

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/183795>

How to cite:

The repository item page linked to above, will contain details on accessing citation guidance from the publisher.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work of researchers of the University of Warwick available open access under the following conditions.

This article is made available under the Creative Commons Attribution 4.0 International license (CC BY 4.0) and may be reused according to the conditions of the license. For more details see: <http://creativecommons.org/licenses/by/4.0/>.



Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

Coordinated transcriptional response to environmental stress by a *Synechococcus* virus

Branko Rihtman^{1,3}, Alberto Torcello-Requena¹, Alevtina Mikhaylina¹, Richard J. Puxty¹,
Martha R. J. Clokie², Andrew D. Millard², David J. Scanlan¹

¹ School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK

² Leicester Centre for Phage Research, Department of Genetics and Genome Biology,
University of Leicester, University Road, Leicester, UK

³ Corresponding author: b.rihtman.1@warwick.ac.uk

Address: Branko Rihtman, School of Life Sciences, University of Warwick; Gibbet Hill
Road, Coventry, CV4 7AL, United Kingdom

Running title: Transcriptome of P-stress phage infection

Abstract

Viruses are a major control on populations of microbes. Often, their virulence is examined in controlled laboratory conditions. Yet, in nature, environmental conditions lead to changes in host physiology and fitness that may impart both costs and benefits on viral success.

Phosphorus (P) is a major abiotic control on the marine cyanobacterium *Synechococcus*.

Some viruses infecting *Synechococcus* have acquired, from their host, a gene encoding a P substrate binding protein (PstS), thought to improve virus replication under phosphate starvation. Yet, *pstS* is uncommon amongst cyanobacterial viruses. Thus, we asked how infections with viruses lacking PstS are affected by P scarcity. We show that production of infectious virus particles of such viruses is reduced in low P conditions. However, this reduction in progeny is not caused by impaired phage genome replication, thought to be a major sink for cellular phosphate. Instead, transcriptomic analysis showed that under low P conditions a PstS-lacking cyanophage increased the expression of a specific gene set that included *mazG*, *hli2*, and *gp43* encoding a pyrophosphatase, a high-light inducible protein and DNA polymerase respectively. Moreover, several of the upregulated genes were controlled by the hosts *phoBR* two-component system. We hypothesise that recycling and polymerization of nucleotides liberates free phosphate and thus allows viral morphogenesis, albeit at lower rates than when phosphate is replete or when phages encode *pstS*. Together, our data shows how phage genomes, lacking obvious P-stress related genes, have evolved to exploit their host's environmental sensing mechanisms to coordinate their own gene expression in response to resource limitation.

Keywords: bacteriophage; cyanobacteria; *Synechococcus*, nutrient limitation; phosphorus.

Introduction

Viruses encounter their hosts under a range of environmental conditions. Classically, infection under sub-optimal conditions leads to lysogeny in temperate phages that have a choice between lytic and lysogenic life cycles [1,2]. Additional studies have shown that the decision-making process leading to a lysogenic lifestyle is not necessarily as straightforward as previously thought and can be the result of other environmental triggers than low nutrient levels, such as high microbial density [3]. In contrast, obligately lytic phages have no choice but to pursue a lytic cycle once they infect their host. Whilst much of phage biology is based on laboratory experimentation under defined nutrient conditions, albeit usually using rich media, relatively little is known of phage infection dynamics under conditions of host nutrient limitation.

For many environmental bacteria, like cyanobacteria, growth under nutrient-limiting conditions is the norm, especially those marine genera that occupy the vast oligotrophic ocean gyres that are particularly impoverished for the macronutrients N and P [4]. Whilst many of these marine cyanobacteria have optimised their physiology, biochemistry, and genomic footprint (see [5,6]) to optimise growth under such harsh surroundings, infection by a phage adds yet another burden on the cell. However, compared to host nutrient stress responses, relatively little is known about how such phage-host interactions play out under such sub-optimal growth conditions for the host, particularly the molecular mechanisms that allow phages to lyse such a “sick” host.

Previous work on bacteriophage T4, infecting *E. coli* cells in stationary phase, has shown that bacteriophages infecting cells deprived of nutrients can employ different strategies to deal with this deprivation, in this case altering life strategies from a regular

infectious mode to one of “hibernation”, where some phage proteins are produced yet there is no complete synthesis of mature virions until nutrients become available again [7]. Another infection strategy is scavenging of the few nutrients available, together with utilisation of host cellular building blocks, which often get degraded in the course of infection [8,9], and production of greatly reduced viral progeny from starved cells [7, 10].

Similar to this *E. coli* work, cyanophage S-PM2d infection of the model marine cyanobacterium *Synechococcus* sp. WH7803 during growth under phosphate-deplete conditions also showed evidence of perturbed infection dynamics, noticeably an extended latent period [11]. In contrast, no significant difference in the timing of cell lysis was observed with cyanophage P-SSM2 infecting the marine *Prochlorococcus* strain NATL2A [12] and cyanophage S-SM1 infecting *Synechococcus* sp. WH8102 [13], following P-deplete host growth. It was suggested that expression of a gene encoding a phage version of the high affinity phosphate periplasmic binding protein PstS was a key factor here, potentially maintaining phosphate uptake at a time when host gene expression was being shut down [12]. Indeed, expression of this P-related auxiliary metabolic gene (AMG), appears to be controlled by the host two-component PhoBR system, comprising a sensor (PhoR) and a DNA binding response regulator (PhoB), see [14] for a review, suggesting phages can even hijack host regulatory networks in order to selectively overexpress specific AMGs under the relevant nutrient conditions [12,13].

Not all phages infecting marine cyanobacteria, which thrive in nutrient-deplete environments, possess obvious orthologues of these host metabolic genes, and, given that such types of sub-optimal infections likely play out in great number in real world environments, like ocean gyres where nutrient impoverished sub-optimal host growth

occurs, it is of general interest to investigate how obligately lytic phages optimise these infections under nutrient limiting conditions.

In this study we used a cyanophage-*Synechococcus* model system to specifically address this question, focusing on a phage that lacks obvious AMG genes that could potentially help it optimize the infection process under P-deplete conditions. Using whole-cell transcriptomics and electrophoretic mobility shift assays, we describe a new network of genes up-regulated in a cyanophage genome during infection of a P-deplete host, but controlled by the host PhoBR system, that are widely distributed amongst cyanophages and potentially give a broad indication of how a variety of cyanophages undergo a lytic infection under sub-optimal host growth conditions.

UNCORRECTED MANUSCRIPT

Materials and Methods

Culture growth conditions

Synechococcus sp. WH7803 (<https://roscoff-culture-collection.org/rcc-strain-details/752>)

cultures were grown in defined artificial seawater (ASW) medium [11], containing either 172 μM K_2HPO_4 for phosphate-replete (ASW+P) growth or transferred at a 1:10 volume into ASW medium lacking phosphate (ASW-P), see below, under continuous illumination at a light intensity of $\sim 10 \mu\text{mol photons m}^{-2} \text{ s}^{-1}$ with constant shaking at ~ 220 rpm. Cyanophage S-PM2d [15,16] was propagated on *Synechococcus* sp. WH7803 grown in ASW+P medium. Upon lysis, the lysed culture was filtered through a 0.22 μm pore size vacuum filter (Corning, Corning, NY, USA) and the lysate concentrated with 10% w/v PEG 8000. Concentrated phages were resuspended in ASW+P or ASW-P, the PEG removed by centrifugation with chloroform and phages further concentrated in the appropriate medium, using Amicon filter columns (50,000 kDa MW cut-off, Merck, UK). The phage titre was measured via a plaque assay method [17]. Infection parameters were measured either using plaque assays, or via one-step growth experiments performed in 96-well plates. To elicit phosphate-deplete infection dynamics daily infection assays were performed using 96-well plates as follows: *Synechococcus* sp. WH7803 (100 ml) was grown in ASW+P medium to an $\text{OD}_{750} = 0.35$. The culture was then transferred into 1 L of fresh ASW+P or ASW-P medium. These freshly transferred cultures were grown under the same conditions as the initial starter culture until reaching an $\text{OD}_{750} = 0.25$. Then, infection assays were performed daily, with *Synechococcus* cell numbers being enumerated by flow cytometry using a FACScan (Becton Dickinson, Franklin, NJ, USA). Each sample count was normalised to counts of multifluorescent beads (Polysciences, Warrington, PA, USA), the concentration of which was prior estimated via fluorescent microscopy. Red and orange fluorescence, accounting for

chlorophyll/allophycocyanin and phycoerythrin, respectively, were measured through FL3 (650 nm) and FL2 (585/42 nm) filters respectively. Cell and bead counts were collected using CellQuest software (Becton Dickinson, UK). P-replete and P-deplete *Synechococcus* sp. WH7803 cells (200 μ l) were aliquoted into a 96-well plate in triplicate and cyanophage S-PM2d of known titre added to give a multiplicity of infection (MOI) of 10. A no virus control was also performed, adding ASW or ASW-P medium (100 μ l) instead of phage, to wells containing the uninfected host grown under either P-replete or P-deplete conditions. The microtitre plate was then incubated under continuous illumination at an intensity of ~ 10 μ mol photons $m^{-2} s^{-1}$ and the OD₇₅₀ measured every two hours using an iMark Microplate reader (Bio-Rad, UK). A delayed infection phenotype was indicated when a lysis delay of more than 2 h in the P-deplete, compared to the P-replete culture, was observed. When cells were shown to be sufficiently P-stressed to exhibit a delayed infection, a large-scale infection experiment was performed, and samples were taken every 3 hours for qPCR and RNAseq experiments (see below).

PCR assays to assess the presence of *pstS* in phage genomes

The presence of *pstS* in phage isolates was assessed via PCR amplification using degenerate primers (see Supp. Table 1). Degenerate primers were designed using the HYDEN program [18]. As an input for the HYDEN script, sequences of putative *pstS* genes from six sequenced cyanophages were used (see Supp. Table 2). PCR amplifications were performed using MyTag Mix (Meridian, Memphis, TN, USA) according to the manufacturer's instructions, over 35 amplification cycles with an annealing temperature of 55°C and elongation time of 30 sec.

qPCR assays

Upon onset of the lysis delay, as detected by the 96-well plate assay described above, *Synechococcus* sp. WH7803 cells growing in either P-replete or P-depleted ASW medium were enumerated using flow cytometry, as described above, and cultures diluted to a final concentration of 5×10^7 cells ml^{-1} in ASW/ASW-P medium respectively. Diluted cultures (5 ml) were aliquoted into polycarbonate tubes and infected with cyanophage S-PM2d at an MOI of 10 in triplicate. Uninfected cultures were used as a control. Samples were taken at time points 0, 2, 4, 6, 8, 10, 12-, 14-, 16-, and 18 hours post-infection. 200 μl of each sample was diluted in 500 μl ASW medium and vacuum filtered through a 0.2 μm pore size polycarbonate Isopore filter (Merck Millipore, Burlington, MA, USA), mounted on a glass filter tower. Whilst still on the filter, samples were washed three times with 1 ml preservation solution (100 mM EDTA, 500 mM NaCl, 10 mM Tris-HCl, pH 8.0). The filters were then inserted into a ribolyser Lysing Matrix E tube (MP Bioproducts, Irvine, CA, USA) and snap-frozen in liquid nitrogen. An additional 100 μl infected culture was fixed with paraformaldehyde (1% final concentration) and stored at 4°C for subsequent cell counting by flow cytometry.

To obtain cell lysates, 650 μl Tris-HCl pH 8.0 was added to each filter-containing ribolyser tube and cells lysed via 3 cycles of 30 s shaking at 30 Hz in a TissueLyser Qiagen (Retsch GmbH, Germany). The tubes were then subjected to centrifugation at 10,000 g and the supernatant snap-frozen in liquid nitrogen and stored at -80°C. To calculate the intracellular phage DNA concentration samples were diluted 1:10 (v/v) and qPCR performed as described elsewhere [19].

Whole-cell transcriptomics

A 1 litre culture of *Synechococcus* sp. WH7803 was grown under P-replete and P-deplete conditions, until a delayed lysis phenotype was observed in the P-deplete compared to the P-replete culture (as detected via the 96-well plate assay, see above). Cells were then enumerated using flow cytometry, as described above, and cultures diluted to 5×10^7 cells ml^{-1} in ASW+P/ASW-P medium, respectively. The ASW+P and ASW-P cultures were then divided into six replicates of 100ml each. Three replicates of each condition were infected with cyanophage S-PM2d at a MOI of 8, whilst the remaining three replicates were left uninfected as the no-phage control. At each time point, 15 ml sample was taken from each replicate and filtered through a 0.2 μm pore size polycarbonate filter. Filters were washed three times using preservation buffer (100 mM EDTA, 500 mM NaCl, 10 mM Tris-HCl, pH 8.0), snap frozen in ribolyser tubes in liquid nitrogen and stored at -80°C .

RNA was extracted and purified according to a previous publication [20]. The RNA concentration was quantified with a Qubit Fluorometer (Invitrogen, Waltham, MA, USA) using the Qubit RNA HS Assay and RNA integrity verified using Bioanalyzer RNA 6000 Pico chips (Agilent, Santa Clara, CA, USA).

Samples for RNA sequencing were sent to the Next Generation Sequencing Facility at the Leeds Institute of Biomedical and Clinical Sciences, St James University Hospital, University of Leeds. Libraries were prepared using the ScriptSeq Complete Kit (Bacteria) and ScriptSeq v2 RNA-seq Kit (Illumina, San Diego, CA, USA). Ribosomal RNA was removed using the RiboZero Kit – Bacteria (Illumina). Paired end sequencing was performed on the HiSeq 3000 platform, producing 150 base reads.

Paired reads were mapped to both the cyanophage S-PM2d and *Synechococcus* sp. WH7803 genomes using the BWA MEM v0.7.13 [21] with default parameters. The resulting SAM files were converted to BAM files and sorted using SAMtools v1.3.1 [22]. To count the reads mapping to specific S-PM2d loci, sorted BAM files were first converted to BED files using the Bedtools bamtobed script [23]. BED files were then used to count the reads mapping to different S-PM2d genes using the Bedtools intersect script with '-c -bed -s' options. Reads were normalised using the RPKM model, giving an estimate of relative gene expression calculated according to the following equation:

$$RPKM = \frac{Read\ Count}{Gene\ length * 10^{-3} * Total\ Reads * 10^{-6}}$$

where Read Count represents the number of reads mapping to a specific locus, Gene Length is the length of that locus and Total Reads represents the number of total reads out of each sample mapping to the S-PM2d and *Synechococcus* sp. WH7803 genomes.

To establish changes in gene expression of S-PM2d genes between the corresponding time points in ASW+P and ASW-P infected cultures, differential gene expression values were calculated using the DESeq2 R package [24] in the R Studio environment. The gene expression fold change values, as well as statistical significance were calculated using the DESeqDataSetFromHTSeqCount function from the DESeq2 package. Only genes with a False Discover Rate (FDR) p value <0.05 were considered to be differentially expressed between the conditions.

Bioinformatics analysis of putative Pho boxes in differentially regulated S-PM2d genes

Promoters of S-PM2d genes identified as differentially expressed under -P conditions, were examined for the presence of putative PhoB-binding motifs. Promoter sequences were

analysed using the Pattern Locator script [25]. The consensus sequence previously identified as a putative Pho box binding site in cyanobacteria - 5'- PyTTAAPyPyT/A-3' - [26] was used to scan the upstream region of these S-PM2d genes and several potential binding sites were identified.

Over-expression of the *Synechococcus* sp. WH7803 PhoB protein in *E. coli*

The *Synechococcus* sp. WH7803 *phoB* gene (Syn_WH7803_01545) was codon optimized for expression in *E. coli* and cloned into the pMAL c4X expression vector (Genscript, Piscataway, NJ, USA) to produce a maltose binding protein (MBP)-PhoB fusion protein. The construct was used to chemically transform NEBExpress competent *E. coli* and transformants selected using 100 µg ml⁻¹ ampicillin. After reaching OD₆₀₀≈0.5 in LB medium PhoB protein expression was induced using 0.3 mM IPTG for 4 h. Cells were then harvested by centrifugation, resuspended in BugBuster lysis buffer (Merck Millipore), and lysed using a French Press. Soluble and insoluble fractions were separated via centrifugation and the overexpressed MBP-PhoB fusion protein purified via an AKTA fast protein liquid chromatography (FPLC) system (GE Healthcare, Chicago, IL, USA) using a MBPTrap column (GE Healthcare). The amount of purified MBP-PhoB was quantified using a Bradford assay (ThermoFisher, Waltham, MA, USA) according to the manufacturer's instructions. To purify the MBP protein alone, which was used as a control in Electrophoretic Mobility Shift Assays (EMSA), over-expressed MBP-PhoB fusion protein was eluted with 10 mM maltose in loading buffer (20 mM Tris, pH 8; 2 mM MgCl₂, 200 mM NaCl) and Factor Xa protease (NEB, USA) added to a final concentration of 10 µg ml⁻¹ in loading buffer containing 2 mM CaCl₂. and incubated at 4°C overnight on a shaker. Factor Xa protease was removed using a HiTrap Q XL column (Cytiva, UK) washed with loading buffer. Separated MBP and PhoB proteins were eluted using a 0-1 M NaCl gradient. Purified MBP eluted at ~300 mM NaCl and PhoB at ~500 mM

NaCl. Fractions were collected, concentrated, and further purified by size-exclusion chromatography in loading buffer using a Superdex 200 10/300 column (Cytiva, UK) and an AKTA fast protein liquid chromatography (FPLC) system (GE Healthcare, Chicago, IL, USA). Note that purified PhoB alone was insoluble as evidenced by the formation of a precipitate in solution after removal of the MBP tag, elution of the protein as an unexpectedly high molecular weight peak on the AKTA column and protein stuck in the well in native PAGE.

Electrophoretic mobility shift assays

Varying amounts of purified *Synechococcus* sp. WH7803 MBP-PhoB fusion protein were incubated with 25 ng DNA fragments encoding phage gene promoter sequences (see Supp. Table 1) in 10 μ l 2x binding buffer (10 mM Tris-HCl pH 8, 50 mM KCl, 2 mM MgCl₂, 5% (v/v) glycerol, 0.05 mg mL⁻¹ BSA, 1 mM DTT, 3 mM spermidine) for 30 min at 23°C. 2 μ l 5x loading buffer (binding buffer, 20% (v/v) glycerol) was added and the binding reaction run on a pre-cast 10% acrylamide gel (Invitrogen) at 100V for 90 min at 4°C. The gel was visualised using an ImageQuant LAS 4000 (GE Healthcare). For binding competition assays, DNA fragments were labelled by amplifying the specific genomic region using FAM-labelled primers (IDT, USA) targeting the upstream regions of S-PM2d134 and S-PM2d004 (Supp. Table 1)

Genome sequencing of cyanophages S-BM1 and S-BM3

An exponentially growing *Synechococcus* sp. WH7803 culture (100 ml) was infected with 100 μ l of filtered cyanophage (S-BM1 or S-BM3) lysate and incubated for several days at 23°C, under illumination at a light intensity of $\sim 10 \mu\text{mol photons m}^{-2} \text{s}^{-1}$ with constant shaking at ~ 220 rpm. Lysates were subsequently centrifuged at 13,000 g to remove cell debris and filtered through 0.22 μm pore size syringe filters. Filtered lysates were then used to extract viral DNA, using a phenol-chloroform extraction method described in [27].

Purified DNA was then used in Nextera library preparations and sequenced using a MiSeq (Illumina, read length 150 bp) by MicrobesNG (<https://microbesng.com/>). Fastq files were trimmed with Sickle v1.33, using default parameters [28], and assembled with SPAdes v3.12.0, with the careful option [29]. In each case, the assembly produced a single phage contig with >100-fold coverage. Single contigs were then annotated with Prokka v.1.12 [30] against a custom database [31,32] derived from previously published bacteriophage genomes.

Phylogenetic analysis

Cyanophage and cyanobacterial PstS sequences were extracted from the NCBI nr database, Cyanorak (<http://abims.sb-roscoff.fr/cyanorak/>) and Cyanobase (<http://genome.microbedb.jp/cyanobase/>), using BLAST. Sequences were aligned using ClustalO [33]. Trees were produced using FastTree 2.1 [34] with default settings, under the Jones-Taylor-Thorton ML model, visualised, and annotated using the iTol phylogenetic trees online annotation tool [35]. *Kyanoviridae* PstS protein sequences were recovered using a uBLAST search against all published *Kyanoviridae* phage genome sequences [31], using the uSEARCH v11 script with the following parameters “-ublast -evaluate 1e-20” [36]. Sequences resulting from this query were added to the list of orthologous proteins, aligned, ordered in a phylogenetic tree, visualised, and annotated as described above. To examine congruency between the *Kyanoviridae* cyanophage core tree and the resulting PstS tree, pairwise leaf-leaf distances were extracted from each tree. Weighted and un-weighted Robinson Foulds distances [37] were calculated using TreeDist in R.

The *Kyanoviridae* cyanophage core genome tree was generated using a list of 57 core genes, as defined in [38] and [39]. Hidden Markov Model (hmm) files of each of the

core genes were generated using HMMER 3.3.2, with hmmbuild with default settings, followed by hmmpress. Thus, generated hmm files were used to query a database of curated cyanophage genomes [31] and MAFFT v7.490 [40] was used to create separate alignments of each core gene homologue using a custom Perl script listed in Supplementary Data 2. The alignments were concatenated using catfasta2phyml.pl (<https://github.com/nylander/catfasta2phyml>), and a phylogenetic tree generated using FastTree 2.1 [34] with default settings, under the Jones-Taylor-Thorton ML model, visualised, and annotated with the iTol phylogenetic trees online annotation tool [35]. The cyanophage DNA polymerase tree was produced by performing a protein BLAST similarity search against a custom database containing protein sequences of cyanophage genomes using an e-value threshold of 1E-7. The list of 466 cyanophage genomes was generated by searching the INPHARED database [31] for words “cyano*”, “Prochlorococcus”, “Synechococcus”, and the names of cyanobacterial genera listed in Supp. Table 2 of a previous publication [41]. The sequences were aligned and the tree was produced in a manner described above.

A random representative list of cyanophages belonging to the *Kyanoviridae* (previously *Myoviridae*) and *Autographiviridae* (previously *Podoviridae*) families (roughly corresponding to myovirus and podovirus morphotypes respectively) was generated, covering different clades of the DNA polymerase tree shown in Supp. Fig. 7 (chosen genomes are marked by a star symbol). For each of those genomes, a Genbank file was generated, containing the DNA polymerase gene and 5 genes upstream and downstream from it, using a custom-made python script listed in Supplementary Data 2. Those GenBank files were used to demonstrate synteny between gene neighbours of the DNA polymerase genes, using Clinker [42].

Results

Evolution of PstS within cultured *Kyanoviridae* cyanophages

To understand the distribution and evolution of the *pstS* gene in cyanophages, we analysed 95 currently available complete genomes from the *Kyanoviridae* cyanophages. Of these, *pstS* was present in 30 isolates (~1/3). Genomes containing *pstS* were distributed across the entire phylogeny of this group (Fig. 1a), suggesting either widespread loss or frequent acquisition of the gene. To understand which, we constructed a phylogenetic tree of cyanophage PstS proteins as well as freshwater and marine host proteins (Fig. 1b). Unlike *E. coli*, *pstS* in cyanobacteria is often found in multi-copy (see [43]). Here, we utilise the Cyanorak database [44] to assign a *sphX* cluster (CK_00001829) and two *pstS* clusters (CK_00000023 and CK_00043821) whilst taxonomically we also identify a discrete β -cyanobacterial, ‘freshwater-dominated’ PstS cluster which encompasses both the *pstS1* and *pstS2* genes characterised experimentally by [43]. Cyanophages possess only *pstS* cluster CK_00043821. The absence of *sphX* or *pstS* cluster CK_00000023 in cyanophages is notable, given that *Synechococcus* hosts frequently possess all three types (themselves often in multiple copy), whereas *Prochlorococcus* hosts only possess the *pstS* CK_00043821 cluster. The freshwater PstS cluster shown in Fig. 1b encompasses proteins encoded by the *pstS1* and *pstS2* genes in *Synechocystis* sp. PCC6803 which are known to have quite different transport kinetics for inorganic phosphate [43]. Thus, it is difficult to predict functional properties of PstS proteins from sequence alone, which precludes understanding why cyanophages have only co-opted the CK_00043821 PstS cluster.

Our phylogeny suggests *pstS* has been acquired twice in marine cyanophages, once from *Synechococcus* and once from *Prochlorococcus* (Fig. 1b). In most cases (24/30) the

donor genus (*Synechococcus* or *Prochlorococcus*), corresponds to the isolation host of the phage, even though isolation host is not phylogenetically conserved (Fig. 1a). Thus, it is likely that these PstS variants facilitate specific interaction with the cognate PstABC membrane bound components of the transport system of their preferred host (see also Zhao et al. [45]). In the six cases where *pstS* has not been acquired from the host genus it is not clear what the host range of these phages are, and therefore the interaction of these “incorrect” PstS types with their host’s PstABC requires further investigation. Within each acquisition event, we assessed the congruency between PstS subtrees and the core phylogeny. For those *pstS* genes acquired from *Prochlorococcus* there was a significant but weak correlation between phylogenetic tree topologies. For *Synechococcus* there was no significant correlation (Fig. 1c), suggesting that once acquired, there is frequent horizontal gene transfer of *pstS* between phages. This may explain their rather patchy distribution across the core phylogenetic tree (Fig. 1a). The distribution of *pstS* in these phage genomes could reflect niche adaptation in low P environments, as has been observed in marine metaviromes [46,47,48]. Similar to previous reports [39], we observed that the *phoA* gene encoding a putative alkaline phosphatase, is only present in phage genomes that also possess a *pstS* gene (Fig. 1a).

PstS-lacking cyanophages show delayed lysis in response to low P

To assess whether the absence of *pstS* in cyanophage genomes more generally correlated with a delayed lysis phenotype [11], we tested PstS-lacking cyanophages isolated from a variety of marine environments under +P/-P conditions (Supp. Table 3). We used both plaque assays (Figs. 2a and 2b), measuring the yield and the potential delay in production of plaques, as well as one-step infection assays, measuring the OD₇₅₀ of phage-infected cells over the latent period of infection that indicates when cell lysis occurs (Fig. 2c and 2d). We

further used degenerate primers targeting the phage *pstS* (Supp. Table 1). to assess the presence of *pstS* in genomes of cyanophages isolated from a variety of marine environments (Supp. Table 3).

We found a strong correlation between the absence of a *pstS* homologue in the phage genome and the presence of the delayed lysis phenotype when infecting under -P conditions. Of the 18 phages tested, 17 lacked *pstS* and showed a severe delay in lysis (see Fig. 2c for one example). In contrast, a newly sequenced phage, S-BM1, which encodes PstS, had identical latent periods under both P deplete and P replete conditions (Supp. Table 3, Fig. 2d) in agreement with previous data for two other *pstS*-possessing cyanophages P-SSM2 and S-SM1 that also show no delayed lysis [12,13]. Therefore, it appears that cyanophage-encoded PstS supports the phosphorus demand of an infected cell to avoid delayed lysis.

Is the DNA replication rate affected in *pstS*-lacking cyanophage infecting a P-deplete host?

Since phage PstS likely controls proper lysis time in response to P availability, we turned our attention to phages lacking *pstS*, in a bid to understand how they were limited. We hypothesized that phages lacking *pstS* would have a decreased DNA replication rate when infecting a P-deplete host. However, for the *pstS*-lacking cyanophage S-PM2d, phage DNA replication rates were not statistically different 2-6 hours post infection between the P-replete and P-deplete conditions (2 Sample T-test values: $t = 0.78$, $p = 0.48$, Fig. 3A and C).

The dramatic impact on phage infection kinetics with a lack of P in the medium on those cyanophages that do not possess *pstS* (Figs. 2a and 2c, Supp. Table 3), points towards a distinct molecular mechanism, either controlled by the host or the phage, by which the infection process is delayed, and the burst size reduced, despite still maintaining similar DNA replication rates (Figs. 3a and 3c). From the host perspective, this mechanism would

reduce the rate of infection of the P-depleted host by the phosphorus-demanding phages, allowing persistence of the host population throughout periods of P-scarcity in the dynamic marine environment. From the cyanophage perspective, the altered kinetics would mean an alternative to undertaking an abortive infection pathway– and even though the infection is significantly affected and sub-optimal, the low number of viral progeny produced would still allow phage persistence, facilitating future host infection once the depleted nutrient(s), in this case P, become available again.

How then, does the cyanophage maintain a DNA replication rate similar to nutrient replete conditions when infecting a P-deplete grown host? Or how does the phage (or indeed host) control the latent period or burst size? To determine this, we undertook a transcriptomics approach comparing gene expression under P-replete/P-deplete host growth.

Differential gene expression in response to P availability

Synechococcus sp. WH7803 grown in either P-replete or P-deplete conditions was infected with cyanophage S-PM2d at a multiplicity of infection of 8, and samples taken for RNA analysis at t= 3h, 6h, and 9h or additionally at t= 12h and 15h post-infection under P-deplete host growth. Sequence coverage and mapping statistics of this RNAseq data are shown in Supp. Table 4. We noted that the relative abundance of transcripts mapping to the *Synechococcus* host genome in the infected samples is significantly lower, compared to the uninfected control (Supp. Fig 1 and Supp. Table 4), pointing to the immediate degradation of the host transcriptome, similar to what has been observed previously [8,19]. This reiterates the importance of AMGs in phage genomes, with host transcription being abruptly halted and phage relying only on host genes expressed in the early stages of infection as well as its own genes to facilitate the metabolic needs of its infection cycle.

We did not observe any differentially expressed host genes in response to P conditions in infected cells (data not shown). We expect this is due to the above-mentioned degradation of host mRNAs early in infection. In contrast, 14 S-PM2d genes were statistically significantly up-regulated 3h post-infection under P-deplete conditions, compared with the P-replete control (Table 1; Supp. Fig. 2). These included six S-PM2d genes located in close proximity on the genome i.e. S-PM2d131 to S-PM2d136 of which S-PM2d133 encodes DNA polymerase (gp43), S-PM2d134 - a UvsX RecA-like recombination protein, S-PM2d135 - a DNA primase-helicase (gp41), and S-PM2d136 - the MazG pyrophosphatase [49]. Most of the other 8 significantly up-regulated genes are annotated as encoding hypothetical proteins apart from S-PM2d118 (gp46), encoding a recombination endonuclease subunit, and S-PM2d172, encoding a high-light inducible protein. There were an additional three genes that were identified as differentially expressed but are probably misannotated in this reference genome (Table 1).

Several of the differentially regulated S-PM2d genes under P-deplete host growth are controlled by the *Synechococcus* PhoBR system

Since these cyanophage S-PM2d genes were specifically upregulated following infection of a P-deplete host, we sought to understand how these genes were up-regulated. Given it is known that the *pstS* gene in cyanophage P-SSM2 infecting marine *Prochlorococcus* is controlled by the host two-component system PhoBR [13] we investigated whether this was the case for these genes as well. Using a Pho box consensus motif previously described for marine *Synechococcus* as comparison [26], bioinformatics and visual inspection of promoter regions showed seven of the 14 up-regulated genes possessed a putative Pho box (see Table 2). To experimentally confirm that some of these putative Pho boxes were functional, the *Synechococcus* sp. WH7803 PhoB protein was over-expressed and purified in *E. coli* as a

maltose binding protein (MBP)-PhoB fusion protein and the purified protein used in electrophoretic mobility shift assays with DNA fragments of S-PM2d upstream regions either i) containing a putative Pho box and differentially expressed (genes S-PM2d004, S-PM2d133, S-PM2d136) ii) containing a putative pho-box, but not differentially expressed (gene S-PM2d130) or iii) lacking a predicted Pho box but differentially expressed (S-PM2d134).

Of the upstream regions of the genes tested, the MBP-PhoB fusion protein bound to the promoters of S-PM2d004 (encoding a hypothetical protein), S-PM2d133 (encoding DNA polymerase), both with a predicted Pho box but also to SPM2d134 (encoding a putative UvsX, RecA-like protein), which lacks a predicted Pho box (Fig 4). These three promoters showed no binding to the purified MBP alone (Supp. Fig. 6). S-PM2d136 showed no binding to the PhoB-MBP fusion (Supp. Fig. 3), despite being differentially expressed in response to low P and possessing a predicted Pho box. However, this gene may be co-transcribed with the upstream genes S-PM2d135 and S-PM2d134, the latter experimentally determined to bind PhoB. Alternatively, one or more of the over-expressed genes may be controlled by another transcription factor that is in turn under PhoB control. Gene S-PM2d130 had a predicted Pho box, but did not bind PhoB (Supp. Fig. 6), nor was differentially expressed in response to low P. Thus, bioinformatics prediction of Pho boxes in phage genomes is not reliable and instead experimental validation of PhoB binding better predicts whether the gene was differentially expressed in response to low P. To confirm the specificity of PhoB binding we performed competition assays in which fluorescently-labelled upstream DNA fragments for S-PM2d004 and S-PM2d134 were competed by identical unlabelled DNA fragments, where the binding of the protein to the promoter was outcompeted by increasing the concentration of the unlabelled promoter-containing DNA fragment (Supp.

Fig. 4a) whilst competition with a non-binding fragment (S-PM2d136) showed no effect (Supp. Fig. 4b).

In order to confirm that it is indeed the predicted Pho box to which the overexpressed PhoB protein is binding, we progressively reduced the size of these fragments to show that for both the S-PM2d004 (Supp. Fig. 5a) and S-PM2d133 (Supp. Fig. 5b) genes only those DNA fragments containing the predicted PhoB binding site resulted in a gel shift.

Discussion

Cyanophage infection represents a potentially important loss factor for picocyanobacterial populations throughout the global ocean via cell lysis (e.g. see [50]) but also has wide ranging implications for estimates of primary production given their ability to directly inhibit CO₂ fixation [51]. Thus, given that cyanophage infections can theoretically occur under every environmental condition possible, it is important we have a mechanistic understanding of how infection dynamics vary as a function of these *in situ* host growth conditions, given this may directly affect the latent period (the time before cell lysis, but also the time during which CO₂ fixation is inhibited), and burst size (the number of phages produced).

Phosphorus availability has previously been shown to affect cyanophage productivity in picocyanobacteria [12,13] but with no observed delayed lysis period. Here, we found a more general relationship between the presence of a *pstS* homologue in phage genomes and the lack of the delayed lysis phenotype under -P conditions (Fig. 1a; Supp. Table 3). The acquisition of *pstS* may therefore enable cyanophages to overcome a bottleneck affecting lysis during phage morphogenesis. Phage *pstS* transcripts can be used to either produce additional copies of the periplasmic P-binding PstS protein, thus increasing the ability to

acquire any available phosphate, or they can supplement the mRNA encoding the host *pstS* which is degraded together with the rest of the host transcriptome during the course of infection ([8]; Supp. Fig. 1), thus providing a continuous supply of the PstS protein in the periplasm of the infected cell. A similar strategy has been shown to be used by cyanophages carrying a photosynthetic gene *psbA* which encodes a functional D1 protein and enables maximal energy production under high-light conditions in the infected cyanobacterial cell [19,52,53].

Phylogenetic examination of the cyanobacterial *pstS* shows strong clustering between the viral genes and the genes of the specific host these viruses infect (Fig. 1b). Thus, *pstS* sequences from *Synechococcus* phages form a sister clade to the *Synechococcus* *pstS* sequences, while the *Prochlorococcus* viral *pstS* sequences group together with the cyanobacterial *Prochlorococcus* *pstS* clade. This clear phylogenetic relationship becomes less obvious when looking at the totality of *Kyanoviridae* phages. Here (Fig. 1a), the P-stress related gene *pstS*, as well as the putative alkaline phosphatase-encoding *phoA*, seem to have evolved numerous times in different phages and do not show clustering that is apparent when looking only at the cyanobacterial and cyanophage *pstS* sequences (Fig. 1b). Additionally, there seems to be no synteny conservation between *pstS* homologs found in different isolated cyanophage genomes, further supporting the multiple origin theory [54]. This observation points to the possibility that adopting host *pstS* is not the only strategy available to phages infecting hosts under P-deplete conditions and that alternative infection strategies, one of which is described in this work, may be more prevalent than we previously thought. The existence of alternative strategies for optimising infection under different P-conditions, the lack of synteny of *pstS*-surrounding genes between cyanophage isolates, and its apparent sporadic and repeated acquisition over ecological timescales, all suggest that,

unlike their cyanobacterial hosts, cyanophages have not evolved into phosphorus-adapted ecotypes.

We have also shown that the delayed lysis phenotype of cyanophage S-PM2d during infection of a P-deplete host [11] appears to be a general phenotype of *pstS* lacking cyanomyoviruses (Supp. Table 3). Given that S-PM2d genome replication rate was comparable during infection of a P-replete or P-deplete host (Fig. 3a and 3c), excludes this as the underlying mechanism. Indeed, we show that under P-deplete conditions, phage genes for DNA replication are specifically upregulated and that phage genomes have evolved to exploit the host's regulatory system to control this. Since the infectious burst size is five-fold lower under P-deplete conditions (Fig. 2; [11]), replicated phage genomes do not appear to get incorporated into infectious virions. How and why cyanophages maintain the synthesis of DNA, which is extremely phosphorus demanding (e.g. almost 50% of cellular phosphorus is in DNA [55]) is initially puzzling. However, cyanophages have an abundant source of nucleotides in the form of the host's chromosome which is immediately degraded upon infection [56], liberating free nucleotides. In *E. coli* phage T4, host chromosome degradation is catalysed by a combination of the endonuclease II product of the *denA* gene and the gene product of *d2a* [57]. Whilst no homologues of these exist in *Kyanoviridae* cyanophages, it is likely that another endonuclease performs this function. In unrelated cyanophages, host mRNAs are degraded in a process thought to be a result of the host's RNase E, with phages protecting their mRNAs by antisense transcription [8]. A similar mechanism may exist in *Kyanoviridae* cyanophages where antisense transcription is observed [8]. Free NTPs can be converted to dNTPs by the activity of the cyanophage-encoded ribonucleotide reductase [58,59]. Together, it is likely there is enough free or recycled nucleotides derived from host nucleic acids to meet the demands of phage chromosome replication. Moreover,

DNA replication itself releases intracellular phosphate, via the pyrophosphatase activity of DNA polymerase [60]. Since *Synechococcus* cells infected under P-deplete conditions produce 12 new cyanophages on average [11,61] and that each S-PM2d genome is approximately 200 kb in length [16], whilst the addition of each nucleotide to the growing DNA chain produces 2 phosphate molecules, we estimate that during the course of DNA replication approximately 10^6 molecules of phosphate are released by the activity of DNA polymerase alone. Previous work has shown that *Synechococcus* cells growing at external phosphate concentrations of 1 μ M accumulate a similar amount of phosphate per cell per hour [62]. Thus, DNA replication may provide a rich source of phosphorus for high P demanding processes of phage morphogenesis.

We also observed that the phage *mazG* (S-PM2d136, Table 1) is up-regulated under P deplete conditions. S-PM2d *mazG* encodes a pyrophosphohydrolase which was initially hypothesised to play a role in suppressing the stringent response in infected cells [63]. However, our previous work showed that the viral MazG does not hydrolyze the alarmone nucleotides ppGpp and pppGpp, but rather hydrolyses nucleotides, with increased affinity for dGTP and dCTP [49]. Since cyanophage S-PM2d has a lower %GC (37.7%; [16]) compared to its *Synechococcus* sp. WH7803 host (60.2%; [64]), it was proposed that the viral MazG has a role in preferentially hydrolysing nucleotides for which it has lower demand, thus providing a potential additional source of intracellular phosphate produced by hydrolysis of less-required nucleotides. This observation further supports our model of amplification of nucleotide metabolism (including DNA replication) processes to provide a novel source of intracellular phosphate. In reality, there may even be an excess of nucleotides for DNA replication. Instead, cyanophages may scavenge phosphate from these to meet other phosphorus demands such as protein synthesis and packaging during the later stages of

infection which have been implicated as the main energetic sink for phage production [19].

Thus, it is possible that the role of the phosphate produced intracellularly during phage DNA replication is to provide the energy for the later stages of the latent phase of infection.

Since viral DNA polymerase seems to play an important role in the P-stress response of this group of cyanophages, we examined the phylogeny of DNA polymerases from published genomes of isolated cyanophages (Supp. Fig. 7), as well as the preservation of synteny of DNA polymerase in these genomes (Supp. Fig. 8). The phylogeny of DNA polymerase seems to align with the cyanophage family phylogeny with the main clades clustering along the family divide, with the *Kyanoviridae* family members possessing the Type B DNA polymerase whereas the *Autographiviridae* contain the Type A DNA polymerase (Supp. Fig. 7). There is no separate clustering between the *Kyanoviridae* DNA polymerases belonging to phages infecting *Prochlorococcus* and *Synechococcus* hosts, pointing towards a possibility of interchangeable acquisition of the gene from the common host of these cyanophages. Noteworthy is the split nature of the DNA polymerase gene found in most of the *Autographiviridae* strains (19/32). In these phages, DNA polymerase is divided into two or three ORFs, found next to each other, often partially overlapping, and encoded in a different reading frame to one another. This feature seems to be conserved amongst cyanophages from this family, as the second and third polymerase fragments all cluster together in a separate branch within the *Autographiviridae* clade. While this feature has been reported before [65], further work is required to establish whether this curious feature has any biological importance.

For both the *Kyanoviridae* and *Autographiviridae* DNA polymerases there is a high degree of synteny conservation within the two clades. In the case of *Kyanoviridae*

polymerases, including the one belonging to cyanophage S-PM2d described in this work, there is the conserved presence of a helicase, UvsX-like recombinase, putative heat-shock protein and MazG homologues surrounding the DNA polymerase (Supp. Fig. 8a). Since we find all of these genes, together with the DNA polymerase, to be transcriptionally overexpressed during infection under P-deplete conditions, it is possible that this synteny is conserved as a part of a general cyanophage adaptation to infection under varying P-stress conditions. Certainly, this strategy appears to have been specifically selected for in S-PM2d given DNA polymerase (S-PM2d133) is under direct control of the host P-stress transcriptional regulation response (Fig. 4). Noteworthy here is that the host DNA polymerase is not under the same P-stress control [66,67]. It is also notable that in the case of some cyanophages infecting both *Synechococcus* and *Prochlorococcus*, the immediate genomic context of the DNA polymerase also contains a *pstS* homologue. This proximity of *pstS* and DNA polymerase in some of these *Kyanoviridae* cyanophage genomes might explain the origin of the Pho-binding motifs in the promoters of these latter genes. The genomes of *Autographiviridae* cyanophages, although also showing a high degree of synteny around their DNA polymerases (Supp. Fig. 8B), contain a somewhat different (compared to *Kyanoviridae* genomes) list of putative genes located in this genomic neighbourhood: ssDNA binding protein, RNaseH, nucleotide kinase, ribonucleotide kinase and RNA polymerase, with the conservation of *mazG* between these families, a gene which has previously been identified as a part of the core genome of *Kyanoviridae* cyanophages [38,39].

Overall, our results indicate that bacteriophages not only adapt to nutrient stress by incorporating host metabolic genes into their own genomes [13], but also acquire stress-specific promoters to exploit host transcription factors related to environmental stress. This

mechanism presents an exciting new model of viral evolution which should be examined in other virus-host models.

UNCORRECTED MANUSCRIPT

Data Availability

The datasets generated during the current study are available as follows: Genome sequences of cyanophages S-BM1 and S-BM3 are available in Genbank under accession numbers OQ319120 and OQ319121, respectively. Paired-end reads used for cyanophage transcriptomics analyses have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB64420.

UNCORRECTED MANUSCRIPT

References:

1. Weinbauer MG. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 2004;28, 127–181.
2. Wahl LM, Betti MI, Dick DW, Pattenden T, Puccini AJ. Evolutionary stability of the lysis-lysogeny decision: why be virulent? Evolution 2019;73: 92-98.
3. 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.
4. Wu J, Sunda W, Boyle EA, Karl DM. Phosphate depletion in the western North Atlantic Ocean. Science 2000; 289: 759-762.
5. Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, *et al.* Ecological genomics of marine picocyanobacteria. Microbiol. Mol. Biol. Rev. 2009;73: 249-299.
6. Biller SJ, Berube PM, Lindell D, Chisholm SW. *Prochlorococcus*: the structure and function of collective diversity. Nature Rev. Microbiol. 2015;13: 13-27.
7. Bryan D, El-Shibiny A, Hobbs Z, Porter J, Kutter EM. Bacteriophage T4 infection of stationary phase *E. coli*: life after log from a phage perspective. Front. Microbiol. 2016;7: a1391.
8. Stazic D, Pekarski I, Kopf M, Lindell D, Steglich C. A novel strategy for exploitation of host RNase E activity by a marine cyanophage. Genetics 2016;203: 1149-1159.
9. Chevallereau A, Blasdel BG, De Smet J, Monot M, Zimmermann M, Kogadeeva M, *et al.* Next-generation “-omics” approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*. PLoS Genetics 2016;12: e1006134.
10. Moebus K. Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. I. Investigations with six phage-host systems. Mar. Ecol. Prog. Ser. 1996;144: 1-12.

11. Wilson WH, Carr NG, Mann NH. The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. *J. Phycol.* 1996;23: 506–516.
12. Lin X, Ding H, Zeng Q. Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. *Environ. Microbiol.* 2016;18: 450–460.
13. Zeng Q, Chisholm SW. Marine viruses exploit their host's two-component regulatory system in response to resource limitation. *Current Biol.* 2012;22: 124–128.
14. Santos-Beneit F. The Pho regulon: a huge regulatory network in bacteria. *Front. Microbiol.* 2015;6:a402.
15. Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, Wheatley PJ, *et al.* The genome of S-PM2, a “photosynthetic” T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* 2005;187: 3188-3200.
16. Puxty RJ, Perez-Sepulveda B, Rihtman B, Evans DJ, Millard AD, Scanlan DJ. Spontaneous deletion of an “ORFanage” region facilitates host adaptation in a “photosynthetic” cyanophage. *PLoS One* 2015;10: e0132642.
17. Millard AD. Isolation of cyanophages from aquatic environments. *In* Clokie MRJ. and Kropinski AM. (Eds) *Bacteriophages: Methods and Protocols, Volume 1: Isolation, Characterization, and Interactions.* Totowa, NJ: Humana Press, 2009. pp. 33–42.
18. Linhart C, Shamir R. The degenerate primer design problem: Theory and applications. *J. Comp. Biol.* 2005;12: 431-456.
19. Puxty RJ, Evans DJ, Millard AD, Scanlan DJ. Energy limitation of cyanophage development: implications for marine carbon cycling. *ISME J.* 2018;12: 1273-1286.

20. Pitt FD, Millard A, Ostrowski M, Dervish S, Mazard S, Paulsen IT, *et al.* A sample-to-sequence protocol for genus targeted transcriptomic profiling: Application to marine *Synechococcus*. *Front. Microbiol.* 2016;7: a1592.
21. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997. 2013 Mar 16.
22. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25: 2078-2079.
23. Quinlan AR, Hall IM. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26: 841–842.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15: 1–21.
25. Mrázek J, Shaohua X. Pattern locator: a new tool for finding local sequence patterns in genomic DNA sequences. *Bioinformatics* 2006;22: 3099-3100.
26. Su Z, Olman V, Xu Y. Computational prediction of Pho regulons in cyanobacteria. *BMC Genomics* 2007;8: 1-12.
27. Rihtman B, Meaden S, Clokie MR, Koskella B, Millard AD. Assessing Illumina technology for the high-throughput sequencing of bacteriophage genomes. *Peer J* 2016;4: e2055.
28. Joshi NA, Fass JN. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) 2011. [Software available at <https://github.com/najoshi/sickle>].
29. Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, *et al.* Assembling genomes and mini-metagenomes from highly chimeric reads. *In Annual International Conference on Research in Computational Molecular Biology, 2013.* pp. 158-170. Springer, Berlin, Heidelberg

30. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30: 2068-2069.
31. Cook R, Brown N, Redgwell T, Rihtman B, Barnes M, Clokie M, *et al.* Infrastructure for a phage reference database: Identification of large-scale biases in the current collection of cultured phage genomes. *Phage* 2021;2: 214-223.
32. Terzian P, Olo Ndela E, Galiez C, Lossouarn J, Pérez Bucio RE, Mom R, *et al.* PHROG: families of prokaryotic virus proteins clustered using remote homology. *NAR Genomics and Bioinformatics*. 2021;3:lqab067.
33. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 2011;7: 539.
34. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. *PloS One* 2010;5: e9490.
35. Letunic I, Bork P. Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 2019;47: W256-W259
36. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26: 2460-2461.
37. Böcker S, Canzar S, Klau, GW. The Generalized Robinson-Foulds Metric. In: Darling, A., Stoye, J. (eds) *Algorithms in Bioinformatics. WABI 2013. Lecture Notes in Computer Science*, 8126. Springer, Berlin, Heidelberg, 2013.
38. Millard AD, Zwirgmaier K, Downey MJ, Mann NH, Scanlan DJ. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ. Microbiol.* 2009;11: 2370-2387.

39. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, *et al.* Genomic analysis of oceanic cyanobacterial myoviruses compared with *Kyanoviridae* myoviruses from diverse hosts and environments. *Environ. Microbiol.* 2010;12: 3035-3056.
40. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 2013;30: 772-780.
41. Walter JM, Coutinho FH, Dutilh BE, Swings J, Thompson FL, Thompson CC. Ecogenomics and taxonomy of Cyanobacteria phylum. *Front. Microbiol.* 2017;8: a2132
42. Gilchrist CL, Chooi YH. Clinker & clustermap: Automatic generation of gene cluster comparison figures. *Bioinformatics* 2021;37: 2473-2475.
43. Pitt FD, Mazard S, Humphreys L, Scanlan DJ. Functional characterization of *Synechocystis* sp. strain PCC 6803 *pst1* and *pst2* gene clusters reveals a novel strategy for phosphate uptake in a freshwater cyanobacterium. *J. Bacteriol.* 2010;192: 3512-3523.
44. Garczarek L, Guyet U, Doré H, Farrant GK, Hoebeke M, Brillet-Guéguen L, *et al.* Cyanorak v2. 1: a scalable information system dedicated to the visualization and expert curation of marine and brackish picocyanobacteria genomes. *Nucleic Acids Res.* 2021;49: D667-76.
45. Zhao F, Lin X, Ca, K, Jiang Y, Ni T, Chen Y, *et al.* Biochemical and structural characterization of the cyanophage-encoded phosphate-binding protein: implications for enhanced phosphate uptake of infected cyanobacteria. *Environ. Microbiol.* 2022;24:3037-3050.

46. Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J.* 2013;7: 1827-1841.
47. Fuchsman CA, Carlson MC, Garcia Prieto D, Hays MD, Rocap G. Cyanophage host-derived genes reflect contrasting selective pressures with depth in the oxic and anoxic water column of the Eastern Tropical North Pacific. *Environ. Microbiol.* 2021;23: 2782-2800.
48. Fuchsman CA, Garcia Prieto D, Hays MD, Cram JA. Associations between picocyanobacterial ecotypes and cyanophage host genes across ocean basins and depth. *PeerJ* 2023;11: e14924
49. Rihtman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MR, Scanlan DJ. Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ. Microbiol. Rep.* 2019;11: 448-455.
50. Carlson MC, Ribalet F, Maidanik I, Durham BP, Hulata Y, Ferrón S, *et al.* Viruses affect picocyanobacterial abundance and biogeography in the North Pacific Ocean. *Nature Microbiol.* 2022;7:570-80.
51. Puxty RJ, Millard AD, Evans D.J, Scanlan DJ. Viruses inhibit CO₂ fixation in the most abundant phototrophs on Earth. *Curr. Biol.* 2016;26:1585-1589.
52. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 2005;438: 86-89.
53. Clokie MR, Shan J, 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.

54. Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JB. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* 2016;499: 219-229.
55. Bertilsson S, Berglund O, Karl DM, Chisholm SW. Elemental composition of marine *Prochlorococcus* and *Synechococcus*: Implications for the ecological stoichiometry of the sea. *Limnol. Oceanogr.* 2003;48: 1721-1731.
56. Kutter E, Bryan D, Ray G, Brewster E, Blasdel B, Guttman B. From host to phage metabolism: hot tales of phage T4's takeover of *E. coli*. *Viruses* 2018;10: 387.
57. Souther A, Bruner R, Elliot J. Degradation of *Escherichia coli* chromosome after infection by bacteriophage T4: Role of bacteriophage gene D2a. *J. Virol.* 1972;10: 979-984.
58. Lundin D, Gribaldo S, Torrents E, Sjöberg BM, Poole AM. Ribonucleotide reduction-horizontal transfer of a required function spans all three domains. *BMC Evol. Biol.* 2010;10: 1-19.
59. Sullivan MB, Coleman ML, Weigle P, Rohwer F, Chisholm SW. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biology.* 2005;3: e144.
60. Kottur J, Nair DT. Pyrophosphate hydrolysis is an intrinsic and critical step of the DNA synthesis reaction. *Nucleic Acids Res.* 2018;46: 5875–5885.
61. Brown CM, Lawrence JE, Campbell DA. Are phytoplankton population density maxima predictable through analysis of host and viral genomic DNA content? *J Mar. Biol. Assoc. UK* 2006;86: 491–498.
62. Kamennaya NA, Geraki K, Scanlan DJ, Zubkov MV. Accumulation of ambient phosphate into the periplasm of marine bacteria is proton motive force dependent. *Nature Comms.* 2020; 11: 2642.

63. Bryan MJ, Burroughs NJ, Spence EM, Clokie MR, Mann NH, Bryan SJ. Evidence for the intense exchange of MazG in marine cyanophages by horizontal gene transfer. *PLoS One* 2008;3: e2048.
64. Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP, *et al.* Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol.* 2008;9: R90.
65. Flores-Urbe J, Philosof A, Sharon I, Fridman S, Larom S, Bèjà, O. A novel uncultured marine cyanophage lineage with lysogenic potential linked to a putative marine *Synechococcus* 'relic' prophage. *Environ. Microbiol. Rep.* 2019;11: 598-604.
66. Martiny AC, Coleman ML, Chisholm SW. Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc. Natl. Acad. Sci. USA* 2006;103: 12552-12557.
67. Tetu, SG, Brahamsha B, Johnson DA, Tai V, Phillippy K, Palenik B, *et al.* Microarray analysis of phosphate regulation in the marine cyanobacterium *Synechococcus* sp. WH8102. *The ISME J.* 2009;3:835-849.

Acknowledgements:

BR was in receipt of a University of Warwick Chancellor's International scholarship and AT by a NERC CENTA PhD studentship. This work was also funded by the European Research Council under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no 883551) to DJS.

Author Contributions:

DJS, ADM, RJP, and MRJC designed the study. BR, AT, AM, and RJP performed experiments and analysed data. BR, RJP, and DJS wrote the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing Interests:

The authors declare that they have no conflict of interest.

UNCORRECTED MANUSCRIPT

Figure legends

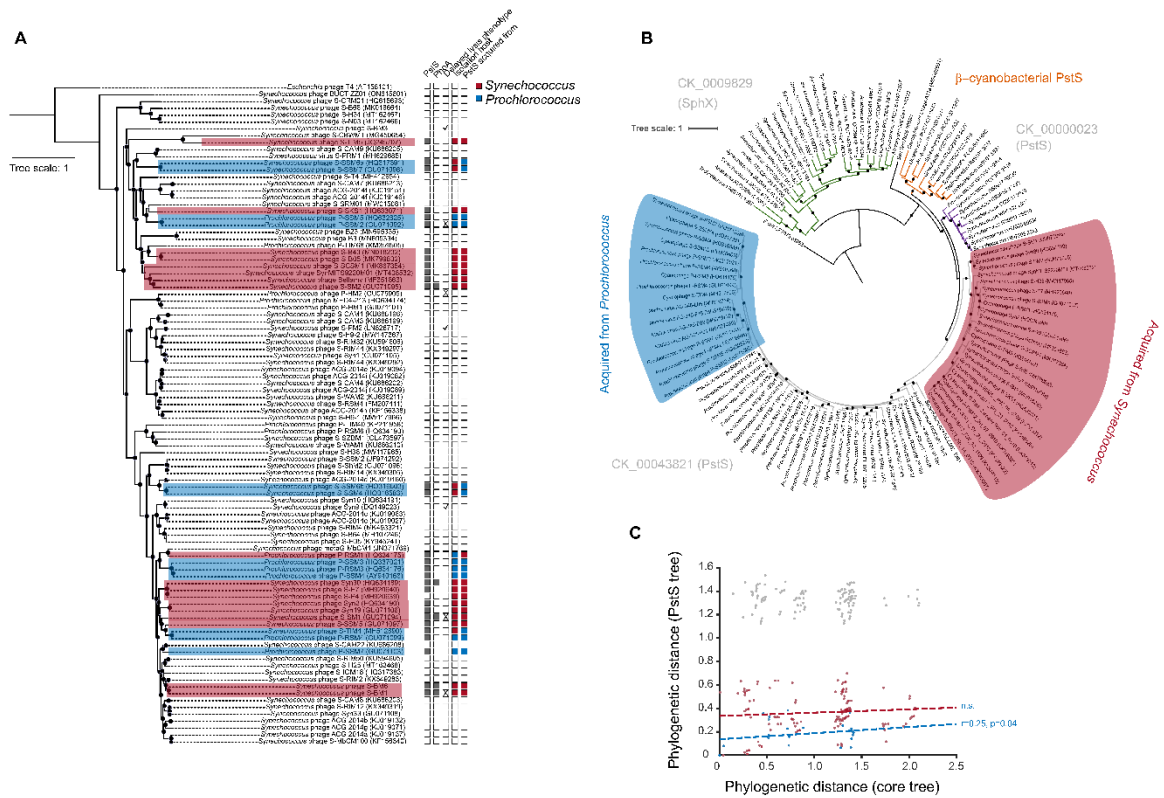


Figure 1: The distribution and evolution of phosphate acquisition AMGs in *Kyanoviridae* cyanophages. (a) Core phylogenetic tree of *Kyanoviridae* cyanophages constructed from 57 gene markers. The tree is rooted on *E. coli* phage T4. The host genus, the presence/absence of *pstS*, *phoA*, and the delayed lysis phenotype is shown for each taxon. Black circles on branch junctions indicate bootstrap values >80% (b). Phylogeny of PstS proteins from cyanophages and cyanobacteria. The tree is rooted with *E. coli* PstS. Black circles on branch junctions indicate bootstrap values >80%. (c) Comparison of the phylogenetic distance of cyanophage PstS proteins and their phylogenetic distance in the core tree. Grey dots indicate correlation between all cyanophage PstS proteins while red and blue signify correlations within the red and blue groups highlighted in (b). Output linear regression statistical tests (Pearson's correlation) are shown next to the line.

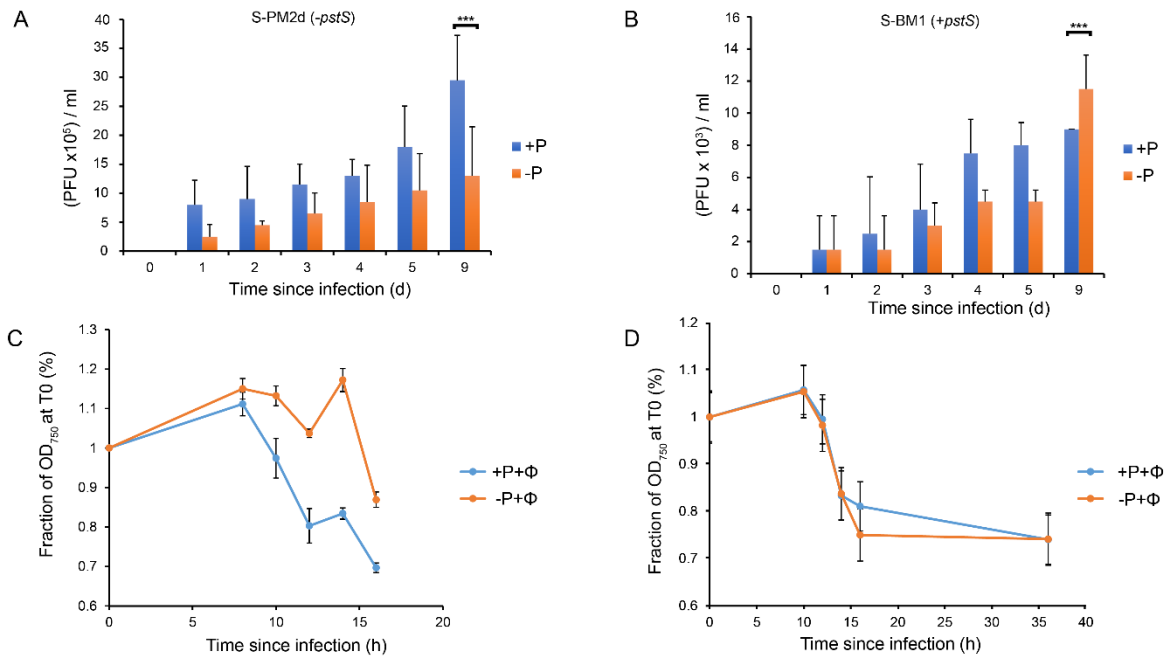


Figure 2: Plaque assay results [(a) and (b)] and OD₇₅₀ values [(c) and (d)] of cyanophage infection of *Synechococcus* sp. WH7803 grown under P-replete and P-deplete conditions. Cyanophage used: (a) and (c) S-PM2d; (b) and (d) S-BM1. Error bars represent the standard error of the average of three replicates. T-tests between the number of plaques at the last time point show statistically significant differences between +P and -P-infected samples, both for S-PM2 (t value: 3.511, p value: 0.025) and S-BM1 (t value: -3.024, p value: 0.039).

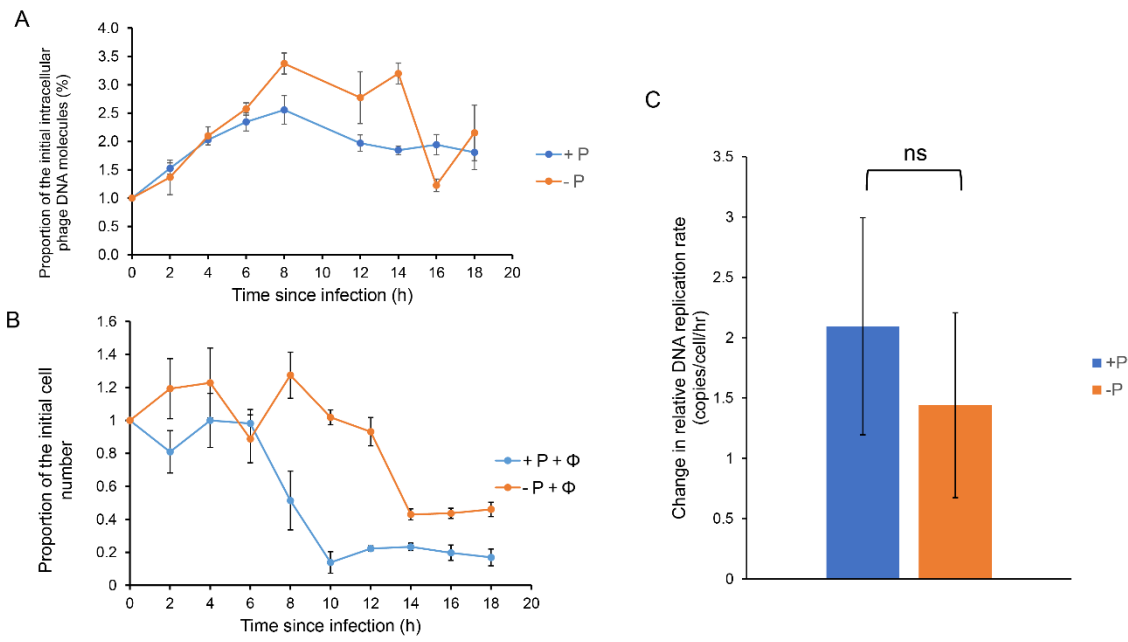
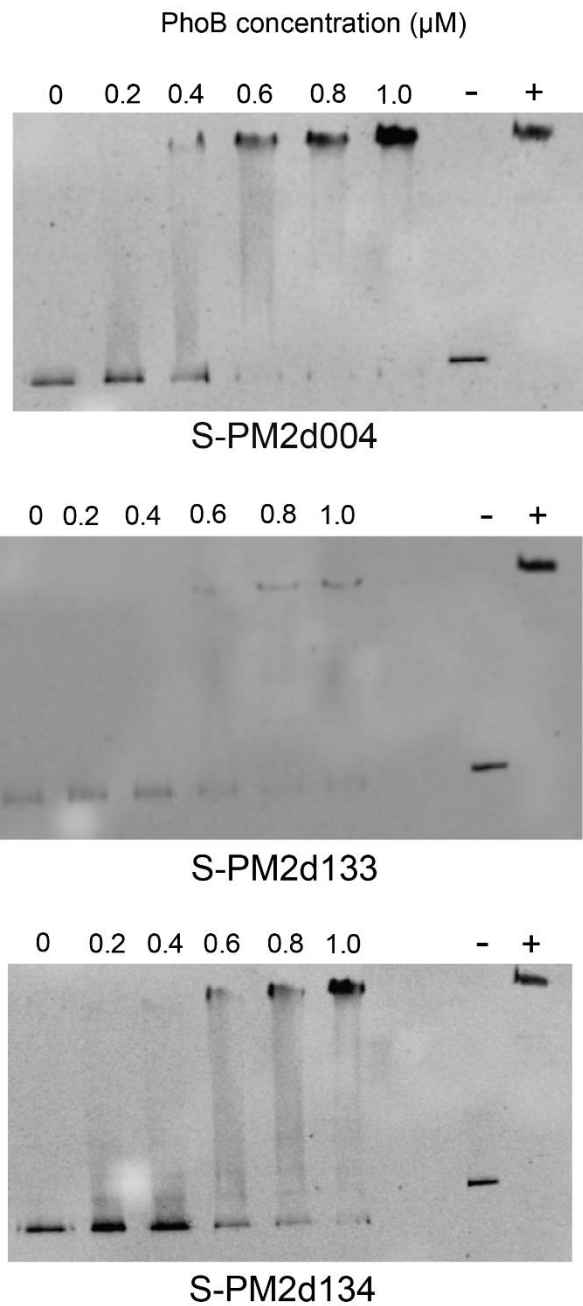


Figure 3: DNA replication rate of cyanophage S-PM2d during infection of *Synechococcus* sp. WH7803 under P-replete and P-deplete conditions. (a) Percentage of the initial amount of intracellular S-PM2d phage DNA over the course of infection under P-replete and P-deplete conditions. (b) The percentage of initial *Synechococcus* sp. WH7803 cell abundance following infection with cyanophage S-PM2d under P-replete and P-deplete conditions. (c) The relative cyanophage DNA replication rate under P-replete and P-deplete conditions. The rate was estimated by calculating the slope of a curve representing the change in the amount of DNA per hour for each of the three replicates, between 2- and 6-hour time points post infection. Error bars represent the standard deviation of three replicates.



MANUSCRIPT

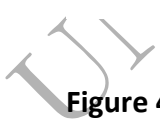


Figure 4: Electrophoretic Motility Shift Assays (EMSA) of purified MBP-PhoB from *Synechococcus* sp. WH7803 with specific cyanophage S-PM2d gene promoters. The concentration of MBP-PhoB protein used ranged from 0-1 μM whilst 25 ng DNA fragment was used in each case. (-): negative control, PhoB with an internal fragment of the

Synechococcus sp. WH7803 *phoB* gene. (+): positive control, PhoB with the promoter region of the *Synechococcus* sp. WH7803 *phoB* gene.

UNCORRECTED MANUSCRIPT

1 **Tables**

2 **Table 1.** Differentially expressed cyanophage S-PM2d genes during infection of *Synechococcus* sp. WH7803 under P-deplete compared to P-
 3 replete conditions.

CDS	Genbank accession number	Product	Gene	Start codon position	Stop codon position	Strand	log2 Fold Change	p value	FDR-adjusted p value	Presence in % of cyanophage genomes
S-PM2d004	CFW42138.1	hypothetical protein		1255	1560	+	0.79	3.06E-03	4.36E-02	24.44% (143/585)
S-PM2d006	CFW42141.1	hypothetical protein		1913	2116	+	1.12	6.28E-04	1.60E-02	2.74% (16/585)
S-PM2d074*1	CFW42190.1	hypothetical protein		14226	14432	-	1.03	1.74E-03	3.06E-02	0.34% (2/585)
S-PM2d241	CFW42194.1	hypothetical protein		14805	15005	+	1.00	1.51E-03	2.88E-02	0.17% (1/585)
S-PM2d115	CFW42285.1	hypothetical protein		75481	75699	+	0.63	4.18E-03	5.03E-02	13.50% (79/585)
S-PM2d118	CFW42290.1	recombination endonuclease subunit	gp46	77014	78744	+	0.98	1.61E-04	6.15E-03	74.70% (437/585)
S-PM2d131	CFW42313.1	hypothetical protein		87134	87601	+	1.12	5.76E-05	4.40E-03	9.40% (55/585)
S-PM2d132	CFW42314.1	hypothetical protein		87598	87849	+	1.04	1.18E-04	5.86E-03	9.23% (54/585)
S-PM2d133	CFW42316.1	DNA polymerase	gp43	87849	90341	+	0.63	2.62E-03	3.99E-02	84.44% (494/585)
S-PM2d134	CFW42318.1	UvsX RecA-like recombination protein	UvsX RecA-like	90352	91383	+	0.84	1.21E-03	2.78E-02	75.04% (439/585)
S-PM2d135	CFW42319.1	DNA primase-helicase	gp41	91343	92755	+	0.67	3.24E-03	4.36E-02	74.53% (436/585)
S-PM2d136	CFW42320.1	Pyrophosphatase	mazG	92757	93164	+	1.33	1.11E-05	2.11E-03	74.53% (436/585)

S-PM2d169	CFW42374.1	hypothetical protein		114418	114768	+	1.42	1.84E-05	2.11E-03	26.67% (156/585)
S-PM2d172	CFW42378.1	High light inducible protein	hli2	115295	115414	+	0.76	1.41E-03	2.88E-02	50.94% (298/585)
S-PM2d190	CFW42403.1	hypothetical protein		132902	133072	+	0.87	3.10E-04	1.01E-02	0.85% (5/585)
S-PM2d250* ²	CFW42449.1	putative ATPase		155233	155376	+	0.71	3.19E-02	1.61E-01	0.17% (1/585)
S-PM2d216* ²	CFW42450.1	hypothetical protein		155436	155594	+	0.91	2.01E-03	3.29E-02	0.17% (1/585)

4 *Indicates the CDS is most likely misannotated in the reference genome since ¹The CDS is on the negative strand and contains no orthologs in
5 cyanophage genomes whilst there is a more probable overlapping ORF on the positive strand which is shared amongst cyanophages but is not
6 differentially expressed. ²These are very small ORFs (<50 aa's), not shared across cyanophage genomes and are located within a tRNA operon.

7

8 **Table 2.** The sequence and position of putative Pho boxes found upstream of the
9 differentially expressed cyanophage S-PM2d genes during infection of *Synechococcus* sp.
10 WH7803 under P-deplete compared to P-replete conditions. Bold nucleotides represent
11 those pertaining to the Pho box, whilst underlined nucleotides are conserved compared to
12 the recognised consensus Pho box binding site (5'-PyTTAAPyPyT/A-3') [26]. NA – no putative
13 Pho box was identified. Position is the number of nucleotides upstream of the putative
14 transcriptional start site of that gene.

Gene name	Gene annotation	Putative Pho Box	Position
S-PM2d004	hypothetical protein	CTTCCTCTCGTTTAACTG ATTATACAATTCTTGGAGA	-62 and -25
S-PM2d006	hypothetical protein	NA	
S-PM2d074	hypothetical protein	NA	NA
S-PM2d115	hypothetical protein	NA	NA
S-PM2d118	recombination endonuclease subunit (gp46)	NA	NA
S-PM2d130	Heat shock protein (Hsp20)	ATTCAAACCTCGCTTAAATA	-31
S-PM2d131	hypothetical protein	TTTGAAGATGGTCTTTTGAAGATCTCACTGG	-138
S-PM2d132	hypothetical protein	CTAATCCAACCTTGATGC	-91
S-PM2d133	DNA polymerase (gp43)	GTTAAAATATCCATATGCA	-162
S-PM2d134	UvsX RecA-like	NA	
S-PM2d135	DNA primase-helicase (gp41)	NA	NA
S-PM2d136	Pyrophosphatase (<i>mazG</i>)	CTTTTGAAGACCTTTCAACTCGCCAAAAGC	-63
S-PM2d169	hypothetical protein	NA	
S-PM2d172	hli2	NA	NA
S-PM2d190	hypothetical protein	NA	NA
S-PM2d216	hypothetical protein	NA	NA
S-PM2d241	hypothetical protein	CTTCCCCGCCGCTGAGGTG	-105
S-PM2d250	putative ATPase	CCTATACTGATCTCAGTTC	-30

15