npg

ORIGINAL ARTICLE

Transcriptome dynamics of a broad host-range cyanophage and its hosts

Shany Doron[1,6], Ayalla Fedida[2,6], Miguel A Hernández-Prieto[3,7], Gazalah Sabehi[2], Iris Karunker[1], Damir Stazic[4], Roi Feingersch[2], Claudia Steglich[4], Matthias Futschik[3,5], Debbie Lindell[2] and Rotem Sorek[1]

[1]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot, Israel; [2]Faculty of Biology, Technion—Israel Institute of Technology, Haifa, Israel; [3]Faculty of Sciences and Technology, University of Algarve, Faro, Portugal; [4]Faculty of Biology, University of Freiburg, Freiburg, Germany and [5]Centre of Marine Sciences, University of Algarve, Faro, Portugal

Cyanobacteria are highly abundant in the oceans and are constantly exposed to lytic viruses. The T4-like cyanomyoviruses are abundant in the marine environment and have broad host-ranges relative to other cyanophages. It is currently unknown whether broad host-range phages specifically tailor their infection program for each host, or employ the same program irrespective of the host infected. Also unknown is how different hosts respond to infection by the same phage. Here we used microarray and RNA-seq analyses to investigate the interaction between the Syn9 T4-like cyanophage and three phylogenetically, ecologically and genomically distinct marine *Synechococcus* strains: WH7803, WH8102 and WH8109. Strikingly, Syn9 led a nearly identical infection and transcriptional program in all three hosts. Different to previous assumptions for T4-like cyanophages, three temporally regulated gene expression classes were observed. Furthermore, a novel regulatory element controlled early-gene transcription, and host-like promoters drove middle gene transcription, different to the regulatory paradigm for T4. Similar results were found for the P-TIM40 phage during infection of *Prochlorococcus* NATL2A. Moreover, genomic and metagenomic analyses indicate that these regulatory elements are abundant and conserved among T4-like cyanophages. In contrast to the near-identical transcriptional program employed by Syn9, host responses to infection involved host-specific genes primarily located in hypervariable genomic islands, substantiating islands as a major axis of phage–cyanobacteria interactions. Our findings suggest that the ability of broad host-range phages to infect multiple hosts is more likely dependent on the effectiveness of host defense strategies than on differential tailoring of the infection process by the phage.
The ISME Journal (2016) **10,** 1437–1455; doi:10.1038/ismej.2015.210; published online 1 December 2015

Introduction

Marine cyanobacteria from the genus *Synechococcus* are major primary producers, contributing significantly to the global carbon cycle (Scanlan, 2003). *Synechococcus* strains are divided into distinct clades based on their phylogeny and physiology (Rocap *et al.*, 2002; Fuller *et al.*, 2003; Scanlan *et al.*, 2009) and have different distribution patterns that correlate with distinctive environmental conditions (Zwirglmaier *et al.*, 2008). Their genomes range in

size from 2.2 to 2.9 Mbp with 50–67% of their genes (~1570) being common to all strains (Dufresne *et al.*, 2008; Scanlan *et al.*, 2009). A substantial number of additional genes are found in only a subset of the strains, which are generally localized in hypervariable genomic islands (Palenik *et al.*, 2003; Dufresne *et al.*, 2008). Island-residing genes are often differentially expressed in response to environmental stressors (Coleman *et al.*, 2006; Thompson *et al.*, 2011a; Stuart *et al.*, 2013; Tetu *et al.*, 2013), as well as in response to viral infection (Lindell *et al.*, 2007).

The oceans are teeming with viruses, with global estimates reaching 10^{30} virus-like particles (Suttle, 2005). They are thought to have a major impact on host mortality (Fuhrman, 1999; Suttle, 2005) and evolution (Avrani *et al.*, 2011; Marston *et al.*, 2012), and on the biogeochemical cycling of matter (Suttle, 2005; Breitbart, 2012). Both broad and narrow host-range phages are known to exist in the oceans (Miller *et al.*, 2003a; Sullivan *et al.*, 2003;

1438

Suttle, 2005; Comeau et al., 2006). Narrow host-range (specialist) phages are assumed to infect highly abundant hosts, whereas broad host-range (generalist) phages are expected to be prevalent when abundances of different host types are low or variable (Woolhouse et al., 2001; Sullivan et al., 2003; Elena et al., 2009; Dekel-Bird et al., 2015). Thus, broad host-range phages infect genetically diverse hosts (Sullivan et al., 2003; Elena et al., 2009; Dekel-Bird et al., 2015) and can respond rapidly to increases in population abundance of a wide range of host types (Suttle, 2005).

Representatives of all three families of tailed double-stranded DNA viruses, the Myoviridae, the Podoviridae and the Siphoviridae, are known to infect marine cyanobacteria. Of these the T4-like cyanomyoviruses are well-known to have a broad host-range (Waterbury and Valois, 1993; Suttle and Chan, 1993; Sullivan et al., 2003; Millard and Mann, 2006; Wang and Chen, 2008; Dekel-Bird et al., 2015). Furthermore, they are considered to be one of the most abundant cyanophages in the oceans based on isolation studies and metagenomics surveys (Marston and Sallee, 2003; Sullivan et al., 2003; Suttle, 2005; Millard and Mann, 2006; Angly et al., 2006; Bench et al., 2007). They are named T4-like cyanophages owing to their morphological and genomic similarity to the enteric Escherichia coli T4 phage (Sullivan et al., 2005; Mann et al., 2005; Weigele et al., 2007; Millard et al., 2009; Sullivan et al., 2010). Their genomes are 174–253 kb in size with 215–334 genes, of which 40–50 are homologous to T4 replication and virion structural genes. In addition, they carry a number of auxiliary metabolic genes that are of (cyano)bacterial origin, such as photosynthesis, carbon metabolism and phosphorus utilization genes (Mann et al., 2005; Sullivan et al., 2005; Weigele et al., 2007; Millard et al., 2009; Sullivan et al., 2010) that are likely to be highly relevant for phage fitness in the environment.

The temporal transcriptional events during phage infection of a single host are known to be highly reproducible (Pène and Uzan, 2000; Uzan, 2009). However, it is currently not known whether the transcriptional program of broad host-range phages is differentially tailored for infection of distinct hosts or whether the same program is maintained irrespective of the host infected. In addition, little is known about the transcriptional response of different hosts to infection by the same phage. This is of particular interest in light of the expectation that generalism comes with tradeoffs and thus differential fitness during infection of diverse hosts (Woolhouse et al., 2001; Elena et al., 2009). Furthermore, there is a growing understanding that different bacteria employ distinct defense strategies in response to phage infection (reviewed in Labrie et al., 2010; Stern and Sorek, 2011) and that these defense mechanisms are often encoded by strain-specific genes

(Jabbar and Snyder, 1984; Riede and Eschbach, 1986; Mosig et al., 1997; Rifat et al., 2008).

The transcriptional program of the ecologically important T4-like cyanophages has yet to be investigated, although studies limited to a few genes have been carried out (Clokie and Mann, 2006; Thompson et al., 2011b). Phage transcriptional programs are generally orchestrated in three expression classes of early, middle and late genes, corresponding to host take-over, replication and virion morphogenesis, respectively (Guttman et al., 2005; Staley et al., 2007). In T4, this is regulated by the sequential modification of the host RNA polymerase and associated sigma factor, leading to the consecutive recognition of different promoter sequences: early T4 promoters resemble the major E. coli promoters recognized by the primary host sigma-70 (σ^{70}) factor; middle promoters have a distinctive motif that are recognized with the aid of two phage proteins, AsiA and MotA; and expression from late promoters requires a phage-encoded sigma factor, gp55 (reviewed in Mosig and Eiserling, 2006). The lack of T4-like middle promoters and the motA and/or asiA genes led to the suggestion that a variety of T4-like phages, including marine cyanophages, have a simpler expression program consisting of only two modules (Desplats et al., 2002; Miller et al., 2003a; Mann et al., 2005; Clokie et al., 2006). However, this hypothesis has not yet been tested.

In this study we investigated the infection process and transcriptional program of Syn9, a broad host-range T4-like cyanophage (Waterbury and Valois, 1993), during infection of three different Synechococcus hosts and assessed their transcriptional response to infection using both RNA-seq and microarray analyses. The three Synechococcus hosts, WH7803, WH8102 and WH8109, are phylogenetically distinct and occupy different oceanic environments (Zwirglmaier et al., 2008; Scanlan et al., 2009). We found that the transcriptional program of the phage is nearly identical irrespective of the host infected and revealed a regulatory program that significantly deviates from the current paradigm for T4-like phages. This transcriptional and regulatory program appears to be common among T4-like cyanophages as a similar infection process was observed for an additional phage, P-TIM40, during infection of Prochlorococcus NATL2A, and similar promoter motifs were identified in numerous sequenced cyanophage genomes and in metagenomes. Our results indicate that the well-known mode of regulation in T4 cannot be taken as the rule among the broader family of T4-like phages. In contrast to the near-identical transcriptional program of the Syn9 phage, transcriptional responses of the three Synechococcus hosts showed considerable heterogeneity, with a large number of responsive genes located in hypervariable genomic islands. This likely reflects different strategies in the hosts' attempts to cope with infection.

Materials and methods

Culture growth and experimental design
Synechococcus strains were grown in an artificial seawater-based medium (Wyman *et al.*, 1985) with modifications as described in Lindell *et al.*, 1998, at 22 °C under cool white light with a 14 : 10 h light: dark cycle at an intensity of 30 µmol photon $m^{-2} s^{-1}$ during the light period. Generation times were: 2.1 ± 0.1 days for WH7803 and WH8102 and 1.4 ± 0.1 days for WH8109. The Syn9 phage was isolated from the Atlantic Ocean (Waterbury and Valois, 1993) and has 232 genes in its genome, of which ~20% are homologs of T4 genes (Weigele *et al.*, 2007). Syn9 was propagated on *Synechococcus* WH8102. Cultures (800 ml) were infected at a multiplicity of infection of 2–3 ~ 2 h after 'lights on' in the morning. The infection cycle was complete 4 h before darkness. Samples were collected at different time points during the lytic cycle (5 min and 0.5, 1, 1.5, 2, 3, 4, 5, 6, 8, 10 h after phage addition), with samples from the appropriate time points used to assess host and phage RNA transcription, host DNA degradation, phage DNA replication and release of infective phage. Three independent infection experiments were carried out on each of the *Synechococcus* hosts for all analyses. Where relevant, results were compared with uninfected control cultures.

Prochlorococcus NATL2A was grown in the seawater-based medium Pro99 (Moore *et al.*, 2007) with Mediterranean seawater under the same conditions described above. Generation time under these conditions was 1.5 ± 0.1 days. The P-TIM40 phage was isolated from the Pacific Ocean and its genome was sequenced as part of this study (see Supplementary Text). Sequencing revealed a 188.6 kb genome and a GC content of 40.7%. Its genome encodes 235 predicted open-reading frames, 141 of which are homologous to Syn9, and one tRNA gene. P-TIM40 contains all 38 T4-like core genes as well as the 25 additional T4-like cyanophage core genes as defined by Sullivan *et al.*, 2010. Auxiliary metabolic genes found in this phage include the *psbA, hli, PTOX, phoH, talC, CP12, mazG, hsp20* and tryptophan halogenase genes. *Prochlorococcus* NATL2A (600 ml) was infected at a multiplicity of infection of two in triplicate experiments and the transcriptional program of the P-TIM40 phage was investigated during the first 3 h of the 10 h latent period, with samples for RNA analysis taken at 0.5, 1, 2 and 3 h after phage addition.

Quantification of infective phages and phage and host genomic DNA
Infective phages in the extracellular medium were quantified using the plaque assay. Samples were filtered over 0.2 µm sterile syringe filters (Tuffryn HT), and serial dilutions of the phage-containing filtrate were plated on lawns of the host in pour-plates of 0.28% ultra-pure low melting point agarose (Invitrogen, Carlsbad, CA, USA) in artificial seawater-based medium, as described in Moore *et al.*, 2007. Extracellular phage DNA was quantified using qPCR for the *g20* gene (see Supplementary Text) on samples diluted 100-fold in 10 mM Tris (pH8).

Intracellular phage and *Synechococcus* DNA was quantified after collection of cells on 0.2 µm pore-sized polycarbonate filters (GE, Fairfield, CT, USA) by filtration (7–10 inch Hg). Cells on the filters were washed three times with sterile seawater and once with 3-ml preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl; pH8) and frozen at −80 °C. A heat lysis method was used to extract DNA from *Synechococcus* cells (Zinser *et al.*, 2006). In brief, the polycarbonate filter with *Synechococcus* cells was immersed in 10 mM Tris pH8 and agitated in a mini-bead beater for 2 min at 5000 rpm without beads. The sample was removed from the filter shards and heated at 95 °C for 15 min and used in triplicate qPCR reactions. The portal protein gene (*g20*) was amplified for Syn9 DNA and the *rbcL* gene was amplified for *Synechococcus* DNA using the TaqMan qPCR procedure. See Supplementary Text for details of the PCR procedures. Standard curves were produced using linear plasmid DNA containing the phage gene and genomic DNA for the host gene. See Supplementary Table S1 for all qPCR primer sequences and Universal Probe Library (Roche Diagnostics, Basel, Switzerland) probe identities used in this study.

RNA extraction
Cultures for RNA extraction (50–75 ml of each infected or control culture) were collected on 0.4-µm pore-sized polycarbonate filters (GE) by filtration (10 inch Hg) and snap frozen in liquid nitrogen in 2 ml PGTX buffer and stored at −80 °C. Total RNA was extracted using the PGTX 95 method (Pinto *et al.*, 2009) with minor modifications. Total nucleic acids were quantified based on absorption at 260 nm and RNA integrity was verified by gel electrophoresis. DNA was removed by DNase I digestion using the Turbo DNA-free kit (Ambion, Foster City, CA, USA). See Supplementary Text for further details of these procedures.

Microarray experimentation
An Affymetrix high-density custom-made tiling array was utilized for determining gene expression of the three *Synechococcus* hosts during infection by Syn9 and of the P-TIM40 phage during infection of *Prochlorococcus*. This array, SYN-PHG, contains perfect match probes on both strands for the three *Synechococcus* hosts at a density of every 35 nt and for the Syn9 and P-TIM40 phages at a density of every 17 nt (see Supplementary Text for more details on the microarray). Three biological replicates were analyzed for each time point after phage addition.

1440

Time points analyzed for Syn9 infection of the three *Synechococcus* strains were 2–5 min and 0.5, 1, 1.5, 2, 3, 4 h after phage addition, and were 0.5, 1, 2, 3 h after phage addition for P-TIM40 infection of *Prochlorococcus* NATL2A. Synthesis of cDNA, labeling, hybridization, staining and scanning were carried out according to Affymetrix protocols for Prokaryotic target RNA (GeneChip Expression Analysis Technical Manual, Prokaryotic; 702232, Rev.3) using 2 µg of total RNA, and were carried out at the Center for Genomic Technologies, Hebrew University of Jerusalem, Israel.

Data were analyzed using the probe signal intensities from the CEL files and the probe sets defined in the CDF library file using the R statistical language with Bioconductor add-on packages. The CEL file contains the signal intensities of all probes on the array for a particular sample, whereas the CDF library file defines the sets of probes that are positioned within each open-reading frame and is used for calculating expression summaries for each gene. Following previous analyses (Lindell *et al.*, 2007), background signal correction, normalization and summarization were performed by Robust Multi-array Averaging from the affy package. Differential expression between infected and uninfected controls was statistically evaluated by a linear model comparing the time-resolved signal intensities from the arrays with uninfected hosts to those from arrays with infected hosts (Limma package; Smyth, 2005). *P*-values were adjusted for multiple testing and converted to false-discovery rates (*q*-values) by the Benjamini–Hochberg approach. A detailed description of the microarray design, experimental execution and analysis is provided in the Supplementary Text.

RNA-seq experimentation
RNA-seq analysis was used for the investigation of the Syn9 transcriptome during infection of the three *Synechococcus* hosts. The cDNA libraries were prepared using Illumina's TruSeq RNA Sample Preparation Kit without the polyA isolation step. Whole-transcriptome libraries were prepared for two replicates of each infected host at four time points (5 min, 0.5, 1, 2 h after phage addition), and for the pooled triplicates of each uninfected host (control) at three time points (0.5, 1, 2 h after the beginning of the experiment). Complementary DNA libraries for bisulfite strand-specific sequencing were prepared as previously described (Edelheit *et al.*, 2013). Libraries were prepared for pooled replicates of the infected *Synechococcus* sp. WH8109 host at three time points (0.5, 1, 2 h) and for uninfected *Synechococcus* sp. WH8109 at the 0.5 h time point. The 5′-end sequencing libraries were prepared as previously described (Wurtzel *et al.*, 2012a, b), with some modifications, and are briefly described in the Supplementary Text. RNA-seq libraries were sequenced using the Illumina HiSeq platform at the Weizmann Institute.

A detailed description of the analyses of the RNA-sequencing results is provided in the Supplementary Text. In brief, the RNA-seq reads were aligned separately for each sample to the reference genomes, and analyzed according to the library type. Reads from the whole-transcriptome library were counted for each gene, normalized to host *rnpB* transcript levels and to gene length. The normalized data were used for clustering the phage genes by their expression profile and are shown in all figures unless otherwise stated. The RPKM normalization was not used for phage gene expression because when most of the genes change their expression at a given condition, the RPKM—which calculates the expression of each gene relative to the sum of all mapped reads—fails to represent the actual changes in expression; therefore, an external reference point was needed. Normalization with a house-keeping gene has been shown to perform well for RNA-seq read counts (Bullard *et al.*, 2010), and the *rnpB* gene was used for this purpose as the transcript levels of this gene were stable over the course of infection (Supplementary Figure S1), as has been shown previously during phage infection of *Prochlorococcus* (Lindell *et al.*, 2007). Transcript levels of this gene are stable across a variety of physiological conditions in cyanobacteria (for example see Mitschke *et al.*, 2011), and hence this gene is often used as an internal control for cyanobacterial expression studies (for examples see references Holtzendorff *et al.*, 2001; Muro-Pastor *et al.*, 2001; Stork *et al.*, 2005; Puerta-Fernández and Vioque, 2011). Reads from the bisulfite strand-specific libraries were used to estimate antisense transcripts (asRNA) length according to the coverage they produced for the phage genome. Reads from the 5′-end libraries were used for determining transcriptional start sites across the phage genome, and were analyzed as previously described (Wurtzel *et al.*, 2012a) with some minor modifications (Supplementary Text). Regions upstream of these transcriptional start sites were scanned for promoter motifs.

A summary of the RNA-sequencing data (number of reads, number of mapped reads and so on) is presented in Supplementary Tables S2–S4. The RNA-seq data are available at http://www.weizmann.ac.il/molgen/Sorek/syn9_browser/ under the 'Data files' tab. The transcriptome coverage data were normalized according to library size for each lane separately, to allow visual comparison between hosts at each time point.

Comparison of microarray and RNA-seq data
A comparison of the RNA-seq and microarray results, together with verification of gene profiles by qRT-PCR was carried out for the Syn9 infection experiments. This showed that the microarrays quickly became saturated for phage transcripts levels, whereas the RNA-seq captured the high dynamic range of the phage transcriptome more

accurately and without saturation (see Supplementary Text, Supplementary Figure S2). However, the microarrays better represented the host's transcriptional response, as the high proportion of phage-derived sequencing reads in the RNA-seq data set largely masked the readout of the host gene expression (see Supplementary Text,Supplementary Figure S2, Supplementary Tables S2–S4). We therefore used the RNA-seq platform to study the dynamics of the phage transcriptome, whereas the microarrays were used to infer relative host transcript levels after infection.

Clustering of phage gene expression
Cluster analyses for Syn9 were performed on normalized transcript levels of phage genes derived from the RNA-seq data. Input data consisted of transcript levels measured at four time points after infection from two independent replicates from each of the three hosts (resulting in 24 expression values per gene). To standardize across hosts prior to clustering, the transcript levels for each gene were adjusted to have a maximum expression equal to 1 and a minimum of 0 in each host independently. Hierarchical clustering was performed using Euclidian distance and average linkage as implemented in the R package stats. See Supplementary Text for Jaccard coefficient analysis of clustering. The dendrogram of Syn9 expression clusters was plotted using the dendrogram function (R package stats), with the gene names at the dendrogram tips colored using the labels_colors function (R package dendextend). The branches were colored using the dendrapply function with attribute 'edgePar' (R package stats). The ovals adjacent to the dendrogram tips were added manually.

Mass spectrometry identification of virion proteins
Phage virion particle proteins were identified following Sabehi et al., 2012. See Supplementary Text for details.

Data deposition in public databases
The genomes of *Synechococcus* sp. WH8109 and cyanophage P-TIM40 were sequenced as part of this study (see Supplementary Text) and were deposited in GenBank under the accession numbers CP006882 and KP211958, respectively.

The mRNA expression data were deposited in GEO under the SuperSeries accession number GSE74922: The microarray and RNA-sequencing data appear under the SubSeries accession numbers GSE63690 and GSE74921, respectively.

Results

*Syn9 phage infection and transcriptome dynamics were nearly identical in the three hosts*
We began by investigating the infection process and transcriptome dynamics of the Syn9 phage during infection of three distinct *Synechococcus* hosts: WH7803, WH8102 and WH8109. These strains belong to different clades and occupy different niches in the oceans (Table 1). Furthermore, 17–24% of their genomes contain strain-specific genes with only about two-thirds of their genes being common among all three hosts (Figure 1a). These strain-specific non-core genes are generally localized to genomic islands (Dufresne et al., 2008) (Supplementary Figure S3).

Despite the extensive genomic differences mentioned above, the progression of the Syn9 infection cycle was similar in all three *Synechococcus* hosts. The lytic cycle was 6–8 h long with a latent period of ~4 h prior to release of the first virions from the infected cells (Figure 1b). Furthermore, a similar yield of infective phage was produced in two of the three hosts, but fewer infective phage were released during infection of WH8109 (Figure 1b), presumably due to a lower ratio of infective to total phage production on this host relative to the other hosts (Supplementary Figure S4). Host genomic DNA was rapidly degraded in all three hosts with levels dropping to 20% of the initial (maximal) level by 1 h after infection (Figure 1c). By this time phage

Table 1 Host strains used in this study

	WH7803	*WH8102*	*WH8109*
Accession	NC_009481	NC_005070	CP_006882
Genome size (Mbp)	2.37	2.43	2.11
Protein-coding genes #	2533	2528	2661
%GC	60.2	59.4	60.1
Genome sequence reference	Dufresne et al., 2008	Palenik et al., 2003	This study
Clade[a]	V	III	II
Distribution of clade[a]	Mesotrophic upwelling regions, low abundance in a broad range of ecosystems.	Ultraoligotrophic open-ocean waters	Oligotrophic environments as well as coastal/continental shelf zones in subtropical/tropical regions.
Lifestyle strategy[a]	'coastal/opportunists'	'open ocean/specialists'	'open ocean/specialists'

[a]Clade is according to references (Rocap et al., 2002; Fuller et al., 2003), clade distribution is from references (Zwirglmaier et al., 2008; Scanlan et al., 2009) and lifestyle strategy is from reference (Dufresne et al., 2008).
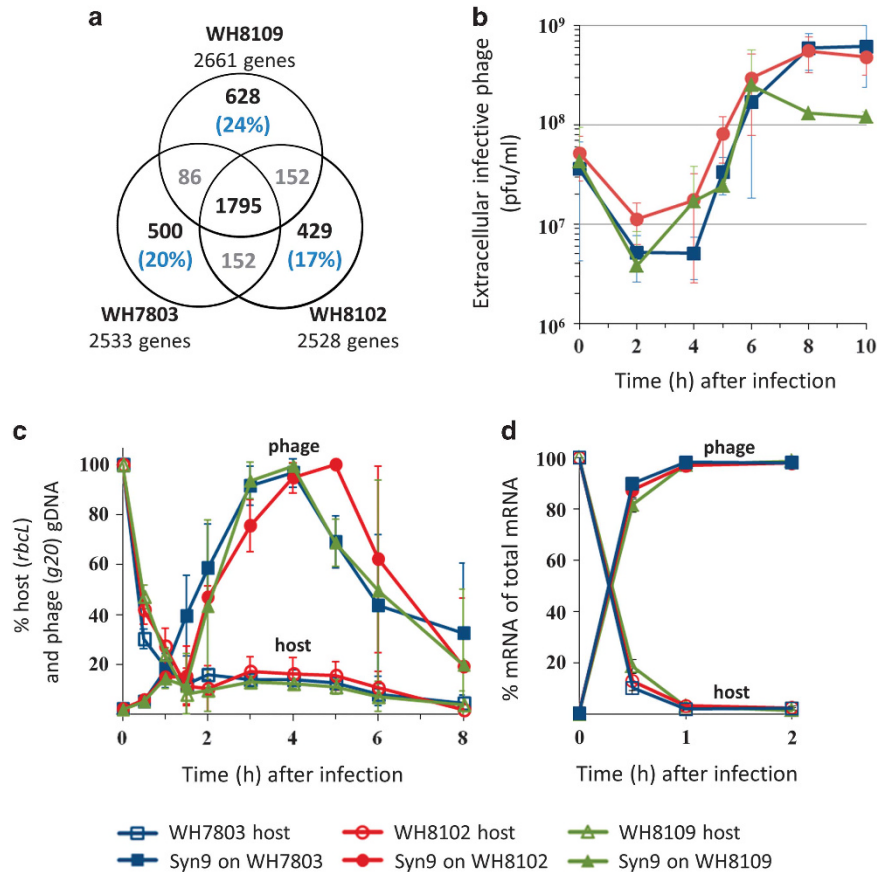
Figure 1 Infection dynamics of the Syn9 phage in three *Synechococcus* hosts. (**a**) Core and accessory genes in the genomes of the three *Synechococcus* hosts. (**b**) The length of the lytic cycle as a function of the number of infective phage released to the extracellular medium, as determined from the plaque assay. Average and s.d. of three biological replicates for WH7803 and WH8102, and two biological replicates for WH8109. (**c**) The timing of intracellular phage gDNA replication (closed symbols) and host gDNA degradation (open symbols) as determined by qPCR targeting the phage *g20* portal protein gene and the host *rbcL* Rubisco gene, respectively. Data are presented as percent of maximal gDNA for each gene. Average and s.d. of three biological replicates. (**d**) Ratios of phage and host mRNA with time after infection as determined from RNA-seq reads that mapped to the phage and host genomes, respectively. Average and range of two biological replicates.

genome replication had begun in all three hosts, although the timing of maximal phage genome replication occurred slightly later during infection of WH8102 (Figure 1c). Furthermore, transcription inside the cell was largely phage derived by 30 min after infection in all three hosts. An almost complete switch to phage transcription had occurred by 2 h after infection, with phage mRNA constituting ~98% of the total cellular mRNA (Figure 1d, Supplementary Tables S2–S4).

The temporal progression of gene expression was practically identical during infection of all three hosts, with largely the same gene cluster patterns found in three temporal expression classes (see below). A very small set of phage genes showed some variation in transcription patterns between hosts; however, none of these observed variations was statistically significant (Supplementary Figure S5). Strikingly, the correlation between transcript levels of individual phage genes when infecting each of the three hosts was extremely high (Figure 2). In fact, nearly identical expression patterns were

observed for almost all phage genes (for example, Figure 2a and b). This is even though expression levels spanned three orders of magnitude between the most highly expressed and the most lowly expressed phage genes (Figure 2c and d). Thus, the Syn9 phage followed a clearly defined and predetermined gene expression program in three temporal classes irrespective of the host infected.

The phage temporal expression program
Three temporal classes of early, middle and late genes were evident for Syn9 when clustering the phage genes according to their expression patterns (Figure 3, Supplementary Figure S6, Supplementary Table S5, Supplementary Text). Each class can be further divided into subclasses, displaying a variety of expression patterns (Figure 3). For example, the expression of some early genes declined 1 h following infection, whereas others remained highly expressed past 1 h (Figure 3, Supplementary Figure S7).
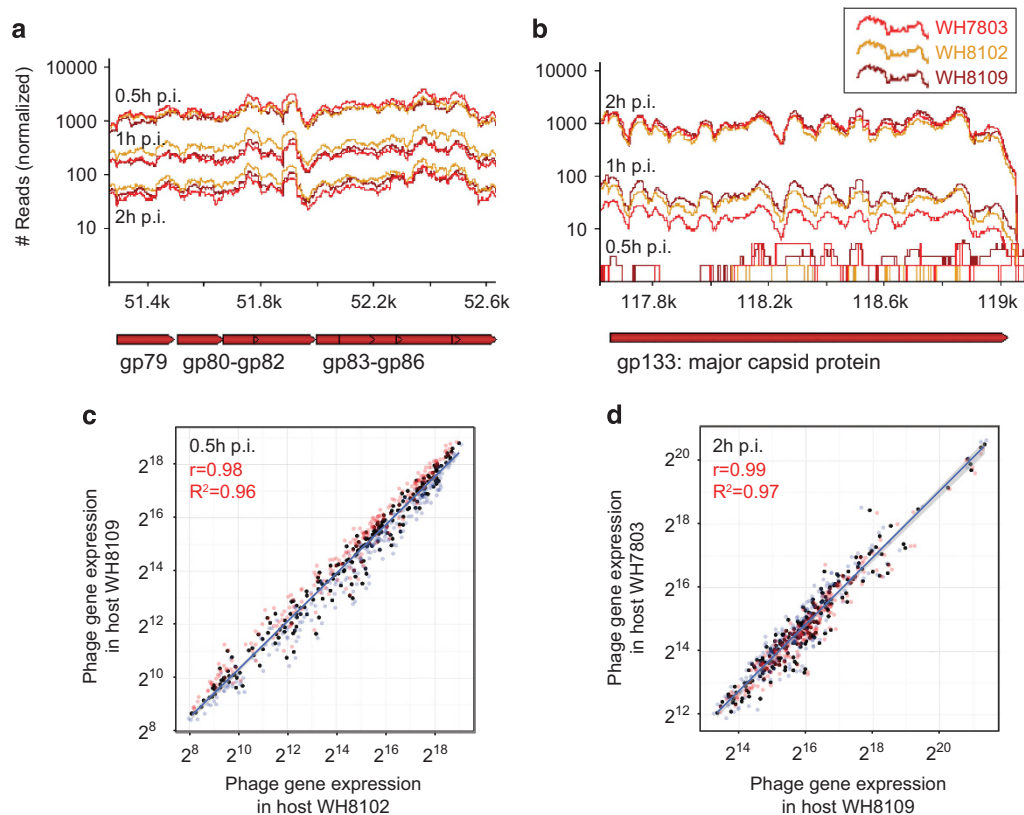
Figure 2 High correlation in phage gene expression across hosts. (**a**, **b**) Gene expression in representative regions of the phage genome showing early (**a**) and late (**b**) phage genes at three different time points post infection (p.i.). Phage gene expression during infection of the *Synechococcus* WH7803, WH8102 and WH8109 hosts is shown with red, orange and brown lines, respectively. *X* axis, position on the phage genome; phage gene products are represented by red block arrows below the *X* axis. The RNA-seq reads (*Y* axis) were normalized to library size. (**c**, **d**) Correlation of phage gene expression for representative hosts and time points: at 0.5 h after infection of *Synechococcus* WH8102 and WH8109 (**c**) and at 2 h after infection of *Synechococcus* WH8109 and WH7803 (**d**). Expression of phage genes averaged over two biological replicates is shown in black and biological replicates are shown in red and blue (normalized to host *rnpB* transcript levels). Pearson correlation coefficients (*r*) using linear values are shown in the respective panels with *P*-values <2.2e-16 for both correlations. Coefficients of determination (*R*^{2}) are also shown in each panel. Pearson correlation coefficients using logged values were 0.99 and 0.96 for (**c**, **d**), respectively. See Supplementary Material for correlations between all hosts and time points (Supplementary Figure S13) as well as for reproducibility between biological replicates (Supplementary Figure S14).

The early genes in Syn9 constitute 15% of the genome. The majority (29 out of 35) were already expressed 5 min after infection (Figure 4a). Most are concentrated in a single 8 kb window on the phage genome and are transcribed from seven operons spanning 30 open-reading frames (Figure 4b, Supplementary Figure S8). None of the early genes are homologous to early genes of T4. Most, however, are homologous to hypothetical proteins found in only a subset of other cyanophages (Weigele *et al.*, 2007; Sullivan *et al.*, 2010). Only three are homologous to genes in the databases that code for protein domains of known function. These are a helicase (*g75*); a gene with AAA and *cbbQ* type domains (*g153*, previously annotated as *cobS* or a putative porphyrin biosynthesis gene); and a gene with a nucleotidyl transferase domain (*g184*). Interestingly, *g184* is found in all T4-like cyanophage genomes, and remains highly expressed for at least 2 h after infection. We found that its protein is present in the virion particle (Supplementary Table S6) and is

probably carried in the capsid and may thus function throughout the infection process.

The middle genes consist of 106 genes, and include the DNA replication, nucleotide metabolism and recombination/repair genes identified by Weigele *et al.*, 2007 as homologs of T4 genes. These include the DNA polymerase, primase, helicase, ribonucleotide reductase and thymidylate synthase genes (Supplementary Table S5). The T4 gp55 sigma factor homolog required for gene expression from late promoters in T4 is also transcribed as a middle gene. The six Syn9 tRNA genes are transcribed as middle genes and would therefore be available for the translation of both middle and late proteins as previously hypothesized (Weigele *et al.*, 2007; Limor-Waisberg *et al.*, 2011; Enav *et al.*, 2012) (see Supplementary Text). In addition, a number of auxiliary metabolic genes coded in the Syn9 genome belong to the middle gene cluster (see below) as do five genes encoding virion particle proteins.
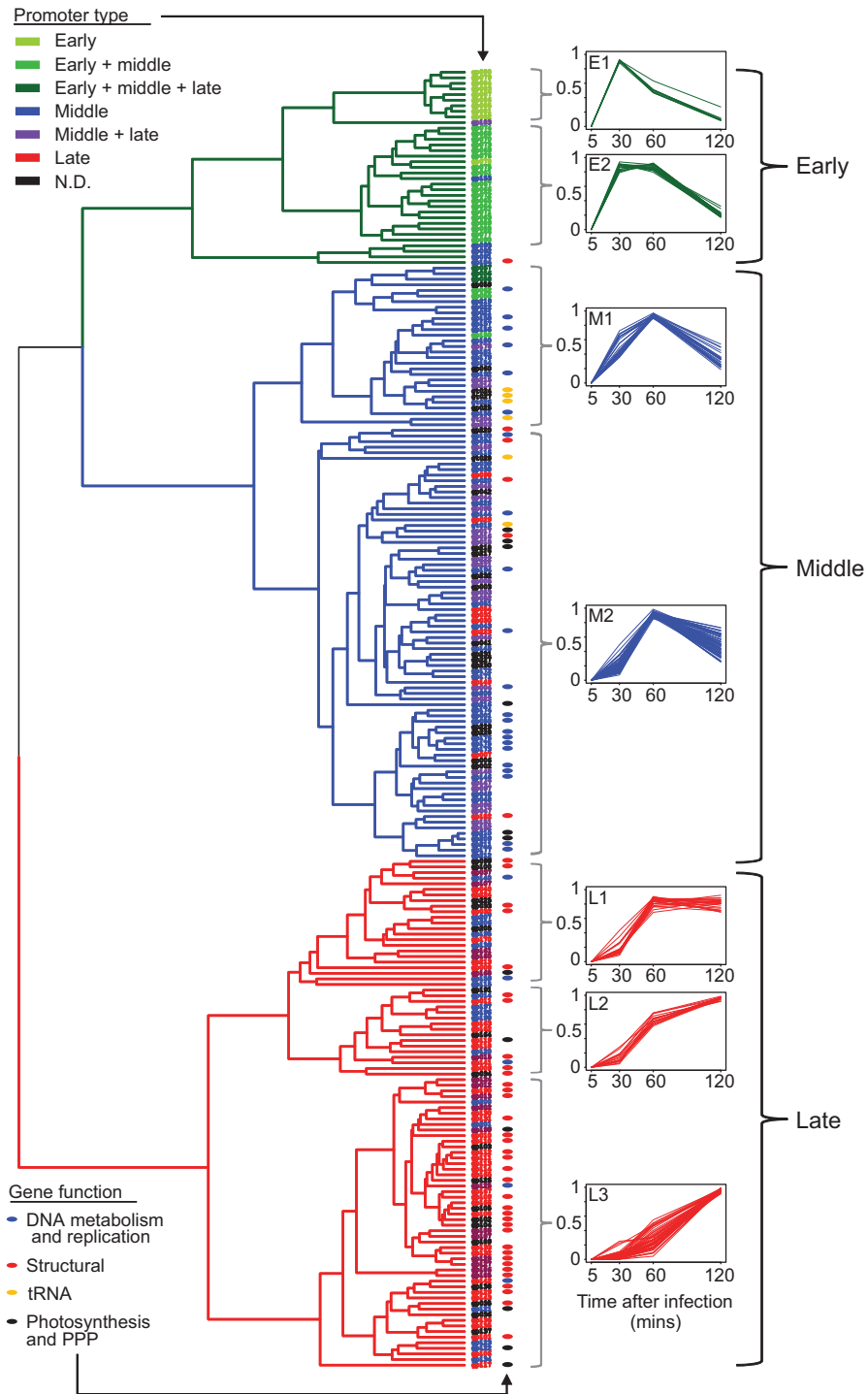
Figure 3 The temporal expression program of the Syn9 phage. Clustering of phage genes by their expression patterns is presented as a dendrogram with early, middle and late expression clusters shown in green, blue and red lines, respectively. Gene names at the dendrogram tips are colored according to the promoter class driving their expression (see legend, top left). The colors of the ovals adjacent to the genes denote the major classes of gene functions (see legend, bottom left). Graphs at the right of the subclusters show expression profiles of the individual genes in that subcluster as a function of time after infection. Subcluster designation at the top left corner of each graph is as in Supplementary Table S5. Relative transcript levels (*Y* axes of expression profile graphs) were normalized to minimum (0) and maximum (1) values for each gene and were averaged across all three hosts. A third early cluster with only four genes is presented in the Supplementary Material (Supplementary Results and Discussion; Supplementary Figure S7). The clustering and the construction of the dendrogram are explained in the Materials and methods under 'Clustering of phage gene expression'.
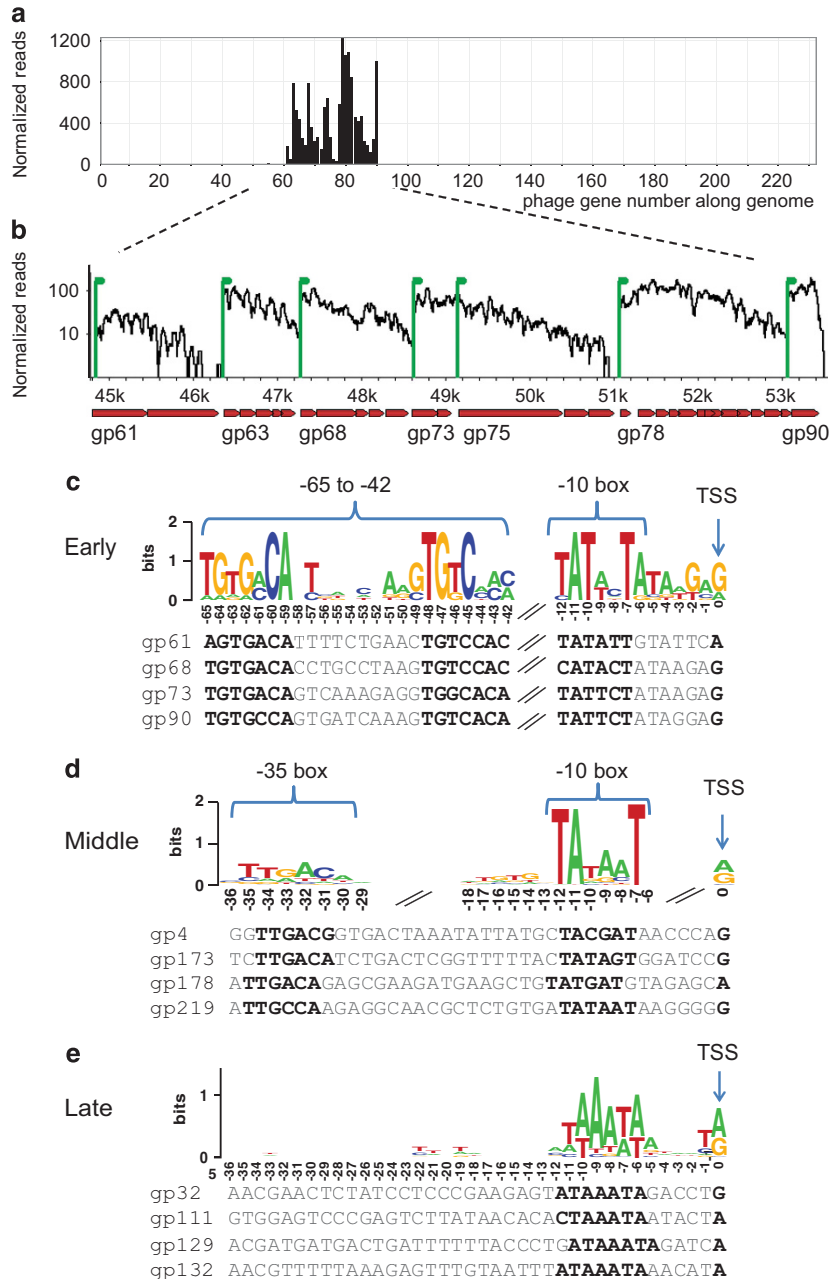
Figure 4 Syn9 early, middle and late promoters. (**a**) Expression of phage genes at 5 min after infection of the *Synechococcus* WH8102 host. (**b**) An expanded view of the region on the phage genome showing expression of the early-phage gene cluster at 5 min after infection. *X* and *Y* axes are as in Figure 2a. Transcription patterns suggest that the genes are organized in several consecutive operons. Green vertical lines represent the transcription start sites (TSS) of these operons, experimentally determined from genome-wide 5′ end RNA-seq. (**c–e**) Promoter logos of the sequences upstream of the TSSs of the phage early (**c**), middle (**d**) and late (**e**) genes. The numbers below the motifs represent the position upstream of the TSS. For (**c**) and (**d**), sequences were aligned by the prominent motifs, separately for each region of the motif. For (**e**), sequences were aligned relative to the TSS. Representative promoter sequences from each class are shown below each consensus motif. Promoter logos were generated from 9 early promoters (**c**), 70 middle promoters (**d**) and 59 late promoters (**e**).

The late-gene cluster consists of 91 genes (Supplementary Table S6) of which at least 32 are structural genes based on mass spectrometry we performed to identify the proteins making up the virion particle (Supplementary Table S6 and see Weigele *et al.*, 2007). These include the capsid, tail and putative tail fiber proteins (Weigele *et al.*, 2007) as well as numerous proteins of unknown function. Other late genes include the DNA-packaging terminase genes, as well as the *regA* translational repressor gene, a few DNA replication and metabolism genes and some auxiliary metabolic genes (see below).

Phage asRNAs were identified from high-throughput transcriptional start site mapping.

Thirty-one asRNAs were reproducibly expressed during infection of all three hosts (Supplementary Table S7). Many were transcribed from middle promoters, whereas a few were transcribed from late promoters. Strand-specific bisulfite sequencing revealed that some asRNAs overlap the 5′-untranslated region of phage genes, whereas others span multiple genes (Supplementary Figure S9). The lack of a predicted ORF suggests that many of the asRNA transcripts have a regulatory function (see Supplementary Text). An asRNA that may be involved in regulating gene expression was first reported in the T4-like cyanophage, S-PM2 (Millard et al., 2010).

The (cyano)bacterial-like auxiliary metabolic genes coded by Syn9 were transcribed in the middle- and late-temporal expression classes. Most of the cyanobacterial-like photosynthesis genes were transcribed in the middle cluster with the DNA replication and metabolism genes (Supplementary Table S5; Figure 3). These include the photosystem II core reaction center genes, *psbA* and *psbD,* two photosynthesis-related *hli* stress-response genes, and the phycoerythrin biosynthesis gene, *cpeT.* Two of these genes (*psbA* and *hli*) are known to be transcribed with DNA replication and metabolism genes in other cyanophages (Lindell et al., 2005; Clokie et al., 2006; Lindell et al., 2007) and are thought to be important for continued photosynthetic activity and the production of ATP and reducing equivalents needed for phage DNA replication and nucleotide biosynthesis (Mann et al., 2003; Lindell et al., 2005; Clokie et al., 2006; Lindell et al., 2007). In addition, two photosynthetic electron transfer genes (*petE* and *PTOX*) clustered with the late genes, with maximal transcript levels detected at 2 h after infection (see Supplementary Text).

A single bacteria-like carbon metabolism gene, the CP12 Calvin cycle inhibitor gene, had maximal transcription 1 h after infection and clustered with the middle genes. This gene is thought to direct carbon flux to the pentose phosphate pathway (PPP) during infection (Thompson et al., 2011b), which, in turn, has been hypothesized to enhance the production of ribose substrates and reducing equivalents for nucleotide biosynthesis during infection (Lindell et al., 2007; Thompson et al., 2011b). In contrast to CP12, the three pentose phosphate pathway genes (*gnd, zwf* and *talC*) clustered with the late genes, different to findings reported in Thompson et al., 2011b (see Supplementary Text). Additional bacterial-like genes, such as the *phoH*-family stress–response gene, the *mazG* potential regulatory gene and the tryptophan halogenase genes, were transcribed with middle genes. However, their function in cyanophages remains unclear.

### Novel regulation of the transcriptional program
To understand how the phage expression program is regulated, we determined the promoters of all phage

transcripts after mapping transcriptional start sites at a single-base resolution. A complete temporal promoter map for all Syn9 phage transcripts can be visualized using the interactive 'Syn9 transcriptome browser' we devised (http://www.weizmann.ac.il/molgen/Sorek/syn9_browser/). Grouping of promoters according to the time of appearance of their associated transcriptional start sites revealed distinct motif signatures that are likely to be responsible for directing the expression of the early, middle and late genes (Figure 4c and e). A very prominent element was found upstream of the early transcripts, composed of two palindromic 7-bp motifs separated by 10 bp, which are found 29–30 bp upstream of a −10 Pribnow box (with a consensus sequence of TGTGACA-N_{10}-TGTCACA-N_{30}-TATACT) (Figure 4c). These results stand in contrast with current knowledge on phage early transcription, because the expression from early-phage promoters is usually regulated by the host core transcriptional machinery, and hence phage early promoters are expected to resemble host σ^{70} promoters (Miller et al., 2003b; Nolan et al., 2006) (see Discussion). In contrast, the promoters of Syn9 middle genes, rather than the early genes, were characterized by the signature of the canonical σ^{70} recognition site, suggesting that it is these genes that are transcribed using the core host transcriptional machinery (Figure 4d). Finally, late promoters show a distinct motif very similar to the T4 late motif (Figure 4e; Miller et al., 2003b; Nolan et al., 2006). Overall we defined nine early promoters, 70 middle promoters and 59 late promoters upstream of phage genes. Approximately 25% of the promoters do not consist of any of these three signals. In these promoters, the motif might be too diverged from the canonical signal to be identified, or, a few of these promoters may be false positives inherent to our method.

Some phage genes presented complex regulation patterns, which can explain the variety of expression profiles observed (Figure 3). Many of these genes were activated by multiple alternative promoters of different temporal classes (Figure 3, Figure 5 and Supplementary Figure S10). For example, some of the early genes have both an early and a middle promoter, which explains how they retain their high expression at 1 h, dropping only at 2 h after infection (Figure 3). Other genes, such as *g107* were transcribed from both middle and late promoters (Figure 5a). Alternative promoters also led to modular operon structures (Figure 5b) and, when within open-reading frames, to multiple protein variants (Figure 5c) (see Supplementary Text).

### Other T4-like cyanophages display a program similar to Syn9
The expression of the Syn9 genome in three transcriptional clusters is different to expectations based on previous hypotheses for cyanophages (Mann et al., 2005; Clokie et al., 2006), and the
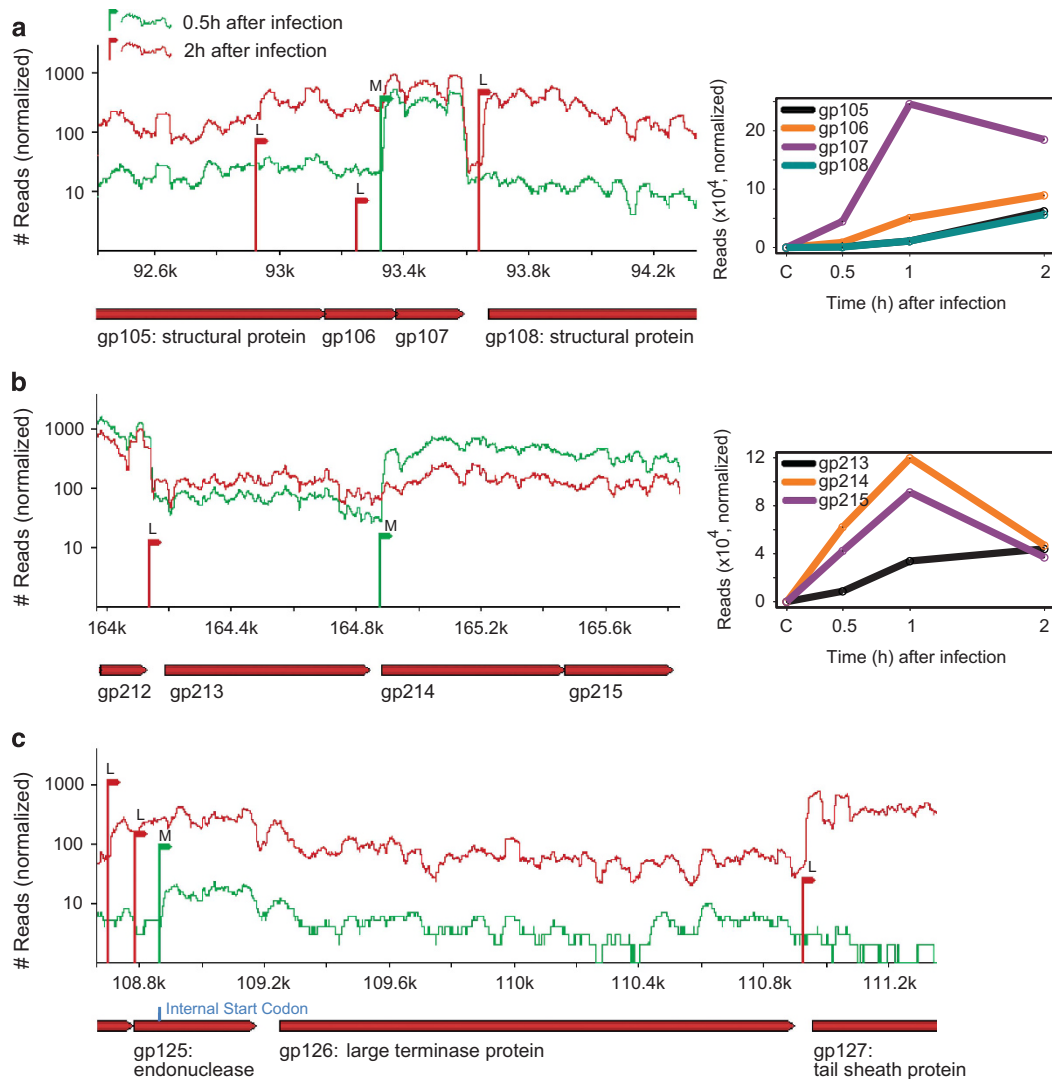
Figure 5 Alternative Syn9 promoters generate complex temporal transcription patterns. Phage gene expression and promoter positions at 0.5 h (green) and 2 h (red) after infection of the *Synechococcus* WH8109 host. *X* and *Y* axes are as in Figure 2a. The letters E, M and L on the TSS vertical lines indicate whether the associated promoter is an early, middle or late promoter, respectively. (**a**) Example of an alternative promoter: a TSS driven by a middle promoter upstream of gene *g107* is replaced by an alternative TSS driven by a late promoter at 2 h after infection. Normalized expression profiles of the genes are shown at the right. (**b**) An example of modular operon structure: *g214* and *g215* are transcribed from a middle promoter at 0.5 h after infection, and from an upstream late promoter at 2 h after infection as a part of the *g213* operon. (**c**) An example of a modular gene: a middle promoter is inside the coding region of *g125*, potentially giving rise to a shorter protein. The blue line indicates the matching internal start codon. A TSS driven by a late promoter appears at 2 h after infection, upstream of the full length *g125* ORF.

regulatory elements that are likely to drive them are unlike those known for T4. Therefore, we wondered whether our findings are specific for Syn9 or are also found among other T4-like cyanophages. To begin addressing this question we analyzed the transcriptional program of P-TIM40, another T4-like cyanophage, during infection of *Prochlorococcus* NATL2A (see Materials and methods and the Supplementary Text for information on this phage). We found that, similar to Syn9, the P-TIM40 genome is transcribed in three expression clusters (Supplementary Figure S11). Also similar to Syn9, the early-gene cluster was made up of short genes of unknown function that were primarily clustered in one region of the genome (Figure 6a). Furthermore, the same distinctive early

promoter motif identified in Syn9 was observed upstream of these early P-TIM40 genes (Figure 6a, Supplementary Figure S12). The middle- and late-gene expression clusters consisted of classical replication and morphogenesis genes, respectively (Supplementary Table S8). In addition, the *psbA, hli, PTOX, talC, phoH* and *mazG* auxiliary metabolic genes were transcribed with middle genes, whereas the *CP12* and tryptophan halogenase genes were transcribed with late genes. The putative promoter sequences upstream of representative replication and morphogenesis genes revealed the host-like σ^{70} motif upstream of middle genes and the classical late motif upstream of late genes, as found for Syn9 (Figure 6c and d).
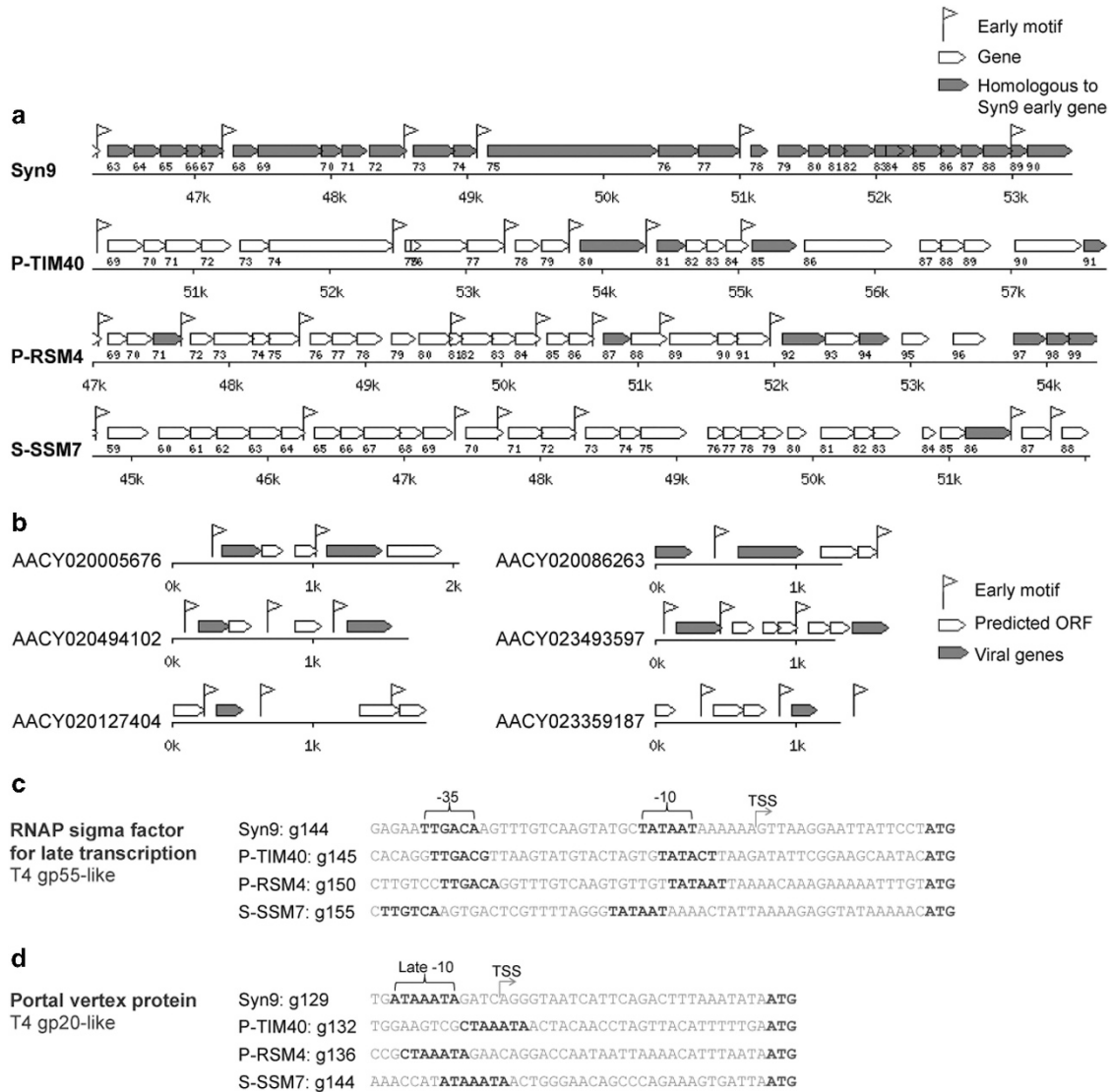
Figure 6 Conservation of the three classes of Syn9 promoter motifs across T4-like cyanophages and environmental samples. (**a**) Examples of regions from sequenced cyanophage genomes showing the Syn9 early promoter motif. The P-TIM40 region consists of P-TIM40 early genes based on expression data analysis (see Supplementary Table S8). Phage genes are represented by block arrows; the genes that are homologous to Syn9 early genes are marked in grey. Locus numbers appear below gene names, as do positions on the phage genome. Phage names are shown on the left. The early promoter motif of each phage is displayed in Supplementary Figure S12. (**b**) Examples of contigs from the GOS project (Rusch *et al.*, 2007) where the early promoter motif was detected (white flag). Genes with a significant BLAST hit to a viral gene are in grey (see Methods). ORFs (block arrows) were predicted using Glimmer (Salzberg *et al.*, 1998). (**c**) Promoter sequences of a classical middle gene, the RNA polymerase sigma factor gene (T4 gp55-like), show the host-like σ^{70} promoter motif in Syn9 and in other representative T4-like cyanophages. This motif was found upstream of additional classical middle genes in the T4-like cyanophages, including the DNA polymerase (T4 gp43-like), DNA primase subunit (T4 gp61-like) and ssDNA-binding protein (T4 gp32-like) genes. (**d**) Promoter sequences of a classical late gene, the portal vertex protein gene (T4 gp20-like), show the late promoter motif in Syn9 and in other representative T4-like cyanophages. This motif was found upstream of additional classical late genes in the T4-like cyanophages, including the tail sheath protein (T4 gp18-like), baseplate hub subunit (T4 gp26-like) and head completion protein (T4 gp4-like) genes. For (**c**, **d**), promoter motifs (bold black type towards the left and middle of the sequences) and the TSS of the Syn9 genes (arrow) are shown. Sequences were aligned relative to the start codon of the gene (bold black type at the right end of the sequences).

Next, we assessed how common these early, middle- and late-promoter motifs are in the genomes of 15 other cyanophages that infect *Prochlorococcus* and *Synechococcus* and represent the diversity of known T4-like marine cyanophages (Sullivan *et al.*, 2010). Multiple copies of the early motif were found in 14 out of the 15 phages (Supplementary Figure S12).

In the vast majority of cases (90%) this promoter motif was located upstream of operon-like clusters of short uncharacterized genes that are localized to one main region of the genome (Figure 6a). In the S-PM2 phage, however, the operons of these putative early genes and their associated early promoter motifs were more

dispersed in the genome relative to the other T4-like cyanophages investigated. Many of these operons contained genes homologous to Syn9 early genes (Figure 6a). These 15 cyanophage genomes also have host-like σ^{70} motif upstream of classical replication genes (Figure 6c) and the late promoter motif upstream of morphogenesis genes (Figure 6d).

We then wanted to determine whether the information-rich early promoter motif could be detected in natural populations of T4-like cyanophages. Indeed, this motif was found on 3035 contigs from the Global Ocean Sampling (GOS) expedition (Rusch et al., 2007). Of these, 924 (30%) contained at least one identifiable phage gene (Figure 6b), a 5.85-fold enrichment ($P < 2.2e-16$, Fisher's exact test) as compared with a random set of similarly sized contigs from the GOS data set. This suggests that these are bonafide phage contigs and that the genes downstream are early-phage genes. It should be noted that no traces of this early motif were detected in cyanobacterial genomes, nor in any other bacterial genome found in Genbank. Thus, the prevalence of this early motif in other T4-like cyanophage genomes and in the GOS data set implies that many T4-like cyanophages in the oceans undergo similar transcriptional and regulatory programs as those found for Syn9 and P-TIM40.

Host-specific transcriptional responses preferentially localize to genomic islands

Although phage gene expression was nearly identical irrespective of the host infected, the transcriptional response of the three hosts showed marked differences (see below). Nonetheless, similar general transcription patterns were observed. An immediate response was observed within 5 min of infection in all hosts, apparent from a transient increase in transcript levels of a small number of genes (Figure 7, Table 2, Supplementary Table S9) and a concomitant decline in transcript levels of another small group of genes (Table 2, Supplementary Table S10). Subsequently, within 0.5–1 h of phage infection, a marked decline of the vast majority of host genes ($> 90\%$) was observed. This massive decline was accompanied by a subsequent increase in transcript levels of some genes (Figure 7, Table 2, Supplementary Table S11), whereas other genes remained unchanged for most of the latent period (Table 2, Supplementary Table S12). Here we define host–response genes as those whose transcript levels underwent an immediate change as well as those that increased or remained unchanged subsequent to this immediate response. These latter two groups of genes seem to have been shielded from the rapid mRNA degradation or transcriptional downregulation, which was the fate of the vast majority of host genes.

Large disparities in the identities of host-response genes were found among the hosts. Most strikingly, of the 270 response genes combined across the three
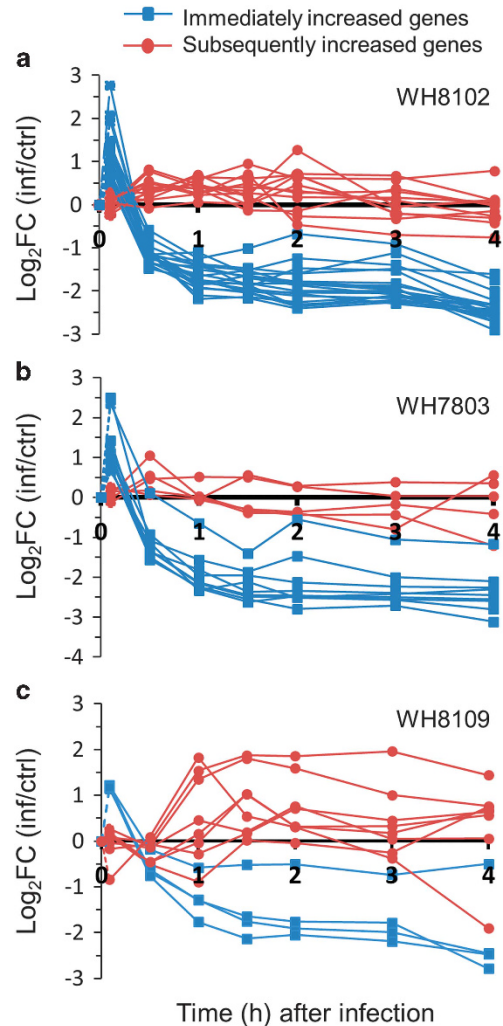
Figure 7 Temporal transcriptional profiles of host genes with an increase in transcript levels following Syn9 infection. (**a**) *Synechococcus* WH8102, (**b**) *Synechococcus* WH7803, (**c**) *Synechococcus* WH8109. Transcript levels, determined from microarrays, are presented as $\log_2$ fold-change (FC) in infected cultures (inf) relative to uninfected cultures (ctrl) over the 4 h latent period of infection. Only genes whose expression levels were significant at a false-discovery rate of $q < 0.05$ are shown. Immediately and subsequently increased transcript levels are shown with blue and red lines, respectively (see Supplementary Tables S9 and S11). Genes with unchanged and immediately declined transcript levels, as well as the majority of genes that subsequently declined, are not shown. Dashed lines indicate the change in transcript levels in the uninfected control cultures prior to infection relative to the infected cultures at 5 min after infection. Average of three biological replicates.

hosts, only two of them were common to all three hosts. These were the carboxysome carbon-fixation gene, *ccmK2* (*csoS1*), which functions in concentrating CO_2 (Supplementary Table S9); and a small heat shock-like chaperone (Supplementary Table S12). The majority of other response genes were host-specific. For example, transcript levels of three additional carbon-fixation genes and two high-light inducible (*hli*) stress-response genes increased only in *Synechococcus* WH8102. Nonetheless, the host-

Table 2 Summary of the number of protein-coding host-response genes and their genomic location[a]

	WH7803 Total in genome	In genomic islands	WH8102 Total in genome	In genomic islands	WH8109 Total in genome	In genomic islands
Immediately increased	9	1 (11%) (NS)	20	5 (25%) (NS)	4	0 (NS)
Immediately declined	2	0 (NS)	41	15 (37%) $P < 0.01$	3	1 (33%) (NS)
Subsequently increased	5	3 (60%) $P < 0.01$	12	11 (92%) $P < 0.0001$	8	2 (25%) (NS)
Unchanged	65	24 (37%) $P < 0.0001$	57	35 (61%) $P < 0.0001$	44	19 (43%) $P < 0.0001$
Response genes (sum of above)	81	28 (35%) $P < 0.0001$	130	66 (51%) $P < 0.0001$	59	22 (37%) $P < 0.0001$
Subsequently declined	2452	ND	2396	ND	2602	ND
Total	2533	438 (17%)	2528	541 (21%)	2661	476 (18%)

Abbreviations: ND, not determined; NS, not significant. Fisher's exact test was used to determine the significance of overrepresentation of these genes in genomic islands. Subsequently declined genes were not included in this analysis as they constitute most of the genome.
[a]Genomic islands in WH7803 and WH8102 were defined as per Dufresne et al., 2008 and the genomic islands in WH8109 were defined in this study (see Supplementary Text).

specific response genes often belonged to the same general functional groups of cell envelope, DNA repair, carbon fixation, respiration, translation (tRNA genes) and nutrient utilization.

An example of genes belonging to a similar functional group yet responding differentially to Syn9 infection is found for nitrogen transport and assimilation genes. In *Synechococcus* WH8102, for which the most marked immediate response was observed (Figure 7, Table 2), transcript levels of 33 nitrogen-related genes declined immediately after infection (Supplementary Table S10). Many belong to the NtcA regulon (Su et al., 2006), including *ntcA* itself, which is the major nitrogen transcriptional regulator in cyanobacteria (Herrero et al., 2001). In contrast to WH8102, transcript levels of a few nitrogen-related genes (*nirA, narB, amt1*) increased significantly in WH7803 immediately after infection, two of which are homologs of genes that declined in WH8102.

Some potential defense-related genes were also among the *Synechococcus* WH8102-response genes. Two of these genes, SYNW0071 and SYNW1946, encode a single-stranded RNA nuclease PIN domain often found in the toxic components of toxin–antitoxin operons in prokaryotes (Anantharaman and Aravind, 2003). The latter gene also encodes a PhoH domain, which is common in marine phages, including cyanophages such as Syn9 (Weigele et al., 2007; Sullivan et al., 2010; Goldsmith et al., 2011). A third gene, SYNW1659, is a conserved hypothetical protein that consists of a domain of unknown function (DUF3387, PF11867), that is associated with type I and type III restriction enzymes (see PF04851 and PF04313).

A number of other genes of unknown function with conserved domains were among the most highly expressed response genes. These included four genes whose transcript levels increased immediately after infection in WH8102 that have a common domain (DUF4278) limited to cyanobacteria and cyanophages. In addition, transcript levels of four homologous genes with the DUF1651 domain (three in WH8102 and one in WH8109) increased subsequently or remained unchanged in response to infection by Syn9. These too appear to be limited to unicellular marine cyanobacteria. Although the function of these two sets of genes remains unknown, they may represent a unique type of cyanobacterial-specific response to phage infection.

Analysis of the genomic localization of the response genes in the three hosts revealed that a large number of them (between 35–51%) are located in hypervariable genomic islands (Table 2). This is a significant overrepresentation with a 2–2.5-fold enrichment for each host ($P < 0.0001$, Fisher's exact test). This was most pronounced for genes with transcript levels that were subsequently increased or remained unchanged during the latent period, especially in *Synechococcus* WH8102 (Table 2). These results imply an important role for *Synechococcus* genomic islands in mediating the host response to phage infection (see Discussion).

Discussion

This study is the first to explore the transcriptional program of a broad host-range phage during infection of different hosts and the response of more than one host to the same phage. Our findings indicate that the ability of a single T4-like cyanophage to infect multiple, genetically diverse host types is not due to specialization of the infection process for the different hosts. Rather, if capable of infecting a particular host, the phage follows a strict, predetermined expression program and is indifferent to the considerable differences in host genome content. Conversely, each of the hosts mounted a gene-specific response to infection even though they were infected by the same

phage. These findings suggest that successful infection of multiple hosts by broad host-range phages, at least within the T4-like phages, is more likely to be dependent on the effectiveness of host defense mechanisms in preventing infection than on the differential tailoring of the infection process by the phage for individual hosts.

Previous work suggested that T4-like cyanophages employ a two-class temporal transcriptional program, including early and late, but not middle, gene expression. This was initially based on bioinformatic analyses reporting the absence of T4-like middle promoter motifs from the S-PM2 and Syn9 genomes, as well as the lack of *motA* and *asiA* genes that are required for middle transcription in T4 (Mann *et al.*, 2005; Weigele *et al.*, 2007). Furthermore, host-like promoters, used for early expression in T4, were identified upstream of classical middle genes, leading to the conclusion that only early- and late-transcription occur in these cyanophages and that, different to T4, DNA replication genes are transcribed as early genes (Mann *et al.*, 2005; Clokie *et al.*, 2006). Subsequent studies that examined transcription of a small subset of S-PM2 and Syn9 genes also suggested that only two temporal expression classes were present in cyanophages (Clokie *et al.*, 2006; Thompson *et al.*, 2011b). However, neither of these studies examined the transcriptional profiles of any of the early genes identified in this present study.

In contrast to the conclusions reached in the above studies, our genome-wide transcriptional investigations of Syn9 and P-TIM40 clearly indicate that three temporal transcriptional classes are used by T4-like cyanophages and that they have three distinct promoter types. They further indicate that, similar to T4, classical DNA replication genes belong to the class of middle gene expression. However, these expression classes are regulated from promoter types different to those known for T4. This appears to be a common phenomenon across T4-like cyanophages, as seen from the presence of Syn9 promoter motifs upstream of similar gene types in diverse T4-like cyanophages: the newly identified prominent early motif is present upstream of clusters of small genes of unknown function; the host-like σ^{70} middle motif is upstream of DNA replication genes and the T4-like late motif is upstream of phage morphogenesis (structural) genes. Thus, our study provides the first detailed description of a cyanomyovirus transcriptional and regulatory program, and revealed a previously unidentified regulatory strategy that deviates from the T4 paradigm. Furthermore, this regulatory program appears to be ancestral to the T4-like marine cyanophage lineage. Moreover, the mechanism of regulation known so well for T4 may actually be the exception rather than the rule among the broad group of T4-like phages (see Discussion in Supplementary Text).

Although we have identified the promoter motifs likely to be directing the three transcriptional classes in T4-like cyanophages, the mechanisms underlying their temporal mode of regulation remain unknown. Particularly intriguing is the finding of the novel early promoter that controls early-gene expression in T4-like cyanophages. This early promoter motif has no resemblance to sequences in cyanobacterial genomes. Therefore, expression from this promoter element may well be enabled by a phage factor packaged in the capsid, especially as expression from this promoter is already observed 5 min after infection. Such factors are known to exist in other phages, for example, the Alt protein of phage T4 (Horvitz, 1974a, b; Koch *et al.*, 1995). What then prevents transcription from the host-like middle promoters during the early phase of infection? One possibility is that phage proteins entering the cell together with phage DNA modify the affinity of the host RNA polymerase for the different sequence motifs; increasing affinity for the early motif while reducing affinity for the middle host-like σ^{70} motif. Alternatively, access of the host RNA polymerase to phage middle promoters may be blocked by different phage proteins bound to the incoming DNA. Such proteins entering into the bacterial cell with the phage DNA are currently unknown for cyanophages, although some candidate proteins have been identified from our mass spectrometry analyses (Supplementary Table S6).

Despite the near-identical transcriptional program employed by the Syn9 phage, each host displayed a largely strain-specific response with genes that are preferentially localized to genomic islands. These included genes from the functional groups of cell envelope, DNA repair, carbon fixation, respiration and nutrient utilization. This may constitute a general response to stress, and these functional groups may be indicative of the kind of stresses experienced during infection. It is also possible that some of these host responses reflect an attempt at defense against infection, as inferred from the identity of some of the response genes. Alternatively, these genes may be induced by the phage to aid in the replication process, with the immediate response induced by phage proteins entering into the cell together with phage DNA. In this scenario, the factors involved would need to be different to those having a role in early-phage gene expression as early-phage promoter motifs are not found upstream of these (or any) cyanobacterial genes. The subsequent response may also be induced by a phage factor, but this would need to be one that is expressed relatively early in the transcriptional program. Rather than transcriptional induction, it is also feasible that the subsequent host-response is due to differential transcript stabilization, mediated by host or phage proteins, as drastic and rapid degradation of host DNA occurred within an hour of infection (Figure 1c). Indeed, both the *E. coli* T4 and T7 phages are known to induce changes in host transcript stability (Marchand *et al.*, 2001; Ueno and Yonesaki, 2004) and duplex RNA has been

proposed to stabilize host transcripts during infection of the marine cyanobacterium *Prochlorococcus* MED4 by the podovirus P-SSP7 (Stazic *et al.*, 2011). Rapid degradation of host DNA is also known to occur during T4 infection (Mosig and Eiserling, 2006).

Until about a decade ago, it was generally thought that phage infection led to a complete shut-down of host transcription (Koerner and Snustad, 1979; Roucourt and Lavigne, 2009). However, the employment of whole-genome transcriptional analyses to assess host responses to infection has shown this not to be the case, with a range of responses in different host-phage systems (Poranen *et al.*, 2006; Lindell *et al.*, 2007; Ravantti *et al.*, 2008; Lavigne *et al.*, 2013, this study). First of all, common across systems is an immediate transient response followed by a subsequent response to infection. Yet the degree to which transcript levels increase, both with respect to the number of genes and their fold change, differ in the different systems. Furthermore, some but not all bacteria experience a massive decline in host gene expression upon infection. However, our findings can be generalized across marine *Synechococcus* (this study) and *Prochlorococcus* (Lindell *et al.*, 2007), with both transient and subsequent increases in transcript levels accompanying an overall decline in host transcript abundance. Furthermore, not only did these same patterns occur in the two different host genera, but also in response to infection by phages belonging to two very different phage families, both T4-like (this study) and T7-like (Lindell *et al.*, 2007) cyanophages. Whether this will also be the situation in response to infection by TIM5-like myoviruses and the siphoviruses in these cyanobacteria remains to be seen.

A most striking commonality in the response of the three *Synechococcus* hosts (this study) as well as that of *Prochlorococcus* (Lindell *et al.*, 2007), was the preferential localization of the response genes in genomic islands. Indeed, responses of marine cyanobacteria to stressful conditions such as nutrient deprivation, high light, oxidative stress and metal toxicity have been documented to emanate from genes located in genomic islands (Coleman *et al.*, 2006; Thompson *et al.*, 2011a; Stuart *et al.*, 2013). Furthermore, as found here, some of the most highly expressed genes in response to these environmental stressors are genes of unknown function located in genomic islands (Martiny *et al.*, 2006; Tolonen *et al.*, 2006; Thompson *et al.*, 2011a). Thus, despite the apparent transient nature of genes in genomic islands (on evolutionary time scales) and the unknown function of many of these genes, they appear to be of great importance for the acclimation and adaptation of cyanobacteria to environmental change.

Numerous lines of evidence indicate that genomic islands are part of the co-evolutionary process between marine cyanobacteria and their phages. First, phages are considered one of the major means by which genes are transferred into genomic islands (Palenik *et al.*, 2003, 2006; Coleman *et al.*, 2006). Second, a number of host-responsive island genes have homologs in phage genomes, including *hli* and *phoH*-family genes, DNA repair and heat shock genes as well as genes of unknown function like the conserved DUF4278 domain-carrying genes ((Lindell *et al.*, 2007); this study). Thus, expression of island genes during infection may have led to their utilization by phages over evolutionary time and their stable incorporation into phage genomes (Lindell *et al.*, 2007). Third, phages serve as an important selective pressure influencing diversity and content of cell-surface genes in genomic islands (Avrani *et al.*, 2011; Avrani and Lindell, 2015), the products of which are utilized by phage for attachment and entry into the host. Combined, these findings highlight the central role of hypervariable genomic islands in cyanobacterial host-phage interactions, serving as a central axis for host response and defense against phage infection, as well as a hotspot for dynamic gene exchange that continually reshapes the genomes of both host and phage.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Anantharaman V, Aravind L. (2003). New connections in the prokaryotic toxin-antitoxin network: relationship

with the eukaryotic nonsense-mediated RNA decay system. *Genome Biol* **4**: R81.

Angly FE, Felts B, Breitbart M, Salamon P, Edwards RA, Carlson C *et al.* (2006). The marine viromes of four oceanic regions. *PLoS Biol* **4**: e368.

Avrani S, Lindell D. (2015). Convergent evolution toward an improved growth rate and a reduced resistance range in *Prochlorococcus* strains resistant to phage. *Proc Natl Acad Sci USA* **112**: E2191–E2200.

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

Bench SR, Hanson TE, Williamson KE, Ghosh D, Radosovich M, Wang K *et al.* (2007). Metagenomic characterization of Chesapeake Bay virioplankton. *Appl Environ Microbiol* **73**: 7629–7641.

Breitbart M. (2012). Marine viruses: truth or dare. *Ann Rev Mar Sci* **4**: 425–448.

Bullard JH, Purdom E, Hansen KD, Dudoit S. (2010). Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**: 94.

Clokie MR, Mann NH. (2006). Marine cyanophages and light. *Environ Microbiol* **8**: 2074–2082.

Clokie MR, Shan J, Bailey S, Jia Y, Krisch HM, West S *et al.* (2006). Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* **8**: 827–835.

Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, Delong EF *et al.* (2006). Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science* **311**: 1768–1770.

Comeau AM, Chan AM, Suttle CA. (2006). Genetic richness of vibriophages isolated in a coastal environment. *Environ Microbiol* **8**: 1164–1176.

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

Desplats C, Dez C, Tétart F, Eleaume H, Krisch HM. (2002). Snapshot of the genome of the pseudo-T-even bacteriophage RB49. *J Bacteriol* **184**: 2789–2804.

Dufresne A, Ostrowski M, Scanlan DJ, Garczarek L, Mazard S, Palenik BP *et al.* (2008). Unraveling the genomic mosaic of a ubiquitous genus of marine cyanobacteria. *Genome Biol* **9**: R90.

Edelheit S, Schwartz S, Mumbach MR, Wurtzel O, Sorek R. (2013). Transcriptome-wide mapping of 5-methyl-cytidine RNA modifications in bacteria, archaea, and yeast reveals m5C within archaeal mRNAs. *PLoS Genet* **9**: e1003602.

Elena SF, Agudelo-Romero P, Lalić J. (2009). The evolution of viruses in multi-host fitness landscapes. *Open Virol J* **3**: 1–6.

Enav H, Béjà O, Mandel-Gutfreund Y. (2012). Cyanophage tRNAs may have a role in cross-infectivity of oceanic *Prochlorococcus* and *Synechococcus* hosts. *ISME J* **6**: 619–628.

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

Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, Scanlan DJ. (2003). Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. *Appl Environ Microbiol* **69**: 2430–2443.

Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA *et al.* (2011). Development of phoH as a novel signature gene for assessing marine phage diversity. *Appl Environ Microbiol* **77**: 7730–7739.

Guttman B, Raya R, Kutter E. (2005) Basic phage biology. In: Kutter E, Sulakvelidze A (eds). *Bacteriophages Biology and Application*. CRC Press: Boca Raton, FL, USA, pp 29–66.

Herrero A, Muro-Pastor AM, Flores E. (2001). Nitrogen control in cyanobacteria. *J Bacteriol* **183**: 411–425.

Holtzendorff J, Partensky F, Jacquet S, Bruyant F, Marie D, Garczarek L *et al.* (2001). Diel expression of cell cycle-related genes in synchronized cultures of *Prochlorococcus* sp. strain PCC 9511. *J Bacteriol* **183**: 915–920.

Horvitz HR. (1974a). Bacteriophage T4 mutants deficient in alteration and modification of the *Escherichia coli* RNA polymerase. *J Mol Biol* **90**: 739–750.

Horvitz HR. (1974b). Control by bacteriophage T4 of two sequential phosphorylations of the alpha subunit of *Escherichia coli* RNA polymerase. *J Mol Biol* **90**: 727–738.

Jabbar MA, Snyder L. (1984). Genetic and physiological studies of an *Escherichia coli* locus that restricts polynucleotide kinase- and RNA ligase-deficient mutants of bacteriophage T4. *J Virol* **51**: 522–529.

Koch T, Raudonikiene A, Wilkens K, Rüger W. (1995). Overexpression, purification, and characterization of the ADP-ribosyltransferase (gpAlt) of bacteriophage T4: ADP-ribosylation of *E. coli* RNA polymerase modulates T4 'early' transcription. *Gene Expr* **4**: 253–264.

Koerner JF, Snustad DP. (1979). Shutoff of host macromolecular synthesis after T-even bacteriophage infection. *Microbiol Rev* **43**: 199–223.

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

Lavigne R, Lecoutere E, Wagemans J, Cenens W, Aertsen A, Schoofs L *et al.* (2013). A multifaceted study of *Pseudomonas aeruginosa* shutdown by virulent podovirus LUZ19. *MBio* **4**: e00061–13.

Limor-Waisberg K, Carmi A, Scherz A, Pilpel Y, Furman I. (2011). Specialization versus adaptation: two strategies employed by cyanophages to enhance their translation efficiencies. *Nucleic Acids Res* **39**: 6016–6028.

Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T *et al.* (2007). Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.

Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. (2005). Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.

Lindell D, Padan E, Post AF. (1998). Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* **180**: 1878–1886.

Mann NH, Clokie MR, Millard AD, Cook A, Wilson WH, Wheatley PJ *et al.* (2005). The genome of S-PM2, a 'photosynthetic' T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* **187**: 3188–3200.

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

Marchand I, Nicholson AW, Dreyfus M. (2001). Bacteriophage T7 protein kinase phosphorylates RNase E and stabilizes mRNAs synthesized by T7 RNA polymerase. *Mol Microbiol* **42**: 767–776.

Marston MF, Pierciey FJ, Shepard A, Gearin G, Qi J, Yandava C et al. (2012). Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci USA* **109**: 4544–4549.

Marston MF, Sallee JL. (2003). Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. *Appl Environ Microbiol* **69**: 4639–4647.

Martiny AC, Coleman ML, Chisholm SW. (2006). Phosphate acquisition genes in *Prochlorococcus* ecotypes: evidence for genome-wide adaptation. *Proc Natl Acad Sci USA* **103**: 12552–12557.

Millard AD, Gierga G, Clokie MR, Evans DJ, Hess WR, Scanlan DJ. (2010). An antisense RNA in a lytic cyanophage links *psbA* to a gene encoding a homing endonuclease. *ISME J* **4**: 1121–1135.

Millard AD, Mann NH. (2006). A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. *J Mar Biol Assoc UK* **86**: 507–515.

Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ. (2009). 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* **11**: 2370–2387.

Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A et al. (2003a). Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* **185**: 5220–5233.

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

Mitschke J, Georg J, Scholz I, Sharma C, Dienst D, Bantscheff J et al. (2011). An experimentally anchored map of transcriptional start sites in the model cyanobacterium *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci USA* **108**: 1–6.

Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D et al. (2007). Culturing the marine cyanobacterium. *Prochlorococcus Limnol Ocean Methods* **5**: 353–362.

Mosig G, Eiserling F. (2006) T4 and related phages: structure and development. In: Calendar R (ed). *The Bacteriophages*. Oxford University Press: New York, NY, USA, pp 225–267.

Mosig G, Yu S, Myung H, Haggård-Ljungquist E, Davenport L, Carlson K et al. (1997). A novel mechanism of virus-virus interactions: bacteriophage P2 Tin protein inhibits phage T4 DNA synthesis by poisoning the T4 single-stranded DNA binding protein, gp32. *Virology* **230**: 72–81.

Muro-Pastor MI, Reyes JC, Florencio FJ. (2001). Cyanobacteria perceive nitrogen status by sensing intracellular 2-oxoglutarate levels. *J Biol Chem* **276**: 38320–38328.

Nolan JM, Petrov V, Bertrand C, Krisch HM, Karam JD. (2006). Genetic diversity among five T4-like bacteriophages. *Virol J* **3**: 30.

Palenik BP, Brahamsha B, Larimer FW, Land M, Hauser L, Chain P et al. (2003). The genome of a motile marine. *Synechococcus Nature* **424**: 1037–1042.

Palenik BP, Ren Q, Dupont CL, Myers GS, Heidelberg JF, Badger JH et al. (2006). Genome sequence of *Synechococcus* CC9311: Insights into adaptation to a coastal environment. *Proc Natl Acad Sci USA* **103**: 13555–13559.

Pinto FL, Thapper A, Sontheim W, Lindblad P. (2009). Analysis of current and alternative phenol based RNA extraction methodologies for cyanobacteria. *BMC Mol Biol* **10**: 79–87.

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

Puerta-Fernández E, Vioque A. (2011). Hfq is required for optimal nitrate assimilation in the Cyanobacterium *Anabaena* sp. strain PCC 7120. *J Bacteriol* **193**: 3546–3555.

Pène C, Uzan M. (2000). The bacteriophage T4 anti-sigma factor AsiA is not necessary for the inhibition of early promoters in vivo. *Mol Microbiol* **35**: 1180–1191.

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

Riede I, Eschbach ML. (1986). Evidence that TraT interacts with OmpA of *Escherichia coli*. *FEBS Lett* **205**: 241–245.

Rifat D, Wright NT, Varney KM, Weber DJ, Black LW. (2008). Restriction endonuclease inhibitor IPI* of bacteriophage T4: a novel structure for a dedicated target. *J Mol Biol* **375**: 720–734.

Rocap G, Distel DL, Waterbury JB, Chisholm SW. (2002). Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. *Appl Environ Microbiol* **68**: 1180–1191.

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

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S et al. (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5**: e77.

Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. (2012). A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proc Natl Acad Sci USA* **109**: 2037–2042.

Salzberg SL, Delcher AL, Kasif S, White O. (1998). Microbial gene identification using interpolated Markov models. *Nucleic Acids Res* **26**: 544–548.

Scanlan DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR et al. (2009). Ecological genomics of marine picocyanobacteria. *Microbiol Mol Biol Rev* **73**: 249–299.

Scanlan DJ. (2003). Physiological diversity and niche adaptation in marine *Synechococcus*. *Adv Microb Physiol* **47**: 1–64.

Smyth GK. (2005). Limma: linear models for microarray data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S (eds). *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*. Statistics for Biology and Health, Springer-Verlag: New York, NY, USA, pp 397–420.

Staley JT, Gunsalus RP, Lory S, Perry JJ. (2007) Viruses. In: *Microbial Life*. Sinauer Associates, Inc.: Sunderland, MA, USA, pp 389–408.

Stazic D, Lindell D, Steglich C. (2011). Antisense RNA protects mRNA from RNase E degradation by RNA-RNA duplex formation during phage infection. *Nucleic Acids Res* **39**: 4890–4899.

Stern A, Sorek R. (2011). The phage-host arms race: shaping the evolution of microbes. *Bioessays* **33**: 43–51.

Stork T, Michel K-P, Pistorius EK, Dietz K-J. (2005). Bioinformatic analysis of the genomes of the cyanobacteria *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942 for the presence of peroxiredoxins and their transcript regulation under stress. *J Exp Bot* **56**: 3193–3206.

Stuart RK, Brahamsha B, Busby K, Palenik B. (2013). Genomic island genes in a coastal marine *Synechococcus* strain confer enhanced tolerance to copper and oxidative stress. *ISME J* **7**: 1139–1149.

Su Z, Mao F, Dam P, Wu H, Olman V, Paulsen IT *et al.* (2006). Computational inference and experimental validation of the nitrogen assimilation regulatory network in cyanobacterium *Synechococcus* sp. WH 8102. *Nucleic Acids Res* **34**: 1050–1065.

Sullivan MB, Coleman ML, Weigele PR, Rohwer F, Chisholm SW. (2005). Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* **3**: e144.

Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR *et al.* (2010). Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* **12**: 3035–3056.

Sullivan MB, Waterbury JB, Chisholm SW. (2003). Cyanophages infecting the oceanic cyanobacterium. *Prochlorococcus Nature* **424**: 1047–1051.

Suttle CA, Chan AM. (1993). Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser* **92**: 99–109.

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

Tetu SG, Johnson DA, Varkey D, Phillippy K, Stuart RK, Dupont CL *et al.* (2013). Impact of DNA damaging agents on genome-wide transcriptional profiles in two marine *Synechococcus* species. *Front Microbiol* **4**: 232.

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

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

Tolonen AC, Aach J, Lindell D, Johnson ZI, Rector T, Steen R *et al.* (2006). Global gene expression of *Prochlorococcus* ecotypes in response to changes in nitrogen availability. *Mol Syst Biol* **2**: 53.

Ueno H, Yonesaki T. (2004). Phage-induced change in the stability of mRNAs. *Virology* **329**: 134–141.

Uzan M. (2009). RNA processing and decay in bacteriophage T4. *Prog Mol Biol Transl Sci* **85**: 43–89.

Wang K, Chen F. (2008). Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ Microbiol* **10**: 300–312.

Waterbury JB, Valois FW. (1993). Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol* **59**: 3393–3399.

Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF *et al.* (2007). Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* **9**: 1675–1695.

Woolhouse ME, Taylor LH, Haydon DT. (2001). Population biology of multihost pathogens. *Science* **292**: 1109–1112.

Wurtzel O, Sesto N, Mellin JR, Karunker I, Edelheit S, Becavin C *et al.* (2012a). Comparative transcriptomics of pathogenic and non-pathogenic *Listeria* species. *Mol Syst Biol* **8**: 583.

Wurtzel O, Yoder-Himes DR, Han K, Dandekar AA, Edelheit S, Greenberg EP *et al.* (2012b). The single-nucleotide resolution transcriptome of *Pseudomonas aeruginosa* grown in body temperature. *PLoS Pathog* **8**: e1002945.

Wyman M, Gregory RP, Carr NG. (1985). Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**: 818–820.

Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ *et al.* (2006). *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol* **72**: 723–732.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D *et al.* (2008). Global phylogeography of marine *Synechococcus* and *Prochlorococcus* reveals a distinct partitioning of lineages among oceanic biomes. *Environ Microbiol* **10**: 147–161.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)