environmental microbiology

Environmental Microbiology (2015) doi:10.1111/1462-2920.12756

Ecological and genetic interactions between cyanobacteria and viruses in a low-oxygen mat community inferred through metagenomics and metatranscriptomics

Alexander A. Voorhies,[1] Sarah D. Eisenlord,[2] Daniel N. Marcus,[1] Melissa B. Duhaime,[3] Bopaiah A. Biddanda,[4] James D. Cavalcoli[5] and Gregory J. Dick[1,3]*

Departments of [1]Earth and Environmental Sciences, [3]Ecology and Evolutionary Biology and [5]Computational Medicine and Bioinformatics and [2]School of Natural Resources and Environment, University of Michigan, Ann Arbor, MI 48109, USA.

[4]Annis Water Resources Institute, Grand Valley State University, Muskegon, MI 49441, USA.

Summary

Metagenomic and metatranscriptomic sequencing was conducted on cyanobacterial mats of the Middle Island Sinkhole (MIS), Lake Huron. Metagenomic data from 14 samples collected over 5 years were used to reconstruct genomes of two genotypes of a novel virus, designated PhV1 type A and PhV1 type B. Both viral genotypes encode and express *nblA*, a gene involved in degrading phycobilisomes, which are complexes of pigmented proteins that harvest light for photosynthesis. Phylogenetic analysis indicated that the viral-encoded *nblA* is derived from the host cyanobacterium, *Phormidium* MIS-PhA. The cyanobacterial host also has two complete CRISPR (clustered regularly interspaced short palindromic repeats) systems that serve as defence mechanisms for bacteria and archaea against viruses and plasmids. One 45 bp CRISPR spacer from *Phormidium* had 100% nucleotide identity to PhV1 type B, but this region was absent from PhV1 type A. Transcripts from PhV1 and the *Phormidium* CRISPR loci were detected in all six metatranscriptomic data sets (three during the day and three at night), indicating that both are transcriptionally active in the environment. These results reveal ecological and genetic interactions between viruses and cyanobacteria at MIS, highlighting the value of parallel analysis of viruses and hosts in understanding ecological interactions in natural communities.

Received 12 June, 2014; accepted 15 December, 2014. *For correspondence. E-mail gdick@umich.edu; Tel. (734) 763 3228; Fax (734) 763 4690.

Introduction

Bacteria and archaea are some of the most diverse and abundant organisms on the planet, yet it is estimated that there are as many as 10 times more viruses (Weinbauer, 2004). Viruses evolve at a rapid rate and strategically avoid defence systems of their microbial hosts (Samson *et al.*, 2013). This forces bacteria and archaea to adapt and drives the evolution of microbial defence mechanisms in what is often described as a co-evolutionary arms race (Weinbauer, 2004; Banfield and Young, 2009; Heidelberg *et al.*, 2009; Rodriguez-Valera *et al.*, 2009; Avrani *et al.*, 2011; Kashtan *et al.*, 2014). Understanding the fundamental dynamics of microbe–virus interactions, particularly defence and defence–avoidance systems, has implications for medicine, industry and environmental microbiology.

Current studies of environmental microbe–virus interactions are hindered by inherent biological properties and methodological limitations. The lack of a universally conserved virus marker gene (Kristensen *et al.*, 2013) to delineate operational taxonomic units in a phylogenetic framework precludes the analysis of viral diversity and community structure in the manner typical of microbial studies. The classification of viral assemblages is further complicated by the mosaic and dynamic nature of viral genomes (Hendrix *et al.*, 1999; Emerson *et al.*, 2012). Also confounding our understanding of the role of viruses in overall ecosystem functioning is the challenge of linking novel viruses to their hosts, which is essential to assess the impact of viral activity on specific microbial populations. Towards this goal, lab-based methods have been developed recently, such as phageFISH (Allers *et al.*, 2013), digital polymerase chain reaction (Tadmor *et al.*, 2011) and viral tagging (Deng *et al.*, 2012).

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd

Metagenomic sequencing can offer further windows into whole communities of coexisting microbes and viruses and yield evidence for host–virus links based on shared gene content (Mizuno *et al*., 2013), signatures of nucleotide compositional patterns shared between viruses and hosts (Andersson and Banfield, 2008; Pride and Schoenfeld, 2008; Dick *et al*., 2009) and CRISPR/cas systems (clustered regularly interspaced short palindromic repeats/CRISPR-associated sequences) (Andersson and Banfield, 2008; Garcia-Heredia *et al*., 2012).

CRISPRs are used by many bacteria and archaea as a mechanism of defence from viruses and plasmids (Makarova *et al*., 2006; Barrangou *et al*., 2007). CRISPRs are loci in the host genome consisting of short sequences of viral or plasmid DNA ('spacers') (Barrangou *et al*., 2007) interleaved between conserved repeat sequences. Adjacent to these spacer/repeat regions are CRISPR-associated (*cas*) genes (Haft *et al*., 2005) which encode proteins needed to acquire new spacers from invading viruses or plasmids (Barrangou *et al*., 2007; Horvath and Barrangou, 2010; Paez-Espino *et al*., 2013; Sun *et al*., 2013) and prevent the production of proteins from invading sources of DNA (Haft *et al*., 2005; Sorek *et al*., 2008; van der Oost *et al*., 2009; Garneau *et al*., 2010; Horvath and Barrangou, 2010). New spacers are acquired from a region of viral or plasmid DNA referred to as the protospacer and are added to one end of the CRISPR (the 'leader' end) upon introduction to viral or plasmid DNA. At this leader end, spacer content reflects the evolution of viruses and dynamics of viral communities (Andersson and Banfield, 2008; Tyson and Banfield, 2008; Paez-Espino *et al*., 2013). Spacers on the 'trailer end' are conserved, perhaps reflecting past selective sweeps, and may guard against reinfection by persistent viruses (Andersson and Banfield, 2008; Tyson and Banfield, 2008; Heidelberg *et al*., 2009; Weinberger *et al*., 2012). CRISPRs have been found in 40% of sequenced bacteria and 90% of sequenced archaea (Grissa *et al*., 2007a). CRISPRs are present in the majority of cyanobacteria (86 of 126 genomes), although absent in all but one member of the widespread marine *Prochlorococcus* and *Synechococcus* subclades (Cai *et al*., 2013).

Another important interaction between viruses and their microbial hosts is the exchange of genes. Viruses contribute to host genome diversification as an agent of lateral gene transfer (Mann *et al*., 2003; Lindell *et al*., 2004; 2005; Anantharaman *et al*., 2014). The presence of host genes in viral genomes is thought in some cases to augment biochemical processes at key metabolic bottlenecks; such genes are referred to as auxiliary metabolic genes (AMG) (Breitbart *et al*., 2007). A wide range of AMGs involved in photosynthetic processes have been recovered from the genomes of viruses of cyanobacteria in various environments (Ignacio-Espinoza and Sullivan, 2012; Hurwitz *et al*., 2013), including photosystem genes (Mann *et al*., 2003; Sullivan *et al*., 2006; Sharon *et al*., 2009), phycobilin biosynthesis genes (Dammeyer *et al*., 2008) and phycobilisome degradation genes (Yoshida *et al*., 2008).

The Middle Island Sinkhole (MIS) in Lake Huron, Michigan, USA, hosts cyanobacterial mats in a low-oxygen (O_2) habitat (Ruberg *et al*., 2008). The mat community is dominated by a single genotype of cyanobacteria that shows evidence of viral predation pressure (Voorhies *et al*., 2012). Because the mats thrive in the presence of low-O_2 and a steep gradient of sulfide, this ecosystem is a valuable analogue of ancient microbial mat systems that represent the earliest evidence for life on Earth (Allwood *et al*., 2006) and profoundly influenced biogeochemical cycles in the Precambrian (Hoehler *et al*., 2001). Cyanobacterial mats continue to play important ecological roles on the modern Earth (Golubic and Abed, 2010; Bolhuis and Stal, 2011), yet little is known about viruses associated with cyanobacterial mats or their potential impacts on cyanobacterial diversity and ecological function. This low-diversity community also offers a system where viral strain variation can be teased apart with sequence-based techniques without the need for isolation and pure culture. Here, we report genome sequences of the virus PhV1 and document genetic and ecological interactions with its cyanobacterial host. We find that the host, *Phormidium* MIS-PhA, encodes multiple CRISPR systems, one of which targets one of the PhV1 genotypes. PhV1 contains a host-derived gene, *nblA*, which breaks down major photosynthetic pigment proteins.

Results

Recovery of two circular viral genomes from MIS metagenomes

A total of 14 MIS cyanobacterial mat samples were collected at seven time points in five different years between 2007 and 2012, including day and night samples in 2011 and 2012 (Table 1). Random shotgun sequencing and *de novo* genome assembly were used to recover complete genomes for two genomic variants of a novel virus, designated as PhV1 type A and PhV1 type B (Fig. 1). Only two of fourteen samples (2011-1D and 2011-2N) independently yielded complete genomes for both type A and B, likely due to the significantly larger number of reads available in those samples (Table 1). Separation of PhV1 genotypes in remaining samples was achieved by mapping reads from each sample to template sequences and visually verifying the reads matched to the correct variant (Fig. S1). Resulting consensus sequences were used to evaluate the completeness of PhV1 genomes in each sample (Fig. S2).

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Table 1. Metagenome sample summary.

Sample	Finger[a]	Prostrate Mat[b]	Day	Night	Illumina gDNA Reads	Illumina cDNA Reads
2007-1D	x	–	x	–	55 742 528	None
2009-1D	–	x	x	–	56 610 020	None
2009-2D	x	–	x	–	58 811 950	None
2009-3D	x	–	x	–	25 542 490	None
2010-1D	x	–	x	–	69 716 858	None
2011-1D	–	x	x	–	98 430 946	None
2011-2N	–	x	–	x	211 878 358	None
2011-3D	x	–	x	–	59 608 800	None
2012-1D	x	–	x	–	25 666 608	11 997 534
2012-2D	x	–	x	–	41 595 684	14 577 110
2012-3D	x	–	x	–	41 318 538	11 524 300
2012-4N	x	–	–	x	42 876 682	12 125 414
2012-5N	x	–	–	x	38 335 454	12 100 438
2012-6N	x	–	–	x	28 207 614	13 079 448

a. Microbial mat structure that is raised off the lake sediments.
b. Microbial mat that lies flat on lake sediments.

PhV1 genomes assembled into circular chromosomes with sizes of 45 kb and 41 kb, encoding 62 and 57 predicted genes in PhV1 type A and B respectively. Annotation revealed primarily conserved genes of unknown function or completely novel coding regions with no known homologues; only 9 genes in type A and 10 genes in type B had predicted functions (Tables S1 and S2). The two genomes share 81% nucleotide identity, but due to extensive insertions or deletions, they share only 44 genes (Fig. 1). The genome size and number of coding genes present in PhV1 is consistent with phage infecting other *Phormidium* species, such as cyanophage Pf-WMP3, a member of the T7 phage supergroup that infects *Phormidium foveolarum* and has a 43 kb genome with 41 genes (Liu *et al*., 2008). However, it should be noted that PhV1 and cyanophage Pf-WMP3 do not share any homologous genes and are genetically unrelated. Although only two genotypes of PhV1 were abundant enough to rebuild complete genomes, there is evidence of other low abundance variants of PhV1 at MIS in the form of single nucleotide polymorphisms, insertions and deletions (Fig. S3).

A total of 13 PhV1 genes had predicted protein sequences with greater than 30% amino acid identity to predicted proteins from a custom database containing all phage genomes available in the NCBI RefSeq database. 23 PhV1 genes had predicted proteins greater than 30% amino acid identity to proteins in the Pacific Ocean Virome (Hurwitz and Sullivan, 2013). The possibility that PhV1 is a plasmid is unlikely based on lack of signature plasmid replication or pilus genes (Ma *et al*., 2012), homologues of which were found at MIS (data not shown), but not in either PhV1 genome. Because no universally conserved plasmid signature genes have been identified, we searched the NCBI plasmid database for PhV1 gene homologues. Only seven PhV1 genes had homology to plasmid genes and all had less than 50% sequence similarity to genes of unknown function that are also present in genetic elements other than plasmids. The lack of plasmid signature genes and higher relative similarity of PhV1 genes to viruses than plasmids strongly suggest that PhV1 is a virus and not a plasmid.

Two lines of evidence indicated that PhV1 targets the dominant cyanobacterium at MIS, *Phormidium* sp. MIS-PhA. First, PhV1 and *Phormidium* shared tetranucleotide frequency patterns (Fig. S4), which accurately distinguish *Phormidium* from other cyanobacteria at MIS (Voorhies *et al*., 2012). Second, the *Phormidium* MIS-PhA genome contained a 45 bp segment of DNA with 100% nucleotide identity to the PhV1 type B genome. This sequence, designated CRISPR.III-B.spacer1, was flanked by CRISPR subtype III-B repeats and is part of a complete CRISPR locus (Fig. 2).

Further, both PhV1 genotypes contained a gene encoding NblA, a protein that guides the proteolytic degradation of phycobilisomes by the Clp protease during conditions of nitrogen starvation in cyanobacteria (Karradt *et al*., 2008; Gao *et al*., 2012; Yoshida-Takashima *et al*., 2012). Our analysis shows that *nblA* is present in 44 of the 83 complete cyanobacterial genomes in the Integrated Microbial Genomes database (https://img.jgi.doe.gov/) as of October 2014 (Table S3). In addition to being present in the viral genomes, *nblA* was also present in the host *Phormidium* genome. Four distinct versions of NblA were encoded by *Phormidium* at MIS, two of which were closely related to those from PhV1 with 67–69% sequence identity at the amino acid level (Fig. 3). Our genome-binning method is unable to resolve whether these *nblA* copies occur in one genome or in separate genotypes.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

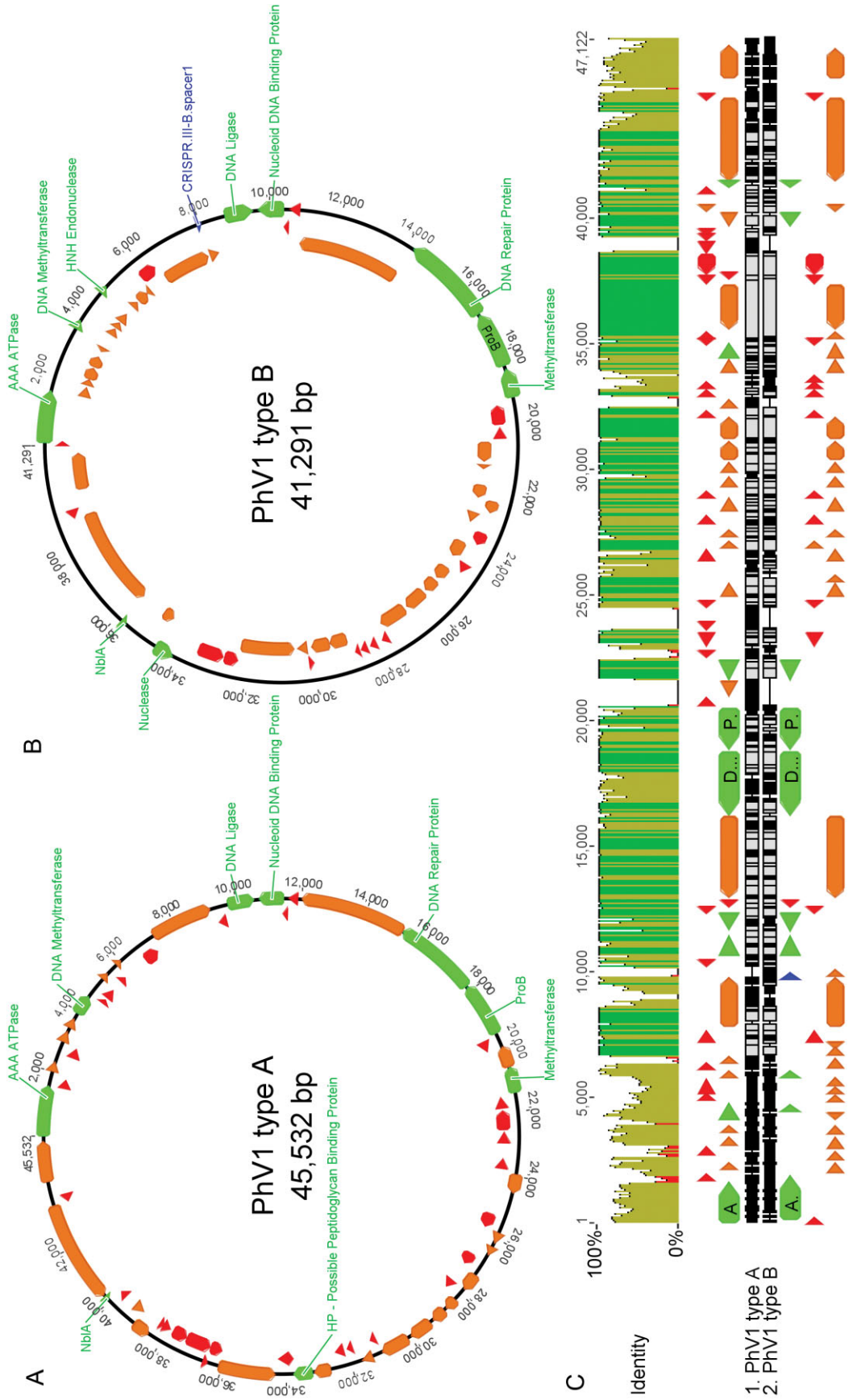

Fig. 1. *Phormidium* phage PhV1 genotypes.
A. Genome map of PhV1 type A with gene annotations.
B. genome map of PhV1 type B with gene annotations.
C. nucleotide alignment of type A and type B.
Green: genes with annotation; orange: hypothetical genes present in the NCBI nr database, but with no annotation; red: novel genes with no homologues in NCBI nr. In the nucleotide alignment, grey areas show nucleotide agreement, black indicates nucleotide disagreement and thin line indicates gaps in the alignment. Nucleotide identity (top of C) is represented by green (100%), gold (99–30%) and red (less than 30%).

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

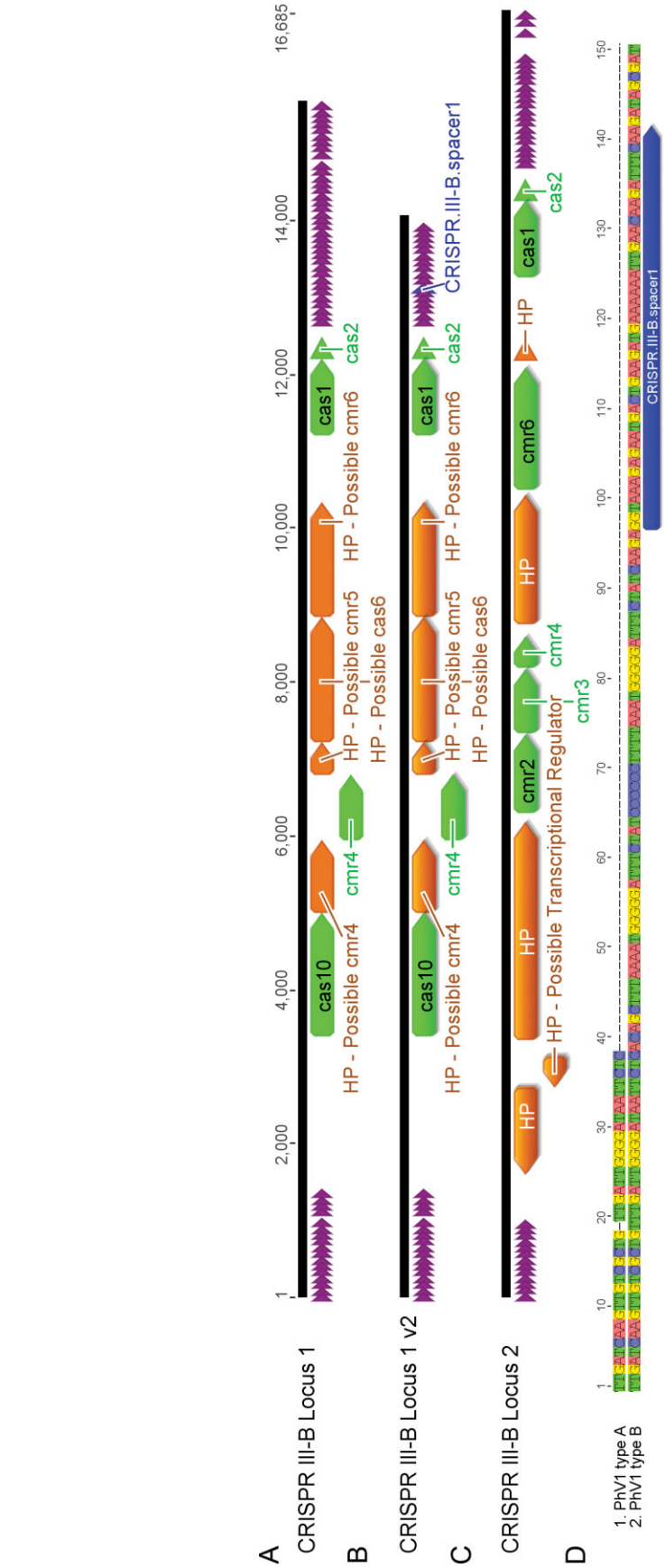

Fig. 2. Gene maps of CRISPR subtype III-B loci belonging to *Phormidium* sp. MIS-PhA.
A. CRISPR III-B.Locus1 with the most abundant spacer order.
B. CRISPR III-B.Locus1.v2 shows alternative spacers and contains CRISPR.III-B.spacer1.
C. CRISPR III-B.Locus2 with the most abundant spacer order.
D. CRISPR spacer from III-B.Locus1.v2 aligned between PhV1 genotypes.
Dashes indicate missing sequence in that region of the genome. Green: genes that could be annotated; orange: genes of unknown function; purple: CRISPR III-B repeat; blue: region that matches a CRISPR III-B spacer.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

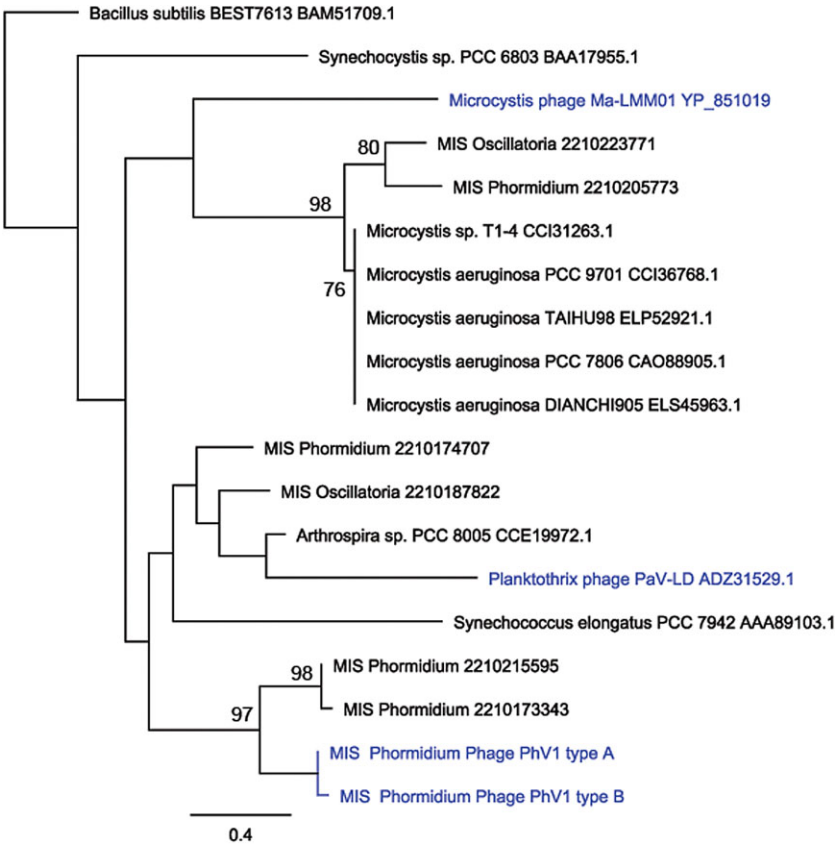

Fig. 3. Maximum likelihood tree of phycobilisome degradation protein NblA from cyanobacteria and their viruses. Bootstrap values below 70 have been removed. Blue text indicates NblA sequences from viral genomes. Sequences from this study are indicated with 'MIS'.

Host defence

Phormidium sp. MIS-PhA was the dominant organism at MIS and the only cyanobacterium present in all data sets analysed here. Based on the presence of conserved *cas* genes and five conserved signature repeats, *Phormidium* sp. MIS-PhA contained two distinct subtype III-B CRISPR loci (as classified by Makarova *et al.*, 2011a,b), designated here as Locus1 and Locus2 (Table S4). Although these two CRISPR loci shared repeats and homologous *cas* genes, the *cas1* genes shared only 88.7% nucleotide identity, and the spacers and gene order were distinct (Fig. 2; Fig. S5). A total of 33 contigs representing these CRISPR loci – but containing different arrangements of 225 unique spacers – were recovered. Locus1 *cas* genes and the most abundant arrangement of spacers were conserved and were reconstructed for all samples (Fig. S6).

The abundance of individual spacers and spacer-containing contigs (henceforth, 'spacer contigs') was highly variable across samples (Fig. 4, Fig. S7). However, permutational multivariate analysis of variance

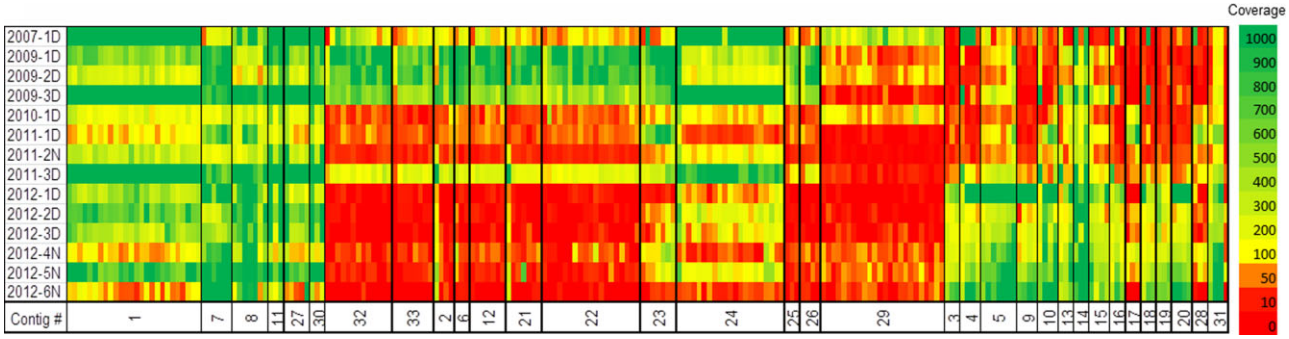

Fig. 4. Heat map of CRISPR III-B spacer abundance across 14 samples from 2007 to 2012. Year and type of sample [day(D)/night(N)] are indicated on the far left. Columns represent individual spacers, and black vertical lines distinguish groups of spacers with conserved order assembled to the same 'spacer contig'.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

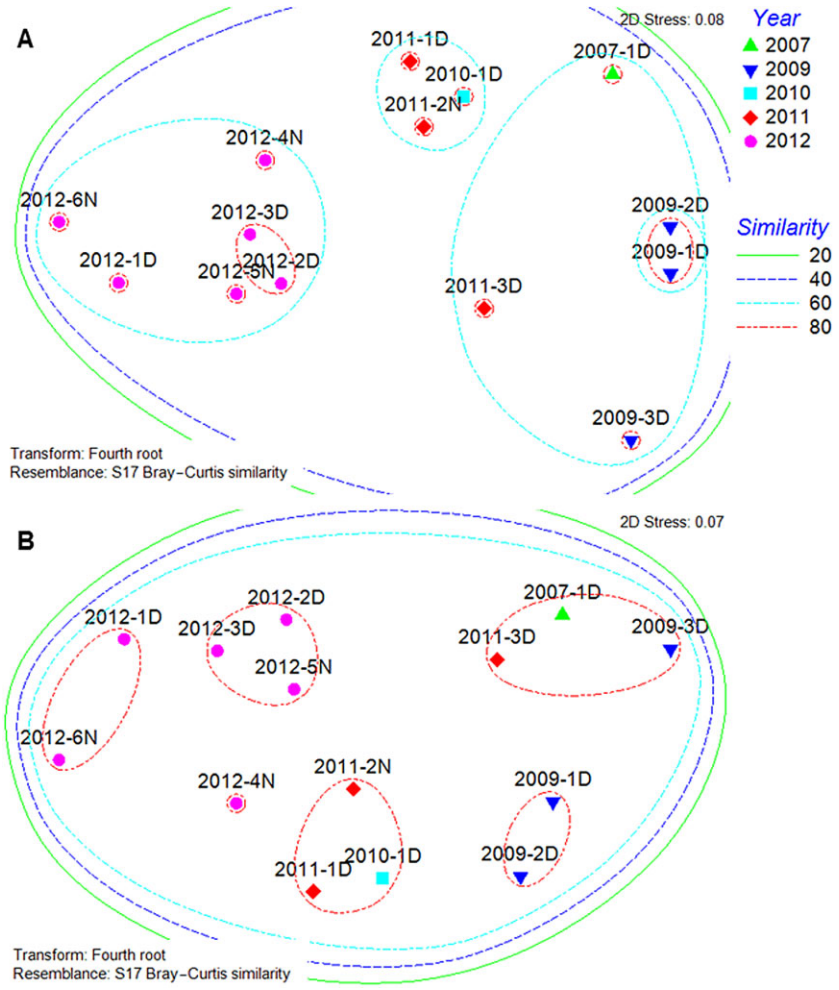

Fig. 5. Non-metric multidimensional scaling (nMDS) plots of the similarity of CRISPR spacer read composition between samples. A. Similarity of abundance of CRISPR conserved order 'spacer contigs'. B. Similarity of abundance of individual CRISPR spacers.

(PERMANOVA) showed that the composition of both individual spacers and spacer contigs were conserved within each sampling year ($P = 0.005$ and 0.001 respectively). Comparison of spacer content between samples by non-metric multidimensional scaling confirmed clustering of samples according to year, with some exceptions, such as sample 2011-3D (Fig. 5). In some cases, the spacer content was also highly similar between years (e.g. 2010 and 2011; 2007 and 2009). These trends were also apparent through principal coordinate analysis (Fig. S8).

Average abundance of spacers calculated from assembled spacer contigs was typically similar to average read abundance calculated from single spacers (Fig. S7), suggesting that spacers are commonly found in the conserved order represented by our spacer contigs. However, several statistical approaches consistently showed that the composition of individual spacers was more conserved across samples than the composition of spacer contigs, indicating variability that was not captured by the contigs. This included analysis of similarity (ANOSIM R statistic 0.668, 0.1% significant for individual spacers; 0.824, 0.1% significant for spacer contigs), CLUSTER analysis (60% and 40% similarity when comparing individual spacers and spacer contigs respectively; Fig. 5) and principal coordinate analyses (58.6% and 53.8% of variation between samples explained by the first axis, which correlated with sample year, for individual spacers and spacer contigs respectively; Fig. S8).

BLASTN was used to compare the 225 unique spacers from CRISPR subtype III-B against both genotypes of PhV1, revealing a single spacer that shares 100% nucleotide identity with the proto-spacer region of PhV1 type B, and was present in half the samples analysed. This spacer, designated CRISPR.III-B.spacer1, belongs to the *Phormidium* sp. MIS-PhA CRISPR Locus1, whereas Locus2 does not possess any spacers matching to either genotype of PhV1. The PhV1 type B proto-spacer that matches CRISPR.III-B.spacer1 is not present in PhV1 type A due to a gap in that region of the type A genome (Fig. 2D).

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Gene expression of PhV1 and host CRISPR loci

Metatranscriptomic sequencing was performed on six samples collected in 2012. cDNA reads were mapped to genome sequences from PhV1 and the *Phormidium* CRISPR loci to determine whether these regions were actively expressing RNA in the environment at the time of collection. Approximately half of PhV1 genes showed some level of expression and the most highly expressed (up to 65× average gene coverage) were from genes of unknown function (Fig. S9). *nblA* and a gene with a predicted peptidoglycan-binding domain were also expressed in multiple samples, albeit at low levels (Fig. S9). These results indicate that the PhV1 viruses were transcriptionally active at the time of sampling.

All six metatranscriptomes from 2012 contained transcripts from both *cas* genes and spacer regions in CRISPR Locus1 (Fig. S10), demonstrating that *Phormidium* was actively transcribing its CRISPR loci in the environment at the time of collection. The spacer/repeat regions in the *Phormidium* genome show more expression than the *cas* genes themselves in all samples. A non-coding region directly adjacent to the upstream series of spacers (located at the ~1500 bp mark in Fig. S10) was the most highly expressed area of Locus1. Transcript data were examined for day/night trends (Figs S9 and S10), but none were detected.

Abundance of viral genotypes and PhV1 targeting CRISPR spacer across samples

The CRISPR.III-B.spacer1 that matches PhV1 type B was recovered from 7 out of 14 samples taken from MIS between 2007 and 2012, including the 2007 sample, one of the three 2011 samples and five out of six samples from 2012 (Fig. S11). Average read abundance showed a decline in abundance of PhV1 type B from 2007 to 2011 and an increase in abundance of PhV1 type A from 2011 to 2012, resulting in a shift in PhV1 dominance from type B to A (Fig. S11). During the same period, the relative abundance of *Phormidium*, based on read abundance of four single-copy housekeeping genes; *L2P*, *L4P*, *secY* and *ychF*, was relatively high in 2007–2010, but declined in 2011–2012. Of the nine samples from 2011 to 2012, only two samples had *Phormidium* relative abundances equivalent to those detected in earlier samples from 2007 to 2010.

Discussion

We used a metagenomic approach to link the dominant member of a cyanobacterial mat community to a highly abundant virus that preys upon it. Genome analysis showed that *Phormidium* phage MIS-PhV1 contained genes that are mostly of unknown function and shared no genes with viruses known to infect *Phormidium* in other environments (Liu *et al*., 2008). No structural gene homologue (e.g. capsid proteins, portal proteins, etc.) was identified in the PhV1 genomes, precluding suggestions towards taxonomic classification. Two distinct genotypes of PhV1 contain large genome rearrangements consisting of insertions, deletions and single nucleotide polymorphisms, and there is evidence that low abundance genotypes coexist in the environment.

Both genotypes of PhV1 encode NblA, which is used by cyanobacteria to degrade phycobilisomes (Collier and Grossman, 1994; Gao *et al*., 2012; Yoshida-Takashima *et al*., 2012). Phycobilisomes are photosynthetic antenna pigments in cyanobacteria that can account for up to one half of total soluble protein in a cell (Bogorad, 1975). The *nblA* gene is thought to be present in all organisms that contain phycobilisomes (Luque and Forchhammer, 2008); our analysis showed that annotated *nblA*s are present in approximately half of complete cyanobacterial genomes. Breakdown of phycobilisomes by NblA in cyanobacteria, such as *Synechococcus* sp. PCC7492, is linked to N stress (Luque *et al*., 2001) and may prevent photodamage during stress or provide resources for the synthesis of proteins required for adaptation (Karradt *et al*., 2008). *nblA* is also in the genomes of viruses of *Microcystis aeruginosa* (Yoshida *et al*., 2008) and *Planktothrix agardhii* (Gao *et al*., 2012). Although the benefit of virally encoded NblA is uncertain, it could provide amino acids for viral replication, or reduce photodamage to viral particles during infection (Gao *et al*., 2012; Yoshida-Takashima *et al*., 2012). Photo-damage may be particularly important for viruses at MIS given the likely slow host growth and viral reproduction rates. Although doubling times for *Phormidium* species at MIS have not been directly measured, the low temperature of the MIS sinkhole (perennially 9°C) and the doubling times of known *Phormidium* (0.07–0.5 d^{-1}, depending on light and nutrient availability) (Litchman, 2000; Litchman *et al*., 2003) suggest the MIS strains are relatively slow growing. Because slower host growth is correlated with longer infection periods (Middelboe, 2000), MIS *Phormidium* viruses may experience long latent periods (time between infection and release of virus progeny), exposing them to photo-oxidative damage in the intracellular environment for longer periods of time. The NblA in *Microcystis* phages is also expressed during infection of a relatively slow-growing host (0.12–0.45 d^{-1}) (van der Westhuizen and Eloff, 1985; Yoshida-Takashima *et al*., 2012). In this system, viral-mediated degradation of host phycobilisomes may improve viral fitness by either dampening host photosynthetic capacity and reducing build-up of reactive oxygen or by providing macromolecules for amino acids needed by the phage during replication. The

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

link between AMGs and latent period has also been noted in cyanophage that infect *Prochlorococcus* and *Synechococcus*: cyanophage containing the *psbA* gene have longer latent periods (> 8 h) than the one phage that lacks this gene (Sullivan *et al*., 2006).

PhV1 *nblA* clusters phylogenetically with *nblA* encoded by the host *Phormidium* sp. MIS-PhA, suggesting that *nblA* was acquired by PhV1 from its host (Fig. 3). This pattern is consistent with other virus-encoded AMGs (Breitbart *et al*., 2007), which are primarily derived from the host or other bacteria (Ignacio-Espinoza and Sullivan, 2012). However, the *nblA* found in *Microcystis* phage Ma-LMM01 clusters separately from the *Microcystis* sp. *nblA* subgroup (Fig. 3), suggesting a more complex history of horizontal gene transfer in some cases. The current lack of available *nblA* sequences for other *Phormidium* species and viruses make definitive determination of the phylogenetic relationships and evolutionary history of *nblA* difficult.

Rearrangements of the PhV1 genome may result in evasion of CRISPR defences, as suggested in previous studies of phage–host dynamics in the environment (Andersson and Banfield, 2008; Tyson and Banfield, 2008; Pride *et al*., 2011). This form of viral evasion of host defences differs from experiments in pure cultures, where rapid point mutation in regions targeted by CRISPRs was observed within 1 week (Sun *et al*., 2013). Indeed, PhV1 type B genomic sequences from 2011 to 2012 contained no conserved mutations in the region targeted by CRISPR.III-B.spacer1, even though type B persists in the environment (at reduced abundance) in all 2011–2012 samples. It is unclear whether the sequence conservation of the proto-spacer region of PhV1 type B was due to strong positive selection on that viral gene, reflects the relatively slow host growth rates at MIS, or some other constraint. Given the high abundance and persistence of PhV1 and multiple CRISPR/cas systems in the *Phormidium* genome, it is also surprising that only one CRISPR spacer matching PhV1 type B and no spacers matching PhV1 type A were found. This differs from previous studies that show rapid spacer acquisition (in as little as 10 min) upon exposure to new viruses (Barrangou *et al*., 2007; Tyson and Banfield, 2008; Sapranauskas *et al*., 2011). The lack of observed mutations in the proto-spacer region of PhV1 and the lack of spacers targeting the abundant virus PhV1 type A imply a more stable host–virus dynamic than has been observed in culture (Barrangou *et al*., 2007; Sun *et al*., 2013) or other environments (Andersson and Banfield, 2008).

Spacer sequences within the CRISPR system are acquired sequentially through time and thus provide a record of exposure to plasmids and viruses (Tyson and Banfield, 2008). Despite low genomic diversity of the dominant cyanobacterial population in MIS mats, *Phormidium* MIS-PhA (Voorhies *et al*., 2012), a remarkably diverse and temporally variable pool of spacers was observed (Fig. 4). Spacers of the CRISPR subtype III-B loci were more consistent within samples collected from the same year than those between sampling years. A core set of spacers were present in all samples and probably provide resistance to ancestral or persistently occurring viruses (Tyson and Banfield, 2008), whereas less abundant spacers changed in abundance over time and likely reflect recently acquired spacers. Similarity in read abundance between CRISPR spacer contigs (representing a conserved spacer order) and individual spacers (which disregards the order spacers are found in) implied that most spacers maintain their order over the time scale of the current study, with a few high abundance spacers responsible for the majority of difference in spacer content from sample to sample and year to year. Recent work showed a strong bias for re-acquisition of spacers from previously sampled proto-spacers (Paez-Espino *et al*., 2013), which may account for the observed highly abundant spacers at MIS.

Metatranscriptomic data revealed that both the virus and the host CRISPR system were transcriptionally active *in situ*. Given the modest sequencing effort and high diversity of spacers, transcriptomic coverage was not sufficient to investigate the expression of individual spacers. The most highly expressed portion of the *Phormidium* CRISPR/cas system was a non-coding region adjacent to the spacers. This region may represent a regulatory non-protein-coding small RNA, which have been identified at high frequencies in other environmental microbial metatranscriptomes and have been implicated in regulatory functions based on their proximity to regulatory genes (e.g. for carbon metabolism and nutrient acquisition) and lack of protein coding (Frias-Lopez *et al*., 2008; Gilbert *et al*., 2008; Shi *et al*., 2009).

During the course of this study, we observed a consistent shift in the abundance of PhV1 genotypes (Fig. S11). The correlated reversal of abundance of PhV1 genotypes is consistent with higher relative fitness of PhV1 type A due to evasion of *Phormidium* CRISPR defence through absence of the proto-spacer region in the type A genome. Indeed, increased abundance of PhV1 type A in 2011 and 2012 correlates with a severe decrease in abundance of *Phormidium* in seven out of nine samples. These correlations are consistent with a 'Kill-the-Winner' scenario (Thingstad, 2000; Rodriguez-Valera *et al*., 2009) in which viruses target the dominant organism, causing its abundance to decrease. Such viral activity could promote increased diversity of cyanobacteria within the MIS microbial mats, where this phage-promoted diversity may manifest at fine phylogenetic scales (Rodriguez-Valera *et al*., 2009; Kashtan *et al*., 2014). However, the diversity of MIS cyanobacterial communities is likely shaped by a balance

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

between top-down controls (e.g. predation, viruses) and bottom-up controls (e.g. nutrients, growth rates), and there may be positive feedback loops that enable *Phormidium* populations to retain their dominance, as proposed in the 'King of the Mountain' hypothesis (Giovannoni *et al*., 2013).

This study linked an abundant virus to its cyanobacterial host using metagenomic sequence data, yielding insights into viral ecology and host interactions in a cyanobacterial mat. Time series data showed that the PhV1 viruses were a persistent and transcriptionally active component of the MIS cyanobacterial mat community over the course of 5 years. The abundance and persistence of viruses carrying genes for the degradation of a major cyanobacterial pigment underscores the high potential of viruses to influence elemental cycling within cyanobacterial mats. The discovery of viral NblA genes in a benthic cyanobacterial microbial mat also reveals a parallel with viruses of the toxin-producing and bloom-forming *P. agardhii* and *M. aeruginosa*. That virus-mediated phycobilisome degradation occurs in such distinct cyanobacterial taxa and habitats suggest that it may be a common feature of virus–cyanobacteria interactions. In the modern world, cyanobacterial mats persist and play important ecological roles largely where grazing pressure is reduced due to extreme environmental conditions that exclude animals (Stal, 2012). On the early Earth, prior to the evolution of animals and when cyanobacterial mats were widespread, the impact of viruses on cyanobacterial mats and therefore global biogeochemical cycling may have been substantial. Thus, quantifying the effect of viruses on the biogeochemical cycling and diversity of cyanobacterial mats in modern analogues of Precambrian cyanobacterial mat systems is an important future goal.

Experimental procedures

Sampling and sample preparation

Microbial mat samples were collected from MIS aboard the *R/V Storm* by NOAA divers between 2007 and 2012. DNA was extracted from whole mats and processed (without amplification) for metagenomic shotgun pyrosequencing as previously described (Voorhies *et al*., 2012). RNA was extracted from 2012 samples, converted to cDNA, and amplified for sequencing as previously described (Frias-Lopez *et al*., 2008). All samples available and preserved in RNA later from MIS were used for metagenomic analysis to maximize sample time points.

Sequencing, assembly, and annotation

Specific parameters for all bioinformatics software can be found in Table S5.

A total of 14 environmental samples (Table 1) were shotgun sequenced on an Illumina Hi Seq 2000 instrument producing paired end reads at the University of Michigan DNA Sequencing Core. Assembly of Illumina shotgun reads was achieved by running three independent VELVET (Zerbino and Birney, 2008) assemblies for each sample with kmer values of 91, 75 and 61 which were then edited with METAVELVET (Namiki *et al*., 2012) and combined using the SAMTOOLS (Li and Durbin, 2009; Li *et al*., 2009) function minimus2. Reads files were dereplicated prior to assembly to remove reads with 100% ID and overlap, and trimmed based on quality scores before assembly.

Separate assemblies were performed for each of the 14 samples, and contiguous sequences from all 14 samples were combined using minimus2 to identify regions of interest, such as CRISPR loci or viral genomes that were present in multiple samples. Reads from each sample were then mapped back onto the templates using GENEIOUS (Biomatters, 2013) (http://www.geneious.com) to create sample specific sequences from the consensus. Assemblies were validated against the previous 454Ti assembly (Voorhies *et al*., 2012), and through extensive manual curation using the genomic viewers the INTEGRATED GENOME VIEWER (IGV) (Robinson *et al*., 2011) and GENEIOUS to visualize reads mapped onto contigs using the BURROWS-WHEELER ALIGNER (BWA) (Li and Durbin, 2009). Gene calling and annotation was performed by the Joint Genome Institute's Integrated Microbial Genomes Expert Review portal (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi), and gene calling was confirmed using PRODIGAL (Hyatt *et al*., 2010). The functional annotations of specific genes of interest were verified manually using BLASTP to the NR, COG, KEGG and Pfam databases, with cut-offs of 150 bitscore, and 60% sequence identity. E-values were not used as a cut-off, but all matches had e-values less than 10^{-30}.

Estimating viral and bacterial abundance

Genomic DNA and cDNA read abundance was calculated by first mapping reads to regions of interest in BWA, and then extracting those reads into a new fastq file. The selected reads were then competitively mapped to genomic templates using GENEIOUS (reads with multiple 100% matches were mapped to all possible regions). Separation of different strains of PhV1, and CRISPRs in the *Phormidium* genome, were performed visually in GENEIOUS based on read coverage and single nucleotide polymorphism patterns. Read abundances were normalized by dividing the number of reads attained per sample, and then averaged over the length of the gene or genome. See Table 1 for sample sizes.

Multivariate statistics

Multivariate statistics were conducted in PRIMER v6 (Clarke and Gorley, 2006) to visualize and analyse the differences in spacer and contig composition in samples from 2007 through 2012. Normalized spacer abundances and normalized contig abundances were fourth root transformed to de-emphasize the most abundant sequences, and then rank similarity matrices were calculated with the Bray–Curtis similarity metric. A one-way analysis of similarity (ANOSIM) with year as the main factor was then calculated with 999 permutations to retrieve the R statistic for community similarity between years

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

(Clarke, 1993). The ANOSIM R statistic ranges from 0 to 1, with values closer to 1 indicating larger differences and values closer to 0 indicating more similarity. Hierarchical clustering, using the group average method, was calculated to visualize the natural grouping of samples. Non-metric multidimensional scaling (NMDS) plots and principal coordinates analyses were created to further visualize differences in the composition of spacers and contigs within and between sampling dates. The multivariate adaptation of ANOVA, PERMANOVA (Anderson, 2001), was then calculated with year as the main factor and Monte Carlo simulations to test the difference between individual spacer and spacer contig abundance across years.

Phylogenetic analyses

Protein sequences were globally aligned with the GENEIOUS alignment tool using a Blosum62 cost matrix. A maximum likelihood tree was created using PHYML (Guindon and Gascuel, 2003) using the Jones, Taylor, and Thornton (JTT) substitution model and were bootstrapped 5000 times; only values of > 70 are reported on the tree.

Viral identification

Potential viral sequences were searched with BLASTP (minimum sequence identity of 30% and minimum bitscore of 50) against two custom databases. E-values were not used as a cut-off, but e-values of positive hits were less than 10^{-5}. One of the databases was created using all available complete phage genomes available on NCBI's RefSeq (as of June 2013), with all non-phage sequences (mostly eukaryotic viruses) removed. The second was a curated database of genes found in viral metagenomes ('The Pacific Ocean Virome'; Hurwitz and Sullivan, 2013). MIS sequences were also compared with a custom database created using all plasmid sequences available on NCBI's RefSeq as of June 2013.

CRISPR identification

CRISPR sequences were identified using CRISPR FINDER ONLINE (Grissa *et al*., 2007b) and *cas* genes were identified using a combination of BLAST and CRISPRdb (Grissa *et al*., 2007a). CRISPR spacers and repeats on contigs with *cas* genes were searched using BLASTN against all metagenomic samples (up to one mismatch over 100% of the spacer/repeat) to identify contigs with valid CRISPRs that contain valid spacer sequences, but lacked *cas* genes due to incomplete assembly. *Cas* genes were identified using BLASTP against a custom database comprising known *cas* genes from NCBI's RefSeq and *cas* genes described by Makarova and colleagues (2011b).

Data access

The sequence data from this study have been submitted to NCBI (http://www.ncbinlm.nih.gov/bioproject) under BioProject identifier PRJNA72255. Individual accession numbers for reads from the 14 metagenomic samples deposited in NCBI's Sequence Read Archive can be found in Table S6. Accession numbers for assembled and annotated sequences from MIS can be found in Table S7. Accession numbers for non-MIS *NblA* sequences can be found in Table S8.

Acknowledgements

We would like to thank Russ Green, Tane Casserly, Stephanie Gandulla, Wayne Lusardi and the NOAA Thunder Bay National Marine Sanctuary for sampling and logistical assistance, the University of Michigan DNA Sequencing Core for DNA sequencing, and Sunit Jain for bioinformatics help. This work was supported by NSF grant EAR 1035955 to GJD, EAR 1035957 to BAB, University of Michigan CCMB Pilot Grant to JDC and GJD and the Scott Turner Award to AAV.

References

Allers, E., Moraru, C., Duhaime, M.B., Beneze, E., Solonenko, N., Barrero-Canosa, J., *et al*. (2013) Single-cell and population level viral infection dynamics revealed by phagefish, a method to visualize intracellular and free viruses. *Environ Microbiol* **15:** 2306–2318.

Allwood, A.C., Walter, M.R., Kamber, B.S., Marshall, C.P., and Burch, I.W. (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* **441:** 714–718.

Anantharaman, K., Duhaime, M.B., Breier, J.A., Wendt, K.A., Toner, B.M., and Dick, G.J. (2014) Sulfur oxidation genes in diverse deep-sea viruses. *Science* **344:** 757–760.

Anderson, M.J. (2001) A new method for non-parametric multivariate analysis of variance. *Austral Ecol* **26:** 32–46.

Andersson, A.F., and Banfield, J.F. (2008) Virus population dynamics and acquired virus resistance in natural microbial communities. *Science* **320:** 1047–1050.

Avrani, S., Wurtzel, O., Sharon, I., Sorek, R., and Lindell, D. (2011) Genomic island variability facilitates prochlorococcus–virus coexistence. *Nature* **474:** 604–608.

Banfield, J.F., and Young, M. (2009) Variety – the splice of life – in microbial communities. *Science* **326:** 1198–1199.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., *et al*. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709–1712.

Biomatters (2013) Geneious. In.

Bogorad, L. (1975) Phycobiliproteins and complementary chromatic adaptation. *Annu Rev Plant Physiol Plant Mol Biol* **26:** 369–401.

Bolhuis, H., and Stal, L.J. (2011) Analysis of bacterial and archaeal diversity in coastal microbial mats using massive parallel 16s rRNA gene tag sequencing. *ISME J* **5:** 1701–1712.

Breitbart, M., Thompson, L.R., Suttle, C.A., and Sullivan, M.B. (2007) Exploring the vast diversity of marine viruses. *Oceanography* **20:** 135–139.

Cai, F., Axen, S.D., and Kerfeld, C.A. (2013) Evidence for the widespread distribution of CRISPR-cas system in the phylum Cyanobacteria. *RNA Biol* **10:** 687–693.

Clarke, K., and Gorley, R.N. (2006) *Primer v6: User Manual/Tutorial*. Plymouth, UK: PRIMER-E.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Clarke, K.R. (1993) Nonparametric multivariate analyses of changes in community structure. *Aust J Ecol* **18:** 117–143.

Collier, J.L., and Grossman, A.R. (1994) A small polypeptide triggers complete degradation of light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J* **13:** 1039–1047.

Dammeyer, T., Bagby, S.C., Sullivan, M.B., Chisholm, S.W., and Frankenberg-Dinkel, N. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol* **18:** 442–448.

Deng, L., Gregory, A., Yilmaz, S., Poulos, B.T., Hugenholtz, P., and Sullivan, M.B. (2012) Contrasting life strategies of viruses that infect photo- and heterotrophic bacteria, as revealed by viral tagging. *Mbio* **3:** e00373-12.

Dick, G.J., Andersson, A.F., Baker, B.J., Simmons, S.L., Yelton, A.P., and Banfield, J.F. (2009) Community-wide analysis of microbial genome sequence signatures. *Genome Biol* **10:** R85.

Emerson, J.B., Thomas, B.C., Andrade, K., Allen, E.E., Heidelberg, K.B., and Banfield, J.F. (2012) Dynamic viral populations in hypersaline systems as revealed by metagenomic assembly. *Appl Environ Microbiol* **78:** 6309–6320.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and DeLong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105:** 3805–3810.

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

Garcia-Heredia, I., Martin-Cuadrado, A.B., Mojica, F.J.M., Santos, F., Mira, A., Anton, J., and Rodriguez-Valera, F. (2012) Reconstructing viral genomes from the environment using fosmid clones: the case of haloviruses. *PLoS ONE* **7:** e33802.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., *et al*. (2010) The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468:** 67–71.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., and Joint, I. (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS ONE* **3:** e3042.

Giovannoni, S., Temperton, B., and Zhao, Y.L. (2013) Sar11 viruses and defensive host strains reply. *Nature* **499:** E4–E5.

Golubic, S., and Abed, R. (2010) Entophysalis mats as environmental regulators. In *Microbial Mats: Modern and Ancient Microorganisms in Stratified Systems*. Seckbach, J., and Oren, A. (eds). Dordrecht, The Netherlands: Springer, pp. 239–251.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007a) The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8:** 172.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007b) CRISPFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* **35:** W52–W57.

Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52:** 696–704.

Haft, D.H., Selengut, J., Mongodin, E.F., and Nelson, K.E. (2005) A guild of 45 CRISPR-associated (cas) protein families and multiple CRISPR/cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1:** 474–483.

Heidelberg, J.F., Nelson, W.C., Schoenfeld, T., and Bhaya, D. (2009) Germ warfare in a microbial mat community: CRISPRs provide insights into the co-evolution of host and viral genomes. *PLoS ONE* **4:** e4169.

Hendrix, R.W., Smith, M.C.M., Burns, R.N., Ford, M.E., and Hatfull, G.F. (1999) Evolutionary relationships among diverse bacteriophages and prophages: all the world's a phage. *Proc Natl Acad Sci USA* **96:** 2192–2197.

Hoehler, T.M., Bebout, B.M., and Des Marais, D.J. (2001) The role of microbial mats in the production of reduced gases on the early earth. *Nature* **412:** 324–327.

Horvath, P., and Barrangou, R. (2010) CRISPR/cas, the immune system of bacteria and archaea. *Science* **327:** 167–170.

Hurwitz, B.L., and Sullivan, M.B. (2013) The pacific ocean virome (pov): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. *PLoS ONE* **8:** e57355.

Hurwitz, B.L., Hallam, S.J., and Sullivan, M.B. (2013) Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14:** R123.

Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11:** 119.

Ignacio-Espinoza, J.C., and Sullivan, M.B. (2012) Phylogenomics of t4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ Microbiol* **14:** 2113–2126.

Karradt, A., Sobanski, J., Mattow, J., Lockau, W., and Baier, K. (2008) Nbla, a key protein of phycobilisome degradation, interacts with clpc, a hsp100 chaperone partner of a cyanobacterial clp protease. *J Biol Chem* **283:** 32394–32403.

Kashtan, N., Roggensack, S.E., Rodrigue, S., Thompson, J.W., Biller, S.J., Coe, A., *et al*. (2014) Single-cell genomics reveals hundreds of coexisting subpopulations in wild prochlorococcus. *Science* **344:** 416–420.

Kristensen, D.M., Waller, A.S., Yamada, T., Bork, P., Mushegian, A.R., and Koonin, E.V. (2013) Orthologous gene clusters and taxon signature genes for viruses of prokaryotes. *J Bacteriol* **195:** 941–950.

Li, H., and Durbin, R. (2009) Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25:** 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., *et al*. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA* **101:** 11013–11018.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine

viruses yield proteins during host infection. *Nature* **438:** 86–89.

Litchman, E. (2000) Growth rates of phytoplankton under fluctuating light. *Freshw Biol* **44:** 223–235.

Litchman, E., Steiner, D., and Bossard, P. (2003) Photosynthetic and growth responses of three freshwater algae to phosphorus limitation and daylength. *Freshw Biol* **48:** 2141–2148.

Liu, X., Kong, S., Shi, M., Fu, L., Gao, Y., and An, C. (2008) Genomic analysis of freshwater cyanophage pf-wmp3 infecting cyanobacterium *Phormidium foveolarum*: the conserved elements for a phage. *Microb Ecol* **56:** 671–680.

Luque, I., and Forchhammer, K. (2008) Nitrogen assimilation and c/n balance sensing. In *The Cyanobacteria: Molecular Biology, Genomics, and Evolution*. Herrero, A., and Flores, E. (eds). Norfolk, UK: Caister Academic Press, pp. 335–382.

Luque, I., Zabulon, G., Contreras, A., and Houmard, J. (2001) Convergence of two global transcriptional regulators on nitrogen induction of the stress-acclimation gene nbla in the *Cyanobacterium synechococcus* sp pcc 7942. *Mol Microbiol* **41:** 937–947.

Ma, Y., Paulsen, I.T., and Palenik, B. (2012) Analysis of two marine metagenomes reveals the diversity of plasmids in oceanic environments. *Environ Microbiol* **14:** 453–466.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006) A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1:** 7.

Makarova, K.S., Aravind, L., Wolf, Y.I., and Koonin, E.V. (2011a) Unification of cas protein families and a simple scenario for the origin and evolution of CRISPR-cas systems. *Biol Direct* **6:** 38.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., *et al*. (2011b) Evolution and classification of the CRISPR-cas systems. *Nat Rev Microbiol* **9:** 467–477.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424:** 741.

Middelboe, M. (2000) Bacterial growth rate and marine virus–host dynamics. *Microb Ecol* **40:** 114–124.

Mizuno, C.M., Rodriguez-Valera, F., Kimes, N.E., and Ghai, R. (2013) Expanding the marine virosphere using metagenomics. *PLoS Genet* **9:** e1003987.

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012) MetaVelvet: an extension of velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res* **40:** e155.

van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M., and Brouns, S.J.J. (2009) CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem Sci* **34:** 401–407.

Paez-Espino, D., Morovic, W., Sun, C.L., Thomas, B.C., Ueda, K., Stahl, B., *et al*. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* **4:** 1430.

Pride, D.T., and Schoenfeld, T. (2008) Genome signature analysis of thermal virus metagenomes reveals archaea and thermophilic signatures. *BMC Genomics* **9:** 420.

Pride, D.T., Sun, C.L., Salzman, J., Rao, N., Loomer, P., Armitage, G.C., *et al*. (2011) Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res* **21:** 126–136.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011) Integrative genomics viewer. *Nat Biotechnol* **29:** 24–26.

Rodriguez-Valera, F., Martin-Cuadrado, A.-B., Rodriguez-Brito, B., Pasic, L., Thingstad, T.F., Rohwer, F., and Mira, A. (2009) Opinion: explaining microbial population genomics through phage predation. *Nat Rev Microbiol* **7:** 828–836.

Ruberg, S.A., Kendall, S.T., Biddanda, B.A., Black, T., Nold, S.C., Lusardi, W.R., *et al*. (2008) Observations of the Middle Island Dinkhole in Lake Huron – a unique hydrogeologic and glacial creation of 400 million years. *Mar Technol Soc J* **42:** 12–21.

Samson, J.E., Magadan, A.H., Sabri, M., and Moineau, S. (2013) Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **11:** 675–687.

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P., and Siksnys, V. (2011) The streptococcus thermophilus CRISPR/cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* **39:** 9275–9282.

Sharon, I., Alperovitch, A., Rohwer, F., Haynes, M., Glaser, F., Atamna-Ismaeel, N., *et al*. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature* **461:** 258–262.

Shi, Y., Tyson, G.W., and DeLong, E.F. (2009) Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266–269.

Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat Rev Microbiol* **6:** 181–186.

Stal, L.J. (2012) Cyanobacterial mats and stromatolites. In *Ecology of Cyanobacteria II: Their Diversity in Space and Time*. Whitton, B.A. (ed.). Heidelberg, Germany: Springer, pp. 65–125.

Sullivan, M.B., Lindell, D., Lee, J.A., Thompson, L.R., Bielawski, J.P., and Chisholm, S.W. (2006) Prevalence and evolution of core photosystem ii genes in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4:** e234.

Sun, C.L., Barrangou, R., Thomas, B.C., Horvath, P., Fremaux, C., and Banfield, J.F. (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* **15:** 463–470.

Tadmor, A.D., Ottesen, E.A., Leadbetter, J.R., and Phillips, R. (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* **333:** 58–62.

Thingstad, T.F. (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr* **45:** 1320–1328.

Tyson, G.W., and Banfield, J.F. (2008) Rapidly evolving CRISPRs implicated in acquired resistance of microorganisms to viruses. *Environ Microbiol* **10:** 200–207.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Voorhies, A.A., Biddanda, B.A., Kendall, S.T., Jain, S., Marcus, D.N., Nold, S.C., *et al*. (2012) Cyanobacterial life at low O_2: community genomics and function reveal metabolic versatility and extremely low diversity in a great lakes sinkhole mat. *Geobiology* **10:** 250–267.

Weinbauer, M.G. (2004) Ecology of prokaryotic viruses. *FEMS Microbiol Rev* **28:** 127–181.

Weinberger, A.D., Sun, C.L., Plucinski, M.M., Denef, V.J., Thomas, B.C., Horvath, P., *et al*. (2012) Persisting viral sequences shape microbial CRISPR-based immunity. *PLoS Comput Biol* **8:** e1002475.

van der Westhuizen, A.J., and Eloff, J.N. (1985) Effect of temperature and light on the toxicity and growth of the blue-green alga *Microcystis aeruginosa* (UV-006). *Planta* **163:** 55–59.

Yoshida, T., Nagasaki, K., Takashima, Y., Shirai, Y., Tomaru, Y., Takao, Y., *et al*. (2008) Ma-lmm01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J Bacteriol* **190:** 1762–1772.

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

Zerbino, D.R., and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18:** 821–829.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Screen captures from Geneious showing reads mapped to PhV1.TypeA. Highlighted bases denote deviation from the template sequence; red arrow denotes low abundance genomic variant matching the template.

Fig. S2. Reads from each sample were mapped to consensus sequences of PhV1 genotypes to evaluate genome completeness in each sample. (A) reads mapped to PhV1 type A, (B) reads mapped to PhV1 type B. Green: genes which could be annotated, orange: hypothetical genes present in NR with no annotation, blue: region that matches a CRISPR III-B spacer, grey: regions with read coverage, black: regions lacking read coverage.

Fig. S3. Screen captures from Geneious showing multiple genomic variants whose reads mapped to PhV1.TypeA. (A–C) Multiple SNP patterns, (D) insertions or deletions. Highlighted bases denote deviation from the template sequence.

Fig. S4. Self organizing ESOM map of the MIS mat community. Each data point represents a 5-kb sequence fragment that has been spatially clustered according to tetranucleotide frequency. Sequences from Phormidium are shown in orange and those from PhV1 are shown in red. Data points from reference genomes included as internal standards and other metagenomic contigs are included but with minimized size because they are not relevant to the current study. To visualize genomic bins, the background color represents the euclidean tetranucleotide distance between data points as shown in the legend. Hence, bins appear as green areas and borders between bins appear as brown or white areas. Note the right and left edges, as well as the top and bottom edges, are contiguous.

Fig. S5. Nucleotide alignment of the Phormidium Type III-B CRISPR region from Locus 1 and Locus 2. Purple: CRISPR repeats with spacers located between repeats, grey: areas of nucleotide agreement, black: nucleotide disagreement.

Fig. S6. Nucleotide alignment of Phormidium CRISPR subtype III-B Locus1 showing sequence conservation in each of the 14 samples. Green: genes which could be annotated, orange: hypothetical genes present in NR with no annotation, purple: CRISPR repeats with spacers located between repeats, grey: regions of nucleotide agreement.

Fig. S7. CRISPR subtype III-B individual spacer and conserved order 'spacer contigs' normalized average read abundance.

Fig. S8. Multivariate analysis of CRISPR spacer contigs and individual spacers. (A) Principal coordinate analyses of CRISPR spacer contigs by year, (B) principal coordinate analyses of individual CRISPR spacers by year, (C) group average analysis of CRISPR spacer contigs by year, (D) group average analysis of individual CRISPR spacers by year.

Fig. S9. cDNA from 2012 samples mapped to PhV1. Note that peaks range from 1× to 65× coverage. (A) PhV1 type A; (B) PhV1 type B. Green: genes which could be annotated, orange: genes with no known function, red: novel genes, blue: region that matches a CRISPR III-B spacer.

Fig. S10. cDNA from 2012 samples mapped to CRISPR Locus1 . (A) Sample 2012-1D, (B) sample 2012-2D, (C) sample 2012-3D, (D) sample 2012-4N, (E) sample 2012-5N; (F) sample 2012-6N, Green: genes which could be annotated, orange: hypothetical genes, purple: CRISPR repeats, with spacers located between repeats. Grey boxes under peaks: individual reads with window cut off at 3× read depth.

Fig. S11. Normalized gDNA read abundance of *Phormidium* and viruses PhV1 type A and B in each sample. (A) abundance of virus PhV1 type A and B based on reads mapped to each genome, (B) abundance of *Phormidium* based on reads mapped to four different single copy housekeeping genes. Blue background shows samples that contain CRISPR.III-B.spacer1 that matches PhV1.TypeB. Samples collected during the day are indicated by D and samples collected at night are indicated by N.

Table S1. PhV1.TypeA annotations.

Table S2. PhV1.TypeB annotations.

Table S3. Occurrence of NblA proteins in finished cyanobacterial genomes.

Table S4. II-B CRISPR repeats.

Table S5. Parameters used for bioinformatics software.

Table S6. Sample read accession information.

Table S7. Sequence accession information.

Table S8. Accession numbers for NblA tree.

© 2015 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*