# Prevalence and Evolution of Core Photosystem II Genes in Marine Cyanobacterial Viruses and Their Hosts

**Matthew B. Sullivan[1☯], Debbie Lindell[1☯], Jessica A. Lee[2], Luke R. Thompson[2], Joseph P. Bielawski[3,4], Sallie W. Chisholm[1,2*]**

**1** Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **2** Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, **3** Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada, **4** Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada

**Cyanophages (cyanobacterial viruses) are important agents of horizontal gene transfer among marine cyanobacteria, the numerically dominant photosynthetic organisms in the oceans. Some cyanophage genomes carry and express host-like photosynthesis genes, presumably to augment the host photosynthetic machinery during infection. To study the prevalence and evolutionary dynamics of this phenomenon, 33 cultured cyanophages of known family and host range and viral DNA from field samples were screened for the presence of two core photosystem reaction center genes, *psbA* and *psbD*. Combining this expanded dataset with published data for nine other cyanophages, we found that 88% of the phage genomes contain *psbA,* and 50% contain both *psbA* and *psbD*. The *psbA* gene was found in all myoviruses and *Prochlorococcus* podoviruses, but could not be amplified from *Prochlorococcus* siphoviruses or *Synechococcus* podoviruses. Nearly all of the phages that encoded both *psbA* and *psbD* had broad host ranges. We speculate that the presence or absence of *psbA* in a phage genome may be determined by the length of the latent period of infection. Whether it also carries *psbD* may reflect constraints on coupling of viral- and host-encoded PsbA–PsbD in the photosynthetic reaction center across divergent hosts. Phylogenetic clustering patterns of these genes from cultured phages suggest that whole genes have been transferred from host to phage in a discrete number of events over the course of evolution (four for *psbA,* and two for *psbD*), followed by horizontal and vertical transfer between cyanophages. Clustering patterns of *psbA* and *psbD* from *Synechococcus* cells were inconsistent with other molecular phylogenetic markers, suggesting genetic exchanges involving *Synechococcus* lineages. Signatures of intragenic recombination, detected within the cyanophage gene pool as well as between hosts and phages in both directions, support this hypothesis. The analysis of cyanophage *psbA* and *psbD* genes from field populations revealed significant sequence diversity, much of which is represented in our cultured isolates. Collectively, these findings show that photosynthesis genes are common in cyanophages and that significant genetic exchanges occur from host to phage, phage to host, and within the phage gene pool. This generates genetic diversity among the phage, which serves as a reservoir for their hosts, and in turn influences photosystem evolution.**

Citation: Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, et al. (2006) Prevalence and evolution of core photosystem II genes in marine cyanobacterial viruses and their hosts. PLoS Biol 4(8): e234. DOI: 10.1371/journal.pbio.0040234

## Introduction

The marine cyanobacteria *Prochlorococcus* and *Synechococcus* are the smallest and most numerous photosynthetic cells in the oceans [1,2]. The abundances of cyanophages (cyanobacterial viruses) that infect these marine cyanobacteria vary over spatial [3–6] and temporal scales [4,7]—patterns shaped by the dynamics of their host cells [4,8]. Cyanophages are double-stranded DNA viruses belonging to three morphologically defined families: Podoviridae, Myoviridae, and Siphoviridae [3–5,9,10]. Among the cyanophages, podoviruses and siphoviruses tend to be very host-specific, whereas myoviruses generally have a broader host range, even across genera [5], and thus are potential vectors for horizontal gene transfer via transduction.

The movement of genes between organisms is an important mechanism in evolution. As agents of gene transfer, phages play a role in host evolution by supplying the host with new genetic material [11–15] and by displacing "host" genes with viral-encoded homologues [16–18]. Phage evolution is in turn influenced by the acquisition of DNA from their hosts [13,19–22] and by the swapping of genes within the phage gene pool [23,24]. Recent evidence suggests that gene flow within the global phage gene pool extends across ecosystems [25–27].

Cyanophage genomes bearing key photosynthesis genes *psbA* and *psbD* provide a notable example of the co-option of "host" genes for phage purposes [13,22,28–30]. The *psbA* and *psbD* genes encode the two photosystem II core reaction

**Academic Editor:** Nancy A. Moran, University of Arizona, United States of America

**Received** February 13, 2006; **Accepted** May 11, 2006; **Published** July 4, 2006

**DOI:** 10.1371/journal.pbio.0040234



**Abbreviations:** HL, high-light adapted; LL, low-light adapted

* To whom correspondence should be addressed. E-mail: chisholm@mit.edu

☯ These authors contributed equally to this work.

center proteins, D1 and D2 (denoted here as PsbA and PsbD, respectively), found in all oxygenic photosynthetic organisms. It has recently been shown that the phage-encoded *psbA* gene is expressed during infection [31,32]. Because maximal cyanophage production is dependent on photosynthesis [31,33], and the host PsbA protein turns over rapidly [34] and declines during infection [31], expression of these phage-encoded genes likely enhances photosynthesis during infection, thus increasing cyanophage fitness.

If photosynthesis genes indeed provide a fitness advantage to cyanophages, one might expect them to be widespread among cyanophage genomes. Through whole or partial genome sequencing, *psbA* has been documented in three *Prochlorococcus* cyanophages (one podovirus and two myoviruses) and five *Synechococcus* myoviruses, whereas *psbD* was found in only some of these phages [13,29,35]. Neither of these genes is found in the *Synechococcus* P60 podovirus genome [36]. A survey of *Synechococcus* myovirus isolates revealed that at least 37 of them contained *psbA* [29], and this gene has also been found in cyanophage genome fragments in seawater samples [37]. Thus, the presence of *psbA* is a common, but not universal, feature in the cyanophages examined to date, most of which have been *Synechococcus* cyanophages.

Using limited genomic sequence data from one *Synechococcus* and three *Prochlorococcus* cyanophages, we suggested that both *psbA* and *psbD* were transferred as whole genes from host to phage multiple times, but not from phage to host [13]. Subsequently, Zeidner et al. [37] analyzed *psbA* data predominantly from field sequences and suggested that genetic exchanges of segments of the gene (intragenic recombination) may have occurred among host and phage copies in both directions [37]. However, this novel and controversial hypothesis requires further investigation with sequences of known organismal origin and using methodology capable of identifying the recombination partners and the directionality of such potential exchanges.

To better describe and understand the phenomenon of photosynthesis genes in cyanophage, we looked for the *psbA* and *psbD* genes in 33 cultured cyanophage isolates that infect *Synechococcus* or *Prochlorococcus* (or both) and analyzed the sequences of these genes in the context of known host ranges of the phage. This dataset allowed us to address the following questions: (1) How prevalent are both *psbA* and *psbD* in cyanophages that infect *Synechococcus* and/or *Prochlorococcus*? and (2) To what extent have photosynthesis genes, or segments thereof, been moved between and among hosts and phages?

## Results/Discussion

### Prevalence of the *psbA* and *psbD* Genes in Cyanophages

The *psbA* gene was amplified from 28 out of the 33 cyanophage isolates examined (Table 1). Combining these findings with published results (Table 1), we find that the *psbA* gene is present in 88% of cyanophage isolates examined, including all myoviruses ($n = 32$) and all five *Prochlorococcus* podoviruses included in this study. However, this gene was not detected in *Prochlorococcus* siphoviruses ($n = 2$) and *Synechococcus* podoviruses ($n = 3$), suggesting that there are some combinations of phage family and host genus that do not lead to incorporation of the *psbA* gene into the phage genome. Six additional phages yielded ambiguous results and were excluded from these analyses (see Materials and Methods for details).

When present, the *psbA* gene is likely to be functional, as there is evidence for the conservation of amino acid sequences through purifying selection [13,37], and the gene is expressed during infection [31,32], implying that this gene confers a fitness advantage to the phages that carry it [13,22,29,31]. Sustained photosynthesis is necessary for maximal phage production [31,33,38], and the long latent period of many freshwater and marine cyanophages (8 h or more; [9,31,33,38]) presumably results in energy- and/or carbon-limitation for phage replication. Thus, cyanophage-encoded *psbA* likely serves to boost the photosynthetic performance of the host during infection, thereby increasing phage production. It is perhaps not coincidental that one of the phages that lacks *psbA*, *Synechococcus* podovirus P60 (Table 1), has a latent period of only 1 h (K. Wang and F. Chen, personal communication), which may be too short for *psbA* expression to be beneficial. Latent period information for marine cyanophages, however, is sparse. It is not known for the *Prochlorococcus* siphoviruses that lack *psbA*, and it has only been shown to be >8 h for a single phage strain from each of the *Synechococcus* myoviruses [39] and *Prochlorococcus* podoviruses [31]. Further, theory [40–43] and experiments [44] suggest that latent period length may be a transient property that rapidly evolves in response to changes in host cell densities. Thus, further exploration of this hypothesis requires analysis of the latent period of many more phage isolates under variable host cell concentrations.

The *psbD* gene was amplified from 15 out of the 33 cyanophage isolates examined (Table 1). Again, combining our data with published findings, we observe that *psbD* is found only in isolates that contain *psbA* and only in myoviruses, but not in all *psbA*-containing myoviruses. Only four of 12 *Prochlorococcus* myoviruses (as defined by original host strain of isolation; Table 1) contained *psbD,* whereas this was the case for 17 of 20 *Synechococcus* myoviruses. Although it is possible that differences in the photosystem II reaction center between *Prochlorococcus* and *Synechococcus* exist (such as differences in the rate of PsbD degradation) and could explain the biased distribution of the *psbD* gene among the myoviruses, there is no evidence that this is the case. The breadth of phage host ranges (as operationally defined in Table 1), however, appears to be a reasonably good predictor of whether a phage will contain *psbD*: 17 of 18 broad-host-range phages encode it, whereas only one out of 21 narrow-host-range phages do so (Table 1). Perhaps broad-host-range phages have co-opted both *psbA* and *psbD* to better ensure the formation of a functional PsbA–PsbD protein complex in the host during infection.

### Origins and Evolutionary History of *psbA* and *psbD* in Cyanophages

To investigate the origins of photosynthesis genes in phages and their hosts, we conducted phylogenetic analyses (using measures to minimize systematic errors; see Materials and Methods) of host and phage *psbA* and *psbD* sequences, including new sequence data for nine *Synechococcus* hosts *(psbA),* 19 *Synechococcus* and *Prochlorococcus* hosts *(psbD),* and 33 phages (both *psbA* and *psbD*). Phylogenetic reconstructions of host *psbA* and *psbD* genes in *Prochlorococcus* showed that well-

**Table 1.** Presence or Absence of *psbA* and *psbD* among *Prochlorococcus* and *Synechococcus* Cyanophages

| Family | Phage Name | Cross-Infection[a] | Original Host[a] | Number of Known Hosts[b] | Host Range Breadth[d] | *psbA* | *psbD* | Genome Sequence Confirmation[e] | Reference for *psbA* and *psbD* Sequence |
|---|---|---|---|---|---|---|---|---|---|
| Podoviridae | P-SSP3 | | *Prochlorococcus* MIT9312 | 2 | Narrow | + | − | − | This study |
| | P-SSP5 | | *Prochlorococcus* MIT9515 | 1 | Narrow | + | − | − | This study |
| | P-SSP6 | | *Prochlorococcus* MIT9515 | 1 | Narrow | + | − | − | This study |
| | P-SSP7 | | *Prochlorococcus* MED4 | 1 | Narrow | + | − | Y | [13] |
| | P-GSP1 | | *Prochlorococcus* MED4 | 1 | Narrow | + | − | − | This study |
| | Syn12 | | *Synechococcus* WH8017 | 2 | Narrow | − | − | − | This study |
| | Syn5 | | *Synechococcus* WH8109 | 1 | Narrow | − | − | Y | This study; P. Weigele, W. Pope, G. Hatfull, R. Hendrix, unpublished data |
| | P60 | | *Synechococcus* WH7803 | 1 | Narrow | − | − | Y | [36] |
| Myoviridae | P-SSM8 | | *Prochlorococcus* MIT9211 | 2[c] | Narrow | + | − | − | This study |
| | P-SSM1 | | *Prochlorococcus* MIT9303 | 3 | Broad | + | + | − | This study |
| | P-RSM4 | | *Prochlorococcus* MIT9303 | 1[c] | Narrow | + | − | − | This study |
| | P-SSM2 | | *Prochlorococcus* NATL1A | 3 | Narrow | + | − | Y | [13] |
| | P-RSM5 | | *Prochlorococcus* NATL1A | 1[c] | Narrow | + | − | − | This study |
| | P-SSM3 | | *Prochlorococcus* NATL2A | 3 | Narrow | + | − | − | This study |
| | P-SSM4 | | *Prochlorococcus* NATL2A | 4 | Broad | +, ID to P-RSM2, P-RSM3 | + | Y | [13] |
| | P-SSM9 | | *Prochlorococcus* NATL2A | 2[c] | Narrow | +, ID to P-SSM12 | − | − | This study |
| | P-SSM10 | | *Prochlorococcus* NATL2A | 1[c] | Narrow | + | − | − | This study |
| | P-SSM12 | | *Prochlorococcus* NATL2A | 2[c] | Narrow | +, ID to P-SSM9 | − | − | This study |
| | P-RSM2 | Δ | *Prochlorococcus* NATL2A | 6 | Broad | +, ID to P-RSM3, P-SSM4 | + | − | This study |
| | P-RSM3 | Δ | *Prochlorococcus* NATL2A | 6 | Broad | +, ID to P-RSM2, P-SSM4 | + | − | This study |
| | S-SM1 | | *Synechococcus* WH6501 | 2 | Narrow | + | + | − | This study |
| | S-ShM1 | | *Synechococcus* WH6501 | 2 | Narrow | + | − | − | This study |
| | S-SSM1 | | *Synechococcus* WH6501 | 2 | Narrow | + | − | − | This study |
| | syn33 | Δ | *Synechococcus* WH7803 | 8 | Broad | + | + | − | This study |
| | S-WHM1 | Δ | *Synechococcus* WH7803 | 5 | Broad | + | + | Y | [29] |
| | S-PM2 | | *Synechococcus* WH7803 | 2 | Broad | + | + | Y | [29] |
| | S-RSM2 | na | *Synechococcus* WH7803 | Unknown | N.D. | + | + | Y | [29] |
| | S-BM4 | na | *Synechococcus* WH7803 | Unknown | N.D. | + | + | Y | [29] |
| | S-RSM88 | na | *Synechococcus* WH7803 | Unknown | N.D. | + | + | Y | [29] |
| | syn9 | Δ | *Synechococcus* WH8012 | 13 | Broad | + | + | Y | This study; P. Weigele, W. Pope, G. Hatfull, R. Hendrix, unpublished data |
| | syn10 | Δ | *Synechococcus* WH8017 | 7 | Broad | +, ID to syn26 | + | − | This study |
| | syn26 | Δ | *Synechococcus* WH8017 | 9 | Broad | +, ID to syn10 | + | − | This study |
| | S-SSM3 | Δ | *Synechococcus* WH8018 | 5[c] | Broad | +, ID to S-SSM5 | + | − | This study |
| | syn30 | Δ | *Synechococcus* WH8018 | 7 | Broad | + | + | − | This study |
| | syn1 | | *Synechococcus* WH8101 | 4 | Broad | + | + | − | This study |
| | S-ShM2 | Δ | *Synechococcus* WH8102 | 9 | Broad | + | + | − | This study |
| | S-SSM2 | Δ | *Synechococcus* WH8102 | 9 | Broad | + | + | − | This study |
| | S-SSM5 | Δ | *Synechococcus* WH8102 | 6[c] | Broad | +, ID to S-SSM3 | + | − | This study |
| | S-SSM6 | Δ | *Synechococcus* WH8109 | 7[c] | Broad | + | − | − | This study |
| | syn19 | Δ | *Synechococcus* WH8109 | 9 | Broad | + | + | − | This study |
| Siphoviridae | P-SS1 | | *Prochlorococcus* MIT9313 | 1 | Narrow | − | − | − | This study |
| | P-SS2 | | *Prochlorococcus* MIT9313 | 1 | Narrow | − | − | − | This study |

Presence is indicated by +, and absence by −.

Phages that contained identical sequences to other phages are noted as "ID to X."

[a]Cultured strain used for isolation of phage from natural seawater samples. Phages are defined as either *Prochlorococcus* or *Synechococcus* phages based on original host of isolation, but many of the myoviruses cross-infect both genera. Those phages that cross-infect both genera are marked "Δ", those that do not are left blank, and those that were not available for testing are marked "na".

[b]The number of host strains infected by each phage out of 21 strains tested [5].

[c]Phages whose host ranges are first reported here: P-SSM8, *Prochlorococcus* MIT9211 and MIT9303; P-RSM4, *Prochlorococcus* MIT9313; P-RSM5, *Prochlorococcus* NATL1A; P-SSM9, *Prochlorococcus* NATL1A and NATL2A; P-SSM10, *Prochlorococcus* NATL2A; P-SSM12, *Prochlorococcus* NATL2A and NATL1A; S-SSM3, *Prochlorococcus* NATL1A and NATL2A, *Synechococcus* WH7803, WH8102 and WH8109; S-SSM5, *Prochlorococcus* MIT9303 and MIT9313, *Synechococcus* WH7803, WH8102, WH8103, and WH8109; S-SSM6, *Prochlorococcus* strains NATL1A, MIT9215, and MIT9211, *Synechococcus* strains WH6501, WH8017, WH8018, and WH8109.

[d]Operationally defined as narrow if a phage infects less than four hosts within a single cluster determined from 16S-23S rRNA internal transcribed spacer clustering [48] and broad if it infects more than four hosts within a cluster or at least two hosts that span more than one cluster. Small variations in this definition did not significantly affect the conclusions made.

[e]Indicates whether the PCR results were corroborated by genome sequencing. In all cases where the genome sequence became available, it confirmed the PCR results.

DOI: 10.1371/journal.pbio.0040234.t001

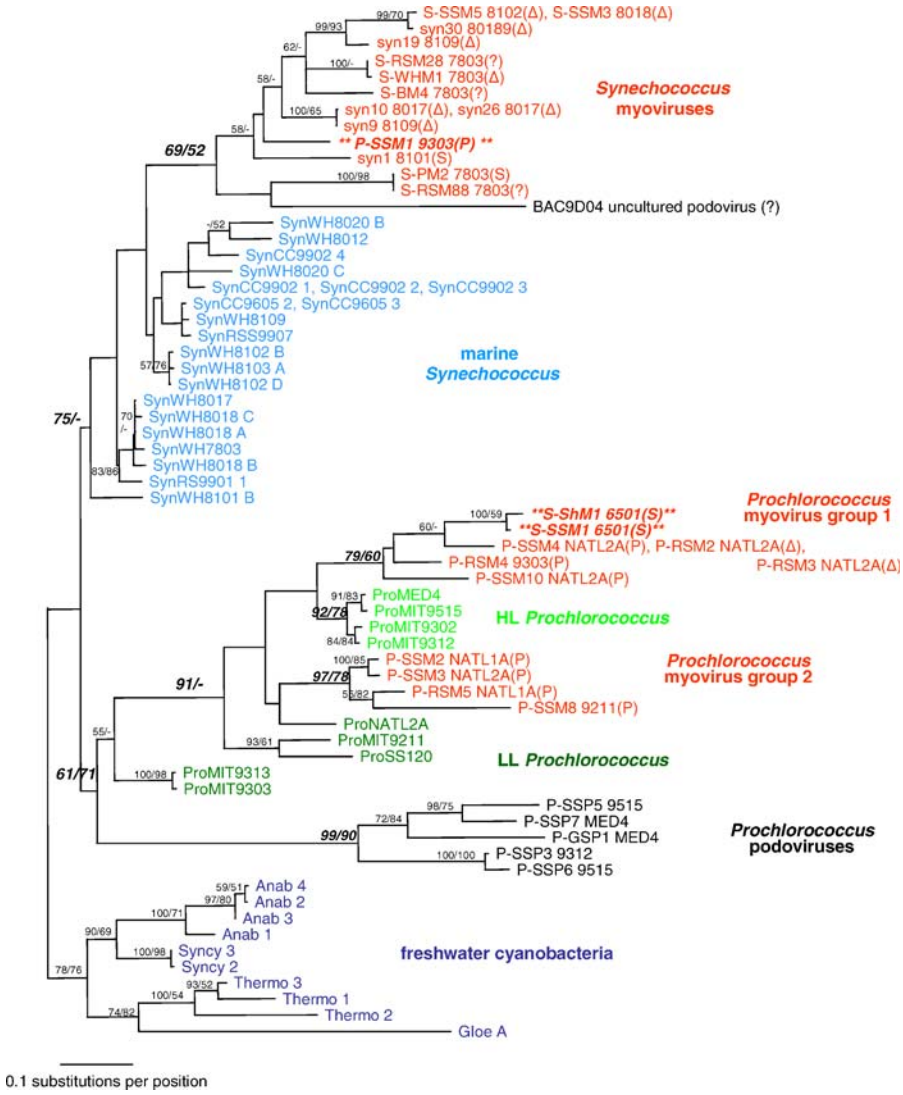


**Figure 1.** Phylogenetic Tree of *psbA* Gene Sequences from Cultured Cyanobacteria and Cyanophages

Phages are listed by their name, followed by their original host. Phages that are known to infect both *Prochlorococcus* and *Synechococcus* hosts are indicated with a "Δ"; those that infect only one genus are labeled either P (infect only *Prochlorococcus* hosts) or S (infect only *Synechococcus* hosts), while those that are unknown are designated with a "?". Phages shown in italics and bracketed with "**" were isolated on hosts that do not belong to the same cluster and are thus exceptions to the general clustering pattern (see text). Taxa are color coded according to the following biological groupings: myoviruses (red), podoviruses (black), marine *Synechococcus* hosts (light blue), marine *Prochlorococcus* hosts (dark green, LL; light green, HL), freshwater cyanobacteria (dark blue). The tree topology was estimated by LogDet analysis of 1st and 2nd codon positions. Sequences where intragenic recombination was detected using other methods (see Materials and Methods) were not included in these phylogenetic analyses. Branch lengths were estimated by maximum likelihood under a model with nonstationary nucleotide frequencies. Numbers at the nodes represent neighbor-joining bootstrapping and maximum likelihood puzzling support. Anab, *Anabaena;* Gloe, Gleobacter; HL, high-light adapted; LL, low-light adapted; Syncy, *Synechocystis;* Thermo, *Thermosynechococcus*.

DOI: 10.1371/journal.pbio.0040234.g001

supported sequence clusters contain only one organism type (Figures 1 and 2), with sequences from high-light adapted (HL) and low-light adapted (LL) *Prochlorococcus* [45] forming discrete clusters. These well-supported *Prochlorococcus* clusters are similar to those observed using other host genes such as *rRNA*, *rpoC1*, and *ntcA* [46–49], indicating that *psbA* and *psbD* have not been transferred between *Prochlorococcus* lineages. In contrast, the *Synechococcus* clusters for both *psbA* and *psbD* are poorly supported, a finding different to that obtained using other highly conserved genes [46–49] and thus may have resulted from genetic exchange between *Synechococcus* lineages.

The *psbA* sequences from *Synechococcus* myoviruses, *Prochlorococcus* myoviruses, and *Prochlorococcus* podoviruses generally formed discrete clusters consistent with their host ranges (Figure 1), suggesting that the transfer of photosynthesis genes from host to phage has been largely limited by host range (but see exceptions discussed below). Although many of these phages are capable of infecting both host genera (denoted as "Δ" in all figures), we designated each cyanophage isolate as a *Prochlorococcus* or *Synechococcus* cyanophage based upon its original host strain of isolation (as mentioned above and in Table 1). Given this designation scheme, it appears that transfers were predominantly from *Prochlorococcus* to their phages and from *Synechococcus* to their phages. This suggests host-range-limited host-to-phage transfer events,

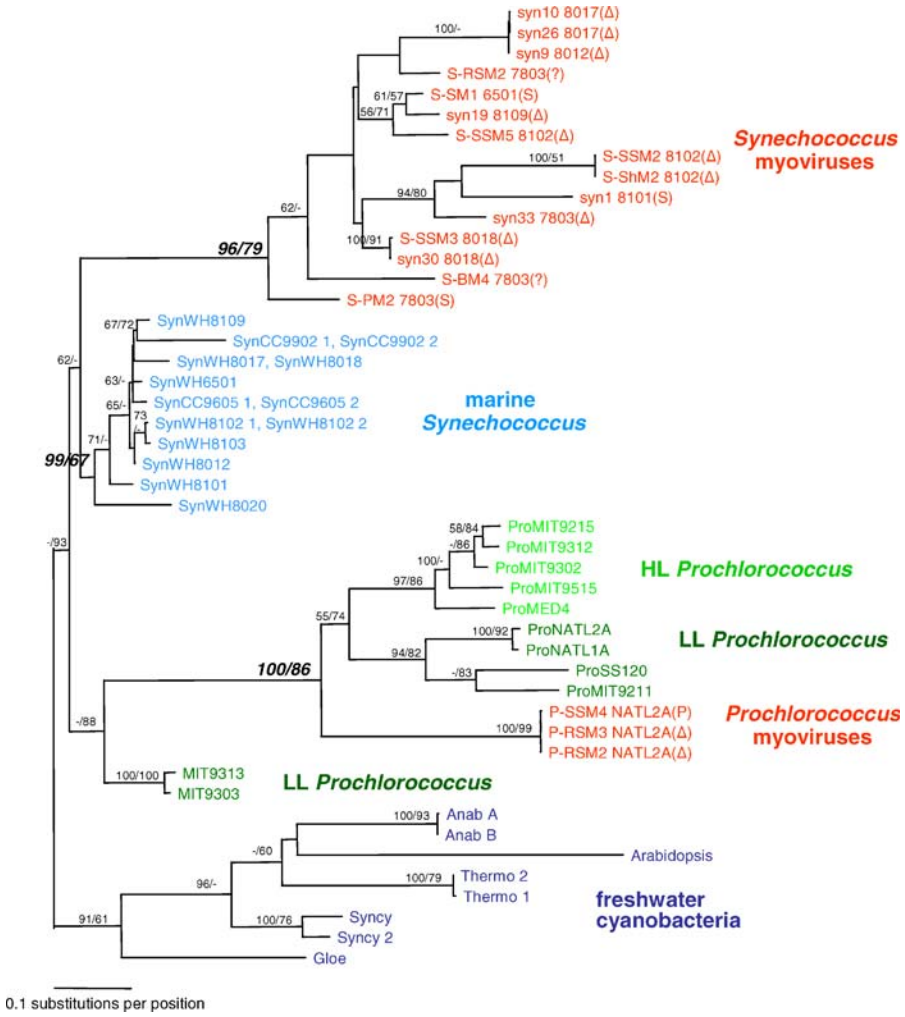


**Figure 2.** Phylogenetic Tree of *psbD* Gene Sequences from Cultured Cyanobacteria and Cyanophages

Details as in Figure 1. Sequences where intragenic recombination was detected using other methods (e.g., P-SSM1) were not included in these phylogenetic analyses.

DOI: 10.1371/journal.pbio.0040234.g002

with subsequent horizontal and vertical transfers occurring among viral lineages.

Two isoforms of the PsbA protein are often found in cyanobacteria [50]. The PsbA.1 (D1.1) isoform is constitutively expressed, whereas the PsbA.2 (D1.2) isoform is upregulated in response to high light and UV stress [51,52]. Many of the differences between the isoforms are found in ten amino acids between position 121 and 312 [50]. Based on which isoform the majority of these ten amino acids were identical to (including glutamine/glutamate at position 130), we determined that PsbA from both *Prochlorococcus* myoviruses and podoviruses are more similar to PsbA.1, the only isoform found in *Prochlorococcus* hosts so far [53] (unpublished data). Although *Synechococcus* hosts encode both isoforms (unpublished data), *Synechococcus* myoviruses encode the stress-responsive PsbA.2 isoform exclusively (unpublished data), which may be particularly beneficial during the stress of infection. These findings are consistent with the hypothesis of host-range-limited transfers of the *psbA* gene (but see exceptions below).

Host-to-phage transfers appear to have occurred at least four times for *psbA* and twice for *psbD*, as seen from the number of discrete clades containing phage-encoded genes in each case (Figures 1 and 2). The four *psbA* gene acquisitions by phage appear to include two transfer events for the *Prochlorococcus* myoviruses (*Prochlorococcus* myovirus group 1 and 2 in Figure 1) and a single event for *Prochlorococcus* podoviruses all from their *Prochlorococcus* hosts, as well as a single event for *Synechococcus* myoviruses from their hosts (Figure 1). The *psbD* gene appears to have been acquired once by both *Synechococcus* and *Prochlorococcus* myoviruses from their respective hosts (Figure 2). Interestingly, the three *Prochlorococcus* myoviruses that contain *psbD* all encode *Prochlorococcus* myovirus group 1 *psbA* sequences, suggesting that this gene was acquired only once by a subset of these myoviruses. Although the specific source is difficult to determine from phylogeny alone, the placement of the *Prochlorococcus* myovirus sequence clusters suggests that *psbA* was derived from either HL *Prochlorococcus* hosts or LL NATL2A-type hosts, while the *psbD* genes could have been acquired from any of the *Prochlorococcus* hosts other than MIT9313/9303. The placement of the *Prochlorococcus* podovirus (*psbA* only) and *Synechococcus* myovirus sequence clusters at the base of the

host and virus clades provides little further information about the source of these phage genes.

We found three exceptions to the above host-constrained evolutionary scenario—i.e., cases where phage *psbA* and *psbD* genes did not cluster with those of their hosts (Figure 1 and Figures S1 and S2) and did not have PsbA isoforms consistent with that of their hosts (unpublished data). These include two narrow host-range *Synechococcus* myoviruses (S-ShM1, S-SSM1), which encode *psbA* sequences most similar to *Prochlorococcus* myoviruses (Figure 1) even to the extent that they encode the PsbA.1 isoform, as well as a *Prochlorococcus* myovirus (P-SSM1) with a *psbA* sequence that is most similar to those from *Synechococcus* myoviruses (Figure 1) and encodes the PsbA.2 isoform as expected for a *Synechococcus* myovirus. Although the latter can cross-infect across *Prochlorococcus* ecotypes, it has not been shown to infect *Synechococcus* [5]. The P-SSM1 phage also encodes *psbD,* which, like its *psbA* gene, is more similar to *Synechococcus psbD* sequences than those of the *Prochlorococcus* host upon which it was isolated (Figure S2; note that this sequence does not appear in Figure 2 because it was a candidate for intragenic recombination; see Materials and Methods). It is likely that these exceptions to the rather consistent host-phage sequence clustering resulted from horizontal transfer events between a broad-host-range donor phage and a limited-host-range recipient phage during coinfection of a single host, i.e., swapping of genes within the phage gene pool [24]. Whole gene transfers within the phage gene pool are likely to be more common than this, but undetectable when occurring within phages that form a discrete phylogenetic cluster. These observations call for caution when using clustering patterns of *psbA* and *psbD* sequences from uncultured phage (obtained from environmental genome data) to identify potential hosts.

## Intragenic Recombination within Core Reaction Center Proteins

The lack of well-supported clade structure in phylogenetic reconstructions for *Synechococcus* host strains when using both *psbA* and *psbD* differs from those constructed using other genes [46–49], which led us to wonder about underlying mechanisms that could be responsible for such a blurred phylogenetic signal. In a recent study, Zeidner et al. [37] showed that *Synechococcus*-phage-like *psbA* sequences from the environment had a patchy %G+C distribution, which they suggest is due to intragenic recombination [37]. Their analyses demonstrated that such recombination had occurred within the inferred-phage clusters and within clusters spanning both phage and host *psbA* sequences. They could not discern, however, whether the signal was caused only by phage-to-phage exchanges, or included phage-to-host exchanges, because the majority of their sequences were of unknown origin (i.e., they were derived from environment clone libraries), and the test employed does not assess the directionality of intragenic recombination events. Our cultured hosts and phages provide an opportunity to assess recombination partners without ambiguity regarding the source of the genes. In addition, the known host ranges of these phages [5] (Table 1), together with the types of recombination tests we have used (see Materials and Methods), allow us to assess the directionality and the pathways through phages and hosts that these recombination events are likely to have taken.

As a first assessment for potential intragenic recombination, we analyzed the %G+C patterns in all of the *psbA* and *psbD* genes (Figures 3 and 4, respectively). *Prochlorococcus* phage genes had similar average %G+C contents to those from their *Prochlorococcus* hosts (39%–46%), whereas those of *Synechococcus* phages had %G+C contents that were lower than those from their *Synechococcus* hosts (46%–51% versus 56%–62%), but not as low as those from *Prochlorococcus* hosts and phages. This intermediate %G+C could be the result of intragenic recombination between variants of the two host lineages. Alternatively, it may reflect the current state of mutational amelioration of the acquired gene from a high %G+C source towards the low genome-wide %G+C of the virus (*Synechococcus* myoviruses S-PM2 and Syn9 both have low genome-wide %G+C; [28]; P. Weigele, W. Pope, G. Hatfull, R. Hendrix, personal communication). If the latter is the case, we might expect such amelioration to be constant across the gene, resulting in an even %G+C distribution pattern.

To help differentiate between these hypotheses, we mapped the %G+C variation across the *psbA* and *psbD* genes using the methodology developed by Zeidner et al. [37]. We detected patchiness of %G+C in *Synechococcus* myovirus *psbA* sequences dispersed along the length of the gene (Figure 3), confirming the findings reported by Zeidner et al. [37]. We also detected %G+C patchiness among *psbA* from *Prochlorococcus* podoviruses, but not from *Prochlorococcus* myoviruses, despite overall similarity of their %G+C content with their *Prochlorococcus* hosts. This suggests that intragenic recombination has occurred among the podoviruses. In addition, patterns of %G+C were not uniform and even markedly clumped across the *psbD* gene from *Synechococcus* myoviruses (Figure 4), with the first segment resembling *Synechococcus* hosts and the last segment resembling *Prochlorococcus* hosts and their phages. Thus, intragenic recombination is likely to be at least partly responsible for the intermediate %G+C content in *Synechococcus* myovirus *psbA* and *psbD* sequences.

Statistical methods for detecting intragenic recombination (see Materials and Methods) revealed strong evidence for its presence in both the *psbA* and *psbD* sequence sets (Tables S1 and S2), but the relative frequency of recombination events was not equal for different groups of hosts and phages. Recombination appears most common among the cyanophages, and more so for *Synechococcus* than *Prochlorococcus* phages. Exchanges were detected between phages that infect both *Synechococcus* and *Prochlorococcus* as well as within myoviruses that infect a single genus *(Synechococcus).* Note that exchanges within a single phylogenetic phage cluster, such as within the *Synechococcus* myoviruses, were undetectable by our previous phylogenetic analyses. Interestingly, our analyses also revealed exchanges between *Prochlorococcus*-specific podoviruses and broad-host-range *Synechococcus* myoviruses, with the *Prochlorococcus* podoviruses serving as the donors (Table S1). Marine cyanobacterial podoviruses contain integrase genes and are thought to have the ability to integrate into the genomes of their hosts as prophages [30] (P. Weigele, W. Pope, G. Hatfull, R. Hendrix, personal communication). If true, genetic exchange could occur between the *Prochlorococcus* prophage and a *Synechococcus* lytic phage—a scenario well accepted in other phage-host systems for genetic exchange [14,15].

Intragenic recombination involving host genes appears less common than phage-to-phage recombination events (Tables S1 and S2). Exchanges between *Synechococcus* and their viruses

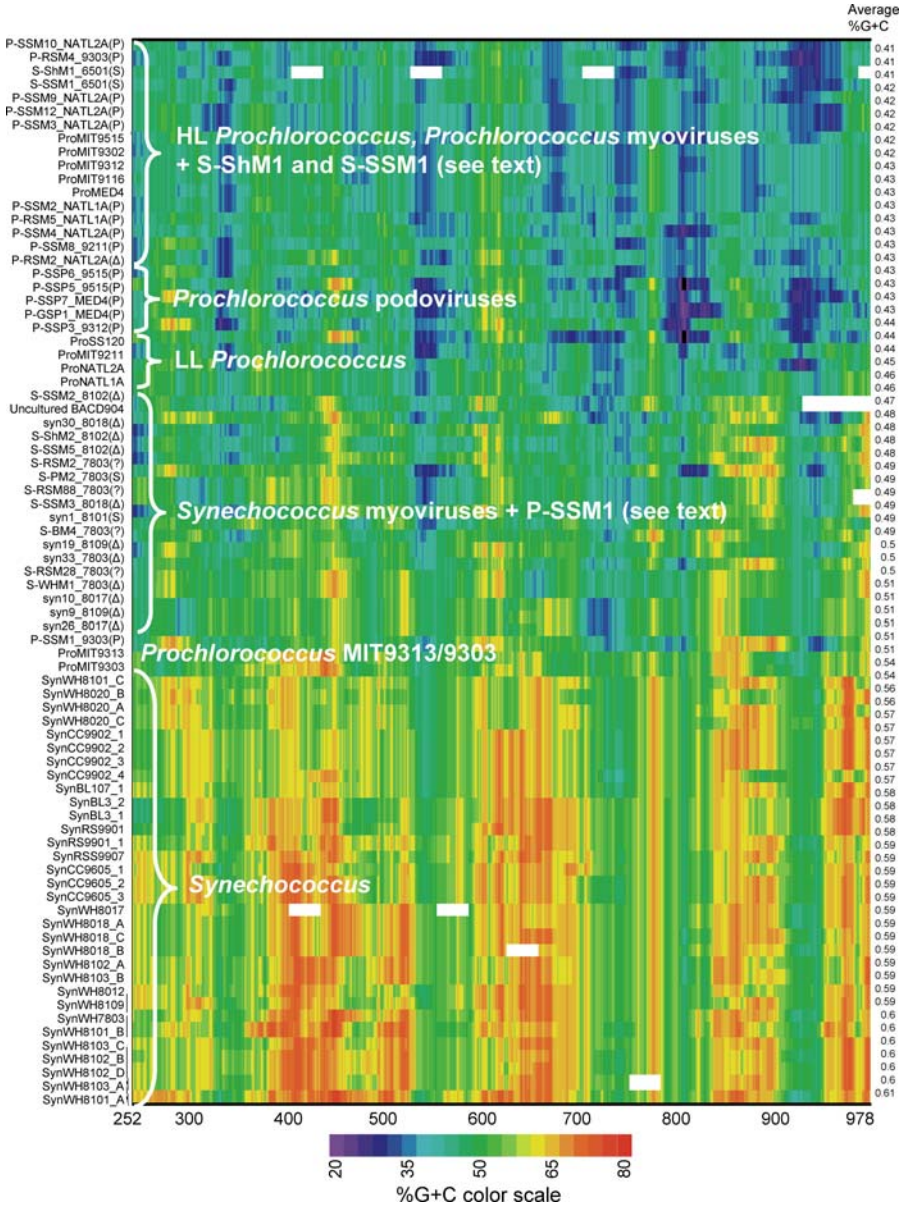


**Figure 3.** Visualization of %G+C Content across the *psbA* Gene

Colors represent the averaged %G+C in sliding windows along the length of the gene (20%–80%); white regions represent windows that included ambiguous bases in which %G+C could not be calculated for that region. The average %G+C content of the amplified sequence is tabulated on the right side of the figure. Phages are listed by phage name followed by their original host. Phages that are known to infect both *Prochlorococcus* and *Synechococcus* hosts are indicated with a "Δ"; those that infect only one genus or the other have no marker, while those that are unknown are designated with a "?". Host names are prefaced with Syn or Pro for *Synechococcus* and *Prochlorococcus* hosts, respectively. Scale indicates nucleotide positions relative to the *psbA* gene sequence in *Thermosynechococcus*.

DOI: 10.1371/journal.pbio.0040234.g003

are evident, however, and appear to have occurred both from host to phage and phage to host for both *psbA* and *psbD*. Although such events were not detected between *Prochlorococcus* and their phages, there were cases where *Prochlorococcus* myoviruses were the recipients of external DNA from an unknown source (i.e., recombination events possibly involving donors outside of our dataset). Thus, phages may be contributing to the intragenic recombination of portions of these genes in *Synechococcus,* perhaps explaining the lack of phylogenetic structure observed in *psbA* and *psbD* trees for *Synechococcus* clusters (but not for *Prochlorococcus* clusters) relative to those obtained when using other phylogenetic markers [46–49]. Presumably, phage-host intragenic exchanges occur via homologous recombination during infection. Clearly, the transfer of DNA will be retained in host lineages only if infection fails to lyse the host (e.g., abortive infection [54]).

Finally, intragenic exchanges among hosts were also occasionally detected, particularly among *Synechococcus* (Tables S1 and S2). This may also play a role in the lack of clade structure among *Synechococcus* strains in the *psbA* and *psbD* trees. Although two possible intragenic recombination events between *Synechococcus* and *Prochlorococcus* were identified, they were resolved as small regions (15–16 bases) and may be false

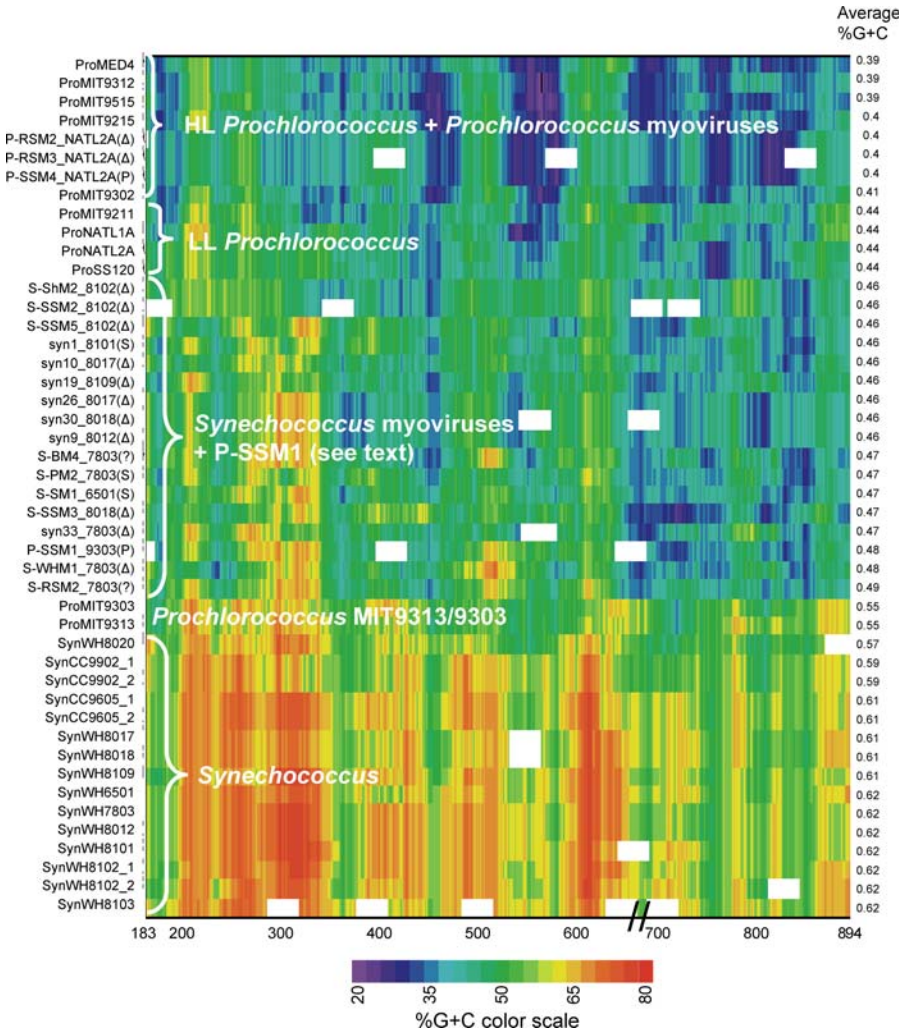


**Figure 4.** Visualization of %G+C Content across the *psbD* Gene

Details as in Figure 3. Note that the 21-nucleotide indel in *Prochlorococcus* hosts and their phages [13] (unpublished data) was excluded from the analysis at the position indicated by the "//" symbol to maximize the data that could be displayed using the sliding window approach.

DOI: 10.1371/journal.pbio.0040234.g004

positives. Host-to-host transfers may have occurred through the uptake of DNA directly from the environment (e.g., via transformation) or through viral intermediates [37]. Such host-to-host intragenic exchanges via viral intermediates presumably occur through generalized transduction [55].

In summary, our findings suggest that the shuffling of segments of *psbA* and *psbD* within the cyanophage gene pool has generated significant photosynthesis gene diversity and serves as an extended reservoir of genetic diversity for their hosts, influencing photosystem evolution.

## *psbA* and *psbD* Gene Diversity in Cultured Isolates Captures Most of the Field Diversity

We next sought to determine how well *psbA* and *psbD* sequence diversity observed in culture collections represents that observed in wild phage populations, and whether additional whole-gene host-to-phage transfer events could be identified from these wild sequences from the phage gene pool. Zeidner et al. [37] had previously examined field diversity of the *psbA* gene sequence from environmental samples where *Synechococcus* strains were the dominant phototroph [37]. Thus, we sought to examine genetic diversity of this gene, as well as that of *psbD,* from an environment where *Prochlorococcus* cells commonly outnumber *Synechococcus* cells by orders of magnitude [56]. To this end, we amplified, cloned, and sequenced *psbA* and *psbD* gene sequences obtained from the viral-sized fraction (0.02–0.2 μm) of two seawater samples within (25 m) and below (75 m) the mixed layer in the Pacific Ocean off the coast of Hawaii (Figures 5 and 6, respectively). The *psbA* and *psbD* sequences from these viral-fraction samples clustered with cultured *Prochlorococcus* cyanophage isolates (with varying levels of support; Figures 5 and 6), but not with *Synechococcus* cyanophages. There was not a notable difference in the phylogenetic placement of the *psbA* or *psbD* clones obtained from within or below the mixed layer. Although this suggests a lack of vertical structure in diversity among the sequence types, we did not sequence these samples to saturation; thus, such conclusions are preliminary.

More than half of the wild *psbA* sequences (42 of 81) form a large cluster with cultured *Prochlorococcus* podoviruses (Figure 5). Within this group, all but one cluster of wild sequences

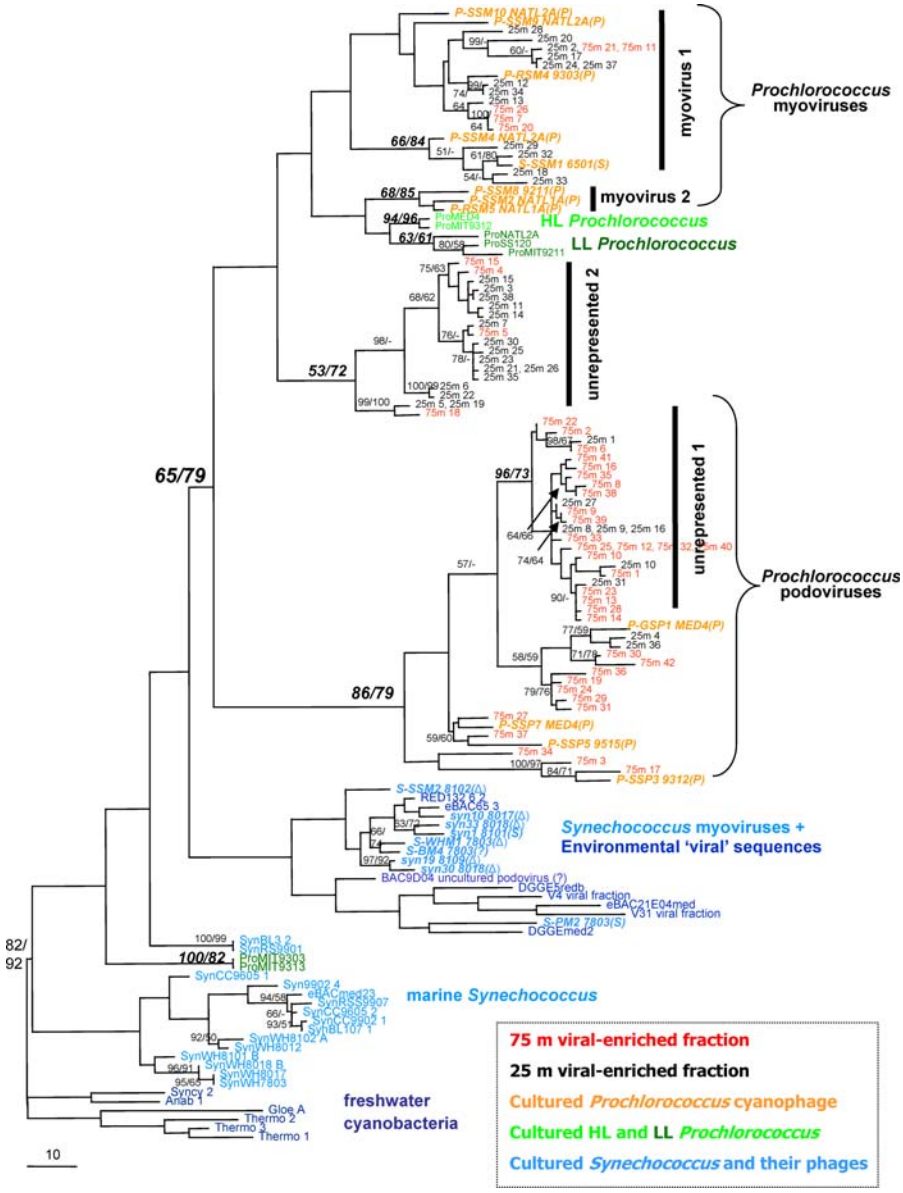


**Figure 5.** Phylogenetic Tree of *psbA* Gene Sequences from Representative Cultured Cyanobacterial and Cyanophage Isolates and Cloned Environmental Sequences from the Hawaii Ocean Time Series Site in the Pacific Ocean

Phylogenetic tree of *psbA* gene sequences and cloned environmental sequences were collected from above (25 m, black) and below (75 m, red) the surface mixed layer at the Hawaii Ocean Time Series site in the Pacific Ocean, a region where *Prochlorococcus* are the dominant phototrophs. Details for naming conventions are as in Figure 1. *Synechococcus* environmental "viral" sequences from [37]. The tree topology was estimated by LogDet analysis of 1st and 2nd codon positions, with branch lengths estimated using stationary nucleotide frequencies.

DOI: 10.1371/journal.pbio.0040234.g005

contain cultured podovirus sequences (Figure 5). The extensive microdiversity in this cluster (labeled "unrepresented 1") was probably derived from within the podovirus gene pool, as evidenced by the presence of podovirus phage isolates in the more basal branches of the cluster. Other *psbA* sequences from the field samples form subclusters that contain cultured *Prochlorococcus* myoviruses and form a large group that also contains *Prochlorococcus* hosts (Figure 5). One cluster ("unrepresented 2" in Figure 5) within this group also lacks sequences from cultured hosts or phages. The basal position of this cluster suggests that these sequences may belong to phages that infect as-yet uncultured *Prochlorococcus* hosts [57] and may represent an additional host-to-phage transfer event. Thus, our work here, together with that of Zeidner et al. [37], suggests that cyanophage culture collections represent much of the naturally occurring *Prochlorococcus* and *Synechococcus* cyanophage *psbA* gene sequence diversity [37].

All *psbD* sequences from wild phages fall into a single well-supported cluster that includes a representative cultured *Prochlorococcus* cyanophage P-SSM4 (Figure 6). This cluster reveals significant microdiversity within the *psbD Prochlorococcus* phage gene pool in the viral-fraction from this Pacific Ocean site and suggests that phages that encode *Prochlorococcus*-phage-like *psbD* genes are perhaps not rare in this environment. The four *Prochlorococcus* cyanophages that

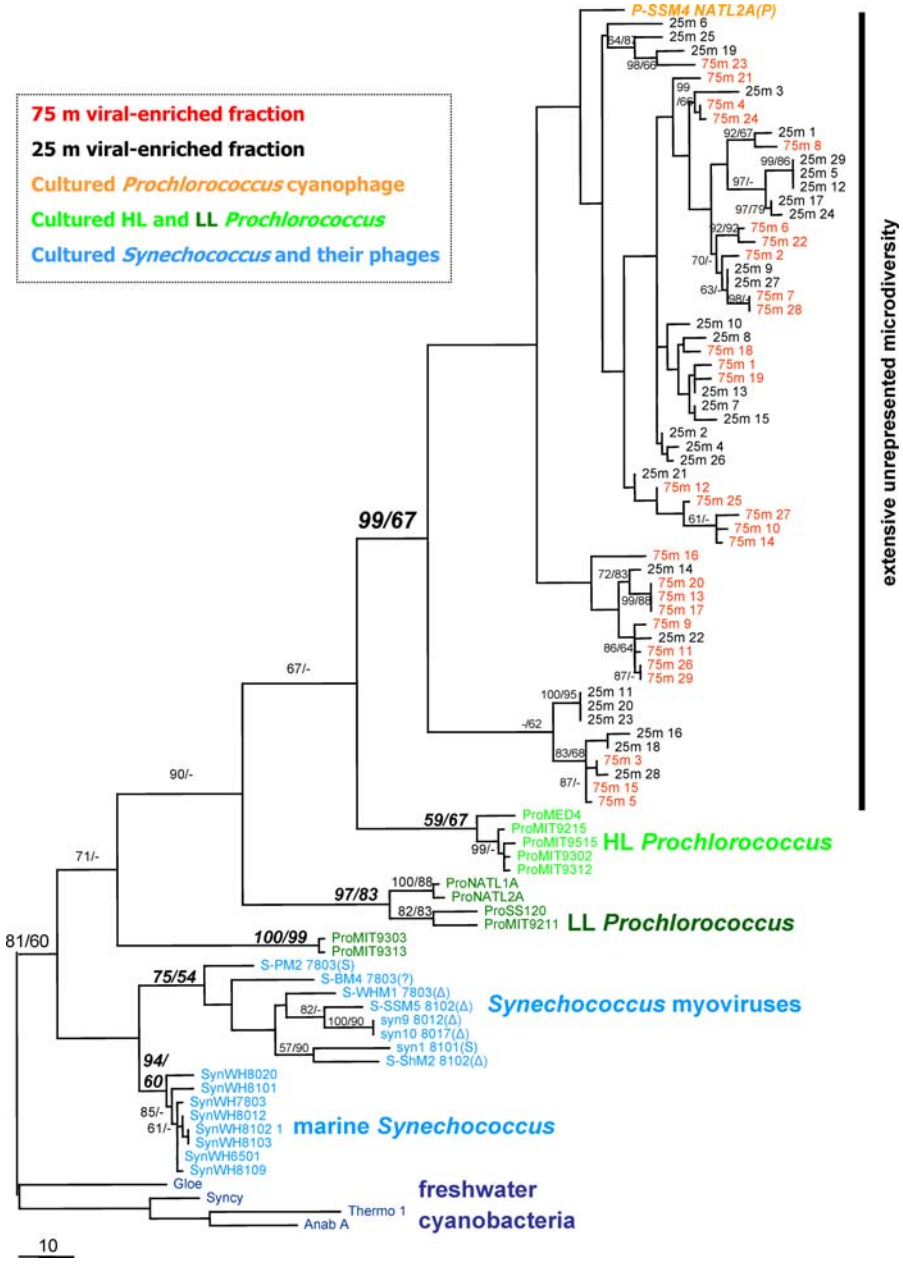


**Figure 6.** Phylogenetic Tree of *psbD* Gene Sequences from Cultured Cyanobacterial and Cyanophage Representatives and Cloned Environmental Sequences from the Pacific Ocean

Details for naming conventions are as in Figure 1, and phylogenetic analyses are as in Figure 5.

DOI: 10.1371/journal.pbio.0040234.g006

contain the *psbD* gene in our culture collection originated from either the Sargasso Sea or the Red Sea; thus, it is perhaps not surprising that the viral-fraction microdiversity from the Pacific Ocean is largely unrepresented in this collection.

## Conclusions

The phage genomic repertoire evolves through the exchange of genetic material from other phages [24] and by co-opting metabolic genes from their hosts [13,20,22]. The prevalence of photosynthesis genes in cyanophages strongly suggests that the capture of these genes provides a significant fitness advantage among certain cyanophage types. Previously, we have shown that the horizontal transfer of *hli* genes from cyanophages to their hosts has likely played a role in driving host niche differentiation [13]. More recently, cyanophages were hypothesized to be involved in partial gene exchanges even for the core photosystem gene *psbA* of their hosts [37]. Here, we show that genetic exchanges involving cyanophages may have influenced the make-up of both of the core photosystem II genes *(psbA* and *psbD)* in *Synechococcus,* whereas this was less apparent for *Prochlorococcus*. Therefore, mounting evidence indicates that host-like genes acquired by phages undergo a period of diversification in

phage genomes and serve as a genetic reservoir for their hosts. Thus, a complex picture of overlapping phage and host gene pools emerges, where genetic exchange across these pools leads to evolutionary change for host and phage. Fully understanding the mechanisms of microbial and phage coevolution clearly requires an improvement in our ability to quantify horizontal gene transfer at the whole and partial gene level and in our ability to accurately estimate the relative fluxes into and out of these pools.

## Materials and Methods

**DNA isolation from cultured hosts and phages and environmental samples.** Eleven strains of *Prochlorococcus,* ten strains of *Synechococcus,* and 38 phages of *Prochlorococcus* and *Synechococcus* (seven podoviruses, 29 myoviruses, and two siphoviruses) were screened for *psbA* and *psbD* sequences for this study. We report here on new *psbA* sequences from nine *Synechococcus* hosts and new *psbD* sequences from 19 *Prochlorococcus* and *Synechococcus* hosts (including two from unpublished *Synechococcus* genomes for strains CC9605 and CC9902; available from http://genome.jgi-psf.org/mic_home.html. The 38 phages screened included seven phage templates for which genome sequences are now available (P-SSM2, P-SSM4, P-SSP7, S-PM2, S-WHM1, Syn5, Syn9), enabling us to validate our PCR amplification findings. Host genomic DNA was extracted using a DNeasy Tissue Kit (Qiagen, Valencia, California, United States). Filtered (0.2 μm, Acrodisc supor membrane syringe filter) phage lysates in Pro99 medium were used as DNA templates for subsequent PCR amplification experiments.

Environmental samples were collected from the Hawaii Ocean Time Series (HOT) on 15 October 2003 at 45°N 158°W from depths of 25 m and 75 m. These samples were filtered through a 0.2-μm filter (Osmonics, Minnetonka, Minnesota, United States, Poretics polycarbonate 25-mm filter) to remove cellular material and substantially enrich for environmental phages. A 100-ml volume of 0.2-μm filtrate was then filtered onto a 0.02-μm filter (Whatman Anotop 25) to collect phage particles and resuspended in 7 ml of a modified SM storage buffer (600 mM NaCl, 8 mM $MgSO_4$-$7H_2O$, 50mM Tris [pH 7.5], 0.04% gelatin).

**Overview of *psbA* and *psbD* screening strategy.** PCR screening for *psbA* and *psbD* across a diverse set of samples presented several challenges. These included variable amplification efficiencies, uncertainty about whether amplicons derived from phage or host, and multiple gene copies in hosts. The amplification strategy was as follows: for each virus and host strain, four PCR reactions were carried out, pooled, and analyzed by gel electrophoresis; if the amplification product was not visible, it was diluted 10-fold and used as template for nested or semi-nested PCR and the resulting products analyzed; if still no product was visible, multiple phage stocks were rescreened. Multiple copies of *psbA* in *Synechococcus* strains were identified by sequencing many clones and were distinguished from sequencing errors as described below. We did not screen for multiple copies of *psbA* from *Prochlorococcus* or multiple copies of *psbD* from either *Synechococcus* or *Prochlorococcus,* as when present, they are generally indistinguishable from each other [58–60].

**Amplification of *psbA* and *psbD.*** PCR reactions were performed with *Taq* DNA polymerase and deoxyribonucleotide triphosphates from New England Biolabs (Beverly, Massachusetts, United States) or Invitrogen (Carlsbad, California, United States) and carried out with a PTC-100 or PTC-200 DNA Engine (MJ Research, Waltham, Massachusetts, United States) or a Robocycler Gradient 96 (Stratagene, La Jolla, California, United States). Template amounts were 10 ng of genomic DNA for *Prochlorococcus* and *Synechococcus,* 1 μl of lysate for cyanophages, and 2 μl of filtrate for environmental samples. PCR primers and amplification reaction conditions are shown in Tables S3 and S4.

The *psbA* gene from all sources was amplified using primer pair *psbA*-F/R [61] and PCR protocol A (Tables S3 and S4). Four reactions were conducted with each template, and the products were pooled and analyzed by agarose gel electrophoresis. Primer *psbA*-R falls on the intron region in S-PM2 [29]. Therefore, for efficient amplification of phage *psbA* genes that may contain introns, and for increased sensitivity, we used the Pro-*psbA*-F/R primer set and protocol B in nested PCR reactions when no PCR product was visible from cyanophage lysates and environmental filtrates. To reduce the incidence of heteroduplex formation, amplification products from environmental samples were subjected to reconditioning PCR [62]: initial PCR products were diluted 1:10, then amplified using protocol A but for only three cycles.

The *psbD* gene from *Prochlorococcus, Synechococcus,* and cyanophages was amplified using primer pair *psbD*-54F/*psbD*-308R and protocol D. However, when product yield was low or absent, semi-nested PCR was carried out as follows. Amplification was first conducted using primer pair *psbD*-26F/*psbD*-308R and protocol C. Four reactions were conducted with each template, the products were pooled, diluted 1:10, and used as templates for a second round of amplification using primer pair *psbD*-54F/*psbD*-308R and protocol D. *psbD* from environmental samples was amplified using primer pair *psbD*-26F/*psbD*-308R and protocol C and subjected to reconditioning PCR as for *psbA* (see above).

In preparation for sequencing, PCR products were either purified directly using the QIAquick PCR Purification Kit (Qiagen) or separated on an agarose gel and then purified using the QIAquick Gel Extraction Kit (Qiagen).

To confirm that the absence of *psbA* or *psbD* PCR products from phage was not simply due to a lack of amplifiable phage DNA, we screened phage lysates for known phage genes: *g20* (for myoviruses) and *DNApol* (for podoviruses). *g20* was amplified using primer pair g20-F/R and protocol E, and *DNApol* using primer pair DNApol-F/R and protocol F, both with 1 μl of lysate. In all cases, a product was obtained, suggesting the phage template DNA was present and amplifiable by PCR (unpublished data).

Six phage lysates yielded PCR products with sequences identical to those of a known host. These six phage lysates include five cyanophages previously described (P-RSP1, P-SSP1, P-SSP2, P-ShM1, P-ShM2; [5]), as well as one cyanophage not previously reported in the literature (P-SSP9; M.B.S. and S.W.C., unpublished data). In these cases we could not eliminate the possibility that the amplicon resulted from host DNA, the amplification of which may be more likely to occur when there is no phage template for this gene. Thus, we excluded these phages from further analyses. In contrast, phages with amplicon sequences identical to those of other phages (indicated as "ID to X" in Table 1) were passed through multiple lysates, and a "fingerprint" phage gene *(g20)* was used to confirm that there was a single phage in the lysate. The *psbA* sequence was then reassayed, increasing our confidence in these results. Even with this precaution, we cannot rule out the possibility of PCR contamination for those few cases where identical sequences were amplified from different phage lysates.

**Cloning and sequencing of PCR products.** The *psbA* gene is often found in multiple distinct copies in marine *Synechococcus* [59], whereas in *Prochlorococcus* the *psbA* gene is either single copy per genome or encodes multiple copies that are nearly identical to each other [60,63,64]. Among cyanophages, the *psbA* gene has only been found in a single copy per genome [28,30]. To allow for the identification of multiple *psbA* gene copies in *Synechococcus* strains, PCR products from *Synechococcus* templates were cloned prior to sequencing. Cloning was performed using the TOPO TA Cloning Kit for Sequencing (Invitrogen) with the pCR4-TOPO vector. Ligation products were transformed into TOP10 competent cells. Plasmid purification and sequencing were conducted by Genaissance Pharmaceuticals (New Haven, Connecticut, United States). Inserts were sequenced from both forward and reverse directions, using the M13F and M13R primer binding sites in the pCR4-TOPO vector.

Approximately ten *psbA* clones were sequenced for each *Synechococcus* strain. The published genome of *Synechococcus* WH8102 provides an example of natural *psbA* diversity in a given strain, as it contains four copies of *psbA:* two copies that are 99.8% identical and a third and fourth copy that are 99.4% and 88% identical, respectively, to the above two *psbA* copies [59]. Considering a *Taq* polymerase error rate of $3 \times 10^{-5}$ per nucleotide per duplication [65], at most one error could be expected in each *psbA* gene sequenced. Thus, sequences were considered identical, and removed from the analysis pool, if they were more than 99.8% identical, to avoid data issues stemming from possible PCR error (sequencing error should be nonexistent because consensus sequences were obtained from forward and reverse sequencing of the clones). Sequence identity levels for nonidentical clones from the remaining dataset ranged from about 60% to 99.0%.

PCR products from genes presumed not to have multiple distinct copies per genome (*psbA* from *Prochlorococcus* and cyanophage; *psbD* from all organisms) were generally sequenced directly (Harvard Medical School Biopolymers Facility [Boston, Massachusetts, United States], Davis Sequencing [Davis, California, United States], or Genaissance Pharmaceuticals). The absence of multiple significant-height peaks at single nucleotide positions in chromatograms from this direct sequencing (unpublished data) confirmed that single products were amplified during PCR. Each strain was sequenced in

both forward and reverse directions, using the same primers used for PCR amplification.

**Sequence analyses.** Previous analyses have raised important concerns about using *psbA* gene sequence datasets that may suffer from large %G+C variability and conflicting phylogenetic signals in phylogenetic reconstructions [37]. To minimize such errors, we followed these steps.

We first performed phylogenetic analyses using sequences from all taxa (80 for *psbA* and 50 for *psbD*) and all codon positions (Figures S1 and S2). Phylogenetic trees were constructed by using distance and maximum likelihood. Neighbor-joining [66] was used to reconstruct a distance tree under the HKY85 model [67]. Maximum likelihood analysis was performed under HKY85 combined with a gamma model for among sites rate variation, assuming eight rate categories with model parameters estimated from the data [68]. Maximum likelihood trees were obtained by quartet puzzling, as implemented in the program TREE-PUZZLE 5.0 [69]. Bootstrap resampling (1,000 pseudoreplicates) was used to measure the relative support for internal branches of the neighbor-joining trees. For quartet puzzling, support was estimated from 25,000 (*psbD* trees) or 50,000 (*psbA* trees) pseudoreplicates.

These analyses resulted in trees with high bootstrap support at many critical nodes (Figures S1 and S2). However, fitting a single tree to large datasets containing conflicting phylogenetic signals can lead to reconstruction artifacts (i.e., systematic errors) that result in high bootstrap support [70,71]. We found, using neighbor-nets [72] constructed by using the SplitsTree2 program [73], within-gene conflicting phylogenetic signals in both the *psbA* and *psbD* datasets as indicated by the box-like structures in neighbor-nets graphs (Figures S3 and S4). Specifically, networks for both genes revealed substantial conflict involving splits between *Synechococcus* strains, their myoviruses, and a complex of sequences comprised of *Prochlorococcus* and their viruses.

We further investigated whether these large datasets could suffer from systematic errors related to: (i) substitution rate variation among lineages [74], (ii) heterogeneous compositional bias among lineages (e.g., %G+C; [75]), and (iii) within-gene heterogeneity in phylogenetic signals [76]. We found significant substitution rate variation among lineages (Table S5) using likelihood ratio tests. In addition, nucleotide frequencies were nonstationary across these data, with significant differences in equilibrium frequencies for clades defined according to organism types (Table S6; [77]). Not surprisingly, the largest divergence in %G+C across taxa was at the 3rd codon positions of both *psbA* and *psbD*.

Zeidner et al. [37] hypothesized intragenic recombination in *psbA* [37]. We attempted to identify this qualitatively through graphical analysis of %G+C and quantitatively using four different tests for intragenic recombination. The %G+C distribution was examined within overlapping sequence windows (a sliding window of 30 nucleotides with a five-nucleotide step) using the GCViz script [37] (available upon request from Dr. Shmoish of Technion–IIT; mshmoish@cs.technion.ac.il) written in the R-language (http://www.r-project.org). Three of the four different tests for within-gene recombination are based on the distribution of substitutions (GeneConv: [78]; MaxChi: [79]; Chimaera: [80]), while the fourth used a phylogenetic approach ("RDP," as implemented in [81]). We considered only those recombination events that satisfied all of the following criteria: (i) results were significant after application of Bonferroni correction for multiple tests, (ii) regions were detected by two or more different methods, and (iii) consensus breakpoints could be estimated for a given region identified using different methods. Once a putative recombination event was detected, we inferred the best candidate donor sequence (that most similar to the recombinant segment) using RDP [81].

In summary, to minimize systematic errors in the ultimate phylogenetic analyses, we first processed the dataset as follows: (i) excluded those sequences having a strong signal for intragenic recombination, (ii) excluded 3rd codon positions, which display the largest differences in %G+C and substitution rates among lineages, and (iii) employed LogDet distances [75] to accommodate compositional heterogeneity (variable %G+C) in the remaining data. These measures proved to be important. The uncorrected dataset grouped lineages according to evolutionary rates and %G+C bias (Figures S1 and S2), whereas the ultimate analysis did not (see Figures 1 and 2). Statistical analysis of the processed dataset under nonhomogenous evolutionary models [77] revealed that the ultimate phylogenetic hypotheses (see Figures 1 and 2) provided a significantly better fit to the data (Table S7). Prior to processing the data, the alternative phylogenies were indistinguishable (Table S7).

## Supporting Information

**Figure S1.** Phylogenetic Analyses Including All *psbA* Gene Sequences from Cultured Cyanobacteria and Cyanophages

Phages are listed by phage name, followed by their original host. Host range information is designated in parentheses. Phages known to infect both *Prochlorococcus* and *Synechococcus* hosts are indicated with a "Δ"; phages that infect only *Prochlorococcus* or *Synechcococcus* are designated by a P or S, respectively; and those host ranges that are unknown have a "?". Phages shown in italics and bracketed with "**" were isolated on hosts that do not belong to the same cluster and are thus exceptions to the general clustering pattern (see text). Taxa are color coded according to the following biological groupings: myoviruses (red), podoviruses (black), marine *Synechococcus* hosts (light blue), marine *Prochlorococcus* hosts (dark green, HL; light green, LL), freshwater cyanobacteria (dark blue). Neighbor-joining tree was inferred under HKY85 mode and using sequences from all taxa and all codon positions. Nucleotide frequencies were assumed to be homogenous across lineages. Numbers at the nodes represent neighbor-joining bootstrapping and maximum likelihood puzzling support. Anab, *Anabaena;* Gloe, Gleobacter; HL, high-light adapted; LL, low-light adapted; Syncy, *Synechocystis;* Thermo, *Thermosynechococcus*.

Found at DOI: 10.1371/journal.pbio.0040234.sg001 (79 KB PPT).

**Figure S2.** Phylogenetic Analyses Including All *psbD* Gene Sequences from Cultured Cyanobacteria and Cyanophages

Details are as in Figure S1.

Found at DOI: 10.1371/journal.pbio.0040234.sg002 (59 KB PPT).

**Figure S3.** Neighbor-Nets Analysis of 80 *psbA* Gene Sequences (including All Cyanophage and Marine Cyanobacterial Sequences Available)

The analysis was conducted under the HKY85 model of substitution using all codon positions. Taxa color coding and abbreviations are as in Figure S1. The box-like appearance in the basal branches of this phylogeny suggests regions of conflicting phylogenetic signals (see Materials and Methods).

Found at DOI: 10.1371/journal.pbio.0040234.sg003 (272 KB PDF).

**Figure S4.** Neighbor-Nets Analysis of 50 *psbD* Gene Sequences (including All Cyanophage and Marine Cyanobacterial Sequences Available)

Taxa color coding and abbreviations are as in Figure S1. Details of the analysis are as in Figure S3.

Found at DOI: 10.1371/journal.pbio.0040234.sg004 (249 KB PDF).

**Table S1.** Consensus Results from Four Tests for Intragenic Recombination within Gene Sequences in Our *psbA* Dataset

The four tests included (1) RDP, (2) GeneConv, (3) MaxChi, and (4) Chimaera (as described in Materials and Methods), and recombination was considered "detected" only when the following criteria were satisfied: (i) similar regions were detected by two or more methods, (ii) all such regions were significant at $p < 0.05$ after a Bonferroni correction for multiple tests, and (iii) consensus breakpoints could be inferred from the results. Thus, "No recombination detected" does not preclude that intragenic recombination could be occurring within the sequence, but rather indicates that our stringent criteria have not identified such an event. While we define phages as either *Prochlorococcus* or *Synechococcus* phages depending on the original host of isolation, we note that many of the myoviruses cross-infect both genera (represented with a "Δ" where known, a "?" where unknown, and no symbol for isolates that do not cross-infect across genera). Consensus breakpoints are relative to nucleotide positions in *Thermosynechococcus psbA*.

Found at DOI: 10.1371/journal.pbio.0040234.st001 (29 KB XLS).

**Table S2.** Consensus Results from Four Tests for Intragenic Recombination within Gene Sequences in Our *psbD* Dataset

Details are as in Table S1.

Found at DOI: 10.1371/journal.pbio.0040234.st002 (28 KB XLS).

**Table S3.** PCR Conditions

Found at DOI: 10.1371/journal.pbio.0040234.st003 (38 KB DOC).

**Table S4.** PCR Primers

Found at DOI: 10.1371/journal.pbio.0040234.st004 (39 KB DOC).

**Table S5.** Likelihood Ratio Tests for Variable Evolutionary Rates among Branches

For both *psbA* and *psbD,* individual sequences exhibiting a signature for intragenic recombination (Tables S1 and S2) were excluded from analysis. Likelihood scores were obtained under a stationary HKY85 model combined with a gamma correction for among-sites rate variation. All model parameters, including nucleotide frequencies, were estimated by using maximum likelihood. Data analysis included all three codon positions. Models were employed as implemented in the baseml program of the PAML package [82]. Tree 1 was obtained by neighbor-joining analysis of LogDet distances estimated from all three codon positions. Tree 2 was obtained by neighbor-joining analysis of LogDet distances estimated from 1st and 2nd codon positions. For both genes, Tree 1 grouped lineages along lines of similarity in evolutionary rates and compositional biases, and Tree 2 did not.

Found at DOI: 10.1371/journal.pbio.0040234.st005 (36 KB DOC).

**Table S6.** Likelihood Ratio Tests for Nonstationary Frequencies among Lineages

$H_0$ denotes the null hypothesis of stationary nucleotide frequencies; this was modeled by specifying one set of nucleotide frequencies for all branches of the tree. $H_1$ denotes the alternative hypothesis of nonstationary nucleotide frequencies; this was modeled by assigning all branches of the tree topology to one of several independent sets of frequency parameters (six sets for *psbA* and five sets for *psbD*). Apart from nucleotide frequencies, $H_0$ and $H_1$ assumed a substitution process equivalent to an HKY85 model combined with a gamma model for among-sites rate variation. The transition/transversion ratio was assumed to be homogenous among branches. $H_1$ represents a user-defined version of the nonhomogenous models of Yang and Roberts [77]. All model parameters, including nucleotide frequencies, were estimated by using maximum likelihood. Data analysis included all three codon positions. Models were employed as implemented in the baseml program of the PAML package [82].
Tree 1 was obtained by neighbor-joining analysis of LogDet distances estimated from all three codon positions. Tree 2 was obtained by neighbor-joining analysis of LogDet distances estimated from 1st and 2nd codon positions. For both genes, Tree 1 grouped lineages along lines of similarity in evolutionary rates and compositional biases, and Tree 2 did not. User-defined sets of frequency parameters for $H_1$ were specified in the tree file (shown below) by using the "branch label" format described in the PAML manual. For both *psbA* and *psbD,* individual sequences exhibiting a signature for intragenic recombination (Tables S1 and S2) were excluded from analysis.

Found at DOI: 10.1371/journal.pbio.0040234.st006 (44 KB DOC).

**Table S7.** Likelihood-Based Statistical Comparison of Competing Evolutionary Hypotheses under a Model of Nonstationary Nucleotide Frequencies

$P_{KH}$ denotes the *p*-value for the KH normal test of [83]. $P_{SH}$ denotes the *p*-value for the SH test [84]. $P_{RELL}$ denotes the RELL bootstrap proportion [83]. Note that although Tree 1 and Tree 2 were not selected independently of the data, neither was selected according to its likelihood score. For both genes, Tree 1 grouped lineages along lines of similarity in evolutionary rates and compositional biases, and Tree 2 did not. For both *psbA* and *psbD,* individual sequences exhibiting a signature for intragenic recombination (Tables S1 and S2) were excluded from analysis. Tree 1 was estimated by a neighbor-joining analysis of LogDet distances from all sites, and Tree 2 was estimated by a neighbor-joining analysis of LogDet distances based on only 1st and 2nd codon positions. Likelihood scores were obtained under nonstationary models of nucleotide frequencies (see Table S5 for additional model details).

Found at DOI: 10.1371/journal.pbio.0040234.st007 (46 KB DOC).

### Accession Numbers

New sequences from cultured cyanobacteria and cyanophages are deposited in GenBank (http://www.ncbi.nlm.nih.gov/Genbank) under accession numbers DQ473647–DQ473719, whereas new environmental sequences are deposited under accession numbers DQ473720–DQ473847.

## Acknowledgments

We thank M. Shmoish, R. Fu, and V. Quinlivan for technical assistance; M. Coleman, M. Osburne, J. Waldbauer, and V. Rich for valuable comments on the manuscript; A. Thompson for collecting field samples; and Z. Johnson, K. Armstrong, and B. Tidor for analysis and discussion of possible PsbA/PsbD interactions. We thank P. Weigele, W. Pope, G. Hatfull, and R. Hendrix for providing unpublished genome sequences (Syn5 and Syn9); and F. Chen for sharing his unpublished phage lytic cycle information (P60) with us.

**Author contributions.** MBS, DL, and SWC conceived and designed the experiments. MBS, DL, JAL, and LRT performed the experiments. MBS, DL, JAL, LRT, and JPB analyzed the data. MBS and DL wrote the paper, with significant contributions from all authors.

**Funding.** This research was supported by grants from the United States Department of Energy (DE-FG02-99ER62814 and DE-FG02-02ER63445), the National Science Foundation and the Gordon and Betty Moore Foundation to SWC, Massachusetts Institute of Technology's Undergraduate Research Opportunities Program funding to JAL, a National Institutes of Health predoctoral training grant in the biological sciences (GM07287–31) to LRT, and a National Sciences and Engineering Research Council (Canada) Discovery Grant (DG 298394) to JPB.

**Competing interests.** The authors have declared that no competing interests exist.

### References

1. Waterbury JB, Watson SW, Valois FW, Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium *Synechococcus*. Can Bull Fish Aquat Sci 214: 71–120.
2. Partensky F, Hess WR, Vaulot D (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. Microbiol Mol Biol Rev 63: 106–127.
3. Suttle CA, Chan AM (1994) Dynamics and distribution of cyanophages and their effects on marine *Synechococcus* spp. Appl Environ Microbiol 60: 3167–3174.
4. Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophage abundant in seawater. Appl Environ Microbiol 59: 3393–3399.
5. Sullivan MB, Waterbury JB, Chisholm SW (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. Nature 424: 1047–1051.
6. Lu J, Chen F, Hodson RE (2001) Distribution, isolation, host specificity, and diversity of cyanophages infecting marine *Synechococcus* spp. in river estuaries. Appl Environ Microbiol 67: 3285–3290.
7. Marston MF, Sallee JL (2003) Genetic diversity and temporal variation in the cyanophage community infecting marine *Synechococcus* species in Rhode Island's coastal waters. Appl Environ Microbiol 69: 4639–4647.
8. Muhling M, Fuller NJ, Millard A, Somerfield PJ, Marie D, et al. (2005) Genetic diversity of marine *Synechococcus* and co-occurring cyanophage communities: Evidence for viral control of phytoplankton. Environ Microbiol 7: 499–508.
9. Wilson WH, Joint IR, Carr NG, Mann NH (1993) Isolation and molecular characterization of five marine cyanophages propogated on *Synechococcus* sp. strain WH 7803. Appl Environ Microbiol 59: 3736–3743.
10. Suttle CA, Chan AM (1993) Marine cyanophages infecting oceanic and coastal strains of *Synechococcus*: Abundance, morphology, cross-infectivity and growth characteristics. Mar Ecol Prog Ser 92: 99–109.
11. Brussow H, Canchaya C, Hardt WD (2004) Phages and the evolution of bacterial pathogens: From genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev 68: 560–602.
12. Faruque SM, Mekalanos JJ (2003) Pathogenicity islands and phages in *Vibrio cholerae* evolution. Trends Microbiol 11: 505–510.
13. Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, et al. (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. Proc Natl Acad Sci U S A 101: 11013–11018.
14. Canchaya C, Proux C, Fournous G, Bruttin A, Brussow H (2003) Prophage genomics. Microbiol Mol Biol Rev 67: 238–276.
15. Casjens S (2003) Prophages and bacterial genomics: What have we learned so far? Mol Microbiol 49: 277–300.
16. Forterre P (1999) Displacement of cellular proteins by functional analogues from plasmids or viruses could explain puzzling phylogenies of many DNA informational proteins. Mol Microbiol 33: 457–465.
17. Filee J, Forterre P, Laurent J (2003) The role played by viruses in the evolution of their hosts: A view based on informational protein phylogenies. Res Microbiol 154: 237–243.
18. Filee J, Forterre P, Sen-Lin T, Laurent J (2002) Evolution of DNA polymerase families: Evidences for multiple gene exchange between cellular and viral proteins. J Mol Evol 54: 763–773.
19. Hendrix RW (1999) Evolution: The long evolutionary reach of viruses. Curr Biol 9: R914–917.
20. Hendrix RW, Lawrence JG, Hatfull GF, Casjens S (2000) The origins and ongoing evolution of viruses. Trends Microbiol 8: 504–508.
21. Juhala RJ, Ford ME, Duda RL, Youlton A, Hatfull GF, et al. (2000) Genomic

sequences of bacteriophages HK97 and HK022: Pervasive genetic mosaicism in the lambdoid bacteriophages. J Mol Biol 299: 27–51.
22. Mann NH, Cook A, Millard A, Bailey S, Clokie M (2003) Bacterial photosynthesis genes in a virus. Nature 424: 741.
23. Botstein D (1980) A theory of modular evolution for bacteriophages. Ann New York Acad Sci 354: 484–491.
24. Hendrix RW, Smith MC, Burns RN, Ford ME, Hatfull GF (1999) Evolutionary relationships among diverse bacteriophages and prophages: All the world's a phage. Proc Natl Acad Sci U S A 96: 2192–2197.
25. Silander OK, Weinreich DM, Wright KM, O'Keefe KJ, Rang CU, et al. (2005) Widespread genetic exchange among terrestrial bacteriophages. Proc Natl Acad Sci U S A 102: 19009–19014.
26. Filee J, Tetart F, Suttle CA, Krisch HM (2005) Marine T4-type bacteriophages, a ubiquitous component of the dark matter of the biosphere. Proc Natl Acad Sci U S A.
27. Breitbart M, Miyake JH, Rohwer F (2004) Global distribution of nearly identical phage-encoded DNA sequences. FEMS Microbiol Lett 236: 249–256.
28. Mann NH, Clokie MR, Millard A, Cook A, Wilson WH, et al. (2005) The genome of S-PM2, a "photosynthetic" T4-type bacteriophage that infects marine *Synechococcus*. J Bacteriol 187: 3188–3200.
29. Millard A, Clokie MR, Shub DA, Mann NH (2004) Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains. Proc Natl Acad Sci U S A 101: 11007–11012.
30. Sullivan MB, Coleman M, Weigele P, Rohwer F, Chisholm SW (2005) Three *Prochlorococcus* cyanophage genomes: Signature features and ecological interpretations. PLoS Biol 3: e144. DOI: 10.1371/journal.pbio.0030144
31. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW (2005) Photosynthesis genes in marine viruses yield proteins during host infection. Nature 438: 8689.
32. Clokie MRJ, Shan J, Bailey S, Jia Y, Krisch HM (2006) Transcription of a 'photosynthetic' T4-type phage during infection of a marine cyanobacterium. Environ Microbiol 8: 827–835.
33. Benson R, Martin E (1981) Effects of photosynthetic inhibitors and light-dark regimes on the replication of cyanophage SM-2. Arch Microbiol 129: 165–167.
34. Adir N, Zer H, Schochat S, Ohad I (2003) Photoinhibition—A historical perspective. Photosynth Res 76: 343–370.
35. Paul JH, Sullivan MB (2005) Marine phage genomics: What have we learned? Curr Opin Biotechnol 16: 299–307.
36. 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.
37. Zeidner G, Bielawski JP, Shmoish M, Scanlan DJ, Sabehi G, et al. (2005) Potential photosynthesis gene recombination between *Prochlorococcus* and *Synechococcus* via viral intermediates. Environ Microbiol 7: 1505–1513.
38. Sherman LA (1976) Infection of *Synechococcus cedrorum* by the cyanophage AS-1M. III. Cellular metabolism and phage development. Virology 71: 199–206.
39. Wilson WH, Carr NG, Mann NH (1996) The effect of phosphate status on the kinetics of cyanophage infection in the oceanic cyanobacterium *Synechococcus* sp. WH7803. J Phycol 32: 506–516.
40. Levin BR, Lenski RE (1983) Coevolution in bacteria and their viruses and plasmids. In: Futuyma DJ, Slatkin M, editors. Coevolution. Sunderland (Massachusetts): Sinauer. pp. 99–127.
41. Wang IN, Dykhuizen DE, Slobodkin LB (1996) The evolution of phage lysis timing. Evol Ecol 10: 545–558.
42. Abedon ST (1989) Selection for bacteriophage latent period length by bacterial density: A theoretical examination. Microb Ecol 18: 79–88.
43. Abedon ST, Herschler TD, Stopar D (2001) Bacteriophage latent-period evolution as a response to resource availability. Appl Environ Microbiol 67: 4233–4241.
44. Abedon ST, Hyman P, Thomas C (2003) Experimental examination of bacteriophage latent-period evolution as a response to bacterial availability. Appl Environ Microbiol 69: 7499–7506.
45. Moore LR, Rocap G, Chisholm SW (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. Nature 393: 464–467.
46. Lindell D, Penno S, Al-Qutob M, David E, Rivlin T, et al. (2005) Expression of the nitrogen stress response gene *ntcA* reveals nitrogen-sufficient *Synechococcus* populations in the oligotrophic northern Red Sea. Limnol and Oceanogr 50: 1932–1944.
47. Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. Appl Environ Microbiol 67: 991–994.
48. Rocap G, Distel D, Waterbury JB, Chisholm SW (2002) Resolution of *Prochlorococcus* and *Synechococcus* ecotypes by using 16S-23S ribosomal DNA internal transcribed spacer sequences. Appl Environ Microbiol 68: 1180–1191.
49. Fuller NJ, Marie D, Partensky F, Vaulot D, Post AF, et al. (2003) Clade-specific 16S ribosomal DNA oligonucleotides reveal the predominance of a single marine *Synechococcus* clade throughout a stratified water column in the Red Sea. Appl Environ Microbiol 69: 2430–2443.
50. Golden SS, Brusslan J, Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. Embo J 5: 2789–2798.
51. Sicora CI, Appleton SE, Brown CM, Chung J, Chandler J, et al. (2006) Cyanobacterial *psbA* families in Anabaena and Synechocystis encode trace, constitutive and UVB-induced D1 isoforms. Biochim Biophys Acta 1757: 47–56.
52. Clarke AK, Soitamo A, Gustafsson P, Oquist G (1993) Rapid interchange between two distinct forms of cyanobacterial photosystem II reaction-center protein D1 in response to photoinhibition. Proc Natl Acad Sci U S A 90: 9973–9977.
53. Hess WR, Weihe A, Loiseaux-de Goer S, Partensky F, Vaulot D (1995) Characterization of the single *psbA* gene of *Prochlorococcus marinus* CCMP1375 (Prochlorophyta). Plant Mol Biol 27: 1189–1196.
54. Calendar R (1988) The bacteriophages. New York: Plenum.
55. Ackermann HW, DuBow MS (1987) Viruses of prokaryotes, Volume 1. General properties of bacteriophages. Boca Raton (Florida): CRC Press.
56. Karl DM (1999) A sea of change: Biogeochemical variability in the North Pacific Subtropical Gyre. Ecosystems 2: 181–214.
57. Zinser ER, Coe A, Johnson ZI, Martiny AC, Fuller NJ, et al. (2006) *Prochlorococcus* ecotype abundances in the North Atlantic Ocean as revealed by an improved quantitative PCR method. Appl Environ Microbiol 72: 723–732.
58. Dufresne A, Salanoubat M, Partensky F, Artiguenave F, Axmann IM, et al. (2003) Genome sequence of the cyanobacterium *Prochlorococcus marinus* SS120, a nearly minimal oxyphototrophic genome. Proc Natl Acad Sci U S A 100: 10020–10025.
59. Palenik B, Brahamsha B, McCarren J, Waterbury J, Allen E, et al. (2003) The genome of a motile marine *Synechococcus*. Nature 424: 1037–1041.
60. Rocap G, Larimer FW, Lamerdin J, Malfatti S, Chain P, et al. (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. Nature 424: 1042–1047.
61. Zeidner G, Preston CM, Delong EF, Massana R, Post AF, et al. (2003) Molecular diversity among marine picophytoplankton as revealed by *psbA* analyses. Environ Microbiol 5: 212–216.
62. Thompson JR, Marcelino LA, Polz MF (2002) Heteroduplexes in mixed-template amplifications: Formation, consequence and elimination by 'reconditioning PCR.' Nucleic Acids Res 30: 2083–2088.
63. Coleman ML, Sullivan MB, Martiny AC, Steglich C, Barry K, et al. (2006) Genomic islands and the ecology and evolution of *Prochlorococcus*. Science 311: 1768–1770.
64. Hess WR, Rocap G, Ting CS, Larimer FW, Stilwagen S, et al. (2001) The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. Photosynth Res 70: 53–71.
65. Acinas SG, Sarma-Rupavtarm R, Klepac-Ceraj V, Polz MF (2005) PCR-induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. Appl Environ Microbiol 71: 8966–8969.
66. Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425.
67. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174.
68. Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. J Mol Evol 39: 306–314.
69. Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504.
70. Phillips MJ, Delsuc F, Penny D (2004) Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 21: 1455–1458.
71. Kennedy M, Holland BR, Gray RD, Spencer HG (2005) Untangling long branches: Identifying conflicting phylogenetic signals using spectral analysis, neighbor-net, and consensus networks. Syst Biol 54: 620–633.
72. Bryant D, Moulton V (2004) Neighbor-net: An agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21: 255–265.
73. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254–267.
74. Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. Syst Zool 27: 401–410.
75. Lockhart PJ, Steel M, Hendy MD, Penny D (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. Mol Biol Evol 11: 605–612.
76. Schierup MH, Hein J (2000) Consequences of recombination on traditional phylogenetic analysis. Genetics 156: 879–891.
77. Yang Z, Roberts D (1995) On the use of nucleic acid sequences to infer early branchings in the tree of life. Mol Biol Evol 12: 451–458.
78. Padidam M, Sawyer S, Fauquet CM (1999) Possible emergence of new geminiviruses by frequent recombination. Virology 265: 218–225.
79. Maynard Smith J (1992) Analyzing the mosaic structure of genes. J Mol Evol 34: 126–129.
80. Posada D, Crandall KA (2001) Evaluation of methods for detecting recombination from DNA sequences: Computer simulations. Proc Natl Acad Sci U S A 98: 13757–13762.
81. Martin D, Rybicki E (2000) RDP: Detection of recombination amongst aligned sequences. Bioinformatics 16: 562–563.
82. Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13: 555–556.
83. Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J Mol Evol 29: 170–179.
84. Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods and combining nonnested models with applications to phylogenetic inference. Mol Biol Evol 16: 1114–1116.