

ARTICLE **OPEN**

Genetic engineering of marine cyanophages reveals integration but not lysogeny in T7-like cyanophages

Dror Shitrit¹, Thomas Hackl², Raphael Laurenceau², Nicolas Raho², Michael C. G. Carlson¹, Gazalah Sabehi¹, Daniel A. Schwartz¹, Sallie W. Chisholm¹ and Debbie Lindell¹ 

© The Author(s) 2021

Marine cyanobacteria of the genera *Synechococcus* and *Prochlorococcus* are the most abundant photosynthetic organisms on earth, spanning vast regions of the oceans and contributing significantly to global primary production. Their viruses (cyanophages) greatly influence cyanobacterial ecology and evolution. Although many cyanophage genomes have been sequenced, insight into the functional role of cyanophage genes is limited by the lack of a cyanophage genetic engineering system. Here, we describe a simple, generalizable method for genetic engineering of cyanophages from multiple families, that we named REEP for REcombination, Enrichment and PCR screening. This method enables direct investigation of key cyanophage genes, and its simplicity makes it adaptable to other ecologically relevant host-virus systems. T7-like cyanophages often carry integrase genes and attachment sites, yet exhibit lytic infection dynamics. Here, using REEP, we investigated their ability to integrate and maintain a lysogenic life cycle. We found that these cyanophages integrate into the host genome and that the integrase and attachment site are required for integration. However, stable lysogens did not form. The frequency of integration was found to be low in both lab cultures and the oceans. These findings suggest that T7-like cyanophage integration is transient and is not part of a classical lysogenic cycle.

The ISME Journal (2022) 16:488–499; <https://doi.org/10.1038/s41396-021-01085-8>

INTRODUCTION

Marine cyanobacteria of the genera *Synechococcus* and *Prochlorococcus* are abundant primary producers that span vast areas of the oceans [1]. Their viruses (cyanophages) are also highly abundant [2–4] and influence cyanobacterial ecology and evolution [5–7], as well as global biogeochemical cycles [8–10]. All currently known cyanophages are tailed, double-stranded DNA viruses belonging to the *Myoviridae*, *Podoviridae* and *Siphoviridae* families [3, 11–15]. Over the last two decades, dozens of cyanophage genomes from isolates and metagenomes have been sequenced [13–21]. Analyses of these genomes have contributed insights into the general properties and lifestyle of marine cyanophages. For example, the T4-like cyanophages from the *Myoviridae* family and the T7-like cyanophages from the *Podoviridae* family share sets of core genes with the *Escherichia coli* T4 and T7 phages, respectively [13, 15, 18, 22]. These genes are involved in virion formation, DNA replication and packaging, transcription and other fundamental processes of the lytic infection cycle. In addition, cyanophage genomes contain many intriguing and unusual features. In particular, they encode a variety of “auxiliary metabolic genes” (AMGs) [10, 14, 22–25] of (cyano)bacterial origin, including those linked to photosynthesis [26, 27], carbon metabolism [28] and nutrient utilization [18, 29], as determined from homology-derived functional assignments. Furthermore, a large reservoir of diverse genes with no putative function is present in cyanophage genomes [16–18, 22], as found

for other viruses in the environment [30]. Some of these genes are conserved in multiple cyanophage types [16, 18, 21, 22] whereas others have no database match, and are termed “ORFans” [30, 31].

Elucidating cyanophage gene functions and their role in infection is key for understanding their ecological and evolutionary interactions with their cyanobacterial hosts. Heterologous expression experiments [32–34], in-vitro biochemical assays [32, 35] and comparisons between phage strains with different gene contents [28, 36] have shed light on the function of some cyanophage genes. However, to elucidate the role of these genes in the infection process and to determine their contribution to cyanophage fitness and evolution it is necessary to generate phage mutants and directly compare infection between phages with and without the gene of interest. Thus, a genetic engineering system that is simple and easily transferable between host-virus systems will provide a powerful tool for investigating the function of a wide range of genes in a suite of ecologically relevant cyanophage-cyanobacterial systems.

Lysogeny is a well-known phenomenon in which a virus remains in a non-lytic state within the host, typically after integrating its DNA into the host chromosome. After integration, phages can remain latent within the host cell as prophages until an environmental trigger leads to their excision from the genome and entry into a lytic cycle inside the cell [37]. The presence of the prophage suppresses infection by additional phages of the same strain and can sometimes even prevent infection by other

¹Faculty of Biology, Technion—Israel Institute of Technology, Haifa, Israel. ²Department of Civil and Environmental Engineering, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA.  email: dlindell@technion.ac.il

Received: 14 March 2021 Revised: 2 August 2021 Accepted: 3 August 2021
Published online: 24 August 2021

phage strains [38]. This superinfection immunity is a result of the transcriptional repression of phage lytic genes [39] or, in some cases, inhibition of adsorption by additional phages [40].

In recent years, the potential importance of lysogeny in the marine environment has come to the fore and its extent and significance is under debate [37, 41, 42]. Integrated phages have been observed in many sequenced bacterial genomes [41–43] and single cells from the environment [41]. Integrase genes, used for integration, are commonly found in viromes and cellular metagenomes [42, 44, 45] and their presence is often considered to infer the ability to integrate and partake in a lysogenic lifestyle [42, 45]. Integration and excision are key processes mediating horizontal gene transfer that shape the genomes of phage and microbial populations in the oceans [37, 46], including cyanophages and cyanobacteria [27, 47, 48].

T7-like cyanophages are generally described as lytic, due to their resemblance to the lytic archetype *Escherichia coli* phage T7 and their lytic behavior during infection [5, 49, 50]. Thus, it is surprising that many T7-like cyanophages carry an integrase gene, in some cases followed by putative attachment sites (*attP*) [13, 16, 17], two lysogeny-associated features that are known to facilitate site-specific integration in lysogenic phages and are absent from T7 itself. This raises the possibility of lysogeny in the T7-like cyanophages. However, intact prophages are absent from marine unicellular *Synechococcus* and *Prochlorococcus* isolates, despite evidence for remnants of phage integration [47, 51, 52]. In addition, efforts to isolate marine *Synechococcus* or *Prochlorococcus* lysogens of any type have failed [3, 5, 52], making it unclear whether this phenomenon occurs for cyanophages in these cyanobacteria.

Here, we developed a method for genetic engineering of cyanophages that is highly effective for two distinct and dominant cyanophage families that infect marine *Synechococcus* and *Prochlorococcus*, the T7-like and T4-like cyanophages. Employing this method, we provide experimental evidence for the integration of T7-like cyanophages into their host's genome despite a lytic infection cycle, and show that both the integrase gene and *attP* site are required for this process. We found that the frequency of integration was low both in the lab and the oceans. Stable lysogens of T7-like cyanophages were not found, however, despite repeated attempts in an experimental setup designed to obtain them even in the absence of superinfection immunity. These results suggest that lysogeny does not occur in T7-like cyanophages and that integration is less regulated and more transient than that described for classical lysogenic phages.

RESULTS AND DISCUSSION

REEP: a genetic engineering system for cyanophages

Genetic engineering of phages requires two stages. The first is the generation of mutations in the phage genome and the second is the isolation of the mutant phages. The genetic engineering system presented here is based on naturally high rates of homologous recombination between plasmid and phage DNA during infection. Following recombination, an enrichment procedure and a PCR screen are used for the isolation of mutant phages. We name this method REEP for REcombination, Enrichment and PCR screening. This is a simple and generalizable method particularly well suited for ecologically relevant host-virus systems for which genetic tools are limited.

High frequency of recombination between cyanophage and plasmid DNA during infection

The first stage of any genetic engineering system is the generation of mutations. In the REEP method, as in various other phage genetic engineering methods [53–55], phage mutations are generated during infection by homologous recombination between cyanophage DNA and recombination templates carried

on replicative plasmids inside cyanobacterial hosts (Fig. 1a). Recombination templates consist of two regions, 200–300 bp long (Table 1), that are homologous to those flanking the target gene in the phage genome. This length of homologous regions was chosen to be long enough to increase the probability for homologous recombination but without cloning an entire phage gene into the recombination template, which may be lethal when introduced into the host. A short foreign DNA TAG sequence (20–60 bp long) is positioned between the homologous regions and replaces the target gene with minimal interruption to the compact phage genome. This TAG sequence is used later for detection of recombinant phages by PCR. The recombination templates are cloned into a broad host-range, mobilizable plasmid in *E. coli* and inserted into marine *Synechococcus* by conjugation [56] to produce recombination host strains (see "Methods" for details). Infection of a recombination host with the wild-type phage results in lysates with a mixture of wild-type and recombinant phages (recombinant-containing lysates).

We tested this recombination system for cyanophages belonging to the two major cyanophage families and investigated loci not expected to be essential for phage reproduction. The host strain used was *Synechococcus* sp. strain WH8109. The generation of recombinant-containing phage lysates was successful for all three cyanophages tested: S-TIP37 and Syn5 from the T7-like cyanophages, and Syn9 from the T4-like cyanophages (Table 1). Recombinant phages were produced at an average frequency of 1.31×10^{-3} (0.1%) for five recombinant phages, ranging from 3.61×10^{-4} to 3.71×10^{-3} , as determined by quantitative PCR (qPCR) (Table 1). This range of frequencies was observed for various deletion and insertion sizes and was not dependent on the family or strain of phage investigated (Table 1). These recombination frequencies were high enough to suggest that screening, rather than selection systems, could be used to isolate recombinant phages. For comparison, recombination rates reported previously for other phages range from nearly 10^{-3} [57] to as low as 10^{-9} [58].

A generalizable method for the screening and isolation of recombinant phages

The next stage in a genetic engineering method is the isolation of recombinants. Although the recombination step described above is quite efficient, detecting recombinant phages by PCR screening thousands of individual plaques would be very time consuming and labor intensive. To overcome this problem, we developed a procedure that combines an enrichment step for recombinant phages and a high throughput PCR screening procedure for detection of recombinant-enriched lysates (Fig. 1b). In this procedure, hundreds of host cultures, grown in 96-well plates, are infected with subsamples of the recombinant-containing lysate at a concentration of 5–10 phages per well. This concentration is pre-determined using the most probable number method (MPN) and ensures lysis of all infected wells. Though most wells contain only wild-type phages, wells with a recombinant phage generate lysates that are generally enriched more than 100-fold to contain 10–20% recombinants rather than the ~0.1% recombinants present in the initial recombinant-containing lysate (Table 1).

Once lysis of the infected cultures is visible, a PCR screen is performed to detect wells enriched with recombinant phages that are identified based on the inserted TAG sequence (Fig. 1b). It is generally sufficient to screen 3–10 96-well plates (180–600 wells, infected with 900–6000 phages) to detect several wells highly enriched with recombinant phages. PCR can be performed on each plate individually or after pooling samples from each row of 10 plates into a single plate and screening the 10 individual wells that went into a positive pooled well. This reduces the number of PCR reactions tenfold, although may reduce the sensitivity of detecting recombinant phages. The contents of the PCR-positive

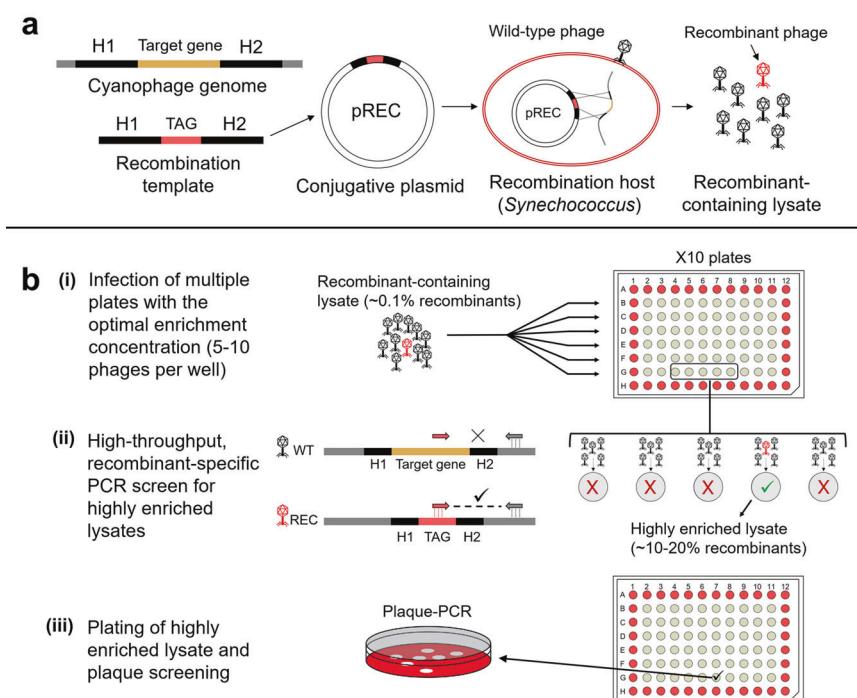


Fig. 1 Diagram of the REEP genetic engineering system for cyanophages. a The process of producing recombinant-containing phage lysates. A recombination template is composed of homologous regions (H1, H2, 200–300 bp each) that flank the target gene to be deleted in the phage genome and a short TAG sequence that will be inserted in place of the target gene. This is cloned into the pREC plasmid and inserted into *Synechococcus* to produce the recombination host. This host is infected with wild-type cyanophages. Homologous recombination that occurs during infection produces a lysate containing wild-type (black) and recombinant (red) phages. **b** Enrichment and screening for the isolation of recombinant phages. (i) Multiple 96-well plates are infected with the recombinant-containing lysate at the optimal enrichment dilution, which is 5–10 phages per well. (ii) Once lysis is complete, the plates are screened by recombinant-specific PCR to detect recombinant phages based on their TAG sequence. Wells in which a recombinant phage was present among the initial 5–10 phages used for infection will now contain a lysate that is highly enriched with recombinants (>100-fold) and will be detected in the PCR screen. (iii) The highly enriched lysates from PCR positive wells are plated for plaque screening and isolation of the recombinant phage.

wells are plated on cyanobacterial lawns for plaque formation and ~10–50 individual plaques are screened by PCR to obtain an isolated recombinant phage that is further purified and taken for validation (see “Methods”).

The REEP method’s power lies in its simplicity. It requires only the ability to insert a plasmid into the bacterial host and homologous recombination, which is a frequent process in many DNA and RNA viruses [59, 60]. Even for viruses where homologous recombination is less frequent, it can be enhanced by expression of a recombination system commonly used for genetic engineering (e.g., the Lambda RED system) [61]. Furthermore, this method does not require any specialized equipment or tools and the enrichment and screening is relatively high throughput when using 96-well plates and multi-channel pipettes. It generates mutants with minimal disruption to the phage genome, using a very short sequence to replace the target gene, and can potentially be used to generate scarless mutations (i.e., without insertion of a TAG, with screening done based on differences in amplicon size). The REEP method was successfully employed on all three cyanophages tested in this study and was efficient in generating a range of deletions and insertions in cyanophages belonging to two distinct phage families with different characteristics, including vast differences in genome size and content. This demonstrates the flexibility of this method, a valuable feature for studying a set of phage strains, as often done in ecologically relevant systems.

Many methods for genetic engineering of bacteriophages exist [61–63]. Although the generation of mutations is quite straightforward, a major obstacle is the isolation of the mutant phages. This is largely due to the lack of generalizable selection markers for

phages, similar to those used in bacteria, such as antibiotic resistance genes. Specialized systems have been developed for well-studied model phages that often rely on strain-specific characteristics [62, 63]. For example, host genes that are essential for phage infection but not for host growth can be used as selection markers. CRISPR-Cas based methods have been proposed as a generalizable approach for counter-selection against wild-type phages [53, 64], as have the use of reporter genes to isolate recombinant phages by visual screening [54, 55]. However, setting up such systems can be difficult and time consuming, especially as they require expression of foreign genes in the host, often entailing the development and optimization of suitable regulatory elements and codon optimization. This is no trivial task when working with non-model organisms with a limited set of molecular tools and is likely to require specific adjustments for each host-virus system of interest. In addition, overexpression of a reporter gene, as well as the insertion of a large fragment into the compact phage genome, may have an effect on the phage phenotype beyond that of the deletion of the gene of interest, confounding results from infection experiments with such mutant phages.

Prior to the development of the REEP method, we examined several other approaches for isolation of recombinant phages without success (see supplementary information for details). First, three different CRISPR-Cas based methods were constructed for targeting and eliminating non-recombinant cyanophages [53, 64]. None of these systems were efficient against cyanophages, despite being adjusted for use in cyanobacteria. In parallel, we attempted to optimize several genes encoding fluorescent proteins for use as reporter genes for visual screening of

Table 1. Frequency of recombinant phages in recombinant-containing lysates for various mutations in different cyanophages.

| Mutant strain name | Parent cyanophage strain | Phage type | Deleted sequence (size) | Inserted sequence (size) | Homology length (left/right) | Recombinant phage frequency (n = 3) |
|------------------------|--------------------------|------------|--|--------------------------|------------------------------|-------------------------------------|
| S-TIP37- Δ int | S-TIP37 | T7-like | gp11- integrase (836 bp) | pRL-TAG (57 bp) | 266/290 bp | 8.48E-04 ± 8.61E-04 |
| S-TIP37- Δ attP | S-TIP37 | T7-like | attP - putative integration (67 bp) | DR-TAG1 (23 bp) | 234/234 bp | 3.71E-03 ± 3.11E-04 |
| S-TIP37-Cm | S-TIP37 | T7-like | Non-coding region downstream of gp45 (22 bp) | proCAT (961 bp) | 268/268 bp | 3.61E-04 ± 2.09E-04 |
| Syn5- Δ int | Syn5 | T7-like | gp17 - integrase (508 bp) | pRL-TAG (57 bp) | 251/226 bp | 6.75E-04 ± 1.77E-04 |
| Syn9- Δ 6.5 | Syn9 | T4-like | Non-coding region downstream of gp6 (14 bp) | DR-TAG1 (23 bp) | 251/274 bp | 9.57E-04 ± 3.77E-04 |
| Average | | | | | | 1.31E-03 ± 1.22E-03 |

recombinant phages, yet none were successful in generating a detectable signal above the autofluorescence of the cyanobacterial host. Another approach was screening for recombinant phages by plaque-hybridization [65], which led to the detection of recombinants, but did not yield viable recombinant cyanophages despite many attempts. Last, cloning of whole phage genomes into fosmid vectors and inserting them into *Synechococcus* by conjugation were attempted, failing in the step of transfer into *Synechococcus*. These unsuccessful attempts further emphasize the advantages of our straightforward method. It has no gene expression involved, is free of potentially confounding effects on mutant phage fitness and phenotype, and is unlikely to require changes to be transferred to other host-virus systems. These advantages make this method an ideal candidate for genetic engineering in viruses for which genetic methods are currently unavailable, especially for ecologically relevant host-virus systems that are not well characterized and have limited genetic tools.

T7-like phage integration into marine cyanobacteria

Integrase genes are often detected in the genomes of cultured phages and bacteria as well as in viromes and metagenomes collected from the oceans [13, 16, 22, 37, 43]. Based on this, many marine phages are hypothesized to have the ability to integrate into the genome of their hosts and to have a lysogenic life cycle [37, 46]. With a genetic system in hand, we set out to experimentally test these hypotheses for T7-like cyanophages. This entailed assessing whether: (1) T7-like cyanophages integrate into the genome of their host; (2) the integrase gene is functional and, together with the *attP* site, mediate integration; and (3) stable integration ensues and superinfection immunity is conferred.

Site-specific integration of T7-like cyanophages occurs during lytic infection

The presence of an integrase gene and a putative phage attachment site (*attP*) in T7-like cyanophages was first reported for the P-SSP7 phage that infects *Prochlorococcus* sp. strain MED4 [13]. The potential *attP* site is downstream of the integrase gene and is identical to a region in a tRNA-Leu gene of the host. Similarly, we found an integrase gene and a comparable putative *attP* site in the genome of S-TIP37 (Fig. 2a), a T7-like cyanophage that infects *Synechococcus* sp. strain WH8109 [66].

We first tested the ability of these two wild-type phages to integrate into the genomes of their hosts over a 10-hour period after phage addition (Fig. 2b, c). Intracellular genomic DNA was subjected to PCR assays designed to amplify across expected integration junctions when the phages integrate into the cyanobacterial tRNA-Leu gene. Both the S-TIP37 and the P-SSP7 phages were found to integrate into their hosts' genomes at the expected site from 2 hours after infection onwards (Fig. 2b, c).

We then tested whether the integrase gene and the *attP* sites are functional and are involved in the integration process. This is particularly important because marine cyanobacteria, including both *Synechococcus* WH8109 and *Prochlorococcus* MED4, code for integrase genes [47, 67], and host enzymes have been shown to mediate phage integration in other systems [68]. We generated two mutant strains of S-TIP37 in which either the integrase gene (Δ int) or the *attP* site (Δ attP) were knocked out. Neither left nor right integration junctions were formed during infection with these mutant S-TIP37 phages (Fig. 2b). These results indicate that both the integrase gene and the *attP* site are essential for integration.

Interestingly, infection by both wild-type and integrase-mutant S-TIP37 phages exhibited infection dynamics and phage progeny production typical of classic lytic phages (Fig. S1). The length of the latent period and phage fitness were no different in the mutant and wild-type S-TIP37 phages (Fig. S1, S2a). Nutrient availability has previously been proposed to affect the frequency of lysogeny [37, 69, 70]. We therefore hypothesized that fewer

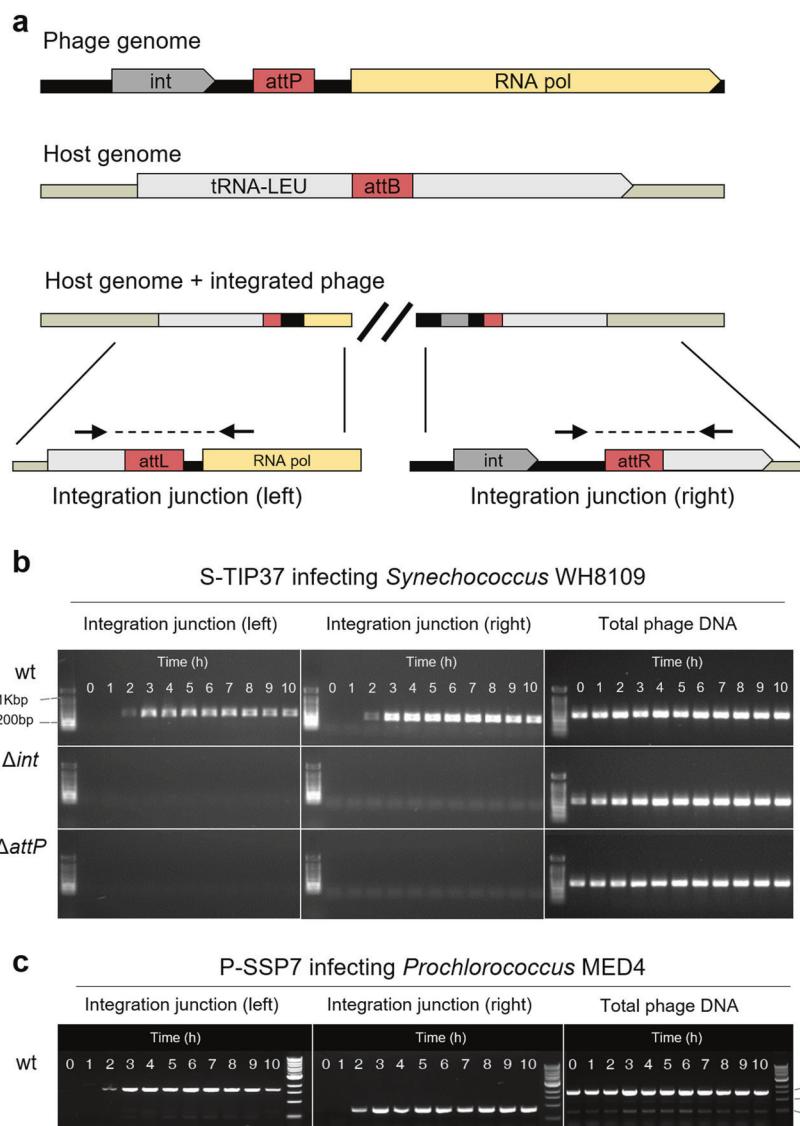


Fig. 2 Site-specific integration of T7-like cyanophages. a Organization of the integration elements found in cyanophages S-TIP37 and P-SSP7, and in their host genomes. Integration junctions formed due to phage integration and the location of the primers (black arrows) used to detect them by PCR are also shown. Int- integrase, attP/B – attachment sites on the phage/bacterial genomes, respectively, attL/attR – left and right attachment sites after integration, respectively. Primers for total phage DNA assessment are located elsewhere on the phage genome (see Table S5). **b** Evidence for phage integration from a PCR assay for detection of integrated phages during infection of *Synechococcus* WH8109 with S-TIP37 phage strains: wild-type (wt), integrase deleted (Δ int) and attP deleted (Δ attP) strains. **c** Integration of wild-type phage P-SSP7 (wt) during infection of *Prochlorococcus* MED4. All results shown here are representative of 3 independent experiments.

phages would be produced (i.e., lower fitness would result) if more cells were lysogenized under nutrient deprivation. However, even when the host was grown in nutrient-depleted medium, wild-type and integration mutants demonstrated similar progeny production (Fig. S2).

The above results raised the question of whether phage integration is a basic part of the lytic infection cycle, occurring in every infected cell. To assess this we quantified the number of infected cells containing an integrated phage using the iPolony method, a method capable of detecting single molecules of viral DNA in infected cells [71]. The percentage of integrated phages correlated with the number of infected cells, increasing as more phages adsorbed and entered the cell and decreasing as phages were released and cells lysed (Fig. 3). However, phage integration occurred only in a small fraction of cells, ranging from 0.04 to 0.32%, even though up to 60% of the cells were infected (Fig. 3b).

Thus, the frequency of integrated phages was low, ranging from 0.3–1% of infected cells for most of the infection cycle (Fig. 3c). These findings indicate that integration is a relatively rare event, and may be transient, rather than a fundamental part of lytic infection in these cyanophages.

S-TIP37 integration is not part of a classic lysogenic cycle

Hallmark features of lysogeny include stable integration into the host genome, a regulatory system that controls the lytic/lysogenic life cycles and superinfection immunity [37, 38, 72]. To test whether integrated T7-like cyanophages possess these features, we first searched for homologs of repressors of lytic infection present in known lysogenic phages. No such genes were identified in T7-like cyanophages. Next, we performed a series of experiments to test whether stable integration of S-TIP37 occurs when infecting its *Synechococcus* WH8109 host. Classic

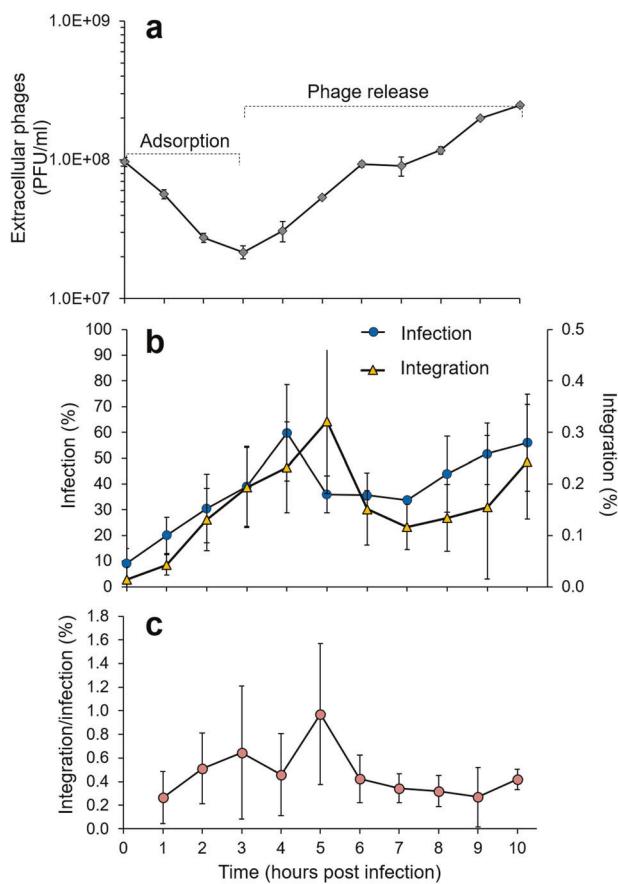


Fig. 3 Dynamics of S-TIP37 infection and integration. a Phage growth curve of the wild-type S-TIP37 phage during infection of *Synechococcus* WH8109. **b** Dynamics of infected cells (blue circles, left Y-axis) and of cells containing integrated wild-type phages (yellow triangles, right Y-axis; only left junctions were tested), determined by single-cell, solid phase PCR. Percent infection and integration were calculated from the number of colonies divided by the number of cells added to the slide. **c** Average ratio of cells with integrated phages to total number of infected cells, calculated by dividing average % integration by average % infection from each timepoint. Integration at timepoint zero was below the limit of detection and was thus not used to calculate the integration/infection ratio. No integration was detected with the S-TIP37 integrase mutant, which served as a negative control in this experiment.

lysogens (e.g., Lambda lysogens of *E. coli*) can be isolated based on their superinfection immunity, by simply plating an infected culture on agar plates and testing the colonies that grow for the presence of a prophage [73]. However, when we attempted this previously for various marine cyanobacteria and their integrase-containing cyanophages (including P-SSP7 and S-TIP37), none of the resistant colonies were found to contain a prophage [6, 74]. We therefore predicted that if stable integration occurs it is extremely rare. To overcome this we used our genetic system to generate a strain of S-TIP37 carrying a chloramphenicol resistance gene (*proCAT*) in a non-coding region of the phage genome. This engineered phage strain, S-TIP37-Cm (Table 1), was allowed to adsorb to *Synechococcus* WH8109 cultures which were then plated on chloramphenicol-containing agar plates. If stable lysogens that provide superinfection immunity exist then chloramphenicol resistant colonies are expected to grow. This approach allows testing a vast number of phage-infected cells because it eliminates colonies that originate from non-infected cells as well as from spontaneous phage resistant mutants, allowing only

prophage-containing colonies that express the chloramphenicol resistance gene to grow. Roughly 3×10^9 cells were plated in five different experiments, which was expected to yield up to 9×10^5 colonies considering the integration frequency of ~0.3% found in the previous experiment (Fig. 3b). However, no chloramphenicol resistant colonies were found. Note that the *proCAT* gene provides chloramphenicol resistance in our pREC replicating plasmids (Fig. 1a) and when expressed from the chromosome of *Synechococcus* strains [75]. This suggests that despite phages integrating into the genome of their host this integration is not stable or that superinfection immunity is not provided by the integrated phage.

We then asked whether stable integration occurs but superinfection immunity is not provided, i.e. that prophage-containing cells are killed as a result of subsequent infections. To address this question, we repeated the above experiments in a manner that removes and separates free phages from host cells, preventing a second round of infection (see "Methods"). Here too, no colonies grew on chloramphenicol plates, despite growth of colonies in the absence of chloramphenicol, which confirmed that a second round of infection and lysis was indeed prevented. Finally, to rule out the possibility that prophages exist in these latter colonies but did not provide resistance to chloramphenicol, we tested 2000 colonies for the presence of prophages by PCR and found none. These findings indicate that stable integration did not occur, irrespective of whether superinfection immunity ensues or not. We note that these experiments do not enable us to ascertain whether superinfection immunity is provided for the short period during which phages are integrated into the host genome.

Our combined findings suggest that the site-specific integration observed here in T7-like cyanophages is not a part of a classical lysogenic lifestyle, but rather is a transient process. This intrinsic lack of stable prophage integration may be due to the absence of regulatory genes that, in well studied lysogen systems, suppress the expression of lytic cycle genes and maintain the phage genome integrated into the host genome [38, 72]. Their absence can result in spontaneous excision of the integrated phage and induction of the lytic cycle, or in lysis by a superinfecting phage. In addition, lytic infection may be initiated by the integrated phage from within the host genome. Such a phenomenon was recently reported in the classic lysogenic phage Lambda, with lytic infection observed from a phage integrated into the host genome, possibly due to low expression of the CII gene coding for the repressor of the lytic cycle [76].

T7-like cyanophage integration in the oceans

We sought to estimate the frequency of integrated T7-like cyanophages in cyanobacteria in the oceans. First, we searched for *attP* sites in 24 publicly available genomes in this group of cyanophages [14, 16, 19, 22, 77] by searching for homologous sequences downstream of integrase-containing cyanophage genomes in cyanobacteria (Table S1). We found that of the 13 cyanophage genomes with integrase genes, 11 also carried an *attP* site (Fig. 4). Consistent with the *attP* sites for P-SSP7 [13] and S-TIP37 (this study), all *attP* sites were perfect matches for regions in cyanobacterial tRNA-Leu genes, which can thus be considered a general cyanobacterial attachment (*attB*) site. Based on their sequences, we divided the *attP* sites into four groups, with groups I and II belonging to phages that infect *Prochlorococcus* and groups III and IV belonging to phages that infect *Synechococcus* (Fig. 4b). Second, we analyzed publicly available metagenomic datasets from six different projects (Table S2), searching for reads of cyanobacterial origin that contain these T7-like cyanophage attachment sites. The number of reads containing "empty" integration sites (i.e., cyanobacterial DNA only) or phage-host integration junctions were used to calculate the integration frequency. For the two dominant groups, I and IV, an integration frequency of 1.6% and 0.08% was found, whereas

a

| Cyanophage | Accession | int | attP | Host strain | Genome publication |
|------------|------------|-----|------|--------------------------------|----------------------|
| P-GSP1 | HQ332140 | | | <i>Prochlorococcus</i> MED4 | Labrie et al. [16] |
| P-HP1 | GU071104 | | | <i>Prochlorococcus</i> NATL2A | Labrie et al. [16] |
| P-RSP2 | HQ332139 | | | <i>Prochlorococcus</i> MIT9302 | Labrie et al. [16] |
| P-RSP5 | GU071102 | | | <i>Prochlorococcus</i> NATL1A | Labrie et al. [16] |
| P-SSP10 | HQ337022 | | | <i>Prochlorococcus</i> NATL2A | Labrie et al. [16] |
| P-SSP2 | GU071107 | + | + | <i>Prochlorococcus</i> MIT9312 | Labrie et al. [16] |
| P-SSP3 | HQ332137 | + | + | <i>Prochlorococcus</i> MIT9312 | Labrie et al. [16] |
| P-SSP5 | GU071100 | | | <i>Prochlorococcus</i> MIT9515 | Unpublished |
| P-SSP6 | HQ634152 | | | <i>Prochlorococcus</i> MIT9515 | Unpublished |
| P-SSP7 | AY939843 | + | + | <i>Prochlorococcus</i> MED4 | Sullivan et al. [13] |
| P-SSP9 | HQ316584 | + | + | <i>Prochlorococcus</i> SS120 | Labrie et al. [16] |
| P60 | AF338467 | | | <i>Synechococcus</i> WH7805 | Chen & Lu [77] |
| S-CBP1 | KC310802 | + | + | <i>Synechococcus</i> CB0101 | Huang et al. [22] |
| S-CBP2 | KC310806 | + | + | <i>Synechococcus</i> CB0208 | Huang et al. [22] |
| S-CBP3 | KC310803 | + | + | <i>Synechococcus</i> CB0101 | Huang et al. [22] |
| S-CBP4 | KC310804 | | | <i>Synechococcus</i> CB0101 | Huang et al. [22] |
| S-CBP42 | KC310805 | + | | <i>Synechococcus</i> WH7803 | Huang et al. [22] |
| S-RIP1 | HQ317388 | | | <i>Synechococcus</i> WH8101 | Unpublished |
| S-RIP2 | HQ317389 | | | <i>Synechococcus</i> WH7803 | Unpublished |
| S-TIP37 | MH540083.1 | + | + | <i>Synechococcus</i> WH8109 | Unpublished |
| S-TIP28 | MZ803112 | + | + | <i>Synechococcus</i> WH8109 | Unpublished |
| S05-C243 | GU943065 | + | + | Unknown (uncultured phage) | Ghai et al. [19] |
| S08-C41 | GU943073 | + | + | Unknown (uncultured phage) | Ghai et al. [19] |
| Syn5 | EF372997 | + | | <i>Synechococcus</i> WH8109 | Pope et al. [14] |

b**c**

| Attachment site group | Bacterial integration sites | | Integration frequency (%) |
|-----------------------|-----------------------------|----------------------------------|---------------------------|
| | Empty (Host DNA only) | Occupied (Integration junctions) | |
| I | 95504 | 1605 | 1.600 |
| II | 111480 | 3 | 0.003 |
| III | 239 | 0 | 0.000 |
| IV | 7334 | 6 | 0.080 |

Fig. 4 Lysogeny- associated elements present in T7-like cyanophage genomes. a Analysis of available T7-like cyanophage genomes for the presence of integrase genes (*int*) and *attP* sites. **b** Sequence comparison and group division of the *attP* sites found in the analyzed genome. Cyanobacterial host genus is indicated by font color of the phage name (green- *Prochlorococcus*, red- *Synechococcus*, black- unknown). **c** Frequency of integration found for cyanophages in the different attachment site groups in oceanic samples.

group II and III showed only 0.003% and 0% integration, respectively (Fig. 4c). Thus, similar to lab findings, T7-like cyanophage integration is rare but detectable in the oceans.

Cell density has been hypothesized to impact lysogeny. However, this is still under debate due to conflicting evidence (summarized by Brum et al. [78]). Nonetheless, our combined findings of rare integration at high cell densities in lab experiments (5×10^7 cells·ml $^{-1}$) as well as at the considerably lower cyanobacterial densities found in the oceans (ranging from $\sim 10^4$ to 3×10^5 cells·ml $^{-1}$ [1]) argues against density being a factor for the T7-like cyanophages.

CONCLUSIONS

The discovery of phages as highly abundant and diverse members of the community in virtually every habitat on Earth [9, 10, 79, 80] has fueled recognition of their ecological importance. Over the past two decades there has been an explosion in the sequence information available for environmental viruses. Yet a substantial portion of viral genomes contains genes of unknown function [16, 17, 22, 30, 36], as well as other genes with putative functions that were unexpected in viral genomes, leading to intriguing hypotheses as to their role in infection [13–15]. The lack of a genetic system broadly applicable to ecologically relevant models has precluded gaining fundamental understanding of the many novel features of these significant members of the natural world. This method now provides the opportunity, in the post-genomics era, to assess the function of interesting virus genes directly in the infection process and their impact on virus fitness and evolution. Thus, the REEP genetic engineering method developed here places marine cyanophages among the few model organisms that are culturable, genetically engineerable and ecologically significant.

The possibility of integration and a potential lysogenic lifestyle in T7-like phages was first hypothesized 15 years ago for T7-like cyanophages [13, 14, 16, 22] and more recently for T7-like pelagiphages [81, 82], based on the presence of integrase genes and *attP* sites in their genomes. In our first use of the REEP method we addressed this hypothesis for the T7-like cyanophages. Our findings unequivocally indicate that these elements are used for integration into the host genome even though these are essentially lytic phages. Indeed, no evidence for stable integration nor lysogeny was found, although we cannot categorically rule it out for all members of this family or under all environmentally relevant conditions. Nonetheless, the absence of intact prophages in cyanobacterial genomes, including hundreds of single-cell genomes from diverse environments [83, 84], provides independent evidence for the lack of classical lysogeny in these phages. Furthermore, lysogeny-derived immunity was never the mechanism of resistance among hundreds of experimentally selected resistant cyanobacterial strains, including against the P-SSP7 [6] and S-TIP37 [74] phages, even though integration occurs at a 1000-fold greater frequency than the appearance of resistance mutations [6]. Thus, although the presence of the integrase gene facilitated integration when an *attP* site was also present, we caution against automatically equating the presence of the gene with a lysogenic lifestyle.

The physiological or evolutionary benefit that marine T7-like cyanophages obtain from carrying these elements and transiently integrating into host genomes remains uncertain. One feasible scenario is that integration provides a safety net for phages that cannot complete their lytic cycle. This may be important when environmental conditions are not suitable for lytic infection. Alternatively, this process may allow defective phages to be rescued through complementation or recombination with other phages infecting the same host cell, and perhaps even facilitate the formation of new genetic combinations, accelerating phage evolution. Clearly, phage integration is important for host and phage evolution through horizontal gene transfer [37, 46].

Transient integration in up to 1% of infections may enhance the frequency of the introduction of new genes into microbial hosts as well as the capture of microbial genes by viruses in the environment.

MATERIALS AND METHODS

Culture growth

Cultures of *Synechococcus* sp. strain WH8109 were grown in artificial seawater (ASW) medium [85], with modifications as described elsewhere [86], at 21°C and a light intensity of 30 μmol photons $\text{m}^{-2} \text{s}^{-1}$. *Prochlorococcus* sp. strain MED4 was grown in Pro99 medium [87], at 18°C and 70 μmol photons $\text{m}^{-2} \text{s}^{-1}$. All cultures were grown under a 14:10 light-dark cycle. Chlorophyll *a* fluorescence (excitation at 440 nm, emission at 660 nm) was used as a proxy for cell density and was measured in 96-well plates using a Synergy Mx Microplate Reader (Biotek). Pour-plating was performed as previously described [56], using the ASW growth medium with low-melting point agarose at a final concentration of 0.28% [87, 88] and the addition of 1 mM sodium sulfite [88]. For isolation of cyanobacterial colonies, a heterotrophic helper strain, *Alteromonas* sp. EZ55, was added to the pour-plate mixture [89]. All bacterial and phage strains used in this work are listed in Table S3. Experimentation was carried out at cell abundances of $\sim 3-5 \times 10^7$ cells/ml for *Synechococcus* WH8109 and *Prochlorococcus* MED4.

Lysate preparation and phage growth curves

Lysates were prepared by infecting cyanobacterial cultures at early logarithmic phase with phage from either a single plaque or a pre-existing liquid lysate. Once full lysis of the culture was observed, lysates were filtered through 0.22 μm syringe filters (Millex-GV, Millipore) and stored in glass tubes at 4°C. Plaque assays were used to determine infective phage concentrations. Lysate samples were diluted and plated in agarose plates containing cyanobacteria at concentrations sufficient to produce lawns, allowing formation of plaques. Phage growth curve experiments were performed by infecting cyanobacterial cultures at a multiplicity of infection (MOI) of 1-3, with subsamples collected over a period of 10 h. To measure the concentration of infective phages in the extracellular medium samples of 0.1 ml were collected, added to 0.9 ml of medium, and filtered through a 0.22 μm syringe filter to remove cyanobacterial cells and the filtrate used in plaque assays.

Construction of plasmids and bacterial conjugation

The plasmid pDS-proCAT (Fig. S3) was used as a backbone for the construction of all recombination plasmids used in this work. This replicative plasmid is a derivative of the broad host-range mobilizable vector pRL1342 (received as a gift from Peter Wolk [90]). It was modified by replacing the original chloramphenicol resistance gene with a cyanobacterial-optimized copy, *proCAT*. This gene was codon-optimized for expression in several *Prochlorococcus* strains and was found to work well in *Synechococcus* strains. For efficient expression, the *proCAT* gene was fused to the *rnpB* promoter and the *atpB* ribosome-binding site of *Prochlorococcus* MED4. Recombination templates were constructed either by the PCR overlap extension method [91] or by a two-step cloning procedure. A short TAG sequence (20–60 bp) was inserted between two regions of homology on the phage genome (~ 250 bp long) (H1 and H2) that flank the region to be deleted (Fig. 1a). Recombination plasmids (pREC) were inserted by electroporation into *E. coli* strain DH10b, an efficient recipient of large plasmids. Plasmids were extracted from PCR-positive colonies using a miniprep kit (NucleoSpin Plasmid EasyPure, MACHEREY-NAGEL) and transformed into *E. coli* SM10 or S17.1, conjugation donor strains carrying the *λpir* gene [92]. Conjugation of plasmids into *Synechococcus* was done as described previously [56] and resulting recombination hosts were grown in the presence of 1 $\mu\text{g}/\text{ml}$ chloramphenicol. Plasmids used in this study are listed in Table S4.

Production of recombinant-containing cyanophage lysates and estimating recombinant frequencies

Recombination plasmids were conjugated into *Synechococcus* sp. strain WH8109 to form recombination hosts. Recombination hosts were infected with wild-type phages at a low MOI (~ 0.01). After lysis of the cultures, lysates were filtered and the 0.2 μm filtrate was stored as described above. The presence of recombinant cyanophages was tested by PCR, using

primers that anneal to the TAG sequence and specifically detect recombinants (Table S5) using 2x Taq PCR mix (Tiangen) and 2 μ l of the lysate (in 20 μ l reactions). Recombinant phage frequency was determined for several representative lysates by real-time qPCR (see below).

Isolation of recombinant cyanophages by enrichment and PCR screening

Recombinant phages were enriched in 96-well plates containing cultures of *Synechococcus* sp. strain WH8109 at early logarithmic growth phase. The optimal phage enrichment concentration was determined by the MPN method. *Synechococcus* cultures grown in a 96-well plate were infected with 10-fold serial dilutions of the recombinant-containing lysate, with each dilution used to infect a single row of wells. The optimal enrichment concentration is operationally determined to be the highest tenfold serial dilution at which lysis is observed in an entire row of wells. This concentration is typically 5–10 phages per well, and is used in the subsequent enrichment step to infect the internal 60 wells of three to ten 96-well plates containing *Synechococcus* cultures. Once the cultures lyse, the wells were screened by PCR for the presence of recombinant phages in 20 μ l reactions using 10 μ l Taq PCR MasterMix (Tiangen), 0.4 μ M primers and 2 μ l of phage lysate (without prior filtration). Samples were run on 1.5% agarose gels with \sim 0.2 μ g·ml $^{-1}$ ethidium bromide in large electrophoresis trays containing 200 wells each. Lysates from PCR-positive wells were filtered through a 0.22 μ m syringe filter and plated. Plaques were purified twice and verified by PCR to obtain pure recombinant phage mutants. Full genome sequencing of all phage strains (Table S3), including the wild-type parental strains, were performed to verify the mutation and to ensure an identical genetic background to the wild-type phage.

Real-time quantitative PCR to measure recombination frequencies

Recombination frequencies were determined by real-time qPCR, in which the copy-number of recombinant and non-recombinant phage DNA was quantified. DNA templates were extracted from phage lysates using Wizard PCR Preps DNA Purification Resin and Minicolumns (Promega), as previously described [93]. Template DNA and primers (0.2 μ M each) were added to LightCycler 480 SYBR Green I Master mix (Roche) and amplified using a LightCycler 480 Real-Time PCR System (Roche). The LightCycler 480 software was used to calculate the number of cycles required to reach the optimal fluorescence threshold (Ct). All amplicons tested by qPCR were first amplified by PCR and cloned into plasmids using the TOPO TA Cloning Kit (Thermo Fischer). Plasmid DNA carrying the tested amplicons was extracted and used to generate calibration curves for calculation of the absolute DNA copy number in each qPCR reaction.

Detection of integrated phages using PCR

Cells were collected over the course of phage growth curve experiments for PCR detection of phage DNA integrated into the cyanobacterial genome. Samples were filtered onto polycarbonate 0.22 μ m pore-size membrane filters (GE), washed twice with growth medium to remove free phages, and washed once with 3 ml of preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl, pH 8) [94, 95]. The filter was flash-frozen in liquid nitrogen and stored at -80°C . A heat lysis method was used to extract DNA from cells collected on filters [94, 95]. Filters were resuspended in 10 mM Tris-HCl solution (pH = 8), shaken in a bead-beater (Mini-BeadBeater, Biospec) for 2 min (3450 oscillations/min) without beads. Samples were heated at 95°C for 15 minutes and the supernatant collected. For detection of integrated phage DNA, 2 μ l of intracellular template DNA were used in 20 μ l PCRs (Taq PCR MasterMix, Tiangen), as described above. Primers specific for detection of total phage DNA and integration junctions were used (Table S5).

Quantification of infected cells and integrated phages using the iPolony method

For single-cell quantification of integrated phages, cultures of the *Synechococcus* WH8109 host were infected with the wild-type S-TIP37 phage (MOI = 3). The S-TIP37 integrase mutant was used as a negative control in infection experiments with the host strain. Samples from phage-infected cultures were collected at different time points, fixed in 0.1% glutaraldehyde, incubated for 20 min in darkness, flash-frozen in liquid nitrogen and stored at -80°C . Samples were sorted using an Influx flow

cytometer (BD) based on their chlorophyll *a* fluorescence (excitation at 488 nm, emission at 580/30 nm) and forward scatter to remove free phages and excess glutaraldehyde [71]. A known number of cells (quantified using the same flow cytometer) were added to iPolony reactions designed for intracellular DNA detection [71]. Briefly, a mixture of cells, PCR reagents and polyacrylamide gel were poured into custom-made glass slides (11.6 μ l final volume). A 15 min cell lysis step was followed by in-gel PCR in a slide thermal cycler (DNA Engine with dual-block slide chamber, Bio-Rad). An acrydite-modified primer (Table S5) covalently binds one PCR strand to the gel, creating local DNA amplification clusters, termed polonies (PCR-colonies) [96]. Polonies were hybridized with a Cy5-modified fluorescently labeled DNA probe (Table S5) after removal of the unbound PCR strand. Hybridized polonies were identified using a GenePix 4000B microarray scanner (Axon Instruments). Plasmids harboring the tested amplicons (with left integration junction or total phage) served as positive controls for PCR amplification and efficiency. Percent integration was adjusted for the efficiency of detection of single virus genome copies which was 33% for S-TIP37 [71].

Testing for stable integration and superinfection immunity

The S-TIP37-Cm strain was constructed by inserting the *proCAT* chloramphenicol resistance gene into the non-coding region between *g45* and *g46* of the S-TIP37 phage. This engineered phage strain was used to infect *Synechococcus* WH8109 cultures at a high MOI of 3. After a 2 h incubation to allow phage adsorption, 10^8 cells per plate were plated in the presence of 1 μ g·ml $^{-1}$ chloramphenicol. A total of 3×10^9 cells was plated in 5 independent experiments.

To differentiate between lack of stable integration and lack of superinfection immunity, in the event of no chloramphenicol resistant colonies, we carried out the above experiment after physically preventing subsequent infection of cells by free phages. Cells were diluted to 10^5 cells·ml $^{-1}$ and sorted to remove free phages [71] remaining after the 2 h adsorption step. In addition, cells were plated at various dilutions with and without chloramphenicol to physically distance potential prophage-containing cells from phages newly released by lytically infected cells. A total of 15 plates were plated at each dilution in 3 independent experiments. The presence of colonies in the absence of chloramphenicol indicated that this procedure prevented infection by remaining or newly released phages. The efficiency of plating of these phage-exposed cultures was approximately one-third of the non-infected control cultures, as expected from the percent infection measured two hours post infection (see Results and Discussion). Over 2000 of these colonies, that grew in the absence of chloramphenicol and originated either from non-infected cells or from lysogens, were tested for the presence of prophage by PCR.

Fitness assays

To assess the fitness of the different phage strains, *Synechococcus* cultures were infected over a 2-day period beginning with a low MOI of 0.001. Phage abundances were determined by plaque assay at the beginning and end of experiment. A 2-day period was used to ensure host cultures did not become a limiting resource for phage production [97]. Phage fitness, expressed as the number of population doublings per day, was calculated as $(\log_2(N_t/N_0))/t$, where t is the assay time in days and N_0 and N_t are the phage concentrations at the beginning and end of the experiment, respectively [97, 98]. Growth of the host culture was measured by chlorophyll *a* fluorescence (see above). For growth under nutrient deprivation (Fig. S2), a low-nutrient ASW medium was used, containing all ingredients but at 10% nutrient concentrations for NaNO₃, NaH₂PO₄, NaHCO₃ and trace metals. Exponentially growing *Synechococcus* cultures were centrifuged for 5 min at 5200 RCF, and resuspended in this low nutrient ASW medium. Nutrient deprivation was verified by reduced chlorophyll *a* fluorescence relative to growth in 100% ASW medium.

Analyses of *attP* sites and estimation of cyanophage integration frequencies in the oceans

To identify potential attachment sites between cyanophages and their putative hosts we compared 727 *Prochlorococcus* and *Synechococcus* genomes (Table S1) against 12 integrase-carrying cyanophage genomes (Fig. 4) using BLAST (BLAST v2.6.0 + : blastn -task blastn -reward 1 -penalty -4 -gapopen 5 -gapextend 2 -perc_identity 94 -eval 10e-5). We filtered hits of 39–44 bp length, collapsed those overlapping by at least 20 bp, and only kept the ones present in more than 10 host genomes and directly downstream of the integrase gene (bedtools v2.27.1) [99]. From the curated alignment of the

corresponding sequences, we identified four distinct 39 bp attachment site motifs, all of which match the first half of a tRNA-Leu gene.

To estimate how many cyanobacteria in the wild carry integrated phages, we screened 1093 publicly available metagenomic libraries from six different marine sequencing projects [100–104] (Table S2) for reads containing a match to one of the four attachment site motifs with up to two differences (BBMap v38.16: bbdtk.sh k = 39 edist = 2) [105]. We determined the locations of putative attachment sites in the cyanobacterial host genomes by BLAST best hit against the four attachment site motifs with a minimum bit score of 52. Based on those data, we generated a database of fragments containing the attachment site motif and 500 bp of up- and downstream flanking regions for all combinations of cyanobacteria and phages containing the same site. We then mapped the pre-screened metagenomic reads to this database using bwa mem v0.7.16a-r1181 and filtered the results using a custom Perl script (alignment length \geq 100 bp, aligned read fraction \geq 95%, overlap with flanking regions \geq 20 bp) [106, 107]. Finally, we obtained counts for “empty” and “occupied” integration sites by counting reads matching to “host-att-host” and “host-att-phage”/“phage-att-host” fragments using a custom R script.

REFERENCES

1. Flombaum P, Gallegos JL, Gordillo RA, Rincón J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc Natl Acad Sci USA*. 2013;110:9824–9.
2. Goldin S, Hulata Y, Baran N, Lindell D. Quantification of T4-like and T7-like cyanophages using the polony method show they are significant members of the viroplankton in the North Pacific Subtropical Gyre. *Front Microbiol*. 2020;11:1210.
3. Waterbury JB, Valois FW. Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. *Appl Environ Microbiol*. 1993;59:3393–9.
4. Marston MF, Sallee JL. Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island’s coastal waters. *Appl Environ Microbiol*. 2003;69:4639–47.
5. Clokie MRJ, Mann NH. Marine cyanophages and light. *Environ Microbiol*. 2006;8:2074–82.
6. Avrani S, Wurtzel O, Sharon I, Sorek R, Lindell D. Genomic island variability facilitates *Prochlorococcus* -virus coexistence. *Nature*. 2011;474:604–8.
7. Marston MF, Piercley FJ, Shepard A, Gearin G, Qi J, Yandava C, et al. Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci USA*. 2012;109:4544–9.
8. Fuhrman JA. Marine viruses and their biogeochemical and ecological effects. *Nature*. 1999;399:541–8.
9. Suttle CA. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol*. 2007;5:801–12.
10. Breitbart M, Bonnain C, Malki K, Sawaya NA. Phage puppet masters of the marine microbial realm. *Nat Microbiol*. 2018;3:754–66.
11. Suttle CA, Chan AM. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar Ecol Prog Ser*. 1993;92:99–109.
12. Sullivan MB, Waterbury JB, Chisholm SW. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature*. 2003;424:1047–51.
13. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW. Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol*. 2005;3:0790–806.
14. Pope WH, Weigele PR, Chang J, Pedulla ML, Ford ME, Houtz JM, et al. Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a “horned” bacteriophage of marine *Synechococcus*. *J Mol Biol*. 2007;368:966–81.
15. Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, et al. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol*. 2007;9:1675–95.
16. Labrie SJ, Frois-Moniz K, Osburne MS, Kelly L, Roggensack SE, Sullivan MB, et al. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol*. 2013;15:1356–76.
17. Huang S, Wang K, Jiao N, Chen F. Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol*. 2012;14:540–58.
18. Sullivan MB, Huang KH, Ignacio-Espinoza JC, Berlin AM, Kelly L, Weigele PR, et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol*. 2010;12:3035–56.
19. Ghai R, Martin-Cuadrado AB, Molto AG, Heredia IG, Cabrera R, Martin J, et al. Metagenome of the Mediterranean deep chlorophyll maximum studied by direct and fosmid library 454 pyrosequencing. *ISME J*. 2010;4:1154–66.
20. Ma Y, Allen LZ, Palenik B. Diversity and genome dynamics of marine cyanophages using metagenomic analyses. *Environ Microbiol Rep*. 2014;6:583–94.
21. Sabehi G, Shaulov L, Silver DH, Yanai I, Harel A, Lindell D. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proc Natl Acad Sci USA*. 2012;109:2037–42.
22. Huang S, Zhang S, Jiao N, Chen F. Comparative genomic and phylogenomic analyses reveal a conserved core genome shared by estuarine and oceanic cyanopodoviruses. *PLoS One*. 2015;10:1–17.
23. Ignacio-Espinoza JC, Sullivan MB. Phylogenomics of T4 cyanophages: lateral gene transfer in the ‘core’ and origins of host genes. *Environ Microbiol*. 2012;14:2113–26.
24. Crummett LT, Puxty RJ, Weihe C, Marston MF, Martiny JBHH. The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology*. 2016;499:219–29.
25. Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: Implications for mechanisms of cyanophage evolution. *Environ Microbiol*. 2009;11:2370–87.
26. Mann NH, Cook A, Millard A, Bailey S, Clokie M. Bacterial photosynthesis genes in a virus. *Nature*. 2003;424:741–741.
27. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci USA*. 2004;101:11013–8.
28. Puxty RJ, Millard AD, Evans DJ, Scanlan DJ. Viruses inhibit CO₂ fixation in the most abundant phototrophs on Earth. *Curr Biol*. 2016;26:1585–9.
29. Kelly L, Ding H, Huang KH, Osburne MS, Chisholm SW. Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J*. 2013;7:1827–41.
30. Yin Y, Fischer D. Identification and investigation of ORFans in the viral world. *BMC Genomics*. 2008;9:24.
31. Clokie MR, Millard AD, Mann NH. T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. *Virol J*. 2010;7:291.
32. Rihtman B, Bowman-Grahl S, Millard A, Corrigan RM, Clokie MRJ, Scanlan DJ. Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ Microbiol Rep*. 2019;11:448–55.
33. Dammeyer T, Bagby SC, Sullivan MB, Chisholm SW, Frankenberger-Dinkel N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr Biol*. 2008;18:442–8.
34. Roitman S, Hornung E, Flores-Uribe J, Sharon I, Feussner I, Béjà O. Cyanophage-encoded lipid desaturases: Oceanic distribution, diversity and function. *ISME J*. 2018;12:343–55.
35. Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci USA*. 2011;108:E757–64.
36. Puxty RJ, Perez-Sepulveda B, Rihtman B, Evans DJ, Millard AD, Scanlan DJ. Spontaneous deletion of an “ORFanage” region facilitates host adaptation in a “photosynthetic” cyanophage. *PLoS One*. 2015;10:e0132642.
37. Howard-Varona C, Hargreaves KR, Abedon ST, Sullivan MB. Lysogeny in nature: mechanisms, impact and ecology of temperate phages. *ISME J*. 2017;11:1511–20.
38. Ranade K, Poteete AR. Superinfection exclusion (*sieB*) genes of bacteriophages P22 and λ. *J Bacteriol*. 1993;175:4712–8.
39. Fogg PCM, Allison HE, Saunders JR, McCarthy AJ. Bacteriophage Lambda: a paradigm revisited. *J Virol*. 2010;84:6876–9.
40. van Houte S, Buckling A, Westra ER. Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol Mol Biol Rev*. 2016;80:745–63.
41. Tuttle MJ, Buchan A. Lysogeny in the oceans: lessons from cultivated model systems and a reanalysis of its prevalence. *Environ Microbiol*. 2020;22:4919–33.
42. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobian-Guemes AG, et al. Lytic to temperate switching of viral communities. *Nature*. 2016;531:466–70.
43. Touchon M, Bernheim A, Rocha EPC. Genetic and life-history traits associated with the distribution of prophages in bacteria. *ISME J*. 2016;10:2744–54.
44. Roux S, Hallam SJ, Woyke T, Sullivan MB. Viral dark matter and virus-host interactions resolved from publicly available microbial genomes. *Elife*. 2015;4: e08490.
45. Luo E, Eppley JM, Romano AE, Mende DR, DeLong EF. Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J*. 2020;14:1304–15.
46. Touchon M, Moura de Sousa JA, Rocha EP. Embracing the enemy: the diversification of microbial gene repertoires by phage-mediated horizontal gene transfer. *Curr Opin Microbiol*. 2017;38:66–73.
47. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, DeLong EF, et al. Genomic islands and the ecology and evolution of *Prochlorococcus*. *Science*. 2006;311:1768–70.

48. Zhaxybayeva O, Gogarten JP, Charlebois RL, Doolittle WF, Papke RT. Phylogenetic analyses of cyanobacterial genomes: quantification of horizontal gene transfer events. *Genome Res.* 2006;16:1099–108.

49. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature.* 2007;449:83–6.

50. Raytcheva DA, Haase-Pettingell C, Piret JM, King JA. Intracellular assembly of cyanophage Syn5 proceeds through a scaffold-containing procapсид. *J Virol.* 2011;85:2406–15.

51. Flores-Uribe J, Philosof A, Sharon I, Fridman S, Larom S, Béjà O. A novel uncultured marine cyanophage lineage with lysogenic potential linked to a putative marine *Synechococcus* 'relic' prophage. *Environ Microbiol Rep.* 2019;11:598–604.

52. Malmstrom RR, Rodrigue S, Huang KH, Kelly L, Kern SE, Thompson A, et al. Ecology of uncultured *Prochlorococcus* clades revealed through single-cell genomics and biogeographic analysis. *ISME J.* 2013;7:184–98.

53. Kiro R, Shitrit D, Qimron U. Efficient engineering of a bacteriophage genome using the type I-E CRISPR-Cas system. *RNA Biol.* 2014;11:42–4.

54. Sarkis GJ, Jacobs WR, Hatfull GF. L5 luciferase reporter mycobacteriophages: a sensitive tool for the detection and assay of live mycobacteria. *Mol Microbiol.* 1995;15:1055–67.

55. Tanji Y, Furukawa C, Na SH, Hijikata T, Miyanaga K, Unno H. *Escherichia coli* detection by GFP-labeled lysozyme-inactivated T4 bacteriophage. *J Biotechnol.* 2004;114:11–20.

56. Brahamsha B. A genetic manipulation system for oceanic cyanobacteria of the genus *Synechococcus*. *Appl Environ Microbiol.* 1996;62:1747–51.

57. Mahichi F, Synnott AJ, Yamamichi K, Osada T, Tanji Y. Site-specific recombination of T2 phage using IP008 long tail fiber genes provides a targeted method for expanding host range while retaining lytic activity. *FEMS Microbiol Lett.* 2009;295:211–7.

58. Le S, He X, Tan Y, Huang G, Zhang L, Lux R, et al. Mapping the tail fiber as the receptor binding protein responsible for differential host specificity of *Pseudomonas aeruginosa* bacteriophages PaP1 and JG004. *PLoS One.* 2013;8:e68562.

59. Murphy KC. Phage recombinases and their applications. *Adv Virus Res.* 2012;83:367–414.

60. Bujarski JJ. Recombination of viruses. In: *Encyclopedia of Virology*. Elsevier; 1999. p. 1446–53.

61. Pires DP, Cleto S, Sillankorva S, Azeredo J, Lu TK. Genetically engineered phages: a review of advances over the last decade. *Microbiol Mol Biol Rev.* 2016;80:523–43.

62. Qimron U, Marintcheva B, Tabor S, Richardson CC. Genomewide screens for *Escherichia coli* genes affecting growth of T7 bacteriophage. *Proc Natl Acad Sci USA.* 2006;103:19039–44.

63. Selick HE, Kreuzer KN, Alberts BM. The bacteriophage T4 insertion/substitution vector system. A method for introducing site-specific mutations into the virus chromosome. *J Biol Chem.* 1988;263:11336–47.

64. Martel B, Moineau S. CRISPR-Cas: an efficient tool for genome engineering of virulent bacteriophages. *Nucleic Acids Res.* 2014;42:9504–13.

65. Dale JW, Greenaway PJ. Identification of recombinant phages by plaque hybridization. In: Walker JM, editor. *Nucleic Acids*. Totowa, NJ: Humana Press; 1984. p. 285–8.

66. Dekel-Bird NP, Avrani S, Sabehi G, Pekarsky I, Marston MF, Kirzner S, et al. Diversity and evolutionary relationships of T7-like podoviruses infecting marine cyanobacteria. *Environ Microbiol.* 2013;15:1476–91.

67. Doron S, Fedida A, Hernández-Prieto MA, Sabehi G, Karunker I, Stazic D, et al. Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J.* 2016;10:1437–55.

68. Krupovic M, Forterre P. Single-stranded DNA viruses employ a variety of mechanisms for integration into host genomes. *Ann NY Acad Sci.* 2015;1341:41–53.

69. Morris RM, Cain KR, Hvorecny KL, Kollman JM. Lysogenic host–virus interactions in SAR11 marine bacteria. *Nat Microbiol.* 2020;5:1011–5.

70. Martínez-García E, Jatsenko T, Kivisaar M, de Lorenzo V. Freeing *Pseudomonas putida* KT2440 of its proviral load strengthens endurance to environmental stresses. *Environ Microbiol.* 2015;17:76–90.

71. Mruwat N, Carlson MCG, Goldin S, Ribalet F, Kirzner S, Hulata Y, et al. A single-cell polony method reveals low levels of infected *Prochlorococcus* in oligotrophic waters despite high cyanophage abundances. *ISME J.* 2021;15:41–54.

72. Oppenheim AB, Kobiler O, Stavans J, Court DL, Adhya S. Switches in bacteriophage lambda development. *Annu Rev Genet.* 2005;39:409–29.

73. Ray U, Sakalka A. Lysogenization of *Escherichia coli* by bacteriophage Lambda: complementary activity of the host's DNA polymerase I and ligase and bacteriophage replication proteins Q and P. *J Virol.* 1976;18:511–7.

74. Greengrass E. Resistance of marine *Synechococcus* to podovirus infection: genetic basis and phenotypic characterization. M.Sc. thesis. Technion - Israel Inst Technol. 2013.

75. Fedida A, Lindell D. Two *Synechococcus* genes, two different effects on cyanophage infection. *Viruses.* 2017;9:136.

76. Shao Q, Trinh JT, McIntosh CS, Christenson B, Balázs G, Zeng L. Lysis-lysogeny coexistence: prophage integration during lytic development. *Microbiol Open.* 2017;6:e00395.

77. Chen F, Lu J. Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol.* 2002;68:2589–94.

78. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J.* 2016;10:437–49.

79. Pratama AA, van Elsas JD. The 'neglected' soil virome—potential role and impact. *Trends Microbiol.* 2018;26:649–62.

80. Srinivasiah S, Bhavsar J, Thapar K, Liles M, Schoenfeld T, Wommack KE. Phages across the biosphere: contrasts of viruses in soil and aquatic environments. *Res Microbiol.* 2008;159:349–57.

81. Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. *Nature.* 2013;494:357–60.

82. Zhao Y, Qin F, Zhang R, Giovannoni SJ, Zhang Z, Sun J, et al. Pelagiphages in the Podoviridae family integrate into host genomes. *Environ Microbiol.* 2018;21:1989–2001.

83. Kashtan N, Roggensack SE, Berta-Thompson JW, Grinberg M, Stepanauskas R, Chisholm SW. Fundamental differences in diversity and genomic population structure between Atlantic and Pacific *Prochlorococcus*. *ISME J.* 2017;11:1997–2011.

84. Berube PM, Biller SJ, Hackl T, Hogle SL, Satinsky BM, Becker JW, et al. Single cell genomes of *Prochlorococcus*, *Synechococcus*, and sympatric microbes from diverse marine environments. *Sci Data.* 2018;5:1–11.

85. Wyman M, Gregory RPF, Carr NG. Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science.* 1985;230:818–20.

86. Lindell D, Padan E, Post AF. Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol.* 1998;180:1878–86.

87. Moore LR, Coe A, Zinser ER, Saito MA, Sullivan MB, Lindell D, et al. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Methods.* 2007;5:353–62.

88. Lindell D. The genus *Prochlorococcus*, phylum cyanobacteria. In: *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer-Verlag Berlin Heidelberg; 2014. p. 829–45.

89. Morris JJ, Kirkegaard R, Szul MJ, Johnson ZI, Zinser ER. Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper" heterotrophic bacteria. *Appl Environ Microbiol.* 2008;74:4530–4.

90. Wolk CP, Fan Q, Zhou R, Huang G, Lechno-Yossef S, Kuritz T, et al. Paired cloning vectors for complementation of mutations in the cyanobacterium *Anabaena* sp. strain PCC 7120. *Arch Microbiol.* 2007;188:551–63.

91. Bryksin A, Matsumura I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques.* 2010;48:463–5.

92. Simon R, Priefer U, Pühler A. A broad host range mobilization system for in vivo genetic engineering: Transposon mutagenesis in gram negative bacteria. *Nat Biotechnol.* 1983;1:784–91.

93. Henn MR, Sullivan MB, Stange-Thomann N, Osburne MS, Berlin AM, Kelly L, et al. Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLoS One.* 2010;5:e9083.

94. Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, Scanlan DJ, et al. *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl Environ Microbiol.* 2006;72:723–32.

95. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature.* 2005;438:86–9.

96. Mita RD, Church GM. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Res.* 1999;27:e34.

97. Schwartz DA, Lindell D. Genetic hurdles limit the arms race between *Prochlorococcus* and the T7-like podoviruses infecting them. *ISME J.* 2017;11:1836–51.

98. Bull JJ, Badgett MR, Wichman HA, Hulsenbeck JP, Hillis DM, Gulati A, et al. Exceptional convergent evolution in a virus. *Genetics.* 1997;147:1497–507.

99. Quinlan AR. BEDTools: the Swiss-army Tool for genome feature analysis. *Curr Protoc Bioinform.* 2014;47:1–34. 11.12.

100. Biller SJ, Berube PM, Dooley K, Williams M, Satinsky BM, Hackl T, et al. Marine microbial metagenomes sampled across space and time. *Sci Data.* 2018;5:180176.

101. Haro-Moreno JM, López-Pérez M, de la Torre JR, Picazo A, Camacho A, Rodriguez-Valera F. Fine metagenomic profile of the Mediterranean stratified

and mixed water columns revealed by assembly and recruitment. *Microbiome*. 2018;6:128.

102. Philosof A, Yutin N, Flores-Uribe J, Sharon I, Koonin EV, Béjà O. Novel abundant oceanic viruses of uncultured marine group II Euryarchaeota. *Curr Biol*. 2017;27:1362–8.

103. Wilson ST, Aylward FO, Ribalet F, Barone B, Casey JR, Connell PE, et al. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocospaera*. *Nat Microbiol*. 2017;2:17118.

104. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. *Science*. 2015;348:1261359.

105. Bushnell B. BBMap: a fast, accurate, splice-aware aligner. 2014;URL: <https://www.osti.gov/servlets/purl/1241166>.

106. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;1303:3997.

107. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.

ACKNOWLEDGEMENTS

We thank Yotam Hulata for his assistance with cell-sorting, Claudia Steglich for advice on the choice of cyanobacterial regulatory elements for construction of the *proCAT* gene cassette, Paul M. Berube, Jessie W. Berta-Thompson and Libusha Kelly for their guidance in using the curated BigRAPA database and Lindell lab members for helpful discussions. We also thank Gal Harel for her assistance with graphic design of the figures. This work was funded by the European Research Council (Grant No. ERC-CoG 646868) and the Simons Foundation SCOPE Awards (Grant Nos. 329108 and 721254) to DL. It was also supported in part by the Simons Foundation (Life Sciences Project Award IDs 337262, 509034SCFY17, 647135, and SCOPE Award ID 329108) and the US National Science Foundation NSF-EDGE grant (1645061) to SWC. This manuscript is a contribution of the Simons Foundation Collaboration on Ocean Processes and Ecology (SCOPE).

AUTHOR CONTRIBUTIONS

DS designed and developed the cyanophage genetic engineering system and performed experiments to characterize wild-type and mutant S-TIP37 cyanophage strains. TH performed metagenomic analyses. RL and NR performed experiments with cyanophage P-SSP7. MCGC performed the iPolony infected cell experiment. DAS

designed and GS constructed and tested the pDS-proCAT plasmid. DL participated in the design of the genetic engineering system, characterization of S-TIP37 cyanophage mutants and pDS-proCAT plasmid design. SWC coordinated the metagenomics analyses and experimentation with P-SSP7. DS and DL wrote the manuscript with contributions from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41396-021-01085-8>.

Correspondence and requests for materials should be addressed to D.L.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021