J Mol Evol (1998) 46:188–201

JOURNAL OF MOLECULAR EVOLUTION

Rapid Diversification of Marine Picophytoplankton with Dissimilar Light-Harvesting Structures Inferred from Sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria)

Ena Urbach,[1] David J. Scanlan,[2] Daniel L. Distel,[3] John B. Waterbury,[3] Sallie W. Chisholm[1]

[1] Ralph M. Parsons Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2] Department of Biological Sciences, University of Warwick, Gibbet Hill Rd. Coventry CV4 7AL, United Kingdom
[3] Woods Hole Oceanographic Institution, Woods Hole MA 02543, USA

Received: 18 January 1997 / Accepted: 18 May 1997

Abstract. Cultured isolates of the unicellular planktonic cyanobacteria *Prochlorococcus* and marine *Synechococcus* belong to a single marine picophytoplankton clade. Within this clade, two deeply branching lineages of *Prochlorococcus*, two lineages of marine A *Synechococcus* and one lineage of marine B *Synechococcus* exhibit closely spaced divergence points with low bootstrap support. This pattern is consistent with a near-simultaneous diversification of marine lineages with divinyl chlorophyll *b* and phycobilisomes as photosynthetic antennae. Inferences from 16S ribosomal RNA sequences including data for 18 marine picophytoplankton clade members were congruent with results of *psbB* and *petB* and *D* sequence analyses focusing on five strains of *Prochlorococcus* and one strain of marine A *Synechococcus.* Third codon position and intergenic region nucleotide frequencies vary widely among members of the marine picophytoplankton group, suggesting that substitution biases differ among the lineages. Nonetheless, standard phylogenetic methods and newer algorithms insensitive to such biases did not recover different branching patterns within the group, and failed to cluster *Prochlorococcus* with chloroplasts or other chlorophyll *b*-containing prokaryotes. *Prochlorococcus* isolated from surface waters of stratified, oligotrophic ocean provinces predominate in a lineage exhibiting low G + C nucleotide frequencies at highly variable positions.

Key words: *Prochlorococcus* — *Synechococcus* — Cyanobacteria — Picophytoplankton — Photosynthetic picoplankton — Prochlorophyte — Molecular evolution — Gene clusters

Introduction

Prochlorococcus and *Synechococcus*, the dominant picophytoplankton in the world's open oceans, are unicellular cyanobacteria with different photosynthetic pigments (Li and Wood 1988; Olson et al. 1990; Campbell et al. 1994). *Prochlorococcus* uniquely possesses divinyl chlorophylls *a* and *b* (8-vinyl chlorophylls *a* and *b*, chl a_2 and chl b_2), and lacks phycobilisomes observable by transmission electron microscopy (Chisholm et al. 1988, 1992). The *Prochlorococcus* chlorophyll *a/b*-binding antenna protein is similar to CP43′, a Chl *a*-binding protein induced by iron stress in other cyanobacteria (LaRoche et al. 1996). *Synechococcus* contains monovinyl chlorophyll *a* (chl a_1), similar to plastids and most cyanobacteria, and light harvesting phycobilisomes with numerous polypeptide subunits and covalently bound bilin chromophores. *Synechococcus* phycobilisomes are homologous to structures in rhodophyte plastids, cyanelles, and most cyanobacteria (Bryant 1991).

* *Present Address:* (E.U.), Department of Microbiology, Nash Hall, Oregon State University, Corvallis, OR 97331, USA; (D.L.D.), Biochemistry, Microbiology, and Molecular Biology Department, 5735 Hitchner Hall, University of Maine, Orono, ME 04469, USA
Correspondence to: E. Urbach

Table 1. Pigmentation of *Prochlorococcus* and *Synechococcus* strains

Strain	chl a_1	chl a_2	chl b_1	chl b_2	b/a_2	PBS	PE	PUB	PEB	Reference
Prochlorococcus										
CCMP1375	−	+	+	+	0.4–2.4	−	+	+	+	a,b
CCMP1378	−	+	−	+	0.05–0.15	−	−	−	−	a,b
GP2	−	+	?	+	0.17–0.18	−	?	?	?	d
MIT9303	−	+	−	+	0.7–1.3	−	?	?	?	c
NATL2	−	+	?	+	?	−	−	−	−	b
SB	−	+	?	+	0.13	−	?	?	?	d
TATL1	−	+	?	+	?	−	−	−	−	b
TATL2	−	+	?	+	?	−	+	?	?	b
Synechococcus										
PCC6301	+	−	−	−	−	+	−	−	−	e
PCC7942	+	−	−	−	−	+	−	−	−	e
PCC6307	+	−	−	−	−	+	−	−	−	e
WH7805	+	−	−	−	−	+	+	−	+	e
WH8101	+	−	−	−	−	+	−	−	−	e
WH8103	+	−	−	−	−	+	+	+	+	e

Phenotypically uncharacterized *Prochlorococcus* strains not included
[a] Moore et al. 1995
[b] Hess et al. 1996
[c] Unpublished data of L. Moore, G. Rocap and S.W. Chisholm
[d] Shimada et al. 1996
[e] Waterbury and Rippka 1989

PBS: Phycobilisomes
PE: Phycoerythrin
PUB: Phycourobilin
PEB: Phycoerythrobilin
?: Not determined

Prochlorococcus marinus strains contain chls a_2 and b_2, carotenoids and a chl *c*-like pigment, and lack phycobilisomes (Chisholm et al. 1992). Chl b_1 and/or a novel phycoerythrin may also be present (Table 1) (Morel et al. 1993; Partensky et al. 1993; Moore et al. 1995; Hess et al. 1996). Type strain CCMP1375 was collected near the bottom of the euphotic zone in the Sargasso Sea, and numerous other strains have been established from different depths and diverse locations (Table 2). Studies investigating the physiological capabilities of a small number of these strains have identified distinct high and low chl b/a_2 ratio phenotypes, supposed to be genetic adaptations to low (deepwater) and high (surface) illumination environments, respectively (Morel et al. 1993; Partensky et al. 1993; Moore et al. 1995; Shimada et al. 1995). Included in the present study are two members of the high chl b/a_2 ratio group, CCMP1375 (chl b/a_2 ratio 0.4–2.4, collected from 120 m) and MIT9303 (chl b/a_2 ratio 0.7–1.3, collected from 100 m), and three low chl b/a_2 ratio strains, CCMP1378 (chl b/a_2 ratio 0.05–0.15, collected from 5 m), SB (chl b/a_2 ratio 0.13, collected from 40 m) and GP2 (b/a_2 ratio 0.17–0.18, collected from 150 m) (Moore et al. 1995; Shimada 1995, unpublished data of L. Moore, G. Rocap, and S.W. Chisholm). *Prochlorococcus marinus* strains are phenotypically diverse and some may be described as new species in the future. We therefore adopt the nomenclature *Prochlorococcus* to refer to chl a_2 and b_2-containing marine picocyanobacteria, rather than designating all strains *P. marinus.*

The genus *Synechococcus* is a provisional taxon including unicellular coccoid to rod-shaped cyanobacteria lacking structured sheaths (Waterbury and Rippka 1989). The designation is typically restricted to strains with chl a_1 and phycobilisome photosynthetic antennae, although the current diagnosis does not exclude *Prochlorococcus.* The genus is divided into six ''strain clusters'' which are intended as provisional genera. Strains from four of these clusters are considered in the present study: marine A (WH7805 and WH8103), marine B (WH8101), *Cyanobium* (PCC6307), and *Synechococcus* (PCC6301 and PCC7942) (Table 2). Strain clusters are distinguished on the basis of morphological and physiological properties, characteristics of the strain isolation sites, G + C nucleotide content, and the presence or absence of phycoerythrin, a phycobiliprotein present in phycobilisomes (Table 1). The marine A cluster is additionally diverse in that ratios of phycourobilin to phycoerythrobilin chromophores differ among phycoerythrins of different strains (Wood 1985; Waterbury et al. 1986).

Among cyanobacteria as a whole, phylogenetic analyses have shown that photosynthetic pigments and the unicellular character of *Prochlorococcus* and *Synechococcus* are unreliable taxonomic indicators. 16S ribosomal RNA (rRNA) sequence analyses have shown that *Prochlorococcus*, green chloroplasts and the chl b_1-containing prokaryotes *Prochloron* and *Prochlorothrix hollandica* are phylogenetically dispersed among the cyanobacteria (Urbach et al. 1992). Comparisons of *rpoC* (Palenik and Haselkorn 1992), *atpBE* (Lockhart et al. 1992a), *psbA* (Kishino et al. 1990; Hess et al. 1995), and 16S rRNA (Turner et al. 1989) gene sequences confirm the polyphyletic origins of these chl *b*-containing taxa and organelles, once supposed to descend from a common ancestor (Lewin 1977). 16S rRNA sequence analyses similarly indicate that members of the unicel-

Table 2. *Prochlorococcus* and *Synechococcus* strains included in this study

Strain	Isolation history					Culture history	Ref.
	Region	Site	Depth	Date	Researcher		
Prochlorococcus							
CCMP1375[T]	Sargasso	28°59′N,64°21′W	120m	5/30/88	B. Palenik	(a.k.a. SS120) cloned from SARG	a
CCMP1378	Mediterranean	43°12′N,61°52′E	5m	1/89	D. Vaulot/ F. Partensky	(a.k.a. Med4) cloned from MED	a
GP2	West Pacific	8°32.5′N,136°31.8′E	150m	9/10/92	A. Shimada	Clonal culture	c
MIT9107	South Pacific	14°60′S,134°60′W	25m	8/8/91	J. Dusenberry/ M. DuRand	Primary culture	
MIT9303	Sargasso	34°45′N,66°11′W	100m	7/15/93	L. Moore	Primary culture	
NATL2	North Atlantic	38°59′N,49°33′W	10m	4/90	F. Partensky	Primary culture*	b
NATL2A	North Atlantic	38°59′N,49°33′W	10m	4/90	F. Partensky	Primary culture*	
PAC1	Tropical Pacific	22°45′N,158°00′W	100m	4/92	L. Campbell	Primary culture	b
SB	West Pacific	35°00.9′N,138°35.8′E	40m	10/21/92	A. Shimada	Clonal culture	d
TATL1	Tropical Atlantic	21°02′N,31°08′W	20m	10/91	F. Partensky	Primary culture	b
TATL2	Tropical Atlantic	20°25′N,31°08′W	30m	10/91	F. Partensky	Primary culture	b
Synechococcus							
Synechococcus-cluster							
PCC6301[R]	Texas freshwater			1952	W.A. Kratz	Clonal, axenic culture	e
PCC7942	California freshwater			1973	K. Floyd	Clonal, axenic culture	e
Cyanobium-cluster							
PCC6307[R]	Wisconsin freshwater			1949	G.C. Gerloff	Clonal, axenic culture	e
marine A-cluster							
WH7805	Sargasso	33°44.8′N,67°30′W	Surface	6/30/78	L. Brand	Clonal, axenic culture	f
WH8103[R]	Sargasso	28°30′N,67°23.5′W	Surface	3/17/81	J. Waterbury	Clonal, axenic culture	f
marine B-cluster							
WH8101	Woods Hole Harbor		Surface	1981	F. Valois	Clonal, axenic culture	f

[a] Chisholm et al. 1992
[b] Parpais et al. 1996
[c] Shimada et al. 1996
[d] Shimada et al. 1995
[e] Waterbury and Rippka 1989
[f] Waterbury et al. 1986
T: Type culture
R: Reference strain
a.k.a.: Also know as
* Isolated from the same enrichment culture

lular cyanobacterial Order Chroococcales are polyphyletic and dispersed among the cyanobacteria (Giovannoni et al. 1988).

According to 16S rRNA sequence phylogenies, one *Prochlorococcus* isolate (SSW5 [descended from SARG]) forms a shallow phylogenetic cluster with two marine A *Synechococcus* strains (WH7805 and WH8103) (Urbach et al. 1992). Similarly, *rpoC* sequence analyses link two *Prochlorococcus* isolates (DV1 [descended from MED] and LG [descended from SARG]) to three marine A *Synechococcus* strains (WH7803, WH7805 and WH8103) (Swift and Palenik 1992; Palenik 1994). In the present work, we employ 17 new 16S rRNA sequences to address questions of whether additional strains of both taxa belong to the *Prochlorococcus*/marine *Synechococcus* group, and whether ancestor–descendant relationships can be inferred for strains from different oceanic provinces, depths of isolation, and pigment phenotypes. To assess the reliability of phylogenetic inferences from rRNA genes, phylogenetic relationships were inferred for five strains of *Prochlorococcus* and one strain of marine *Synechococcus* using *psbB* and *petB* and *D* sequences. *psbB* encodes the chl *a*-binding antenna protein CP47, and *petB* and *D* encode the *b*-type cytochrome and subunit IV polypeptides of the photosynthetic b_6/f complex, respectively (Vermaas et al. 1978; Widger and Cramer 1991). These protein-encoding genes were chosen because they were likely to show more variability than 16S rRNA sequences, and also because they are single-copy, photosynthetic organism-specific loci, easily amplified with degenerate PCR primers and not implicated in gene exchange. Phylogenetic relationships inferred by established methods were compared to results from new methods insensitive to nucleotide substitution biases.

Methods

Cell Culture and DNA Isolation

Prochlorococcus were grown in modified K/10-Cu media: strains NATL2, TATL1, TATL2, and PAC1 were grown as described by Scanlan et al. (1996) and the remaining strains as in Chisholm et al. (1992). DNA was prepared from dense cultures as described (Urbach et al. 1992; Scanlan et al. 1996), or by a modification of the protocol of Li et al. (1991). In the modified Li protocol, 10^6 to 10^8 *Prochlorococcus* were concentrated by spin filtration (0.2 μm-pore Ultrafree-MC, Millipore)

at 2000 *g* and washed twice with cell suspension buffer (0.5 M NaCl, 10 mM Tris, pH 8.0, 10 mM EDTA). Cells were resuspended in a final volume of 100-μl cell suspension buffer and lysed by addition of 12 μl 0.5 M DTT and 6 μl 10 M NaOH, with 10 min at 65°C. After addition of 120 μl neutralization buffer (90 mM Tris, pH 8.0, plus 0.5 mol/l HCl), DNA was precipitated by the addition of 1 μl 20 mg/ml glycogen (Boehringer) and 0.6 ml cold 100% ethanol, resuspended in TE (10 mM Tris, pH 8.0, 1 mM EDTA) and stored at −20°C.

Synechococcus strains were grown in S/N Medium according to Waterbury et al. (1986). DNA and RNA were prepared by CsCl density gradient centrifugation (Sambrook et al. 1989).

PCR Amplification, Cloning and Sequencing

16S rRNA genes from *Prochlorococcus* strains were amplified using oxygenic phototroph-specific primers and protocols as described (Urbach et al. 1992). Sequences for strains CCMP1375, CCMP1378, NATL2A, MIT9107, and MIT9303 were obtained by direct sequencing of PCR products, either by a solid-phase protocol (Hultman et al. 1989) or using an ABI automated DNA sequencer. Sequences for strains NATL2, PAC1, TATL1, and TATL2 were obtained from cloned PCR products. Products from at least two independent PCR reactions were cloned into the TA vector (Invitrogen) and then subcloned into M13 for single-stranded sequencing using Sequenase (USB). At least two independent clones were completely sequenced for each strain, using forward and reverse PCR primers and internal 16S rRNA primers. In the case of strain TATL1, initial partial sequencing (ca 300 bp from the forward primer) identified nine clones with the TATL1A sequence and one with the TATL1B sequence. The TATL1B clone and two TATL1A clones were sequenced completely.

16S rRNA sequences for *Synechococcus* strains WH8103 and WH7805 were determined using reverse transcriptase (Lane et al. 1985). Sequences for strains WH8101 and PCC6307 were determined by automated DNA sequencing of PCR products amplified using bacterial-specific primers (Lane 1991).

psbB and *petB/D* sequences were amplified using primer pairs PPSBB1353 (5′GTIGCIGGIACIATGTGGTA)/PPSBB1928R (5′GCRTGICCRAAIGTRAACCA) and PPETBD314 (5′ATGATGGTIYTIATGATGAT)/PPETBD1160R (5′CCRTARTARTTRTGICCCAT), respectively (I denotes inosine). *psbB* amplification reactions used 2.25 mM $MgCl_2$ and the temperature profile: 2 min 94°C; 1 min 94°C, 1 min 51°C, 1 min 72°C (5 cycles); 1 min 94°C, 1 min 56°C, 1 min 72°C (10 cycles); 1 min 94°C, 1 min 62°C, 1 min 72°C (25 cycles); 10 min 72°C. *petB/D* amplifications used 1.5 mM $MgCl_2$ and temperature profile: 2 min 94°C; 1 min 94°C, 1 min 45°C, 1 min 72°C (5 cycles); 1 min 94°C, 1 min 50°C, 1 min 72°C (10 cycles); 1 min 94°C, 1 min 57°C, 1 min 72°C (25 cycles); 10 min 72°C. 100-μl PCR reactions contained $MgCl_2$ as indicated, plus 1 × Mg-free Taq polymerase buffer (Promega), 200 μM each dNTP, 100 nM each primer, and 5 units Taq polymerase (Promega). PCR amplicons from three or more replicate reactions were pooled, separated by electrophoresis in 1% agarose gels, and purified from gel slices using GeneClean (Bio 101). Sequences were determined using amplification primers and internal primers PPSBB1527 (5′ITTYTAYGAYTAYGTIGG), PPSBB 1508R (5′CCIAGRTARTGRTARAAIGC), PPSBB1898R (5′CGAAAIACICCRTC), and PPETBD532R (5′CCIACRCTYTCICCICC). Most sequences were determined using the solid-phase protocol (Hultman et al. 1989). Sequences for strain MIT9303 were obtained by automated sequencing of pooled, cloned PCR products as described (Urbach et al. 1997).

Sequence Alignment and Analysis

Sequences were aligned by hand using the Genetic Data Environment (provided by Stephen Smith) according to conserved regions of rRNA primary and secondary structures and amino acid translations (Lang and Haselkorn 1989; Widger and Cramer 1991). Base compositions were calculated using GCG (Devereux et al. 1990). Phylogenetic calculations were performed using PHYLIP version 3.5c (Felsenstein 1993), PAUP* (Swofford personal communication, neighbor-joining, DNA parsimony, LogDet and DNA maximum likelihood) and GG95 (Galtier and Gouy 1995). Bootstrap analyses examined 1000 resampled data sets.

16S rRNA gene sequence analyses included all unambiguously aligned and determined nucleotide sites. Parallel analyses omitted hypervariable regions (Gray et al. 1984; Dams et al. 1988). Neighbor-joining calculations (Saitou and Nei 1987) employed Kimura two-parameter genetic distances (Kimura 1980). DNA parsimony and associated bootstrap calculations employed PAUP*'s heuristic search and MULPARS options. The DNA maximum likelihood calculation used nucleotide frequencies estimated from the data, transition:transversion ratio 2:1, and was initiated with a tree inferred by parsimony. LogDet calculations used invariant nucleotide fractions 0.01–0.5. Bootstrap proportions reported are from calculations assuming an invariant fraction of 0.1. GG95 calculations used nucleotide frequencies and transition:transversion ratios estimated from the data. A parallel analysis examined relationships among *Prochlorococcus* CCMP1375, CCMP1378, MIT9107, and MIT9303, *Synechococcus* WH8103, *Synechococystis* PCC6803 and *Prochlorothrix hollandica*, specifying equilibrium G + C frequencies equal to third codon position values in *psbB* and *petB/D* sequences.

psbB and *petB/D* phylogenetic relationships were inferred from concatenated sequences, and also from the individual genes. Amino acid sequence phylogenies for *psbB* and *petB/D* were inferred using PHYLIP's PROTPARS, PROTML, and PROTDIST programs, with the Dayhoff substitution matrix specified for PROTDIST. GG95 genetic distances were calculated from first two codon position nucleotide data using transition:transversion ratios estimated from the data and equilibrium G + C frequencies set to third codon position values. Distance trees were inferred by neighbor-joining.

Tests for Recombination Among Gene Sequences

The possibility of intragenic recombination among 16S rRNA gene sequences was investigated by comparison of phylogenetic trees inferred for each of seven hypervariable regions (Gray et al. 1984; Dams et al. 1988) by parsimony. Regions considered in the analyses were nucleotide positions 30–48, 227–287, 380–438, 615–644, 783–812, 894–929, and 1015–1068 (*E. coli* numbering). Recombination among *psbB* and *petB/D* sequences was assessed by the method of Sawyer (1989) using the program MULTICOMP (Reeves et al. 1994).

Results and Discussion

DNA Sequence Characteristics

Thirteen 16S rRNA genes were sequenced from 12 cultured isolates of *Prochlorococcus*, and four sequences were obtained from four strains of *Synechococcus* (Table 3). *Prochlorococcus* TATL1, a nonclonal isolate, yielded sequences TATL1A and TATL1B for different clones in a 16S rRNA PCR product library. Since clonal *Prochlorococcus* cultures appear to contain only one 16S rRNA gene copy (unpublished Southern hybridization data of D.J. Scanlan), the TATL1 sequences probably derive from different lineages coexisting in the culture. The *Prochlorococcus* CCMP1375 sequence is identical to the

Table 3. Genbank accession numbers and sequencing template types for *Prochlorococcus* and *Synechococcus* sequences included in this study

	16S rRNA		*psbB*		*petB/D*	
Strain	Accession number	Template type	Accession number	Template type	Accession number	Template type
Prochlorococcus						
CCMP1375	X63140	PCR product	AF001481	PCR product	AF001487	PCR product
CCMP1378	AF001466	PCR product	AF001482	PCR product	AF001488	PCR product
GP2	AF001472	Cloned PCR product	—	—	—	—
MIT9107	AF001468	PCR product	AF001484	PCR product	AF001490	PCR product
MIT9303	AF001469	PCR product	AF001485	Pooled PCR product clones	AF001491	Pooled PCR product clones
NATL2	AF001470	Cloned PCR product	—	—	—	—
NATL2A	AF001467	PCR product	AF001483	PCR product	AF001489	PCR product
PAC1	AF001471	Cloned PCR product	—	—	—	—
SB	AF001473	Cloned PCR product	—	—	—	—
TATL1A	AF001474	Cloned PCR product	—	—	—	—
TATL1B	AF001475	Cloned PCR product	—	—	—	—
TATL2	AF001476	Cloned PCR product	—	—	—	—
Synechococcus						
PCC6301	X03538	Genomic clone	—	—	—	—
PCC6307	AF001477	PCR product	—	—	—	—
PCC7942	—	—	Z14087	Genomic clone	X64768	Genomic clone
WH7805	AF001478	rRNA	—	—	—	—
WH8103	AF001479	rRNA	AF001486	PCR product	AF001492	PCR product
WH8101	AF001480	PCR product	—	—	—	—

sequence from strain SSW5, subcultured from the same primary culture (SARG) (Urbach et al. 1992).

psbB and *petB* and *D* sequences were determined for five strains of *Prochlorococcus* (CCMP1375, CCMP1378, MIT9107, MIT9303, and NATL2A) and one strain of marine A *Synechococcus* (WH8103) (Table 3). These single-copy genes are unique to photosynthetic organisms (Vermaas and Ikeuchi 1991) and, therefore, easily amplified from *Prochlorococcus* cultures containing heterotrophic bacteria. *petB* and *D* are neighboring cistrons with a consistent tandem orientation in oxygenic photosynthetic prokaryotes and organelles (Vermaas and Ikeuchi 1991; Greer and Golden 1992). Regions of conserved amino acid sequence provide PCR priming sites for amplification of a fragment containing the 3′ end of *petB*, the 5′ end of *petD*, and the highly variable intergenic region between them (the amplified locus will be referred to as ''*petB/D*''). The *psbB* and *petB/D* loci are not close to each other on cyanobacterial chromosomes (Vermaas and Ikeuchi 1991).

The six *psbB* and *petB/D* sequences appear to encode functional proteins. Their amino acid translations align with previously published sequences (Lang and Haselkorn 1989; Widger and Cramer 1991). They contain conserved heme-binding amino acid residues in *petB* (Widger and Cramer 1991; Hope 1993), as well as *psbB* residues implicated in PSII assembly, photoautotrophic growth, and susceptibility to photoinhibition (Haag et al. 1993). At both loci, nucleotide sequence differences occur most frequently at third codon positions (Table 4).

Comparison of *petB/D* and *psbB* sequences for the five *Prochlorococcus* strains and *Synechococcus* WH8103 reveals that third codon positions and intergenic regions are saturated with nucleotide substitutions and (in the intergenic region) insertions and deletions. Third codon position identities range from 30.1% (MIT9107 and WH8103) to 56.6% (NATL2A and MIT9107) for *psbB* and from 23.7% (MIT9107 and WH8103) to 60.3% (MIT9107 and CCMP1375) for *petB/D*, too small to give reliable phylogenetic information (Table 4). *Prochlorococcus* and marine A *Synechococcus petB/D* intergenic regions range from 34 to 90 basepairs in length, with identifiable sequence similarity for only a 23 basepair region shared by one pair of isolates (CCMP1378 and MIT9107, Figure 1). Intergenic region sequences were, therefore, omitted from phylogenetic analyses, and third codon positions were used only to estimate equilibrium G + C frequencies for GG95 calculations.

Phylogenetic Analysis

In the 16S rRNA sequence phylogeny, strains of *Prochlorococcus* and marine A, marine B and *Cyanobium*-Cluster *Synechococcus* form a monophyletic group (Fig. 2). Included in this lineage are three sequences (SAR6, SAR7, and SAR139) recovered by direct cloning from Sargasso Sea picoplankton (Giovannoni et al. 1990; Britschgi and Giovannoni 1991), which typically contains both *Prochlorococcus* and marine A *Synechococcus* cells (Olson et al. 1990). *Cyanobium*-Cluster strain *Synechococcus* PCC6307, a freshwater isolate, branches

Table 4. Fractional identity (X 1000) for sequences of *psbB* (below the diagonal) and *petB/D* (above) for cultured strains of *Prochlorococcus synechococcus*, and other members of the cyanobacterial radiation

	1	2	3	4	5	6	7	8	9	10
		843	813	782	739	694	696	684	704	727
1 *Prochlorococcus* CCMP1378		**970**	**947**	**936**	**932**	**902**	**886**	**864**	**875**	**845**
		588	*542*	*473*	*351*	*275*	*313*	*321*	*359*	*489*
	821		813	830	742	684	704	678	686	711
2 *Prochlorococcus* MIT9107	**967**		**943**	**943**	**924**	**905**	**886**	**848**	**879**	**830**
	524		*550*	*603*	*374*	*237*	*336*	*336*	*298*	*473*
	787	799		792	754	716	722	668	719	709
3 *Prochlorococcus* NATL2A	**914**	**914**		**928**	**932**	**902**	**879**	**845**	**879**	**845**
	530	*566*		*519*	*397*	*344*	*405*	*313*	*397*	*435*
	741	739	765		759	719	694	643	696	681
4 *Prochlorococcus* CCMP1375	**881**	**884**	**893**		**932**	**902**	**875**	**833**	**879**	**818**
	458	*446*	*506*		*412*	*351*	*328*	*260*	*328*	*405*
	685	691	697	705		800	724	719	742	691
5 *Prochlorococcus* MIT9303	**833**	**833**	**836**	**836**		**955**	**902**	**864**	**890**	**860**
	386	*404*	*416*	*440*		*489*	*366*	*427*	*443*	*351*
	663	651	665	659	745		757	739	762	661
6 *Synechococcus* WH8103	**830**	**824**	**821**	**804**	**869**		**902**	**867**	**886**	**822**
	325	*301*	*349*	*367*	*494*		*466*	*481*	*511*	*336*
	629	616	598	627	653	667		737	762	676
7 *Synechcoccus* PCC7942	**777**	**768**	**741**	**756**	**768**	**786**		**875**	**902**	**848**
	331	*307*	*307*	*367*	*422*	*428*		*458*	*481*	*328*
	588	608	612	616	618	659	713		767	661
8 *Synechocystis* PCC6803	**744**	**750**	**747**	**750**	**747**	**783**	**839**		**871**	**848**
	271	*319*	*337*	*343*	*355*	*410*	*458*		*557*	*282*
	608	627	614	627	675	685	715	767		701
9 *Prochlorothrix hollandica*	**753**	**756**	**756**	**744**	**789**	**798**	**830**	**887**		**879**
	313	*367*	*325*	*392*	*446*	*458*	*482*	*524*		*344*
	604	618	600	612	568	574	592	633	610	
10 *Zea maize* chloroplast	**729**	**726**	**714**	**723**	**696**	**744**	**741**	**783**	**771**	
	349	*398*	*367*	*386*	*307*	*229*	*289*	*331*	*283*	

Plain type: all codon positions, n = 503 (*psbB*), n = 375 (*petB/D*)
Bold type: first two codon positions, n = 336 (*psbB*), n = 244 (*petB/D*)
Italic type: third codon positions, n = 167 (*psbB*), n = 131 (*petB/D*)

at the base of this lineage, with marine *Synechococcus* and *Prochlorococcus* strains forming a more recently evolved, marine picophytoplankton clade. Bootstrap resampling strongly supports placement of the freshwater strain at the base of this group in neighbor-joining analyses. *Synechococcus* PCC6301, a member of the freshwater *Synechococcus*-Cluster, is outside the group (though linked to marine picophytoplankton and *Cyanobium* by some analyses, the association receives low bootstrap support). Consistent with previous results, the phylogeny implies that all analyzed *Prochlorococcus* strains share more recent common ancestry with the phycobilisome-containing *Cyanobium*-Cluster (PCC6307) than with other chl *b*-containing prokaryotes (*Prochloron* sp. and *Prochlorothrix hollandica*) or organelles (*Marchantia polymorpha* and *Cyanophora paradoxa* organelles) (Urbach et al. 1992). This phylogenetic pattern is consistent in trees including additional members of the cyanobacterial clade (not shown).

Within the shallow clade containing *Prochlorococcus* and marine *Synechococcus* strains (the marine picophytoplankton clade), it is not possible to distinguish well-supported, unitary lineages for phycobilisome- and chl b_2-containing isolates, or for marine A and B *Synechococcus*. Six deeply branching lineages: a group containing SAR6 and most *Prochlorococcus* isolates, SAR7, *Prochlorococcus* MIT9303, *Synechococcus* WH7805, a group containing *Synechococcus* WH8103 plus SAR139, and *Synechococcus* WH8101 receive poor bootstrap support for their relative branching order (Lineages 1–6, Fig. 2). Different branching orders are inferred for these lineages by different phylogenetic methods, but in no instance is any grouping of these lineages supported by >70% of 1000 bootstrap analyses. Excluding the cloned sequences, which are potentially chimeras combining sequences from different lineages, does not result in greater support for groupings of the remaining branches (not shown).

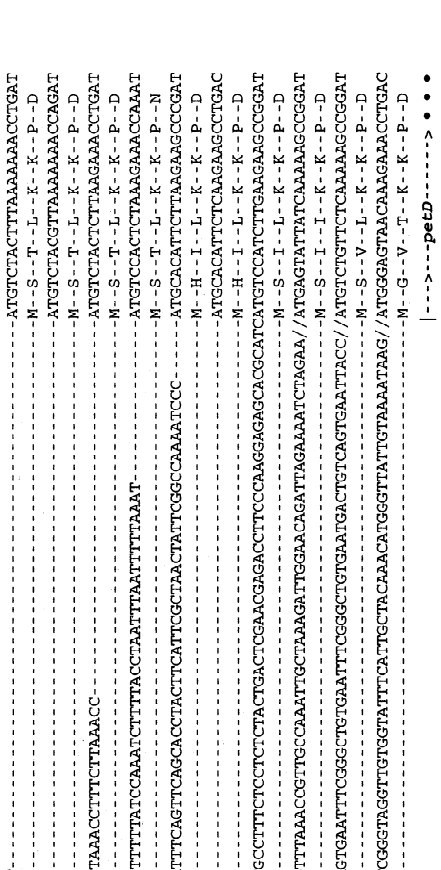

Fig. 1. Comparison of *petB/D* intergenic region sequences for cultured strains of *Prochlorococcus*, *Synechococcus*, and other members of the cyanobacterial radiation. Coding regions are aligned according to amino acid translations, and an alignment gap is inserted into the *Prochlorococcus* MIT9107 intergenic region to accentuate its similarity to the *Prochlorococcus* CCMP1378 sequence. Otherwise, intergenic regions are arbitrarily represented as continuous with *petB*. Intergenic region sequences from *Synechocystis* PCC6803, *Prochlorothrix hollandica*, and *Zea maize* chloroplast are truncated.

Phylogenetic relationships inferred from concatenated sequences of *psbB* and *petB/D* are consistent with relationships inferred from 16S rRNA sequences (Fig. 3, Fig. 2). Individual *psbB* and *petB/D* data sets gave congruent results as well (not shown). As in the 16S rRNA phylogeny, Lineage 1 is well supported by bootstrap proportions. In addition, a phylogenetic association between *Prochlorococcus* MIT9303 and *Synechococcus* WH8103 receives greater support in the protein phylogeny than in 16S rRNA trees.

Evidence Against Gene Transfer

Phylogenetic analyses of *Prochlorococcus* and *Synechococcus* sequences suggest that gene transfers among these lineages are infrequent. The congruence among *psbB*, *petB/D*, and 16S rRNA phylogenies suggests that gene transfers among marine picophytoplankton lineages are infrequent at these loci. Recent *rbcL* sequence phylogenies imply an interphylum gene transfer, grouping *Prochlorococcus* and marine *Synechococcus* with a diverse assemblage of proteobacteria (Shimada et al. 1995; Delwiche and Palmer 1996). However, well-supported elements of the *rbcL* and 16S rRNA gene trees are consistent with the marine picophytoplankton forming a deep branch within the cyanobacteria (Delwiche and Palmer 1996) (Fig. 2). Thus, several proteobacteria appear to have acquired a cyanobacterial *rbcL*, but marine picophytoplankton gene complements were apparently unaffected by the event.

Rapid Diversification of Marine Picoplankton

Low bootstrap confidence in the branching order among picophytoplankton lineages (Fig. 2) suggests that *Prochlorococcus* and *Synechococcus* lineages are unresolved due to rapid, contemporaneous diversification of both taxa. Other potential sources of low bootstrap confidence, insufficient sequence information, and conflicting patterns of sequence similarity within the data sets, may be ruled out by the following: The possibility that the data sets contain insufficient sequence information to resolve the ambiguous branchings is contradicted by strong bootstrap support for more recently evolved clades (e.g., Lineages 1 and 5, Fig. 2; Lineage 1, Fig. 3), and for deeper branchings (e.g., above and below *Synechococcus* PCC6307, Fig. 2; the marine picophytoplankton clade, Fig. 3). Conflicting patterns of sequence similarity caused by nucleotide substitution biases (Lockhart et al. 1992b) are contradicted by phylogenetic calculations insensitive to such biases (LogDet and GG95), which fail to infer strong support for the ambiguous branchings (Fig. 2 and 3). Conflicting phylogenetic patterns attributable to mosaic evolution (Maynard Smith et al. 1991; Sneath 1993; Cilia et al. 1996) are ruled out for 16S rRNA sequences by comparison of

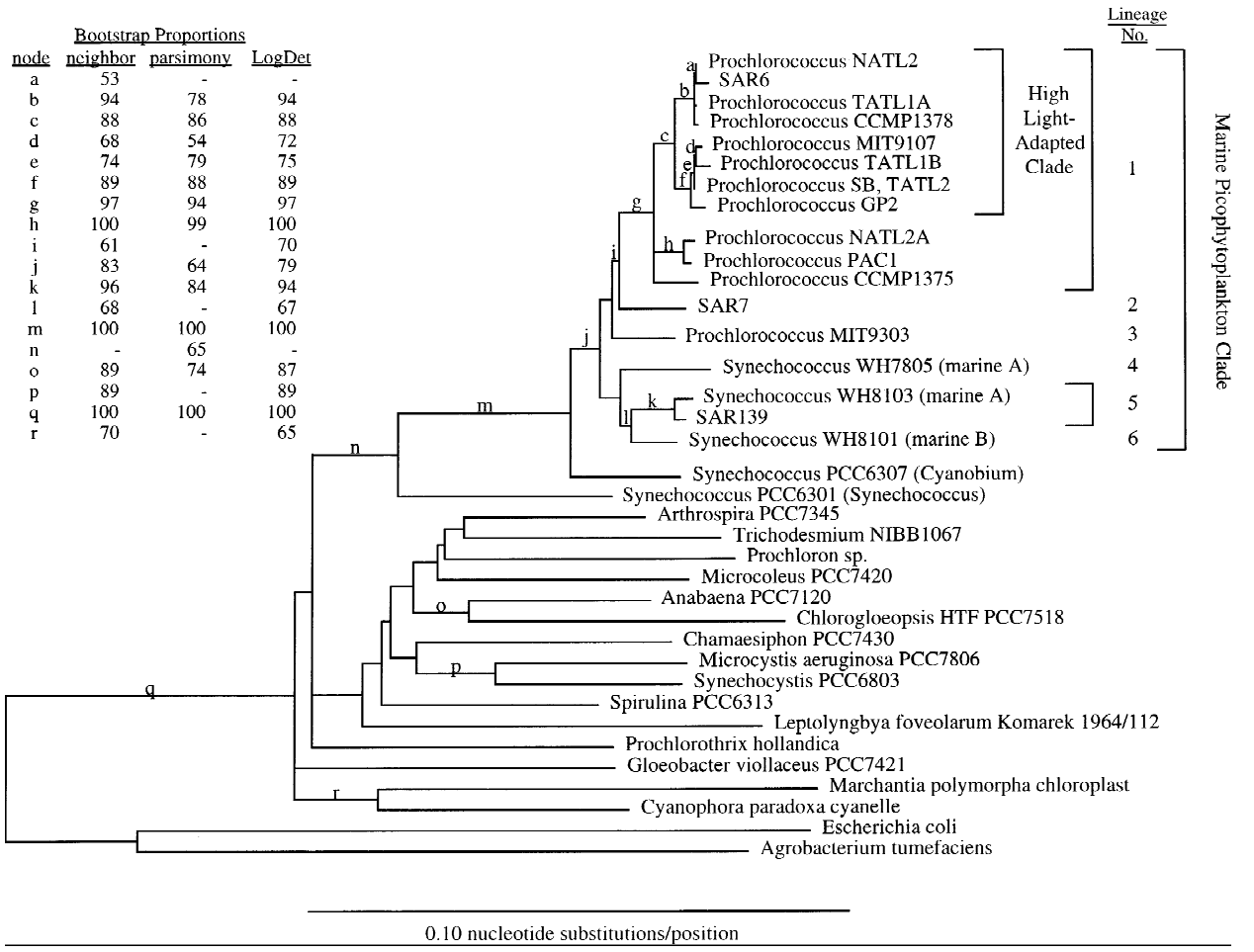

Fig. 2. Phylogenetic relationships among cultured strains of *Prochlorococcus, Synechococcus,* cloned sequences from Sargasso Sea bacterioplankton and other members of the cyanobacterial lineage inferred from 16S rRNA gene sequences. *Escherichia coli* and *Agrobacterium tumefaciens* are outgroups. The phylogenetic framework was inferred by neighbor joining, considering all 958 unambiguously aligned and determined nucleotides between *E. coli* positions 146 and 1312. Frameworks inferred by parsimony, LogDet, GG95 and maximum likelihood were congruent at nodes which received >80% bootstrap support in neighbor-joining analyses. Bootstrap values below 50% and for nodes with zero branch length are omitted.

phylogenetic branching patterns inferred for each of seven hypervariable regions, and for *psbB* and *petB/D* sequences by statistical analyses according to the method of Sawyer (1989). Neither set of analyses reveals localized regions of anomalous phylogenetic pattern. Thus, neither insufficient sequence information nor conflicting phylogenetic patterns is the principal source of low bootstrap confidence for deep branchings in the marine picophytoplankton clade. While these analyses do not rule out other potential confounding factors, such as differential retention of multiple rRNA operons in different lineages, they are consistent with the idea that *Prochlorococcus* and marine *Synechococcus* lineages are poorly resolved because they arose in rapid succession.

Coexistence with diverse populations of cyanophage viruses contributes to the genetic diversity of local marine *Synechococcus* populations (Waterbury and Valois 1993). Pressure to develop resistance to diverse cyanophage may also have contributed to picophytoplankton diversification on a larger evolutionary scale.

Relationships Among Synechococcus *Strains*

On the basis of evolutionary relationships inferred for representative strains, it appears that most *Synechococcus* strain clusters can be distinguished phylogenetically (Fig. 2). The freshwater *Cyanobium*-Cluster PCC6307 is outside the marine picophytoplankton clade, but is clearly descended from a shared ancestor. Freshwater *Synechococcus*-Cluster isolate PCC6301 branches outside the group, within the general cyanobacterial radiation. Within the marine picophytoplankton clade are both obligately marine (marine A, WH7805 and WH8103) and facultatively marine (marine B, WH8101) strains, with different phycobiliprotein phenotypes falling into different, unresolved lineages. High phycourobilin marine A WH8103, no phycourobilin marine A WH7805, and no phycoerythrin marine B WH8101 form independent lineages within the clade. Sequence and pigment data for additional strains will be required to determine whether strains with similar phenotypes form distinct

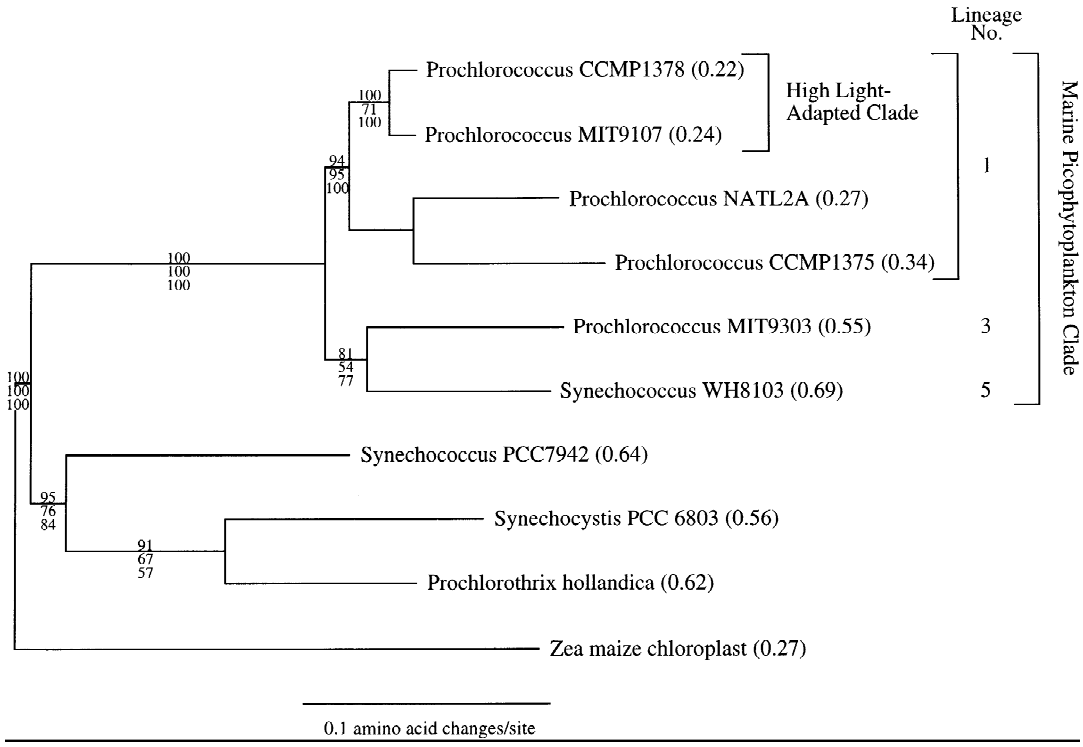

Fig. 3. Phylogenetic relationships among cultured strains of *Prochlorococcus, Synechococcus* and other members of the cyanobacterial lineage inferred from concatenated sequences of *psbB* and *petB/D.* The phylogenetic framework was inferred from amino acid translations by protein maximum likelihood. Bootstrap proportions are from protein distance, protein parsimony and LogDet calculations, from top to bottom, respectively. A branching order in accord with nodes receiving >80% bootstrap support in neighbor-joining analyses was inferred by GG95. Bootstrap proportions below 50% are omitted. Numbers in parentheses are third codon position G + C frequencies.

lineages. However, phylogenetic relationships are consistent with provisional genus-level distinctions between *Synechococcus, Cyanobium,* and the marine strain clusters. Among the marine picoplankton strains, high and low phycourobilin marine A *Synechococcus* may merit separate taxonomic status on par with marine B *Synechococcus.*

Relationships Among Prochlorococcus *Strains*

The 16S rRNA gene phylogeny reveals a recently arising ''high light-adapted clade'' containing eight *Prochlorococcus* sequences from isolates collected at or above 40 m, and one strain (GP2) from 150 m (Fig. 2, Table 1). More deeply branching *Prochlorococcus* lineages include one strain from 30 m (NATL2A) and three from 100 m or greater (PAC1, CCMP1375, and MIT9303). Available pigment data correlate with the phylogeny, in accord with the hypothesis of Goericke and Repeta (1993) and Moore et al. (1995): low chl b/a_2 ratio strains CCMP1378, SB and GP2 belong to the high light-adapted clade, while the high chl b/a_2 ratio strains CCMP1375 and MIT9303 are outside the group. It therefore appears that the recently arisen clade may be best adapted to conditions near the surface of the euphotic zone. A similar relationship between phylogeny and pigment phenotype has been observed for other isolates of *Prochlorococcus* (unpublished data of L. Moore, G. Rocap, and S.W. Chisholm). It is surprising that such a good correlation can be detected between the depth of isolation, pigment phenotype, and phylogenetic branching pattern, as physical mixing processes should tend to move cells from the depths at which they thrive. Isolation protocols that mimic *in situ* light intensities may have selected for different phenotypes in isolates recovered from different depths.

The ability to synthesize phycoerythrin does not follow the phylogenetic pattern. Phycoerythrin-containing strain TATL2 is more closely related to strains that do not bind a *Prochlorococcus* phycoerythrin gene probe (NATL2, CCMP1378, and TATL1) (Hess et al. 1996) than it is to phycoerythrin-containing CCMP1375 (Fig. 2, Table 2). Phylogenetic relationships, therefore, suggest that, while low chl b/a_2 ratio *Prochlorococcus* strains appear similar due to descent from a common ancestor, phycoerythrin-containing strains appear similar due to differential retention (or transfer) of phycoerythrin genes among *Prochlorococcus* lineages. Pigment data for uncharacterized strains will shed additional light on patterns of pigment inheritance.

Closely related *Prochlorococcus* strains have been isolated from geographically distant sites: tropical Atlantic strain TATL2 and Pacific strain SB have identical 16S rRNA sequences, and are closely related (>99.1%

Table 5. Fractional identity of 16S rRNA sequences (X 1000)

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 *Prochlorococcus* NATL2													
2 *Prochlorococcus* TATL1A	999												
3 SAR6	998	997											
4 *Prochlorococcus* CCMP1378	999	998	997										
5 *Prochlorococcus* MIT9107	992	991	990	991									
6 *Prochlorococcus* TATL1B	991	990	989	990	997								
7 *Prochlorococcus* SB	992	991	990	991	998	997							
8 *Prochlorococcus* GP2	993	992	991	992	995	994	997						
9 *Prochlorococcus* NATL2A	985	984	983	986	981	980	983	983					
10 *Prochlorococcus* PAC1	986	985	984	987	982	981	984	984	997				
11 *Prochlorococcus* CCMP1375	985	984	983	984	985	982	985	985	982	983			
12 SAR7	975	976	974	974	975	972	973	971	970	971	975		
13 *Prochlorococcus* MIT9303	973	972	971	974	973	970	973	971	977	976	978	976	
14 *Synechococcus* WH7805	962	962	960	963	962	959	962	960	962	963	971	971	977
15 *Synechococcus* WH8103	967	968	966	968	967	963	967	965	966	967	970	979	981
16 SAR139	967	968	966	968	967	963	967	965	966	967	970	980	980
17 *Synechococcus* WH8101	962	963	960	962	962	959	962	960	962	961	968	975	975
18 *Synechococcus* PCC6307	957	956	955	958	956	955	958	956	957	956	961	956	969
19 *Synechococcus* PCC6301	910	910	908	911	907	906	909	907	910	909	908	907	921
20 *Arthrospira* PCC7345	882	883	880	882	880	879	881	879	884	885	879	880	883
21 *Trichodesmium* NIBB1067	862	863	860	861	866	863	866	864	862	862	865	866	867
22 *Prochloron* sp.	860	861	859	859	857	856	859	857	862	864	857	861	863
23 *Microcoleus* PCC7420	875	875	873	874	877	876	878	875	873	873	872	874	873
24 *Anabaena* PCC7120	880	880	878	880	879	878	881	879	882	882	882	878	886
25 *Chlorogloeopsis* HTF PCC7518	854	855	852	854	854	853	855	853	856	858	854	857	859
26 *Chamaesiphon* PCC7430	881	882	879	880	881	880	882	880	878	878	876	873	873
27 *Microcystis aeruginosa* PCC7806	873	873	871	872	873	871	874	872	876	878	877	872	871
28 *Synechocystis* PCC6803	872	872	871	871	872	871	873	871	872	872	868	871	871
29 *Spirulina* PCC6313	887	887	885	887	884	883	886	884	887	889	885	886	889
30 Leptolyngbya foveolarum Komarek 1964/112	865	865	864	864	865	864	866	864	864	865	861	867	861
31 *Prochlorothrix hollandica*	886	886	884	886	885	884	887	885	889	891	888	884	888
32 *Marchantia polymorpha* chloroplast	842	841	840	843	843	841	843	841	844	843	838	839	841
33 *Cyanophora paradoxa* cyanelle	877	878	875	877	878	877	880	878	877	875	870	874	874
34 *Gloeobacter violaceus* PCC7421	879	878	878	878	876	875	878	876	882	882	881	882	888
35 *Escherichia coli*	777	777	776	778	777	776	777	775	784	783	777	781	778
36 *Agrobacterium tumefaciens*	779	780	778	778	782	781	782	779	780	782	778	787	782

identity) to Mediterranean and North Atlantic strains, in addition to other Pacific and tropical Atlantic isolates (high light-adapted clade, Figure 1, Table 5). On the other hand, Sargasso Sea isolates CCMP1375 and MIT9303 fall into different, deeply branching lineages (97.9% identity) and North Atlantic strains NATL2 and NATL2A, cultured from a single water sample, are more distantly related (98.5% identity) than are NATL2A and Pacific strain PAC1 (99.6% identity). There is thus no correlation between genetic distance and the geographic distance between sites of *Prochlorococcus* culture isolation. This result is consistent with the findings of an RFLP study of *Prochlorococcus* cultures (Scanlan et al. 1996).

Phylogenetic Affiliations of Environmental Clones

Studies investigating uncultured microorganisms in the sea have recovered 16S rRNA gene clones belonging to the marine picophytoplankton clade (Giovannoni et al. 1990; Britschgi and Giovannoni 1991; Schmidt et al. 1991; DeLong et al. 1993; Fuhrman et al. 1993). The longest of these sequences are included in our analyses (Fig. 2). Clones SAR6 and SAR139 are closely related to *Prochlorococcus* NATL2 and *Synechococcus* WH8103, respectively, and presumably derive from cells with similar phenotypes. Clone SAR7 is phylogenetically unaffiliated with either *Prochlorococcus* or *Synechococcus* cultured strains. Inference of a pigment phenotype for this clone must await sequences from additional, closely related isolates or results of *in situ* hybridization studies.

Nucleotide Substitution Biases

Picophytoplankton third codon positions and intergenic regions are highly heterogeneous in G + C content (Table 6). Third codon position G + C frequencies range from 0.26 to 0.66 for *psbB*, from 0.18 to 0.73 for *petB/D*, and from 0.17 to 0.51 for intergenic regions, with each strain exhibiting similar biased tendencies at all three loci. *Prochlorococcus* and marine *Synechococcus* sequences

Table 5. Extended

14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
976																						
973	995																					
969	979	983																				
959	959	962	965																			
918	915	918	922	914																		
873	876	878	886	881	897																	
863	865	868	877	871	886	923																
858	861	865	875	861	877	915	902															
868	862	864	872	870	879	919	910	904														
882	880	880	890	882	887	912	907	901	914													
854	859	860	863	856	876	892	886	882	895	913												
862	870	871	875	881	884	901	894	890	904	903	894											
868	870	872	879	873	894	903	888	897	910	900	881	903										
865	875	876	880	880	902	898	888	904	899	902	882	916	933									
888	889	892	900	889	909	914	903	919	899	912	887	902	913	929								
862	864	864	861	856	886	876	861	864	880	886	856	887	884	901	887							
886	886	887	895	892	914	889	884	878	876	882	859	897	895	896	898	877						
838	835	837	842	846	852	866	873	870	865	874	852	857	857	856	866	830	861					
863	875	877	884	875	887	895	886	888	882	894	880	891	882	881	896	857	873	884				
879	883	887	888	892	883	890	891	884	876	877	870	876	871	871	891	849	880	861	882			
773	776	777	775	776	764	769	760	765	766	764	767	760	760	757	767	755	766	775	760	780		
780	781	784	780	788	771	777	784	788	781	772	773	768	773	780	784	770	770	769	775	796	797	

in the literature also exhibit similar, strain-specific biases (Table 7). It therefore appears that consistent nucleotide frequency biases apply to large segments of each genome, in coding and noncoding regions. Thus, mutational biases, rather than selection via intracellular tRNA frequencies (Post et al. 1979; Ikemura 1982), are the likely causes. We explored the pattern of nucleotide substitution biases in our phylogenetic trees by labeling the *psbB* and *petB/D* concatenated sequence phylogeny according to third codon position G + C frequencies, which result in part from these biases (Lockhart and Penny 1992; Lockhart et al. 1992c) (Fig. 3). It has been suggested that the apparent polyphyly of chl *b*-containing prokaryotes and organelles is an artifact of dissimilar substitution biases (Lockhart et al. 1992b, c). However, *Prochlorococcus* NATL2A and maize chloroplast sequences fail to form a phylogenetic cluster, despite identity of their third codon position G + C frequencies. Thus, earlier conclusions of chl *b* polyphyly are reinforced.

Within the marine picophytoplankton clade, there is a correspondence between the branching pattern and third codon position G + C frequencies (Fig. 3). The high light-adapted clade (CCMP1378 and MIT9107) exhibits low frequencies (0.22 and 0.24), with more distantly related isolates exhibiting progressively higher values. *Synechococcus* WH8103, the most divergent strain from *Prochlorococcus* CCMP1378, has a third codon position G + C frequency of 0.69. Since phylogenetic inferences using methods insensitive to substitution biases confirm the pattern inferred by standard methods, the gradient is an intriguing pattern which likely reflects substitution bias differences among marine picophytoplankton lineages.

Taxonomy of Prochlorococcus *and* Synechococcus

Molecular phylogenies for *Prochlorococcus* and *Synechococcus* strains may be used to refine cyanobacterial taxonomies to reflect evolutionary relationships. Branching patterns and bootstrap proportions support proposed

Table 6. Base composition of third codon positions in *psbB*, *petB/D* and the *petB/D* intergenic region; G + C values and numbers in parentheses are fractional values

	A	C	G	T	n	G + C
psbB						
Prochlorococcus CCMP1378	63 (0.38)	23 (0.14)	19 (0.11)	62 (0.37)	167	0.26
Prochlorococcus MIT9107	59 (0.35)	27 (0.16)	17 (0.10)	64 (0.38)	167	0.26
Prochlorococcus NATL2A	53 (0.32)	29 (0.17)	14 (0.08)	71 (0.43)	167	0.26
Prochlorococcus CCMP1375	45 (0.27)	41 (0.25)	17 (0.10)	64 (0.38)	167	0.35
Prochlorococcus MIT9303	24 (0.14)	59 (0.35)	28 (0.17)	56 (0.34)	167	0.52
Synechococcus WH8103	10 (0.06)	69 (0.41)	41 (0.25)	47 (0.28)	167	0.66
Synechococcus PCC7942	22 (0.13)	67 (0.40)	35 (0.21)	43 (0.26)	167	0.61
Synechocystis PCC6803	27 (0.16)	55 (0.33)	31 (0.19)	54 (0.32)	167	0.52
Prochlorothrix hollandica	19 (0.11)	64 (0.38)	32 (0.19)	52 (0.31)	167	0.58
Zea maize chloroplast	57 (0.34)	18 (0.11)	25 (0.15)	67 (0.40)	167	0.26
petB/D						
Prochlorococcus CCMP1378	50 (0.38)	6 (0.05)	17 (0.13)	58 (0.44)	131	0.18
Prochlorococcus MIT9107	50 (0.38)	6 (0.05)	21 (0.16)	54 (0.41)	131	0.21
Prochlorococcus NATL2A	43 (0.33)	14 (0.11)	23 (0.18)	51 (0.39)	131	0.28
Prochlorococcus CCMP1375	37 (0.28)	18 (0.06)	24 (0.18)	52 (0.40)	131	0.32
Prochlorococcus MIT9303	15 (0.11)	41 (0.31)	35 (0.27)	40 (0.31)	131	0.58
Synechococcus WH8103	9 (0.07)	59 (0.45)	37 (0.28)	26 (0.20)	131	0.73
Synechococcus PCC7942	13 (0.10)	49 (0.37)	39 (0.30)	30 (0.23)	131	0.67
Synechocystis PCC6803	16 (0.12)	46 (0.35)	36 (0.27)	33 (0.25)	131	0.63
Prochlorothrix hollandica	13 (0.10)	45 (0.34)	43 (0.33)	30 (0.23)	131	0.67
Zea maize chloroplast	43 (0.33)	14 (0.11)	23 (0.18)	51 (0.39)	131	0.28
Intergenic region between *petB* and *petD*						
Prochlorococcus CCMP1378	18 (0.43)	6 (0.14)	1 (0.02)	17 (0.40)	42	0.17
Prochlorococcus MIT9107	14 (0.41)	9 (0.26)	3 (0.09)	8 (0.24)	34	0.35
Prochlorococcus NATL2A	23 (0.39)	13 (0.22)	1 (0.02)	22 (0.37)	59	0.24
Prochlorococcus CCMP1375	27 (0.34)	11 (0.14)	8 (0.10)	34 (0.42)	80	0.24
Prochlorococcus MIT9303	24 (0.27)	26 (0.29)	9 (0.10)	31 (0.34)	90	0.39
Synechococcus WH8103	8 (0.21)	14 (0.36)	6 (0.05)	11 (0.28)	39	0.51
Synechococcus PCC7942	22 (0.23)	28 (0.29)	25 (0.26)	21 (0.22)	96	0.55
Synechocystis PCC6803	41 (0.32)	31 (0.24)	22 (0.17)	36 (0.28)	130	0.41
Prochlorothrix hollandica	88 (0.21)	101 (0.24)	114 (0.27)	124 (0.29)	427	0.50
Zea maize chloroplast	62 (0.34)	22 (0.12)	36 (0.20)	62 (0.34)	182	0.32

Table 7. G + C nucleotide frequencies for marine picophytoplankton sequences in GenBank

Strain	Gene	Accession No.	Positions	Length (bp)	G + C frequency	Reference
Prochlorococcus						
MED	*rpoC*	Z11159	3rd codon pos.	204	0.20	Palenik and Haselkorn 1992
TATL1	*pstS*	U75514	3rd codon pos.	331	0.22	Scanlan, unpublished
GP2	*rbcL*	D21833	3rd codon pos.	396	0.35	Shimada et al. 1995
SARG	Noncoding clone GATA/1200	Z37732	All	1142	0.37	Lorenz et al. 1995b
SARG	Noncoding clone GATA/1100	Z37732	All	1147	0.29	Lorenz et al. 1995b
SARG	*rpoC*	Z11160	3rd codon pos.	204	0.32	Palenik and Haselkorn 1992
CCMP1375	*asd*[a]	Z68126	3rd codon pos.	56	0.23	Lorenz et al. 1995a, b
CCMP1375	*aspA*	Z80110	3rd codon pos.	309	0.27	Hess, unpublished
CCMP1375	*dapA*[a]	Z68126	3rd codon pos.	304	0.21	Lorenz et al. 1995a, b
CCMP1375	*dnaA*	U44977	3rd codon pos.	462	0.25	Richter, unpublished
CCMP1375	*psbA*	Z49201	3rd codon pos.	361	0.38	Hess et al. 1995
Synechococcus						
WH7803	*pstS*	X71359	3rd codon pos.	326	0.76	Scanlan et al. 1996
WH7805	*rpoC*	L34062	3rd codon pos.	204	0.72	Palenik 1994
WH8103	*rpoC*	L34063	3rd codon pos.	204	0.79	Palenik 1994
WH8103	*mpeA*	M91809	3rd codon pos.	166	0.70	de Lorimer et al. 1992
WH8103	*mpeB*	M91809	3rd codon pos.	179	0.69	de Lorimer et al. 1992

MED: parent strain of CCMP1378

SARG: parent strain of CCMP1375

[a] Does not include nucleotide positions in overlap region

genus-level distinctions among the *Synechococcus* strain clusters *Synechococcus, Cyanobium,* and marine A plus marine B (allied with *Prochlorococcus* in the marine picophytoplankton clade). Branching patterns within the marine picophytoplankton clade further suggest that high phycourobilin marine A, no phycourobilin marine A, and marine B *Synechococcus* are closely related taxa of equal rank, perhaps each representing a single genus. Similarly, independent lineages of *Prochlorococcus* may represent different taxa. For example, low chl b/a_2 ratio *Prochlorococcus* strains may form a monophyletic taxon distinct from *Prochlorococcus marinus.* Additional phenotypic and sequence data will be required to test the generality of these taxonomic conclusions.

Acknowledgments. The authors thank A. Shimada for unpublished sequences, L. Campbell and F. Partensky for *Prochlorococcus* cultures, J. Newman, C. Dias, and P.-N. Xu for help with DNA sequencing, and M. Gouy, N. Galtier, and P.R. Reeves for computer programs. We also thank S.J. Giovannoni for support during the latter portion of this work, B. Lanoil for critical reading of the manuscript and M. Sogin for helpful discussions. This work was supported in part by grants from the National Science Foundation to S.W.C. (OCE-9223793, 9302529, 9022285, 9012117; BSR-9020254; DIR-9101361) and to J.B.W. and D.L.D. (OCE-9315895). D.J.S. is a Royal Society University Research Fellow and was also funded by the NERC under the Plankton Reactivity in the Marine Environment (PRIME) Thematic initiative. This is PRIME publication no. 24.

References

Britschgi TB, Giovannoni SJ (1991) Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. Appl Environ Microbiol 57:1707–1713

Bryant DA (1991) Cyanobacterial phycobilisomes: progress toward complete structural and functional analysis via molecular genetics. In: Bogorad L, Vasil IK (eds) The photosynthetic apparatus: molecular biology and operation. Academic Press, San Diego, CA, pp 257–300

Campbell L, Nolla HA, Vaulot D (1994) The importance of *Prochlorococcus* to community structure in the central North Pacific ocean. Limnol Oceanogr 39:954–961

Chisholm SW, Frankel SL, Goericke R, Olson RJ, Palenik B, Waterbury JB, West-Johnsrud L, Zettler ER (1992) *Prochlorococcus marinus nov.* gen. *nov.* sp.: a marine prokaryote containing divinyl chlorophyll *a* and *b.* Arch Microbiol 157:297–300

Chisholm SW, Olson RJ, Zettler ER, Goericke R, Waterbury J, Welschmeyer N (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. Nature 334:340–343

Cilia V, Lafay B, Christen R (1996) Sequence heterogeneities among 16S ribosomal RNA sequences, and their effect on phylogenetic analyses at the species level. Mol Biol Evol 13:451–461

Dams E, Hendriks L, Van de Peer Y, Neefs JM, Smits G, Vandenbempt I, De Wachter R (1988) Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res 16 (suppl.):r87–r173

de Lorimer RM, Chen C-CJ, Glazer AN (1992) Sequence comparison of two highly homologous phycoerythrins differing in bilin composition. Plant Mol Biol 20:353–356

DeLong EF, Franks DG, Alldredge AL (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. Limnol Oceanogr 38:924–934

Delwiche CF, Palmer JD (1996) Rampant horizontal transfer and duplication of Rubisco genes in eubacteria and plastids. Mol Biol Evol 13:873–882

Devereux J, Haeberli P, Smithies O (1990) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Felsenstein J (1993) PHYLIP v. 3.5c. University of Washington, Seattle, WA.

Fuhrman JA, McCallum K, Davis AA (1993) Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. Appl Environ Microbiol 59:1294–1302

Galtier N, Gouy M (1995) Inferring phylogenies from DNA sequences of unequal base compositions. Proc Natl Acad Aci USA 92:11317–11321

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990) Genetic diversity in Sargasso Sea bacterioplankton. Nature 345:60–62

Giovannoni SJ, Turner S, Olsen GJ, Barnes S, Lane DJ, Pace NR (1988) Evolutionary relationships among cyanobacteria and green chloroplasts. J Bacteriol 170:3584–3592

Goericke R, Repeta DJ (1993) Chromatographic analysis of divinylchlorophylls *a* and *b* in samples from the subtropical north Atlantic Ocean. Mar Ecol Prog Ser

Gray MW, Sankoff D, Cedergren RJ (1984) On the evolutionary descent of organisms and organelles: a global phylogeny based on a highly conserved structural core in small subunit ribosomal RNA. Nucleic Acids Res 12:5837–5852

Greer KL, Golden SS (1992) Conserved relationship between psbH and petBD genes: presence of a shared upstream element in *Prochlorothrix hollandica.* Plant Mol Biol 19:355–365

Haag E, Eaton-Rye JJ, Renger G, Vermaas WFJ (1993) Functionally important domains of the large hydrophilic loop of CP47 as probed by oligonucleotide-directed mutagenesis in *Synechocystis* sp. PCC6803. Biochem 32:4444–4454

Hess WR, Partensky F, van der Staay GWM, Garcia-Fernandez JM, Borner T, Vaulot D (1996) Coexistence of phycoerythrin and a chlorophyll *a/b* antenna in a marine prokaryote. Proc Natl Acad Sci USA 93:11126–11131

Hess WR, Weihe A, Loiseaux-De Goer S, Partensky F, Vaulot D (1995) Characterization of the single psbA gene of *Prochlorococcus marinus* CCMP 1375 (Prochlorophyta). Plant Mol Biol 27: 1189–1196

Hope AB (1993) The chloroplast cytochrome *bf* complex: a critical focus on function. Biochim Biophys Acta 1143:1–22

Hultman T, Stahl S, Hornes E, Uhlen M (1989) Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. Nucleic Acids Res 17:4937–4946

Ikemura T (1982) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optional for the *E. coli* translational system. J Mol Biol 151:389–409

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 31:151–160

Lane DJ (1991) 16S/23/S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds) Nucleic acid techniques in bacterial systematics. John Wiley and Sons, New York, N.Y.

Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR (1985) Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. Proc Natl Acad Sci USA 82:6955–6959

Lang JD, Haselkorn R (1989) Isolation, sequence and transcription of the gene encoding the photosystem II chlorophyll-binding protein, CP-47, in the cyanobacterium *Anabaena* 7120. Plant Mol Biol 13: 441–456

LaRoche J, van der Staay GWM, Partensky F, Ducret A, Aebersold R, Li R, Golden SS, Hiller RG, Wrench PM, Larkum AWD, Green BR

(1996) Independent evolution of the prochlorophyte and green plant chlorophyll *a/b* light-harvesting proteins. Proc Natl Acad Sci USA 93:15244–15248

Lewin RA (1977) *Prochloron*, type genus of the Prochlorophyta. Phycologia 16:217

Li H, Cui X, Arnheim A (1991) Analysis of DNA sequence variation in single cells. Methods 2:49–59

Li WKW, Wood M (1988) Vertical distribution of North Atlantic ultraphytoplankton: analysis by flow cytometry and epifluorescence microscopy. Deep Sea Res 35:1615–1638

Lockhart PJ, Beanland, Howe CJ, Larkum AWD (1992a) Sequence of Prochloron didemni atpBE and the inference of chloroplast origins. Natl Acad Sci USA 89:2742–2746

Lockhart PJ, Howe CJ, Bryant DA, Beanland TJ, Larkum AWD (1992b) Substitutional bias confounds inference of cyanelle origins from sequence data. J Mol Evol 34:153–162

Lockhart PJ, Penny D (1992) The problem of GC content, evolutionary trees and the origins of chl-*a/b* photosynthetic organelles: are the prochlorophytes a eubacterial model for higher plant photosynthesis? In: Murata N (ed) Research in Photosynthesis, vol. III. Kluwer Academic, pp 499–505

Lockhart PJ, Penny D, Hendy MD, Howe CJ, Beanland TJ, Larkum AWD (1992c) Controversy on chloroplast origins. FEBS Lett 301: 127–131

Lorenz M, Boerner T, Hess WR (1995a) Molecular cloning and characterization of a dihydropicolinate synthase (DHPS) gene from the photoautotrophic prokaryote *Prochlorococcus marinus* CCMP1375 (Prochlorophyta). Endocyt Cell Res 11:59–68

Lorenz M, Partensky F, Boerner T, Hess WR (1995b) Sequencing of RAPD fragments amplified from the genome of the prokaryote *Prochlorococcus marinus* (Prochlorophyta). Biochem Mol Biol Internat 36:705–713

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1991) How clonal are bacteria? Proc Natl Acad Sci 90:4384–4388

Moore LR, Goericke R, Chisholm SW (1995) The comparative physiology of *Synechococcus* and *Prochlorococcus:* influence of light and temperature on growth, pigments, fluorescence and absorptive properties. Mar Ecol Prog Ser 116:259–275

Morel A, Ahn Y-H, Partensky F, Vaulot D, Claustre H (1993) *Prochlorococcus* and *Synechococcus:* a comparative study of their size, pigmentation and related optical properties. J Mar Res 51:617–649

Olson RJ, Chisholm SW, Zettler ER, Altabet MA, Dusenberry JA (1990) Spatial and temporal distributions of prochlorophyte picoplankton in the North Atlantic Ocean. Deep-sea Res 37:1033–1051

Palenik B (1994) Cyanobacterial community structure as seen from RNA polymerase gene sequence analysis. Appl Environ Microbiol 60:3212–3219

Palenik B, Haselkorn R (1992) Multiple evolutionary origins of prochlorophytes, the chlorophyll *b*-containing prokaryotes. Nature 355:265–267

Parpais J, Marie D, Partensky F, Morin P, Vaulot D (1996) Effect of phosphorus starvation on the cell cycle of the photosynthetic prokaryote *Prochlorococcus* spp. Mar Ecol Prog Ser 132:265–274

Partensky R, Hoepffner N, Li WKW, Ulloa O, Vaulot D (1993) Photoacclimation of *Prochlorococcus* sp. (Prochlorophyta) strains isolated from the north Atlantic and the Mediterranean Sea. Plant Physiol 101:285–296

Post LE, Strycharz GD, Nomura H, Lewis H, Dennis PP (1979) Nucleotide sequence of the ribosomal protein gene cluster adjacent to the gene for RNA polymerase subunit b in *Escherichia coli.* Proc Natl Acad Sci USA 76:1697–1701

Reeves PR, Farnell L, Lan R (1994) MULTICOMP: a program for preparing sequence data for phylogenetic analysis. Comput Applic Biosci 10:281–284

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual Second Edition. Cold Spring Harbor Laboratory Press, Plainview, NY

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Scanlan DJ, Hess WR, Partensky F, Newman J, Vaulot D (1996) High degree of genetic variation in *Prochlorococcus* (Prochlorophyta) revealed by RFLP analysis. Eur J Phycol 31:1–9

Schmidt TM, DeLong EF, Pace NR (1991) Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol 173:4371–4378

Shimada A (1995) Seasonal appearance of *Prochlorococcus* in Suruga Bay, Japan in 1992–1993. J Oceanogr 51:289–300

Shimada A, Kanai S, Maruyama T (1995) Partial sequence of ribulose-1,5-bisphosphate carboxylase/oxygenase and the phylogeny of *Prochloron* and *Prochlorococcus* (Prochlorales). J Mol Evol 40: 671–677

Shimada A, Maruyama T, Miyachi S (1996) Vertical distributions and photosynthetic action spectra of two oceanic picophytoplankters, *Prochlorococcus marinus* and *Synechococcus* sp. Mar Biotechnol 127:15–23

Sneath PHA (1993) Evidence from *Aeromonas* for genetic crossing-over in ribosomal sequences. Int J Syst Bacteriol 43:626–629

Swift H, Palenik B (1992) Prochlorophyte evolution and the origin of chloroplasts: morphological and molecular evidence. In: Lewin RA (ed) Origins of Plastids. Chapman and Hall, pp 123–139

Turner S, Burger-Wiersma T, Giovannoni SJ, Mur LR, Pace NR (1989) The relationship of a prochlorophyte *Prochlorothrix hollandica* to green chloroplasts. Nature 337:380–382

Urbach E, Daniels B, Salama MS, Sandine WE, Giovannoni SJ (1997) The *ldh* phylogeny for environmental isolates of *Lactococcus lactis* is consistent with rRNA genotypes but not with phenotypes. Appl Environ Microbiol 63:694–702

Urbach E, Robertson D, Chisholm SW (1992) Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. Nature 355:267–269

Vermaas WF, Williams JG, Arntzen CJ (1978) Sequencing and modification of *psbB*, the gene encoding the CP-47 protein of photosystem II in the cyanobacterium *Synechocustis* 6803. Plant Mol Biol 8:317–329

Vermaas WFJ, Ikeuchi M (1991) Photosystem II. In: Bogorad L, Vasil IK (ed.) The photosynthetic apparatus: molecular biology and operation. Academic Press, NY, pp 26–111

Waterbury JB, Rippka R (1989) Subsection 1. Order *Croococcales* Wettstein 1924, emend. Rippka et al., 1979. In: Staley JT, Bryant MP, Pfennig N, Holt JG (ed.) Bergey's manual of systematic bacteriology, vol. 3. Williams and Wilkins, Baltimore, pp 1728–1746

Waterbury JB, Valois FS, Franks DG (1986) Biological and ecological characterization of the marine unicellular cyanobacterium. In: Photosynthetic picoplankton. Can Bull Fish Aquatic Sci 214:71–120

Waterbury JB, Valois FW (1993) Resistance to co-occurring phages enables marine *Synechococcus* communities to coexist with cyanophages abundant in seawater. Appl Environ Microbiol 59:3393–3399

Widger WR, Cramer WA (1991) The b_6/f complex. In: Bogorad L, Vasil IK (ed.) The photosynthetic apparatus: molecular biology and operation. Academic Press, NY, pp 149–224

Wood AM (1985) Photosynthetic apparatus of marine ultraphytoplankton adapted to natural light fields. Nature 316:253–254