International Journal of Systematic and Evolutionary Microbiology (2001), **51**, 861–871 Printed in Great Britain

Phylogenetic analyses of *Synechococcus* strains (cyanobacteria) using sequences of 16S rDNA and part of the phycocyanin operon reveal multiple evolutionary lines and reflect phycobilin content

Microbiology Laboratory, National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaraki, 305-0053, Japan

Bronwyn R. Robertson,† Naoaki Tezuka and Makoto M. Watanabe

Author for correspondence: Bronwyn R. Robertson. Tel: +61 2 9385 2030. Fax: +61 2 9385 1050. e-mail: b.robertson@unsw.edu.au

The genus *Synechococcus* (cyanobacteria), while containing morphologically similar isolates, is polyphyletic and organisms presently classified as such require reclassification into several independent genera. Studies based on analysis of 16S rRNA gene sequences have shown that members of the genus *Synechococcus* are affiliated to three of seven deeply branching cyanobacterial lineages. In addition, some strains do not appear to be associated with any of these lineages and may represent novel clades. In this report, a cyanobacterial phylogeny based on 16S rDNA sequences, including 14 newly sequenced *Synechococcus* isolates, is presented. One newly sequenced *Synechococcus* strain (PCC 7902) did not have any close relatives amongst cyanobacterial isolates currently contained in 16S rDNA sequence databases and was only loosely affiliated to a cyanobacterial lineage in which no other *Synechococcus* strains were found. Three hot-spring *Synechococcus* isolates, including two that were newly sequenced in this study (PCC 6716 and PCC 6717), formed an additional cyanobacterial lineage. These results indicated that *Synechococcus* species are affiliated to five of eight deeply branching cyanobacterial lineages. Part of the phycocyanin (PC) gene sequence (*cpc*), including the intergenic spacer (IGS) between *cpcB* and *cpcA* and the corresponding flanking regions (*cpcBA*-IGS), was used to investigate relationships between closely related *Synechococcus* isolates. Previously described PCR primers did not amplify this region from the majority of strains under investigation, so a new set of primers was designed that allowed amplification and sequencing of the *cpcBA*-IGS and flanking regions from 38 *Synechococcus* species. Phylogenetic analysis of this region was largely consistent with that obtained from 16S rDNA sequence analysis and revealed a relationship between the primary PC DNA sequence and the phycobilin content of cells.

Keywords: cyanobacteria, phycocyanin, phylogeny, 16S rDNA, *Synechococcus*

INTRODUCTION

Cyanobacteria represent one of the major bacterial phyla, being a large, diverse and widespread group inhabiting most of the Earth's environments. They play an important role as primary producers (Chisholm *et al*., 1988; Waterbury *et al*., 1979), but are also capable of extensive growth, resulting in bloom events that can cause a significant threat to human and animal health (Carmichael, 1992). The current classification of cyanobacteria relies heavily upon morphological observations such as cell size, shape and arrangement (filamentous, colonial or single cells), coloration and the presence of characters such as gas vacuoles and a sheath (Castenholz & Waterbury,

†**Present address:** Clive and Vera Ramaciotti Centre for Gene Function Analysis, School of Biochemistry and Molecular Genetics, Faculty of Life Sciences, The University of New South Wales, Sydney, NSW 2052, Australia.

Abbreviations: PC, phycocyanin; PE, phycoerythrin; IGS, intergenic spacer.

The GenBank accession numbers for the 16S rDNA sequences determined in this paper are AF216942–AF216955. The GenBank accession numbers for the sequences of the *cpcBA*-IGS and flanking regions determined in this paper are AF223428–AF223465.

1989). A number of unicellular, non-beocyte-forming species of cyanobacteria, however, lack an abundance of complex morphological characteristics and have been classified together in a single order, the Chroococcales. Although morphologically similar, it has been shown that this order contains organisms originating from several evolutionarily distinct and deeply branching groups within the cyanobacterial phylum (Stanier *et al*., 1971). *Synechococcus* is one member genus of the Chroococcales, being characterized as unicellular, rod-shaped to coccoid organisms, less than 3 μm in diameter, with division by binary fission into equal halves in one plane (Holt *et al*., 1994).

It has been widely suggested that organisms presently classified in the genus *Synechococcus*, sometimes referred to as a 'culture group', be further divided into a number of separate genera (Herdman *et al*., 1979; Honda *et al*., 1999; Rippka *et al*., 1979; Wilmotte & Stam, 1984). Six strain clusters of *Synechococcus* have been defined by Waterbury & Rippka (1989) based primarily on differences in G+C content (mol%) and habitat (freshwater or marine). These strain clusters are intended to represent provisional genera (Urbach *et al*., 1998) and are designated as follows: *Cyanobacterium* cluster, *Synechococcus* cluster, *Cyanobium* cluster, Marine cluster A, Marine cluster B and Marine cluster C (Holt *et al*., 1994). *Synechococcus* isolates maintained at the Pasteur Culture Collection (PCC), however, are divided into five clusters that do not correlate simply with the six clusters described above (Rippka & Herdman, 1992). The *Cyanobacterium* cluster is analogous to PCC cluster 4. Marine cluster C contains strains from PCC cluster 3. The *Cyanobium* cluster contains strains from PCC cluster 5 as well as strain PCC 6603, which has not yet been classified in a PCC cluster. The *Synechococcus* cluster, however, contains strains from three PCC clusters, namely clusters 1, 2 and 3. Marine cluster A contains strains from the Woods Hole Culture Collection, as does Marine cluster B, which also includes a single PCC strain classified in cluster 5. These differing clustering systems can contribute to confusion when dealing with *Synechococcus* strains. At present, identification of *Synechococcus* species is unclear. A recent study of cyanobacterial phylogeny using 16S rRNA gene sequence analysis further emphasized that the genus *Synechococcus* is polyphyletic and not a natural taxon, with members found in three of the seven groups identified, as well as outside these seven groups (Honda *et al*., 1999). It is clear that there are numerous limitations with the current taxonomy of *Synechococcus* at the genus level and, additionally, detailed studies to determine species classification of these bacteria are required.

In order to elucidate natural relationships among the *Synechococcus* culture group further, genetic characterization is required on a wide range of isolates and the use of multiple genetic markers is desirable. In addition to analysis of 16S rRNA gene sequences (Giovannoni *et al*., 1988; Honda *et al*., 1999; Neilan *et al*., 1997; Otsuka *et al*., 1998; Turner *et al*., 1999; Urbach *et al*., 1998; Wilmotte, 1994), genes encoding the major light-harvesting accessory pigment proteins, particularly the phycocyanin (PC) operon (*cpc*), including the intergenic spacer (IGS) between *cpcB* and *cpcA* and the corresponding flanking regions (*cpcBA*-IGS), have been targeted for phylogenetic studies of cyanobacteria (Barker *et al*., 1999; Bolch *et al*., 1996; Hayes & Barker, 1997; Neilan *et al*., 1995).

The aim of this study was to investigate the phylogeny of *Synechococcus* isolates cultured from Japanese lakes as well as *Synechococcus* isolates from the PCC (Paris, France) using PCR amplification and DNA sequencing of the 16S rDNA and the *cpcBA*-IGS and flanking regions. The 16S rDNA sequences from 14 *Synechococcus* isolates and the *cpcBA*-IGS sequences, including the 3′ flanking region of *cpcB* and the 5′ flanking region of *cpcA*, were determined from 38 *Synechococcus* isolates. Phylogenetic analysis of the 16S rDNA sequences further reinforced the polyphyletic nature of this taxon. Analysis of DNA sequences from the *cpcBA*-IGS and flanking regions revealed strain clusters that were in agreement with those obtained for the 16S rDNA data and supported the validity of this fragment for phylogenetic studies on this group. *cpcBA*-IGS sequence analysis enabled a large group of closely related *Synechococcus* species to be divided into seven clusters that reflected phycobilin content, as indicated by cell colour, and *in vivo* and phycobilin pigment absorption spectra. These results point to a phylogenetic relationship between the PC gene sequence and phycobilin content in this group of isolates.

METHODS

Cyanobacterial strains. A list of strains investigated is given in Table 1. PS strains (Provisional Strain Identification Number; $n = 29$) were obtained from the culture collection of the National Institute for Environmental Studies (NIES), Tsukuba, Japan, and, with one exception, were isolated from lakes in Japan. PCC strains ($n = 9$) were obtained from the PCC. Strains were cultured in C, CA, CB or CT media for PS strains (Watanabe & Hiroki, 1997) or BG-11 media for PCC strains (Rippka & Herdman, 1992). All PS strains were grown at 15 °C, whereas PCC strains were grown at either 20 or 37 °C with illumination of approximately 50 μE m^{-2} s^{-1} from daylight fluorescent lamps with a 12:12 h light:dark cycle.

Determination of *in vivo* and phycobilin pigment absorption spectra. Exponential phase cultures were used for *in vivo* and phycobilin pigment absorption spectra determinations using the method described by Suda *et al*. (1998). *In vivo* pigment absorption spectra measurements were performed between 350 and 850 nm using a Hitachi U-3000 spectrophotometer with a 60 mm diameter integrating sphere device. For phycobilin pigment extraction, 10 ml culture was centrifuged at 3600 r.p.m. for 15 min and the supernatant was removed. Cells were then washed once in saline EDTA and resuspended in potassium phosphate buffer (pH 6·5). Cells were broken using a model 3110BX Mini-Beadbeater

Table 1. *Synechococcus* strains used in this study

PS, Provisional Strain Identification Number, National Institute for Environmental Studies (Japan); PCC, Pasteur Culture Collection (France). PS strains were grown in medium C except for PS-680, which was grown in medium CA, PS-672, PS-673, PS-674, PS-716 and PS-726, which were grown in medium CB, and PS-720, which was grown in CT. PCC strains were grown in BG-11 medium. Information on PCC strains was taken from Rippka & Herdman (1992). Group designations are based on the *cpcBA*-IGS sequence as described in the text. Sequences for other strains used in the phylogenetic analyses are available through GenBank or the Ribosomal Database Project (Maidak *et al.*, 1999).

Strain	Origin/collection date (PS strains)	Phycobilin type (PC/PE)	IGS length (bp)	GenBank accession no.	
				cpcBA-IGS	16S rDNA
Group A					
PCC 7009	Low salinity brine pond, CA, USA	PC	60	AF223460	AF216945
PCC 7918	Pond, Gunnarskulla, Kirkkonummi, Finland	PC	56	AF223462	AF216947
PCC 6904	Stream, CA, USA	PC	45	AF223459	AF216944
PCC 7920	Pond, Tolsa, Kirkkonummi, Finland	PC	43	AF223463	AF216948
Group B					
PS-729	Lake Misuzu, Nagano, Japan (May 1991)	PE	46	AF223451	
PS-727	Lake Nakatsuna, Nagano, Japan (May 1991)	PE	46	AF223447	
PS-725	Lake Kasumigaura, Ibaraki, Japan (Feb. 1991)	PE	42	AF223446	
PS-728	Lake Megami, Nagano, Japan (May 1991)	PE	42	AF223437	
PS-720	Lake Nojiri, Nagano, Japan (May 1991)	PE	42	AF223450	
PS-714	Lough Neagh, Northern Ireland, UK (Mar. 1991)	PE	42	AF223448	
Group C					
PS-684	Lake Abashiri, Hokkaido, Japan (May 1991)	PC	40	AF223454	
PS-679	Lake Abashiri, Hokkaido, Japan (May 1991)	PC	40	AF223453	
PS-681	Lake Kasumigaura, Ibaraki, Japan (Feb. 1991)	PC	42	AF223443	
Group D					
PS-685	Lake Ushikunuma, Ibaraki, Japan (Jun. 1991)	PC	54	AF223455	
PS-672	Lake Junsainuma, Hokkaido, Japan (Jun. 1991)	PC	51	AF223433	
PS-676	Lake Teganuma, Chiba, Japan (Jun. 1991)	PC	54	AF223439	
PS-675	Lake Himenuma, Hokkaido, Japan (May 1991)	PC	54	AF223452	
PS-678	Lake Kojima, Okayama, Japan (Jul. 1991)	PC	54	AF223441	
PS-673	Lake Harutori, Hokkaido, Japan (May 1991)	PC	54	AF223434	
PS-674	Lake Harutori, Hokkaido, Japan (May 1991)	PC	54	AF223438	
PS-677	Lake Ushikunuma, Ibaraki, Japan (Jun. 1991)	PC	54	AF223440	
Group E					
PS-722	Lake Biwa, Shiga, Japan (Sep. 1990)	PE	38	AF223435	
PS-716	Lake Kizaki, Nagano, Japan (May 1991)	PE	38	AF223449	
PS-726	Lake Ohnuma, Hokkaido, Japan (Jun. 1991)	PE	38	AF223442	
PS-721	Lake Biwa, Shiga, Japan (Jun. 1991)	PE	38	AF223431	AF216954
PS-719	Lake Kizaki, Nagano, Japan (Apr. 1991)	PE	38	AF223445	
PS-723	Lake Biwa, Shiga, Japan (Sep. 1990)	PE	38	AF223432	AF216955
PS-717	Lake Biwa, Shiga, Japan (Sep. 1990)	PE	38	AF223430	AF216953
Group F					
PCC 7943	Unknown	PC	51	AF223464	AF216949
Group G					
PCC 6717	Hot spring, Yellowstone National Park, USA	PC	64	AF223458	AF216943
PCC 6716	Hunter's Hot Spring, OR, USA	PC	64	AF223457	AF216942
Ungrouped					
PS-683	Lake Sagami, Kanagawa, Japan (Apr. 1991)	PC	79	AF223444	
PCC 9005	Pond, Bangkok, Thailand	PC	89	AF223465	AF216950
PS-724	Lake Tsukui, Kanagawa, Japan (Apr. 1991)	PE	97	AF223436	
PS-680	Lake Biwa, Shiga, Japan (Sep. 1990)	PC	99	AF223428	AF216951
PS-715	Lake Akan, Hokkaido, Japan (May 1991)	PE	49	AF223429	AF216952
PS-718	Lake Biwa, Shiga, Japan (Sep. 1990)	PE	56	AF223456	
PCC 7902	Polar bear hair, San Diego Zoo, CA, USA	PC	148	AF223461	AF216946

(Biospec Products) with 0·3 mm beads, 4 × 40 s at 5000 r.p.m. on ice. The broken cell suspension was centrifuged at 3300 r.p.m. for 10 min at 5 °C and the supernatant was subjected to ultra-centrifugation at 50000 r.p.m. for 1 h at 5 °C. The absorption spectrum of the supernatant was measured as above except that the integrating device was not used.

DNA extraction. DNA was prepared from exponential phase culture using a modification of the bacterial lysis method of Tillett & Neilan (2000). A 2 ml aliquot of mid- to late-exponential phase culture was pelleted by centrifugation at 15000 r.p.m. for 5 min. The supernatant was decanted and the pellet was resuspended in 1 ml XS buffer containing 0·01 % (w/v) potassium ethyl xanthogenate (Fluka Chemika), 0·8 M ammonium acetate, 0·1 M Tris/HCl (pH 8·0), 20 mM EDTA (pH 8·0) and 1 % (w/v) SDS and mixed by vortexing for 10 s. The solution was incubated at 70 °C for 30 min, vortexed for 10 s and cooled on ice for 30 min. The tube was centrifuged at 15000 r.p.m. and 4 °C for 10 min and then the supernatant, containing genomic DNA, was carefully removed into a new microcentrifuge tube. The DNA was precipitated by the addition of an equal volume of 2-propanol and incubated at room temperature for 10 min. The DNA was collected by centrifugation at 15000 r.p.m. for 20 min at 4 °C. The resulting pellet was washed with 70 % ethanol, dried using a vacuum desiccator and then resuspended in 200 µl TE buffer, pH 8·0 (10 mM Tris/HCl, 1 mM EDTA). An aliquot of DNA was diluted 1/10 in distilled water for use in PCR applications.

PCR amplification and sequencing. PCR was carried out in 0·2 ml tubes with a final volume of 50 µl and contained 1 × *Ex Taq* buffer, 1 U *TaKaRa Ex Taq* (TaKaRa Biomedicals), 200 µM each dNTP, 2·5 mM $MgCl_2$, 10 pmol each oligonucleotide primer and approximately 50 ng genomic DNA. For amplification of the 16S rDNA, primers F27(UFP) and R1492(URP) (Neilan *et al.*, 1997) were included in the reaction with an amplification profile consisting of an initial denaturation at 95 °C for 3 min followed by 30 cycles of 94 °C for 10 s, 55 °C for 10 s and 72 °C for 90 s. This was followed by a final elongation step of 72 °C for 8 min. Ramp rates between temperatures were set at 1 °C s^{-1} using a PCR personal thermal cycler (TaKaRa Biomedicals). For amplification of the *cpcBA*-IGS and flanking regions, primers PCβF and PCαR were initially adopted using the conditions described by Neilan *et al.* (1995). Subsequently, primers cpcBF(UFP) and cpcAR (URP) were utilized for amplification of this region (see Results). PCR products were separated by electrophoresis through 1·5 % agarose in TAE buffer (40 mM Tris/acetate, 1 mM EDTA) and photographed under UV transillumination after staining with ethidium bromide.

PCR-amplified DNA was purified prior to being used as a template in sequencing reactions by using the QIAquick PCR Purification kit (Qiagen) according to the manufacturer's protocol. Sequencing reactions were carried out using the ABI BigDye Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems) according to the manufacturer's protocol except that the final volume was 10 µl and contained 2 µl BigDye mix and 2 µl *half*BD sequencing reagent (GENPAK). Sequencing products were separated on a model 310 DNA sequencer machine and analysed using programs contained in the INHERIT package (PE Applied Biosystems). Primers used for 16S rDNA sequencing reactions were 27F, 119F, 341R, 530F, 685R, 942F, 1096R, 1221R, 1241F and 1494R, as described previously (Neilan *et al.*, 1997; Weisburg *et al.*, 1991, 1985). In all cases, both strands of DNA were sequenced with contiguous overlaps.

Phylogenetic inferences. Sequences were aligned using the GCG program PILEUP, version 8 (Genetics Computer Group, Program Manual for the Wisconsin Package, 1994) and the multiple-sequence alignment and profile alignment tools in the CLUSTAL X package (Thompson *et al.*, 1994). Aligned sequences were corrected manually and nucleotide positions that contained ambiguities (insertions, deletions or unknown bases) were removed from further analysis. Genetic distances, corrected for multiple base changes by the method of Jukes & Cantor (1969), were calculated using the DNADIST program in the PHYLIP package, version 3.57c (Felsenstein, 1989). The phylogenetic tree was reconstructed by the neighbour-joining method of Saitou & Nei (1987) and the tree was plotted using NJPLOT in the CLUSTAL X package. Parsimony analysis was performed using DNAPARS in the PHYLIP package. Bootstrap values for both the distance and parsimony phylogenetic trees were obtained from analysis of 100 resamplings of the data set created using the programs SEQBOOT (Felsenstein, 1985) and CONSENSE from the PHYLIP package. With the exception of CLUSTAL X, programs used in the sequence manipulation and phylogenetic analyses were made available through the Australian National Genomic Information Service (ANGIS, Sydney University, Australia).

RESULTS

Determination of *in vivo* and phycobilin pigment absorption spectra

In vivo and phycobilin absorption spectra were determined for 29 PS strains. Two types of spectra were observed for both the *in vivo* and the phycobilin determinations. PC-type absorption spectra were obtained for 13/29 strains. In these strains, *in vivo* absorption spectra showed a peak at 630 nm that was considered to be PC (Fig. 1a). This was confirmed in the phycobilin absorption spectra for the same strains, which showed a peak at 621 nm (Fig. 1b). Phycoerythrin (PE)-type absorption spectra were obtained for the remaining 16/29 strains. Here, the *in vivo* absorption spectra showed a major peak at 570 nm and a small peak at 625 nm. These were considered to be PE and PC, respectively (Fig. 1c). This was again confirmed by the phycobilin absorption spectra, which showed a major peak at 560 nm and a small peak at 620 nm (Fig. 1d). The results for individual strains are given in Table 1 as being either PC- or PE-type. All PCC strains investigated for this study were blue-green in colour and were assigned to the PC-type group without determining their *in vivo* or phycobilin absorption spectra.

Amplification of *cpcBA*-IGS

DNA was successfully amplified from only one of five PS strains initially tested using primers PCβF and PCαR and conditions described by Neilan *et al.* (1995). Altering the annealing temperature of the reaction to 45 °C (from 50 °C) and titrating the $MgCl_2$ concentration between 2·0 and 3·5 mM did not allow the remaining four DNA samples to be amplified. With an

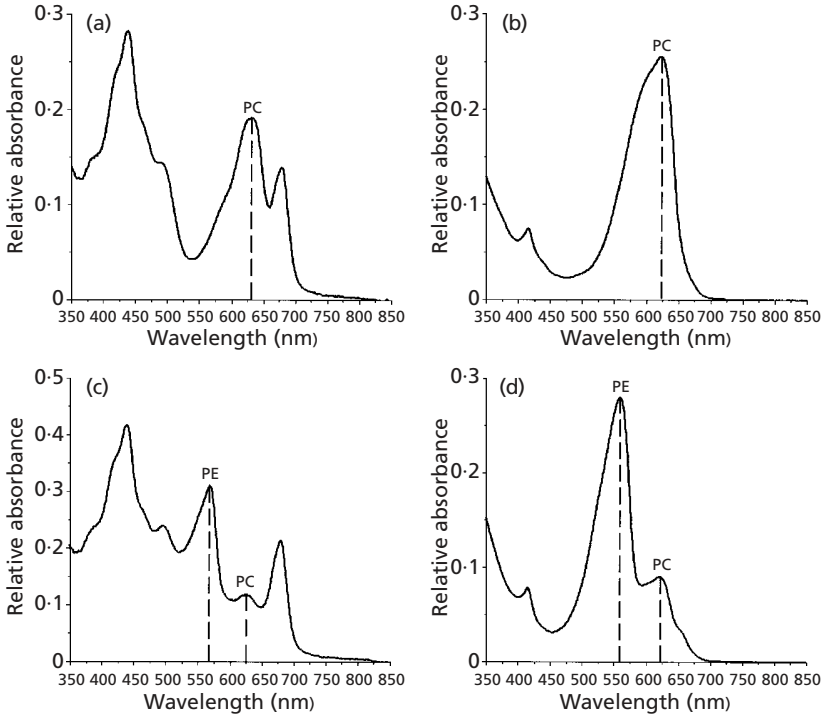

Fig. 1. Examples of typical *in vivo* and phycobilin absorption spectra obtained for PS strains. PC-type *in vivo* (a) and phycobilin (b) spectra were from isolate PS-679, whereas PE-type *in vivo* (c) and phycobilin (d) spectra were from isolate PS-725.

annealing temperature of 45 °C and a $MgCl_2$ concentration of 2·5 mM, a single fragment was amplified from only four of 29 PS strains. As this previously described primer set was not suitable for amplification of DNA from most of the PS strains, a new primer set was designed for PCR amplification of the *cpcBA*-IGS and flanking regions.

Oligonucleotide primers were designed from PC sequences available in GenBank from members of the Chroococcales, including *Synechococcus*, *Synechocystis* and *Microcystis* species. Sequences were aligned using the multiple-sequence alignment tool in the CLUSTAL X package (Thompson *et al.*, 1994) and primers were selected from conserved regions within the *cpcB* and *cpcA* genes to amplify a region encompassing the *cpcBA*-IGS. The selection of bases at non-conserved positions was based on a bias towards the majority of *Synechococcus* sequences. Primers were synthesized with a universal sequencing primer (either forward or reverse) attached to the 5′ end to enable subsequent sequencing to be carried out utilizing these non-degenerate primers. The forward primer, designated cpcBF(UFP) [5′-TAG TGT AAA ACG ACG GCC AGT TG(C/T) (C/T)T(G/T) CGC GAC ATG GA], was selected from a region overlapping the PCβF primer of Neilan *et al.* (1995) beginning at position 8 from the 5′ end of this primer and finishing 5 bases downstream from the 3′ end of PCβF (Table 2). The 5′ end of the reverse primer, designated cpcAR(URP) [5′-TAG CAG GAA ACA GCT ATG ACG TGG TGT A(G/A)G GGA A(T/C)T T], was located 95 bases upstream from the 3′ end of PCαR. The expected size of the amplification fragment from these two primers was, therefore, 100 bases smaller than if the previously published primer set had been employed successfully. Primer cpcBF(UFP) contained degeneracy at three positions, whereas cpcAR(URP) had two positions with degeneracy. Part of the alignment on which these primers were based is shown in Table 2.

The optimized PCR employed an annealing temperature of 50 °C. The thermocycling was otherwise carried out as for the 16S rDNA amplification except that 35 cycles were performed. Two PCC strains (PCC 7918 and PCC 7902) could not be amplified under these conditions, but were successfully amplified when the annealing temperature was reduced to 45 °C. Amplification products varied in size between approximately 500 and 600 bp, with most having a product of approximately 500 bp. PS-680, PS-683, PS-724, PCC 7902 and PCC 9005 had visibly larger amplification products than all other isolates (data not shown).

Sequencing and phylogenetic analyses

Near-complete 16S rDNA sequences were determined for five PS *Synechococcus* isolates that were available in axenic culture as well as for nine PCC *Synechococcus* strains. GenBank accession numbers are given in Table 1. A phylogenetic tree was reconstructed from an alignment of 58 sequences, including 30 *Synechococcus* isolates, representing the major lineages of cyanobacteria with *Escherichia coli* as the outgroup (Fig. 2). The corrected sequence alignment, which formed the basis of the phylogenetic analyses, corresponded to positions 46–1455 according to the *E. coli* numbering system and was 1332 bases in length after gaps and ambiguous positions had been removed.

Synechococcus sp. PCC 6716 and PCC 6717 clustered

Table 2. Phycocyanin primer target sites

Asterisks (*) indicate positions that were conserved for all sequences used in the alignment. DNA sequences are written in the 5′ to 3′ orientation. The reverse primer was synthesized as the reverse complement to the sequence shown here.

Organism	Accession no.	Forward primer target site	Reverse primer target site
Synechococcus sp. PCC 7002	K02659	GGCTGCTTGTCTTCGCGACATGGA	AAGTTCCCCTACACCAC
Synechococcus sp.	K02660	GGCTGCTTGTCTTCGCGACATGGA	AAGTTCCCCTACACCAC
Synechococcus sp. EWAG174	AJ003180	GGCTGCTTGTTTACGCGACATGGA	AAATTCCCCTACACCAC
Synechococcus sp. PCC 7942	AB008546	GGCGGCTTGTCTGCGCGACATGGA	AAGTTCCCCTACACCAC
Synechococcus sp. PCC 6301	M94218	GGCGGCTTGTCTGCGCGACATGGA	AAGTTCCCCTACACCAC
Synechococcus sp. WH8020	M95289	GGCTGCTTGCCTGCGCGACATGGA	AAGTTCCCCTACACCAC
Synechococcus sp.	X59809	GGCTGCATGCCTGCGAGACATGGA	AAATTCCCCTACACCAC
Synechococcus sp.	M95288	GGCAGCTTGCCTTCGTGACATGGA	AAGTTCCCCTACACCAC
'*Synechococcus elongatus*' (Toray)	D13173	GGCTGCCTGCCTGCGCGACATGGA	AAATTCCCCTACACCAC
Synechocystis sp. PCC 9413	AF068771	GGCTGCTTGCTTGCGCGACATGGA	AAGTTCCCTTACACCAC
Synechocystis sp. PCC 6701	M33820	GGCTGCTTGTCTTCGTGACATGGA	AAATTCCCTTACACCAC
Microcystis aeruginosa PCC 7820	AJ003169	GGCTGCTTGTTTACGCGACATGGA	AAGTACCCCTACACCAC
Chroococcus dispersus	AJ003186	GGCTGCTTGTTTACGCGACATGGA	AAATTCCCCTATACCAC
		*** ** ** * ** ********	** * *** ** *****
Primer consensus sequence target sites			
PCβF		GGCTGCTTGTTTACGCGACA	
cpcBF/cpcAR		TGYYTKCGCGACATGGA	AARTTCCCYTACACCAC

with '*Synechococcus elongatus*' (Toray), which remained ungrouped in the cyanobacterial phylogeny of seven lineages proposed by Honda *et al.* (1999). These three sequences shared greater than 98·7% sequence similarity. With the addition of our two sequences, it is proposed that this cluster be called group 8 as it was supported by 100% bootstrap resamplings in both distance and parsimony analysis and was the deepest branching group after *Gloeobacter violaceus*. *Synechococcus* sp. PCC 7902, sequenced in this study, was loosely affiliated with group 4, which previously did not contain any *Synechococcus* strains. The majority (21/30) of *Synechococcus* strains, including all five PS and six of the nine newly sequenced PCC strains, were located in group 6, which divided, with a high degree of certainty, into three subgroups (a–c). The three strains forming group 6a were not included in the study by Honda *et al.* (1999). These isolates originated from saline lakes in eastern Antarctica (Rankin, 1998). The separation of sequences in group 6c, with one additional sequence determined here, confirms what has been shown previously. In total, *Synechococcus* isolates were found in five of eight major cyanobacterial lineages based on 16S rDNA sequence analysis.

The DNA sequence for the *cpcBA*-IGS and flanking regions was determined for 29 PS strains and the same nine PCC strains for which the 16S rDNA sequence was determined in this investigation. GenBank accession numbers are shown in Table 1. The length of IGS sequences varied between 38 and 148 bp (Table 1, Fig. 3). Initially, 50 *cpcBA*-IGS sequences were used in the phylogenetic analysis. These included all sequences available from strains used in the 16S rDNA analysis ($n = 20$), which were from groups 4, 5, 6b, 6c and 8, the remaining PS strains ($n = 24$) and six additional cyanobacterial species from the genera *Synechococcus* (WH8020), *Synechocystis* (PCC 6701 and PCC 9413), *Microcystis* (PCC 7820), *Calothrix* (PCC 7601) and *Pseudanabaena* (PCC 7409). The tree resulting from this analysis indicated that sequences known or assumed to originate from 16S rDNA lineages outside groups 6 and 8 clustered accordingly (data not shown).

Sequences from strains outside 16S rDNA groups 6 and 8 were omitted from further analysis, with PCC 7002 being retained as the outgroup. The sequence obtained from PCC 7902 was therefore the only new PC fragment sequence determined here that was omitted from further phylogenetic analysis. In this analysis, 16/41 strains, not including the outgroup, were the same as those analysed in the 16S rDNA analysis. A phylogenetic tree was constructed based on an unambiguous alignment of 350 bases (Fig. 3). Due to the highly variable nature of the IGS sequence and the resulting alignment difficulties, these 350 bases originated almost exclusively from within the coding regions of *cpcB* and *cpcA* flanking the IGS. The tree was divided into seven groups, designated A–G. These divisions were consistent with the lengths of the IGS of individual isolates (Table 1, Fig. 3). The division of the more distantly related groups was in agreement with the previously determined 16S rDNA-based phylogeny. Group G was the deepest branching group and included isolates identical to those seen in group 8 of the 16S rDNA-based tree. Similarly, group F contained members identical to 16S rDNA group 6c. Both of these groups were supported by bootstrap values of 100% for distance and parsimony analyses.

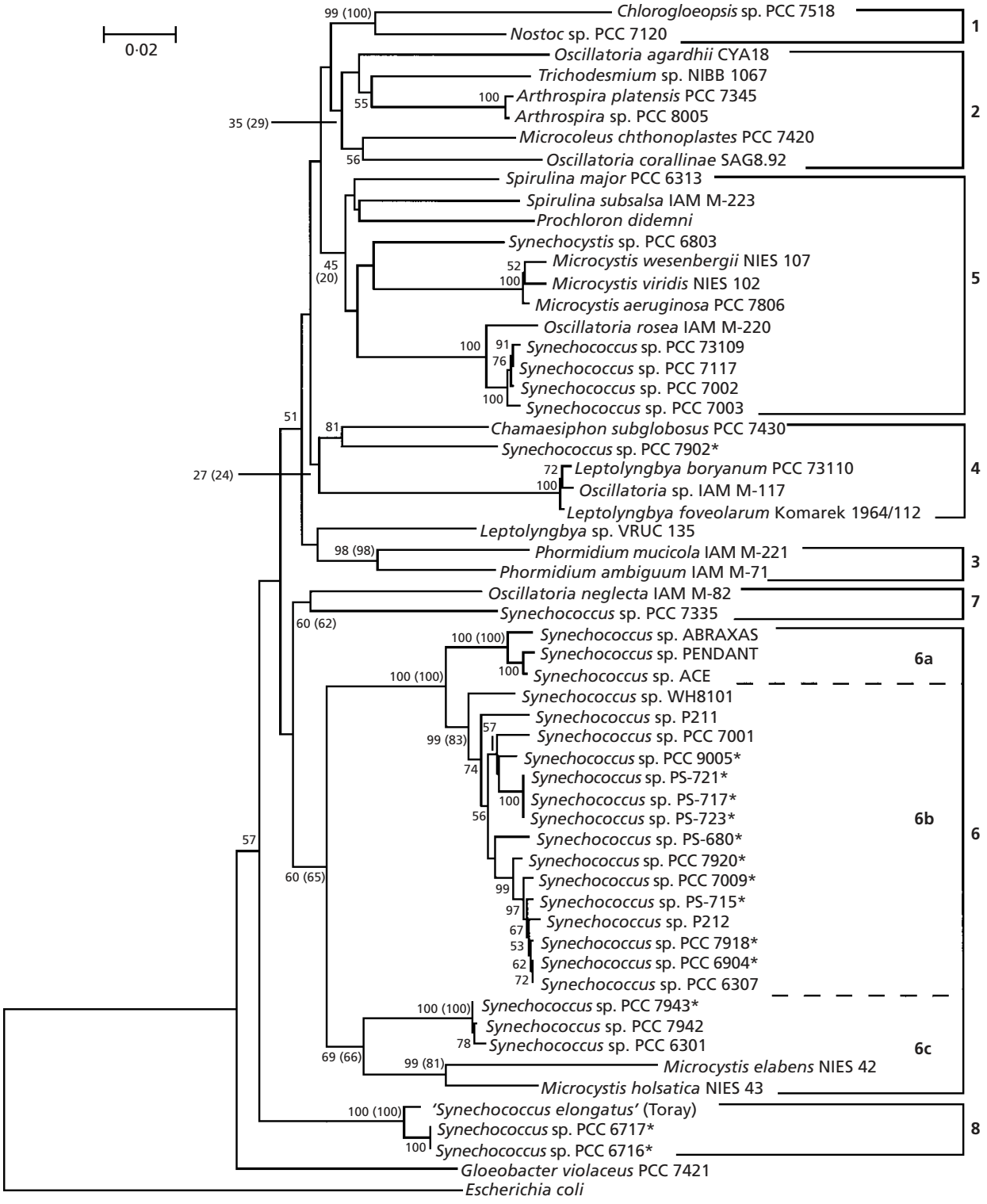

Fig. 2. Phylogenetic tree based on analysis of near-complete cyanobacterial 16S rDNA sequences with *E. coli* as the outgroup. Bar, 0·02 substitutions per site. Significant bootstrap values from the distance analysis (> 50 %) and those for major lineages are given at the nodes. Numbers in parentheses indicate bootstrap values for the major lineages obtained from parsimony analysis. Asterisks indicate strains for which the 16S rDNA sequence was determined in this study.

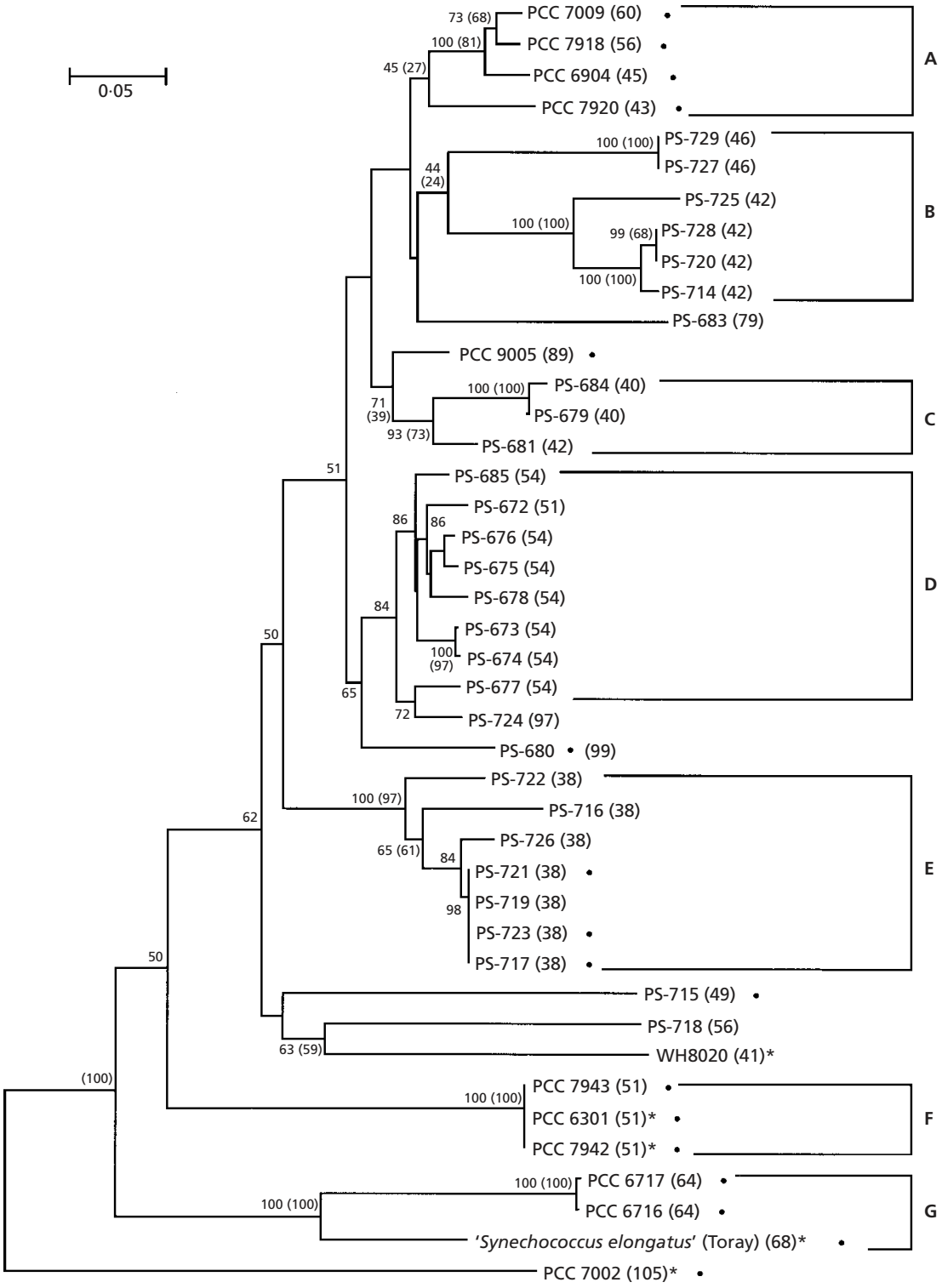

Fig. 3. Phylogenetic tree based on analysis of sequences of the *cpcBA*-IGS and flanking regions from 42 *Synechococcus* strains. Numbers in parentheses following strain designation are the lengths (bp) of IGS sequences alone. Bar, 0·05 substitutions per site. Significant bootstrap values (> 50 %) and those for the major groups from distance and parsimony analysis (in parentheses) are indicated at the nodes. Asterisks indicate strains for which the *cpcBA*-IGS sequence was obtained from GenBank. Bullet points indicate strains that were also included in the phylogenetic tree reconstructed from 16S rDNA sequences (Fig. 2).

Ten of the remaining sequences in the *cpcBA*-IGS analysis were from isolates found in group 6b in the 16S rDNA-based phylogeny, where all 16S rDNA sequences shared sequence similarities of $\geqslant 96{\cdot}2\,\%$. Separation of these sequences revealed similarities to the 16S rDNA tree topology, as outlined below (Fig. 3).

Group A contained four PCC strains that were found in the lower half of group 6b. Two of these strains (PCC 7009 and PCC 6904) are classified as being in the *Cyanobium* strain cluster and PCC cluster 5 and all four strains had PC-type *in vivo* and phycobilin absorption spectra (Table 1). The group was not supported by significant bootstrap values and the IGS length varied between 43 and 60 bp (Table 1, Fig. 3); however, the subcluster of three strains was highly supported. Group B contained six PS strains for which no 16S rDNA data were available. These strains had PE-type *in vivo* and phycobilin absorption spectra (Table 1). Again, the entire cluster was not well supported by bootstrap analysis, but the two subgroups were each supported by 100 % bootstrap values in both distance and parsimony analysis. Group C contained three PS strains, all having PC-type *in vivo* and phycobilin absorption spectra (Table 1). This group was significantly supported by bootstrap values for both distance and parsimony analysis. Group D contained eight strains, for which no 16S rDNA data were available. All strains had PC-type absorption spectra and seven of these eight strains had IGS sequences of 54 bp (Table 1, Fig. 3). Group E contained seven PS isolates, all of which had PE-type spectra and IGS sequences of 38 bp (Table 1), and was well supported by bootstrap analyses. The three strains in this group for which 16S rDNA data were also obtained during this study were located in the upper half of group 6b in that analysis. Seven strains remained ungrouped in the *cpcBA*-IGS tree, namely PS-683, PCC 9005, PS-724, PS-680, PS-715, PS-718 and WH8020.

DISCUSSION

The tree topology obtained from phylogenetic analysis of cyanobacterial 16S rDNA sequences was in agreement with that described by Honda *et al.* (1999) and supported the idea that members of the genus *Synechococcus* are polyphyletic in origin and should be divided into several different genera. For simplicity, the same numbering system for cyanobacterial phylogenetic lineages proposed previously (Honda *et al.*, 1999) has been adopted. However, the addition of an eighth lineage is proposed (Fig. 2).

It was found that 16S rDNA sequences from two *Synechococcus* strains examined during this study, PCC 6716 and PCC 6717, clustered with '*Synechococcus elongatus*' (Toray) (Fig. 2). This result indicates that the '*Synechococcus* cluster' is not monophyletic as stated in previous studies (Honda *et al.*, 1999; Urbach *et al.*, 1998), since PCC 6716 and PCC 6717 are classified within this cluster along with PCC 6301 and others. Furthermore, *Bergey's Manual of Systematic Bacteriology* states that PCC 6716 and PCC 6717 are probably members of the same species as PCC 6301 and numerous other strains (Waterbury & Rippka, 1989). This is clearly not the case, and PCC 6716 and PCC 6717 should be classified in a different genus from that of PCC 6301. The PCC clustering system does separate these two strains (PCC 6716 and PCC 6717) correctly from other members of the '*Synechococcus* cluster' by placing them in cluster group 2 with other '*Synechococcus* cluster' members placed in PCC clusters 1 and 3. This separation better reflects the relationships identified here. Analysis of the *cpcBA*-IGS region reinforced the observation that PCC 6717, PCC 6716 and '*Synechococcus elongatus*' (Toray) are closely related to one another and clearly separated from other *Synechococcus* species where they formed the deepest branching group (group G; Fig. 3).

'*Synechococcus elongatus*' (Toray), PCC 6716 and PCC 6717 appear to be a very early branch of the cyanobacterial phylum. All three strains were isolated from hot springs in either Japan (Toray strain from Beppu, Kyushu) or North America (PCC 6716 from Hunter's Hot Spring, OR, and PCC 6717 from Fairy Springs, Yellowstone National Park). Further studies on these three strains may be particularly useful in unravelling the early evolution of cyanobacteria and their relationship with other bacteria. This cluster may represent one of the more ancient cyanobacterial forms and supports the idea that the *Synechococcus*-like morphology represents an early morphological type of cyanobacteria, from which other more complex morphologies have evolved. Such a hypothesis is further supported by the presence of this morphology type in five of the eight lineages presented here.

16S rDNA group 6 was the most complex in terms of *Synechococcus* phylogeny. Our analysis showed a clear division into three groups (a–c), which were well supported by bootstrap analysis (Fig. 2). Group c included three members from the *Synechococcus* cluster (and PCC group 1); PCC 6301, PCC 7942 and PCC 7943. The group also included sequences from two *Microcystis* species, *Microcystis elabens* NIES 42 and *Microcystis holsatica* NIES 43 (Neilan *et al.*, 1997), which were isolated from Lake Kasumigaura, Japan, in 1974 (Watanabe & Hiroki, 1997). The position of these *Microcystis* species is unclear, with fairly low percentage sequence similarities to the *Synechococcus* strains in this group, as well as to each other, as shown by the long branches on which they are found. These two *Microcystis* species were classified according to the system of Geitler (1932) in that, when isolated, they formed loose aggregates held together by amorphous capsular material, which is characteristic of natural populations of *Microcystis*. They do not, however, possess gas vacuoles. Subsequently, Stanier *et al.* (1971) proposed that only *Microcystis* spp. with gas vacuoles should be retained in this genus. Under this classification system, NIES 42 and NIES 43 would

not remain as *Microcystis* species, a view strongly supported by 16S rDNA analysis presented here and by others (Honda *et al*., 1999; Neilan *et al*., 1997). The same three *Synechococcus* strains from group 6c made up group F in the *cpcBA*-IGS analysis. Group F was clearly separated from all other species on a long branch, indicating a high level of sequence divergence relative to the other groups (Fig. 3). The three isolates PCC 6301, PCC 7942 and PCC 7943 probably represent members of the same species, as indicated by earlier DNA–DNA hybridization studies (Wilmotte & Stam, 1984).

Group 6a included three members that were not included in the previous study of Honda *et al*. (1999) and was supported by 100% bootstrap values in both distance and parsimony analysis (Fig. 2). These three isolates originated from saline lakes in the Vestfold Hills of eastern Antarctica (Rankin, 1998). A recent phylogenetic study of these organisms showed that they formed a separate cluster from a number of Arctic *Synechococcus* isolates, also from saline lakes (including P211 and P212), and *Synechococcus* strains from Marine clusters A and B (PCC 7805, WH8103 and WH8101) (Vincent *et al*., 1999). This suggests that these isolates in group 6a represent an additional lineage separate from other known *Synechococcus* species isolated from saline environments.

The third group, 6b, contained 15 *Synechococcus* strains from at least two different cluster groups (Marine cluster B and the *Cyanobium* cluster). This group was supported by 99 and 83% bootstrap values in distance and parsimony analysis, respectively, and all sequences in this cluster had sequence similarities of over 96·2% (Fig. 2). It is predicted, from examination of previous studies, that Marine cluster A species PCC 7805 and WH8103 would also fall in this group, as would *Prochlorococcus marinus* (Honda *et al*., 1999; Vincent *et al*., 1999).

The utility of the 16S rDNA sequence was limited in delineation of subgroups with group 6b. The *cpcBA*-IGS analysis, however, yielded useful information in this regard. The separation of *cpcBA*-IGS group A supports an expanded *Cyanobium* cluster division including PCC 7918 and PCC 7920, as well as the previously classified PCC 7009 and PCC 6904 (Fig. 3). This clustering was present, but not clearly defined in the 16S rDNA analysis (Fig. 2). PS-715 presents an anomaly in this case, as it clustered with the *Cyanobium* group in the 16S rDNA analysis (Fig. 2) but remained ungrouped in the *cpcBA*-IGS analysis (Fig. 3).

In summary, *Synechococcus* isolates found outside 16S rDNA group 6 should be considered as misclassified strains and renamed accordingly. These included PCC 7902 found in group 4, PCC 73109, PCC 7117, PCC 7002 and PCC 7003 located in group 5 and PCC 7335 in group 7. *Synechococcus* strains making up group 8 are also clearly a separate genus and should be reclassified as such. Our results indicated the following divisions within group 6, which contained the majority of *Synechococcus* isolates. At the highest level, three divisions exist, represented by groups 6a, 6b and 6c. These divisions are likely to be at the generic level. Group 6b contained numerous isolates and warrants division into a number of groups, probably at the species level. Analysis of the *cpcBA*-IGS fragment for this purpose is a useful tool. Our results revealed a good correlation with 16S rDNA data presented here and with some of the *Synechococcus* and PCC strain clusters. Analysis of sequences from additional representatives of marine isolates is required to discuss properly the status of these organisms in relation to group 6b as a whole.

One striking feature of the analysis of the *cpcBA*-IGS and flanking regions was that each group contained isolates with either PC-type or PE-type absorption spectra (Table 1, Fig. 3). This indicated that there is an evolutionary relationship between the primary DNA sequence of the PC operon and the possession or expression of PE. This result differs from that reported by Otsuka *et al*. (1998), who saw no relationship between 16S rDNA similarity and phycobilin pigment composition in a group of *Microcystis* species. A strong relationship was also observed between the length of the IGS sequences and phylogenetic groups based on analysis of *cpcBA*-IGS and flanking regions (Table 1, Fig. 3).

The phylogenetic analyses presented here, based on DNA sequences from a large number of *Synechococcus* isolates for both 16S rDNA and *cpcBA*-IGS and flanking regions, reinforce the notion of *Synechococcus* as an unnatural taxon with polyphyletic origins and show the utility of the PC locus for examining closely related species within this group of cyanobacteria. Clearly, the genus *Synechococcus* should be revised extensively, taking into consideration that the phylogenetic diversity seen among this group as it stands, as inferred by 16S rDNA sequences, may be almost as wide as it is for all cyanobacteria. In our analysis, five of eight groups representing the major cyanobacterial lineages contained *Synechococcus* isolates. It is suggested that the *Synechococcus*-like morphology may represent an ancestral morphology type from which other morphologies have evolved.

ACKNOWLEDGEMENTS

The authors are grateful to Brett Neilan (The University of New South Wales, Australia) for his comments on the manuscript in preparation. This work was supported by the Science and Technology Agency, Japan, through a Research Fellowship to B. R. R.

REFERENCES

Barker, G. L. A., Hayes, P. K., O'Mahony, S. L., Vacharapiyasophon, P. & Walsby, A. E. (1999). A molecular and phenotypic analysis of *Nodularia* (cyanobacteria) from the Baltic Sea. *J Phycol* **35**, 931–937.

Bolch, C. J. S., Blackburn, S. I., Neilan, B. A. & Grewe, P. M. (1996). Genetic characterisation of strains of cyanobacteria using PCR–

RFLP of the *cpcBA* intergenic spacer and flanking regions. *J Phycol* **32**, 445–451.

Carmichael, W. W. (1992). Cyanobacteria secondary metabolites – the cyanotoxins. *J Appl Bacteriol* **72**, 445–459.

Castenholz, R. W. & Waterbury, J. B. (1989). Group I Cyanobacteria. In *Bergey's Manual of Systematic Bacteriology*, pp. 1710–1728. Edited by J. T. Staley, M. P. Bryant, N. Pfennig & J. G. Holt. Baltimore: Williams & Wilkins.

Chisholm, S. W., Olsen, R. J., Zettler, E. R., Goericke, R., Waterbury, J. B. & Welschmeyer, N. A. (1988). A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**, 340–343.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.

Felsenstein, J. (1989). PHYLIP – Phylogeny inference package (version 3.2). *Cladistics* **5**, 164–166.

Geitler, L. (1932). *Cyanophyceae*. Reprinted 1971, Johnson, New York edn. Leipzig: Akad Verslagsses.

Giovannoni, S. J., Turner, S., Olsen, G. J., Barns, S., Lane, D. J. & Pace, N. R. (1988). Evolutionary relationships among cyanobacteria and green chloroplasts. *J Bacteriol* **170**, 3584–3592.

Hayes, P. K. & Barker, G. L. A. (1997). Genetic diversity within Baltic Sea populations of *Nodularia* (cyanobacteria). *J Phycol* **33**, 919–923.

Herdman, M., Janvier, M., Waterbury, J. B., Rippka, R., Stanier, R. Y. & Mandel, M. (1979). Deoxyribonucleic acid base composition of cyanobacteria. *J Gen Microbiol* **111**, 63–71.

Holt, J. G., Krieg, N. R., Sneath, P. H. A., Staley, J. T. & Williams, S. T. (1994). Group 11. Oxygenic phototrophic bacteria. In *Bergey's Manual of Determinative Bacteriology*, 9th edn, pp. 377–425. Edited by J. G. Holt. Baltimore: Williams & Wilkins.

Honda, D., Yokota, A. & Sugiyama, J. (1999). Detection of seven major evolutionary lineages in Cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains. *J Mol Evol* **48**, 723–739.

Jukes, T. H. & Cantor, C. R. (1969). Evolution of protein molecules. In *Mammalian Protein Metabolism*, vol. 3, pp. 21–132. Edited by H. N. Munro. New York: Academic Press.

Maidak, B. L., Cole, J. R., Parker, J. & 11 other authors (1999). A new version of the RDP (Ribosomal Database Project). *Nucleic Acids Res* **27**, 171–173.

Neilan, B. A., Jacobs, D. & Goodman, A. E. (1995). Genetic diversity and phylogeny of toxic cyanobacteria determined by DNA polymorphisms within the phycocyanin locus. *Appl Environ Microbiol* **61**, 3875–3883.

Neilan, B. A., Jacobs, D., Del Dot, T., Blackall, L. L., Hawkins, P. R., Cox, P. T. & Goodman, A. E. (1997). rRNA sequences and evolutionary relationships among toxic and nontoxic cyanobacteria of the genus *Microcystis*. *Int J Syst Bacteriol* **47**, 693–697.

Otsuka, S., Suda, S., Li, R., Watanabe, M., Oyaizu, H., Matsumoto, S. & Watanabe, M. M. (1998). 16S rDNA sequences and phylogenetic analyses of *Microcystis* strains with and without phycoerythrin. *FEMS Microbiol Lett* **164**, 119–124.

Rankin, L. M. (1998). *The ecology and taxonomy of Synechococcus from saltwater lakes in the Vestfold Hills, Antarctica*. PhD thesis, University of Tasmania, Australia.

Rippka, R. & Herdman, M. (1992). *Pasteur Culture Collection of Cyanobacterial Strains in Axenic Culture*. Paris: Institut Pasteur.

Rippka, R., Deruelles, J., Waterbury, J. B., Herdman, M. & Stanier, R. Y. (1979). Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* **111**, 1–61.

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

Stanier, R. Y., Kunisawa, R., Mandel, M. & Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol Rev* **35**, 171–205.

Suda, S., Liu, Y., He, J., Hu, Z., Hiroki, M. & Watanabe, M. M. (1998). Morphological, biochemical and physiological characteristics of *Lyngbya hieronymusii* var. *hieronymusii* (Oscillatoriales, Cyanobacteria). *Phycol Res* **46** (Suppl.), 51–55.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tillett, D. & Neilan, B. A. (2000). Xanthogenate nucleic acid isolation from cultured and environmental cyanobacteria. *J Phycol* **36**, 251–258.

Turner, S., Pryer, K. M., Miao, V. P. W. & Palmer, J. D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* **46**, 327–338.

Urbach, E., Scanlan, D. J., Distel, D. L., Waterbury, J. B. & Chisholm, S. W. (1998). Rapid diversification of marine picophytoplankton with dissimilar light-harvesting structures inferred from sequences of *Prochlorococcus* and *Synechococcus* (Cyanobacteria). *J Mol Evol* **46**, 188–201.

Vincent, W. F., Bowman, J. P., Rankin, L. M. & McMeekin, T. A. (1999). Phylogenetic diversity of picocyanobacteria in Arctic and Antarctic ecosystems. In *Microbial Biosystems: New Frontiers. Proceedings of the 8th International Symposium on Microbial Ecology*. Edited by C. R. Bell, M. Brylinsky & P. Johnson-Green. Halifax, Nova Scotia: Atlantic Canada Society for Microbial Ecology.

Watanabe, M. M. & Hiroki, M. (1997). *NIES Collection. List of Strains, Microalgae and Protozoa*. Japan: National Institute for Environmental Studies and Environment Agency.

Waterbury, J. B. & Rippka, R. (1989). Subsection I. Order Chroococcales Wettstein 1924, emend. Rippka *et al.* 1979. In *Bergey's Manual of Systematic Bacteriology*, pp. 1728–1746. Edited by J. T. Staley, M. P. Bryant, N. Pfennig & J. G. Holt. Baltimore: Williams & Wilkins.

Waterbury, J. B., Watson, S. W., Guillard, R. R. L. & Brand, L. E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* **277**, 293–294.

Weisburg, W. G., Oyaizu, Y., Oyaizu, H. & Woese, C. R. (1985). Natural relationship between *Bacteroides* and *Flavobacteria*. *J Bacteriol* **164**, 230–236.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**, 697–703.

Wilmotte, A. (1994). Molecular evolution and taxonomy of the cyanobacteria. In *The Molecular Biology of Cyanobacteria*, pp. 1–25. Edited by D. A. Bryant. Dordrecht: Kluwer Academic Publishers.

Wilmotte, A. M. R. & Stam, W. T. (1984). Genetic relationships among cyanobacterial strains originally designated as '*Anacystis nidulans*' and some other *Synechococcus* strains. *J Gen Microbiol* **130**, 2737–2740.