LETTERS
DOI: 10.1038/s41564-017-0002-9
nature microbiology

A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells

Svetlana Fridman[1], José Flores-Uribe[1], Shirley Larom[1], Onit Alalouf[1], Oded Liran[2], Iftach Yacoby[2], Faris Salama[1], Benjamin Bailleul[3], Fabrice Rappaport[3], Tamar Ziv[4], Itai Sharon[5,6], Francisco M. Cornejo-Castillo[7], Alon Philosof[1], Christopher L. Dupont[8], Pablo Sánchez[7], Silvia G. Acinas[7], Forest L. Rohwer[9], Debbie Lindell[1]* and Oded Béjà[1]*

Cyanobacteria are important contributors to primary production in the open oceans. Over the past decade, various photosynthesis-related genes have been found in viruses that infect cyanobacteria (cyanophages). Although photosystem II (PSII) genes are common in both cultured cyanophages and environmental samples[1–4], viral photosystem I (vPSI) genes have so far only been detected in environmental samples[5,6]. Here, we have used a targeted strategy to isolate a cyanophage from the tropical Pacific Ocean that carries a PSI gene cassette with seven distinct PSI genes (*psaJF*, *C*, *A*, *B*, *K*, *E*, *D*) as well as two PSII genes (*psbA*, *D*). This cyanophage, P-TIM68, belongs to the T4-like myoviruses, has a prolate capsid, a long contractile tail and infects *Prochlorococcus* sp. strain MIT9515. Phage photosynthesis genes from both photosystems are expressed during infection, and the resultant proteins are incorporated into membranes of the infected host. Moreover, photosynthetic capacity in the cell is maintained throughout the infection cycle with enhancement of cyclic electron flow around PSI. Analysis of metagenomic data from the *Tara* Oceans expedition[7] shows that phages carrying PSI gene cassettes are abundant in the tropical Pacific Ocean, composing up to 28% of T4-like cyanomyophages. They are also present in the tropical Indian and Atlantic Oceans. P-TIM68 populations, specifically, compose on average 22% of the PSI-gene-cassette carrying phages. Our results suggest that cyanophages carrying PSI and PSII genes are likely to maintain and even manipulate photosynthesis during infection of their *Prochlorococcus* hosts in the tropical oceans.

Oxygenic photosynthetic organisms, including cyanobacteria, use four integral membrane complexes: photosystem II (PSII), b_6f, photosystem I (PSI) and ATPase, to convert sunlight into chemical energy. Linear electron transfer begins with PSII catalysing the transfer of electrons, derived from water, to the plastoquinone pool. PSI transfers electrons from plastocyanin/cytochrome c_6 to ferredoxin, thereby generating the reducing power needed for CO_2 fixation in the form of NADPH. The b_6f complex then couples electron transfer from the reduced plastoquinone (QH_2) to the oxidized plastocyanin/cytochrome c_6 to the pumping of protons in the lumen. Finally, the electrochemical proton gradient drives ATP formation by the ATPase complex.

About 90% of cyanophage isolates carry the PSII *psbA* gene coding for the labile D1 protein[4,8], and it is thought that its expression increases phage fitness[9–12]. In addition to the *psbA* gene, many cyanophages carry genes encoding the PSII D2 protein (*psbD*) or different genes of the photosynthetic electron transport chain, such as plastocyanin (*petE*), ferredoxin (*petF*) and plastoquinol terminal oxidase (PTOX)[1,3,8,13,14].

Using environmental metagenomics, we previously suggested that cyanophages from the oceans contain gene cassettes coding for PSI reaction centre proteins[6]. Subsequent environmental surveys showed that these gene cassettes are diverse and present in tropical and subtropical waters[15]. Two main types of viral PSI gene cassettes were identified: one containing seven PSI genes, *psaJF*, *C*, *A*, *B*, *K*, *E* and *D* (*vPSI-7*) and the other containing four PSI genes, *psaD*, *C*, *A* and *B* (*vPSI-4*)[6,15]. Cyanophages containing the *vPSI-7* gene organization were hypothesized to shift host photosynthesis toward a cyclic electron flow mode[6].

Despite evidence from environmental DNA for the existence of cyanophages encoding PSI gene cassettes, no cyanophage carrying PSI genes has been isolated so far. Indeed, this phage was dubbed the 'mystery phage' by Puxty and co-authors[8]. Here, we have developed a targeted isolation strategy in an attempt to find such a PSI-carrying cyanophage (Supplementary Fig. 1). This strategy relied on PCR screening of lysates and plaques for PSI gene arrangements unique to environmental viral DNA[6], primarily adjacent *psaC* and *psaA* genes. To this end, 18 different *Prochlorococcus* and marine *Synechococcus* strains were used to screen viral concentrates collected from the Equatorial Pacific Ocean off the Line Islands (Caroline Island 9.9° S, 150.2° W). Phage lysates containing viral PSI gene arrangements were observed only when the high-light-adapted (HL) *Prochlorococcus* MIT9515 strain served as the

[1]Faculty of Biology, Technion-Israel Institute of Technology, Haifa 32000, Israel. [2]Department of Molecular Biology and Ecology of Plants, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel. [3]Institut de Biologie Physico-Chimique, UMR 7141 CNRS and Université Pierre et Marie Curie, 13 rue Pierre et Marie Curie, 75005 Paris, France. [4]Smoler Proteomics Center, Technion-Israel Institute of Technology, Haifa 32000, Israel. [5]Migal Galilee Research Institute, Kiryat Shmona 11016, Israel. [6]Tel Hai College, Upper Galilee 12210, Israel. [7]Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), CSIC, 08003 Barcelona, Spain. [8]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, CA 92037, USA. [9]Department of Biology, San Diego State University, San Diego, CA 92182, USA. Fabrice Rappaport is deceased. *e-mail: dlindell@tx.technion.ac.il; beja@tx.technion.ac.il

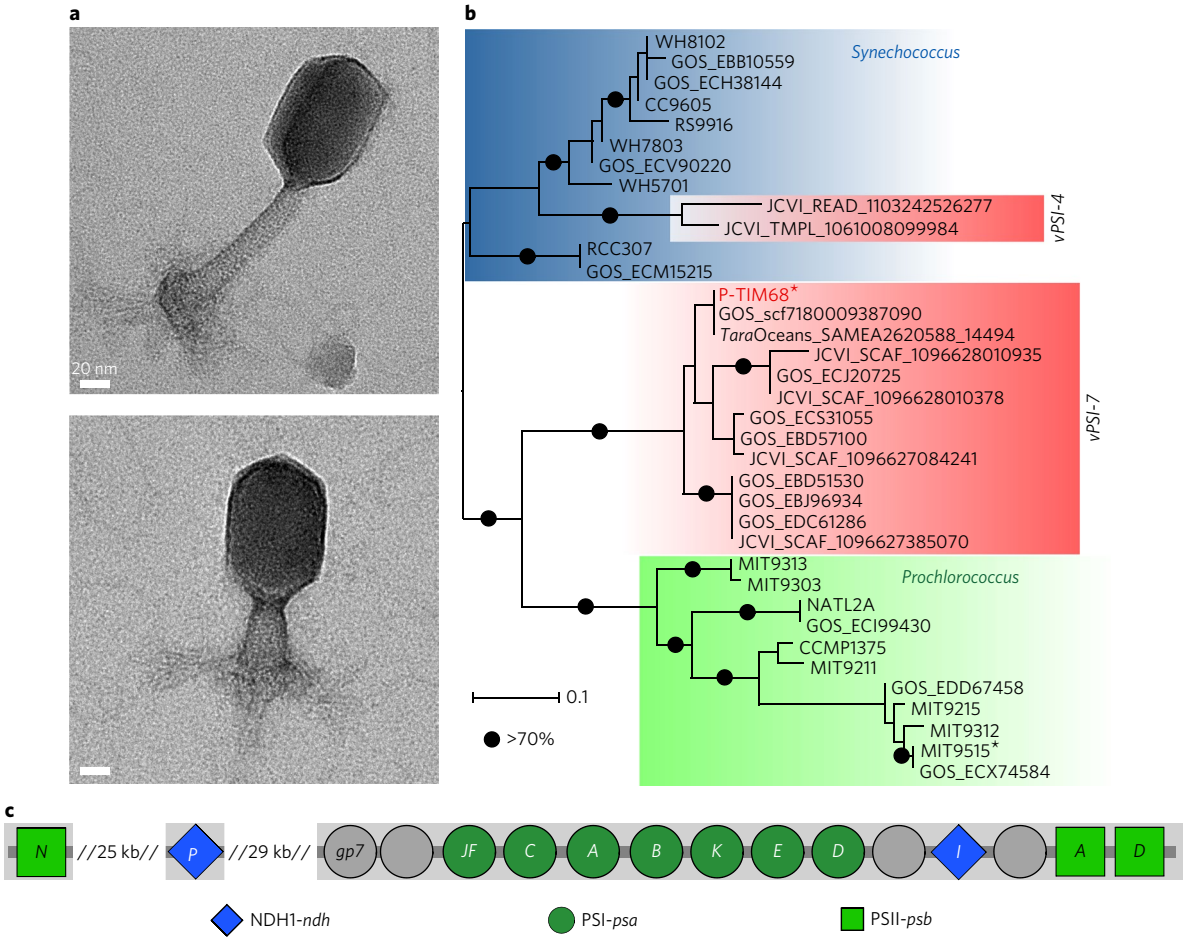

Fig. 1 | P-TIM68 gene organization, phylogeny and morphology. a, Representative (out of 13) negative-stain TEM micrographs of the P-TIM68 phage with relaxed and contracted tails. Scale bars, 20 nm. **b**, A maximum-likelihood phylogenetic tree of PsaA proteins obtained from the GOS expedition, marine cyanobacteria and the P-TIM68 phage. Bold type denotes genes belonging to cultured isolates, and background colours denote *Synechococcus* (blue), *Prochlorococcus* (green) and putative phages (red). Black filled circles indicate bootstrap values higher than 70%. The MIT9515 host and the P-TIM68 phage are labelled with asterisks. PsaA proteins from *Trichodesmium* and *Synechocystis* were used as outgroups and are not shown. **c**, Schematic physical map of the photosynthesis region in the P-TIM68 genome: PSI and PSII genes (green), NDH-1 genes (blue) and viral *gp7* and hypothetical genes (grey).

host (see Methods and Supplementary Fig. 1). A phage from these lysates, P-TIM68, was plaque-purified and characterized.

P-TIM68 has a long contractile tail indicative of myoviruses and a prolate capsid (Fig. 1a). It has a 197 kb genome with 34.3 G+C%, 246 predicted open reading frames and no tRNA genes (Supplementary Fig. 2a). The P-TIM68 genome contains a suite of DNA replication and virion structural genes characteristic of the archetype T4 (Supplementary File 1) that are homologous to those found in other T4-like cyanophages[13,14]. This phage clusters within clade II of the T4-like myocyanophages (based on the *g20* gene), and genomic comparison to a closely related T4-like cyanomyophage from that clade, S-SM2, showed that there is a high degree of synteny and per cent identity between the structural, replication and nucleotide metabolism genes of the two phages (Supplementary Fig. 2b).

The P-TIM68 phage genome contains PSI genes arranged in a cassette, with the fused *psaJF* gene followed by the *psaC*, *A*, *B*, *K*, *E* and *D* genes (Fig. 1c and Supplementary Fig. 2a). This is the same arrangement as that previously observed for the metagenomic *vPSI-7* cassette[6]. Furthermore, a phylogenetic tree constructed with the PsaA protein clearly places it within the *vPSI-7* cluster of viral PsaA from environmental DNA (Fig. 1b).

In addition to the PSI genes, the genome of P-TIM68 contains three PSII genes, *psbA* and *psbD* genes coding for the core D1 and D2 proteins, and a *psbN* gene homologue coding for a small protein required for the repair and assembly of PSII reaction centres[16]. It also contains two genes coding for type I NAD(P)H dehydrogenase (NDH-1) subunits (the *ndhP* and *ndhI* genes) and four high-light-inducible protein (*hli*) genes arranged in tandem.

P-TIM68 also carries genes involved in carbon metabolism. Two genes that encode for aldolases are shared between the Calvin cycle and the pentose phosphate pathway[17]: *talC*, which is readily found in different cyanophage genomes[18], and a fructose-1,6-bisphosphate aldolase class I gene that has not previously been found in phage genomes but was recently reported in metagenomic scaffolds of suspected viral origin[5]. In addition, P-TIM68 contains a CP12 gene homologue that codes for a small protein predicted to inhibit the Calvin cycle[17,19].

The photosynthesis region of the P-TIM68 phage genome is made up of the seven-gene PSI cassette, one of the NDH-1 genes and the two reaction centre genes of PSII (Fig. 1c and Supplementary Fig. 2b). This region is flanked by hypothetical genes and is downstream of a structural protein (the baseplate wedge initiator protein, gp7), rather than the *nrdA* and *nrdB* ribonucleotide reductase genes found to neighbour previously reported PSI environmental scaffolds[6]. Analyses of data from the *Tara* Oceans[7] and Global Ocean Sampling (GOS)[20,21] expeditions revealed the presence of two scaffolds

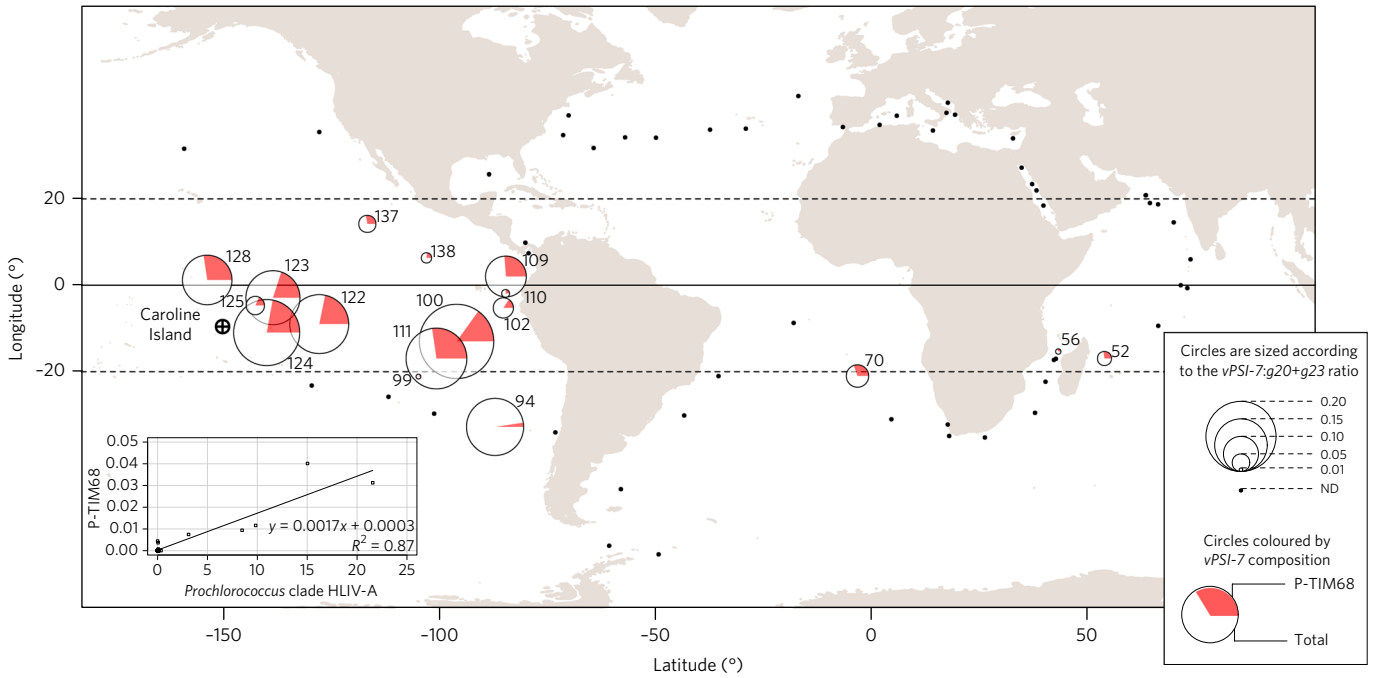

Fig. 2 | Global distribution and relative abundance of P-TIM68 and other *vPSI-7* phages. The area of each circle is scaled to represent the fraction of *vPSI-7* phages (based on the viral *psaA* and *psaB* genes) from the total cyanomyophages (based on the viral *g20* and *g23* genes) at different *Tara* Oceans stations. ND, not detected (stations with fewer than 20 *vPSI-7* reads). Caroline Island, from which P-TIM68 was isolated, is marked with a circled black cross. Proportions of P-TIM68 phage from *vPSI-7* phages at each *Tara* Oceans station are shown in red. Exact numbers of reads recruited at each depth, and the filtering scheme, are provided in Supplementary Table 1. The equator is shown as a solid line and latitudes 20° N and 20° S are shown as dashed lines. Station numbers are shown for stations in which *vPSI-7* reads were detected. For clarity, stations 122–125 are not shown at their exact locations. Station coordinates and normalized read recruitment numbers are provided in Supplementary Table 1. Inset: P-TIM68 abundance (represented as the ratio of P-TIM68 out of total cyanomyophages) versus *Prochlorococcus* HLIV-A clade abundance[29]. Abundances of P-TIM68 and *Prochlorococcus* HLIV-A were significantly correlated ($P < 1 \times 10^{-15}$; the regression line, regression equation and R^2 value are shown). Station coordinates and normalized read recruitment numbers are provided in Supplementary Table 1.

with exactly the same gene organization as in P-TIM68 over their entire lengths of 3.9 and 17.8 kb (Supplementary File 2). These scaffolds showed high identity (98%) to P-TIM68 at the DNA level over the region of the PSI gene cassette, indicating that phages very similar to P-TIM68 are present in the environment.

We next assessed how widespread the P-TIM68 phage is in the oceans. Recruitment of reads from the *Tara* Oceans expedition to the *vPSI-7* segment of P-TIM68 showed that P-TIM68 sequences (estimated using ≥98% nucleotide identity) are present in tropical waters in the Pacific, Indian and Atlantic Oceans (Supplementary Table 1). We then estimated the abundance of P-TIM68 (≥98% identity) and other *vPSI-7* (≥80% identity) phages in the *Tara* Oceans data set (Fig. 2a) relative to all T4-like cyanomyophages (using the T4-like *g20* and *g23* marker genes[22]). The *vPSI-7* phages were particularly prevalent in the southern region of the tropical Pacific Ocean, often reaching 15–28% of total cyanomyophages (Fig. 2a and Supplementary Table 1). The P-TIM68 genotype composed on average 22% of these *vPSI-7* phages (average from all stations containing more than 20 *psaA* reads; proportions are labelled in red in Fig. 2a).

Characterization of the infection cycle of the P-TIM68 phage revealed a long latent period of 10–12 h (Supplementary Fig. 3) and a burst size of 12 ± 3 ($n = 3$) infective phages per cell. Similarly, other T4-like cyanophages have latent periods ranging from 4 to 18 h (refs [19,23,24]) and burst sizes that range from 21 to 45 phages per cell[23,25,26]. The phage PSI and PSII genes were transcribed at the same time during the infection cycle, with maximal transcript levels at 6–8 h after infection (Fig. 3a). These photosynthesis genes were transcribed together with middle genes involved in phage genome replication (Fig. 3a), as found for PSII genes in other cyanophages[24,27].

At the protein level, unique peptides from phage PSI (PsaJF, PsaA, PsaB, PsaK, PsaE and PsaD) and PSII (PsbA, PsbD and PsbN) proteins were detected in host membranes 8 h after infection (Supplementary Table 2). Quantification of host- and phage-specific photosynthesis peptides showed that levels of the host PSII PsbA protein (D1) decreased during infection to about 50% of their initial levels (Fig. 3b). This corresponded with an increase of the viral homologue to 43% of the total D1 proteins in infected cells by the end of the latent period (Supplementary Fig. 4a), thus almost completely restoring the amount of D1 proteins to original levels. A similar trend was previously observed for a *Prochlorococcus* T7-like podovirus (P-SSP7) carrying a single PSII gene, although the viral D1 reached about 10% of total D1 proteins[9], implying a general role for viral PSII genes in T4- and T7-like cyanophages.

The host PSI PsaA protein, in comparison, showed no significant decrease (Fig. 3c). The viral PSI PsaA protein, however, increased steadily (Fig. 3c), reaching about 6% of the total PSI PsaA proteins by the end of the latent period (Supplementary Fig. 4b). Because only 70% of host cells were infected in these experiments, the level of phage PsaA protein corresponds to 8.6% of the total PsaA in infected cells. This is a small but significant level of PSI proteins that we hypothesize supplements host PSI proteins during infection. It is possible that even more phage PSI proteins replace host proteins under higher light conditions similar to those found in the surface layers of tropical waters.

To characterize the host's photosynthetic activity during P-TIM68 infection, the relative maximum electron transfer rate ($rETR_{max}$) and PSI reduction rate were measured. *Prochlorococcus* cells were grown under a photon flux density of 10 µmol photons m^{-2} s^{-1}.

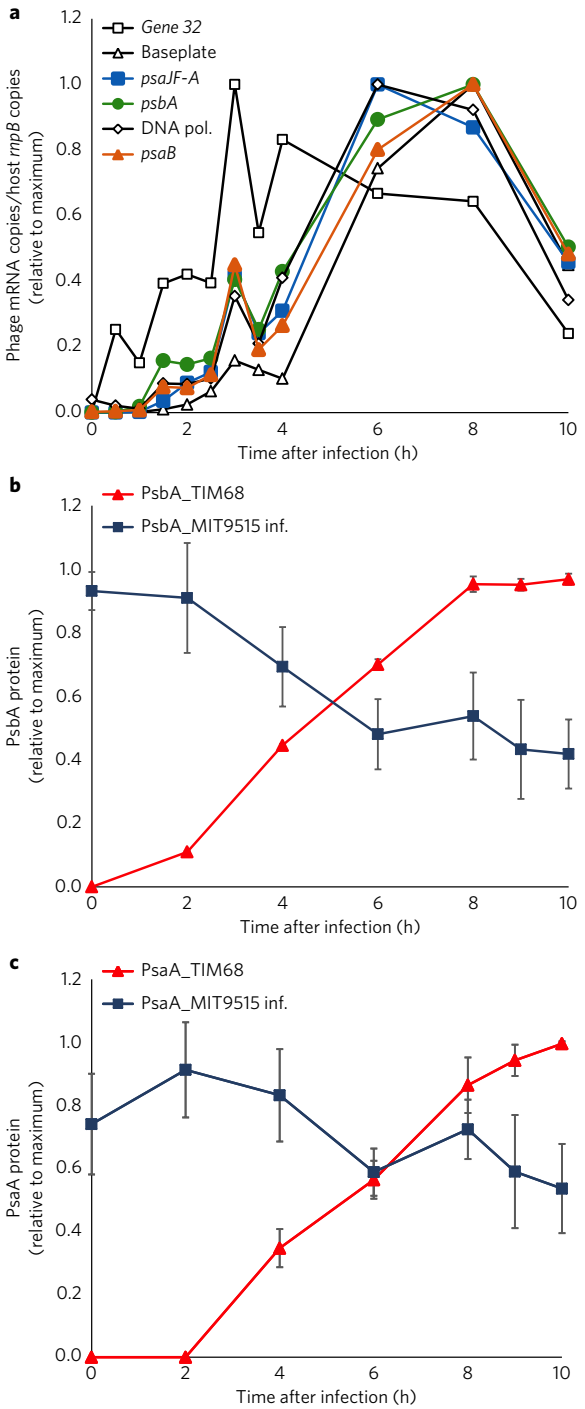

Fig. 3 | Expression of P-TIM68 PSI and PSII genes in infected *Prochlorococcus* MIT9515 cells. a, mRNA transcript levels of representative viral early (*gene 32*), middle (DNA polymerase) and late (baseplate) genes and viral PSII (*psbA*) and PSI (*psaB* and *psaJF*) genes during the 10 h latent period. Results shown are from a single biological replicate and are representative of all three independent experiments. **b**,**c**, Levels of phage and host PsbA (PSII) (**b**) and PsaA (PSI) (**c**) protein levels during the latent period of infection. Data shown in **b** and **c** were normalized to expression in uninfected control cells. Error bars denote standard deviation (s.d.) of three biological experiments. *P* values for the paired samples *t*-test of the first two time points and the last time point are <0.001 for the decrease in host PSII PsbA levels and 0.016 for the decrease in host PSI PsaA levels. All data are presented relative to maximal levels for each gene and protein independently. Protein data not normalized to uninfected control cells are shown in Supplementary Fig. 4.

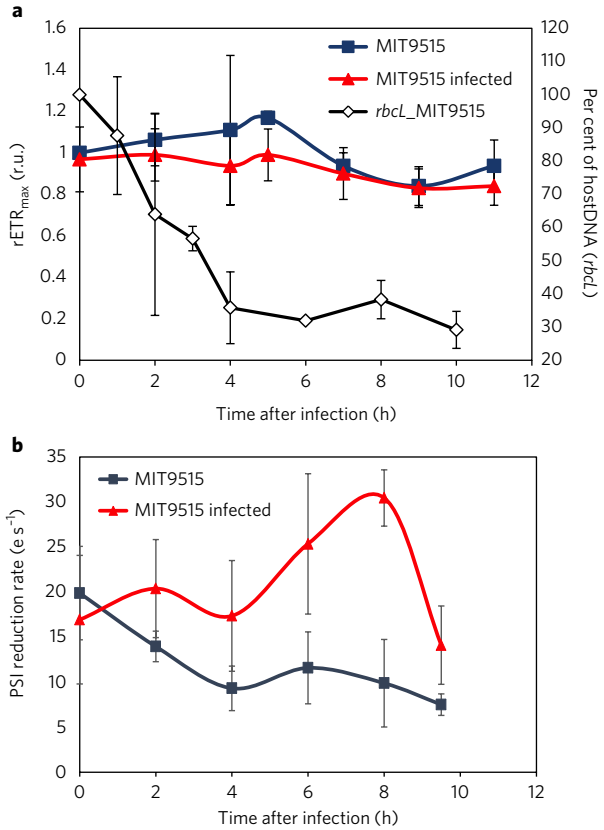

Fig. 4 | Changes in PSII and PSI photosynthetic parameters during infection of *Prochlorococcus* MIT9515 with phage P-TIM68. a, $rETR_{max}$ and host DNA degradation. $rETR_{max}$ is normalized to the $rETR_{max}$ of the control *Prochlorococcus* culture before infection and corresponds to the mean of four independent biological samples ± s.d. Timing of host DNA degradation is presented as per cent of maximal genomic DNA. Intracellular host DNA was quantified by real-time PCR targeting of the host *rbcL* gene ($n = 3$ independent biological samples ± s.d.); r.u., relative units. **b**, PSI reduction rate, mean of three independent biological replicates ± s.d; e s^{-1}, electrons per second.

Throughout the 10 h latent period of P-TIM68, $rETR_{max}$ remained constant while host chromosome degradation clearly indicated viral takeover of the host (Fig. 4a). These results are consistent with previous findings of continued PSII activity during infection of viruses carrying PSII genes[9,19,28]. In contrast, when the PSI reduction rate was monitored (by absorbance changes at 705 nm), an increased rate of electron flow was detected in infected cells from 6 to 8 hours post infection (Fig. 4b).

The photosynthetic measurements of infected relative to uninfected cells revealed that PSII activity remained unchanged during P-TIM68 infection ($rETR_{max}$ in Fig. 4a) while the rate of PSI reduction increased. These results suggest that a PSII-independent electron flow to PSI, that is, an increased cyclic electron flow around PSI, is induced during infection. Such increased cyclic electron flow could generate the extra ATP required for phage DNA replication and nucleotide biosynthesis. However, an alternative explanation was also considered: even if the $rETR_{max}$ remains constant during infection, the electron flow from PSII could have increased when normalized per PSI if the PSII/PSI stoichiometry or the absorption cross-section of PSII increases during infection. However, chlorophyll fluorescence emission spectra at low temperature (77 K) showed no changes in PSII/PSI stoichiometry between infected and uninfected cells (Supplementary Fig. 5a). In addition,

absorption cross-section of PSII did not change during infection (Supplementary Fig. 5b). Therefore, these results suggest that there is increased cyclic electron flow around PSI as a result of P-TIM68 infection.

Previous metagenomic data have indicated the likely existence of PSI-carrying phages in the environment, but the isolation of P-TIM68 provides definitive proof that the PSI cassettes are of cyanophage origin and that phage encoding proteins capable of forming an entire PSI complex are present in the oceans and can infect *Prochlorococcus*. Furthermore, the physiological characterization of this phage showed that infected cells maintain and even increase photosynthesis during the infection cycle.

The restricted distribution pattern of the PSI-carrying phages to tropical waters is intriguing and directly correlates to the distribution of the uncultured HLIII-A and HLIV-A *Prochlorococcus* clades that are unique to this high-nutrient, low-chlorophyll (HNLC), iron-depleted region[29] (Fig. 2 inset shows the correlation between *Prochlorococcus* HLIV-A and P-TIM68 abundances). If, indeed, members of one or both of these *Prochlorococcus* lineages are the natural hosts for *vPSI-7* phages, this would explain their limited distribution pattern. In addition, it was recently shown that the ratio of PSI/PSII declines under iron-limiting conditions in freshwater cyanobacteria[30]. Although highly speculative, perhaps *vPSI-7* phages are compensating for such a decline in infected hosts under the iron-limiting conditions found in this environment. Future experiments under high-light and iron-limited conditions will be important for elucidating the full impact of the *vPSI-7*-carrying phages on their *Prochlorococcus* hosts in such environments.

In summary, our combined observations that P-TIM68 and other *vPSI-7* phages are abundant in the environment, together with their ability to maintain and even manipulate the mode of host photosynthesis in their *Prochlorococcus* hosts, suggest that viral PSI genes play a key functional role in oceanic photosynthesis in tropical waters.

Methods

Bacterial strains. The *Synechococcus* and *Prochlorococcus* strains used in this study were *Synechococcus* spp. strains WH7803, WH7805, CC9311, WH8102, WH8109, CC9605, BL107, RS9916, WH5701, RS9917, RCC307 and CC9902, and *Prochlorococcus* spp. strains MED4, MIT9215, MIT9515, MIT9312, NATL2A and SS120. *Synechococcus* strains were grown in an artificial seawater-based medium (ASW)[31] with modifications as described previously[32]. *Prochlorococcus* strains were grown in the natural seawater-based medium Pro99[33]. Cultures were grown at 21 °C under cool white light with a 14:10 h light:dark cycle at an intensity of 10–14 μmol photons m^{-2} s^{-1} during the light period. The *Prochlorococcus* MIT9515 strain, on which the P-TIM68 phage was isolated, belongs to the HLI clade of *Prochlorococcus* and was isolated from the Equatorial Pacific (5.7° S, 107.1° W) at 15 m depth by the Chisholm group (http://proportal.mit.edu/genome/id=272/).

Phage isolation strategy. Viral concentrates from the Pacific Line Islands were collected in April 2009 from the uninhabited Caroline (Millennium) Island and were prepared according to ref. [34]. Cyanobacteria were challenged with these concentrates in tenfold serial dilutions in 96-well plates. Wells showing lysis were then analysed by PCR for the presence of the *psaC-psaA* PSI gene arrangement found in the putative cyanophages but not known for cyanobacteria. Lysates from wells positive for the viral *psaC-psaA* were used to challenge the host again (Supplementary Fig. 1) and were then tested for additional unique viral PSI gene combinations (*psaJK-psaC, psaJF-psaA* and *psaB-psaD)*. The highest-dilution lysate positive for these combinations was then used in four rounds of plaque purification and PCR screening until all plaques were positive for the viral *psaC-psaA* gene arrangement. Although lysis was observed in all cyanobacterial strains tested in the first challenge, only lysed wells from *Prochlorococcus* MIT9515 gave positive PCR signals for the *psaC-psaA* combination and underwent subsequent isolation procedures.

The degenerate primers used for the PCR reaction of *psaC-psaA* were PsaC(MVPW[D/E]G)-fwd [5′-atggtnccntgggaygg-3′] and PsaA(GWLADP)-rev [5′-ggrtcngcnarccancc-3′] [35]. PCR conditions were set according to ref. [36]. Primers and PCR conditions for reactions *psaJF-psaC*, *psaJF-psaA* and *psaB-psaD* were according to ref. [6].

Transmission electron microscopy. A cyanophage lysate was filtered using a 0.2-μm-pore-sized filter and then brought to a final concentration of ~10^9 p.f.u. ml^{-1} by centrifugation using a 100 kDa molecular-weight cutoff (~10 nm sized pores) centrifugal filter (Millipore). The phage lysate was then CsCl-purified according to ref. [37]. For transmission electron microscopy (TEM) analysis, 20 μl of the CsCl-purified phage concentrate was placed on a formvar/carbon-coated 100-mesh grid. After 20 min at room temperature, excess liquid was removed and phages were negatively stained with 1% uranyl acetate for 1 min. Uranyl acetate was removed and the grid was air-dried for 10 min before overnight desiccation under vacuum. Phages were examined at an accelerating voltage of 200 kV using a FEI Tecnai G2 T20 S-Twin TEM.

Sequencing and annotations. The P-TIM68 phage was sequenced using Illumina dual index MiSeq using a Nextera XT library preparation kit. Raw reads were deposited in the GenBank Sequence Read Archive (BioSample accession no. SAMN06618437) and *de novo* assembled using the velvet 1.2.10 (ref. [38]) program with a k-mer size = 73. Assembly coverage was 1,790-fold. The genome assembled into a single circular contig as expected for linear T4-like phage genomes that are circularly permuted and terminally redundant[39]. Misassemblies were detected through manual scanning of mapped reads to the assembled genome. Read mapping was performed using Bowtie[40] (parameters -q -n 2 -e 200 -p 10 –-best). The genome of P-TIM68 was deposited in GenBank under accession no. KM359505. Gene calling was carried out using Prodigal 2.6.1 (ref. [41]) followed by annotation using blast[42] (program blastall, parameters –p blastX -e 1e-5 –F F –b 1 –v 1) against NCBI's nr database. Hits were curated manually.

Infection characterization. Infection dynamics were determined from the appearance of phage in the extracellular medium with time. Exponentially growing *Prochlorococcus* MIT9515 cells were grown in triplicate at 21 °C in continuous light (14 μmol photons m^{-2} s^{-1}) and infected with P-TIM68 (at a multiplicity of infection (MOI) of 0.1). Quantification of phage DNA in the extracellular medium (0.2 μm filtrate) was determined by real-time PCR for the P-TIM68 portal protein (*g20*) gene using the real-time procedure described in the 'Quantitative PCR protocol' section of the Methods. Burst size was calculated according to ref. [37].

Host gDNA degradation. Exponentially growing *Prochlorococcus* MIT9515 cells were infected with P-TIM68 (MOI of 3). *Prochlorococcus* gDNA was quantified after collection of cells on 0.2-μm-pore-sized polycarbonate filters (GE), washed with sterile seawater followed by 3 ml of preservation solution (10 mM Tris, 100 mM EDTA, 0.5 M NaCl; pH 8.0) and frozen at −80 °C. DNA was extracted by the heat lysis method[43]. The RuBisCO large subunit gene (*rbcL*) was amplified for *Prochlorococcus* gDNA using the real-time procedure described in the 'Quantitative PCR protocol' section of the Methods.

Gene expression. *RNA extraction. Prochlorococcus* cells were grown at 21 °C to mid-logarithmic phase and were infected with P-TIM68 at an MOI of 3. Samples were collected every 30 min for the first 4 h after infection and then every 2 h until 10 h after infection. The experiment was repeated three times. A 9 ml volume of infected or uninfected cells was filtered onto 0.22 μm Durapore filters (Millipore). The filters were transferred to tubes containing 1 ml RNAlater (Ambion), frozen, and stored at −80 °C. Total RNA was extracted using a mirVana RNA isolation kit (Ambion). Genomic DNA was removed using a Turbo DNA-free kit (Ambion). cDNA was made using a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems).

Several phage genes were chosen for transcript analysis by real-time PCR: the *psaJF-A*, *psaB* and *psbA* photosynthetic genes, the *g32* early gene[24], the *g147* DNA polymerase middle gene and the *g85* baseplate late gene. Gene numbers are as per the P-TIM68 genome. Results were normalized to the amount of host *rnpB* gene in infected cells.

Quantitative PCR protocol. Each real-time PCR reaction contained 1× Roche universal probe library (UPL) master mix (LightCycler 480 Probes Master, Roche), 80 nM UPL hydrolysis probe, 500 nM primers and cDNA template in a total volume of 25 μl. Reactions were carried out on a LightCycler 480 Real-Time PCR System (Roche). The program used was one step of 95 °C for 15 min followed by 45 cycles of amplification, each including 10 s denaturation at 95 °C and a 30 s combined annealing and elongation step at 60 °C, at the end of which plate fluorescence was read (FAM; Ex/Em 465/510 nm). A list of genes, primers and probes used in real-time PCR assays is provided in Supplementary Table 3.

Shotgun proteomics. Prochlorococcus cells (1.5 l) were grown at 21 °C to mid-logarithmic phase and were infected with P-TIM68 (MOI of 3). Cells were collected by centrifugation for 30 min at 13,300*g*, 4 °C. The pellet was resuspended in buffer A (50 mM Tris/HCl, 5 mM $MgCl_2$•$6H_2O$, 10 mM NaCl, pH 8.0), frozen in liquid nitrogen and stored at −80 °C. The pellets were thawed at room temperature and were supplemented with buffer A containing protease inhibitors (Calbiochem protease inhibitor cocktail set VI). Cells were disrupted by five passes in the EmulsiFlex-C3 homogenizer (Avestin) at 100–1,500 bar at 4 °C. The suspension was centrifuged for 10 min at 13,300*g*, 4 °C, to remove cell debris and centrifuged again for 30 min at 46,000*g*, 4 °C, to collect the membranes. Membranes were washed once with 1 ml of buffer A and resuspended in 1 ml of buffer A containing 400 mM sucrose, flash-frozen in liquid nitrogen and stored at −80 °C.

Membranal proteins were separated on SDS–PAGE. The proteins in the gel were reduced with 3 mM dithiothreitol (DTT; 60 °C for 30 min), modified with 10 mM iodoacetamide in 100 mM ammonium bicarbonate (in the dark at room temperature for 30 min) and digested in 10% acetonitrile and 10 mM ammonium bicarbonate with modified trypsin (Promega) at a 1:10 enzyme-to-substrate ratio, overnight at 37 °C.

The resulting tryptic peptides were resolved by reverse-phase chromatography on a 0.075 × 200 mm fused-silica capillary column (J&W) packed with Reprosil reversed phase material (Dr Maisch). The peptides were loaded onto a homemade capillary column (25 cm, 75 µm inner diameter (ID)) packed with Reprosil C18-Aqua (Dr Maisch) in solvent A (0.1% formic acid in water). The peptide mixture was resolved at a flow rate of 0.15 µl min^{-1} with a linear gradient, 5–28%, of 95% acetonitrile with 0.1% formic acid for 30 min followed by a gradient of 28–95% for 15 min and finally 15 min at 95% acetonitrile with 0.1% formic acid. Mass spectrometry (MS) was performed in positive mode (m/z 350–1800) with repetitive cycles of full MS scans followed by high-collision-induced dissociation (at 35 normalized collision energy) of the 10 most dominant ions (>1 charges) selected from the first MS scan. A dynamic exclusion list was used with exclusion duration of 20 s.

The MS data were analysed using Proteome Discoverer 1.4 software with the Sequest algorithm against *Prochlorococcus* MIT9515 and P-TIM68 proteomes allowing a 1% false discovery rate.

Targeted proteomics. Prochlorococcus MIT9515 cells (300 ml) were grown at 21 °C to mid-logarithmic phase and were infected with P-TIM68 (MOI of 3). Samples (30 ml) were collected at each time point and centrifuged for 30 min at 13,300*g*, 4 °C washed in Pro99 medium, and centrifuged again. Pellets were frozen in liquid nitrogen and stored at −80 °C. Three biological repeats were carried out and seven time points were analysed during the infection process from both infected and uninfected cells.

To quantify the amount of PS proteins during the infection process, targeted proteomics with isotope-labelled peptides was used (JPT Peptide Technologies). The sequences of these peptides were unique to either the viral or the cyanobacterial proteins (Supplementary Table 4; isotopically labelled (C^{13} N^{15}) R or K at the C-termini of the peptides are highlighted in bold in the table).

Samples were resuspended in 200 µl of SDC buffer (2% wt/vol sodium deoxycholic acid in 50 mM ammonium bicarbonate), vortexed and sonicated to complete disruption of the pellet (~4 min).

Protein samples were reduced with 3 mM DTT for 45 min at 54 °C and modified with 12 mM iodoacetamide (30 min at room temperature). Samples were diluted with 50 mM ammonium bicarbonate at a 1:1 ratio, trypsinized at a 1:50 enzyme-to-protein ratio and incubated overnight at 37 °C. An additional trypsinization was carried out for 4 h. The samples were then spiked with a mix of heavy labelled peptides (Supplementary Table 4, 1 pmol of each peptide/5–10 µg extract). The samples were acidified with 1.5% formic acid to remove the SDC buffer before sample analysis by MS. The resulting peptides were desalted using C18 SpinColumns (Harvard Apparatus), dried, and resuspended in 0.1% formic acid.

The peptides were loaded onto a homemade capillary column (25 cm, 75 µm ID) packed with Reprosil C18-Aqua (Dr Maisch) in solvent A. The peptide mixture was resolved at a flow rate of 0.15 µl min^{-1} with a linear gradient (5 to 28%) of 95% acetonitrile with 0.1% formic acid for 120 min followed by a gradient of 28–95% for 5 min and 25 min at 95% acetonitrile with 0.1% formic acid. MS was performed as stated above.

The ratios between the isotopically labelled peptides and the unlabelled peptides were calculated based on the peak area of the extracted ion currents (XICs). Absolute quantities were determined by dividing the amount of heavy peptides with the normalized ratio (corrected for the extract concentration). In addition, the MS data were analysed and quantified using MaxQuant software 1.5.1.2 (www.maxquant.org) for peak identification and quantitation using the Andromeda search engine, against *Prochlorococcus* MIT9515 and P-TIM68 proteomes allowing a 1% false discovery rate.

Phylogenetic analysis. Multiple sequence alignment of the PsaA protein (94 shared amino acid positions; Supplementary File 3) was constructed using ClustalX v2.1 (ref. [44]). The alignment contains only 94 shared amino acids positions due to the short length of some of the GOS *psaA* reads used. Maximum-likelihood (ML) phylogenetic trees were constructed using the phylogeny.fr pipeline[45], which included PhyML v3.0 (ref. [46]) and the WAG substitution model for amino acids[47]. One hundred bootstrap replicates were conducted.

Photosynthetic measurements. Fluorescence-based photosynthetic parameters were measured with an in-house assembled fluorescence set-up[48]. The maximal quantum yield of PSII was calculated as $(F_m - F_0)/F_m$, where F_m and F_0 are the fluorescence levels in dark adapted samples and after a saturating light pulse, respectively[49]. The quantum yield of PSII (ΦPSII) in the light was calculated as $(F_m' - F)/F_m'$, where F and F_m' are the steady-state and maximum fluorescence intensities in light-acclimated cells, respectively. The relative photosynthetic electron transfer rate $rETR_{PSII}$ was calculated as ΦPSII*I, where I is the irradiance and was measured at ten irradiances (steps of 30 s illumination for each light irradiance). $rETR_{max}$ was defined as the highest value of the light saturation curve of $rETR_{PSII}$.

Absorption spectroscopy measurements were performed with whole cells of *Prochlorococcus* MIT9515 using an in-house assembled spectrophotometer as reported previously[50]. The probe light in the visible region ($\lambda > 410$ nm) was generated by a frequency-tripled Nd:YAG laser (Surelite II, Continuum) pumping a computer-controlled optical parametric oscillator (Panther, Continuum). The samples were excited at 700 nm with a laser pulse generated by another optical parametric oscillator (SLOPO, Continuum) at an intensity of ~100 µJ per pulse. Cyclic photosynthetic electron flow activity was measured using a JTS-10 spectrophotometer (Bio-Logic Science Instruments SAS) set as in ref. [51]. This was achieved by following absorption changes at 705 nm, reflecting the oxidation/reduction of P700 (the reaction centre of PSI). The detecting flashes were provided by white light-emitting diodes (LEDs) narrowed to 705 ± 10 nm, using an interference filter. Actinic light was generated by an orange-red LED (639 nm) ring situated in front of the sample. The negative rise in the absorbance of P700 was fitted with a simple first-order exponential model in Matlab to extract the rate constant in units of s^{-1}.

***vPSI-7* and P-TIM68 recruitment from *Tara* Oceans stations.** *Tara* Oceans expedition microbiome and virome metagenomic data sets[7,52] were obtained from EBI-ENA (accession no.: EBI-ENA PRJEB402) on 7 August 2015. Paired-end Illumina reads from each of the 399 data sets retrieved were mapped onto an index built from 207 gene sequences using Bowtie2 (v2.2.6)[53] set to the option '-a –very-sensitive-local'. The sequences, in FASTA format, include the following genes: viral PSI *psaA* and *psaB*, viral PSII *psbA and psbD*, cyanomyophage and pelagiphage *g20* and cyanophage *g23* (Supplementary Table 5). Bowtie2 alignments were sorted and converted to BAM format using SAMTools[54]. BAM files were processed using BP-1.py, a python script based on HTSeq-count[55] that filters the alignments by their per cent identity, with a cutoff value of 80% identity over an alignment length of at least 60 nt. Filtered reads were used as the query for BLASTN (v2.2.31) with default parameters. The database was composed of 2,499 cyanobacterial and plastid genomes or contigs from NCBI RefSeq[56] release 74 (accessed 2 February 2016) as well as the viral sequences used for bowtie mapping. Reads with cyanobacteria or plastid best hits were discarded. The number of viral reads for each gene at 80% and 98% identity was normalized according to ref. [57], and the *g20* final count was obtained by subtracting the pelagiphage *g20* normalized counts.

The list of genes used for Bowtie2 mapping and the genomes comprising the BLASTN database are provided in Supplementary Table 5. The python script BP-1.py used to process the BAM files is available at https://github.com/BejaLab/p-tim68/.

Correlation analysis between *Prochlorococcus* HLIV-A and cyanophage P-TIM68. A total of 33 samples for the bacterial fraction (>0.22 µm) were used to perform a linear regression analysis between the abundance of *Prochlorococcus* HLIV-A[29] and the abundance of P-TIM68, using the 'lm' function implemented in the R stats package. The relative abundance of P-TIM68 was calculated as the P-TIM68:*g20*+*g23* ratio, determined from the normalized counts for *psaA* from P-TIM68 at an identity cutoff of 97% and the sum of normalized counts from *g20* and *g23* at an identity cutoff of 80% (Supplementary Table 1 and section '*vPSI-7* and P-TIM68 recruitment from *Tara* Oceans stations'). The TARA_052 sample was excluded from the regression analysis because it was considered an outlier based on the Cook's distance ($D_i > 1$). Because the abundances of *Prochlorococcus* from clades HLIII-A and HLIV-A are directly and strongly correlated ($R^2 = 0.97$), a correlation between P-TIM68 and HLIV-A abundances is also representative of a correlation with HLIII-A.

Reassembly of *Tara* Oceans scaffolds. *Tara* Oceans reads for the viral fraction (<0.22 µm) were assembled using the IDBA-UD assembler[58] with default parameters. The resulting assemblies went through an automatic mis-assembly detection and correction procedure using in-house software (Sharon *et al.*, in preparation). The procedure for mis-assembly correction takes advantage of read mapping to the assembled scaffolds and contigs as follows. First, reads from which an assembly was generated were mapped to the assembly using read-mapping software (bowtie2; ref. [53]) with the (--sensitive flag) to generate a SAM file of mapped reads. Next, local mis-assemblies such as single nucleotide polymorphisms and short insertions/deletions were identified by locating read mapping that consistently differed from the assemblies. Finally, we used paired-end information to identify global mis-assemblies by locating regions in which only one read of each pair was mapped while the other was flanking, or mapped to a different sequence or region on the same sequence. Where sufficient numbers of such pairs existed, the sequences were broken into two sequences.

Data availability. The data that support the findings of this study are available from the corresponding author upon request. The genome of P-TIM68 has been deposited in GenBank under accession code KM359505 and the raw reads have been deposited in the GenBank Sequence Read Archive (BioSample accession code SAMN06618437).

Received: 5 February 2017; Accepted: 28 June 2017;
Published online: 7 August 2017

References

1. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
2. Millard, A., Clokie, 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).
3. Lindell, D. et al. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc. Natl Acad. Sci. USA* **101**, 11013–11018 (2004).
4. 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).
5. Sharon, I. et al. Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J.* **5**, 1178–1190 (2011).
6. Sharon, I. et al. Photosystem-I gene cassettes are present in marine virus genomes. *Nature* **461**, 258–262 (2009).
7. Brum, J. et al. Patterns and ecological drivers of ocean viral communities. *Science* **348**, 1261498 (2015).
8. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Shedding new light on viral photosynthesis. *Photosynth. Res.* **126**, 71–97 (2015).
9. 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).
10. Clokie, M. R. J. et al. Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ. Microbiol.* **8**, 827–835 (2006).
11. Bragg, J. G. & Chisholm, S. W. Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS ONE* **3**, e3550 (2008).
12. Hellweger, F. L. Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. *Environ. Microbiol.* **11**, 1386–1394 (2009).
13. Millard, A. D., Zwirglmaier, K., Downey, M. J., Mann, N. H. & Scanlan, D. J. Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ. Microbiol.* **11**, 2370–2387 (2009).
14. Sullivan, M. B. et al. Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ. Microbiol.* **12**, 3035–3056 (2010).
15. Roitman, S. et al. Closing the gaps on the viral photosystem-I *psaDCAB* gene organization. *Environ. Microbiol.* **17**, 5100–5108 (2015).
16. Torabi, S. et al. PsbN is required for assembly of the photosystem II reaction center in *Nicotiana tabacum*. *Plant Cell* **26**, 1183–1199 (2014).
17. 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).
18. Ignacio-Espinoza, J. C. & Sullivan, M. B. Phylogenomics of T4 cyanophages: lateral gene transfer in the 'core' and origins of host genes. *Environ. Microbiol.* **14**, 2113–2126 (2012).
19. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Viruses inhibit CO_2 fixation in the most abundant phototrophs on Earth. *Curr. Biol.* **26**, 1585–1589 (2016).
20. Rusch, D. B. et al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through the eastern tropical Pacific. *PLoS Biol.* **5**, e77 (2007).
21. 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).
22. Comeau, A. M., Arbiol, C. & Krisch, H. M. Gene network visualization and quantitative synteny analysis of more than 300 marine T4-like phage scaffolds from the GOS metagenome. *Mol. Biol. Evol.* **27**, 1935–1944 (2010).
23. Wang, K. & Chen, F. Prevalence of highly host-specific cyanophages in the estuarine environment. *Environ. Microbiol.* **10**, 300–312 (2008).
24. Doron, S. et al. Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J.* **10**, 1437–1455 (2016).
25. Wilson, W. H., Carr, N. G. & Mann, N. H. The effect of phosphate status on the kinetics of cyanophage infection of the oceanic cyanobacterium *Synechococcus* sp. WH7803. *J. Phycol.* **32**, 506–516 (1996).
26. Kirzner, S., Barak, E. & Lindell, D. Variability in progeny production and virulence of cyanophages determined at the single-cell level. *Environ. Microbiol. Rep* **8**, 605–613 (2016).
27. Lindell, D. et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**, 83–86 (2007).
28. Clokie, M. R. J. & Mann, N. H. Marine cyanophages and light. *Environ. Microbiol.* **8**, 2074–2082 (2006).
29. Farrant, G. K. et al. Delineating ecologically significant taxonomic units from global patterns of marine picocyanobacteria. *Proc. Natl Acad. Sci. USA* **113**, E3365–E3374 (2016).
30. Fraser, J. M. et al. Photophysiological and photosynthetic complex changes during iron starvation in *Synechocystis* sp. PCC 6803 and *Synechococcus elongatus* PCC 7942. *PLoS ONE* **8**, e59861 (2013).
31. Wyman, M., Gregory, R. P. & Carr, N. G. Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**, 818–820 (1985).
32. Lindell, D., Padan, E. & Post, A. F. Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J. Bacteriol.* **180**, 1878–1886 (1998).
33. Moore, L. R. et al. Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol. Oceanogr. Methods* **5**, 353–362 (2007).
34. Haas, A. F. et al. Unraveling the unseen players in the ocean—a field guide to water chemistry and marine microbiology. *J. Vis. Exp.* **5**, e52131 (2014).
35. Béjà, O., Fridman, S. & Glaser, F. Viral clones from the GOS expedition with an unusual photosystem-I gene cassette organization. *ISME J.* **6**, 1617–1620 (2012).
36. Hevroni, G., Enav, H., Rohwer, F. & Béjà, O. Diversity of viral photosystem-I *psaA* genes. *ISME J.* **9**, 1892–1898 (2015).
37. Sabehi, G. et al. A novel lineage of myoviruses infecting cyanobacteria is widespread in the oceans. *Proc. Natl Acad. Sci. USA* **109**, 2037–2042 (2012).
38. Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829 (2008).
39. Casjens, S. R. & Gilcrease, E. B. Determining DNA packaging strategy by analysis of the termini of the chromosomes in tailed-bacteriophage virions. *Methods Mol. Biol.* **502**, 91–111 (2009).
40. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
41. Hyatt, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119 (2010).
42. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
43. Zinser, E. R. et al. *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl. Environ. Microbiol.* **72**, 723–732 (2006).
44. Larkin, M. A. et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
45. Dereeper, A. et al. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res.* **36**, W465–W469 (2008).
46. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
47. Whelan, S. & Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **18**, 691–699 (2001).
48. Rappaport, F., Béal, D., Joliot, A. & Joliot, P. On the advantages of using green light to study fluorescence yield changes in leaves. *Biochim. Biophys. Acta* **1767**, 56–65 (2007).
49. Genty, B., Briantais, J. M. & Baker, N. R. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. Acta* **990**, 87–92 (1989).
50. Béal, D., Rappaport, F. & Joliot, P. A new high-sensitivity 10-ns time-resolution spectrophotometric technique adapted to *in vivo* analysis of the photosynthetic apparatus. *Rev. Sci. Instrum.* **70**, 202–207 (1999).
51. Joliot, P. & Joliot, A. Quantification of cyclic and linear flows in plants. *Proc. Natl Acad. Sci. USA* **102**, 4913–4918 (2005).
52. Sunagawa, S. et al. Structure and function of the global ocean microbiome. *Science* **348**, 1261359 (2015).
53. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
54. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
55. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
56. O'Leary, N. A. et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).
57. Sharon, I., Pati, A., Markowitz, V. M. & Pinter, R. Y. in *Research in Computational Molecular Biology* Vol. 5541 (ed. Batzoglou, S.) 496–511 (Springer, Berlin, Heidelberg, 2009).
58. Peng, Y., Leung, H. C., Yiu, S. M. & Chin, F. Y. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**, 1420–1428 (2012).

Acknowledgements

The authors thank N. Keren, N. Adir and J. Golbeck for their insight regarding photosynthesis in cyanobacteria, O. Kleifeld for preliminary proteomics results, I. Pekarsky and M. Rosenberg for help with TEM imaging and Béjà and Lindell laboratory members for continuous discussions. The authors also thank L. Garczarek for providing cyanobacteria abundance data. This work was funded by a European Commission ERC Advanced Grant (no. 321647), the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme FP7/2007-2013/ under

REA Grant Agreement No. 317184, an Israel Science Foundation grant (no. 580/10) and the Louis and Lyra Richmond Memorial Chair in Life Sciences to O.B., a European Commission ERC starting grant (no. 203406) to D.L. and the Technion's Lorry I. Lokey Interdisciplinary Center for Life Sciences and Engineering and the Russell Berrie Nanotechnology Institute. This is contribution number 54 of *Tara* Oceans. This paper is dedicated to the memory of F.R. (CNRS), who sadly passed away before the paper was finalized.

Author contributions

S.F., D.L. and O.B. designed the project and the experiments. S.F. isolated the phage and, together with S.L. and O.A., performed laboratory experiments. F.L.R. collected the phage concentrate. T.Z. performed proteomics. J.F.-U., I.S., A.P., C.L.D., F.M.C.-C., P.S., S.G.A. and O.B. performed bioinformatic analyses. S.L., O.L., I.Y., F.S., B.B. and F.R. performed photosynthetic measurements. O.B. and D.L. wrote the manuscript with contributions from all authors to data analysis, figure generation and the final manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at doi:10.1038/s41564-017-0002-9.

Reprints and permissions information is available at www.nature.com/reprints.

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

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