J. Eukaryot. Microbiol., 46(4), 1999 pp. 327–338

Investigating Deep Phylogenetic Relationships among Cyanobacteria and Plastids by Small Subunit rRNA Sequence Analysis[1]

SEÁN TURNER[a], KATHLEEN M. PRYER[b], VIVIAN P. W. MIAO[c] and JEFFREY D. PALMER[a]

[a]*Department of Biology, Indiana University, Bloomington, Indiana 47405, USA, and*
[b]*Department of Botany, The Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, USA, and*
[c]*TerraGen Diversity Inc., 2386 East Mall, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z3*

ABSTRACT. Small subunit rRNA sequence data were generated for 27 strains of cyanobacteria and incorporated into a phylogenetic analysis of 1,377 aligned sequence positions from a diverse sampling of 53 cyanobacteria and 10 photosynthetic plastids. Tree inference was carried out using a maximum likelihood method with correction for site-to-site variation in evolutionary rate. Confidence in the inferred phylogenetic relationships was determined by construction of a majority-rule consensus tree based on alternative topologies not considered to be statistically significantly different from the optimal tree. The results are in agreement with earlier studies in the assignment of individual taxa to specific sequence groups. Several relationships not previously noted among sequence groups are indicated, whereas other relationships previously supported are contradicted. All plastids cluster as a strongly supported monophyletic group arising near the root of the cyanobacterial line of descent.

Key Words. Kishino-Hasegawa test, maximum likelihood, relative likelihood support (RLS) scores.

IT is now widely accepted that plastids are derived from one or more endosymbiotic events involving the uptake of an oxygenic photosynthetic bacterium of the cyanobacterial line of descent by one or more nonphotosynthetic eukaryotic hosts (Delwiche and Palmer 1997; Douglas 1994; Gray 1992; Loiseaux-de Göer 1994). Previous phylogenetic studies using small subunit rRNA sequence data suggest that extant cyanobacteria, oxychlorobacteria (prochlorophytes), and plastids are the products of an explosive evolutionary radiation, the deep-branching order of which has been poorly resolved in published analyses (Douglas and Turner 1991; Ferris et al. 1996; Garcia-Pichel, Nübel and Muyzer 1998; Ishida, Yokota and Sugiyama 1997; Kane et al. 1997; Nelissen et al. 1994; Nelissen et al. 1995; Nelissen et al. 1996; Reeves 1996; Wilmotte 1994; Wilmotte, Van der Auwera and De Wachter 1993; Wilmotte et al. 1992). On the other hand, several relationships among these taxa have been established with fairly strong support. Plastids consistently form a monophyletic group with strong support and bear no relation to any of the three chlorophyll *b*-containing oxychlorobacteria, which are themselves not closely related to one another. *Gloeobacter violaceus* appears to be the deepest-branching line of descent among the cyanobacteria (Nelissen et al. 1995; but see Ferris et al. 1996). This result is supported by maximum likelihood analyses of sequences for *tuf*A genes (Delwich, Kuhsel and Palmer 1995; Köhler et al. 1997) and by distance analyses of group I introns in $tRNA^{fMet}$ (Paquin, et al. 1997). However, it should be noted that these other studies included far fewer cyanobacterial taxa than did analyses of SSU rRNAs. The deep placement of *G. violaceus* is consistent with other, nonmolecular data, including a lack of photosynthetic thylakoids (Rippka, Waterbury and Cohen-Bazire 1974), an unusual ultrastructural arrangement of phycobiliproteins (Guglielmi, Cohen-Bazire and Bryant 1981), and the absence of sulfoquinovosyl diacylglycerol, a lipid that is otherwise found in the photosynthetic membranes of cyanobacteria and plastids (Selstam and Campbell 1996).

In an effort to understand better the relationships among the deep-branching lines of descent that comprise the cyanobacterial radiation, particularly in the context of plastid phylogeny, we have determined partial and nearly complete sequences of 27 cyanobacterial small subunit rRNA genes (SSU rRNAs) and used these in conjunction with other SSU rRNA sequences in phylogenetic analyses employing maximum likelihood tree inference methods.

MATERIALS AND METHODS

Gene amplification and sequence determination. A list of the organisms for which SSU rRNA sequences were determined in this study as well as sequences from other sources is given in Table 1. Cell pellets of nostocalean, pleurocapsalean, and oscillatorialean strains of cyanobacteria were broken up using a ground glass homogenizer prior to DNA extraction. Genomic DNA was extracted by two methods. The first was a variation of the small-scale preparation method of Porter (1988) modified to include a polysaccharide precipitation step as described by Wilmotte et al. (1992). In the second method, DNA was isolated with the Puregene kit (Gentra Systems, Minneapolis, MN) following the manufacturer's protocol.

SSU rRNA genes were amplified and sequenced by the polymerase chain reaction (PCR) in two ways. In the first method, approximately 1.5 kb of the SSU rRNA gene was amplified from genomic DNA by PCR using primers 8F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-GGTTACCTTGTTACGACTT-3′). Amplifications were carried out in 50 μL reactions under standard conditions on a PTC 200 DNA Engine thermal cycler (MJ Research, Watertown, MA). The reaction mixture typically contained 1.0 U of *Taq* Polymerase and 10× PCR buffer (Gibco BRL Life Technologies, Grand Island, NY), 0.04 mM of each deoxynucleotide, 600 nM of each amplification primer, ca. 50 ng of genomic template DNA, and purified water to volume. Temperature and cycling conditions were as follows: one 95° C denaturation cycle for 3 min, followed by 35 cycles of 95° C denaturation for 15 s, primer annealing at 49° C for 15 s, and elongation at 42° C for 90 s. Four μL of the amplified products were visualized on 1.5% agarose minigels and the remainder was purified using 30,000 NMWL low-binding, regenerated cellulose membrane filter units (Millipore). Agarose plugs were sometimes taken of weak PCR products and reamplified at 51° C using the same conditions. Both strands of purified PCR products were directly sequenced in 10 μL reactions using the following primers: 8F, 357F (5′-CTCCTACGGGAGGCAGCAG-3′), 515F (5′-GTGCCAGCMGCCGCGGTAA-3′), 906F (5′-GAAACTTAAAKGAATTG-3′), 1100F (5′-GGCAACGAGCGMGACCC-3′), 1237F (5′-GGGCTACACACGYGCWAC-3′), 338R (5′-CTGCTGCCTCCCGTAGGAGT-3′), 519R (5′-GWATTACCGCGGCKGCTG-3′), 907R (5′-CCGTCAATTCCTTTRAGTTT-3′), 1100R (5′-AGGGTTGCGCTCGTTG-3′), 1391R (5′-GACGGGCGGTGTGTRCA-3′), and 1492R. Cycle sequencing was conducted using

Corresponding Author: Seán Turner—Telephone number: 812-855-4234; FAX number: 812-855-6705; email: sturner@bio.indiana.edu

[1] Invited presentation delivered during a joint meeting of the Society of Protozoologists, American Phycological Society, and the International Society of Evolutionary Protistology, Flagstaff, AR, August 7, 1998.

Table 1. List of taxa, SSU rRNA sequence accession numbers, and references and sources used.

Taxon	Accession number[a]	Reference & (source)[b]
CYANOBACTERIA[c]		
Section I (Chroococcales)		
Cyanobacterium stanieri PCC 7202^T	AF132782	this work (JW)
Cyanobium gracile PCC 6307^T	AF001477	(Urbach et al. 1998)
Cyanothece PCC 7424*	AF132932	this work (JW)
Cyanothece ATCC 51142	AF132771	this work (LS)
Dactylococcopsis salina PCC 8305^T	AJ000711	(Garcia-Pichel et al. 1998)
Gloeobacter violaceus PCC 7421^T	AF132790	this work (JW & PCC)
[*Gloeobacter* PCC 8105]	AF132792	this work (PCC)
Gloeocapsa PCC 73106*	AF132784	this work (JW)
Gloeothece membranacea PCC 6501^T	X78680	(Nelissen et al. 1995)
Microcystis aeruginosa PCC 7941^T	U40340	(Neilan et al. 1997)
Synechococcus elongatus PCC 6301^T (formerly *Anacystic nidulans*)	AF132776, Z82780, K01982	this work (JW; Rudi et al. 1997; Tomioka et al. 1981)
[*Synechococcus* PCC 7942]	AF132930	this work (SG)
Synechococcus PCC 7002*	AJ000716	(Garcia-Pichel et al. 1998)
Synechococcus "*elongatus*"	D83715	n.a.[d]
Synechococcus lividus C1 (= ATCC 700243)	AF132772	this work (DW)
Synechococcus C9 (= ATCC 700244)	AF132773	this work (DW)
Synechococcus P1 (= ATCC 700245)	AF132774	this work (DW)
[*Synechococcus* P2 (= ATCC 700246)]	AF132775	this work (DW)
Synechococcus WH 8101	AF001480	(Urbach et al. 1998)
Synechocystis PCC 6803	D64000	(Kaneko et al. 1995)
Section II (Pleurocapsales)		
Pleurocapsa PCC 7516*	X78681	(Nelissen et al. 1995)
Stanieria cyanosphaera PCC 7437^T	AF132931	this work (PCC)
Xenococcus PCC 7305*	AF132783	this work (PCC)
Section III (Oscillatoriales)		
Arthrospira PCC 8005	X70769	(Nelissen et al. 1994)
Geitlerinema PCC 7105*	AF132780	this work (PCC)
Leptolyngbya PCC 73110	AF132785	this work (PCC)
	X84810	(Nelissen et al. 1996)
["*Plectonema boryanum*" UTEX 485]	AF132793	this work (MC)
Leptolyngbya PCC 7375*	AF132786	this work (PCC)
Leptolyngbya VRUC 135	X84809	(Nelissen et al. 1996)
Lyngbya aestuarii PCC 7419^T	AJ000714	(Garcia-Pichel et al. 1998)
Microcoleus chthonoplastes PCC 7420^T	X70770	(Nelissen et al. 1995)
"*Oscillatoria rosea*" IAM M-220	AB003164	(Ishida 1997)
"*Oscillatoria neglecta*" IAM M-82	AB003168	(Ishida 1997)
Oscillatoria sancta PCC 7515^T	AF132933	this work (PCC)
Oscillatoria cf. *corallinae* CJ1 (= SAG 8.92)	X84812	(Nelissen et al. 1996)
"*Phormidium mucicola*" IAM M-221	AB003165	(Ishida et al. 1997)
Planktothrix prolifica NIVA-CYA 18 (formerly *Oscillatoria aghardii*)	X84811, Y12679	(Nelissen et al. 1996; Rudi et al. 1997)
"*Plectonema*" F3	AF091110	(Turner 1997)
Pseudanabaena PCC 6903	AF132778	this work (PCC)
[*Pseudanabaena* PCC 7402]	AF132787	this work (PCC)
Pseudanabaena PCC 7367*	AF091108	(Turner 1997)
Pseudanabaena PCC 7409	A. Wilmotte, pers comm.	(Nelissen et al. 1995)
Spirulina major PCC 6313^T	X75045	(Nelissen et al. 1994)
Sprulina subsalsa IAM M-223	AB003166	(Ishida et al. 1997)
Trichodesmium NIBB 1067	X70767	(Wilmotte et al. 1994)
Section IV (Nostocales)		
Anabaena NIES 19	AF091150	n.a.
(= *A. cylindrica* PCC 7122^T)	RDP	(Giovannoni et al. 1988)
Calothrix desertica PCC 7102^T	AF132779	this work (PCC)
Cylindrospermum ATCC 29204 (= *C. stagnale* PCC 7417^T)	AF132789	this work (ATCC)
Nodularia BCNOD9427	AJ224447	n.a.
Nostoc punctiforme PCC 73102^T	AF027655	(Miao et al. 1997)
Nostoc PCC 7120 (formerly *Anabaena* PCC 7120)	X59559	(Ligon et al. 1991)
Scytonema hofmanni PCC 7110^T	AF132781	this work (JW)
Section V (Stigonemateles)		
Chlorogloeopsis fritschii PCC 6718	AF132777	this work (JW)
Chlorogloeopsis PCC 7518	X68780	(Wilmotte et al. 1993)
Fischerella muscicola PCC 7414^T	AF132788	this work (PCC)

Table 1. Cont.

Taxon	Accession number[a]	Reference & (source)[b]
OXYCHLOROBACTERIA (PROCHLOROPHYTA)[c]		
Prochlorococcus MIT 9312	AF053398	(Moore et al. 1998)
Prochloron didemni	X63141	(Urbach et al. 1992)
Prochlorothrix hollandica	AF132792	this work (GB)
PHOTOSYNTHETIC PLASTIDS[c]		
Chlorophyta		
Chlorella vulgaris C-27	AB001684	(Wakasugi et al. 1997)
Klebsormidium flaccidum SAG 335-2b	X75522	n.a.
Embryophyta		
Marchantia polymorpha	X04465	(Ohyama et al. 1986)
Glaucocystophyta (Glaucophyta)		
Cyanophora paradoxa UTEX 555	U30821	(Stirewalt et al. 1995)
Glaucocystis nostochinearum SAG 45.88	X82496	(Helmchen et al. 1995)
Rhodophyta		
Porphyra purpurea str. Avonport	U38804	Reith & Munholland 1995)
Cryptophyta		
Guillardia theta (formerly *Cryptomonas* str. Φ)	X56806	(Douglas & Turner 1991)
Haptophyta (Prymnesiophyta)		
Emiliania huxleyi PML 92D	X82156	(Medlin et al. 1995)
Bacillariophyta		
Odontella sinensis	Z67753	(Kowallik et al. 1995)
Phaeophyta		
Pylaiella littoralis	X14873	(Markowicz et al. 1988)

[a] GenBank/EMBL/DDBJ electronic database accession numbers. RDP = Ribosomal Database Project (Maidak et al. 1996).

[b] Sources for materials used in this study: ATCC: culture from the American Type Culture Collection, Manassas, Virginia, U.S.A. DW: culture from David Ward, Montana State University, Boseman, Montana, U.S.A. GB: genomic DNA from George Bullerjahn, Bowling Green State University, Bowling Green, Ohio, USA. JW: culture from John Waterbury, Woods Hole Oceanographic Institute, Woods Hole, Massachusetts, USA. LS: genomic DNA from Louis Sherman, Purdue University, West Lafayette, Indiana, USA. MC: genomic DNA from Martina Celerin, Indiana University, Bloomington, Indiana, USA. PCC: culture from the Pasteur Culture Collection, Institut Pasteur, Paris, France. SG: genomic DNA from Susan Golden, Texas A&M University, College Station, Texas, USA.

[c] Taxonomic scheme according to Burger-Wiersma et al. (1989), Chisholm et al. (1992), Lewin (1989), Rippka and Herdman (1992), Waterbury and Rippke (1989). T = type strain. * = reference strain. Organisms with names in quotes are likely to have been misidentified. Strains with indented names in square brackets have SSU rRNA similarity values >99% relative to the strain immediately above them.

[d] n.a.: literature reference not available.

[e] Taxonomic scheme according to Lee (1989).

dRhodamine Dye Terminator reagents and a PE-ABI 377 automated DNA sequencer (Perkin Elmer-Applied Biosystems, Foster City, CA). Sequence fragments were edited and assembled into contigs using Sequencher 3.0 (Gene Codes).

In the second method, ca. 2 ng of genomic DNA were used for gene amplification using the primers 16S.0007.F21 and 16S.1511R.21 as described in Miao, Rabenau and Lee (1997). Amplification conditions were a 94° C denaturation cycle for 2 min, followed by 28 cycles of 94° C denaturation for 40 s, primer annealing at 55° C for 44 s, and elongation at 72° C for 2 min. A final 5 min extension step was carried out at 72° C. Reactions were processed using the Qiaquick PCRClean Up kit (Qiagen, Santa Clarita, CA) following the manufacturer's protocol. Direct sequencing of PCR products was carried out as in the first method using primers 16S.0007.F21, 16S0515.F16, 16S.1511.R21 (Miao et al. 1997), 16S.0538.R21 (5′-SCTATTACCGCGGCAGCTGGC-3′), and 16S.1100.F16 (5′-CAACGAGCGCAACCCT-3′). Sequence fragments were edited and assembled into contigs using AssemblyLign (Oxford Molecular Group, Beaverton, OR). Sequences obtained in this study have been assigned GenBank accession numbers AF132771–AF132793 and AF132930–AF132933.

Sequence alignment and phylogenetic analysis. SSU rRNA gene sequences corresponding to positions 27 to 1,491 of the *Escherichia coli* SSU rRNA sequence (accession J01859) were aligned manually and the alignment verified by secondary structure analysis of each sequence using that of *Synechococcus* PCC 6301 as template (Gutell 1993). The alignment contained a total of 1,448 positions, including sequence gaps, of which 1,377 were considered to be unambiguously aligned and used for subsequent phylogenetic analysis. Of these, 695 were "invariant" and 478 were "phylogenetically informative." Phylogenetic trees were inferred by a maximum likelihood method under the evolutionary model of Felsenstein (1981) modified to incorporate a user-defined transition/transversion (ti/tv) ratio. This modified model has been part of the PHYLIP package of programs since 1984 and is hereafter referred to as the F84 model (Felsenstein 1993). The initial ti/tv ratio (1.48) was estimated from the sequence data using the program PUZZLE (version 4.0) under the Tamura-Nei model of evolution with parameter estimation set at "approximate" (Strimmer, Goldman and von Haeseler 1997; Strimmer and von Haeseler 1996; Tamura and Nei 1993).

Initial maximum likelihood trees were inferred with the program fastDNAml (Olsen et al. 1994). Trees were inferred ten times with randomized order of sequence input using the "jumble" option and "local" branch-swapping only (equivalent to nearest-neighbor interchange). The three trees with the best log-

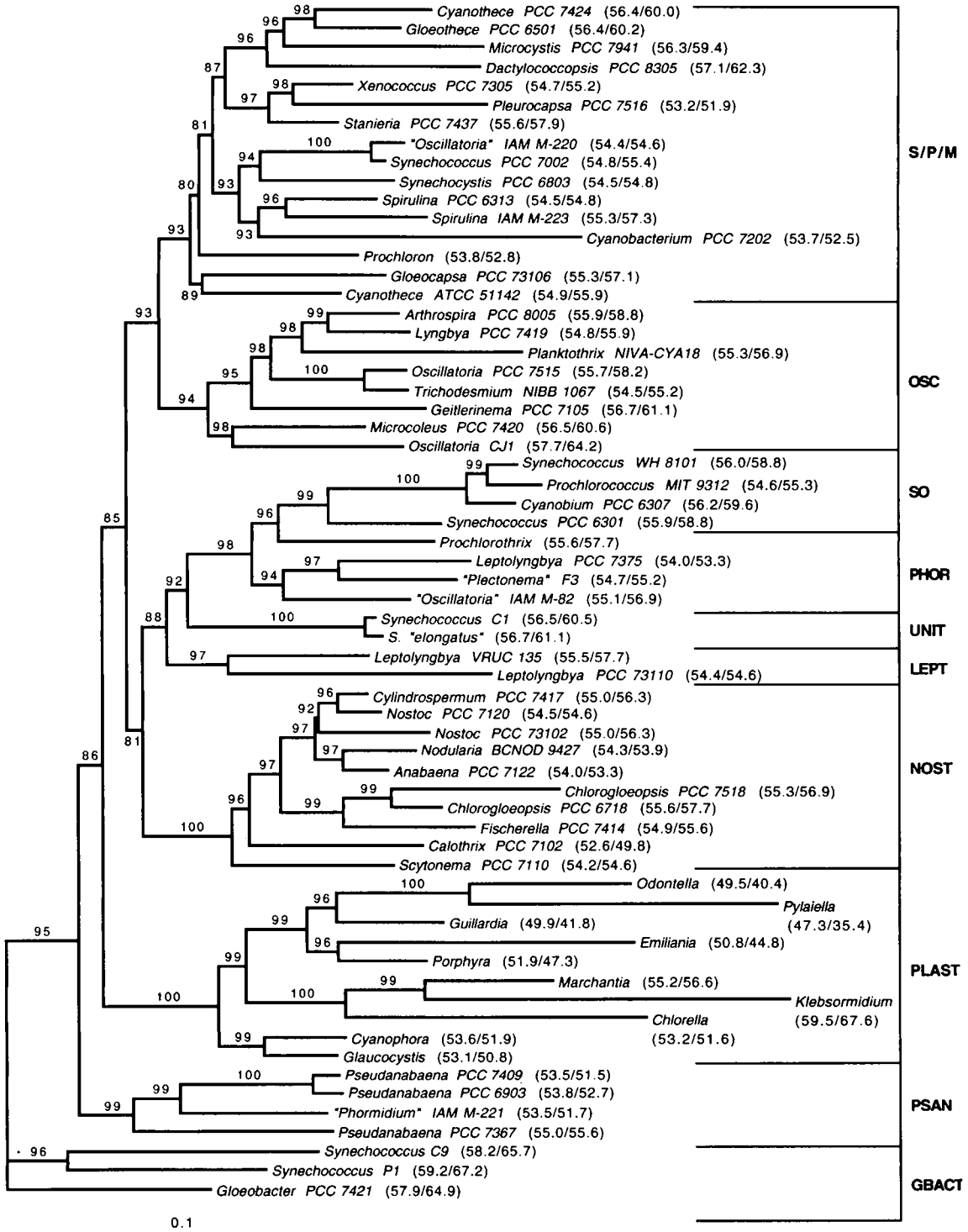

Fig. 1. Unrooted phylogenetic tree of 53 cyanobacterial and 10 plastid SSU rRNA sequences inferred by maximum likelihood analysis with correction for site-to-site variability of change. Lengths of horizontal branches correspond to the number of substitutions per sequence position as indicated by the scale bar. Specific epithets are provided only for those taxa the identity of which would otherwise be ambiguous (*S. "elongatus"*). Names in quotes designate strains likely to have been taxonomically misidentified. Full names are given in Table 1. Numbers in parentheses denote the (G + C)% of all positions utilized in the analysis (first value) or of phylogenetically informative positions only (second value). Numbers on branches represent relative likelihood support (RLS) scores rounded down to the nearest percentile. SSU rRNA sequence groups as defined by Turner (1997) are indicated by brackets on the right with abbreviations as follows: GBACT—*Gloeobacter* sequence group; LEPT—*Leptolyngbya* sequence group; NOST—*Nostoc* sequence group; OSC—*Oscillatoria* sequence group; PHOR—*Phormidium* sequence group; PSAN—*Pseudanabaena* sequence group; PLAST—"plastid" sequence group; SO—*Synechococcus* sequence group; S/P/M—*Synechocystis/Pleurocapsa/Microcystis* sequence group; UNIT—"unicellular thermophilic" sequence group. For further details, see text.

likelihood scores were then each used as a starting tree for further optimization using the "restart" option and "global" branch-swapping (equivalent to subtree-pruning-regrafting). The same optimal tree was found in all three cases.

Evolutionary models that incorporate variability in rates of change among sequence positions will often necessitate different values in certain parameters relative to models that do not (Wakeley 1994). Consequently, the estimated ti/tv ratio was reoptimized with the program PAUP* (version 4.0) (Swofford 1997) by inputting the maximum likelihood tree topology for estimation of parameters under the F84 evolutionary model assuming a discrete Gamma distribution with four categories of site-to-site variability of change. The resulting estimated ti/tv ratio (1.82) and the maximum likelihood tree were then submitted to DNArates, a program that computes a maximum likelihood estimate of the evolutionary rate of change for each sequence position and partitions the positions into, in this instance, 35 rate categories. "Invariant" positions are given a rate of change half that of positions that change only once on the input tree (Pracht, S., Overbeek, R. & Olsen, G. J., pers. commun.). Maximum likelihood trees were then inferred de novo as before using the "categories" option of fastDNAml with the aforementioned rate categories as well as the reoptimized ti/tv ratio. This allows for a more accurate estimation of branch lengths in any given tree, and hence a more accurate estimation of overall evolutionary distances among the sequences. The new optimal tree was then used as input for a second round of rates estimation and maximum likelihood trees inferred a third time in order to ensure that the topology of the optimal tree was stable. The programs fastDNAml and DNArates are available through the internet from the Ribosomal Database Project (http://www.cme.msu.edu/RDP).

Confidence in the topological features of the final maximum likelihood tree was determined by the computation of relative likelihood support (RLS) scores for the individual branches (Jermiin et al. 1997). The optimal maximum likelihood tree was used as a starting tree for global branch-swapping with rate categories as before and using the "keep" option of fastDNAml. The best 1,000 trees as determined by log-likelihood scores were retained and analyzed by the Kishino-Hasegawa (KH) test (Kishino and Hasegawa 1989). The resulting output was analyzed by the program TreeCons (version 1.0) in conjunction with the CONSENSE program of PHYLIP to compute a majority-rule consensus tree and associated RLS scores under a standardized, exponential weighting scheme at both the 5% and 1% significance levels for the KH test (Jermiin et al. 1997). Under this scheme ("Class V" in the TreeCons program) the contribution of an individual tree in the inference of the consensus tree is weighted in an exponential manner on the basis of its standardized difference in log likelihood compared to the optimal tree.

For display purposes, trees were graphically modified with the programs TreeView (version 1.5) (Page 1996) and MacClade (version 3.05) (Maddison and Maddison 1992). The various data analyses were run on Apple Power Macintosh 9600/300, Apple Power Macintosh 9600/200MP, and Sun SPARC server 1000 computers.

RESULTS

Of the 27 strains of cyanobacteria and oxychlorobacteria for which complete or partial sequence data were generated in this study, five were found to have similarity values greater than 99% relative to other strains (Table 1). These were *Gloeobacter* PCC 8105, *Pseudanabaena* PCC 7402, *Plectonema* UTEX 485, *Synechococcus* P2, and *Synechococcus* PCC 7942. As there is no gain in phylogenetic information by incorporating virtually duplicate sequences in tree inference, these were excluded from phylogenetic analyses.

For four strains, partial sequences generated in this study or others were found either to be in conflict at one or more positions with previously published sequences or to resolve ambiguous nucleotide assignments in them. For strains *Anabaena cylindrica* PCC 7122, *Leptolyngbya* PCC 73110, *Planktothrix prolifica* NIVA-CYA 18, and *Synechococcus elongatus* PCC 6301, signature nucleotide analysis and other considerations were used to construct consensus sequences for phylogenetic analysis.

A maximum likelihood analysis that included correction for site-to-site variability in rates of change resulted in the unrooted phylogenetic tree of SSU rRNA sequences from 53 cyanobacteria and 10 plastids shown in Fig. 1. A single pass with the DNArates program was sufficient to generate rate categories for the inference of an optimal tree with a stable topology. A second pass resulted in an optimal tree with an identical topology and slight improvement in log likelihood score due to further optimization of branch lengths.

Alternative topologies were examined by a global search using the first-pass optimal tree as a starting point for branch swapping. Of the 14,281 trees generated, the best 1,000 based on log likelihood values were kept for examination by the KH test. Of these, 522 were not rejected as being significantly different from the best tree at the 5% level of significance. The majority-rule consensus tree computed from an individually standardized and weighted contribution of these 523 best trees was identical in topology to the optimal tree. The relative likelihood support (RLS) scores for the internal branches of the consensus tree are shown in Fig. 1. When the significance level of the KH test was set at 1%, 883 alternative trees were not rejected when compared to the optimal tree. Inclusion of the additional trees in computation of the consensus tree resulted in an identical topology and RLS scores that differed from the previous values by no more than a single percentile (not shown). Consequently, one can conclude that computation of the consensus tree is not significantly affected by the exclusion of these additional trees.

All the RLS scores in Fig. 1 are greater than or equal to 80%. These RLS values are consistently much higher when compared to values for similar branches from bootstrap analysis (Nelissen et al. 1995; Wilmotte 1994) or quartet puzzling (Turner 1997). However, it is important to note that RLS scores are fundamentally different from these other values. The KH test is based on Neyman-Pearson hypothesis testing (Kishino and Hasegawa 1989). It addresses the question as to whether the original data can be said to support alternative hypotheses (tree topologies) at a level that is considered not significantly different in a statistical sense from the best hypothesis (optimal phylogenetic tree).

On the other hand, bootstrap analysis serves as a measure of the internal consistency of the sequence data. That is, it is a measure of the expectation of finding the same taxonomic groupings in subsequent phylogenetic trees if one were to obtain additional sequence data that were similar in nature to the original data and analyzed in the same manner (Felsenstein 1993). Alternatively, a consensus tree inferred from bootstrap replicates can be viewed as an average of multiple trees each independently inferred with randomized character weighting under the condition that the weights in each case be nonnegative integers that sum to the number of characters. It has been demonstrated both by computer simulation studies and on the basis of theoretical considerations that the bootstrap method as implemented in most available phylogenetics programs is biased toward the 50% value (Felsenstein and Kishino 1993; Hil-

lis and Bull 1993; Zharkikh and Li 1992). Hence, it will, in principle, consistently underestimate support values between 50% and 100%. Moreover, the degree of underestimation is correlated with the number of competing topologies (Zharkikh and Li 1995). Consequently, specific groupings in phylogenetic trees with relatively large numbers of taxa are liable to receive relatively low support with the bootstrap technique.

Reliability values derived from the quartet puzzling method are based on a consensus of equally weighted trees each of which is derived by a random addition of taxa with placement on a specific branch of the growing tree determined by a penalty score. This score is based on the consistency of all pertinent quartets of taxa involving the taxon to be added and the taxa already on the tree (Strimmer and von Haeseler 1996; Strimmer, Goldman and von Haeseler 1997). However, the resulting consensus tree is not guaranteed to be optimal (Cao, Adachi and Hasegawa 1998). The sequence data of this study were used to derive 1,000 quartet puzzled trees inferred under an F84 model of evolution and with site-to-site variability of change modeled under a discrete Gamma distribution with four rate categories. Examination of these trees by the KH test showed 885 to be significantly different from the optimal tree within this set at the 5% level of significance. Moreover, when the topology derived from the fastDNAml analysis with 35 rate categories was compared to the 1,000 puzzled trees under the same model by which the puzzled trees were inferred, all 1,000 were found to be significantly different from the optimal tree at the 0.1% significance level (not shown). In short, reliability values from quartet puzzling are based at least in part on trees that would be rejected from inclusion in the computation of RLS scores.

These facts and the different statistical considerations addressed by the KH test relative to bootstrapping and quartet puzzling in the context of consensus tree inference make the mutual comparison of these methods difficult. The statistical justification for the use of RLS scores to establish confidence in phylogenies is discussed in detail by Jermiin et al. (1997).

Figure 2 is a depiction of the tree in Fig. 1 in the form of a cladogram. As in previous analyses, the taxa fall into 10 monophyletic "sequence groups" (Turner 1997). Although the tree inferred in this study is unrooted, the root of the cyanobacterial line of descent likely falls on the branch leading to *Gloeobacter violaceus* PCC 7421 for reasons discussed in the Introduction. However, the lack of additional molecular and nonmolecular data for other strains in the *Gloeobacter* (GBACT) group makes the exact location of the root uncertain. For purposes of further discussion, it will be assumed that the root of the cyanobacterial tree lies either within the *Gloeobacter* (GBACT) group or on the branch leading to it.

Because computation of RLS scores involves only those alternative topologies that are not rejected by the KH test, higher values are expected when compared to bootstrapping or quartet puzzling. However, it must be noted that not all acceptable topologies are sampled by a heuristic search based on a branch-swapping algorithm and the large number of taxa in this study makes an exhaustive search impractical. Inclusion of a greater number of alternative topologies at the same level of significance might result in lower RLS scores. An additional consideration is that scores for branches within individual sequence groups may be artificially high due to taxonomic sampling bias. That is, incorrect evolutionary relationships can receive high support values on phylogenetic trees when analyses are limited to relatively small numbers of taxa for any given group (Adachi and Hasegawa 1996; Lecointre, Philippe and Le Guyader 1993; Philippe 1997; Philippe and Douzery 1994). Consequently, the branching order in the individual sequence groups is best studied independently using a larger number of ingroup taxa, and the support values shown in Fig. 1 for relationships within these groups should be interpreted with appropriate caution.

Besides these specific caveats, the usual ones concerning interpretation of molecular phylogenetic analyses should be kept in mind. The analyses presented here result in a gene tree that is not guaranteed to accurately reflect the relationships among the organisms themselves, and support values, regardless of how they are generated, are more a measure of the consistency of the analytical method used with the data in hand rather than a direct measure of the historical accuracy of evolutionary relationships implied by a given tree topology.

Regardless of methodology, the choice of support values that are considered by the investigator to establish confidence in topological features of phylogenetic trees is largely arbitrary. Such values can range from as conservative as 100% (strict consensus) to as liberal as 50% (majority-rule consensus) or can be any value between. The method employed here utilizes a selective criterion such that topologies that are considered significantly inferior to the optimal tree are excluded from the computation of support scores. This is in contrast to quartet puzzling and bootstrap resampling, as most commonly implemented, whereby all trees inferred from individual iterations are given equal weight in the construction of a majority-rule consensus tree regardless of their likelihood relative to the optimal tree. Consequently, a relatively high cutoff value is in order for interpretation of the RLS scores presented in this study. Thus, if those branches in Fig. 2 with RLS scores less than 90% are considered poorly supported, their truncation leads to the cladogram shown in Fig. 3A. The 10 individual sequence groups are retained as are several relationships among them. The *Oscillatoria* (OSC) sequence group is sister to the *Synechocystis/Pleurocapsa/Microcystis* (S/P/M) sequence group (RLS= 93%). Similarly, the supergroup comprised of the *Synechococcus* (SO) and *Phormidium* (PHOR) sequence groups is maintained, as is its association with the "unicellular thermophilic" (UNIT) sequence group as a sister clade (RLS = 92%). This placement of the "unicellular thermophilic" (UNIT) group is the only major change among sequence group relationships resulting from correction for site-to-site variability in the sequence data relative to results obtained in the absence of such correction. In the latter case, this group branches, artifactually deeply we presume, between the *Gloeobacter* (GBACT) and *Pseudanabaena* (PSAN) sequence groups (not shown).

Other topological features in Fig. 1, 2 do not hold up under the 90% rule, resulting in a radiation of six lines of descent from a common ancestral node (Fig. 3A). Particularly noteworthy is the placement of the *Pseudanabaena* (PSAN) and "plastid" (PLAST) sequence groups relative to their previous positions. In Fig. 1 and 2 these are the two deepest-branching lines of descent after the *Gloeobacter* (GBACT) sequence group but form part of the aforementioned radiation in Fig. 3A. That is, their particularly deep placement is not supported here in contrast to reports using other methods of sequence analysis and fewer cyanobacterial taxa (Nelissen et al. 1994)]. The other intergroup associations not supported under the 90% rule include the *Leptolyngbya* (LEPT) sequence group as sister clade to the *Synechococcus/Phormidium/* "unicellular thermophilic" (SO/PHOR/UNIT) supergroup, and this combined supergroup as sister clade to the *Nostoc* (NOST) sequence group. From this analysis, the evolutionary relationships among these six lines of descent are considered to be indeterminate and may or may not be accurately reflected in the branching order shown in Fig. 1. Moreover, a number of relationships within the S/P/M sequence group collapse under the 90% rule, consistent with the particularly weak support for several intragroup clades within this assemblage as found by quartet puzzling (Turner 1997).

Fig. 2. Phylogenetic tree of 53 cyanobacterial and 10 plastid small subunit (SSU) rRNA sequences (as in Fig. 1) represented as an unrooted cladogram. Dark circles are on those branches with relative likelihood support (RLS) scores less than 90%. For further details, see legend of Fig. 1.

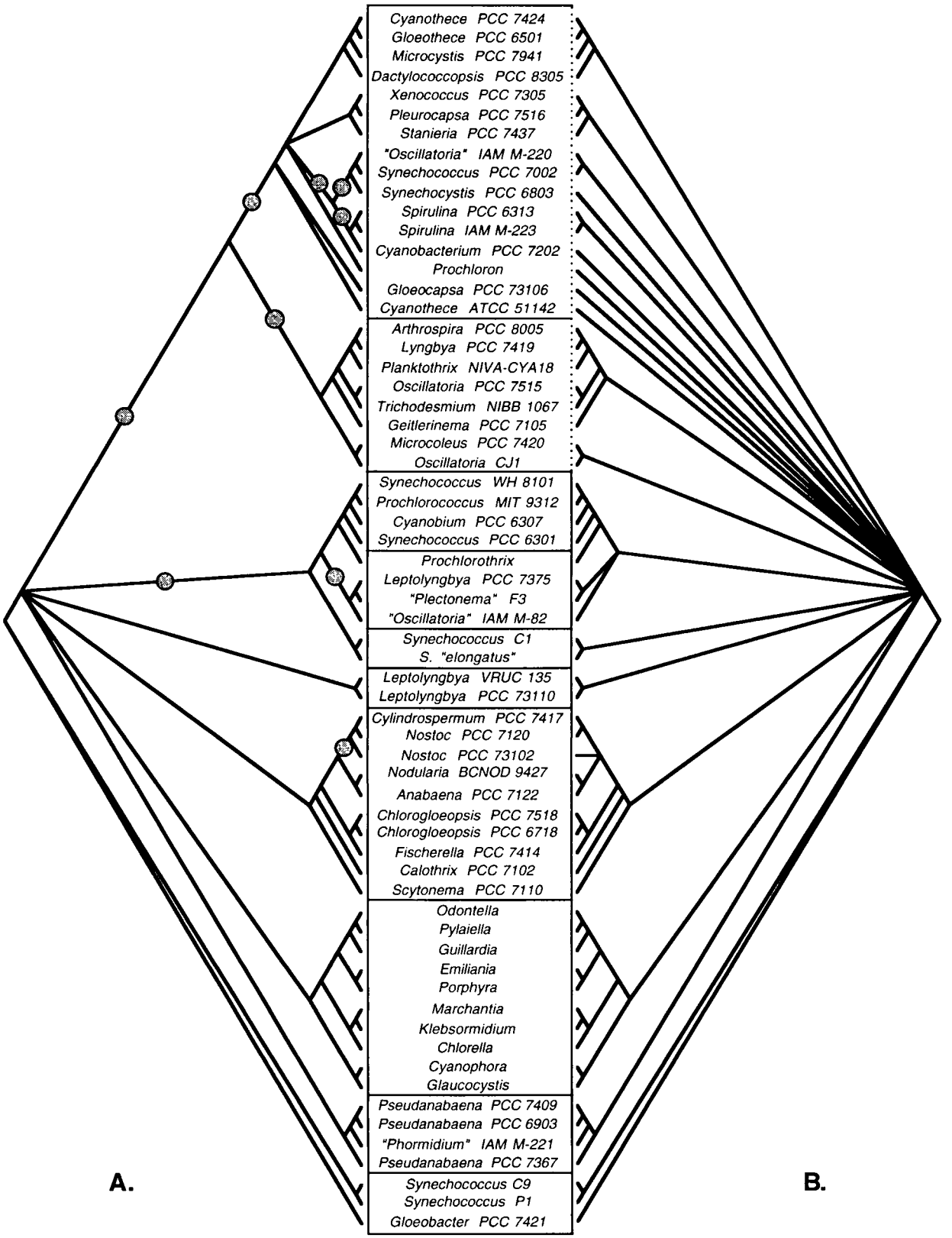

Fig. 3. Unrooted cladograms of 53 cyanobacterial and 10 plastid small subunit (SSU) rRNA sequences (as in Fig. 2) with branches truncated using particular relative likelihood support (RLS) scores as given thresholds. **A**. Truncation of branches with RLS < 90%. Light circles are on those branches such that 90% ≤ RLS < 95%. **B**. Truncation of branches with RLS < 95%. Loss of integrity for the OSC and S/P/M sequence groups is indicated by a broken line along their demarcating boxes. For further details, see legends of Fig. 1 and 2.

A more conservative interpretation would be to truncate those branches of Fig. 1 and 2 that receive less than 95% relative likelihood support (Fig. 3B). Under the 95% rule, all intergroup associations collapse with the exception of that involving members of the *Synechococcus* (SO) and *Phormidium* (PHOR) sequence groups. However, the individual integrity of these two groups is lost. Loss of structure can be seen to occur in the *Nostoc* (NOST) sequence group as well. Most striking is the loss of monophyly for the *Oscillatoria* (OSC) and S/P/M sequence groups. In the latter case, some individual strains are reduced to behaving as unique lines of descent radiating from an ancestral node shared by nearly all sequences in the tree.

The imposition of a 95% cutoff for RLS scores is likely to be overly conservative in this instance because there is independent evidence within the SSU rRNA sequence data supporting the monophyly of the S/P/M sequence group. In helix 45 of the secondary structure of all cyanobacterial SSU rRNAs, there is a three basepair minihelix upstream of three consecutive single-stranded adenosine nucleotides (Van de Peer et al. 1997). All members of the S/P/M sequence group, including those not shown in the phylogenetic trees presented here, have three nucleotides between these two features, while all other cyanobacterial and plastid sequences, with a single exception, have two nucleotides. These three positions were among those not included in the phylogenetic analysis due to the ambiguity inherent in aligning sequence gaps. Consequently, this single-base insertion, virtually unique to the members of the S/P/M group among cyanobacteria, can be considered an independent SSU rRNA character that supports the monophyly of the group inferred from analysis of the set of aligned SSU rRNA positions.

DISCUSSION

Previous analyses of SSU rRNA sequences from taxonomically diverse samplings of cyanobacteria generally have yielded phylogenetic trees with little support for most internal nodes of moderate or greater depth. In all these cases, however, phylogenetic tree inference was carried out using distance matrix and/or maximum parsimony methods, and in all cases branch support was estimated by bootstrap resampling (Douglas and Turner 1991; Ferris et al. 1996; Garcia-Pichel, Nübel and Muyzer 1998; Ishida, Yokota and Sugiyama 1997; Kane et al. 1997; Nelissen et al. 1994; Nelissen et al. 1995; Nelissen et al. 1996; Reeves 1996; Wilmotte 1994; Wilmotte et al. 1992; Wilmotte, et al. 1993; Wilmotte, Neefs and De Wachter 1994).

In this study, a maximum likelihood method of tree inference was used. While more computationally intensive than other methods, maximum likelihood has several advantages: 1) it is a character-based method (as is maximum parsimony) and so makes fuller use of information in the data than do distance methods; 2) it is more robust to the artifact of "long-branch attraction" than are other methods, including maximum parsimony (Hillis et al. 1994; Huelsenbeck 1998; Kuhner and Felsenstein 1994); 3) compared to some distance methods, it is more robust to violation of the assumption of evolutionary independence of characters (Tillier and Collins 1995). Furthermore, correction was made in this study for site-to-site variability in rates of change among sequence characters. Failure to account for such variability is one of the leading causes of "long-branch attraction" artifacts in phylogenetic tree inference (Olsen 1987).

The membership of individual taxa to particular sequence groups as seen here is in accordance with previous studies (see, for example, Turner 1997; Wilmotte 1994). However, the use of RLS scores in the determination of confidence levels reveals some intergroup relationships that previously had not received strong support by bootstrap analysis. The *Oscillatoria* (OSC) and S/P/M sequence groups appear to be sister clades, and the "unicellular thermophilic" (UNIT) sequence group becomes a basal member of the *Synechococcus/Phormidium* (SO/PHOR) supergroup.

Conversely, the deep branching points of the *Pseudanabaena* (PSAN) and "plastid" (PLAST) sequence groups found by others is not supported (Nelissen et al. 1995; Turner 1997). Rather, when a 90% support rule is invoked, they are interpreted as part of a radiation involving at least six lines of descent from an ancestral node near the root of the cyanobacterial tree (Fig. 3A). Although there appears to be some overall improvement in this study relative to others in elucidating the evolutionary relationships among cyanobacterial groups, the persistent occurrence of a radiation-like topology in cyanobacterial SSU rRNA trees may be an indication of an explosive evolutionary history of these taxa. Alternatively, SSU rRNA gene sequences may simply lack the information necessary to infer detailed relationships among deep-branching events. The use of other molecular markers is an obvious step to be taken in the investigation of this phenomenon.

Here, as in previous publications, the SSU rRNA sequences of photosynthetic plastids form a monophyletic group with strong support (Bhattacharya and Medlin 1995; Bhattacharya and Schmidt 1997; Douglas and Turner 1991; Nelissen et al. 1995). Because plastid genomes appear to be subject to mutation pressure leading to a decrease in guanine plus cytosine content ((G + C)%), it has been suggested that the monophyly of photosynthetic plastids in SSU rRNA trees is an artifact attributable to a convergence in base composition (Beanland and Howe 1992; Howe et al. 1992; Lockhart et al. 1992a; Lockhart et al. 1994). While compositional bias has been shown to lead to artifactual groupings *within* the plastid clade in some instances (Egea and Lang-Unnasch 1995; Lockhart et al. 1994; McFadden, Gilson and Waller 1995), no convincing evidence has yet been presented that the monophyly of plastids *as a group* in SSU rRNA trees is the product of compositional bias. Such an artifact is unlikely to be the case in this study. The range of (G + C)% for the set of SSU rRNA plastid sequences investigated here is greater than, and includes, that of the cyanobacterial sequence positions (47%–60% vs. 53%–59%, respectively). If the comparison is limited to phylogenetically informative sequence positions, the result is the same (35%–68% vs. 50%–67%, respectively) (Fig. 1). This fact is in diametric contradiction to the hypothesis of plastid sequence monophyly being attributable to convergence in (G + C)%. Rather, the plastid SSU rRNAs form a monophyletic group despite having a highly variable (G + C)%. In particular, the range of (G + C)% of the plastid sequences from glaucocystophytes (*Cyanophora paradoxa* and *Glaucocystis nostochinearum*) and the chlorophyll *b*-containing "green" plastids (*Chlorella vulgaris, Klebsormidium flaccidum,* and *Marchantia polymorpha*) is virtually indistinguishable from that of the cyanobacterial sequences (53%–60%, all positions; 52%–68%, phylogenetically informative positions). Among the plastid sequences incorporated into this study, low (G + C)% values are restricted to "red" plastids (*Porphyra purpurea, Guillardia theta, Emiliania huxleyi, Odontella sinensis,* and *Pylaiella littoralis*) (47%–52%, all positions; 35%–47%, phylogenetically informative positions) (Fig. 1).

An alternative hypothesis for apparent plastid monophyly would be that of "long-branch attraction" although the disparity in (G + C)% among the plastid sequences indirectly argues against this. That is, even though there is a greater apparent rate of change in plastid rRNA sequences relative to those of cyanobacteria, the differences in mutational bias among the plastids would imply that homoplasy among them would ac-

tually be decreased rather than increased overall. Moreover, the use of a maximum likelihood tree inference method in conjunction with correction for site-to-site variability in evolutionary rate is among the best available methodologies to guard against the occurrence of "long-branch attraction" artifacts. Finally, an additional check was carried out as follows. Plastid sequences were pruned from the final tree. The resulting cyanobacterial tree was then used with the "restart" option of fastDNAml and "global" branch-swapping to add subsets of the plastid sequence set in three ways: the glaucocystophytes alone, the "green" plastids alone, and the "red" plastids alone. In all three instances, the point of attachment of the plastid sequences in the optimal tree was the same as when all 10 plastid sequences were used in the analysis (Fig. 1). The fact that subsets of plastid sequences independently have the same branch point as when all the plastid sequences are used together argues against their common location being due to an artifactual mutual attraction. However, we cannot rule out the possibility that each of these three plastid lineages is independently artifactually attracted to the same location on the tree.

In previously inferred phylogenetic trees, a peculiarity pertaining to the position of the plastid clade within the cyanobacterial lineage is the relative deepness of the group (Bhattacharya and Medlin 1995; Bhattacharya and Schmidt 1997; Nelissen et al. 1995). The implication is that the lineage leading to photosynthetic organelles is older than most other cyanobacterial lines of descent. An interpretation based on a 90% RLS rule for the results presented here leads to the conclusion that while the "plastid" (PLAST) sequence group may not branch as deeply as in some previous analyses, it nonetheless composes one of six approximately equally deep lines of descent from which the clear majority of modern cyanobacteria and plastids are derived (Fig. 3A). For the reasons discussed in the preceding paragraphs, this location in the inferred SSU rRNA tree is unlikely to be an artifact of the analytical methodology.

ACKNOWLEDGMENTS

We thank George Bullerjahnn, Martina Celerin, Susan Golden, Louis Sherman, David Ward, and John Waterbury for providing genomic DNA or cellular material used in this study. Thanks are also extended to Annick Wilmotte for providing the SSU rRNA sequence of *Pseudanabaena* PCC 7409, to Dave Swofford for making test versions of PAUP* available, and to François Lutzoni for reading and commenting on an earlier version of the manuscript. VPWM thanks Joe McDermott and Xiang Li for providing primers 16s.0538.R1 and 16s.1100.F16. KMP gratefully acknowledges support from the Pritzker Endowment for Biosystematic Research and thanks Jeremy Kirchman and Elizabeth Grismer for technical assistance with DNA amplification and sequencing in the Field Museum's Pritzker Laboratory for Molecular Systematics and Evolution. This work was partly supported by grants from the NSF (DEB-9318594) and NIH (RO1-GM35087) to JDP.

LITERATURE CITED

Adachi, J. & Hasegawa M. 1996. Instability of quartet analyses of molecular sequence data by the maximum likelihood method: the Cetacea/Artiodactyla relationships. *Mol. Phylogenet. Evol.*, **6**:72–76.

Beanland, T. J. & Howe, C. J. 1992. The inference of evolutionary trees from molecular data. *Comp. Biochem. Physiol.*, **102B**:643–659.

Bhattacharya, D. & Medlin, L. 1995. The phylogeny of plastids: a review based on comparisons of small subunit ribosomal RNA coding regions. *J. Phycol.*, **31**:489–498.

Bhattacharya, D. & Schmidt, H. A. 1997. Division *Glaucocystophyta*. *Pl. Syst. Evol.*, Su[*Suppl.*] **11**:139–148.

Burger-Wiersma, T., Stal, L. J. & Mur, L. R. 1989. *Prochlorothrix hollandica* gen. nov., sp. nov., a filamentous oxygenic photoautotrophic procaryote containing chlorophylls *a* and *b*: assignment to *Prochlorotrichaceae* fam. nov. and order Prochlorales Florenzano, Balloni, and Materassi 1986, with emendation of the ordinal description. *Int. J. Syst. Bacteriol.*, **39**:250–257.

Cao, Y., Adachi, J. & Hasegawa, M. 1998. Comment on the quartet puzzling method for finding maximum-likelihood tree topologies. *Mol. Biol. Evol.*, **15**:87–89.

Chisholm, S. W., Frankel, S. L. Goericke, R., Olson, R. J., Palenik, B., Waterbury, J. B., West-Johnsrud, L. & Zettler, E. 1992. *Prochlorococcus marinus* nov. gen. nov. sp.: an oxyphototrophic marine prokaryote containing divinyl chlorophyll *a* and *b*. *Arch. Microbiol.*, **157**:297–300.

Delwiche, C. F. & Palmer, J. D. 1997. The origin of plastids and their spread via secondary symbiosis. *Pl. Syst. Evol.*, [*Suppl.*] **11**:53–86.

Delwiche, C. F., Kuhsel, M. & Palmer, J. D. 1995. Phylogenetic analysis of *tufA* sequences indicates a cyanobacterial origin of all plastids. *Mol. Phylogenet. Evol.*, **4**:110–128.

Douglas, S. E. 1994. Chloroplast origins and evolution. *In*: Bryant, D. A. (ed.), The Molecular Biology of Cyanobacteria. Kluwer, Amsterdam. p. 91–118.

Douglas, S. E. & Turner, S. 1991. Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. *J. Mol. Evol.*, **33**:267–273.

Egea, N. & Lang-Unnasch, L. 1995. Phylogeny of the large extrachromosomal DNA of organisms in the phylum Apicomplexa. *J. Euk. Microbiol.*, **42**:679–684.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, **17**:368–376.

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

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle, Washington, USA.

Felsenstein, J. & Kishino, H. 1993. Is there something wrong with the bootstrap on phylogenies? A reply to Hillis and Bull. *Syst. Biol.*, **42**: 193–200.

Ferris, M. J., Ruff-Roberts, A. L., Kopczynski, E. D., Bateson, M. M. & Ward, D. M. 1996. Enrichment culture and microscopy conceal diverse thermophilic *Synechococcus* populations in a single hot spring microbial mat habitat. *Appl. Environm. Microbiol.*, **62**:1045–1050.

Garcia-Pichel, F., Nübel, U. & Muyzer, G. 1998. The phylogeny of unicellular, extremely halotolerant cyanobacteria. *Arch. Microbiol.*, **169**:469–482.

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.

Gray, M. W. 1992. The endosymbiont hypothesis revisited. *Int. Rev. Cytol.*, **141**:233–357.

Guglielmi, G., Cohen-Bazire, G. & Bryant, D. 1981. The structure of *Gloeobacter violaceus* and its phycobilisomes. *Arch. Microbiol.*, **129**: 181–189.

Gutell, R. R. 1993. Collection of small subunit (16S- and16S-like) ribosomal RNA structures. *Nucl. Acids Res.*, **21**:181–189.

Helmchen, T., Bhattacharya, D. & Melkonian, M. 1995. Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. *J. Mol. Evol.*, **41**:303–210.

Hillis, D. M. & Bull, J. J. 1993. An empirical test of bootstrapping as a measure of assessing confidence in phylogenetic analysis. *Syst. Biol.*, **42**:182–192.

Hillis, D. M., Huelsenbeck, J. P. & Swofford, D. L. 1994. Hobgoblin of phylogenetics? *Nature*, **69**:363–364.

Howe, C. J., Beanland, T. J., Larkum, A. W. D. & Lockhart, P. J. 1992. Plastid origins. *Trends Ecol. Evol.*, **7**:378–383.

Huelsenbeck, J. P. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.*, **47**:519–537.

Ishida, T., Yokota, A. & Sugiyama, J. 1997. Phylogenetic relationships of filamentous cyanobacterial taxa inferred from 16S rRNA sequence divergence. *J. Gen. Appl. Microbiol.*, **43**:237–241.

Jermiin, L. S., Olsen, G. J., Mengersen, K. L. & Easteal, S. 1997. Majority-rule consensus of phylogenetic trees obtained by maximum-likelihood analysis. *Mol. Biol. Evol.*, **14**:1296–1302.

Kane, M., Maeda, H., Fukunaga, T. & Nishi, K. 1997. Molecular phylogenetic relationship between strains of cyanobacterial picoplankton in Lake Biwa, Japan. *J. Mar. Biotechnol.*, **5**:41–45.

Kaneko, T., Tanaka, A., Sato, S., Kotani, H., Sazuka, T., Miyajima, N., Sugiura, M. & Tabata, S. 1995. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. I. Sequence features in the 1 Mb region from map positions 64% to 92% of the genome. *DNA Res.*, **2**:153–166.

Kishino, H. & Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in *Hominoidea*. *J. Mol. Evol.*, **29**:170–179.

Köhler, S. Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. & Roos, D. S. 1997. A plastid of probable green algal origin in apicomplexan parasites. *Science*, **275**: 1485–1489.

Kowallik, K. V., Stoebe, B., Schaffran, I., Kroth-Panic, P. & Freier, U. 1995. The chloroplast genome of a chlorophyll *a* + *c*-containing alga, *Odontella sinensis*. *Pl. Mol. Biol. Reporter*, **13**:336–342.

Kuhner, M. K. & Felsenstein, J. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.*, **11**:459–468.

Lecointre, G., Philippe, H., Van Le, H. L. & Le Guyader, H. 1993. Species sampling has a major impact on phylogenetic inference. *Mol. Phylogenet. Evol.*, **2**:205–224.

Lee, R. E. 1989. Phycology, 2d ed. Cambridge University Press, Cambridge.

Lewin, R. A. 1989. Group II. Order *Prochlorales* Lewin 1977. *In:* Stalcy, J. T., Bryant, M. P., Pfennig, N. & Holt, J. G. (ed.), Bergey's Manual of Systematic Bacteriology. Williams & Wilkins, Baltimore. **3**:1709–1802.

Ligon, P. J. B., Meyer, K. G., Martin, J. A. & Curtis, S. E. 1991. Nucleotide sequence of a 16S rRNA gene from *Anabaena* sp. strain PCC 7120. *Nucl. Acids Res.*, **19**:4553–4553.

Lockhart, P. J., Beanland, T. J., Howe, C. J. & Larkum, A. W. D. 1992a. Sequence of *Prochloron didemni atbBE* and the inference of chloroplast origins. *Proc. Natl. Acad. Sci. USA*, **89**:2742–2746.

Lockhart, P., Steel, M., Hendy, M. & Penny, D. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.*, **11**:605–612.

Lockhart, P. J., Howe, C. J., Bryant, D. A., Beanland, T. J. & Larkum, A. W. D. 1992b. Substitutional bias confounds inference of cyanelle origins from sequence data. *J. Mol. Evol.*, **34**:153–162.

Loiseaux-de Göer, S. 1994. Plastid lineages. *Progr. Phycol. Res.*, **10**: 137–177.

Maddison, W. P. & Maddison, D. R. 1992. MacClade: analysis of phylogeny and character evolution. Version 3.0. Sinauer, Sunderland, Massachusetts.

Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J. & Woese, C. R. 1996. The ribosomal data base project (RDP). *Nucl. Acids Res.*, **24**:82–85.

Markowicz, Y., Loiseaux-de Göer, S. & Mache, R. 1988. Presence of a 16S rRNA pseudogene in the bi-molecular plastid genome of the primitive brown alga *Pylaiella littoralis*. Evolutionary implications. *Curr. Genet.*, **14**:599–608.

McFadden G. I., Gilson, P. R. & Waller, R. F. 1995. Molecular phylogeny of chlorarachniophytes based on plastid rRNA and *rbcL* sequences. *Arch. Protistenk.*, **145**:231–239.

Medlin, L. K., Cooper, A., Hill, C., Wrieden-Prigge, S. & Wellbrock, U. 1995. Phylogenetic position of the *Chromista* plastids from 16S rDNA coding regions. *Curr. Genet.*, **28**:560–565.

Miao, V. P. W., Rabenau, A. & Lee, A. 1997. Cultural and molecular characterization of photobionts of *Peltigera membranacea*. *Lichenologist*, **29**:571–586.

Moore, L. R., Rocap, G. & Chisholm, S. W. 1998. Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature*, **393**:464–467.

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*. *Internat. J. Syst. Bacteriol.*, **47**:693–697.

Nelissen, B., De Baere, R., Wilmotte, A. & De Wachter, R. 1996. Phylogenetic relationships of nonaxenic filamentous cyanobacterial strains based on 16S rRNA sequence analysis. *J. Mol. Evol.*, **42**:194–200.

Nelissen, B., Van de Peer, Y., Wilmotte, A. & De Wachter, R. 1995. An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol. Biol. Evol.*, **12**:1166–1173.

Nelissen, B., Wilmotte, A., Neefs, J.-M. & De Wachter, R. 1994. Phylogenetic relationships among filamentous helical cyanobacteria investigated on the basis of 16S ribosomal RNA gene sequence analysis. *Syst. Appl. Microbiol.*, **17**:206–210.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S. I., Inokuchi, H. & Ozeki, H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Pl. Mol. Biol. Reporter*, **4**:148–175.

Olsen, G. J. 1987. Earliest phylogenetic branchings: comparing rRNA-based evolutionary trees inferred with various techniques. *Cold Spring Harbor Symp. Quant. Biol.*, **52**:829–837.

Olsen, G. J., Matsuka, H., Hagstrom, R. & Overbeek, R. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. *Computer Applic. Biosci.*, **10**: 41–48.

Page, R. D. M. 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applic. Biosci.*, **12**: 357–358.

Paquin, B., Kathe, S. D., Nierzwiki-Bauer, S. A. & Shub, D. A. 1997. Origin and evolution of Group I introns in cyanobacterial tRNA genes. *J. Bacteriol.*, **179**:6798–6806.

Philippe, H. 1997. Rodent monophyly: pitfalls of molecular phylogenies. *J. Mol. Evol.*, **45**:712–715.

Philippe, H. & Douzery, E. 1994. The pitfalls of molecular phylogeny based on four species, as illustrated by the Cetacea/Artiodactyla relationships. 1994. *J. Mammal. Evol.*, **2**:133–152.

Porter, R. D. 1988. DNA transformation. *Meth. Enzymol.*, **167**:703–712.

Reith, M. & Munholland, J. 1995. Complete nucleotide sequence of the *Porphyra purpurea* chloroplast genome. *Pl. Mol. Biol. Reporter*, **13**: 333–335.

Reeves, R. H. 1996. 16S ribosomal RNA and the molecular phylogeny of the cyanobacteria. *Beih. Nova Hedwigia*, **112**:55–67.

Rippka, R. & Herdman, H. 1992. Pasteur Culture Collection of Cyanobacteria Catalogue & Taxonomic Handbook. 1. Catalogue of Strains. Institut Pasteur, Paris.

Rippka, R., Waterbury, J. & Cohen-Bazire, G. 1974. A cyanobacterium which lacks thylakoids. *Arch. Microbiol.*, **100**:419–436.

Rudi, K., Skulberg, O. M., Larsen, F. & Jakobsen, K. S. 1997. Strain characterization and classification of oxyphotobacteria in clone cultures on the basis of 16S rRNA sequences from the variable regions V6, V7, and V8. *Appl. Environm. Microbiol.*, **63**:2593–2599.

Selstam, E. & Campbell, D. 1996. Membrane lipid composition of the unusual cyanobacterium *Gloeobacter violaceus* sp. PCC 7421, which lacks sulfoquinovosyl diacylglycerol. *Arch. Microbiol.*, **166**:132–135.

Stirewalt, V. L., Michalowski, C. B., Löffelhardt, W., Bohnert, H. J. & Bryant, D. A. 1995. Nucleotide sequence of the cyanelle genome from *Cyanophora paradoxa*. *Pl. Mol. Biol. Reporter*, **13**:327—332

Strimmer, K. & von Haeseler, A. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.*, **13**:964–969.

Strimmer, K., Goldman, N. & von Haeseler, A. 1997. Bayesian probabilities and quartet puzzling. *Mol. Biol. Evol.*, **14**:210–211.

Swofford, D. L. 1997. "PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods), Version 4.0." Program and Documentation. Sinauer, Sunderland, Massachusetts.

Tamura, K. & Nei, M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, **10**:512–526.

Tillier, E. R. M. & Collins, R. A. 1995. Neighbor joining and maximum likelihood with RNA sequences: addressing the interdependence of sites. *Mol. Biol. Evol.*, **12**:7–15.

Tomioka, N., Shinozaki, K. & Sugiura, M. 1981. Molecular cloning and characterization of ribosomal RNA genes from a blue-green alga, *Anacystis nidulans*. *Mol. Gen. Genet.*, **184**:359–363.

Turner, S. 1997. Molecular systematics of oxygenic photosynthetic bacteria. *Pl. Syst. Evol.* [*Suppl.*],**11**:13–52.

Urbach, E., Robertson, D. L. & Chisholm, S. W. 1992. Multiple evolutionary origins of prochlorophytes within the cyanobacterial radiation. *Nature,* **355**:267–270.

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.

Van de Peer, Y., Jansen, J., De Rijk, P. & De Wachter, R. 1997. Database on the structure of small ribosomal subunit RNA. *Nucl. Acids Res.,* **25**:111–116.

Wakasugi, T., Nogai, T., Kapoor, M., Sugita, M., Ito, M., Ito, S., Tsudzuki, J., Nakashima, K., Tsudzuki, T., Suzuki, Y., Hamada, A., Ohto, T., Inamura, A., Yoshinaga, K. & Sugiura, M. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: the existence of genes possibly involved in chloroplast division. *Proc. Natl. Acad. Sci. USA,* **94**:5967–5972.

Wakeley, J. 1994. Substitution-rate variation among sites and the estimation of transition bias. *Mol. Biol. Evol.,* **11**:436–442.

Waterbury, J. B. & Rippka, R. 1989. Subsection I, order *Chroococcales* Wettstein 1924, emend. Rippka et al., 1979. *In*: Staley, J. T., Bryant, M. P., Pfennig, N. & Holt, J. G. (ed.), Bergey's Manual of Systematic Bacteriology. Williams & Wilkins, Baltimore. **3**:1746–1770.

Wilmotte, A. 1994. Molecular evolution and taxonomy of the cyanobacteria. *In:* Bryant, D. A. (ed.), The Molecular Biology of Cyanobacteria. Kluwer, The Netherlands. p. 1–25.

Wilmotte, A., Neefs, J.-M. & De Wachter, R. 1994. Evolutionary affiliation of the marine nitrogen-fixing cyanobacterium *Trichodesmium* sp. strain NIBB 1067, derived by 16S ribosomal RNA sequence analysis. *Microbiology,* **140**:2159–2164.

Wilmotte, A., Van der Auwera, G. & De Wachter, R. 1993. Structure of the 16S ribosomal RNA of the thermophilic cyanobacterium *Chlorogloeopsis* HTF (*'Mastigocladus laminosus HTF'*) strain PCC7518, and phylogenetic analysis. *FEBS Lett.,* **317**:96–100.

Wilmotte, A., Turner, S., Van de Peer, Y. & Pace, N. R. 1992. Taxonomic study of marine oscillatoriacean strains (cyanobacteria) with narrow trichomes. II. Nucleotide sequence analysis of the 16S ribosomal RNA. *J. Phycol.,* **28**:828–838.

Zharkikh, A. & Li, W.-H. 1992. Statistical properties of bootstrap estimation of phylogenetic variability from nucleotide sequences. I. Four taxa with a molecular clock. *Mol. Biol. Evol.,* **9**:1119–1147.

Zharkikh, A. & Li, W.-H. 1995. Estimation of confidence in phylogeny: the complete-and-partial bootstrap technique. *Mol. Phylogenet. Evol.,* **4**:44–63.