



**HAL**  
open science

## **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.

► **To cite this version:**

Svetlana Fridman, José Flores-Uribe, Shirley Larom, Onit Alalouf, Oded Liran, et al.. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected Prochlorococcus cells. *Nature Microbiology*, 2017, 2 (10), pp.1350-1357. 10.1038/s41564-017-0002-9 . hal-04006860

**HAL Id: hal-04006860**

**<https://hal.science/hal-04006860>**

Submitted on 27 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **A novel myovirus encoding both photosystem-I and II**  
2 **proteins enhances cyclic electron flow in infected**  
3 ***Prochlorococcus* cells**

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

8 <sup>1</sup>Faculty of Biology, and <sup>4</sup>Smoler Proteomics Center, Technion-Israel Institute of Technology,  
9 Haifa 32000, Israel. <sup>2</sup>Department of Molecular Biology and Ecology of Plants, George S. Wise  
10 Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel. <sup>3</sup>Institut de Biologie  
11 Physico-Chimique, UMR 7141 CNRS and Université Pierre et Marie Curie, 13 rue Pierre et  
12 Marie Curie, 75005 Paris, France. <sup>5</sup>Migal Galilee Research Institute, Kiryat Shmona, 11016,  
13 Israel. Tel Hai College, Upper Galilee 12210, Israel. <sup>6</sup>Department of Marine Biology and  
14 Oceanography, Institute of Marine Sciences (ICM), CSIC, Barcelona, Spain. <sup>7</sup>Microbial and  
15 Environmental Genomics, J. Craig Venter Institute, La Jolla, California 92037, USA.  
16 <sup>8</sup>Department of Biology, San Diego State University, San Diego, California 92182, USA.

17

18 <sup>§</sup>This paper is dedicated to the memory of Prof. Fabrice Rappaport (CNRS), who sadly  
19 passed away before the paper was finalized.

20

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

23

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

49           Oxygenic photosynthetic organisms, including cyanobacteria, use four  
50 integral membrane complexes: PSII, *b<sub>6</sub>f*, PSI and ATPase, to convert sunlight  
51 into chemical energy. Linear electron transfer begins with PSII catalyzing the  
52 transfer of electrons, derived from water, to the plastoquinone pool. PSI  
53 transfers electrons from plastocyanin/cytochrome *c<sub>6</sub>* to ferredoxin thereby  
54 generating reducing power needed for CO<sub>2</sub> fixation in the form of NADPH.  
55 Then, the *b<sub>6</sub>f* complex couples the electron transfer from the reduced  
56 plastoquinone (QH<sub>2</sub>) to the oxidized plastocyanin/cytochrome *c<sub>6</sub>* to the  
57 pumping of protons in the lumen. Lastly, the electrochemical proton gradient  
58 drives ATP formation by the ATPase complex.

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

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

74           Despite evidence from environmental DNA for the existence of  
75 cyanophages encoding PSI gene cassettes, no cyanophage carrying PSI  
76 genes has been isolated to date. Indeed this phage was dubbed the “mystery  
77 phage” by Puxty et al.<sup>8</sup>. Here we developed a targeted isolation strategy in an  
78 attempt to find such a PSI-carrying cyanophage (Supplementary Fig. S1).  
79 This strategy relied on PCR screening of lysates and plaques for PSI gene  
80 arrangements unique to environmental viral DNA<sup>6</sup>, primarily adjacent *psaC*  
81 and *psaA* genes. To this end, eighteen different *Prochlorococcus* and marine  
82 *Synechococcus* strains were used to screen viral concentrates collected from  
83 the Equatorial Pacific Ocean off the Line Islands (Caroline Island 9.9°S,  
84 150.2°W). Phage lysates containing viral PSI gene arrangements were  
85 observed only when the high-light-adapted (HL) *Prochlorococcus* MIT9515  
86 strain served as the host (see Methods and Supplementary Fig. S1). A phage  
87 from these lysates, P-TIM68, was plaque purified and characterized.

88           P-TIM68 has a long contractile tail indicative of myoviruses and a  
89 prolate capsid (Fig. 1a). It has a 197 kb genome with 34.3 G+C%, 246  
90 predicted open reading frames and no tRNA genes (Supplementary Fig. S2a).  
91 The P-TIM68 genome contains a suite of DNA replication and virion structural  
92 genes characteristic of the archetype T4 (supplementary File S1) that are  
93 homologous to those found in other T4-like cyanophages<sup>13,14</sup>. This phage  
94 clusters within clade II of the T4-like myocyanophages (based on the *g20*  
95 gene) and genomic comparison to a closely related T4-like cyanomyophage  
96 from that clade, S-SM2, showed that there is a high degree of synteny and  
97 percent identity between the structural, replication and nucleotide metabolism  
98 genes of the two phages (Supplementary Fig. S2b).

99           The P-TIM68 phage genome contains PSI genes arranged in a  
100 cassette with the fused *psaJF* gene followed by the *psaC*, *A*, *B*, *K*, *E* and *D*  
101 genes (Fig. 1c; Supplementary Fig. S2a). This is the same arrangement as  
102 that previously observed for the metagenomic *vPSI-7* cassette<sup>6</sup>. Furthermore,  
103 a phylogenetic tree constructed with the PsaA protein clearly places it within  
104 the *vPSI-7* cluster of viral PsaA from environmental DNA (Fig. 1b).

105           In addition to the PSI genes, the genome of P-TIM68 contains three  
106 PSII genes, *psbA* and *psbD* genes coding for the core D1 and D2 proteins,  
107 and a *psbN* gene homolog coding for a small protein required for the repair  
108 and assembly of PSII reaction centers<sup>16</sup>. It also contains two genes coding for  
109 type I NAD(P)H dehydrogenase (NDH-1) subunits (the *ndhP* and *ndhI* genes),  
110 and four High-Light-Inducible protein (*hli*) genes arranged in tandem.

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

119           The photosynthesis region of the P-TIM68 phage genome is made up  
120 of the 7-gene PSI cassette, one of the NDH-1 genes and the two reaction  
121 center genes of PSII (Fig. 1c and Supplementary Fig. S2b). This region is  
122 flanked by hypothetical genes and is downstream of a structural protein (the

123 baseplate wedge initiator protein - gp7), rather than the *nrdA* and *nrdB*  
124 ribonucleotide reductase genes found to neighbor previously reported PSI  
125 environmental scaffolds<sup>6</sup>. Analyses of data from the *Tara Oceans*<sup>7</sup> and the  
126 Global Ocean Sampling (GOS)<sup>20,21</sup> expeditions revealed the presence of two  
127 scaffolds with the exact same gene organization as in P-TIM68 over their  
128 entire length of 3.9 and 17.8 kb (supplementary File S2). These scaffolds  
129 showed high identity (98%) to P-TIM68 at the DNA level over the region of the  
130 PSI gene cassette, indicating that phages very similar to P-TIM68 are present  
131 in the environment.

132 We next assessed how widespread the P-TIM68 phage is in the  
133 oceans. Recruitment of reads from the *Tara Oceans* expedition to the *vPSI-7*  
134 segment of P-TIM68, showed that P-TIM68 sequences (estimated using  
135  $\geq 98\%$  nucleotide identity) are present in tropical waters in the Pacific, Indian  
136 and Atlantic Oceans (supplementary Table S1). We then estimated the  
137 abundance of P-TIM68 ( $\geq 98\%$  identity) and other *vPSI-7* ( $\geq 80\%$  identity)  
138 phages in the *Tara Oceans* dataset (Fig. 2a) relative to all T4-like  
139 cyanomyophages (using the T4-like *g20* and *g23* marker genes<sup>22</sup>). The *vPSI-*  
140 *7* phages were particularly prevalent in the southern region of the tropical  
141 Pacific Ocean, often reaching 15-28% of total cyanomyophages (Fig. 2a,  
142 supplementary Table S1). The P-TIM68 genotype composed on average 22%  
143 of these *vPSI-7* phages (average from all stations containing more than 20  
144 *psaA* reads; proportions labeled in red in Fig. 2a).

145 Characterisation of the infection cycle of the P-TIM68 phage revealed a  
146 long latent period of 10-12 hours (Supplementary Fig. S3) and a burst size of  
147  $12 \pm 3$  ( $n=3$ ) infective phages/cell. Similarly, other T4-like cyanophages have

148 latent periods ranging from 4-18 h<sup>19,23,24</sup> and burst sizes that range from 21-45  
149 phages/cell<sup>23,25,26</sup>. The phage PSI and PSII genes were transcribed at the  
150 same time during the infection cycle, with maximal transcript levels at 6-8  
151 hours after infection (Fig. 3a). These photosynthesis genes were transcribed  
152 together with middle genes involved in phage genome replication (Fig. 3a), as  
153 found for PSII genes in other cyanophages<sup>24,27</sup>.

154 At the protein level, unique peptides from phage PSI (PsaJF, PsaA,  
155 PsaB, PsaK, PsaE and PsaD) and PSII (PsbA, PsbD and PsbN) proteins  
156 were detected in host membranes 8 hours after infection (supplementary  
157 Table S2). Quantification of host- and phage- specific photosynthesis peptides  
158 showed that levels of the host PSII PsbA protein (D1) decreased during  
159 infection to about 50% of their initial levels (Fig. 3b). This corresponded with  
160 an increase of the viral homolog to 43% of the total D1 proteins in infected  
161 cells by the end of the latent period (Supplementary Fig. S4a), thus almost  
162 completely restoring the amount of D1 proteins to its original levels. A similar  
163 trend was previously observed for a *Prochlorococcus* T7-like podovirus (P-  
164 SSP7) carrying a single PSII gene, although the viral D1 reached about 10%  
165 of total D1 proteins<sup>9</sup>, implying a general role for viral PSII genes in T4- and  
166 T7-like cyanophages.

167 The host PSI PsaA protein, in comparison, showed no significant  
168 decrease (Fig. 3c). The viral PSI PsaA protein, however, increased steadily  
169 (Fig. 3c) reaching about 6% of the total PSI PsaA proteins by the end of the  
170 latent period (supplementary Fig. S4b). Since only 70% of host cells were  
171 infected in these experiments, the level of phage PsaA protein corresponds to  
172 8.6% of total PsaA in infected cells. This is a small but significant level of PSI

173 proteins that we hypothesize supplements host PSI proteins during infection.  
174 It is possible that even more phage PSI proteins replace host proteins under  
175 higher light conditions similar to those found in surface layers of tropical  
176 waters.

177 To characterize the host's photosynthetic activity during P-TIM68  
178 infection, relative maximum electron transfer rate ( $rETR_{max}$ ) and PSI reduction  
179 rate were measured. *Prochlorococcus* cells were grown under a photon flux  
180 density of  $10 \mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ . Throughout the 10 hour latent period of P-  
181 TIM68,  $rETR_{max}$  remained constant while host chromosome degradation  
182 clearly indicated viral takeover of the host (Fig. 4a). These results are  
183 consistent with previous findings of continued PSII activity during infection of  
184 viruses carrying PSII genes<sup>9,19,28</sup>. In contrast, when the PSI reduction rate was  
185 monitored, by absorbance changes at 705 nm, an increased rate of electron  
186 flow was detected in infected cells from 6-8 hours post infection (Fig. 4b).

187 The photosynthetic measurements of infected relative to uninfected  
188 cells revealed that PSII activity remained unchanged during P-TIM68 infection  
189 ( $rETR_{max}$  in Fig. 4a) while the rate of PSI reduction increased. These results  
190 suggest that a PSII-independent electron flow to PSI, *i.e.* an increased cyclic  
191 electron flow around PSI, is induced during infection. Such increased cyclic  
192 electron flow could generate extra ATP required for phage DNA replication  
193 and nucleotide biosynthesis. However, an alternative explanation was also  
194 considered: even if the  $rETR_{max}$  remains constant during infection, the  
195 electron flow from PSII could have increased *when normalized per PSI* if the  
196 PSII/PSI stoichiometry or the absorption cross-section of PSII increases  
197 during infection. However, chlorophyll fluorescence emission spectra at low

198 temperature (77 Kelvin) showed no changes in PSII/PSI stoichiometry  
199 between infected and uninfected cells (supplementary Fig. S5a). In addition,  
200 absorption cross-section of PSII did not change during infection  
201 (supplementary Fig. S5b). Therefore, these results suggest that there is  
202 increased cyclic electron flow around PSI as a result of P-TIM68 infection.

203         While previous metagenomic data had indicated the likely existence of  
204 PSI carrying phages in the environment, the isolation of P-TIM68 provides  
205 definitive proof that the PSI cassettes are of cyanophage origin and that a  
206 phage encoding proteins capable of forming an entire PSI complex are  
207 present in the oceans and can infect *Prochlorococcus*. Furthermore, the  
208 physiological characterisation of this phage showed that infected cells  
209 maintain and even increase photosynthesis during the infection cycle.

210         The restricted distribution pattern of the PSI carrying phages to tropical  
211 waters is intriguing and directly correlates to the distribution of the uncultured  
212 HLIII-A and HLIV-A *Prochlorococcus* clades that are unique to this High-  
213 Nutrient, Low-Chlorophyll (HNLC), iron depleted region<sup>29</sup> (see inset in Fig. 2  
214 for the correlation between *Prochlorococcus* HLIV-A and P-TIM68  
215 abundances). If indeed members of one or both of these *Prochlorococcus*  
216 lineages are the natural hosts for *vPSI-7* phages, this would explain their  
217 limited distribution pattern. In addition, it was recently shown that the ratio of  
218 PSI/PSII declines under iron limiting conditions in freshwater cyanobacteria<sup>30</sup>.  
219 Although highly speculative, perhaps *vPSI-7* phages are compensating for  
220 such a decline in infected hosts under the iron-limiting conditions found in this  
221 environment. Future experiments under high-light and iron-limited conditions,

222 will be important for elucidating the full impact of the *vPSI-7* carrying phages  
223 on their *Prochlorococcus* hosts in such environments.

224           In summary, our combined observations that P-TIM68 and other *vPSI-*  
225 *7* phages are abundant in the environment, together with their ability to  
226 maintain and even manipulate the mode of host photosynthesis in their  
227 *Prochlorococcus* hosts, suggest that viral PSI genes play a key functional role  
228 in oceanic photosynthesis in tropical waters.

## 229 **Methods**

230 **Bacterial strains.** *Synechococcus* and *Prochlorococcus* strains used in this  
231 study were *Synechococcus* spp. strains WH7803, WH7805, CC9311,  
232 WH8102, WH8109, CC9605, BL107, RS9916, WH5701, RS9917, RCC307,  
233 and CC9902; and *Prochlorococcus* spp. strains MED4, MIT9215, MIT9515,  
234 MIT9312, NATL2A, and SS120. *Synechococcus* strains were grown in an  
235 artificial sea water-based medium (ASW)<sup>31</sup> with modifications as described  
236 previously<sup>32</sup>. *Prochlorococcus* strains were grown in the natural seawater-  
237 based medium Pro99<sup>33</sup>. Cultures were grown at 21 °C under cool white light  
238 with a 14:10-h light:dark cycle at an intensity of 10-14  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$   
239 during the light period. The *Prochlorococcus* MIT9515 strain, on which the P-  
240 TIM68 phage was isolated, belongs to the HLI clade of *Prochlorococcus* and  
241 was isolated from the Equatorial Pacific (5.7°S, 107.1°W) at 15m depth by the  
242 Chisholm group (<http://proportal.mit.edu/genome/id=272/>).

243

244 **Phage isolation strategy.** Viral concentrates from the Pacific Line Islands  
245 were collected in April 2009 from the uninhabited Caroline (Millennium) Island  
246 and were prepared according to Haas et al.<sup>34</sup>. Cyanobacteria were challenged  
247 with these concentrates in 10-fold serial dilutions in 96-well plates. Wells  
248 showing lysis were then analyzed by PCR for the presence of the *psaC-psaA*  
249 PSI gene arrangement found in the putative cyanophages but not known for  
250 cyanobacteria. Lysates from wells positive for the viral *psaC-psaA* were used  
251 to challenge the host again (Supplementary Fig. S1) and were then tested for  
252 additional unique viral PSI gene combinations (*psaJK-psaC*, *psaJF-psaA* and

253 *psaB-psaD*). The highest dilution lysate positive for these combinations was  
254 then used in 4 rounds of plaque purification and PCR screening until all  
255 plaques were positive for the viral *psaC-psaA* gene arrangement. While lysis  
256 was observed in all cyanobacterial strains tested in the first challenge, only  
257 lysed wells from *Prochlorococcus* MIT9515 gave positive PCR signals for the  
258 *psaC-psaA* combination and underwent subsequent isolation procedures.

259 Degenerate primers used for the PCR reaction of *psaC-psaA* were  
260 PsaC(MVPW[D/E]G)-fwd [5'-atggtncntggaygg-3'] and PsaA(GWLADP)-rev  
261 [5'-ggrtcngcncarccancc-3']<sup>35</sup>. PCR conditions were according to Hevroni et  
262 al.<sup>36</sup>. Primers and PCR conditions for reactions *psaJF-psaC*, *psaJF-psaA* and  
263 *psaB-psaD* were according to Sharon et al.<sup>6</sup>.

264

265 **Transmission Electron Microscopy (TEM).** A cyanophage lysate was  
266 filtered using a 0.2  $\mu\text{m}$  pore-sized filter and then brought to a final  
267 concentration of  $\sim 10^9$  pfu/mL by centrifugation using a 100 KDa molecular  
268 weight cutoff (ca. 10 nm sized pores) centrifugal filter (Millipore). The phage  
269 lysate was then CsCl-purified according to Sabehi et al.<sup>37</sup>. For TEM analysis,  
270 20  $\mu\text{l}$  of the CsCl-purified phage concentrate were placed on a formvar/carbon  
271 coated 100-mesh grid. After 20 min at room temperature, excess liquid was  
272 removed, and phages were negatively stained with 1% uranyl acetate for 1  
273 min. Uranyl acetate was removed, and the grid was air-dried for 10 min before  
274 overnight desiccation under vacuum. Phages were examined at an  
275 accelerating voltage of 200-kV using a FEI Tecnai<sup>TM</sup> G2 T20 S-Twin TEM.

276

277 **Sequencing and annotations.** The P-TIM68 phage was sequenced using  
278 Illumina dual index MiSeq using Nextera XT library preparation kit. Raw reads  
279 were deposited in GenBank Sequence Read Archive (BioSample accession  
280 SAMN06618437) and *de-novo* assembled using the velvet 1.2.10<sup>38</sup> program  
281 with a k-mer size=73. Assembly coverage was 1790-fold. The genome  
282 assembled into a single circular contig as expected for linear T4-like phage  
283 genomes that are circularly permuted and terminally redundant<sup>39</sup>.  
284 Misassemblies were detected through manual scanning of mapped reads to  
285 the assembled genome. Read mapping was performed using Bowtie<sup>40</sup>  
286 (parameters -q -n 2 -e 200 -p 10 --best). The genome of P-TIM68 was  
287 deposited in GenBank under accession number KM359505. Gene calling was  
288 carried out using Prodigal 2.6.1<sup>41</sup> followed by annotation using blast<sup>42</sup>  
289 (program blastall, parameters -p blastX -e 1e-5 -F F -b 1 -v 1) against  
290 NCBI's nr database. Hits were manually curated.

291

292 **Infection Characterization.** Infection dynamics were determined from the  
293 appearance of phage in the extracellular medium with time. Exponentially  
294 growing *Prochlorococcus* MIT9515 cells were grown in triplicate at 21°C in  
295 continuous light (14  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ) and infected with P-TIM68 (at a  
296 multiplicity of infection (MOI) of 0.1). Quantification of phage DNA in the  
297 extracellular medium (0.2  $\mu\text{m}$  filtrate) was determined by real-time PCR for the  
298 P-TIM68 portal protein (*g20*) gene using the real-time procedure described in  
299 the "Quantitative PCR Protocol" section of the Methods. Burst size was  
300 calculated according to Sabehehi et al.<sup>37</sup>.

301

302 **Host gDNA degradation:** Exponentially growing *Prochlorococcus* MIT9515  
303 cells were infected with P-TIM68 (MOI of 3). *Prochlorococcus* gDNA was  
304 quantified after collection of cells on 0.2 µm pore-sized polycarbonate filters  
305 (GE), washed with sterile seawater followed by 3 ml of preservation solution  
306 (10 mM Tris, 100 mM EDTA, 0.5M NaCl; pH 8.0) and frozen at -80°C. DNA  
307 was extracted by the heat lysis method<sup>43</sup>. The RuBisCO large subunit gene  
308 (*rbcL*) was amplified for *Prochlorococcus* gDNA using the real-time procedure  
309 described in the “Quantitative PCR Protocol” section of the Methods.

310

311 **Gene expression. RNA extraction:** *Prochlorococcus* cells were grown at  
312 21°C to mid-logarithmic phase and were infected with P-TIM68 at a multiplicity  
313 of infection (MOI) of 3. Samples were collected every 30 min for the first 4  
314 hours after infection and then every 2 hours until 10 hours after infection. The  
315 experiment was repeated 3 times. Nine mL of infected or uninfected cells  
316 were filtered onto 0.22 µm Durapore filters (Millipore). The filters were  
317 transferred to tubes containing 1 mL of RNAlater (Ambion), frozen and stored  
318 at -80°C. Total RNA was extracted using the mirVana RNA isolation kit  
319 (Ambion). Genomic DNA was removed using the Turbo DNA-free kit  
320 (Ambion). cDNA was made using the High-Capacity cDNA Reverse  
321 Transcription Kit (Applied Biosystems).

322 Several phage genes were chosen for transcript analysis by real-time  
323 PCR: the *psaJF-A*, *psaB* and *psbA* photosynthetic genes, the *g32* early  
324 gene<sup>24</sup>, the *g147* DNA polymerase middle gene and the *g85* baseplate late

325 gene. Gene numbers are as per the P-TIM68 genome. Results were  
326 normalized to the amount of host *mpB* gene in infected cells.

327

328 **Quantitative PCR Protocol:** Each real-time PCR reaction contained 1X  
329 Roche universal probe library (UPL) master mix (LightCycler® 480 Probes  
330 Master, Roche), 80 nM UPL hydrolysis probe, 500 nM primers and cDNA  
331 template in a total volume of 25  $\mu$ l. Reactions were carried out on a  
332 LightCycler 480 Real-Time PCR System (Roche). The program used was 1  
333 step of 95 °C for 15 min followed by 45 cycles of amplification, each including  
334 10 sec denaturation at 95 °C, and a 30 sec combined annealing and  
335 elongation step at 60 °C, at the end of which plate fluorescence was read  
336 (FAM; Ex/Em 465/510 nm). A list of genes, primers and probes used in real-  
337 time PCR assays is shown in Table 1.

338

339

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

Target gene	Sequence 5' – 3'	Probe 5' – 3'
Region from <i>psaJF</i> to <i>psaA</i>	Fwd: TGTAAGTATCCCCTAGAT Rev: TGCTTACAGGGACAGGATTTTT	CTCCACCA
<i>psaB</i>	Fwd: CATAACACAGGCAGGTTCTAC Rev: CTTGGTTTCAGATGTAATTGAG	CTGGTGGT
<i>psbA</i>	Fwd: AGCATGCTTCGTTGTTGCATTTATT Rev: AGAAATGATGTTGTTACCATAGAGA	GCCATCAA
DNA polymerase	Fwd: AGTGGCACAGAAGAGGATGTAAT Rev: GGAGTTCCTTTGCATAGATTG	CTGGTGGT
Base plate	Fwd: TGCATCACAAAATAGGGCAGT Rev: TAGATTCTAGGCACTAACGCTTCA	GCAGCAGA

Early gene32	Fwd: CACATTTATCACTTAAGATGGATGGT Rev: TCTCAGGGTGAGTCCCAAAT	CAGCAGGT
<i>rnpB</i>	Fwd: AAGTCCGGCTCCATA Rev: GCACTATCCTCACGGTTGC	TTCCCAGT
<i>rbcL</i>	Fwd: GTCAGATATGCCAATCATCATGC Rev: GCCAAGCCAGTATTTGCAGT	CTGGTGGT
<i>g20</i>	Fwd: CCTAAGGAGCAACTACACCA Rev: TGATGTGCAATATGTAATTGAAT	TGGTGGAA

341

342 **Shotgun proteomics:** *Prochlorococcus* cells (1.5 L) were grown at 21°C to  
343 mid-logarithmic phase and were infected with P-TIM68 (MOI of 3). Cells were  
344 collected by centrifugation for 30 min at 13,300 x g, 4°C. The pellet was  
345 resuspended in buffer A (50 mM Tris/HCl, 5 mM MgCl<sub>2</sub>·6H<sub>2</sub>O, 10 mM NaCl,  
346 pH 8.0), frozen in liquid nitrogen and stored at -80°C. The pellets were thawed  
347 at room temperature and were supplemented with buffer A containing  
348 protease inhibitors (Calbiochem protease inhibitor cocktail set VI). Cells were  
349 disrupted by 5 passes in the EmulsiFlex-C3 homogenizer (Avestin) at 100-  
350 1500 bar at 4°C. The suspension was centrifuged for 10 min at 13,300 x g,  
351 4°C, to remove cell debris, and centrifuged again for 30 min at 46,000 x g,  
352 4°C, to collect the membranes. Membranes were washed once with 1 mL of  
353 buffer A, and resuspended in 1 mL of buffer A containing 400 mM sucrose,  
354 flash frozen in liquid nitrogen and stored at -80°C.

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

361 The resulting tryptic peptides were resolved by reverse-phase  
362 chromatography on a 0.075 X 200-mm fused silica capillary column (J&W)  
363 packed with Reprosil reversed phase material (Dr Maisch GmbH, Germany).  
364 The peptides were loaded onto a homemade capillary column (25 cm, 75  
365 micron ID) packed with Reprosil C18-Aqua (Dr Maisch GmbH, Germany) in  
366 solvent A (0.1% formic acid in water). The peptide mixture was resolved at a  
367 flow rate of 0.15  $\mu$ l/min with a linear gradient, 5 to 28%, of 95% acetonitrile  
368 with 0.1% formic acid for 30 minutes followed by a gradient of 28 to 95% for  
369 15 min and finally 15 minutes at 95% acetonitrile with 0.1% formic acid. Mass  
370 spectrometry was performed in a positive mode (m/z 350–1800) with  
371 repetitive cycles of full MS scans followed by high collision induced  
372 dissociation (at 35 normalized collision energy) of the 10 most dominant ions  
373 (>1 charges) selected from the first MS scan. A dynamic exclusion list was  
374 used with exclusion duration of 20 s.

375 The mass spectrometry data was analyzed using Proteome Discoverer  
376 1.4 software with Sequest algorithm against *Prochlorococcus* MIT9515 and P-  
377 TIM68 proteomes allowing a 1% false discovery rate.

378

379 **Targeted proteomics:** *Prochlorococcus* MIT9515 cells (300 mL) were grown  
380 at 21°C to mid-logarithmic phase and were infected with P-TIM68 (MOI of 3).  
381 Thirty mL samples were collected at each time point and centrifuged for 30  
382 min at 13,300 x g, 4°C washed in Pro99 medium, and centrifuged again.  
383 Pellets were frozen in liquid nitrogen and stored at -80°C. Three biological  
384 repeats were carried out and seven time points were analyzed during the  
385 infection process from both infected and uninfected cells.

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

392

393 **Table 2: Isotopically Labeled peptides used in this study**

	P-TIM68	MIT9515
PsbA	ETTGLESQNYGYK	ETTETESQNYGYK
PsaB	DLGYSFPCDGPGR	DLGYAFPCDGPGR
PsaA	LAPTIQPR	VAPTIQPR
PsaJF	LLETNVTVSPR	

394

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

398 Protein samples were reduced with 3 mM DTT for 45 min at 54°C and  
 399 modified with 12 mM iodoacetamide (30 min at room temperature). Samples  
 400 were diluted with 50 mM ammonium bicarbonate at a 1:1 ratio, trypsinized at  
 401 a 1:50 enzyme to protein ratio and incubated overnight at 37°C. An additional  
 402 trypsinization was done for 4 hours. The samples were then spiked with a mix  
 403 of heavy labeled peptides (Table 2, 1  $\mu$ mole of each peptide/5-10  $\mu$ g extract).  
 404 The samples were acidified with 1.5% formic acid in order to remove the SDC

405 buffer prior to sample analysis by mass spectrometry. The resulting peptides  
406 were desalted using C18 SpinColumns (Harvard Apparatus), dried and  
407 resuspended in 0.1% formic acid.

408 The peptides were loaded onto a homemade capillary column (25 cm,  
409 75 micron ID) packed with Reprosil C18-Aqua (Dr Maisch GmbH, Germany)  
410 in solvent A. The peptide mixture was resolved at a flow rate of 0.15  $\mu$ l/min  
411 with a linear gradient (5 to 28%) of 95% acetonitrile with 0.1% formic acid for  
412 120 minutes followed by a gradient of 28 to 95% for 5 min and 25 minutes at  
413 95% acetonitrile with 0.1% formic acid. Mass spectrometry was performed as  
414 stated above.

415 The ratios between the isotopically labeled peptides and the unlabeled  
416 peptides were calculated based on the peak area of the extracted ion currents  
417 (XICs). Absolute quantities were determined by dividing the amount of heavy  
418 peptides with the normalized ratio (corrected for the extract concentration). In  
419 addition the mass spectrometry data were analyzed and quantified using the  
420 MaxQuant software 1.5.1.2 ([www.maxquant.org](http://www.maxquant.org)) for peak identification and  
421 quantitation using the Andromeda search engine, against *Prochlorococcus*  
422 MIT9515 and P-TIM68 proteomes allowing a 1% false discovery rate.

423

424 **Phylogenetic analysis.** Multiple sequence alignment of the PsaA protein (94-  
425 shared amino acids positions; supplementary File S3) was constructed using  
426 ClustalX v2.1<sup>44</sup>. The alignment contains only 94-shared amino acids positions  
427 due to the short length of some of the GOS *psaA* reads used. Maximum-  
428 likelihood (ML) phylogenetic trees were constructed using the phylogeny.fr

429 pipeline<sup>45</sup> which included PhyML v3.0<sup>46</sup> and the WAG substitution model for  
430 amino acids<sup>47</sup>. One hundred bootstrap replicates were conducted.

431

432 **Photosynthetic measurements.** Fluorescence-based photosynthetic  
433 parameters were measured with an in-house assembled fluorescence setup<sup>48</sup>.  
434 The maximal quantum yield of PSII was calculated as  $(F_m - F_0)/F_m$ , where  $F_m$   
435 and  $F_0$  are the fluorescence levels in dark adapted samples and after a  
436 saturating light pulse, respectively<sup>49</sup>. The quantum yield of PSII ( $\Phi_{PSII}$ ) in the  
437 light was calculated as  $(F_m' - F)/F_m'$ , where  $F$  and  $F_m'$  are the steady-state and  
438 maximum fluorescence intensities in light-acclimated cells, respectively. The  
439 relative photosynthetic electron transfer rate  $rETR_{PSII}$  was calculated as  
440  $\Phi_{PSII} * I$ , where  $I$  is the irradiance, and was measured at ten irradiances (steps  
441 of 30 seconds illumination for each light irradiance). The  $rETR_{max}$  was defined  
442 as the highest value of the light saturation curve of the  $rETR_{PSII}$ .

443 Absorption spectroscopy measurements were performed with whole  
444 cells of *Prochlorococcus* MIT9515 using an in-house assembled  
445 spectrophotometer as reported previously<sup>50</sup>. The probe light in the visible  
446 region ( $\lambda > 410$  nm) is generated by a frequency-tripled Nd:YAG laser  
447 (Surelite II; Continuum) pumping a computer controlled optical parametric  
448 oscillator (Panther; Continuum). The samples were excited at 700 nm with a  
449 laser pulse generated by another optical parametric oscillator (SLOPO,  
450 continuum) at an intensity of about 100  $\mu J$ /pulse. Cyclic photosynthetic  
451 electron flow activity was measured using a JTS-10 spectrophotometer (Bio-  
452 Logic Science Instruments SAS, France) set as in Joliot & Joliot<sup>51</sup>. 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  $705 \pm 10$  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 [ $\text{sec}^{-1}$ ].

459

460 **vPSI-7 and P-TIM68 recruitment from Tara Oceans stations.** Tara Oceans  
461 expedition microbiome and virome metagenomic datasets<sup>7,52</sup> 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)<sup>53</sup> 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<sup>54</sup>. BAM files were processed using BP-1.py, a python script  
470 based on HTSeq-count<sup>55</sup> 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<sup>56</sup> 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

478 according to Sharon et al.<sup>57</sup> and *g20* final count was obtained by subtracting  
479 the pelagiphage *g20* normalized counts.

480 The list of genes used for the Bowtie2 mapping and the genomes  
481 comprising the BLASTN database can be found in Supplementary Table S3.  
482 The python script BP-1.py used to process the BAM files is available at  
483 <https://github.com/BejaLab/p-tim68/>.

484

485 **Correlation analysis between *Prochlorococcus* HLIV-A and cyanophage**  
486 **P-TIM68.** A total of 33 samples for the bacterial fraction (>0.22  $\mu\text{m}$ ) were used  
487 to perform a linear regression analysis between the abundance of  
488 *Prochlorococcus* HLIV-A<sup>29</sup> and the abundance of P-TIM68, using the “lm”  
489 function implemented in the R stats package. The relative abundance of P-  
490 TIM68 was calculated as P-TIM68:*g20+g23* ratio, determined from the  
491 normalized counts for *psaA* from P-TIM68 at identity cut-off 97% and the sum  
492 of normalized counts from *g20* and *g23* at identity cut-off of 80%;  
493 Supplementary Table S1 and above “*vPSI-7* and P-TIM68 recruitment from  
494 *Tara* Oceans stations” method section). The TARA\_052 sample was excluded  
495 from the regression analysis because it was considered an outlier based on  
496 the Cook’s distance ( $D_i > 1$ ). Since the abundances of *Prochlorococcus* from  
497 clades HLIII-A and HLIV-A are directly and strongly correlated ( $R^2=0.97$ ), a  
498 correlation between P-TIM68 and HLIV-A abundances is also representative  
499 of a correlation with HLIII-A.

500

501 **Reassembly of *Tara Oceans* scaffolds.** *Tara* Ocean reads for the viral  
502 fraction ( $<0.22 \mu\text{m}$ ) were assembled using the idba-ud assembler<sup>58</sup> with  
503 default parameters. Resulting assemblies went through an automatic mis-  
504 assembly detection and correction procedure using in-house software  
505 (Sharon et al., in preparation). The procedure for mis-assembly correction  
506 takes advantage of read mapping to the assembled scaffolds and contigs as  
507 follows. First, reads from which an assembly was generated were mapped to  
508 the assembly using read mapping software (bowtie2<sup>53</sup>) with the (--sensitive  
509 flag) to generate a SAM file of mapped reads. Next, local mis-assemblies  
510 such as SNPs and short insertions/deletions were identified by locating read  
511 mapping that consistently differed to the assemblies. Finally, we used paired-  
512 end information in order to identify global mis-assemblies by locating regions  
513 in which only one read of each pair is mapped while the other is flanking, or  
514 mapped to a different sequence or region on the same sequence. Given that  
515 sufficient numbers of such pairs exist the sequences are broken into two  
516 sequences in these case.

517

518

519 **ACKNOWLEDGMENTS**

520 We thank Nir Keren, Noam Adir and John Golbeck for their insight regarding  
521 photosynthesis in cyanobacteria, Oded Kleifeld for preliminary proteomics  
522 results, Irena Pekarsky and Mira Rosenberg for help with TEM imaging, and  
523 Béjà and Lindell lab members for continuous discussions. We also thank  
524 Laurence Garczarek for providing cyanobacteria abundance data. This work  
525 was funded by a European Commission ERC Advanced Grant (no. 321647),  
526 an Israel Science Foundation grant (no. 580/10), and the Louis and Lyra  
527 Richmond Memorial Chair in Life Sciences to O.B., a European Commission  
528 ERC Starting Grant (no. 203406) to D.L., and the Technion's Lorry I. Lokey  
529 Interdisciplinary Center for Life Sciences and Engineering and the Russell  
530 Berrie Nanotechnology Institute. This is contribution number **X** of the *Tara*  
531 Oceans.

532

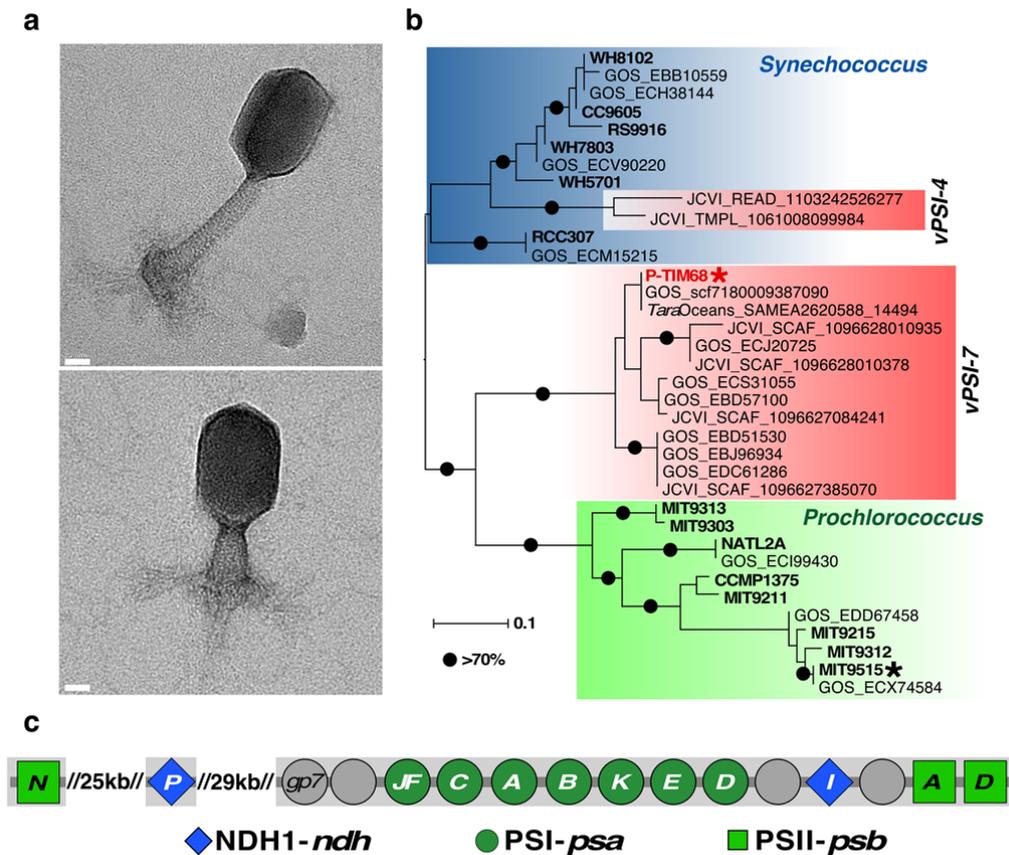
533 **Author Contributions**

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

541

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

544

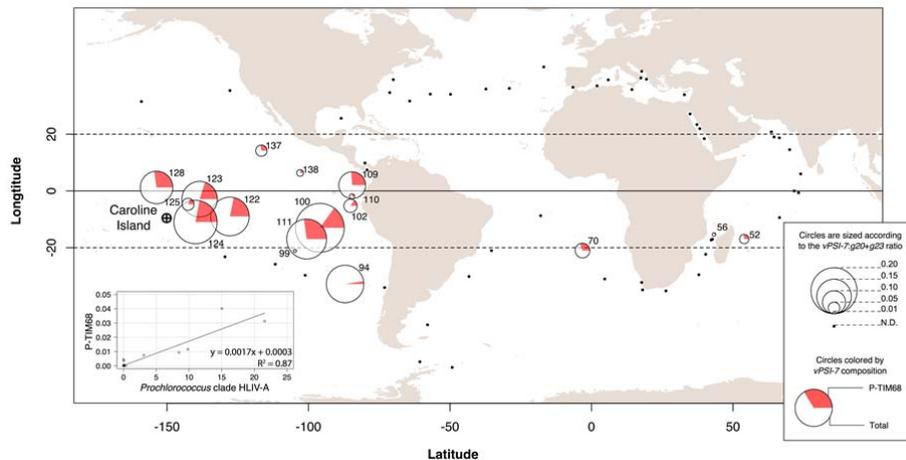
545 **Figure Legends**

546

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

557

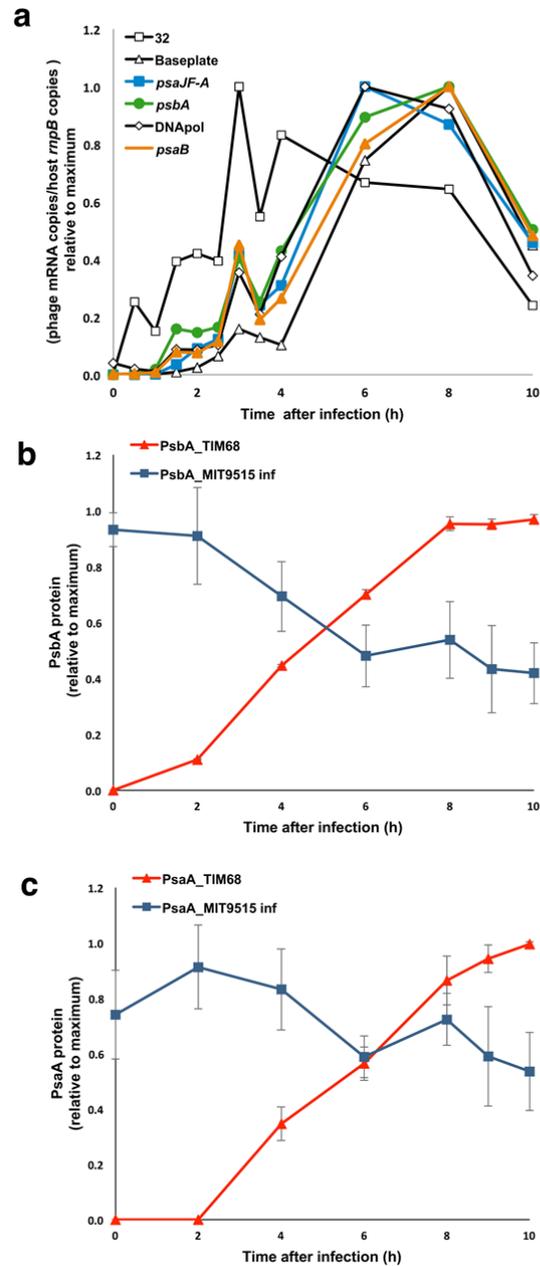
558



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 20°N and 20°S 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<sup>29</sup>.  
 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



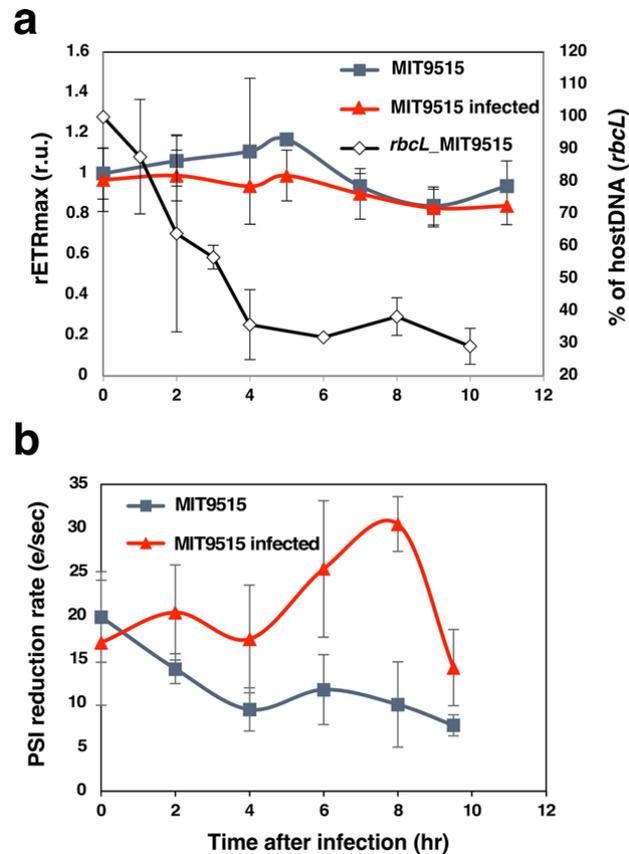
577

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.

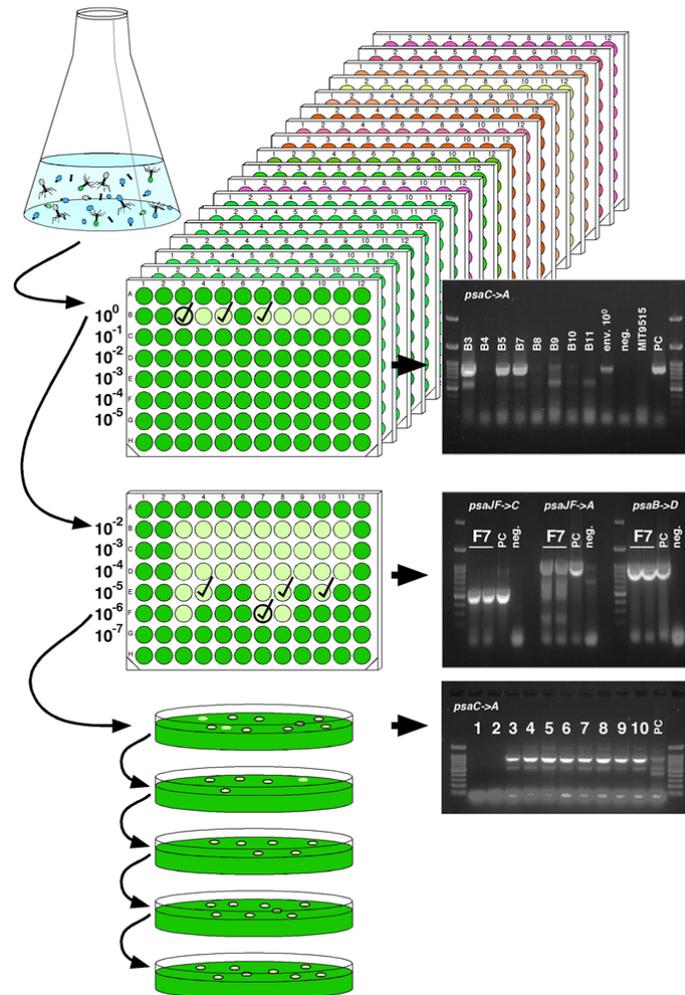
591



592

593 **Figure 4. Changes in PSII and PSI photosynthetic parameters during infection of**  
 594 ***Prochlorococcus* MIT9515 with phage P-TIM68. (a)** rETR<sub>max</sub> and host DNA degradation;  
 595 rETR<sub>max</sub> is normalized to the rETR<sub>max</sub> 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

601 **Supplementary Figures**

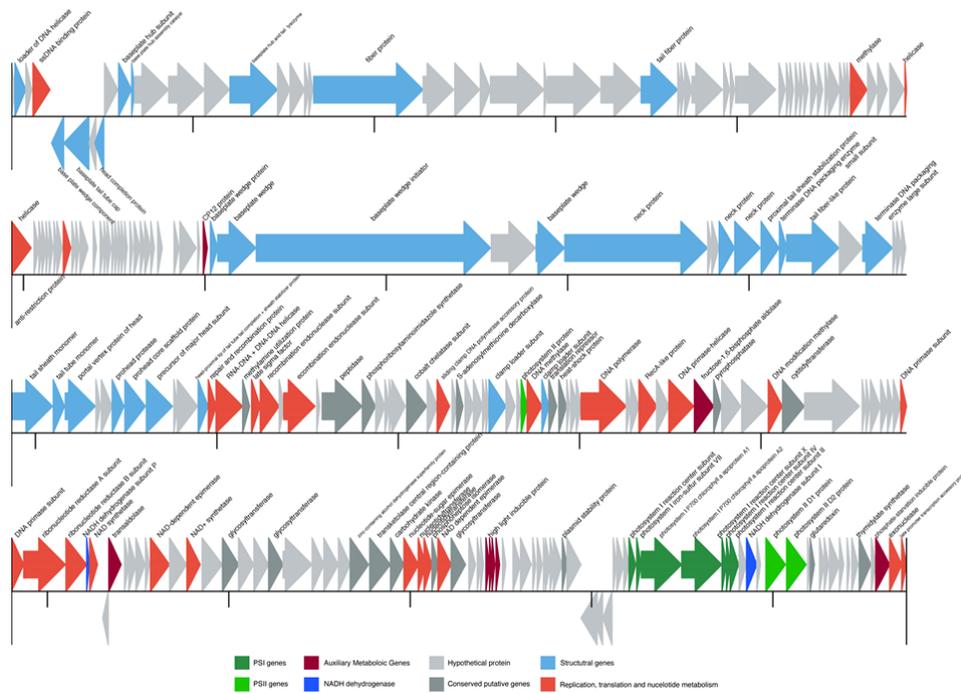
602

603 **Supplementary Figure S1. P-TIM68 isolation strategy.** Top, various *Synechococcus* and  
 604 *Prochlorococcus* strains were challenged with serial dilutions of environmental phage  
 605 concentrate from the Pacific Ocean in 96-well plates. Wells showing lysis were screened for  
 606 the putative viral gene organization, *psaC-psaA*. One lysate (labeled with thick borders) out of  
 607 3 positive wells (✓) from the *Prochlorococcus* MIT9515 plate was serially diluted and used to  
 608 challenge fresh host cultures. Wells were then screened for other putative viral PSI gene  
 609 arrangements, *psaJF-psaC*, *psaJF-psaA* and *psaB-psaD*, found in putative cyanophages but  
 610 not in cyanobacteria. A lysate from one of the positive wells from the highest dilution (labeled  
 611 with thick borders) was then used to infect *Prochlorococcus* MIT9515 on semi-solid media.  
 612 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<sup>6</sup> [genbank #  
618 GQ268816] as the template; the negative control (neg.) was without template addition.

619

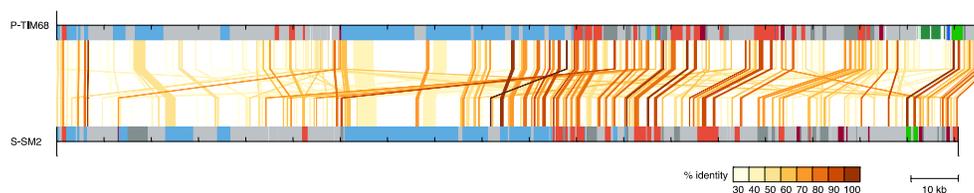
620 a



621

622

623 b



624

625 **Supplementary Figure S2. Synteny and genome map of P-TIM68.** (a) Genome map of the626 P-TIM68 phage. (b) Comparison of synteny between P-TIM68 and the *Synechococcus*

627 myophage S-SM2. The color code of the putative gene functions is the same as in part (a).

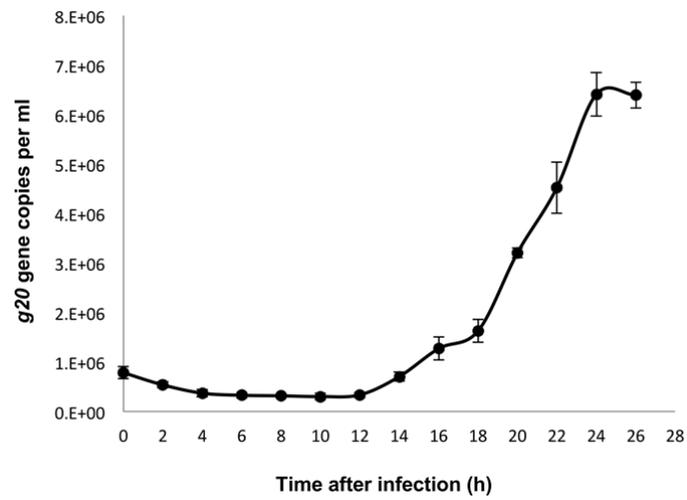
628 Homologous genes are connected with colored lines according to % protein identity.

629

630

631

632



633

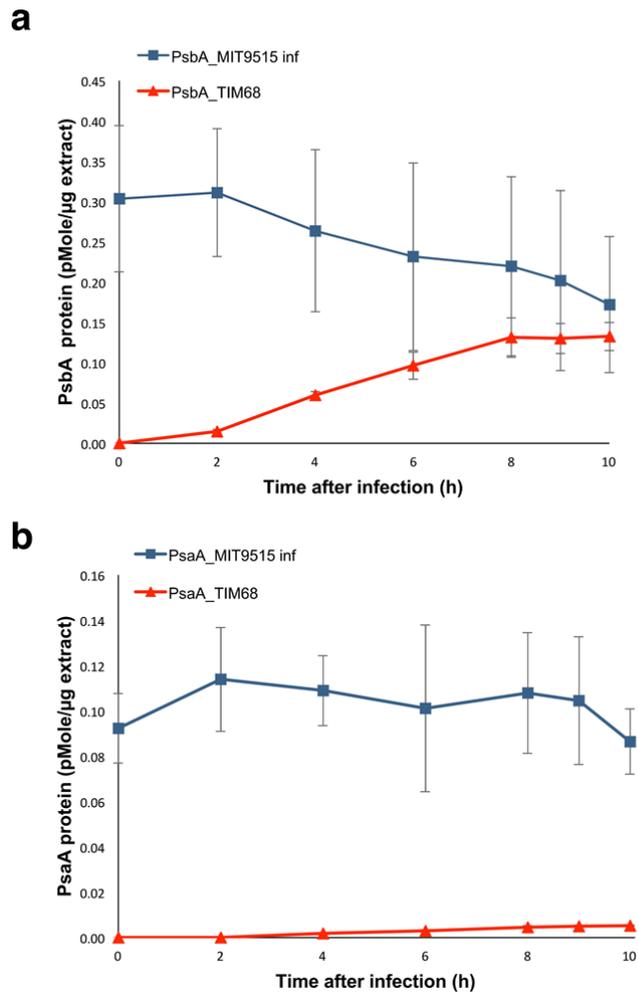
634 **Supplementary Figure S3. Infection of *Prochlorococcus* MIT9515 by the P-TIM68 phage.**

635 Infection dynamics determined from infective phage in the extracellular medium with time

636 after phage addition at 0 h. Average and S.D. of three biological replicates.

637

638

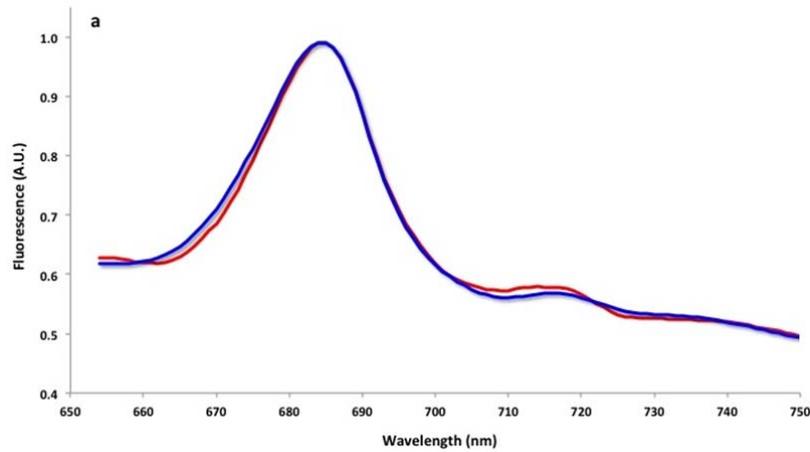


639

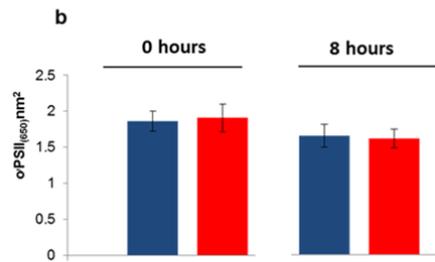
640 **Supplementary Figure S4. Expression of P-TIM68 PSI and PSII proteins in infected**641 ***Prochlorococcus* cells. Levels of phage and host (a) PsbA (PSII) and (b) PsaA (PSI) protein**642 **concentrations. Error bars denote S.D. from three replicate experiments. Graphs of protein**643 **levels normalized to those in uninfected control cells can be found in Fig. 3.**

644

645



646



647

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

657

658

659 **Supplementary Tables**

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

662

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

	Unique viral peptide
<b>Photosystem II</b>	
PsbA	ETTGLSQNYGYK QGILQGWPFCEWVTSTNNR
PsbD	AFDPTQDEETYSMITANR
PsbN	AQEQDY PQHK HPVMTQK
<b>Photosystem I</b>	
PsaA	GLYEFLAESR PQEDFVDKNPVPVSTEK LAPTIQPR NSQLIPDK EILDGQNGFPQGVTHR LHASLYNVSDVFGGSTTMVGGK HSSDLQENSR
PsaB	DFDPTLNK DFDPTLNKDNVLDLDR DLGYSFPCDGPGR FNQDLASDPTR TPIANIAGWR IWDPHFGGATDAYTQAGSTFPVNR QILIEPIFAQWIQAASGK EILEAHTPPR
PsaD	DGVVSER IGQNPVVELK NGEVEFLYPK EQGFELPTGGAALMNEGDNIMYFAR KEQCLALGTQFK TYAETYAITWTSPK FNINGFNPQR NTGNLIDPR TDLKPK
PsaE	IKRPESYWFNECGK
PsaJF	LLETNVTVSPR CILTSAVWPAAEHFEATK AFPVYNELANGK YLIVVR
PsaK	TIGASSEGLQLPKPK

665

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

## 668 REFERENCES

669

- 670 1 Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Bacterial  
671 photosynthesis genes in a virus. *Nature* **424**, 741 (2003).
- 672 2 Millard, A., Clokie, M. R. J., Shub, D. A. & Mann, N. H. Genetic organization  
673 of the *psbAD* region in phages infecting marine *Synechococcus* strains.  
674 *Proc Natl Acad Sci U S A* **101**, 11007-11012 (2004).
- 675 3 Lindell, D. *et al.* Transfer of photosynthesis genes to and from  
676 *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **101**, 11013-11018  
677 (2004).
- 678 4 Sullivan, M. B. *et al.* Prevalence and evolution of core photosystem II genes  
679 in marine cyanobacterial viruses and their hosts. *PLoS Biol* **4**, e234  
680 (2006).
- 681 5 Sharon, I. *et al.* Comparative metagenomics of microbial traits within  
682 oceanic viral communities. *ISME J* **5**, 1178-1190 (2011).
- 683 6 Sharon, I. *et al.* Photosystem-I gene cassettes are present in marine virus  
684 genomes. *Nature* **461**, 258-262 (2009).
- 685 7 Brum, J. *et al.* Patterns and ecological drivers of ocean viral communities.  
686 *Science* **348**, 1261498 (2015).
- 687 8 Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Shedding new light on  
688 viral photosynthesis. *Photosynth Res* **126**, 71-97, doi:10.1007/s11120-  
689 014-0057-x (2015).
- 690 9 Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W.  
691 Photosynthesis genes in marine viruses yield proteins during host  
692 infection. *Nature* **438**, 86-89 (2005).
- 693 10 Clokie, M. R. J. *et al.* Transcription of a 'photosynthetic' T4-type phage  
694 during infection of a marine cyanobacterium. *Environ Microbiol* **8**, 827-  
695 835 (2006).
- 696 11 Bragg, J. G. & Chisholm, S. W. Modeling the fitness consequences of a  
697 cyanophage-encoded photosynthesis gene. *PLoS ONE* **3**, e3550 (2008).
- 698 12 Hellweger, F. L. Carrying photosynthesis genes increases ecological fitness  
699 of cyanophage in silico. *Environ Microbiol* **11**, 1386-1394,  
700 doi:10.1111/j.1462-2920.2009.01866.x (2009).
- 701 13 Millard, A. D., Zwirgmaier, K., Downey, M. J., Mann, N. H. & Scanlan, D. J.  
702 Comparative genomics of marine cyanomyoviruses reveals the  
703 widespread occurrence of *Synechococcus* host genes localized to a  
704 hyperplastic region: implications for mechanisms of cyanophage  
705 evolution. *Environ Microbiol* **11**, 2370-2387 (2009).
- 706 14 Sullivan, M. B. *et al.* Genomic analysis of oceanic cyanobacterial  
707 myoviruses compared with T4-like myoviruses from diverse hosts and  
708 environments. *Environ Microbiol* **12**, 3035-3056 (2010).
- 709 15 Roitman, S. *et al.* Closing the gaps on the viral photosystem-I *psaDCAB*  
710 gene organization. *Environ Microbiol* **17**, 5100-5108 (2015).
- 711 16 Torabi, S. *et al.* PsbN is required for assembly of the photosystem II  
712 reaction center in *Nicotiana tabacum*. *Plant Cell* **26**, 1183-1199 (2014).

- 713 17 Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection  
714 of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci U S A* **108**,  
715 e757-764 (2011).
- 716 18 Ignacio-Espinoza, J. C. & Sullivan, M. B. Phylogenomics of T4 cyanophages:  
717 lateral gene transfer in the 'core' and origins of host genes. *Environ*  
718 *Microbiol* **14**, 2113-2126 (2012).
- 719 19 Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Viruses inhibit CO<sub>2</sub>  
720 fixation in the most abundant phototrophs on Earth. *Curr Biol* **26**, 1585-  
721 1589, doi:10.1016/j.cub.2016.04.036 (2016).
- 722 20 Rusch, D. B. *et al.* The Sorcerer II Global Ocean Sampling expedition: I, The  
723 northwest Atlantic through the eastern tropical Pacific. *PLoS Biol* **5**, e77  
724 (2007).
- 725 21 Dupont, C. L. *et al.* Genomes and gene expression across light and  
726 productivity gradients in eastern subtropical Pacific microbial  
727 communities. *ISME J* **9**, 1076-1092, doi:10.1038/ismej.2014.198 (2015).
- 728 22 Comeau, A. M., Arbiol, C. & Krisch, H. M. Gene network visualization and  
729 quantitative synteny analysis of more than 300 marine T4-like phage  
730 scaffolds from the GOS metagenome. *Mol Biol Evol* **27**, 1935-1944 (2010).
- 731 23 Wang, K. & Chen, F. Prevalence of highly host-specific cyanophages in the  
732 estuarine environment. *Environ Microbiol* **10**, 300-312 (2008).
- 733 24 Doron, S. *et al.* Transcriptome dynamics of a broad host-range  
734 cyanophage and its hosts. *ISME J* **10**, 1437-1455,  
735 doi:10.1038/ismej.2015.210 (2016).
- 736 25 Wilson, W. H., Carr, N. G. & Mann, N. H. The effect of phosphate status on  
737 the kinetics of cyanophage infection of the oceanic cyanobacterium  
738 *Synechococcus* sp. WH7803. *J Phycol* **32**, 506-516 (1996).
- 739 26 Kirzner, S., Barak, E. & Lindell, D. Variability in progeny production and  
740 virulence of cyanophages determined at the single-cell level. *Environ*  
741 *Microbiol Rep*, doi:10.1111/1758-2229.12409 (2016).
- 742 27 Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and  
743 host reveal features of co-evolution. *Nature* **449**, 83-86 (2007).
- 744 28 Clokie, M. R. J. & Mann, N. H. Marine cyanophages and light. *Environ*  
745 *Microbiol* **8**, 2074-2082 (2006).
- 746 29 Farrant, G. K. *et al.* Delineating ecologically significant taxonomic units  
747 from global patterns of marine picocyanobacteria. *Proc Natl Acad Sci U S A*  
748 **113**, E3365-3374, doi:10.1073/pnas.1524865113 (2016).
- 749 30 Fraser, J. M. *et al.* Photophysiological and photosynthetic complex changes  
750 during iron starvation in *Synechocystis* sp. PCC 6803 and *Synechococcus*  
751 *elongatus* PCC 7942. *PLoS One* **8**, e59861,  
752 doi:10.1371/journal.pone.0059861 (2013).
- 753 31 Wyman, M., Gregory, R. P. & Carr, N. G. Novel role for phycoerythrin in a  
754 marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**, 818-820,  
755 doi:10.1126/science.230.4727.818 (1985).
- 756 32 Lindell, D., Padan, E. & Post, A. F. Regulation of *ntcA* expression and nitrite  
757 uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* **180**,  
758 1878-1886. (1998).
- 759 33 Moore, L. R. *et al.* Culturing the marine cyanobacterium *Prochlorococcus*.  
760 *Limnol Oceanogr Methods* **5**, 353-362 (2007).

- 761 34 Haas, A. F. *et al.* Unraveling the unseen players in the ocean - a field guide  
762 to water chemistry and marine microbiology. *J Vis Exp*, e52131,  
763 doi:10.3791/52131 (2014).
- 764 35 Bèjà, O., Fridman, S. & Glaser, F. Viral clones from the GOS expedition with  
765 an unusual photosystem-I gene cassette organization. *ISME J* **6**, 1617-  
766 1620 (2012).
- 767 36 Hevroni, G., Enav, H., Rohwer, F. & Bèjà, O. Diversity of viral photosystem-I  
768 *psaA* genes. *ISME J* **9**, 1892-1898 (2015).
- 769 37 Sabehi, G. *et al.* A novel lineage of myoviruses infecting cyanobacteria is  
770 widespread in the oceans. *Proc Natl Acad Sci U S A* **109**, 2037-2042,  
771 doi:10.1073/pnas.1115467109 (2012).
- 772 38 Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read  
773 assembly using de Bruijn graphs. *Genome Res* **18**, 821-829,  
774 doi:10.1101/gr.074492.107 (2008).
- 775 39 Casjens, S. R. & Gilcrease, E. B. Determining DNA packaging strategy by  
776 analysis of the termini of the chromosomes in tailed-bacteriophage  
777 virions. *Methods Mol Biol* **502**, 91-111, doi:10.1007/978-1-60327-565-  
778 1\_7 (2009).
- 779 40 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-  
780 efficient alignment of short DNA sequences to the human genome.  
781 *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
- 782 41 Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation  
783 initiation site identification. *BMC Bioinformatics* **11**, 119,  
784 doi:10.1186/1471-2105-11-119 (2010).
- 785 42 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local  
786 alignment search tool. *J Mol Biol* **215**, 403-410 (1990).
- 787 43 Zinser, E. R. *et al.* *Prochlorococcus* ecotype abundances in the North  
788 Atlantic Ocean as revealed by an improved quantitative PCR method. *Appl*  
789 *Environ Microbiol* **72**, 723-732, doi:10.1128/AEM.72.1.723-732.2006  
790 (2006).
- 791 44 Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**,  
792 2947-2948 (2007).
- 793 45 Dereeper, A. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-  
794 specialist. *Nucleic Acids Res* **36**, W465-469 (2008).
- 795 46 Guindon, S. *et al.* New algorithms and methods to estimate Maximum-  
796 Likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*  
797 **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).
- 798 47 Whelan, S. & Goldman, N. A general empirical model of protein evolution  
799 derived from multiple protein families using a maximum-likelihood  
800 approach. *Mol Biol Evol* **18**, 691-699 (2001).
- 801 48 Rappaport, F., Béal, D., Joliot, A. & Joliot, P. On the advantages of using  
802 green light to study fluorescence yield changes in leaves. *Biochim. Biophys.*  
803 *Acta* **1767**, 56-65 (2007).
- 804 49 Genty, B., Briantais, J. M. & Baker, N. R. The relationship between the  
805 quantum yield of photosynthetic electron transport and quenching of  
806 chlorophyll fluorescence. *Biochim. Biophys. Acta* **990**, 87-92 (1989).
- 807 50 Béal, D., Rappaport, F. & Joliot, P. A new high-sensitivity 10-ns time-  
808 resolution spectrophotometric technique adapted to in vivo analysis of  
809 the photosynthetic apparatus. *Rev Sci Instrum* **70**, 202-207 (1999).

- 810 51 Joliot, P. & Joliot, A. Quantification of cyclic and linear flows in plants. *Proc*  
811 *Natl Acad Sci U S A* **102**, 4913-4918, doi:10.1073/pnas.0501268102  
812 (2005).
- 813 52 Sunagawa, S. *et al.* Structure and function of the global ocean microbiome.  
814 *Science* **348**, 1261359 (2015).
- 815 53 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2.  
816 *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 817 54 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.  
818 *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352  
819 (2009).
- 820 55 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work  
821 with high-throughput sequencing data. *Bioinformatics* **31**, 166-169  
822 (2015).
- 823 56 O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI:  
824 current status, taxonomic expansion, and functional annotation. *Nucleic*  
825 *Acids Res* **44**, D733-745, doi:10.1093/nar/gkv1189 (2016).
- 826 57 Sharon, I., Pati, A., Markowitz, V. M. & Pinter, R. Y. in *Research in*  
827 *Computational Molecular Biology* Vol. 5541 (ed S. Batzoglou) 496-511  
828 (Springer Berlin / Heidelberg, 2009).
- 829 58 Peng, Y., Leung, H. C., Yiu, S. M. & Chin, F. Y. IDBA-UD: a de novo assembler  
830 for single-cell and metagenomic sequencing data with highly uneven  
831 depth. *Bioinformatics* **28**, 1420-1428, doi:10.1093/bioinformatics/bts174  
832 (2012).
- 833 59 Schreiber, U., Klughammer, C. & Kolbowski, J. High-end chlorophyll  
834 fluorescence analysis with the MULTI-COLOR-PAM. I. Various light  
835 qualities and their applications. *PAM Appl. Notes* **1**, 1-19 (2011).

836