The ISME Journal (2017) 11, 284–295

www.nature.com/ismej

OPEN

ORIGINAL ARTICLE

Regulation of infection efficiency in a globally abundant marine *Bacteriodetes* virus

Cristina Howard-Varona[1,6], Simon Roux[2,6], Hugo Dore[3], Natalie E Solonenko[2,6], Karin Holmfeldt[4], Lye M Markillie[5], Galya Orr[5] and Matthew B Sullivan[1,2,6,7]

[1]*Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ, USA;* [2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA;* [3]*Département de biologie, ENS Lyon, Lyon, France;* [4]*School of Natural Sciences, Linnaeus University, Kalmar, Sweden and* [5]*Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory (PNNL), Richland, WA, USA*

Bacteria impact humans, industry and nature, but do so under viral constraints. Problematically, knowledge of viral infection efficiencies and outcomes derives from few model systems that over-represent efficient lytic infections and under-represent virus–host natural diversity. Here we sought to understand infection efficiency regulation in an emerging environmental *Bacteroidetes*–virus model system with markedly different outcomes on two genetically and physiologically nearly identical host strains. For this, we quantified bacterial virus (phage) and host DNA, transcripts and phage particles throughout both infections. While phage transcriptomes were similar, transcriptional differences between hosts suggested host-derived regulation of infection efficiency. Specifically, the alternative host overexpressed DNA degradation genes and underexpressed translation genes, which seemingly targeted phage DNA particle production, as experiments revealed they were both significantly delayed (by >30 min) and reduced (by >50%) in the inefficient infection. This suggests phage failure to repress early alternative host expression and stress response allowed the host to respond against infection by delaying phage DNA replication and protein translation. Given that this phage type is ubiquitous and abundant in the global oceans and that variable viral infection efficiencies are central to dynamic ecosystems, these data provide a critically needed foundation for understanding and modeling viral infections in nature.

The ISME Journal (2017) **11,** 284–295; doi:10.1038/ismej.2016.81; published online 17 May 2016

Introduction

Bacteria are well recognized to impact human health and disease (Heintz and Mair, 2014), industrial processes (Luo *et al.*, 2015) and natural ecosystems (Falkowski *et al.*, 2008). More recently, viruses are also gaining relevance as they modulate these processes through cell lysis, horizontal gene transfer and metabolic reprogramming during infection (Fuhrman, 1999; Wommack and Colwell, 2000; Hurwitz *et al.*, 2013; Salmond and Fineran, 2015). Although most bacteria are thought to be infected by viruses (phages) (Weinbauer, 2004; Hyman and Abedon, 2012), and the extent of viral impacts depends on the efficiency (e.g., fraction of infected/lysed cells, latent period) and outcome (e.g., lysis, lysogeny) of phage–host interactions, these have largely been studied only in a few phyla, for example, Actinobacteria, Firmicutes and γ-Proteobacteria, that under-represent environmental diversity (Holmfeldt *et al.*, 2013). Further, even though lytic phages have variable efficiencies depending on host and environmental factors (You *et al.*, 2002; Wang, 2006), infection data derives from common approaches, for example, plaque assays and spot tests, that favor investigating *efficient* (e.g., large fraction of infected/lysed cells and short latent periods) phage–host interactions (Dang and Sullivan, 2014).

In contrast, the lack of model systems to study *inefficient* infections (e.g., those with lower fraction of infected/lysed cells and long latent periods) is problematic, as natural ecosystems are dynamic and environmental conditions impact host diversity, density and/or physiology, as well as the efficiency of phage–host interactions (You *et al.*, 2002; Abedon *et al.*, 2003; Shao and Wang, 2008; Stocker, 2012; Mojica and Brussaard, 2014; Storms *et al.*, 2014; Zeglin, 2015). To date, work on infection efficiencies is mostly theoretical (Abedon *et al.*, 2001; You *et al.*, 2002;

Correspondence: MB Sullivan Sullivan, Department of Microbiology and Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH 43210, USA.
E-mail: mbsulli@gmail.com
[6]Current address: Department of Microbiology, The Ohio State University, Columbus, OH, USA.
[7]Current address: Department of Civil, Environmental and Geodetic Engineering, The Ohio State University, Columbus, OH, USA.
Received 14 January 2016; revised 29 March 2016; accepted 3 April 2016; published online 17 May 2016

Bragg and Chisholm, 2008) and there is a need for new phage—host model systems, particularly those displaying inefficient infections and that are relevant in nature.

One such system is podovirus ϕ38:1 infecting *Cellulophaga baltica* (Holmfeldt *et al*., 2014; Dang *et al*., 2015) of the *Bacteroidetes* phylum. This host phylum is abundant in the human gut and the oceans (Kirchman, 2002; Gomez-Pereira *et al*., 2010), and the phage is the only cultured representative for one of the four globally abundant virus types in the oceans (Roux *et al*., 2015a). Podovirus ϕ38:1 is also well characterized as one of a collection of *Cellulophaga baltica* phages that have been examined using genomics, structural proteomics, quantitative host-range assays and infection dynamics (Holmfeldt *et al*., 2014; Dang *et al*., 2015). Such analyses revealed that ϕ38:1 has a broad host range as it infects 13 of the 21 isolated *C. baltica* strains, contrasting other phages from the same collection, which have a narrower host range given that they can only infect 1–2 strains (Holmfeldt *et al*., 2014). Two of these 13 infected host strains, NN016038 and no. 18, are nearly identical genetically, as they display identical 16S-rRNA gene sequences (Holmfeldt *et al*., 2007) and share 93% of their genes with average nucleotide identity of >99.99% (this study). As well, such strains are physiologically nearly identical, as they have the same growth, size and morphology (Dang *et al*., 2015). Despite the large similarities between the host strains, ϕ38:1 infection differs markedly as follows. Infection of the original host used for isolation (strain NN016038) is fast and efficient, as ~60% of the cells are infected and lysed within 70 min (Dang *et al*., 2015). Contrastingly, infection of alternative host strain no. 18 under identical experimental conditions (i.e., same phage stock and titre, growth and infection parameters) is inefficient given its reduced adsorption, fewer (~30%) infected cells and a much longer latent period and cell lysis (~11 h latent period by plaque assay or 150 min by phageFISH) (Dang *et al*., 2015). Beyond the inefficient adsorption, the mechanisms driving intracellular inefficiency of infection, if any, are unknown.

Here we complemented these prior investigations to characterize the intracellular infection dynamics of variably efficient phage–host interactions by temporally quantifying DNA (via quantitiatve PCR (qPCR)), RNA (via RNA-sequencing (RNA-seq)) and phage particles (via electron microscopy (EM)) throughout ϕ38:1 infection of each *C. baltica* host. These findings advance understanding of how viral infection efficiency varies across nearly identical bacterial hosts and are foundational for unveiling mechanisms of phage–host interactions in nature.

Materials and methods

Data availability

Reads from RNA-seq data are available at the Sequence Read Archive under the study number SRP066570. All protocols can be found at https://u.osu.edu/viruslab/protocols/. Scripts are in https://bitbucket.org/MAVERICLab/rnaseq-cba38-1-phage-host-time-series as well as in http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/iVirus/Cellulophaga_Transcriptomics, where additionally the Supplementary Data set S1 can be found. The latter contains all statistical and functional analyses for the RNA-seq data, phage and bacterial genome analyses and the qPCR and transmission electron microscopy analyses.

Phage–host manipulations

Growth and infections of *C. baltica* host strains NN016038 (original) and no. 18 (alternative) with ϕ38:1 were conducted as described previously (Holmfeldt *et al*., 2014; Dang *et al*., 2015). Briefly, cells were grown in Marine Luria Bertani medium to a density of 10^8 cells per ml in early logarithmic stage. They were then infected with ϕ38:1 in 50 ml at a multiplicity of infection of 1.1 and 6.1 for N016038 and no. 18, respectively, which maximized the fraction of infected cells (75% and 38% for the original and alternative host infections, respectively) as determined previously (Dang *et al*., 2015). Such differences were not problematic or likely to drive the differences in infection efficiency, as a previously determined wide span of multiplicity of infections (e.g., 0.1, 3, 6) resulted in similar infection dynamics (Holmfeldt *et al*., 2014; Dang *et al*., 2015). Included were three infections and three no-phage controls where Marine Sodium Magnesium buffer was added instead of phage. After 15 min of phage–host adsorption time, replicates were diluted 10-fold in Marine Luria Bertani and time 0 ('0 min') was established. Phage and cell abundance (via viable cell plating) were enumerated as described previously (Holmfeldt *et al*., 2014).

Whole-genome comparison of both hosts

Genomes from *C. baltica* NN016038 and no. 18 were compared through the online tool http://enve-omics.ce.gatech.edu/ani/http://enve-omics.ce.gatech.edu/ani/ to obtain their average nucleotide identity. Genomes were further compared with Blastp (threshold of 50 on bit score, 0.001 on *e*-value and 99% of amino-acid identity of BLAST hit). For each genome, a protein not detected in the other was considered as 'unique' to this host. Genes unique to each host were surveyed for the presence of phage defense genetic features, including CRISPR genes, restriction modification, abortive infection mechanisms and toxin/antitoxin genes (Samson *et al*., 2013), as well as bacteriophage exclusion genes (Goldfarb *et al*., 2015) (Supplementary Dataset S1, tab *GenomicDifferencesHosts* and Supplementary Table S1).

Quantitative PCR
Samples for qPCR were collected (1 ml) for the original host (at 0, 15, 30, 45, 60, 75 and 90 min) and the alternative host (at 0, 15, 30, 45, 60, 120 and 150 min), centrifuged at 20 817 *g* for 5 min, the supernatant removed and pellets were stored at −80 °C until processing. Phage gene *gp021*, which was previously used for phage detection throughout infection (Holmfeldt *et al*., 2014), and bacterial housekeeping genes *M667_14370* (*C. baltica* N016038) and *M666_05390* (*C. baltica* no. 18), which have high, constant expression over time (i.e., *edgeR* false discovery rate >0.05; Supplementary Dataset S1, tabs *RNASeq_original_host* and *RNASeq_alternative_host*), were chosen as targets to be amplified by qPCR using the QuantiTect SYBR Green PCR Kit (Qiagen, Valencia, CA, USA; cat. no. 204145). Results are found in the Supplementary Data set S1 tab *qPCR*.

Transmission electron microscopy
Samples (10 ml) were collected throughout each infection at 0, 45, 60 and 75 min (original host, *C. baltica* NN016038 infection) and at 0, 90, 120, 150 and 240 min (alternative host, *C. baltica* no. 18 infection), centrifuged for 10 min at 10 000 *g* to remove the supernatant and then the pellet was fixed onto grids (200 mesh copper grids with carbon-stabilized formvar support; Ted Pella, Redding, CA, USA), which were analyzed as described previously (Brum *et al*., 2005), to determine the frequency of visibly infected cells using a transmission electron microscope (CM12, Philips, Eindhoven, The Netherlands) to count the phage particles within cells. Complete protocols are available at http://u.osu.edu/viruslab/protocols/#TransmissionElectronMicroscopy and raw data in the Supplementary Data set S1, tab TEM.

RNA extractions
Samples for RNA were collected (1–2 ml) from minutes 0, 15, 30, 45, 60 and 120 (original host) and from minutes 0, 20, 40, 60 and 120 (alternative host), and spun down at 20 817 *g* for 5 min. The supernatant was then removed and tubes were flash-frozen in liquid nitrogen and stored at −80 °C until the extraction. RNA extractions were performed with the RNeasy Mini Kit (Qiagen; cat. no. 74104), DNase treated with TURBO DNA-free (Ambion, Carlsbad, CA, USA; AM1907) and concentrated to 20 µl with the RNA Clean and Concentrator-5 Kit (Zymo Research Corporation, Irvine, CA, USA; R1015). The RNA integrity number and concentration of the extracted RNA was determined via the Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA, USA). For all samples, nomenclature includes the time followed by infection (I) or control (C) and the replicate (R) number. For example, biological replicate 1 of the infection at 15 min is '15 min I R1'.

RNA-seq, read mapping to host and phage genomes and coverage calculations
Libraries of complementary DNA were prepared for SOLiD 5500 XL (Applied Biosystems, Foster City, CA, USA) sequencing according to the manufacturer's protocol (Life Technologies, Carlsbad, CA, USA). All sequencing data is available at the Sequence Read Archive under study number SRP066570. The 50 bp reads were mapped to the genomes of the corresponding host, *C. baltica* NN016038 (GenBank no. CP009887.1) or *C. baltica* #18 (GenBank # CP009976), and phage ϕ38:1 (GenBank no. NC_021796.1) using Bowtie2 v.2.14. Coverage of the phage and host genomes was calculated as the number of base pairs mapped to a genome divided by the genome length (Supplementary Tables S2).

Phage genome reorganization and annotation
The publicly available phage genome (GenBank no. NC_021796.1) was reordered to accommodate the transcriptional pattern, that is, by making the first gene represented coincide with the first gene transcribed and can be found at http://mirrors.iplantcollaborative.org/browse/iplant/home/shared/iVirus/Cellulophaga_Transcriptomics. New predicted functional annotations were incorporated by comparing unknown gene products to publicly available protein sequences via Blastx, Blastn, BlastP, InterProScan and NCBI Conserved Domain and overlapping results were selected as an approximation of the protein's function, which was designated as 'putative'.

Statistical analyses of the phage and host transcriptomes
Scripts for the following analyses are provided at https://bitbucket.org/MAVERICLab/rnaseq-cba38-1-phage-host-time-series.

Values for $\log_2$RPKM (reads per kilobase of transcript per million mapped reads) and results from differential expression (DE) analyses for all phage and both host genes can be found in the Supplementary Dataset S1, tabs *RNASeq_original_host, RNASeq_phage_on_OriginalHost*, *RNA Seq_alternative_host, RNASeq_phage_on_AlternativeHost*, *RNASeq OriginalHostCategories*, *RNASeqAlternativeHost Categories*, *RNASeqOriginalHostPhageDefense* and *RNA SeqAlternHostPhageDefense*.

Normalization of the read counts, calculation of RPKM and heat map representation. Read counts of phage and host were normalized separately using the R software package *edge*R in Bioconductor (Robinson *et al*., 2010). The resulting matrices were used for calculating the RPKM values (Mortazavi

et al., 2008) for all genes and for DE analyses (see below).

Sample clustering. After normalization and calculation of $\log_2$RPKM, host-infected and control samples were hierarchically clustered via Pearson's correlation using the pvclust package (Suzuki and Shimodaira, 2006) in R. Node support was calculated through multiscale resampling (10 000 bootstraps; Supplementary Figure S3). From the original host infection, samples '0 min I R3', '15 min I R3', '30 min I R3' and '120 min C R3' were discarded owing to the lack of clustering with samples from the same time point and type (Supplementary Figure S3a).

DE analyses. DE was calculated between (i) infected and control host samples at every time point, (ii) between time points of host control samples and (iii) between time points of phage-infected samples. Statistical package *edge*R (Robinson *et al.*, 2010) was used and genes with a false discovery rate <0.05 were considered as DE. Further, the genes obtained as DE within time points (e.g., 0 min vs the others) in host control samples were removed from the list of genes differentially expressed between infected and control. Although there were none in the original host, there were 13 (0.3% of total genes; Table 2) in the alternative host. Fold change (in $\log_2$FC) for those differentially expressed genes was calculated as the difference in expression (i.e., $\log_2$RPKM) between infected and control. To compare the two infections, the RPKM obtained from infected samples was assumed to derive from the sum of infected and uninfected cells such that $\mathrm{RPKM_i} = a \times \mathrm{RPKM}_x + (1-a) \times \mathrm{RPKM_c}$, where $\mathrm{RPKM_i}$ includes the total RNA-seq signal obtained from 'infected' samples (which includes infected and uninfected cells), RPKM_x is the expression of just the infected cells (and what needs to be solved for in the previous equation), a is the fraction of infected cells and RPKM_c is the expression of non-infected (control) cells. The resulting normalized $\log_2$FC equaled $\log_2\mathrm{RPKM}_x - \log_2\mathrm{RPKM_c}$. Genes were overexpressed if normalized $\log_2\mathrm{FC} > 0$ and underexpressed if normalized $\log_2\mathrm{FC} < 0$.

Temporal expression clusters. Phage genes from infected samples were clustered according to their temporal expression profile (based on $\log_2$RPKM). All time points were compared against each other using *edge*R (see above), and the resulting genes with a false discovery rate <0.05 were hierarchically clustered to determine the number of stable clusters as described previously (Lindell *et al.*, 2007). Briefly, the scaled $\log_2$RPKM data set was hierarchically clustered using Pearson's correlation and resampled using the clusterStab R package (Smolkin and Ghosh, 2003).

Functional group assignment to differentially expressed host genes
Host genes found DE between infected and control samples at every time point (see above) were placed into one of 14 functional categories based on MG-RAST (Meyer *et al.*, 2008) (i.e., amino-acid metabolism, cell cycle and division, cell signaling, cell wall remodeling, DNA metabolism, energy, motility, nucleotide metabolism, prophages and transposons, transport, protein metabolism, RNA metabolism, stress, unknown) by grouping similar functions into the same category using R (Figure 4 and Supplementary Dataset S1, tabs *RNASeqOriginalHostCategories* and *RNASeqAlternativeHostCategories*).

Codon adaptation index
The codon usage table for the phage and both the original and the alternative hosts as well as the codon adaptation index) (Sharp and Li, 1987) were calculated with the EMBOSS software (Rice *et al.*, 2000).

Results and discussion

RNA-seq of phage ϕ38:1 and two similar C. baltica host strains with contrasting lytic infections
Phage ϕ38:1 (a 70 kb double-stranded DNA podovirus) was previously shown to differentially infect two nearly identical *C. baltica* host strains, NN016038 (original) and no. 18 (alternative), which could be at least partially explained by differential adsorption efficiencies (Dang *et al.*, 2015). Here we assessed whether other features of the phage or host genomes or transcriptomes might also contribute to variability in infection efficiency across these hosts.

Genomically, both host strains are 100% identical at the 16S rRNA level (Holmfeldt *et al.*, 2007) and share 93% of their protein-coding genes with an average nucleotide identity of 99.99% (see Materials and methods). Among the unique genes in each host, no intact phage defense mechanisms were observed when surveyed for CRISPR, complete bacteriophage exclusion cassettes, restriction modification, abortive infection or toxin/antitoxin genes (Supplementary Table S1, tab *GenomicDifferencesHosts*) (Samson *et al.*, 2013; Goldfarb *et al.*, 2015).

Beyond adsorption and common defense mechanisms, differences in expression of phage or host genes could be involved in determining differences in infection efficiency. To assess this hypothesis, gene expression of phage and the *C. baltica* hosts were temporally quantified (Dang *et al.*, 2015) via time-resolved, whole-genome transcriptome sequencing (Figure 1). For each phage–host pair, three infections and no-phage controls were diluted 10-fold to minimize secondary infections after 15 min

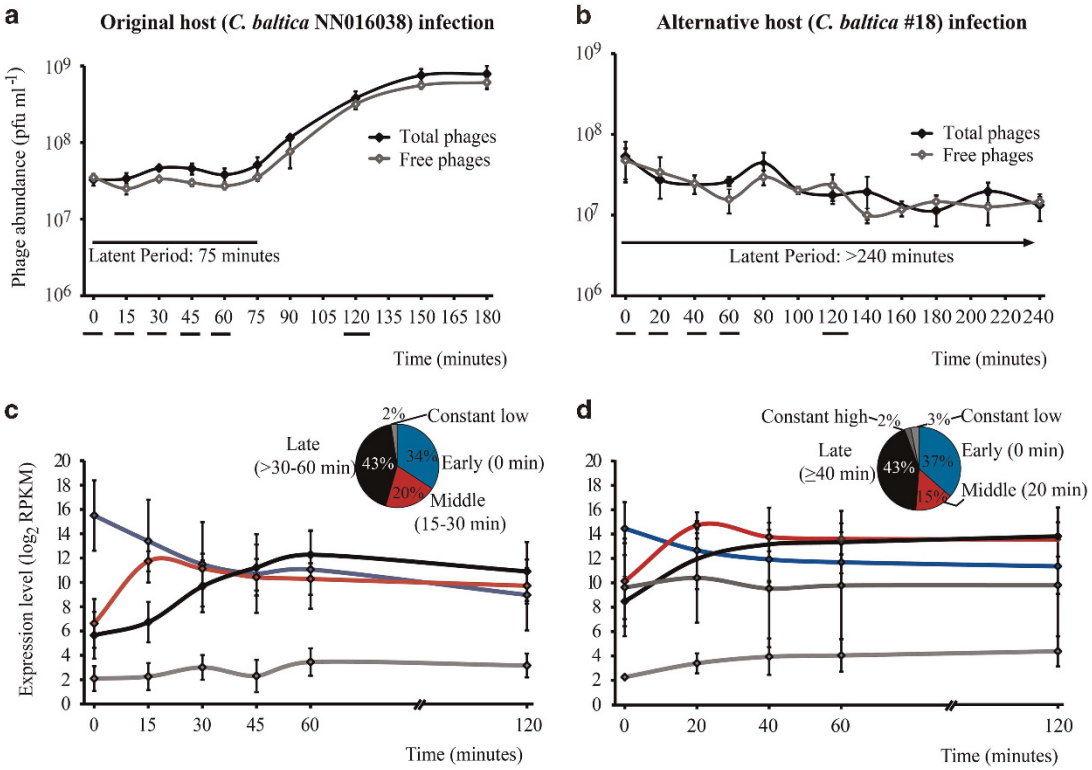

Figure 1 Infection dynamics of phage ϕ38:1 on its two *C. baltica* host strains. (**a**) Efficiently lytic infection on the original host displays a 75 min infection cycle via one-step growth curve, whereas (**b**) the inefficiently lytic infection shows no significant phage production over 4 h. (**c**) The efficiently lytic phage expresses genes in *early*, *middle*, *late* and *constant low* temporal groups, similar to when infecting the alternative host (**d**), where genes are also expressed in *constant high* temporal groups. Represented are the average and standard error of the biological triplicates (**a** and **b**) or the genes in each group (**c** and **d**). PFU, plaque forming units. Underlined are the time points assayed in the whole-genome transcriptional analyses.

of phage–host adsorption when time '0' was established (see Materials and methods). The observed infection dynamics were consistent with the previous work (Dang *et al*., 2015), with 75% infected cells and a 75 min latent period (two-tailed *t*-test, $P<0.05$; Figure 1a) on the original host, and 38% infected cells with no significant change in phage abundance over 4 h on the alternative host (two-tailed *t*-test, $P>0.05$; see Materials and methods and Figure 1b).

Transcriptome data were obtained via Sequencing by Oligonucleotide Ligation and Detection (SOLiD; see Materials and methods), generating 1.7×10^7–1.4×10^8 and 1.6×10^7–1.1×10^8 reads per sample in the original and alternative host infections, respectively, which covered both genomes in each experiment $\geqslant 40$ times on average (Supplementary Tables S2 and Supplementary Figure S1). Strand-specific sequencing provided expression of the phage intergenic regions, which enabled identification of six putative phage non-coding RNAs (Supplementary Figure S2 and Supplementary Table S4), which are now routinely discovered in phage transcriptome sequencing experiments and commonly have gene regulatory functions (Doron *et al*., 2015; Lin *et al*., 2015). Phage and host transcriptomes were further analyzed to explore the biology that drives efficient vs inefficient phage–host interactions as described in the following sections.

Universalities of lytic phage infection

The phage transcriptome during the efficiently lytic infection on the original host. The phage transcriptome when infecting the original host revealed that most phage genes were expressed in three temporally regulated groups—*early* (34% of total genes), *middle* (20.3% of total genes) and *late* (43.1% of total genes)—and a few genes with constant temporal expression—*constant low* (2.4% of total genes) (Figure 1c and Table 1). These gene groups were largely consecutive to each other and physically arranged such that early, middle and late progressed from one end to another in the phage genome (Figure 2). Exceptions to these patterns in ϕ38:1 included three genes with unknown functions that were expressed as middle and located within the late genes (Figure 2 and Supplementary Table S4), presumably prolonging expression or degradation of the transcripts until late infection, as posited for other phage–host systems (Lindell *et al*., 2007; Pavlova *et al*., 2012; Ceyssens *et al*., 2014).

Functional gene annotations provided insight into how ϕ38:1 phage infection progresses throughout an efficient lytic infection (Figure 2 and Supplementary Table S4). Early genes functioned in take-over to express phage genes and control host metabolism

Table 1 Summary statistics of phage ϕ38:1's transcriptome during the infection on *C. baltica* original (NN016038) and alternative (no. 18) host strains (FDR < 0.05)

	ϕ*38:1 on* C. baltica *NN016038*			ϕ*38:1 on* C. baltica *no. 18*		
	Number of genes	*% of total genes*	*Highest expression*	*Number of genes*	*% of total genes*	*Highest expression*
Early	42	34.2	0.0–15 min	45	36.6	0 min
Middle	25	20.3	15–30 min	18	14.6	20 min
Late	53	43.1	45–60 min	53	43.1	>40 min
Constant low	3	2.4	—	4	3.3	—
Constant high	—	—	—	3	2.4	—

Abbreviations: *C. baltica*, *Cellulophaga baltica*; FDR, false discovery rate.

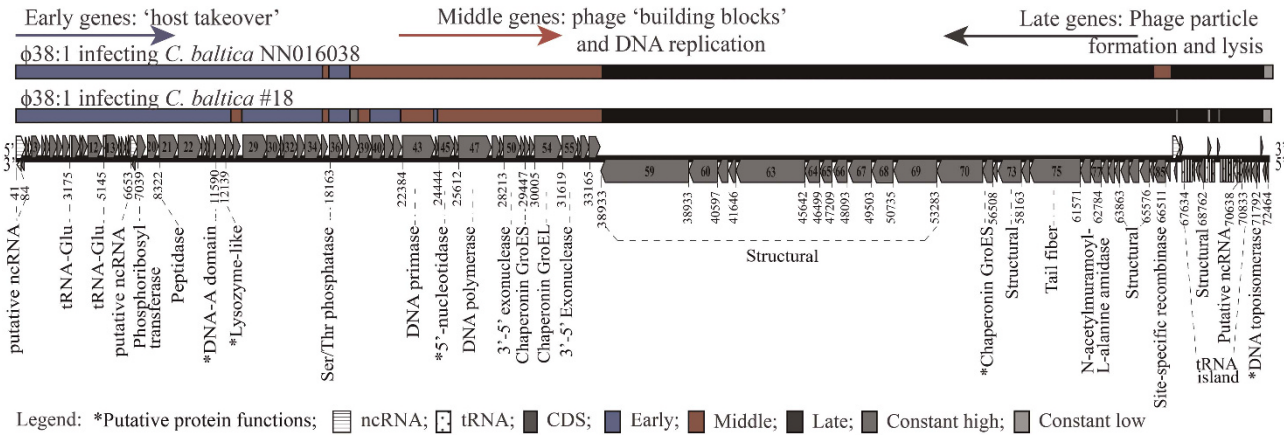

Figure 2 Whole-genome transcriptome of phage ϕ38:1 infecting its two *C. baltica* host strains. Genes in the plus strand (5′–3′) are represented above the genome line (black) and genes in the minus strand (3′–5′) are below the line. Some numbers inside the genes, representing gene product (gp) identifiers, as well as their function are displayed for guidance. Complete information can be found in the Supplementary Table S4 and Supplementary Dataset S1.

(e.g., ribosyltransferase *gp079*; Serine/Threonine phosphatase *gp036*). Additionally, some of these genes (lysozyme (*gp027*) (Rossmann *et al*., 2004), an α-mannosidase (*gp030*) (Prehm and Jann, 1976) and an S-layer-related protein (*gp034*) (Mann *et al*., 2005)) likely prepared the host cell to defend against superinfection via cell wall modifications. Alternatively, the lysozyme-like proteins may instead bind to RNA polymerase, as observed in coliphage T7 (Omcallister and Wu, 1978; Cheetham and Steitz, 2000), and function to benefit the phage via modulating host gene expression. As infection progressed, middle genes helped provide the phage 'building blocks' for phage reproduction, as the phage expressed genes that appeared to modulate nucleic acid and protein catabolism (e.g., exonuclease *gp050*; proteasome *gp039*, respectively), phage DNA replication (e.g., DNA primase *gp043*) and energy metabolism (e.g., phosphomannomutase *gp041*; NAD-dependent epimerase/dehydratase *gp043*). Finally, late genes promoted viral particle formation (e.g., major capsid protein *gp067*; tail fiber *gp071*) and cell lysis (e.g., *N*-acetylmuramoyl-L-alanine amidase *gp077*; Figure 2 and Supplementary Table S4).

Taken together these patterns are consistent with the temporal regulation that is common among lytic phages. Particularly, these phages appear to strongly regulate gene expression via either two ('early' and 'late'; Loskutoff *et al*., 1973; Pene *et al*., 1973; Poranen *et al*., 2006; Fallico *et al*., 2011; Ainsworth *et al*., 2013) or three ('early', 'middle'/'intermediate' and 'late'; Lindell *et al*., 2007; Legendre *et al*., 2010; Pavlova *et al*., 2012; Lavigne *et al*., 2013; Ceyssens *et al*., 2014; Doron *et al*., 2015; Lin *et al*., 2015) groups of physically proximal genes with similar functions. Such functions are tailored to the progression of the infection, with early host takeover and phage genome replication followed by late particle formation and cell lysis. Such patterns have now been observed across a wide range of phage systems, from classic phages (e.g., lambda (Lobocka *et al*., 2004), T4 (Miller *et al*., 2003) and PRD1 (Poranen *et al*., 2006)) to pathogens (e.g., *Pseudomonas* phiLuz19 (Lavigne *et al*., 2013)) and environmental viruses (e.g., giant viruses like mimivirus (Legendre *et al*., 2010) and *Pseudomonas* phiKZ (Ceyssens *et al*., 2014)) or the marine cyanophages (PSSP7 (Lindell *et al*., 2007), Syn9 (Doron *et al*., 2015) and PSSM2 (Lin *et al*., 2015)). The characterization of these phages has revealed a relative universality in regulating phage gene expression during a lytic infection, regardless of the host and the environment.

Phage transcriptome on the alternative host is similar to that on the original host. While the

alternative host infection dynamics was significantly delayed relative to that in the original host (Figure 1), the phage transcriptome was remarkably similar across infection of both hosts (Figures 1 and 2). First, phage genes were similarly spatiotemporally regulated as 36.6%, 14.6% and 43.1% of the total genes were expressed as early, middle and late, respectively, with the few remaining expressed constantly as constant high (2.4% of total genes) or constant low (3.3% of total genes) (Figures 1 and 2, Table 1 and Supplementary Table S4). Second, nearly all (90.2%) of the phage genes were expressed in the same temporal categories observed when infecting the original host (Figure 2 and Supplementary Table S4). Only one of the remaining genes (*gp085*) could be functionally annotated, as a putative site-specific recombinase. Since ϕ38:1 does not integrate its genome into the alternative host chromosome DNA (Holmfeldt *et al*., 2014), *gp085* might function either as an integrase in another host or be involved in phage DNA replication, similar to other members of this gene family (Cox, 2001). Third, phage gene expression was nearly identical across both hosts by the end of the efficiently lytic infection (i.e., 60 min; Figure 1). Thus, even though expression of genes such as *gp085* was delayed in the alternative host compared with the original host, potentially contributing to the inefficient infection, the overall similarities in genome-wide phage expression when infecting each host suggests that the phage was unlikely to have driven the infection inefficiencies across these hosts.

Host transcriptomes drive infection efficiency

While the phage transcriptomes were similar between infections, the host transcriptomes were strikingly different. First, relative to the uninfected controls, early host gene expression was globally underexpressed in the original host relative to the alternative host (Figure 3). The former suggests that the phage successfully redirected original host metabolic machinery towards making phage progeny, and is consistent with observations in some coliphage and cyanophage systems (Koerner and Snustad, 1979; Lindell *et al*., 2007; Fallico *et al*., 2011; Doron *et al*., 2015). Second, fewer genes were differentially expressed at each time point in the original (0.7–6.3%) vs the alternative (6–12%) host, with more diversity of genes differentially expressed in the alternative host (22.1%) than in the original (9.4%) host (Table 2 and Supplementary Figure S4). Small host responses to lytic phage infection, such as those observed here for the original host, are common among other genome-wide transcriptome studies of lytic model systems (Poranen *et al*., 2006; Ravantti *et al*., 2008; Fallico *et al*., 2011; Ainsworth *et al*., 2013; Lavigne *et al*., 2013; Ceyssens *et al*., 2014) and suggest a fast host takeover by the phage. However, the larger transcriptional response to phage infection observed in the alternative host is

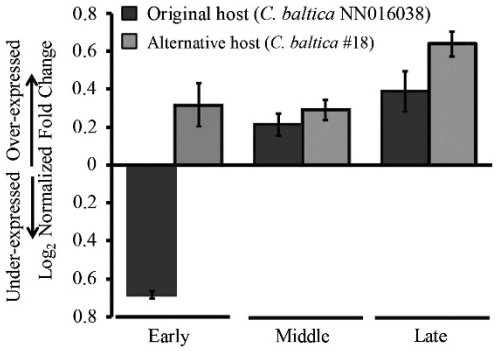

Figure 3 Transcriptional response of *C. baltica* to phage ϕ38:1 infection. Plotted is the average expression of all differentially expressed genes during infection relative to control (fold change), normalized by fraction of infected cells. Error bars represent the standard deviation of the biological replicates.

uncommon and comparable only to a lytic cyanophage that strongly depends upon its host's metabolic machinery (Lindell *et al*., 2007). Thus, either ϕ38:1 is more dependent on the host's metabolic machinery in the alternative host or this host more strongly defends against phage infection.

Functional analyses of differentially expressed host genes (Figure 4) helped discriminate between those two hypotheses, specifically suggesting benefits to phage reproduction in the original but not in the alternative host, as follows. First, the fold change of expression between infected and control was larger by 60 min (end of infection) in the original host than by 120 min in the alternative host (Figure 4), suggesting that the genes differentially expressed in the alternative host were not expressed at optimal levels as in the original host. Second, each stage of phage infection (i.e., early, middle and late) seemed supported by beneficial functions in the original host, but not in the alternative host (Figure 4 and Supplementary Dataset S1, tabs *RNASeqOriginalHostCategories*, *RNASeqAlternativeHostCategories*). For example, during early infection, the original host overexpressed genes involved in RNA, nucleotide and amino-acid metabolism, whereas the alternative host overexpressed a tRNA-Pro gene (protein metabolism category). This tRNA-Pro gene was expressed in the original host, but not differentially expressed during phage infection, which suggests that the gene product is not critical for the efficient infection. Most underexpressed genes in the original host were involved in the stress response, likely indicating the phage shutting down host defenses against infection (Figure 4 and Supplementary Dataset S1, tabs *RNASeqOriginalHostCategories*, *RNASeqAlternativeHostCategories*). Consistent with this, no such genes were underexpressed in the alternative host, where infection was challenged by host defenses and the infection was inefficient.

During middle gene expression, the original host overexpressed RNA (e.g., RNAP gene), nucleotide

Table 2 Summary statistics of the differentially expressed genes in *C. baltica* original (NN016038) and alternative (no.18) host strains (FDR < 0.05)

	C. baltica *NN016038*								C. baltica *no.18*						
	I vs C						*C vs C*	*Total DE*[a]	*I vs C*					*C vs C*	*Total DE*[a,b]
	0 min	*15 min*	*30 min*	*45 min*	*60 min*	*120 min*	*All*[a]		*0 min*	*20 min*	*40 min*	*60 min*	*120 min*	*All*[a]	
Genes	26	238	117	105	94	127	0	354	252	267	356	229	456	13	844
% of total genes	0.7	6.3	3.1	2.8	2.5	3.4	0.0	9.4	6.6	7.0	9.3	6.0	12.0	0.5	22.1

Abbreviations: C, control; *C. baltica*, *Cellulophaga baltica*; DE, differential expression; FDR, false discovery rate; I, infected.
[a]Non-redundant genes DE throughout the infection.
[b]After removing the genes DE in control samples.

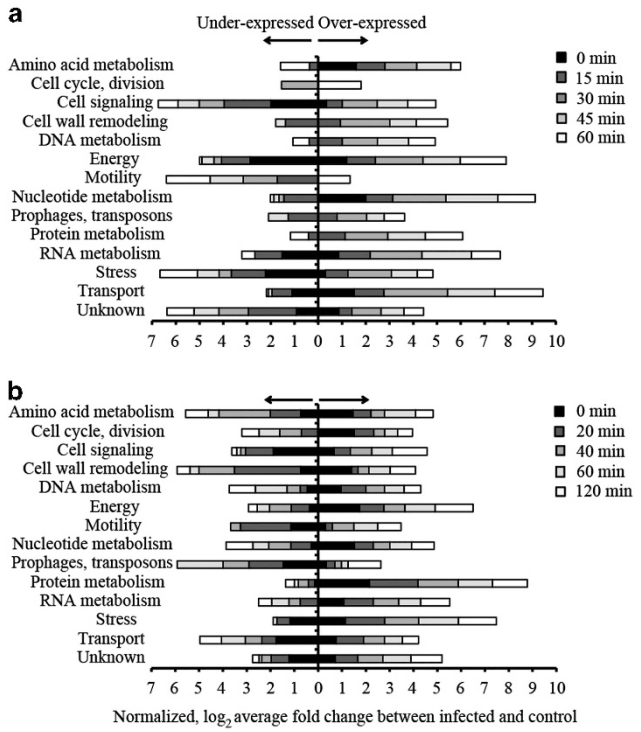

Figure 4 Host *C. baltica*'s response to phage ϕ38:1 infection. (**a**) Original (NN016038) and (**b**) alternative (no.18) host's transcriptional categories during infection, normalized to account for the fraction of infected cells. Detailed figures for DNA and protein metabolism are represented in Figure 5.

and DNA metabolism genes (Figure 4a), presumably to make phage transcripts and DNA (Brown and Bidle, 2014). Modulating the expression of RNAP in the original host was expected, as ϕ38:1 does not encode its own RNAP, and thus must repurpose the host RNAP towards transcribing phage genes. In contrast, the alternative host did not overexpress any of these functions and instead most highly expressed the tRNA-Asp gene again (which was not differentially expressed in the original host) and stress response genes (Figure 4b and Supplementary Dataset S1, tabs *RNASeqOriginalHostCategories*, *RNASeqAlternativeHostCategories*). Among underexpressed genes, the largest effect was in motility in both hosts and cell wall biosynthesis in the alternative host, suggesting the phage seeks to mitigate the host's defenses against infection or prevent unnecessary energy consumption in processes not beneficial to phage infection.

Finally, host transcriptomes during late infection again revealed cellular environments responding differently in the efficient vs inefficient infections. The original host overexpressed transport and energy genes (Figure 4a), perhaps reflecting the energetic demand of the lysis process that would follow. In addition, underexpression of stress genes suggested a phage-derived need to prevent host metabolic shutdown (Brissette *et al.*, 1990) before phage-controlled cell lysis, as in coliphage M13 (Karlsson *et al.*, 2005). In contrast, the alternative host overexpressed genes involved in stress and protein metabolism (tRNA-Asp) genes (at 60 min) while underexpressing amino-acid metabolism genes (Figure 4b), none of which aligned with the transcriptional response observed in the efficient infection. Additionally, the recurring DE of tRNA genes in the alternative host was not driven by any obvious codon bias across the hosts (Supplementary Figure S5), which is consistent with their near-complete lack of variation in %G+C (34.7% and 38.1% in both hosts and the phage, respectively), and suggests that such transcriptional pattern is unlikely to explain the infection differences between the hosts.

Inefficient infections are driven by host defenses against phage DNA and protein

At this point, it appeared that inefficient phage infection was driven, not by differential phage gene expression, but by host defenses throughout the infection against phage. To further test this hypothesis we sought to evaluate possible host responses against phage beyond transcription, such as DNA and protein production (Figure 5), that could be differentially transcribed in the efficient vs inefficient infections. First, investigation of core host genes potentially involved in phage defense (e.g., CRISPRs, restriction-modification, bacteriophage exclusion, abortive infection and toxin/antitoxin;

Goldfarb *et al*., 2015; Samson *et al*., 2013) based on their annotation suggested the presence of methyltransferases and endonucleases that could be part of a restriction-modification system (Supplementary Table S1, tabs *RNASeqOriginalHostPhageDefense* and *RNASeqAlternHostPhageDefense*) with only one gene (endonuclease *M666_14090*) differentially expressed in the alternative host (Supplementary Table S1, tab *RNASeqAlternHostPhageDefense*). This suggested differential regulation of phage DNA production in the alternative host, previously supported by phageFISH data, indicating that phage DNA replication occurred in both hosts, but took nearly twice as long in the alternative host (Dang *et al*., 2015). Second, following this lead, transcriptomics revealed that the original host overexpressed DNA degradation genes (e.g., nucleases) by middle infection (Figure 5a), which was coincident with a 63% reduction (range: 15–90%) of host relative to phage DNA from middle to late infection (Figure 5b). This suggested that the original host DNA was efficiently degraded, presumably to enable phage DNA replication given that (i) degraded host DNA can be recycled for phage DNA replication (Koerner and Snustad, 1979; Lavigne *et al*., 2013), and (ii) the phage expressed genes for replicating its genome (Figure 2) as phage DNA increased and host DNA decreased (Figure 5b). Here, the alternative host also overexpressed its DNA degradation genes early, and then underexpressed them late (including endonuclease *M666_14090*) (Figure 5a). However, phage DNA in the alternative host increased late instead of early as it did in the original host, was $>50\%$ less abundant and plateaued later (by 60 min) as compared with the original host (Figure 5b). These findings suggest that the alternative host contributed to the inefficient infection by specifically reducing and delaying phage DNA production.

Third, the expression pattern of host protein translation genes suggested that the alternative host defended against phage protein production as well, and experimental measurements following the formation of viral particles confirmed this. Specifically, the original host over-expressed translation genes from middle to late infection (Figure 5c), at which time phage particles were increasingly detected by EM (per cell averages of $\geqslant 9.4$ particles at 60–75 min; Figure 5d). In contrast, throughout the inefficient infection, the alternative host under-expressed translation genes (Figure 5c), and phage particle abundances were delayed (by >30 min) and reduced (by $>60\%$) as detected by EM (per cell averages of 3.5 particles at 90 min; Figure 5d). As all viruses are thought to depend on host translational machinery to make viral particles (Walsh and Mohr, 2011), the timing of over-expressed translation genes in the original host coinciding with the formation of phage particles presumably was designed to help make such particles, whereas the failure to be over-expressed in the alternative host contributed to yet another inefficiency of that infection.

An emerging paradigm of the biology of phage infection

Taken together, these data and previous findings (Holmfeldt *et al*., 2014) suggest that phage ϕ38:1 infection inefficiency is multidimensional as it derives initially from reduced phage adsorption to

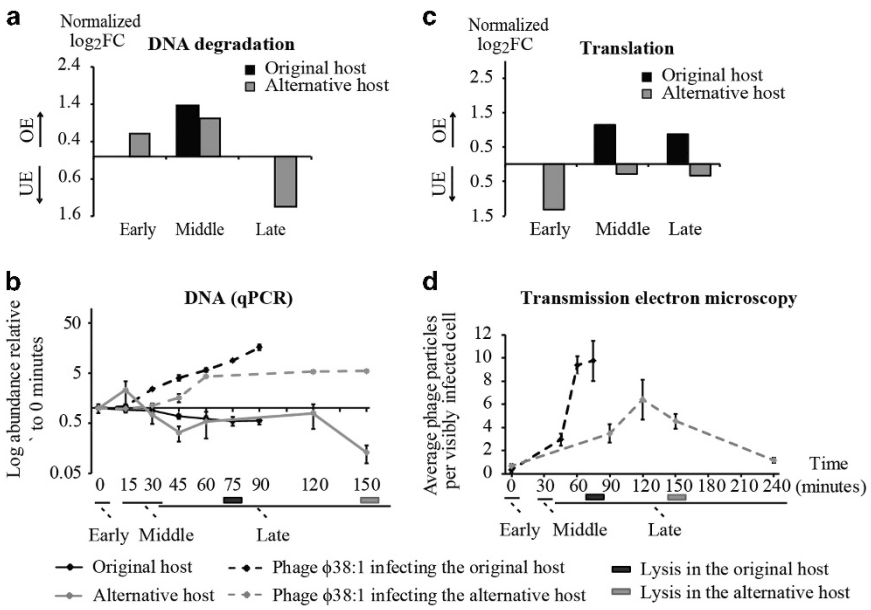

Figure 5 Regulation of the efficient and inefficient *C. baltica* phage–host interactions. Dynamics of (**a** and **b**) DNA replication and of (**c** and **d**) protein production in the original and alternative *C. baltica* hosts' infection with phage ϕ38:1. Gene expression is represented as the difference between infected and control (fold change, FC), normalized by the different fraction of infected cells in each host (normalized $\log_2$FC), resulting in overexpression (OE) or underexpression (UE). qPCR data is represented relative to time 0 min. For all analyses, the average of three biological replicates and their standard error are displayed.

the alternative host, but then intracellularly from failure to shut down host defenses in contrast to the original host. This likely leads to activating a stress response in the alternative host throughout middle and late stages of infection that is not observed in the original host, where stress genes are instead underexpressed early (Supplementary Figure S6). Consequently, throughout infection, such defenses in the alternative host succeeded at delaying synthesis of new phage genomes and virions. This complements current knowledge of bacterial defense mechanisms against viral infections, largely focused at early stages of infection, where phage attack is stymied through cell surface receptor modification and various ways of phage DNA degradation (Samson *et al.*, 2013). Instead, in our system, host defenses do not eradicate the virus; rather, they collectively and persistently lead to an inefficient infection, impacting phage adsorption, DNA and protein production.

As sequencing costs drop and non-traditional model systems are more routinely explored, knowledge of the intracellular dynamics of phage–host interaction are coming to light. For example, recent findings in a marine cyanophage infecting a broad range of hosts also showed similar phage transcription across the hosts and that host expression drove the differences between infections (Doron *et al.*, 2015). Taken together, with our findings, this places new emphasis on the role of the host cellular environment (measured here though the transcriptome) in enabling phage infection across diverse hosts. That phage gene expression is similar regardless of the host genetic background might also represent an emerging paradigm as it has now been observed in phage–bacterial systems ranging from heterotrophs to cyanobacteria. Additionally, our experimentally supported functional analyses in an environmentally relevant *Bacteroidetes* virus provide a mechanistic foundation for better understanding the efficiency of viral infections across hosts.

Conclusions

Phage–host interactions are fundamental for understanding microbial ecosystems, yet research to date has largely ignored the inefficient phage–host interactions that are likely common in the environment. The model system investigated here represents one of the most abundant phage types in the global oceans (Roux *et al.*, 2015a) and has a previously characterized broad host range (Holmfeldt *et al.*, 2014) with which to analyze intraspecies phage–host interactions, as well as how one phage can differentially infect two nearly identical host strains. As knowledge of bacterial and viral diversity, abundance and ecological relevance advances (e.g., Rinke *et al.*, 2013; Roux *et al.*, 2015b), it becomes critical to expand mechanistic understanding of virus–host interaction types beyond those displayed by traditional model systems. Developing environmental model systems has been experimentally challenging owing to slow growth and a lack of background knowledge and genetic tools. However, emerging technologies, such as genome-wide transcriptomics, offer new and powerful windows into the mechanistic underpinnings of environmentally relevant virus–host interactions. Taken together with new approaches to quantitatively survey viral diversity (Duhaime and Sullivan, 2012; Solonenko *et al.*, 2013) and virus–host interactions in the environment (reviewed in Dang and Sullivan, 2014), the field is poised to develop the foundational knowledge critical to model and understand the ecological impacts of virus–host interactions (Weitz *et al.*, 2015), as well as guide efforts to manipulate infections that will improve bacterial-based biotechnological productions and human health (Rohwer and Segall, 2015).

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We thank members from the Tucson Marine Phage Lab at the University of Arizona, especially Lacey Orsini, Sarah Schwenck, Clayton Pierce, Cassidy Danbury and Mario Moreno for technical support; Dr W Day from the Arizona Health Sciences Center Imaging Core facility for TEM support; and Dr H Mitchell at PNNL for initial bioinformatics support. Partial support for CHV came from the NIH Graduate Training Grant in Biochemistry and Molecular Biology T32 GM008659. SR was partly supported by a grant to the UA Ecosystem Genomics Institute through the UA Technology and Research Initiative Fund and the Water, Environmental, and Energy Solutions Initiative. Part of the research was performed using the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research and located at PNNL. This work was also funded by the Gordon and Betty Moore Foundation grants (GBMF no. 2631, 3790) and a DOE EMSL User Award (no. 47930) to MBS.

References

Abedon ST, Herschler TD, Stopar D. (2001). Bacteriophage latent-period evolution as a response to resource availability. *Appl Environ Microbiol* **67**: 4233–4241.

Abedon ST, Hyman P, Thomas C. (2003). Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. *Appl Environ Microbiol* **69**: 7499–7506.

Ainsworth S, Zomer A, Mahony J, Van Sinderen D. (2013). Lytic infection of *Lactococcus lactis* by bacteriophages Tuc2009 and c2 triggers alternative transcriptional host responses. *Appl Environ Microbiol* **79**: 4786–4798.

Bragg JG, Chisholm SW. (2008). Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS One* **3**: e3550.

Brissette JL, Russel M, Weiner L, Model P. (1990). Phage shock protein, a stress protein of *Escherichia coli*. *Proc Natl Acad Sci USA* **87**: 862–866.

Brown CM, Bidle KD. (2014). Attenuation of virus production at high multiplicities of infection in *Aureococcus anophagefferens*. *Virology* **466–467**: 71–81.

Brum JR, Steward GF, Jiang SC, Jellison R. (2005). Spatial and temporal variability of prokaryotes, viruses, and viral infections of prokaryotes in an alkaline, hypersaline lake. *Aquatic Microbial Ecol* **41**: 247–260.

Ceyssens PJ, Minakhin L, Van Den Bossche A, Yakunina M, Klimuk E, Blasdel B *et al.* (2014). Development of giant bacteriophage varphiKZ is independent of the host transcription apparatus. *J Virol* **88**: 10501–10510.

Cheetham GM, Steitz TA. (2000). Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases. *Curr Opin Struct Biol* **10**: 117–123.

Cox MM. (2001). Historical overview: searching for replication help in all of the rec places. *Proc Natl Acad Sci USA* **98**: 8173–8180.

Dang VT, Howard-Varona C, Schwenck S, Sullivan MB. (2015). Variably lytic infection dynamics of large Bacteroidetes podovirus phi38:1 against two *Cellulophaga baltica* host strains. *Environ Microbiol* **17**: 4659–4671.

Dang VT, Sullivan MB. (2014). Emerging methods to study bacteriophage infection at the single-cell level. *Front Microbiol* **5**: 724.

Doron S, Fedida A, Hernandez-Prieto MA, Sabehi G, Karunker I, Stazic D *et al.* (2015). Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J*; e-pub ahead of print 1 December 2015; doi:10.1038/ismej.2015.210.

Duhaime MB, Sullivan MB. (2012). Ocean viruses: Rigorously evaluating the metagenomic sample-to-sequence pipeline. *Virology* **434**: 181–186.

Falkowski PG, Fenchel T, Delong EF. (2008). The microbial engines that drive Earth's biogeochemical cycles. *Science* **320**: 1034–1039.

Fallico V, Ross RP, Fitzgerald GF, Mcauliffe O. (2011). Genetic response to bacteriophage infection in *Lactococcus lactis* reveals a four-strand approach involving induction of membrane stress proteins, D-alanylation of the cell wall, maintenance of proton motive force, and energy conservation. *J Virol* **85**: 12032–12042.

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

Goldfarb T, Sberro H, Weinstock E, Cohen O, Doron S, Charpak-Amikam Y *et al.* (2015). BREX is a novel phage resistance system widespread in microbial genomes. *EMBO J* **34**: 169–183.

Gomez-Pereira PR, Fuchs BM, Alonso C, Oliver MJ, Van Beusekom JE, Amann R. (2010). Distinct flavobacterial communities in contrasting water masses of the north Atlantic Ocean. *ISME J* **4**: 472–487.

Heintz C, Mair W. (2014). You are what you host: microbiome modulation of the aging process. *Cell* **156**: 408–411.

Holmfeldt K, Howard-Varona C, Solonenko N, Sullivan MB. (2014). Contrasting genomic patterns and infection strategies of two co-existing Bacteroidetes podovirus genera. *Environ Microbiol* **16**: 2501–2513.

Holmfeldt K, Middelboe M, Nybroe O, Riemann L. (2007). Large variabilities in host strain susceptibility and phage host range govern interactions between lytic marine phages and their Flavobacterium hosts. *Appl Environ Microbiol* **73**: 6730–6739.

Holmfeldt K, Solonenko N, Shah M, Corrier K, Riemann L, Verberkmoes NC *et al.* (2013). Twelve previously unknown phage genera are ubiquitous in global oceans. *Proc Natl Acad Sci USA* **110**: 12798–12803.

Hurwitz BL, Hallam SJ, Sullivan MB. (2013). Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* **14**: R123.

Hyman P, Abedon ST. (2012). Smaller fleas: viruses of microorganisms. *Scientifica (Cairo)* **2012**: 734023.

Karlsson F, Malmborg-Hager AC, Albrekt AS, Borrebaeck CA. (2005). Genome-wide comparison of phage M13-infected vs. uninfected *Escherichia coli*. *Can J Microbiol* **51**: 29–35.

Kirchman DL. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.

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

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.

Legendre M, Audic S, Poirot O, Hingamp P, Seltzer V, Byrne D *et al.* (2010). mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res* **20**: 664–674.

Lin X, Ding H, Zeng Q. (2015). Transcriptomic response during phage infection of a marine cyanobacterium under phosphorus-limited conditions. *Environ Microbiol*.

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.

Lobocka MB, Rose DJ, Plunkett G III, Rusin M, Samojedny A, Lehnherr H *et al.* (2004). Genome of bacteriophage P1. *J Bacteriol* **186**: 7032–7068.

Loskutoff DJ, Pene JJ, Andrews DP. (1973). Gene expression during the development of *Bacillus subtilis* bacteriophage phi 29. I. Analysis of viral-specific transcription by deoxyribonucleic acid-ribonucleic acid competition hybridization. *J Virol* **11**: 78–86.

Luo Y, Li B-Z, Liu D, Zhang L, Chen Y, Jia B *et al.* (2015). Engineered biosynthesis of natural products in heterologous hosts. *Chem Soc Rev* **44**: 5265–5290.

Mann NH, Clokie MR, Millard A, 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.

Meyer F, Paarmann D, D'souza M, Olson R, Glass EM, Kubal M *et al.* (2008). The metagenomics RAST server —a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* **9**: 386.

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

Mojica KD, Brussaard CP. (2014). Factors affecting virus dynamics and microbial host–virus interactions in marine environments. *FEMS Microbiol Ecol* **89**: 495–515.

Mortazavi A, Williams BA, Mccue K, Schaeffer L, Wold B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**: 621–628.

Omcallister WT, Wu HL. (1978). Regulation of transcription of the late genes of bacteriophage T7. *Proc Natl Acad Sci USA* **75**: 804–808.

Pavlova O, Lavysh D, Klimuk E, Djordjevic M, Ravcheev DA, Gelfand MS *et al.* (2012). Temporal regulation of gene expression of the *Escherichia coli* bacteriophage phiEco32. *J Mol Biol* **416**: 389–399.

Pene JJ, Murr PC, Barrow-Carraway J. (1973). Synthesis of bacteriophage phi 29 proteins in Bacillus subtilis. *J Virol* **12**: 61–67.

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.

Prehm P, Jann K. (1976). Enzymatic action of coliphage omega8 and its possible role in infection. *J Virol* **19**: 940–949.

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

Rice P, Longden I, Bleasby A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**: 276–277.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431–437.

Robinson MD, Mccarthy DJ, Smyth GK. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Rohwer F, Segall AM. (2015). In retrospect: a century of phage lessons. *Nature* **528**: 46–48.

Rossmann MG, Mesyanzhinov VV, Arisaka F, Leiman PG. (2004). The bacteriophage T4 DNA injection machine. *Curr Opin Struct Biol* **14**: 171–180.

Roux S, Brum J, Dutilh BE, Sunagawa S, Duhaime MB, Loy A *et al.* (2015a). Ecogenomics of uncultivated globally abundant ocean viruses. *Nature* (in review). doi:10.1101/053090.

Roux S, Hallam SJ, Woyke T, Sullivan MB. (2015b). Viral dark matter and virus–host interactions resolved from publicly available microbial genomes. *Elife* **4**: e08490.

Salmond GP, Fineran PC. (2015). A century of the phage: past, present and future. *Nat Rev Microbiol* **13**: 777–786.

Samson JE, Magadan AH, Sabri M, Moineau S. (2013). Revenge of the phages: defeating bacterial defences. *Nat Rev Microbiol* **10**: 675–687.

Shao Y, Wang IN. (2008). Bacteriophage adsorption rate and optimal lysis time. *Genetics* **180**: 471–482.

Sharp PM, Li W-H. (1987). The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* **15**: 1281–1295.

Smolkin M, Ghosh D. (2003). Cluster stability scores for microarray data in cancer studies. *BMC Bioinform* **4**: 36.

Solonenko SA, Ignacio-Espinoza JC, Alberti A, Cruaud C, Hallam S, Konstantinidis K *et al.* (2013). Sequencing platform and library preparation choices impact viral metagenomes. *BMC Genom* **14**: 320.

Stocker R. (2012). Marine microbes see a sea of gradients. *Science* **338**: 628–633.

Storms ZJ, Brown T, Cooper DG, Sauvageau D, Leask RL. (2014). Impact of the cell life-cycle on bacteriophage T4 infection. *FEMS Microbiol Lett* **353**: 63–68.

Suzuki R, Shimodaira H. (2006). Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540–1542.

Walsh D, Mohr I. (2011). Viral subversion of the host protein synthesis machinery. *Nat Rev Microbiol* **9**: 860–875.

Wang IN. (2006). Lysis timing and bacteriophage fitness. *Genetics* **172**: 17–26.

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

Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A *et al.* (2015). A multitrophic model to quantify the effects of marine viruses on microbial food webs and ecosystem processes. *ISME J* **9**: 1352–1364.

Wommack KE, Colwell RR. (2000). Virioplankton: Viruses in aquatic ecosystems. *Microbiol Mol Biol Rev* **64**: 69–114.

You L, Suthers PF, Yin J. (2002). Effects of *Escherichia coli* physiology on growth of phage T7 *in vivo* and *in silico*. *J Bacteriol* **184**: 1888–1894.

Zeglin LH. (2015). Stream microbial diversity in response to environmental changes: review and synthesis of existing research. *Front Microbiol* **6**: 454.

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