-gene promoters showed a distinct motif, it was still very similar to the SigA recognition motif as well as early- and middle-gene promoters (**Figure 2-10C**). Late transcription of T4-like phages is generally independent of the host σ^{70}, and is instead mediated by a phage-encoded σ factor. Hence, phage late-gene promoters are expected to resemble T4 late promoters (43, 65, 75). Therefore, in contrast to current knowledge on phage transcriptional regulation, these results suggest that all phases of Ma-LMM01 gene transcription are dependent on host σ^{70}. Overall, I defined four early-gene promoters, 97 middle-gene promoters, and four late-gene promoters upstream of phage genes. Transcriptomic read mapping pattern of viral genes supported the results of promoter prediction in each expression classes (**Figure 2-11**). Furthermore, I determined that the most plausible Ma-LMM01 early-, middle-, and late-gene promoter sequences were TAGNNNN$_{16\text{-}18}$YATANT, TTNNNNN$_{10\text{-}25}$TANNNT, and WTNNANN$_{16\text{-}22}$TATTMT, respectively (**Figure 2-10**).

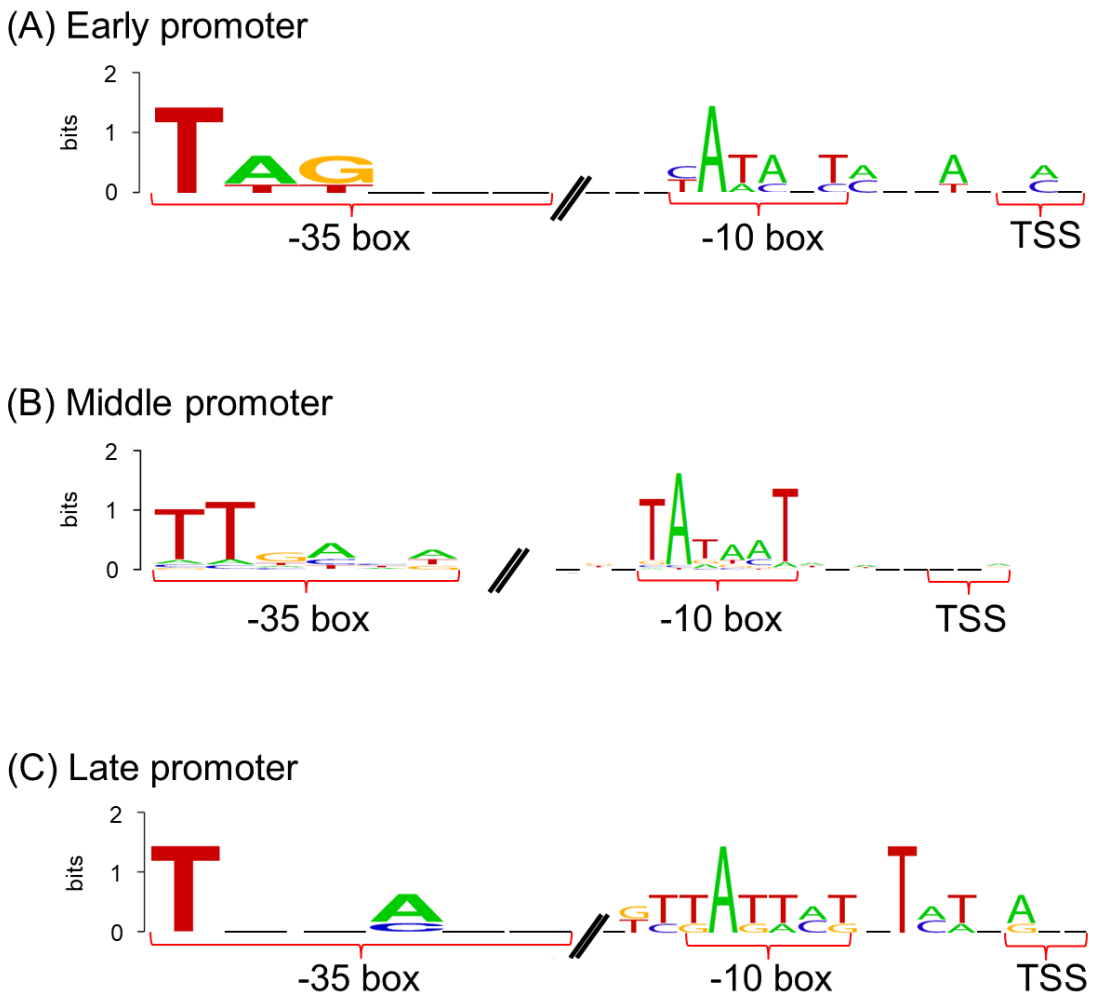


Figure 2-10. Ma-LMM01 early, middle, and late promoters. Promoter logos of the sequences upstream of the transcription start sites of the phage early (A), middle (B), and late (C) genes. Promoter logos were generated from four early-gene promoters (A), 97 middle-gene promoters (B), and four late-gene promoters (C).

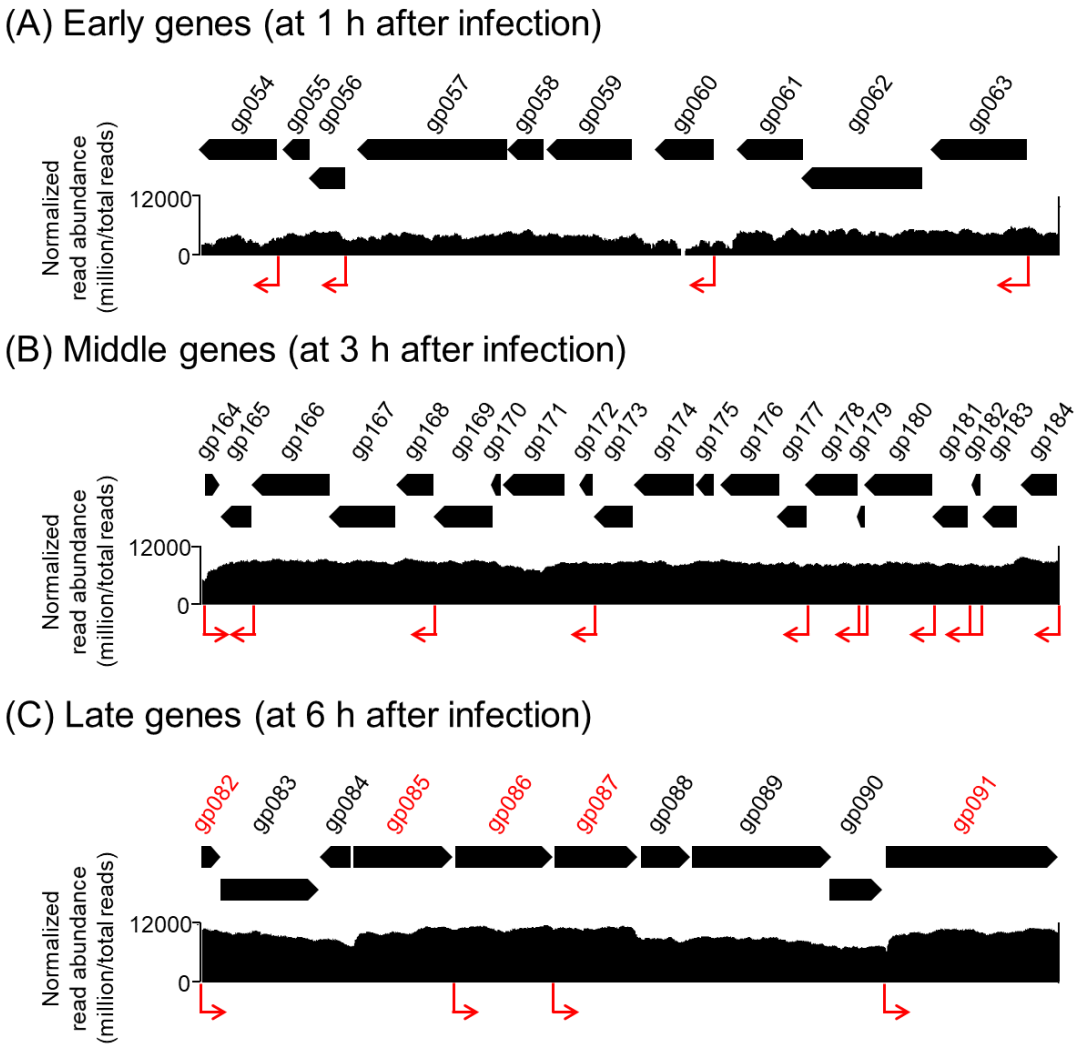


Figure 2-11. Transcriptomic read mapping pattern of early, middle, and late genes. Part of genome-wide overview of reads mapped to the Ma-LMM01 genome at samples taken 1 h (A), 3 h (B), and 6 h (C) after infection, which are composed of early, middle, late genes respectively. Genes shown in red represent the viral late genes (C). The visualization analyses for each time point were conducted independently. Read abundance was normalized by Integrative Genomics Viewer function (million per total read count). Red arrows indicate the position of predicted early, middle, and late promoter sequence *in silico* analysis.

Discussion

This study is the first report on both the transcriptional profile of cyanophage Ma-LMM01 during host infection, and the whole transcriptional response of *M. aeruginosa* to the phage infection. The results revealed that Ma-LMM01 infection did not affect the transcriptional levels of most host genes (99.83%) at any point during the infection process (**Figure 2-5**). Honda *et al*. (2014) previously investigated transcriptional alterations of host genes involved in cellular processes during Ma-LMM01 infection (4 housekeeping genes, 7 stress response genes, 3 carbonhydrate metabolic genes, and 5 photosynthetic genes). In this analysis, no remarkable change in expression levels of these genes is observed (84). Thus, unlike other marine T4-cyanophages (74, 75, 77), an incomplete switch to phage transcription was observed in the infected cells throughout the Ma-LMM01 infection process. This result suggested that Ma-LMM01-encoded proteins did not cause the changes in host transcriptional activity as seen in T4 phage (43, 65). Ma-LMM01 lacks phage-encoded proteins that are important for switching from host to phage transcription, including AsiA (anti-σ factor), MotA (activator for middle transcription), gp33 (co-activator for late transcription), and gp55 (T4-encoded σ factor) (43, 49). Therefore, these results coincide with the genomic features of Ma-LMM01.

Similar results were also obtained for the PRD1 and PRR1 phage during infection of *E. coli* K-12 (106) and *Pseudomonas aeruginosa* PAO1 (107), respectively. These phages are thought to down-regulate host protein synthesis, mainly by controlling ppGpp concentration (106) and ribosomal protein synthesis (107), respectively, to channel host resources for viral reproduction. However, none of the host genes involved in protein synthesis showed a significant change in expression during Ma-LMM01 infection (**Figure 2-5**). An infection profile that does not affect host transcriptional levels

might be advantageous for Ma-LMM01 to ensure viral propagation without the induction of host defense systems. *M. aeruginosa* possesses the highest number of putative antiviral defense systems of any prokaryote or archeal species examined to date (64), including CRISPR-Cas systems, restriction-modification systems, Toxin-Antitoxin (TA) systems, and abortive infection systems. In particular, putative TA genes are highly abundant in the genome of *M. aeruginosa* (396 TA genes of 492 total defense genes) (64). A TA system is generally composed of a stable toxin protein and an unstable antitoxin protein or small RNA, meaning that it is essential to continuously synthesize antitoxin to neutralize the toxin (108). Therefore, phage-mediated host transcriptional shutoff may robustly induce programmed cell death caused by TA systems in this species. Of the distinct defense genes, the type I-D CRISPR-associated protein Cas10d/Csc3 gene showed significant change in transcriptional level during infection (**Table 2-3**). In CRISPR-Cas systems, the incorporation of foreign DNA fragments into the CRISPR array, mediated by Cas1 and Cas2, first occurs in the adaptation stage. CRISPR RNAs (cRNAs), which are transcribed from the CRISPR array in the expression and interference stages, then function as guides to specifically target and cleave the nucleic acids of cognate viruses or plasmids with the aid of the Cas proteins. Multiprotein crRNA-effector complexes including Cas10d/Csc3 protein mediate the processing and interference stages of type I-D CRISPR-Cas system (109). However, in the current study, the expression of other CRISPR-Cas system related genes were not significantly altered during infection, indicating that CRISPR-Cas systems may not be effective against Ma-LMM01 infection even though *M. aeruginosa* NIES-298 possess spacers matching for this viral genome (61). This fact supported the hypothesis that infection profile without affecting host transcriptional levels might enable Ma-LMM01 to escape the highly abundant host defense systems during infection.

Unknown non-enzymatic peptides are recently thought to be significantly important for viruses to take-over host metabolism (110). Among numerous Ma-LMM01- genes with unknowns (49), 10 early genes observed in this study (**Figure 2-7**) may be important to repress host responses to phage infection. Host transcriptional responses in the early infection in other marine cyanophages (75, 101) were not observed in this study, which supports this idea. Another possibility is that *M. aeruginosa* may control or limit the highly abundant TA systems so as not to frequently induce programmed cell death. Also, the function of up-regulated membrane protein gene remains to be understood although bacterial cell surface-related genes are associated with blocking phage adsorption (111).

Like other T4-like marine cyanophages, Ma-LMM01 showed three temporal expression classes: early, middle, and late (**Figure 2-7**). The qRT-PCR based phage expression patterns supported this result (**Figure 2-4**). In addition, promoter motifs linked to each of the expression classes were similar to those of cyanobacterial primary σ factor SigA recognition-like sequences, although they differed slightly from each other (**Figure 2-10**). Transcription in cyanobacteria is controlled by the host RNA polymerase core enzyme in combination with heterogeneous σ factors that are assigned into groups 1 (SigA), 2 (SigB–E), and 3 (SigF–J) (112). Group 1 and 3 σ factors are essential for cell viability and survival under stress, respectively, whereas group 2 σ factors are nonessential for cell viability. These findings suggest that Ma-LMM01 does not utilize alternative σ factors but favors the primary σ factor SigA for gene expression. This hypothesis coincides with the viral genomic features and host transcriptional responses observed during Ma-LMM01 infection. However, the homogenous promoter sequences raise the question as to how Ma-LMM01 controls the three temporal expression patterns during infection. One possibility is that the slight differences among promoter sequences

may contribute to distinguishing between the three promoter types as well as the host promoters (43). Another possibility is that early-gene products (Gp054–Gp063) may be involved in the regulation of viral expression patterns during Ma-LMM01 infection. In general, phage-encoded early products usually modify the host RNA polymerase complex, and then switch viral expression classes (43). However, Ma-LMM01 early-gene products may employ a novel mechanism to control their expression patterns, as phage gene products usually down-regulate host transcription (43).

In addition to the scenario that Ma-LMM01 maintained host transcriptional activity as described above, there is an alternative possibility that decomposition as well as production of host transcripts did not occur, and thereby host transcripts were apparently stable. In T4 phage, host transcription is halted shortly after infection even though viral reproduction are independent of host transcripts at all (43, 65), which supports this scenario. In either scenario, it is possible that Ma-LMM01 encounters nutrient limitation during infection, particularly in molecules such as nucleic acids and amino acids that are required for viral reproduction. In general, bacteriophages exploit the host metabolism to establish an efficient infection cycle and redirect host cell components, including metabolic substrates and the machinery for replication, transcription, and translation, towards the production of new virions (65). In T4 phage, for example, nucleotide precursors for DNA replication are generated from host DNA degradation, and host transcription is down-regulated by the sequential modification of host transcriptional machinery (65). In addition, T4-like marine cyanophages redirect carbon flux from the Calvin cycle to the pentose phosphate pathway, maintaining host photosynthesis by using AMGs (77, 79). However, Ma-LMM01 lacks the defined genes that are required for the acquisition of precursors for their replication and virion

morphogenesis (49). One possibility is that viral DNA replication and virion synthesis proceed gradually during Ma-LMM01 infection using the remaining precursors inside the infected cells. This hypothesis is supported by the observations that middle-gene transcriptional levels remained elevated up to 3–6 h post-infection, and late-gene transcriptional levels increased gradually during the latent period (**Figure 2-7**). Also, *M. aeruginosa* has a larger genome (4.92 Mb) than marine cyanobacteria such as *Prochlorococcus* and *Synechococcus* (1.64–2.86 Mb) (113). This suggests that unlike in other marine cyanobacteria, nucleotide precursors are highly abundant in *Microcystis* cells, and can therefore be exploited by Ma-LMM01 for its own DNA replication. Furthermore, phage-encoded NblA may provide amino acid precursors that are required for viral protein synthesis during Ma-LMM01 infection. The degradation of phycobilisomes, catalyzed by NblA, is thought to provide a pool of resources that can be reused by cyanobacteria during nutrient limitation (114). In addition, phage-encoded NblA is active in *Planktothrix* phage PaV-LD, and degrades the host phycobilisomes (82). Also, Yoshida *et al*. (2008) previously reported that Ma-LMM01-encoded *nblA* is expressed in the infected culture and that host phycobilisomes are degraded during Ma-LMM01 infection. These findings suggest that Ma-LMM01-encoded NblA contributes to sustaining amino acid pools to prevent nutrient limitation. Phage-encoded NblA and host heat-shock proteins such as GroES and hspA are thought to contribute the maintenance of photosynthesis activity during Ma-LMM01 infection (84). According to this idea, transcriptional levels of *GroES*, *GroEL* and heat-shock protein gene significantly increased at 6 h post-infection (**Table 2-3**). Thus, Ma-LMM01 may provide the precursors and energy for their reproduction by maintaining photosynthetic apparatus and degradation of phycobilisome. In addition, T4 phage halts phage development until

appropriate nutrients become available in the stationary phase *E. coli* cells although DNA replication is completed (115). Similarly, T4-like marine cyanophage P-SSM2 responds to phosphate limited conditions and maintains the host phosphate uptake rate during infection by controlling host PhoR/PhoB two-component signal transduction system (76, 116). Ma-LMM01 which encounters nutrient limitation during infection may possess similar mechanisms that adjust infection process and host physical states in response to cellular conditions as seen in these phages. Another possibility is that Ma-LMM01 may control host metabolism in translational levels to provide the precursors such as nucleotides and amino acids for their reproduction. Indeed, cellular adaptation for the production of phage progeny is thought to be more active at the translational or posttranslational level in *Lactococcus lactis* phage Tuc2009 and c2 (117) and *Pseudomonas aeruginosa* phage LUZ19 (118), in which the minimal transcriptional response is observed during infection. Translational dynamics during Ma-LMM01 infection will help us to further understand the viral impacts on host physiology.

In conclusion, Ma-LMM01 employs an infection program in the apparently stable host transcriptional levels, and uses the host core σ factor SigA while avoiding host defense systems. This type of infection is a novel example of adaptation based on host defense systems to ensure efficient viral reproduction, and differs from that seen in other marine T4-like cyanophages. Future work is needed to explore whether other types of cyanophage infecting *M. aeruginosa* show the same transcriptome dynamics and infection program.

Chapter 3

Co-occurrence of broad and narrow host-range viruses infecting the toxic bloom-forming cyanobacterium *Microcystis aeruginosa*

Abstract

Viruses play important roles in regulating the abundance and composition of bacterial populations in aquatic ecosystems. The toxic bloom-forming cyanobacterium *Microcystis aeruginosa* interacts with diverse cyanoviruses, resulting in their population diversification. However, current knowledge of the genomes from these viruses and their infection programs are limited to a sole *Microcystis* virus Ma-LMM01. Here, I investigated the genomic information and transcriptional dynamics of *Microcystis*-interacting viruses using metagenomic and metatranscriptomic approaches. I identified three novel phylogenetic viral groups: Group I (including Ma-LMM01), II (high abundance and transcriptional activity), and III (new lineages). The Group II viruses interacted with all three phylogenetically distinct *Microcystis* population types (phylotypes), whereas the Group I and III viruses interacted with only one or two phylotypes, indicating the co-occurrence of broad (Group II) and narrow (Group I and III) host-range viruses in the bloom. These viruses showed peak transcriptional levels during daytime regardless of their genomic differences. Interestingly, *M. aeruginosa* expressed antiviral defense genes against viral infection, unlike that seen with a Ma-LMM01 infection in a previous culture experiment. Given that broad host-range viruses often induce antiviral responses within alternative hosts, these findings suggest that Group II viruses are major drivers for the diversification of *Microcystis* populations.

3. Broad and narrow host range *Microcystis* viruses

Introduction

Microcystis aeruginosa is one of the most pervasive bloom-forming cyanobacteria found in freshwater ecosystems throughout the world (1). Some strains of this species produce the potent hepatotoxin called microcystin (119–121), and the persistent blooms caused by such strains pose serious problems for humans who use water resources from such impaired sources for drinking water, recreational activities, and fisheries (1, 122). It is important, therefore, to identify and characterize the environmental and biological factors that affect *Microcystis* blooms.

This cyanobacterium possesses the highest number of putative antiviral defense systems among all prokaryote or archeal species examined as of 2011 (64, 123). Of antiviral defense systems, CRISPR-Cas systems, which are composed of short direct repeats separated by unique sequences (spacers) and the CRISPR-associated genes, incorporate foreign DNA fragments such as viruses into the leader-end of the CRISPR loci as a new spacer (59). Thus, the spacer arrangement (CRISPR array) provides a sequence memory of the invasion of exogenous genetic elements like viruses, or provides a targeted defense against subsequent invasion by the corresponding invaders (59). Recent studies on the *Microcystis* CRISPR array revealed that this cyanobacterium has been challenged by diverse communities of cyanoviruses in natural environments (60, 61, 63), suggesting that *Microcystis* cyanoviruses are one of the important factors that determine bloom dynamics and termination. However, to date, the only isolated *Microcystis* viruses are Ma-LMM01 (50) and MaMV-DC (57) (*Microcystis* virus Ma-LMM01 according to the International Committee on Taxonomy of Viruses (124)) although the establishment of laboratory virus-host systems would invaluably augment our understanding of the interactions occurring between *M. aeruginosa* and the viruses

targeting it.

Microcystis virus Ma-LMM01 and MaMV-DC are *Myoviridae* family members with very narrow host ranges, and they are known only to infect *M. aeruginosa* strains NIES-298 (50) and FACHB-524 (57) among the tested strains, respectively. They are also phylogenetically distinct from other known T4-like marine cyanoviruses (50, 57). Ma-LMM01 lacks the viral-encoded proteins that are important for switching from host to viral transcription and that are needed to acquire the precursors for replication and virion morphogenesis, as seen in T4-like marine cyanoviruses (43, 49, 73). In chapter 2, I revealed that Ma-LMM01 achieves three temporal expression classes comprising early, middle and late periods without affecting the host's physiology to escape host defense systems while maintaining host photosynthesis (125). Interestingly, Ma-LMM01-matching spacers are present in very low abundances in natural *Microcystis* populations (10/995 spacers) (61), suggesting that numerous uncharacterized cyanoviruses exist that could affect the bloom dynamics and termination process in *Microcystis*.

Advances in next-generation sequencing and sequence assembly techniques have allowed us to access viral metagenomics (virome) data to assess the genome content and architecture of uncultured viruses. This data offer the possibility to gain unique insights into the main viral families in the environment. Indeed, recent studies have revealed the viral diversity, viral habitat distribution, numerous uncharacterized viral lineages, and the viral-host interactions that occur in nature (126–128).

To date, metatranscriptomic studies have been conducted in freshwater environments during *Microcystis* blooming (56, 129–131). These studies have only focused on *Microcystis* (and in some case other bacterial) gene expression profiles in environmental samples. However, few studies have been conducted on the gene

expression profiles of *Microcystis* viruses in the environment (100). A previous study revealed that the abundance of tail sheath gene *gp091* mRNA from Ma-LMM01 within host cells reached peak levels during the daytime, falling to the lowest levels at midnight (100). Despite the importance of cyanoviruses in *Microcystis* bloom dynamics, no comprehensive analyses have been done to test for the existence of other *Microcystis* viruses, or to examine the whole transcriptional dynamics of both *M. aeruginosa* and its viruses in the environment.

Consequently, in this study, I identified uncharacterized *Microcystis* cyanoviruses using metagenomic approaches, and then investigated their transcriptional dynamics in the environment to better understand their impact on *Microcystis* blooms.

Materials and Methods

Sampling

Samples were prepared from a time-series of nine freshwater samples from the surface at an offshore point at Hirosawanoike Pond (Kyoto, Japan; 35° 026′ N 135° 690′ E) every 3 h over a 24 h period on October 19 and 20, 2017. On these days, the sunrise and sunset times were 06:06 and 17:18 (October 19) and 06:07 and 17:16 (October 20). For the metatranscriptomic analysis, samples (100 mL) were collected on 3.0-μm pore-size polytetrafluoroethylene membrane filters (Millipore, Billerica, MA, USA), resuspended in 5 mL of stop solution (phenol: ethanol, 5: 95 v/v) for transport, and stored at −80°C (132). At each time point, these procedures were accomplished within 20 min to prevent transcriptional shifts and RNA degradation (87). For the virome analysis, the 1 L of freshwater collected at 18:00 on October 19 and 06:00 on October 20 was prefiltered through 142 mm 3.0-μm pore-size polycarbonate membrane filters (Millipore),

and then filtered sequentially through 0.22-μm pore-size Sterivex filters (Millipore). The filtrates were stored at 4°C prior to treatment. The viruses in the filtrates were concentrated after incubation in 10% polyethylene glycol 8000–1 M NaCl, purified using a CsCl density centrifugation step, and stored at −80°C (49). The freshwater samples (100 mL) and the filtrates (viral fraction; 1 or 15 mL) were collected at each time point to quantify the numbers of *Microcystis* cells, viral particles and *gp091* abundance. Viral particle densities were measured using epifluorescence microscopy (Nikon ECLIPSE E800; Nikon, Tokyo, Japan) with SYBR Gold staining (Molecular Probes, Eugene, OR, USA). The abundances of *Microcystis* cells (PC-IGS) and Ma-LMM01 particles (*gp091*) were determined using a previously described quantitative polymerase chain reaction analysis method (100).

DNA/RNA extraction and sequencing

For the virome analysis, DNA extraction from the two purified viral fractions was performed using the previously described xanthogenate–SDS method (88). For the metatranscriptomic analysis, total RNA was extracted from 1 mL of the stored samples as described previously (85). DNA was removed using TURBO DNase (Ambion, Austin, TX, USA). The Ribo-Zero rRNA removal kit (Bacteria) (Illumina, San Diego, CA, United States) was used to deplete the ribosomal RNA. Genomic DNA and ribosomal RNA depletion were checked using a reverse transcription-PCR assay with DNA-depleted RNA samples served as the non-reverse transcribed controls and Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA), respectively. The rRNA-depleted RNA was purified using Agencourt RNAClean XP beads (Beckman Coulter Genomics, Danvers, MA, USA), and then converted to double-stranded cDNA using the PrimeScript Double Stranded cDNA Synthesis Kit (TaKaRa Bio, Otsu, Japan). Virome

and cDNA libraries were prepared using the Nextera XT DNA Sample Preparation Kit (Illumina), and then sequenced using the MiSeq Reagent Kit v3 (2 × 300-bp read lengths for the virome and 2 × 75-bp read lengths for the cDNA; Illumina) on the Illumina MiSeq platform.

Virome analysis

Total reads from the samples collected at 18:00 on October 19 and 06:00 on October 20 were co-assembled using SPAdes 3.9.1 with default k-mer lengths (**Table 3-1a**) (90). Decontamination of the prokaryotic sequences was performed using VirSorter 1.0.3 (133). To identify the putative viral contigs derived from *Microcystis* viruses, I performed a host prediction of viral contigs (≥ 10 kb) using the CRISPR-Cas system, viral tRNA matches (128) and hexanucleotide frequency similarity (134). For host assignment using the CRISPR-Cas system, a nonredundant CRISPR-Cas spacer database of 3881 sequences was created from the published data (60, 61, 63, 135) and spacer sequences from the available genomes of 29 *M. aeruginosa* isolates using the CRISPR Recognition Tool (136). All nonredundant spacers were queried against all the viral contigs I obtained in this study using the BLASTn-short function from the BLAST+ package with the following parameters: an e-value threshold of 1.0×10^{-10} and a percentage identity of 95% (128). For the host assignment using viral tRNA matches, tRNA identification from the 29 *M. aeruginosa* isolates and two complete environmental genomes (135) was performed using ARAGORN v1.2.38 and the '-t' option (137). A nonredundant tRNA database of 129 sequences was also queried against all the viral contigs using the BLASTn-short function from the BLAST+ package with the following parameter: 100% length and sequence identity (128). The calculation of hexanucleotide frequency similarity between the viral and host sequences was conducted using VirHostMatcher

(dissimilarity score threshold ≤ 0.15) (134). The proteomic tree, gene annotations and genomic alignment views were constructed using the ViPTree server (138). To further determine whether shared proteins among other known viruses appeared on the viral contigs identified in this study, translated ORFs from each viral contig were searched against the hidden Markov model profiles downloaded from the prokaryotic Virus Orthologous Groups (pVOGs) database, which provides 9,518 orthologous groups shared among bacterial and archaeal viruses (139) using hmmscan (140) with an e-value threshold of 1.0×10^{-3}. The genome-wide similarity score (S_G) cutoff for clustering was set to ≥ 0.15 (viral genus-level cutoff) according to a previous study (127). Maximum-likelihood (ML) analysis of the terminase large subunit (TerL), ribonucleotide reductase α and β subunit genes, and internal transcribed spacers (ITSs) was performed using the MEGA package (141). Bootstrap resamplings were conducted for 100 replications in the ML analysis. Viral reads derived from samples at 18:00 and 06:00 were aligned separately to the *Microcystis* viral genomes using bowtie2 (92) with the option "-score-min L,0,-0.3" after quality-control screening.

***M. aeruginosa* 11-30S32 genome sequencing and analysis**

Genomic DNA extraction from *M. aeruginosa* 11-30S32 was performed using the xanthogenate-SDS method and phenol/chloroform/isoamyl alcohol procedures in combination, as described previously (88, 89). A paired-end library was then prepared using the Nextera XT DNA Sample Prep Kit (Illumina) according to the manufacturer's instructions. The paired-end library was sequenced using the MiSeq Reagent Kit v3 (2 × 150-bp read length; Illumina) and the Illumina MiSeq platform, and assembled using SPAdes ver.3.7.0 (90). Open reading frames (ORFs) were predicted using GenemarkS (91), and the predicted ORFs were annotated by blastp analysis against the National

Center for Biotechnology Information (NCBI) non-redundant database (nr) (*E*-value threshold of $1e^{-3}$).

Metatranscriptomic data processing

Reads from each library were aligned separately to the reference genomes using bowtie2 (92) with the option "-score-min L,0,-0.3". Eukaryotic, archeal, and bacterial rRNA reads were removed from the total reads prior to read mapping using the SortmeRNA 2.1 software package (142). On average, 1.6 million reads were recovered from cDNA libraries at each time point (**Table 3-1b**). Reads from the whole transcriptome library were counted for each genome or gene. *Microcystis* and viral transcript counts were normalized as FPKM (fragments per kilobase of exon per million mapped reads) and the read counts from *M. aeruginosa* NIES-843 *rnpB* (as a proxy for host cell density) (100). The heat map analyses for the *Microcystis* toxin–antitoxin gene expression profiles were conducted using the heatmap.2 function in the R/Bioconductor package "gplots".

Recruitment of Lake Erie metatranscriptomic data for *Microcystis* viruses

I collected seven metatranscriptomic data from early, middle, and late blooms in western Lake Erie in North America (130). Considering that the average nucleotide identity is 86.1% between Ma-LMM01 and MaMV-DC (58), single-end reads were aligned separately to the reference genomes using bowtie2 (92) with the option "-score-min L,0,-0.9".

Public data

The *M. aeruginosa* 11-30S32 genome sequence was deposited with the DNA Data Bank of Japan (DDBJ) Mass Submission System (MSS) under the accession numbers BHVU01000001–BHVU01000814. The *Microcystis* viral contigs assembled from the virome reads were deposited with DDBJ Nucleotide Sequence Submission

System under accession numbers LC425512–LC425526. The virome and metatranscriptomic data were deposited in the DDBJ Sequence Read Archive under accession numbers DRR151114–DRR151124.

Results and Discussion

Identification and characterization of novel *Microcystis* viruses

To identify uncharacterized *Microcystis* viruses in the environment, I first sequenced two virome samples collected from Hirosawanoike Pond at 18:00 on October 19 and 06:00 on October 20, 2017. The PC-IGS gene copy numbers of *M. aeruginosa* fluctuated from 4.2×10^5–1.7×10^6 copies/mL during sampling (**Figure 3-1a**). Sequence reads from each virome library were co-assembled into 960 contigs (≥ 10 kb) (**Table 3-1a**). No *Microcystis* viral contigs were identified using viral tRNA matches or hexanucleotide frequency similarity as the search criteria. However, 15 viral contigs possessing *Microcystis* protospacers were identified using the CRISPR spacer-based host prediction (**Table 3-2**). Of these viral contigs, one contig, NODE34, possessed only one protospacer sequence (**Table 3-2**), and the interaction with *M. aeruginosa* was supported by genome similarity. This viral contig showed high sequence similarity to that of the isolated *Microcystis* virus Ma-LMM01 (S_G = 0.94, where the S_G value is 1 when two genomes in a comparison are identical and decreases to 0 when sequence similarity is not detected by tblastx (127)). In contrast, 13 viral contigs possessed multiple protospacer sequences (up to 23) (**Table 3-2**), which was strongly indicative of an interaction with *M. aeruginosa*. The remaining NODE982 viral contig, which possessed only one protospacer, was close to NODE656 (S_G = 0.56; ≥ 0.2937, threshold for infecting host organisms that are evolutionarily related at the genus level (127)), which possessed three protospacers

(**Table 3-2**). Hereafter, I refer to these 15 viral contigs as *Microcystis* viral genomes (MVGs).

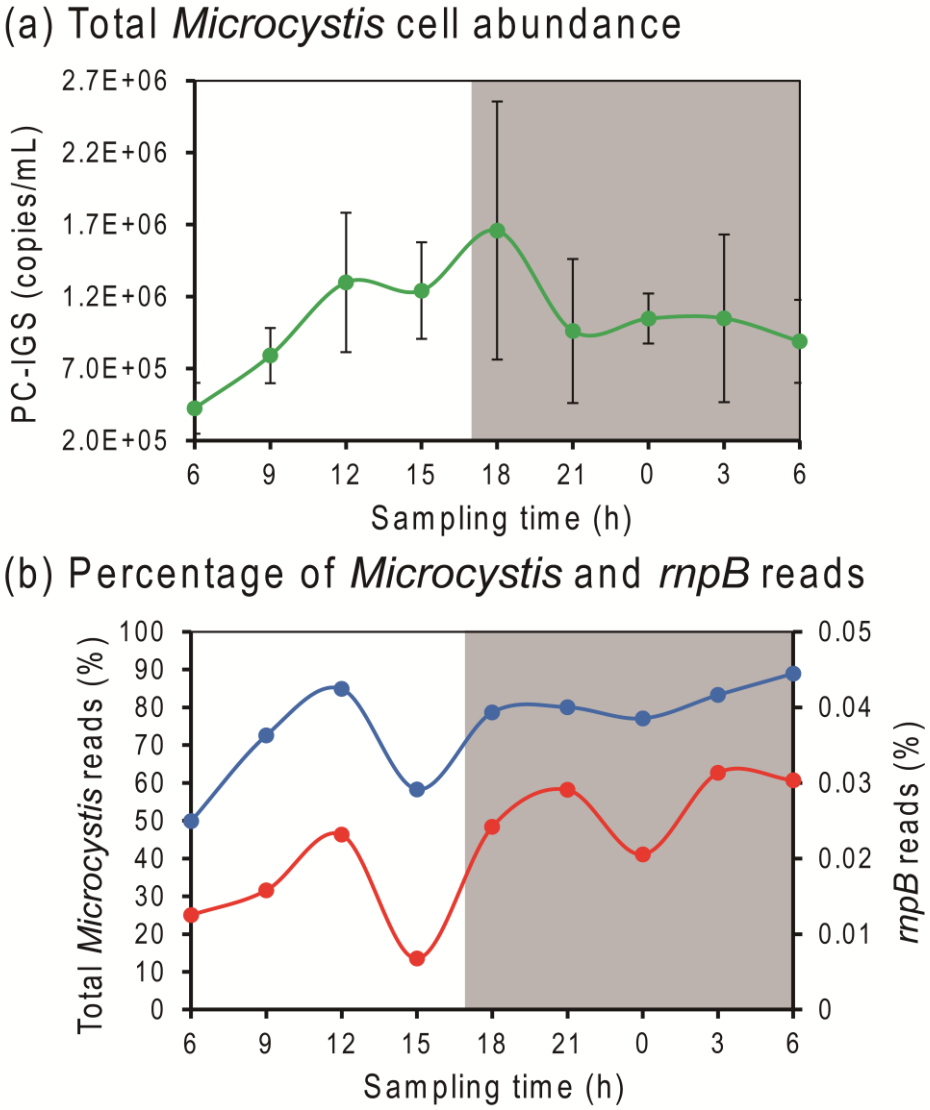


Figure 3-1. Sequence biases relating to the total *Microcystis* cell abundance at each sampling time point. (a) Total *Microcystis* cell abundance was determined by quantitative polymerase chain reaction analysis. (b) Total *Microcystis* and *rnpB* read percentages are shown in blue and red, respectively. Shaded areas indicate the periods of darkness.

Table 3-1. Summary of the sequencing data generated in this study.

(a) Virome analysis

Total Paired reads	No. of contigs (≥10 kb)	Host prediction		
		spacer	tRNA	HNF
10,799,129	960	15	0	0

(b) Metatranscriptome analysis

		6	9	12	15	18	21	0	3	6
Paired reads		2,032,721	1,841,909	2,114,583	1,952,772	2,261,942	2,556,467	2,895,684	2,786,238	2,652,467
Q30 filtered		1,901,924	1,727,679	1,971,646	1,833,019	2,119,398	2,421,849	2,745,471	2,638,588	2,508,493
rRNA removed		1,286,580	1,064,152	540,139	1,109,518	1,149,867	2,059,964	2,376,395	2,353,558	2,350,183
Mapped reads	Host	641,622	772,501	458,460	646,746	904,494	1,648,724	1,831,685	1,960,147	2,090,628
	Virus	6,714	3,163	1,950	4,163	6,631	8,833	16,631	18,712	10,504

Table 3-2. Summary of the spacer sequences identical to viral contigs assembled in this study.

Group	contig_id (≧10 kb)	CRISPR type (CT)	identity (%)	length	mismatch	gap	qstart	qend	sstart	send	e-value	bitscore
Group I	MVG_NODE34	DA14_genomic.fna_CRISPR.157	100	37	0	0	1	37	288	252	7.49E-13	69.4
	MVG_NODE47	NIES98_genomic.fna_CRISPR.172	100	34	0	0	1	34	42328	42295	3.07E-11	63.9
		NaRes975_genomic.fna_CRISPR.20	97.6	42	1	0	1	42	16601	16642	7.15E-14	73.1
		PCC9807_genomic.fna_CRISPR.20	100	40	0	0	1	40	24996	24957	1.89E-14	75
	MVG_NODE620	CT73-4	100	33	0	0	1	33	6761	6793	9.58E-11	62.1
		CT100-1	100	34	0	0	1	34	1154	1121	2.87E-11	63.9
		LE3_genomic.fna_CRISPR.75	100	37	0	0	1	37	1607	1571	7.49E-13	69.4
		PCC7005_genomic.fna_CRISPR.3	100	35	0	0	1	35	9700	9666	8.55E-12	65.8
		SPC777_genomic.fna_CRISPR.70	100	36	0	0	1	36	6098	6063	2.53E-12	67.6
		SPC777_genomic.fna_CRISPR.88	100	37	0	0	1	37	9550	9514	8.81E-13	69.4
	MVG_NODE869	NIES298_genomic.fna_CRISPR.70	97.4	38	1	0	1	38	7015	6978	1.03E-11	65.8
		PCC7005_genomic.fna_CRISPR.3	100	35	0	0	1	35	7205	7171	8.55E-12	65.8
		SPC777_genomic.fna_CRISPR.70	100	36	0	0	1	36	3291	3256	2.53E-12	67.6
Group II	MVG_NODE331	NIES298_genomic.fna_CRISPR.42	100	33	0	0	1	33	11053	11085	9.58E-11	62.1
		NIES298_genomic.fna_CRISPR.79	100	37	0	0	1	37	18423	18459	7.49E-13	69.4
		NIES298_genomic.fna_CRISPR.82	100	35	0	0	1	35	478	512	8.55E-12	65.8
		NIES298_genomic.fna_CRISPR.86	100	35	0	0	1	35	2605	2571	8.55E-12	65.8
		NIES843_genomic.fna_CRISPR.129	100	36	0	0	1	36	9331	9296	2.53E-12	67.6
		NIES843_genomic.fna_CRISPR.180	100	36	0	0	1	36	6504	6469	2.53E-12	67.6
		NIES98_genomic.fna_CRISPR.175	100	35	0	0	1	35	14820	14786	8.55E-12	65.8
		NIES98_genomic.fna_CRISPR.200	100	37	0	0	1	37	4758	4794	7.49E-13	69.4
		PCC7941_genomic.fna_CRISPR.74	100	34	0	0	3	36	2264	2231	3.28E-11	63.9
		PCC9701_genomic.fna_CRISPR.119	100	34	0	0	1	34	18768	18735	2.87E-11	63.9
		PCC9701_genomic.fna_CRISPR.140	100	34	0	0	1	34	14580	14547	2.87E-11	63.9
		PCC9809_genomic.fna_CRISPR.34	100	34	0	0	2	35	2786	2753	3.07E-11	63.9
	MVG_NODE375	CT48-1	100	34	0	0	1	34	4692	4659	2.87E-11	63.9
		CT48-4	100	37	0	0	1	37	5865	5901	7.49E-13	69.4
		CACIAM03_genomic.fna_CRISPR.88	100	41	0	0	1	41	7312	7352	5.53E-15	76.8
		LE3_genomic.fna_CRISPR.47	100	35	0	0	1	35	3918	3884	8.55E-12	65.8
		LE3_genomic.fna_CRISPR.57	100	35	0	0	1	35	557	523	8.55E-12	65.8
		LE3_genomic.fna_CRISPR.101	97.3	37	1	0	1	37	17679	17643	4.10E-11	63.9
		NIES298_genomic.fna_CRISPR.59	100	34	0	0	1	34	18711	18678	2.87E-11	63.9
		NIES298_genomic.fna_CRISPR.162	100	35	0	0	1	35	1563	1597	8.55E-12	65.8
		NIES298_genomic.fna_CRISPR.193	97.6	42	1	0	1	42	3499	3458	7.15E-14	73.1
		NIES298_genomic.fna_CRISPR.194	100	35	0	0	1	35	5500	5466	1.03E-11	65.8
		NIES298_genomic.fna_CRISPR.199	100	36	0	0	1	36	5976	5941	2.53E-12	67.6
		NIES298_genomic.fna_CRISPR.202	97.4	38	1	0	1	38	5106	5069	1.20E-11	65.8
		NIES44_genomic.fna_CRISPR.46	100	35	0	0	1	35	1228	1262	8.55E-12	65.8
		NIES843_genomic.fna_CRISPR.62	100	34	0	0	1	34	19612	19645	2.87E-11	63.9
		NIES843_genomic.fna_CRISPR.70	100	36	0	0	1	36	6291	6326	2.53E-12	67.6
		NIES98_genomic.fna_CRISPR.15	100	39	0	0	1	39	7293	7331	7.15E-14	73.1
		NIES98_genomic.fna_CRISPR.282	97.6	42	1	0	1	42	10794	10835	7.15E-14	73.1
		PCC7941_genomic.fna_CRISPR.84	100	35	0	0	1	35	8049	8015	8.55E-12	65.8
		PCC9701_genomic.fna_CRISPR.45	100	44	0	0	1	44	3517	3474	1.30E-16	82.4
		PCC9701_genomic.fna_CRISPR.54	100	35	0	0	1	35	2373	2407	8.55E-12	65.8
		TAIHU98_genomic.fna_CRISPR.39	97.4	38	1	0	3	40	5069	5106	1.14E-11	65.8
		DA14_genomic.fna_CRISPR.49	100	36	0	0	1	36	4963	4998	2.53E-12	67.6
		TA09_genomic.fna_CRISPR.13	100	35	0	0	1	35	17048	17082	8.55E-12	65.8
	MVG_NODE382	CT99-2	100	39	0	0	1	39	16539	16501	6.47E-14	73.1
		DIANCHI905_genomic.fna_CRISPR.59	97.3	37	1	0	3	39	10408	10372	3.89E-11	63.9
		NIES843_genomic.fna_CRISPR.91	100	34	0	0	1	34	10980	10947	2.87E-11	63.9
		NIES98_genomic.fna_CRISPR.21	100	34	0	0	1	34	5253	5220	2.87E-11	63.9

Table 3-2. continued.

Group	contig_id (≧10 kb)	CRISPR type (CT)	identity (%)	length	mismatch	gap	qstart	qend	sstart	send	e-value	bitscore
Group II	MVG_NODE382	NIES98_genomic.fna_CRISPR.37	100	37	0	0	4	40	13059	13023	8.81E-13	69.4
		NIES98_genomic.fna_CRISPR.50	100	39	0	0	2	40	15082	15120	7.15E-14	73.1
		NIES98_genomic.fna_CRISPR.66	100	34	0	0	1	34	1978	2011	2.87E-11	63.9
		NIES98_genomic.fna_CRISPR.213	100	35	0	0	1	35	3112	3146	9.12E-12	65.8
		NIES98_genomic.fna_CRISPR.215	100	35	0	0	1	35	7575	7609	8.55E-12	65.8
		NIES98_genomic.fna_CRISPR.223	100	35	0	0	1	35	7710	7744	8.55E-12	65.8
		NIES98_genomic.fna_CRISPR.241	100	36	0	0	1	36	8335	8370	2.53E-12	67.6
		NaRes975_genomic.fna_CRISPR.217	100	35	0	0	1	35	19914	19880	8.55E-12	65.8
		PCC7941_genomic.fna_CRISPR.120	100	36	0	0	1	36	1061	1096	2.53E-12	67.6
		PCC9443_genomic.fna_CRISPR.197	100	33	0	0	1	33	2683	2651	9.58E-11	62.1
		PCC9717_genomic.fna_CRISPR.51	100	35	0	0	1	35	12175	12209	8.55E-12	65.8
		PCC9717_genomic.fna_CRISPR.52	100	34	0	0	1	34	1879	1846	2.87E-11	63.9
		PCC9717_genomic.fna_CRISPR.71	100	40	0	0	1	40	11874	11913	1.89E-14	75
		PCC9717_genomic.fna_CRISPR.136	100	34	0	0	1	34	7537	7570	2.87E-11	63.9
		PCC9717_genomic.fna_CRISPR.228	100	37	0	0	1	37	13265	13229	7.49E-13	69.4
		SPC777_genomic.fna_CRISPR.124	100	35	0	0	1	35	1059	1093	8.55E-12	65.8
		SPC777_genomic.fna_CRISPR.125	100	37	0	0	1	37	3117	3081	7.49E-13	69.4
	MVG_NODE385	CT39-3	100	35	0	0	1	35	5112	5146	8.55E-12	65.8
		CT62-3	100	35	0	0	1	35	15902	15868	8.55E-12	65.8
		CT67-2	100	34	0	0	1	34	17220	17187	2.87E-11	63.9
		CT67-3	100	34	0	0	1	34	6496	6463	2.87E-11	63.9
		CT79-3	100	34	0	0	1	34	15982	16015	2.87E-11	63.9
		CT103-2	100	36	0	0	1	36	10359	10324	2.53E-12	67.6
		CHAOHU1326_genomic.fna_CRISPR.49	97.4	38	1	0	1	38	3059	3022	1.03E-11	65.8
		NIES298_genomic.fna_CRISPR.75	97.4	38	1	0	1	38	4091	4128	1.03E-11	65.8
		NIES98_genomic.fna_CRISPR.9	100	38	0	0	2	39	13634	13597	2.33E-13	71.3
		NIES98_genomic.fna_CRISPR.12	100	43	0	0	2	44	1124	1166	4.68E-16	80.5
		NIES98_genomic.fna_CRISPR.185	100	38	0	0	1	38	15638	15601	2.20E-13	71.3
		NaRes975_genomic.fna_CRISPR.158	100	34	0	0	1	34	6597	6564	2.87E-11	63.9
		NaRes975_genomic.fna_CRISPR.161	98.1	52	1	0	1	52	11496	11547	2.91E-19	91.6
		NaRes975_genomic.fna_CRISPR.165	97.3	37	1	0	1	37	7148	7184	3.48E-11	63.9
		PCC9443_genomic.fna_CRISPR.472	100	34	0	0	3	36	6595	6628	3.28E-11	63.9
		PCC9701_genomic.fna_CRISPR.150	100	33	0	0	1	33	7971	7939	9.58E-11	62.1
		PCC9717_genomic.fna_CRISPR.226	100	42	0	0	1	42	10672	10631	1.54E-15	78.7
		DA14_genomic.fna_CRISPR.22	100	39	0	0	5	43	9977	10015	8.51E-14	73.1
Group III	MVG_NODE378	CT57-1	100	34	0	0	1	34	15782	15815	2.87E-11	63.9
		CT66-3	100	34	0	0	3	36	11854	11821	3.28E-11	63.9
		NIES98_genomic.fna_CRISPR.157	100	34	0	0	1	34	932	899	2.87E-11	63.9
	MVG_NODE562	PCC9443_genomic.fna_CRISPR.95	97.3	37	1	0	1	37	6624	6660	3.89E-11	63.9
		PCC9717_genomic.fna_CRISPR.162	100	33	0	0	1	33	6892	6924	9.58E-11	62.1
		DA14_genomic.fna_CRISPR.153	100	35	0	0	1	35	8161	8127	8.55E-12	65.8
		DA14_genomic.fna_CRISPR.177	97.6	42	1	0	1	42	12742	12783	7.15E-14	73.1
	MVG_NODE577	CT46-1	100	35	0	0	1	35	674	640	8.55E-12	65.8
		LE3_genomic.fna_CRISPR.13	97.6	41	1	0	1	41	403	363	2.57E-13	71.3
		PCC9443_genomic.fna_CRISPR.108	100	34	0	0	1	34	14638	14605	3.07E-11	63.9
		PCC9717_genomic.fna_CRISPR.163	100	35	0	0	1	35	6143	6177	8.55E-12	65.8
		SPC777_genomic.fna_CRISPR.79	100	36	0	0	3	38	14312	14347	2.85E-12	67.6
	MVG_NODE636	CT63-4	97.4	38	1	0	1	38	2585	2548	1.03E-11	65.8
		CT84-2	100	35	0	0	1	35	11672	11706	8.55E-12	65.8
		NIES298_genomic.fna_CRISPR.68	100	34	0	0	1	34	2118	2151	2.87E-11	63.9
		NIES298_genomic.fna_CRISPR.112	100	33	0	0	1	33	1437	1405	9.58E-11	62.1
		NIES98_genomic.fna_CRISPR.138	100	34	0	0	1	34	3248	3215	2.87E-11	63.9
		NIES98_genomic.fna_CRISPR.139	100	34	0	0	1	34	4060	4027	2.87E-11	63.9
	MVG_NODE656	PCC9443_genomic.fna_CRISPR.220	100	35	0	0	1	35	8625	8591	8.55E-12	65.8
		PCC9807_genomic.fna_CRISPR.87	100	34	0	0	3	36	7657	7690	3.28E-11	63.9
		DA14_genomic.fna_CRISPR.130	100	34	0	0	1	34	5574	5541	2.87E-11	63.9
	MVG_NODE671	CT57-4	100	35	0	0	1	35	2118	2152	8.55E-12	65.8
		PCC7941_genomic.fna_CRISPR.26	97.6	42	1	0	1	42	389	430	7.15E-14	73.1
	MVG_NODE982	NIES843_genomic.fna_CRISPR.29	100	34	0	0	2	35	4237	4270	3.07E-11	63.9

The fifteen MVGs referred to above were largely divisible into three groups; namely, Group I (MVG_NODE34, NODE47, NODE620 and NODE869), Group II (MVG_NODE331, NODE375, NODE382 and NODE385) and Group III (MVG_NODE378, NODE562, NODE577, NODE636, NODE656, NODE671 and NODE982) based on their genome similarities using ViPTree (**Figure 3-2**) (138).

The four viral Group I contig sizes ranged from 10,954 to 48,147 bp and contained 21–49 predicted protein-coding genes (**Table 3-3**). According to sequence similarity, MVG_NODE34 (S_G = 0.94) and NODE47 (S_G = 0.84) were derived from close relatives of *Microcystis* virus Ma-LMM01 and had 28.5–29.7% and 25.3–26.4% of the genome length of *Microcystis* virus Ma-LMM01, respectively (**Figure 3-3, Table 3-3**) (49, 50, 57, 58). These two viral contigs contained the middle genes required for DNA replication, recombination/repair and nucleotide metabolism, in addition to viral structural genes (125). Reconstruction of these Ma-LMM01-related MVGs from the virome reads corroborated the quality of the sequence assembly and host prediction. In addition, the other two Group I MVGs, which were similar each other (MVG_NODE620 and NODE869; S_G = 0.80), shared three protein homologs (thymidylate synthase, ribonucleotide reductase α and β subunit) with Ma-LMM01 with 40–50% identity (**Figure 3-3, Table 3-3**). The low similarity to *Microcystis* virus Ma-LMM01 (S_G = 0.11 for MVG_NODE620; 0.14 for NODE869) revealed that MVG_NODE620 and NODE869 were derived from novel *Microcystis* viruses in the environment. This finding was also supported by the phylogenetic trees for ribonucleotide reductase α and β subunits (**Figure 3-4**).

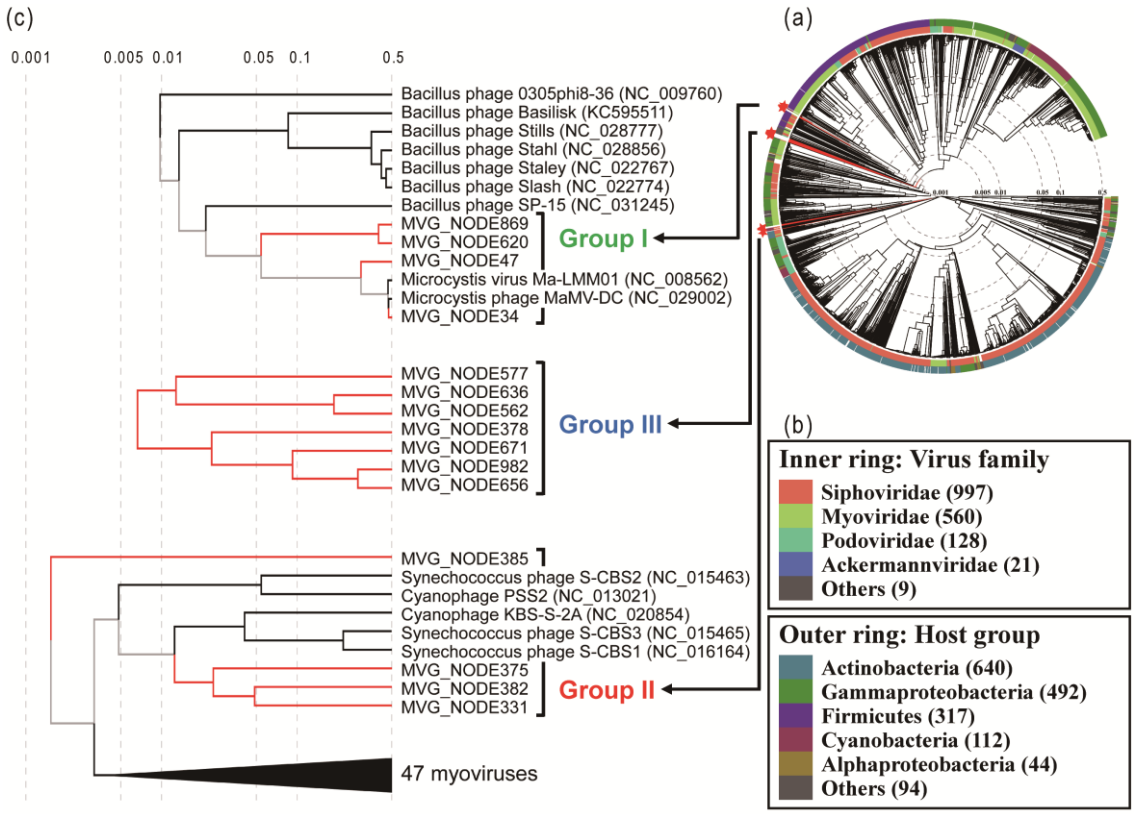


Figure 3-2. Proteomic tree of 15 *Microcystis* viral genomes identified in this study and 1730 prokaryotic dsDNA viruses. (a) Whole proteomic tree including the related prokaryotic dsDNA viruses (1730 reference genomes) generated by ViPTree server ver.1.5. The dendrogram represents the proteome-wide similarity relationships between 15 *Microcystis* viral genomes (MVGs) and reference viral genomes (RVGs). Branches are colored red (MVGs) or black (RVGs), and branch lengths are logarithmically scaled from the root of the entire proteomic tree. (b) Rings outside the dendrogram, from inside to outside, represent viral family classifications and taxonomic groups of known hosts, respectively. (c) Enlarged view of the proteomic tree including areas of the MVGs and two reference genomes of *Microcystis* viruses. Seven *Bacillus* viruses that show weak genomic similarity to Group I viral genomes and five cyanoviruses and 47 myoviruses that form clades with Group II viruses are also shown.

Table 3-3. Gene annotation of 15 *Microcystis* viral genomes identified in this study.

contig_id	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE34	MVG_NODE34_1	99	203	2	0	47	249	1	203	1.30E-109	403	[YP_851176] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_2	98.2	392	7	0	1	392	1	392	2.40E-218	766	[YP_009217831] putative helicase [Microcystis phage MaMV-DC]
	MVG_NODE34_3	95.7	93	4	0	1	93	1	93	1.10E-47	196	[YP_009217830] hypothetical protein MaMVDC_146 [Microcystis phage MaMV-DC]
	MVG_NODE34_4	87	108	12	1	1	108	1	106	8.30E-45	187	[YP_009217829] hypothetical protein MaMVDC_145 [Microcystis phage MaMV-DC]
	MVG_NODE34_5	78	132	25	2	1	131	1	129	9.10E-52	210	[YP_009217828] hypothetical protein MaMVDC_144 [Microcystis phage MaMV-DC]
	MVG_NODE34_6	92.8	69	5	0	1	69	1	69	1.20E-30	140	[YP_009217826] hypothetical protein MaMVDC_142 [Microcystis phage MaMV-DC]
	MVG_NODE34_7	96.5	57	2	0	1	57	1	57	6.30E-24	117	[YP_851166] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_9	87.3	63	4	1	1	63	1	59	1.90E-23	116	[YP_851163] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_10	93.9	99	6	0	1	99	1	99	1.60E-51	209	[YP_009217821] hypothetical protein MaMVDC_137 [Microcystis phage MaMV-DC]
	MVG_NODE34_11	82.5	63	10	1	1	62	1	63	2.20E-21	109	[YP_009217819] hypothetical protein MaMVDC_135 [Microcystis phage MaMV-DC]
	MVG_NODE34_12	52	98	21	1	4	75	2	99	2.20E-21	109	[YP_009217821] hypothetical protein MaMVDC_137 [Microcystis phage MaMV-DC]
	MVG_NODE34_13	43.7	71	25	3	1	69	11	68	6.60E-05	54	[YP_851155] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_14	95.2	63	3	0	1	63	1	63	3.60E-27	128	[YP_851151] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_15	98.5	65	1	0	1	65	1	65	2.90E-29	135	[YP_009217817] hypothetical protein MaMVDC_133 [Microcystis phage MaMV-DC]
	MVG_NODE34_16	98	354	7	0	1	354	35	388	7.60E-202	711	[YP_009217816] DNA primase [Microcystis phage MaMV-DC]
	MVG_NODE34_17	98.2	165	3	0	1	165	2	166	9.80E-94	349	[YP_009217815] hypothetical protein MaMVDC_131 [Microcystis phage MaMV-DC]
	MVG_NODE34_18	97.2	107	3	0	1	107	7	113	1.40E-52	213	[YP_851146] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_19	97.4	266	7	0	1	266	1	266	2.60E-154	552	[YP_009217813] putative Seryl-tRNA synthetase [Microcystis phage MaMV-DC]
	MVG_NODE34_20	97	165	4	1	1	164	1	165	3.50E-83	314	[YP_009217812] hypothetical protein MaMVDC_128 [Microcystis phage MaMV-DC]
	MVG_NODE34_21	93.8	112	7	0	1	112	1	112	4.30E-57	228	[YP_851143] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_22	97.4	305	8	0	1	305	1	305	3.80E-169	602	[YP_851142] putative Fe-S oxidoreductase [Microcystis virus Ma-LMM01]
	MVG_NODE34_23	88	183	19	1	1	183	2	181	1.30E-85	322	[YP_851140] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_24	98.5	344	5	0	1	344	37	380	3.20E-197	695	[YP_009217807] hypothetical protein MaMVDC_123 [Microcystis phage MaMV-DC]
	MVG_NODE34_25	100	148	0	0	1	148	1	148	4.30E-81	307	[YP_851138] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_26	100	164	0	0	1	164	72	235	2.40E-84	318	[YP_009217805] hypothetical protein MaMVDC_121 [Microcystis phage MaMV-DC]
	MVG_NODE34_27	97.4	2867	74	0	1	2867	32	2898	0.00E+00	5447	[YP_851134] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_28	97.2	211	6	0	1	211	1	211	1.00E-115	423	[YP_851133] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_29	98.1	576	11	0	1	576	17	592	0.00E+00	1120	[YP_851132] terminase [Microcystis virus Ma-LMM01]
	MVG_NODE34_30	95.8	239	10	0	1	239	24	262	2.00E-125	456	[YP_851131] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_31	95.5	156	7	0	1	156	1	156	3.20E-76	291	[YP_009217799] hypothetical protein MaMVDC_115 [Microcystis phage MaMV-DC]
	MVG_NODE34_32	98.4	122	2	0	1	122	1	122	2.00E-62	245	[YP_009217798] hypothetical protein MaMVDC_114 [Microcystis phage MaMV-DC]
	MVG_NODE34_33	98.4	382	6	0	1	382	1	382	2.30E-210	739	[YP_009217796] hypothetical protein MaMVDC_112 [Microcystis phage MaMV-DC]
	MVG_NODE34_34	90	641	54	4	1	633	14	652	6.2e-312	1078	[YP_851126] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_35	98.2	433	8	0	1	433	94	526	1.10E-243	850	[YP_851125] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_36	95.7	94	4	0	1	94	18	111	2.70E-43	182	[YP_851124] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_37	97.1	819	24	0	1	819	22	840	0.00E+00	1564	[YP_851123] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_38	95.1	1359	67	0	1	1359	1	1359	0.00E+00	2518	[YP_009217792] putative tail protein [Microcystis phage MaMV-DC]
	MVG_NODE34_39	98.9	185	2	0	1	185	1	185	2.40E-100	371	[YP_009217791] hypothetical protein MaMVDC_107 [Microcystis phage MaMV-DC]
	MVG_NODE34_40	99.4	332	2	0	1	332	6	337	4.90E-192	678	[YP_851120] putative lysine/ornithine N-monooxygenase [Microcystis virus Ma-LMM01]
	MVG_NODE34_41	94.4	576	32	0	1	576	1	576	0.00E+00	1090	[YP_851119] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE34_42	89.7	224	23	0	1	224	1	224	1.30E-104	386	[WP_103112954] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE34_43	97.4	271	7	0	1	271	1	271	1.10E-150	540	[YP_009217787] hypothetical protein MaMVDC_103 [Microcystis phage MaMV-DC]

Table 3-3. continued.

contig_id	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE47	MVG_NODE47_1	33.6	134	86	2	51	183	51	182	1.10E-07	64	[F8TVG3_9CAUD] SubName: Full=Uncharacterized protein gp11 {ECO:0000313|EMBL:AEI71159.1}
	MVG_NODE47_2	30	130	85	2	9	135	98	224	7.40E-09	67	[WP_079498353] hypothetical protein [Burkholderia sp. YR277]
	MVG_NODE47_3	67	264	62	3	1	239	155	418	1.00E-79	304	[WP_016516706] pentapeptide repeat-containing protein [Microcystis aeruginosa]
	MVG_NODE47_4	98.9	185	2	0	1	185	43	227	3.30E-105	387	[YP_009217713] hypothetical protein MaMVDC_29 [Microcystis phage MaMV-DC]
	MVG_NODE47_5	93.1	58	4	0	1	58	53	110	1.70E-21	109	[YP_851049] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_6	81.4	59	11	0	22	80	1	59	4.40E-17	95	[YP_851047] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_7	96.6	506	17	0	1	506	1	506	6.80E-283	981	[YP_851044] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_8	97.8	183	4	0	1	183	14	196	5.10E-95	354	[YP_851043] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_9	100	158	0	0	1	158	7	164	1.20E-83	316	[YP_009217709] hypothetical protein MaMVDC_25 [Microcystis phage MaMV-DC]
	MVG_NODE47_10	98.5	66	1	0	44	109	2	67	3.40E-30	138	[YP_851040] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_11	98.2	384	7	0	1	384	23	406	8.70E-218	764	[YP_851039] serine/threonine protein phosphatase [Microcystis virus Ma-LMM01]
	MVG_NODE47_12	94.4	54	3	0	1	54	487	540	3.40E-22	112	[WP_015166972] type I restriction modification DNA specificity protein [Synechococcus sp. PCC 7502]
	MVG_NODE47_13	99.3	270	2	0	1	270	1	270	2.30E-145	522	[YP_851038] prophage antirepressor [Microcystis virus Ma-LMM01]
	MVG_NODE47_14	62.4	125	20	2	1	99	3	126	5.00E-29	134	[YP_851037] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_15	95	242	12	0	1	242	1	242	1.00E-124	453	[YP_851035] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_16	96.7	491	16	0	1	491	1	491	3.70E-278	965	[YP_851034] putative thymidylate synthase [Microcystis virus Ma-LMM01]
	MVG_NODE47_17	79.2	307	51	5	1	306	1	295	3.80E-124	452	[YP_009217703] tail collar protein [Microcystis phage MaMV-DC]
	MVG_NODE47_18	100	68	0	0	1	68	1	68	7.40E-33	147	[YP_851032] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_19	97.5	119	3	0	1	119	1	119	1.10E-57	230	[YP_009217701] hypothetical protein MaMVDC_17 [Microcystis phage MaMV-DC]
	MVG_NODE47_20	93.8	192	5	1	1	185	1	192	6.00E-99	367	[YP_009217700] hypothetical protein MaMVDC_16 [Microcystis phage MaMV-DC]
	MVG_NODE47_21	92.9	85	6	0	1	85	1	85	7.10E-36	157	[YP_009217699] hypothetical protein MaMVDC_15 [Microcystis phage MaMV-DC]
	MVG_NODE47_22	85.5	55	8	0	1	55	1	55	3.60E-19	102	[A0A2H6BYX5_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:GBF00292.1}
	MVG_NODE47_23	95.6	677	27	2	1	677	23	696	0.00E+00	1315	[YP_851028] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_24	72.9	133	36	0	1	133	1	133	1.60E-43	183	[YP_009217695] hypothetical protein MaMVDC_11 [Microcystis phage MaMV-DC]
	MVG_NODE47_25	86.8	114	8	1	1	114	26	132	8.40E-53	213	[YP_851026] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_26	96.5	173	6	0	1	173	1	173	2.20E-93	348	[YP_851024] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_27	98.8	162	2	0	1	162	1	162	6.70E-87	327	[YP_009217693] prophage antirepressor [Microcystis phage MaMV-DC]
	MVG_NODE47_28	98.9	354	4	0	1	354	1	354	3.80E-193	682	[YP_851022] recA recombinase [Microcystis virus Ma-LMM01]
	MVG_NODE47_29	95.2	145	7	0	1	145	1	145	1.00E-71	276	[YP_009217691] hypothetical protein MaMVDC_7 [Microcystis phage MaMV-DC]
	MVG_NODE47_30	46.4	1094	519	14	2	1037	31	1115	2.80E-274	953	[T2JRY2_CROWT] RecName: Full=Ribonucleoside-diphosphate reductase {ECO:0000256|RuleBase:RU003410}; EC=1.17.4.1 {ECO:0000256|RuleBase:RU0
	MVG_NODE47_31	95.3	64	3	0	1	64	1	64	5.20E-26	124	[YP_009217689] phycobilisome degradation protein NblA [Microcystis phage MaMV-DC]
	MVG_NODE47_32	91.4	186	16	0	1	186	1	186	6.20E-88	330	[YP_009217688] hypothetical protein MaMVDC_4 [Microcystis phage MaMV-DC]
	MVG_NODE47_33	98.9	264	3	0	1	264	1	264	2.80E-148	532	[YP_009217687] hypothetical protein MaMVDC_3 [Microcystis phage MaMV-DC]
	MVG_NODE47_34	99.4	348	2	0	1	348	1	348	1.30E-201	710	[YP_009217686] ribonucleoside-diphosphate reductase beta subunit [Microcystis phage MaMV-DC]
	MVG_NODE47_35	94.7	797	38	2	1	793	1	797	0.00E+00	1535	[YP_009217685] rIIA-like protein [Microcystis phage MaMV-DC]
	MVG_NODE47_36	97.8	271	6	0	1	271	1	271	1.20E-149	537	[YP_851198] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_37	99.6	223	1	0	1	223	1	223	8.10E-120	437	[YP_851197] phoH-like phosphate starvation-inducible protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_38	98.1	214	4	0	1	214	11	224	4.90E-116	424	[YP_851195] dUTPase [Microcystis virus Ma-LMM01]
	MVG_NODE47_39	93.7	427	27	0	1	427	1	427	2.40E-227	796	[YP_009217850] 3'-5' exonuclease [Microcystis phage MaMV-DC]
	MVG_NODE47_40	96.9	320	10	0	1	320	1	320	6.00E-176	624	[YP_009217849] DNA polymerase [Microcystis phage MaMV-DC]
	MVG_NODE47_41	97.8	184	4	0	1	184	1	184	3.50E-99	367	[YP_009217848] hypothetical protein MaMVDC_164 [Microcystis phage MaMV-DC]
	MVG_NODE47_42	98.8	324	4	0	1	324	44	367	6.20E-176	624	[YP_851190] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_43	94.4	124	7	0	1	124	1	124	3.20E-60	238	[YP_009217846] hypothetical protein MaMVDC_162 [Microcystis phage MaMV-DC]
	MVG_NODE47_44	92.5	107	8	0	1	107	1	107	5.90E-46	191	[WP_024969052] DUF1822 domain-containing protein [Microcystis aeruginosa]
	MVG_NODE47_45	96.9	359	11	0	2	360	20	378	2.60E-205	722	[YP_851188] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE47_46	96.4	251	9	0	1	251	1	251	1.60E-139	503	[YP_851187] uracil-DNA glycosylase [Microcystis virus Ma-LMM01]
	MVG_NODE47_47	42.3	142	76	2	50	185	11	152	7.30E-26	124	[WP_052704494] hypothetical protein [Halomonas sp. S2151]
	MVG_NODE47_48	95.9	363	15	0	1	363	1	363	2.20E-199	703	[YP_009217840] DNA polymerase III subunit gamma and tau [Microcystis phage MaMV-DC]
	MVG_NODE47_49	97.8	225	5	0	1	225	4	228	7.30E-124	450	[YP_851182] hypothetical protein [Microcystis virus Ma-LMM01]

Table 3-3. continued.

contig_id	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE620	MVG_NODE620_1	48.3	60	31	0	9	68	1	60	1.50E-09	70	[WP_019496354] XRE family transcriptional regulator [Calothrix sp. PCC 7103]
	MVG_NODE620_2	61.1	95	34	2	3	94	45	139	9.90E-22	110	[WP_103111144] PadR family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE620_3	51.4	144	65	3	5	147	2	141	6.50E-29	134	[WP_107670356] hypothetical protein [Cyanothece sp. BG0011]
	MVG_NODE620_5	46.3	82	36	1	16	89	48	129	3.80E-13	82	[WP_068072364] hypothetical protein [Rhizobiales bacterium CCH9-A3]
	MVG_NODE620_6	40.2	246	145	1	16	261	18	261	1.90E-43	184	[WP_013157695] hypothetical protein [Meiothermus silvanus]
	MVG_NODE620_9	36.7	139	85	2	35	173	50	185	9.90E-14	84	[WP_061432464] PEP-CTERM sorting domain-containing protein [Microcystis aeruginosa]
	MVG_NODE620_10	82.3	344	61	0	10	353	1	344	9.40E-168	597	[WP_103672886] ribonucleotide-diphosphate reductase subunit beta [Microcystis aeruginosa]
	MVG_NODE620_11	51.2	82	40	0	7	88	108	189	8.30E-16	91	[WP_046661967] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE620_12	71.5	752	196	6	2	750	34	770	0.00E+00	1084	[WP_008206357] MULTISPECIES: ribonucleoside-diphosphate reductase subunit alpha [Microcystis]
	MVG_NODE620_13	33.6	116	63	2	12	127	3	104	2.10E-08	66	[WP_089097346] hypothetical protein [Geobacillus galactosidasius]
	MVG_NODE620_17	61.3	507	162	5	15	512	16	497	2.40E-174	620	[WP_012594585] thymidylate synthase [Cyanothece sp. PCC 8801]
	MVG_NODE620_19	87.3	63	8	0	1	63	1	63	4.10E-23	115	[WP_002791926] MULTISPECIES: NblA protein [Microcystis]
MVG_NODE869	MVG_NODE869_2	47.3	91	39	2	7	89	40	129	3.50E-14	85	[WP_068072364] hypothetical protein [Rhizobiales bacterium CCH9-A3]
	MVG_NODE869_3	41.3	252	143	2	13	261	12	261	9.90E-45	188	[WP_013157695] hypothetical protein [Meiothermus silvanus]
	MVG_NODE869_6	34.3	172	86	6	10	170	2644	2799	1.90E-12	79	[WP_035327852] T9SS C-terminal target domain-containing protein [Dokdonia donghaensis]
	MVG_NODE869_7	78.4	37	8	0	1	37	121	157	5.80E-09	68	[WP_002786755] Similar to tr|P74732|P74732 (fragment) [Microcystis aeruginosa]
	MVG_NODE869_8	84.2	304	48	0	10	313	1	304	7.90E-149	534	[WP_103672886] ribonucleotide-diphosphate reductase subunit beta [Microcystis aeruginosa]
	MVG_NODE869_9	50	82	41	0	7	88	108	189	2.90E-16	92	[WP_046661967] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE869_13	71.8	752	194	6	2	750	29	765	0.00E+00	1083	[A0A1X9LA75_MICAE] RecName: Full=Ribonucleoside-diphosphate reductase {ECO:0000256|RuleBase:RU003410}; EC=1.17.4.1 {ECO:0000256|RuleBase:F
	MVG_NODE869_14	47.4	57	29	1	12	68	5	60	6.50E-05	54	[A0A0F9DJX4_9ZZZZ] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:KKL54136.1}
	MVG_NODE869_19	61.6	427	150	5	1	423	81	497	2.60E-149	537	[WP_096646102] thymidylate synthase [Calothrix brevissima]
MVG_NODE331	MVG_NODE331_1	46.4	138	72	1	7	144	7	142	1.50E-22	113	[WP_061431364] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE331_2	55.4	570	240	7	1	561	48	612	2.50E-180	640	[WP_072319651] phage terminase large subunit family protein [Microcystis aeruginosa]
	MVG_NODE331_3	36.1	285	154	9	4	274	2	272	7.10E-35	156	[A0A257BC92_9BACT] RecName: Full=Methyltransferase {ECO:0000256|RuleBase:RU362026}; EC=2.1.1.- {ECO:0000256|RuleBase:RU362026}
	MVG_NODE331_5	65	123	41	1	18	140	2	122	5.70E-41	174	[YP_009217746] hypothetical protein MaMVDC_62 [Microcystis phage MaMV-DC]
	MVG_NODE331_7	29.8	191	109	5	2	167	1	191	2.60E-09	69	[AP014012_20] unnamed protein product [uncultured Mediterranean phage uvMED]
	MVG_NODE331_8	49.3	432	211	5	1	428	57	484	3.90E-114	420	[WP_061431361] phage portal protein [Microcystis aeruginosa]
	MVG_NODE331_9	49	600	288	8	13	609	9	593	6.10E-151	543	[A0A139GPB2_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:KXS92027.1}
	MVG_NODE331_11	35.1	94	51	4	4	91	4	93	5.20E-02	45	[K7YBG1_9VIRU] SubName: Full=Putative phage head-tail attachment protein {ECO:0000313|EMBL:AFX83904.1}
	MVG_NODE331_14	31.6	237	149	7	15	243	2	233	1.10E-19	105	[A0A2D2W2E0_9CAUD] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:ATS92314.1}
	MVG_NODE331_15	29.9	137	67	4	7	143	6	113	3.30E-04	52	[WP_061431348] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE331_16	23.7	1337	871	33	2044	3287	1582	2862	8.00E-80	309	[WP_006519205] tape measure domain protein [Leptolyngbya sp. PCC 7375]

Table 3-3. continued.

contig_id	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE375	MVG_NODE375_1	74.5	141	35	1	85	224	6	146	2.00E-57	230	[I4G2J4_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:CCI02155.1}
	MVG_NODE375_2	33.2	319	179	9	6	299	2	311	2.00E-32	148	[WP_061433468] helix-turn-helix domain-containing protein [Microcystis aeruginosa]
	MVG_NODE375_3	64.8	71	25	0	1	71	61	131	4.90E-21	108	[WP_106458557] hypothetical protein [Aphanotheca hegewaldii]
	MVG_NODE375_4	40	65	33	2	9	71	41	101	1.40E-04	53	[WP_002769191] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_6	37.9	253	136	6	105	344	2	246	4.60E-35	157	[WP_103672878] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_9	78.6	112	24	0	4	115	8	119	3.80E-42	178	[WP_061431330] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_10	47.1	138	70	2	7	141	22	159	7.50E-26	127	[WP_069347105] DUF3102 domain-containing protein [Scytonema millei]
	MVG_NODE375_13	49.1	55	28	0	22	76	21	75	5.50E-04	51	[WP_089127810] XRE family transcriptional regulator [Tolypothrix sp. NIES-4075]
	MVG_NODE375_14	62.8	78	27	1	1	78	4	79	5.80E-17	94	[WP_024971272] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_15	35	337	201	5	4	323	37	372	7.50E-57	229	[WP_006518787] hypothetical protein [Leptolyngbya sp. PCC 7375]
	MVG_NODE375_16	32.1	131	80	3	47	173	22	147	3.40E-06	59	[CP012688_1452] hypothetical protein [Ralstonia solanacearum]
	MVG_NODE375_17	46.9	32	17	0	76	107	8	39	1.80E-02	46	[WP_085333019] IS6 family transposase [Clostridium botulinum]
	MVG_NODE375_20	60.7	901	330	7	7	888	6	901	0.00E+00	1080	[WP_061431341] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_22	45.7	289	138	8	3	272	23	311	8.70E-55	221	[YP_009217757] hypothetical protein MaMVDC_73 [Microcystis phage MaMV-DC]
	MVG_NODE375_23	30.4	237	153	6	15	244	2	233	1.10E-17	99	[A0A2D2W2E0_9CAUD] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:ATS92314.1}
	MVG_NODE375_24	62.3	334	121	2	2	332	3	334	4.00E-117	429	[WP_061431344] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_25	44.7	132	71	2	1	131	1	131	3.10E-23	115	[WP_061431345] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE375_26	51.7	288	135	1	1	288	1	284	5.90E-84	319	[WP_061431346] DUF2163 domain-containing protein [Microcystis aeruginosa]
	MVG_NODE375_27	39.9	193	101	6	478	659	7	195	3.20E-23	119	[L8M1U9_9CYAN] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:ELS03183.1}
MVG_NODE382	MVG_NODE382_1	94.6	444	24	0	1	444	1	444	4.80E-242	845	[WP_061431341] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_2	90.6	64	6	0	1	64	1	64	6.40E-24	117	[WP_061431340] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_3	98.9	94	1	0	1	94	1	94	7.10E-44	184	[WP_061431339] XRE family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE382_4	98.7	76	1	0	1	76	1	76	6.80E-34	151	[WP_061431338] XRE family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE382_5	98.9	92	1	0	1	92	1	92	7.20E-44	184	[WP_061431337] XRE family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE382_6	100	73	0	0	1	73	1	73	1.90E-33	149	[WP_061431336] XRE family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE382_7	93.2	59	4	0	1	59	1	59	1.20E-22	113	[WP_061431335] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_8	81.1	111	21	0	1	111	10	120	8.80E-42	177	[WP_061431334] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_9	96	50	2	0	21	70	45	94	2.10E-19	102	[WP_061431333] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_10	94.2	411	24	0	1	411	1	411	1.70E-225	790	[WP_061431332] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_11	86.9	505	27	3	1	497	1	474	3.90E-243	849	[WP_061431331] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_12	95.9	123	5	0	1	123	1	123	2.70E-59	235	[WP_061431330] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_13	97.4	77	2	0	1	77	21	97	6.10E-35	154	[WP_061431329] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_14	93.1	379	23	2	1	378	1	377	2.70E-200	706	[WP_061431328] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_15	78	168	11	1	1	142	1	168	1.30E-61	243	[WP_061431369] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_16	92.9	84	6	0	1	84	1	84	1.20E-35	156	[WP_061431368] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_19	83.8	210	28	3	1	207	1	207	3.70E-86	325	[WP_061431367] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_20	90.9	121	11	0	1	121	1	121	2.80E-56	225	[WP_061431365] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_21	98.2	168	3	0	1	168	1	168	9.70E-86	323	[WP_061431364] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_22	97.9	565	12	0	1	565	48	612	0.00E+00	1145	[WP_072319651] phage terminase large subunit family protein [Microcystis aeruginosa]
	MVG_NODE382_23	96.9	194	6	0	1	194	1	194	2.70E-100	371	[WP_061431362] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_25	97.7	900	15	1	1	894	1	900	0.00E+00	1727	[WP_061431361] phage portal protein [Microcystis aeruginosa]
	MVG_NODE382_26	61.5	634	203	10	1	629	1	598	5.80E-205	722	[WP_061431360] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE382_27	100	26	0	0	63	88	81	106	9.00E-07	60	[WP_061431360] hypothetical protein [Microcystis aeruginosa]

Table 3-3. continued.

contig_id	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE385	MVG_NODE385_1	89.8	108	11	0	1	108	1	108	1.20E-48	200	[WP_103113469] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_2	87.3	110	13	1	1	109	21	130	9.80E-46	190	[WP_103113468] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_3	59.4	64	26	0	1	64	1	64	2.00E-14	86	[WP_103113467] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_4	77.8	126	27	1	1	126	1	125	2.50E-49	202	[WP_103113466] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_5	71.5	256	72	1	1	256	1	255	1.00E-99	371	[WP_103113465] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_6	79.4	218	42	3	1	216	1	217	1.70E-90	339	[WP_103113464] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_7	68.9	476	113	2	1	442	1	475	2.10E-181	643	[A0A2I9CIL4_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:GBF00299.1}
	MVG_NODE385_8	34.1	123	51	4	11	127	2	100	1.00E-02	47	[A0A2I9CT04_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:GBF00300.1}; Flags: Fragment
	MVG_NODE385_9	54.7	364	131	9	19	351	19	379	3.20E-99	370	[WP_002785200] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_10	78	127	27	1	1	127	1	126	1.00E-50	206	[WP_103113461] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_11	88.6	790	88	1	1	788	1	790	0.00E+00	1445	[WP_103113459] DUF3854 domain-containing protein [Microcystis aeruginosa]
	MVG_NODE385_13	91.4	116	10	0	1	116	1	116	1.50E-57	229	[WP_103113457] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_15	64.2	81	27	1	1	79	1	81	2.80E-19	102	[WP_103113456] XRE family transcriptional regulator [Microcystis aeruginosa]
	MVG_NODE385_16	74.2	236	51	3	1	236	15	240	1.60E-85	323	[WP_103113484] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_18	80.9	68	13	0	1	68	1	68	1.80E-26	126	[WP_103113483] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_19	86.2	65	9	0	1	65	1	65	1.30E-24	120	[WP_103113482] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_20	88	108	10	2	1	108	1	105	7.00E-44	184	[WP_103113481] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_21	40.9	66	34	3	3	65	12	75	5.10E-05	55	[WP_103672815] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_22	71.4	63	18	0	1	63	15	77	2.40E-18	99	[WP_103113479] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_23	91	67	6	0	1	67	1	67	3.00E-26	125	[WP_103113477] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_24	35.5	110	58	4	24	123	75	181	1.10E-07	64	[WP_090299768] hypothetical protein [Ensifer sp. YR511]
	MVG_NODE385_25	85.2	128	19	0	1	128	20	147	9.50E-57	226	[WP_103112618] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_26	87.7	358	43	1	10	366	8	365	6.20E-186	658	[WP_103113473] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_27	62.9	97	28	2	4	94	3	97	2.70E-27	129	[WP_107667102] HNH endonuclease [Cyanothece sp. BG0011]
	MVG_NODE385_28	83	106	18	0	1	106	15	120	4.40E-46	191	[WP_103113472] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_29	50.6	89	37	1	9	90	5	93	1.20E-14	87	[WP_103113471] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_30	52.5	257	108	5	6	249	5	260	4.80E-61	244	[WP_002785205] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE385_32	80.8	78	15	0	1	78	110	187	5.20E-26	124	[WP_103113469] hypothetical protein [Microcystis aeruginosa]
MVG_NODE378	MVG_NODE378_1	28.9	159	93	5	73	230	642	781	1.80E-02	49	[WP_012498275] massive surface protein MspE [Mycoplasma arthritidis]
	MVG_NODE378_2	31.9	232	146	4	5	228	8	235	5.30E-25	122	[WP_006519269] hypothetical protein [Leptolyngbya sp. PCC 7375]
	MVG_NODE378_3	40.3	119	61	3	16	128	92	206	2.80E-16	92	[YP_005098312] unnamed protein product [Synechococcus phage S-CBS4]
	MVG_NODE378_5	72.4	116	32	0	1	116	28	143	4.40E-46	191	[WP_055077309] hypothetical protein [Pseudanabaena sp. 'Roaring Creek']
	MVG_NODE378_7	46.8	331	121	7	1	284	35	357	3.50E-69	270	[A0A257SMV5_9GAMM] RecName: Full=Methyltransferase {ECO:0000256|RuleBase:RU362026}; EC=2.1.1.- {ECO:0000256|RuleBase:RU362026}
	MVG_NODE378_8	36	386	228	10	8	386	26	399	1.70E-55	225	[WP_017290089] hypothetical protein [Leptolyngbya boryana]
	MVG_NODE378_10	45.6	272	115	6	4	252	468	729	4.10E-58	232	[A0A0K1S2D0_9CHRO] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:AKV68153.1}
	MVG_NODE378_11	28.3	99	65	1	42	134	544	642	5.10E-02	45	[J9GCW8_9SPIT] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:EJY88181.1}
	MVG_NODE378_12	27	296	200	7	700	992	4	286	2.80E-21	113	[WP_048757496] DUF2163 domain-containing protein [Afipia felis]
	MVG_NODE378_13	48.8	168	75	6	16	180	28	187	9.60E-31	143	[WP_106908154] PEP-CTERM sorting domain-containing protein [Microcystis sp. MC19]
	MVG_NODE378_14	29.3	297	164	8	215	503	134	392	1.00E-18	103	[A0A2D6XBD8_9ARCH] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:MAH48245.1}
	MVG_NODE378_15	37.5	379	219	7	236	605	183	552	7.70E-61	243	[WP_023067563] helicase [Lyngbya aestuarii]
	MVG_NODE378_22	29.2	72	51	0	15	86	12	83	3.90E-05	55	[K9SNF8_9CYAN] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:AFY71516.1}
MVG_NODE562	MVG_NODE562_4	55.6	232	50	3	4	182	348	579	4.80E-64	251	[WP_016514926] serine/threonine protein kinase [Microcystis aeruginosa]
	MVG_NODE562_8	33.3	105	64	2	1	99	10	114	6.30E-08	64	[YP_009217745] hypothetical protein MaMVDC_61 [Microcystis phage MaMV-DC]
	MVG_NODE562_11	31.2	125	76	3	13	131	21	141	2.60E-09	69	[A0A2E0YG91_9RHOB] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:MAT87182.1}

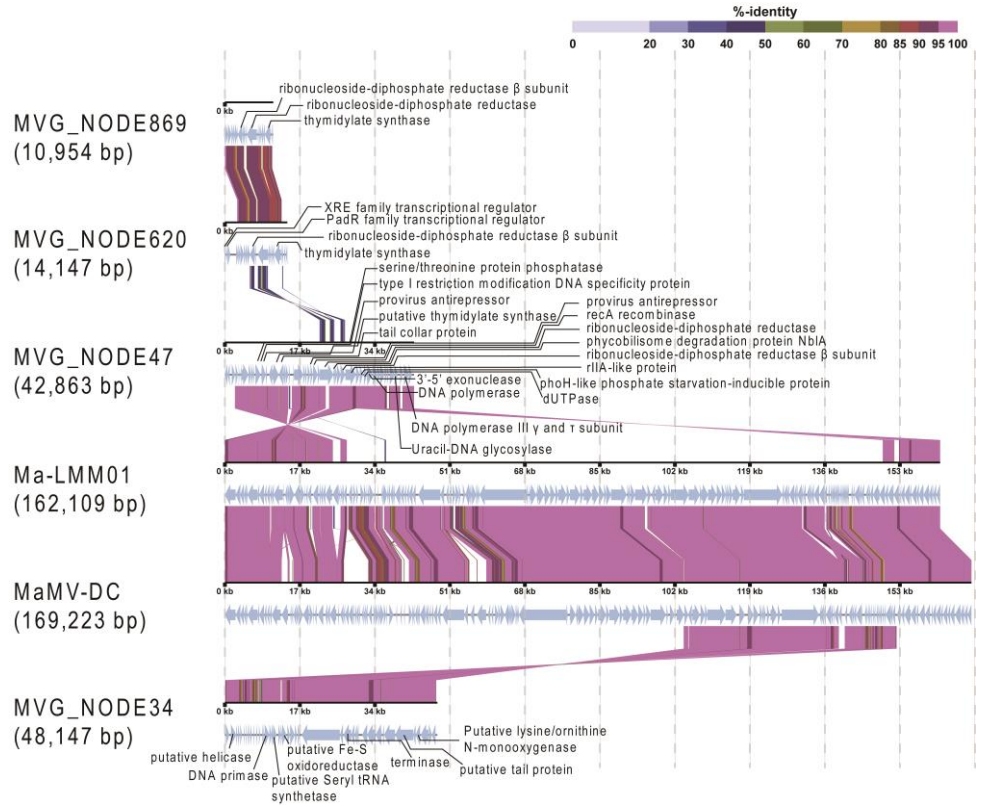
Table 3-3. continued.

	gene_id	identity (%)	alignment length	mismatch	gap	query start	query end	subject start	subject end	e-value	bit score	annotation
MVG_NODE577	MVG_NODE577_9	42.4	151	68	3	3	139	1	146	1.80E-21	110	[WP_103137304] hypothetical protein [Nostoc sp. CENA543]
	MVG_NODE577_14	45.3	402	167	8	12	412	92	441	7.20E-89	336	[A0A139XBM2_9CYAN] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:KYC42108.1}
	MVG_NODE577_15	89.7	97	10	0	1	97	1	97	7.30E-41	174	[YP_009217718] XRE-family like protein [Microcystis phage MaMV-DC]
	MVG_NODE577_16	88.9	280	31	0	1	280	1	280	3.90E-138	498	[WP_079205764] IS5 family transposase [Microcystis aeruginosa]
	MVG_NODE577_17	93.6	156	10	0	1	156	1	156	2.10E-80	305	[YP_009217719] hypothetical protein MaMVDC_35 [Microcystis phage MaMV-DC]
	MVG_NODE577_18	82.6	259	45	0	23	281	188	446	1.20E-123	450	[WP_069475170] TIR domain-containing protein [Microcystis aeruginosa]
	MVG_NODE577_19	61.1	175	63	2	4	173	3	177	3.00E-55	221	[YP_009217724] serine/threonine protein kinase [Microcystis phage MaMV-DC]
	MVG_NODE577_20	50.9	53	25	1	35	86	36	88	5.20E-07	61	[A0A2D6TTJ9_9PROT] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:MAH05250.1}
MVG_NODE636	MVG_NODE636_2	36.2	105	61	2	1	99	10	114	3.40E-09	69	[YP_009217745] hypothetical protein MaMVDC_61 [Microcystis phage MaMV-DC]
	MVG_NODE636_5	63.7	113	38	2	1	113	1	110	2.20E-32	146	[YP_851143] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE636_6	29.6	125	78	3	13	131	21	141	3.70E-08	65	[A0A2E0YG91_9RHOB] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:MAT87182.1}
	MVG_NODE636_9	31.8	110	55	4	15	108	303	408	6.60E-02	45	[A0A2T7IH55_THEOR] SubName: Full=Methyltransferase {ECO:0000313|EMBL:PVC56329.1}
MVG_NODE656	MVG_NODE656_1	100	190	0	0	1	190	1	190	1.30E-104	386	[YP_009217726] hypothetical protein MaMVDC_42 [Microcystis phage MaMV-DC]
	MVG_NODE656_2	89.3	150	16	0	6	155	7	156	7.70E-70	270	[YP_009217725] hypothetical protein MaMVDC_41 [Microcystis phage MaMV-DC]
	MVG_NODE656_3	90.7	246	23	0	1	246	1	246	3.30E-134	485	[YP_851051] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE656_4	76.6	158	33	1	41	194	3	160	8.80E-68	264	[YP_851051] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE656_5	87.5	136	17	0	1	136	1	136	4.20E-65	254	[YP_009217723] hypothetical protein MaMVDC_39 [Microcystis phage MaMV-DC]
	MVG_NODE656_6	94.1	68	4	0	1	68	1	68	9.30E-28	130	[YP_009217722] hypothetical protein MaMVDC_38 [Microcystis phage MaMV-DC]
	MVG_NODE656_7	96.2	130	5	0	1	130	1	130	1.10E-63	250	[YP_009217721] hypothetical protein MaMVDC_37 [Microcystis phage MaMV-DC]
	MVG_NODE656_9	77.9	77	16	1	1	76	1	77	2.40E-23	116	[YP_009217717] hypothetical protein MaMVDC_33 [Microcystis phage MaMV-DC]
	MVG_NODE656_12	61.7	206	74	2	1	201	1	206	7.20E-65	254	[A0A139GJR8_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:KXS90421.1}
	MVG_NODE656_14	32.1	589	365	14	10	584	65	632	8.70E-65	257	[WP_103672900] DUF3987 domain-containing protein [Microcystis aeruginosa]
MVG_NODE671	MVG_NODE671_1	88.1	67	8	0	80	146	12	78	3.00E-26	125	[WP_080604598] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE671_2	71.1	83	24	0	3	85	14	96	8.10E-24	117	[WP_002767180] hypothetical protein [Microcystis aeruginosa]
	MVG_NODE671_3	39.5	261	143	6	3	248	4	264	5.20E-34	152	[WP_045057041] hypothetical protein [Aliterella atlantica]
	MVG_NODE671_10	90.2	173	17	0	1	173	31	203	9.00E-84	316	[YP_009217834] hypothetical protein MaMVDC_150 [Microcystis phage MaMV-DC]
	MVG_NODE671_11	28.1	739	485	28	20	743	32	739	2.40E-54	222	[WP_021171276] phage tail protein [Sporomusa ovata]
	MVG_NODE671_12	82.4	74	13	0	1	74	1	74	2.30E-26	126	[YP_851058] hypothetical protein [Microcystis virus Ma-LMM01]
	MVG_NODE671_15	33.5	221	132	6	1	219	1	208	9.00E-26	126	[A0A2K2ZUX5_HALP7] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:PNW35306.1}
	MVG_NODE671_17	36.2	116	63	4	116	221	11	125	1.20E-09	71	[YP_009217723] hypothetical protein MaMVDC_39 [Microcystis phage MaMV-DC]
	MVG_NODE671_19	77	126	28	1	30	154	10	135	1.70E-50	206	[YP_009217723] hypothetical protein MaMVDC_39 [Microcystis phage MaMV-DC]
	MVG_NODE671_20	60.4	134	50	2	1	131	1	134	7.20E-36	157	[YP_009217722] hypothetical protein MaMVDC_38 [Microcystis phage MaMV-DC]
	MVG_NODE671_21	95.7	47	2	0	1	47	84	130	4.90E-16	91	[YP_009217721] hypothetical protein MaMVDC_37 [Microcystis phage MaMV-DC]
MVG_NODE982	MVG_NODE982_2	32.5	231	131	8	1	229	1	208	4.50E-23	117	[A0A2K2ZUX5_HALP7] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:PNW35306.1}
	MVG_NODE982_5	65.7	134	44	1	1	132	1	134	8.80E-42	177	[YP_009217722] hypothetical protein MaMVDC_38 [Microcystis phage MaMV-DC]
	MVG_NODE982_6	54.5	110	44	3	6	115	1	104	9.50E-20	104	[YP_009217714] hypothetical protein MaMVDC_30 [Microcystis phage MaMV-DC]
	MVG_NODE982_8	81.8	77	13	1	1	76	1	77	1.20E-25	123	[YP_009217717] hypothetical protein MaMVDC_33 [Microcystis phage MaMV-DC]
	MVG_NODE982_11	58.3	206	81	2	1	201	1	206	1.10E-60	240	[A0A139GJR8_MICAE] SubName: Full=Uncharacterized protein {ECO:0000313|EMBL:KXS90421.1}
	MVG_NODE982_13	30.8	590	374	13	10	592	39	601	2.20E-60	243	[WP_015179913] DUF3987 domain-containing protein [Oscillatoria nigro-viridis]

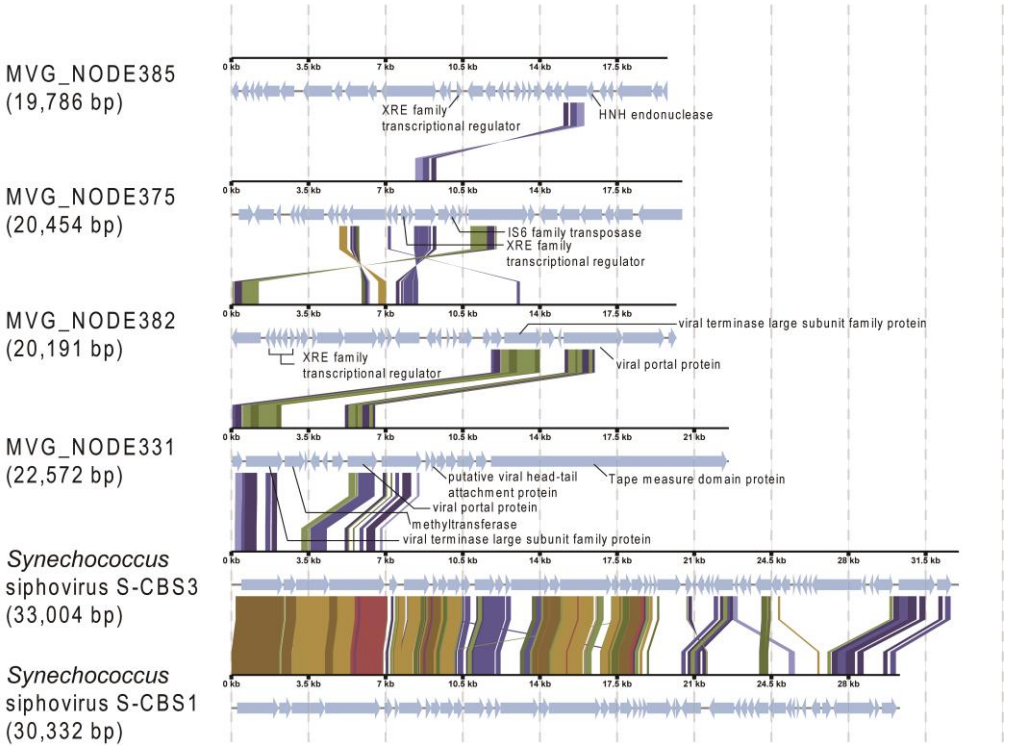
Genes with no hits are not listed.

3. Broad and narrow host range *Microcystis* viruses

(a) Group I



(b) Group II



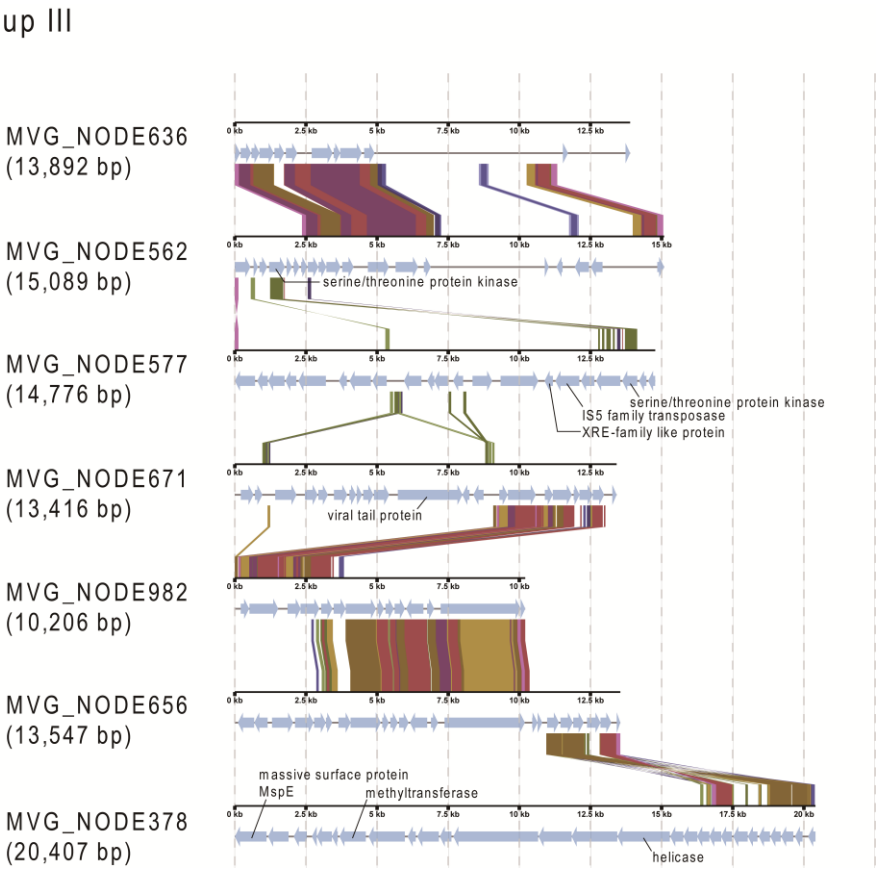
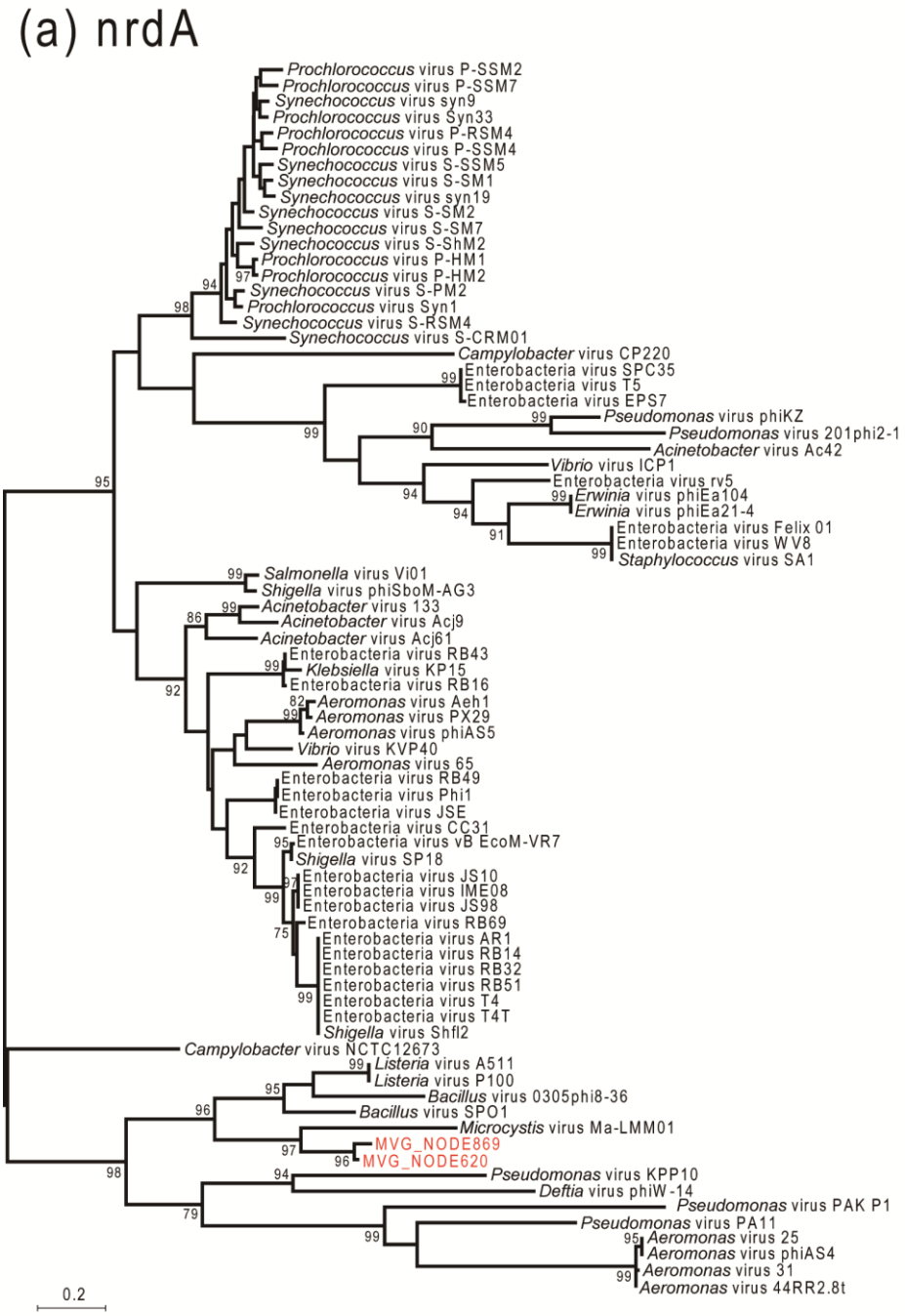


Figure 3-3. Genome map of *Microcystis* viral genomes in Group I, Group II and Group III. Putative gene functions are indicated for each *Microcystis* viral genome. All tBLASTx alignments are represented by colored lines between two genomes. The color scale represents the tBLASTx percent identity.

3. Broad and narrow host range *Microcystis* viruses



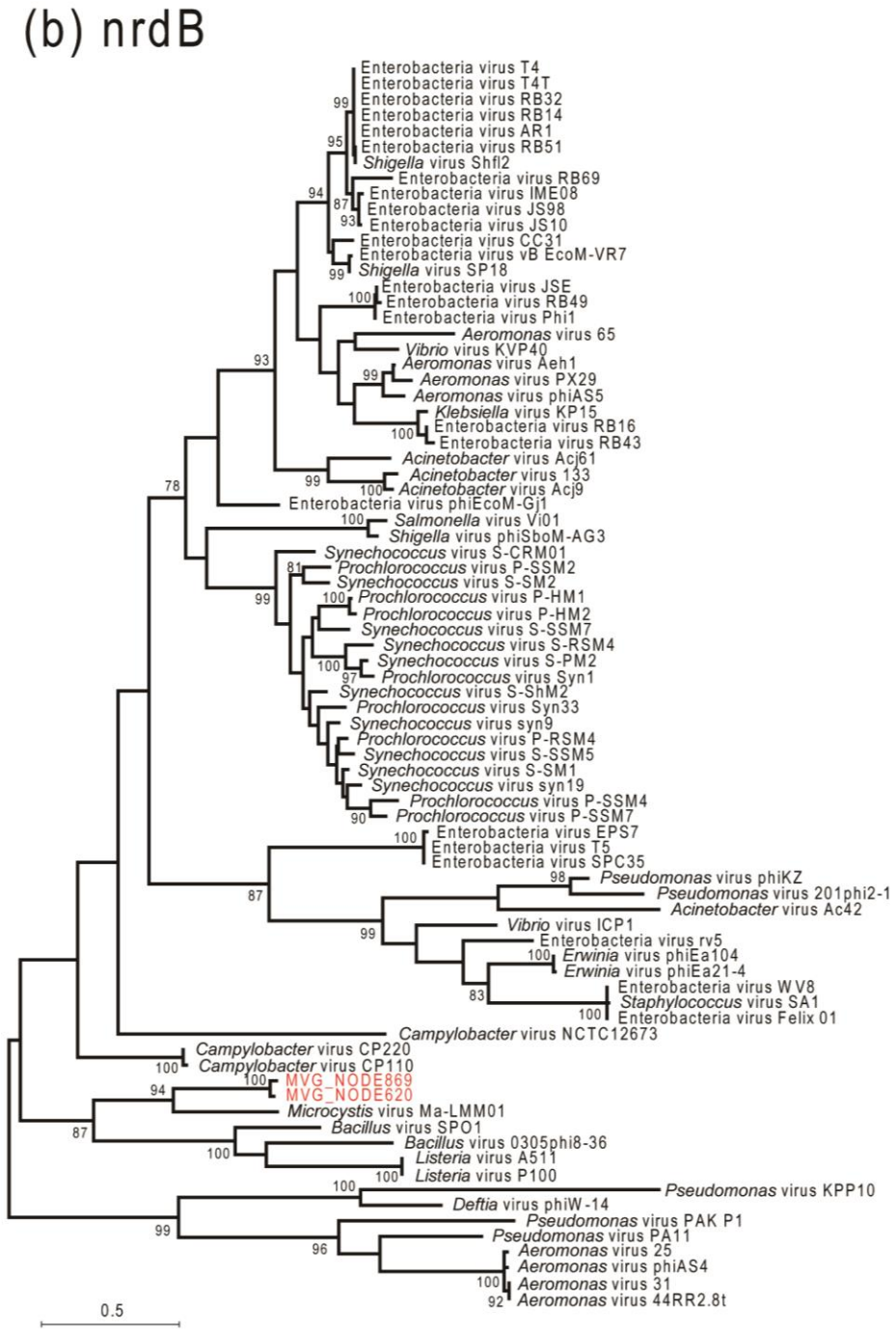


Figure 3-4. Maximum-likelihood tree of ribonucleotide reductase α (nrdA) and β (nrdB) subunit genes. The tree contains the protein sequences used in a previous study (shown in black characters) (143) and encoded in MVG_NODE620 and NODE869 (shown in red characters). The scale bar refers to the estimated number of amino acid substitutions per site. Numbers close to the nodes represent bootstrap percentages above 75%.

The four viral contigs in Group II ranged in size from 19,786 to 22,572 bp and contained 16–32 predicted protein-coding genes (**Table 3-3**). Of these, MVG_NODE331, NODE375 and NODE382 shared one to five homologs (e.g., TerL, viral portal protein and hypothetical protein) with each other. These MVGs partly shared sequence similarities with cyanobacterial siphoviruses like P-SS2 (144), S-CBS1, S-CBS2, S-CBS3 (145) and KBS-2A (146) (**Figure 3-3**). In *Synechococcus* siphovirus S-CBS1 and S-CBS3, the viral structural genes are conserved on the left arm of their genomes (145). MVG_NODE331 and MVG_NODE382 shared three (TerL, viral portal protein and major capsid protein) and two (TerL and viral portal protein) homologs with *Synechococcus* siphovirus S-CBS3, respectively (**Figure 3-3, Table 3-3**). Phylogenetic analysis of the *terL* gene showed that the two MVGs formed a sister clade with *Synechococcus* siphovirus S-CBS1 and S-CBS3 (**Figure 3-5**). MVG_NODE375 displayed partial sequence similarity (50–60% identity) with the lysozyme gene from *Synechococcus* siphovirus S-CBS3 (**Figure 3-3**). Interestingly, MVG_NODE385 shared only one homolog (hypothetical protein) with MVG_NODE375 and MVG_NODE382 (**Figure 3-3, Table 3-3**, S_G = 0.011 and 0.026, respectively).

The seven viral Group III contigs ranged in size from 10,206 to 20,407 bp and contained 12–27 predicted protein-coding genes (**Table 3-3**). None of these shared genome-wide sequence similarities with known viruses in the current database (**Figure 3-3**). Although blast analysis against the NCBI-nr database revealed few detectable homologues (3.70–36.4% of ORFs) in each MVG (**Table 3-3**), 33.3–50.0% of the ORFs were predicted to encode viral proteins in four MVGs (MVG_NODE378, NODE656, NODE671 and NODE982) by further searches against pVOGs (139) (**Table 3-4**). These MVGs partially overlapped with each other, suggesting that they might be derived from

the same viruses. In contrast, only 10.5–25.0% of the ORFs were predicted to encode viral proteins in the other MVGs (MVG_NODE562, NODE577 and NODE636) (**Table 3-4**). Considering that pVOGs mainly comprise the Caudovirales order (147) and include few freshwater viromes (139), this result is possibly due to such database biases. Furthermore, the CsCl gradient centrifugation step and decontamination of prokaryotic sequences by VirSorter provided support that the Group III contigs were derived from viral sequences, and not cellular organisms. These findings strongly suggested that the Group III MVGs are new viral lineage members.

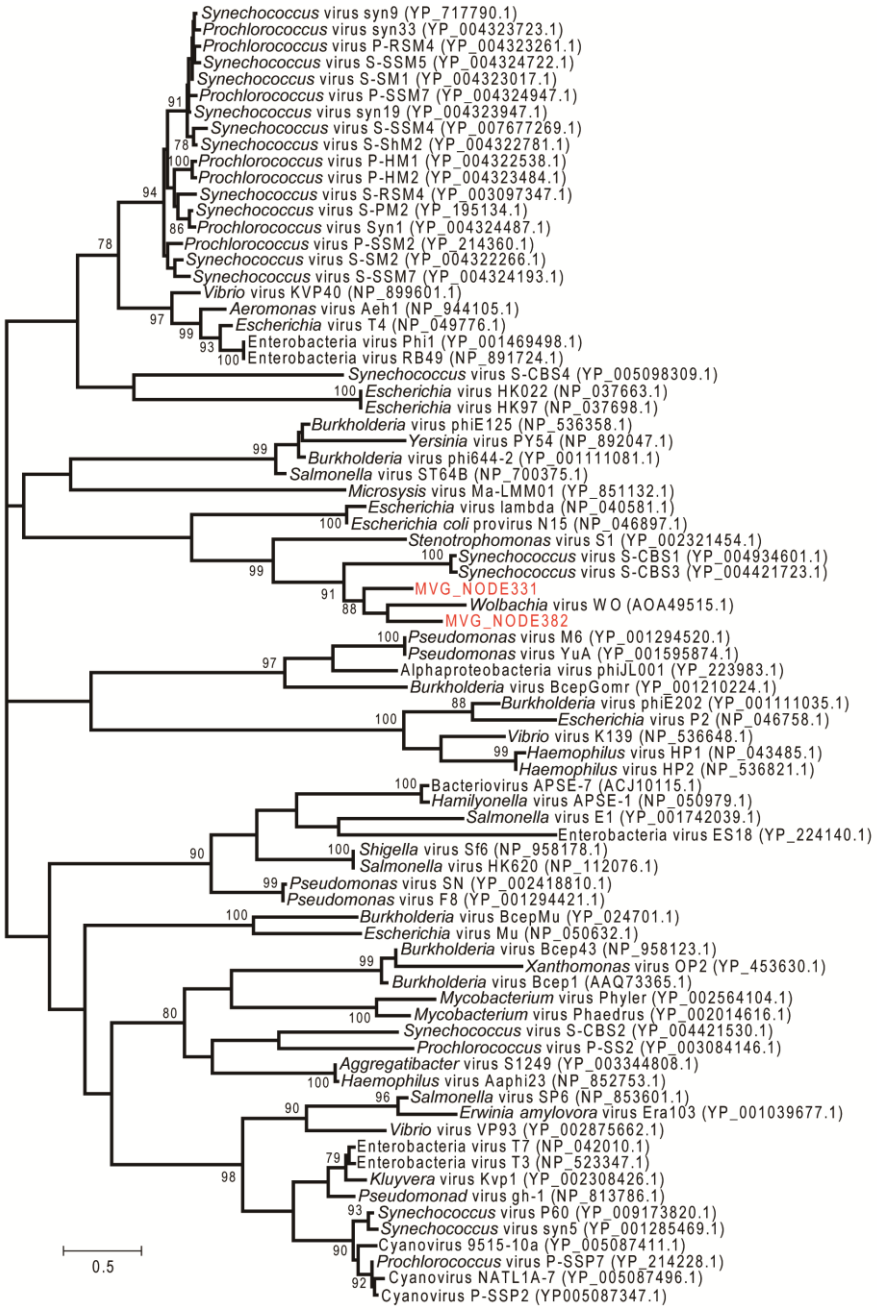


Figure 3-5. Maximum-likelihood tree of TerL (terminase large subunit) genes. The tree contains protein sequences encoded in MVG_NODE331 and NODE382 (shown in red characters). The scale bar refers to the estimated number of amino acid substitutions per site. Numbers close to the nodes represent bootstrap percentages above 75%.

Table 3-4. Putative viral genes in *Microcystis* viral genomes determined by hidden Markov model profiles from the prokaryotic Virus Orthologous Groups (pVOGs) database.

contig_ID	ORF	function	vPOG	e-value
MVG_NODE34	1	hypothetical protein	VOG9956	8.50E-123
	2	DNA helicase	VOG0377	9.10E-51
	4	hypothetical protein	VOG2063	0.00046
	7	hypothetical protein	VOG10935	2.90E-26
	11	hypothetical protein	VOG10935	2.90E-10
	16	DNA primase	VOG4551	2.60E-14
	19	hypothetical protein	VOG5541	2.10E-88
	22	putative Fe-S oxidoreductase	VOG9839	3.60E-183
	23	hypothetical protein	VOG1093	1.30E-10
	24	exonuclease	VOG4692	5.50E-11
	26	hypothetical protein	VOG0283	4.40E-11
	29	terminase large subunit	VOG1886	1.50E-151
	32	hypothetical protein	VOG4127	8.50E-05
	38	hypothetical protein	VOG0080	8.90E-103
	39	hypothetical protein	VOG0079	2.70E-29
	41	hypothetical protein	VOG6811	2.70E-20
	43	hypothetical protein	VOG5296	9.20E-73
MVG_NODE47	2	hypothetical protein	VOG7296	9.40E-06
	3	hypothetical protein	VOG3671	8.20E-59
	11	serine/threonine protein phosphatase	VOG0156	1.90E-79
	13	hypothetical protein	VOG4552	1.60E-33
	16	thymidylate synthase	VOG4561	1.50E-05
	17	hypothetical protein	VOG5169	1.10E-05
	23	hypothetical protein	VOG3108	2.60E-233
	24	hypothetical protein	VOG4741	1.10E-37
	27	prophage antirepressor	VOG10917	1.70E-91
	28	putative DNA helicase	VOG0025	4.90E-19
	30	ribonucleoside triphosphate reductase, alpha chain	VOG2368	1.70E-156
	31	phycobilisome degradation protein NblA	VOG7606	1.30E-30
	32	hypothetical protein	VOG9316	2.10E-73
	34	ribonucleotide reductase subunit β	VOG4562	1.80E-66
	35	protector from prophage-induced early lysis	VOG4594	9.80E-54
	36	hypothetical protein	VOG4590	5.40E-25
	37	P-starvation inducible protein	VOG0058	8.90E-55
	38	deoxyuridine 5'-triphosphate nucleotidohydrolase	VOG0085	3.40E-26
	39	DNA polymerase	VOG0026	1.20E-15
	40	DNA polymerase	VOG0026	2.00E-31
	47	hypothetical protein	VOG7140	1.80E-11
	48	clamp loader subunit	VOG0996	2.20E-49
MVG_NODE620	6	hypothetical protein	VOG5169	3.50E-40
	10	ribonucleotide reductase subunit β	VOG4562	1.10E-66
	12	ribonucleotide reductase	VOG0088	5.90E-147
	14	hypothetical protein	VOG5532	7.40E-05
	17	thymidylate synthase	VOG0092	2.50E-21
	19	phycobilisome degradation protein NblA	VOG7606	3.50E-12
MVG_NODE869	3	hypothetical protein	VOG5169	5.70E-41
	8	ribonucleotide reductase subunit β	VOG4562	3.10E-64
	13	ribonucleotide reductase	VOG0088	8.20E-147
	15	hypothetical protein	VOG5532	0.00043
	19	thymidylate synthase	VOG0092	1.20E-22

Host-virus interactions

To investigate the host range of the above-named viruses, I conducted a phylogenetic analysis on the 15 *Microcystis* strains for which both ITS sequences and CRISPR spacers were available. Consequently, these *M. aeruginosa* strains could be largely divided into the following three groups: type I (CACIAM03, DA14, NIES298, NIES98, PCC7941, PCC9443, PCC9717, PCC9807, SPC777, TAIHU98, TA09), type II (NIES843, PCC9809) and type III (NIES44, PCC9701)) (**Figure 3-6**) (60, 148). According to the ITS phylotyping of *Microcystis* strains, Group I viruses only interacted with seven (DA14, NIES298, NIES98, PCC9807, SPC777) type I strains (**Figure 3-6, Table 3-5**). Group III viruses interacted with eight (DA14, NIES298, NIES98, PCC7941, PCC9443, PCC9717, PCC9807, SPC777) type I strains and one (NIES843) type II strain (**Figure 3-6, Table 3-5**). In contrast, Group II viruses interacted with all type I–III strains except for strain PCC9807 (**Figure 3-6, Table 3-5**). Also, each Group II virus apparently possessed more highly abundant protospacers, parts of which have been acquired repeatedly by the same strains, more than the other groups (up to 23; MVG_NODE375) (**Table 3-2**). These suggested that Group II viruses could interact with the broad range of *Microcystis* strains more frequently than others and have more impact on the bloom than the other groups.

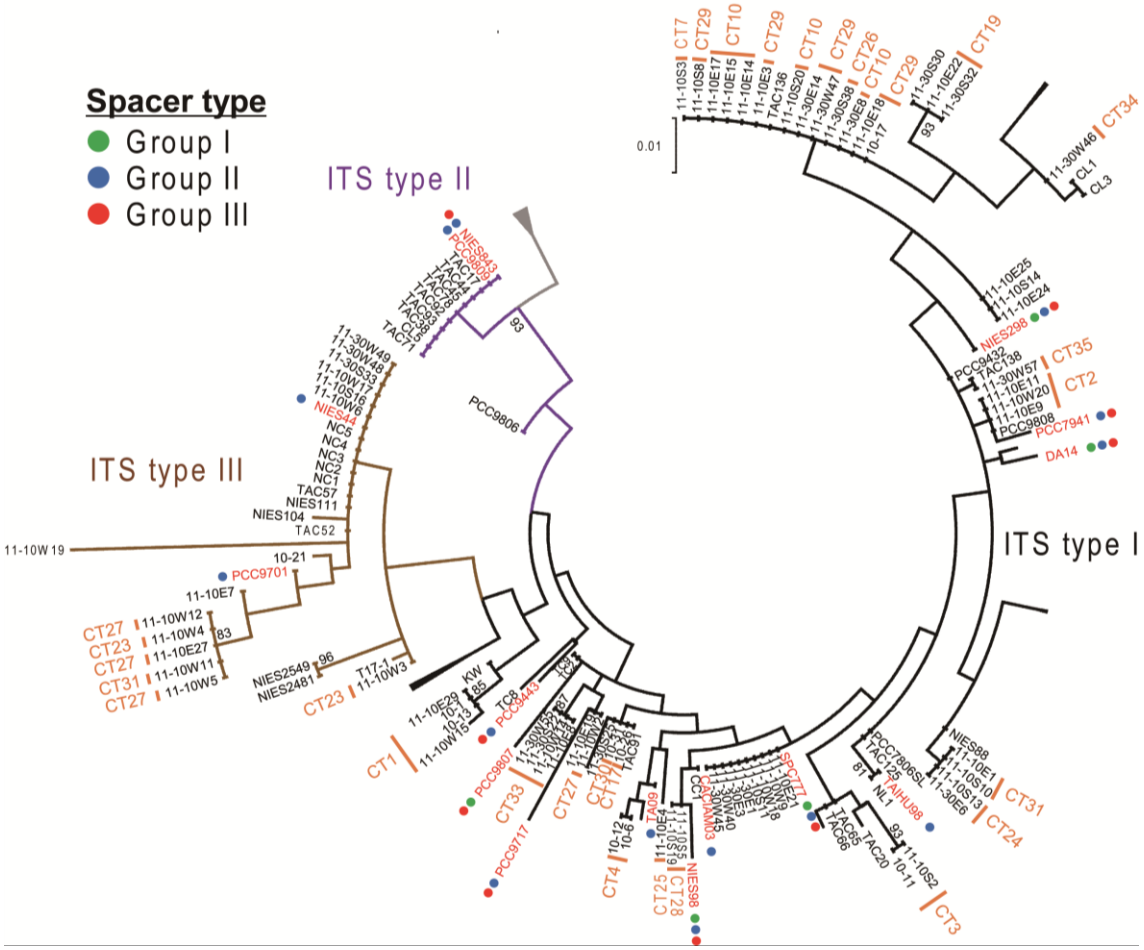


Figure 3-6. Maximum-likelihood tree of the internal transcribed spacer (ITS) sequences from *M. aeruginosa* strains. The tree contains the nucleotide sequences used in previous studies (shown in black characters) (60, 149) and this study (shown in red characters). Strains belonging to the three phylogenetic groups (ITS Cluster I, II and III) were defined in previous studies (60, 149). ITS clusters I, II and III branches are shown in black, purple and brown, respectively. *Microcystis* strains possessing Group I, II and III viral spacers are represented by green, blue and red circles, respectively. Some of the strains used in this study (PCC7005, NaRes975, LE3, DIANCHI905 and CHAOHU1326) are not included because their ITS sequences could not be determined. CRISPR types defined by a previous study (60) are shown in orange characters. The scale bar refers to the estimated number of nucleic acid substitutions per site. Numbers close to the nodes represent bootstrap percentages above 75%.

Table 3-5. Potential host strains for 15 *Microcystis* viruses identified in this study.

contig_id	viral group	potential host strain	host ITS type
MVG_NODE34	Group I	DA14	type I
MVG_NODE47		NIES98	type I
		NaRes975	N.D.
		PCC9807	type I
MVG_NODE620		LE3	N.D.
		PCC7005	N.D.
		SPC777	type I
MVG_NODE869		NIES298	type I
		PCC7005	N.D.
		SPC777	type I
MVG_NODE331	Group II	NIES298	type I
		NIES843	type II
		NIES98	type I
		PCC7941	type I
		PCC9701	type III
		PCC9809	type II
MVG_NODE375		CACIAM03	type I
		LE3	N.D.
		NIES298	type I
		NIES44	type III
		NIES843	type II
		NIES98	type I
		PCC7941	type I
		PCC9701	type III
		TAIHU98	type I
		DA14	type I
		TA09	type I
MVG_NODE382		DIANCHI905	N.D.
		NIES843	type II
		NIES98	type I
		NaRes975	N.D.
		PCC7941	type I
		PCC9443	type I
		PCC9717	type I
		SPC777	type I
MVG_NODE385		CHAOHU1326	N.D.
		NIES298	type I

Table 3-5. continued.

contig_id	viral group	potential host strain	host ITS type
MVG_NODE385	Group II	NIES98	type I
		NaRes975	N.D.
		PCC9443	type I
		PCC9701	type III
		PCC9717	type I
		DA14	type I
MVG_NODE378	Group III	NIES98	type I
MVG_NODE562		PCC9443	type I
		PCC9717	type I
		DA14	type I
MVG_NODE577		LE3	N.D.
		PCC9443	type I
		PCC9717	type I
		SPC777	type I
MVG_NODE636		NIES298	type I
		NIES98	type I
MVG_NODE656		PCC9443	type I
		PCC9807	type I
		DA14	type I
MVG_NODE671		PCC7941	type I
MVG_NODE982		NIES843	type II

N.D. indicates the *Microcystis* strains for which the ITS sequences were not determined.

Transcriptional dynamics of *M. aeruginosa* in the environment

Next, I conducted metatranscriptome analyses to investigate the diurnal expression dynamics of *Microcystis* genes. The sampling of the *Microcystis* bloom generated a total of 9 metatranscriptome samples spread over the day-night transition from the Hirosawanoike Pond. After removing the remaining rRNA reads, 49.9–89.0% of the total paired reads were mapped to *M. aeruginosa* 30S32 genome that is the most dominant genotype (CRISPR type 19 isolate) (62) in Hirosawanoike Pond (**Table 3-1**). Thus, the *Microcystis* read percentages fluctuated greatly during 24 h sampling (**Table 3-1, Figure 3-1b**). This was mainly associated with the accumulation of *Microcystis* cells at the water surface during the night (100), thereby potentially having a large impact on the metatranscriptome results. Temporal changes in the percentage of *rnpB* reads well reflected the sequencing biases derived from the cell densities per sample (**Figure 3-1b**) (100). Therefore, all read counts were normalized as FPKM and *rnpB* counts to capture the transcriptional dynamics within the cells correctly. After normalizing the transcriptional levels, the photosynthesis genes showed the highest transcriptional levels during the daytime, as described previously (129) (**Figure 3-7a**). The expression levels of the TCA genes also increased during the light/dark transition, as was seen in a previous culture experiment (149) (**Figure 3-7b**). These results suggested that *M. aeruginosa* activates various cellular metabolisms including photosynthesis during the daytime, and then consumes the resultant products at night. This coincides with the observation that *Microcystis* cells move vertically towards the water surface by becoming lighter at night (100).

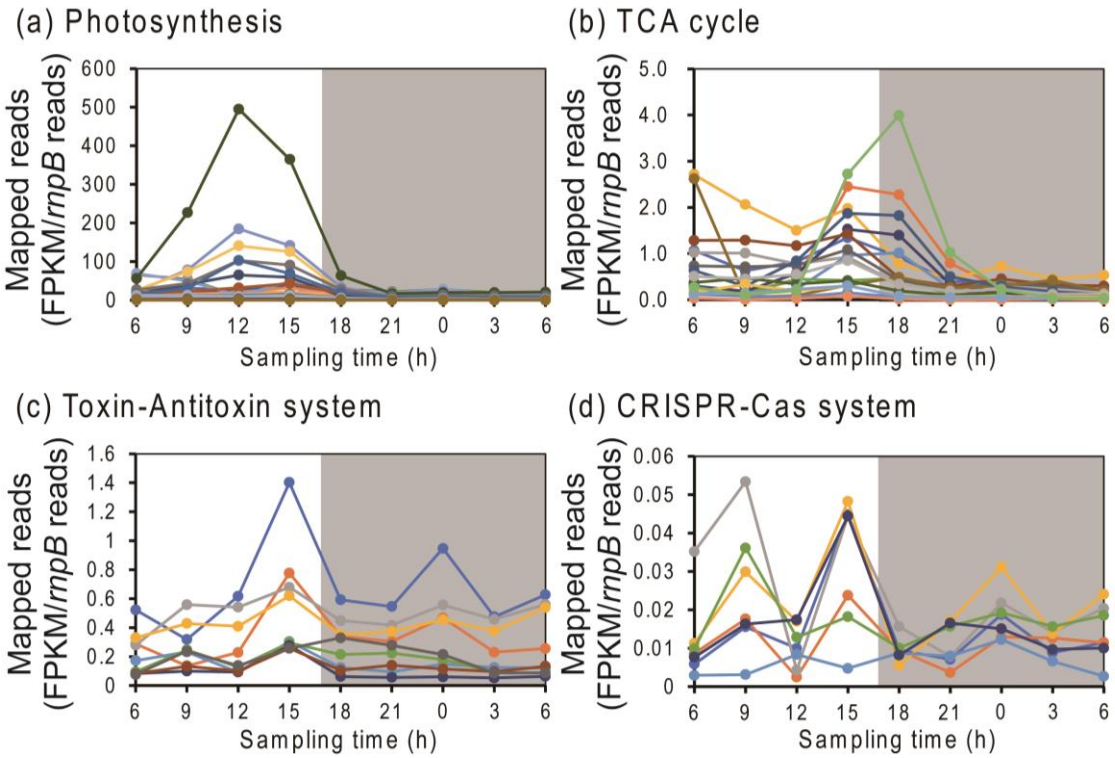


Figure 3-7. Temporal gene expression patterns of *M. aeruginosa* in the environment. (a) Photosynthesis, (b) TCA cycle, (c) Toxin-antitoxin system, (d) CRISPR-Cas system (a, b, d: all genes defined in the KEGG Orthology database are indicated, c: only the top 10 genes showing the highest transcriptional abundances at 15:00 h are indicated). The y-axis represents FPKM (fragments per kilobase per mapped million reads) normalized by *rnpB* reads as a proxy for *Microcystis* cell density. Shaded areas indicate the periods of darkness.

Transcriptional dynamics of *M. aeruginosa* viruses in the environment

I further investigated the transcriptional dynamics of *Microcystis* viruses. After read processing, 0.41–1.05% of the total paired reads were mapped to *Microcystis* virus Ma-LMM01 and the 960 viral contigs (≥ 10 kb), including the MVGs (**Table 3-1**). In the case of Ma-LMM01, for example, the viral reads derived from the samples at 15:00 hours were mapped to 52.7% of the total genes (97/184; **Figure 3-8**), suggesting that the metatranscriptome analysis could capture the whole transcriptional dynamics of each

MVG, not a specific gene expression per se.

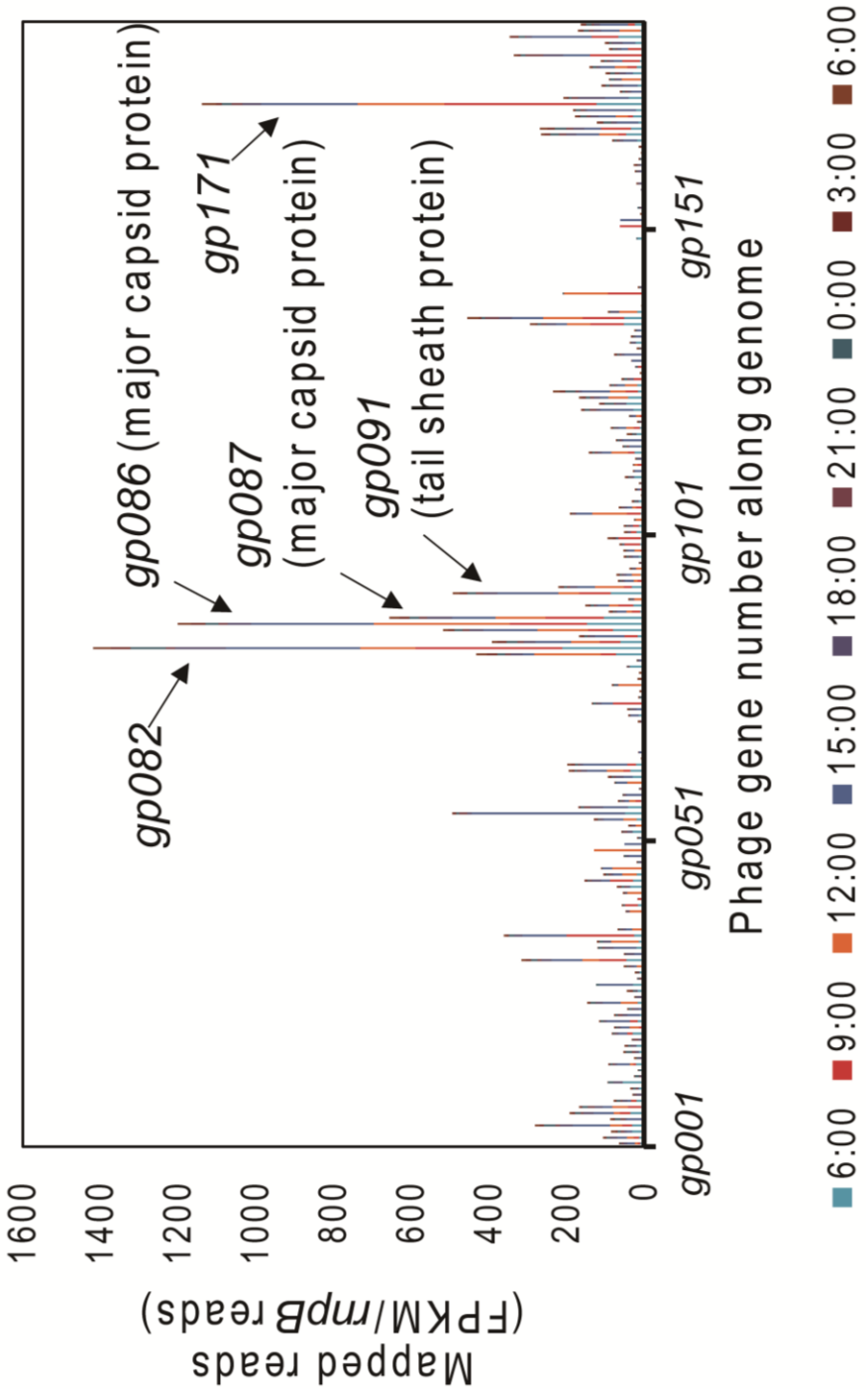


Figure 3-8. Transcriptomic read mapping pattern of Ma-LMM01 at each sampling time point. The x-axis represents gene positions in the Ma-LMM01 genome. The y-axis represents FPKM (fragments per kilobase per mapped million reads) normalized by *mpB* reads as a proxy for *Microcystis* cell density.

Transcripts were observed in all of the MVGs, clearly indicating that the MVGs actively infect their hosts. The transcriptional dynamics of the MVGs increased gradually during the daytime, peaking at 12:00 or 15:00 hours (**Figure 3-9**). The transcriptional noise peaks at 06:00 hours were derived from the high expression levels of host-like genes (e.g. MAE_RS01135; data not shown). Group II virus, particularly MVG_NODE385, showed the highest transcriptional activities among the *Microcystis* viruses (**Figure 3-9**). However, the transcriptional activities of the MVGs did not coincide with the abundance rank of viral reads mapping to each contig (**Figure 3-10**), suggesting that progeny productivity is determined by not only transcriptional activity but other factors also (further discussed in the conclusion section). Conversely, all the contigs showed minimal transcriptional activity at night (**Figure 3-9**). A previous study reported that Ma-LMM01 *gp091* (tail sheath protein) transcripts peaked during the daytime and that *gp091* DNA copy numbers in the host cell fractions peaked in the afternoon, followed by an increase of *gp091* DNA copy numbers in the nighttime viral fractions (100). Furthermore, other studies have reported that marine cyanoviruses levels increase and cause high mortality rates of their host cells at midnight (150–152). The observed dynamics of *gp091* DNA copy numbers concorded with the results of the previous study (**Figure 3-11**) (100). Also, the viral metagenomic read abundances of the MVGs derived from samples at 06:00 were higher than those derived from samples at 18:00 (**Figure 3-10**). These suggested that all MVG viruses are released from *Microcystis* cells at midnight in the freshwater environment.

Strikingly, Group I, II and III viruses were included in both groups showing transcriptional peak levels at 12:00 (P_{12}) and 15:00 (P_{15}) (**Figure 3-9**). Therefore, I compared the genomic features of MVG_NODE620 (P_{15}) with those of MVG_NODE869

(P_{12}) to investigate whether the slightly different patterns depended on their gene contents or not. These Group I MVGs showed high sequence similarities (S_G = 0.80; **Figure 3-3**) with each other and shared part of the Ma-LMM01 middle gene homologs (not containing early and late gene homologs). This suggested that the observed transcriptome dynamics do not depend on the gene expression classes (or certain genes) but on the combination of viruses and host (*M. aeruginosa*) populations in the environment. Thus, the observed transcriptional pattern of each MVG in this study reflected the actual transcription dynamics of *Microcystis* viruses in the infected cells.

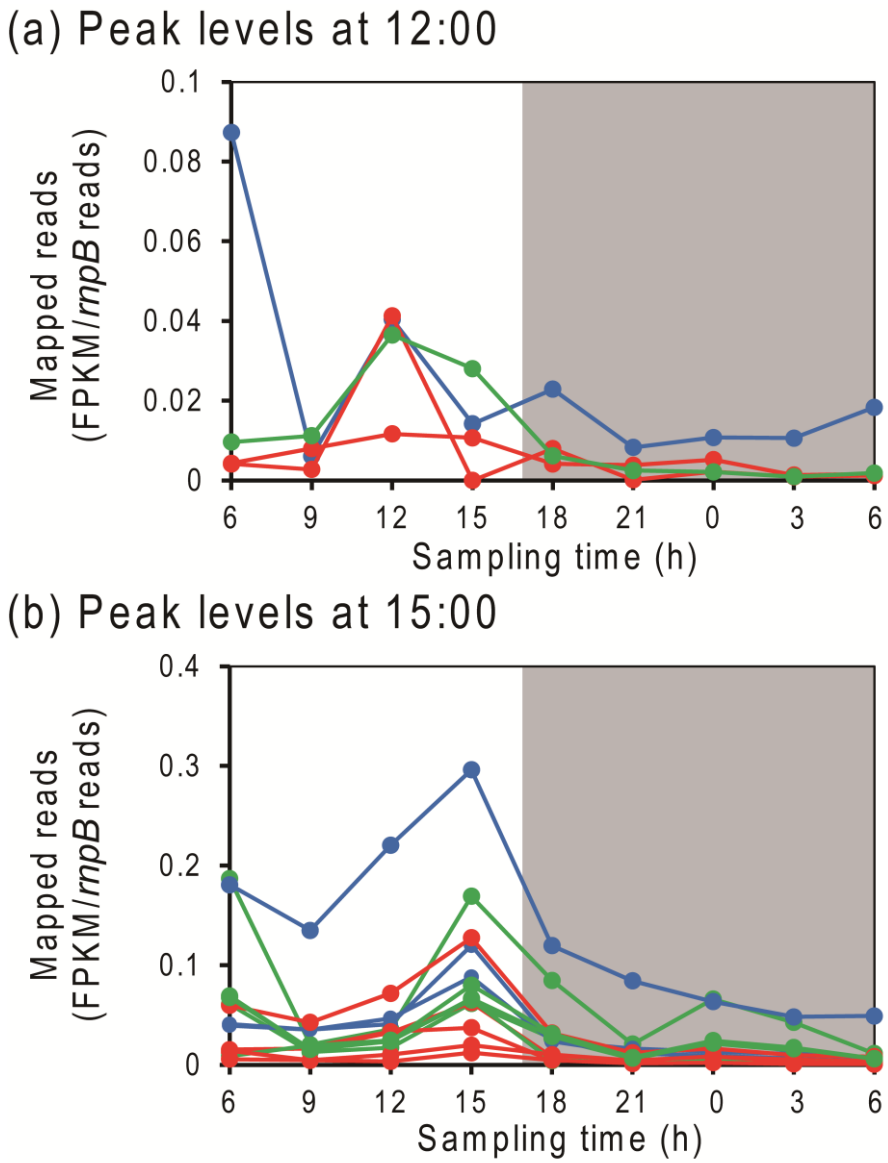


Figure 3-9. Temporal transcriptional dynamics of 15 *Microcystis* viruses in the environment. Group I, II and III viruses are shown in green, blue and red, respectively. (a) MVG_NODE869 (Group I), NODE382 (Group II), NODE636, NODE671 (Group III); (b) MVG_NODE34, NODE47, NODE620, Ma-LMM01, MaMV-DC (Group I), MVG_NODE331, NODE375, NODE385 (Group II), MVG_NODE378, NODE562, NODE577, NODE656, NODE982 (Group III). The y-axis represents FPKM (fragments per kilobase per mapped million reads) normalized by *rnpB* reads as a proxy for *Microcystis* cell density. Shaded areas indicate the periods of darkness.

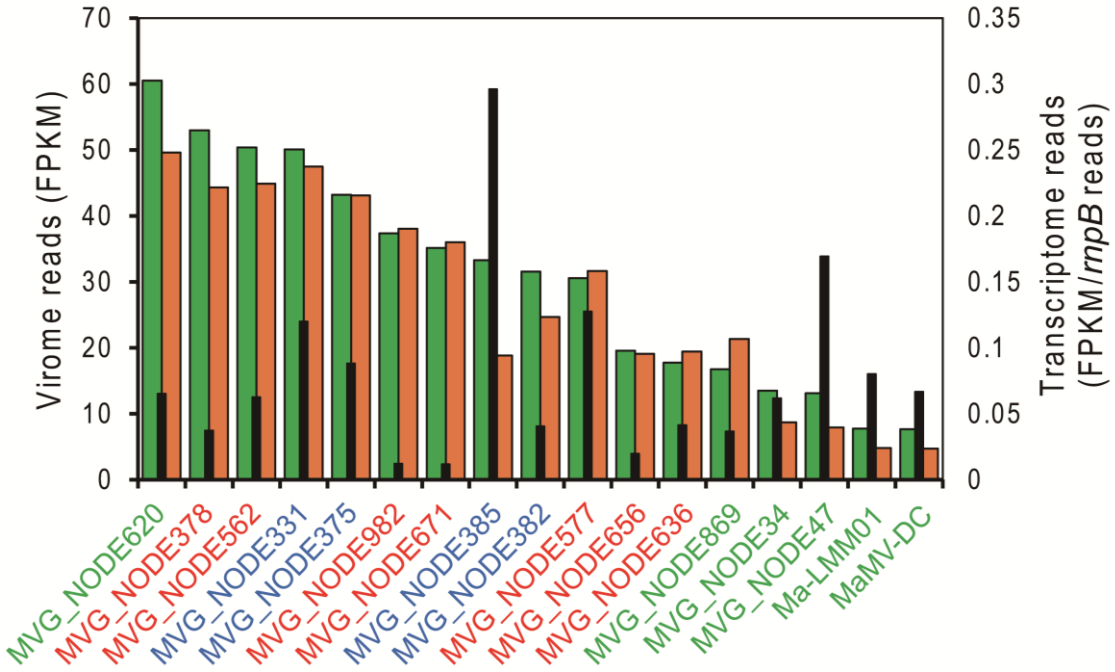


Figure 3-10. Virome and transcriptome read abundances of *Microcystis* viral genomes including Ma-LMM01 and MaMV-DC. The x-axis is the ranked value for each *Microcystis* viral genome. The y-axis represents FPKM (fragments per kilobase per mapped million reads) or FPKM normalized by *rnpB* reads as a proxy for *Microcystis* cell density. The virome abundances at 18:00 and 06:00 are shown in orange and green bars, respectively. Transcriptome read abundances at 12:00 or 15:00 are shown in black bars. Green, blue, and red characters indicate Group I, II, and III viruses, respectively.

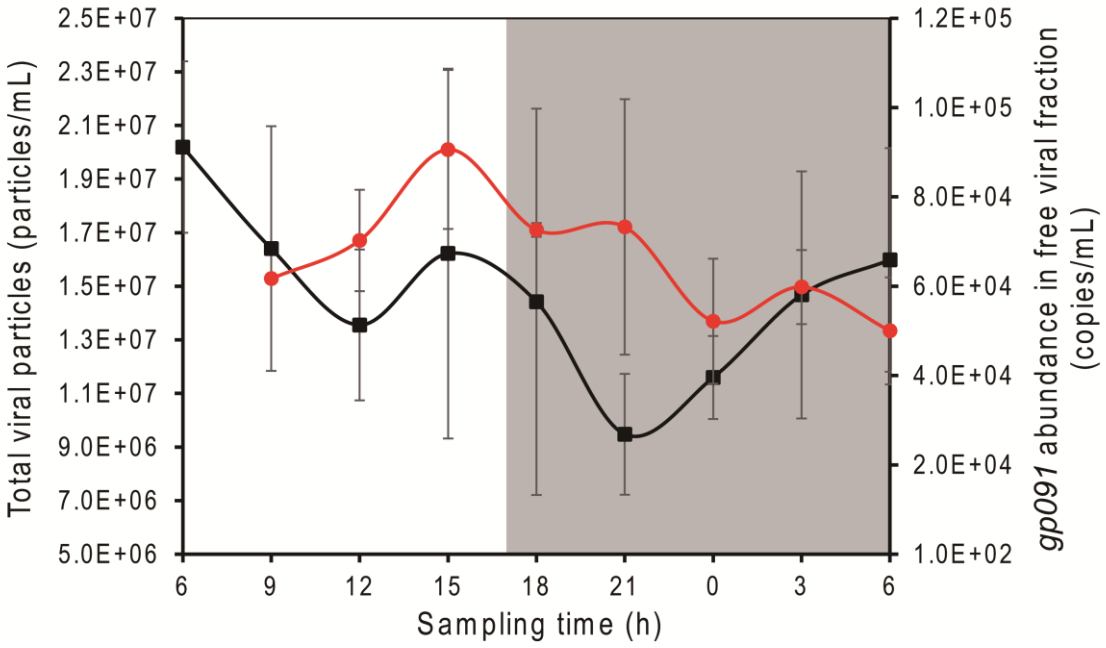


Figure 3-11. Diel changes in viral abundance in the free viral fraction. Total viral particle numbers and the abundances of Ma-LMM01 *gp091* gene are shown in black and red, respectively. Total viral particle numbers were determined by direct counts using microscopy with SYBR Gold. The abundance of *gp091* was determined by quantitative polymerase chain reaction analysis. Shaded area indicates the period of darkness.

I next examined the viral gene expression patterns to address whether viral gene expression classes (i.e., early, middle, or late (125)) could be captured in the environment. Of all the mapped reads in the Ma-LMM01 genome, 1.58–6.69%, 51.3–75.9% and 17.7–32.5% of the metatranscriptome reads were mapped to early, middle and late Ma-LMM01 genes, respectively, at each time point. The total read abundance for each gene expression class increased gradually during the daytime, peaked at 15:00, and then drastically decreased during the daytime/night transition (**Figure 3-12**). Unlike in the synchronized culture experiment (125), Ma-LMM01 did not clearly show three temporal expression classes (early, middle and late) in the environment (**Figure 3-12**). Considering the results

of the culture experiment (125), it was possible that at least two distinct transcriptional peaks corresponding to middle and late genes were observed in the environment. Therefore, this result suggested that there is a slightly different infection stage in each infected cell in the environment.

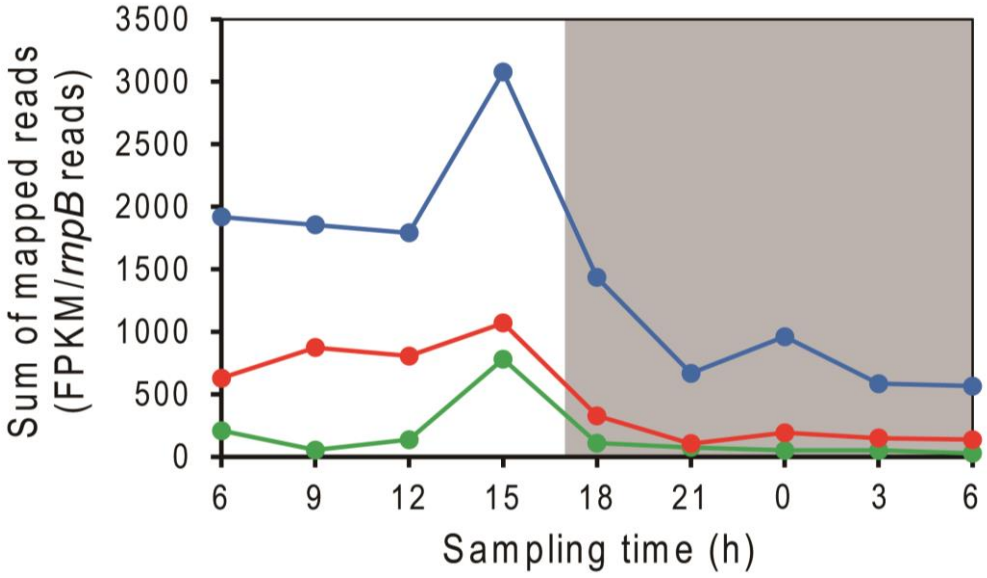


Figure 3-12. Temporal expression patterns of Ma-LMM01 viral genes in the environment. Early, middle and late genes defined in a previous culture experiment are shown in green, blue and red, respectively. The y-axis represents FPKM (fragments per kilobase per mapped million reads) normalized by *rnpB* reads as a proxy for *Microcystis* cell density. Shaded area indicates the period of darkness.

The implication for *Microcystis* antiviral responses and its viral infection profiles

M. aeruginosa reaches high cell density in the environment during the bloom (**Figure 3-1a**) (100), suggesting that this cyanobacterium can frequently be attacked by its diverse viruses. Also, this cyanobacterium possesses highly abundant host defense systems, especially, the toxin–antitoxin system (64). However, the metatranscriptome analysis revealed that the antiviral defense genes showed no constitutive expression

against the viral infection during the daytime (**Figure 3-7c and d**). The expression patterns of the toxin–antitoxin and CRISPR-Cas system-related genes seem to reach peak levels at 15:00, which might correspond to the late infection stage (**Figure 3-7c and d, Figure 3-13**). The culture-based study in chapter 2 revealed that such gene expression patterns did not occur in the control culture or in the Ma-LMM01-infected culture (125) (**Figure 3-13**). Therefore, this result indicated that *M. aeruginosa* could respond to viral infections via the expression of antiviral defense genes such as the toxin–antitoxin and CRISPR-Cas systems in the environment. Considering that higher expression levels of antiviral defense genes were observed in the environmental samples than in the culture experiment (**Figure 3-13**), *M. aeruginosa* may generally have more effective antiviral defense mechanisms that can be induced by infection with diverse viral species/strains than strain NIES-298; the Ma-LMM01-like infection profile, which does not cause significant changes in host transcriptional levels (125), is not common for *Microcystis* viruses.

Recruitment of metatranscriptome reads to *Microcystis* viral genomes in Lake Erie

Spacers matching Group I–III viruses were also acquired by *Microcystis* strains isolated in other countries (**Figure 3-14**), suggesting that similar *Microcystis* viruses exist in other freshwater ecosystems throughout the world. Therefore, I recruited the available seven metatranscriptomic read sets derived from the early, middle and late bloom in Lake Erie (130) to the MVGs. Although the transcriptional activities of the MVGs were largely influenced by the sampling time (**Figure 3-9**), the metatranscriptomic reads from Hirosawanoike Pond were apparently more preferentially recruited onto the MVGs than those from Lake Erie (**Figure 3-15**). In a previous study, Kimura *et al*. (2013) observed

that rapid gene diversification of *Microcystis* virus Ma-LMM01 *gp091* had occurred in even the same freshwater environments through host-virus interactions (80). Therefore, these results suggested that *Microcystis* viruses have diversified their genomes through host–virus co-evolution independently in each freshwater environment although similar viruses are also found in Lake Erie.

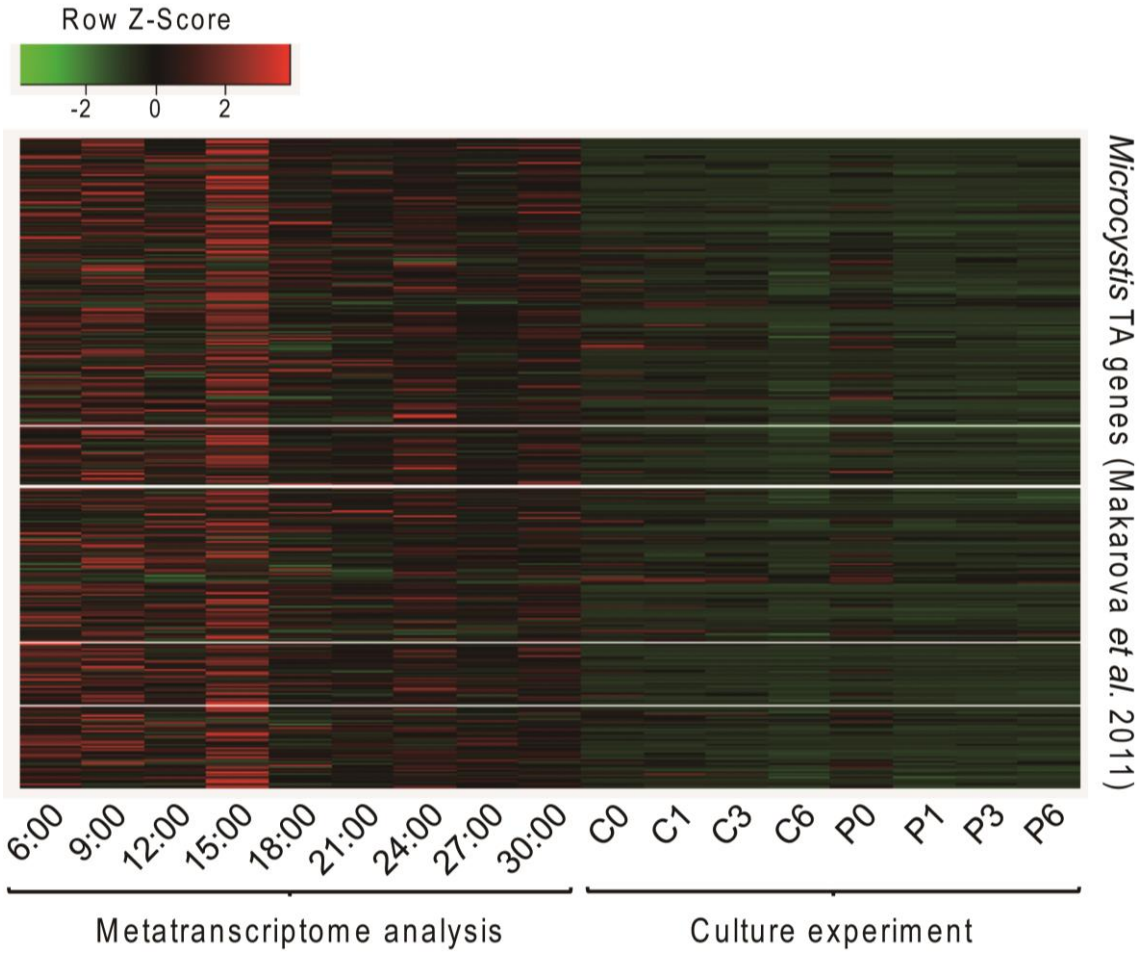


Figure 3-13. Heat map of *Microcystis* toxin–antitoxin system genes in the environment and a culture experiment. The toxin-antitoxin system genes listed in Makarova *et al*. 2011 were used in this analysis. The *Microcystis* transcriptome data for the infected (P0-P6) and control cultures (C0-C6) were obtained from Morimoto *et al*. 2018. The color gradient shows the unchanged (green) and enriched (red) gene transcripts.

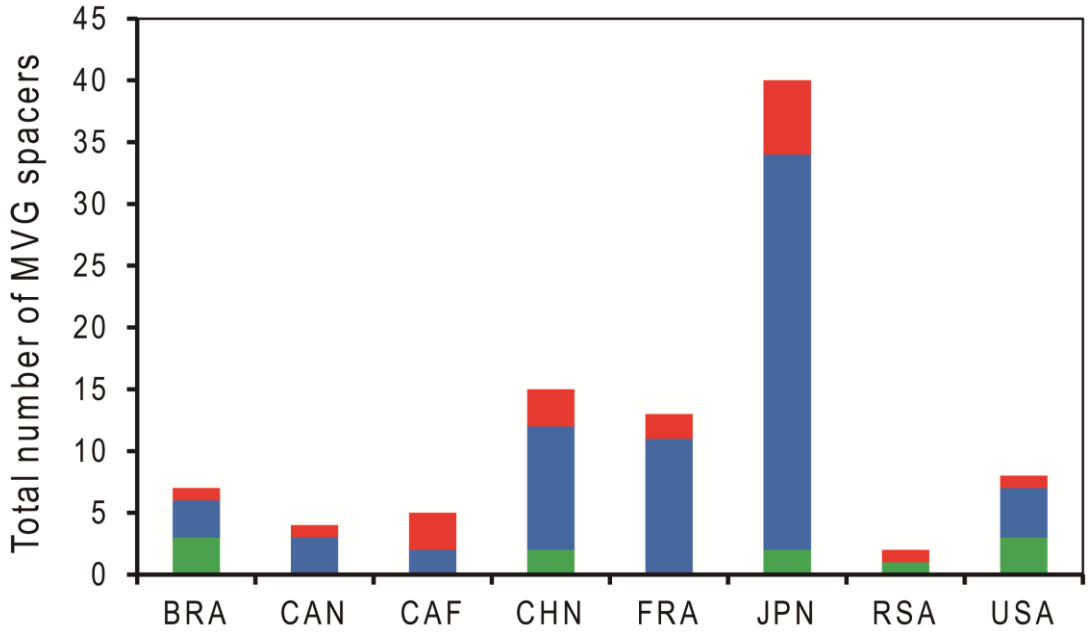


Figure 3-14. Abundance of spacers corresponding to Group I, II and III viruses in the isolated *Microcystis* genomes from each country. The x-axis represents the country from which the *Microcystis* strains possessing MVG spacers were isolated. BRA; Brazil, CAN; Canada, CAF; Central African Republic, CHN; China, FRA; France, JPN; Japan, RSA; South Africa, USA; United States of America. The y-axis represents the total number of MVG spacers. Green, blue, and red bars represent the number of Group I, II and III viral spacers, respectively.

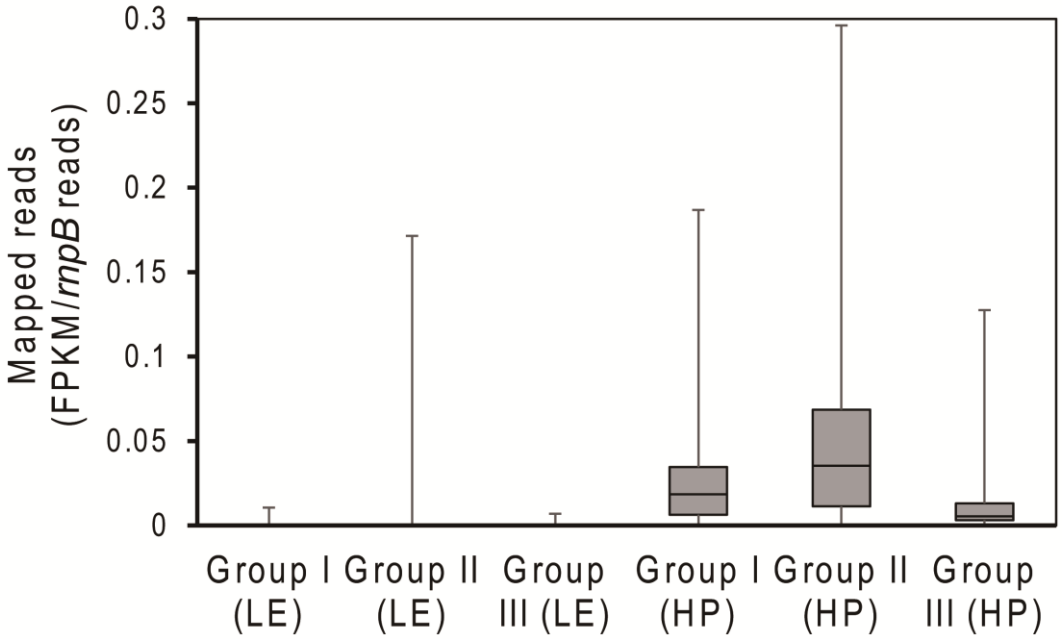


Figure 3-15. *Microcystis* viral transcript abundances for Lake Erie (LE) and Hirosawanoike Pond (HP). Sequence reads obtained from seven metatranscriptomes in Lake Erie or metatranscriptomes at each time point in Hirosawanoike Pond were mapped onto *Microcystis* viral genomes. Boxes represent the first, median and third quartiles. The y-axis represents FPKM (fragments per kilobase per mapped million reads) normalized by *rnpB* reads as a proxy for *Microcystis* cell density.

Conclusion

The interactions occurring between the toxic bloom-forming cyanobacterium *M. aeruginosa* and its infecting viruses have generated diverse host populations that possess different CRISPR arrays in their genomes (60–62). However, an important knowledge gap exists between such observations and the known Ma-LMM01 infection profile because this only isolated virus can escape from the highly abundant host defense systems (125). In this study, I revealed the existence of diverse *Microcystis* cyanoviruses, which included a new viral lineage Group III viruses. Comparing the abundant virome reads, such novel viruses are thought to be more prevalent than *Microcystis* virus Ma-LMM01 (or the corresponding contigs; MVG_NODE34 and NODE47). Notably, the Group II viruses were found to interact with the broad range of *Microcystis* strains more frequently than the Group I and III viruses that interacted with a narrower range of strains. This is supported by the highest transcriptional levels being in Group II virus. The highest transcriptional levels of these *Microcystis* viruses occurred at 12:00 or 15:00 regardless of their genomic differences. Considering that MVG_NODE620 and MVG_NODE869, which shared the Ma-LMM01 middle gene homologs, showed different expression patterns, this result suggested that diverse combinations of *M. aeruginosa* and its viruses exist in the environment. In addition, Ma-LMM01 did not show the distinct three temporal expression classes of early, middle, and late genes in the environment, suggesting that viral gene expression might differ slightly within each infected cell. On the other hand, *M. aeruginosa* was found to express antiviral defense genes such as the toxin–antitoxin and CRISPR-Cas systems in the environment, which allowed it to defend itself against viral infections. This indicated that Ma-LMM01-like infection profile, which does not affect the host's transcriptional levels to escape antiviral defense systems, is not common

in *Microcystis* viruses. Generally, narrow host-range viruses are thought to infect highly abundant hosts, whereas broad-host range viruses are assumed to infect low abundance hosts (75, 153). According to the previous study (62), the *Microcystis* bloom at Hirosawanoike Pond comprises at least 16 major and other rare CRISPR genotypes, supporting the co-occurrence of narrow- and broad-host range *Microcystis* viruses. Additionally, the transcriptional activities of the MVGs did not necessarily reflect the abundances of each viral contig in the environment (**Figure 3-10**). Given that viruses with broad-host ranges often induce different antiviral responses within each host strain (75, 154), this observation suggested that antiviral gene expression in *M. aeruginosa* inhibits viral multiplication, especially broad-range viruses like Group II. Collectively, these findings suggest that Group II viruses are the major drivers of *Microcystis* population diversification, whereas Group I and III viruses contribute to the control of *Microcystis* abundance and composition. The isolation and characterization of *Microcystis* viruses such as the Group II and III viruses described in the present study will expand our knowledge about other infection profiles in *Microcystis* viruses. Future work on the seasonal dynamics of MVGs and their hosts will also help us to further understand the viral impact on *Microcystis* blooms and their population dynamics.

Chapter 4

Integration and outlook

Toxic bloom-forming cyanobacterium *Microcystis aeruginosa* has the highest number of putative antiviral defense genes and interacts with diverse viruses in the environment, resulting in the diversification of their population. Despite such potential significance of cyanoviruses in *Microcystis* blooms, little is known about even the whole host transcriptional responses and infection process during a sole *Microcystis* virus Ma-LMM01 infection. Also, low proportion of Ma-LMM01-matching spacers suggested that numerous uncharacterized *Microcystis* viruses exist in the environment, however, no comprehensive studies have been done to investigate for the existence of other *Microcystis* viruses, or both *Microcystis* and its viral transcriptional dynamics.

In chapter 2, I first investigated the infection process and transcriptional program of Ma-LMM01, and assessed host transcriptional responses to infection using RNA-seq analysis. Strikingly, almost all of the host genes did not show a significant change in expression during Ma-LMM01 infection, however, like other lytic dsDNA viruses including marine cyanoviruses, Ma-LMM01 transcriptional programs are orchestrated in three expression classes: early (host-takeover), middle (replication), and late (virion morphogenesis). In addition, cyanobacterial primary σ factor SigA recognition-like sequences were found in the upstream region of each class genes, whereas viral specific motifs were not found. These findings suggested that unlike other known T4-like phages, Ma-LMM01 achieves three gene expression patterns without changing host promoter activity by exploiting SigA for its transcription. This type of infection may be advantageous in allowing Ma-LMM01 to escape the highly abundant host defense systems while maintaining host photosynthesis.

In chapter 3, I further investigated the genomic information and transcriptional dynamics of *Microcystis* viruses using cross-omics analysis. Virome approach revealed that three novel phylogenetic viral groups: Group I (including Ma-LMM01), II (high abundance and transcriptional activity), and III (new lineages). Of these, the Group II viruses interacted with all three phylogenetically distinct *Microcystis* phylotypes, whereas the Group I and III viruses interacted with only one or two phylotypes. This indicated the co-occurrence of broad (Group II) and narrow (Group I and III) host-range viruses in the bloom. All these viruses showed the highest transcriptional levels during daytime regardless of their genomic differences. Interestingly, metatranscriptomic approach also revealed that *M. aeruginosa* expressed antiviral defense genes against viral infection, unlike that seen with a Ma-LMM01 infection in chapter 2. Given that broad host-range viruses often induce antiviral responses within alternative hosts, these findings suggested that Group II viruses are major drivers for the diversification of *Microcystis* populations, whereas Group I and III viruses contribute to the control of *Microcystis* abundance and composition.

These studies expand our knowledge about the infection profiles, host responses, the genomic features of other *Microcystis* viruses, and the potential ecological roles of broad- and narrow-host range viruses in the bloom. Future work on the seasonal dynamics of each viral group and their hosts will also help us to further understand the viral impact on *Microcystis* blooms and their population dynamics as well as the isolation and characterization of them.

Acknowledgements

First of all, I am profoundly grateful to Professor Yoshihiko Sako for giving me a chance to study in this field, insightful comments, suggestions, and encouragement. I would like to express my gratitude to Associate Professor Takashi Yoshida for spending a lot of time to discuss with me and supporting me patiently. They gave me many opportunities to grow up as a researcher such as the collaborative project with National Research Centre in Lithuania.

I am also grateful to Professor Shigeki Sawayama for thoughtful words and reviewing this dissertation. I would like to thank Professor Hiroyuki Ogata for his constructive comments.

Furthermore, I would like to show my grateful appreciate to Dr. Sigitas Sulcius who is one of the respectable researchers. His approaches and attitudes on the research works encouraged me greatly and changed my way of thinking about the studies.

I thank all my colleagues in Laboratory of Marine Microbiology for their support and encouragement. Especially, I give a special thanks to Mr. Yuto Fukuyama and Mr. Kimiho Omae for valuable discussion, technical advices and mutual help to each other. I am also deeply grateful to Dr. Shigeko Kimura, Dr. Takashi Honda, Dr. Yosuke Nishimura, and Dr. Masao Inoue for giving me helpful comments and thoughtful attention. Many thanks for Mr. Kento Tominaga, Mr. Tatsuhiro Isozaki, Ms. Mai Nakata, Ms. Aya Sasaki, Ms. Nana Haruki, Mr. Masaya Fukuda, Mr. Naohiro Yoshida, and Mr. Satoshi Nakagawa for their supports. With many thanks to Mr. Florian Prodinger, we have discussed and done several experiments together.

Last, I appreciate sincerely to my parents and grandparents for giving me all supports and encouragements.

Acknowledgements

My doctoral works were supported by Grants-in Aids for Scientific Research (B) (No. 17H03850) and challenging Exploratory Research (No. 26660171) from the Japan Society for the Promotion of Science (JSPS), The Canon Foundation (No. 203143100025), JSPS Scientific Research on Innovative Areas (No. 16H06437), and the Bilateral Open Partnership Joint Research Project (Japan-Lithuania Research Cooperative Program) “Research on prediction of environmental change in Baltic Sea based on comprehensive metagenomic analysis of microbial viruses”.

References

1. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, Paerl HW. 2016. A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, *Microcystis* spp. *Harmful Algae* 54: 4–20.

2. Michalak AM, Anderson EJ, Beletsky D, Boland S, Bosch NS, Bridgeman TB, Chaffin JD, Cho K, Confesor R, Daloglu I, DePinto J V., Evans MA, Fahnenstiel GL, He L, Ho JC, Jenkins L, Johengen TH, Kuo KC, LaPorte E, Liu X, McWilliams MR, Moore MR, Posselt DJ, Richards RP, Scavia D, Steiner AL, Verhamme E, Wright DM, Zagorski MA. 2013. Record-setting algal bloom in Lake Erie caused by agricultural and meteorological trends consistent with expected future conditions. *Proc Natl Acad Sci* 110: 6448–6452.

3. Ullah H, Nagelkerken I, Goldenberg SU, Fordham DA. 2018. Climate change could drive marine food web collapse through altered trophic flows and cyanobacterial proliferation. *PLoS Biol* 16: e2003446.

4. Visser PM, Verspagen JMH, Sandrini G, Stal LJ, Matthijs HCP, Davis TW, Paerl HW, Huisman J. 2016. How rising CO_2 and global warming may stimulate harmful cyanobacterial blooms. *Harmful Algae* 54: 145–159.

5. Paerl HW, Huisman J. 2008. Blooms like it hot. *Science* 320: 57–58.

6. Kosten S, Huszar VLM, Bécares E, Costa LS, van Donk E, Hansson LA, Jeppesen E, Kruk C, Lacerot G, Mazzeo N, De Meester L, Moss B, Lürling M, Nõges T, Romo S, Scheffer M. 2012. Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob Chang Biol* 18: 118–126.

7. Huisman J, Codd GA, Paerl HW, Ibelings BW, Verspagen JMH, Visser PM. 2018. Cyanobacterial blooms. *Nat Rev Microbiol* 16: 471–483.

8. Reynolds CS, Rogers DA. 1976. Seasonal variations in the vertical distribution and buoyancy of *Microcystis aeruginosa* Kütz. emend. Elenkin in Rostherne Mere, England. *Hydrobiologia* 48: 17–23.

9. Walsby AE. 1994. Gas vesicles. *Microbiol Mol Biol Rev* 58: 94–144.

10. Pfeifer F. 2012. Distribution, formation and regulation of gas vesicles. *Nat Rev Microbiol* 10: 705–715.

11. Rinehart KL, Harada KI, Namikoshi M, Chen C, Harvis CA, Munro MHG, Blunt JW, Mulligan PE, Beasley VR, Dahlem AM, Carmichael WW. 1988. Nodularin, microcystin, and the configuration of Adda. *J Am Chem Soc* 110: 8557–8558.

12. Tillett D, Dittmann E, Erhard M, Dohren von H, Borner T, Neilan BA. 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol* 7: 753–764.

13. Puddick J, Prinsep MR, Wood SA, Kaufononga SAF, Cary SC, Hamilton DP. 2014. High levels of structural diversity observed in microcystins from *Microcystis* CAWBG11 and characterization of six new microcystin congeners. *Mar Drugs* 12: 5372–5395.

14. Rinehart KL, Namikoshi M, Choi BW. 1994. Structure and biosynthesis of toxins from blue-green algae (cyanobacteria). *J Appl Phycol* 6: 159–176.

15. Goldberg J, Huang H, Kwon Y, Greengard P, Nairn AC, Kuriyan J. 1995. Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-1. *Nature* 376: 745–753.

16. Maynes JT, Luu HA, Cherney MM, Andersen RJ, Williams D, Holmes CFB, James MNG. 2006. Crystal structures of protein phosphatase-1 bound to motuporin and dihydromicrocystin-LA: elucidation of the mechanism of enzyme inhibition by cyanobacterial toxins. *J Mol Biol* 356: 111–120.

17. Feurstein D, Kleinteich J, Heussner AH, Stemmer K, Dietrich DR. 2010. Investigation of microcystin congener-dependent uptake into primary murine neurons. *Environ Health Perspect* 118: 1370–1375.

18. Feurstein D, Holst K, Fischer A, Dietrich DR. 2009. Oatp-associated uptake and toxicity of microcystins in primary murine whole brain cells. *Toxicol Appl Pharmacol* 234: 247–255.

19. Milutinović A, Živin M, Zorc-Pleskovič R, Sedmak B, Šuput D. 2003. Nephrotoxic effects of chronic administration of microcystins -LR and -YR. *Toxicon* 42: 281–288.

20. Stewart I, Seawright AA, Shaw GR. 2008. Cyanobacterial poisoning in livestock, wild mammals and birds - an overview. *Adv Exp Med Biol* 619: 613–637.

21. Falconer IR. 2005. Is there a human health hazard from microcystins in the drinking water supply? *Acta Hydrochim Hydrobiol* 33: 64–71.

22. Wood SA, Dietrich DR. 2011. Quantitative assessment of aerosolized cyanobacterial toxins at two New Zealand lakes. *J Environ Monit* 13: 1617–1624.

23. Poste AE, Hecky RE, Guildford SJ. 2011. Evaluating microcystin exposure risk through fish consumption. *Environ Sci Technol* 45: 5806–5811.

24. Azevedo SMF., Carmichael WW, Jochimsen EM, Rinehart KL, Lau S, Shaw GR, Eaglesham GK. 2002. Human intoxication by microcystins during renal dialysis treatment in Caruaru-Brazil. *Toxicology* 181–182: 441–446.

25. WHO. 2003. Cyanobacterial toxins: microcystin-LR in drinking water. In: Organization W.H. (Eds.), Background Document for Preparation of WHO Guidelines for Drinking-water Quality. World Health Organization, Geneva.

26. Paerl HW, Gardner WS, Mccarthy MJ, Peierls BL, Wilhelm SW. 2014. Algal blooms : noteworthy nitrogen. Science 346: 175.

27. Paerl HW, Huisman J. 2009. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environ Microbiol Rep* 1: 27–37.

28. Paerl HW, Xu H, Hall NS, Rossignol KL, Joyner AR, Zhu G, Qin B. 2015. Nutrient limitation dynamics examined on a multi-annual scale in Lake Taihu, China: implications for controlling eutrophication and harmful algal blooms. *J Freshw Ecol* 30: 5–24.

29. Paerl H. 2014. Mitigating harmful cyanobacterial blooms in a human- and climatically-impacted world. *Life* 4: 988–1012.

30. Chaffin JD, Bridgeman TB. 2014. Organic and inorganic nitrogen utilization by nitrogen-stressed cyanobacteria during bloom conditions. *J Appl Phycol* 26: 299–309.

31. Orihel DM, Bird DF, Brylinsky M, Chen H, Donald DB, Huang DY, Giani A, Kinniburgh D, Kling H, Kotak BG, Leavitt PR, Nielsen CC, Reedyk S, Rooney RC, Watson SB, Zurawell RW, Vinebrooke RD, Smith REH. 2012. High microcystin concentrations occur only at low nitrogen-to-phosphorus ratios in nutrient-rich Canadian lakes. *Can J Fish Aquat Sci* 69: 1457–1462.

32. Smith VH. 1983. Low nitrogen to phosphorus ratios favor dominance by blue-green algae in lake phytoplankton. *Science* 221: 669–671.

33. Xie L, Xie P, Li S, Tang H, Liu H. 2003. The low TN:TP ratio, a cause or a result of *Microcystis* blooms? *Water Res* 37: 2073–2080.

34. Peng G, Wilhelm SW, Lin S, Wang X. 2018. Response of *Microcystis aeruginosa* FACHB-905 to different nutrient ratios and changes in phosphorus chemistry. *J Oceanol Limnol* 36: 1040–1052.

35. Marinho MM, De Oliveira E Azevedo SMF. 2007. Influence of N/P ratio on competitive abilities for nitrogen and phosphorus by *Microcystis aeruginosa* and *Aulacoseira distans*. *Aquat Ecol* 41: 525–533.

36. Nalewajko C, Murphy TP. 2001. Effects of temperature, and availability of nitrogen and phosphorus on the abundance of *Anabaena* and *Microcystis* in Lake Biwa, Japan: an experimental approach. *Limnology* 2: 45–48.

37. Paerl HW, Otten TG. 2013. Harmful cyanobacterial blooms: causes, consequences, and controls. *Microb Ecol* 65: 995–1010.

38. Sunda WG, Graneli E, Gobler CJ. 2006. Positive feedback and the development and persistence of ecosystem disruptive algal blooms. *J Phycol* 42: 963–974.

39. Smayda TJ. 2008. Complexity in the eutrophication-harmful algal bloom relationship, with comment on the importance of grazing. *Harmful Algae* 8: 140–151.

40. Wilson AE, Sarnelle O, Tillmanns AR. 2006. Effects of cyanobacterial toxicity and morphology on the population growth of freshwater zooplankton: meta-analyses of laboratory experiments. Limnol Oceanogr 51:1915–1924.

41. Yang Z, Kong F, Shi X, Cao H. 2006. Morphological response of *Microcystis aeruginosa* to grazing by different sorts of zooplankton. *Hydrobiologia* 563: 225–230.

42. Suttle CA. 2005. Viruses in the sea. *Nature* 437: 356–361.

43. Roucourt B, Lavigne R. 2009. The role of interactions between phage and bacterial proteins within the infected cell: a diverse and puzzling interactome. *Environ Microbiol* 11: 2789–2805.

44. Fuhrman JA. 1999. Marine viruses and their biogeochemical and ecological effects. *Nature* 399: 541–548.

45. Suttle CA. 2007. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* 5: 801–812.

46. Phlips EJ, Monegue RL, Aldridge FJ. 1990. Cyanophages which impact bloom-forming cyanobacteria. *J Aquat plant Manag* 28: 92–97.

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

48. Manage PM, Kawabata Z, Nakano S. 1999. Seasonal changes in densities of cyanophage infectious to *Microcystis aeruginosa* in a hypereutrophic pond. Hydrobiologia 411: 211–216.

49. 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.

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

51. Takashima Y, Yoshida T, Yoshida M, Shirai Y, Tomaru Y, Takao Y, Hiroishi S, Nagasaki K. 2007. Development and application of quantitative detection of cyanophages phylogenetically related to cyanophage Ma-LMM01 infecting *Microcystis aeruginosa* in fresh water. *Microbes Env* 22: 207–213.

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

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

54. Mankiewicz-Boczek J, Jaskulska A, Pawełczyk J, Gągała I, Serwecińska L, Dziadek J. 2016. Cyanophages infection of *Microcystis* bloom in lowland dam reservoir of Sulejów, Poland. *Microb Ecol* 71: 315–325.

55. Stough JMA, Tang X, Krausfeldt LE, Steffen MM, Gao G, Boyer GL, Wilhelm SW. 2017. Molecular prediction of lytic vs lysogenic states for *Microcystis* phage: metatranscriptomic evidence of lysogeny during large bloom events. *PLoS One* 12: e0184146.

56. Steffen MM, Belisle BS, Watson SB, Boyer GL, Bourbonniere RA, Wilhelm SW. 2015. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Appl Environ Microbiol* 81: 3268–3276.

57. Ou T, Li S, Liao X, Zhang Q. 2013. Cultivation and characterization of the MaMV-DC cyanophage that infects bloom-forming cyanobacterium *Microcystis aeruginosa*. *Virol Sin* 28: 266–271.

58. Ou T, Gao XC, Li SH, Zhang QY. 2015. Genome analysis and gene nblA identification of *Microcystis aeruginosa* myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *J Gen Virol* 96: 3681–3697.

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

60. Kuno S, Sako Y, Yoshida T. 2014. Diversification of CRISPR within coexisting genotypes in a natural population of the bloom-forming cyanobacterium *Microcystis aeruginosa*. *Microbiology* 160: 903–916.

61. Kuno S, Yoshida T, Kaneko T, Sako Y. 2012. Intricate interactions between the bloom-forming cyanobacterium *Microcystis aeruginosa* and foreign genetic elements, revealed by diversified clustered regularly interspaced short palindromic repeat (CRISPR) signatures. *Appl Environ Microbiol* 78: 5353–5360.

62. Kimura S, Uehara M, Morimoto D, Yamanaka M, Sako Y, Yoshida T. 2018. Incomplete selective sweeps of *Microcystis* population detected by the leader-end CRISPR fragment analysis in a natural pond. *Front Microbiol* 9: article 425.

63. Yang C, Lin F, Li Q, Li T, Zhao J. 2015. Comparative genomics reveals diversified CRISPR-Cas systems of globally distributed *Microcystis aeruginosa*, a freshwater bloom-forming cyanobacterium. *Front Microbiol* 6: article 394.

64. Makarova KS, Wolf YI, Snir S, Koonin E V. 2011. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *J Bacteriol* 193: 6039–6056.

65. Miller ES, Kutter E, Mosig G, Arisaka F, Kunisawa T, Ruger W. 2003. Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67: 86–156.

66. Luke K, Radek A, Liu X, Campbell J, Uzan M, Haselkorn R, Kogan Y. 2002. Microarray analysis of gene expression during bacteriophage T4 infection. *Virology* 299: 182–191.

67. Rohrer H, Zillig W, Mailhammer R. 1975. ADP-ribosylation of DNA-dependent RNA polymerase of *Escherichia coli* by an NAD+: protein ADP-ribosyltransferase from Bacteriophage T4. *FEBS J* 60: 227–238.

68. Goff CG, Setzer J. 1980. ADP ribosylation of *Escherichia coli* RNA polymerase is nonessential for Bacteriophage T4 development. *J Virol* 33: 547–549.

69. Marshall P, Sharma M, Hinton DM. 1999. The bacteriophage T4 transcriptional activator MotA accepts various base-pair changes within its binding sequence. *J Mol Biol* 285: 931–44.

70. Sharma M, Marshall P, Hinton DM. 1999. Binding of the bacteriophage T4 transcriptional activator, MotA, to T4 middle promoter DNA: evidence for both major and minor groove contacts. *J Mol Biol* 290: 905–915.

71. Sanson B, Hu R-M, Troitskaya E, Mathy N, Uzan M. 2000. Endoribonuclease RegB from bacteriophage T4 is necessary for the degradation of early but not middle or late mRNAs. *J Mol Biol* 297: 1063–1074.

72. Tiemann B, Depping R, Gineikiene E, Kaliniene L, Nivinskas R, Rüger W. 2004. ModA and ModB, two ADP-ribosyltransferases encoded by bacteriophage T4: catalytic properties and mutation analysis. *J Bacteriol* 186: 7262–7272.

73. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, DeFrancesco AS, Kern SE, Thompson LR, Young S, Yandava C, Fu R, Krastins B, Chase M, Sarracino D, Osburne 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.

74. Clokie MRJ, Shan J, Bailey S, Jia Y, Krisch HM, West S, Mann NH. 2006. Transcription of a “photosynthetic” T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* 8: 827–835.

75. Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, Feingersch R, Steglich C, Futschik M, Lindell D, Sorek R. 2016. Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* 10: 1437–1455.

76. Lin X, Ding H, Zeng Q. 2016. Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. *Environ Microbiol* 18: 450–460.

77. Thompson LR, Zeng Q, Chisholm SW. 2016. Gene expression patterns during light and dark infection of Prochlorococcus by cyanophage. *PLoS One* 11: e0165375.

78. Mann NH, Clokie MRJ, 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.

79. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, Chisholm SW. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci* 108: E757–E764.

80. Kimura S, Sako Y, Yoshida T. 2013. Rapid *Microcystis* cyanophage gene diversification revealed by longand short-term genetic analyses of the tail sheath gene in a natural pond. *Appl Environ Microbiol* 79: 2789–2795.

81. Yoshida-Takashima Y, Yoshida M, Ogata H, Nagasaki K, Hiroishi S, Yoshida T. 2012. Cyanophage infection in the bloom-forming cyanobacteria *Microcystis aeruginosa* in surface freshwater. *Microbes Environ* 27: 350–355.

82. Gao E-B, Gui J-F, Zhang Q-Y. 2012. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. *J Virol* 86: 236–245.

83. Kasai F, Kawachi M, Erata M, Watanabe MM. 2004. NIES-collection list of strains seventh edition 2004 microalgae and protozoa. National Institute for Environmental Studies, Tsukuba.

84. Honda T, Takahashi H, Sako Y, Yoshida T. 2014. Gene expression of *Microcystis aeruginosa* during infection of cyanomyovirus Ma-LMM01. *Fish Sci* 80: 83–91.

85. Yoshida T, Maki M, Okamoto H, Hiroishi S. 2005. Coordination of DNA replication and cell division in cyanobacteria *Microcystis aeruginosa*. *FEMS Microbiol Lett* 251: 149–154.

86. Wyshak G, Detre K. 1972. Estimating the number of organisms in quantal assays. *Appl Microbiol* 23: 784–90.

87. Stewart FJ. 2013. Preparation of microbial community cDNA for metatranscriptomic analysis in marine plankton. *Methods Enzymol*, 1st ed. 531: 187–218.

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

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

90. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19: 455–477.

91. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29: 2607–2618.

92. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9: 357–360.

93. Hammer Ø, Harper DAT a. T, Ryan PD. 2001. PAST: paleontological statistics software package for education and data analysis. *Palaeontol Electron* 4: article 9.

94. Robinson TJ, Thorvaldsdóttir H, Winckler W, Guttman M, Lander SE, Getz G, Mesirov PJ. 2011. Integrative genomics viewer. *Nat Biotechnol* 29: 24–26.

95. Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genomic Biol* 11: R106.

96. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: article 550.

97. Gentleman R, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JYH, Zhang J. 2004. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5: R80.

98. Larkin MA, Blackshields G, Brown NP, Chenna R, Mcgettigan PA, Mcwilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

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

100. Kimura S, Yoshida T, Hosoda N, Honda T, Kuno S, Kamiji R, Hashimoto R, Sako Y. 2012. Diurnal infection patterns and impact of *Microcystis* cyanophages in a Japanese pond. *Appl Environ Microbiol* 78: 5805–5811.

101. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM, Chisholm SW. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449: 83–86.

102. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Kerrie B, DeLong EF, Chisholm SW. 2006. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* 311: 1768–1770.

103. Thompson AW, Huang K, Saito MA, Chisholm SW. 2011. Transcriptome response of high- and low-light-adapted *Prochlorococcus* strains to changing iron availability. *ISME J* 5: 1580–1594.

104. Stuart RK, Brahamsha B, Busby K, Palenik B. 2013. Genomic island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper and oxidative stress. *ISME J* 7: 1139–49.

105. Tetu SG, Johnson DA, Varkey D, Phillippy K, Stuart RK, Dupont CL, Hassan KA, Palenik B, Paulsen IT. 2013. Impact of DNA damaging agents on genome-wide transcriptional profiles in two marine *Synechococcus* species. *Front Microbiol* 4: article 232.

106. Poranen MM, Ravantti JJ, Grahn AM, Gupta R, Auvinen P, Bamford DH. 2006. Global changes in cellular gene expression during bacteriophage PRD1 infection. *J Virol* 80: 8081–8088.

107. Ravantti JJ, Ruokoranta TM, Alapuranen AM, Bamford DH. 2008. Global transcriptional responses of *Pseudomonas aeruginosa* to phage PRR1 infection. *J Virol* 82: 2324–2329.

108. Lobato-Márquez D, Díaz-Orejas R, Portillo FG. 2016. Toxin-antitoxins and bacterial virulence. *FEMS Microbiol Rev* 40: 592–609.

109. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, van der Oost J, Backofen R, Koonin E V. 2015. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* 13: 722–736.

110. De Smet J, Zimmermann M, Kogadeeva M, Ceyssens PJ, Vermaelen W, Blasdel B, Bin Jang H, Sauer U, Lavigne R. 2016. High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection. *ISME J* 10: 1823–1835.

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

112. Imamura S, Asayama M. 2009. Sigma factors for cyanobacterial transcription. *Gene Regul Syst Bio* 3: 65–87.

113. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, Post AF, Hagemann M, Paulsen I, Partensky F. 2009. Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* 73: 249–299.

114. Collier JL, Grossman AR. 1994. A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J* 13: 1039–1047.

115. Bryan D, El-Shibiny A, Hobbs Z, Porter J, Kutter EM. 2016. Bacteriophage T4 infection of stationary phase *E. coli*: life after log from a phage perspective. *Front Microbiol* 7: article 1391.

116. Zeng Q, Chisholm SW. 2012. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Curr Biol* 22: 124–128.

117. Ainsworth S, Zomer A, Mahony J, Sinderen D van. 2013. Lytic infection of *Lactococcus lactis* by bacteriophages Tuc2009 and C2 triggers alternative transcriptional host responses. *Appl Environ Microbiol* 79: 4786–4798.

118. Lavigne R, Lecoutere E, Wagemans J, Cenens W, Aertsen A, Schoofs L, Landuyt B, Paeshuyse J, Scheer M, Schobert M, Ceyssens P-J. 2013. A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *mBio* 4: e00061-13.

119. MacKintosh C, Beattie KA, Klumpp S, Cohen P, Codd GA. 1990. Cyanobacterial microcystin-LR is a potent and specific inhibitor of protein phosphatases 1 and 2A from both mammals and higher plants. *FEBS* 264: 187–192.

120. Nishiwaki-Matsushima R, Ohta T, Nishiwaki S, Suganuma M, Kohyama K, Ishikawa T, Carmichael WW, Fujiki H. 1992. Liver tumor promotion by the cyanobacterial cyclic peptide toxin microcystin-LR. *J Cancer Res Clin Oncol* 118: 420–424.

121. Yoshizawa S, Matsushima R, Watanabe MF, Harada K ichi, Ichihara A, Carmichael WW, Fujiki H. 1990. Inhibition of protein phosphatases by microcystis and nodularin associated with hepatotoxicity. *J Cancer Res Clin Oncol* 116: 609–614.

122. Otten TG, Paerl HW. 2015. Health effects of toxic cyanobacteria in U.S. drinking and recreational waters: our current understanding and proposed direction. *Curr Environ Heal reports* 2: 75–84.

123. Koonin E V., Makarova KS, Wolf YI. 2017. Evolutionary genomics of defense systems in archaea and bacteria. *Annu Rev Microbiol* 71: 233–261.

124. Krupovic M, Dutilh BE, Adriaenssens EM, Wittmann J, Vogensen FK, Sullivan MB, Rumnieks J, Prangishvili D, Lavigne R, Kropinski AM, Klumpp J, Gillis A, Enault F, Edwards RA, Duffy S, Clokie MRC, Barylski J, Ackermann H-W, Kuhn JH. 2016. Taxonomy of prokaryotic viruses: update from the ICTV bacterial and archaeal viruses subcommittee. *Arch Virol* 161: 1095–1099.

125. Morimoto D, Kimura S, Sako Y, Yoshida T. 2018. Transcriptome analysis of a bloom-forming cyanobacterium *Microcystis aeruginosa* during Ma-LMM01 phage infection. *Front Microbiol* 9: article 2.

126. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537: 689–693.

127. Nishimura Y, Watai H, Honda T, Mihara T, Kimiho O, Roux S, Blanc-Mathieu R, Yamamoto K, Hingamp P, Sako Y, Sullivan MB, Goto S, Ogata H, Yoshida T. 2017. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* 2: e00359-16.

128. Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, Rubin E, Ivanova NN, Kyrpides NC. 2016. Uncovering Earth's virome. *Nature* 536: 425–430.

129. Penn K, Wang J, Fernando SC, Thompson JR. 2014. Secondary metabolite gene expression and interplay of bacterial functions in a tropical freshwater cyanobacterial bloom. *ISME J* 8: 1866–1878.

130. Meyer KA, Davis TW, Watson SB, Denef VJ, Berry MA, Dick GJ. 2017. Genome sequences of lower Great Lakes *Microcystis* sp. reveal strain-specific genes that are present and expressed in western Lake Erie blooms. *PLoS One* 12: e0183859.

131. Chen Z, Zhang J, Li R, Tian F, Shen Y, Xie X, Ge Q, Lu Z. 2018. Metatranscriptomics analysis of cyanobacterial aggregates during cyanobacterial bloom period in Lake Taihu, China. *Environ Sci Pollut Res* 25: 4811–4825.

132. Sato N. 1995. A family of cold-regulated RNA-binding protein genes in the cyanobacterium *Anabaena variabilis* M3. *Nucleic Acids Res* 23: 2161–2167.

133. Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3: e985.

134. Ahlgren NA, Ren J, Lu YY, Fuhrman JA, Sun F. 2017. Alignment-free d2∗ oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res* 45: 39–53.

135. Li Q, Lin F, Yang C, Wang J, Lin Y, Shen M, Park MS, Li T, Zhao J. 2018. A large-scale comparative metagenomic study reveals the functional interactions in six bloom-forming *Microcystis*-epibiont communities. *Front Microbiol* 9: article 746.

136. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P. 2007. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics* 8: article 209.

137. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32: 11–16.

138. Nishimura Y, Yoshida T, Kuronishi M, Uehara H, Ogata H, Goto S. 2017. ViPTree: the viral proteomic tree server. *Bioinformatics* 33: 2379–2380.

139. Grazziotin AL, Koonin E V., Kristensen DM. 2017. Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Res* 45: D491–D498.

140. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7: e1002195.

141. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33: 1870–1874.

142. Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28: 3211–3217.

143. Dwivedi B, Xue B, Lundin D, Edwards RA, Breitbart M. 2013. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol Biol* 13: article 33.

144. 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.

145. 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.

146. Ponsero AJ, Chen F, Lennon JT, Wilhelm SW. 2013. Complete genome sequence of cyanobacterial siphovirus KBS2A. *Genome Announc* 1: e00472-13.

147. Kristensen DM, Waller AS, Yamada T, Bork P, Mushegian AR, Koonin E V. 2013. Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J Bacteriol* 195: 941–950.

148. Otsuka S, Suda S, Li R, Watanabe M, Oyaizu H, Matsumoto S, Watanabe MM. 1999. Phylogenetic relationships between toxic and non-toxic strains ofthe genus *Microcystis* based on 16S to 23S internal transcribedspacer sequence. *FEMS Microbiol Lett* 172: 15–21.

149. Straub C, Quillardet P, Vergalli J, de Marsac NT, Humbert JF. 2011. A day in the life of *Microcystis aeruginosa* strain PCC 7806 as revealed by a transcriptomic analysis. *PLoS One* 6: e16208.

150. Ribalet F, Swalwell J, Clayton S, Jiménez V, Sudek S, Lin Y, Johnson ZI, Worden AZ, Armbrust EV. 2015. Light-driven synchrony of *Prochlorococcus* growth and mortality in the subtropical Pacific gyre. *Proc Natl Acad Sci* 112: 8008–8012.

151. Clokie MRJ, Millard AD, Mehta JY, Mann NH. 2006. Virus isolation studies suggest short-term variations in abundance in natural cyanophage populations of the Indian Ocean. *J Mar Biol Assoc United Kingdom* 86: 499–505.

152. Yoshida T, Nishimura Y, Watai H, Haruki N, Morimoto D, Kaneko H, Honda T, Yamamoto K, Hingamp P, Sako Y, Goto S, Ogata H. 2018. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. *ISME J* 12: 1287–1295.

153. Dekel-Bird NP, Sabehi G, Mosevitzky B, Lindell D. 2015. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. *Environ Microbiol* 17: 1286–1299.

154. Howard-Varona C, Roux S, Dore H, Solonenko NE, Holmfeldt K, Markillie LM, Orr G, Sullivan MB. 2017. Regulation of infection efficiency in a globally abundant marine *Bacteriodetes* virus. *ISME J* 11: 284–295.

Publication list

1. Morimoto D, Kimura S, Sako Y and Yoshida T. Transcriptome analysis of a bloom-forming cyanobacterium *Microcystis aeruginosa* during Ma-LMM01 phage infection. (2018) *Front. Microbiol.* 9: 1-13.
2. Morimoto D, Tominaga K, Nishimura Y, Yoshida N, Kimura S, Sako Y and Yoshida T. Co-occurrence of broad and narrow host range viruses infecting a toxic bloom-forming cyanobacterium *Microcystis aeruginosa* revealed by cross-omics analysis. (2018) *submitted*.
3. Yoshida T, Nishimura Y, Watai H, Haruki N, Morimoto D, Kaneko H, Honda T, Yamamoto K, Hingamp P, Sako Y, Goto S, Ogata H. Locality and diel cycling of viral production revealed by a 24 h time course cross-omics analysis in a coastal region of Japan. (2018) *ISME J.* 12: 1287-1295
4. Kimura S, Uehara M, Morimoto D, Yamanaka M, Sako Y, Yoshida T. Incomplete selective sweeps of *Microcystis* population detected by the leader-end CRISPR fragment analysis in a natural pond. (2018) *Front. Microbiol.* 9: 1-9.
5. Honda T, Morimoto D, Sako Y, Yoshida T. LexA binds to transciption regulatory site of cell division gene *ftsZ* in toxic cyanobacterium *Microcystis aeruginosa*. (2018) *Mar. Biotechnol.* 20: 549-556.
6. Yoshida T, Morimoto D, Kimura S. DNA Traffic in the Environment: Bacteria-virus interactions (Chapter 5). *Springer Book*. (accepted)