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

Svetlana Fridman, José Flores-Uribe, Shirley Larom, Onit Alalouf, Oded Liran, Iftach Yacoby, Faris Salama, Benjamin Bailleul, Fabrice Rappaport, Tamar Ziv, et al.

A novel 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], Francisco M. Cornejo-Castillo[6], Alon Philosof[1], Christopher L. Dupont[7], Pablo Sánchez[6], Silvia G. Acinas[6], Forest L. Rohwer[8], Debbie Lindell[1]* & Oded Béjà[1]*

[1]Faculty of Biology, and [4]Smoler Proteomics Center, 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. [5]Migal Galilee Research Institute, Kiryat Shmona, 11016, Israel. Tel Hai College, Upper Galilee 12210, Israel. [6]Department of Marine Biology and Oceanography, Institute of Marine Sciences (ICM), CSIC, Barcelona, Spain. [7]Microbial and Environmental Genomics, J. Craig Venter Institute, La Jolla, California 92037, USA. [8]Department of Biology, San Diego State University, San Diego, California 92182, USA.

[§]This paper is dedicated to the memory of Prof. Fabrice Rappaport (CNRS), who sadly passed away before the paper was finalized.

*Correspondence and requests for materials should be addressed to D.L. (dlindell@tx.technion.ac.il) and O.B. (beja@tx.technion.ac.il).

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). While photosystem-II (PSII) genes are common in both cultured cyanophages and environmental samples[1-4], viral photosystem-I (vPSI) genes have only been detected in environmental samples to date[5,6]. Here we 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: PSII, b_6f, PSI and ATPase, to convert sunlight into chemical energy. Linear electron transfer begins with PSII catalyzing the transfer of electrons, derived from water, to the plastoquinone pool. PSI transfers electrons from plastocyanin/cytochrome c_6 to ferredoxin thereby generating reducing power needed for CO_2 fixation in the form of NADPH. Then, the b_6f complex couples the electron transfer from the reduced plastoquinone (QH_2) to the oxidized plastocyanin/cytochrome c_6 to the pumping of protons in the lumen. Lastly, 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 photosystem-I (PSI) reaction center 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,D* (*vPSI-7*), and the other containing four PSI genes, *psaD,C,A,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 to date. Indeed this phage was dubbed the “mystery phage” by Puxty et al.[8]. Here we developed a targeted isolation strategy in an attempt to find such a PSI-carrying cyanophage (Supplementary Fig. S1). 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, eighteen 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 host (see Methods and Supplementary Fig. S1). 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. S2a). The P-TIM68 genome contains a suite of DNA replication and virion structural genes characteristic of the archetype T4 (supplementary File S1) 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 percent identity between the structural, replication and nucleotide metabolism genes of the two phages (Supplementary Fig. S2b).

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; Supplementary Fig. S2a). 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 homolog coding for a small protein required for the repair and assembly of PSII reaction centers[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 7-gene PSI cassette, one of the NDH-1 genes and the two reaction center genes of PSII (Fig. 1c and Supplementary Fig. S2b). 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 neighbor previously reported PSI environmental scaffolds[6]. Analyses of data from the *Tara* Oceans[7] and the Global Ocean Sampling (GOS)[20,21] expeditions revealed the presence of two scaffolds with the exact same gene organization as in P-TIM68 over their entire length of 3.9 and 17.8 kb (supplementary File S2). 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 S1). We then estimated the abundance of P-TIM68 (≥98% identity) and other *vPSI-7* (≥80% identity) phages in the *Tara* Oceans dataset (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, supplementary Table S1). The P-TIM68 genotype composed on average 22% of these *vPSI-7* phages (average from all stations containing more than 20 *psaA* reads; proportions labeled in red in Fig. 2a).

Characterisation of the infection cycle of the P-TIM68 phage revealed a long latent period of 10-12 hours (Supplementary Fig. S3) and a burst size of 12±3 (n=3) infective phages/cell. Similarly, other T4-like cyanophages have

latent periods ranging from 4-18 h[19,23,24] and burst sizes that range from 21-45 phages/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 hours 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 hours after infection (supplementary Table S2). 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 homolog to 43% of the total D1 proteins in infected cells by the end of the latent period (Supplementary Fig. S4a), thus almost completely restoring the amount of D1 proteins to its 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. S4b). Since only 70% of host cells were infected in these experiments, the level of phage PsaA protein corresponds to 8.6% of 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 surface layers of tropical waters.

To characterize the host's photosynthetic activity during P-TIM68 infection, 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$\cdot m^{-2}\cdot s^{-1}$. Throughout the 10 hour 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 6-8 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, *i.e.* an increased cyclic electron flow around PSI, is induced during infection. Such increased cyclic electron flow could generate 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 Kelvin) showed no changes in PSII/PSI stoichiometry between infected and uninfected cells (supplementary Fig. S5a). In addition, absorption cross-section of PSII did not change during infection (supplementary Fig. S5b). Therefore, these results suggest that there is increased cyclic electron flow around PSI as a result of P-TIM68 infection.

While previous metagenomic data had indicated the likely existence of PSI carrying phages in the environment, the isolation of P-TIM68 provides definitive proof that the PSI cassettes are of cyanophage origin and that a phage encoding proteins capable of forming an entire PSI complex are present in the oceans and can infect *Prochlorococcus*. Furthermore, the physiological characterisation 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] (see inset in Fig. 2 for 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. *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 sea water-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$\cdot m^{-2}\cdot s^{-1}$ 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 15m 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 Haas et al.[34]. Cyanobacteria were challenged with these concentrates in 10-fold serial dilutions in 96-well plates. Wells showing lysis were then analyzed 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. S1) 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 4 rounds of plaque purification and PCR screening until all plaques were positive for the viral *psaC-psaA* gene arrangement. While 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.

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 according to Hevroni et al.[36]. Primers and PCR conditions for reactions *psaJF-psaC*, *psaJF-psaA* and *psaB-psaD* were according to Sharon et al.[6].

Transmission Electron Microscopy (TEM). A cyanophage lysate was filtered using a 0.2 *μ*m pore-sized filter and then brought to a final concentration of ~10^9 pfu/mL by centrifugation using a 100 KDa molecular weight cutoff (ca. 10 nm sized pores) centrifugal filter (Millipore). The phage lysate was then CsCl-purified according to Sabehi et al.[37]. For TEM analysis, 20 *μ*l of the CsCl-purified phage concentrate were 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 Nextera XT library preparation kit. Raw reads were deposited in GenBank Sequence Read Archive (BioSample accession SAMN06618437) and *de-novo* assembled using the velvet 1.2.10[38] program with a k-mer size=73. Assembly coverage was 1790-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 number KM359505. Gene calling was carried out using Prodigal 2.6.1[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 manually curated.

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 Sabehi et al.[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, 0.5M NaCl; pH 8.0) and frozen at -80^{o}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^{o}C to mid-logarithmic phase and were infected with P-TIM68 at a multiplicity of infection (MOI) of 3. Samples were collected every 30 min for the first 4 hours after infection and then every 2 hours until 10 hours after infection. The experiment was repeated 3 times. Nine mL of infected or uninfected cells were filtered onto 0.22 µm Durapore filters (Millipore). The filters were transferred to tubes containing 1 mL of RNAlater (Ambion), frozen and stored at -80°C. Total RNA was extracted using the mirVana RNA isolation kit (Ambion). Genomic DNA was removed using the Turbo DNA-free kit (Ambion). cDNA was made using the 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 1X 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 *µ*l. Reactions were carried out on a LightCycler 480 Real-Time PCR System (Roche). The program used was 1 step of 95 °C for 15 min followed by 45 cycles of amplification, each including 10 sec denaturation at 95 °C, and a 30 sec 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 shown in Table 1.

Table 1: Real-time PCR primers and probes used in this study

Target gene	Sequence 5' – 3'	Probe 5' – 3'
Region from *psaJF* to *psaA*	Fwd: TGTAAAGTATCCCCTAGAT Rev: TGCTTACAGGGACAGGATTTTT	CTCCACCA
psaB	Fwd: CATACACACAGGCAGGTTCTAC Rev: CTTGGTTTCAGATGTAATTGAG	CTGGTGGT
psbA	Fwd: AGCATGCTTCGTTGTTGCATTTATT Rev: AGAAATGATGTTGTTACCATAGAGA	GCCATCAA
DNA polymerase	Fwd: AGTGGCACAGAAGAGGATGTAAT Rev: GGAGTTCCCTTTGCATAGATTG	CTGGTGGT
Base plate	Fwd: TGCATCACAAAATAGGGCAGT Rev: TAGATTCTAGGCACTAACGCTTCA	GCAGCAGA

Early gene32	Fwd: CACATTTATCACTTAAGATGGATGGT Rev: TCTCAGGGTGAGTCCCAAAT	CAGCAGGT
rnpB	Fwd: AAGTCCGGCTCCCATA Rev: GCACTATCCTCACGGTTGC	TTCCCAGT
rbcL	Fwd: GTCAGATATGCCAATCATCATGC Rev: GCCAAGCCAGTATTTGCAGT	CTGGTGGT
g20	Fwd: CCTAAGGAGCAACTACACCA Rev: TGATGTGCAATATGTAATTGAAT	TGGTGGAA

Shotgun proteomics: *Prochlorococcus* cells (1.5 L) were grown at 21^{o}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 x g, 4^{o}C. The pellet was resuspended in buffer A (50 mM Tris/HCl, 5 mM $MgCl_2 \cdot 6H_2O$, 10 mM NaCl, pH 8.0), frozen in liquid nitrogen and stored at -80^{o}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 5 passes in the EmulsiFlex-C3 homogenizer (Avestin) at 100-1500 bar at 4^{o}C. The suspension was centrifuged for 10 min at 13,300 x g, 4^{o}C, to remove cell debris, and centrifuged again for 30 min at 46,000 x g, 4^{o}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^{o}C.

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

The resulting tryptic peptides were resolved by reverse-phase chromatography on a 0.075 X 200-mm fused silica capillary column (J&W) packed with Reprosil reversed phase material (Dr Maisch GmbH, Germany). The peptides were loaded onto a homemade capillary column (25 cm, 75 micron ID) packed with Reprosil C18-Aqua (Dr Maisch GmbH, Germany) in solvent A (0.1% formic acid in water). The peptide mixture was resolved at a flow rate of 0.15 *µ*l/min with a linear gradient, 5 to 28%, of 95% acetonitrile with 0.1% formic acid for 30 minutes followed by a gradient of 28 to 95% for 15 min and finally 15 minutes at 95% acetonitrile with 0.1% formic acid. Mass spectrometry was performed in a 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 mass spectrometry data was analyzed using Proteome Discoverer 1.4 software with 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^{o}C to mid-logarithmic phase and were infected with P-TIM68 (MOI of 3). Thirty mL samples were collected at each time point and centrifuged for 30 min at 13,300 x g, 4^{o}C washed in Pro99 medium, and centrifuged again. Pellets were frozen in liquid nitrogen and stored at -80^{o}C. Three biological repeats were carried out and seven time points were analyzed during the infection process from both infected and uninfected cells.

In order to quantify the amount of the photosystem proteins during the infection process, targeted proteomics with isotope labeled peptides was used (JPT Peptide Technologies). The sequence of these peptides were unique to either the viral or the cyanobacterial proteins (Table 2). The isotopically labeled (C^{13} N^{15}) R or K at the C-termini of the peptides are highlighted in bold.

Table 2: Isotopically Labeled peptides used in this study

	P-TIM68	MIT9515
PsbA	ETTGLESQNYGY**K**	ETTETESQNYGY**K**
PsaB	DLGYSFPCDGPG**R**	DLGYAFPCDGPG**R**
PsaA	LAPTIQP**R**	VAPTIQP**R**
PsaJF	LLETNVTVSP**R**	

Samples were resuspended in 200 *µ*l of SDC Buffer (2%(w/v) sodium deoxycholic acid in 50 mM ammonium bicarbonate), vortexed and sonicated to complete disruption of the pellet (approx. 4min).

Protein samples were reduced with 3 mM DTT for 45 min at 54ºC 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^{o}C. An additional trypsinization was done for 4 hours. The samples were then spiked with a mix of heavy labeled peptides (Table 2, 1 *p*mole of each peptide/5-10 µg extract). The samples were acidified with 1.5% formic acid in order to remove the SDC

buffer prior to sample analysis by mass spectrometry. 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 micron ID) packed with Reprosil C18-Aqua (Dr Maisch GmbH, Germany) in solvent A. The peptide mixture was resolved at a flow rate of 0.15 *µ*l/min with a linear gradient (5 to 28%) of 95% acetonitrile with 0.1% formic acid for 120 minutes followed by a gradient of 28 to 95% for 5 min and 25 minutes at 95% acetonitrile with 0.1% formic acid. Mass spectrometry was performed as stated above.

The ratios between the isotopically labeled peptides and the unlabeled 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 mass spectrometry data were analyzed and quantified using the 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 acids positions; supplementary File S3) was constructed using ClustalX v2.1[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[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 setup[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 seconds illumination for each light irradiance). The $rETR_{max}$ was defined as the highest value of the light saturation curve of the $rETR_{PSII}$.

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 (λ > 410 nm) is 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 about 100 μJ/pulse. Cyclic photosynthetic electron flow activity was measured using a JTS-10 spectrophotometer (Bio-Logic Science Instruments SAS, France) set as in Joliot & Joliot[51]. This was

453 achieved by following absorption changes at 705 nm, reflecting the
454 oxidation/reduction of P700. The detecting flashes were provided by white
455 LEDs narrowed to 705 ± 10 nm, using interference filter. Actinic light came
456 from orange-red leds (639 nm) ring situated in front of the sample. The
457 negative rise in the absorbance of P700 was fitted with a simple first order
458 exponential model in Matlab in order to extract the rate constant in [sec^{-1}].

459

460 ***vPSI-7* and P-TIM68 recruitment from *Tara* Oceans stations.** *Tara* Oceans
461 expedition microbiome and virome metagenomic datasets[7,52] were obtained
462 from EBI-ENA (Accession: EBI-ENA: PRJEB402) on August 7th 2015. Paired-
463 end Illumina reads from each of the 399 datasets retrieved were mapped onto
464 an index built from 207 gene sequences using Bowtie2 (v2.2.6)[53] set to the
465 option "-a –very-sensitive-local". The sequences, in FASTA format, include
466 the following genes: viral PSI *psaA* and *psaB*, viral PSII *psbA and psbD*,
467 cyanomyophage and pelagiphage *g20*, and cyanophage *g23* (Supplementary
468 Table S3). Bowtie2 alignments were sorted and converted to BAM format
469 using SAMTools[54]. BAM files were processed using BP-1.py, a python script
470 based on HTSeq-count[55] that filters the alignments by their percent-identity,
471 with a cutoff value of 80% identity over an alignment length of at least 60 nt.
472 Filtered reads were used as query for BLASTN (v2.2.31) with default
473 parameters. The database was composed of 2,499 cyanobacterial and plastid
474 genomes or contigs from the NCBI RefSeq[56] release 74 (accessed on
475 February 2nd 2016) as well as the viral sequences used for the bowtie
476 mapping. Reads with cyanobacteria or plastid best hits were discarded. The
477 number of viral reads for each gene at 80% and 98% identity was normalized

according to Sharon et al.[57] and *g20* final count was obtained by subtracting the pelagiphage *g20* normalized counts.

The list of genes used for the Bowtie2 mapping and the genomes comprising the BLASTN database can be found in Supplementary Table S3. 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 P-TIM68:*g20*+*g23* ratio, determined from the normalized counts for *psaA* from P-TIM68 at identity cut-off 97% and the sum of normalized counts from *g20* and *g23* at identity cut-off of 80%; Supplementary Table S1 and above "*vPSI-7* and P-TIM68 recruitment from *Tara* Oceans stations" method section). 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). Since 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* Ocean reads for the viral fraction (<0.22 *μ*m) were assembled using the idba-ud assembler[58] with default parameters. 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[53]) with the (--sensitive flag) to generate a SAM file of mapped reads. Next, local mis-assemblies such as SNPs and short insertions/deletions were identified by locating read mapping that consistently differed to the assemblies. Finally, we used paired-end information in order to identify global mis-assemblies by locating regions in which only one read of each pair is mapped while the other is flanking, or mapped to a different sequence or region on the same sequence. Given that sufficient numbers of such pairs exist the sequences are broken into two sequences in these case.

ACKNOWLEDGMENTS

We thank Nir Keren, Noam Adir and John Golbeck for their insight regarding photosynthesis in cyanobacteria, Oded Kleifeld for preliminary proteomics results, Irena Pekarsky and Mira Rosenberg for help with TEM imaging, and Béjà and Lindell lab members for continuous discussions. We also thank Laurence Garczarek for providing cyanobacteria abundance data. This work was funded by a European Commission ERC Advanced Grant (no. 321647), 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 X of the *Tara* Oceans.

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.

541

542 **Competing financial interests.** The authors declare no competing financial
543 interests.

544

Figure Legends

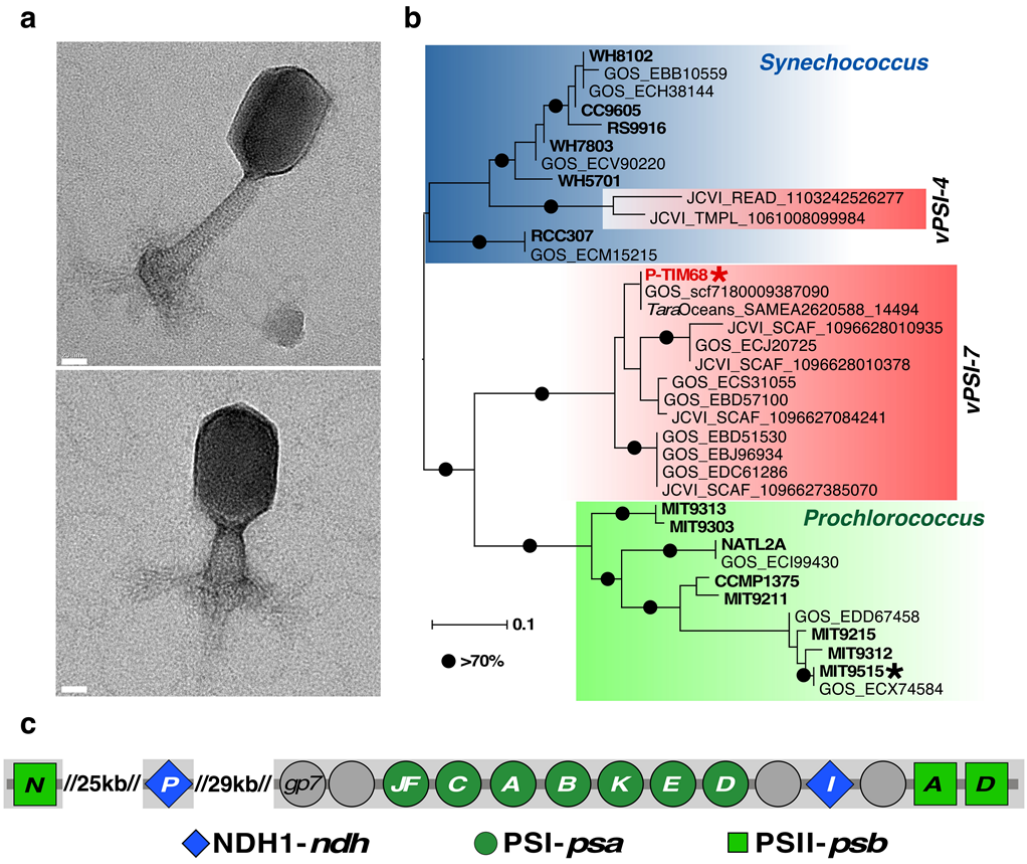
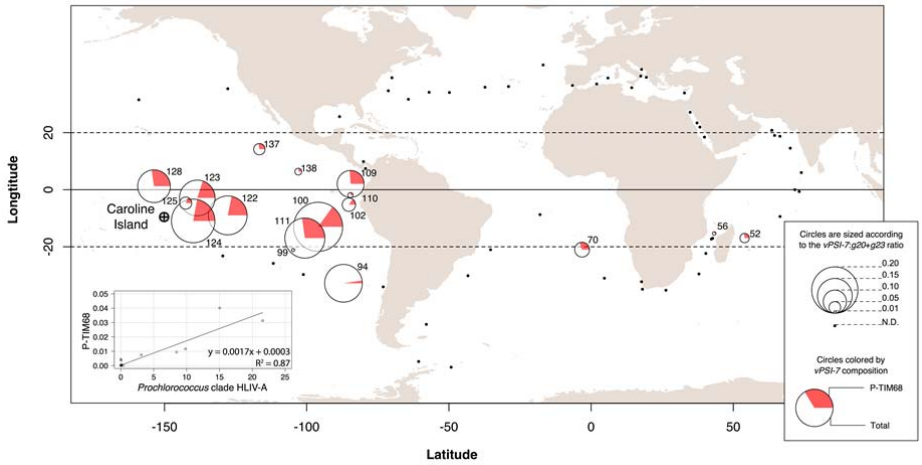


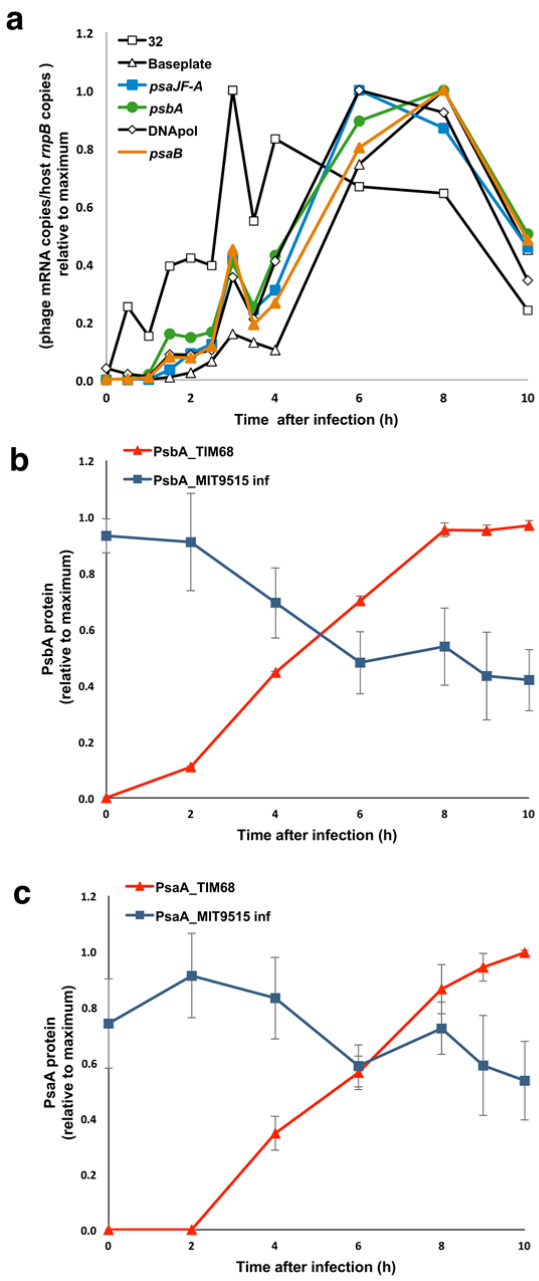
Figure 1. P-TIM68 gene organization, phylogeny and morphology. (**a**), A schematic physical map of the photosynthesis region in the P-TIM68 genome: PSI and PSII genes (green), NDH-1 genes (blue), and viral *gp7* and hypothetical genes (grey). (**b**), A maximum-likelihood phylogenetic tree of PsaA proteins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



559

560 **Figure 2. Global distribution and relative abundance of P-TIM68 and other *vPSI-7***
561 **phages.** The area of each circle is scaled to represent the fraction of *vPSI-7* phages (based
562 on the viral *psaA* and *psaB* genes) out of the total cyanomyophages (based on the viral *g20*
563 and *g23* genes) at different *Tara* Oceans stations. N.D. stands for not detected (stations with
564 less than 20 *vPSI-7* reads). Caroline Island, from which P-TIM68 was isolated, is marked with
565 a circled black cross. Proportions of P-TIM68 phage out of *vPSI-7* phages at each *Tara*
566 Oceans station is shown in red. Exact numbers of reads recruited at each depth and the
567 filtering scheme can be found in supplementary Table S1. The equator is shown as a solid
568 line while latitudes 20ºN and 20ºS are shown as dashed lines. Station numbers are shown for
569 stations in which *vPSI-7* reads were detected. For clarity, stations 122-125 are not shown at
570 their exact locations. Station coordinates as well as normalized read recruitment numbers can
571 be found in supplementary Table S1. Inset – P-TIM68 abundance (represented as the ratio of
572 P-TIM68 out of total cyanomyophages) versus *Prochlorococcus* HLIV-A clade abundance[29].
573 Abundances of P-TIM68 and *Prochlorococcus* HLIV-A were significantly correlated ($P < 10^{-15}$;
574 regression line, regression equation and R^2 value are shown). Station coordinates as well as
575 normalized read recruitment numbers can be found in supplementary Table S1.

576



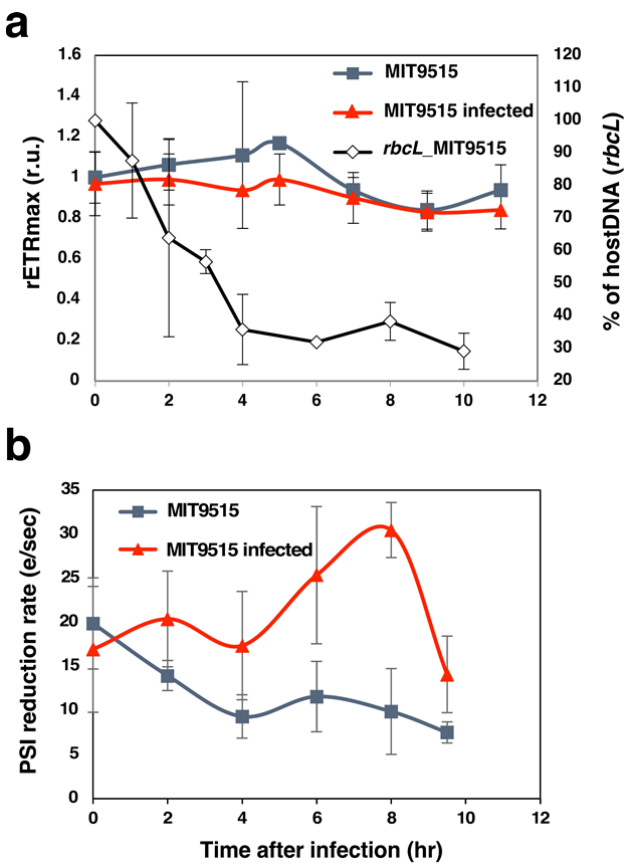
577

578

579 **Figure 3. Expression of P-TIM68 PSI and PSII genes in infected *Prochlorococcus***
580 **MIT9515 cells. (a)** mRNA transcript levels of representative viral early (*gene 32*), middle
581 (DNA polymerase), and late (baseplate) genes, and viral PSII (*psbA*) and PSI (*psaB* and
582 *psaJF*) genes during the 10 hour latent period. Results shown are from a single biological
583 replicate and are representative of all 3 independent experiments. Levels of phage and host

584 **(b)** PsbA (PSII) and **(c)** PsaA (PSI) protein levels during the latent period of infection. The
585 data shown in panels **b** and **c** were normalized to expression in uninfected control cells. Error
586 bars denote S.D. of three biological experiments. P values for the paired samples t-test of the
587 first two time points and the last time point are <0.001 for the decrease in host PSII PsbA
588 levels, and 0.016 for the decrease in the host PSI PsaA levels. All data are presented relative
589 to maximal levels for each gene and protein independently. Protein data not normalized to
590 uninfected control cells are shown in supplementary Fig. S4.

591

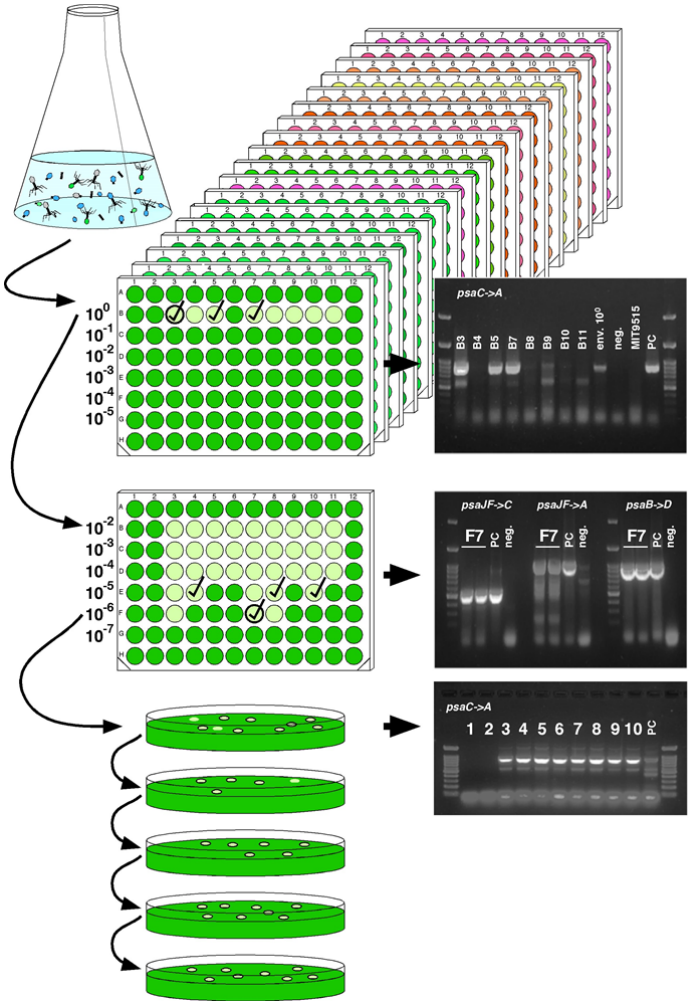


592

593 **Figure 4. Changes in PSII and PSI photosynthetic parameters during infection of**
594 ***Prochlorococcus* MIT9515 with phage P-TIM68. (*a*)** $rETR_{max}$ and host DNA degradation;
595 $rETR_{max}$ is normalized to the $rETR_{max}$ of the control *Prochlorococcus* culture before infection,
596 and corresponds to the mean of 4 independent biological samples ± S.D. Timing of host DNA
597 degradation is presented as percent of maximal genomic DNA. Intracellular host DNA was
598 quantified by real-time PCR targeting the host *rbcL* gene, (n=3 independent biological
599 samples ± S.D.) **(*b*)** PSI reduction rate, mean of 3 independent biological replicates ± S.D.

600

Supplementary Figures

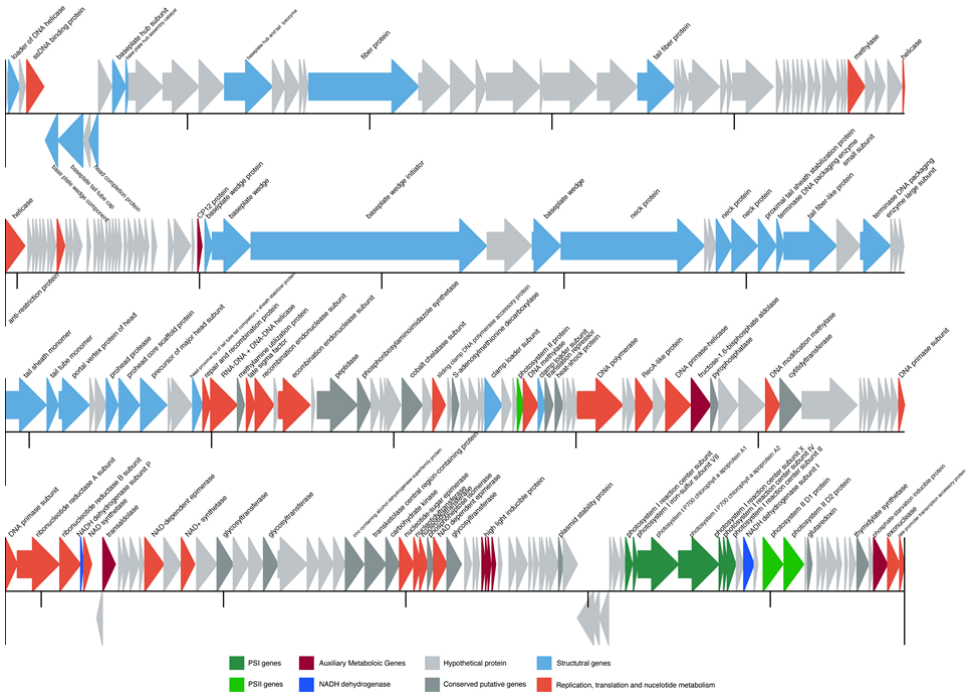


Supplementary Figure S1. P-TIM68 isolation strategy. Top, various *Synechococcus* and *Prochlorococcus* strains were challenged with serial dilutions of environmental phage concentrate from the Pacific Ocean in 96-well plates. Wells showing lysis were screened for the putative viral gene organization, *psaC-psaA*. One lysate (labeled with thick borders) out of 3 positive wells (√) from the *Prochlorococcus* MIT9515 plate was serially diluted and used to challenge fresh host cultures. Wells were then screened for other putative viral PSI gene arrangements, *psaJF-psaC, psaJF-psaA* and *psaB-psaD,* found in putative cyanophages but not in cyanobacteria. A lysate from one of the positive wells from the highest dilution (labeled with thick borders) was then used to infect *Prochlorococcus* MIT9515 on semi-solid media. Four rounds of plaque purification were carried out until all plaques were positive for the

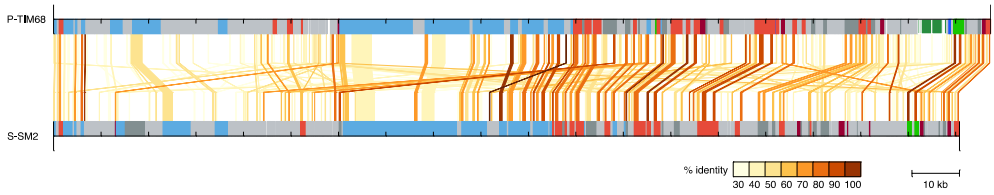
613 *psaC-psaA* gene arrangement (plaques showing a positive signal are marked with thick
614 borders). Colors shown for the 96-well plates indicate the colors of the different
615 cyanobacterial cultures used (wells showing lysis are represented in lighter colors). The front
616 plate represents *Prochlorococcus* MIT9515. The positive control (PC) for the PCRs were
617 performed with a PCR amplicon of a 6.2 kb environmental PSI gene cassette[6] [genbank #
618 GQ268816] as the template; the negative control (neg.) was without template addition.

619

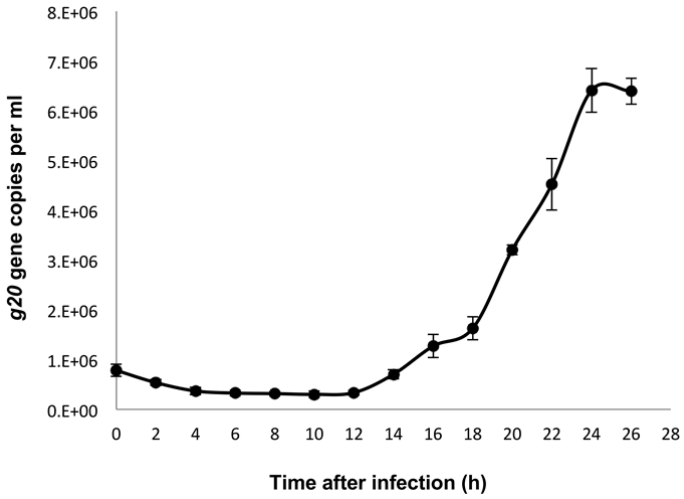
a



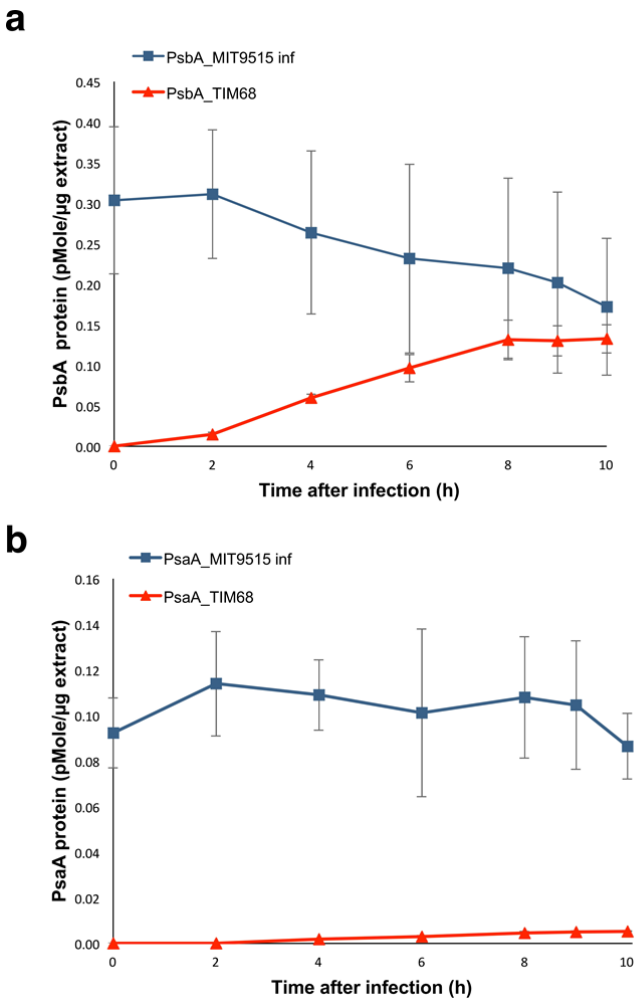
b



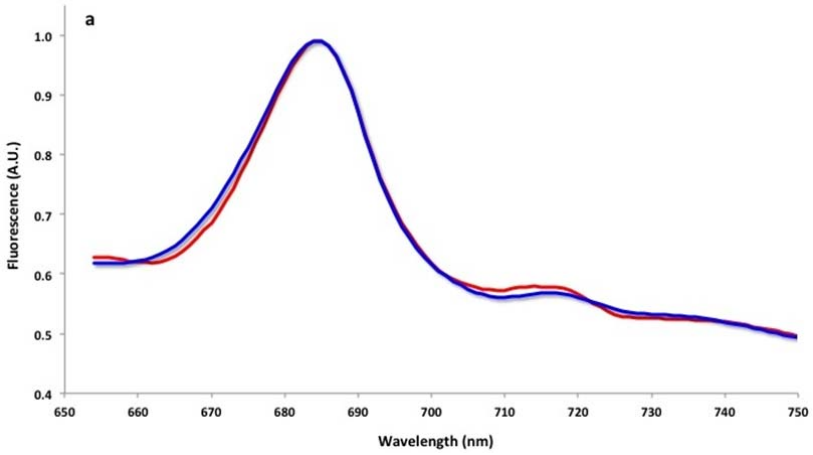
Supplementary Figure S2. Synteny and genome map of P-TIM68. (**a**) Genome map of the P-TIM68 phage. (**b**) Comparison of synteny between P-TIM68 and the *Synechococcus* myophage S-SM2. The color code of the putative gene functions is the same as in part (a). Homologous genes are connected with colored lines according to % protein identity.

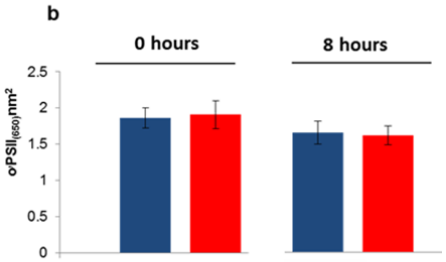


Supplementary Figure S3. Infection of *Prochlorococcus* MIT9515 by the P-TIM68 phage. Infection dynamics determined from infective phage in the extracellular medium with time after phage addition at 0 h. Average and S.D. of three biological replicates.



Supplementary Figure S4. Expression of P-TIM68 PSI and PSII proteins in infected *Prochlorococcus* cells. Levels of phage and host **(a)** PsbA (PSII) and **(b)** PsaA (PSI) protein concentrations. Error bars denote S.D. from three replicate experiments. Graphs of protein levels normalized to those in uninfected control cells can be found in Fig. 3.





Supplementary Figure S5. (a) Chlorophyll fluorescence emission spectra at low temperature (77 Kelvin). Relative fluorescence emission spectra of chlorophyll in whole *Prochlorococcus* MIT9515 cells [Infected (red) or control (blue)] at 8 h after infection. Spectra were normalized for the peak value at 682 nm. (**b**) **Absorption-cross section of PSII**. *Prochlorococcus* MIT9515 cells [Infected (red) or control (blue)] were kept in the dark for 10 minutes prior to a single turnover flash at an intensity of 1600 [μmole photons$\cdot m^{-2}\cdot sec^{-1}$] for 1 second using a Pulse Amplitude Modulation Fluorometer (PSI, FL-3500). Absorption cross section of PSII ($\sigma^{PSII}_{(650)}$) was calculated according to Schreiber et al.[59]. Mean values ± SD of the fluorescence ratios were calculated from 5 independent measurements.

Supplementary Tables

Supplementary Table S1. Counts of reads that were recruited to the vPSI-7 and P-TIM68 photosynthesis gene cassettes from the different *Tara* Oceans stations. Excel file.

Supplementary Table S2. Unique P-TIM68 tryptic peptides detected by mass spectrometry (membranal preparation; 8 h after infection).

	Unique viral peptide
Photosystem II	
PsbA	ETTGLESQNYGYK QGILQGWPQFCEWVTSTNNR
PsbD	AFDPTQDEETYSMITANR
PsbN	AQEQDYPQHK HPVMTQK
Photosystem I	
PsaA	GLYEFLAESR PQEDFVDKNPVPVSTEK LAPTIQPR NSQLIPDK EILDGQNGFPQGVTHR LHASLYNVSDVFGGSTTMVGGK HSSDLQENSR
PsaB	DFDPTLNK DFDPTLNKDNVLDR DLGYSFPCDGPGR FNQDLASDPTTR TPIANIAGWR IWDPHFGQGATDAYTQAGSTFPVNR QILIEPIFAQWIQAASGK EILEAHTPPR
PsaD	DGVVSER IGQNPNPVELK NGEVEFLYPK EQGFELPTGGAALMNEGDNIMYFAR KEQCLALGTQFK TYAETYAITWTSPK FNINGFNPQR NTGNLIDPR TDLKPK
PsaE	IKRPESYWFNECGK
PsaJF	LLETNVTVSPR CILTSAVWPAEAHFEATK AFPVYNELANGK YLIVVR
PsaK	TIGASSEGLQLPKPK

Supplementary Table S3. Accession numbers of PsbA, *g20* and *g23* proteins used to calculate vPSI-7 abundance in Supplementary Table S1. Excel file.

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 U S A* **101**, 11007-11012 (2004).

3 Lindell, D. *et al.* Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **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, doi:10.1007/s11120-014-0057-x (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, doi:10.1111/j.1462-2920.2009.01866.x (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 U S A* **108**, e757-764 (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, doi:10.1016/j.cub.2016.04.036 (2016).

20 Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: I, The 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, doi:10.1038/ismej.2014.198 (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, doi:10.1038/ismej.2015.210 (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*, doi:10.1111/1758-2229.12409 (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 U S A* **113**, E3365-3374, doi:10.1073/pnas.1524865113 (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, doi:10.1371/journal.pone.0059861 (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, doi:10.1126/science.230.4727.818 (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*, e52131, doi:10.3791/52131 (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 U S A* **109**, 2037-2042, doi:10.1073/pnas.1115467109 (2012).

38 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821-829, doi:10.1101/gr.074492.107 (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, doi:10.1007/978-1-60327-565-1_7 (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, doi:10.1186/gb-2009-10-3-r25 (2009).

41 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**, 119, doi:10.1186/1471-2105-11-119 (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, doi:10.1128/AEM.72.1.723-732.2006 (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**, W465-469 (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, doi:10.1093/sysbio/syq010 (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 U S A* **102**, 4913-4918, doi:10.1073/pnas.0501268102 (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, doi:10.1038/nmeth.1923 (2012).

54 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (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**, D733-745, doi:10.1093/nar/gkv1189 (2016).

57 Sharon, I., Pati, A., Markowitz, V. M. & Pinter, R. Y. in *Research in Computational Molecular Biology* Vol. 5541 (ed S. Batzoglou) 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, doi:10.1093/bioinformatics/bts174 (2012).

59 Schreiber, U., Klughammer, C. & Kolbowski, J. High-end chlorophyll fluorescence analysis with the MULTI-COLOR-PAM. I. Various light qualities and their applications. *PAM Appl. Notes* **1**, 1-19 (2011).