Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism

Luke R. Thompson[a,1], Qinglu Zeng[b,1], Libusha Kelly[b], Katherine H. Huang[b], Alexander U. Singer[c], JoAnne Stubbe[a,d,2], and Sallie W. Chisholm[a,b,2]

Departments of [a]Biology and [b]Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario Canada M5G 1L6; and [d]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by James M. Tiedje, Michigan State University, East Lansing, MI, and approved July 29, 2011 (received for review February 15, 2011)

Cyanophages infecting the marine cyanobacteria *Prochlorococcus* and *Synechococcus* encode and express genes for the photosynthetic light reactions. Sequenced cyanophage genomes lack Calvin cycle genes, however, suggesting that photosynthetic energy harvested via phage proteins is not used for carbon fixation. We report here that cyanophages carry and express a Calvin cycle inhibitor, CP12, whose host homologue directs carbon flux from the Calvin cycle to the pentose phosphate pathway (PPP). Phage CP12 was coexpressed with phage genes involved in the light reactions, deoxynucleotide biosynthesis, and the PPP, including a transaldolase gene that is the most prevalent PPP gene in cyanophages. Phage transaldolase was purified to homogeneity from several strains and shown to be functional in vitro, suggesting that it might facilitate increased flux through this key reaction in the host PPP, augmenting production of NADPH and ribose 5-phosphate. Kinetic measurements of phage and host transaldolases revealed that the phage enzymes have k_{cat}/K_m values only approximately one third of the corresponding host enzymes. The lower efficiency of phage transaldolase may be a tradeoff for other selective advantages such as reduced gene size: we show that more than half of host-like cyanophage genes are significantly shorter than their host homologues. Consistent with decreased Calvin cycle activity and increased PPP and light reaction activity under infection, the host NADPH/NADP ratio increased two-fold in infected cells. We propose that phage-augmented NADPH production fuels deoxynucleotide biosynthesis for phage replication, and that the selection pressures molding phage genomes involve fitness advantages conferred through mobilization of host energy stores.

coevolution | photosynthesis | virus

The use of host-like metabolic genes by viruses is a central theme in the coevolution of viruses and their hosts, and the particular metabolic genes carried by viruses provide clues to the mechanisms of viral replication. Bacteriophages typically carry multiple genes for nucleic acid biosynthesis and replication (1, 2), critical processes for replication of the phage genome. Cyanophages (cyanobacterial dsDNA bacteriophages) infecting the marine cyanobacteria *Prochlorococcus* and *Synechococcus* carry a much richer cache of metabolic genes than most sequenced phages, encoding elements of photosynthesis (3), the pentose phosphate pathway (PPP) (4), and phosphate acquisition (5). Such genes have been termed "auxiliary metabolic genes" (6) because they are thought to provide supplemental support to key steps in host metabolism of significance to phage, thereby fostering a more successful infection.

Photosynthetic light reaction genes in phage genomes have attracted particular attention because photosystems are sophisticated macromolecular complexes with functions in core cyanobacterial metabolism, and they are not traditionally associated with the phage infection process. Nevertheless, physiological studies of cyanophage infection have shown that a phage copy of *psbA*, which encodes the D1 protein essential for photosystem II (PSII), is transcribed and translated during infection (7–9). This result suggested that phage gene products may integrate into host photosystems and contribute to host photosynthesis during infection (7), possibly fueling phage dNTP biosynthesis (10, 11). It has long been recognized that both light and the photosynthetic light reactions are necessary for optimal cyanophage production (7, 12–14). However, in the host, the energy (i.e., ATP) and reducing equivalents (i.e., NADPH) generated by the light reactions are typically used for carbon fixation by the Calvin cycle. The presence of genes encoding Calvin cycle enzymes has not been reported for any sequenced cyanophages (5, 15–21), suggesting that phage use the ATP and NADPH generated by the light reactions of photosynthesis for other functions.

Clues to the role of photosynthesis during infection have been provided by our observation that some cyanophages carry a gene for CP12 (21), an inhibitor of the Calvin cycle. CP12 is an intrinsically unstructured protein that in plants (22) and cyanobacteria (23) binds and inhibits two key enzymes in the Calvin cycle (phosphoribulokinase, PRK, *prkB* gene; and glyceraldehyde-3-phosphate dehydrogenase, GAPDH, *gap2* gene). CP12 binding has the effect of directing carbon flux away from glucose synthesis and toward the PPP, where glucose 6-P is oxidized by NADP to ribose 5-P and NADPH (Fig. 1*A*). We were prompted to look for CP12 in cyanophages after our studies of the light–dark cycle of *Prochlorococcus* (24) indicated its probable role in regulating the Calvin cycle and PPP in these cyanobacteria. We showed that host *cp12* is maximally expressed at night in *Prochlorococcus* (24), consistent with its functioning to direct carbon flux away from the Calvin cycle and toward the PPP in the dark.

Cyanophages that encode CP12 also encode proteins for the PPP, the light reactions, and dNTP biosynthesis (21). Of particular interest among these phage-encoded functions is a putative transaldolase, widely represented in cyanophage genomes (5), which could boost the host PPP during infection. As a whole, the set of carbon metabolism genes carried by cyanophages is consistent with the hypothesis that, under phage infection, the ATP and NADPH produced by photosynthesis are not used by the Calvin cycle but rather are used to fuel phage dNTP biosynthesis. We now report our genomics and metagenomics studies, our enzymological studies of transaldolase, and our analyses of gene

Author contributions: L.R.T., Q.Z., L.K., K.H.H., J.S., and S.W.C. designed research; L.R.T., Q.Z., L.K., K.H.H., and A.U.S. performed research; L.R.T., Q.Z., L.K., K.H.H., and A.U.S. analyzed data; and L.R.T., J.S., and S.W.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3HJZ). The sequences reported in this paper have been deposited in the GenBank database [GU071107 (P-SSP2), GU071104 (P-HP1), and GU071102 (P-RSP5)].

[1]L.R.T. and Q.Z. contributed equally to this work.

[2]To whom correspondence may be addressed: chisholm@mit.edu or stubbe@mit.edu.

See Author Summary on page 16147.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1102164108/-/DCSupplemental.

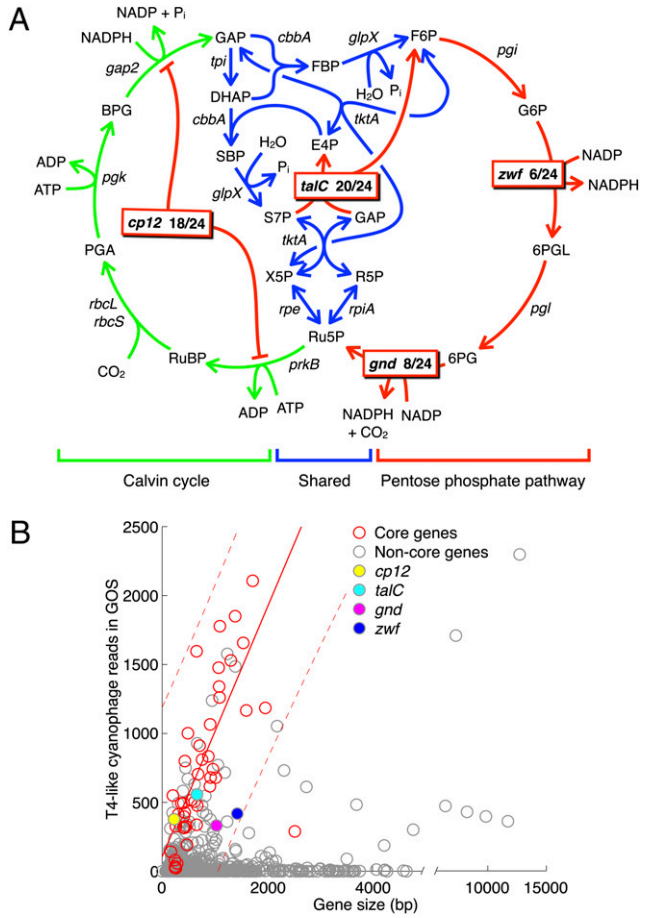

Fig. 1. (*A*) The PPP and Calvin cycle in cyanobacteria showing genes carried by cyanophages. Genes found in phages are boxed and denoted with the number of genomes among 24 in which they are found. CP12 (*cp12*) is grouped with the PPP, because it shuts off the competing Calvin cycle, inhibiting PRK (*prkB*) and GAPDH (*gap2*). Both pathways run clockwise, with the shared reactions (blue) running downward (Calvin cycle) or upward (PPP), and therefore the two pathways are not expected to run concurrently. Transaldolase (*talC* in phage, *talB* in host) is the only reaction in the intersecting part of the pathways that is needed only in one direction. Arrows show proposed directionality of the reactions, and one-way arrows do not necessarily indicate irreversibility. P, phosphate; (genes) *cbbA*, fructose-1,6-bis-P/sedoheptulose-1,7-bis-P aldolase; *glpX*, fructose-1,6/sedoheptulose-1,7-bisphosphatase; *pgi*, P-glucose isomerase; *pgk*, P-glycerate kinase; *pgl*, 6-P-gluconolactonase; *prkB*, P-ribulokinase; *rbcLS*, ribulose-1,5-bis-P carboxylase/oxygenase; *rpe*, ribulose-5-P epimerase; *rpiA*, ribulose-5-P isomerase; *tktA*, transketolase; *tpi*, triose-P isomerase; (substrates) BPG, 2,3-bis-P-glycerate; DHAP, DHA P; E4P, erythrose 4-P; FBP, fructose 1,6-bis-P; F6P, fructose 6-P; GAP, glyceraldehyde 3-P; G6P, glucose 6-P; PGA, 3-P-glyceric acid; R5P, ribose 5-P; RuBP, ribulose 1,5-bis-P; Ru5P, ribulose 5-P; SBP, sedoheptulose 1,7-bis-P; 6PG, 6-P-gluconate; 6PGL, 6-P-gluconolactone; S7P, sedoheptulose 7-P; and X5P, xylulose 5-P. (*B*) Relative abundance of T4-like cyanophage genes in the GOS database. For each gene found in a T4-like cyanophage (red circles represent genes found in all sequenced genomes, i.e., core genes, and gray circles represent noncore genes), the number of times it was observed as a sequence read is plotted as a function of gene size. Core genes show a linear increase with gene size in the number of observed reads, as expected, as the likelihood of cloning and sequencing fragments of larger genes is greater than for smaller genes. The linear regression (with 95% confidence intervals) of this pattern is shown for core genes.

expression and cellular redox balance in uninfected and infected cells, which support this hypothesis.

Results and Discussion

CP12 and PPP Genes in Cultured and Wild Cyanophages. We first examined the genetic potential of all sequenced cyanophages for photosynthesis and carbon metabolism and then measured the frequency of these genes in wild cyanophage populations. Twenty-four cultured cyanophage genomes representing the three major morphotypes (T4-like myoviruses, T7-like podoviruses, and siphoviruses) were analyzed for the presence and frequency of all host photosynthesis and carbon metabolism genes (Table S1). Notably, a focused search (*Methods*) for all Calvin cycle genes confirmed their absence in all 24 cyanophage genomes, supporting the suspicion that cyanophage do not exploit this part of host metabolism. The sequence analysis did reveal, however, the presence of the Calvin cycle inhibitor gene, *cp12*, in 18 of 24 cyanophages (Fig. 1*A* and Table S1), including all three major morphotypes. All cyanophage CP12 sequences (Fig. S1) contain a conserved C-terminal CxxxPxxxxC motif found in all cyanobacterial and nearly all plant CP12 sequences (25) and a pattern of hydrophilic residues, characteristic of intrinsically unstructured proteins (26). Unequivocal determination of the function of CP12 in marine cyanobacteria and cyanophage, however, awaits in vitro characterization of these proteins.

In addition to CP12, whose presence in phage has implications for promoting PPP flux in the host, three genes for PPP enzymes reported previously (4, 18) were found: transaldolase (*talC*), glucose-6-phosphate dehydrogenase (*zwf*), and 6-phosphogluconate dehydrogenase (*gnd*; Fig. 1*A*). We reveal here that these genes and *cp12* display a hierarchical pattern of gene abundance (Table S1), which, as we discuss later, provides hints to the selection pressures on phage. Most abundant are *talC* and *cp12*, found in 20 and 18 of the 24 genomes, respectively, followed by *gnd* and *zwf*, found in eight and six, respectively (Fig. 1*A* and Table S1). *cp12* is found in most T4-like phages and only once in available T7-like phages and siphoviruses; *talC* is widespread in both T4-like and T7-like phages; and *gnd* and *zwf* are found only in T4-like phages isolated on *Synechococcus* (Table S1).

The abundance of phage DNA fragments in environmental metagenomic databases, such as the Global Ocean Sampling (GOS) database (27), allowed us to analyze the prevalence of CP12 and PPP genes in wild phage populations, focusing on T4-like cyanophages, for which there are the most reference genomes (17) (Table S1). Interestingly, the prevalence pattern of *cp12* and the PPP genes among cultured phages was largely recapitulated in the GOS metagenome (Fig. 1*B*). *cp12* and *talC* (found in 16/17 T4-like cyanophage genomes) had abundances similar to "core genes," i.e., genes found in every T4-like cyanophage genome, whereas *gnd* and *zwf* (found in 8/17 and 6/17) were less abundant than most core genes (Fig. 1*B*). This trend was not dependent on the number of different orthologues used to recruit sequences from the GOS database (*SI Methods*). As a side note, our metagenomic analysis was designed to probe only for PPP/Calvin cycle-related genes identified in cultured cyanophage genomes, and therefore we did not find evidence of fructose-1,6-bisphosphate aldolase, shared by the two pathways, which was recently reported in GOS scaffolds of suspected cyanophage origin (28) and represents an additional phage enzyme for the host PPP.

The emergent prevalence pattern suggests not only that the host PPP is important to infecting phage, but that some steps in the PPP may be bottlenecks more often than others, and therefore phage genes for those steps are under stronger selection. Transaldolase especially may be a key metabolic bottleneck under infection. Transaldolase, although reversible, is the only reaction in the nonoxidative PPP (blue plus transaldolase, Fig. 1*A*) that is not also shared by the Calvin cycle, required only in the direction that regenerates glucose 6-P for NADPH production (blue and red, clockwise, Fig. 1*A*). In our previous studies over the light–dark cycle of host *Prochlorococcus* MED4, all of the shared Calvin cycle/PPP genes had maximal mRNA levels in the morning, whereas the host transaldolase gene had maximal mRNA in the evening, with the other PPP genes (24).

By 10:00 AM, the level of host transaldolase protein had decreased to half its maximum level, whereas all other proteins in the PPP and Calvin cycle remained relatively constant over this period (29). Thus, if phage exploit host photosynthetic light reactions by infecting during the day, expression of their own transaldolase may be critical to overcoming low levels of the host enzyme at this time.

Coordinated Transcription of Phage CP12 and Metabolic Enzymes. In support of our hypothesis, the expression of *talC* has already been demonstrated for the T7-like cyanophage P-SSP7 (9) during infection. We now report the expression of phage *cp12*, *zwf*, and *gnd* during infection, together with *psbA*, encoding PSII protein D1, and *nrdA*/*nrdB*, encoding ribonucleotide reductase (RNR), the enzyme responsible for synthesis of deoxynucleotides from nucleotides. Our hypothesis about the coordinated activity of auxiliary metabolic genes for dNTP biosynthesis suggests that all these genes should have similar expression profiles and be expressed early in the infection cycle. To explore this, we examined gene expression patterns in two yet-unstudied T4-like myoviruses: Syn9, infecting *Synechococcus* WH8109 (Fig. 2); and P-HM2, infecting *Prochlorococcus* MED4 (Fig. S2 *A* and *C*). Both phages carry *cp12*, and in both phages it was coexpressed with *talC*, *psbA*, and *nrdA*/*nrdB*. In Syn9, these genes were also coexpressed with *zwf* and *gnd* (Fig. 2). As expected, each of these metabolic genes was expressed with the T4-like "early genes," which tend to encode enzymes for phage replication, rather than with T4-like "late genes," which usually encode structural proteins (2), thus positioning the gene products to participate in the takeover of host metabolism.

The coordinated early expression of genes involving host photosynthesis, central carbon metabolism, and dNTP biosynthesis in these T4-like phages is reminiscent of the pattern observed in the T7-like phage P-SSP7 (9). However, in contrast to T4-like phages, in T7-like phages, gene expression is dictated by gene order rather than by early and late promoters (30). In P-SSP7, *talC* is the last gene in the genome but was expressed much earlier than expected based on its terminal position, so that it was coexpressed with *psbA* and *nrdJ* (i.e., RNR) (9). Thus, although the mechanisms that dictate timing of gene expression in these two phage types are different, the resultant coordination of expression is similar.

What differs among cyanophages, then, is not the timing of metabolic gene expression—evidence suggests they are generally expressed early—but which metabolic genes they carry. For example, P-SSP7 and, in total, five of six T7-like cyanophages do not carry *cp12*, and no T7-like phages have *zwf* or *gnd* (Table S1). The PPP gene that emerges as most important for phage is transaldolase (*talC*). Although still not universal, *talC* is found in more cyanophages (20 of 24) than any other PPP gene. We therefore conducted a detailed enzymological study of phage and host transaldolases to confirm their functions and measure their kinetic parameters. We asked whether *talC* encodes a functional transaldolase, and if so, whether it is kinetically more efficient than the host transaldolase.

Comparative Kinetics of Phage and Host Transaldolases. The cyanophage transaldolase (TalC) differs significantly in sequence from the host transaldolase (TalB), which led us to question not only why phages encode transaldolase, but why they encode one so different from the host enzyme. Multiple sequence alignments reveal that TalC has a number of deletions relative to TalB (Fig. S3*A*) but no change in the active site residues essential for catalysis (Fig. S3*A*) (31, 32). TalC more closely resembles the *Escherichia coli* fructose-6-phosphate (F6P) aldolase, FsaA (33), than host TalB (Fig. S3 *A* and *B*). Therefore, we sought to confirm that phage TalCs have transaldolase activity, not F6P aldolase activity. We then compared their kinetic properties to

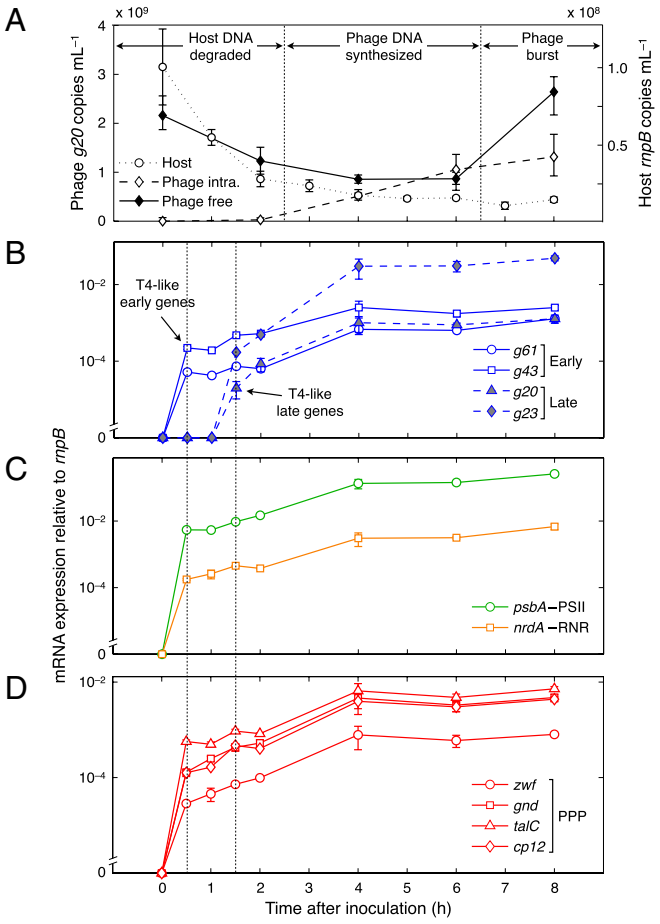

Fig. 2. Cyanophage Syn9 infection of *Synechococcus* WH8109, showing expression of a phage gene for the Calvin cycle inhibitor CP12 alongside other phage metabolic genes. Error bars represent SDs of three technical and three biological replicates. (*A*) Degradation of host gDNA and replication of phage gDNA (free and intracellular) as determined by qPCR of chromosomal genes in host (*rnpB*) and phage (*g20*). (*B*) mRNA of T4-like early genes *g61* (DNA primase) and *g43* (DNA polymerase) and late genes *g20* (portal protein) and *g23* (major coat protein). (*C*) mRNA of PSII gene *psbA* (D1 subunit) and RNR gene *nrdA* (α subunit). (*D*) mRNA of PPP genes *zwf* (glucose-6-P dehydrogenase), *gnd* (6-P-gluconate dehydrogenase), *talC* (transaldolase), and *cp12* (Calvin cycle inhibitor CP12).

those of their corresponding host TalBs, as their structural differences might be accompanied by kinetic differences. We anticipated that phage transaldolase might possess kinetic advantages relative to host transaldolase and thus augment the host's PPP during infection.

Purification and assay. As a first step in TalC and TalB characterization, the genes from three phages and their corresponding hosts were cloned and the proteins expressed and purified to approximately 90% homogeneity (Fig. S4*A*). Many different vectors and expression systems were investigated before settling on N- or C-terminal 6×His-tagged constructs (Tables S3 and S4) to facilitate purification and a cold-adapted expression strain wherein small amounts of soluble protein could be produced (*SI Methods*). The first task was to establish whether TalC is a transaldolase or a F6P aldolase. As shown in Fig. 1*A* and **1**, transaldolase is central to the PPP and catalyzes the reversible, three-carbon transfer of dihydroxyacetone (DHA) from F6P (C1-3 of F6P) to erythrose 4-phosphate (E4P) to generate sedoheptulose 7-phosphate (S7P) and glyceraldehyde 3-phosphate (GAP). In contrast, F6P aldolase catalyzes the conversion of F6P to GAP and

DHA. Thus, for transaldolase activity, both F6P and E4P are required, whereas for F6P aldolase, only the former is required (33).

$$\text{F6P} + \text{E4P} \rightleftharpoons \text{GAP} + \text{S7P} \quad [1]$$

The assay for transaldolase activity (34) involves two coupling enzymes: triosephosphate isomerase (TPI) and glycerol-3-phosphate dehydrogenase (G3PDH). TPI converts GAP (as shown in **1**) to DHA phosphate, which is then reduced to glycerol 3-phosphate by G3PDH using an NADH cofactor, whose oxidation is monitored spectrophotometrically by a decrease in A_{340}. The results of these experiments (Fig. S4*B*) reveal that all three TalCs and all three TalBs required E4P for turnover and are thus transaldolases. This result is consistent with a role for TalC in the host PPP (Fig. 1*A*), with **1** proceeding from right to left.

Structures of TalC and TalB. Further comparison of host and phage enzymes has been possible by structural determination of *Prochlorococcus* MIT9312 TalB (1.90-Å resolution by X-ray crystallography; Table S5) and homology modeling of cyanophage P-SSP7 TalC using sequence alignments and the *Thermotoga maritima* TalC structure [Protein Data Bank (PDB) ID no. 1VPX; 41% identity]. An overlay of the two structures (*SI Methods*) reveals a superimposable α_8/β_8-barrel including the conserved active site residues (Fig. S3*C*) previously demonstrated to be important in catalysis (31, 32). The major distinguishing feature between these transaldolases is the number and orientation of exterior α-helices, with TalB containing more external α-helices relative to TalC's more compact structure. To reveal the quaternary structures of TalB and TalC, the same two representatives (MIT9312 TalB and P-SSP7 TalC) were examined by size-exclusion chromatography (Fig. S3*D*). Comparison of the elution profiles with a standard curve generated from globular proteins of known molecular weight revealed that TalB appears to form a monomer and TalC a pentamer. Most other TalBs are reported to be dimers in solution (35), but it is interesting to note that a monomeric state is also observed in our crystal structure. Most other TalCs form decamers in solution (35), consisting of a dimer of doughnut-shaped pentamers (31). Thus, both host and phage transaldolases have unusual quaternary structures.

Kinetic properties of TalC and TalB. Finally, to assess the potential of phage transaldolase to substitute for or even outperform the host transaldolase, we compared the kinetic parameters of the three phage TalCs and three host TalBs by using the coupled assay. The results are summarized in Table 1. In general, the k_{cat} and k_{cat} / K_m values for host transaldolases are approximately threefold higher than for phage transaldolases. Our hypothesis was that the TalCs might have higher efficiencies than the TalBs. However, in all three phage–host pairs of TalC and TalB, this was shown not to be the case. We consider three explanations for why a phage would carry a kinetically less efficient enzyme than its host. First, for phage, the acquisition and use of host-like metabolic genes must always involve fitness tradeoffs: the gene may contribute to a more successful infection, but it also increases the size of the genome that must be replicated (11). Larger gene size could reduce phage fitness, as phage genome size is partially limited by capsid size (36). The decreased length of the phage *talC* gene (average, 659 bp) relative to the host *talB* gene (average, 1,098 bp) may be a result of this tradeoff. Supporting this argument, many of the phage metabolic genes with host orthologues are shorter than their host orthologues (Fig. 3 and Table S6). More specifically, comparing phage and host gene sizes separately for each of 24 shared orthologues (*SI Methods*), 14 orthologues were shorter in phages than in hosts, one orthologue was longer in phages than in hosts, and nine orthologues were not statistically different in size (Fig. 3 and Table S6). This is consistent with the general trend that phage genes are, on average, shorter than host genes (37, 38), but this pairwise comparison between phage and host genes with the same functions is more directly relevant to the evolutionary tradeoffs in encoding a particular gene function. Second, phage transaldolase gene dosage may exceed that of the host and compensate for a less active enzyme. Many copies of the phage genome are synthesized during the latent period of infection, each encoding and potentially expressing transaldolase. Third, as we have already mentioned, changes in host TalB levels over the light–dark cycle could render the host PPP less active during the day, and a separate phage TalC could potentially circumvent this regulation, regardless of its activity. Unfortunately, the genetic tools required to address the effects of differing enzyme properties in vivo are not yet available for these phages and hosts. Whatever the underlying pressures, phage transaldolase is functional and has been conserved across a wide range of cyanophages, consistent with a key role in host metabolism under infection.

Proposed Model of Cyanobacterial Metabolism Under Phage Infection. Expression of phage genes for the PPP, the light reactions, and inhibition of the Calvin cycle early in the infection process—combined with evidence that at least one of the gene products (transaldolase) is functional in vitro—implies that these metabolic pathways are augmented in infected cells (Fig. 1*A*). We hypothesize that, when a cyanophage infects a host cell, the Calvin cycle is down-regulated by phage-encoded CP12, preventing energy and reducing power produced by the phage-augmented light reactions from being used to fix carbon dioxide, such that phage infection decouples the light reactions from the Calvin cycle. Thus, two pathways that under the light–dark cycle are normally offset by 12 h—the light reactions and the PPP—potentially occur simultaneously in the host cell. We propose that the combined ATP and NADPH from these processes (along with ribose 5-P from the PPP) fuels phage dNTP biosynthesis, particularly via the phage-encoded RNR, which requires NADPH as the terminal electron donor to reduce NDPs to dNDPs.

Table 1. Comparison of kinetic parameters of three TalBs and three TalCs

TalB/TalC	k_{cat}, s^{-1}	K_m, mM		k_{cat}/K_m, s^{-1} mM^{-1}	
		Fructose 6-P	Erythrose 4-P	Fructose 6-P	Erythrose 4-P
Prochlorococcus NATL2A TalB	15.2 ± 0.7	1.1 ± 0.2	0.11 ± 0.02	13.6 ± 2.5	134 ± 29
Prochlorococcus MED4 TalB	14.9 ± 0.4	1.5 ± 0.2	0.15 ± 0.02	9.8 ± 1.2	103 ± 14
Prochlorococcus MIT9312 TalB	20.8 ± 1.3	1.0 ± 0.1	0.10 ± 0.04	20.8 ± 2.3	206 ± 75
Cyanophage P-SSM2 TalC	3.6 ± 0.3	0.7 ± 0.1	0.08 ± 0.01	5.4 ± 0.7	48 ± 9
Cyanophage P-SSM4 TalC	5.6 ± 0.4	1.3 ± 0.2	0.20 ± 0.05	4.5 ± 0.9	29 ± 7
Cyanophage P-SSP7 TalC	5.9 ± 0.2	1.6 ± 0.2	0.07 ± 0.01	3.6 ± 0.5	86 ± 18

Kinetic parameters (k_{cat}, K_m, and k_{cat} / K_m) were measured at 25 °C in 50 mM Gly-Gly (pH 8.0), 15 mM $MgCl_2$, 200 μM NADH, 0.6 U TPI, 0.06 U G3PDH, 0.1–10.0 mM F6P, and 0.01–1.00 mM E4P. *Prochlorococcus* TalB assays also contained 10 mM DTT.

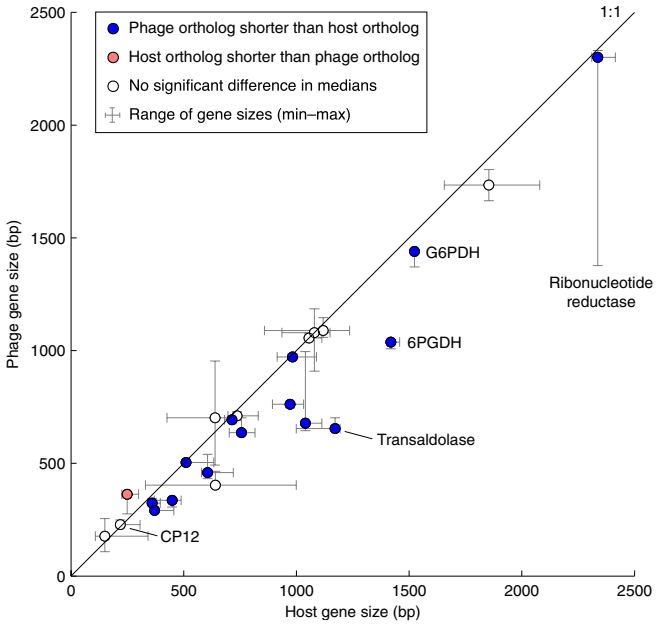

Fig. 3. Comparison of phage and host gene sizes for shared orthologues. For each gene orthologue found in at least three cyanophage genomes and at least three *Prochlorococcus* or *Synechococcus* genomes, the median gene size across all phages and across all hosts is plotted. Bars represent the range of sizes observed. The line shows a 1:1 ratio, whereby phage and host gene sizes are identical. Median gene sizes for each host–phage pair were compared by using the Mann–Whitney–Wilcoxon test ($P < 0.05$) with correction for false discovery rate (*SI Methods*). Blue circles indicate genes that were shorter in phages than in hosts, red circles those that were shorter in hosts than in phages, and unfilled circles those that were statistically the same size in phages and hosts.

The hypothesis that cyanophage need to boost dNTP synthesis because dNTPs are limiting during infection is supported by differences in phage and host molecular composition and by host physiology. As described in Table S7, the protein/DNA ratio of a cyanobacterial host cell is 20 times the protein/DNA ratio of a T4 phage; therefore, the metabolic fluxes required for phage reproduction relative to host reproduction are biased toward DNA synthesis. However, the host chromosome (2 Mbp), even if completely digested, would be insufficient to supply all of the dNTPs needed by progeny phage: assuming an average burst size of 100 phage cell^{-1} [range of 40–150 phage cell^{-1} (39, 40)] and a phage genome size of 200 kbp phage^{-1} (Table S1), the host chromosome could supply only 10% of the dNTPs in the progeny phage. Phytoplankton living in low-nutrient environments are thought to have limited dNTP pools (10, 41). Indeed, for many marine viruses, the majority of phage dNTPs are derived from the host chromosome (41) and host genome size limits burst size in dNTP equivalents (10). Modeling of the phage infection process (11) suggests that phage infecting hosts with small genomes like *Prochlorococcus* are more likely to encode auxiliary metabolic genes for dNTP synthesis. Cyanophages are able to yield significantly higher burst sizes than predicted by their host genomes (10), possibly resulting from an ability to help the host produce large quantities of dNTPs de novo.

Dynamics of NADPH and Its Derivatives. In our model of phage-augmented nucleotide biosynthesis, the shared Calvin cycle/PPP reactions are proposed to run in the direction of the PPP (clockwise in Fig. 1*A*) such that ribose 5-P provides the sugars for new dNTPs. However, much of the pentose phosphate produced is recycled back to glucose 6-P for additional NADPH production. Meanwhile, the light reactions of photosynthesis produce ATP and additional NADPH for making dNTPs. Our model therefore makes the following predictions: first, phage-infected cells produce more NADPH than uninfected cells; second, light is required for optimal NADPH production under infection; and third, phage-encoded proteins are partially responsible for this increase in NADPH production. To test these predictions, we grew *Prochlorococcus* MED4 cultures in constant light, inoculated with cyanophage P-HM2 or spent medium, and returned to constant light or shifted to darkness. At regular intervals following inoculation, we measured ratios of NADPH and its derivatives (NADP, NADH, NAD) by using an enzymatic assay (23), and we measured gene expression and phage replication dynamics by quantitative PCR (qPCR). Additionally, to better understand what phage in the wild might encounter when infecting cells during the day or night, we synchronized *Prochlorococcus* MED4 on a 24-h light–dark cycle, took samples at regular intervals over a period of 48 h, and measured ratios of NADPH and its derivatives by using the enzymatic assay. We now describe the results of these experiments and compare them with those predicted by our model.

First, our model predicts that, in order for infected cells to increase de novo production of dNTPs, they require increased NADPH, which provides reducing equivalents for RNR (42). The NADPH/NADP ratio should therefore increase in infected cells, as phage shift metabolism to provide more NADPH for nucleotide reduction. In accordance with our prediction, the NADP(H) pool became more reduced during infection in the light relative to uninfected cells, with the NADPH/NADP ratio increasing from approximately 0.7 to approximately 1.3 over the first 6 h and then remaining unchanged (Fig. 4). The elevated NADPH/NADP ratio indicates an even greater increase in NADPH production than apparent from the ratio, as the added demand for dNTPs under infection requires increased NADPH consumption.

Second, as a major source of NADPH in cyanobacteria is the light reactions of photosynthesis, which are required for maximal phage production (7, 12), we predicted that the increase in NADPH/NADP ratio would be dampened or nullified under infection in the dark, accompanied by decreased phage production and gene expression. Indeed, the NADPH/NADP ratio decreased in infected and uninfected cultures in the dark (Fig. 4). The decrease in NADPH/NADP ratio in infected cultures in the dark coincided with stunted phage genomic DNA (gDNA) replication and host gDNA degradation in the dark (Fig. S2*B*) relative to infection in the light (Fig. S2*A*). We also observed more than 100-fold lower phage gene expression (on average for *cp12*, *talC*, *psbA*, *nrdA*/*nrdB*) in the dark (Fig. S2*D*) than in the light (Fig. S2*C*). Our results show that phage infecting in the dark are unable to degrade the host chromosome, transcribe their genes fully, or make progeny phage. The oxidizing effect of darkness on the NADPH/NADP ratio may help explain why photosynthesis is necessary for optimal phage production: without the photosynthetic light reactions to generate NADPH, glucose stores could quickly become exhausted, leaving no source of NADPH to fuel phage production.

To assess the applicability of these results to wild cyanophage populations, we measured ratios of NADPH and its derivatives in uninfected cultures synchronized on a 24-h light–dark cycle, which simulates the light conditions of the surface ocean. In our measurements over two 24-h light–dark cycles, uninfected cells were more reduced in the light and more oxidized in the dark (Fig. S5). Interestingly, the change was observed in the NADH/NAD ratio rather than in the NADPH/NADP ratio, and thus further work is needed to fully explain these redox dynamics. Nevertheless, it is clear that darkness causes *Prochlorococcus* to become more oxidized, and this may be an important factor limiting phage replication.

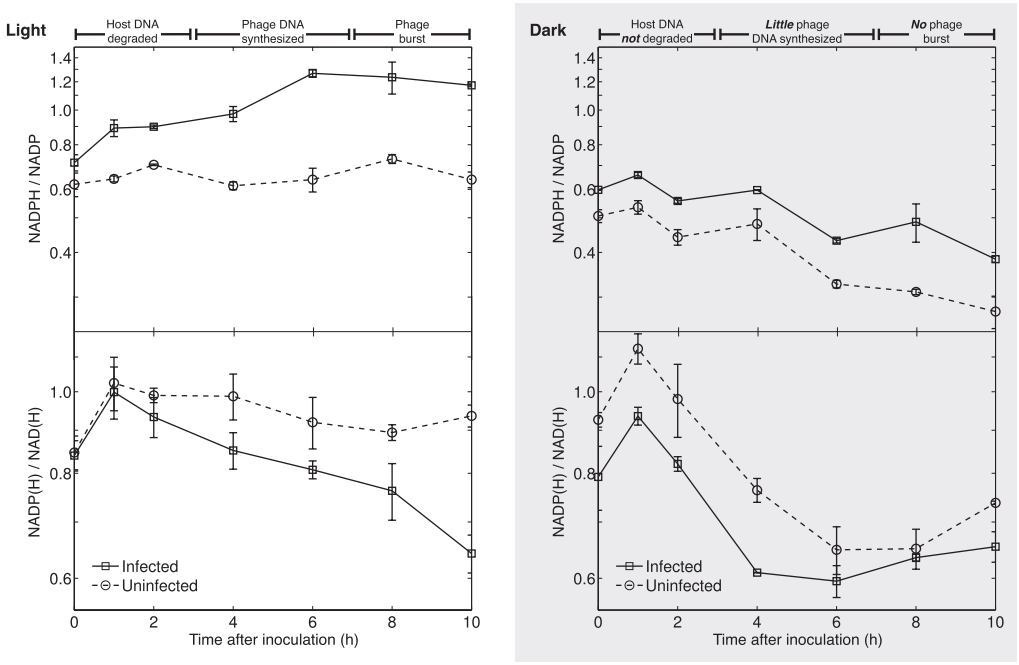

Fig. 4. Dynamics of the intracellular NADPH/NADP ratio (*Top*) and NADP(H)/NAD(H) ratio (*Bottom*) during infection of *Prochlorococcus* MED4 by cyanophage P-HM2 in the light or in the dark. Error bars represent SDs of two technical and two biological replicates.

Finally, because phage auxiliary metabolic genes are conserved in the phage and have roles in the PPP and its production of NADPH, the encoded proteins are predicted to be directly involved in NADPH production. A corollary of this prediction is that intracellular conditions under infection should be compatible with high activity of the phage-encoded enzymes and regulatory proteins. This is particularly relevant to CP12, which not only affects NADPH levels via its action on the Calvin cycle, but is itself regulated by levels of NADPH and its derivatives. Although the detailed regulatory mechanisms in cyanophage have not yet been determined, in plants and cyanobacteria, the complex of CP12 with PRK and GAPDH (Fig. 1*A*) is stabilized by NAD or NADH [i.e., NAD(H)] and destabilized by NADP or NADPH [i.e., NADP(H)] (22, 23, 43). [The ratio NADP(H)/NAD(H) therefore refers to the phosphorylation state of the pool comprising NADPH and its derivatives, whereas the ratios NADPH/NADP and NADH/NAD refer to the redox state of this pool.] Thus, we examined not only the ratio of NADPH/NADP but also the ratio of NADP(H)/NAD(H).

Based on our understanding of CP12 regulation in other systems (22, 23, 43), for phage CP12 to be active during infection, NAD(H) is predicted to increase relative to NADP(H) in infected cells, such that the NADP(H)/NAD(H) ratio would decrease. The NADP(H)/NAD(H) ratio is predicted to decrease likewise during nighttime of the day–night cycle of uninfected cells, as host CP12 is putatively active at night. As shown in Fig. 4, infected cells in the light indeed showed a decrease in NADP(H)/NAD(H), which was also seen in both infected and uninfected cells in the dark. Similarly, during the dark portion of the light–dark cycle, uninfected cells displayed a decrease in their NADP(H)/NAD(H) ratio (Fig. S5). Phage infection in the light therefore appears to mimic the effect of darkness on light–dark-synchronized *Prochlorococcus*, leading to conditions that favor CP12 activity [low NADP(H), high NAD(H)] and therefore dark metabolism, whereby the PPP is activated and the Calvin cycle is deactivated, resulting in NADPH production.

Further Considerations and Conclusions. The evidence presented here supports our hypothesis that cyanophage gene products redirect host metabolism to increase dNTP biosynthesis. Demonstrating the causal role of phage auxiliary metabolic genes in this process, however, will require genetic tools as yet unavailable for this host–phage system. Even when these tools are in hand, however, multipronged approaches like the one described here will be essential for developing a complete understanding of the forces that shape the coevolution of host and phage in the oceans.

Our analysis has focused on the need for NADPH to fuel dNTP biosynthesis for phage replication, but clearly, additional raw materials for making dNTPs could be limiting, such as ATP, carbon skeletons, nitrogen for purine and pyrimidine bases, and phosphate linkages. There appear to be cyanophage adaptations for many of these other dNTP inputs. ATP is needed at various steps in dNTP biosynthesis and could be produced via the photosynthetic electron transport chain, assisted by well documented phage-encoded gene products for photosystem II along with more recently discovered phage proteins for photosystem I (44) and NAD(P)H dehydrogenase (45). Ribose 5-P, the sugar skeleton for both purine and pyrimidine deoxynucleotides, is also produced by the PPP, and phage augmentation of this pathway as argued here likely produces ribose 5-P in parallel with NADPH. Nitrogen is needed for both purine and pyrimidine bases, and phage-induced nitrogen stress is suggested by the presence of multiple NtcA promoters in T4-like cyanophages, with phage possibly exploiting the host N-stress response system to drive gene expression (21). Finally, phage carry genes involved in host phosphate acquisition, which has clear implications for dNTP production in P-limited oceanic regions (21).

A phage reproductive strategy reflects selective pressures on both phage and host. As viewed by the genes phage carry, this strategy is the result of opposing selective forces: the cost of maintaining a gene in a size-limited genome versus the benefit of encoding a novel metabolic function. However, this strategy also reflects the physiology and environment of the host—in this case a photoautotroph living in low-nutrient waters. The fact that all three cyanophage types carry genes for the PPP despite having

evolved from completely separate phage lineages (46) points to the pressures brought on by their common host type and its austere habitat. This remarkable example of convergent evolution in phage demonstrates the unifying pressures of infecting an obligate photoautotrophic host with a very small genome, a host that alone may not be able to supply the DNA building blocks a phage needs to replicate effectively.

Methods

Sequences and Gene Annotation. Twenty-four genomes from cyanophages infecting *Prochlorococcus* and marine *Synechococcus* were included in the analyses (Table S1), three of which are being introduced in the present study. We searched the 24 cyanophage genomes for each Calvin cycle and PPP gene, along with known cyanophage genes in photosynthetic electron transport and nucleotide biosynthesis, using the genome annotations. To check for the possibility of uncalled or miscalled genes in the annotations, we also searched the 24 genomes by using TBLASTN with default parameters, an E-value cutoff of 10^{-5}, and using all *Prochlorococcus* or *Synechococcus* Calvin cycle and PPP genes as queries. In cases of positive hits for genes not reported in the annotations, multiple sequence alignments were used to confirm the presence of key conserved residues. Hydrophobicity plots were used to compare hydrophobicity patterns of CP12 sequences. More details are provided in *SI Methods*.

Metagenomic Analyses. Sequence reads from the GOS database (27) were recruited to cyanophage gene clusters from 17 T4-like cyanophage genomes defined previously (21). Reads were kept only if the top five BLAST hits were to the same gene cluster and only if the read did not have a better hit when blasted against an exhaustive set of marine genomes and the GenBank database. To test whether the size (i.e., number of orthologues) of a gene cluster adversely affected the number of sequence reads that could be recruited, we randomly excluded different orthologues from *talC*, *cp12*, *gnd*, and *zwf* gene clusters and repeated the recruitment; we found that as few as one orthologue of each gene was sufficient to recruit most of the reads recruited by the full gene cluster. Additional details are provided in *SI Methods*.

Infection of *Synechococcus* WH8109 by Cyanophage Syn9. Choice of strains was based on the fact that cyanophage Syn9 carries all four of the PPP genes of interest (*cp12*, *talC*, *zwf*, *gnd*; Table S1), and it readily infects *Synechococcus* WH8109, which has a sequenced genome. Log-phase *Synechococcus* WH8109 in the light was inoculated with cyanophage Syn9 at a multiplicity of infection (MOI) of 3. Three infected and three uninfected replicates were maintained. Uninfected controls were given spent medium instead of phage lysate. Samples were taken at regular intervals to be analyzed for gDNA and RNA. For phage and host gDNA quantification, samples were filtered with 0.2-μm polycarbonate filters, and the filtrate was used for extracellular phage gDNA quantification, whereas the retentate was used for intracellular phage gDNA quantitation and host gDNA quantification. For RNA, *Synechococcus* culture was harvested by centrifugation, flash-frozen in liquid nitrogen, and stored at −80 °C. Detailed descriptions of materials and methods are provided in *SI Methods*.

Infection of *Prochlorococcus* MED4 by Cyanophage P-HM2. Choice of strains was based on the fact that the transcriptional (24) and proteomic (29) dynamics of MED4 grown on a light–dark cycle have been studied extensively, and cyanophage P-HM2 was isolated on MED4 and carries the most prevalent auxiliary metabolic genes (*cp12*, *talC*, *psbA*; Table S1). Procedures were identical to those described above except that the MOI was 1, duplicate experiments were done in which bottles were moved to the light or the dark after inoculation (two infected and two uninfected replicates for each treatment), and culture aliquots were centrifuged for pyridine nucleotide extraction. Detailed descriptions of materials and methods are provided in *SI Methods*.

qPCR and RT-PCR. qPCR primers (Table S2) were designed from the genomes of cyanophages P-HM2 and Syn9, *Prochlorococcus* MED4, and *Synechococcus* WH8109, and were tested by using gDNA and shown to have specific and concentration-dependent amplification of target DNA. For mRNA quantification, RNA was extracted from lysed *Prochlorococcus* or *Synechococcus* cell pellets, gDNA degraded with DNase I, and first-strand cDNA synthesized from the RNA. For both gDNA and mRNA quantitation, copies were measured by using the QuantiTect SYBR Green PCR Kit (Qiagen) with a LightCycler 480 Real-Time PCR System (Roche Diagnostics). Relative copy numbers of each cDNA over the time course were determined by the $\Delta\Delta C_T$ method, with *rnpB* (RNA component of ribonuclease P) as the internal calibrator gene. Detailed descriptions of materials and methods are provided in *SI Methods*.

Cloning, Expression, and Purification of Transaldolases. Transaldolases from *Prochlorococcus* and cyanophages were PCR-amplified, cloned into pET vectors (Tables S3 and S4) encoding N- or C-terminal 6×His-tags, and transformed into BL21 (DE3) expression strains. Large-scale cultures of these strains were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated for 15 to 35 h at 25 °C or 13 °C. The cells were harvested and lysed by French press or sonication. The recombinant proteins were then purified to homogeneity by Ni-NTA affinity chromatography (elution with imidazole) and dialyzed into 50 mM Gly-Gly. Detailed descriptions of materials and methods are provided in *SI Methods*.

Transaldolase Assay and Determination of Kinetic Parameters. Transaldolase activity was measured by using a coupled assay as described previously (34) and in *SI Methods*. Enzymes and reagents were from Sigma-Aldrich. Kinetic data were fitted to the Michaelis–Menten equation to determine kinetic parameters as described in *SI Methods*.

Crystal Structure Determination. *Prochlorococcus* MIT9312 TalB was crystallized and its structure solved by X-ray crystallography with molecular replacement as described in *SI Methods*. The atomic coordinates and structure factors for *Prochlorococcus* MIT9312 TalB have been deposited in the PDB (accession no. 3HJZ).

Structure Homology Modeling and Alignment. A homology model was built for cyanophage P-SSP7 TalC using the structure of *T. maritima* TalC (PDB accession no. 1VPX) with SWISS-MODEL software, and this structure was overlaid with the structure of *Prochlorococcus* MIT9312 TalB by using UCSF Chimera software, both as described in *SI Methods*.

SEC Determination of Transaldolase Quaternary Structure. SEC was performed with a Superose 12 column (10 × 300 mm; GE Healthcare) attached to a Waters 2487 HPLC detector. Gel filtration molecular weight standards were run at the beginning of each experiment and used to determine the molecular weights of *Prochlorococcus* MIT9312 TalB and cyanophage P-SSP7 TalC as described in *SI Methods*.

Measurement of Pyridine Nucleotides in *Prochlorococcus*. Pyridine nucleotides NAD, NADH, NADP, and NADPH were extracted fresh from *Prochlorococcus* cultures and measured enzymatically as described previously (23) and in *SI Methods*.

ACKNOWLEDGMENTS. The authors thank Jacob Waldbauer and Erik Zinser for valuable discussions on carbon metabolism and its regulation; Matthew Sullivan for providing phages for genome sequencing; Debbie Lindell for providing *Synechococcus* WH8109 gene sequences; Brianne Holmbeck and Sara Roggensack for assistance in maintaining cultures; Hong Cui and Xiaohui Xu for help in crystallization and data collection; Allison Ortigosa, Daniela Hristova, Mohammad Seyedsayamdost, Joseph Cotruvo, Kenichi Yokoyama, Sumit Chakraborty, Alexa Price-Whelan, Arne Materna, Katya Frois-Moniz, Sébastien Rodrigue, Kolea Zimmerman, Daniel Sher, and Nadav Kashtan for advice on technical issues; James Calvin for statistical assistance; Marcia Osburne for comments on the manuscript; and Jeffrey Palm for assistance in manuscript preparation. This research was funded in part by the Gordon and Betty Moore Foundation (GBMF) through a grant to the Broad Institute. Samples G1170, G1171, and G1172 were sequenced and assembled at the Broad Institute. This study was supported by GBMF, the Department of Energy (Genomics:GTL Program), the National Science Foundation (Center for Microbial Oceanography: Research and Education), and a National Institutes of Health Training Grant (to L.R.T.).

1. De Waard A, Paul AV, Lehman IR (1965) The structural gene for deoxyribonucleic acid polymerase in bacteriophages T4 and T5. *Proc Natl Acad Sci USA* 54:1241–1248.
2. Miller ES, et al. (2003) Bacteriophage T4 genome. *Microbiol Mol Biol Rev* 67: 86–156.
3. Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Marine ecosystems: Bacterial photosynthesis genes in a virus. *Nature* 424:741.
4. Millard A, Clokie MRJ, Shub DA, Mann NH (2004) Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA* 101: 11007–11012.
5. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* 3:e144.

6. Breitbart M, Thompson LR, Suttle CA, Sullivan MB (2007) Exploring the vast diversity of marine viruses. *Oceanography (Wash DC)* 20:135–139.
7. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438:86–89.
8. Clokie MRJ, et al. (2006) Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. *Environ Microbiol* 8:827–835.
9. Lindell D, et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449:83–86.
10. Brown C, Lawrence J, Campbell D (2006) Are phytoplankton population density maxima predictable through analysis of host and viral genomic DNA content? *J Mar Biol Assoc UK* 86:491–498.
11. Bragg JG, Chisholm SW (2008) Modeling the fitness consequences of a cyanophage-encoded photosynthesis gene. *PLoS ONE* 3:e3550.
12. Mackenzie JJ, Haselkorn R (1972) Photosynthesis and the development of blue-green algal virus SM-1. *Virology* 49:517–521.
13. Sherman LA (1976) Infection of *Synechococcus cedrorum* by the cyanophage AS-1M. III. Cellular metabolism and phage development. *Virology* 71:199–206.
14. Kao CC, Green S, Stein B, Golden SS (2005) Diel infection of a cyanobacterium by a contractile bacteriophage. *Appl Environ Microbiol* 71:4276–4279.
15. Chen F, Lu J (2002) Genomic sequence and evolution of marine cyanophage P60: a new insight on lytic and lysogenic phages. *Appl Environ Microbiol* 68:2589–2594.
16. Mann NH, et al. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* 187:3188–3200.
17. Pope WH, et al. (2007) Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: a "horned" bacteriophage of marine *synechococcus*. *J Mol Biol* 368:966–981.
18. Weigele PR, et al. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* 9:1675–1695.
19. Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ (2009) Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: implications for mechanisms of cyanophage evolution. *Environ Microbiol* 11:2370–2387.
20. Sullivan MB, et al. (2009) The genome and structural proteome of an ocean siphovirus: a new window into the cyanobacterial 'mobilome'. *Environ Microbiol* 11: 2935–2951.
21. Sullivan MB, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035–3056.
22. Wedel N, Soll J, Paap BK (1997) CP12 provides a new mode of light regulation of Calvin cycle activity in higher plants. *Proc Natl Acad Sci USA* 94:10479–10484.
23. Tamoi M, Miyazaki T, Fukamizo T, Shigeoka S (2005) The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J* 42:504–513.
24. Zinser ER, et al. (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS ONE* 4:e5135.
25. Groben R, et al. (2010) Comparative sequence analysis of CP12, a small protein involved in the formation of a Calvin cycle complex in photosynthetic organisms. *Photosynth Res* 103:183–194.
26. Tompa P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533.
27. Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5:e77.
28. Sharon I, et al. (2011) Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J* 5:1178–1190.
29. Waldbauer JR (2009) Molecular biogeochemistry of modern and ancient marine microbes. PhD Thesis (Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Woods Hole, MA).
30. Bräutigam AR, Sauerbier W (1973) Transcription unit mapping in bacteriophage T7. I. In vivo transcription by Escherichia coli RNA polymerase. *J Virol* 12:882–886.
31. Thorell S, Schürmann M, Sprenger GA, Schneider G (2002) Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *J Mol Biol* 319:161–171.
32. Schneider S, Sandalova T, Schneider G, Sprenger GA, Samland AK (2008) Replacement of a phenylalanine by a tyrosine in the active site confers fructose-6-phosphate aldolase activity to the transaldolase of *Escherichia coli* and human origin. *J Biol Chem* 283:30064–30072.
33. Schürmann M, Sprenger GA (2001) Fructose-6-phosphate aldolase is a novel class I aldolase from *Escherichia coli* and is related to a novel group of bacterial transaldolases. *J Biol Chem* 276:11055–11061.
34. Bergmeyer HU, Gawehn K, Grassl M (1974) Enzymes as biochemical reagents. *Methods of Enzymatic Analysis*, ed Bergmeyer HU (Academic, New York), Vol 1, pp 425–522.
35. Samland AK, Sprenger GA (2009) Transaldolase: From biochemistry to human disease. *Int J Biochem Cell Biol* 41:1482–1494.
36. Russell PW, Müller UR (1984) Construction of bacteriophage ϕX174 mutants with maximum genome sizes. *J Virol* 52:822–827.
37. McGrath S, van Sinderen D (2007) *Bacteriophage: Genetics and Molecular Biology* (Caister, Norfolk, UK).
38. Hatfull GF, et al. (2010) Comparative genomic analysis of 60 Mycobacteriophage genomes: Genome clustering, gene acquisition, and gene size. *J Mol Biol* 397: 119–143.
39. Wilson W, Carr N, Mann N (1996) The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH 7803. *J Phycol* 32:506–516.
40. Thompson LR (2010) Auxiliary metabolic genes in viruses infecting marine cyanobacteria. PhD Thesis (Massachusetts Institute of Technology, Cambridge, MA).
41. Wikner J, Vallino J, Steward G, Smith DC, Azam F (1993) Nucleic acids from the host bacterium as a major source of nucleotides for three marine bacteriophages. *FEMS Microbiol Ecol* 12:237–248.
42. Jordan A, Reichard P (1998) Ribonucleotide reductases. *Annu Rev Biochem* 67:71–98.
43. Wedel N, Soll J (1998) Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc Natl Acad Sci USA* 95:9699–9704.
44. Sharon I, et al. (2009) Photosystem I gene cassettes are present in marine virus genomes. *Nature* 461:258–262.
45. Alperovitch-Lavy A, et al. (2010) Reconstructing a puzzle: Existence of cyanophages containing both photosystem-I and photosystem-II gene suites inferred from oceanic metagenomic datasets. *Environ Microbiol* 13:24–32.
46. Glazko G, Makarenkov V, Liu J, Mushegian A (2007) Evolutionary history of bacteriophages with double-stranded DNA genomes. *Biol Direct* 2:36.

Supporting Information

Thompson et al. 10.1073/pnas.1102164108

SI Methods

Sequences and Gene Annotation. Twenty-four genomes from cyanophages infecting *Prochlorococcus* and marine *Synechococcus* were included in the analyses (Table S1), three of which are being introduced in this study. Complete genome sequences and annotations of published genomes were downloaded from Integrated Microbial Genomes (1) and GenBank (2). New cyanophage genomes P-SSP2 (G1170), P-HP1 (G1171), and P-RSP5 (G1172) were sequenced and assembled at the Broad Institute by using the method of Henn et al. (3) and annotated using the annotation pipeline described by Sullivan et al. (4).

We searched the 24 cyanophage genomes for each Calvin cycle and PPP gene, along with known phage genes in photosynthetic electron transport and nucleotide biosynthesis, using the genome annotations. To check for the possibility of uncalled or miscalled genes in the annotations, we also searched the 24 genomes using TBLASTN with default parameters, an E-value cutoff of 10^{-5}, and by using all *Prochlorococcus* or *Synechococcus* Calvin cycle and PPP genes as queries. In cases of positive hits for genes not reported in the annotations, multiple sequence alignments were used to confirm the presence of key conserved residues.

Hydrophobicity plots were used to compare hydrophobicity patterns of hypothetical cyanophage CP12 sequences and known CP12 sequences from plants, algae, and cyanobacteria, by using the method of Kyte and Doolittle (5) with a window size of 11 residues. Before plotting hydrophobicity, sequences were aligned using MUSCLE (6) and then trimmed to remove all gapped positions.

Metagenomic Analyses. The GOS database (7) was downloaded from CAMERA (8). The database contained 9,893,120 sequences from 78 unique sites and 8,047,788,530 bp of sequence, with an average read length of 813 bp. GOS reads were first recruited to a protein database containing 1,306 previously defined cyanophage gene clusters (4) from 17 T4-like cultured cyanophage genomes using BLASTX with an e-value cutoff of 10^{-4}. Gene clusters are referred to as "core" if they are found in all 17 genomes. To improve the fidelity of our recruitment, all reads were required to have their top five hits be to the same gene cluster.

Each read meeting these criteria was then compared with a nucleotide database containing 11,770 sequences representing marine bacteria and phage genomes to recruit each read to its closest identifiably homologous marine genome. This database includes genomes of sequenced marine isolates, including Gordon and Betty Moore Foundation Marine Microbial Initiative genomes, National Center for Biotechnology Information marine isolates, and cultured cyanophages available from the GenBank (2) and CAMERA (8) databases using BLASTN. All reads with a best hit to a cyanophage sequence (82 total sequences) were then compared with all sequences in the National Center for Biotechnology Information nucleotide database (downloaded June 10, 2010) using BLASTN with default parameters to ensure that there were no better hits to a nonphage sequence in the nucleotide database.

To test whether genes with more representatives among the phage genomes would, as a result, be able to recruit proportionally more reads from GOS, we repeated the analysis with six modified cyanophage gene cluster databases in which orthologues of *talC*, *cp12*, *gnd*, and *zwf* were randomly excluded. Databases 1 and 2 contained one orthologue each of the four genes chosen randomly, databases 3 and 4 contained three orthologues each, and databases 5 and 6 contained five orthologues each. We found that the number or orthologues of these genes used to recruit GOS reads had little effect on the final counts of identified genes, with *talC* and *cp12* most abundant, followed by *gnd* and *zwf*, in all analyses.

Infection of *Synechococcus* WH8109 by Cyanophage Syn9. *Synechococcus* WH8109 was maintained in SN medium (9) made with 75% filtered seawater from the Environmental Systems Laboratory (Woods Hole, MA). Salts and metals for SN medium were from Sigma-Aldrich. Cultures were grown in a "sunbox," a modified Percival Scientific I-35LL plant growth chamber with a 24-h light–dark cycle consisting of 5 h of increasing light from 0 to 320 μE m^{-2} s^{-1}, 5 h of 320 μE m^{-2} s^{-1}, 4 h of decreasing light from 320 to 0 μE m^{-2} s^{-1}, and 10 h of dark. Temperature was maintained at 24 ± 0.2 °C. On the day of infection, 2 h before dark, log-phase *Synechococcus* WH8109 (1×10^8 cells mL^{-1} by flow cytometry; Influx; Cytopeia-BD) was inoculated with cyanophage Syn9 [3×10^8 infective phage mL^{-1} by the most probable number assay (10)], resulting in an MOI of 3. Three replicate infected cultures and three replicate uninfected control cultures of 1 L each were maintained. Uninfected controls were given spent medium instead of phage lysate. Both spent medium and phage lysate were filtered through 0.2-μm polycarbonate filters (Millipore) before addition. Following addition of phage lysate or spent medium, bottles were transferred to constant light of 50 μE m^{-2} s^{-1}.

Samples were taken at regular intervals for RNA and gDNA quantification. For RNA, 1-mL samples were centrifuged at 15,000 × *g* for 2 min at 4 °C, the supernatant aspirated, and the cell pellet flash-frozen in liquid nitrogen and stored at −80 °C. For phage and host gDNA quantification, 100-μL samples were filtered with 0.2-μm polycarbonate filters. The filtrate was diluted 1:1,000 for extracellular phage gDNA quantification. For intracellular phage and host gDNA quantification, the filter was washed with three 1-mL volumes of preservation solution (10 mM Tris-HCl, 100 mM EDTA, 500 mM NaCl, pH 8.0) and flash-frozen; the cells were subsequently resuspended in 650 μL 10-mM Tris-HCl (pH 8.0) by agitation in a Mini-Beadbeater (BioSpec), the supernatant heated to 95 °C for 15 min, then diluted 1:100.

Infection of *Prochlorococcus* MED4 by Cyanophage P-HM2. Axenic *Prochlorococcus* MED4 was maintained in Pro99 medium (11) made with filtered Sargasso seawater. Salts and metals for Pro99 medium were from Sigma-Aldrich. Cultures were grown in constant light of 90 μE m^{-2} s^{-1} with cool white fluorescent lamps or constant dark. Temperature was maintained at 19 to 22 °C. Log-phase *Prochlorococcus* MED4 (4×10^7 cells mL^{-1}) was inoculated with cyanophage P-HM2 (4×10^7 infective phage mL^{-1}), resulting in an MOI of 1. Cell concentration was determined by flow cytometry, and phage concentration was determined by the most probable number assay (10). For both light and dark experiments, which were conducted on separate days, two replicate infected cultures and two replicate uninfected control cultures of 2 L each were maintained. Uninfected controls were given spent medium instead of phage lysate. Both spent medium and phage lysate were filtered through 0.2-μm polycarbonate filters (Millipore) before addition.

Following inoculation, cultures were placed in a dark incubator or returned to the light incubator. Samples were taken at regular intervals to be analyzed for RNA, gDNA, and pyridine nucleotides. For phage and host gDNA quantification, 100-μL samples were filtered with 0.2-μm polycarbonate filters. The filtrate was

diluted 1:1,000 for extracellular phage gDNA quantification. For intracellular phage and host gDNA quantification, the filter was washed with three 1-mL volumes of preservation solution (10 mM Tris-HCl, 100 mM EDTA, 500 mM NaCl, pH 8.0) and flash-frozen; the cells were subsequently resuspended in 650 μL 10 mM Tris-HCl (pH 8.0) by agitation in a Mini-Beadbeater, the supernatant heated to 95 °C for 15 min, then diluted 1:100. For RNA and pyridine nucleotide samples, 200 mL axenic *Prochlorococcus* culture was harvested by centrifugation at 15,000 × *g* for 10 min at 4 °C, decanted, resuspended in approximately 1 mL supernatant, divided into equal aliquots into four tubes, and centrifuged again at 15,000 × *g* for 5 min at 4 °C. Samples for RNA (two tubes) were flash-frozen in liquid nitrogen and stored at −80 °C. Samples for pyridine nucleotides (two tubes) were extracted fresh as described later.

qPCR and RT-PCR. qPCR primers were designed from the genomes of cyanophage P-HM2 (4), *Prochlorococcus* MED4 (12), cyanophage Syn9 (13), and *Synechococcus* WH8109 (GenBank) using Primer3 (14) with a GC clamp of at least 2 bp, yielding products of 150 to 200 bp. Sequences are given in Table S2. Primers were tested using gDNA and were shown to have specific and concentration-dependent amplification of target DNA.

Cell pellets were resuspended in 100 μL of 10 mM Tris-HCl (pH 8.0), 100 U RNase inhibitor (SUPERASE-In; Ambion), and 15,000 U lysozyme (Ready-Lyse; Epicentre), and incubated at 37 °C for 30 min, after which 15,000 U additional lysozyme were added, followed by 30 min at 37 °C. RNA was extracted from this lysate by using the Mini RNA Isolation II Kit (Zymo Research), and RNA was eluted with nuclease-free water. This RNA was treated with 6 U Turbo DNase I (Ambion). cDNA was made from this DNase-treated RNA using the iScript cDNA Synthesis Kit (Bio-Rad); before cDNA synthesis, the reaction mixture lacking reverse transcriptase was heated to 65 °C for 5 min and then cooled on ice.

gDNA or cDNA copies were quantified by using the Quanti-Tect SYBR Green PCR Kit (Qiagen) with a LightCycler 480 Real-Time PCR System (Roche Diagnostics). qPCR reactions contained 0.5 μM forward and reverse primers and approximately 0.5 ng μL^{-1} cDNA. The amplification reaction consisted of an initial activation step of 15 min at 95 °C, followed by 50 cycles of 15 s at 95 °C (denaturation), 30 s at 56 °C (annealing), and 30 s at 72 °C (extension), followed by extension for 5 min at 72 °C, followed by a melting curve from 50 to 90 °C. Threshold cycle (C_T) of amplification was determined by the second derivative maximum method. Concentrations of phage and host gDNA over the time course were determined with standard curves of log(concentration of standard) versus C_T. Relative copy numbers of each cDNA over the time course were determined by the $\Delta\Delta C_T$ method (15), with *rnpB* (RNA component of ribonuclease P) as the internal calibrator gene (16).

Measurement of Pyridine Nucleotides in *Prochlorococcus*. Pyridine nucleotides NAD, NADH, NADP, and NADPH were extracted and measured enzymatically as described previously (17–19). NADH and NADPH standards were from Calbiochem, and all other reagents and enzymes were from Sigma-Aldrich. Pyridine nucleotides were extracted as follows: fresh cell pellets were resuspended in 200 μL 100 mM HCl, 500 mM NaCl (for determination of NAD and NADP), or 200 μL 100 mM NaOH, 500 mM NaCl (for determination of NADPH and NADH); these resuspensions were then heated at 95 °C for 5 min and centrifuged at 15,000 × *g* for 5 min at 4 °C, and the supernatants removed, flash-frozen in liquid nitrogen, and stored at −80 °C. NADH and NADPH standards of 20, 50, 100, 200, 500, and 1,000 nM were prepared in 100 mM NaOH, 500 mM NaCl from stock solutions whose concentrations were determined by A_{340}.

All assays were carried out at 30 °C in 200-μL reactions. Concentrated master solutions were made such that combining 180 μL master solution with 20 μL sample or unknown would yield the following final concentrations: 100 mM bicine (pH 8.0), 4 mM EDTA, 1.66 mM phenazine ethosulfate, and 0.42 mM 3-(4,5-dimethyl-2-thiazolyl)-2,5-diphenyl-2H-tetrazolium bromide. For determination of NAD and NADH, master solutions also contained 10% ethanol and 0.2 U alcohol dehydrogenase (final concentrations). For determination of NADP and NADPH, master solutions also contained 5 mM glucose 6-phosphate and 0.2 U glucose-6-phosphate dehydrogenase (final concentrations). Assays were initiated by combining 180 μL master solution with 20 μL *Prochlorococcus* extract or 20 μL NADH or NADPH standard. All assays were performed in duplicate. Time-dependent increases in A_{550} were monitored using an Ultramark Microplate Reader (Bio-Rad) for approximately 20 min. Absorbance data were smoothed by using the robust LOESS method (span, 20%), and rates were calculated by linear regression analysis, implemented with MATLAB software. Standard curves were then used to calculate NAD, NADH, NADP, or NADPH concentrations, from which relevant ratios were calculated.

Sequence Alignment and Phylogenetics. Transaldolase protein sequences were downloaded from GenBank (2) or pulled directly from genome sequences. Multiple sequence alignments were built by using ClustalW 1.83 (20), with alignment of transaldolase sequences assisted by the structure-based alignment of Thorell et al. (21). For phylogenetic analysis, positions with gaps were removed if they were present in at least 50% of sequences. A tree was built by using the maximum likelihood algorithm implemented by PhyML (22), and statistical tests of branches were done using aLRT (approximate likelihood-ratio test) parametric statistics with χ^2-based parametric branch supports (23). The tree was then midpoint rooted and displayed using TreeView (24).

Gene Size Comparison of Shared Phage and Host Orthologues. Sizes of genes predicted to encode the same functions in cyanophages and their hosts were compared. Predicted protein functions shared by phage and host genomes were identified and gene sizes obtained from clusters of orthologous groups (COG) functions on the ProPortal Web site (http://proportal.mit.edu). Gene orthologues were included in the analysis only if they were found in at least three cyanophage genomes and at least three *Prochlorococcus* or *Synechococcus* genomes, and only if they were enriched in cyanophages relative to noncyanophages, as defined by Sullivan et al. (4). For the phage class Ia RNR, made from one catalytic subunit (α, encoded by *nrdA*) and a second essential subunit that contains the metallo-cofactor (β, encoded by *nrdB*), only the length of *nrdA* was used. For phycoerythrobilin synthase, two genes are required in the host for the same activity carried out by one gene in phage, and thus the sum of the two host gene sizes was used (plus sign in Table S6). Median phage and host gene sizes were compared using the Mann–Whitney–Wilcoxon rank-sum test (25). Differences were considered significant if $P < 0.05$, corrected for false discovery rate using the Benjamini–Hochberg method for multiple comparisons (26). Additional information on the COGs used and statistical tests is found in Table S6.

Materials for Cloning, Expression, Purification, and Assay of Transaldolase Proteins. Phusion High-Fidelity Polymerase was from New England Biolabs. Champion pET Directional TOPO vectors and *E. coli* BL21 Star (DE3) One Shot chemically competent cells were from Invitrogen. BL21-CodonPlus (DE3)-RIPL competent cells and ArcticExpress (DE3) competent cells were from Stratagene. F6P (∼98% by enzymatic assay; lot no.

015K7013), E4P (~60% by enzymatic assay; lot no. 115K3789), rabbit-muscle TPI (4,400 U/mg; lot no. 035K7457), and rabbit-muscle G3PDH (270 U/mg; lot no. 035K7457) were from Sigma-Aldrich. IPTG and 1,4-DTT were from Promega. DNase and Complete Mini EDTA-free protease inhibitor tablets were from Roche. BugBuster protein extraction reagent was from Novagen. Gel filtration molecular weight standards were from Bio-Rad. NADH was from Calbiochem.

Cloning of Recombinant Transaldolases. *Prochlorococcus* and cyanophage gDNA were isolated using a Qiagen DNeasy Blood and Tissue kit (cells) or used directly (phage) and amplified by PCR by using Phusion High-Fidelity Polymerase, with primer sequences given in Tables S3 and S4. Amplicons were cloned into Champion pET Directional TOPO vector pET100 or pET101 or plasmid p15TvLic. Plasmid p15TvLic is a modified version of vector pET-15b from Novagen in which the tobacco etch virus (TEV) protease cleavage site replaces the thrombin cleavage site and a double stop codon is inserted downstream of the BamHI site (27). Additional information on constructs is contained in Tables S3 and S4. Sequences of *talB* and *talC* constructs were confirmed by DNA sequencing at the Massachusetts Institute of Technology Biopolymers Laboratory. Cyanophage *talC* pET101 constructs were transformed into *E. coli* BL21 Star (DE3) One Shot chemically competent cells, *Prochlorococcus* MIT9312 *talB* p15TvLic construct was transformed into BL21-CodonPlus (DE3)-RIPL competent cells, and *Prochlorococcus talB* pET100 constructs were transformed into ArcticExpress (DE3) competent cells.

Expression and Purification of *Prochlorococcus* TalB. The expression and purification of *Prochlorococcus* MED4 TalB is described; similar expression and purification protocols were used for *Prochlorococcus* MIT9312 TalB and *Prochlorococcus* NATL2A TalB. Cells carrying the pET constructs were grown in Luria–Bertani medium containing 100 μg/mL ampicillin and 20 μg/mL gentamicin, with shaking at 37 °C. When OD_{600} reached 0.5, temperature was reduced to 13 °C and cultures incubated until OD_{600} was 1.0 (~3 h), at which time they were induced with 0.5 mM IPTG. Following 35 h of growth at 13 °C, cells were harvested by centrifugation for 10 min at 3,000 × *g*. Typical yield was 4 g cell paste per liter of culture.

Cell paste (16 g) was resuspended in 80 mL of buffer A [50 mM Gly-Gly (pH 8.0), 500 mM NaCl, and 5% glycerol] with 10 mM imidazole, 1 mg/mL lysozyme, 10 U/mL DNase, and two Complete Mini EDTA-free protease inhibitor tablets. This resuspension was lysed using a French pressure cell press (Thermo Scientific) at 14,000 psi, and the lysate was centrifuged for 10 min at 40,000 × *g*. The supernatant fraction was incubated with Ni-NTA resin (16 mL; Qiagen) and buffer A with 10 mM imidazole in a final volume of 160 mL. The slurry was poured into a column, and material unbound to the resin was allowed to flow through, followed by washes of 40 column volumes of buffer A containing 10 mM imidazole. Bound proteins were eluted with a 100 × 100-mL linear gradient of 10 to 300 mM imidazole in Buffer A. TalB eluted at 120 mM imidazole, and fractions with high protein content (based on A_{280}) or high activity were pooled and concentrated with Amicon Ultra-15 centrifugal filter units (Millipore). Concentrated protein was diluted 1:100 to reduce the salt concentration, and this solution was loaded onto a Q Sepharose Fast Flow anion-exchange column (5.5 × 6.5 cm, 150 mL) preequilibrated with 50 mM Gly-Gly (pH 8.0) containing 5 mM NaCl. Protein was eluted by using a 250 × 250-mL linear gradient of 5 to 500 mM NaCl in 50 mM Gly-Gly (pH 8.0). TalB eluted at 250 mM NaCl, and fractions with high A_{280} or high activity were pooled and concentrated. The concentrated protein was loaded onto a Sephadex G-25 size-exclusion column (2.5 × 41 cm, 200 mL) preequilibrated with 50 mM Gly-Gly (pH 8.0). Fractions with high A_{280} were pooled and concentrated. Purified TalB aliquots were stored in 10% glycerol at −80 °C. Protein concentrations were determined according to a previously described method (28).

Purification of *Prochlorococcus* MIT9312 TalB was similar to that described earlier with the following exceptions. Protein expression was induced overnight at 18 °C. Cells were lysed by sonication rather than French press. Following purification by Ni-NTA affinity chromatography, the protein was dialyzed and concentrated in a buffer with 10 mM Hepes (pH 7.5), 300 mM NaCl, and 0.5 mM Tris(2-carboxyethyl)phosphine (TCEP).

Expression and Purification of Cyanophage TalC. The expression and purification of cyanophage P-SSM2 TalC is described; nearly identical expression and purification protocols were used for cyanophage P-SSM4 TalC and cyanophage P-SSP7 TalC. Cells carrying the pET constructs were grown in Luria–Bertani medium containing 100 μg/mL ampicillin, with shaking at 37 °C. When OD_{600} reached 0.7, temperature was reduced to 25 °C and cultures incubated until OD_{600} was 1.0 (~1 h), at which time they were induced with 0.5 mM IPTG. Following 15 h of growth at 25 °C, cells were harvested by centrifugation for 10 min at 3,000 × *g*. Typical yield was 5 g cell paste per liter of culture.

Cell paste (20 g) was resuspended in 100 mL of buffer A with 20 mM imidazole, 1 mg/mL lysozyme, 10 U/mL DNase, and two Complete Mini EDTA-free protease inhibitor tablets. This resuspension was lysed by using a French pressure cell press at 14,000 psi, and the lysate was centrifuged for 10 min at 40,000 × *g*. The supernatant fraction was incubated with Ni-NTA resin (40 mL) and buffer A with 20 mM imidazole in a final volume of 400 mL. The slurry was poured into a column and material not bound to the resin was allowed to flow through, followed by washes with 40 column volumes of buffer A containing 20 mM imidazole. Bound proteins were eluted with a 200 × 200-mL linear gradient of 20 to 500 mM imidazole. TalC eluted at 160 mM imidazole, and fractions with high protein content (based on A_{280}) or high activity were pooled and concentrated with Amicon Ultra-15 centrifugal filter units. The concentrated protein was transferred into 50 mM Gly-Gly (pH 8.0) by using dialysis with a Slide-A-Lyzer dialysis cassette. Purified TalC aliquots were stored in 10% glycerol at −80 °C. Protein concentrations were determined according to a previously published method (28).

Transaldolase Assay. Transaldolase activity was measured by using a coupled assay as described previously (29). A typical assay in a final volume of 500 μL contained 50 mM Gly-Gly (pH 8.0), 15 mM $MgCl_2$, 10 mM F6P, 1 mM E4P, 0.2 mM NADH, 10 mM DTT, 0.6 U TPI, 0.06 U G3PDH, and transaldolase (~0.5 μg or 0.005 U). F6P and E4P concentrations were determined by using endpoint assays. NADH consumption was measured by A_{340} (ε of 6.2 mM^{-1} cm^{-1}) with a Cary 3 UV–visible spectrophotometer (Varian) or an Ultramark Microplate Reader (Bio-Rad). Cuvettete path length for Cary 3 assays was 1 cm, whereas path length for microtiter plate wells was determined empirically with NADH standards, and a correction factor of 1.785 was applied to Ultramark A_{340} measurements to make them comparable to Cary 3 A_{340} measurements. For assays using the Cary 3 spectrophotometer, a solution containing Gly-Gly, $MgCl_2$, F6P, E4P, NADH, and DTT (480 μL, final concentrations as described earlier) was equilibrated at 25 °C and monitored for 1 min to confirm no change in A_{340}. A solution of TPI and G3PDH (10 μL, final concentrations as described earlier) was added and any change in A_{340} allowed to dissipate for approximately 5 min. Transaldolase (10 μL) was added, and the change in A_{340} monitored for 5 min. For assays using the Ultramark reader, two microtiter plates were used, and the assay volume was reduced to 200 μL. In plate A, buffer, F6P, and transaldolase (180 μL, final concentrations as described earlier) were premixed and in-

cubated at 25 °C for 10 min. In plate B, E4P, NADH, TPI, and G3PDH (20 μL, final concentrations as described earlier) were incubated at room temperature for 10 min. The contents of plate A (180 μL) were added to plate B to initiate the assay.

Determination of Kinetic Parameters. Kinetic parameters were determined using Eq. **S1**, where F6P (10 mM) or E4P (1 mM) was kept constant and the other substrate was varied from 0.05 to 20 × K_m. Kinetic data were fitted to Eq. **S1** by using linear least-squares analysis implemented with MATLAB software (MathWorks).

$$v = \frac{V_{max}[S]}{K_m + [S]} \quad \textbf{[S1]}$$

Crystallization Conditions. *Prochlorococcus* MIT9312 TalB (19 mg/mL) in 0.2 M Hepes (pH 7.5), 500 mM NaCl, and 0.5 mM TCEP was transferred to 10 mM Hepes (pH 7.5), 300 mM NaCl, and 0.5 mM TCEP by dialysis and concentrated to 19 mg/mL. Crystallization was performed at room temperature (21 °C), and a homemade preparation of TEV protease was added in a ratio of protease to TalB of 1:20 to remove the His-tag from TalB; the protease was not removed before crystallization. Crystallization trials were performed by using hanging-drop vapor diffusion with an optimized sparse matrix crystallization screen (30). The crystal used for the data collection (Table S5) was obtained by using a crystallization liquor containing final concentrations of 20% PEG10K and 0.1 M Hepes (pH 7.5). Crystals were cryoprotected in 20% PEG10K, 0.1 M Hepes (pH 7.5), and 10% ethylene glycol, then rinsed in Paratone-N oil (Hampton Research) and immediately flash-frozen in liquid nitrogen and stored in liquid nitrogen before data collection.

Data Collection, Structure Determination, and Refinement. Diffraction data were collected at 100 °K on a Rigaku Micromax-007 rotating anode generator equipped with Osmic mirrors. Diffraction data were recorded on an R-Axis IV++ detector and integrated and scaled by using HKL2000 (31). The structure of *Prochlorococcus* MIT9312 TalB was solved by molecular replacement by using the coordinates of human TALDO1 (PDB accession no. 1F05) (32) as the initial model. The program PHASER (33), as part of the CCP4 program suite (34), was used to find the position of TalB monomer in the unit cell. The model was then improved by alternate cycles of manual building and water-picking by using COOT (35) and restrained refinement against a maximum-likelihood target with 5% of the reflections randomly excluded as an R_{free} test set. All refinement steps were performed by using REFMAC (36) in the CCP4 program suite. Only one residue (residue 333) of the 333-residue protein was omitted in the model because of poor electron density. The final model contains one molecule of TalB, 341 water molecules, and 45 atoms from other small molecules and ions, and was refined to an R_{work} and R_{free} of 15.6% and 20.3%, respectively. Data collection, phasing, and structure refinement statistics are summarized in Table S5. The Ramachandran plot generated by PROCHECK (37) showed excellent stereochemistry overall, with 100% of the residues in the most favored and additional allowed regions. The atomic coordinates and structure factors for *Prochlorococcus* MIT9312 TalB have been deposited in PDB (accession no. 3HJZ).

Structure Homology Modeling and Alignment. Homology models were built by using the ClustalW multiple sequence alignment described earlier, and existing crystal structures using the alignment mode of the SWISS-MODEL workspace (38, 39). Sequences of TalC from cyanophages P-SSM2, P-SSM4, and P-SSP7 were modeled by using the structure of *T. maritima* TalC (Joint Center for Structural Genomics; PDB accession no. 1VPX), and sequences of TalB from *Prochlorococcus* MED4 and NATL2A were modeled by using the structure of *Prochlorococcus* MIT9312 (present study; PDB accession no. 3HJZ).

For visualization of superimposed 3D structures, structure alignments and molecular graphics images were produced using the University of California, San Francisco, Chimera package (40). Alignments were done using the MatchMaker tool in Chimera, with best-aligning pairs of chains aligned using the Needleman–Wunsch algorithm and the BLOSUM-30 scoring matrix, depending on the average sequence identity among pairs of sequences. Raytraced images were produced with Persistence of Vision Raytracer (version 3.6) software (http://www.povray.org/).

SEC Determination of Oligomerization State. SEC was performed by using a Superose 12 column (10 × 300 mm; GE Healthcare) attached to a Waters 2487 HPLC detector. Gel filtration molecular weight standards were vitamin B12 (1.35 kDa), myoglobin (17 kDa), ovalbumin (44 kDa), γ-globulin (158 kDa), and thyroglobulin (670 kDa). The elution buffer was 50 mM Gly-Gly (pH 8.0), 150 mM NaCl. Molecular mass standards were run at the beginning of each experiment. TalB or TalC (40 μL, 10 mg/mL) was injected onto the column. The flow rate was 0.5 mL/min. A_{280} was monitored. Plots of log(molecular weight standard) vs. retention time were used to estimate molecular weights of TalB and TalC.

1. Markowitz VM, et al. (2006) The integrated microbial genomes (IMG) system. *Nucleic Acids Res* 34(database issue):D344–D348.
2. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL (2008) GenBank. *Nucleic Acids Res* 36(database issue):D25–D30.
3. Henn MR, et al. (2010) Analysis of high-throughput sequencing and annotation strategies for phage genomes. *PLoS ONE* 5:e9083.
4. Sullivan MB, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035–3056.
5. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157:105–132.
6. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797.
7. Rusch DB, et al. (2007) The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol* 5:e77.
8. Seshadri R, Kravitz SA, Smarr L, Gilna P, Frazier M (2007) CAMERA: A community resource for metagenomics. *PLoS Biol* 5:e75.
9. Waterbury J, Willey J (1988) Isolation and growth of marine planktonic cyanobacteria. *Methods Enzymol* 167:100–105.
10. Tillett HE (1987) Most probable numbers of organisms: Revised tables for the multiple tube method. *Epidemiol Infect* 99:471–476.
11. Moore L, et al. (2007) Culturing the marine cyanobacterium *Prochlorococcus*. *Limnol Oceanogr Methods* 5:353–362.
12. Rocap G, et al. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* 424:1042–1047.
13. Weigele PR, et al. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* 9:1675–1695.
14. Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386.
15. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29:e45.
16. Fernández-González B, Martínez-Férez IM, Vioque A (1998) Characterization of two carotenoid gene promoters in the cyanobacterium *Synechocystis* sp. PCC 6803. *Biochim Biophys Acta* 1443:343–351.
17. Maciejewska U, Kacperska A (1987) Changes in the level of oxidized and reduced pyridine nucleotides during cold acclimation of winter rape plants. *Physiol Plant* 69:687–691.
18. Leonardo MR, Dailly Y, Clark DP (1996) Role of NAD in regulating the *adhE* gene of *Escherichia coli*. *J Bacteriol* 178:6013–6018.
19. Tamoi M, Miyazaki T, Fukamizo T, Shigeoka S (2005) The Calvin cycle in cyanobacteria is regulated by CP12 via the NAD(H)/NADP(H) ratio under light/dark conditions. *Plant J* 42:504–513.
20. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.

21. Thorell S, Schürmann M, Sprenger GA, Schneider G (2002) Crystal structure of decameric fructose-6-phosphate aldolase from *Escherichia coli* reveals inter-subunit helix swapping as a structural basis for assembly differences in the transaldolase family. *J Mol Biol* 319:161–171.
22. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696–704.
23. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55:539–552.
24. Page RDM (2002) Visualizing phylogenetic trees using TreeView. *Curr Protoc Bioinformatics* Chapter 6:Unit 6.2.
25. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18:50–60.
26. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* 57:289–300.
27. Zhang RG, et al. (2001) Structure of *Thermotoga maritima* stationary phase survival protein SurE: A novel acid phosphatase. *Structure* 9:1095–1106.
28. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.
29. Bergmeyer HU, Gawehn K, Grassl M (1974) Enzymes as biochemical reagents. In *Methods of Enzymatic Analysis*, ed Bergmeyer HU (Academic, New York), Vol Vol 1, pp 425–522.
30. Kimber MS, et al. (2003) Data mining crystallization databases: Knowledge-based approaches to optimize protein crystal screens. *Proteins* 51:562–568.
31. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M (2006) HKL-3000: The integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr D Biol Crystallogr* 62:859–866.
32. Thorell S, Gergely P Jr., Banki K, Perl A, Schneider G (2000) The three-dimensional structure of human transaldolase. *FEBS Lett* 475:205–208.
33. McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61:458–464.
34. Collaborative Computational Project Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50:760–763.
35. Emsley P, Cowtan K (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60:2126–2132.
36. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53: 240–255.
37. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J Appl Cryst* 26:283–291.
38. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385.
39. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22: 195–201.
40. Pettersen EF, et al. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25:1605–1612.

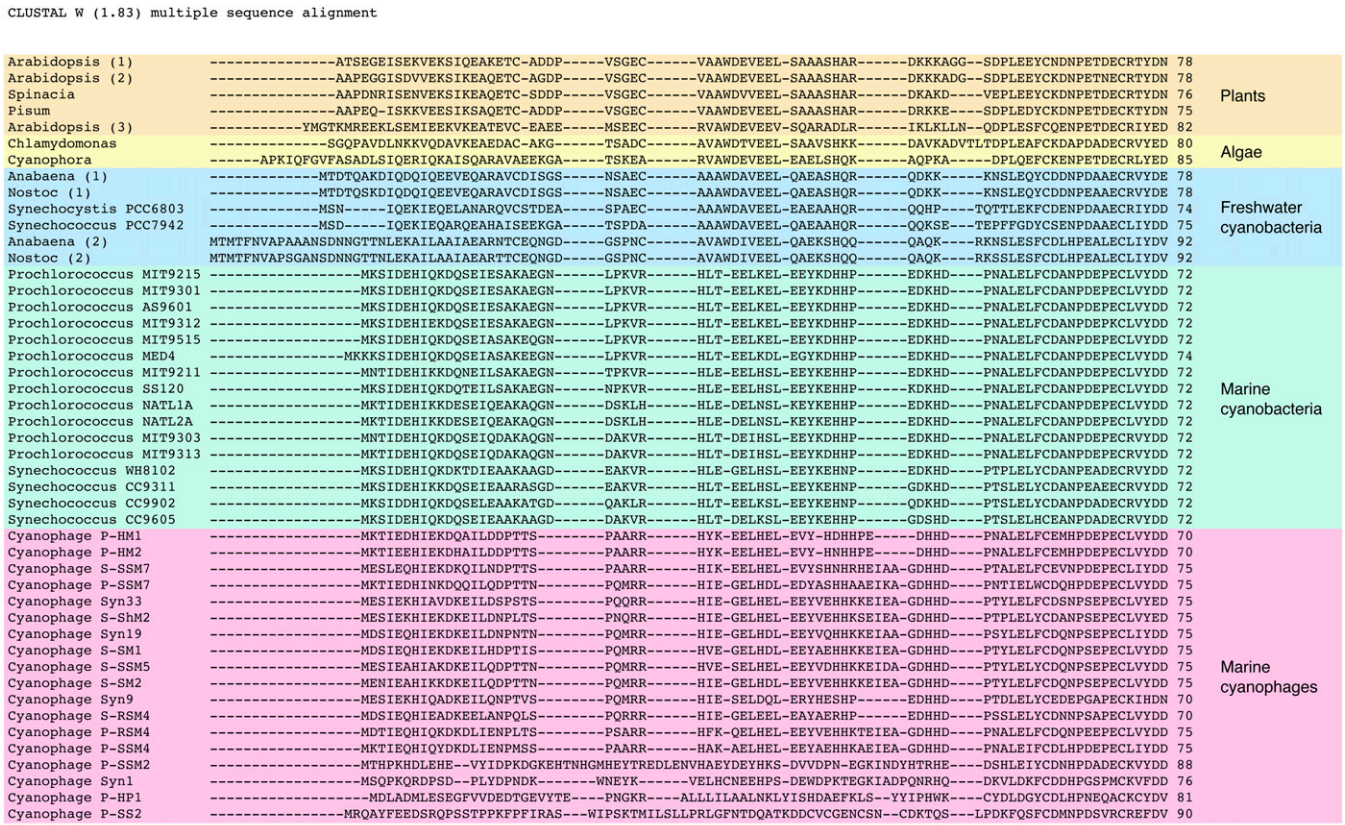

Fig. S1. Multiple sequence alignment of CP12 sequences from cyanophages, cyanobacteria, algae, and plants. Despite amino acid identities between CP12 orthologues of less than 10%, all sequences exhibit a similar pattern of predominantly hydrophilic residues, characteristic of intrinsically unstructured proteins like CP12, and all sequences contain the C-terminal motif CxxxPxxxxC, which is conserved in nearly all CP12 proteins.

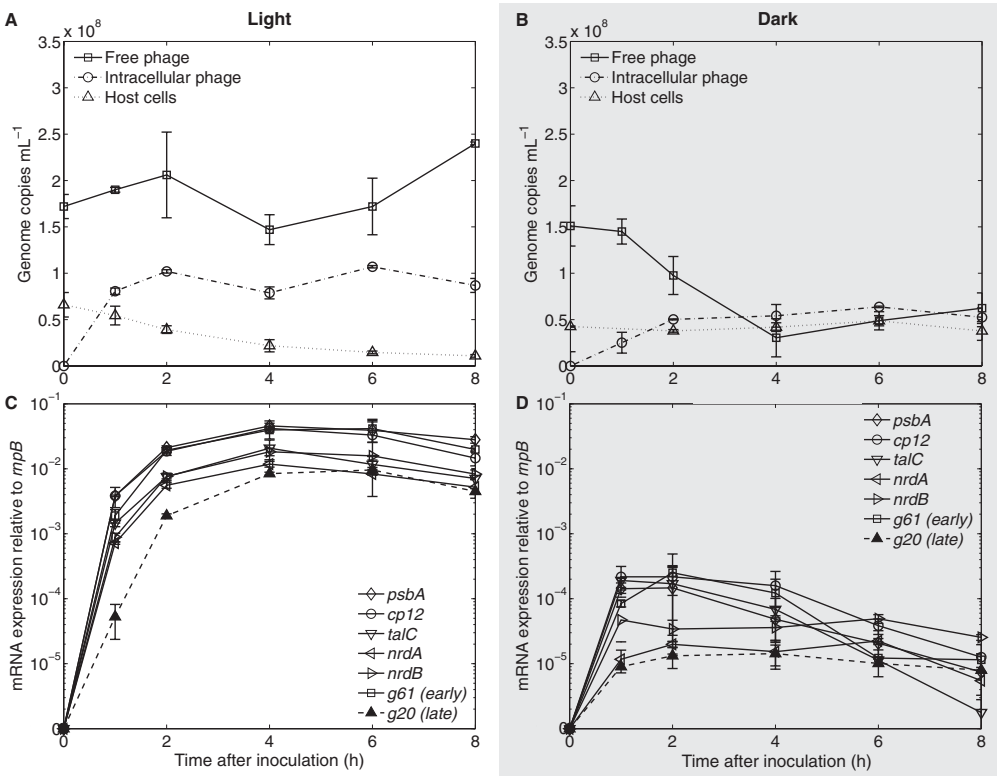

Fig. S2. Infection dynamics and mRNA levels of *Prochlorococcus* MED4 infected by cyanophage P-HM2 in the light and in the dark. Error bars represent SDs of three technical and two biological replicates. (*A* and *B*) Degradation of host gDNA and replication of phage gDNA (free and intracellular) as determined by qPCR of chromosomal genes. Host DNA degradation and phage replication in the dark were severely stunted relative to infection in the light. (*C* and *D*) Expression of phage metabolic genes as determined by RT-qPCR. Known T4-like phage early (*g61*, DNA primase) and late (*g20*, portal protein) genes were used to calibrate the timing of expression, showing that all phage metabolic genes were expressed early. In the light, *cp12* was coexpressed with genes for photosystem II (*psbA*), the PPP (*talC*), and RNR (*nrdA*, *nrdB*). In the dark, these genes were also coexpressed but at levels 100 to 1,000 fold lower than in the light.

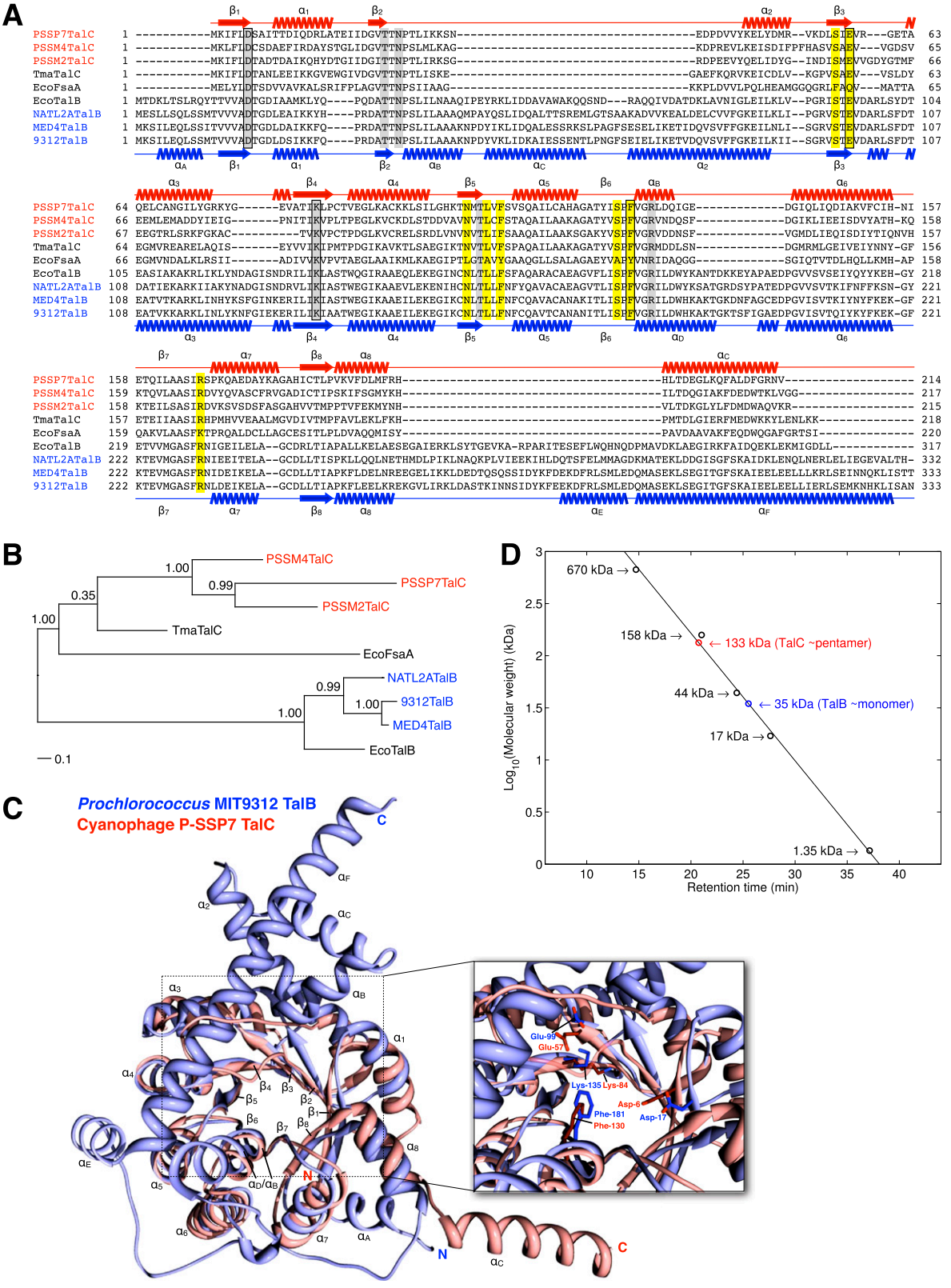

Fig. S3. Sequence alignment, phylogenetic tree, crystal structures, and molecular weights of cyanophage and *Prochlorococcus* transaldolases. Cyanophage TalCs are colored red, and *Prochlorococcus* TalBs are colored blue. (*A*) Multiple sequence alignment, generated starting from the structure-based alignment of Thorell et al. (21). Active-site residues are colored according to whether they are identical (gray) or variable (yellow), and residues pictured in the structure in *C* are outlined in black. Secondary structure elements of cyanophage P-SSP7 TalC and *Prochlorococcus* MIT9312 TalB are annotated above and below the alignment and are named as described previously (21). (*B*) Maximum-likelihood tree generated from alignment, with gaps removed, midpoint-rooted with χ^2-based parametric branch supports. (*C*) Structures of *Prochlorococcus* TalB and cyanophage TalC subunits. Naming of α-helices and β-strands matches the labels in *A*. Although the arrangement of exterior helices varies between the two proteins, the core α/β-barrel structure is conserved. The active site structure (*Inset*) is

Legend continued on following page

also conserved, including the arrangement of the Schiff base-forming lysine (Lys-135, MIT9312 TalB numbering), proton-donating/accepting glutamate (Glu-99) and aspartate (Asp-17), and specificity-determining phenylalanine (Phe-181). (*D*) Molecular weights of *Prochlorococcus* MIT9312 TalB and cyanophage P-SSP7 TalC determined by size-exclusion chromatography. Retention times for TalB are shown in blue, TalC in red, and molecular weight standards in black. TmaTalC, *T. maritima* TalC; PSSM2TalC, cyanophage P-SSM2 TalC; PSSM4TalC, cyanophage P-SSM4 TalC; PSSP7TalC, cyanophage P-SSP7 TalC; EcoFsaA, *E. coli* FsaA; EcoTalB, *E. coli* TalB; NATL2ATalB, *Prochlorococcus* NATL2A TalB; MED4TalB, *Prochlorococcus* MED4 TalB; and 9312TalB, *Prochlorococcus* MIT9312 TalB.

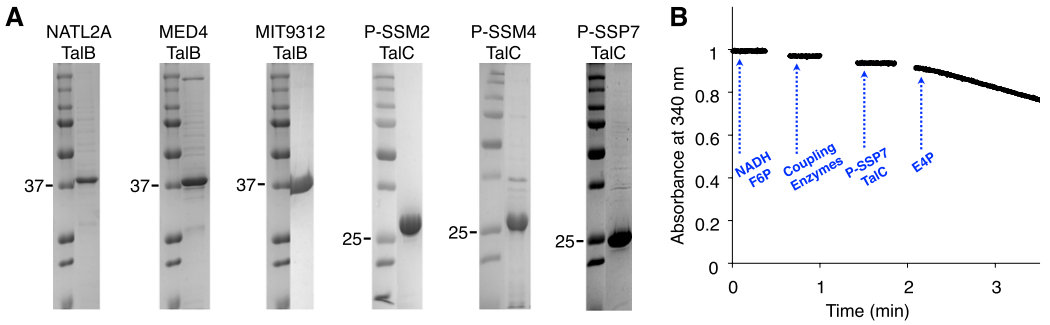

Fig. S4. SDS/PAGE and transaldolase activity of *Prochlorococcus* TalB and cyanophage TalC. (*A*) SDS/PAGE gels of purified TalB and TalC loaded with 2 to 10 μg protein. Molecular weight standards were run in each gel; standards of 37 kDa or 25 kDa are labeled. (*B*) A typical assay demonstrating transaldolase activity, showing A_{340} (NADH) decrease over time as an indicator, via coupling enzymes, of transaldolase activity (*SI Methods*). No appreciable activity was detected without the addition of E4P, which in transaldolase serves as an acceptor substrate for DHA. This is shown here for cyanophage P-SSP7 TalC and was demonstrated for all cyanophage and *Prochlorococcus* transaldolases tested.

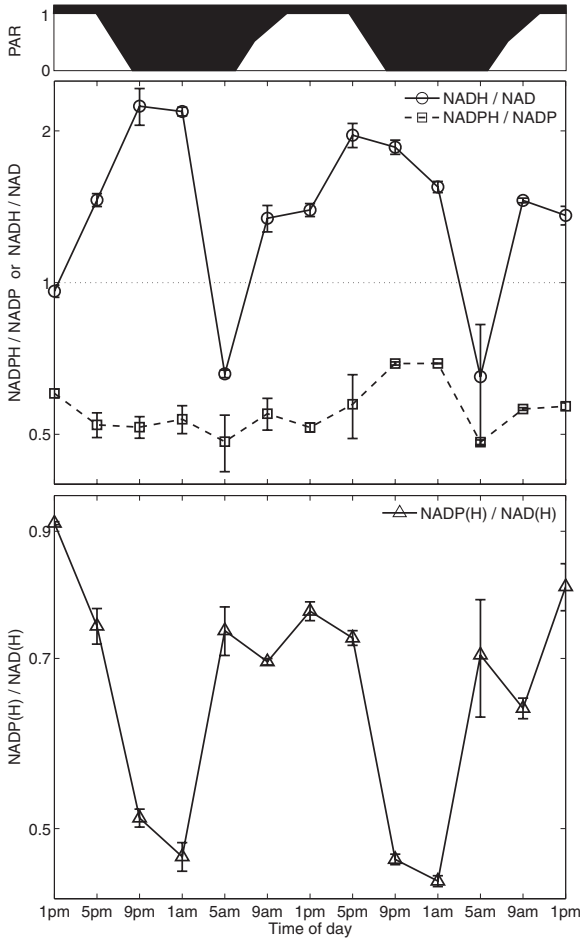

Fig. S5. Redox (*Top*) and phosphorylation state (*Bottom*) dynamics of pyridine nucleotides NADPH, NADP, NADH, and NAD in *Prochlorococcus* MED4 over two 24-h light–dark cycles. Relative photosynthetically available radiation (PAR) is shown above the graphs. Error bars represent SDs of two technical and two biological replicates.

Table S1. Distribution of all PPP, select photosynthetic electron transport, and select nucleotide biosynthesis genes in 24 cyanophage genomes, with additional information for three new T7-like podovirus genomes

							PPP				Photosynthetic electron transport						Nucleotide biosynthesis		
Strain	Type	Original host	Location isolated	Size, kbp	GC, %	Ref.	*talC*	*cp12*	*gnd*	*zwf*	*psbA*	*psbD*	*petE*	*petF*	PTOX	*hli*	*nrd*	*cobS*	*thyX*
P-SS2	S	*Pro.* MIT9313	Atlantic slope waters	107.5	52.3	1	—	x	—	—	—	—	—	—	—	—	*J*	—	x
P-SSP7	T7	*Pro.* MED4	Sargasso Sea	45.0	38.8	2	x	—	—	—	x	—	—	—	—	x	*J*	—	—
P-SSP2[a]	T7	*Pro.* MIT9312	Sargasso Sea	45.9	37.9	Present study	x	—	—	—	x	—	—	—	—	x	*J*	—	—
P-RSP5[b]	T7	*Pro.* NATL1A	Red Sea	47.7	38.7	Present study	x	—	—	—	x	—	—	—	—	x	*J*	—	—
P-HP1[c]	T7	*Pro.* NATL2A	Hawai'i	47.5	39.9	Present study	x	x	—	—	x	—	—	—	—	x	*J*	—	—
P60	T7	*Syn.* WH7803	Georgia coastal river	47.9	53.2	3	—	—	—	—	—	—	—	—	—	—	*J*	—	x
Syn5	T7	*Syn.* WH8109	Sargasso Sea	46.2	55.0	4	—	—	—	—	—	—	—	—	—	—	*J*	—	x
P-HM2	T4	*Pro.* MED4	Hawai'i	183.8	38.0	5	x	x	—	—	x	x	—	—	—	x	*AB*	x	x
P-HM1	T4	*Pro.* MED4	Hawai'i	181.0	38.0	5	x	x	—	—	x	x	—	—	x	x	*AB*	x	x
P-SSM4	T4	*Pro.* NATL2A	Sargasso Sea	178.2	36.7	2	x	x	—	—	x	x	—	—	x	x	*AB*	x	x
P-RSM4	T4	*Pro.* MIT9303	Red Sea	176.4	38.0	5	x	x	—	—	x	—	—	—	x	x	*AB*	x	x
P-SSM7	T4	*Pro.* NATL1A	Sargasso Sea	182.2	37.0	5	x	x	—	—	x	—	—	x	—	x	*AB*	x	x
P-SSM2	T4	*Pro.* NATL1A	Sargasso Sea	252.4	35.5	2	x	x	—	—	x	—	x	x	—	x	*AB*	x	x
S-PM2	T4	*Syn.* WH7803	English Channel	196.3	37.8	6	—	—	—	—	x	x	—	—	—	x	*AB*	x	x
S-SSM7	T4	*Syn.* WH8109	Sargasso Sea	232.9	39.0	5	x	x	—	—	x	—	x	—	x	x	*AB*	x	x
Syn33	T4	*Syn.* WH7803	Gulf Stream	174.4	40.0	5	x	x	—	—	x	x	x	x	—	x	*AB*	x	x
S-ShM2	T4	*Syn.* WH8102	Atlantic shelf waters	179.6	41.0	5	x	x	x	—	x	x	x	—	—	x	*AB*	x	x
Syn1	T4	*Syn.* WH8101	Woods Hole	191.2	41.0	5	x	x	x	—	x	x	x	—	—	x	*AB*	x	x
S-SM1	T4	*Syn.* WH6501	Atlantic slope waters	178.5	41.0	5	x	x	x	x	x	x	x	—	x	x	*AB*	x	x
S-SSM5	T4	*Syn.* WH8102	Sargasso Sea	176.2	40.0	5	x	x	x	x	x	x	x	—	x	x	*AB*	x	x
Syn9	T4	*Syn.* WH8012	Woods Hole	177.3	40.5	7	x	x	x	x	x	x	x	—	x	x	*AB*	x	x
Syn19	T4	*Syn.* WH8109	Sargasso Sea	175.2	41.0	5	x	x	x	x	x	x	x	—	x	x	*AB*	x	x
S-RSM4	T4	*Syn.* WH8103	Red Sea	194.5	41.1	8	x	x	x	x	x	x	x	x	x	x	*AB*	x	x
S-SM2	T4	*Syn.* WH8017	Atlantic slope waters	190.8	40.0	5	x	x	x	x	x	x	x	x	x	x	*AB*	x	x

Phage strains are classified by phage type (S, siphovirus; T7, T7-like podovirus; T4, T4-like myovirus) and host of isolation (*Pro.*, *Prochlorococcus*; *Syn.*, *Synechococcus*). Additional source water information for new genomes: a, P-SSP2 isolated from 31°48′N 64°16′W at 120 m on September 28, 1995; b, P-RSP5 isolated from 29°28′N 34°55′E at 130 m on September 13, 2000; c, P-HP1 isolated from 22°45′N 158°00′W at 25 m on March 8, 2006. Gene products: *talC*, transaldolase; *cp12*, CP12; *gnd*, 6-phosphogluconate dehydrogenase; *zwf*, glucose-6-phosphate dehydrogenase; *psbA*, photosystem II protein D1; *psbD*, photosystem II protein D2; *petE*, plastocyanin; *petF*, ferredoxin; PTOX, plastoquinol terminal oxidase; *hli*, high-light inducible protein; *nrdJ*, class II RNR; *nrdAB*, class Ia RNR; *cobS*, cobalt chelatase; and *thyX*, thymidylate synthase.

1. Sullivan MB, et al. (2009) The genome and structural proteome of an ocean siphovirus: A new window into the cyanobacterial 'mobilome'. *Environ Microbiol* 11:2935–2951.
2. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol* 3:e144.
3. Chen F, Lu J (2002) Genomic sequence and evolution of marine cyanophage P60: A new insight on lytic and lysogenic phages. *Appl Environ Microbiol* 68:2589–2594.
4. Pope WH, et al. (2007) Genome sequence, structural proteins, and capsid organization of the cyanophage Syn5: A "horned" bacteriophage of marine *synechococcus*. *J Mol Biol* 368:966–981.
5. Sullivan MB, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035–3056.
6. Mann NH, et al. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus* strains. *J Bacteriol* 187:3188–3200.
7. Weigele PR, et al. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* 9:1675–1695.
8. Millard AD, Zwirglmaier K, Downey MJ, Mann NH, Scanlan DJ (2009) Comparative genomics of marine cyanomyoviruses reveals the widespread occurrence of *Synechococcus* host genes localized to a hyperplastic region: Implications for mechanisms of cyanophage evolution. *Environ Microbiol* 11:2370–2387.

Table S2. qPCR primer sequences used in this study

Strain/gene	Forward primer	Reverse primer
Cyanophage Syn9		
g61	5′-GGTTTGGGTATCAGGGAAGG-3′	5′-AACATCAGCACCACACATCG-3′
g43	5′-GAAGTTGGAGCCTTTCATCG-3′	5′-ACCTCACACCCTCACTGTCC-3′
g20	5′-AATTGAAATCCGCAATGAGC-3′	5′-CATAGCGGGATCCATTTCC-3′
g23	5′-AACCTACGAGCAAGCAGACG-3′	5′-ATTGCCTTCAGGTCTTGTGC-3′
psbA	5′-CGGTGGGTCACTTTTCTCG-3′	5′-CGACCGAAGTAACCATGAGC-3′
nrdA	5′-CTGGGCATTGGTTTTATTGG-3′	5′-CCTTTTTCCATTGCCATACG-3′
zwf	5′-TTCTCCATCGTCTGGATTGG-3′	5′-GCAATCCTGCTTCTTTGAGG-3′
gnd	5′-CTAAGGTGGCTGAGCTTTGG-3′	5′-ACAGCAGCGTGAACAGTCC-3′
talC	5′-CCCGAGCTTATTGCTACTGC-3′	5′-AATCTGCTGCCATACCAAGC-3′
cp12	5′-CATCGAAAAGCACATTCAGG-3′	5′-CCTCGCAGTAGAGCTCAAGG-3′
Synechococcus WH8109		
rnpB	5′-GCCGATCTCTTTGAGTGTCG-3′	5′-GCTCTTACCGCACCTTTGC-3′
Cyanophage P-HM2		
g61	5′-AGGAACCATTGGAAAGAACG-3′	5′-TCTCGCTTCATCGTTAGTGG-3′
g20	5′-CGTAGAGAAGGTGGCAGAGG-3′	5′-GACCTTCCGATGTTAAATTGC-3′
psbA	5′-ATCCCGCAGTAAAGGTCTCC-3′	5′-AGAAAGAACCCGCAACTGG-3′
nrdA	5′-TGGGATACAGTTCTTCGATGG-3′	5′-TCAGCACACTTTTTGACATGC-3′
nrdB	5′-GAATGTGTGGCAGTTGATGG-3′	5′-AAGGTCGTCATATGCTTTGG-3′
talC	5′-CGATGTTAGAGGAGGCAAGC-3′	5′-TGATCAATGCAAGACCATCG-3′
cp12	5′-AATCGAAGAACACATTGAAAAGG-3′	5′-AAAATAGTTCGAGTGCGTTGG-3′
Prochlorococcus MED4		
rnpB	5′-AAAGCAGGAGAGGCAATCG-3′	5′-TTAGGCGGTATGTTTCTGTGG-3′

Table S3. PCR primer sequences used in this study for cloning and expression of transaldolase genes

Gene	Vector	Forward primer	Reverse primer
MIT9312 *talB*	p15TvLic	5′-TTGTATTTCCAGGGCATGAAATCAATTTTAG AACAATTGTC-3′	5′-CAAGCTTCGTCATCAGTTGGCAGAAATTAATTTATGA TTTTTCA-3′
MED4 *talB*	pET100	5′-CACCATGAAATCAATTTTAGAACAATTATC-3′	5′-CTAAGTTGTCGAAATTAATTTTTGATTATTAATTTC-3′
NATL2A *talB*	pET100	5′-CACCATGGAATCCCTGCTGAGTCAGCTGTC-3′	5′-TCAGTGAGTTAGGGCAACTTCTCC-3′
P-SSM2 *talC*	pET101	5′-CACCATGAAAATCTTTTTAGATACTGCC-3′	5′-ACGCTTAACCTGAGCCC-3′
P-SSM4 *talC*	pET101	5′-CACCATGAAACTATTTTTAGATTGTTCAG-3′	5′-TCCTCCTACAAGTTTAGTCCAATC-3′
P-SSP7 *talC*	pET101	5′-CACCATGAAGATATTTCTGGATTCAG-3′	5′-GACATTTCTGCCAAAATCTAAGGC-3′

MIT9312, MED4, and NATL2A are *Prochlorococcus* strains; P-SSM2, P-SSM4, and P-SSP7 are cyanophage strains.

Table S4. Expression vectors used in this study for cloning and expression of transaldolase genes

Vector	Source	His-tag	Protein	Size, kDa
p15TvLic	Novagen and ref. 1	*N*-MGSSHHHHHHSSGENLYFQ↓G…	MIT9312 TalB	39.7
pET100	Invitrogen	*N*-MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDHPFT…	MED4 TalB	41.3
pET100	Invitrogen	*N*-MRGSHHHHHHGMASMTGGQQMGRDLYDDDDKDHPFT…	NATL2A TalB	40.6
pET101	Invitrogen	…KGELNSKLEGKPIPNPLLGLDSTRTGHHHHHH-C	P-SSM2 TalC	27.4
pET101	Invitrogen	…KGELNSKLEGKPIPNPLLGLDSTRTGHHHHHH-C	P-SSM4 TalC	27.1
pET101	Invitrogen	…KGELNSKLEGKPIPNPLLGLDSTRTGHHHHHH-C	P-SSP7 TalC	27.2

MIT9312, MED4, and NATL2A are *Prochlorococcus* strains; P-SSM2, P-SSM4, and P-SSP7 are cyanophage strains. The TEV cleavage site of MIT9312 TalB is marked with a down arrow.

1. Zhang RG, et al. (2001) Structure of *Thermotoga maritima* stationary phase survival protein SurE: A novel acid phosphatase. *Structure* 9:1095–1106.

Table S5. Data collection and refinement statistics for the *Prochlorococcus* MIT9312 TalB structure (PDB accession no. 3HJZ)

Data collection	Statistic
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c*, Å	42.8, 80.4, 97.9
Wavelength, Å	1.54178
Resolution, Å	40.19–1.9 (1.97–1.9)
R_{merge}, %*	0.083 (0.372)
$I/\sigma I$	22.25 (5.7)
Completeness, %	96.1 (68.3)
Redundancy	7.0 (5.5)
Refinement	
Resolution, Å	40.19–1.95
No. reflections	26,218
R_{work}, %†	15.6
R_{free}, %‡	20.3
No. atoms	
Protein	2,723
Water	341
Other	45
B-factors, Å^2	
Overall	16.5
Protein	15.1
Water	25.9
Other	35.0
Rmsd	
Bond lengths, Å	0.017
Bond angles, °	1.45
Ramachandran plot	
In most favored regions, %	93.1
In additionally allowed regions, %	6.9
In generously allowed or disallowed regions, %	0

Values in parentheses are for the highest-resolution shell.

*$R_{merge} = \sum_{hkl} |I - \langle I \rangle| / \sum_{hkl} I$.

†$R_{work} = \sum |F_{obs} - F_{calc}| / \sum |F_{obs}|$, where F_{obs} and F_{calc} are the observed and the calculated structure factors, respectively.

‡R_{free} is calculated by using 5% of total reflections randomly chosen and excluded from the refinement.

Table S6. Comparison of gene size between shared orthologues in cyanophages and host cyanobacteria

	Phage			Host			MWW/BH test		
Function	COG	*N*	Median	COG	*N*	Median	*P*	(*k*/*m*)*α	Significant?
Transaldolase	PhCOG205	22	654	CyCOG2913	24	1171.5	4.93×10^{-9}	2.08×10^{-3}	Yes
RNR	PhCOG2, 956	23	2301	CyCOG3062	24	2337	1.42×10^{-8}	4.17×10^{-3}	Yes
PhoH	PhCOG237	18	762	CyCOG3795	24	972	3.69×10^{-8}	6.25×10^{-3}	Yes
6PGDH	PhCOG 964	9	1038	CyCOG3200	24	1419	2.15×10^{-7}	8.33×10^{-3}	Yes
CpeT	PhCOG19, 762	13	459	CyCOG5460	27	606	4.06×10^{-7}	1.04×10^{-2}	Yes
Plastocyanin	PhCOG447	12	324	CyCOG2975	23	360	8.65×10^{-7}	1.25×10^{-2}	Yes
PurM	PhCOG112, 651	12	678	CyCOG4144	24	1039.5	1.02×10^{-6}	1.46×10^{-2}	Yes
G6PDH	PhCOG969	7	1440	CyCOG3565	24	1524	6.85×10^{-6}	1.67×10^{-2}	Yes
SpeD	PhCOG1089	7	336	CyCOG4127	24	448.5	7.30×10^{-5}	1.88×10^{-2}	Yes
Carboxylesterase	PhCOG496	5	363	CyCOG3327	29	249	4.74×10^{-4}	2.08×10^{-2}	Yes
Ferredoxin	PhCOG225	5	291	CyCOG3354	24	370.5	4.84×10^{-4}	2.29×10^{-2}	Yes
PTOX	PhCOG448	10	504	CyCOG2691, 10759	11	510	7.78×10^{-4}	2.50×10^{-2}	Yes
Phycoerythrobilin synthase	PhCOG423	4	636	CyCOG4148+4147	44	756	1.09×10^{-3}	2.71×10^{-2}	Yes
Heme oxygenase	PhCOG711	3	693	CyCOG4149	25	714	9.56×10^{-3}	2.92×10^{-2}	Yes
PstS	PhCOG174	10	972	CyCOG3129	40	982.5	1.48×10^{-2}	3.13×10^{-2}	Yes
PcyA	PhCOG20	4	711	CyCOG3177	24	736.5	3.63×10^{-2}	3.33×10^{-2}	No
Carbamoyltransferase	PhCOG396	4	1734	CyCOG3762	9	1854	4.20×10^{-2}	3.54×10^{-2}	No
Cobalt chelatase	PhCOG154	18	1089	CyCOG3041	24	1119	1.82×10^{-1}	3.75×10^{-2}	No
CP12	PhCOG18	15	228	CyCOG5346	27	219	2.29×10^{-1}	3.96×10^{-2}	No
Thymidylate synthase	PhCOG235	21	702	CyCOG2593, 5422, 8687	30	639	2.86×10^{-1}	4.17×10^{-2}	No
PsbA	PhCOG223	23	1080	CyCOG2573, 9386	69	1080	3.16×10^{-1}	4.38×10^{-2}	No
PsbD	PhCOG1166	13	1056	CyCOG3662	35	1056	4.33×10^{-1}	4.58×10^{-2}	No
High light-inducible protein	PhCOG213, 494	39	177	CyCOG2411, 2443, 3634, 3829, 3918, 3919, 3938, 4037, 5220, 7013, 9518	259	150	5.94×10^{-1}	4.79×10^{-2}	No
MazG	PhCOG182	18	403.5	CyCOG3367, 4240	48	640.5	7.97×10^{-1}	5.00×10^{-2}	No

For each protein function, the relevant COGs from ProPortal are given, along with the number *N* of genes in the COG(s) and the median gene size in bp. Mann–Whitney–Wilcoxon *P* values are given, and the rows are ordered by *P* value (smallest to largest). The expression (*k*/*m*)*α represents the rank *k* of the *P* value divided by the total number of tests *m* (24) multiplied by the *P* value cutoff α (0.05). Using the Benjamini–Hochberg correction for false discovery rate, the medians are considered significantly different up through the largest value for *k* such that $P < (k/m)*\alpha$.

Table S7. Protein and DNA content of a typical *Prochlorococcus* cell and a typical T4-like phage

Content	*Prochlorococcus* (strain MED4)	Phage (bacteriophage T4)
Protein	1.8×10^8 aa cell^{-1} 109 Da aa^{-1} = 2.0×10^{10} Da cell^{-1}	8.2×10^7 Da (head) 2.0×10^7 Da (baseplate and tail tube) 0.7×10^7 Da (tail fibers) = 1.1×10^8 Da virion^{-1}
DNA	1.66×10^6 bp cell^{-1} 615 Da bp^{-1} = 1.0×10^9 Da cell^{-1}	1.72×10^5 bp virion^{-1} 615 Da bp^{-1} = 1.1×10^8 Da virion^{-1}
Protein/DNA ratio	20.	1.0

Estimates are provided for *Prochlorococcus* MED4 and bacteriophage T4, which closely resembles T4-like cyanophages in its ultrastructure. The estimate of *Prochlorococcus* MED4 amino acid composition is from Waldbauer (1): the midpoint of the range (0.75×10^8 to 2.9×10^8) of amino acids per MED4 cell was used, and the average amino acid mass of the 51 most abundant proteins in MED4 was used. The copy numbers and molecular weights of T4 proteins from Leiman et al. (2) were used to calculate the total masses of the head, baseplate, tail tube, and tail fiber proteins. The chromosome size of T4 is 102% of its genome size (1.69×10^5 bp) because of terminal redundancy.

1. Waldbauer JR (2009) Molecular biogeochemistry of modern and ancient marine microbes. PhD Thesis (Massachusetts Institute of Technology and Woods Hole Oceanographic Institution, Woods Hole, MA).
2. Leiman PG, Kanamaru S, Mesyanzhinov VV, Arisaka F, Rossmann MG (2003) Structure and morphogenesis of bacteriophage T4. *Cell Mol Life Sci* 60:2356–2370.

Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism

Luke R. Thompson[a,1], Qinglu Zeng[b,1], Libusha Kelly[b], Katherine H. Huang[b], Alexander U. Singer[c], JoAnne Stubbe[a,d,2], and Sallie W. Chisholm[a,b,2]

Departments of [a]Biology and [b]Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario Canada M5G 1L6; and [d]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

AUTHOR SUMMARY

Approximately half of all photosynthesis on Earth takes place in the ocean, and a sizable portion of that is carried out by tiny (about 1 μm in diameter) unicellular cyanobacteria, or blue-green algae, called *Prochlorococcus* and *Synechococcus*, which often reach densities of 100 million per liter of seawater. Even more abundant are viruses, called cyanophage, that can attach to these "host" cells, inject their DNA, and use the host's biochemical machinery to make more phage. The phage then break open the host cell, killing it, and release their progeny into the seawater. One of the ways cyanophages are thought to take over their hosts is via the use of host-like metabolic genes—acquired over eons of intimate coevolution—that encode enzymes that boost metabolic steps that might be bottlenecks in the phage reproduction process. An apparent contradiction arises, however, in that many cyanophages carry genes for both photosynthesis and a carbon metabolism pathway called the pentose phosphate pathway, and these two processes achieve opposite goals, with photosynthesis "fixing" carbon and the pentose phosphate pathway "burning" carbon. Are host-like metabolic genes in cyanophage steering host metabolism toward fixing carbon or toward burning carbon? In this work, we used a combination of physiology, enzymology, and sequencing approaches to address this question. Our evidence suggests that cyanophages direct host metabolism to mobilize carbon stores, burning but not fixing carbon to fuel the synthesis of DNA building blocks and phage replication.

The first clues to the strategy used by cyanophages for directing host metabolism came from the analysis of cyanophage genomes. These genomes revealed the widespread presence of the "auxiliary metabolic genes" mentioned earlier, which encode proteins similar to those used in host metabolism. These genes tend to be absent from viruses infecting noncyanobacterial hosts, suggesting that they are a specific adaptation suited to infecting cyanobacteria. The metabolic pathways that these phage genes are associated with in host cells include the light reactions of photosynthesis (1), the pentose phosphate pathway (1), nutrient acquisition pathways (2), and pathways for nucleotide biosynthesis that are also commonly found in noncyanophages (2).

To make sense of host cyanobacterial metabolism as it relates to phage infection, we first constructed a simple model of host metabolism (Fig. P1*A*), facilitated by an analysis of gene expression patterns over the light–dark cycle of *Prochlorococcus* (3). In this model, cyanobacteria use the light reactions of photosynthesis to harness light energy (*hv*) and split water to produce NADPH (electron carrier) and ATP (energy carrier), using these metabolites to fix CO_2 in the Calvin cycle (green) and produce glucose 6-phosphate during the day. At night, this sugar is oxidized in the pentose phosphate pathway (red) to NADPH and ribose 5-phosphate, which can be used for carbon skeletons or recycled back through the pathway. Finally, nucleotide biosynthesis takes the NADPH and ribose 5-phosphate produced by the pentose phosphate pathway, combined with NADPH and ATP produced by the light reactions of photosynthesis, to produce nucleotides, the DNA building blocks of chromosomes. In essence, taking energy from the sun, electrons from water, and

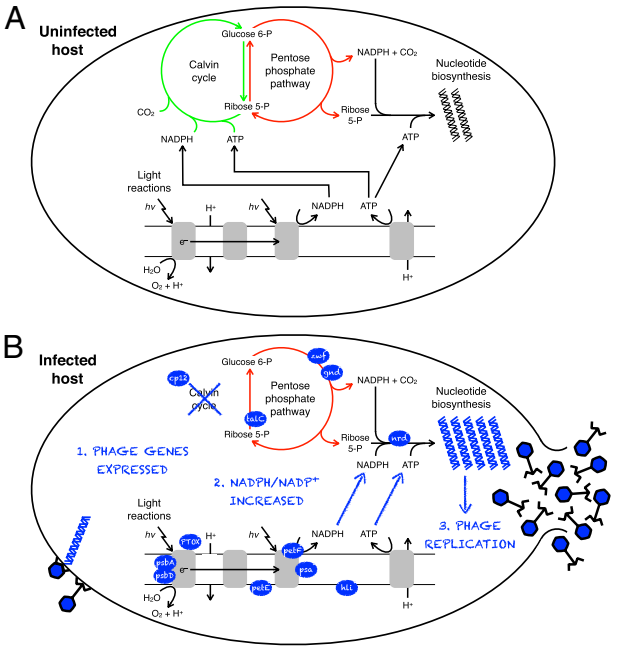

Fig. P1. Model of cyanobacterial metabolism during cyanophage infection. (*A*) In uninfected cells, four interrelated pathways combine to make DNA building blocks (nucleotides): the light reactions of photosynthesis, the Calvin cycle (the so-called "dark reactions" of photosynthesis), the pentose phosphate pathway, and nucleotide biosynthesis. (*B*) In infected cells, host-like genes (blue ovals) acquired by phage over evolutionary time are proposed to augment or inhibit key steps in host metabolism, leading to increased nucleotide biosynthesis for phage reproduction.

Author contributions: L.R.T., Q.Z., L.K., K.H.H., J.S., and S.W.C. designed research; L.R.T., Q.Z., L.K., K.H.H., and A.U.S. performed research; L.R.T., Q.Z., L.K., K.H.H., and A.U.S. analyzed data; and L.R.T., J.S., and S.W.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID code 3HJZ). The sequences reported in this paper have been deposited in the GenBank database [GU071107 (P-SSP2), GU071104 (P-HP1), and GU071102 (P-RSP5)].

[1]L.R.T. and Q.Z. contributed equally to this work.

[2]To whom correspondence may be addressed: chisholm@mit.edu or stubbe@mit.edu.

See full research article on page E757 of www.pnas.org.

Cite this Author Summary as: PNAS 10.1073/pnas.1102164108.

carbon from carbon dioxide, cyanobacteria are able to make sugar and transform it into nucleotides for replicating their genomes.

In the context of this simplified view of host metabolism, we noticed something peculiar about the collection of host-like genes found in cyanophage genomes: Genes for the light reactions and the pentose phosphate pathway were commonly found, as were those for nucleotide biosynthesis; however, no genes for the Calvin cycle were apparent, despite a search of more than 20 phage genomes. Although the presence of cyanophage photosynthesis proteins initially led to the proposal of increased carbon fixation in the sea (4), the absence of Calvin cycle genes and the presence of pentose phosphate genes suggested that an alternative model was required. Our model (Fig. P1*B*) suggests that NAPDH and ATP, the products of the light reactions, and redirection of carbon flux away from the Calvin cycle and through the pentose phosphate pathway to generate ribose 5-phosphate and NADPH, together would lead to enhanced nucleotide biosynthesis required for phage replication.

In support of this hypothesis was the discovery of a gene for a Calvin cycle inhibitor, CP12, in many phage genomes (5). CP12 is a host regulatory protein that binds to and inhibits two enzymes in the Calvin cycle, limiting carbon flux through this pathway and favoring flux through the pentose phosphate pathway. Thus, we propose that phages use the CP12 that they encode to shut down the Calvin cycle in the host, and redirect the products of the light reactions to their own ends.

Here we reveal that the *cp12* gene is widespread in both cultured and wild cyanophages, and is expressed during infection, concurrently with phage genes involved in the pentose phosphate pathway, genes for the light reactions of photosynthesis, and nucleotide biosynthesis. Thus, cyanophage infection appears to short-circuit host carbon metabolism, shutting down the Calvin cycle while boosting the three other pathways. As shown in Fig. P1*B*, cyanophages carry many genes (blue ovals) with putative functions in host metabolism. According to our model, the light reactions, aided by phage-encoded proteins, lead to increased production of NADPH and ATP. These storage molecules are not used to power carbon fixation because the Calvin cycle is blocked by phage-encoded CP12, forcing carbon flux through the pentose phosphate pathway, boosted by phage-encoded enzymes. NADPH, ATP, and ribose 5-phosphate produced by these processes are used to power phage nucleotide biosynthesis.

Several additional pieces of evidence support this model. NADPH, an electron carrier, is a product of the light reactions and the pentose phosphate pathway, and NADP is its electron-poor version. We show that, under infection, the ratio of NADPH to NADP increases. This result is compatible with increased activity of the light reactions and the pentose phosphate pathway for production of NADPH, the source of electrons for the synthesis of DNA building blocks. Furthermore, calculations imply that replicating cyanophages require significant de novo production of these nucleotides, and our model suggests how this might occur.

Finally, we studied the properties of the phage transaldolase enzyme, whose gene is the most widespread pentose phosphate pathway gene in cyanophages. The phage transaldolase is significantly shorter than the host transaldolase, and therefore we wondered why this enzyme would have been evolutionarily selected by phages over the host enzyme. We proposed that the phage transaldolase would have a higher activity than the host transaldolase. We therefore cloned, expressed, and purified several phage and host transaldolase enzymes, and measured their kinetic properties in vitro. We found that, to our surprise, the phage enzymes in fact had one third the kinetic efficiencies of the host enzymes. We therefore sought other explanations for phage use of this smaller enzyme. We hypothesized that the smaller size of the phage gene could be a significant benefit to the phage, with its small genome and limited resources from the host for replication. Indeed, we found that many phage genes are significantly shorter than their host orthologues, providing an evolutionary explanation for differences in gene sizes of shared phage–host orthologues in general.

The marine cyanobacteria–cyanophage system has proven to be an intriguing example of coevolution. In this study, we have attempted to tie the various evidentiary threads together to unveil the selective pressures that have led cyanophages to incorporate a particular set of host-like genes into their genomes. The key finding that cyanophages express the Calvin cycle inhibitor CP12 suggests that there are strong selective pressures for cyanophages to redirect host metabolism toward mobilizing energy, electrons, and carbon skeletons for DNA biosynthesis. Ultimate validation of this model will require a genetic system for phage and host, which is at present not available for these strains.

1. Millard A, Clokie MRJ, Shub DA, Mann NH (2004) Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. *Proc Natl Acad Sci USA* 101: 11007–11012.
2. Sullivan MB, Coleman ML, Weigele P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. *PLoS Biol* 3:e144.
3. Zinser ER, et al. (2009) Choreography of the transcriptome, photophysiology, and cell cycle of a minimal photoautotroph, *Prochlorococcus*. *PLoS ONE* 4:e5135.
4. Sharon I, et al. (2007) Viral photosynthetic reaction center genes and transcripts in the marine environment. *ISME J* 1:492–501.
5. Sullivan MB, et al. (2010) Genomic analysis of oceanic cyanobacterial myoviruses compared with T4-like myoviruses from diverse hosts and environments. *Environ Microbiol* 12:3035–3056 10.1111/j.1462–2920.2010.02280.x.