Oceanic cyanobacterial photosynthesis is negatively affected by viral NblA proteins

Omer Nadel[1], Rawad Hanna[1], Andrey Rozenberg[1], Dror Shitrit[1], Ran Tahan[1], Irena Pekarsky[1], Oded Béjà[1,2]*, Oded Kleifeld[1]*, Debbie Lindell[1]*

[1]Faculty of Biology, Technion – Israel Institute of Technology, Haifa 3200003, Israel. [2]The Nancy and Stephen Grand Technion Energy Program (GTEP), Technion – Israel Institute of Technology, Haifa 3200003, Israel.

*e-mail: beja@technion.ac.il; okleifeld@technion.ac.il; dlindell@technion.ac.ii

Marine picocyanobacteria are abundant photosynthetic organisms of global importance. They coexist in the ocean with cyanophages, viruses that infect cyanobacteria. Cyanophages carry many auxiliary metabolic genes acquired from their hosts that are thought to redirect host metabolism for the phage's benefit[1–5]. One such gene is *nblA* which is present in multiple cyanophage families[2,6–8]. Under nutrient deprivation the cyanobacterial NblA is responsible for inducing proteolytic degradation of the phycobilisome[9–11], the large cyanobacterial photosynthetic light harvesting complex. This increases the pool of amino acids available for essential tasks[11], serving as a survival mechanism[12]. Ectopic expression of different cyanophage *nblA* genes results in host pigment protein degradation[9,6]. However, the benefit of the cyanophage-encoded NblA for the cyanophage and the broader impact on the host are unknown. Here, using a recently developed genetic manipulation system for marine cyanophages[14], we reveal that cyanophage NblA significantly accelerates the cyanophage infection cycle, directs degradation of the host phycobilisome and other proteins and reduces host photosynthetic light harvesting efficiency. Metagenomic analysis revealed that cyanophages carrying *nblA* are widespread in the oceans and compose 35% and 65% of oceanic T7-like cyanophages in the surface and deep photic

zones, respectively. Our results show a large benefit of the *nblA* gene to the cyanophage while exerting a negative effect on the host photosynthetic apparatus and host photosynthesis. These findings suggest that cyanophage *nblA* has an adverse global impact on light harvesting by oceanic picocyanobacteria.

The light-harvesting antennae of most cyanobacteria are composed of phycobilisomes, gigantic membrane-associated soluble supercomplexes, located on the exterior side of the thylakoid membrane[9,15]. Cyanobacteria capture and funnel solar energy through their phycobilisomes towards the photosynthetic reaction centers[16] where the primary energy conversion reactions of photosynthesis occur. Typical phycobilisomes of marine *Synechococcus* are built of multiple subunits that absorb light at different wavelengths, including an allophycocyanin (APC) core and peripheral rods composed of disks of different phycobiliproteins, phycocyanin (PC) and phycoerythrin-I and -II (PEI, PEII) as well as linker proteins that connect rod subunits to each other and to the core and attach the phycobilisome to the thylakoid membrane[17,18]. Each phycobiliprotein is composed of α and β subunits that form heterodimers[10]. These proteins bind the pigment chromophores phycocyanobilin, phycoerythrobilin (PEB) and phycourobilin (PUB), which absorb light at different wavelengths[19]. Due to their large size, phycobilisomes can comprise up to half of the soluble protein content of a cell[15]. *Prochlorococcus* is the most abundant photosynthetic organism on Earth[20]. The vast majority of *Prochlorococcus* lineages have a chlorophyll-based transmembranal light harvesting complex rather than the typical phycobilisome[21–23]. *Prochlorococcus* does, however, code for one or two subunits of phycoerythrin (PE)[24,25], found at low levels in cells[26]. Interestingly, some low-light adapted *Prochlorococcus* lineages from anoxic marine zones possess both chlorophyll-based and phycobilisome light harvesting complexes[20].

Cyanophage NblA is beneficial for infection

Evidence collected over the past two decades suggests that cyanophages possess host-like auxiliary metabolic genes that are thought to influence key processes in host metabolism during infection and enhance phage replication[1–4]. These include genes that are thought to benefit the phage by maintaining essential cellular metabolic processes as well as those that impair host activities to redirect resources towards phage progeny production[3,4]. The *nblA* gene codes for a proteolysis adaptor protein found in multiple cyanophage families[2,6–8,13,27], and is particularly prevalent in T7-like cyanophages (family *Autographiviridae*). In nutrient deprived cyanobacteria, the cyanobacterial NblA is responsible for inducing degradation of the phycobilisome complex, resulting in clear loss of pigmentation[9–11]. Here, we use S-TIP37, a typical T7-like cyanophage isolated from the Red Sea that infects the open ocean *Synechococcus* sp. strain WH8109[28], as our model system to study cyanophage-encoded *nblA*. We began by examining whether the cyanophage *nblA* gene is transcribed and translated during the infection process. We found high *nblA* transcript and protein levels from two hours post-infection onwards (Extended Data Fig. 1). This expression pattern was distinct to that of a host copy of the *nblA* gene which displayed fairly consistent transcript and protein levels (Extended Data Fig. 1). The phage *nblA* expression pattern was similar to that of the phage DNA polymerase gene, and is thus expressed with phage genes involved in DNA metabolism and replication and prior to virion morphogenesis genes[29,30].

Next, we assessed the influence of the *nblA* gene on cyanophage infection. For this we constructed *nblA* cyanophage deletion mutants (Δ*nblA*) using the REEP approach[14] (see Methods). Infection of the host with the Δ*nblA* mutant phage revealed visibly less pigmentation loss than when the host cells were infected with the wild-type (WT) phage (Fig. 1a), indicating a clear phenotype for the Δ*nblA* mutant during infection.

To investigate the importance of the cyanophage NblA to the infection process, we performed phage growth curve experiments. The Δ*nblA* mutant demonstrated a substantially longer latent period (the period of infection prior to the first release of phage

90 progeny from cells) of 8-10 h compared to the 3 h latent period of the WT[31] (Fig. 1b). In
91 addition, synthesis of phage proteins increased with time during infection with the WT
92 phage (Fig. 1c, Suppl. Data File 1), whereas most phage protein production ceased after 3
93 h during infection with the Δ*nblA* mutant with significantly fewer virion morphogenesis
94 proteins produced in the latter (Fig. 1c). These findings suggest a direct link between phage
95 protein production and NblA expression and indicate that NblA is advantageous to the WT
96 phage. Since cyanobacterial NblAs are known to degrade phycobilisomes during nutrient
97 stress[9–11], and to release amino acids critical for cell survival[11,12], we hypothesize that the
98 phage NblA has a similar role during infection but that the release of amino acids is used
99 for building new virions.

100 The gene density in the region of *nblA* in the phage genome is high, thus to
101 verify that the observed phenotype is a result of the lack of the *nblA* gene and not
102 due to potential side-effects caused by the deletion (see Methods), we performed
103 a rescue experiment. This was achieved by infecting *Synechococcus* sp. strain
104 WH8109 cells that were ectopically expressing the cyanophage *nblA* gene under an
105 inducible expression system[32,33] with the Δ*nblA* mutant phage (Fig. 1d). Indeed, expression
106 of the cyanophage NblA restored the timing of infection to that found for the WT
107 cyanophage (Fig. 1d). The shorter latent period for the WT phage suggests that the
108 *nblA* gene would improve phage fitness by allowing it to carry out more rounds of
109 infection. Our findings confirm that the *nblA* gene plays an important function in
110 cyanophage infection and that it likely enhances phage fitness.

111 Cyanophage NblA would be expected to impact host photosynthesis if it directs
112 phycobilisome degradation since phycobilisomes transfer energy to the photochemical
113 reaction centers of the photosystems (PSI or PSII)[16,34]. We assessed the photosynthetic
114 performance of PSII by measuring fluorescence at room temperature after exciting
115 cultures with blue light (495 nm) which preferentially excites the most peripheral disks of
116 the phycobilisome rods consisting of PEII. The apparent photochemical quantum yield of
117 PSII (QY_p) was then calculated. While uninfected cells maintained a steady PSII quantum
118 yield, infection with both the WT and Δ*nblA* mutant phages resulted in a considerable

119 decline (Fig. 1e). Notably, this decline reached low levels by 8 h after infection with the
120 Δ*nblA* mutant, a delay of 4 h relative to infection with the WT cyanophage. This indicates
121 that the cyanophage NblA negatively impacts the host's PSII photochemical quantum yield
122 by 50% over the period of infection. This is likely to have a dramatic effect on the ability of
123 the host cell to produce the energy and reducing equivalents necessary for many cellular
124 processes. This diminished photosynthetic performance could be attributed to targeted
125 phycobilisome degradation, accelerated cell lysis, or a combination of both factors.

126

127 Cyanophage NblA causes degradation of host phycobilisomes

128 Cyanobacterial NblAs are known to be involved in phycobilisome degradation. To assess
129 whether the cyanophage NblA has a similar effect on host phycobilisome integrity, we
130 monitored whole-cell fluorescence emission spectra as a measure of the energy transfer
131 pathway from phycobilisomes to PSII. Samples collected at different time points were
132 excited at the wavelength for PUB, which binds mainly to the peripheral PEII disk of the
133 phycobilisome, and the corresponding fluorescence emission spectra were measured.
134 Infection with the WT cyanophage led to a rapid and substantial decrease in the
135 chlorophyll *a* (Chl) peak with time (Fig. 2a, Extended Data Fig. 2a). This decline was first
136 obvious at 3-h post-infection (Fig. 2a), prior to the onset of cell lysis (Fig. 1b). In contrast,
137 infection by the Δ*nblA* knockout cyanophage led to a more gradual decrease in Chl
138 fluorescence (Fig. 2a, Extended Data Fig. 2a). These findings indicate that energy transfer
139 from PE to Chl was rapidly interrupted during infection with the WT phage, but
140 considerably less so with the Δ*nblA* mutant phage.

141 Whole-cell absorbance spectra and proteomic analysis provide a means to
142 assess the levels of host phycobilisome subunits during infection. At 12 hours post-
143 infection with the WT cyanophage, we observed a flattening of the peaks
144 corresponding to both the PUB and PEB chromophores, which are attached to the
145 PEI and PEII proteins. A less prominent decrease in these peaks was observed during
146 infection with the Δ*nblA* mutant (Fig. 2b), suggestive of NblA degradation of the PE
147 proteins by the WT phage. Proteomic analysis during infection with the WT phage revealed

148 significantly lower levels of the α subunits of PEII, PEI and PC rods (CpeA, MpeA and RpcA,
149 respectively), as well as two rod linker proteins (MpeC and MpeE) relative to infection with
150 the Δ*nblA* mutant at 4-h post-infection (Suppl. Data File 1). A 10%-30% reduction was
151 observed for some of the β subunits, but this was not statistically significant. In
152 contrast, three linker proteins closer to the core[35] that connect PC to APC and both
153 PC and APC to the thylakoid membrane (ApcC, CpcG1, CpcG2) and one rod linker
154 protein (CpeC) had higher protein levels during infection with the WT relative to the Δ*nblA*
155 mutant (Suppl. Data File 1). This may be indicative of attempts by the host to refurbish the
156 phycobilisome core structure in response to the reduction in active phycobilisome
157 rods. Combined, these findings provide further evidence for the role of cyanophage
158 NblA in phycobilisome disassembly and degradation.

159 Next, we wished to determine the fate of the antenna complex during infection. To
160 achieve this, we isolated phycobilisomes from host cells six hours post-infection and
161 compared them to those in uninfected cells. Fractionation of phycobilisomes on a sucrose
162 gradient followed by fluorescence spectra analysis (Fig. 2c) allowed us to differentiate
163 between assembled (lower, high density, pink bands) and disassembled complexes (upper,
164 low density, orange bands) (see Extended Data Fig. 3). While intact and disassembled
165 phycobilisomes were apparent in uninfected host cells, only disassembled phycobilisomes
166 were present in cells infected by the WT cyanophage (Fig. 2c). Cells infected by the Δ*nblA*
167 mutant cyanophage were composed of both intact and degraded phycobilisomes (Fig. 2c),
168 indicative of only partial phycobilisome disassembly.

169 The findings above are consistent with a role for NblA in disassembly and
170 degradation of the phycobilisome during infection with the WT cyanophage. However,
171 these findings could also be due, at least partially, to the beginning of cell lysis from
172 4-h onwards during infection with the WT phage. To more directly assess the impact
173 of phage NblA on host phycobilisomes without the confounding myriad effects of
174 phage infection, we ectopically expressed the phage NblA in host cells without
175 infection. First, we measured whole cell fluorescence spectra which yielded a
176 decline in the energy transfer from PE to Chl, seen by a reduction in the Chl

177 fluorescence peak after excitation at PUB, relative to cells without induction of NblA
178 expression (Fig. 2d). This was, however, to a lesser extent than that found during
179 infection with the WT phage (Fig. 2a). Second, ectopic expression of the phage NblA
180 resulted in some reduction in the PEB and PUB absorbance peaks (Fig. 2e). Third, we
181 assessed whether the cyanophage NblA could restore phycobilisome disassembly during
182 infection with the Δ*nblA* cyanophage mutant. Indeed, NblA expression led to a decrease in
183 the Chl fluorescence peak when excited at PUB, declining to levels similar to those during
184 infection with the WT cyanophage by 3-4-h post-infection (right panels of Extended
185 Data Fig. 4). Note that the partial decline in Chl fluorescence at 4-h post-infection
186 in the control without induction (left panels of Extended Data Fig. 4) is likely caused
187 by leaky expression in this inducible system[33]. Taken together, our findings suggest
188 a clear effect of the cyanophage NblA on energy transfer from PE to Chl, consistent
189 with a role in phycobilisome disassembly and degradation during infection.

190 Next, we assessed the protein targets of the cyanophage NblA protein by
191 following the production of new protein cleavages in host cells ectopically
192 expressing the viral NblA. For this, we analyzed proteins with newly produced N-
193 termini using mass spectrometry[36] at 24 and 48 h post-induction (at which time maximum
194 protein induction occurs[33]). Many new cleavages were observed (Suppl. Data File 3).
195 Among these, 209 cleavage events in 107 proteins were markedly more abundant (≥2-fold)
196 following viral NblA expression, including 81 cleavages in 59 proteins at 24 hours and 138
197 cleavages in 68 proteins at 48 hours post-induction (Fig. 3b, Suppl. Data File 3). These
198 include several cleavages of phycobilisome subunits (Fig. 3a, Extended Data Fig. 5).
199 At 24 h post-induction, four phycobiliproteins were cleaved (Fig. 3b,c); the
200 peripheral PEII disk α subunit (MpeA), the phycocyanin β subunit RpcB**,** the rod
201 linker MpeE and the ApcE linker responsible for energy transfer from the APC core
202 to the chlorophylls of the photosynthetic reaction centers in the thylakoid
203 membranes[9]. At 48 h post-induction, many of the new proteolytic cleavages were
204 of phycobilisome subunits (Fig. 3a,b). These included the same three PEII, PEI and
205 PC rod disk α subunits that displayed low levels during infection with the WT phage
206 (MpeA, CpeA and RpcA) (see above) as well as the three PE rod disk β subunits

207 (MpeB, CpeB and RpcB), one core protein (ApcB) and three linker proteins
208 connecting between disks in the rods (CpeC, MpeD and MpeE) (Fig. 3b,c, Suppl.
209 Data File 3). These proteolytic cleavages spanned almost all phycobilisome rod
210 proteins, and ApcB of the phycobilisome core (Fig. 3c). These results provide
211 further evidence that the cyanophage NblA protein leads to the degradation and
212 disassembly of the cyanobacterial phycobilisome antenna complex. They also
213 suggest that the cyanophage NblA directs disassembly in a stepwise fashion,
214 initially to disconnect the host phycobilisome from the thylakoid membrane, and to
215 degrade peripheral PEII disks, followed by cleavage of the more internal PEI and
216 PC disks. This is similar to the process known for cyanobacterial NblA during
217 nutrient stress in which phycobilisome degradation begins at the peripheral rod
218 disks and moves inwards towards the phycobilisome core[11]. Intriguingly, Ma et al.[37]
219 reported that infection by a cyanophage lacking a detectable *nblA* gene leads to the
220 release of PE peptides, suggesting that cyanophages without *nblA* may have an
221 alternative mechanism for degrading outer PE rods.

222 To our surprise, ectopic expression of the cyanophage NblA led to proteolytic
223 processing of additional photosynthesis-related proteins beyond phycobilisome
224 proteins (Extended Data Fig. 5, Suppl. Data File 3). These include those related to
225 PSI (PsaC, PsaK), the PSII oxygen evolution complex (PsbO, PbsU), and carbon
226 fixation (CcmK2, CsoS2). This suggests that the cyanophage NblA negatively
227 impacts host photosynthesis through proteolytic degradation not only of
228 phycobilisome components but also of reaction center proteins and proteins
229 needed for carbon fixation. Moreover, a set of proteins with functions other than
230 those related to photosynthesis were also cleaved as a result of expression of the
231 cyanophage NblA (Extended Data Fig. 5). These include proteins involved in
232 translation, with the majority of cleavages being in ribosomal proteins at 24 h post-
233 induction (Fig. 3a,b). Translation initiation and elongation factors were also cleaved
234 48 h post-induction (Fig. 3a,b, Suppl. Data File 3). Other key cellular proteins
235 cleaved include proteins involved in carbon metabolism, a subunit of the cellular
236 RNA polymerase (RpoC2), the cell division protein FtsZ, as well as porin and
237 transporter proteins (Fig. 3b, Suppl. Data File 3). It is not known whether host NblA

238 proteins also direct the degradation of these proteins. Cleavage of these proteins
239 would thus disrupt multiple cellular processes during infection, in addition to
240 photosynthesis.

241 In freshwater cyanobacteria, NblA proteins function by binding to
242 phycobilisome subunits as well as to the ClpC chaperone responsible for recruiting
243 the protease that induces proteolytic degradation of the phycobilisome[11,38]. Since
244 proteolytic enzymes often have preferred cleavage sites, we investigated cleavage
245 sites after ectopic expression of the cyanophage NblA in the *Synechococcus* host
246 (Fig. 3d). At 24 h post-induction, we detected cleavages primarily between medium-
247 sized, polar or partially polar residues (Met, Gln, Thr) and small, polar or
248 nucleophilic residues (Ser, Thr, His), while at 48 h post-induction an additional
249 cleavage site was detected between Lys and Arg residues (Fig. 3d), providing an
250 explanation for the additional cleaved proteins at 48 h post-induction. This could be
251 a result of NblA recruiting different proteases or that the different substrates are
252 cleaved by the same proteolytic system (presumably the Clp protease complex)
253 that has broad cleavage specificity.

254

255 Cyanophage *nblA* is common in the oceans

256 T7-like cyanophages are very abundant in the oceans[31,39]. However, the prevalence
257 of the *nblA* gene in this group is largely unknown. Therefore, to understand the
258 environmental significance of our findings, we first assessed the distribution of *nblA*
259 genes in assembled genomes of both isolated and environmental T7-like
260 cyanophages. Our search revealed that 46% of the complete non-redundant T7-
261 like cyanophage genomes code for *nblA* genes (Extended Data Fig. 6). *nblA*-
262 encoding cyanophages have been isolated on picocyanobacterial hosts from the
263 three main lineages: marine *Synechococcus* subcluster 5.1 (*Parasynechococcus*[40]
264 in the broad sense), *Synechococcus* subcluster 5.2 (*Cyanobium* and related
265 genera) and *Prochlorococcus* (the *Prochlorococcus* collective) (Fig. 4), even though
266 the vast majority of the latter do not have phycobilisomes as their light-harvesting
267 antennae[21,23]. We also found this gene to be present in a rare prophage residing in

268 a *Synechococcus* subcluster 5.1 genome, AG-670-B23[41]. Thus, the cyanophage
269 *nblA* gene is quite common in the genomes of T7-like cyanophages, infecting both
270 *Synechococcus* and *Prochlorococcus.*

271 A close examination showed two types of *nblA* genes with differential
272 distribution among the phylogenetic clades of T7-like cyanophages (Fig. 4 and
273 Extended Data Figs. 6,7). *nblA* genes of one type are dominant and are present in
274 clade B cyanophages (to which S-TIP37 belongs) and to a lesser extent in clade A.
275 These *nblA* genes are consistently located downstream of the exonuclease gene
276 and frequently have up to a 110-nt overlap with it. *nblA* genes of another type are
277 restricted to some P-RSP2-like cyanophages (designated here as clade R) and are
278 located downstream of an HNH endonuclease gene. Neither *nblA* gene type was
279 found in the genomes of cyanophages from the newly discovered clade C (Fig. 4).

280 To assess the relationships between NblA proteins from the T7-like
281 cyanophages and those of their hosts, we performed a targeted search for NblA
282 homologs among marine picocyanobacteria. To our surprise, we found *nblA* genes
283 not only in marine *Synechococcus*, but also in many *Prochlorococcus* low-light-
284 adapted ecotypes (Extended Data Fig. 7,8). While marine *Synechococcus* typically
285 possesses several *nblA* genes, with e.g. *Synechococcus* sp. strain WH8109 having
286 five intact genes and a pseudogene, *Prochlorococcus* strains have a single *nblA*
287 gene. These findings indicate that *nblA* genes are also found in cyanobacteria that
288 do not possess phycobilisomes as well as in the cyanophages that infect them.

289 NblAs from the marine picocyanobacteria are substantially divergent from
290 the well-characterized freshwater cyanobacterial NblAs (as exemplified by the
291 chromosomal NblA from *Nostoc* sp. strain PCC 7120) with respect to sequence, yet
292 their predicted structures preserve the classical dimeric NblA architecture
293 (Extended Data Fig. 8). However, NblA from the T7-like cyanophages demonstrate
294 higher overall sequence similarity to the well-characterized cyanobacterial NblAs
295 rather than to those of their hosts (Extended Data Fig. 8). The T7-like cyanophage
296 NblAs cluster together in a phylogenetic network, suggesting that they represent a
297 monophyletic group (see Extended Data Fig. 7). Together these findings suggest

298 that the acquisition of *nblA* by T7-like cyanophages occurred in the distant
299 evolutionary past. This is in contrast to *nblA* from freshwater cyanophages infecting
300 *Microcystis* and *Planktothrix* which have high similarity to *nblA* genes from the
301 corresponding cyanobacterial groups[6,8,27].

302 We now turned to assessing the relative abundance of T7-like cyanophages
303 carrying the *nblA* gene in the oceans. We made use of the consistent genomic
304 association of the dominant *nblA* type with the exonuclease gene, a core gene
305 found in all T7-like cyanophages sequenced so far[43,44], to interrogate the read data
306 from the Global Ocean Viromes (GOV) metagenomic dataset. This allowed us to
307 determine both the presence or absence of the *nblA* gene and the relative
308 abundance of the different clades from the exonuclease gene. Our quantification
309 strategy yielded general distribution patterns of T7-like cyanophages similar to
310 previously reported studies. Cyanophages belonging to clades B and C dominated
311 the global oceans[31,43], while clade A and R phages were relatively rare. The vast
312 majority of cyanophages from clade B possess the *nblA* gene (Fig. 5), with ~72%
313 of those at the surface, and 89% of those deeper in the photic zone at the deep
314 chlorophyll maximum (DCM), carrying the gene. The proportion of *nblA*-encoding
315 cyanophages among the less abundant clade A group was much lower, averaging
316 ~24% at the surface and 10% at the DCM. In accordance with the distribution of the
317 *nblA* genes in complete genomes, none of the clade C and R cyanophages had
318 *nblA* genes downstream of the exonuclease gene. Since the *nblA* gene is located
319 in a different genomic position in clade R, our quantification method could not be
320 applied to them. Overall ~35% and 65% of the T7-like cyanophages (from clades
321 A, B and C) coded for the *nblA* gene in the surface and at the DCM, respectively.
322 These findings indicate that *nblA* is widespread amongst the abundant clade B
323 cyanophages both at the surface and deeper in the photic zone.

324 Among the T7-like cyanophage isolates that code for the *nblA* gene, half
325 infect marine *Synechococcus* strains while the rest infect phycobilisome-less
326 *Prochlorococcus* strains (Fig. 4). Both groups of cyanobacteria are prevalent in
327 surface waters, whereas *Prochlorococcus* is considerably more abundant than

328 marine *Synechococcus* at the DCM[21,45]. Thus, we anticipate that many of the clade
329 B phages at the DCM infect *Prochlorococcus*. Since the vast majority of
330 *Prochlorococcus* lineages have a different type of light harvesting antenna to the
331 phycobilisomes of most cyanobacteria[21–23], we expect that NblA of the phages
332 infecting *Prochlorococcus* direct the degradation of other proteins. This possibility
333 is supported by the finding that multiple low-light-adapted *Prochlorococcus*
334 ecotypes code for an *nblA* gene (Extended Data Fig. 7), even though they do not
335 have phycobilisomes. Furthermore, our proteomics data indicate that NblA-directed
336 degradation is not restricted to phycobilisome proteins, even for a clade B phage
337 that infects *Synechococcus* (Fig. 3a,b, Extended Data Fig. 5). Thus, other
338 photosynthesis-related proteins and proteins involved in other cellular processes,
339 are likely targeted by many *nblA*-encoding cyanophages deep in the photic zone of
340 the oceans.

341 Cyanobacteria acclimate to low light to maximize their ability to harvest
342 light[41]. This is achieved through increasing the number and size of their antenna
343 complexes[46–48] as well as the number of their photosystems[49]. As such, more
344 cyanobacterial resources are bound up in the photosynthetic apparatus at depth.
345 Thus, we propose that the ability of cyanophages to degrade the photosynthetic
346 apparatus deep in the photic zone would provide a considerable advantage through
347 the release of amino acids for use in cyanophage progeny production and may
348 explain why relatively more cyanophages at depth carry the *nblA* gene.

349 Interestingly, the general distribution of the dominant *nblA* type in T7-like
350 cyanophages parallels that for the *psbA* gene (coding for the PSII reaction center
351 protein D1). Both genes are present in the majority of clade B cyanophages,
352 sporadically in clade A, and are absent from clade C. However, they do not always
353 co-occur (see Fig. 4 and Extended Data Fig. 6, see also[28]). The *nblA* gene is more
354 common in T7-like cyanophages than the *psbA* gene, with 36% of the complete
355 genomes (including environmentally assembled genomes) across all clades coding
356 for *psbA* compared to 46% for *nblA* (see above). Since the *psbA* gene is considered
357 to be one of the most common auxiliary metabolic genes in cyanophage

358 genomes[2,3,28,43,44], our results suggest that the cyanophage-encoded *nblA* gene is
359 very common in oceanic environments as well.

360 An intriguing aspect of T7-like cyanophages that carry both *psbA* and *nblA*
361 is that *psbA* is thought to enable continued photosynthetic energy production during
362 infection[3,50,51], while *nblA* degrades the complex that harvests the light funneled to
363 the photosystems (this study). How can these potentially conflicting functions for
364 two different auxiliary metabolic genes in the same cyanophage be reconciled? We
365 propose that these genes are most beneficial to the phage under different
366 conditions. The *psbA* gene has been hypothesized to be more important under
367 conditions of high light[52,53] where more photodamage to the D1 protein is expected.
368 In contrast, *nblA* is likely to be more important to the phage under low light
369 conditions where more of the cell's resources are invested in building the
370 photosynthetic apparatus (see above). In addition, under high light conditions,
371 degradation of the phycobilisome by cyanophage NblA could serve to reduce
372 photodamage to D1[7], such that photosynthesis can continue at some level despite
373 less light being harvested by the phycobilisome. Thus, a balance between the
374 activities of these two proteins is likely needed at high light.

375 The vast abundance of clade B T7-like cyanophages in the oceans coding
376 for the *nblA* gene raises the possibility that they have a global impact on
377 cyanobacterial photosynthesis. This group of cyanophages is considerably more
378 abundant deep in the photic zone than at the surface[31,55] where a very large
379 percentage of them code for *nblA* (Fig. 5). While it is difficult to quantify the global
380 impact of a single gene, we have performed some rough calculations to estimate
381 the potential effect of viral *nblA* on picocyanobacterial light harvesting. We base our
382 calculation on the findings that the cyanophage gene caused a 50% reduction in
383 PSII photosynthetic performance (Fig. 1e), that 1-15% of marine cyanobacterial
384 cells are infected by T7-like cyanophages at a given time[39,56,57], and that 35-65% of
385 marine T7-like cyanophages code for *nblA* in different layers of the photic zone (Fig.
386 5) (see Methods). We estimate that 0.2-1.1% and 1.6-4.9% of cyanobacterial
387 photosynthesis in the surface and DCM layers, respectively, could be impacted by

388 *nblA*. These estimations suggest that, collectively, T7-like cyanophages carrying
389 the *nblA* gene have a global impact, reducing picocyanobacterial photosynthetic
390 light harvesting by 0.2-5% in the upper oceans.

391

392 Concluding remarks

393 Auxiliary metabolic genes in viruses[58,59] are a widespread phenomenon, especially
394 in marine cyanophages[1–4]. Previous studies have investigated the putative function
395 of several such genes using *in-vitro* biochemical approaches[60,61] or ectopic
396 expression[6,13,61,62]. Here we combined a recently developed cyanophage genetic
397 engineering system[14], together with inducible ectopic expression[33] and state-of-the-
398 art N-termini proteomics[36] to directly examine the role of *nblA*, a gene coding for a
399 small proteolysis adaptor, in cyanophage infection and its impact on the host
400 photosynthetic apparatus. While NblA in cyanobacteria serve as a stress response
401 mechanism under nutrient deprivation and mediate the controlled degradation of
402 the phycobilisome photosynthetic antenna[9–11], we propose that its cyanophage
403 counterpart is important not for cell survival, but for virion synthesis. Our results
404 further reveal that the cyanophage NblA directs degradation not only of host
405 phycobilisomes, but also of a suite of other proteins, including important core
406 photosynthesis and house-keeping proteins. Thus, this small auxiliary metabolic
407 gene is likely to have a large impact on both the host's light harvesting efficiency,
408 energy production, and on other essential cellular processes.

409 The implications of the *nblA* gene for the cyanophage are dramatic with an
410 approximately 3-fold shorter infection cycle, suggestive of considerably improved
411 fitness. These findings, together with genomic and metagenomic results of high
412 abundance and widespread distribution of cyanophages carrying the *nblA* gene in
413 the oceans (this study), and in freshwater ecosystems[7,8,13], suggests that these
414 cyanophages have a negative global effect of up to 5% on the amount of light
415 harvested by oceanic cyanobacteria.

416

417 ## References

418 1. Breitbart, M., Thompson, L. R., Suttle, C. A. & Sullivan, M. B. Exploring the vast
419 diversity of marine viruses. *Oceanography* **20**, 135–139 (2007).

420 2. Gao, E.-B., Huang, Y. & Ning, D. Metabolic genes within cyanophage genomes:
421 implications for diversity and evolution. *Genes* **7**, 80 (2016).

422 3. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Shedding new light on viral
423 photosynthesis. *Photosynth. Res.* **126**, 71–97 (2015).

424 4. Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral
425 infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, 21–34 (2020).

426 5. Tian, F. *et al.* Prokaryotic-virus-encoded auxiliary metabolic genes throughout the
427 global oceans. *Microbiome* **12**, 159 (2024).

428 6. Nadel, O. *et al.* An uncultured marine cyanophage encodes an active phycobilisome
429 proteolysis adaptor protein NblA. *Environ. Microbiol. Rep.* **11**, 848–854 (2019).

430 7. Yoshida, T. *et al.* Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates
431 diverse cyanophage genome strategies. *J. Bacteriol.* **190**, 1762–1772 (2008).

432 8. Gao, E.-B., Gui, J.-F. & Zhang, Q.-Y. A novel cyanophage with a cyanobacterial
433 nonbleaching protein A gene in the genome. *J. Virol.* **86**, 236–245 (2012).

434 9. Collier, J. L. & Grossman, A. R. A small polypeptide triggers complete degradation of
435 light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J.* **13**,
436 1039–1047 (1994).

437 10. Bienert, R., Baier, K., Volkmer, R., Lockau, W. & Heinemann, U. Crystal structure of
438 NblA from *Anabaena* sp. PCC 7120, a small protein playing a key role in
439 phycobilisome degradation. *J. Biol. Chem.* **281**, 5216–5223 (2006).

440 11. Sendersky, E. *et al.* The proteolysis adaptor, NblA, is essential for degradation of the
441 core pigment of the cyanobacterial light-harvesting complex. *Plant J. Cell Mol. Biol.*
442 **83**, 845–852 (2015).

443 12. Schwarz, R. & Grossman, A. R. A response regulator of cyanobacteria integrates
444 diverse environmental signals and is critical for survival under extreme conditions.
445 *Proc. Natl. Acad. Sci. U. S. A.* **95**, 11008–11013 (1998).

446 13. Ou, T., Gao, X.-C., Li, S.-H. & Zhang, Q.-Y. Genome analysis and gene *nblA*
447 identification of *Microcystis aeruginosa* myovirus (MaMV-DC) reveal the evidence for
448 horizontal gene transfer events between cyanomyovirus and host. *J. Gen. Virol.* **96**,
449 3681–3697 (2015).

450 14. Shitrit, D. *et al.* Genetic engineering of marine cyanophages reveals integration but
451 not lysogeny in T7-like cyanophages. *ISME J.* **16**, 488–499 (2022).

452 15. Grossman, A. R., Schaefer, M. R., Chiang, G. G. & Collier, J. L. Environmental effects
453 on the light-harvesting complex of cyanobacteria. *J. Bacteriol.* **175**, 575–582 (1993).

454 16. Ueno, Y., Aikawa, S., Kondo, A. & Akimoto, S. Energy transfer in cyanobacteria and
455 red algae: confirmation of spillover in intact megacomplexes of phycobilisome and
456 both photosystems. *J. Phys. Chem. Lett.* **7**, 3567–3571 (2016).

457 17. Ong, L. J. & Glazer, A. N. Phycoerythrins of marine unicellular cyanobacteria. I. Bilin
458 types and locations and energy transfer pathways in *Synechococcus* spp.
459 phycoerythrins. *J. Biol. Chem.* **266**, 9515–9527 (1991).

460 18. Six, C. *et al.* Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.:
461 a comparative genomics study. *Genome Biol.* **8**, R259 (2007).

462 19. Glazer, A. N. Light guides: Directional energy transfer in a photosynthetic antenna. *J.*
463 *Biol. Chem.* **264**, 1–4 (1989).

464 20. Ulloa, O. *et al.* The cyanobacterium *Prochlorococcus* has divergent light-harvesting
465 antennae and may have evolved in a low-oxygen ocean. *Proc. Natl. Acad. Sci.* **118**,
466 e2025638118 (2021).

467 21. Partensky, F., Hess, W. R. & Vaulot, D. *Prochlorococcus*, a marine photosynthetic
468 prokaryote of global significance. *Microbiol. Mol. Biol. Rev.* **63**, 106–127 (1999).

469 22. Ting, C. S., Rocap, G., King, J. & Chisholm, S. W. Cyanobacterial photosynthesis in the
470 oceans: the origins and significance of divergent light-harvesting strategies. *Trends*
471 *Microbiol.* **10**, 134–142 (2002).

472 23. Wiethaus, J., Busch, A. W. U., Dammeyer, T. & Frankenberg-Dinkel, N.
473 Phycobiliproteins in *Prochlorococcus marinus*: Biosynthesis of pigments and their
474 assembly into proteins. *Eur. J. Cell Biol.* **89**, 1005–1010 (2010).

475 24. Hess, W. R. *et al.* The photosynthetic apparatus of Prochlorococcus: Insights through
476 comparative genomics. *Photosynth. Res.* **70**, 53–71 (2001).

477 25. Becker, J. W. *et al.* Novel isolates expand the physiological diversity of
478 Prochlorococcus and illuminate its macroevolution. *mBio* **15**, e03497-23 (2024).

479 26. Steglich, C., Mullineaux, C. W., Teuchner, K., Hess, W. R. & Lokstein, H. Photophysical
480 properties of *Prochlorococcus marinus* SS120 divinyl chlorophylls and phycoerythrin
481 in vitro and in vivo. *FEBS Lett.* **553**, 79–84 (2003).

482 27. Meza-Padilla, I., McConkey, B. J. & Nissimov, J. I. Structural models predict a
483 significantly higher binding affinity between the NblA protein of cyanophage Ma-
484 LMM01 and the phycocyanin of *Microcystis aeruginosa* NIES-298 compared to the
485 host homolog. *Virus Evol.* **10**, veae082 (2024).

486 28. Dekel-Bird, N. P. *et al.* Diversity and evolutionary relationships of T7-like podoviruses
487 infecting marine cyanobacteria. *Environ. Microbiol.* **15**, 1476–1491 (2013).

488 29. Lindell, D. *et al.* Genome-wide expression dynamics of a marine virus and host reveal
489 features of co-evolution. *Nature* **449**, 83–86 (2007).

490 30. Huang, S., Sun, Y., Zhang, S. & Long, L. Temporal transcriptomes of a marine
491 cyanopodovirus and its *Synechococcus* host during infection. *MicrobiologyOpen* **10**,

492 e1150 (2021).

493 31. Maidanik, I. *et al.* Cyanophages from a less virulent clade dominate over their sister
494 clade in global oceans. *ISME J.* **16**, 2169–2180 (2022).

495 32. Nakahira, Y., Ogawa, A., Asano, H., Oyama, T. & Tozawa, Y. Theophylline-dependent
496 riboswitch as a novel genetic tool for strict regulation of protein expression in
497 cyanobacterium *Synechococcus elongatus* PCC 7942. *Plant Cell Physiol.* **54**, 1724–
498 1735 (2013).

499 33. Tahan, R., Shitrit, D. & Lindell, D. A functional cyanophage thioredoxin increases
500 competitive phage fitness. In prep (2024).

501 34. Biswas, A., Akhtar, P., Lambrev, P. H. & van Stokkum, I. H. M. Energy transfer from
502 phycobilisomes to photosystem I at room temperature. *Front. Plant Sci.* **14**, (2024).

503 35. Kondo, K., Mullineaux, C. W. & Ikeuchi, M. Distinct roles of CpcG1-phycobilisome and
504 CpcG2-phycobilisome in state transitions in a cyanobacterium *Synechocystis* sp. PCC
505 6803. *Photosynth. Res.* **99**, 217–225 (2009).

506 36. Hanna, R. *et al.* In-depth characterization of apoptosis N-terminome reveals a link
507 between caspase-3 cleavage and posttranslational N-terminal acetylation. *Mol. Cell.*
508 *Proteomics* **22**, 100584 (2023).

509 37. Ma, X., Coleman, M. L. & Waldbauer, J. R. Distinct molecular signatures in dissolved
510 organic matter produced by viral lysis of marine cyanobacteria. *Environ. Microbiol.*
511 **20**, 3001–3011 (2018).

512 38. Karradt, A., Sobanski, J., Mattow, J., Lockau, W. & Baier, K. NblA, a key protein of
513 phycobilisome degradation, interacts with ClpC, a HSP100 chaperone partner of a
514 cyanobacterial Clp protease. *J. Biol. Chem.* **283**, 32394–32403 (2008).

515 39. Carlson, M. C. G. *et al.* Viruses affect picocyanobacterial abundance and
516 biogeography in the North Pacific Ocean. *Nat. Microbiol.* **7**, 570–580 (2022).

517 40. Komárek, J., Johansen, J. R., Šmarda, J. & Strunecký, O. Phylogeny and taxonomy of
518 *Synechococcus*-like cyanobacteria. *Fottea* **20**, 171–191 (2020).

519 41. Berube, P. M. *et al.* Single cell genomes of *Prochlorococcus*, *Synechococcus*, and
520 sympatric microbes from diverse marine environments. *Sci. Data* **5**, 180154 (2018).

521 42. Biller, S. J., Berube, P. M., Lindell, D. & Chisholm, S. W. *Prochlorococcus*: the structure
522 and function of collective diversity. *Nat. Rev. Microbiol.* **13**, 13–27 (2015).

523 43. Cai, L. *et al.* Abundant and cosmopolitan lineage of cyanopodoviruses lacking a DNA
524 polymerase gene. *ISME J.* **17**, 252–262 (2023).

525 44. Labrie, S. J. *et al.* Genomes of marine cyanopodoviruses reveal multiple origins of
526 diversity. *Environ. Microbiol.* **15**, 1356–1376 (2013).

527 45. Flombaum, P. *et al.* Present and future global distributions of the marine
528 Cyanobacteria *Prochlorococcus* and *Synechococcus*. *Proc. Natl. Acad. Sci. U. S. A.* **110**,
529 9824–9829 (2013).

530 46. Sanfilippo, J. E., Garczarek, L., Partensky, F. & Kehoe, D. M. Chromatic acclimation in
531 cyanobacteria: a diverse and widespread process for optimizing photosynthesis.
532 *Annu. Rev. Microbiol.* **73**, 407–433 (2019).

533 47. Kolodny, Y. *et al.* Phycobilisome light-harvesting efficiency in natural populations of
534 the marine cyanobacteria *Synechococcus* increases with depth. *Commun. Biol.* **5**, 727
535 (2022).

536 48. Garczarek, L., Hess, W. R., Holtzendorff, J., van der Staay, G. W. M. & Partensky, F.
537 Multiplication of antenna genes as a major adaptation to low light in a marine
538 prokaryote. *Proc. Natl. Acad. Sci.* **97**, 4098–4101 (2000).

539 49. Herranen, M., Tyystjärvi, T. & Aro, E.-M. Regulation of photosystem I reaction center
540 genes in *Synechocystis* sp. strain PCC 6803 during light acclimation. *Plant Cell Physiol.*
541 **46**, 1484–1493 (2005).

542 50. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. Marine ecosystems:
543 bacterial photosynthesis genes in a virus. *Nature* **424**, 741 (2003).

544 51. Lindell, D., Jaffe, J. D., Johnson, Z. I., Church, G. M. & Chisholm, S. W. Photosynthesis
545 genes in marine viruses yield proteins during host infection. *Nature* **438**, 86–89
546 (2005).

547 52. Bragg, J. G. & Chisholm, S. W. Modeling the fitness consequences of a cyanophage-
548 encoded photosynthesis gene. *PLoS ONE* **3**, e3550 (2008).

549 53. Hellweger, F. L. Carrying photosynthesis genes increases ecological fitness of
550 cyanophage in silico. *Environ. Microbiol.* **11**, 1386–1394 (2009).

551 54. Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole.
552 *Cell* **177**, 1109-1123.e14 (2019).

553 55. Goldin, S., Hulata, Y., Baran, N. & Lindell, D. Quantification of T4-like and T7-like
554 cyanophages using the polony method show they are significant members of the
555 virioplankton in the North Pacific Subtropical Gyre. *Front. Microbiol.* **11**, 1210 (2020).

556 56. Gilbert, N. E. *et al.* Seasonal enhancement of the viral shunt catalyzes a subsurface
557 oxygen maximum in the Sargasso Sea. 2025.01.23.634377 Preprint at
558 https://doi.org/10.1101/2025.01.23.634377 (2025).

559 57. Shopen Gochev, C. *et al.* Cold surface waters of the sub-Antarctic Pacific Ocean
560 support high cyanophage abundances and infection levels. *Environ. Microbiol.* **27**,
561 e70031 (2025).

562 58. Bidle, K. D. & Vardi, A. A chemical arms race at sea mediates algal host–virus
563 interactions. *Curr. Opin. Microbiol.* **14**, 449–457 (2011).

564 59. Anantharaman, K. *et al.* Sulfur oxidation genes in diverse deep-sea viruses. *Science*
565 **344**, 757–760 (2014).

566 60. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of
567 cyanobacterial host carbon metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E757-64

(2011).

61. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W. & Frankenberg-Dinkel, N. Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr. Biol.* **18**, 442–448 (2008).

62. Rihtman, B. *et al.* Cyanophage MazG is a pyrophosphohydrolase but unable to hydrolyse magic spot nucleotides. *Environ. Microbiol. Rep.* **11**, 448–455 (2019).

Figures

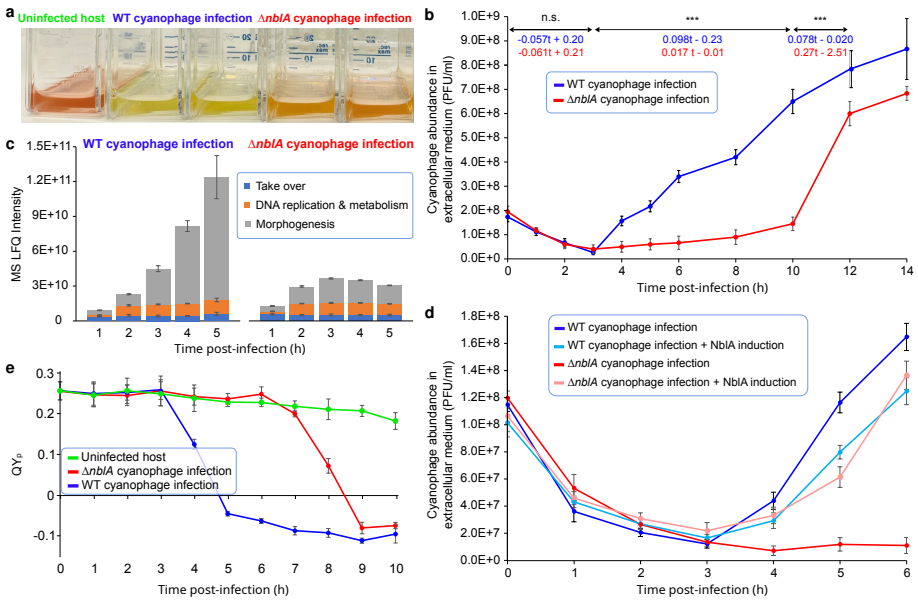


Fig. 1. Influence of the *nblA* gene on S-TIP37 cyanophage infection dynamics. a, Culture pigmentation phenotypes in uninfected *Synechococcus* sp. strain WH8109, and in WH8109 6 h post-infection when infected with WT S-TIP37 or the Δ*nblA* mutant S-TIP37 cyanophage. **b**, Cyanophage growth curves of the WT and Δ*nblA* mutant cyanophages, *n*=3. **c**, Proteomic characterization of different functional groups of phage proteins with time after infection of the cyanobacterium with the WT (left) and Δ*nblA* (right) cyanophages. Linear models were fit for every functional group and a significant contribution of the factor "infection type" was found for the morphogenesis genes (FDR-adjusted *p*-value < 0.001). For further statistical details refer to Suppl. Data File 2, *n*=3. **d**, Cyanophage infection dynamics of the WT and Δ*nblA* mutant phages when infecting the wild-type host or the host ectopically expressing the cyanophage NblA, *n*=3. **e**, Apparent photochemical quantum yield of PSII (QY_p) in cells infected with the WT or the *ΔnblA* cyanophages and in uninfected control cells after preferential excitation of the phycobilisome at 495 nm (the maximal absorbance of PEII). Infection was performed at a multiplicity of infection (MOI) of 5, *n*=3. For **b**,

summary of the linear regression fits are provided for the three near-linear segments: formulas of the relative cyanophage abundance for WT (in blue) and *ΔnblA* cyanophages (in red) as a function of time are provided alongside the significance levels for the factor "infection type" (type II Wald *F* test): n.s. – non-significant, *** – FDR-adjusted p-value < 0.001, see Suppl. Data File 2 for further details. For **c**, genes are grouped into three clusters and the cluster names reflect functions of the majority of the corresponding genes: take over (genes STIP37_1-12 in GenBank record MH540083.1), DNA replication and metabolism (genes STIP37_13-23 and STIP37_55), and morphogenesis (genes STIP37_24-54). The *nblA* gene is located between genes STIP37_18 and STIP37_19. The plotted values in panels **b**-**e** correspond to averages and the whiskers reflect standard deviations.

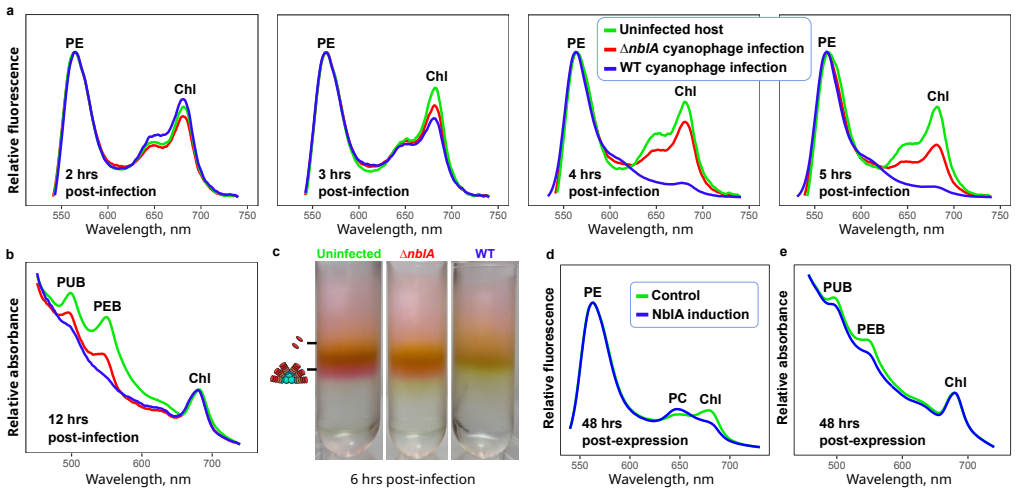
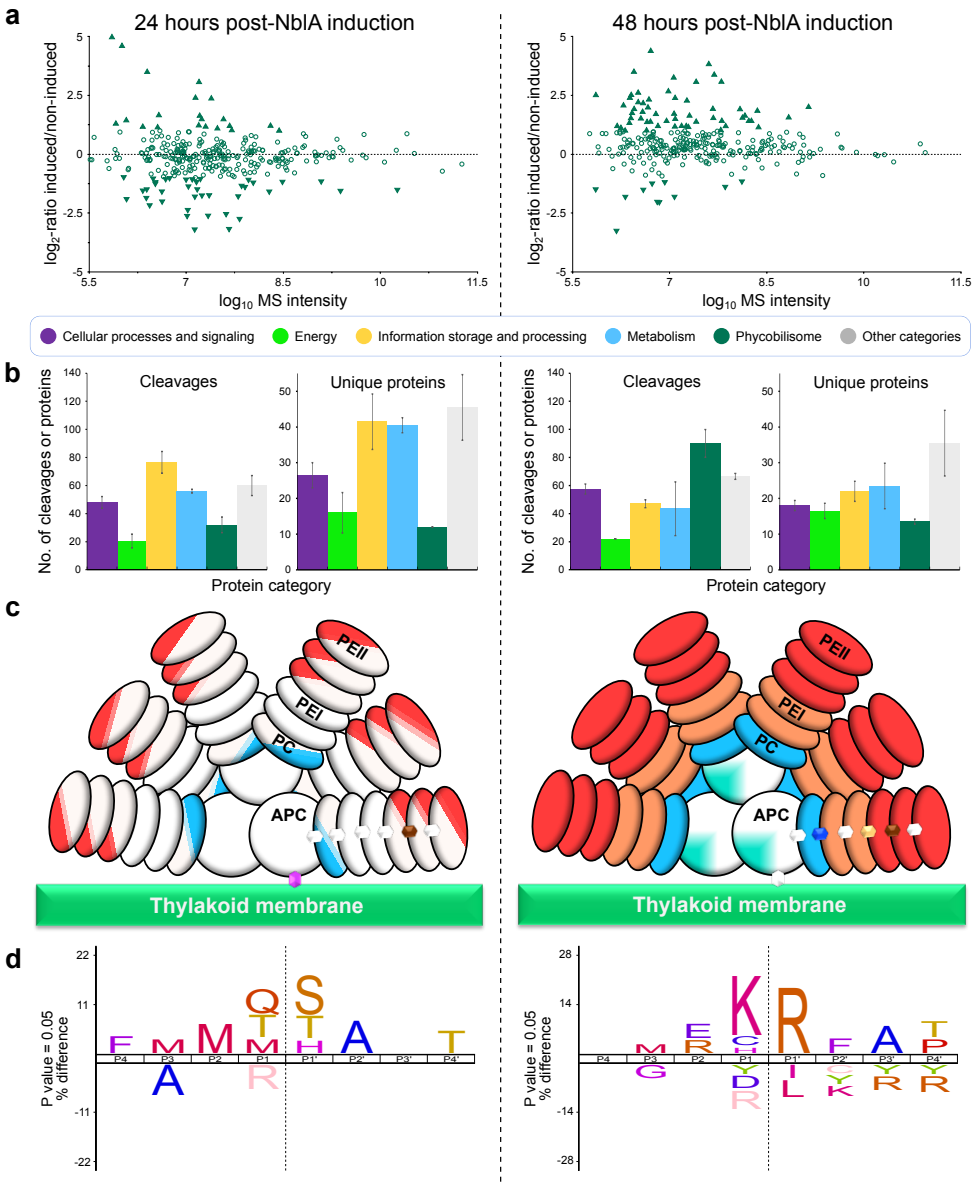


Fig. 2. Effect of cyanophage NblA on the cyanobacterial host photosynthetic antennae. a, Fluorescence emission spectra, **b**, absorbance spectra, and **c**, isolated phycobilisomes, of uninfected *Synechococcus* sp. strain WH8109 cells (green), and WH8109 infected by the WT (blue) or Δ*nblA* mutant (red) S-TIP37 cyanophages at different time points post-infection. In **c**, the lower density (upper) bands are of disassembled phycobilisome subunits, and the higher density (lower) bands are of intact phycobilisomes (see Extended Data Fig. 3). **d**, Fluorescence,

614 and **e**, absorbance spectra of WH8109 cells 48-h after induced expression of the
615 cyanophage NblA protein (blue) compared to the non-induced control (green).
616 Fluorescence and absorbance spectra are representatives of 5 biological
617 replicates. Averages and standard deviations are shown in Extended Data Fig. 2.
618 Measurements were taken at room temperature, and peaks of Chl, PC and PE
619 peaks are indicated. Fluorescence measurements were normalized on the PE peak
620 and absorbance measurements were normalized on the Chl peak. In **a** and **d**, the
621 PE peak at 560 nm is made up of PEI and PEII, to which the PEB and PUB
622 chromophores bind.

623



624

625 **Fig. 3. Proteolytic patterns in host cells upon ectopic expression of the**
626 **cyanophage NblA.** Proteomic analysis of newly cleaved proteins after expression
627 of the cyanophage NblA in *Synechococcus* sp. strain WH8109 cells for 24 h (left
628 panels) and 48 h (right panels), as compared to non-induced cells, *n*=2. **a**, MA plots
629 showing the abundance ratios of neo-N-terminal peptides from phycobilisome
630 proteins plotted against their MS signal intensity. This signal intensity serves as an
631 indicator of the peptides' overall abundances. Filled triangles indicate neo-N-
632 terminal peptides from proteins that show absolute $\log_2$-ratios ≥1 in both
633 replicates. Triangles with upward-pointing and downward-pointing apexes

634 represent peptides more and less abundant, respectively, upon NblA ectopic
635 expression relative to the control. **b**, The numbers of cleavages and unique cleaved
636 proteins in each category that exhibited a significant abundance change upon
637 NblA expression. Proteins are categorized according to a modified classification of
638 COG functional groups (see Methods), with photosynthesis-related proteins other
639 than phycobilisome proteins included in the "Energy" group. **c**, Phycobilisome
640 cartoons showing proteolytic cleavages in the phycobilisome subunits making up
641 the discs. Colors indicate phycobiliproteins and linker proteins that underwent
642 cleavages in APC (turquoise), PC (blue), PEI (orange) and PEII (red), the linker
643 proteins between the discs, the core and the thylakoid membrane (various colors).
644 Semi-colored discs represent cleavages detected only in the PEII alpha subunits
645 and PC beta subunits. Quarter-colored cores represent cleavages detected only in
646 the APC beta subunits. Fully colored discs indicate cleavages detected in both the
647 alpha and beta subunits. **d**, Preferred cleavage motifs upon ectopic expression of
648 the cyanophage NblA. The amino acid sequence logo is for neo-N-terminal
649 peptides that were identified only after S-TIP37 NblA expression and that had
650 abundance ratios that were at least 2-fold higher in the NblA expressing cells.
651 Cleavage sites are marked with a dashed line.

652

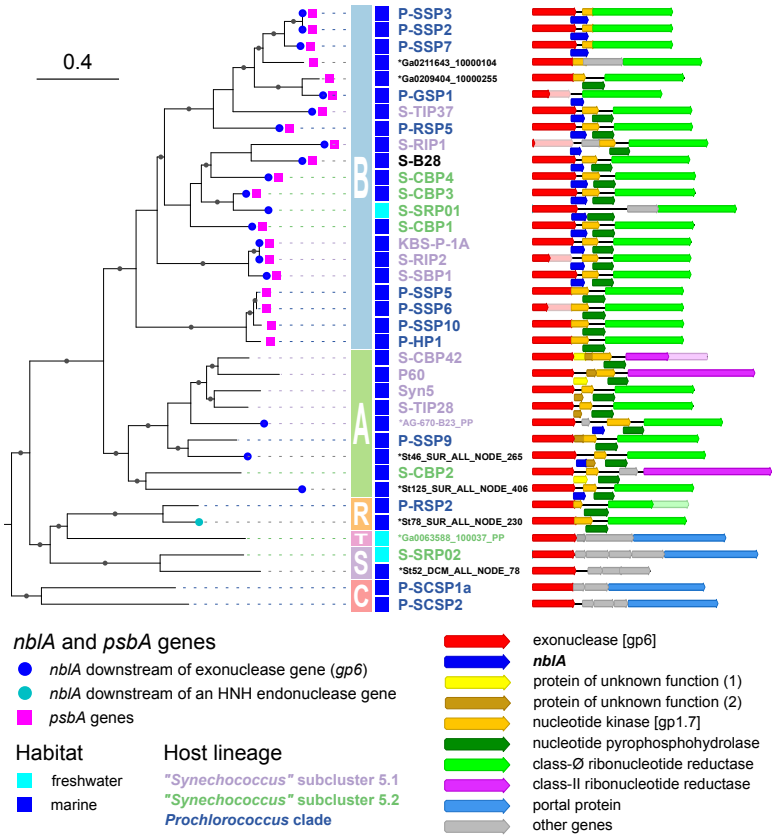


Fig. 4. Distribution of *nblA* genes among isolated T7-like cyanophages and selected environmental phage genomes. Phylogenetic tree is based on concatenation of nine core genes (see Methods), is outgroup-rooted and was pruned to include the chosen genomes (see Extended Data Figs. 5 and Suppl. Data Files 4, 5 for the complete phylogeny of the cyanophages and the associated metadata). The genomes are subdivided into previously delineated clades A, B and C, as well as the newly defined minor clades: R (P-RSP2-like phages), S (S-SRP02-like phages) and T (no isolated representatives). For each representative, the genomic region containing the exonuclease gene and downstream of it, is shown. Frequently-appearing orthologous genes are indicated with color. Asterisks indicate environmental genomes. Solid circles mark branches with ultrafast bootstrap support values ≥95.

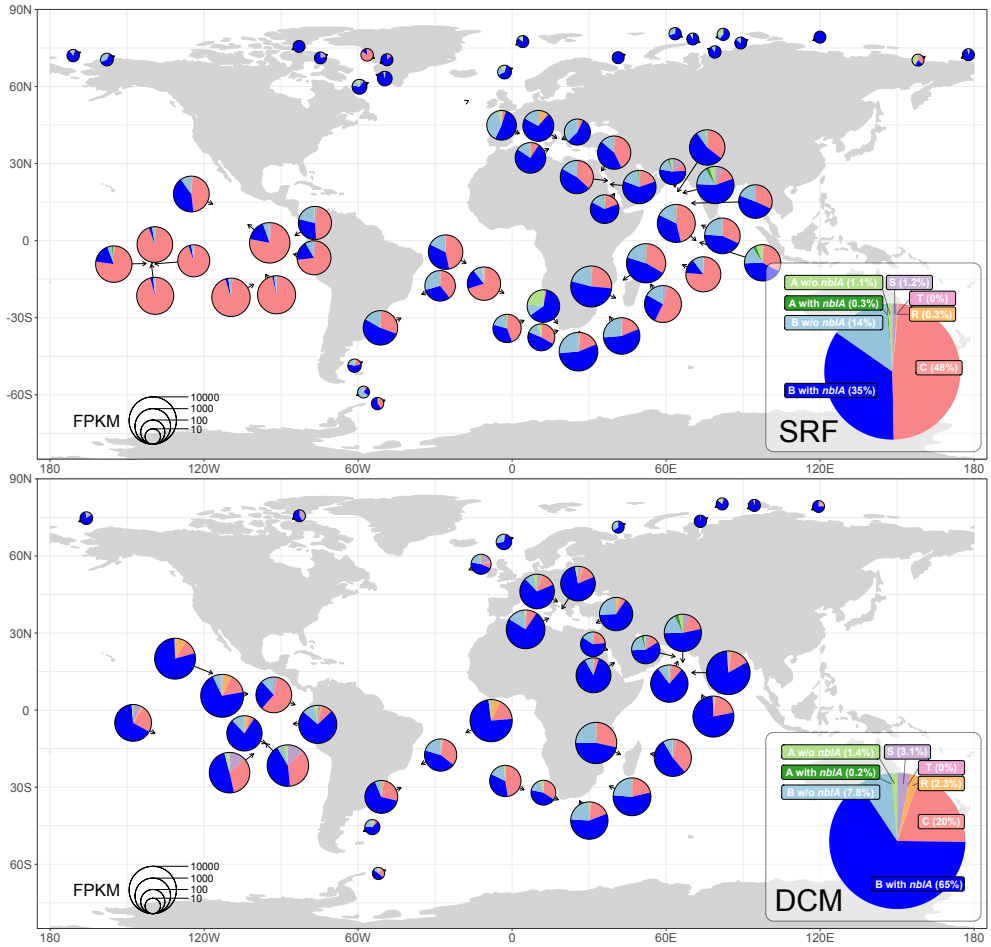


Fig. 5. Global distribution of T7-like cyanophages with and without *nblA* genes. **a**, Surface waters. **b**, Deep chlorophyll maximum (DCM). Quantifications are based on fragments per kilobase per million total reads (FPKM) abundances for exonuclease genes with and without downstream *nblA* genes and assigned to one of the five clades of T7-like cyanophages. Raw data and exonuclease-containing contigs for read recruitment were retrieved from the Global Ocean Viromes 2.0 dataset (GOV 2.0)[54]. The insets show the relative abundances of the different clades of cyanophages across the data (based on the FPKM values added together) and serve as the legend for the colors on the maps.

677 **Methods**

678 **Cyanobacterial growth**

679 S*ynechococcus* sp. strain WH8109 was grown in artificial seawater (ASW) medium[63], with
680 modifications as described in Lindell et al.[64]. Cultures were grown at 21 °C and at a light
681 intensity of 20 µmol photons m^{-2} s^{-1} , under a 14:10 light-dark cycle. Pour-plates were
682 obtained with ASW medium with low-melting point agarose at a final concentration of
683 0.28% with additional 1 mM of sodium sulfite. Heterotrophic helper strain, *Alteromonas*
684 sp. EZ55, was added to the pour-plate mixture for isolation of *Synechococcus* colonies[65].

685

686 **Quantitative reverse transcriptase real-time PCR (qRT-PCR)**

687 For RNA extraction, a 1 ml culture sample was harvested by centrifugation at 4 °C, 15,000*g*
688 for 15 minutes and the pellets were flash frozen in liquid nitrogen. Cell pellets were thawed
689 and incubated with lysozyme L6876-5G (SIGMA ALDRICH) at a final concentration of 30
690 mM, and 200 units of RNase inhibitor (Murine BioLabs) for 60 minutes at 37 °C. An equal
691 volume of lysis buffer was added and cell debris were centrifuged at 4 °C, 16,000*g* for 1
692 minute. Nucleic acids in the supernatant were precipitated with an equal volume of 95%
693 ethanol and centrifuged as described above. RNA wash buffer (500 µl) from Monarch Total
694 RNA Miniprep Kit (NEB #T2010) was added, and the sample centrifuged again for 30
695 seconds. The supernatant was discarded and the step was repeated. DNase I reaction
696 buffer (5 µl) and 4 units of DNase I from the TURBO DNA-free™ Kit were added to the 45
697 µl sample and incubated for 10 minutes at 4 °C. Samples were incubated with 0.5 µl of 0.5
698 M of EDTA, pH 8.0, for 10 minutes at 75 °C.

699 Total RNA was reverse transcribed into cDNA using the LunaScript® RT SuperMix
700 Kit (New England Biolabs, E3010). The reaction mixture was prepared in a total volume of
701 20 µl, 4 µl of LunaScript RT SuperMix containing random hexamers, and 16 µl of nuclease-
702 free water, with the RNA sample added to reach the final volume. As a control, a similar

703 mixture was prepared, but without addition of the reverse transcriptase enzyme. The
704 reaction was incubated at 25 °C for 2 minutes, 55 °C for 10 minutes, and 95 °C for 1 minute.

705 Quantitative PCR (qPCR) reactions were prepared using LightCycler 480 SYBR Green
706 I Master mix from Roche, combined with 0.2 µM of each primer and the *nblA*/*rnpB* DNA
707 template. The reactions were run on a LightCycler 480 Real-Time PCR System. Cycle
708 threshold fluorescence values for each reaction were determined using LightCycler 480
709 software. To quantify DNA copy numbers, a standard curve was generated by running 10-
710 fold serial dilutions of the template and correlating cycle threshold values to known DNA
711 concentrations. See Suppl. Data File 6.

712 The genes whose expression levels were analyzed with qPCR were as follows: host
713 *nblA2* gene (Syncc8109_1607, see explanation below, primers 5'-
714 GCGATCAAGCGGTCAATCAAC-3' and 5'-CTCTCTGCCGCACGTAGAGG-3'), host *rnpB* gene
715 (Syncc8109_0157, primers 5'-CATCGGCGGTGTGTTTCT-3' and 5'-CAGGCTTGCTGGGT-3'), S-
716 TIP37 *nblA* gene (primers 5'-TTCCCGAGGCAGACAAGAG-3' and 5'-
717 TAATGGGATGGTGACTCGGC-3'), S-TIP37 DNA polymerase gene (STIP37_17B, primers 5'-
718 TGAGCTACTACGCAACAGGC-3' and 5'-AGCGCGATCATTCAGGGAAG-3'). The *Synechococcus*
719 sp. strain WH8109 *nblA* gene chosen for qPCR is the one we reported previously[6], although
720 a more refined remote homology search with hhsearch[66] using a custom NblA profile,
721 reveals that the genome carries four additional *nblA*-like genes and an *nblA* pseudogene.
722 In order to clarify which homolog has the highest structural similarity to previously
723 characterized proteins and is thus most likely to have a function similar to that of
724 freshwater cyanobacterial NblAs, we obtained the structures for the corresponding
725 monomers with ColabFold[67] and searched them against the Protein Data Bank with
726 Foldseek[68]. The best match indeed obtained for NblA2 (Syncc8109_1607), with the highest
727 similarity to NblA protein from *Nostoc* sp. PCC 7120[10] (PDB: 1OJH.E, probability 0.94, TM-
728 score 0.6621).

729

730 **Cyanophage infection experiments**

731 Prior to infection, *Synechococcus* sp. strain WH8109 was grown in liquid medium to mid-
732 log growth to ~1×10^8 cells ml^{-1}. Infection experiments were initiated by adding the S-TIP37
733 cyanophage strains at a multiplicity of infection (MOI) of 5. Infection dynamics were
734 determined by collecting samples at hourly time intervals during the initial six hours,
735 followed by sampling every two hours thereafter. Samples were filtered over a 0.22 µm
736 syringe filter and the filtrate containing free cyanophages was plated to determine the
737 number of infective cyanophages using the plaque assay[69] in semi-solid pour plates (see
738 above). Statistical analysis of the infection course was performed with segmented linear
739 regression by selecting near-linear ranges in the infection curves and obtaining linear fits
740 with factors time post-inoculation, cyanophage type (wild-type or mutant) and their
741 interaction and replicate as a random effect using lme4 v. 1.1-31[70]. Significance of the
742 model terms was tested with the Anova function from the car package v. 3.1-1[71].
743 Distributions of the residuals were checked with QQ-plots.

744

745 **Spectral measurements**

746 Absorbance spectra were measured at room temperature using a double beam Shimadzu
747 spectrophotometer equipped with deuterium and halogen lamps. Spectral data were
748 obtained at a constant bandpass with a resolution of 1 nm. Absorbance spectra were
749 normalized to chlorophyll (680 nm). Fluorescence emission spectra were collected using
750 the Jobin-Yvon Horiba Spectrofluorometer. Samples were placed in 1 ml micro quartz
751 cuvettes, excited at 495 nm and fluorescence emission measured from 500 nm to 750 nm
752 and normalized to the fluorescence intensity at the emission maximum of PEI (562-564
753 nm).

754 Apparent photochemical quantum yield of PSII (QY_p), or efficiency of light
755 utilization by PSII, was determined by measuring fluorescence emission, after dark
756 adaptation of 15 minutes, with and without the addition of (3-(3,4-dichlorophenyl)-1,1-
757 dimethylurea (DCMU) to a final concentration of 10 µM. DCMU blocks electron transfer
758 between the primary quinone electron acceptor (Q_A) and the secondary quinone electron

759 acceptor (Q_B) on the acceptor side of PSII, causing the reaction centers to remain in a
760 closed state[72]. Samples were excited at 495 nm and fluorescence emission was measured
761 from 500 nm to 750 nm. The apparent PSII quantum yield was calculated as described by
762 Schoffman & Keren[73], from the onset of infection up to ten hours post-infection, at which
763 point all values fell below zero. QY_P was calculated as $QY_P = (QY_{F+DCMU} - QY_{F-DCMU}) / QY_{F+DCMU}$.
764 Absorbance-corrected fluorescence quantum yield (QY_F) was calculated as $QY_F = F/f_{495}$
765 where F is the integrated area under the fluorescence emission spectrum, and f_{495} is the
766 fraction of incident light actually absorbed at 495 nm (f_{495} = 1 - 10^-A_{495}; A_{495} is the
767 measured absorbance at 495 nm).

768 Quantification of the absorption levels of the different pigments was performed by
769 subtracting the background with the package baseline v. 1.3-4[74] using the 'modpolyfit'
770 method, identifying the peaks of PUB and PEB and chlorophyll and scaling the intensities
771 to the intensity of the chlorophyll peak. Differences between the uninfected cells and cells
772 infected with the wild-type and mutant cyanophages 12 h post-infection were analyzed
773 with linear models and Tukey's post-hoc tests. Fluorescence emission spectra were scaled
774 to the maximum intensity (intensity of the PE peak) to obtain the relative emission
775 intensity of the chlorophyll peak. Since the fluorescence measurements were performed
776 repeatedly for the same batches, a mixed model was built to analyze the relative
777 fluorescence of chlorophyll as a function of infection (uninfected cells and two types of
778 cyanophages), time post-infection, their interaction taking into account the random effect
779 of batch with lmer4. Distributions of the residuals were checked with QQ-plots. Anova
780 function from the car package was used to obtain analysis of deviance tables and *p*-values
781 for the model terms, Tukey's test was performed as post-hoc in both cases.

782

783

784

785 **Phycobilisome extraction**

786 *Synechococcus* sp. strain WH8109 cells were harvested by centrifugation at 6000*g* for 10
787 minutes at 4 °C, 6 hours post-infection and kept at -80 °C. Pellets were thawed at room
788 temperature, resuspended with 7.5 M K-Phosphate, pH 7.5, and homogenized. Cells were
789 disrupted with a microfluidizer (Microfluidics Corporation HC-2000) at 80 psi. The
790 supernatant was collected after centrifugation at 20,000*g*, for 30 minutes at 4 °C and
791 loaded onto a 0.25-1.25 M linear sucrose gradient in 7.5 M K-Phosphate buffer, pH 7.5.
792 Samples were spun down in an ultracentrifuge at 170,000*g* for 18 hours at 4 °C. Fractions
793 were collected with a syringe and analyzed by fluorometry at 77K and spectrophotometry
794 as described above.

795

796 **Construction of S-TIP37 *nblA* deletion mutant**

797 The open reading frame (ORF) encoding the *nblA* gene is 237 bp long (NCBI accession
798 MH540083.1, located at 12,467-12,706 nt) and was previously identified by us in the
799 genome of S-TIP37 based on a match to our custom NblA protein profile[6]. Profile-profile
800 searches with hhsearch via the HHpred server[75] for this protein yield high-scoring matches
801 to cyanobacterial NblA sequences and Pfam NblA profile (PF04485). Analogously, folding
802 the protein with ColabFold[67] produces a typical dimeric structure of NblA with the two
803 monomers composed of two helices each and unstructured terminal regions (see Ext. Data
804 Fig. 7).

805 The *nblA* mutant was produced following Shitrit et al[14]. Since phages gain random
806 mutations readily, we constructed two independent S-TIP37 Δ*nblA* mutant cyanophages
807 to ensure that the observed phenotype was directly related to the lack of the *nblA* gene.
808 Briefly, the *nblA* ORF was used to prepare a construct containing a recombination template
809 for viral homologous recombination. We replaced 137 bp of the *nblA* gene from position
810 12,572 to 12,707 with a tag of 23 bp. Since the *nblA* gene has an overlap of 94 nt with the
811 upstream exonuclease ORF *exo*, the region chosen for the deletion covered the part of
812 *nblA* not overlapping this gene. However, this region covered the part of *nblA* overlapping
813 a downstream ORF coding for a protein homologous to gp1.7, a nucleotide kinase, of the

814 T7 *Escherichia coli* phage, resulting in deletion of 41 nt at the 5' end of the gene. (This short
815 gene is under-annotated in T7-like cyanophage genomes and was not originally annotated
816 in S-TIP37). The construct was cloned into the replicative pRL-proCAT[14] plasmid and
817 conjugated into *Synechococcus* sp. strain WH8109[14,76]. The strain expressing the
818 recombination template was infected with the WT S-TIP37. The lysate was filtered over a
819 0.22 µm syringe filter to remove cell debris. The presence of recombinant phages was
820 verified by PCR with one primer for the inserted TAG sequence and one in the phage
821 genome; 5'-TGGTGATCAGACCGATGGG-3' (forward) and 5'-
822 GAGCTCATAGCAAAGAAGACGTC-3' (reverse).

823 Enrichment and PCR screening for recombinant cyanophages was carried out in 96-
824 well plates containing the *Synechococcus* host[14]. Wells containing recombinant
825 cyanophages were filtered and plated on semi-solid medium. Single phages were plaque
826 purified twice and the whole genome of each purified mutant phage clone was sequenced
827 alongside the corresponding WT. Illumina MiSeq trimmed read pairs were received and
828 the genomes were assembled *de novo* with Spades v. 3.14.1[77]. Both Δ*nblA* mutants had
829 additional mutations in their genomes compared to their corresponding wild-type
830 genomes (see Suppl. Data File 7 for the genome sequences). Non-synonymous mutations
831 in the first mutant relative to the first WT had a T>C substitution at position 11,142
832 (position according to reference genome MH540083.1) leading to a V208A mutation in the
833 gene for DNA polymerase nucleotidyl transferase subunit (accession AXF42115.1) and an
834 A>C substitution at position 37,714 in gene 42 coding for a putative 2OG-Fe(II) oxygenase
835 (accession AXF42102.1) leading to a K34T change. The second Δ*nblA* mutant had two non-
836 synonymous differences compared to its wild type in gene 40 coding for a putative tail
837 fiber protein (accession AXF42100.1): an A>C substitution at position 36,856 in the wild-
838 type genome leading to a D454A change (relative to the reference genome) and an A>G
839 substitution at 36,888 in the Δ*nblA* genome leading to a N465D change. Since both Δ*nblA*
840 mutants had additional mutations compared to the wild-type, infection and spectral
841 experiments were performed with both mutants and gave the same results. Results with
842 the second mutant are shown. Rescue and proteomics experiments were performed with

843 the second mutant. The mutant does not express the putative nucleotide kinase. The
844 relevance of this gene to the phenotype was addressed in the rescue experiments through
845 ectopic expression of *nblA* but not the putative kinase (see below section).

846

847 **Ectopic expression of the cyanophage *nblA***

848 A theophylline translational induction system was used for expression of the cyanophage
849 NblA protein in *Synechococcus* sp. strain WH8109[32,33]. A riboswitch sequence that
850 theophylline binds to and the full length S-TIP37 *nblA* gene were cloned into the pRL-
851 proCAT replicative plasmid downstream of the *rnpB* promoter following Tahan et al[33]. and
852 was conjugated into *Synechococcus* sp. strain WH8109[14,76]. The construct consists of the
853 regions from 12,467 to 12,703 nt relative to the reference S-TIP37 genome (GenBank
854 accession MH540083.1) and six additional amino acids that were added at the NblA C-
855 terminus (Gly, Ser, Tyr, Ser, Val, Thr) due to the cloning process. Since the *nblA* gene
856 overlaps two neighboring genes, the insert contains the last 95 nt of the 3' region of the
857 putative exonuclease gene and the first 37 nt of the 5' region of the putative nucleotide
858 kinase gene. The latter corresponds to 12 out of 94 amino acids of the protein and
859 proteomic analysis verified that this protein was not present in the induction experiments.
860 As such, rescue experiments reinstated expression of NblA but not the putative nucleotide
861 kinase.

862 Expression experiments were carried out with the exponentially growing WH8109
863 conjugant. Theophylline (0.3 mM), dissolved in ASW, was added to induce translation of
864 the NblA protein. Expression of proteins under this theophylline-inducible system is leaky,
865 with differences between induced and non-induced expression detected from 7.5 h after
866 addition of theophylline that increased until 48 h post-induction[33]. For rescue experiments
867 theophylline was added 8 h prior to infection with cyanophages. For assessing the effect
868 of NblA on *Synechococcus* sp. strain WH8109 protein cleavage, samples were collected 24
869 and 48 h post-induction. Induced samples were compared to non-induced samples. Each

870 expression experiment was accompanied by a control construct, carrying the pRL-proCAT
871 plasmid containing the riboswitch but lacking a downstream gene.

872

873 **Proteomic analyses**

874 ***Protein extraction:*** Cyanobacteria were harvested and washed three times using 50 mM
875 of HEPES, pH 7.5. After the final centrifugation the cyanobacterial pellet was resuspended
876 with 8 M guanidine hydrochloride (GuHCl), 100 mM HEPES, pH 7.5, and was heated at 95
877 °C for 10 min. Following this, the samples were sonicated using VialTweeter (Hielscher) at
878 maximum amplitude, 70% cycle time for 5 min, then heated for another 5 min at 95 °C to
879 ensure maximal extraction and denaturation. Cellular debris were pelleted at 18,000*g* for
880 10 min, and the clear supernatant containing proteins was transferred to a new tube.
881 Protein concentrations of each sample were measured with BCA, prior to splitting each
882 sample taking 10 μg for total proteome analysis and 50 μg for N-terminome analysis.

883

884 ***Total proteomics sample preparation:*** Proteins were reduced with 5 mM DL-dithiothreitol
885 (DTT) at 65 °C for 30 min then cooled to room temperature (RT) and alkylated with 12.5
886 mM chloroacetamide (CAA) for 30 min in the dark. Guanidinium concentration was diluted
887 to 1 M using 100 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8,
888 and trypsin (Promega, V115A) was added at 1:100 (w/w) ratio for overnight at 37 °C. Next,
889 trypsin was quenched with 1% formic acid and the samples were desalted using C18 stage
890 tips (Empore 66883-U) before subjecting them to mass spectrometry (MS) analysis.

891

892 ***N-termini proteomics sample preparation:*** N-terminal enrichment was done using an
893 optimized Hydrophobic tagging-assisted N-termini enrichment (HYTANE) procedure as
894 described[36]. Proteins were reduced and alkylated with DTT and CAA as described above.
895 Then, different isotopes (heavy and light) of formaldehyde were used to label primary
896 amines in every two comparative samples. The labeling was done with 40 mM

897 formaldehyde and 20 mM of sodium cyanoborohydride at 37 °C overnight. Formaldehyde
898 leftovers were quenched using 100 mM glycine for 1 h at 37 °C before mixing the heavy
899 and light labeled samples together. The samples were diluted to reduce GuHCl
900 concentration to 1 M using 100 mM of HEPES pH 8 before adding trypsin at 1:100 (w/w)
901 and overnight incubation at 37 °C. The trypsin was then quenched with 1% (final) of formic
902 acid, and the samples were subjected to desalting using OASIS-HLB columns (Waters). The
903 elution was done using 60% of acetonitrile, 0.1% formic acid and the samples were dried
904 using a speed-vac. The peptides were resuspended using 100 µl of 100 mM HEPES pH 7
905 and had undecanal solution (20 mg/ml prepared in ethanol) added to them at w/w ratio
906 of 1:50 (protein:undecanal), followed by the addition of sodium cyanoborohydride at a
907 final concentration of 20 mM. Undecanal tagging was done for 2 hours at 50 °C while
908 renewing the sodium cyanoborohydride after 1 h. The samples were then centrifuged at
909 16,000*g* for 5 min and the supernatant was transferred to a new tube before drying in a
910 speed vac. The dried samples were resuspended in 500 µl of 2% acetonitrile, 0.1% formic
911 acid and subjected again to OASIS-HLB column. Elution was done using 60% of acetonitrile,
912 0.1% formic acid and the samples were dried in a speed-vac before resuspending again
913 and subjecting them to MS.
914
915 ***LC-MS analysis:*** Desalted samples were subjected to LC-MS analysis using an Orbitrap
916 Exploris 480 coupled with an EvoSep One HPLC. Samples were introduced onto the EvoTip,
917 which was then washed twice with 20 µl of 0.1% formic acid. The washed peptides
918 remained wet, maintained by applying 150 µl of 0.1% formic acid atop of the EvoTip until
919 MS analysis. The samples loaded on the EvoTips underwent chromatographic separation
920 on a 15 cm × 150 µm analytical column, filled with 1.9 µm C18 beads (EV1106). Peptides
921 were separated over an 88-minute gradient according to the manufacturer's standard
922 method. Full MS scans acquired in positive ion mode, scanning from 300 to 1800 m/z, were
923 recorded at a resolution of 120,000, by a data-dependent mode selecting the top 20 ions
924 with high energy collisional dissociation (HCD) fragmentation ion at a resolution of 17,500
925 and with dynamic exclusion enabled.

926

927 ***MS data analysis:*** For total proteomics, data analysis was conducted using FragPipe v. 22.0
928 (https://fragpipe.nesvilab.org/) using DDA+ mode[78] with the default Label-Free
929 Quantification Match Between Runs (LFQ-MBR) settings. The search enzyme was set to
930 trypsin, allowing up to two missed cleavages. Variable modifications included methionine
931 oxidation and protein N-terminal acetylation, while carbamidomethylation of cysteine was
932 set as a fixed modification. For N-terminomics data, raw files were first converted to MzML
933 format and then analyzed using the Trans-Proteomic Pipeline version 6.3[79] with the Comet
934 search engine v. 2023_01 rev2. The search enzyme was set to Semi-ArgC, allowing up to
935 two missed cleavages. Variable modifications included methionine oxidation and the mass
936 difference between heavy and light dimethylation at lysine residues or the N-termini of
937 peptides. Fixed modifications included carbamidomethylation of cysteine and light
938 dimethylation at lysine residues or peptides' N-termini.

939 Relative peptide quantifications were performed using XPRESS, with parameters
940 set to mass tolerance of 20 ppm, a minimum of three chromatogram points for
941 quantitation, and the number of isotopic peaks to sum set to zero. Post-analysis table
942 creation, cleavage motif extraction, and ratio normalization were performed using an in-
943 house script[80]. The search database included all *Synechococcus* sp. strain WH8109 proteins
944 in Uniprot (Taxon ID: 166314, both reviewed and unreviewed) in addition to the vector
945 added genes and standard contaminant proteins. For viral infection samples, this database
946 was expanded to also include all *Synechococcus* T7-like phage S-TIP37 (Taxon ID:1332145)
947 protein sequences from Uniprot and the NblA sequence. In the analysis of the N-
948 terminomics data, we considered three classes of blocked N-terminal peptides: ORF N-
949 termini, neo-N-termini generated by internal proteolysis (SemiN1), and proteolysis-
950 adjacent peptides truncated before the first arginine (SemiN2). Only NblA-induced
951 cleavage events that were identified in both biological replicates and showed >2-fold
952 abundance increases in NblA-expressing cells compared to the control in each replicate
953 were considered for downstream analysis. To bin host proteins into functional groups, we
954 utilized the COG functional category assignments for *Synechococcus* sp. strain WH8109

955 proteins in the eggNOG database v. 5.0[81]. The proteins were assigned to one of the broad
956 functions based on a modified classification of COG functional groups as follows:
957 phycobilisome components — Phycobilisome; other photosynthesis-related genes and
958 other proteins assigned to COG functional category C (energy production and conversion)
959 — Energy; COG functional categories J, A, K, L and B — Information Storage and Processing;
960 COG functional categories D, Y, V, T, M, N, Z, W, U and O — Cellular Processes and
961 Signaling; COG functional categories G, E, F, H, I, P and Q — Metabolism; COG functional
962 categories R, S and proteins assigned to multiple categories — Other.

963 Phage proteins were classified into functional categories based on the functions of
964 the majority of genes in three genomic and expression clusters[29,44]. Statistical analysis of
965 the dynamics of abundance of the different functional groups of phage proteins in the total
966 proteomics data as a function of time post-infection onset was performed as follows. For
967 each gene cluster cumulative MS LFQ intensities were used as a response variable in a
968 linear model with time post-infection and infection type (by WT or ΔnblA mutant) and their
969 interaction as predictor terms. Distributions of the residuals were checked with QQ-plots.
970 Detailed results of the statistical analysis are provided in Suppl. Data File 2.

971

972 **Bioinformatic analyses**

973 ***Analysis of T7-like cyanophage genomes*:** Genomes of cultured T7-like cyanophages and
974 outgroup phages (coliphage T7 and pelagiphages HTVC011P and HTVC019P) were
975 collected from GenBank. Complete genomes and genomic fragments of environmental T7-
976 like cyanophages and cyanoprophages were obtained from JGI IMG/VR v. 4.1[82] and Global
977 Ocean Virome v. 2.0 assemblies[54]. To collect genomic fragments containing exonuclease
978 genes with sufficiently long downstream regions, a conserved region in the C-terminal part
979 of the exonuclease (corresponding to residues 86-221 in the exonuclease protein from S-
980 TIP37 [RefSeq accession YP_009807515.1]) was chosen and homologous regions from a
981 reference set of cyanophages were used as queries for tblastn from NCBI blast v. 2.15.0[83].
982 Fragments were retained when they had matches at least 400 bp long, identity of ≥40%,

983 bit score of ≥400 and at least 500 bp of downstream sequence available. To collect
984 genomes for phylogenetic analysis, out of the contigs with the exonuclease gene, we
985 extracted sequences ≥30 Kbp long and dereplicated them with dRep v. 3.4.5[84] at the ANImf
986 identity level of 0.95. Genes were predicted with prodigal v. 2.6.3[85] and phylogeny was
987 reconstructed with phylophlan v. 3.0.2[86] (diamond v. 2.1.8[87], MAFFT v. 7.475[88], trimAl v.
988 1.4.1[89], IQ-TREE v. 2.1.2[90]) based on concatenated alignments of nine core genes: primase-
989 helicase, exonuclease, portal protein (head-to-tail adaptor), head assembly protein, major
990 capsid protein, tail tubular proteins A and B, small and large terminase subunits. The genes
991 were extracted by searching with diamond blastp using representative protein sequences
992 as queries from two distantly related cyanophages, P60 and S-SRP02: P60_gp14 (primase-
993 helicase is missing from S-SRP02), P60_gp18 and SSRP02_p034, P60_gp26 and
994 SSRP02_p038, P60_gp27 and SSRP02_p039, P60_gp28 and SSRP02_p040, P60_gp29 and
995 SSRP02_p041, P60_gp30 and SSRP02_p042, P60_gp40 and SSRP02_p011, P60_gp49 and
996 SSRP02_p012, respectively. IQ-TREE was run with "-m LG --alrt 1000 -pers 0.2 -nstop 500"
997 with 1,000 ultrafast bootstrap replicates[91] and the resulting tree was outgroup-rooted with
998 the ingroup defined as the branch encompassing the known cyanophages. The
999 cyanophage genomes were classified based on the resulting phylogeny into six clades: the
1000 previously recognized clades A, B and C, the newly defined clades R (for P-RSP2-like
1001 phages), S (S-SRP02-like phages) and T (represented by GOV contig
1002 Station52_DCM_ALL_NODE_78). See Suppl. Data File 4 for metadata of the genomes
1003 chosen to represent each of the clades. Two outlier genomes were found to cause spurious
1004 clustering and were excluded from the phylogenetic analysis: IMGVR_UViG
1005 3300032116_000204 and IMGVR_UViG 3300029337_000223. The incidence of *nblA* and
1006 *psbA* genes in the genomes was assessed by searching ORFs (between stop codons) with
1007 hmmsearch from HMMER v. 3.4[92] using a custom NblA profile (see below) and the PsbA
1008 profile TIGR01151.1 from NCBI Protein Family Models. The incidence and type of DNA
1009 polymerase in the genomes was assessed by searching the ORFs with blastp using the
1010 sequences of the exonuclease (YP_009807513.1) and polymerase (YP_009807514.1)
1011 subunits of the split polymerase of S-TIP37 with an E-value threshold of 1e-10. A version

1012 of the tree focusing on representative genomes as shown in the main text, was obtained
1013 by trimming the full tree. The trees were visualized with ggtree v. 3.2.0[93]. The full
1014 phylogenetic tree without the outgroups is available in Suppl. Data File 5.

1015

1016 ***Picocyanobacterial genomes:*** Metadata and genome sequences for picocyanobacteria
1017 were downloaded from Cyanorak v. 2[94] and additional *Prochlorococcus* genomes [20,25] were
1018 downloaded from Integrated Microbial Genomes & Microbiomes[95]. Genes were predicted
1019 with prodigal and GeneMarkS-2 v. 1.14_1.25[96] with the predictions merged using
1020 gffcompare v. 0.12.6[97].

1021

1022 ***Analysis of NblAs:*** The NblA HMM profile used to search for *nblA* genes in T7-like
1023 cyanophage genomes was obtained from an alignment of the previously released NblA
1024 sequences[6] and manually curated NblA sequences from representative T7-like
1025 cyanophages. The searches for the NblA hits were performed without heuristic filters and
1026 the results were filtered using an empirically determined full score threshold of 20. The
1027 same NblA profile was also initially used to search for *nblA* genes in the protein sequences
1028 from picocyanobacteria. Upon discovering that many of their NblAs yield sub-significant
1029 hits, we built a separate NblA profile specifically targeting picocyanobacterial NblAs based
1030 on the sequences collected in the first round of the search. Notice that due to the high
1031 divergence of the NblAs from picocyanobacteria, many of them avoided detection in our
1032 previous screen of cyanobacterial genomes using a general NblA protein profile[6]. The two
1033 NblA protein profiles are available in Suppl. Data File 10.

1034 A NeighborNet network was constructed to visualize similarity relationships
1035 between NblA proteins of different origins. NblAs from T7-like cyanophages and
1036 picocyanobacteria were combined with cyanobacterial sequences assigned in UniProt r.
1037 2025_2[98] to Pfam profile PF04485, as well as selected NblA sequences from non-T7-like
1038 cyanophages. To reduce redundancy, the sequences were clustered at 90% identity level
1039 with CD-HIT v. 4.8.1[99]. The cluster representatives were aligned with MAFFT in automated

1040 mode, and trimAl was used to trim the alignment (strict mode, minimum column block size
1041 of 15). The resulting trimmed alignment was used as input to SplitsTree v. 4.17.0[100] which
1042 generated a NeighborNet network based on uncorrected distances. The network was
1043 visualized with tanggle v. 1.8.0[101].

1044 Sequence logos were generated for different groups of NblAs using DiffLogo v.
1045 2.26.0[102]. The original MAFFT alignment was trimmed in a similar way as above but only
1046 regions outside of the core NblA domain were trimmed. The fasta file was converted to
1047 A3M format (matching columns with <50% gaps) and secondary structure predictions were
1048 added with tools from HH-suite v. 3.3.0[66]. For the sequence logos, the remaining gaps were
1049 coded as a separate state X.

1050 Representative structures of NblA dimers were predicted with ColabFold v. 1.5.5
1051 (based on AlphaFold2)[67].

1052

1053 ***Analysis of global distribution of the cyanophage* nblA*s:*** For the analysis of the
1054 distribution of *nblA* downstream of the exonuclease gene, we took all of the initially
1055 collected environmental contigs with T7-cyanophage type exonucleases and extracted
1056 regions containing the 3' end of the exonuclease gene with up to 100 bp upstream and up
1057 to 1,000 bp downstream. These fragments were searched against exonuclease sequences
1058 from the genomes used for phylogeny using blastx and clades were assigned to fragments
1059 with best hits to cyanophages at identity level of ≥60%. The fragments coming from the
1060 GOV2.0 assemblies were used for recruitment of GOV raw data with bwa v. 0.7.17[103].
1061 Quantification was performed with featureCounts from subread v. 2.0.6[104]. As the
1062 exonuclease and *nblA* genes frequently overlap, we selected strictly defined windows for
1063 read quantification instead of ORF boundaries: the exonuclease was represented by the
1064 above-mentioned conservative region at the 3' end of the gene. For the *nblA* gene we
1065 chose the location corresponding to the core NblA region based on the hmmsearch
1066 matches. Genes assigned to the same cyanophage clade were grouped into metafeatures
1067 for quantification and the resulting per-station mapped read pair counts were converted

1068 to Fragments Per Kilobase per Million reads (FPKM) values by dividing them by the median
1069 feature length (in Kbp) and the total number of the reads per sample (in millions).
1070 Sequences used for recruitment and quantification results in tabular format are available
1071 in Suppl. Data File 9.

1072 The bioinformatic workflow was implemented in snakemake[105], python and R.

1073

1074 **Calculations of the ocean-wide impact of viral *nblA* genes on light harvesting by**
1075 **picocyanobacteria**

1076 In order to estimate the ocean-wide impact of viral *nblA* genes on light harvesting by
1077 picocyanobacteria, we used the following data: A 100% reduction in photosynthetic
1078 performance for 50% of the time attributable to the cyanophage *nblA* gene (Fig. 1e); 1-
1079 15% of marine cyanobacterial cells are infected by T7-like cyanophages at a given
1080 time[39,56,57]; and 35% (surface waters) to 65% (DCM zone) of marine T7-like cyanophages
1081 code for the *nblA* gene (Fig. 5)**.** The impact of viral NblA proteins on marine cyanobacterial
1082 photosynthesis in surface water was calculated as follows: 50% reduction in
1083 photosynthetic light harvesting performance, with 1-6% of cells infected and 35% of T7-
1084 like cyanophages carrying *nblA* genes giving an estimated impact of 0.175-1.05% on the
1085 photosynthetic performance of the picocyanobacteria. The impact on marine
1086 cyanobacterial photosynthesis at the DCM was estimated as follows: 50% reduction in
1087 photosynthetic light harvesting performance, with 5-15% of cells infected and 65% of T7-
1088 like cyanophages carrying *nblA* genes giving an estimated impact of 1.625-4.875% on the
1089 photosynthetic performance of picocyanobacteria.

1090

1091 **References for Methods**

1092 63. Wyman, M., Gregory, R. P. F. & Carr, N. G. Novel role for phycoerythrin in a marine
1093 cyanobacterium, *Synechococcus* strain DC2. *Science* **230**, 818–820 (1985).

1094 64. Lindell, D., Padan, E. & Post, A. F. Regulation of *ntcA* expression and nitrite uptake in
1095 the marine *Synechococcus* sp. strain WH 7803. *J. Bacteriol.* **180**, 1878–1886 (1998).

1096 65. Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I. & Zinser, E. R. Facilitation of
1097 robust growth of *Prochlorococcus* colonies and dilute liquid cultures by "helper"
1098 heterotrophic bacteria. *Appl. Environ. Microbiol.* **74**, 4530–4534 (2008).

1099 66. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein
1100 annotation. *BMC Bioinformatics* **20**, 473 (2019).

1101 67. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods*
1102 **19**, 679–682 (2022).

1103 68. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat.*
1104 *Biotechnol.* **42**, 243–246 (2024).

1105 69. Kirzner, S., Barak, E. & Lindell, D. Variability in progeny production and virulence of
1106 cyanophages determined at the single-cell level. *Environ. Microbiol. Rep.* **8**, 605–613
1107 (2016).

1108 70. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models
1109 using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).

1110 71. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, Thousand Oaks
1111 CA, 2019).

1112 72. Larom, S., Salama, F., Schuster, G. & Adir, N. Engineering of an alternative electron
1113 transfer path in photosystem II. *Proc. Natl. Acad. Sci.* **107**, 9650–9655 (2010).

1114 73. Schoffman, H. & Keren, N. Function of the IsiA pigment–protein complex in vivo.
1115 *Photosynth. Res.* **141**, 343–353 (2019).

1116 74. Liland, K. H., Almøy, T. & Mevik, B.-H. Optimal choice of baseline correction for
1117 multivariate calibration of spectra. *Appl. Spectrosc.* **64**, 1007–1016 (2010).

1118 75. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein
1119 homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248
1120 (2005).

1121 76. Brahamsha, B. A genetic manipulation system for oceanic cyanobacteria of the genus
1122 *Synechococcus*. *Appl. Environ. Microbiol.* **62**, 1747–1751 (1996).

1123 77. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to
1124 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

1125 78. Yu, F., Deng, Y. & Nesvizhskii, A. I. MSFragger-DDA+ enhances peptide identification
1126 sensitivity with full isolation window search. *Nat. Commun.* **16**, 3329 (2025).

1127 79. Deutsch, E. W. *et al.* Trans-Proteomic Pipeline: robust mass spectrometry-based
1128 proteomics data analysis suite. *J. Proteome Res.* **22**, 615–624 (2023).

1129 80. Hanna, R., Rozenberg, A., Lavy, T. & Kleifeld, O. Increasing the coverage of the N-
1130 terminome with LysN amino terminal enrichment (LATE). *Methods Enzymol.* **686**, 1–
1131 28 (2023).

1132 81. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically

1133 annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic*
1134 *Acids Res.* **47**, D309–D314 (2019).

1135 82. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus
1136 genomes within a framework of extensive functional, taxonomic, and ecological
1137 metadata. *Nucleic Acids Res.* **51**, D733–D743 (2023).

1138 83. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421
1139 (2009).

1140 84. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate
1141 genomic comparisons that enables improved genome recovery from metagenomes
1142 through de-replication. *ISME J.* **11**, 2864–2868 (2017).

1143 85. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1144 identification. *BMC Bioinformatics* **11**, 119 (2010).

1145 86. Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from
1146 metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).

1147 87. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life
1148 scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

1149 88. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
1150 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

1151 89. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
1152 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–
1153 1973 (2009).

1154 90. Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic
1155 inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

1156 91. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2:
1157 improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).

1158 92. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195
1159 (2011).

1160 93. Xu, S. *et al.* Ggtree: A serialized data object for visualization of a phylogenetic tree
1161 and annotation data. *iMeta* **1**, e56 (2022).

1162 94. Garczarek, L. *et al.* Cyanorak v2.1: a scalable information system dedicated to the
1163 visualization and expert curation of marine and brackish picocyanobacteria genomes.
1164 *Nucleic Acids Res.* **49**, D667–D676 (2021).

1165 95. Chen, I.-M. A. *et al.* IMG/M v.5.0: an integrated data management and comparative
1166 analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **47**,
1167 D666–D677 (2019).

1168 96. Lomsadze, A., Gemayel, K., Tang, S. & Borodovsky, M. Modeling leaderless
1169 transcription and atypical genes results in more accurate gene prediction in
1170 prokaryotes. *Genome Res.* **28**, 1079–1089 (2018).

1171 97. Pertea, G. & Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **9**,
1172 ISCB Comm J-304 (2020).

1173 98. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025.
1174 *Nucleic Acids Res.* **53**, D609–D617 (2025).

1175 99. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of
1176 protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

1177 100. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary
1178 studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).

1179 101. Schliep, K. *et al. Tanggle: Visualization of Phylogenetic Networks*. (2023).
1180 doi:10.18129/B9.bioc.tanggle.

1181 102. Nettling, M. *et al.* DiffLogo: a comparative visualization of sequence motifs. *BMC*
1182 *Bioinformatics* **16**, 387 (2015).

1183 103. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
1184 MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).

1185 104. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose
1186 program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–
1187 930 (2014).

1188 105. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Research* **10**, 33
1189 (2021).

1190 106. Sabehi, G. & Lindell, D. The P-SSP7 cyanophage has a linear genome with direct
1191 terminal repeats. *PLOS ONE* **7**, e36710 (2012).

1192

1193 **Acknowledgments.** We thank F. Salama for biophysical guidance, M. Suissa-
1194 Szlejf for help in setting up the phycobilisome isolation gradients, R. Edrei for help
1195 with spectroscopic measurements, G. Sabehi for discussions on infection
1196 experiments, S. Goldin for help with qPCR calculations, S. Larom for technical
1197 support, N. Keren and C. Kranzler for discussions on photosynthesis experiments,
1198 N. Keren for help in performing spectroscopy, and N. Frankenberg-Dinkel for
1199 comments on the manuscript. This work was funded by the European Research
1200 Council (ERC Advanced Grant 321647 to O.B. and ERC Consolidator Grant
1201 646868 to D.L.), the Nancy and Stephen Grand Technion Energy Program (GTEP),
1202 Israel Science Foundation grants to O.B. (143/18) and O.K. (1515/23), and the
1203 Simons Foundation Life Science Award 735081 to D.L. O.B. holds the Louis and

Lyra Richmond Chair in Life Sciences and D.L. holds the Dresner Chair in Life Sciences and Medicine.

Author contributions. O.N. conceived the project, performed viral work, phycobilisome isolation and experimentation, and together with O.B., O.K. and D.L. designed the experiments. I.P. performed phage growth curve experiments. A.R. performed bioinformatic analyses. R.H. and O.K. performed proteomic analyses. D.S., R.T. and D.L. developed the methods for constructing the mutant phage and the cyanobacterial NblA expression strain. O.B. coordinated the project. O.N., A.R., O.B., O.K., and D.L. wrote the paper, which was critically revised and approved by all authors.

Competing interests. The authors declare that they have no conflicts of interest.

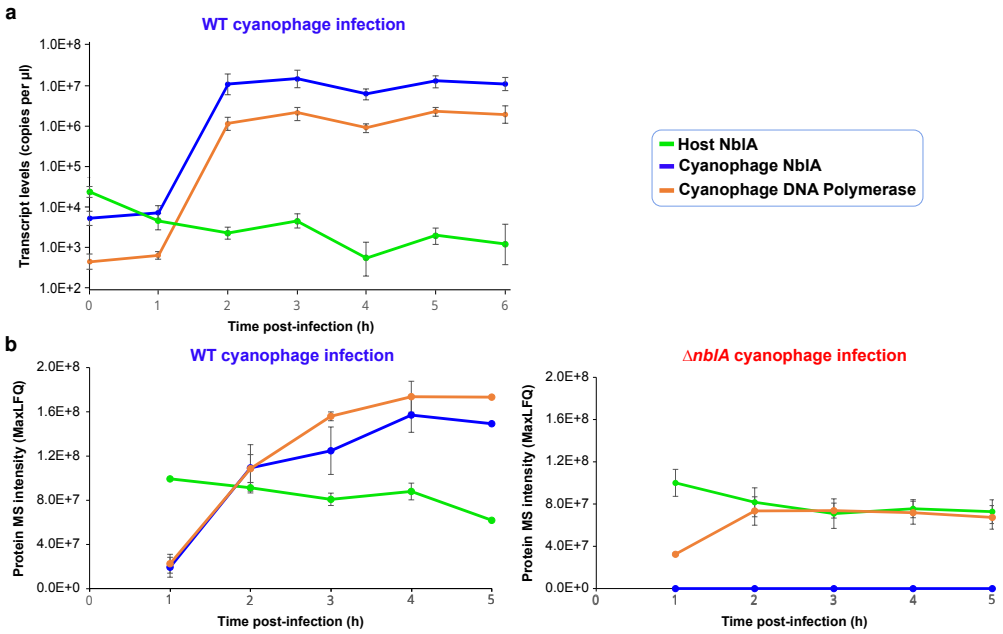
Data availability

All data are available in the main text or the Supplementary Data. Intermediate results of the bioinformatic analyses are available from the GitHub repository https://github.com/BejaLab/autographiviridae. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the following dataset identifiers: PXD057452 (whole proteome data) and PXD057454 (N-terminomics data) [the reviewers can receive access using the following tokens: ZqvdIvO680Np (project PXD057452) and NpEFI3ZdaRSA (project PXD057454)].

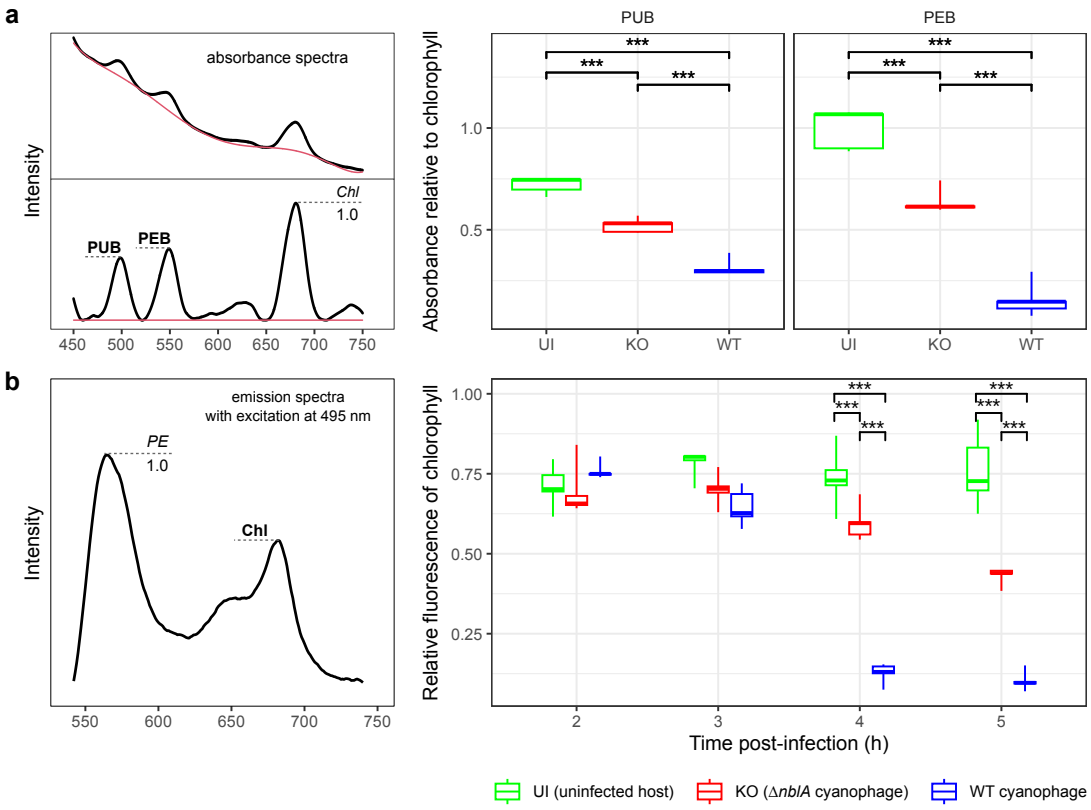
Code availability

The code used for the bioinformatic analyses is available from the GitHub repository https://github.com/BejaLab/autographiviridae.

Extended data figures

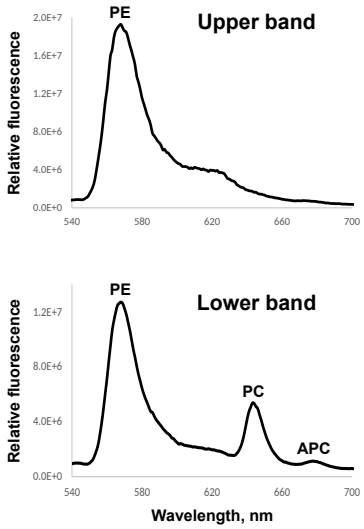


Extended Data Fig. 1. Expression of the cyanophage *nblA* gene during S-TIP37 infection of *Synechococcus* sp. strain WH8109. a, Transcript levels of S-TIP37 cyanophage *nblA* and DNA polymerase genes and one of the host *nblA* genes during infection with WT cyanophage determined by qRT-PCR. Transcript levels are normalized to the *rnpB* gene. Among the host *nblA* genes, *nblA2* was chosen as the gene coding for the NblA protein with the highest structural similarity to previously characterized NblAs (see Methods). **b**, Protein mass spectrometry intensity levels of the S-TIP37 cyanophage NblA, DNA polymerase, and host NblA6 proteins during infection with WT (left) or Δ*nblA* mutant (right) cyanophages, determined by LC-MS/MS analysis. NblA6 was the only host NblA protein detected. Data are presented as mean ± standard deviation of three independent biological replicates, *n*=3.

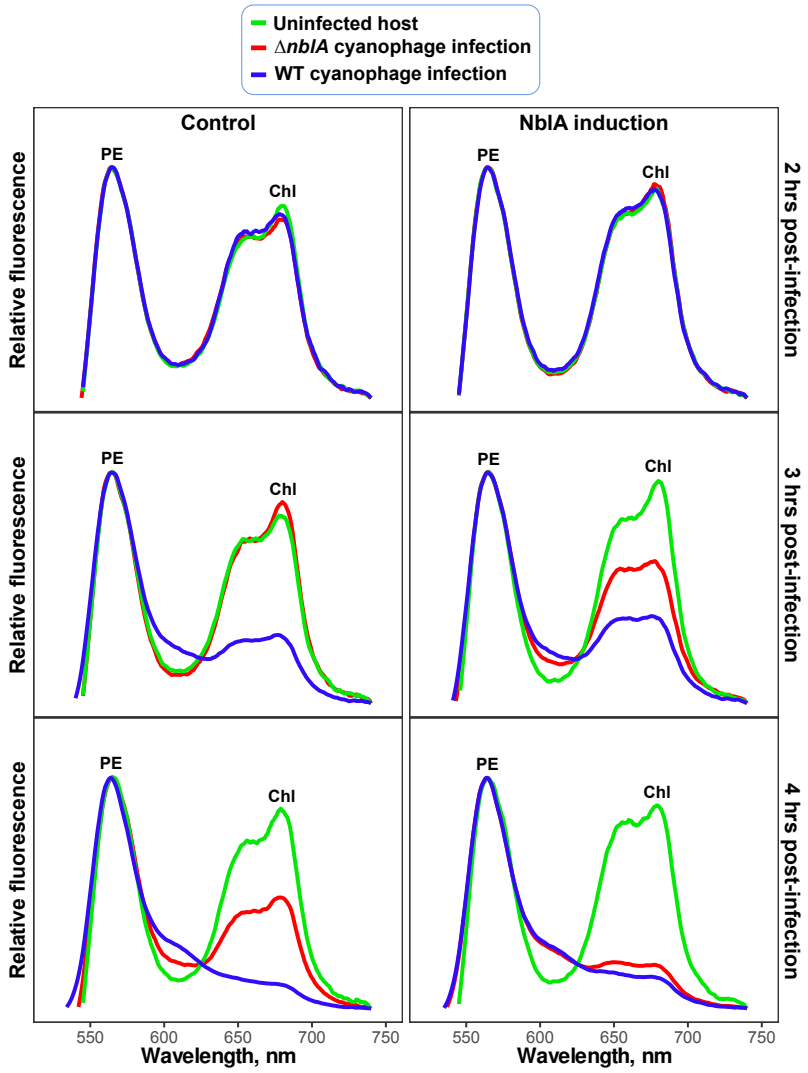


Extended Data Fig. 2. Spectroscopic characteristics of cells infected with wild-type and Δ*nblA* mutant S-TIP37 cyanophages. a, Relative absorbance of PUB and PEB at 12 h post-infection. Relative absorbance was determined as follows (see the panel to the left): the absorbance spectra were background-subtracted (red line) and the intensities were scaled to the height of the Chl peak. UI, uninfected, KO, infection with the Δ*nblA* mutant, and WT, infection with the wild-type phage. **b**, Relative chlorophyll fluorescence at different time points during infection. The samples were excited at 495 nm (excitation maximum of PUB), emission spectra were obtained and the relative height of the chlorophyll peak was calculated (see the panel to the left). The PE peak at 560 nm is made up of PEI and PEII, to which the PEB and PUB chromophores bind. In both panels, asterisks denote significant pairs of differences in the Tukey post-hoc test: *** — p-value < 0.001 (after FDR adjustment in panel b). Boxplots reflect: minimum-maximum range (vertical line), interquartile range (rectangle) and median (horizontal line). The results are based on five biological replicates of each condition: uninfected cells and cells infected

with two cyanophages, Δ*nblA* (KO) and wild-type (WT) phages, sampled repeatedly for measurement. For further results of the statistical analyses refer to Suppl. Data File 2. Representative spectra are shown in Fig. 2a,b.



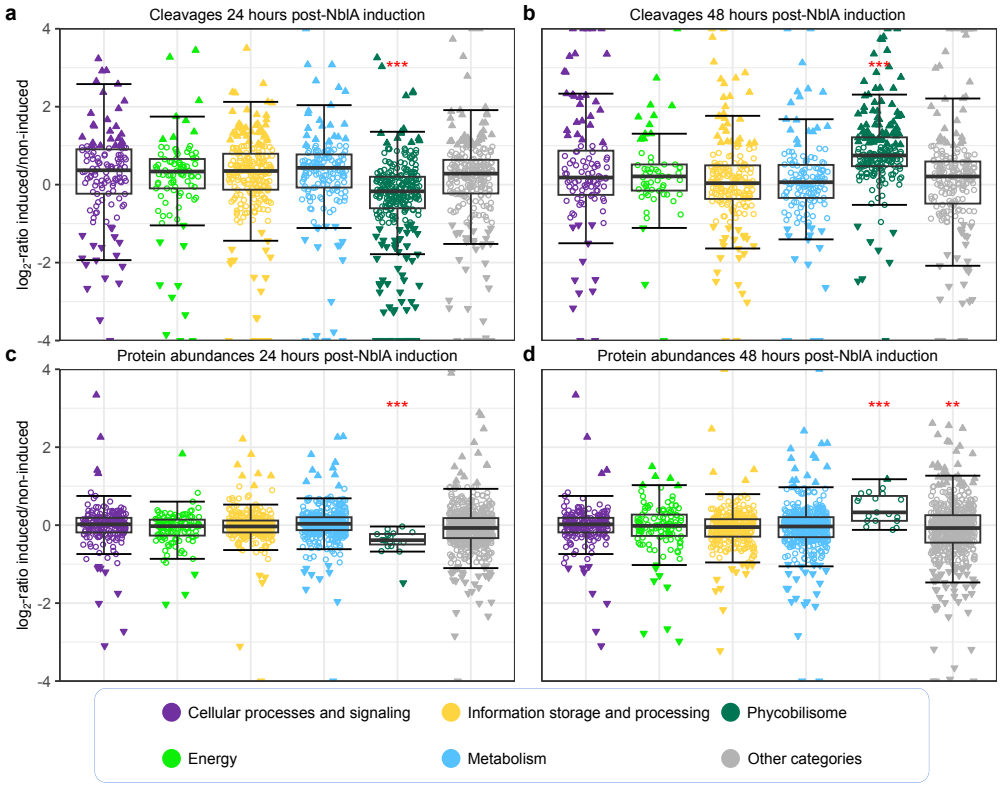
Extended Data Fig. 3. Fluorescence spectral analysis of fractionated phycobilisomes. Fluorescence emission spectra at 77K of fractionated phycobilisomes from uninfected *Synechococcus* sp. strain WH8109. The lower density, upper band (see Fig. 2c) contains disassembled phycobilisome complexes, whereas the higher density, lower band (see Fig. 2c) contains assembled phycobilisome complexes with intact energy transfer from PE to PC and APC.



1275

1276 **Extended Data Fig. 4. Ectopic expression of cyanophage NblA in *Synechococcus***
1277 **sp. strain WH8109 during infection.** Fluorescence emission spectra of uninfected
1278 cells (green), cyanobacteria infected by the WT cyanophage (blue) and by the
1279 Δ*nblA* mutant cyanophage (red) at different time points post-infection. Control
1280 experiments (the panels to the left) were performed without the addition of the
1281 inducer. Ectopic expression was induced in the host 8 h prior to addition of the
1282 phages (NblA induction, the panels to the right). Measurements were taken at
1283 room temperature and normalized using the PE peak.

1284

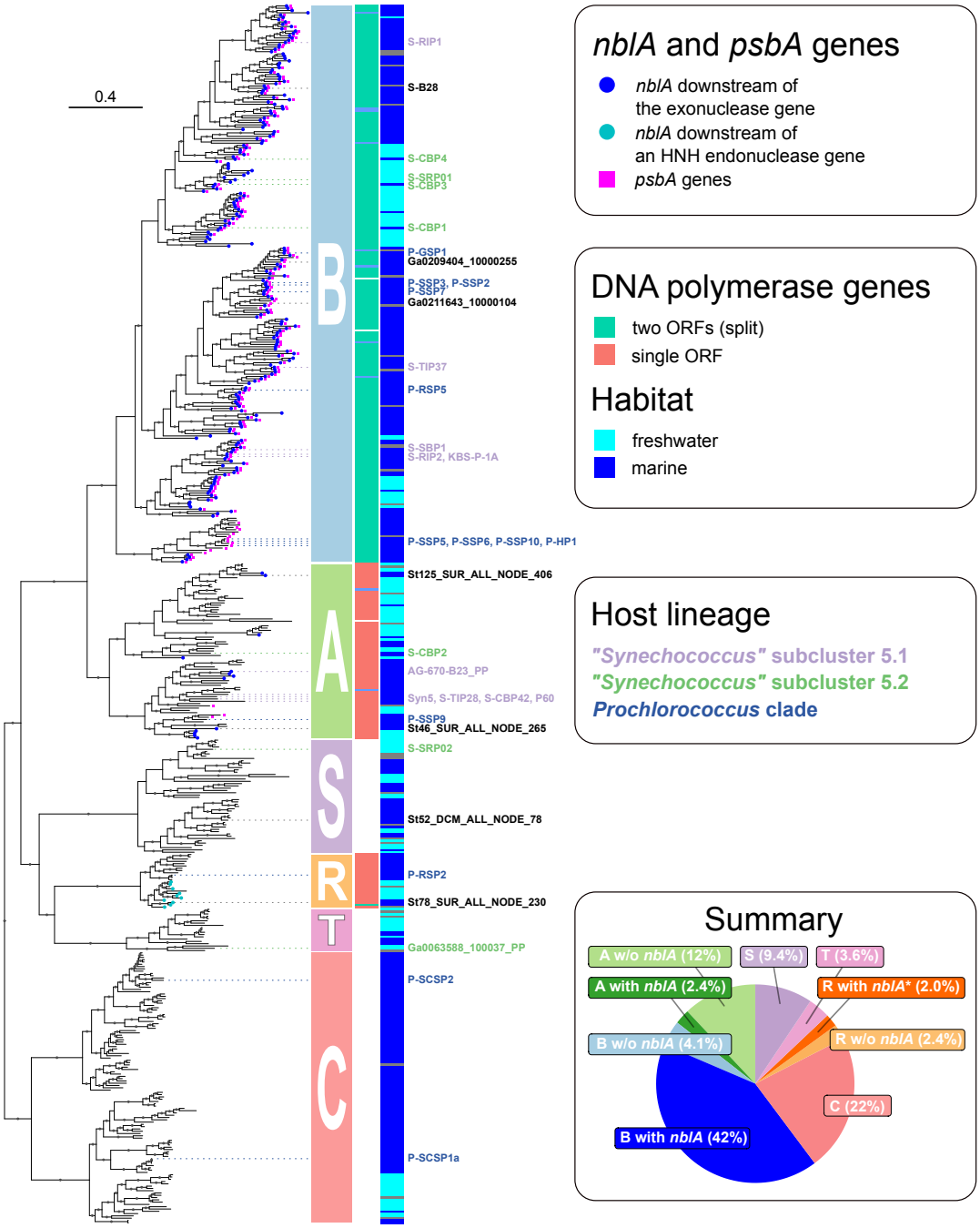


1285

1286 **Extended Data Fig. 5. Cleavage and proteome changes upon ectopic expression**
1287 **of cyanophage NblA in *Synechococcus* sp. strain WH8109.** Proteomic analysis of
1288 cleavage events (**a** and **b**) and total protein abundance changes (**c** and **d**) following
1289 ectopic expression of the cyanophage NblA in *Synechococcus* sp. strain WH8109
1290 cells for 24 hours (**a** and **c**) and 48 hours (**b** and **d**), compared to non-induced
1291 controls (n = 2). The top panels show the abundance ratios of neo-N-terminal
1292 peptides identified by N-terminomics, calculated as the ratio of signal intensities
1293 between induced and non-induced samples, reflecting changes in peptide
1294 abundance.The bottom panels show the abundance ratios for proteins in the total
1295 proteomic analysis based on the intensities of tryptic peptides. Proteins are
1296 categorized according to a modified classification of COG functional groups (see
1297 Methods). Filled triangles indicate neo-N-terminal peptides and proteins with
1298 absolute $\log_2$-ratios ≥1 in both replicates. Triangles with upward-pointing and
1299 downward-pointing apexes represent peptides more and less abundant,
1300 respectively, upon NblA ectopic expression relative to the control. Red asterisks

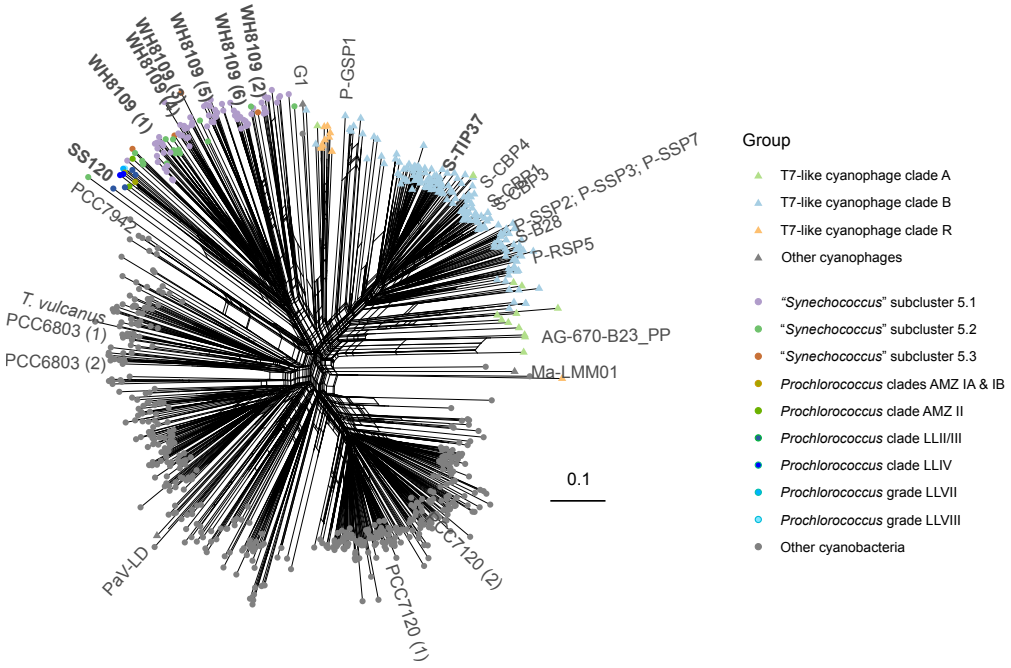
1301 indicate significant deviations of the category-wise mean ratios from zero (two-
1302 sided one-sample t-test) after FDR adjustment across all comparisons: *** —
1303 adjusted p-value ≤ 0.001, ** — adjusted p-value ≤ 0.01.

1304

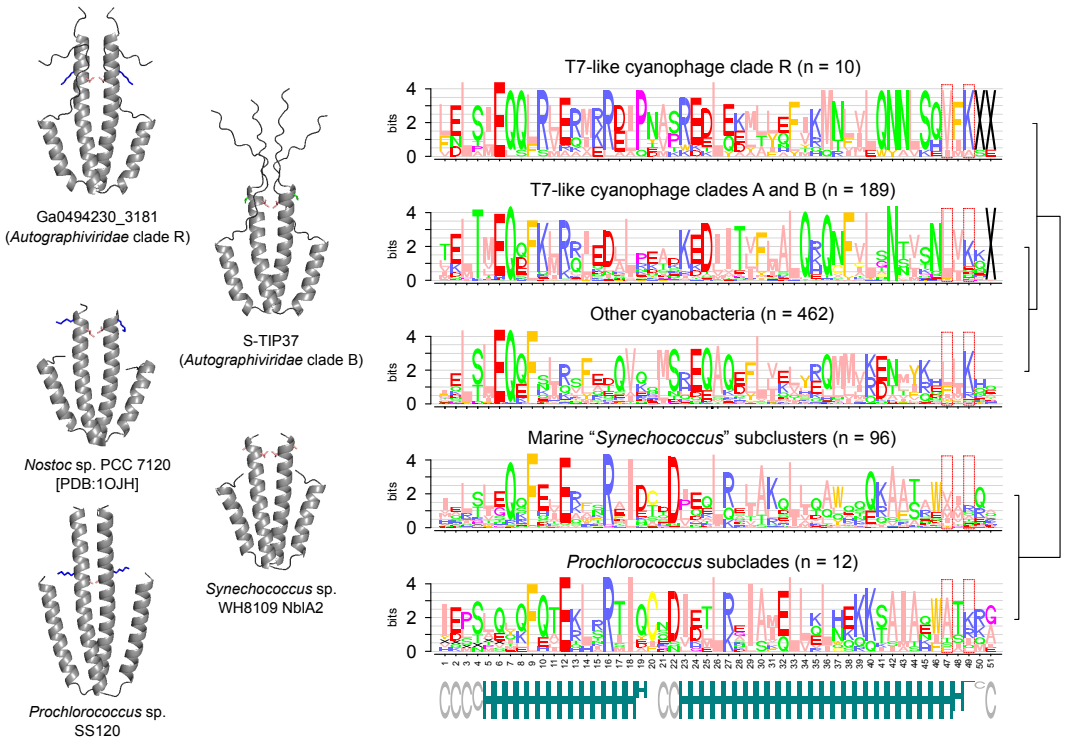


1305

1306 **Extended Data Fig. 6. Phylogeny and distribution of *nblA* genes among T7-like**
1307 **cyanophages.** Phylogenetic tree was built based on concatenation of nine core
1308 genes coding for: primase-helicase, exonuclease, portal protein (head-tail
1309 adaptor), head assembly, major capsid protein, tail tubular proteins A and B, and
1310 small and large terminase subunits. Phage genomes used in the phylogeny were
1311 recruited among cultured T7-like cyanophages and from environmental sources:
1312 Global Ocean Virome v. 2 assemblies and JGI IMG/VR v. 4.1. The environmental
1313 genomes were recruited by searching for the exonuclease gene, selecting
1314 genomes at least 30 kb long, dereplicating them at 0.95 ANImf identity level and
1315 requiring the presence of at least eight core genes. The tree was outgroup-rooted
1316 by inclusion of coliphage T7, pelagiphage HTVC011P, pelagiphage HTVC019P and
1317 environmental phages falling outside of the cyanophage cluster. Phages were
1318 categorized into previously recognized clades A (P60-like), B (S-TIP37-like) and C
1319 (P-SCSP1a/2-like) and newly-delineated clades R (P-RSP2-like), S (S-SRP02-like)
1320 and T (Ga0063588_100037_PP-like, no cultured representatives). Presence and
1321 type of DNA polymerase genes is indicated with the clades dominated by phages
1322 falling in one of the following categories: no DNA polymerase gene (clades C, S, T),
1323 DNA polymerase coded in a single ORF (clades A and R) and a split DNA polymerase
1324 gene[106] with the exonuclease domain coded in a separate ORF (clade B). Clade
1325 representatives are colored with labels according to the host's phylogenetic
1326 affiliation if known. Note that each of the three picocyanobacterial lineages
1327 represents an assemblage of several genus-level taxa. The incidence of *nblA*
1328 (circles) and *psbA* (squares) genes is indicated with symbols at the tips of the
1329 corresponding branches and the *nblA* genes of two categories are differentiated:
1330 those located downstream of the exonuclease gene (as in S-TIP37) and the *nblA*
1331 gene downstream of an HNH endonuclease gene (in a subclade of clade R). No
1332 clade C, S or T genomes were found to code for *nblA* and are thus not separated
1333 into two groups in the summary pie chart. Solid circles mark branches with
1334 ultrafast bootstrap support of ≥95. See Suppl. Data File 7 for the source data.



Extended Data Fig. 7. Phylogenetic network of NblA proteins from the T7-like cyanophages, marine picocyanobacteria and other cyanophages and cyanobacteria. The NblA protein sequences were aligned, trimmed and a NeighborNet network was obtained based on uncorrected distances. The proteins are categorized by the corresponding cyanophage or cyanobacterial group. NblA was found in the low-light adapted *Prochlorococcus* ecotypes LLII/LLIII, LLIV, LLVII, LVIII, AMZIA, AMZIB, and AMZII but not LLI and AMZIII nor any of the high-light-adapted ecotypes. NblAs from selected cyanophages and model cyanobacteria with the NblAs from S-TIP37, *Synechococcus* sp. strain WH8109 and *Prochlorococcus* sp. strain SS120 are highlighted in bold. The network and trimmed sequences of cluster representatives in NEXUS format are available in Suppl. Data File 8 and metadata and untrimmed sequences and aligned sequences of the core NblA region are available in Suppl. Data File 11.



Extended Data Fig. 8. Structures and protein profiles of NblA proteins from T7-like cyanophages and cyanobacteria. Right: dimeric structures of selected representatives of the NblA proteins. All of the structures, except that of the NblA from *Nostoc* sp. strain PCC 7120 [PDB:1OJH[10]] were obtained with ColabFold (AlphaFold2)[67]. Left: sequence logos generated based on a trimmed alignment of clustered NblA sequences. Numbers of cluster representatives per group are provided in parentheses. Secondary structure predictions based on the whole alignment are shown at the bottom (H: helix, C: coil) with the height proportional to the prediction score. Positions known to mediate phycobiliprotein binding[38] are indicated: protein profile positions 48 (residues L51 in *Nostoc* sp. PCC 7120 and L64 in cyanophage S-TIP37) and 50 (residues K53 in PCC 7120 and T66 in S-TIP37). Note that a variation is found between the NblAs from marine cyanobacteria and their cyanophages at these two positions, which could be indicative of a functional difference between them.

Supplementary data files

Supplementary Data File 1. Mass spectrometry analysis of unlabeled proteins during the infection by S-TIP37. Excel spreadsheet.

Supplementary Data File 2. Details of statistical analyses. Excel spreadsheet.

Supplementary Data File 3. Whole proteome and Neo N-termini enrichment characterization by mass spectrometry of *Synechococcus* WH8109 due to ectopic expression of the S-TIP37 NblA. Excel spreadsheet.

Supplementary Data File 4. Non-redundant T7-like cyanophage genomes used for phylogeny and additional metadata for representative T7-like cyanophage genomes. Excel spreadsheet.

Supplementary Data File 5. Rooted and annotated phylogenetic tree of T7-like cyanophages. jtree format.

Supplementary Data File 6. qPCR and mass spectrometry data for Extended Data Fig. 1a and b. Excel spreadsheet.

Supplementary Data File 7. Genome assemblies of the wild-type and mutant S-TIP37 cyanophages. Fasta format.

Supplementary Data File 8. Trimmed amino acid sequences and annotated phylogenetic network of NblA proteins from cyanobacteria, T7-like cyanophages and other selected cyanophages. Nexus format.

Supplementary Data File 9. Exonuclease and *nblA* quantification data: sequences used for read recruitment and regions defined for quantification of exonuclease genes and downstream *nblA*, if present. Excel spreadsheet.

Supplementary Data File 10. NblA protein profiles: a general profile used for the searches in cyanophage genomes and a picocyanobacterial-specific profile. HMM file.

1391 **Supplementary Data File 11.** NblA proteins collected from T7-like cyanophages
1392 and picocyanobacteria and additional NblA sequences used for reference. Excel
1393 spreadsheet.