

ARTICLE

<https://doi.org/10.1038/s41467-019-09106-z>

OPEN

# Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes

Ella T. Sieradzki<sup>1</sup>, J. Cesar Ignacio-Espinoza<sup>1</sup>, David M. Needham<sup>1</sup>, Erin B. Fichot<sup>1</sup> & Jed A. Fuhrman<sup>1</sup>

Viruses provide top-down control on microbial communities, yet their direct study in natural environments was hindered by culture limitations. The advance of bioinformatics enables cultivation-independent study of viruses. Many studies assemble new viral genomes and study viral diversity using marker genes from free viruses. Here we use cellular metatranscriptomics to study active community-wide viral infections. Recruitment to viral contigs allows tracking infection dynamics over time and space. Our assemblies represent viral populations, but appear biased towards low diversity viral taxa. Tracking relatives of published T4-like cyanophages and pelagiphages reveals high genomic continuity. We determine potential hosts by matching dynamics of infection with abundance of particular microbial taxa. Finally, we quantify the relative contribution of cyanobacteria and viruses to photosystem-II *psbA* (reaction center) expression in our study sites. We show sometimes >50% of all cyanobacterial+viral *psbA* expression is of viral origin, highlighting the contribution of viruses to photosynthesis and oxygen production.

<sup>1</sup>Department of Biological Sciences, University of Southern California, 3616 Trousdale Pkwy, Los Angeles 90089 CA, USA. Correspondence and requests for materials should be addressed to J.A.F. (email: [Fuhrman@usc.edu](mailto:Fuhrman@usc.edu))

Marine viruses are an important component of biogeochemical cycles in the ocean, both as a top-down control on microbial populations<sup>1–5</sup> and as converters of particulate organic carbon to dissolved organic carbon by cell lysis, termed the viral shunt<sup>4,6,7</sup>. Viruses that infect bacteria, or phages, have also been shown to promote genomic diversity in their hosts<sup>8–12</sup>. Until recently, the bulk of the research on marine viruses focused on phages that infect cyanobacteria (cyanophages) that include T4-like myoviruses, T7-like podoviruses, and siphoviruses<sup>13–19</sup>. However, cultivation of hosts and mining of metagenomic and metaviromic datasets revealed abundant phages infecting heterotrophic hosts<sup>20–25</sup>, some of which even have demonstrated diel cycles<sup>26</sup>.

The lack of a universal marker gene for phages, in concert with extremely limited reference sequences with which to study viral metagenomes, made it much more difficult to study them compared to bacteria and archaea; thus, many studies instead focused on general enumeration and estimation of late-infection rates via microscopy, tracking phages with cultivable hosts, or studying marker genes within major phage groups (e.g. myoviruses using the *g20* or *g23* genes)<sup>27,28</sup>. While the addition of more cultured phage genomes to public databases enhances our understanding of these fascinating organisms<sup>29</sup>, comprehensive analysis would ultimately require more knowledge on culture conditions for hosts, and the actual cultivation of hundreds to thousands of potential hosts, all very challenging. The rise of next-generation sequencing and metagenomics made the study of environmental phages more accessible<sup>6,26,30–32</sup>. In recent years new techniques for the study of viruses, such as vSAGs, polonies, and viral-BONCAT, were introduced, and promise to move the field further into the realm of culture-independent research<sup>31,33–35</sup>.

Most cyanomyoviruses and some cyanopodoviruses have been shown to carry horizontally transferred auxiliary metabolic genes (AMGs) that presumably maintain the photosynthetic machinery functional during infection, and may even divert energy into nucleotide production instead of carbon fixation<sup>36–43</sup>. Viral *psbA* genes coding for Photosystem II (PS-II) photosynthetic reaction center protein D1 have been shown to be widespread and discernible from the host versions<sup>40,44</sup>. This protein has a short lifetime, and if not replenished, PS-II function fails<sup>37,41,45</sup>. Because viruses generally shut down synthesis of new mRNA coded by the host genome, expression of the viral version of this gene is required to maintain photosynthesis, and it is expressed almost throughout the cyanophage latent period<sup>8,40</sup>.

We have recently become particularly interested in direct evidence of active viral infections. During lytic infection of bacterial hosts, viral genes are generally thought to be expressed as polycistronic mRNA corresponding to sequential phases of infection<sup>46–49</sup>. Thus, active viral infections in mixed natural communities could be tracked by assembling and characterizing multigene transcripts from cellular microbial metatranscriptomes. While some such transcripts may represent host operons or other vectors, recently developed methods can reasonably well identify sequences of viral origin<sup>50,51</sup>. Here we use that approach to characterize active viral infections via marine metatranscriptomic and metagenomic analyses and show temporal and spatial patterns in the diversity and activity of such infections.

Here we study surface seawater from different seasons over three sites across the San Pedro Channel, Southern California, USA: The Port of Los Angeles (POLA), Santa Catalina Island Two Harbors (CAT), and the San Pedro Ocean Time-series (SPOT). These sites are within a transect of 37 km and represent a gradient of human impact with POLA being the most impacted and SPOT resembling open ocean conditions<sup>52</sup>. At all three sites, free virus-like particles, counted in whole seawater, outnumber

bacteria and archaea roughly 10:1 (Supplementary Figure 1). For metagenomic and metatranscriptomic analyses, we examine only the 0.2–1  $\mu\text{m}$  size-fraction, which includes most free-living bacteria, archaea, and some picoeukaryotes.

We show that most assembled viruses in our dataset appear only ephemerally, have no close relatives and most of them represent phages infecting heterotrophic bacteria. Read recruitment to curated cyanophage genomes indicates presence of many close relatives creating a pattern of genomic continuity. Host and phage spatiotemporal dynamics are used to match an assembled cyanophage with a *Synechococcus* strain. Finally, we reveal a high contribution of cyanophages to gene abundance and expression of photosystem-II.

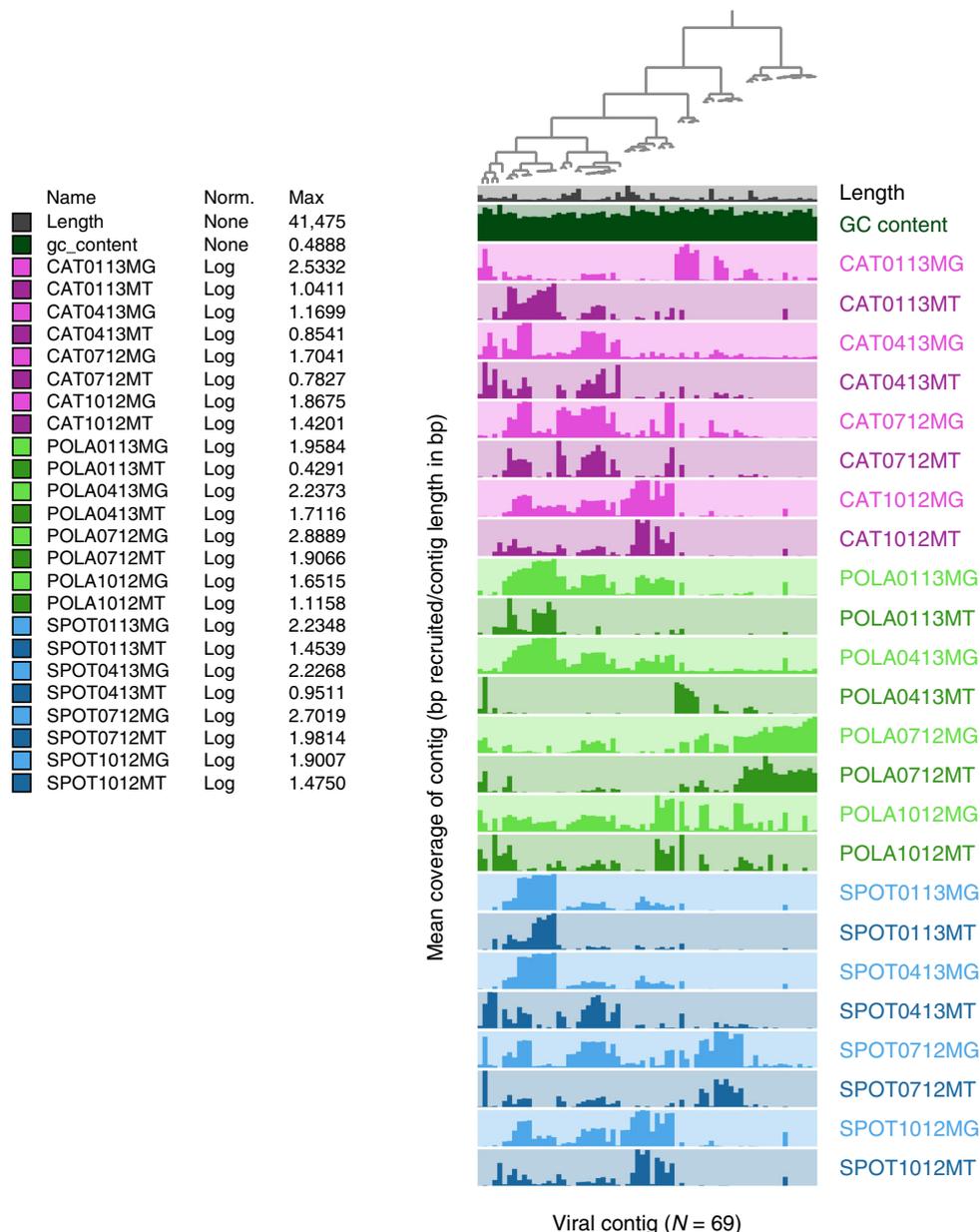
## Results

**Spatiotemporal patterns of active infections.** Via assembly of metatranscriptomes, we obtained 1455 contigs longer than 5 kbp, of which 61 (3.9%) were characterized as viral using Virsorter and VirFinder (see methods and Supplementary Data 1). Additionally, a cross-assembly of the metatranscriptomic viral contigs with metagenomes of the same samples ( $n = 12$ ) yielded nine more contigs (mean length 26,563 bp) characterized as viral. Most of the contigs represent dsDNA viruses ( $n = 68$ ) as apparent from their presence in metagenomes (evaluated by read recruitment, see methods), but one appeared to be an RNA virus possibly infecting a eukaryotic host. This contig contained an RNA-dependent-RNA-polymerase whose nearest match in NCBI non-redundant database (NR) was marine Antarctic phytoplankton RNA virus PAL\_E4<sup>53</sup>.

These 69 viral contigs revealed varied patterns of presence (in metagenomes) and activity (in metatranscriptomes) in the three sites over a year (Fig. 1). Some regional patterns were evident, e.g. some viral contigs were unique to the Port of LA (Fig. 1), and POLA always differed from the other sites in expression of viral contigs (by Bray–Curtis dissimilarity) more than SPOT and CAT differed from each other (Supplementary Figure 2B). This pattern corresponds to the difference in microbial abundance and heterotrophic production between the port and the other sites (Supplementary Figure 1), and to patterns of microbial community composition and activity by amplicon single variants of 16S-rRNA (ASV, Supplementary Figure 2C, D).

We did note that for most sites and dates, the recruitment patterns in metatranscriptomes (presumably mRNA from viruses within the latent period) very generally reflected the metagenomic recruitment (presumably DNA in late infection and assembled viruses, when the highest amount of viral DNA is present in cells), with notable exceptions of POLA and SPOT in April and CAT in January, where the most transcriptionally active contigs had relatively low DNA as shown by recruitment in the metagenomes (Fig. 1).

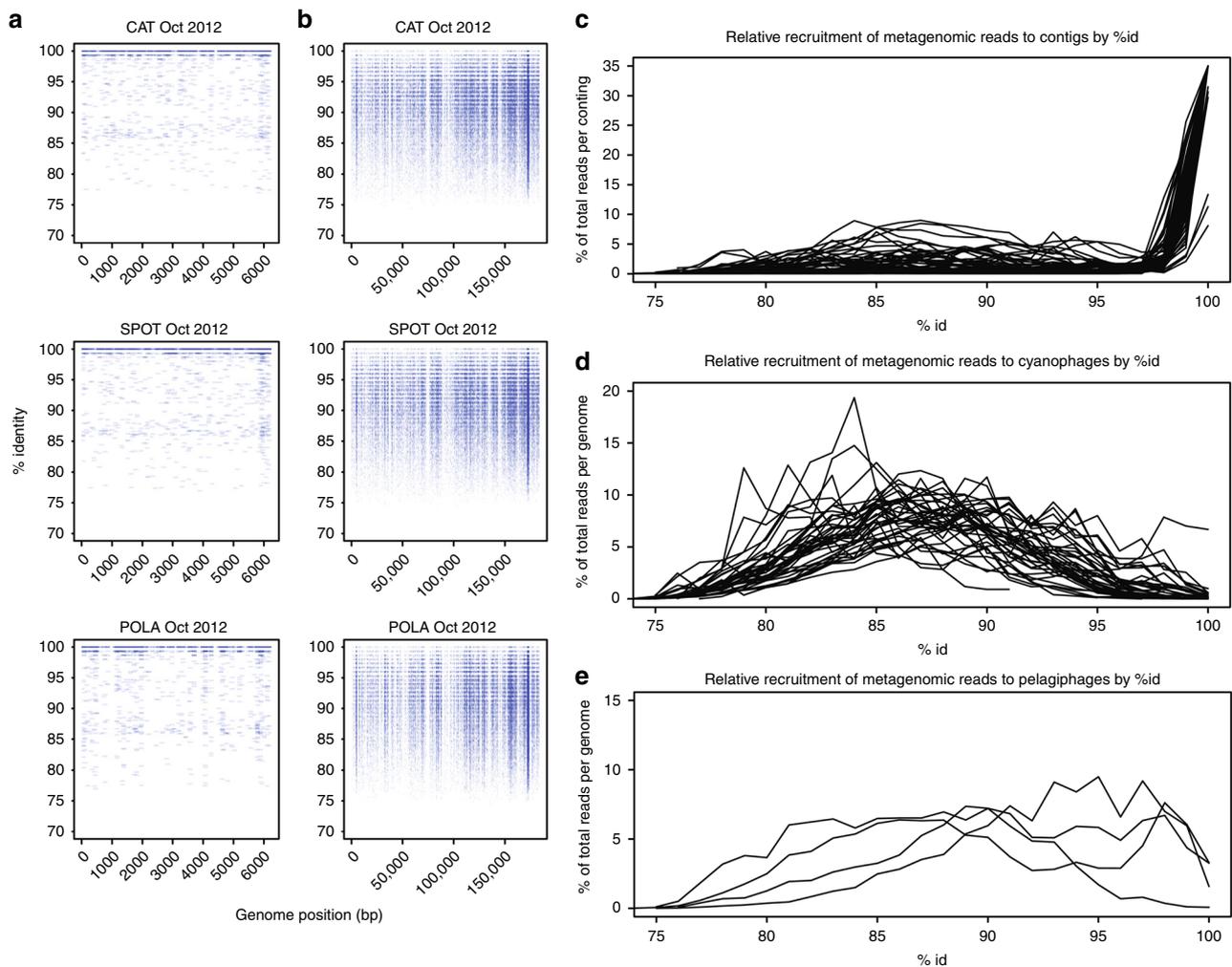
Persistent infections (mean coverage  $\geq 0.75\times$  in at least three out of four metatranscriptomes per site), comprising seven out of 69 contigs, were found mostly at CAT and SPOT (Supplementary Data 1). Two contigs were persistent in all three sites, and none were persistent only at POLA. Non-persistent, i.e. ephemeral, infections dominated the assembled landscape, as 62 out of 69 of the contigs only appeared in 0–2 out of four dates per site. Bray–Curtis dissimilarity of the relative abundance of viral contigs presence (DNA) and expression (RNA) within each site as well as between sites was almost exclusively between 70–100%. High community-level dissimilarity indicates that even within site, most viral infections which we detected as instantaneous snapshots at the time of sampling, were localized and sporadic on the time scale we studied (Supplementary Figure 2A, B).



**Fig. 1** Dynamics of active infections over seasons and sites. Mean coverage, as a proxy of relative abundance, of 69 viral contigs across three sites (Port of LA—POLA, San Pedro Ocean Time-series—SPOT and Two Harbors—CAT) and four seasons (July 2012, October 2012, January 2013, and April 2013) in metagenomes (MG) and metatranscriptomes (MT). Each column represents one contig and each row represents one sample. The bar heights are relative to the highest log-transformed mean coverage within each sample, detailed in the legend column Max. Mean coverage, based on bowtie2 mapping, was calculated excluding contig positions whose coverage depth was in the top 25% of coverage, as they may be biased by non-specific recruitment localized to short, highly conserved portions of the contig (Supplementary Figure 3). The dendrogram denotes hierarchical clustering of the contigs by abundance<sup>79</sup>

**Genomic diversity of actively infecting phages.** To investigate the diversity of phages in our system, we mapped reads at varying identity percentages to the assembled active viral contigs as well as to published, curated cyanophage and pelagiphage genomes (see Supplementary Data 1 for accession numbers). The assembled viral contigs appeared to be biased toward low population-level microdiversity (i.e. more clonal) viruses (Fig. 2a, c). All the viral contigs we assembled in this study appear to have many nearly identical relatives but few moderately close ones as shown by recruitment plots (most recruitment at 98–100% identity and little recruitment at 90–97%, Fig. 2c). Read recruitment to published genomes of cultured pelagiphages was continuous along most of the genome (similar to Fig. 2a) at up to 100% identity (Fig. 2e) with high mean coverage (Supplementary Data 1), yet we

did not assemble any moderately-complete pelagiphage genome as determined by ORF annotations (Supplementary Data 2). In contrast, the published cyanophage genomes had little or no exact read recruitment in our samples, but rather recruitment centering broadly around 85–90% identity, suggesting cyanophages in our region are related to the cultured ones but are not identical (Fig. 2d). In fact, only three of our assembled contigs were recognizable as putative cyanophages, despite the ease in identifying cyanophages by similarity to known genomes or the presence of photosynthesis-related genes. Only one contig represented a potential partial pelagiphage genome. The low rate of success in assembling these phage types was surprising considering that their hosts—cyanobacteria and the SAR11 clade—represented a significant fraction of the microbial community in



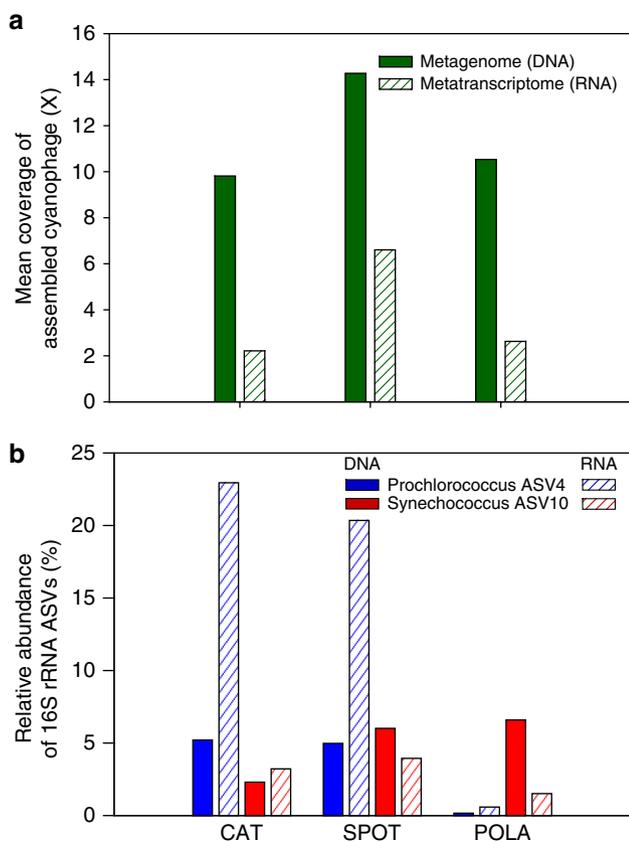
**Fig. 2** Genomic diversity of assembled and curated phages. Metagenomic read recruitment (minimum 100 bp hit length) from October samples (single sample per site) to **a** an assembled cyanophage contig, and **b** *Prochlorococcus* phage P-HM2 genome. The highest recruitment to the assembled contig in **a** was at 99–100% identity (high density near 100% is not fully evident from the graph due to overlaps, see **c**), whereas P-HM2 revealed a genomic continuum ranging from 80% to ~98% identity. **c–e** Recruitment as a function of percent identity of reads to assembled contigs (**c**), curated cyanophages (**d**) and curated pelagiphages (**e**) demonstrated that assembled contigs mostly recruited at 100% ID and had few moderately close relatives (**c**) whereas published reference genomes of 45 cyanophages revealed moderately close relatives of those genomes but few matches near 100% (**d**), and four pelagiphages had continuous recruitment implying a much less discrete spectrum of diversity (**e**). See Supplementary Data 1 for accession numbers of reference genomes

all sites at any given time, although more so at SPOT and CAT (Supplementary Figure 4). We recognize that the recruitment to contigs depends on how successful we were at assembling viral contigs, and to see if we could obtain contigs of cyanophage or pelagiphage by more sophisticated assembly methods, we also tried applying a subsampling-based assembly approach (see methods) on two samples: SPOT from October, which had the highest relative abundance of cyanobacteria, and POLA from April, which had the highest relative abundance of viral *psbA* (see below). While yielding three additional abundant heterotrophic phages, this method did not yield additional contigs containing any cyanophage marker genes (Fig. 1, Supplementary Data 3).

A previous report indicated that *Synechococcus* phage genomes occur in discrete clouds with a discontinuity in recruitment below ~95% identity<sup>54</sup>. While this pattern existed for some cyanophage reference genomes, having gaps in coverage at ~90–95% consistent with that idea, it was by no means the rule in our data for either those genomes or for the reference pelagiphage genomes (Figs. 2b, d, e).

Finally, the recruitment plots revealed a common pattern of high coverage of short regions within a genome or contig at up to 100% identity, whereas the rest of the genome or contig was only recruited at lower percent ID if at all (examples in Supplementary Figure 3); however some samples showed recruitment all along the same contig or genome (Supplementary Figure 3).

**T4-like cyanomyoviruses.** Few of the assembled viral contigs contained the myoviral marker gene *g23* (major capsid protein), which was identified using an HMM (Hidden Markov Model) search of the predicted open reading frames (ORFs) (Supplementary Data 2). In order to determine whether cyanomyoviruses were actively infecting and simply did not assemble, we examined the unassembled reads, and assigned translated reads identified by the same Gp23-HMM to published protein sequences and assembled Gp23 ORFs of cyanomyoviruses and myoviruses infecting heterotrophic bacteria (see Supplementary Table 1 for accession numbers and Supplementary Figure 5 for a maximum likelihood tree of Gp23 proteins). Only two Gp23 ORFs from



**Fig. 3** Matching assembled cyanophage with *Synechococcus* host using expression patterns. Presence and activity of the assembled cyanophage and its potential hosts in October 2012: **a** Mean coverage (after removing the top 25% most abundant contig positions) of the assembled cyanophage, **b** Amplicon single variants (ASVs) relative gene and transcript abundance of 16S-rRNA of the two most abundant cyanobacterial ASVs: *Prochlorococcus* ASV4 and *Synechococcus* ASV10. Note the near-absence of *Prochlorococcus* in POLA, in contrast to *Synechococcus* and the phage RNA, leading us to infer that the phage infected *Synechococcus*. Each group of bar plots, denoted by site name, represents data from a single sample

contigs cross-assembled with metagenomes were placed within cyanomyoviral clades. One had no recruitment in metatranscriptomes, implying that it was not actively infecting its host, and the other was persistent at SPOT but with low coverage (contigs 13 and 15 in Supplementary Data 1). Recruitment to Gp23 was generally higher at POLA except in January, and most of it was attributed to T4-like heterophages (phages infecting heterotrophic bacteria) rather than cyanophages. Published reference cyanophage proteins accounted for about 10–20% of total Gp23 transcripts, and there was minimal recruitment to assembled putative cyanomyoviruses, most of it at POLA in July (Supplementary Figure 6). Thus, it appears that cyanomyovirus genomes did not assemble.

**Cyanophage host matching.** One assembled contig that represented with high certainty, a partial cyanophage genome, was a putative cyanopodovirus (T7-like). This contig contained genes coding for photosystem-II protein D1 (*psbA*) and high-light induced protein (*hli*) that are reportedly widespread in cyanophages<sup>28</sup>. The putative cyanophage represented by this contig was actively transcribed (presumably infecting its host) in all three sites only in October 2012 (Fig. 3a). The cyanobacterial community by 16S-rRNA was dominated in October by two single amplicon variants (ASVs): one *Synechococcus* and one

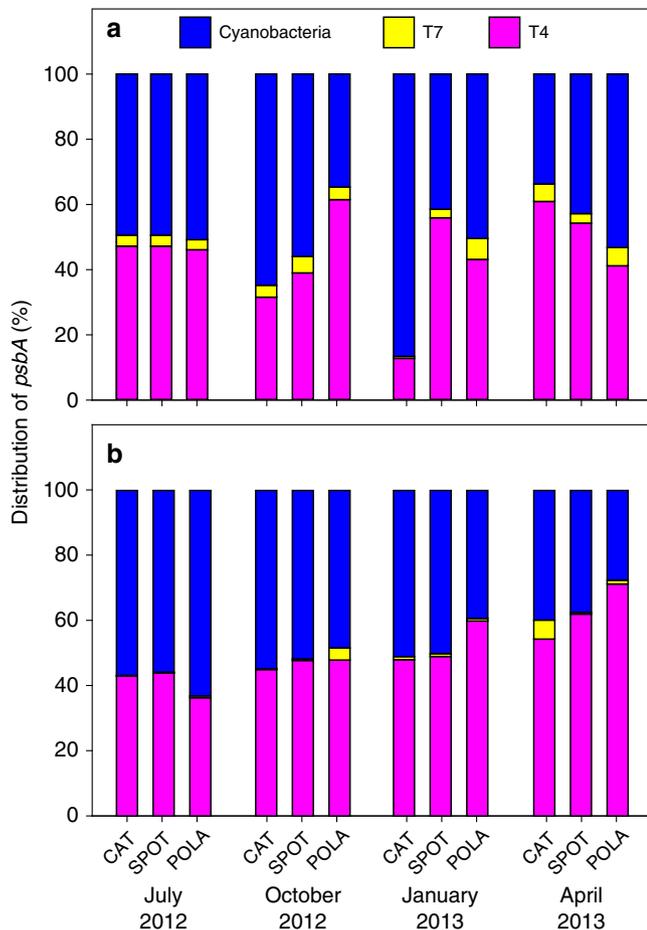
*Prochlorococcus*. Both ASVs were present at SPOT and CAT in October, but only *Synechococcus* was present at POLA (Fig. 3b). Thus, we propose that this assembled contig represents a phage that infects *Synechococcus* ASV 10, which is 100% identical over the V4-V5 hypervariable region of 16S-rRNA to *Synechococcus* CC9902 of clade IV. On a phylogenetic tree of PS-II D1 proteins, the translated PS-II D1 of this phage clustered closely with a different cyanopodovirus isolated on *Synechococcus* (Supplementary Figure 7).

**Expression of viral and bacterial *psbA*.** The presence of *psbA* within the assembled cyanophage prompted us to survey expression of this gene in unassembled reads, which revealed comparable expression by cyanophages and cyanobacteria even as the relative abundance of *psbA* out of the entire metatranscriptome varied between samples (Supplementary Figure 8). Sharon et al.<sup>44</sup> previously showed that viral *psbA* gene variants can outnumber cyanobacterial *psbA* genes in metagenomes from the Mediterranean, and that viral gene expression is evident. We extended this to quantitatively partition gene expression into bacterial contribution from *Synechococcus* and *Prochlorococcus* and viral contribution from cyanomyoviruses and cyanopodoviruses, as evident from placing translated reads onto our PS-II D1 phylogenetic tree (Supplementary Figure 7). We found that *psbA* transcripts of T4-like cyanomyovirus origin generally accounted for roughly 50% ( $51 \pm 10\%$ ) of cyanobacterial and cyanophage *psbA*, at a ratio of  $1.2 \pm 0.6$  (mean  $\pm$  standard deviation) (Fig. 4b). On several occasions, the viral variant exceeded the cyanobacterial variant in recruited read counts (Fig. 4).

In both metagenomes and metatranscriptomes, there was minor consistent recruitment to T7-like cyanopodovirus *psbA*. However, in every sample the contribution of T7-like cyanopodoviruses was very low compared to that of T4-like cyanomyoviruses.

## Discussion

Our determination of persistent vs. ephemeral infections depended on the ability to track read recruitment to contigs, hence it refers only to those viruses represented in contigs. While most of the active infections by viruses we assembled here were ephemeral, the minority that were persistent appeared in the more oligotrophic sites (CAT and SPOT) and no virus was persistently infecting its host at the heavily impacted POLA. This spatial distinction was probably due to varying anthropogenic and riverine inputs at POLA, which do not affect the other sites and can cause large, stochastic shifts in the microbial community of the port leading to differences in phage communities. Moniruzzaman et al.<sup>55</sup> also recently demonstrated dominance of ephemeral dynamics in infections of marine single-cell eukaryotes during an algal bloom, which is a drastically dynamic system. Conversely, Aylward et al. showed the occurrence of many persistent viruses in the relatively stable system of the Hawaii Ocean Time-series<sup>26</sup>. As the San Pedro Channel is a dynamic, seasonal system<sup>56</sup>, it is not surprising that infections varied somewhat over time. An important consideration in determining persistence, which relies on read recruitment to contigs, is the extent that assembled contigs represent the entire viral community. Our results suggest that many viruses may be missed because they are not well represented in the contigs we were able to generate. Recruitment of reads to our assembled viral contigs and to published genomes of cyanophages and pelagiphages indicated that the assemblies originated from viruses with low microdiversity, whereas pelagiphages and cyanophages revealed genomic continuity reflecting high microdiversity. Due to high variation interfering with creation of contigs, this microdiversity probably hindered contig



**Fig. 4** Phage and cyanobacteria expression of the PS-II reaction center. Relative abundance of *psbA* of cyanobacteria (*Synechococcus* and *Prochlorococcus*), T4-like cyanophages, and T7-like cyanophages out of total *psbA* in **a** metagenomes and **b** metatranscriptomes. Each stacked bar chart represents a single sample

assembly from local representatives of those groups<sup>33</sup>. Sampling time of day may have been another significant hurdle for cyanophages (see below). Regardless of the reason, lack of assembled cyanophages and pelagiphages may have contributed to the low fraction of traceable persistent infections out of all active infections, as cyanobacteria and the SAR11 clade consistently make up a significant portion of the prokaryotic community at SPOT and CAT, and therefore their phages are likely to be persistently infecting them as well. Many recruitment plots contained a pattern of coverage of a short sequence across a nucleotide identity range of 70–100%. This pattern highlights two issues: (1) some genes are so conserved or so often laterally transferred that their partial sequences cannot be used to identify whether a certain phage is present and (2) that mean coverage of contigs could be highly biased by these conserved regions, which needs to be considered when evaluating abundance of the contigs and for coverage-based binning of genomes. As an example of this consideration, we ignored the 25% most highly-recruited positions within each contig when calculating the contig mean coverage. We also note that widely-used recruitment algorithms (e.g. bowtie, bwa) only map reads with a local or end-to-end match at a very high percent identity and would therefore miss moderately close relatives to the query sequence that may be relevant to questions about phage ecology. We were surprised not to find multiple and abundant cyanomyoviral contigs, because such cyanophages are reported to be some of the most common

dsDNA viruses in the ocean<sup>57</sup>. T4-like phages can comprise up to 25% of total virus-like particles<sup>58</sup> and the majority of total plaque-forming cyanophages<sup>59</sup>. Additionally, we know that the San Pedro Channel supports a diverse community of myoviruses and cyanobacteria<sup>27,60</sup>. Some of the problem could be high microdiversity ‘breaking’ the assemblies as noted above. That being said, there is some evidence suggesting that sampling time of day may be a factor. Cyanophage infections were shown to peak in the late afternoon in the central Pacific Ocean<sup>26</sup>. This is a reasonable mechanism considering that energy harvesting in the hosts is performed during the day, and then potentially exploited by the cyanophages<sup>39,40,43,61</sup>. This idea becomes even more plausible when discussing phages that carry photosystem genes, as there would be no fitness advantage to keeping those genes in the viral genome if infection did not take place in daylight. These genes are widely distributed in cyanomyoviruses<sup>62</sup>, implying that in those phages the latent period would be especially likely to coincide with daylight. As our samples were collected early in the morning, it is possible that we were not able to assemble genomes of cyanomyoviruses from the metatranscriptomes because they were in the early phase of the lytic cycle and most of the genomes were not yet transcribed. Low recruitment to T4-like cyanomyovirus Gp23 protein compared to T4-like heterophages indicates that at the time of sampling few cyanophages were at the late phase of infection in which capsid protein genes are expressed.

Another theory on the timing of infection by cyanomyoviruses is that late infection coincides with the S/G2 phases in the host cell cycle, in which the host genome was already replicated but the cell hasn’t divided yet, the logic being that there would be double the amount of nucleotides available for synthesis of phage genomes<sup>63</sup>. This is indeed the case in the T4 phage of *Escherichia coli*<sup>64</sup>. Evidence from cultures<sup>65,66</sup> and from environmental samples<sup>67</sup> indicate that for at least some strains of *Prochlorococcus* the host cells divide at dawn, which would imply that the bulk of phage transcription occurs during the night, and may also explain why we were unable to assemble cyanomyovirus genomes from our metatranscriptomes.

Similar theoretical arguments can be made regarding pelagiphages, which depend on light-energy harvesting using proteorhodopsin<sup>68,69</sup>. Indeed, at least one pelagiphage was previously shown to have diel variation of transcription<sup>26</sup>. In an attempt to circumvent the issue with cyanophage and pelagiphage assembly we recruited unassembled metagenomic and metatranscriptomics reads to curated cyanophage and pelagiphage genomes found in the RefSeq database. While the exact published genomes themselves were not present in our samples, we posit that other T4-like cyanophages closely-related to those published were present and persistently infecting their hosts. The broad recruitment covering a wide range of sequence identity to those genomes implies high genomic diversity of these phages. With proper sampling timing, and as assembly methods continue to improve, assembly of organisms with high-microdiversity may become more computationally manageable over multiple samples, and might reveal the exact cyanophage genomes present in samples like ours. Still, the pattern of a genomic continuum in cyanophages and pelagiphages, i.e. with few breakpoints or gaps in genomic identity between many close relatives, should remain relevant and informative.

Matching viral contigs and hosts is an ongoing challenge, but we were able to use genetic information and regional abundance patterns to make a likely match between a cyanopodovirus and a strain of *Synechococcus*. Here we take advantage of presence of the gene *psbA* in the assembled cyanopodovirus contig. This gene was presumably horizontally transferred to viral ancestors of cyanophages from their hosts. Despite its history of horizontal transfers, *psbA* is considered a good indicator of host genus<sup>19</sup>.

Additionally, cyanopodoviruses tend to have a narrow host range compared to cyanomyoviruses<sup>14,19</sup>. Therefore, it is somewhat easier to use community dynamics to match a host to a podovirus.

Statistical correlations between environmental bacteria and viruses are usually hard to interpret, as both positive and negative time-delayed correlations could indicate a host-phage relationship, depending on time scales. Demonstrating an active infection via transcription simplifies the issue, as an actively infecting phage must have host cells present at the time of sampling.

Many cyanophages contain a variety of genes that maintain photosynthetic activity in the host during infection, from spare parts for photosynthetic reaction centers through regulation and optimization of those apparatus<sup>70</sup>. In particular, viruses were shown to maintain photosystem II function during infection under light conditions in order to maintain continuous supply of energy to the host, as transcription of host genes drops during infection and PS-II proteins have a short lifetime<sup>13,37,39–41,61</sup>. Phages that contain the *psbA* gene probably derive fitness advantages from it compared to phages that do not<sup>34,43</sup>.

We show here that expression of viral *psbA* was comparable to cyanobacterial *psbA* year-round. This information can be used to roughly estimate the proportion of infected cyanobacteria from our *psbA* data and to compare it to previously published infection rates. During host infection, the number of phage mRNA molecules of *psbA* increases quickly early in the latent period of infection and becomes the main source of *psbA* transcripts in the cell<sup>8,39,40,43,61</sup>. Under the assumption that the average *psbA* expression is comparable in infected (viral origin) and uninfected (bacterial origin) cells, we can use the viral expression of *psbA* vs. the bacterial expression to give a rough estimate of the fraction of cyanobacterial cells infected with cyanophages. What we observe in the sample is a comparable expression of T4-like *psbA* and cyanobacterial *psbA*, which suggests that on average roughly half of the cyanobacteria are infected. This is in accordance with the high end of published estimates for marine cyanobacteria<sup>4,57,71,72</sup>, and admittedly is very rough. But it confirms the idea that infection is an important part of cyanobacterial ecology. If indeed the lack of assembled cyanomyovirus genomes resulted from sampling in the early phase of lytic infection, we may even be underestimating viral *psbA* expression.

Gene abundance and expression of viral *psbA* of T4-like origin was always much higher than T7-like *psbA* in our samples. This may represent a real higher abundance of actively infecting but not assembled T4-like cyanophages. *psbA* is expressed throughout the lytic phase, including during early infection, which could account for its high expression despite the lack of cyanophage genome assembly. Other contributing factors may include the more specific host range reported for cyanopodoviruses compared to cyanomyoviruses<sup>14,73,74</sup> or that only clade B of T7-like cyanophages carries the *psbA* gene as opposed to nearly all T4-like cyanophages<sup>19,62,73</sup>.

As no temperate marine T4-like and T7-like cyanophages have been reported, suggesting they are generally lytic<sup>75,76</sup>, presence of their *psbA* genes in our metagenomes may result from viral genomic DNA in cells at the later stages of infection, pseudolysogeny, phages that adsorbed to cells or particles, or any free viruses incidentally caught on the filter. In other systems, lysogeny may account for contig-wide recruitment in metagenomes but not in metatranscriptomes.

Extending metatranscriptomic methods as recently applied to marine eukaryotic<sup>55,77,78</sup> and prokaryotic<sup>26</sup> viral infection, we show the power of multiple approaches to track viral infection and dynamics within the broad picoplankton community, using metatranscriptomes of the cellular fraction, with particular examples in the cyanobacteria. Read recruitment is an excellent

way to track particular viruses, but this requires genomes or large contigs and the vast majority of viruses have not been isolated, nor genomes sequenced. While ephemeral and more clonal viral genomes assemble relatively easily from metatranscriptomes, and thus can be tracked by read recruitment, the more common and persistent viruses with many close relatives assembled poorly. For these, the use of marker genes is especially important. The observed infection dynamics can sometimes be used in combination with microbial community structure and viral marker genes to deduce a host. Finally, use of metagenomes and metatranscriptomes provides an insight into quantifiable viral contribution to photosynthesis.

## Methods

**Sample collection.** Surface seawater was collected by bucket on 15 July 2012, 19 October 2012, 9 January 2013, and 24 April 2013 in three locations: The Port of Los Angeles (33°42.75'N 118°15.55'W), the San Pedro Ocean Time-series (33°33.00'N 118°24.01'W), and Two Harbors, Santa Catalina Island (33°27.18'N 118°28.51'W). Water was collected between 7 am and 12 noon. Duplicate samples of 20 l were filtered in each location through an 80 µm mesh followed by a glass fiber syringe prefilter (Gelman, 4523) which collected the >1 µm size fraction and a 0.2 µm PES Sterivex filter (Millipore, SVGPB1010), which collected the free-living size fraction. The duration of filtration was 15–20 min, immediately after which RNAlater (Thermo-Fisher, AM7020) was added to each filter and filters were capped and flash frozen no more than 5 min post-filtration.

Viral and cellular counts as well as heterotrophic production were measured on the same day. For detailed methods please see Supplementary Methods.

**Library preparation.** RNAlater was removed from all filters prior to nucleic acid extraction with a syringe. DNA and RNA were extracted simultaneously from Sterivex filters by bead-beating using 0.1 mm glass beads added into the sterivex shell, followed by an AllPrep kit (Qiagen, 80204). An internal standard (ERCC RNA Spike-In Mix, Thermo-Fisher 4456740) was added into the lysate after bead-beating for quality assurance. RNA was enriched for mRNA with RiboZero (Illumina, MRZB12424). Resulting mRNA was reverse transcribed using SuperScript-III (Invitrogen, 18080–051). DNA and cDNA were physically sheared with Covaris m2 and size-selected for products larger than 300 bp with magnetic AMPure beads (Beckman-Coulter, A63881) at a ratio of 0.7 beads to product. RNA libraries were prepared and barcoded using NEBNext Ultra Directional RNA library Prep Kit for Illumina (New England Biolabs, E74205). DNA libraries were prepared and bar-coded with Ovation UltraLow Library Prep V2 (Nugen, 0344). Metagenomes were sequenced on Illumina HiSeq 2 × 125 bp or 2 × 150 bp. Metatranscriptomes were sequenced on Illumina HiSeq 2 × 250 bp.

**Read processing and assembly.** Raw metagenomic and metatranscriptomic reads were quality trimmed, and residual ribosomal reads as well as the internal standard were removed informatically. Merged reads from each sample separately were assembled via several methods with or without subsampling (see Supplementary Methods). Only 1455 contigs larger than 5 kbp were further analyzed.

**Identification and annotation of viral contigs.** Viral contigs were identified by running VirSorter version 1.0.3<sup>50</sup> using RefSeq on the CyVerse platform on all contigs >5 kbp and only contigs classified as category 1 or category 2 were considered. VirSorter is a well-established and reliable tool, but it has an inherent database bias and performs better on longer contigs because they are more likely to contain hallmark viral genes. Thus, all assembled contigs >5 kbp (regardless of VirSorter results) were also ranked using VirFinder<sup>51</sup>, which relies on k-mer signatures, and only contigs ranking higher than 0.85 were considered further. ORFs were predicted and annotated within those contigs, and hit taxonomy was used toward assigning a contig as viral or non-viral.

Read mapping to the viral contigs was used to describe temporal and spatial dynamics. For a detailed description please see Supplementary Methods.

**Microbial community composition analysis.** The V4-V5 regions of the 16S-rRNA coding gene were amplified from DNA and cDNA from all samples using the 515(N)-F and 926-R primers, and sequenced on an Illumina MiSeq 2 × 300 bp (UC Davis genome center) along with a negative controls and even and staggered mock communities. Reads were quality-trimmed, assembled and processed through the minimum entropy decomposition (MED) pipeline to produce amplicon single variants (ASVs) and calculate beta diversity (Supplementary Methods).

**Analysis of PS-II D1 protein sequences.** We built a maximum likelihood tree from a set of curated PS-II D1 protein sequences and placed the translated ORFs from our assembled cyanophages in it. The same set was used to build a hidden

Markov model (HMM) of PS-II D1 to which we mapped reads from all metagenomes and metatranscriptomes (see Supplementary Methods).

**Analysis of Gp23 protein sequences.** Metatranscriptomic and metagenomic reads were searched with blastx against a set of T4-like clusters of orthologous groups (COGs) with an E-value threshold of  $10^{-5}$ . Metatranscriptomic reads of 89,768 and 134,995 metagenomic reads were annotated as Gp23. An HMM and a maximum-likelihood tree of Gp23 were built as described in the Supplementary Methods for PS-II D1.

**Recruitment to phage genomes.** The four currently available full pelagiphage genomes were downloaded from NCBI and concatenated with assembled viral contigs from metatranscriptomes the metagenomes as well as with published cyanophage genomes downloaded from NCBI RefSeq (accession numbers in Supplementary Figure 1). Metagenomic and metatranscriptomics reads were searched against this dataset with blastn using default settings. For metagenomes only hits longer than 100 bp were retained, and for metatranscriptomes only hits longer than 200 bp. Hits were then plotted against the genomes using R core package.

**Reporting summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

All raw data can be found on EMBL-ENA under project number [PRJEB12234](https://www.ebi.ac.uk/ena/record/PRJEB12234). Raw metatranscriptomics sequences accession numbers are [ERS1864892](https://www.ncbi.nlm.nih.gov/nuccore/ERS1864892)-[ERS1864903](https://www.ncbi.nlm.nih.gov/nuccore/ERS1864903), and negative control library sequences accession number is [ERR2089009](https://www.ncbi.nlm.nih.gov/nuccore/ERR2089009). Raw metagenomic sequences accession numbers are [ERS1869885](https://www.ncbi.nlm.nih.gov/nuccore/ERS1869885)-[ERS1869896](https://www.ncbi.nlm.nih.gov/nuccore/ERS1869896) and negative control accession number is [ERS1872073](https://www.ncbi.nlm.nih.gov/nuccore/ERS1872073). The 69 assembled viral contigs can be found in Genbank under Bioproject [PRJNA472807](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA472807) accession numbers [QKOA01000001](https://www.ncbi.nlm.nih.gov/nuccore/QKOA01000001)-[QKOA01000069](https://www.ncbi.nlm.nih.gov/nuccore/QKOA01000069).

Received: 31 October 2018 Accepted: 21 February 2019

Published online: 12 March 2019

### References

- Proctor, L. M. & Fuhrman, J. A. Viral mortality of marine bacteria and cyanobacteria. *Nature* **343**, 60 (1990).
- Fuhrman, J. A. & Noble, R. T. Viruses and protists cause similar bacterial mortality in coastal seawater. *Limnol. Oceanogr.* **40**, 1236–1242 (1995).
- Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
- Suttle, C. A. Marine viruses—major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**, 801–812 (2007).
- Breitbart, M. Marine viruses: truth or dare. *Ann. Rev. Mar. Sci.* **4**, 425–448 (2012).
- Weitz, J. S. & Wilhelm, S. W. Ocean viruses and their effects on microbial communities and biogeochemical cycles. *F1000 Biol. Rep.* **4**, 17 (2012).
- Roux, S. et al. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* **537**, 689–693 (2016).
- Lindell, D. et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).
- Avrani, S., Wurtzel, O., Sharon, I., Sorek, R. & Lindell, D. Genomic island variability facilitates Prochlorococcus–virus coexistence. *Nature* **474**, 604 (2011).
- Marston, M. F. et al. Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc. Natl. Acad. Sci. USA* **109**, 4544–4549 (2012).
- Martiny, J. B. H., Riemann, L., Marston, M. F. & Middelboe, M. Antagonistic coevolution of marine planktonic viruses and their hosts. *Ann. Rev. Mar. Sci.* **6**, 393–414 (2014).
- Fedida, A. & Lindell, D. Two *Synechococcus* genes, two different effects on Cyanophage infection. *Viruses* **9**, (2017).
- Suttle, C. A. & Chan, A. M. Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: abundance, morphology, cross-infectivity and growth characteristics. *Mar. Ecol. Prog. Ser.* **92**, 99–109 (1993).
- Sullivan, M. B., Waterbury, J. B. & Chisholm, S. W. Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**, 1047–1051 (2003).
- Mann, N. H. et al. The genome of S-PM2, a ‘photosynthetic’ T4-type bacteriophage that infects marine *Synechococcus* strains. *J. Bacteriol.* **187**, 3188–3200 (2005).
- Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**, e144 (2005).
- Clokic, M. R. J., Millard, A. D. & Mann, N. H. T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. *Virology* **7**, 291 (2010).
- Labrie, S. J. et al. Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ. Microbiol.* **15**, 1356–1376 (2013).
- Dekel-Bird, N. P. et al. Diversity and evolutionary relationships of T 7-like podoviruses infecting marine cyanobacteria. *Environ. Microbiol.* **15**, 1476–1491 (2013).
- Zhao, Y. et al. Abundant SAR11 viruses in the ocean. *Nature* **494**, 357–360 (2013).
- Brum, J. R. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
- Paez-Espino, D. et al. Uncovering Earth’s virome. *Nature* **536**, 425 (2016).
- Nishimura, Y. et al. Environmental viral genomes shed new light on virus–host interactions in the ocean. *mSphere* **2**, e00359-16 (2017).
- Duhaime, M. B. et al. Comparative omics and trait analyses of marine *Pseudoalteromonas* phages advance the phage OTU concept. *Front. Microbiol.* **8**, 1241 (2017).
- Zheng, Q., Chen, Q., Xu, Y., Suttle, C. A. & Jiao, N. A virus infecting marine photoheterotrophic Alphaproteobacteria (*Citromicrobium* spp.) Defines a new lineage of ssDNA viruses. *Front. Microbiol.* **9**, 1418 (2018).
- Aylward, F. O. et al. Diel cycling and long-term persistence of viruses in the ocean’s euphotic zone. *Proc. Natl. Acad. Sci. USA* **114**, 11446–11451 (2017).
- Chow, C.-E. T. & Fuhrman, J. A. Seasonality and monthly dynamics of marine myovirus communities. *Environ. Microbiol.* **14**, 2171–2183 (2012).
- Adriaenssens, E. M. & Cowan, D. A. Using signature genes as tools to assess environmental viral ecology and diversity. *Appl. Environ. Microbiol.* **80**, 4470–4480 (2014).
- Perez Sepulveda, B. et al. Marine phage genomics: the tip of the iceberg. *FEMS Microbiol. Lett.* **363**, fnw158 (2016).
- Ignacio-Espinoza, J. C., Solonenko, S. A. & Sullivan, M. B. The global virome: not as big as we thought?. *Curr. Opin. Virol.* **3**, 566–571 (2013).
- Brum, J. R. & Sullivan, M. B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**, 147–159 (2015).
- Roux, S. et al. Towards quantitative viromics for both double-stranded and single-stranded DNA viruses. *PeerJ* **4**, e2777 (2016).
- Martinez-Hernandez, F. et al. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat. Commun.* **8**, 15892 (2017).
- Baran, N., Goldin, S., Maidanik, I. & Lindell, D. Quantification of diverse virus populations in the environment using the polony method. *Nat. Microbiol.* **3**, 62–72 (2018).
- Pasulka, A. L. et al. Interrogating marine virus–host interactions and elemental transfer with BONCAT and nanoSIMS-based methods. *Environ. Microbiol.* **20**, 671–692 (2018).
- Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokic, M. Marine ecosystems: bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
- Bailey, S., Clokic, M. R. J., Millard, A. & Mann, N. H. Cyanophage infection and photoinhibition in marine cyanobacteria. *Res. Microbiol.* **155**, 720–725 (2004).
- Millard, A., Clokic, M. R. J., Shub, D. A. & Mann, N. H. Genetic organization of the psbAD region in phages infecting marine *Synechococcus* strains. *Proc. Natl. Acad. Sci. USA* **101**, 11007–11012 (2004).
- Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89 (2005).
- Clokic, M. R. J. et al. Transcription of a ‘photosynthetic’ T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8**, 827–835 (2006).
- Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Shedding new light on viral photosynthesis. *Photosynth. Res.* **126**, 71–97 (2015).
- Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Viruses inhibit CO<sub>2</sub> fixation in the most abundant phototrophs on earth. *Curr. Biol.* **26**, 1585–1589 (2016).
- Fridman, S. et al. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat. Microbiol.* **2**, 1350–1357 (2017).
- Sharon, I. et al. Viral photosynthetic reaction center genes and transcripts in the marine environment. *Isme. J.* **1**, 492–501 (2007).
- Yao, D. C. I., Brune, D. C. & Vermaas, W. F. J. Lifetimes of photosystem I and II proteins in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS Lett.* **586**, 169–173 (2012).
- Franklin, N. C. & Bennett, G. N. The N protein of bacteriophage lambda, defined by its DNA sequence, is highly basic. *Gene* **8**, 107–119 (1979).
- Schneider, G. J. & Haselkorn, R. Characterization of two early promoters of cyanophage N-i. *Virology* **167**, 160–165 (1988).
- Campbell, A. Comparative molecular biology of lambdoid phages. *Annu. Rev. Microbiol.* **48**, 193–222 (1994).

49. Miller, E. S. et al. Bacteriophage T4 genome. *Microbiol. Mol. Biol. Rev.* **67**, 86–156 (2003).
50. Roux, S., Enault, F., Hurwitz, B. L. & Sullivan, M. B. VirSorter: mining viral signal from microbial genomic data. *PeerJ* **3**, e985 (2015).
51. Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A. & Sun, F. VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome* **5**, 69 (2017).
52. Sieradzki, E. T., Fuhrman, J. A., Rivero-Calle, S. & Gómez-Consarnau, L. Proteorhodopsins dominate the expression of phototrophic mechanisms in seasonal and dynamic marine picoplankton communities. *PeerJ* **6**, e5798 (2018).
53. Miranda, J. A., Culley, A. I., Schvarcz, C. R. & Steward, G. F. RNA viruses as major contributors to Antarctic viroplankton. *Environ. Microbiol.* **18**, 3714–3727 (2016).
54. Deng, L. et al. Viral tagging reveals discrete populations in *Synechococcus* viral genome sequence space. *Nature* **513**, 242–245 (2014).
55. Moniruzzaman, M. et al. Virus-host relationships of marine single-celled eukaryotes resolved from metatranscriptomics. *Nat. Commun.* **8**, 16054 (2017).
56. Cram, J. A. et al. Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years. *Isme. J.* **9**, 563–580 (2015).
57. Williamson, S. J. et al. Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* **7**, e42047 (2012).
58. Matteson, A. R. et al. High abundances of cyanomyoviruses in marine ecosystems demonstrate ecological relevance. *FEMS Microbiol. Ecol.* **84**, 223–234 (2013).
59. Dekel-Bird, N. P., Sabehi, G., Mosevitzky, B. & Lindell, D. Host-dependent differences in abundance, composition and host range of cyanophages from the Red Sea. *Environ. Microbiol.* **17**, 1286–1299 (2015).
60. Needham, D. M., Sachdeva, R. & Fuhrman, J. A. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *Isme. J.* **11**, 1614–1629 (2017).
61. Thompson, L. R. et al. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc. Natl Acad. Sci. USA* **108**, E757–E764 (2011).
62. Sullivan, M. B. et al. Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* **4**, e234 (2006).
63. Clokie, M. R. J. & Mann, N. H. Marine cyanophages and light. *Environ. Microbiol.* **8**, 2074–2082 (2006).
64. Storms, Z. J., Brown, T., Cooper, D. G., Sauvageau, D. & Leask, R. L. Impact of the cell life-cycle on bacteriophage T4 infection. *FEMS Microbiol. Lett.* **353**, 63–68 (2014).
65. Zinser, E. R. et al. Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *prochlorococcus*. *PLoS One* **4**, e5135 (2009).
66. Holtzendorff, J. et al. Genome streamlining results in loss of robustness of the circadian clock in the marine cyanobacterium *Prochlorococcus marinus* PCC 9511. *J. Biol. Rhythms* **23**, 187–199 (2008).
67. Ottesen, E. A. et al. Pattern and synchrony of gene expression among sympatric marine microbial populations. *Proc. Natl Acad. Sci. USA* **110**, E488–E497 (2013).
68. Giovannoni, S. J. et al. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438**, 82–85 (2005).
69. Giovannoni, S. J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Ann. Rev. Mar. Sci.* **9**, 231–255 (2017).
70. Hurwitz, B. L. & U'Ren, J. M. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin. Microbiol.* **31**, 161–168 (2016).
71. Wommack, K. E. & Colwell, R. R. Viroplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
72. Heldal, M. & Bratbak, G. Production and decay of viruses in aquatic environments. *Mar. Ecol. Prog. Ser.* **72**, 205–212 (1991).
73. Wang, K. & Chen, F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ. Microbiol.* **10**, 300–312 (2008).
74. Millard, A. D. & Mann, N. H. A temporal and spatial investigation of cyanophage abundance in the Gulf of Aqaba, Red Sea. *J. Mar. Biol. Assoc. U.K.* **86**, 507–515 (2006).
75. Clokie, M. R. J., Millard, A. D. & Mann, N. H. T4 genes in the marine ecosystem: studies of the T4-like cyanophages and their role in marine ecology. *Virology* **7**, 291 (2010).
76. Martin, E. & Benson, R. in *The Bacteriophages* (ed. Calendar, R.) 607–645 (Springer US, New York, NY, 1988).
77. Zeigler Allen, L. et al. The baltic sea virome: diversity and transcriptional activity of DNA and RNA viruses. *mSystems* **2**, e00125-16 (2017).
78. Dupont, C. L. et al. Genomes and gene expression across light and productivity gradients in eastern subtropical Pacific microbial communities. *Isme. J.* **9**, 1076–1092 (2015).
79. Eren, A. M. et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**, e1319 (2015).

## Acknowledgements

The authors would like to thank R. Sachdeva, N. Ahlgren, A. Parada, L. Berdjeb, E. Graham, M. Lee, J. Ren, F. Sun and T. Delmont for insightful discussions and advice on bioinformatic analyses. We thank C. Roney-Garcia, the Sundiver crew and the USC Wrigley Institute of Environmental Studies for logistic support. This work was supported by NSF grant 1136818, Gordon and Betty Moore Foundation Marine Microbiology Initiative grant GBMF3779 and Norma and Jerol Sonosky summer fellowship to E.T.S.

## Author contributions

E.T.S. participated in work design, data acquisition, performed data analysis and interpretation and wrote and revised the manuscript. J.C.I.-E. Contributed to data analysis and interpretation and reviewed the manuscript. D.M.N. assisted in data acquisition and interpretation and reviewed the manuscript. E.B.F. assisted in data acquisition and analysis and reviewed the manuscript. J.A.F. Conceived the project, designed and supervised the work, assisted in interpretation and reviewed the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <https://doi.org/10.1038/s41467-019-09106-z>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Journal peer review information:** *Nature Communications* thanks Bonnie Hurwitz and the other anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

**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) 2019