Oceanic photosynthesis is directly affected by cyanophage 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[2,6–9]. Under nutrient deprivation the cyanobacterial NblA is responsible for inducing proteolytic degradation of the phycobilisome[10–12], the large cyanobacterial photosynthetic light harvesting complex. This increases the pool of amino acids available for essential tasks[12], serving as a survival mechanism[13]. Ectopic expression of different cyanophage *nblA* genes results in host pigment protein degradation[9,7,6]. 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 cyanophages[14], we reveal that cyanophage NblA significantly accelerates the cyanophage infection cycle, directs degradation of the host phycobilisome and other photosynthetic proteins and reduces host photosynthetic light harvesting efficiency. Furthermore, 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 *nblA*-encoding cyanophages have a global effect on the amount of light harvested by oceanic picocyanobacteria.

The light-harvesting antennae of most cyanobacteria, including marine *Synechococcus,* are composed of phycobilisomes (PBSs), gigantic membrane-associated soluble supercomplexes, located on the exterior side of the thylakoid membrane[10,15]. Cyanobacteria capture and funnel solar energy through their PBSs towards the photosynthetic reaction centers[16] where the primary energy conversion reactions of photosynthesis occur. Typical PBSs 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 proteins (PEI, PEII) as well as linker proteins that connect rod subunits to each other and to the core and attach the PBS to the thylakoid membrane[17,18]. Each phycobiliprotein is composed of α and β subunits that form heterodimers[11]. These proteins bind the pigment chromophores phycocyanobilin, phycoerythrobilin and phycourobilin, which absorb light at different wavelengths[19]. Due to their large size, PBSs can comprise up to half of the soluble protein content of a cell[15]. *Prochlorococcus* is one of the few cyanobacteria that have a light harvesting complex different to the phycobilisome[20–22].

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]. The *nblA* gene codes for a proteolysis adaptor protein found in multiple cyanophage families[2,6–9], and is particularly prevalent in T7-like cyanophages

(family *Autographiviridae*). In nutrient deprived cyanobacteria, the cyanobacterial NblA is responsible for inducing degradation of the PBS complex, resulting in clear loss of pigmentation[10–12]. Here, we use S-TIP37, a typical T7-like cyanophage isolated from the Red Sea that infects the open ocean *Synechococcus* sp. strain WH8109[23], 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 throughout (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[24,25].

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 progeny from the cells) of 10 h compared to the 3 h latent period of the WT[26] (Fig. 1b). In addition, synthesis of phage proteins increased with time during infection with the WT phage (Fig. 1c, Suppl. Data File 1), whereas most phage protein production ceased after 3 h during infection with the Δ*nblA* mutant with significantly fewer virion morphogenesis proteins produced in the latter (Fig. 1c). These findings suggest a direct link between phage protein production and NblA expression and indicate that NblA is advantageous to the WT phage. Since cyanobacterial NblAs are known to degrade PBSs during

91 nutrient stress[10–12], and to release amino acids critical for cell survival[12,13], we
92 hypothesize that the phage NblA has a similar role during infection but that the
93 release of amino acids is used for building new virions.

94 To verify that the observed phenotype is a result of the lack of the *nblA*
95 gene, we performed a rescue experiment. This was achieved by infecting
96 *Synechococcus* sp. strain WH8109 cells that were ectopically expressing the
97 cyanophage *nblA* gene under an inducible expression system[27,28] with the Δ*nblA*
98 mutant phage (Fig. 1d). Indeed, such expression of the cyanophage NblA
99 restored the timing of infection to that found for the WT cyanophage (Fig. 1d).
100 These findings confirm that the *nblA* gene plays an important function in
101 cyanophage infection.

102 Cyanophage NblA would be expected to impact host photosynthesis if it
103 directs PBS degradation. PBSs transfer energy to the photochemical reaction
104 centers of the photosystems (PSI or PSII)[16,29]. Therefore, we measured the
105 functional efficiency of PSII light utilization using fluorescence (at room
106 temperature) as an indicator. The effective PSII absorption cross-sectional area
107 was calculated after exciting at the most peripheral disks of the PBS, PEII (495
108 nm) (Fig. 1e). While uninfected cells maintained a steady PSII cross-sectional
109 area, infection with both the WT and Δ*nblA* mutant phages resulted in a
110 considerable decline (Fig. 1e). Notably, this decline during infection with the Δ*nblA*
111 mutant occurred with a delay of 4 h relative to infection with the WT cyanophage.
112 This delay translates into a 50% longer period in which photosynthetic efficiency
113 remained unaffected during infection with the Δ*nblA* mutant. These findings
114 indicate that the cyanophage NblA has a substantial negative impact on the host's
115 functional efficiency of photosynthetic light utilization during infection. This
116 diminished photosynthetic performance could be attributed to targeted PBS
117 degradation, accelerated cell lysis, or a combination of both factors.

118

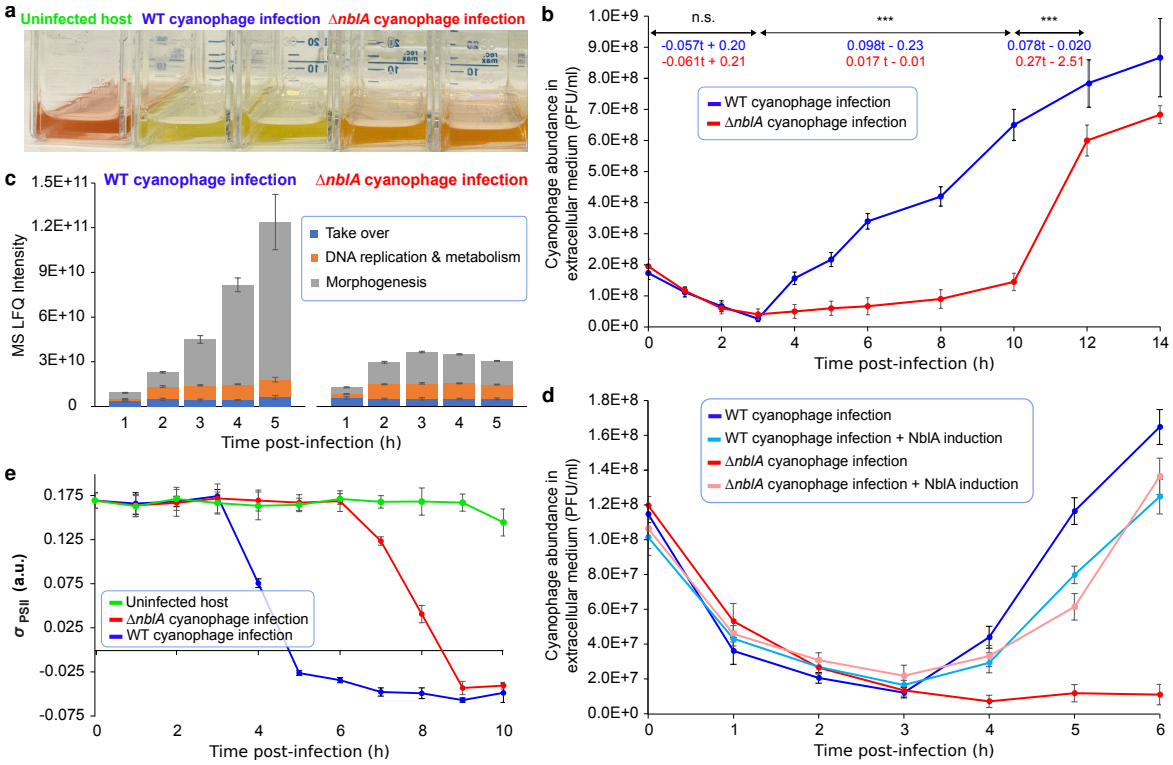


Fig. 1. Influence of the *nblA* gene on S-TIP37 cyanophage infection dynamics. a, Culture pigmentation phenotypes in uninfected *Synechococcus* sp. strain WH8109, and in WH8109 6 h post-infection when infected with WT S-TIP37 or the Δ*nblA* mutant S-TIP37 cyanophage. **b**, Cyanophage growth curves of the WT and Δ*nblA* mutant cyanophages. **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. **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. **e**, Effective PSII absorption cross-section (σPSII) from cells infected with the WT or the Δ*nblA* cyanophages and in uninfected control cells after preferential excitation at the PBS, at 495 nm (the maximal absorbance of PEII). Infection was performed at a multiplicity of infection (MOI) of 5. 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 Δ*nblA* 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. All experiments were performed in triplicates.

Cyanophage NblA causes degradation of host PBS

Cyanobacterial NblAs are known to be involved in PBS degradation. To assess whether the cyanophage NblA has a similar effect on host PBS integrity, we

149 monitored whole-cell fluorescence emission spectra as a measure of the energy
150 transfer pathway from the PBS to PSII. Samples collected at different time points
151 were excited at the peripheral PEII disk of the PBS and the corresponding
152 fluorescence emission spectra were measured. Infection with the WT cyanophage
153 led to a rapid and substantial decrease in the chlorophyll *a* (Chl) peak with time
154 (Fig. 2a, Extended Data Fig. 2a). This decline was first obvious at 3-h
155 post-infection (Fig. 2a), prior to the onset of cell lysis (Fig. 1b). In contrast,
156 infection by the Δ*nblA* knockout cyanophage led to a more gradual decrease in
157 Chl fluorescence (Fig. 2a, Extended Data Fig. 2a). These findings indicate that
158 energy transfer from PE to Chl was rapidly interrupted during infection with the
159 WT phage, but considerably less so with the Δ*nblA* mutant phage.

160 Whole-cell absorbance spectra and proteomic analysis provide a means to
161 assess the levels of host PBS subunits during infection. At 12 hours
162 post-infection, we observed a flattening of both PEI and PEII absorbance peaks,
163 as a result of infection by the WT cyanophage. A less prominent decrease in
164 these peaks was observed during infection with the Δ*nblA* mutant (Fig. 2b),
165 suggestive of NblA degradation of the PE proteins. Proteomic analysis during
166 infection with the WT phage revealed lower levels of the α subunits of PEII, PEI
167 and PC rods (CpeA, MpeA and RpcA respectively), as well as two rod linker
168 proteins (MpeC and MpeE) relative to infection with the Δ*nblA* mutant at 4-h
169 post-infection (Suppl. Data File 3). In contrast, three linker proteins closer to the
170 core[30] that connect PC to APC and both PC and APC to the thylakoid membrane
171 (ApcC, CpcG1, CpcG2) and one rod linker protein (CpeC) had higher protein
172 levels during infection with the WT relative to the Δ*nblA* mutant (Suppl. Data File
173 1) which may be indicative of compensation by the host to a reduction in active
174 PBSs. Combined, these findings provide further evidence for the role of
175 cyanophage NblA in PBS disassembly and degradation.

176 Next, we wished to determine the fate of the antenna complex during
177 infection. To achieve this, we isolated PBSs from host cells infected with WT and
178 Δ*nblA* cyanophages six hours post-infection and compared them to those in

uninfected cells. Fractionation of PBSs from uninfected cells on a sucrose gradient followed by fluorescence spectra analysis (Fig. 2c) allowed us to differentiate between assembled (lower, high density, pink bands) and disassembled complexes (upper, low density, orange bands) (see Extended Data Fig. 3). While intact and disassembled PBSs were apparent in uninfected host cells, only disassembled PBSs were present in cells infected by the WT cyanophage (Fig. 2c). Cells infected by the Δ*nblA* mutant cyanophage were composed of both intact and degraded PBSs (Fig. 2c), indicative of only partial PBS disassembly.

The combined findings above are consistent with a role for NblA in disassembly and degradation of the PBS during infection with the WT cyanophage. However, these findings could also be due, at least partially, to the beginning of cell lysis from 4-h onwards during infection with the WT phage. To more directly assess the impact of phage NblA on host PBSs without the confounding myriad effects of phage infection, we measured whole cell fluorescence spectra in host cells ectopically expressing the phage NblA (without infection). These measurements yielded a decline in the energy transfer from PE to Chl, seen by a reduction in the Chl fluorescence peak after excitation at PEII, relative to cells without induction of NblA expression (Fig 2d). Furthermore, ectopic expression of the phage NblA resulted in a reduction in the PEI and PEII absorbance peaks (Fig. 2e). These results are similar to those found during infection with the WT phage (compare Fig. 2a to 2d and Fig. 2b to 2e). Taken together, our findings suggest a clear effect of the cyanophage NblA on energy transfer from PE to Chl and in degradation of the host antenna complexes.

We used ectopic expression of cyanophage NblA in host cells to determine whether it could restore PBS disassembly during infection with the Δ*nblA* cyanophage mutant. Indeed, NblA expression led to a decrease in the Chl fluorescence peak when excited at the peripheral PEII, declining to levels similar to those during infection with the WT cyanophage by 3-4-h post-infection (right panels of Extended Data Fig. 4). Note that the partial decline in Chl fluorescence

at 4-h post-infection in the control treatment without induction (left panels of Extended Data Fig. 4) is likely caused by leaky expression in this inducible system[28]. These findings confirm that cyanophage NblA interrupts energy transfer from PE to Chl, consistent with a role in PBS disassembly and degradation during infection.

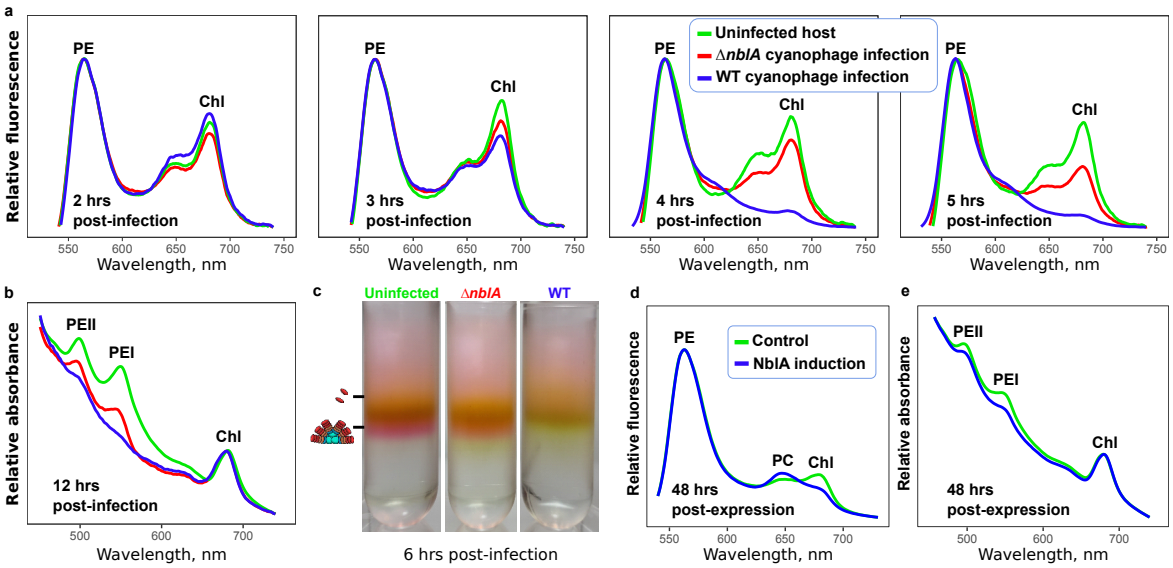


Fig. 2. Effect of cyanophage NblA on the cyanobacterial host photosynthetic antennae. a, Fluorescence emission spectra, **b**, absorbance spectra, and **c**, isolated PBSs, of uninfected *Synechococcus* sp. strain WH8109 cells (green), and WH8109 infected by the WT (blue) or Δ*nblA* mutant (red) S-TIP37 cyanophages at different time points post-infection. In (c), the lower density (upper) bands are of disassembled PBS subunits, and the higher density (lower) bands are of intact PBSs (see Extended Data Fig. 3). **d**, Fluorescence, and **e**, absorbance spectra of WH8109 cells 48-h after induced expression of the cyanophage NblA protein (blue) compared to the non-induced control (green). Fluorescence and absorbance spectra are representatives of 5 biological replicates. Averages and standard deviations are shown in Extended Data Fig. 2. Measurements were taken at room temperature, and peaks of Chl, PC and PE peaks are indicated. Fluorescence measurements were normalized on the PE peak and absorbance measurements were normalized on the Chl peak. In (a) and (d) the PE peak at 560 nm is made up of PEI and PEII.

Next, we assessed the protein targets of the cyanophage NblA protein by following the production of new protein cleavages in host cells ectopically expressing the viral NblA. For this, we analyzed proteins with newly produced N-termini using mass spectrometry[31] at 24 and 48 h post-induction (at which time maximum protein induction occurs[28]). Many new cleavages were observed (Fig. 3a). Among these, 79 proteins were cleaved exclusively upon viral NblA

236 overexpression, with 32 and 59 newly cleaved proteins detected at 24 and 48
237 hours post-induction, respectively (Fig. 3b, Suppl. Data File 3). At 24 h
238 post-induction, these cleaved proteins included two phycobiliproteins (Fig. 3b,c);
239 the peripheral PEII disk α subunit (MpeA), and the ApcE linker responsible for
240 energy transfer from the APC core to the chlorophylls of the photosynthetic
241 reaction centers in the thylakoid membrane[10]. At 48 h post-induction, the vast
242 majority of new proteolytic cleavages were of phycobilisome subunits (Fig. 3a,b).
243 These included the same three PEII, PEI and PC rod disk α subunits that
244 displayed low levels during infection with the WT phage (MpeA, CpeA and RpcA)
245 (see above) as well as the two PE rod disk β subunits (MpeB and CpeB) and two
246 linker proteins connecting between disks in the rods (CpeC, CpeE) (Fig. 3b,c,
247 Suppl. Data File 3). These proteolytic cleavages spanned almost all PBS rod
248 proteins, but not PBS core proteins (Fig. 3c). These results provide further
249 evidence that the cyanophage NblA protein leads to the degradation and
250 disassembly of the cyanobacterial PBS antenna complex. They also suggest that
251 the cyanophage NblA directs disassembly in a stepwise fashion, initially to
252 disconnect the host PBS from the thylakoid membrane, and to degrade peripheral
253 PEII disks, followed by cleavage of the more internal PEI and PC disks. This is
254 similar to the process known for cyanobacterial NblA during nutrient stress in
255 which PBS degradation begins at the peripheral rod disks and moves inwards
256 towards the PBS core[12].

257 To our surprise, ectopic expression of the cyanophage NblA led to
258 proteolytic processing of additional photosynthesis-related proteins beyond PBS
259 proteins (Suppl. Data File 3). These include those related to PSI (PsaC, PsaK),
260 the PSII oxygen evolution complex (PsbO, PbsU), and carbon fixation (CcmK2,
261 CsoS2). This suggests that the cyanophage NblA negatively impacts host
262 photosynthesis through the proteolytic degradation not only of PBS components
263 but also of reaction center proteins and proteins needed for carbon fixation.
264 Moreover, a set of proteins with functions other than those related to
265 photosynthesis were also cleaved as a result of expression of the cyanophage
266 NblA. These include proteins involved in translation, with the majority of cleavages

being in ribosomal proteins at 24 h post-induction (Fig. 3a,b). In addition, translation initiation and elongation factors were also cleaved 48 h post-induction (Fig. 3a,b, Suppl. Data File 3). Other key cellular proteins cleaved include proteins involved in carbon metabolism, a subunit of the cellular RNA polymerase (RpoC2), the cell division protein FtsZ, as well as porin and transporter proteins (Fig. 3b, Suppl. Data File 3). It is not known whether host NblA proteins also direct the degradation of these proteins. Cleavage of these proteins would thus disrupt multiple cellular processes during infection, in addition to those involved in photosynthesis.

In freshwater cyanobacteria, NblA proteins function by binding to PBS subunits as well as to the ClpC chaperone responsible for recruiting the protease that induces proteolytic degradation of the PBS[12,32]. Since proteolytic enzymes often have preferred cleavage sites, this motivated us to investigate cleavage sites after ectopic expression of the cyanophage NblA in the *Synechococcus* host (Fig. 3d). At 24 h post-induction, we detected cleavages primarily between asparagine (N) and serine (S) residues, while at 48 h post-induction an additional cleavage site was detected between lysine (K) and arginine (R) residues (Fig. 3d), providing an explanation for the additional cleaved proteins at 48 h post-induction. This could be a result of NblA recruiting different proteases or that the different substrates are cleaved by the same proteolytic system (presumably the Clp protease complex) that has broad cleavage specificity. Combined, these results suggest that the native cyanophage NblA functions to induce the cleavage of a suite of PBS and photosynthesis-related proteins as well as of other key cellular processes during infection with the WT phage.

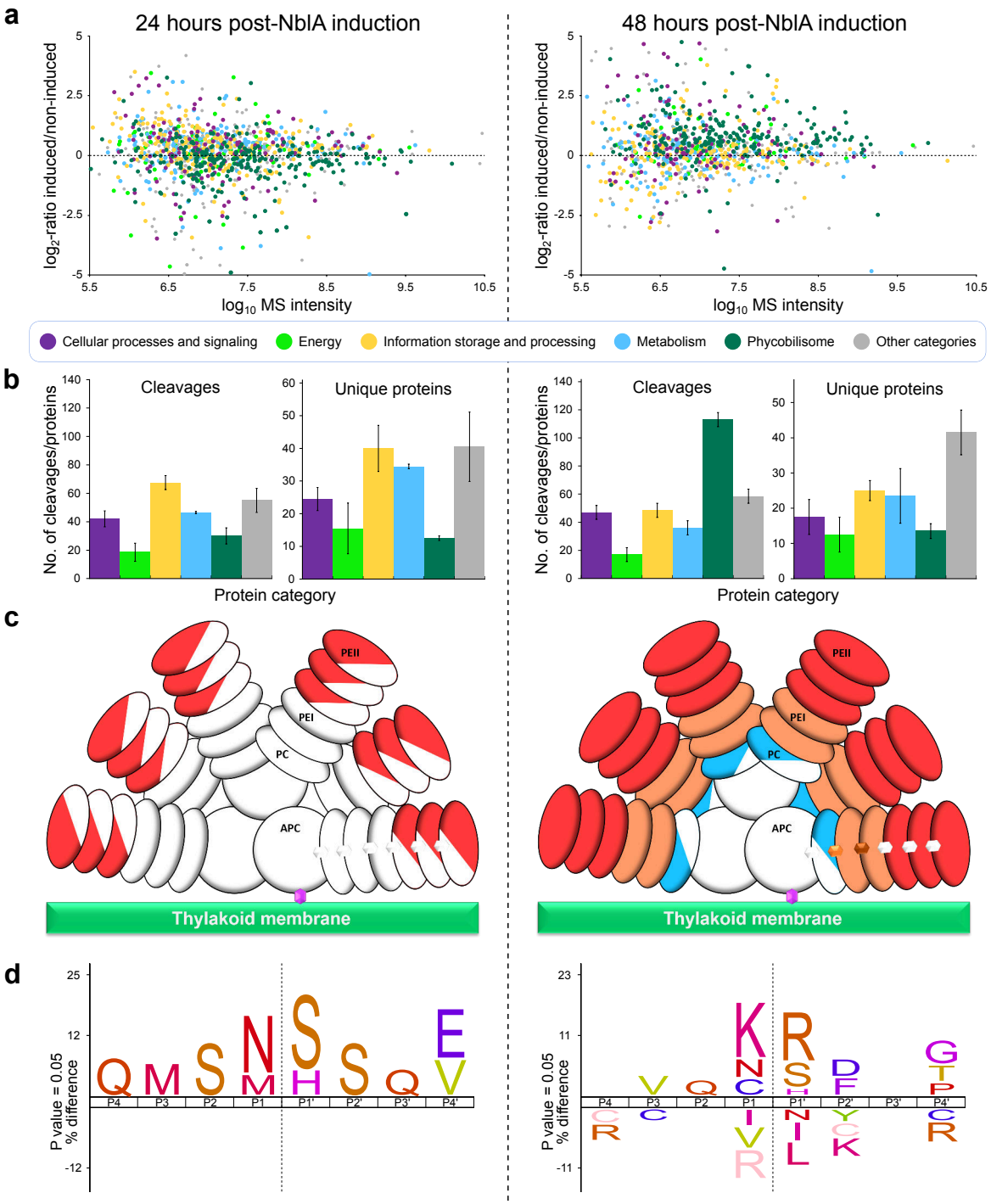


Fig. 3. Proteolytic patterns in host cells upon ectopic expression of the cyanophage NblA. Proteomic analysis of newly cleaved proteins after expression of the cyanophage NblA in *Synechococcus* sp. strain WH8109 cells for 24 h (left panels) and 48 h (right panels), as compared to non-induced cells (n=2). **a**, MA plots showing the abundance ratios of neo-N-terminal peptides, grouped by protein categories and plotted against their MS signal intensity. This signal intensity serves as a rough indicator of the peptides' overall abundance. Proteins are categorized according to a modified classification of COG functional groups (see Methods), with photosynthesis-related proteins other than PBS proteins included in the "Energy" group. **b**, The numbers of cleavages and unique cleaved

proteins in each category that exhibited a significant abundance change upon NblA expression. **c,** PBS cartoons showing proteolytic cleavages in the PBS subunits making up the discs. Colors indicate phycobiliproteins and linker proteins that underwent cleavages in PC (blue), PEI (orange) and PEII (red), the linker proteins between the discs (brown) and between the core and the thylakoid membrane (purple). Semi-colored discs represent cleavages detected only in the alpha subunits. Fully colored discs indicate cleavages detected in both the alpha and beta subunits. **d**, Preferred cleavage motifs upon ectopic expression of the cyanophage NblA. The amino acid sequence logo is for neo-N-terminal peptides that were identified only after S-TIP37 NblA expression and that had abundance ratios that were at least 2-fold higher in the NblA expressing cells. Cleavage sites are marked with a dashed line.

Cyanophage *nblA* is common in the oceans

T7-like cyanophages are very abundant in the oceans[26,33]. However, the prevalence of the *nblA* gene in this group is largely unknown. Therefore, to understand the environmental significance of our findings, we first assessed the distribution of *nblA* genes in assembled genomes of both isolated and environmental T7-like cyanophages. Our search revealed that 46% of the complete non-redundant T7-like cyanophage genomes code for *nblA* genes (Extended Data Fig. 5). *nblA*-encoding cyanophages have been isolated on marine picocyanobacterial hosts from the three main lineages: marine *Synechococcus* subcluster 5.1 (*Parasynechococcus*[34] in the broad sense), *Synechococcus* subcluster 5.2 (*Cyanobium* and related genera) and *Prochlorococcus* (the *Prochlorococcus* collective) (Fig. 4), even though the latter do not have phycobilisomes as their light-harvesting antennae[20,22]. We also found this gene to be present in a rare prophage residing in a *Synechococcus* subcluster 5.1 genome, AG-670-B23[35]. Thus, the cyanophage *nblA* gene is quite common in the genomes of T7-like cyanophages, infecting both *Synechococcus* and *Prochlorococcus.*

A close examination showed two types of *nblA* genes with differential distribution among the phylogenetic clades of T7-like cyanophages (Fig. 4, Extended Data Fig. 5 and Extended Data Fig. 6). *nblA* genes of the first type are dominant and are present in clade B cyanophages (to which S-TIP37 belongs) and to a lesser extent in clade A. These *nblA* genes are consistently located downstream of the exonuclease gene and frequently have up to a 110-nt overlap

336 with it. *nblA* genes of the second type are restricted to some P-RSP2-like
337 cyanophages (designated here as clade D) and are located downstream of an
338 HNH endonuclease gene. Neither *nblA* gene type was found in the genomes of
339 cyanophages from the newly discovered clade C (Fig. 4). Interestingly, one of the
340 amino acid positions (corresponding to residue L51 in NblA from *Nostoc*
341 (*Anabaena*) sp. PCC 7120) with the highest divergence between the two NblA
342 types has been previously implicated in interactions with phycobiliproteins[32] (see
343 Extended Data Fig. 7) and could be indicative of a functional difference between
344 them. The differences in the amino acid sequences of the two *nblA* gene types,
345 their location in the phage genomes, and their non-overlapping distribution among
346 the cyanophage clades, suggest that the *nblA* gene has been acquired at least
347 twice during T7-like cyanophage evolution.

348 It is noteworthy that the general distribution of the dominant *nblA* type in
349 T7-like cyanophages parallels that for the *psbA* gene (coding for the PSII reaction
350 center protein D1). Both genes are present in the majority of clade B
351 cyanophages, sporadically in clade A, and absent from clade C. However, they do
352 not always co-occur (see Fig. 4 and Extended Data Fig. 5, see also[23]).
353 Interestingly, the *nblA* gene is more common in T7-like cyanophages than the
354 *psbA* gene, with 36% of the complete genomes (including environmentally
355 assembled genomes) across all clades coding for *psbA* compared to 46% for *nblA*
356 (Extended Data Fig. 5). Since the *psbA* gene is considered to be one of the most
357 common auxiliary metabolic genes in cyanophage genomes[2,3,23,36,37], our results
358 suggest that the cyanophage-encoded *nblA* gene is very common in oceanic
359 environments as well.

360 We now turned to assessing the relative abundance of T7-like
361 cyanophages carrying the *nblA* gene in the oceans. We made use of the
362 consistent genomic association of the dominant *nblA* type with the exonuclease
363 gene, a core gene found in all T7-like cyanophages sequenced so far[36,37], to
364 interrogate the read data from the Global Ocean Viromes (GOV) metagenomic
365 dataset. This allowed us to determine both the presence or absence of the *nblA*

366 gene and the relative abundance of the different clades from the exonuclease
367 gene itself.

368 Our quantification strategy yielded general distribution patterns of T7-like
369 cyanophages similar to previously reported studies. Cyanophages belonging to
370 clades B and C dominated the global oceans[26,37], while clade A and D phages
371 were relatively rare. The vast majority of cyanophages from clade B possess the
372 *nblA* gene (Fig. 5), with ~72% of those at the surface, and 89% of those deeper in
373 the photic zone at the deep chlorophyll maximum (DCM), carrying the gene. The
374 proportion of *nblA*-encoding cyanophages among the less abundant clade A
375 group was much lower, averaging ~24% at the surface and 10% at the DCM. In
376 accordance with the distribution of the *nblA* genes in complete genomes
377 described above, none of the clade C and D cyanophages had *nblA* genes
378 downstream of the exonuclease gene. Since the *nblA* gene is located in a
379 different genomic position in the relatively rare clade D, our quantification method
380 could not be applied to them. Overall ~35% and 65% of the T7-like cyanophages
381 (from clades A, B and C) coded for the *nblA* gene in the surface and at the DCM,
382 respectively. These findings indicate that *nblA* is widespread amongst the
383 abundant clade B cyanophages both at the surface and deeper in the photic zone.

384 T7-like cyanophages from clade B that code for the *nblA* gene infect
385 picocyanobacteria with the PBS antenna (marine *Synechococcus* lineages) or
386 without PBS (*Prochlorococcus*) (Fig. 4). Both groups of cyanobacteria are found in
387 surface waters, whereas *Prochlorococcus* is considerably more abundant than
388 marine *Synechococcus* at the DCM[20,38]. Thus, we anticipate that many of the
389 clade B phages at the DCM infect *Prochlorococcus*. Since *Prochlorococcus* has a
390 different type of light harvesting antenna to the phycobilisomes of most
391 cyanobacteria[20–22], NblA of the phages infecting *Prochlorococcus* would be
392 expected to direct the degradation of other proteins. Indeed, our proteomics data
393 indicate that NblA-directed degradation is not restricted to PBS proteins, even for
394 a clade B phage that infects *Synechococcus* (Fig. 3a,b). Thus, other
395 photosynthesis-related proteins and proteins involved in other cellular processes,

are likely targeted by many *nblA*-encoding cyanophages deep in the photic zone of the oceans.

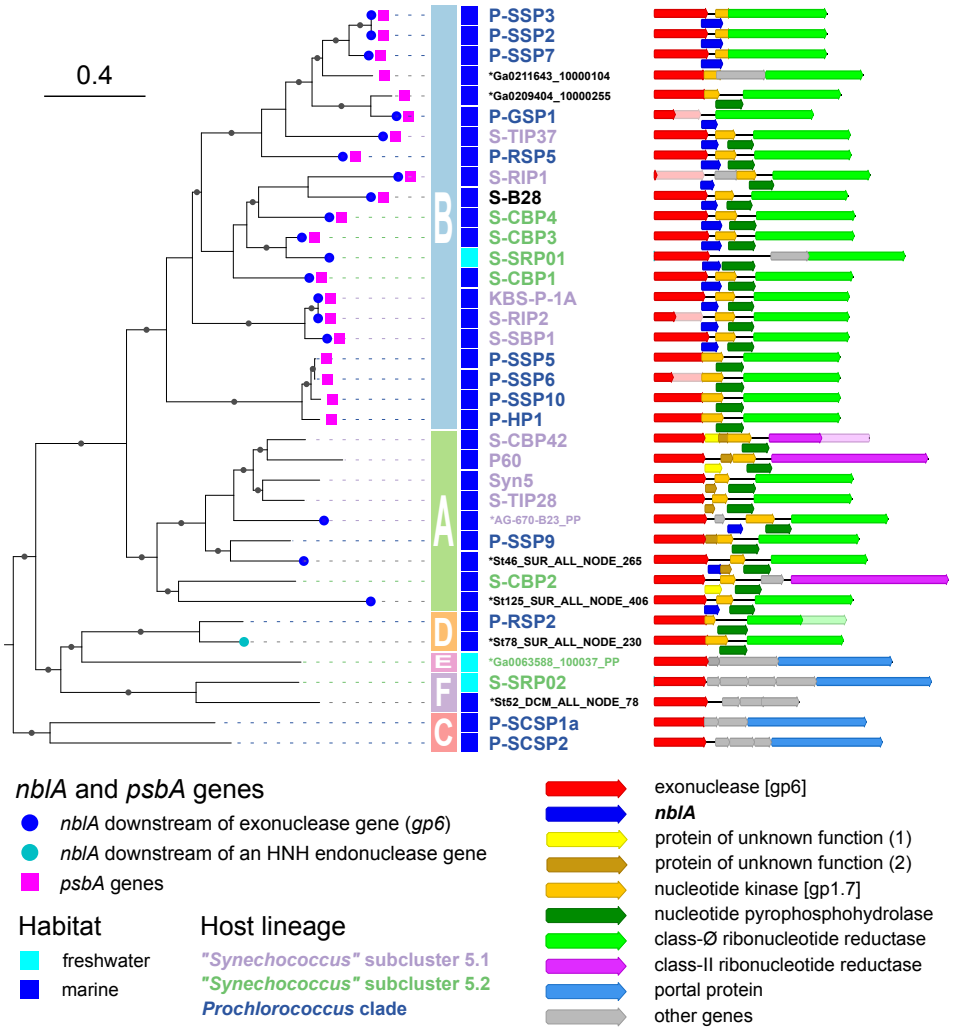


Fig. 4. Distribution of *nblA* genes among isolated T7-like cyanophages and selected environmental phage genomes. Phylogenetic tree is based on concatenation of nine core genes (see Methods), is outgroup-rooted and was pruned to include the chosen genomes (see Extended Data Figs. 5 and Suppl. Data Files 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: D (P-RSP2-like phages), E (no isolated representatives) and F (S-SRP02-like phages). 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.

Cyanobacteria acclimate to low light to maximize their ability to harvest light[39,40]. This is achieved through increasing the number and size of their antenna

complexes[39,40] as well as the number of their photosystems[41,42]. As such, more of the cyanobacterial cell's resources will be bound up in the photosynthetic apparatus at depth. Thus, we expect that the ability of cyanophages to degrade photosynthetic apparatus deep in the photic zone would provide a considerable advantage for the release of amino acids for use in cyanophage progeny production and may explain why relatively more clade B cyanophages at depth carry the *nblA* gene.

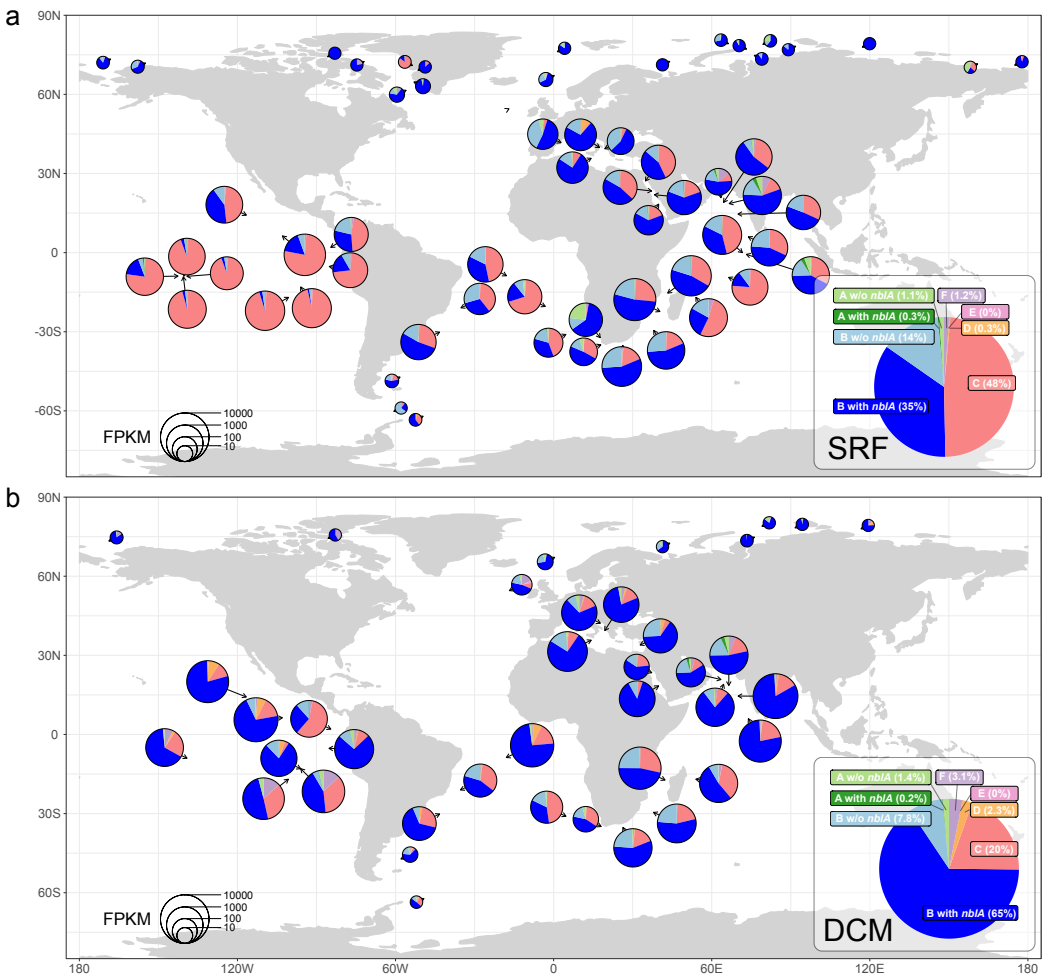


Fig. 5. Global distribution of T7-like cyanophages with and without *nblA* genes. a, Surface waters. **b**, Deep chlorophyll maximum (DCM). Quantifications are based on fragments per kilobase per million total reads (FPKM) abundances for exonuclease genes with and without downstream *nblA* genes and assigned to one of the five clades of *Autographiviridae*. Raw data and exonuclease-containing contigs for read recruitment were retrieved from the Global Ocean Viromes 2.0 dataset (GOV 2.0)[43]. 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.

429 Clade B cyanophages are considerably more abundant deeper in the
430 photic zone than at the surface[26,44]. This, together with the higher incidence of
431 *nblA*-genes among them, indicates that on a global scale, *nblA*-encoding
432 cyanophages vastly outnumber the *nblA*-less members of this dominant clade.
433 These findings suggest that cyanophage NblA-induced degradation of host PBS
434 and other photosynthesis-related proteins described above, is very common
435 among marine T7-like cyanophages in the photic zone in general and in the
436 deeper layers of the sunlit ocean in particular.

437

438 Concluding remarks

439 Auxiliary metabolic genes in viruses[45,46] are a widespread phenomenon, especially
440 in marine cyanophages[1–4]. Previous studies have investigated the putative
441 function of several such genes using in-vitro biochemical approaches[47,48] or
442 ectopic expression[6,7,48,49]. Here we combined a recently developed cyanophage
443 genetic engineering system[14], together with inducible ectopic expression[28] and
444 state-of-the-art N-termini proteomics[31] to directly examine the role of *nblA*, a gene
445 coding for a small proteolysis adaptor, in cyanophage infection and its impact on
446 the host photosynthetic apparatus. While NblA in cyanobacteria serve as a stress
447 response mechanism under nutrient deprivation and mediate the controlled
448 degradation of the PBS photosynthetic antenna[10–12], our findings suggest that its
449 cyanophage counterpart is used not for cell survival, but for virion synthesis. Our
450 results further reveal that the cyanophage NblA directs degradation not only of
451 host phycobilisomes, but also of a suite of other proteins, including important core
452 photosynthesis and house-keeping proteins. Thus, this small auxiliary metabolic
453 gene is likely to have a large impact on both the host's light harvesting efficiency
454 and on other essential cellular processes. The implications of the *nblA* gene for
455 the cyanophage are dramatic with a 3-fold shorter infection cycle. These findings,
456 together with genomic and metagenomic findings of high abundance and
457 widespread distribution of cyanophages carrying the *nblA* gene in the oceans (this
458 study), and in freshwater ecosystems[7–9], suggests that these cyanophages have a

global effect on the amount of light harvested and carbon fixed by oceanic cyanobacteria.

References

1. Breitbart, M., Thompson, L. R., Suttle, C. A. & Sullivan, M. B. Exploring the vast diversity of marine viruses. *Oceanography* **20**, 135–139 (2007).
2. Gao, E.-B., Huang, Y. & Ning, D. Metabolic genes within cyanophage genomes: implications for diversity and evolution. *Genes* **7**, 80 (2016).
3. Puxty, R. J., Millard, A. D., Evans, D. J. & Scanlan, D. J. Shedding new light on viral photosynthesis. *Photosynth. Res.* **126**, 71–97 (2015).
4. Zimmerman, A. E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev. Microbiol.* **18**, 21–34 (2020).
5. Tian, F. *et al.* Prokaryotic-virus-encoded auxiliary metabolic genes throughout the global oceans. *Microbiome* **12**, 159 (2024).
6. Nadel, O. *et al.* An uncultured marine cyanophage encodes an active phycobilisome proteolysis adaptor protein NblA. *Environ. Microbiol. Rep.* **11**, 848–854 (2019).
7. Ou, T., Gao, X.-C., Li, S.-H. & Zhang, Q.-Y. Genome analysis and gene *nblA* identification of *Microcystis aeruginosa* myovirus (MaMV-DC) reveal the evidence for horizontal gene transfer events between cyanomyovirus and host. *J. Gen. Virol.* **96**, 3681–3697 (2015).
8. Yoshida, T. *et al.* Ma-LMM01 infecting toxic *Microcystis aeruginosa* illuminates diverse cyanophage genome strategies. *J. Bacteriol.* **190**, 1762–1772 (2008).
9. Gao, E.-B., Gui, J.-F. & Zhang, Q.-Y. A novel cyanophage with a cyanobacterial nonbleaching protein A gene in the genome. *J. Virol.* **86**, 236–245 (2012).
10. Collier, J. L. & Grossman, A. R. A small polypeptide triggers complete degradation of

483 light-harvesting phycobiliproteins in nutrient-deprived cyanobacteria. *EMBO J.* **13**, 1039–1047
484 (1994).
485 11. Bienert, R., Baier, K., Volkmer, R., Lockau, W. & Heinemann, U. Crystal structure of NblA from
486 *Anabaena* sp. PCC 7120, a small protein playing a key role in phycobilisome degradation. *J.*
487 *Biol. Chem.* **281**, 5216–5223 (2006).
488 12. Sendersky, E. *et al.* The proteolysis adaptor, NblA, is essential for degradation of the core
489 pigment of the cyanobacterial light-harvesting complex. *Plant J. Cell Mol. Biol.* **83**, 845–852
490 (2015).
491 13. Schwarz, R. & Grossman, A. R. A response regulator of cyanobacteria integrates diverse
492 environmental signals and is critical for survival under extreme conditions. *Proc. Natl. Acad.*
493 *Sci. U. S. A.* **95**, 11008–11013 (1998).
494 14. Shitrit, D. *et al.* Genetic engineering of marine cyanophages reveals integration but not
495 lysogeny in T7-like cyanophages. *ISME J.* **16**, 488–499 (2022).
496 15. Grossman, A. R., Schaefer, M. R., Chiang, G. G. & Collier, J. L. Environmental effects on the
497 light-harvesting complex of cyanobacteria. *J. Bacteriol.* **175**, 575–582 (1993).
498 16. Ueno, Y., Aikawa, S., Kondo, A. & Akimoto, S. Energy transfer in cyanobacteria and red algae:
499 confirmation of spillover in intact megacomplexes of phycobilisome and both photosystems.
500 *J. Phys. Chem. Lett.* **7**, 3567–3571 (2016).
501 17. Ong, L. J. & Glazer, A. N. Phycoerythrins of marine unicellular cyanobacteria. I. Bilin types and
502 locations and energy transfer pathways in *Synechococcus* spp. phycoerythrins. *J. Biol. Chem.*
503 **266**, 9515–9527 (1991).
504 18. Six, C. *et al.* Diversity and evolution of phycobilisomes in marine *Synechococcus* spp.: a
505 comparative genomics study. *Genome Biol.* **8**, R259 (2007).
506 19. Glazer, A. N. Light guides. *J. Biol. Chem.* **264**, 1–4 (1989).

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

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

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

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

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

25. Huang, S., Sun, Y., Zhang, S. & Long, L. Temporal transcriptomes of a marine cyanopodovirus and its *Synechococcus* host during infection. *MicrobiologyOpen* **10**, e1150 (2021).

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

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

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

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

30. Kondo, K., Mullineaux, C. W. & Ikeuchi, M. Distinct roles of CpcG1-phycobilisome and

CpcG2-phycobilisome in state transitions in a cyanobacterium Synechocystis sp. PCC 6803. *Photosynth. Res.* **99**, 217–225 (2009).

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

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

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

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

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

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

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

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

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

40. Kolodny, Y. *et al.* Phycobilisome light-harvesting efficiency in natural populations of the

555 marine cyanobacteria *Synechococcus* increases with depth. *Commun. Biol.* **5**, 727 (2022).

556 41. Bonisteel, E. M. *et al.* Strain specific differences in rates of Photosystem II repair in
557 picocyanobacteria correlate to differences in FtsH protein levels and isoform expression
558 patterns. *PLOS ONE* **13**, e0209115 (2018).

559 42. Kolodny, Y. *et al.* Marine cyanobacteria tune energy transfer efficiency in their light-harvesting
560 antennae by modifying pigment coupling. *FEBS J.* **288**, 980–994 (2021).

561 43. Gregory, A. C. *et al.* Marine DNA viral macro- and microdiversity from pole to pole. *Cell* **177**,
562 1109-1123.e14 (2019).

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

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

568 46. Anantharaman, K. *et al.* Sulfur oxidation genes in diverse deep-sea viruses. *Science* **344**,
569 757–760 (2014).

570 47. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial
571 host carbon metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **108**, E757-64 (2011).

572 48. Dammeyer, T., Bagby, S. C., Sullivan, M. B., Chisholm, S. W. & Frankenberg-Dinkel, N. Efficient
573 phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Curr. Biol.* **18**, 442–448
574 (2008).

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

Methods

Cyanobacterial growth

S*ynechococcus* sp. strain WH8109 was grown in artificial seawater (ASW) medium[50], with modifications as described in Lindell et al.[51]. Cultures were grown at 21 °C and at a light intensity of 20 µmol photons m^{-2} s^{-1} , under a 14:10 light-dark cycle. Pour-plates were obtained with ASW medium with low-melting point agarose at a final concentration of 0.28% with additional 1 mM of sodium sulfite. Heterotrophic helper strain, *Alteromonas* sp. EZ55, was added to the pour-plate mixture for isolation of *Synechococcus* colonies[52].

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

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

Total RNA was reverse transcribed into cDNA using the LunaScript® RT SuperMix Kit (New England Biolabs, E3010). The reaction mixture was prepared in a total volume of 20 µl, 4 µl of LunaScript RT SuperMix containing random hexamers, and 16 µl of nuclease-free water, with the RNA sample added to reach the final volume. As a control, a similar mixture was prepared, but without addition of the reverse transcriptase enzyme. The reaction was incubated at 25 °C for 2 minutes, 55 °C for 10 minutes, and 95 °C for 1 minute.

Quantitative PCR (qPCR) reactions were prepared using LightCycler 480 SYBR Green I Master mix from Roche, combined with 0.2 µM of each primer and the *nblA*/*rnpB* DNA template. The reactions were run on a LightCycler 480 Real-Time PCR System. Cycle threshold fluorescence values for each reaction were determined using LightCycler 480 software. To quantify DNA copy numbers,

614 a standard curve was generated by running 10-fold serial dilutions of the template
615 and correlating cycle threshold values to known DNA concentrations. See Suppl.
616 Data File 6.

617 The genes whose expression levels were analyzed with qPCR were as
618 follows: host *nblA2* gene (Syncc8109_1607, see explanation below, primers
619 5'-GCGATCAAGCGGTCAATCAAC-3' and 5'-CTCTCTGCCGCACGTAGAGG-3'),
620 host *rnpB* gene (Syncc8109_0157, primers 5'-CATCGGCGGTGTGTTTCT-3' and
621 5'-CAGGCTTGCTGGGT-3'), S-TIP37 *nblA* gene (primers
622 5'-TTCCCGAGGCAGACAAGAG-3' and 5'-TAATGGGATGGTGACTCGGC-3'),
623 S-TIP37 DNA polymerase gene (STIP37_17B, primers
624 5'-TGAGCTACTACGCAACAGGC-3' and 5'-AGCGCGATCATTCAGGGAAG-3').
625 The *Synechococcus* sp. strain WH8109 *nblA* gene chosen for qPCR is the one we
626 reported previously in [6], although a more refined remote homology search with
627 hhsearch [53] using a custom NblA profile, reveals that the genome carries four
628 additional *nblA*-like genes and an *nblA* pseudogene. In order to clarify which
629 homolog has the highest structural similarity to previously characterized proteins
630 and is thus most likely to have a function similar to that of freshwater
631 cyanobacterial NblAs, we obtained the structures for the corresponding
632 monomers with ColabFold[54] and searched them against the Protein Data Bank
633 with Foldseek[55]. The best match indeed obtained for NblA2 (Syncc8109_1607),
634 with the highest similarity to NblA protein from *Nostoc* sp. PCC 7120[11] (PDB:
635 1OJH.E, probability 0.94, TM-score 0.6621).

636

637 **Cyanophage infection experiments**

638 Prior to infection, *Synechococcus* sp. strain WH8109 were grown in liquid medium
639 to mid-log growth to ~1 × 10^8 cells ml^{-1}. Infection experiments were initiated by
640 adding the S-TIP37 cyanophage strains at a multiplicity of infection (MOI) of 5.
641 Infection dynamics were determined by collecting samples at hourly time intervals
642 during the initial six hours, followed by sampling every two hours thereafter.
643 Samples were filtered over a 0.22 µm syringe filter and the filtrate containing free
644 cyanophages was plated to determine the number of infective cyanophages using
645 the plaque assay[56] in semi-solid pour plates (see above). Statistical analysis of the
646 infection course was performed with segmented linear regression by selecting
647 near-linear ranges in the infection curves and obtaining linear fits with factors time
648 post-inoculation, cyanophage type (wild-type or mutant) and their interaction and
649 replicate as a random effect using lme4 v. 1.1-31[57]. Significance of the model
650 terms was tested with the Anova function from the car package v. 3.1-1[58].
651 Distributions of the residuals were checked with QQ-plots.

652

653 **Spectral measurements**

654 Absorbance spectra were measured at room temperature using a double beam
655 Shimadzu spectrophotometer equipped with deuterium and halogen lamps.
656 Spectral data were obtained at a constant bandpass with a resolution of 1 nm.
657 Absorbance spectra were normalized to chlorophyll (680 nm). Fluorescence
658 emission spectra were collected using the Jobin-Yvon Horiba Spectro-fluorometer.
659 Samples were placed in 1 ml micro quartz cuvettes, excited at 495 nm and
660 fluorescence emission measured from 500 nm to 750 nm and normalized to the
661 fluorescence intensity at the emission maximum of PEI (562-564 nm).

662 Relative PSII absorption cross-section (σPSII) was determined by
663 measuring fluorescence emission, after dark adaptation of 15 minutes, following
664 the addition of (3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to a final
665 concentration of 10 μM. DCMU blocks electron transfer between the primary
666 quinone electron acceptor (Q_A) and the secondary quinone electron acceptor (Q_B)
667 on the acceptor side of PSII, causing the reaction centers to remain in a closed
668 state[59]. This allows for the measurement of the ability of PSII reaction centers to
669 absorb photons transferred from antenna complexes, thereby representing the
670 efficiency of light utilization by PSII. Samples were excited at 495 nm and
671 fluorescence emission measured from 500 nm to 750 nm. The relative σPSII was
672 calculated by integrating the area under the fluorescence emission spectrum, as
673 described by Schoffman et al[60] from the onset of infection up to nine hours
674 post-infection, at which point values fell below zero.

675 Quantification of the absorption levels of the different pigments was
676 performed by subtracting the background with the package baseline v. 1.3-4 [61]
677 using the 'modpolyfit' method, identifying the peaks of PEII, PEI and chlorophyll
678 and scaling the intensities to the intensity of the chlorophyll peak. Differences
679 between the uninfected cells and cells infected with the wild-type and mutant
680 cyanophages 12 h pos-infection were analyzed with linear models and Tukey's
681 post-hoc tests. Fluorescence emission spectra were scaled to the maximum
682 intensity (intensity of the phycoerythrin peak) to obtain the relative emission
683 intensity of the chlorophyll peak. Since the fluorescence measurements were
684 performed repeatedly for the same batches, a mixed model was built to analyze
685 the relative fluorescence of chlorophyll as a function of infection (uninfected cells
686 and two types of cyanophages), time post-infection, their interaction taking into
687 account the random effect of batch with lmer4. Distributions of the residuals were
688 checked with QQ-plots. Anova function from the car package was used to obtain

689 analysis of deviance tables and *p*-values for the model terms, Tukey's test was
690 performed as post-hoc in both cases.

691

692 **PBS extraction**

693 *Synechococcus* sp. strain WH8109 cells were harvested by centrifugation at
694 6000*g* for 10 minutes at 4 °C, 6 hours post-infection and kept at -80 °C. Pellets
695 were thawed at room temperature, resuspended with 7.5 M K-Phosphate, pH 7.5,
696 and homogenized. Cells were disrupted with a microfluidizer (Microfluidics
697 Corporation HC-2000) at 80 psi. The supernatant was collected after
698 centrifugation at 20,000*g*, for 30 minutes at 4 °C and loaded onto a 0.25-1.25 M
699 linear sucrose gradient in 7.5 M K-Phosphate buffer, pH 7.5. Samples were spun
700 down in an ultracentrifuge at 170,000*g* for 18 hours at 4 °C. Fractions were
701 collected with a syringe and analyzed by fluorometry at 77K and
702 spectrophotometry as described above.

703

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

705 The ORF encoding the *nblA* gene is 237 bp long (NCBI accession MH540083.1,
706 located at 12,467-12,706 nt, overlapping the last 94 nt of the exonuclease gene)
707 and was previously identified by us in the genome of S-TIP37 based on a match
708 to our custom NblA protein profile[6]. Profile-profile searches with hhsearch via the
709 HHpred server[62] for this protein yield high-scoring matches to cyanobacterial NblA
710 sequences and Pfam NblA profile (PF04485). Analogously, folding the protein with
711 ColabFold[54] produces a typical dimeric structure of NblA with the two monomers
712 composed of two helices each and unstructured terminal regions.

713 The *nblA* mutant was produced following Shitrit et al[14]. Since phages gain
714 random mutations readily, we constructed two independent S-TIP37 Δ*nblA* mutant
715 cyanophages to ensure that the observed phenotype was directly related to the
716 lack of the *nblA* gene. Briefly, the *nblA* ORF was used to prepare a construct
717 containing a recombination template for viral homologous recombination. We
718 replaced 137 bp of the *nblA* gene from position 12,572 to 12,707 (downstream of
719 the region overlapping with the exonuclease gene) with a tag of 23 bp. The
720 construct was cloned into the replicative pRL-proCAT[14] plasmid and conjugated
721 into *Synechococcus* sp. strain WH8109[14,63]. The strain expressing the
722 recombination template was infected with the WT S-TIP37. The lysate was filtered
723 over a 0.22 µm syringe filter to remove cell debris. The presence of recombinant
724 phages was verified by PCR with one primer for the inserted TAG sequence and

725 one in the phage genome; 5'-TGGTGATCAGACCGATGGG-3' (forward) and
726 5'-GAGCTCATAGCAAAGAAGACGTC-3' (reverse).

727 Enrichment and PCR screening for recombinant cyanophages was carried
728 out in 96-well plates containing the *Synechococcus* host[14]. Wells containing
729 recombinant cyanophages were filtered and plated on semi-solid medium. Single
730 phages were plaque purified twice and the whole genome of each purified mutant
731 phage clone was sequenced alongside the corresponding WT. Illumina MiSeq
732 trimmed read pairs were received and the genomes were assembled *de novo* with
733 Spades v. 3.14.1[64]. Both Δ*nblA* mutants had additional mutations in their genomes
734 compared to their corresponding wild-type genomes (see Suppl. Data File 7 for
735 the genome sequences). Non-synonymous mutations in the first mutant relative to
736 the first WT had a T>C substitution at position 11,142 (position according to
737 reference genome MH540083.1) leading to a V208A mutation in the gene for DNA
738 polymerase nucleotidyl transferase subunit (accession AXF42115.1) and an A>C
739 substitution at position 37,714 in gene 42 coding for a putative 2OG-Fe(II)
740 oxygenase (accession AXF42102.1) leading to a K34T change. The second
741 Δ*nblA* mutant had two non-synonymous differences compared to its wild type in
742 gene 40 coding for a putative tail fiber protein (accession AXF42100.1): an A>C
743 substitution at position 36,856 in the wild-type genome leading to a D454A
744 change (relative to the reference genome) and an A>G substitution at 36,888 in
745 the Δ*nblA* genome leading to a N465D change. Since both Δ*nblA* mutants had
746 additional mutations compared to the wild-type, infection and spectral experiments
747 were performed with both mutants and gave the same results. Results with the
748 second mutant are shown. Rescue and proteomics experiments were performed
749 with the second mutant.

750

751 **Ectopic expression of the cyanophage *nblA***

752 A theophylline translational induction system was used for expression of the
753 cyanophage NblA protein in *Synechococcus* sp. strain WH8109[27,28]. A riboswitch
754 sequence that theophylline binds to and the full length S-TIP37 *nblA* gene were
755 cloned into the pRL-proCAT replicative plasmid downstream of the *rnpB* promoter
756 following Tahan et al[28]. and was conjugated into *Synechococcus* sp. strain
757 WH8109[14,63]. Expression experiments were carried out with the exponentially
758 growing WH8109 conjugant. Theophylline (0.3 mM), dissolved in ASW, was
759 added to induce translation of the NblA protein. Expression of proteins under this
760 theophylline-inducible system is leaky, with differences between induced and
761 non-induced expression detected from 7.5 h after addition of theophylline that
762 increased until 48 h post-induction[28]. For rescue experiments theophylline was

763 added 8 h prior to infection with cyanophages. For assessing the effect of NblA on
764 *Synechococcus* sp. strain WH8109 protein cleavage, samples were collected 24
765 and 48 h post-induction. Induced samples were compared to non-induced
766 samples. Each expression experiment was accompanied by a control construct,
767 carrying the pRL-proCAT plasmid containing the riboswitch but lacking a
768 downstream gene.

769

770 **Proteomic analyses**

771 ***Protein extraction:*** Cyanobacteria were harvested and washed three times using
772 50 mM of HEPES, pH 7.5. After the final centrifugation the cyanobacterial pellet
773 was resuspended with 8 M guanidine hydrochloride (GuHCl), 100 mM HEPES, pH
774 7.5, and was heated at 95 °C for 10 min. Following this, the samples were
775 sonicated using VialTweeter (Hielscher) at maximum amplitude, 70% cycle time
776 for 5 min, then heated for another 5 min at 95 °C to ensure maximal extraction
777 and denaturation. Cellular debris were pelleted at 18,000*g* for 10 min, and the
778 clear supernatant containing proteins was transferred to a new tube. Protein
779 concentrations of each sample were measured with BCA, prior to splitting each
780 sample taking 10 µg for total proteome analysis and 50 µg for N-terminome
781 analysis.

782

783 ***Total proteomics sample preparation:*** Proteins were reduced with 5 mM
784 DL-dithiothreitol (DTT) at 65 °C for 30 min then cooled to room temperature (RT)
785 and alkylated with 12.5 mM chloroacetamide (CAA) for 30 min in the dark.
786 Guanidinium concentration was diluted to 1 M using 100 mM
787 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 8, and trypsin
788 (Promega, V115A) was added at 1:100 (w/w) ratio for overnight at 37 °C. Next,
789 trypsin was quenched with 1% formic acid and the samples were desalted using
790 C18 stage tips (Empore 66883-U) before subjecting them to mass spectrometry
791 (MS) analysis.

792

793 ***N-termini proteomics sample preparation:*** N-terminal enrichment was done
794 using an optimized Hydrophobic tagging-assisted N-termini enrichment (HYTANE)
795 procedure as described[31]. Proteins were reduced and alkylated with DTT and
796 CAA as described above. Then, different isotopes (heavy and light) of
797 formaldehyde were used to label primary amines in every two comparative
798 samples. The labeling was done with 40 mM formaldehyde and 20 mM of sodium
799 cyanoborohydride (CNBr) at 37 °C overnight. Formaldehyde leftovers were

quenched using 100 mM glycine for 1 h at 37 °C before mixing the heavy and light labeled samples together. The samples were diluted to reduce GuHCl concentration to 1 M using 100 mM of HEPES pH 8 before adding trypsin at 1:100 (w/w) and overnight incubation at 37 °C. The trypsin was then quenched with 1% (final) of formic acid, and the samples were subjected to desalting using OASIS-HLB columns (Waters). The elution was done using 60% of acetonitrile, 0.1% formic acid and the samples were dried using a speed-vac. The peptides were resuspended using 100 µl of 100 mM HEPES pH 7 and had undecanal solution (20 mg/ml prepared in ethanol) added to them at w/w ratio of 1:50 (protein:undecanal), followed by the addition of CNBr at a final concentration of 20 mM. Undecanal tagging was done for 2 hours at 50 °C while renewing the CNBr after 1 h. The samples were then centrifuged at 16,000*g* for 5 min and the supernatant was transferred to a new tube before drying in a speed vac. The dried samples were resuspended in 500 µl of 2% acetonitrile, 0.1% formic acid and subjected again to OASIS-HLB column. Elution was done using 60% of acetonitrile, 0.1% formic acid and the samples were dried in a speed-vac before resuspending again and subjecting them to MS.

LC-MS analysis: Desalted samples were subjected to LC-MS analysis using an Orbitrap Exploris 480 coupled with an EvoSep One HPLC. Samples were introduced onto the EvoTip, which was then washed twice with 20 µl of 0.1% formic acid. The washed peptides remained wet, maintained by applying 150 µl of 0.1% formic acid atop of the EvoTip until MS analysis. The samples loaded on the EvoTips underwent chromatographic separation on a 15 cm × 150 µm analytical column, filled with 1.9 µm C18 beads (EV1106). Peptides were separated over an 88-minute gradient according to the manufacturer's standard method. Full MS scans acquired in positive ion mode, scanning from 300 to 1800 m/z, were recorded at a resolution of 120,000, by a data-dependent mode selecting the top 20 ions with high energy collisional dissociation (HCD) fragmentation ion at a resolution of 17,500 and with dynamic exclusion enabled.

MS data analysis: For total proteomics, data analysis was conducted using FragPipe v. 22.0 (https://fragpipe.nesvilab.org/) using DDA+ mode[65] with the default Label-Free Quantification Match Between Runs (LFQ-MBR) settings. The search enzyme was set to trypsin, allowing up to two missed cleavages. Variable modifications included methionine oxidation and protein N-terminal acetylation, while carbamidomethylation of cysteine was set as a fixed modification. For N-terminomics data, raw files were first converted to MzML format and then analyzed using the Trans-Proteomic Pipeline version 6.3[66] with the Comet search engine v. 2023_01 rev2. The search enzyme was set to Semi-ArgC, allowing up to two missed cleavages. Variable modifications included methionine oxidation and

the mass difference between heavy and light dimethylation at lysine residues or the N-termini of peptides. Fixed modifications included carbamidomethylation of cysteine and light dimethylation at lysine residues or peptides' N-termini.

Relative peptide quantifications were performed using XPRESS, with parameters set to mass tolerance of 20 ppm, a minimum of three chromatogram points for quantitation, and the number of isotopic peaks to sum set to zero. Post-analysis table creation, cleavage motif extraction, and ratio normalization were performed using an in-house script[67]. The search databases included all *Synechococcus* sp. strain WH8109 proteins in Uniprot (Taxon ID: 166314, both reviewed and unreviewed) in addition to the vector added genes and standard contaminant proteins. For viral infection samples, this database was expanded to also include all *Synechococcus* T7-like phage S-TIP37 (Taxon ID:1332145) protein sequences from Uniprot and the NblA sequence. To bin host proteins into functional groups, we utilized the COG functional category assignments for *Synechococcus* sp. strain WH8109 proteins in the eggNOG database v. 5.0[68]. The proteins were assigned to one of the broad functions based on a modified classification of COG functional groups as follows: phycobilisome components — Phycobilisome; other photosynthesis-related genes and other proteins assigned to COG functional category C (energy production and conversion) — Energy; COG functional categories J, A, K, L and B — Information Storage and Processing; COG functional categories D, Y, V, T, M, N, Z, W, U and O — Cellular Processes and Signaling; COG functional categories G, E, F, H, I, P and Q — Metabolism; COG functional categories R, S and proteins assigned to multiple categories — Other.

Phage proteins were classified into functional categories based on the functions of the majority of genes in three genomic and expression clusters[24,36]. Statistical analysis of the dynamics of abundance of the different functional groups of phage proteins in the total proteomics data as a function of time post-infection onset was performed as follows. For each gene cluster cumulative MS LFQ intensities were used as a response variable in a linear model with time post-infection and infection type (by WT or ΔnblA mutant) and their interaction as predictor terms. Distributions of the residuals were checked with QQ-plots. Detailed results of the statistical analysis are provided in Suppl. Data File 2.

Bioinformatic analyses

Genomes of cultured *Autographiviridae* cyanophages and outgroup phages (coliphage T7 and pelagiphages HTVC011P and HTVC019P) were collected from GenBank. Complete genomes and genomic fragments of environmental T7-like

878 cyanophages and cyanoprophages were obtained from JGI IMG/VR v. 4.1[69] and
879 Global Ocean Virome v. 2.0 assemblies[43]. To collect genomic fragments
880 containing exonuclease genes with sufficiently long downstream regions, a
881 conserved region in the C-terminal part of the exonuclease (corresponding to
882 residues 86-221 in the exonuclease protein from S-TIP37 [RefSeq accession
883 YP_009807515.1]) was chosen and homologous regions from a reference set of
884 cyanophages were used as queries for tblastn from NCBI blast v. 2.15.0[70].
885 Fragments were retained when they had matches at least 400 bp long, identity of
886 ≥40%, bit score of ≥400 and at least 500 bp of downstream sequence available.
887 To collect genomes for phylogenetic analysis, out of the contigs with the
888 exonuclease gene, we extracted sequences ≥30 Kbp long and dereplicated them
889 with dRep v. 3.4.5[71] at the ANImf identity level of 0.95. Genes were predicted with
890 prodigal v. 2.6.3[72] and phylogeny was reconstructed with phylophlan v. 3.0.2[73]
891 (diamond v. 2.1.8[74], MAFFT v. 7.475[75], trimAl v. 1.4.1[76], IQ-TREE v. 2.1.2[77]) based
892 on concatenated alignments of nine core genes: primase-helicase, exonuclease,
893 portal protein (head-to-tail adaptor), head assembly protein, major capsid protein,
894 tail tubular proteins A and B, small and large terminase subunits. The genes were
895 extracted by searching with diamond blastp using representative protein
896 sequences as queries from two distantly related cyanophages, P60 and S-SRP02:
897 P60_gp14 (primase-helicase is missing from S-SRP02), P60_gp18 and
898 SSRP02_p034, P60_gp26 and SSRP02_p038, P60_gp27 and SSRP02_p039,
899 P60_gp28 and SSRP02_p040, P60_gp29 and SSRP02_p041, P60_gp30 and
900 SSRP02_p042, P60_gp40 and SSRP02_p011, P60_gp49 and SSRP02_p012,
901 respectively. IQ-TREE was run with "-m LG --alrt 1000 -pers 0.2 -nstop 500" with
902 1,000 ultrafast bootstrap replicates[78] and the resulting tree was outgroup-rooted
903 with the ingroup defined as the branch encompassing the known cyanophages.
904 The cyanophage genomes were classified based on the resulting phylogeny into
905 six clades: the previously recognized clades A, B and C, the newly defined clades
906 D (for P-RSP2-like phages), E (represented by GOV contig
907 Station52_DCM_ALL_NODE_78) and F (S-SRP02-like phages). See Suppl. Data
908 File 4 for metadata of the genomes chosen to represent each of the clades. Two
909 outlier genomes were found to cause spurious clustering and were excluded from
910 the phylogenetic analysis: IMGVR_UViG 3300032116_000204 and IMGVR_UViG
911 3300029337_000223. The incidence of *nblA* and *psbA* genes in the genomes was
912 assessed by searching ORFs (between stop codons) with hmmsearch from
913 HMMER v. 3.4[79] using a custom NblA profile and the PsbA profile TIGR01151.1
914 from NCBI Protein Family Models. The NblA HMM profile was obtained from an
915 alignment of the previously released NblA sequences[6] and manually curated NblA
916 sequences from representative T7-like cyanophages. The searches for the NblA
917 hits were performed without heuristic filters and the results were filtered using an

918 empirically determined full score threshold of 20. The incidence and type of DNA
919 polymerase in the genomes was assessed by searching the ORFs with blastp
920 using the sequences of the exonuclease (YP_009807513.1) and polymerase
921 (YP_009807514.1) subunits of the split polymerase of S-TIP37 with an E-value
922 threshold of 1e-10. A version of the tree focusing on representative genomes as
923 shown in the main text, was obtained by trimming the full tree. The trees were
924 visualized with ggtree v. 3.2.0[80]. The full phylogenetic tree without the outgroups is
925 available in Suppl. Data File 5.

926 Unrooted phylogenetic tree of the NblA protein sequences from the
927 *Autographiviridae* cyanophages was obtained by aligning them with MAFFT,
928 trimming with trimAl (automated1 mode) and reconstructing the phylogeny with
929 IQ-TREE with 200 bootstrap replicates. The tree is available in Suppl. Data File 8.

930 For the analysis of the distribution of *nblA* downstream of the exonuclease
931 gene, we took all of the initially collected environmental contigs with
932 T7-cyanophage type exonucleases and extracted regions containing the 3' end of
933 the exonuclease gene with up to 100 bp upstream and up to 1,000 bp
934 downstream. These fragments were searched against exonuclease sequences
935 from the genomes used for phylogeny using blastx and clades were assigned to
936 fragments with best hits to cyanophages at identity level of ≥60%. The fragments
937 coming from the GOV2.0 assemblies were used for recruitment of GOV raw data
938 with bwa v. 0.7.17[81]. Quantification was performed with featureCounts from
939 subread v. 2.0.6[82]. As the exonuclease and *nblA* genes frequently overlap, we
940 selected strictly defined windows for read quantification instead of ORF
941 boundaries: the exonuclease was represented by the above-mentioned
942 conservative region at the 3' end of the gene. For the *nblA* gene we chose the
943 location corresponding to the core NblA region based on the hmmsearch
944 matches. Genes assigned to the same cyanophage clade were grouped into
945 metafeatures for quantification and the resulting per-station mapped read pair
946 counts were converted to Fragments Per Kilobase per Million reads (FPKM)
947 values by dividing them by the median feature length (in Kbp) and the total
948 number of the reads per sample (in millions). Sequences used for recruitment and
949 quantification results in tabular format are available in Suppl. Data File 9.

950 To compare the cyanobacterial NblA with the two clades of NblA proteins
951 from T7-like cyanophages, we used differential logos created with DiffLogo v.
952 2.26.0[83]. For this analysis, UniRef90 representative proteins assigned to Pfam
953 profile PF04485 and taxon "*Cyanobacteriota*" from UniProt r. 2024_02, was
954 combined together with the cyanophage NblAs and aligned with MAFFT in
955 automatic mode. The resulting fasta alignment was converted to A3M (allowing for
956 10% alignment gaps at most per position). Secondary structure was predicted

with addss.pl from HH-suite v. 3.3.0[53]. Structural models for representative cyanophage NblA proteins were obtained with ColabFold v. 1.5.5[54].

The bioinformatic workflow was implemented in snakemake[84], python and R.

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 [85] partner repository with the following dataset identifiers: PXD057452 (whole proteome data) and PXD057454 (N-terminomics data).

Code availability

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

Acknowledgments. We thank M. Suissa-Szlejf for help in setting up the PBS isolation gradients, R. Edrei for help with spectroscopic measurements, G. Sabehi for discussions on infection experiments, S. Goldin for help with qPCR calculations, and N. Keren for discussions on photosynthesis experiments and help in performing spectroscopy, and N. Frankenberg-Dinkel for comments on the manuscript. O.B. is grateful to the Atmosphere and Ocean Research Institute (U.Tokyo, Japan) for the stay during the work on the manuscript. This work was funded by the European Research Council (ERC Advanced Grant 321647 to O.B. and ERC Consolidator Grant 646868 to D.L.), the Nancy and Stephen Grand Technion Energy Program (GTEP), Israel Science Foundation grants to O.B. (143/18) and O.K. (1515/23), and the 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, PBS 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.

References for Methods

50. Wyman, M., Gregory, R. P. F. & Carr, N. G. Novel role for phycoerythrin in a marine cyanobacterium, *Synechococcus* strain DC2. *Science* **230**, 818–820 (1985).
51. Lindell, D., Padan, E. & Post, A. F. Regulation of *ntcA* expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J. Bacteriol.* **180**, 1878–1886 (1998).
52. Morris, J. J., Kirkegaard, R., Szul, M. J., Johnson, Z. I. & Zinser, E. R. Facilitation of robust growth of *Prochlorococcus* colonies and dilute liquid cultures by “helper” heterotrophic bacteria. *Appl. Environ. Microbiol.* **74**, 4530–4534 (2008).
53. Steinegger, M. *et al.* HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics* **20**, 473 (2019).
54. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
55. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* **42**, 243–246 (2024).
56. Kirzner, S., Barak, E. & Lindell, D. Variability in progeny production and virulence of cyanophages determined at the single-cell level. *Environ. Microbiol. Rep.* **8**, 605–613 (2016).
57. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
58. Fox, J. & Weisberg, S. *An R Companion to Applied Regression*. (Sage, Thousand Oaks CA,

1017 2019).

1018 59. Larom, S., Salama, F., Schuster, G. & Adir, N. Engineering of an alternative electron transfer

1019 path in photosystem II. *Proc. Natl. Acad. Sci.* **107**, 9650–9655 (2010).

1020 60. Schoffman, H. & Keren, N. Function of the IsiA pigment–protein complex in vivo. *Photosynth.*

1021 *Res.* **141**, 343–353 (2019).

1022 61. Liland, K. H., Almøy, T. & Mevik, B.-H. Optimal choice of baseline correction for multivariate

1023 calibration of spectra. *Appl. Spectrosc.* **64**, 1007–1016 (2010).

1024 62. Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology

1025 detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).

1026 63. Brahamsha, B. A genetic manipulation system for oceanic cyanobacteria of the genus

1027 *Synechococcus*. *Appl. Environ. Microbiol.* **62**, 1747–1751 (1996).

1028 64. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to

1029 single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).

1030 65. Yu, F., Deng, Y. & Nesvizhskii, A. I. MSFragger-DDA+ Enhances Peptide Identification Sensitivity

1031 with Full Isolation Window Search. 2024.10.12.618041 Preprint at

1032 https://doi.org/10.1101/2024.10.12.618041 (2024).

1033 66. Deutsch, E. W. *et al.* Trans-Proteomic Pipeline: robust mass spectrometry-based proteomics

1034 data analysis suite. *J. Proteome Res.* **22**, 615–624 (2023).

1035 67. Hanna, R., Rozenberg, A., Lavy, T. & Kleifeld, O. Increasing the coverage of the N-terminome

1036 with LysN amino terminal enrichment (LATE). *Methods Enzymol.* **686**, 1–28 (2023).

1037 68. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically

1038 annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.*

1039 **47**, D309–D314 (2019).

1040 69. Camargo, A. P. *et al.* IMG/VR v4: an expanded database of uncultivated virus genomes within

1041 a framework of extensive functional, taxonomic, and ecological metadata. *Nucleic Acids Res.*
1042 **51**, D733–D743 (2023).
1043 70. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
1044 71. Olm, M. R., Brown, C. T., Brooks, B. & Banfield, J. F. dRep: a tool for fast and accurate genomic
1045 comparisons that enables improved genome recovery from metagenomes through
1046 de-replication. *ISME J.* **11**, 2864–2868 (2017).
1047 72. Hyatt, D. *et al.* Prodigal: prokaryotic gene recognition and translation initiation site
1048 identification. *BMC Bioinformatics* **11**, 119 (2010).
1049 73. Asnicar, F. *et al.* Precise phylogenetic analysis of microbial isolates and genomes from
1050 metagenomes using PhyloPhlAn 3.0. *Nat. Commun.* **11**, 2500 (2020).
1051 74. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using
1052 DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
1053 75. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
1054 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
1055 76. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: a tool for automated
1056 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973
1057 (2009).
1058 77. Minh, B. Q. *et al.* IQ-TREE 2: new models and efficient methods for phylogenetic inference in
1059 the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
1060 78. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving
1061 the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
1062 79. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput. Biol.* **7**, e1002195 (2011).
1063 80. Xu, S. *et al.* Ggtree: A serialized data object for visualization of a phylogenetic tree and
1064 annotation data. *iMeta* **1**, e56 (2022).

1065 81. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1066 Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
1067 82. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1068 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
1069 83. Nettling, M. *et al.* DiffLogo: a comparative visualization of sequence motifs. *BMC*
1070 *Bioinformatics* **16**, 387 (2015).
1071 84. Mölder, F. *et al.* Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).
1072 85. Perez-Riverol, Y. *et al.* The PRIDE database resources in 2022: a hub for mass
1073 spectrometry-based proteomics evidences. *Nucleic Acids Res.* **50**, D543–D552 (2022).
1074 86. Sabehi, G. & Lindell, D. The P-SSP7 cyanophage has a linear genome with direct terminal
1075 repeats. *PLOS ONE* **7**, e36710 (2012).

1076

1077 Supplementary data files

1078 **Supplementary Data File 1.** Mass spectrometry analysis of unlabeled proteins
1079 during the infection of S-TIP37. Excel spreadsheet.

1080 **Supplementary Data File 2.** Details of statistical analyses. Excel spreadsheet.

1081 **Supplementary Data File 3.** S-TIP37 NblA Neo N-termini enrichment
1082 characterization by mass spectrometry. Excel spreadsheet.

1083 **Supplementary Data File 4.** Non-redundant *Autographiviridae* cyanophage
1084 genomes used for phylogeny and additional metadata for representative
1085 *Autographiviridae* genomes. Excel spreadsheet.

1086 **Supplementary Data File 5.** Rooted and annotated phylogenetic tree of the
1087 *Autographiviridae* cyanophages. jtree format.

1088 **Supplementary Data File 6.** qPCR and mass spectrometry data for Extended
1089 Data Fig. 1a and b. Excel spreadsheet.

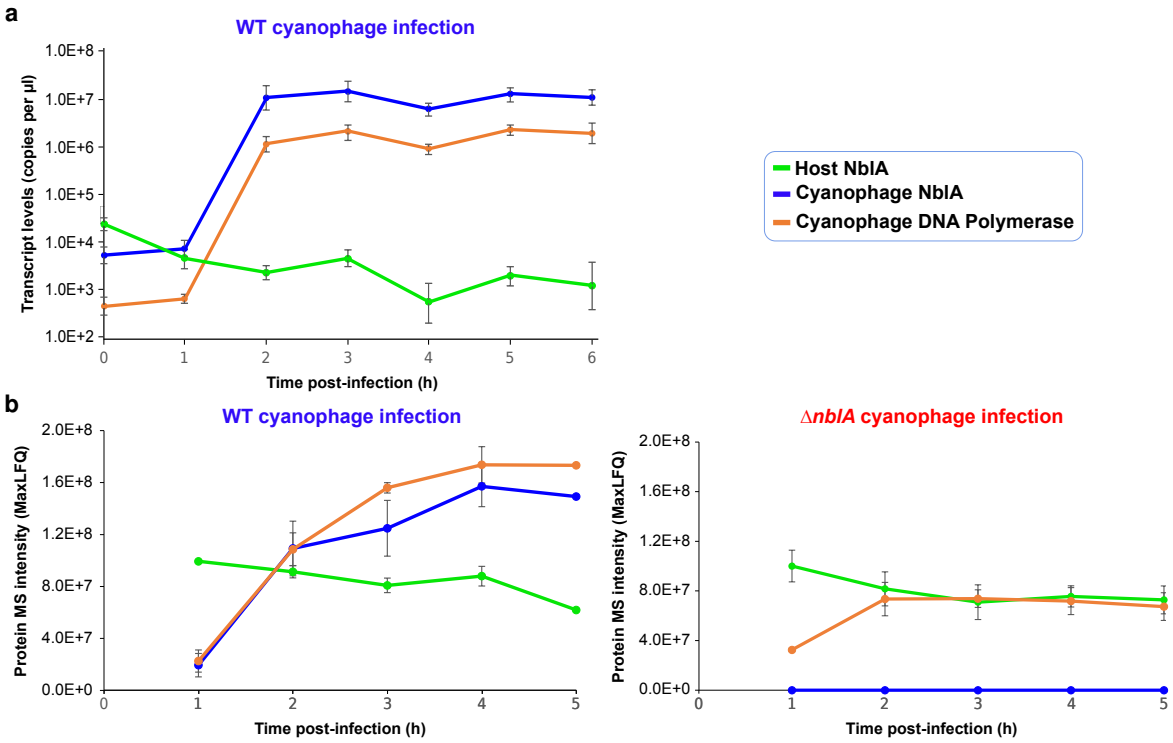
1090 **Supplementary Data File 7.** Genome assemblies of the wild-type and mutant
1091 S-TIP37 cyanophages. Fasta format.

1092 **Supplementary Data File 8.** Amino acid sequences and annotated phylogenetic
1093 tree of NblA from the *Autographiviridae* cyanophages. jtree format.

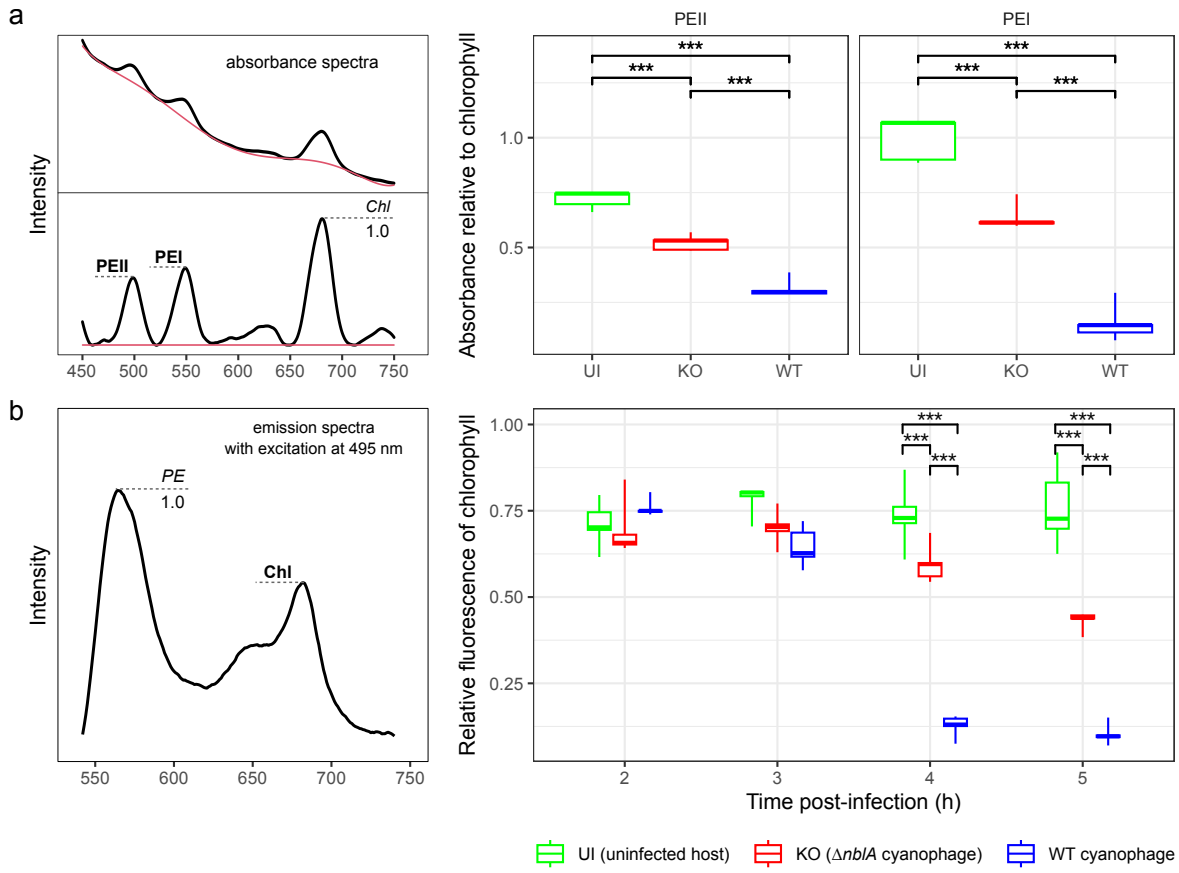
1094 **Supplementary Data File 9.** Exonuclease and *nblA* quantification data:
1095 sequences used for read recruitment and regions defined for quantification of
1096 exonuclease genes and downstream *nblA*, if present. Excel spreadsheet.

1097

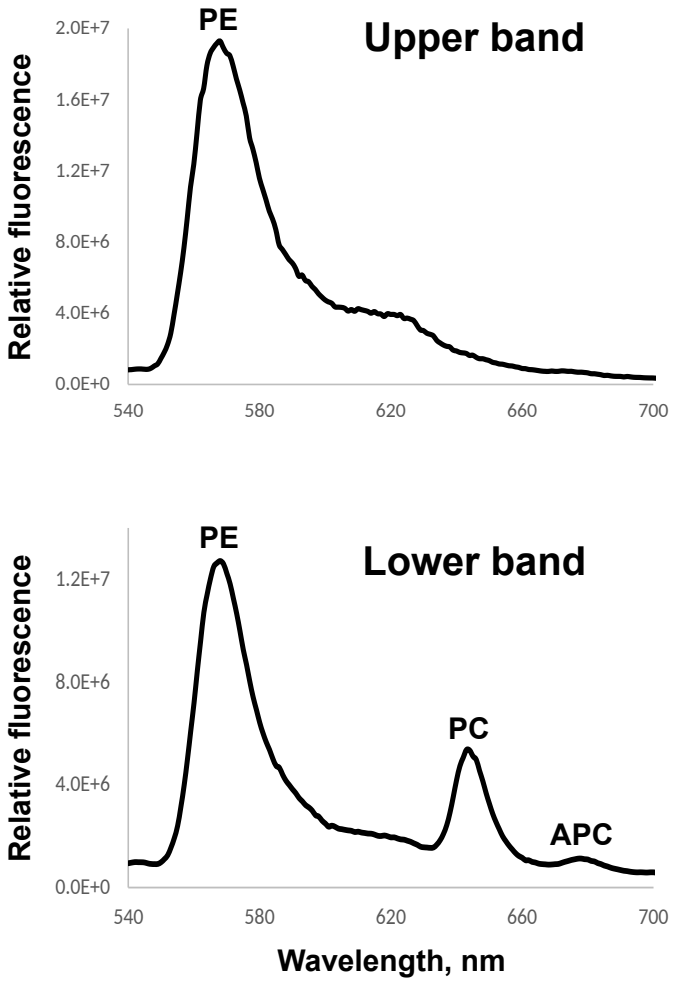
Extended data figures and tables



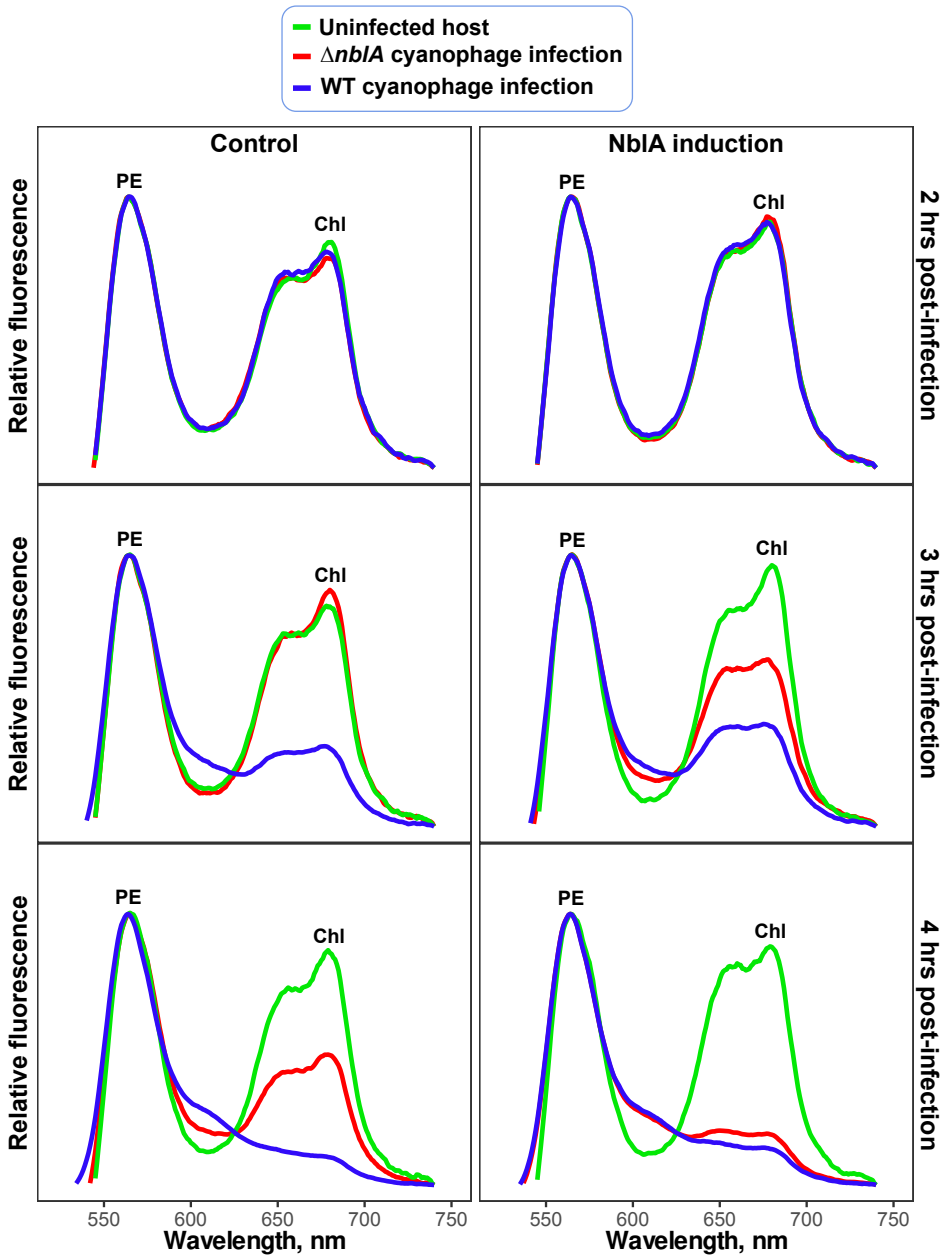
Extended Data Fig. 1. Expression of the cyanophage *nblA* gene during S-TIP37 infection of *Synechococcus* sp. strain WH8109. a, Transcript levels of S-TIP37 cyanophage *nblA* and DNA polymerase genes and one of the host *nblA* genes during infection with WT cyanophage determined by qRT-PCR. Transcript 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 PEII and PEI 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 PEII), emission spectra were obtained and the relative height of the chlorophyll peak was calculated (see the panel to the left). 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.

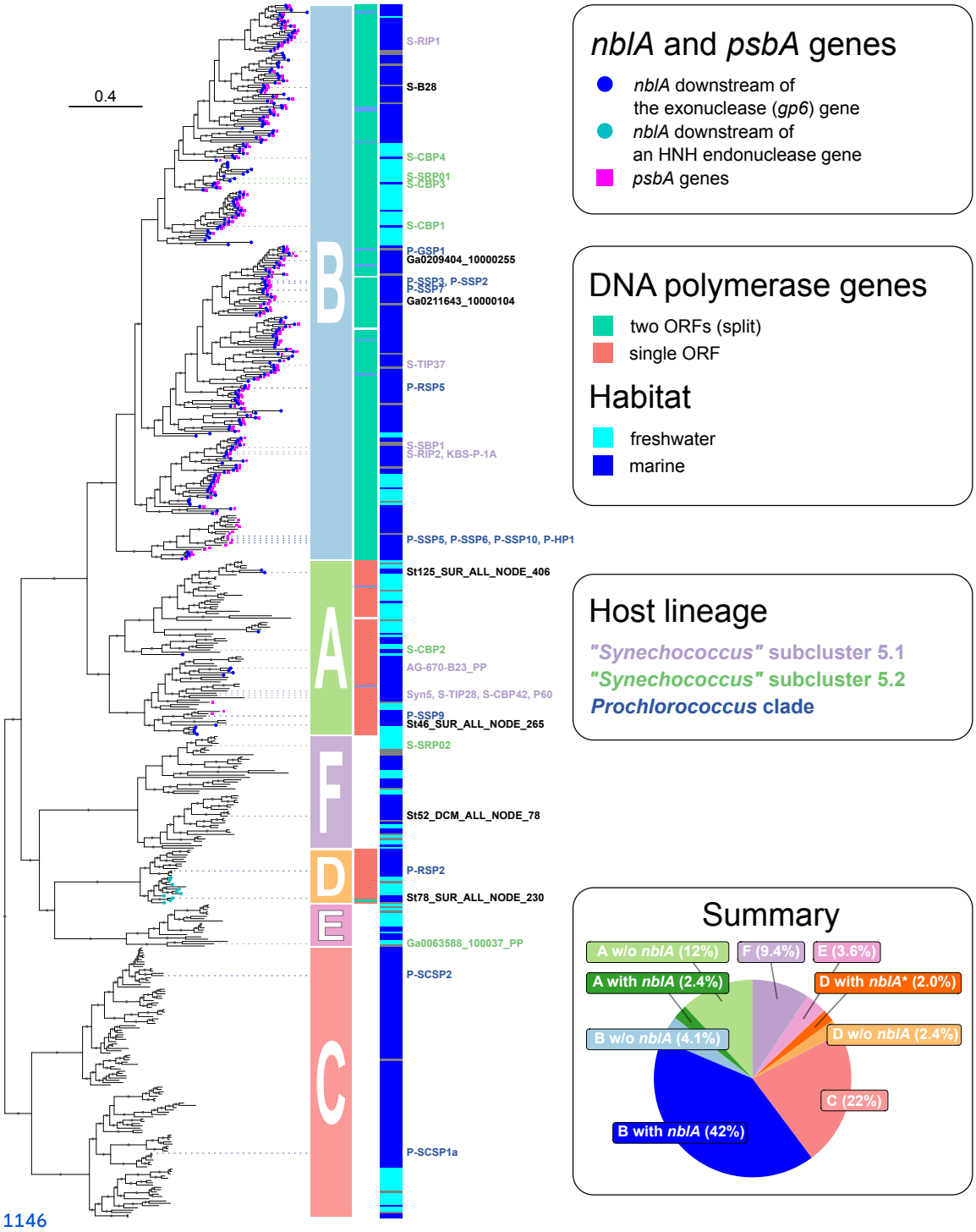


Extended Data Fig. 3. Fluorescence spectral analysis of fractionated PBSs. Fluorescence emission spectra at 77K of fractionated PBSs from uninfected *Synechococcus* sp. strain WH8109. The lower density, upper band (see Fig. 2c) contains disassembled PBS complexes, whereas the higher density, lower band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



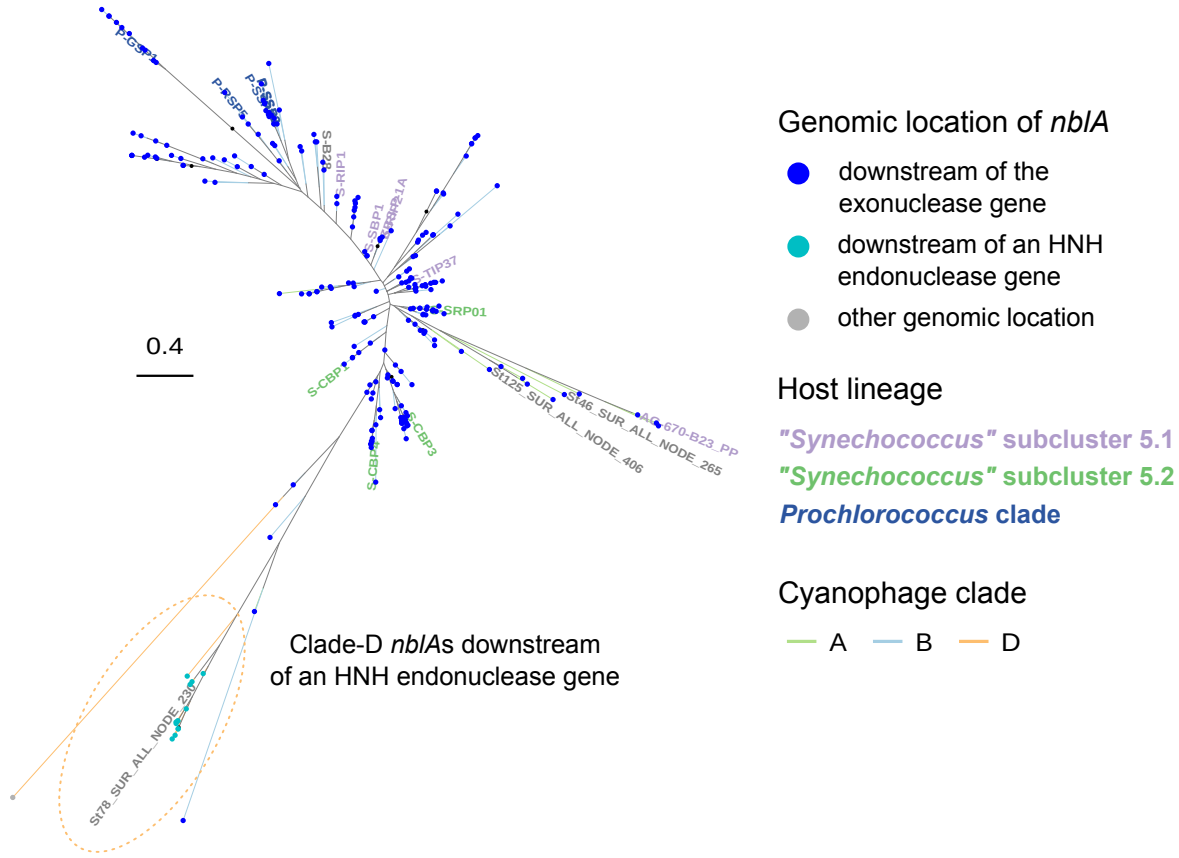
1138

1139 **Extended Data Fig. 4. Ectopic expression of cyanophage NblA in *Synechococcus***
1140 **sp. strain WH8109 during infection.** Fluorescence emission spectra of uninfected cells
1141 (green), cyanobacteria infected by the WT cyanophage (blue) and by the Δ*nblA* mutant
1142 cyanophage (red) at different time points post-infection. Control experiments (the panels
1143 to the left) were performed without the addition of the inducer. Ectopic expression was
1144 induced in the host 8 h prior to addition of the phages (NblA induction, the panels to the
1145 right). Measurements were taken at room temperature and normalized using the PE peak.

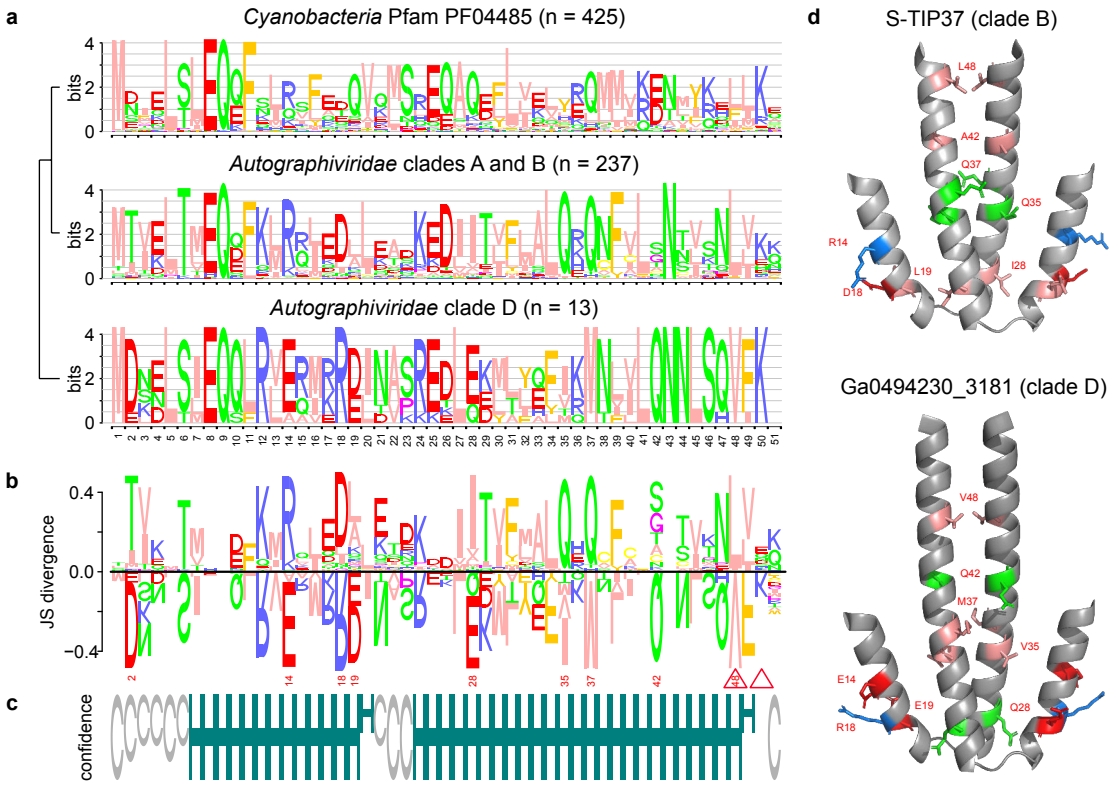


Extended Data Fig. 5. Phylogeny and distribution of *nblA* genes among T7-like cyanophages. Phylogenetic tree was built based on concatenation of nine core genes coding for: primase-helicase, exonuclease, portal protein (head-tail adaptor), head assembly, major capsid protein, tail tubular proteins A and B, and small and large terminase subunits. Phage genomes used in the phylogeny were recruited among cultured T7-like cyanophages and from environmental sources: Global Ocean Virome v. 2

assemblies and JGI IMG/VR v. 4.1. The environmental genomes were recruited by searching for the exonuclease gene, selecting genomes at least 30 kb long, dereplicating them at 0.95 ANImf identity level and requiring the presence of at least eight core genes. The tree was outgroup-rooted by inclusion of coliphage T7, pelagiphage HTVC011P, pelagiphage HTVC019P and environmental phages falling outside of the cyanophage cluster. Phages were categorized into previously recognized clades A (P60-like), B (S-TIP37-like) and C (P-SCSP1a/2-like) and newly-delineated clades D (P-RSP2-like), E (Ga0063588_100037_PP-like, no cultured representatives) and F (S-SRP02-like). Presence and type of DNA polymerase genes is indicated with the clades dominated by phages falling in one of the following categories: no DNA polymerase gene (clades C, E, F), DNA polymerase coded in a single ORF (clades A and D) and a split DNA polymerase gene[86] with the exonuclease domain coded in a separate ORF (clade B). Clade representatives are colored with labels according to the host's phylogenetic affiliation if known. Note that each of the three picocyanobacterial lineages represents an assemblage of several genus-level taxa. The incidence of *nblA* (circles) and *psbA* (squares) genes is indicated with symbols at the tips of the corresponding branches and the *nblA* genes of two categories are differentiated: those located downstream of the exonuclease gene (as in S-TIP37) and the *nblA* gene downstream of an HNH endonuclease gene (in a subclade of clade D). No clade C, E or F genomes were found to code for *nblA* and are thus not separated into two groups in the summary pie chart. Solid circles mark branches with ultrafast bootstrap support of ≥95. See Suppl. Data File 7 for the source data.



Extended Data Fig. 6. Phylogenetic tree of *nblA* genes from the *Autographiviridae*. *nblA* genes are categorized by their genomic location, host lineages (when known) and the corresponding cyanophage clade assignments (see Extended Data Fig. 5). The group of genes representing *nblA*s downstream of the endonuclease gene from clade D cyanophages is highlighted. Solid circles indicate edges with ultrafast bootstrap support of 100. See Suppl. Data File 4 for the source data.



Extended Data Fig. 7. NblA proteins from the *Autographiviridae*. a, Sequence logos of the two types of NblA proteins encoded in genomes of T7-like cyanophages compared to the cyanobacterial NblAs matching Pfam profile PF04485. **b**, Difference logo for the NblAs from clades A and B (top) *relative to* clade D (bottom). Numbers in red below the logo denote positions with Jensen–Shannon divergence >0.9 and triangles indicate positions known to mediate phycobiliprotein binding[32]: 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). **c**, Secondary structure predictions for the corresponding positions of the trimmed alignment. C, turn, H, helix. The height reflects confidence scores. **d**, Dimeric structures of representative *Autographiviridae* NblAs from the two types with highlighted residues corresponding to the positions indicated in (**b**). Terminal unstructured regions are not shown.