Genetic organization of the *psbAD* region in phages infecting marine *Synechococcus* strains

Andrew Millard[†], Martha R. J. Clokie[†], David A. Shub[‡], and Nicholas H. Mann[†§]

[†]Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, United Kingdom; [‡]Department of Biological Sciences, University at Albany, State University of New York, Albany, NY 12222

Edited by W. Ford Doolittle, Dalhousie University, Halifax, Canada, and approved June, 15, 2004 (received for review March 2, 2004)

The discovery of the genes *psbA* and *psbD*, encoding the D1 and D2 core components of the photosynthetic reaction center PSII (photosystem II), in the genome of the bacteriophage S-PM2 (a cyanomyovirus) that infects marine cyanobacteria begs the question as to how these genes were acquired. In an attempt to answer this question, it was established that the occurrence of the genes is widespread among marine cyanomyovirus isolates and may even extend to podoviruses. The phage *psbA* genes fall into a clade that includes the *psbA* genes from their potential *Synechococcus* and *Prochlorococcus* hosts, and thus, this phylogenetic analysis provides evidence to support the idea of the acquisition of these genes by horizontal gene transfer from their cyanobacterial hosts. However, the phage *psbA* genes form distinct subclades within this lineage, which suggests that their acquisition was not very recent. The *psbA* genes of two phages contain identical 212-bp insertions that exhibit all of the canonical structural features of a group I self-splicing intron. The different patterns of genetic organization of the *psbAD* region are consistent with the idea that the *psbA* and *psbD* genes were acquired more than once by cyanomyoviruses and that their horizontal transfer between phages via a common phage gene pool, as part of mobile genetic modules, may be a continuing process. In addition, genes were discovered encoding a high-light inducible protein and a putative key enzyme of dark metabolism, transaldolase, extending the areas of host-cell metabolism that may be affected by phage infection.

Strains of unicellular cyanobacteria *Synechococcus* and *Prochlorococcus* are abundant in the world's oceans, and they dominate the prokaryotic component of the picophytoplankton. Together, they contribute 32–89% of primary production in oligotrophic regions of the oceans (1–4). The recent discovery (5) that a phage infecting marine *Synechococcus* strains encodes key photosynthetic genes has important implications for our understanding of the effect of phage infection on the photosynthetic picophytoplankton physiology and consequent impact on major biogeochemical cycles. The acquisition of photosynthesis genes by a phage, presumably by horizontal gene transfer, begs the questions of (*i*) how common the possession of photosynthesis genes by such phages is, (*ii*) where the acquired genes were from, and (*iii*) whether the acquisition was a single rare ancestral event or a common phenomenon in the oceans.

The cyanobacterial picoplankton, together with all organisms capable of oxygenic photosynthesis, possess two photosynthetic reaction centers, PSI and PSII (photosystems I and II). The PSII complex is a large protein–pigment assembly in the thylakoid membrane that catalyses the light-dependent oxidation of water to molecular oxygen. At the core of PSII lies a heterodimer of two related proteins, D1 and D2, which binds the pigments and cofactors necessary for primary photochemistry. During active photosynthesis, D1 and, to a lesser extent, D2 turn over rapidly as a result of photodamage and are replaced by newly synthesized polypeptides in a repair cycle. When the rate of photoinactivation and damage of D1 exceeds the capacity for repair, photoinhibition occurs, resulting in a decrease in the maximum efficiency of PSII photochemistry (6). Environmental stresses, such as UV-B, may enhance the rate of damage such that photoinhibition may occur at light fluence rates that would not normally exert this effect (7). Fluence rates of both photosynthetically active radiation and UV light in oceanic ecosystems may be high enough, particularly in surface waters, to cause photoinhibition (8).

Viruses in general, and bacteriophages in particular, have been shown to be abundant in marine ecosystems and are thought to exert major biogeochemical and ecological effects (9). Viral infection of marine unicellular cyanobacteria was first reported in 1990 (10, 11), and isolates of these cyanoviruses were first characterized in 1993 (12–14). Recently, it was shown that one of these phages, S-PM2, which is capable of infecting *Synechococcus* strains, carries copies of the *psbA* and *psbD* genes encoding the D1 and D2 proteins of PSII (5). It is speculated that the expression of phage-encoded D1 and D2 proteins in infected cells would permit a continued PSII repair cycle to operate after host protein synthesis had been shut down, thus maintaining the photosynthetic activity of the cells and concomitant oxygen evolution and ensuring the provision of energy for extended viral replication. The D1 protein encoded by S-PM2 is similar to the D1 proteins of marine *Synechococcus* sp. WH8102, and indeed, homology can be detected at the DNA sequence level, suggesting that S-PM2 acquired the gene horizontally from its *Synechococcus* host. By establishing whether the presence of *psbA* genes in cyanophage genomes is a widespread phenomenon and, if so, whether gene organization is similar in geographically distinct isolates, it becomes possible to establish whether the acquisition was a single rare ancestral event or is a common phenomenon in the oceans.

Materials and Methods

Isolation, Propagation, and Maintenance of Bacteriophage Strains. Phages were isolated and propagated by using *Synechococcus* sp. WH7803, grown in artificial sea water, as described by Wilson *et al.* (15). The isolation details of the phages characterized in detail are shown in Table 1, and the same information for phages screened only for the presence of *psbA* genes is provided in Table 2, which is published as supporting information on the PNAS web site.

PCRs, Southern Blotting, and DNA Sequencing. DNA from cyanophages was extracted by using the method described by Wilson *et al.* (14). Homologs of *psbA* in other cyanophages were initially detected by Southern blotting. A hybridization probe was prepared from a 525-bp PCR product from S-PM2, which was amplified by using the following primers: S-PM2F1S, 5′-GCTGCTTCTCTTGATGAGTG-3′; and S-PM2R2S, 5′-AGTGTAGCGAACGAGAGTTG-3′. The PCRs were carried

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: HLIP, high-light inducible protein; BAC, bacterial artificial chromosome.

Data deposition: The sequences reported in this article have been deposited in the EMBL database (accession nos. AJ628768, AJ628769, AJ628858, AJ629075, AJ629221, and AJ630128).

[§]To whom correspondence should be addressed. E-mail: n.h.mann@warwick.ac.uk.

EVOLUTION

Table 1. Phages used in this study

Phage	Locale	Latitude and longitude	Depth, m	Date collected	Source
S-PM2*	English Channel	50°18′N, 4°12′W	0	September 23, 1992	Ref. 14
S-WHM1	Woods Hole Harbor	41°31′N, 71°40′W	0	August 11, 1992	Ref. 14
S-RSM2	Gulf of Aqaba	29°28′N, 34°55′ E	0	March 14, 1994	W. Wilson, personal communication
S-RSM28	Gulf of Aqaba	29°28′N, 34°55′E	50	May 11, 1999	This study
S-RSM88	Gulf of Aqaba	29°28′N, 34°55′ E	125	April 19, 1999	This study
S-BM4	Coastal Bermuda	32°17′N, 64°53′W	0	August 25, 1995	Ref. 14

*Formerly designated S-PS1.

out in a total volume of 50 μl, containing 200 μM dNTPs, 2 mM $MgCl_2$, 50 nM primers, 2 μl of S-PM2 DNA, 2 units of *Taq* polymerase, and 1× enzyme buffer (Helena Biosciences). Amplification conditions were as follows: 92°C for 2 min, 30 cycles of 92°C for 30 sec, 47°C for 30 sec, 72°C for 30 sec, with a final extension of 4 min at 72°C. The PCR product was gel-extracted and labeled with [α-^{32}P]ATP by using DNA polymerase I Klenow fragment in 1× labeling buffer (Promega) at 25°C for 2 h. Unincorporated nucleotides were removed by purification through a Sephadex G-25 column (16). The labeled S-PM2 *psbA* PCR product was used as a hybridization probe against cyanophage DNA that had been digested with *Eco*RI and run on a 0.75% agarose gel. Transfer to a nylon membrane (Amersham Biosciences) was achieved by the method described by Sambrook *et al.* (16). Blots were prehybridized in buffer (5× Denhardt's solution/6× 0.15 M NaCl/10 mM phosphate, pH 7.4/1 mM EDTA (SSPE)/0.5% SDS (wt/vol) for 60 min. Hybridization was at 65°C overnight in fresh buffer (5× Denhardt's solution/6× SSPE/0.5% SDS, wt/vol). Membranes were washed three times for 10 min in 2× SSC/0.1% (wt/vol) SDS at 63°C; twice in 1× SSC/0.1% (wt/vol) SDS at 63°C for 10 min; and four times in 0.1× SSC/0.1% (wt/vol) SDS at 63°C for 5 min.

The primers for used screening and sequencing the *psbA* regions of cyanophage genomes were based on those of Zeidner *et al.* (17). The reverse primer designed in this study spanned the intron-insertion site in the *psbA* gene of cyanophage S-PM2 and was shortened to omit the region corresponding to the first exon. Therefore, primers could amplify *psb*A genes whether or not they contained an intron. The following primers were used: *psbAF*, 5′-GTNGAYATHGAYGGNATHMGNGARCC-3′; and *psbAR* (2), 5′-GGRAARTTRTGNGC-3′. PCRs were carried in a total volume of 25 μl, containing 200 μM dNTP, 2 mM MgCl2/50 nM primers, 2 μl of template, 1.5 units of *Taq* polymerase, and 1× enzyme buffer. Amplification conditions were as follows: 92°C for 2 min, 10 cycles of 92°C for 30 sec, 64°C (−1°C per cycle) for 30 sec, and 68°C for 1 min. There was then an extension of 2 min at 68°C, followed by 25 cycles of 92°C for 30 sec, 56.5°C for 30 sec, and 68°C for 1 min. The final extension was at 68°C for 4 min. Cyanophage DNA was screened for the presence of contaminating *Synechococcus* DNA by the use of primers specific to 16S ribosomal DNA from oxygenic phototrophs. The following primers were used: CYA106F, 5′-CGGACGGGTGAGTAACGCGTG-3′ (18); CYA781R(A), 5′-GACTACTGGGGTATCTAATCCCATT-3′ (19); and OXY1313R(B), 5′-GACTACAGGGGTATCTAATCCCTTT-3′ (18). The conditions used were the same as those used by Nubel *et al.* (19).

It was not possible to amplify the entire *psbA* and *psbD* gene regions of cyanophages by PCR directly because the N-terminal end of *psbA* is not sufficiently conserved to allow the design of degenerate primers. Therefore, to sequence this region and to identify genes upstream from the *psbA* and *psbD* regions, primer walking was used. The genomic DNA required for direct sequencing was produced by using the GenomiPhi DNA amplification kit (Amersham Biosciences) according to the manufacturer's instructions, yielding 4–7 μg of DNA from 15 ng of starting material. For the sequencing reaction, 1 μg of genomic DNA was added to 5 pmol of primer, and the total volume was made up to 6 μl with sterile water. Sequence data were imported into SEQMAN (DNASTAR, Madison, WI). Primers specific to the new sequence were designed by using PRIMER DESIGNER 3.0 (Scientific and Educational Software, Durham, NC), and the process was repeated until the sequence information for the genes upstream and downstream from the *psbA* and *psbD* regions had been obtained. Primers were synthesized commercially by TAG Newcastle (Newcastle, U.K.).

DNA Sequence and Phylogenetic Analysis. Analysis of DNA sequences was carried by the identification of ORFs by using ORF FINDER (available at www.ncbi.nlm.nih.gov/gorf/gorf.html) and subsequent BLAST searching (20). The subcellular localization of proteins was predicted by using PSORT-B (21). Sequences were aligned by using CLUSTALW (22), and the output was checked and corrected manually.

Phylogenetic trees were constructed by using PAUP* 4.0 (23) and MRBAYES (24). In the case of MRBAYES, trees were determined by using 500,000 iterations with a sample frequency of 100 with a burn-in of 50. All analyses in PAUP* were performed by using branch-and-bound searches, with the "collapse" option and "furthest" addition sequence selected. For nucleotide alignments, any gaps in the data matrix were treated as missing data, and indels (insertions or deletions) were coded separately and appended to the sequence data matrix. Coding of indels was binary (deletions, 0; insertions, 1). Data were analyzed by using parsimony and distance methods in PAUP* 4.0, with no weighting or ordering imposed on the characters. Support for clades was estimated by means of nonparametric bootstrap analyses, as implemented in PAUP* 4.0 by using 1,000 replicates.

Results

Bacteriophages capable of infecting marine phycoerythrin-containing *Synechococcus* strains were first characterized in 1993, and most isolates were myoviruses with icosahedral heads and contractile tails (12–14). The myovirus S-PM2, which was originally isolated from sea water from the English Channel by using *Synechococcus* sp. WH7803 as a host, has been shown to encode the *psbA* and *psbD* genes specifying the D1 and D2 proteins of PSII (5). To determine whether the acquisition of the *psbA* and *psbD* genes by myoviruses infecting marine *Synechococcus* strains was a rare or common event, it was decided to establish whether the *psbA* gene could be detected in other marine cyanophage isolates and, if so, to compare the genetic organization of the *psbA* region in geographically distinct isolates and multiple isolates from the same location.

The PCR product from phage S-PM2 was used initially as a hybridization probe in Southern blots of phage DNAs, including S-WHM1, S-BM4, and S-RSM2, which have geographically distinct isolation sites and gave positive signals for the presence

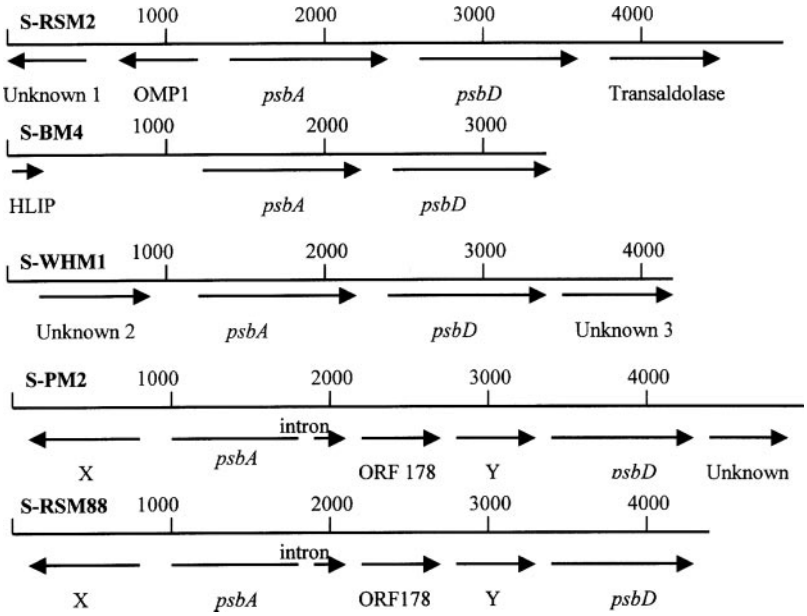

Fig. 1. Genetic organization of the *psbAD* region in the following five cyanomyoviruses infecting *Synechococcus* sp. WH7803: S-RSM2 , S-BM4, S-WHM1, S-PM2, and S-RSM88 (GenBank accession nos. AJ628768, AJ628858, AJ628769, AY329638, and AJ629075, respectively). OMP, putative outer membrane protein. Genes X and Y encode homologues of two proteins of unknown function in both S-RSM88 and S-PM2. All other unknown genes exhibit no similarity to each other.

of a *psbA* gene (data not shown), which was subsequently confirmed by DNA sequencing. A large number of myovirus isolates from the Red Sea were then screened for the occurrence of a *psbA* gene by using the degenerate PCR primers (data not shown). In total, the combination of Southern blotting and PCR screening indicated that 37 of 68 isolates (all myoviruses) carried a copy of the *psbA* gene, suggesting that the possession of *psbA* genes by marine cyanomyoviruses is a common phenomenon. Two phage isolates, in addition to S-PM2, yielded a larger PCR product, indicative of the presence of the putative intron. In the case of S-RSM88, this intron was confirmed by DNA sequencing. Five of the strains giving a positive signal for the presence of the *psbA* gene were selected for more detailed analysis. The S-RSM2, S-RSM28, and S-RSM88 strains are isolates obtained from the Gulf of Aqaba in the Red Sea, whereas S-BM4 and S-WHM1 were obtained from Bermuda and Woods Hole Harbor, respectively.

To investigate the genetic organization of the phages in the region of the *psbA* gene, the regions upstream and downstream were sequenced. Initially, the PCR products obtained by using the *psbA* primers were sequenced and the sequence was extended from the original PCR primers. Primers were then designed to sequence walk in either direction from *psbA* and to make the sequence double-stranded. All of the phages studied carried the *psbD* gene in addition to *psbA*. The genetic organization of the *psbAD* region of each of the phages is shown in Fig. 1. S-RSM88 exhibits an identical organization to S-PM2 with *psbA* and *psbD* being separated by two ORFs. Indeed, with the exception of a single-base-pair substitution in ORF 178, the DNA sequences of S-PM2 and S-RSM88 are identical in the 3,555 bp from the beginning of *psbA* to the end of *psbD*. This similarity extends to the presence of a group I intron in identical positions in the two *psbA* genes, inserted between codons 334 and 335. Although comprising only 212 nucleotides, the intron can be folded into a secondary structure containing all of the canonical group I structural features (Fig. 2) (25, 26). The similarity between S-PM2 and S-RSM88 also extends upstream from *psbA*. Thus, it would appear that the *psbA* and *psbD* genes in these two phages might occur as part of a conserved module that, given the considerable geographical separation of their isolation sites, is mobile.

Analysis of the genetic organization of the *psbA* region of phages S-RSM2, S-BM4, and S-WHM1, however, gives contrasting results (Fig. 1). In all these phages, the *psbA* and *psbD* genes are immediately adjacent to each other, but the *psbA* and *psbD* genes differ in nucleotide sequence from each other. Each phage is quite distinct in the ORFs adjacent to *psbAD*. In the case of S-RSM2, upstream from *psbA* and on the opposite strand are two genes, one of which is predicted by PSORT (21) to have the characteristics of membrane protein. Downstream of *psbD* is an ORF encoding a putative transaldolase, complete with MipB domain. However, the closest scoring match is to a transaldolase

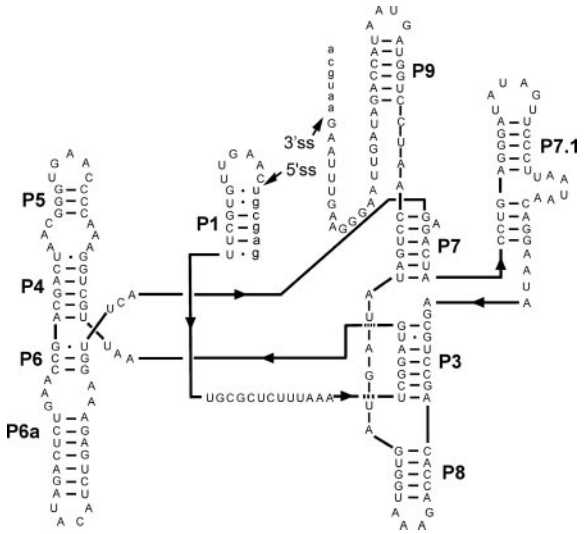

Fig. 2. Secondary structure of the group I intron in *psbA*. Exon sequences are given in lowercase letters, and intron sequences are given in uppercase letters. Arrows indicate splice sites (ss). Bold lines show connections between intron structure domains, with pointers indicating the 5′ to 3′ direction.

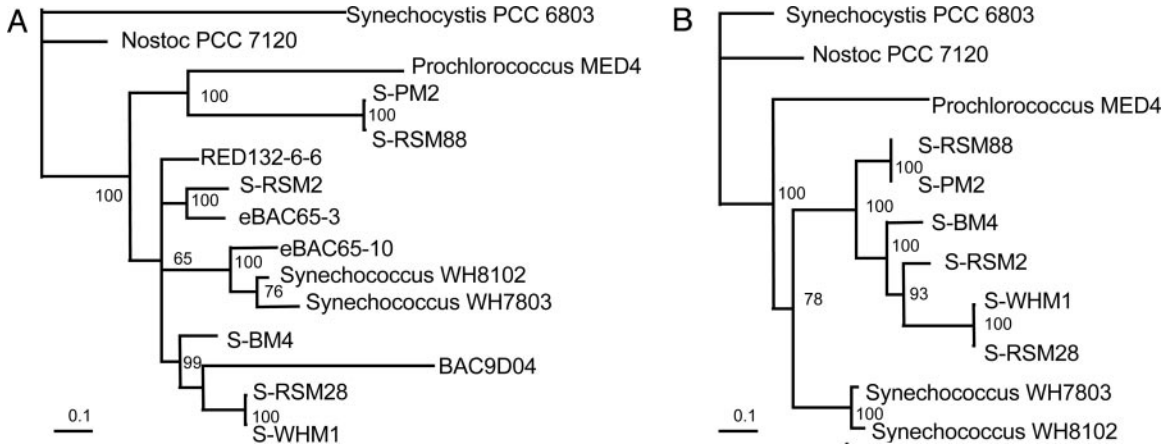

Fig. 3. Phylogenetic analysis using MRBAYES of the *psbA* genes (*A*) and *psbD* genes (*B*) from the following six marine cyanomyoviruses: S-RSM2, S-BM4, S-WHM1, S-PM2, S-RSM28, and S-RSM88 (GenBank accession nos. AJ628768, AJ628858, AJ628769, AY329638, AJ629221, and AJ629075, respectively) and the *psbA* gene (*A*) from a putative podovirus BAC9D04 (GenBank accession no. AY45612). Both *psbA* and *psbD* sequence data were available for *Synechococcus* sp. WH8102, *Synechococcus* sp. WH7803 (*psbD*; F. Partensky and D. Scanlan, personal communication), *Prochlorococcus marinus* MED 4 and the freshwater cyanobacteria *Nostoc* sp. PCC 7120 and *Synechocystis* sp. PCC 6803 (*psbA*1) (GenBank accession nos. NC 005070, AF156980, BX572091, NC 003272, and BA000022, respectively). Additional sequences from Zeidner *et al.* (17) were also included for the *psbA* analysis. These sequences were as follows: RED132-6-6, eBAC65-10, and eBAC65-3 (GenBank accession nos. AY176632, AY176622, and AY176623, respectively); corresponding *psbD* sequences are not available. Trees are unrooted and were generated from nucleotide alignments in the case of both *psbA* and *psbD*; data sets were 1,369 and 1,077 nt, respectively. Clade support values are shown at the nodes of the clades.

from *Caulobacter crescentus* (*E* value, $8e^{-51}$) and not to cyanobacterial transaldolases. It is worth noting that transaldolase is a key enzyme of the oxidative pentose phosphate pathway, which is the primary route for dark oxidative metabolism in cyanobacteria (27). The *psbAD* region in S-WHM1 is bounded by two ORFs encoding polypeptides with no significant similarities to any entries in the protein databases. The polypeptide encoded by the ORF adjacent to *psbA* has a predicted secretion signal sequence. Upstream of the *psbA* gene in phage S-BM4 is an ORF encoding a polypeptide with similarity to a high-light inducible protein (HLIP). HLIP genes are found in both marine *Synechococcus* and *Prochlorococcus* strains, and on the basis of cluster analysis, some of the *hli* genes may be specific to marine cyanobacteria (28). In the case of the freshwater cyanobacterium *Synechocystis* sp. PCC6803, HLIPs have been implicated in the adaptation to variations in light intensity (29).

Phylogenetic analysis of the *psbA* and *psbD* genes was carried out by the following three different methods: the maximum parsimony and distance methods of PAUP* and MRBAYES, which employs a Bayesian approach. All three methods gave essentially the same results, and only results obtained with MRBAYES are presented here. A phylogenetic analysis of the phage-encoded *psbA* genes and other cyanobacterial *psbA* genes at the amino acid level gives no resolution of taxa (tree not shown). However, analysis at the nucleotide level (Fig. 3*A*) shows that there is a major clade with 100% support, which contains all of the phage sequences and the sequences from marine *Synechococcus* and *Prochlorococcus* strains. The *psbA* sequences from the freshwater cyanobacteria *Nostoc* and *Synechocystis* form an outgroup. Within the marine clade, the phage *psbA* genes form several distinct groups. The first group contains the identical S-RSM88 and S-PM2, with a clade support value of 100%. Sister to this group, with 100% clade support, is the *psbA* gene from *Prochlorococcus* MED4. A further distinct phage group supported with a clade support value of 100% contains the S-RSM28 and S-WHM1 *psbA* genes. Sister to this group, but with low clade support, is the *psbA* gene from a putative podophage (BAC9D04) (30), identified as such by the similarity of genes adjacent to *psbA* to genes in the marine cyanopodovirus P60 (data not shown), although this phage does not possess a copy of *psbA* (31). The phage S-RSM2 occupies a group with 100% clade support with eBAC65-3, which encodes a *psbA* gene from an environmental bacterial artificial chromosome (BAC) library. However, the BAC library was prepared in such way that phages would not be excluded (Oded Béja, personal communication), and thus, the origin of the gene is unclear. The final phage involved in this study, S-BM4, represents a sister group of S-RSM28 and S-WHM1. The two *Synechococcus* strains form a group with the putative *Synechococcus psbA* eBAC65-10.

A phylogenetic analysis of *psbD* genes from the same sources (with the exception of those from BAC libraries) shows an only slightly different picture of the relationships from groupings obtained with *psbA*. The first thing to note is that there is again a major clade with clade support of 100% that contains all of the phage, *Synechococcus*, and *Prochlorococcus psbD* genes. There is also good support (78%) both for a clade containing *Synechococcus* and the phages and for a subgroup containing all of the phages (100%). However, the S-PM2 and S-RSM88 phages no longer form a sister group with *Prochlorococcus* MED4 as they do for *psbA*. Within the phage, clades S-BM4, S-RSM2, S-WHM1, and S-RSM28 form a group with 100% support, and within this group, the two phages S-WHM1 and S-RSM28 form a well supported (93%) subgroup as they did for the *psbA* analysis.

Discussion

The phage *psbA* genes fall into a clade that includes the *psbA* genes from their potential *Synechococcus* and *Prochlorococcus* hosts, supporting the idea that these genes were indeed acquired horizontally from their hosts and this is in agreement with the results obtained by Lindell *et al.* (32) for *Prochlorococcus* phages. A similar observation was made for *psbD*. However, the phage *psbA* genes form distinct subclades within this lineage, which suggests that their original acquisition was not very recent or that there is a very strong selection pressure on the *psbA* gene in the phage as compared with the organismal context. Another possibility, albeit less likely only because of a number of *Synechococcus psbA* genes that have now been sequenced, is that representative *psbA* genes from the actual cyanobacterial sources have not been characterized. The *psbD* genes show much

better resolution than the *psbA* genes, which may suggest that *psbD* has a more ancient ancestry within the phage lineage, leading to two well defined radiations. An alternative explanation might be that there are tighter functional constraints on the evolution of *psbA* than there are on the evolution of *psbD*.

The different patterns of genetic organization of the *psbAD* region in the different phages suggest either that the *psbA* and *psbD* genes were acquired more than once by cyanomyoviruses or that horizontal transfer of these genes among phages through a common cyanomyoviral gene pool may be a common process. Indeed, these two options are not mutually exclusive. Thus, the original acquisition of the genes by phages may not have been recent, but when acquired, transfer of the genes among phages may be commonplace. An alternative explanation might be that the *psbA* and *psbD* genes were acquired in a single event and that subsequent horizontal transfer events have disrupted the linkage and might also have replaced the original alleles by homologous recombination. In this context, it is worth noting that it has been suggested that all of the double-stranded DNA-tailed phages might share common ancestry and that all double-stranded DNA phage genomes are mosaics with access, by horizontal exchange, to a large common genetic pool (33). It has been shown that phage S-PM2 shares a genetic module with the ecologically and evolutionarily distant coliphage T4 (34), and the fact that a putative podovirus (BAC9D04) (30) carries a copy of the *psbA* gene tends to confirm the breadth of this gene pool. The presence of *psbA* genes in cyanopodoviruses is confirmed by sequence analysis of a cyanopodovirus infecting *Prochlorococcus* strains (32). The difference in genetic organization between these phages, in which *psbA* and *psbD* are immediately adjacent to each other, and their hosts, in which these two genes are widely separated (35–37), also suggests that the two genes may have been acquired independently.

Another observation that supports the idea of a dynamic gene pool, which includes photosynthesis genes, is the near identity of the *psbAD* region in two geographically distinct cyanomyovirus isolates, S-PM2 and S-RSM88. This result is consistent with the spread of a mobile "*psbAD* module" through the common phage gene pool and is supported by the fact that the two phages have quite different genome sizes (A. Millard, personal communication). Although no introns were observed in the *Prochlorococcus* phages studied by Lindell *et al.* (32), the *psbA* genes of these two phages form a single clade with *Prochlorococcus* MED4, raising the possibility that, although these two phages will infect *Synechococcus* strains, they may have ultimately acquired their *psbA* genes from *Prochlorococcus*.

The origin of the group I intron in the *psbA* gene of S-PM2 and S-RSM88 is not clear. Group I introns have been observed in many cyanobacteria but so far only in tRNA genes. Furthermore, these tRNA introns belong to subfamily IC3 and show no sign of relationship to the subfamily IA1 phage *psbA* introns (26). Interestingly, however, the *psbA* gene of *Chlamydomonas moewusii* chloroplasts contains two group I introns and the *psbA* gene of *C. reinhardtii* has four group I introns: intron 1 of *C. moewusii* and intron 4 of *C. reinhardtii* (both closely related to group IA1) sharing the same insertion site (38, 39). However, none of the five sites of group I intron insertion correspond to the placement of the intron in the phage *psbA* genes. Although capable of self-splicing *in vitro*, splicing of the introns in *psbA* of *C. reinhardtii* is stimulated by light *in vivo* (40), raising the interesting possibility that they participate in posttranscriptional regulation of gene expression.

Remarkably, some of the chloroplast introns are highly similar to introns in bacteriophages, including phage T4 (39), and the cyanophages described here bear a striking resemblance, both in morphology and in aspects of genome composition, to phage T4 and its relatives (34). Collectively, these correlations suggest a long-standing trafficking of introns between phages and *psbA* genes.

The high degree of sequence identity of the *psbAD* cassettes of S-PM2 and S-RSM88 suggests a fairly recent lateral transfer. Many group I introns contain ORFs that encode endonucleases with specificity for the intronless DNA of genes that are homologous to their sites of insertion. By cleaving these intronless genes, the introns become duplicatively inserted into vacant sites during DNA repair, which is a gene conversion event (with coconversion of flanking DNA) that has been called "homing" (41), and the enzymes that initiate this process are called "homing endonucleases." Intron homing, with coconversion of flanking DNA, would provide a ready explanation for the near identity of the *psbAD* regions of S-PM2 and S-RSM88, except that the minimal phage introns described here do not contain any ORFs capable of encoding an endonuclease. How could these introns be mobile?

The answer may reside in the ORF178 immediately downstream of *psbA* in both phages, whose deduced protein product closely resembles another optional phage ORF, gene 13.5 of *Yersinia* phage ΦYe03–12 (42), and other phage proteins of unknown function, all of which show some similarity to a portion of phage T4 gp49 (endonuclease VII), which is involved in recombination and packaging of phage DNA. Although it is not related in sequence to any of the established homing endonucleases, the active center of gp49 has been reported to be structurally homologous to members of one family of homing endonucleases (43).

Interestingly, a homing endonuclease does not need to reside in an intron to promote homing. In some bacteriophages, genes whose products are related to intronic endonucleases are inserted intercistronically, between conserved genes that are adjacent in closely related phages (44–46). Several of these homologs have been shown to have endonuclease activity, cleaving other phage DNAs close to the endonuclease gene-insertion site. Repair of cleaved DNA inserts the endonuclease gene by gene conversion, which is exactly analogous to intron homing. In this "intronless homing," the endonuclease gene retains its function as a mobile element without being associated with an intron, and the coconversion tract of flanking DNA can extend several kilobases from the site of cleavage (47, 48). It is possible that the protein encoded by ORF178 next to the phage *psbA* gene functions in this way, by mobilizing the entire *psbAD* cassette for lateral transfer between related bacteriophages.

The idea that the acquisition by marine cyanophages of genes encoding components of the photosynthetic reaction center confers an advantage gains further support from the discovery of an HLIP-encoding gene, *hli*, in phage S-BM4. HLIPs have been implicated in the adaptation to variations in light intensity (29). Furthermore, multiple *hli* genes and other photosynthetic genes have been discovered in the genomes of phages infecting *Prochlorococcus* strains (32). It has been suggested (5) that the expression of virus-encoded D1 and D2 proteins in infected cells would permit a continued PSII repair cycle to operate after host protein synthesis had been shut down, thus maintaining the photosynthetic activity of the cells and concomitant oxygen evolution, as well as ensuring the provision of energy for extended viral replication. However, high light intensity leading to photoinhibition is not likely to be an important factor in the lower levels of the euphotic zone; therefore, any fitness benefits conferred on phage by carriage of the *psbA* gene are only likely to be significant in the surface layers of the oceans. These considerations would apply to the other photosynthesis genes. Cyanobacteria employ the oxidative pentose phosphate pathway as their source of maintenance energy in the dark (27), and the discovery of a gene encoding a putative transaldolase, a key enzyme of this pathway, suggests that some phages may be capable of influencing dark metabolism as well. Interestingly, it

has been proposed (49) that lytic phages infecting marine *Synechococcus* strains may be capable of establishing a pseudolysogenic state in cells that are nutrient-stressed. It should not be assumed that these genes play a role solely in the infection–lysis pathway but that they also might be expressed on a longer-term basis during the quasistable relationship of pseudolysogeny established in response to the nutrient-limited conditions, which apply in the oligotrophic central regions of the oceans.

The patchiness of the distribution of the *psbA* gene in cyanophages poses a question regarding ideas about its contribution to phage fitness. The single most important selection pressure on the evolution of phages is the density of infectable hosts. This factor, in the long term, will determine the length of the latent period and also whether lysogenic or lytic phages are likely to predominate. The current hypothesis regarding the contribution of the *psbA* gene to phage fitness assumes that the latent period is long enough for photoinhibition to restrict phage replication in the absence of a repair cycle. There is no direct method, to our knowledge, of determining what proportion of a natural cyanobacterial assemblage is susceptible to infection by cooccurring phages. It is very likely that different ecotypes within the assemblage will be susceptible to infection by different subsets of phages. Thus, one cyanobacterial ecotype constituting a significantly higher proportion of the assemblage might be susceptible to infection by phages with a short latent period and no *psbA* gene, whereas another ecotype at lower abundance may be infected by phages with a longer latent period, which would benefit from the carriage of the *psbA* gene. The situation is further complicated by the possibility of pseudolysogeny. Furthermore, it is quite conceivable that the host range of some cyanophages may extend to members of the heterotrophic bacterial community where possession of *psbA* would confer no benefit.

This work was supported by a Natural Environment Research Council grant (to N.H.M.) and National Institutes of Health Grant AI 57158 (to D.A.S.). A.M. was supported by a Natural Environment Research Council Studentship.

1. Goericke, R. & Welschmeyer, N. A. (1993) *Deep Sea Res. Part I Oceanogr. Res. Pap.* **40,** 2283–2294.
2. Li, W. K. W. (1995) *Mar. Ecol. Prog. Ser.* **122,** 1–8.
3. Liu, W. B., Nolla, H. A. & Campbell, L. (1997) *Aquat. Microb. Ecol.* **12,** 39–47.
4. Veldhuis, M. J. W., Kraay, G. W., VanBleijswijk, J. D. L. & Baars, M. A. (1997) *Deep Sea Res. Part I Oceanogr. Res. Pap.* **44,** 425–449.
5. Mann, N. H., Cook, A., Millard, A., Bailey, S. & Clokie, M. (2003) *Nature* **424,** 741–741.
6. Aro, E. M., Virgin, I. & Andersson, B. (1993) *Biochim. Biophys. Acta* **1143,** 113–134.
7. Campbell, D., Eriksson, M. J., Oquist, G., Gustafsson, P. & Clarke, A. K. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 364–369.
8. Piazena, H., Perez-Rodrigues, E., Hader, D. P. & Lopez-Figueroa, F. (2002) *Deep-sea Res. II, Top. Stud. Oceanogr.* **49,** 3513–3528.
9. Fuhrman, J. A. (1999) *Nature* **399,** 541–548.
10. Proctor, L. M. & Fuhrman, J. A. (1990) *Nature* **343,** 60–62.
11. Suttle, C. A., Chan, A. M. & Cottrell, M. T. (1990) *Nature* **347,** 467–469.
12. Suttle, C. A. & Chan, A. M. (1993) *Mar. Ecol. Prog. Ser.* **92,** 99–109.
13. Waterbury, J. B. & Valois, F. W. (1993) *Appl. Environ. Microbiol.* **59,** 3393–3399.
14. Wilson, W. H., Joint, I. R., Carr, N. G. & Mann, N. H. (1993) *Appl. Environ. Microbiol.* **59,** 3736–3743.
15. Wilson, W. H., Carr, N. G. & Mann, N. H. (1996) *J. Phycol.* **32,** 506–516.
16. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
17. Zeidner, G., Preston, C. M., Delong, E. F., Massana, R., Post, A. F., Scanlan, D. J. & Beja, O. (2003) *Environ. Microbiol.* **5,** 212–216.
18. West, N. J., Schonhuber, W. A., Fuller, N. J., Amann, R. I., Rippka, R., Post, A. F. & Scanlan, D. J. (2001) *Microbiology* **147,** 1731–1744.
19. Nubel, U., GarciaPichel, F. & Muyzer, G. (1997) *Appl. Environ. Microbiol.* **63,** 3327–3332.
20. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J. H., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
21. Gardy, J. L., Spencer, C., Wang, K., Ester, M., Tusnady, G. E., Simon, I., Hua, S., deFays, K., Lambert, C., Nakai, K. & Brinkman, F. S. L. (2003) *Nucleic Acids Res.* **31,** 3613–3617.
22. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997) *Nucleic Acids Res.* **24,** 4876–4882.
23. Swofford, D. L. (1998) PAUP*, Phylogenetic Analysis Using Parsimony (and Other Methods) (Sinauer, Sunderland, MA).
24. Huelsenbeck, J. P. & Ronquist, F. (2001) *Bioinformatics* **17,** 754–755.
25. Cech, T. R., Damberger, S. H. & Gutell, R. R. (1994) *Nat. Struct. Biol.* **1,** 273–280.
26. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
27. Schmetterer, G. (1994) in *The Molecular Biology of Cyanobacteria*, ed. Bryant, D. A. (Kluwer, Dordrecht, The Netherlands), pp. 409–435.
28. Bhaya, D., Dufresne, A., Vaulot, D. & Grossman, A. (2002) *FEMS Microbiol. Lett.* **215,** 209–219.
29. Havaux, M., Guedeney, G., He, Q. F. & Grossman, A. R. (2003) *Biochim. Biophys. Acta* **1557,** 21–33.
30. Zeidner, G. & Beja, O. (2004) *Environ. Microbiol.* **6,** 528–534.
31. Chen, F. & Lu, J. R. (2002) *Appl. Environ. Microbiol.* **68,** 2589–2594.
32. Lindell, D., Sullivan, M. B., Johnson, Z. I., Tolonen, A. C., Rohwer, F. & Chisholm, S. W. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 11013–11018.
33. Hendrix, R. W., Smith, M. C. M., Burns, R. N., Ford, M. E. & Hatfull, G. F. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2192–2197.
34. Hambly, E., Tetart, F., Desplats, C., Wilson, W. H., Krisch, H. M. & Mann, N. H. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 11411–11416.
35. Dufresne, A., Salanoubat, M., Partensky, F., Artiguenave, F., Axmann, I. M., Barbe, V., Duprat, S., Galperin, M. Y., Koonin, E. V., Le Gall, F., *et al*. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 10020–10025.
36. Rocap, G., Larimer, F. W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N. A., Arellano, A., Coleman, M., Hauser, L., Hess, W. R., *et al*. (2003) *Nature* **424,** 1042–1047.
37. Palenik, B., Brahamsha, B., Larimer, F. W., Land, M., Hauser, L., Chain, P., Lamerdin, J., Regala, W., Allen, E. E., McCarren, J., *et al*. (2003) *Nature* **424,** 1037–1042.
38. Turmel, M., Boulanger, J. & Lemieux, C. (1989) *Nucleic Acids Res.* **17,** 3875–3887.
39. Holloway, S. P., Deshpande, N. N. & Herrin, D. L. (1999) *Curr. Genet.* **36,** 69–78.
40. Deshpande, N. N., Bao, Y. J. & Herrin, D. L. (1997) *RNA* **3,** 37–48.
41. Dujon, B. (1989) *Gene* **82,** 91–114.
42. Pajunen, M. I., Kiljunen, S. J., Soderholm, M. E. L. & Skurnik, M. (2001) *J. Bacteriol.* **183,** 1928–1937.
43. Raaijmakers, H., Toro, I., Birkenbihl, R., Kemper, B. & Suck, D. (2001) *J. Mol. Biol.* **308,** 311–323.
44. Sharma, M., Ellis, R. L. & Hinton, D. M. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6658–6662.
45. Gorbalenya, A. E. (1994) *Protein Sci.* **3,** 1117–1120.
46. Shub, D. A., Goodrichblair, H. & Eddy, S. R. (1994) *Trends Biochem. Sci.* **19,** 402–404.
47. Belle, A., Landthaler, M. & Shub, D. A. (2002) *Genes Dev.* **16,** 351–362.
48. Liu, Q. Q., Belle, A., Shub, D. A., Belfort, M. & Edgell, D. R. (2003) *J. Mol. Biol.* **334,** 13–23.
49. Mann, N. H. (2003) *FEMS Microbiol. Rev.* **27,** 17–34.