

RESEARCH ARTICLE

# Molecular prediction of lytic vs lysogenic states for *Microcystis* phage: Metatranscriptomic evidence of lysogeny during large bloom events

Joshua M. A. Stough<sup>1</sup>, Xiangming Tang<sup>2</sup>, Lauren E. Krausfeldt<sup>1</sup>, Morgan M. Steffen<sup>3</sup>, Guang Gao<sup>2</sup>, Gregory L. Boyer<sup>4</sup>, Steven W. Wilhelm<sup>1\*</sup>

**1** Department of Microbiology, University of Tennessee, Knoxville, TN, United States of America, **2** State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography & Limnology, Chinese Academy of Sciences, Nanjing, PR China, **3** Department of Biology, James Madison University, Harrisonburg, VA, United States of America, **4** College of Environmental Science and Forestry, The State University of New York, Syracuse, NY, United States of America

\* [wilhelm@utk.edu](mailto:wilhelm@utk.edu)



**OPEN ACCESS**

**Citation:** Stough JMA, Tang X, Krausfeldt LE, Steffen MM, Gao G, Boyer GL, et al. (2017) Molecular prediction of lytic vs lysogenic states for *Microcystis* phage: Metatranscriptomic evidence of lysogeny during large bloom events. PLoS ONE 12 (9): e0184146. <https://doi.org/10.1371/journal.pone.0184146>

**Editor:** Connie Lovejoy, Laval University, CANADA

**Received:** May 15, 2017

**Accepted:** August 18, 2017

**Published:** September 5, 2017

**Copyright:** © 2017 Stough et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information file. Transcriptomic data are archived in the MG-RAST server as noted within the paper.

**Funding:** This work was supported by the National Natural Science Foundation of China, grants 41471040 and 41230744 to GG, the Division of Integrative Organismal Systems, grant 1451528 to SWW, the Division of Environmental Biology, grant 1240870 to SWW, the Gordon and Betty Moore Foundation, grant EMS#4971 to SWW and by the

## Abstract

*Microcystis aeruginosa* is a freshwater bloom-forming cyanobacterium capable of producing the potent hepatotoxin, microcystin. Despite increased interest in this organism, little is known about the viruses that infect it and drive nutrient mobilization and transfer of genetic material between organisms. The genomic complement of sequenced phage suggests these viruses are capable of integrating into the host genome, though this activity has not been observed in the laboratory. While analyzing RNA-sequence data obtained from *Microcystis* blooms in Lake Tai (*Taihu*, China), we observed that a series of lysogeny-associated genes were highly expressed when genes involved in lytic infection were down-regulated. This pattern was consistent, though not always statistically significant, across multiple spatial and temporally distinct samples. For example, samples from Lake Tai (2014) showed a predominance of lytic virus activity from late July through October, while genes associated with lysogeny were strongly expressed in the early months (June–July) and toward the end of bloom season (October). Analyses of whole phage genome expression shows that transcription patterns are shared across sampling locations and that genes consistently clustered by co-expression into lytic and lysogenic groups. Expression of lytic-cycle associated genes was positively correlated to total dissolved nitrogen, ammonium concentration, and salinity. Lysogeny-associated gene expression was positively correlated with pH and total dissolved phosphorous. Our results suggest that lysogeny may be prevalent in *Microcystis* blooms and support the hypothesis that environmental conditions drive switching between temperate and lytic life cycles during bloom proliferation.

Kenneth & Blaire Mossman Endowment to SWW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Viruses are one of the most potent drivers of nutrient cycles, horizontal gene transfer, and microbial evolution in aquatic ecosystems [1, 2]. Bacteriophage play an important role in microbial communities by lysing primary producers and heterotrophic bacteria, releasing nutrients from biomass [3]. Moreover, due to their density-dependent infection, viruses are thought to reduce the competitive advantages of some of the most prolific organisms—the “kill-the-winner” hypothesis [4]. Phage genomes also can encode auxiliary metabolic genes that serve to augment host metabolism during infection, considerably altering the functional potential of entire populations within the microbial community [5, 6]. Despite their recognized importance, much of the potential of viruses remains uncharacterized, highlighting a crucial need for examination of the role they play across ecosystems.

*Microcystis aeruginosa* has repeatedly been identified as a nuisance bloom-former in freshwater systems over the last several decades [7]. It has come to the forefront of public attention as the primary agent in blooms worldwide and for its ability to produce a potent hepatotoxin, originally known as “Fast-Death Factor” [8], but now known as microcystin [9, 10]. Recent impacts include the shutdown of the public water supply to the City of Toledo (Ohio) during the *Microcystis* bloom in 2014 [11], and the considerable accumulation of toxic algal biomass in Lake Tai, China (*Taihu* in Chinese) [12, 13]. While significant strides have been made describing the ecology [14–16], physiology [17–19], and genetics [20–22] of *Microcystis*, little is known about the effect of phage on *Microcystis* ecology. To date, only 11 viruses infecting *M. aeruginosa* have ever been brought into culture [23–28], of which only 2 have sequenced genomes [29, 30], and each of these isolates has subsequently been lost to science. *Microcystis* phage Ma-LMM01, classified as an unassigned myovirus, has been the best studied. The availability of Ma-LMM01’s full genome sequence has led to analyses of distribution (*via* PCR and qPCR-based techniques) and some characterization of its genetic regulation [31, 32].

Ma-LMM01 appears to have been host specific in lab studies, targeting *M. aeruginosa* at the strain level [27]. This has led to the hypothesis that phage play a role in modulating dominant strains during blooms [33]. Ecologically, one gene from this virus (*gp91*), encoding a viral tail sheath and present in the genomes of both *Microcystis* phages Ma-LMM01 and MaMV-DC [29, 30], has been used *via* qPCR to suggest *Microcystis*-specific phage particles can be present at concentrations  $>10,000 \text{ mL}^{-1}$  of lake water [31, 34]. These virus densities and a projected high level of host specificity suggest the potential for long-term predator-prey coevolution between virus and host, a trait generally associated with temperate phage [35]. They also suggest that bloom events of susceptible *Microcystis* cells should quickly succumb to phage infection [4].

Beyond an ability to infect and lyse *Microcystis*, the Ma-LMM01 genome encodes machinery necessary for lysogeny and induction, including 3 transposases, a serine recombinase, and 2 prophage anti-repressors. In addition, one transposase (*gp135*) and the recombinase (*gp136*) make up a 2-gene mobile genetic element called IS607, originally identified in *Helicobacter pylori*, and has led some to hypothesize that these genes further act independently as a transposon [36, 37]. Although there is an absence of lysogenic activity with *Microcystis* observed in the laboratory, expression of these genes has been documented in environmental samples (although they were not tied to lysogeny [38]). Taken together, the presence of lysogeny-associated genes within *Microcystis* and the implied protection against superinfection might explain how this genus can come to dominate freshwater ecosystems and escape Hutchinson’s *Paradox of the Plankton* [39] or the “kill-the-winner” phenomenon [4].

During analyses of metatranscriptomic data from *Microcystis* blooms in Lake Tai, we observed expression of phage-encoded lysogeny-associated genes that negatively correlated

with expression of genes consistent with lytic infection and phage replication. Regulation of these putative lysogenic genes appears to be strongly associated with specific environmental conditions in the water column. Based on these observations, we hypothesize that a phage lysogenizes the *Microcystis* bloom community in a matter constrained by nitrogen and phosphorus availability.

## Materials and methods

### Sample collection and survey of environmental conditions

No specific permissions were required for these locations/activities as Lake Tai is open access water. Research did not involve endangered or protected species. Samples were obtained from Lake Tai over the course of five months during the *M. aeruginosa* bloom in 2014 and have been used in conjunction with several other experiments [12]. Surface water samples were collected monthly from June to October from 11 different locations across the lake (S1 Table). From all stations and dates 35 samples were selected (based on the quality and quantity of extracted RNA) were submitted for RNA-seq. Samples from Lake Tai (25–180 mLs) were collected on 0.2- $\mu\text{m}$  nominal pore-size Sterivex™ (EMD Millipore Corporation, Darmstadt, Germany) and preserved for transport by adding ~ 2 mL of RNAlater (ThermoFisher Scientific, Waltham, MA).

Water column depth and Secchi depth (SD) were measured using a water depth gauge (Uwitec, Austria) and Secchi disk, respectively. Water temperature, electrical conductivity (EC), pH, dissolved oxygen (DO) and phycocyanobilin (PC) were measured *in situ* using a multiparameter water quality sonde (YSI 6600 V2, Yellow Springs Instruments Inc., USA). Total nitrogen (TN), total dissolved nitrogen (TDN), ammonium ( $\text{NH}_4$ ), nitrate ( $\text{NO}_3$ ), total phosphorus (TP), total dissolved phosphorus (TDP), orthophosphate ( $\text{PO}_4$ ), total dissolved solids (TDS), and chlorophyll *a* (chl *a*) were all measured according to standard methods [40].

Cyanobacterial toxins were determined using liquid chromatography coupled with mass spectroscopy as previously described [41]. Fourteen common microcystin congeners were determined by reverse phase liquid chromatography (microcystins RR, dRR, mRR, hYR, YR, LR, mLr, dLR, AR, FR, LA, LW, LF, WR and R-NOD) using a Waters ZQ4000 mass spectrometer coupled with a photodiode array spectrometer. Microcystins were all quantified against a microcystin-LR standard, and their presence confirmed using diagnostic ADDA UV signatures. We also looked for anatoxin-a (ATX), homoanatoxin-a, cylindrospermopsin (CYL) and deoxycylindrospermopsin in these extracts using HPLC coupled with mass selective (LCMS) or tandem mass (LC-MS/MS: Waters TQD) detection, and quantified against respective standards. Method detection limits were dependent on the volume filtered, ranging from 0.1–0.3  $\mu\text{g}$  MC-LR / L and were less than 0.01  $\mu\text{g}$ /L for anatoxin-a, cylindrospermopsin, and their variants.

### RNA extraction and sequencing

Total RNA was extracted using the MOBIO PowerWater (now Qiagen DNeasy PowerWater) DNA isolation kit for Sterivex (Qiagen, San Diego, CA) modified and optimized for RNA isolation. RNA concentration and purity were determined using a NanoDrop™ ND-1000 spectrophotometer. Extracted RNA was tested for DNA contamination by running a polymerase chain reaction using universal bacterial 16S rDNA primers 27F and 1522R (sensitivity ~ 10 gene copies per sample). The On-Spin Column DNase I Kit (MO BIO Laboratories) was used for DNA removal, with the modification that DNase was allowed to sit for up to 30 min to increase the efficiency of DNA removal. Purified RNA samples were shipped to the Hudson Alpha Institute Genomic Services Laboratory (Huntsville, AL) for rRNA reduction, using the

Ribo-Zero Gold Epidemiology rRNA removal kit, and sequencing on the Illumina HiSeq™ platform using a paired-end 125 bp flow cell.

## RNA-seq data processing

Raw sequences were processed using the CLC Genomics Workbench v. 9.5.4 suite (QIAGEN, Hilden, Germany). Bases below 0.03 error score cutoff were trimmed. Samples were subjected to a subsequent *in silico* rRNA reduction using the SortmeRNA 2.0 software package [42]. Filtered paired-reads were competitively mapped to cyanobacterial and phage genomes (S2) with a 0.9 read-length fraction and 0.9 identity-fraction cutoffs. Transcripts were enumerated as read pairs mapped within the open reading frames of individual genes, and counts normalized by library size (unless noted). Paired reads with ends mapping to different genomes were not included in downstream analyses or counts. Sequence information has been deposited in MG-RAST database under the study Lake\_Taihu\_metatranscriptome\_project (sample IDs in S1 Table).

## Phylogenetic analysis

Reference sequences from Proteobacteria, Cyanobacteria, and phage identified by sequence alignment as IS607 regions in [36] were downloaded from NCBI (S3 Table). IS607 reference sequences were aligned in MEGA 7.0.14 software [43] using the MUSCLE algorithm [44] and this alignment was then used to generate a maximum likelihood tree with a Shimodaira-Hasegawa-like approximate likelihood ratio test branch validation using PhyML [45]. The reference sequences were then aligned with RNA-seq reads mapping to the Ma-LMM01 IS607 region in HMMER v. 3.1 (hmmer.org). Reads from the alignment were placed the reference tree using pplacer [46]. Quantity of reads placed on the tree was visualized as branch width using the guppy software package [47].

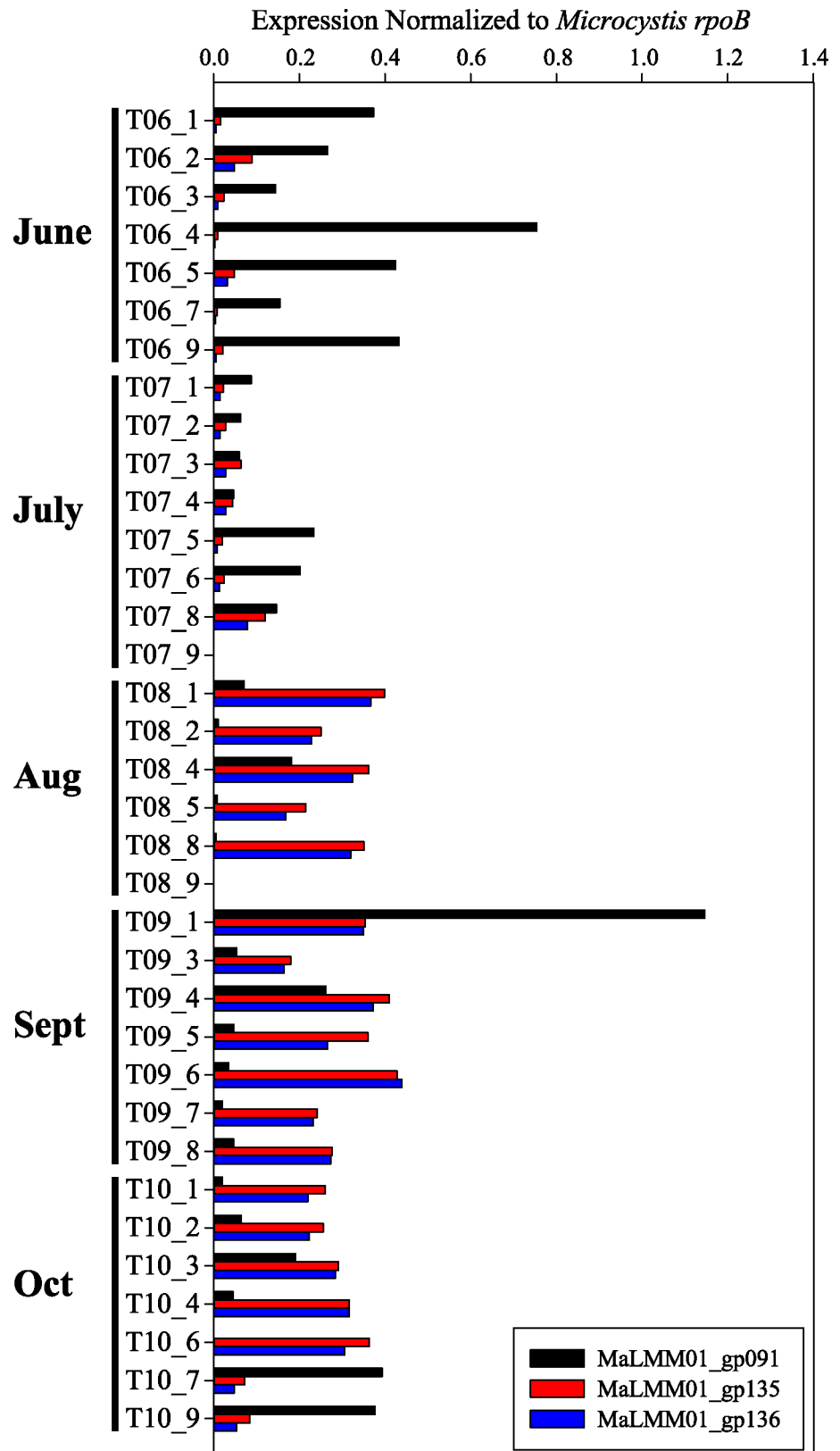
## Statistical analysis

*Microcystis* phage Ma-LMM01 gene read counts were  $\log_2(x+10)$  transformed and Pearson correlation values were calculated in R Statistics [48] using the Hmisc R package [49]. Mapped read counts per gene were normalized to expression of *M. aeruginosa* NIES-843 *rpoB* (as a proxy for host cell density) and plotted using the SigmaPlot software package (Systat Software, Chicago, IL). Whole genome expression was determined by counting reads mapped within gene regions on the Ma-LMM01 reference genome, which were normalized by library size, square root transformed, and used to generate a Bray-Curtis dissimilarity matrix and non-metric multidimensional scaling (nMDS) plots in the PRIMER7 software suite [50]. Associated environmental variables were correlated with Bray-Curtis dissimilarity distributions and plotted as vectors on the nMDS. The relationship between environmental variables and expression of the phage genome was determined using the BEST analysis [50]. The co-occurrence of expression of whole genome expression was grouped using the CLUSTER function using the Pearson correlation coefficient as the index of association with a 0.1 p-value cutoff. The results of this analysis were visualized in a dendrogram, all in PRIMER7.

## Results

### Differential expression of genes from *Microcystis*-infecting phage

Normalized expression of the Ma-LMM01-like tail sheath (*gp091*), transposase (*gp135*), and site-specific recombinase (*gp136*) observed in Lake Tai are shown in Fig 1. Of the 35 samples, 2 (T07\_9 and T08\_9) exhibited negligible expression of phage and host genes and have been



**Fig 1. Lytic and lysogenic gene expression by station.** Spatial and temporal gene expression of lytic and lysogenic genes from *Microcystis*-phage in Lake Tai. Expression of the *Microcystis* phage Ma-LMM01 phage

viral tail sheath (*gp091*, black), transposase (*gp135*, red), and recombinase (*gp136*, blue) normalized by expression of *Microcystis aeruginosa* RNA polymerase B (*rpoB*) observed in the Lake Tai dataset.

<https://doi.org/10.1371/journal.pone.0184146.g001>

removed from subsequent analyses. In the remaining 33 samples, 16 showed more abundant expression of *gp091* relative to *gp135*, with a ratio ranging from 1.21 to 79-fold, implying that lytic infection was dominant. These samples were collected during the earlier months (June and July) of the bloom season, with the exception of T09\_1, T10\_7, and T10\_9, which were collected during September and October. The remaining samples showed expression of the *gp135* and *gp136* to be greater than the expression of *gp091*, implying the *Microcystis* community was, at least to some degree, lysogenized. These samples primarily occurred during the months of August, September, and October. Statistically, sample location within the lake did not relate to expression patterns, with each station exhibiting periods with dominance of lytic or putative-lysogenic transcripts almost in equal measure across all five months. Tail sheath expression was significantly and negatively correlated with both transposase (Fig 2,  $\rho = -0.53$ ,  $p = 0.0017$ ) and also recombinase abundance ( $\rho = -0.57$ ,  $p = 0.0001$ ). Transposase and recombinase were very highly correlated ( $\rho = 0.98$ ,  $p = 0$ ,  $R = 0.986$  on a linear function fit), suggesting tightly coordinated co-expression.

### *Microcystis*-phage genome expression

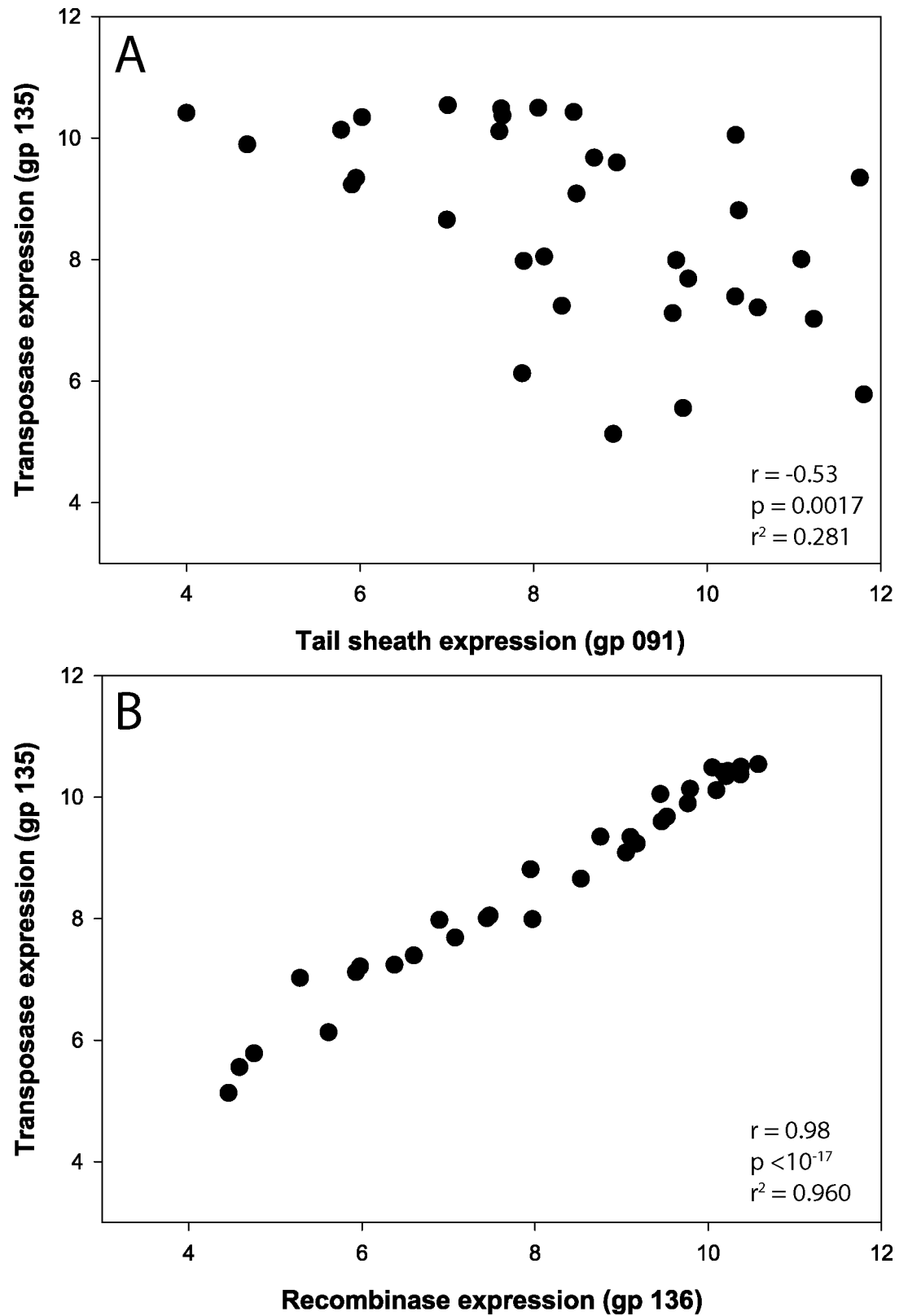
As a proxy for *in situ* expression of all *Microcystis* phage genes, we recruited environmental transcripts to the Ma-LMM01 genome. Results observed from samples collected in Lake Tai, and organized by hierarchical clustering, are represented in Fig 3. Each of the genes for both datasets generally fell into one of three major clusters. The first cluster includes all the genes potentially involved in lysogeny, including all three transposases (*gp031* and *gp032*—collapsed in branch A, *gp135*), the serine recombinase (*gp136*), and two hypothetical proteins (*gp171*, *gp067*).

The second cluster is predominantly made up of genes involved in phage packaging and cell lysis. It contains 60 genes, including 2 encoding lysozymes (*gp069*—collapsed in branch W, and *gp095*—collapsed in branch X) and the genes for DNA terminase (*gp118*—collapsed in branch DD), DNA primase (*gp134*—collapsed in branch AA), and a putative Fe/S oxidoreductase (*gp128*—collapsed in branch AA), which are the only ORFs with functions assigned. These genes exhibit high correlation values ( $\rho \geq 0.7$ ), of which 48 are significantly co-expressed ( $p \leq 0.1$ ) with at least one other gene.

The third cluster is the largest, and is made up of genes whose products are associated with nucleotide metabolism, DNA replication, and the structural components of the phage. It is made up of 112 genes including the viral tail sheath (*gp091*, collapsed in branch T), phage-encoded RecA (collapsed in branch S), the phycobilisome degradation protein NblA (collapsed in branch N), and a rIIA-like protein (collapsed in branch P). Viral tail sheath expression was highly correlated with genes *gp088* and *gp092*, which were predicted by protein size to encode viral tail tube proteins. Genes *gp086* and *gp087* also clustered with the tail sheath, which are believed to encode major head proteins for the phage particle.

### Environmental drivers of phage gene expression

A non-metric multidimensional scaling (nMDS) plot of the Bray-Curtis dissimilarity analysis of phage genome expression in Lake Tai is shown in Fig 4. Samples were distributed in a continuum across the *x*-axis, forming two primary clusters where phage gene expression was at least 60% similar. The position of samples along the *x*-axis corresponds significantly to the



**Fig 2. Tail sheath, transposase, and recombinase coexpression.** Co-expression of genes associated with putative lytic and lysogenic infections in Lake Tai. A. Scatterplot comparing expression of Ma-LMM01 viral tail sheath

(*gp091*, x-axis) to viral transposase (*gp135*, y-axis). Expression values are absolute read abundance  $\log_2$  normalized and demonstrate the negative relationship between the putative lytic (*gp091*) and lysogenic (*gp135*) infection markers. B. Scatterplot comparing expression of Ma-LMM01 recombinase gene (*gp136*, x-axis) to viral transposase (*gp135*, y-axis), both putative markers of lysogenic infection of *Microcystis*.

<https://doi.org/10.1371/journal.pone.0184146.g002>

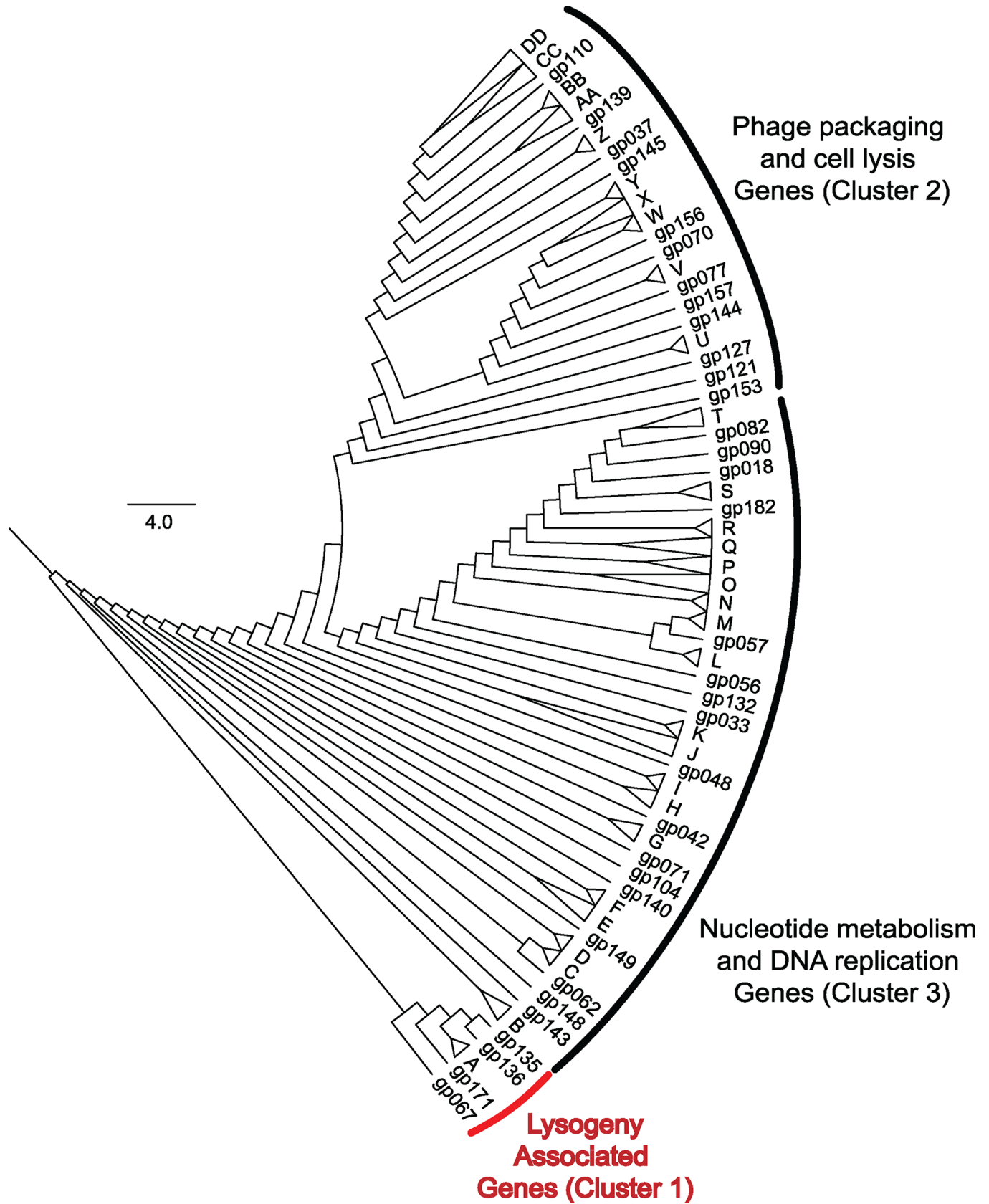
ratio of tail sheath to transposon (*gp135*) expression (a similar trend was observed when the ratio of tail sheath to recombinase (*gp136*) expression is plotted on the samples). Vectors for environmental variables are plotted on the nMDS, showing that pH (towards lysogenic) and concentration of total dissolved solids (towards lytic) contributed most significantly to position along the x-axis. Total dissolved nitrogen and phosphorous also contributed to position along the x-axis, driving the position of samples towards greater expression of lytic genes or putative lysogenic genes, respectively. The dissolved oxygen concentration and water temperature also contributed, though more significantly to position along the y-axis. The BEST analysis of environmental variable contribution to expression of the entire phage genome confirmed these associations, and determined that water temperature, pH, and concentration of total dissolved solids, phosphorous, nitrogen, and oxygen concentrations were responsible for 33% of the variation in gene expression ( $p = 0.05$ ). Phage gene expression was not correlated to toxin concentration (*gp091*: microcystin  $\mu\text{g/L}$ ,  $\rho = -0.19$ ,  $p = 0.2956$ ; *gp135*: microcystin,  $\rho = 0.29$ ,  $p = 0.099$ ; *gp136*: microcystin,  $\rho = 0.25$ ,  $p = 0.1623$ ).

## Discussion

We surveyed community metatranscriptomes from natural populations of *M. aeruginosa* at “bloom densities” to describe the physiology and ecology of *Microcystis*, and in the process identified active phage infections by the *Microcystis* phage Ma-LMM01. We have analyzed this data in light of available nutrient concentrations, toxin levels, and environmental conditions to predict how lake chemistry and climate influenced *Microcystis* phage gene expression. Our observations suggest that expression across the entire phage genome appears to have switched between the expression of genes involved in active viral replication (*i.e.*, the lytic cycle), and the expression of genes that have been proposed to allow the phage to integrate into the host genome (*i.e.*, lysogeny). Lastly, we found that the expression of phage genes appears to have been strongly associated with total dissolved solids and pH as well as the availability of nutrients, specifically the relative abundance of nitrogen and phosphorous. These observations have given rise to three distinct hypotheses: 1.) These correlations and co-occurrences are the product of random chance; 2.) The pattern of gene expression represents a novel physiological interaction (the purpose of which is currently unclear) between this phage and its host and was independent of lysogeny; 3.) The results indicate that *Microcystis* phage were actively switching between lytic and lysogenic cycles. We address these conclusions below within the context of factors that drive freshwater microbial communities.

The possibility that observed patterns in phage gene expression were the result of random chance is not supported by our analyses. The observation of similar expression patterns across Lake Tai suggests the mechanism by which *Microcystis*-infecting phage regulate gene expression has been largely conserved and is important for this virus's survival. Previous attempts to describe Ma-LMM01 transcriptional regulation in the laboratory relied on ORF orientation in the virus genome sequence, which yielded two general groups of genes: an “early” gene region containing 144 genes that were suggested to be responsible for nucleotide metabolism and genome replication, and a late gene region, encoding the remaining 40 genes, believed to encode phage structural components [29]. A subsequent study used q-rtPCR in culture to measure transcripts of the viral tail sheath (*gp091*), a putative late gene, and the gene for the



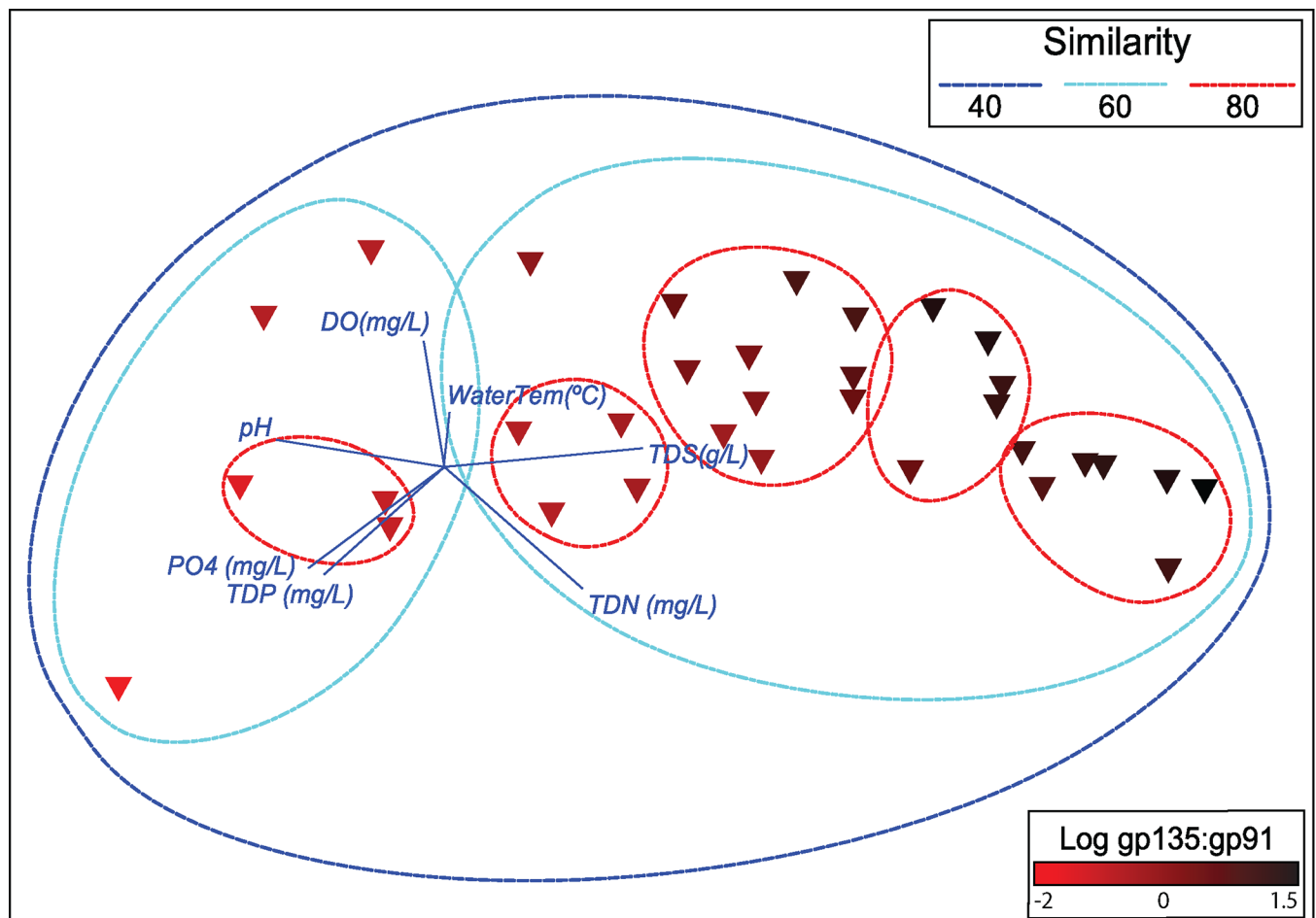


**Fig 3. Ma-LMM01 whole genome coexpression.** Cluster analysis of statistically co-expressed *Microcystis*-phage gene expression (based on Ma-LMM01 genome) in Lake Tai. Individual branches represent genes correlated with the expression of others. Transcript sets are collapsed and labeled with a letter where expression patterns were statistically indistinguishable (see S4 Table for genes contained in collapsed branches).

<https://doi.org/10.1371/journal.pone.0184146.g003>

phycobilisome degradation protein (NblA), a putative early region gene [32]. They observed a temporal separation of expression between the two genes and hypothesized that the larger gene regions were consistent with early/late phage gene expression, a regulation strategy observed in other cyanomyoviruses [51]. The disconnect between the regions identified by Yoshida and colleagues with our clustering is not surprising: in dealing with natural populations (unlike lab studies), we were most likely dealing with non-synchronous infections. Indeed, that there are statistically relevant relationships within the expression data suggests there are strong environmental controls on lytic vs lysogenic decisions.

A second potential explanation for our observations, that switching between expression states is unrelated to lysogeny, remains plausible. Much of the gene expression we attributed to genome integration originates in the virus' three putative-transposases (*gp031*, *gp032*, *gp135*)



**Fig 4. Environmental contribution to whole genome expression.** A. Non-metric multidimensional scaling plot of Bray-Curtis dissimilarity between *Microcystis*-infecting phage whole genome expression for Lake Tai. Read abundance was normalized by library size and square root transformed. Ellipses represent minimum similarity between samples at the 40%, 60%, and 80% levels. Symbols have been colored based on the  $\log_2$  transformed *gp091:gp135* expression ratio to denote lytic (black) vs lysogenic (red) dominated states. Environmental variables identified in the BEST analysis have been correlated (Pearson) with similarity between samples and plotted as vectors, indicating the direction on the 2-dimensional plane with which they correlated.

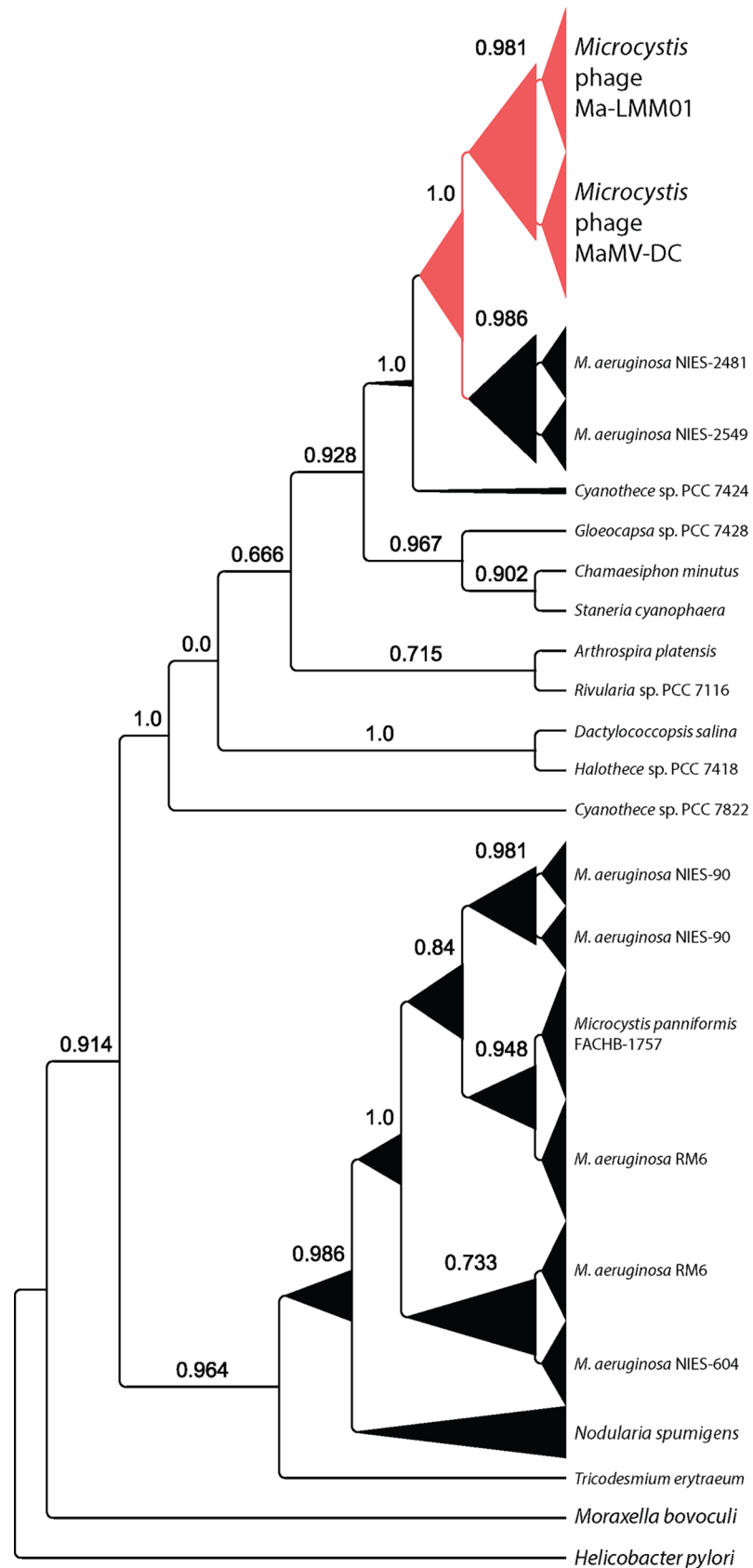
<https://doi.org/10.1371/journal.pone.0184146.g004>

and the recombinase (*gp136*), all of which have some homologues in different strains of *M. aeruginosa* and other cyanobacteria. Transposase *gp135* belongs to a potential family of mobile elements, IS607, which was originally identified in *Helicobacter pylori* [37]. IS607 representatives encode a corresponding serine recombinase (*gp136* in Ma-LMM01) and together, this gene pair is widespread amongst sequenced cyanobacteria [36]. While these genes can be phylogenetically resolved across the length of the insertion sequence, determining the genomic origin of short sequencing reads is more challenging. Our pplacer phylogenetic tree (Fig 5) demonstrates the majority of reads were identified as viral in origin, but the dearth of sequenced phage genomes related to Ma-LMM01 makes it difficult to evaluate the consistency of IS607 in viruses. Additionally, the IS607 encoded serine recombinase is atypical in structure amongst other similar enzymes. The DNA-binding and catalytic domains are flipped in orientation, resulting in a recombinase that acts *via* a modified mechanism, leading to a significant reduction in insertion site specificity [52]. At the outset, it is not known how this would influence activity of the insertion sequence in the context of viral infection, nor how it could play a role in lysogeny, but we speculate that decreased binding specificity might better allow integration of the virus into the notoriously plastic *M. aeruginosa* genome [20, 22]. It should also be noted that the presence of insertion sequences in phage genomes are very rare, as they can negatively impact virus survival [53].

That observed shifts in *Microcystis*-phage gene expression represent active genome integration (lysogeny) are the most consistent with our observations and those in other systems. Moreover, that this process is tied to nutrient availability in the water column gives this observation significant ecological relevance. The formation of a lysogen would explain why putative lysogenic genes are conserved in the phage genome in a variety of geographic locations [38]. There is a broad literature suggesting that phage have adapted to replicate or integrate depending on the conditions that favor the growth or senescence of their particular host [54–57]. Nutrient availability has long been associated with the formation of prophage in environmental systems, though it is generally thought to inhibit induction indirectly by limiting the material available to produce viral progeny, rather than by direct sensing for lysis-lysogeny decision making [58]. In better characterized phage-host systems, such as Lambda phage, the richness of the growth medium modulates signals in host metabolism that influence the lysis-lysogeny decision [59]. Unfortunately, our ability to determine the mechanism of action from metatranscriptomic data is limited, and the lack of similarity to better characterized phage systems, such as Lambda phage, makes comparisons with *Microcystis* phage difficult to draw at this time. This is further complicated by the current unavailability of *Microcystis*-infecting phage for controlled studies. That said, it is clear from the consensus of the scientific community that we cannot discount the importance of this (and similar) environmental molecular studies [60].

That *Microcystis* blooms can proliferate to massive densities [61] and yet somehow escape infection by the community of abundant phage [62] remains a perplexing ecological problem. This may be explained by the ability to resist infection by lytic viruses due to lysogen-induced resistance to superinfection. Indeed, while many observations lie in contrast, other studies that have suggested a “Piggyback-the-winner” model [63], which proposes that the spread of viral genomic material is best served by lysogenizing rapidly growing host cells that can persist at high densities. Clues to how this occurs mechanistically may lie in the uncharacterized genes co-expressed with the transposase and recombinase, namely *gp171* and *gp067*. While neither of these genes have close hits in the NCBI database, their implied relationship with the putative lysogenic genes suggests involvement in prophage maintenance. However, without culture work to identify their function, this remains speculation.

We observed that *Microcystis* phage gene expression could consistently be detected in *Microcystis* blooms and that a dramatic shift expression of lytic vs lysogenic gene groups was



**Fig 5. Phylogenetic distribution of IS607 reads.** Bootstrapped phylogenetic tree of the mobile element IS607 where branch widths indicate abundance of Lake Tai dataset reads mapping to that branch. Branches belonging to the *Microcystis* phage are colored red.

<https://doi.org/10.1371/journal.pone.0184146.g005>

tied to environmental cues. Although the cause and effect of these cues needs further study, we hypothesize that *Microcystis*-infecting phage may actively integrate into the host genome—a state that can be distinguished from the lytic cycle *via* the relative transcription of *gp091* and *gp135*. While these new observations need continued validation and a better resolution of mechanistic controls, this study demonstrates that phage may have a strong influence population dynamics of this harmful bloom forming species.

## Supporting information

**S1 Table. Sampling information.** Name, date, time, and location of each of the samples taken from Lake Tai and used for metatranscriptomic sequencing. Also recorded is the environmental data collected for each sample, including water temperature during sampling (WaterTem °C), electric conductivity (EC), concentration of total dissolved solids (TDS), salinity (Sal), pH, nephelometric turbidity unit (NTU), YSI chlorophyll (YSI-CHL), phycocyanin (PC), dissolved oxygen (DO), Secchi depth (SD), depth of the lake at the sampling site (WaterDep), total nitrogen concentration (TN), total dissolved nitrogen concentration (TDN), ammonium concentration (NH<sub>4</sub>), total phosphorous concentration (TP), total dissolved phosphorous (TDP), phosphate concentration (PO<sub>4</sub>), and chlorophyll A concentration (CHLa). (XLSX)

**S2 Table. Read mapping statistics.** Statistics from mapping metatranscriptomic reads to reference genomes from each of the Lake Tai samples. Cells show the total number of reads mapped to each of the given genomes during competitive read mapping with a 0.9 similarity cutoff and a 0.9 length fraction cutoff. (XLSX)

**S3 Table. Reference sequence information.** Reference sequences used in Fig 5 and their accession numbers. (XLSX)

**S4 Table. Gene branches collapsed in cluster diagram.** Branches collapsed in Fig 3. Each of the collapsed branches is identified by the letter from Fig 3, and beneath are listed the genes whose expression patterns were statistically indistinguishable by hierarchical clustering. (XLSX)

## Acknowledgments

We gratefully acknowledge Taihu Laboratory for Lake Ecosystem Research (TLER), Jingchen Xue and Dr. Keqiang Shao for helping with sample collection and water chemical analysis. We also thank Drs. Gary LeCleir, Hans Paerl, Silvia Newell, Mark McCarthy, Guangwei Zhu, Boqiang Qin and Ferdi Hellweger for assistance and valuable discussions.

## Author Contributions

**Conceptualization:** Joshua M. A. Stough, Steven W. Wilhelm.

**Data curation:** Joshua M. A. Stough, Lauren E. Krausfeldt, Morgan M. Steffen, Steven W. Wilhelm.

**Formal analysis:** Joshua M. A. Stough, Xiangming Tang, Lauren E. Krausfeldt, Morgan M. Steffen, Gregory L. Boyer, Steven W. Wilhelm.

**Funding acquisition:** Guang Gao, Steven W. Wilhelm.

**Investigation:** Joshua M. A. Stough, Xiangming Tang, Lauren E. Krausfeldt, Morgan M. Steffen, Guang Gao, Gregory L. Boyer, Steven W. Wilhelm.

**Methodology:** Joshua M. A. Stough, Xiangming Tang, Lauren E. Krausfeldt, Morgan M. Steffen, Gregory L. Boyer, Steven W. Wilhelm.

**Project administration:** Steven W. Wilhelm.

**Resources:** Guang Gao, Steven W. Wilhelm.

**Writing – original draft:** Joshua M. A. Stough, Xiangming Tang, Morgan M. Steffen, Steven W. Wilhelm.

**Writing – review & editing:** Joshua M. A. Stough, Steven W. Wilhelm.

## References

1. Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Heldal M, et al. Global-scale processes with a nanoscale drive: the role of marine viruses. *ISME J.* 2008; 2: 575–578. <https://doi.org/10.1038/ismej.2008.31> PMID: 18385772
2. Weitz JS, Wilhelm SW. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol Rep.* 2012; 4: 17. <https://doi.org/10.3410/B4-17> PMID: 22991582
3. Wilhelm SW, Suttle CA. Viruses and nutrient cycles in the sea. *Bioscience.* 1999; 49: 781–788.
4. Thingstad TF, Lignell R. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquat Microb Ecol.* 1997; 13: 19–27. <https://doi.org/10.3354/ame013019>
5. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature.* 2016; 537: 689–693. <https://doi.org/10.1038/nature19366> PMID: 27654921
6. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci U S A.* 2011; 108: E757–E764. <https://doi.org/10.1073/pnas.1102164108> PMID: 21844365
7. Harke MJ, Steffen MM, Gobler CJ, Otten TG, Wilhelm SW, Wood SA, et al. A review of the global ecology, genomics, and biogeography of the commonly toxic cyanobacterium, *Microcystis*. *Harmful Algae.* 2016; 54: 4–20. <https://doi.org/10.1016/j.hal.2015.12.007> PMID: 28073480
8. Bishop CT, Anet EFLJ, Gorham PR. Isolation and identification of the Fast-Death Factor in *Microcystis aeruginosa* NRC-1. *Can J Biochem Physiol.* 1959; 37: 453–471. <https://doi.org/10.1139/o59-047> PMID: 13638864
9. Carmichael W. Toxic *Microcystis* and the environment. *Toxic microcystis.* 1996:1–11.
10. Brittain SM, Wang J, Babcock-Jackson L, Carmichael WW, Rinehart KL, Culver DA. Isolation and characterization of Microcystins, cyclic heptapeptide hepatotoxins from a Lake Erie strain of *Microcystis aeruginosa*. *J Great Lakes Res.* 2000; 26: 241–249. [https://doi.org/10.1016/s0380-1330\(00\)70690-3](https://doi.org/10.1016/s0380-1330(00)70690-3)
11. Steffen MM, Davis TW, Stough JMA, McKay RM, Bullerjahn GS, Krausfeldt LE, et al. Lesson from 2014 Lake Erie bloom: business as usual for the western basin? *Environ Sci Technol.* 2017; 51: 6745–6755. <https://doi.org/10.1021/acs.est.7b00856>
12. Krausfeldt LE, Tang X, van de Kamp J, Gao G, Bodrossy L, Boyer GL, et al. Spatial and temporal variability in the nitrogen cyclers of hypereutrophic Lake Taihu. *FEMS Microbiol Ecol.* 2017; 93: fix024. <https://doi.org/10.1093/femsec/fix024> PMID: 28334116
13. Qin BQ, Zhu GW, Gao G, Zhang YL, Li W, Paerl HW, et al. A drinking water crisis in Lake Taihu, China: linkage to climatic variability and lake management. *Environ Manag.* 2010; 45: 105–112. <https://doi.org/10.1007/s00267-009-9393-6> PMID: 19915899
14. Brunberg AK. Contribution of bacteria in the mucilage of *Microcystis* spp. to benthic and pelagic production in a hypereutrophic lake. *FEMS Microbiol Ecol.* 1999; 29: 13–22.
15. Steffen MM, Belisle BS, Watson SB, Boyer GL, Bourbonniere RA, Wilhelm SW. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Appl Environ Microbiol.* 2015; 81:3268–3276. <https://doi.org/10.1128/AEM.04101-14> PMID: 25662977

16. Steffen MM, Li Z, Effler TC, Hauser LJ, Boyer GL, Wilhelm SW. Comparative metagenomics of toxic freshwater cyanobacteria bloom communities on two continents. *PloS one*. 2012; 7: e44002. <https://doi.org/10.1371/journal.pone.0044002> PMID: 22952848
17. Harke MJ, Jankowiak JG, Morrell BK, Gobler CJ. Transcriptomic responses in the bloom-forming cyanobacterium *Microcystis* induced during exposure to zooplankton. *Appl Environ Microbiol*. 2017; 83: e02832–16. <https://doi.org/10.1128/AEM.02832-16> PMID: 28003198
18. Kromkamp J, Konopka A, Mur LR. Buoyancy in light-limited continuous cultures of *Microcystis aeruginosa*. *J Plankt Res*. 1988; 10: 171–183.
19. Shen H, Niu Y, Xie P, Tao MIN, Yang XI. Morphological and physiological changes in *Microcystis aeruginosa* as a result of interactions with heterotrophic bacteria. *Freshw Biol*. 2011; 56: 1065–1080. <https://doi.org/10.1111/j.1365-2427.2010.02551.x>
20. Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M, et al. Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. *DNA Res*. 2007; 14: 247–256. <https://doi.org/10.1093/dnares/dsm026> PMID: 18192279
21. Yamaguchi H, Suzuki S, Tanabe Y, Osana Y, Shimura Y, Ishida K-i, et al. Complete genome sequence of *Microcystis aeruginosa* NIES-2549, a bloom-forming cyanobacterium from Lake Kasumigaura, Japan. *Gen Announc*. 2015; 3. <https://doi.org/10.1128/genomeA.00551-15> PMID: 26021928
22. Steffen MM, Dearth SP, Dill BD, Li Z, Larsen KM, Campagna SR, et al. Nutrients drive transcriptional changes that maintain metabolic homeostasis but alter genome architecture in *Microcystis*. *ISME J*. 2014; 8: 2080–2092. <https://doi.org/10.1038/ismej.2014.78> PMID: 24858783
23. Ou T, Li S, Liao X, Zhang Q. Cultivation and characterization of the MaMV-DC cyanophage that infects bloom-forming cyanobacterium *Microcystis aeruginosa*. *Virol Sin*. 2013; 28: 266–271. <https://doi.org/10.1007/s12250-013-3340-7> PMID: 23990146
24. Hargreaves KR, Anderson NJ, Clokie MRJ. Recovery of viable cyanophages from the sediments of a eutrophic lake at decadal timescales. *FEMS Microbiol Ecology*. 2013; 83: 450–456. <https://doi.org/10.1111/1574-6941.12005> PMID: 22963199
25. Watkins SC, Smith JR, Hayes PK, Watts JEM. Characterisation of host growth after infection with a broad-range freshwater cyanopodophage. *PLoS ONE*. 2014; 9: 8. <https://doi.org/10.1371/journal.pone.0087339> PMID: 24489900
26. Tucker S, Pollard P. Identification of cyanophage Ma-LBP and infection of the cyanobacterium *Microcystis aeruginosa* from an Australian subtropical lake by the virus. *Appl Environ Microbiol*. 2005; 71: 629–635. <https://doi.org/10.1128/AEM.71.2.629-635.2005> PMID: 15691911
27. Yoshida T, Takashima Y, Tomaru Y, Shirai Y, Takao Y, Hiroishi S, et al. Isolation and characterization of a cyanophage infecting the toxic cyanobacterium *Microcystis aeruginosa*. *Appl Environ Microbiol* 2006; 72: 1239–1247. <https://doi.org/10.1128/AEM.72.2.1239-1247.2006> PMID: 16461672
28. Mankiewicz-Boczek J, Jaskulska A, Paweczyk J, Gagala I, Serwecinska L, Dziadek J. Cyanophages infection of *Microcystis* bloom in lowland dam reservoir of Sulejw, Poland. *Microb Ecol*. 2016; 71: 315–325. <https://doi.org/10.1007/s00248-015-0677-5> PMID: 26403721
29. Yoshida T, Nagasaki K, Takashima Y, Shirai Y, Tomaru Y, Takao Y, et al. Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol*. 2008; 190:1762–1772. <https://doi.org/10.1128/JB.01534-07> PMID: 18065537
30. Ou T, Gao XC, Li SH, Zhang QY. 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*. 2015; 96: 3681–3697. <https://doi.org/10.1099/jgv.0.000290> PMID: 26399243
31. Rozon RM, Short SM. Complex seasonality observed amongst diverse phytoplankton viruses in the Bay of Quinte, an embayment of Lake Ontario. *Freshw Biol*. 2013; 58: 2648–2663. <https://doi.org/10.1111/fwb.12241>
32. Yoshida-Takashima Y, Yoshida M, Ogata H, Nagasaki K, Hiroishi S, Yoshida T. Cyanophage infection in the bloom-forming cyanobacteria *Microcystis aeruginosa* in surface freshwater. *Microbes Environ*. 2012; 27: 350–355. <https://doi.org/10.1264/jsme2.ME12037> PMID: 23047146
33. Yoshida M, Yoshida T, Kashima A, Takashima Y, Hosoda N, Nagasaki K, et al. Ecological dynamics of the toxic bloom-forming cyanobacterium *Microcystis aeruginosa* and its cyanophages in freshwater. *Appl Environ Microbiol*. 2008; 74: 3269–3273. <https://doi.org/10.1128/AEM.02240-07> PMID: 18344338
34. Takashima Y, Yoshida T, Yoshida M, Shirai Y, Tomaru Y, Takao Y, et al. Development and application of quantitative detection of cyanophages phylogenetically related to cyanophage Ma-LMM01 infecting *Microcystis aeruginosa* in fresh water. *Microbes and Environ*. 2007; 22: 207–213. <https://doi.org/10.1264/jsme2.22.207>

35. Bobay LM, Rocha EPC, Touchon M. The adaptation of temperate bacteriophages to their host genomes. *Mol Biol Evol.* 2013; 30: 737–751. <https://doi.org/10.1093/molbev/mss279> PMID: 23243039
36. Kuno S, Yoshida T, Kamikawa R, Hosoda N, Sako Y. The distribution of a phage-related insertion sequence element in the cyanobacterium, *Microcystis aeruginosa*. *Microbes Environ.* 2010; 25: 295–301. <https://doi.org/10.1264/jsme2.ME10125> PMID: 21576885
37. Kersulyte D, Mukhopadhyay AK, Shirai M, Nakazawa T, Berg DE. Functional organization and insertion specificity of IS607, a chimeric element of *Helicobacter pylori*. *J Bacteriol.* 2000; 182: 5300–5308. <https://doi.org/10.1128/jb.182.19.5300-5308.2000> PMID: 10986230
38. Steffen MM, Belisle BS, Watson SB, Boyer GL, Bourbonniere RA, Wilhelm SW. Metatranscriptomic evidence for co-occurring top-down and bottom-up controls on toxic cyanobacterial communities. *Appl Environ Microbiol.* 2015; 81: 3268–3276. <https://doi.org/10.1128/AEM.04101-14> PMID: 25662977
39. Hutchinson GE. The paradox of the plankton. *American Nat.* 1961; 95: 137–145.
40. Jin X, Tu Q. The standard methods for observation and analysis in lake eutrophication. Chinese Environmental Science Press, Beijing. 1990; 240.
41. Boyer GL. The occurrence of cyanobacterial toxins in New York lakes: Lessons from the MERHAB-Lower Great Lakes. *Lake Reserv Manag.* 2007; 23: 153–160.
42. Kopylova E, Noe L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics.* 2012; 28: 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611> PMID: 23071270
43. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016; 33: 1870–1874. <https://doi.org/10.1093/molbev/msw054> PMID: 27004904
44. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl Acids Res.* 2004; 32: 1792–1797. <https://doi.org/10.1093/nar/gkh340> PMID: 15034147
45. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate Maximum-Likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59: 307–21. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
46. Matsen FA, Kodner RB, Armbrust EV. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinform.* 2010; 11: 16. <https://doi.org/10.1186/1471-2105-11-538> PMID: 21034504
47. Matsen FA, Evans SN. Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLoS ONE.* 2013; 8: 15. <https://doi.org/10.1371/journal.pone.0056859> PMID: 23505415
48. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.
49. Harrell Jr. FE. Hmisc: Harrell miscellaneous. 2016. <https://cran.r-project.org/web/packages/Hmisc/index.html>
50. Clark KR, Gorley RN. PRIMER v7: User manual/tutorial. Plymouth: PRIMER-E; 2015.
51. Clokie MRJ, Shan JY, Bailey S, Jia Y, Krisch HM, West S, et al. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol.* 2006; 8: 827–835. <https://doi.org/10.1111/j.1462-2920.2005.00969.x> PMID: 16623740
52. Boocock MR, Rice PA. A proposed mechanism for IS607-family serine transposases. *Mob DNA.* 2013; 4: 9. <https://doi.org/10.1186/1759-8753-4-9>
53. Sakaguchi Y, Hayashi T, Kurokawa K, Nakayama K, Oshima K, Fujinaga Y, et al. The genome sequence of *Clostridium botulinum* type C neurotoxin-converting phage and the molecular mechanisms of unstable lysogeny. *Proc Natl Acad Sci U S A.* 2005; 102:17472–17477. <https://doi.org/10.1073/pnas.0505503102> PMID: 16287978
54. Miller R, Day M. Contribution of lysogeny, pseudolysogeny, and starvation to phage ecology. In: Abedon S, editor. *Bacteriophage ecology: population growth, evolution, and impact of bacterial viruses.* Cambridge: Cambridge University Press; 2008. p. 114–144.
55. Payet JP, Suttle CA. To kill or not to kill: The balance between lytic and lysogenic viral infection is driven by trophic status. *Limnol Oceanogr.* 2013; 58: 465–474. <https://doi.org/10.4319/lo.2013.58.2.0465>
56. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* 2016; 10: 437–449. <https://doi.org/10.1038/ismej.2015.125> PMID: 26296067
57. Paul JH. Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.* 2008; 2: 579–589. <https://doi.org/10.1038/ismej.2008.35> PMID: 18521076



58. McDaniel L, Paul JH. Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine *Synechococcus* species. *Appl Environ Microbiol*. 2005; 71: 842–850. <https://doi.org/10.1128/AEM.71.2.842-850.2005> PMID: 15691939
59. Wilson HR, Yu DG, Peters HK, Zhou JG, Court DL. The global regulator RNase III modulates translation repression by the transcription elongation factor N. *EMBO J*. 2002; 21: 4154–4161. <https://doi.org/10.1093/emboj/cdf395> PMID: 12145215
60. Simmonds P, Adams MJ, Benko M, Breitbart M, Brister JR, Carstens EB, et al. Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Micro*. 2017; 15: 161–168. <https://doi.org/10.1038/nrmicro.2016.177> PMID: 28134265
61. Steffen MM, Belisle BS, Watson SB, Boyer GL, Wilhelm SW. Status, causes and controls of cyanobacterial blooms in Lake Erie. *J Great Lakes Res*. 2014; 40: 215–225. <https://doi.org/10.1016/j.jglr.2013.12.012>
62. Long AM, Short SM. Seasonal determinations of algal virus decay rates reveal overwintering in a temperate freshwater pond. *ISME J*. 2016; 10: 1602–1612 <https://doi.org/10.1038/ismej.2015.240> PMID: 26943625
63. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, et al. Lytic to temperate switching of viral communities. *Nature*. 2016; 531: 466–470. <https://doi.org/10.1038/nature17193> PMID: 26982729