J Mol Evol (1991) 32:379–395

Journal of
Molecular Evolution


# Sequence Analysis and Phylogenetic Reconstruction of the Genes Encoding the Large and Small Subunits of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase from the Chlorophyll *b*-Containing Prokaryote *Prochlorothrix hollandica*

Clifford W. Morden* and Susan S. Golden

Department of Biology, Texas A&M University, College Station, TX 77843, USA

**Summary.** Prochlorophytes similar to *Prochloron* sp. and *Prochlorothrix hollandica* have been suggested as possible progenitors of the plastids of green algae and land plants because they are prokaryotic organisms that possess chlorophyll *b* (chl *b*). We have sequenced the *Prochlorothrix* genes encoding the large and small subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco), *rbcL* and *rbcS*, for comparison with those of other taxa to assess the phylogenetic relationship of this species. Length differences in the large subunit polypeptide among all sequences compared occur primarily at the amino terminus, where numerous short gaps are present, and at the carboxy terminus, where sequences of *Alcaligenes eutrophus* and non-chlorophyll *b* algae are several amino acids longer. Some domains in the small subunit polypeptide are conserved among all sequences analyzed, yet in other domains the sequences of different phylogenetic groups exhibit specific structural characteristics. Phylogenetic analyses of *rbcL* and *rbcS* using Wagner parsimony analysis of deduced amino acid sequences indicate that *Prochlorothrix* is more closely related to cyanobacteria than to the green plastid lineage. The molecular phylogenies suggest that plastids originated by at least three separate primary endosymbiotic events, i.e., once each leading to green algae and land plants, to red algae, and to *Cyanophora paradoxa*. The *Prochlorothrix* rubisco genes show a strong GC bias, with 68% of the third codon positions being G or C. Factors that may affect the GC content of different genomes are discussed.

**Key words:** *rbcLS* — Operon — Rubisco — Prochlorophyte — Endosymbiosis — Polyphyletic plastid origin — Codon usage — Chlorophyll *b* — Phycobilins

## Introduction

The concept of an endosymbiotic origin of the chloroplast from a bacterial progenitor was first proposed over 100 yr ago (Schimper 1883) although only recently has the technology been available to test such a theory. Although this view is now generally accepted (Gray 1989), the source of the plastid progenitor among different plastid lineages is still questioned. One of the distinguishing traits among these different lineages is the mechanism by which they harvest light for photochemical activity. Chlorophyll *a* (chl *a*) is ubiquitous among oxygenic photosynthetic organisms. Cyanobacteria and many nongreen algae utilize phycobilin pigments (e.g., phycocyanin and phycoerythrin) in a proteinaceous phycobilisome on the thylakoid surface to harvest light. In contrast, chlorophytes and metaphytes (green algae and land plants, respectively) contain chlorophylls *a* and *b* as light-harvesting pigments. Based on this and other evidence, Cavalier-Smith (1982) proposed a monophyletic origin for plastids from a cyanobacterium, with subsequent evolution of the symbiont giving rise to the diversity of other

* *Present address:* Department of Biology, Indiana University, Bloomington, IN 47405, USA
*Offprint requests to:* S. S. Golden

plastid types. However, others contend that plastids have been derived from multiple endosymbiotic events involving other bacteria (Sagan 1967; Lewin 1981; Whatley 1981; Whatley and Whatley 1981) or eukaryotic algae (Gibbs 1978, 1981; Whatley 1981; Whatley and Whatley 1981) in addition to cyanobacteria.

Several studies have shown that chloroplasts of green algae and land plants are more closely related to cyanobacteria than to other photosynthetic bacterial lineages (Woese 1987; Giovannoni et al. 1988). *Cyanophora paradoxa,* a flagellated protozoan often referred to as a cyanophyte, has been considered by some a likely candidate as the original endosymbiont (Cavalier-Smith 1982; Maxwell et al. 1986). The plastid (cyanelle) of *Cyanophora* shares several characteristics with cyanobacteria, including phycobiliproteins within the cyanelle (Bryant et al. 1985; Lemaux and Grossman 1985) and peptidoglycans (residual cell wall) on the cyanelle membrane (Aitken and Stanier 1979). However, the cyanelle chromosome is similar in size and gene content to plastid genomes (Lambert et al. 1985; Breiteneder et al. 1988) and most cyanelle proteins are presumably encoded by nuclear genes and transported into the organelle (Bayer et al. 1990). These features suggest an intermediate phylogenetic position between cyanobacteria and green chloroplasts (Maxwell et al. 1986). Others propose that these characters indicate an evolutionary history separate from that of green plastids and possibly within the lineage of red algae (Whatley and Whatley 1981; Lambert et al. 1985; Breiteneder et al. 1988).

The prochlorophytes *Prochloron* sp. and the recently discovered *Prochlorothrix hollandica* are photosynthetic bacteria known to contain chl *b* rather than phycobiliproteins as an accessory light-harvesting pigment (Lewin 1975a,b; Burger-Wiersma et al. 1986, 1989). Cytological and physiological characteristics of prochlorophytes point toward their being potential intermediates in the evolution from cyanobacteria to green chloroplasts. The arrangement of thylakoids is similar to that of green plastids (in stacks) rather than that of cyanobacteria (concentric layers; Whatley and Whatley 1981; Burger-Wiersma et al. 1986). Freeze-fracture studies of the thylakoids (Miller et al. 1988) and characteristics of photosynthetic electron transport (Burger-Wiersma and Post 1989) in *Prochlorothrix* show remarkable similarity to the chloroplasts of green plants. However, antibodies from the chl *a/b* binding protein complexes from plants do not produce any cross-reactions with proteins from *Prochlorothrix* (Bullerjahn et al. 1987) and lipid biochemistry of *Prochlorothrix* is more similar to that of prokaryotes than eukaryotes (Volkman et al. 1988).

Phylogenetic analysis using rRNA oligonucleotide catalogues could not resolve the position of *Prochloron* in relation to cyanobacteria and chloroplasts except to conclude that it is associated with these groups (Woese 1987; Bremer and Bremer 1989). Phylogenetic analyses of *Prochlorothrix* have yielded conflicting results. We (Morden and Golden 1989a) analyzed the *psbA* genes from *Prochlorothrix* and found that it has a seven-amino acid deletion at the carboxy terminus of the deduced protein relative to all cyanobacterial sequences examined, a feature that previously had been observed only in plant *psbA* sequences (Curtis and Haselkorn 1984). Subsequent analysis of the deduced protein sequences indicated that *Prochlorothrix* was intermediate to cyanobacteria and chloroplasts or on a branch associated with *Synechococcus* 7942 (= *Anacystis nidulans* R2) adjacent to the chloroplast lineage (Morden and Golden 1989a,b). In contrast, Turner et al. (1989) sequenced segments of the 16S rRNA of *Prochlorothrix* and also found it to be most closely related to *A. nidulans* but not on a branch associated with the plastid lineage. Moreover, cyanobacteria shown to be closely related to green chloroplasts on the basis of 16S rRNA sequences (Turner et al. 1989) were found to be distantly related when *psbA* was analyzed (Morden and Golden 1989a,b). Further analysis of the *psbA* sequence with maximum likelihood favored the 16S rRNA tree by Turner et al. (1989) but could not conclusively exclude the *psbA* tree when the gap was included in the analysis (Kishino et al. 1990).

The gene (*rbcL*) encoding the large subunit (LSU) of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) increasingly is being used for phylogenetic analysis (Palmer et al. 1988). Reasons for this are that it (1) has been shown to be a reliable indicator of relationships (Ritland and Clegg 1987), (2) is relatively large (over 1400 bp), (3) has a reasonably slow evolutionary rate, and (4) is being widely sequenced, providing a large data base for further phylogenetic work. The gene encoding the small subunit (SSU) of rubisco (*rbcS*) also has been used as a phylogenetic indicator (Martin et al. 1983, 1985). However, its utility in phylogenetic analysis is limited by its smaller size (between 300 and 400 bp) and a more rapid rate of base substitution.

There are two fundamentally different types of rubisco holoenzymes with respect to polypeptide subunit composition. In some purple nonsulfur bacteria ($\alpha$-purple bacteria of Woese 1987) rubisco is an oligomer of identical large subunits, such as the $L_2$ rubisco of *Rhodospirillum rubrum* (Tabita and McFadden 1974) and form II rubisco of *Rhodobacter sphaeroides* (Gibson and Tabita 1977). In most other purple bacteria, and in cyanobacteria and eukaryotes that contain rubisco, the enzyme is a hexadecamer of eight large and eight small subunits

($L_8S_8$). The primitive condition of gene organization when both subunits are present in the holoenzyme is an operon in which *rbcL* is upstream of *rbcS* and the two genes are separated by a short spacer region. The *rbcLS* operon has been found in α-, β-, and γ-purple bacteria (Andersen and Caton 1987; Hallenbeck and Kaplan 1988; Viale et al. 1989), cyanobacteria (Shinozaki et al. 1983; Nierzwicki-Bauer et al. 1984), and in the plastid genomes of non-chl *b*-containing algae (Starnes et al. 1985; Boczar et al. 1989; Douglas and Durnford 1989; Hwang and Tabita 1989; Valentin and Zetsche 1989; Douglas et al. 1990). In green algae and land plants, *rbcL* is retained in the plastid genome, but *rbcS* has been translocated to the nuclear genome where it comprises a gene family with two or more different copies of the gene (Ellis 1985; Goldschmidt-Clermont and Rahire 1986). In addition to differences between nuclear- and operon-encoded *rbcS* genes in regions that encode domains of the mature protein (Wolter et al. 1988), nuclear genes also encode a 40–60-amino acid transit peptide upstream of the mature polypeptide that is cleaved off after transfer into the chloroplast (Ellis 1985).

We show here that rubisco genes of *Prochlorothrix* exist as an operon similar to that previously described for bacteria and non-chl *b* algae. The derived amino acid sequences from *rbcL* and *rbcS* were compared to those of other taxa to reconstruct the phylogeny of rubisco genes and to make inferences concerning the endosymbiosis of plastids.

## Materials and Methods

*Prochlorothrix hollandica* cultures were grown as described by Bullerjahn et al. (1987). Mixed cultures of *Prochlorothrix* contained unicellular heterotrophic bacteria; hence prochlorophyte filaments were enriched prior to DNA or RNA extraction (axenic cultures of the organism, isolated by R. Lewin, are now available). To enrich the filaments, the pellicle was rinsed several times in growth medium and disrupted with a tissue homogenizer. Cells were combined with two volumes 50% Percoll in 0.05 M Mes pH 7.0, centrifuged at 10,000 × *g* for 30 min at 4°C, and collected as a green layer in the top of the gradient. Cells were washed again in growth medium and pelleted by centrifugation at 3200 rpm for 10 min. Total DNA was isolated by alkaline lysis of bacterial cells and CsCl buoyant density gradient centrifugation using standard methods described by Maniatis et al. (1982). Total RNA was isolated using methods previously described by Golden et al. (1987) and modified by Schaefer and Golden (1989).

*Escherichia coli* strain DH5αMCR (*mcrA, mcrB*) (Bethesda Research Laboratories, BRL, Gaithersburg, MD) was the host for all plasmids and strain ER1458 (Y1090, *mcrB*) was the host for bacteriophage λ clones. Growth and infection followed standard procedures as described by Maniatis et al. (1982). Most restriction and modifying enzymes were purchased from BRL or Boehringer Mannheim Biochemicals (Indianapolis, IN) and used as directed by the manufacturers.

A recombinant bacteriophage library of *Eco*RI fragments from total *Prochlorothrix* DNA was constructed in the vector λGT-10 (Promega, Madison, WI) and was screened by plaque hybridization (Maniatis et al. 1982) on DuPont colony/plaque screen membranes (BRL). The cloned An602 fragment from *Anabaena* 7120 (Curtis and Haselkorn 1983), which contains most of the *rbcL* gene, was radiolabeled with α-$^{32}$P-deoxynucleotides using a random primer DNA labeling kit (BRL) and used as a heterologous gene probe. Membranes were hybridized with the probe at 60°C in 5× SSPE (1× SSPE is 0.18 M NaCl, 0.01 M Na-phosphate, 1 mM EDTA) (Maniatis et al. 1982), 1% SDS and washed at the same temperature in 5× SSPE, 0.1% SDS. Positive plaques from autoradiography were rescreened and insert DNA was subcloned into pBGS18 (Spratt et al. 1986) to create pAM449. Most of the *rbcS* gene was missing from this clone.

A second library containing genomic *Bam*HI fragments in λL47.1 (Maniatis et al. 1982) was screened to obtain the entire *rbcS* gene. A 320-bp *Bam*HI/*Eco*RI fragment from pAM449 that spans the *rbcL–rbcS* spacer region and the amino-terminal portion of *rbcS* (Fig. 1) was labeled and used as a homologous probe under hybridization conditions described above. A 1.65-kb DNA fragment containing the *rbcS* gene was identified and cloned into pBGS19 (Spratt et al. 1986) to create pAM636.

Fragments from pAM449 and pAM636 were subcloned into pBGS18 or 19 (Spratt et al. 1986) and pUC18 or 19 (Yanisch-Perron et al. 1985) for DNA sequencing. Plasmid DNA was isolated by a boiling miniprep method (Maniatis et al. 1982) and further purified by binding to glass fines in the presence of a chaotropic salt solution (Golden et al. 1987). Double-stranded templates were prepared for sequencing by the following method: an equal volume of 5 M LiCl was added and the sample was incubated at −20°C for 5 min. Precipitates were removed by microcentrifugation for 15 min, and the supernatant solution was treated with 1 μg RNase A per 100 μl sample. DNA was precipitated by adding one-sixth volume of 40% polyethylene glycol, incubating overnight at 4°C, and collecting the pellet by microcentrifugation for 15 min at 4°C. The pellet was washed with 70% ethanol and dried. The DNA strands were denatured in 0.2 M NaOH, 0.2 mM EDTA pH 8.0, and precipitated with ethanol. A Sequenase kit (United States Biochemical Corporation, Cleveland, OH) was used to perform dideoxy chain-termination sequencing reactions on the double-stranded templates. Sequencing reactions were completed from overlapping clones to account for all bases on both DNA strands. Regions for which subclones were not available were spanned by generating nested exonuclease III deletions with an Erase-a-base kit (Stratagene, La Jolla, CA) and synthesizing 17-base oligonucleotide primers corresponding to known regions of the *rbcL* gene sequence.

The 5′ end of the *rbcLS* message was identified by S1 nuclease protection and primer extension. Each reaction contained 100 μg of total RNA. S1 protection was performed as described by Tumer et al. (1983). A fragment labeled at the *Xba*I site at −376 of Fig. 2 and extending approximately 1.6-kb upstream to a *Hinc*II site was used in the analysis. An A+G sequencing ladder (Maxam and Gilbert 1980) of the same labeled fragment was used to determine the length of the protected band. For primer extension, performed according to Kassavetis and Geiduschek (1982), a 17-bp primer was synthesized complementary to the sense strand between positions −366 to −350 of the sequence shown in Fig. 2. The extended product was measured against a sequencing ladder generated by the dideoxynucleotide chain-termination reaction of the same primer on a DNA template.

DNA sequence assemblage, gene translation to determine the deduced protein sequence, sequence alignments, and percent similarities and identities were performed using the University of Wisconsin Genetics Computer Group (UWGCG) program (Devereux et al. 1984; Devereux 1989). The organisms to whose *rbcL* and *rbcS* genes those of *Prochlorothrix* were compared are listed in Table 1. These species represent the major photosynthetic lineages with the exception of green sulfur and nonsulfur bacteria

**Table 1.** Gene sequences of *rbcL* and *rbcS* used in this study for comparison to those of *Prochlorothrix hollandica*

| Division/class | Taxon | LHP[a] | *rbcL*[b] | *rbcS*[b] | References |
|---|---|---|---|---|---|
| Purple bacteria[c] | (α) *Rhodospirillum rubrum* | BC | x | | Nargang et al. (1984) |
| | (β) *Alcaligenes eutrophus* | BC | x | x | Andersen and Caton (1987) |
| | (γ) *Chromatium vinosum* | BC | x | x | Viale et al. (1989) |
| Cyanobacteria | *Anabaena* 7120 | P | x | x | Curtis and Haselkorn (1983); Nierzwicki-Bauer et al. (1984) |
| | *Anacystis nidulans* | P | x | x | Shinozaki et al. (1983); Voordouw et al. (1987) |
| Cyanophyte | *Cyanophora paradoxa* | P | | x | Starnes et al. (1985) |
| Rhodophyte | *Porphyridium aerugineum* | P | x | | Valentin and Zetsche (1989) |
| Cryptophyte | *Cryptomonas* Φ | P, *c* | x | x | Douglas and Durnford (1989); Douglas et al. (1990) |
| Chromophyte | *Olisthodiscus luteus* | *c* | | x | Boczar et al. (1989) |
| Euglenophyte | *Euglena gracilis* | *b* | x | x | Gingrich and Hallick (1985a,b); Chan et al. (1990) |
| Chlorophyte | *Acetabularia mediterranea* | *b* | | x | Schneider et al. (1989) |
| | *Chlamydomonas reinhardtii* | *b* | x | x | Dron et al. (1982); Goldschmidt-Clermont and Rahire (1986) |
| | *Chlorella ellipsoidea* | *b* | x | | Yoshinaga et al. (1988) |
| Metaphyte[c] | *Marchantia polymorpha* | *b* | x | | Ohyama et al. (1986) |
| | *Pinus tunbergii* | *b* | | x | Yamamoto et al. (1988) |
| | *Nicotiana tabacum* | *b* | x | x | Shinozaki and Sugiura (1982); Müller et al. (1983) |
| | *Helianthus annuus* | *b* | | x | Waksman and Freyssinet (1987) |
| | *Lemna gibba* | *b* | | x | Stiekema et al. (1983) |
| | *Oryza sativa* | *b* | x | x | Moon et al. (1987); Xie and Wu (1988) |

[a] Abbreviations for light-harvesting pigments (LHP) are as follows: BC, bacteriochlorophyll *a*; P, phycobiliproteins; *c*, chlorophylls *a* and *c*; *b*, chlorophylls *a* and *b*

[b] Use of the gene sequence in this study is denoted by "x." Specific *rbcS* clones from publications with multiple sequences are M1 (*Acetabularia*), *rbcS1* (*Chlamydomonas*), pLgSSU (*Lemna*)

[c] Nomenclature of purple bacteria follows that of Woese (1987). Use of the term metaphyte is from Gray (1989); metaphytes are also referred to as land plants in the text

for which these data are not available. Meagher et al. (1989) have shown evidence of gene conversion within the *rbcS* gene family of a given species and as such only a single sequence from species having nuclear-encoded *rbcS* genes was used. Muto and Osawa (1987) have shown that nucleotide sequences (and amino acid sequences to a lesser degree) reflect the total GC content of a genome. Amino acid sequences were used to reduce bias that may be introduced because of similar GC pressure among the organisms studied (bacterial, plastid, and nuclear encoded) and to avoid homoplasy introduced from variation in the wobble position of codons. Phylogenetic relationships among the taxa were inferred using the branch and bound algorithm of PAUP (Swofford 1989). Bootstrap analysis with 100 replications of the global branch swapping algorithm was completed to determine the confidence level of branches within the tree (Felsenstein 1985).

## Results

### *Characterization of the rbcLS Operon*

Southern blots of *Prochlorothrix* DNA probed with an *rbcL*-containing fragment from *Anabaena* 7120 indicated that the gene is present on a 7-kb *Eco*RI fragment and a 4-kb *Bam*HI fragment (data not shown). The 7-kb *Eco*RI fragment was isolated from a λGT-10 library, mapped (Fig. 1), and shown by sequence analysis to contain an entire *rbcL* open reading frame of 1413 bp encoding a polypeptide of 470 amino acids (Fig. 2). Also contained on this clone was part of the *rbcS* gene located 255 bp downstream of the *rbcL* terminus. A fragment of this clone was used as a homologous probe to identify from another library an overlapping 1.65-kb *Bam*HI fragment that contained the entire *rbcS* gene (Fig.

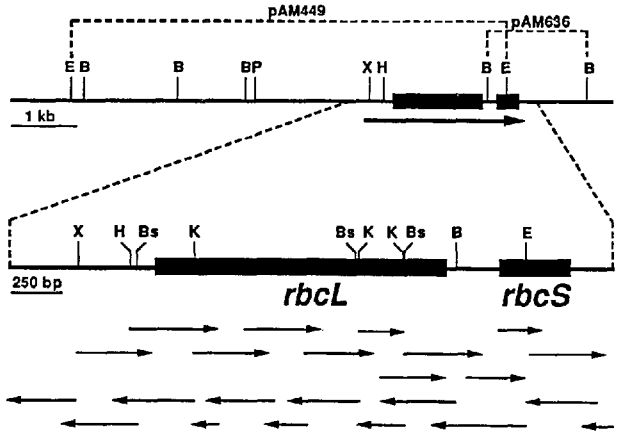


**Fig. 1.** Restriction map and sequencing strategy for *rbcL* and *rbcS* from *Prochlorothrix hollandica*. The 7-kb *Eco*R I fragment containing the entire *rbcL* gene and the 1.65-kb *Bam*H I fragment containing the entire *rbcS* gene (in pAM449 and pAM636, respectively) are shown with restriction sites used in subcloning. Solid boxes represent open reading frames and the heavy arrow indicates the dicistronic transcript. Arrows at the bottom of the figure indicate the direction and length of sequencing runs. Restriction enzyme abbreviations: B, *Bam*H I; Bs, *Bst*E II; E, *Eco*RI; H, *Hinc* II; K, *Kpn* I; P, *Pst* I; X, *Xba* I.

```
-500
ACCGGGCGAT CGAGAAAATC TTAATTAAAA ATGATCGTCT CGTGATACCT TAATTGATGA GCAAGTAATG GGACTGAGTG TTTAGGAGTA TCGACCTAAA
-400
CATCTAAAAA AAATTCTGTG GGAATCTAGA GTCTCTTTGG TATTTCCCAG GTCATTTTTT TCTAGGCTTC ACTCTCTTGG TAACGTTAAG TTAATCAACA
-300
TTTTGATTCC TTAGCTTTAT TGGTGGAGAC TGCTAACTTG TTGGGGTTCT AGGGGACTTA AATCGCTCCG GAACTCAAAC AAACTTAAAT TTTTATGGAG
-200
TGGGAACTGG GGCAGTAACA TCATCACTGC GATGCGATTG CTGAGGCAAG CCCAGTTGTT GACCCGTCCA GTGGAGGTGG CTCCGGTTAC CTCATCGATT
-100
TCAGATTTTT TGTGCCCATA ACCGATGTTT GATGCCGCAG TCAAGTCCTA TGGCCTTTCT TTCGTGCTGC GGTGATCGCT GAAATTCAAG GAGATATCCA
                                                                                                            81
ATG GCA GTA CAG ACC AAA GGC TAT CAG GCC GGT GTA AAA GAC TAC CGC CTG ACC TAC TAC ACC CCC GAA TAC ACC CCC AAG
 M  A  V  Q  T  K  G  Y  Q  A  G  V  K  D  Y  R  L  T  Y  Y  T  P  E  Y  T  P  K
                                                                                                           162
GAC ACA GAC CTG CTG GCT TGT TTC CGC ATG ACC CCC CAG CCC GGT GTC CCC CCC GAA GAA GCT GGT GCT GCG GTT GCT GCT
 D  T  D  L  L  A  C  F  R  M  T  P  Q  P  G  V  P  P  E  E  A  G  A  A  V  A  A
                                                                                                           243
GAA TCT TCC ACC GGT ACC TGG ACC ACC GTT TGG ACT GAC CTT CTG ACC GAC CTC GAT CGC TAC AAA GGT CGT TGC TAT GAA
 E  S  S  T  G  T  W  T  T  V  W  T  D  L  L  T  D  L  D  R  Y  K  G  R  C  Y  E
                                                                                                           324
GTG GAG CCA GTG CCG GGT GAA GAC AAC CAG TAC TTC TGC TTT GTG GCC TAT CCC TTG GAC CTT TTT GAA GAA GGT TCT GTC
 V  E  P  V  P  G  E  D  N  Q  Y  F  C  F  V  A  Y  P  L  D  L  F  E  E  G  S  V
                                                                                                           405
ACC AAC ATT CTG ACC TCC ATC GTC GGT AAC GTG TTT GGC TTC AAA GCC CTG CGT GCC CTG CGT TTG GAA GAT ATC CGC TTC
 T  N  I  L  T  S  I  V  G  N  V  F  G  F  K  A  L  R  A  L  R  L  E  D  I  R  F
                                                                                                           486
CCC ATC GCC CTG GTC AAG ACC TTC CAA GGT CCT CCC CAC GGG ATT CAG GTG GAG CGC GAT CGC TTG AAC AAG TAT GGT CGT
 P  I  A  L  V  K  T  F  Q  G  P  P  H  G  I  Q  V  E  R  D  R  L  N  K  Y  G  R
                                                                                                           567
CCC CTG TTG GGT TGT ACC ATC AAG CCC AAG CTC GGT CTG TCT GCC AAG AAC TAC GGT CGT GCC GTT TAC GAG TGT CTC CGG
 P  L  L  G  C  T  I  K  P  K  L  G  L  S  A  K  N  Y  G  R  A  V  Y  E  C  L  R
                                                                                                           648
GGT GGT CTG GAC TTC ACC AAA GAC GAC GAG AAC ATC AAC TCC CAG CCC TTC ATG CGC TGG CGC GAT CGC TTC CTC TTT GTG
 G  G  L  D  F  T  K  D  D  E  N  I  N  S  Q  P  F  M  R  W  R  D  R  F  L  F  V
                                                                                                           729
CAG GAA GCC ATT GAG AAA GCC CAG GCT GAA ACC GGT GAA GTC AAA GGT CAC TAC CTC AAC GTA ACC GCC GCC ACC TGC GAA
 Q  E  A  I  E  K  A  Q  A  E  T  G  E  V  K  G  H  Y  L  N  V  T  A  A  T  C  E
                                                                                                           810
GAA ATG CTG AAG CGG GCC GAG TTT GCC AAG GAA ATT GGC ACC CCC ATC ATC ATG CAT GAC TTC CTG ACC GGT GGT TTC ACC
 E  M  L  K  R  A  E  F  A  K  E  I  G  T  P  I  I  M  H  D  F  L  T  G  G  F  T
                                                                                                           891
GCC AAC ACC ACC CTG GCC CAC TAT TGC CGC GAC AAC GGC CTG CTG CTC CAC ATT CAC CGC GCC ATG CAC GCT GTG ATT GAC
 A  N  T  T  L  A  H  Y  C  R  D  N  G  L  L  L  H  I  H  R  A  M  H  A  V  I  D
                                                                                                           972
CGT CAG CGC ATC CAC GGG ATT CAC TTC CGC GTG TTA GCC AAG TGT CTG CGT CTG TCC GGT GGT GAC CAC CTC CAC TCC GGT
 R  Q  R  I  H  G  I  H  F  R  V  L  A  K  C  L  R  L  S  G  G  D  H  L  H  S  G
                                                                                                          1053
ACC GTG GTG GGC AAA CTG GAA GGT GAG AAG GAC ATC ACC CTC GGT TTT GTG GAC CTG ATG CGG GAA GAT CAC ATT GAA GAA
 T  V  V  G  K  L  E  G  E  K  D  I  T  L  G  F  V  D  L  M  R  E  D  H  I  E  E
                                                                                                          1134
GAT CGC TCA CGC GGT GTG TTC TTC ACC CAG GAT TGG GCT TCC ATG CCT GGT GTC ATG CCC GTA GCT TCC GGT GGT ATC CAC
 D  R  S  R  G  V  F  F  T  Q  D  W  A  S  M  P  G  V  M  P  V  A  S  G  G  I  H
                                                                                                          1215
GTG TGG CAC ATG CCC GCC CTG GTG GAA ATC TTC GGC GAT GAC TCT TGC CTC CAG TTT GGT GGT GGT ACC TTG GGT CAC CCC
 V  W  H  M  P  A  L  V  E  I  F  G  D  D  S  C  L  Q  F  G  G  G  T  L  G  H  P
                                                                                                          1296
TGG GGT AAC GCG CCT GGT GCA ACG GCT AAC CGG GTT GCC CTG GAA GCC TGT ATC CAA GCC CGT AAC GAA GGC CGC GAC CTC
 W  G  N  A  P  G  A  T  A  N  R  V  A  L  E  A  C  I  Q  A  R  N  E  G  R  D  L
                                                                                                          1377
ATG CGT GAA GGT GGC GAT GTC ATC CGC GAG GCT TGC AAG TGG AGT CCT GAG CTG GCT GTG GCT TGC GAA CTG TGG AAG GAA
 M  R  E  G  G  D  V  I  R  E  A  C  K  W  S  P  E  L  A  V  A  C  E  L  W  K  E
                                   1413                                          1463
ATC AAG TTT GAG TTC GAG GCC ATC GAC ACT CTG TAG GTCTGAACCC TAGACCCGAC AGCCTTTGGT TTGATGAACC CAGGGGATCC
 I  K  F  E  F  E  A  I  D  T  L  *
                                                                                                          1563
TGGCAAGCGG TGGTTGAGGC GATCGCCTGC CAGGGTATCC CCAGCCCACC TCACTACAGG GGCAAATTTC TGATTCAGGT CAAACTACTA GCTATGGCTA
                                                                                                          1663
GGTTTTAAGC CTGCCTATCG GGATCTGTTC TCTGATATCT GGGGTTAGGT CTCCCGATCT GCCCCATCCA ACGCAAGACA CACGGAAGAC GTAAGGATTA
                                                                                                          1746
AAGCC ATG AAA ACT CTG CCC AAA GAG CGT CGC TAC GAA ACC CTT TCC TAC CTG CCC CCC CTG AGC GAT CAG CAA ATT GCT CGC
      M  K  T  L  P  K  E  R  R  Y  E  T  L  S  Y  L  P  P  L  S  D  Q  Q  I  A  R
                                                                                                          1827
CAG ATT GAG TAC ATG GTG CGC GAA GGC TAT ATT CCC GCC GTG GAA TTC AAC GAA GAT TCC GAC GCG ACC ACC TGC TAC TGG
 Q  I  E  Y  M  V  R  E  G  Y  I  P  A  V  E  F  N  E  D  S  D  A  T  T  C  Y  W
                                                                                                          1908
ACC ATG TGG AAG TTG CCC CTG TTC CAC GCC ACT TCT ACC CAA GAA GTG TTG GGC GAA GTG CGC GAG TGC CGC ACC GAA TAC
 T  M  W  K  L  P  L  F  H  A  T  S  T  Q  E  V  L  G  E  V  R  E  C  R  T  E  Y
                                                                                                          1989
CCC AAC TGC TAC ATC CGC GTA GTT GGT TTC GAC AAC ATC AAG CAG TGT CAG TCC GTG AGC TTC ATC GTT CAC AAG CCC AAC
 P  N  C  Y  I  R  V  V  G  F  D  N  I  K  Q  C  Q  S  V  S  F  I  V  H  K  P  N
                                                                                2078
CGT TAC TAA GGTTTGGGTT GTAAATCCCA GACCTGCGAG GGACTCAGGG CCGCTTAGCT AGCCCTAGCC CCTCCCCAGA CCCTGTATCC
 R  Y  *
                                                                                                          2178
CCGTTAACCT ACCCAAAGAT GGGGGACTGC GTACAACCCT GTGCTGCTAG TTCCGAACGT TTTGGAGAAA TAATGGGGAT ATGGGGCTGT TTTATTGGAA
```

**Fig. 2.** *Prochlorothrix hollandica* DNA and deduced amino acid sequences from 500 bp upstream of *rbcL* to 180 bp downstream of *rbcS*. Numbering is relative to the "A" residue of the *rbcL* start codon. Transcription start sites mapped by S1 nuclease protection and primer extension analysis are shown in bold at positions −444 and −442, respectively. Potential "−10" and "−35" promoter elements are double underlined preceding the transcription start site. The potential ribosome binding site preceding the start codon of each gene is underlined. Amino acid translation is shown beneath the DNA sequence using the single letter code.

**Table 2.** Percent *rbcL* sequence identity among the different photosynthetic lineages

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. *Alcaligenes* | | 65 | 58 | 63 | 63 | 60 | 61 | 51 | 53 | 52 | 51 | 53 | 54 |
| 2. *Chromatium* | 55 | | 68 | 74 | 73 | 52 | 53 | 60 | 63 | 61 | 60 | 63 | 62 |
| | **73** | | | | | | | | | | | | |
| 3. *Anabaena* | 58 | 75 | | 77 | 76 | 60 | 60 | 72 | 76 | 74 | 73 | 72 | 71 |
| | **73** | **86** | | | | | | | | | | | |
| 4. *Anacystis* | 59 | 76 | 83 | | 81 | 57 | 58 | 67 | 73 | 70 | 67 | 70 | 69 |
| | **74** | **88** | **91** | | | | | | | | | | |
| 5. *Prochlorothrix* | 59 | 76 | 85 | 88 | | 57 | 58 | 69 | 73 | 72 | 69 | 71 | 71 |
| | **76** | **88** | **93** | **94** | | | | | | | | | |
| 6. *Cryptomonas* | 69 | 52 | 55 | 55 | 57 | | 80 | 60 | 63 | 63 | 60 | 59 | 59 |
| | **82** | **69** | **71** | **72** | **74** | | | | | | | | |
| 7. *Porphyridium* | 71 | 53 | 57 | 57 | 58 | 85 | | 61 | 63 | 63 | 62 | 61 | 61 |
| | **83** | **72** | **74** | **74** | **76** | **92** | | | | | | | |
| 8. *Euglena* | 56 | 72 | 81 | 80 | 85 | 54 | 54 | | 79 | 79 | 79 | 76 | 76 |
| | **73** | **84** | **88** | **89** | **91** | **71** | **73** | | | | | | |
| 9. *Chlamydomonas* | 57 | 74 | 83 | 82 | 86 | 54 | 54 | 92 | | 84 | 79 | 77 | 76 |
| | **74** | **85** | **90** | **91** | **92** | **71** | **73** | **94** | | | | | |
| 10. *Chlorella* | 58 | 73 | 83 | 83 | 86 | 54 | 55 | 89 | 93 | | 81 | 78 | 78 |
| | **74** | **85** | **91** | **90** | **92** | **71** | **73** | **92** | **96** | | | | |
| 11. *Marchantia* | 57 | 73 | 83 | 81 | 84 | 55 | 56 | 87 | 90 | 90 | | 83 | 82 |
| | **74** | **85** | **91** | **91** | **92** | **71** | **73** | **93** | **94** | **94** | | | |
| 12. *Nicotiana* | 57 | 72 | 82 | 80 | 83 | 55 | 56 | 86 | 88 | 88 | 91 | | 86 |
| | **74** | **84** | **90** | **90** | **92** | **72** | **74** | **92** | **93** | **93** | **95** | | |
| 13. *Oryza* | 56 | 71 | 81 | 81 | 83 | 54 | 55 | 85 | 88 | 88 | 91 | 92 | |
| | **73** | **84** | **90** | **90** | **91** | **70** | **73** | **92** | **94** | **93** | **95** | **95** | |

Numbers above the diagonal represent the percent identities among nucleotides, numbers below the diagonal are percent identities at the amino acid level (normal type-face) and percent similarity with conservative substitutions (in bold)

1). Sequence analysis of *rbcS* showed it to be 330 bp long, encoding a polypeptide of 109 amino acids (Fig. 2). Hybridization of the *Anabaena rbcL* probe to *Prochlorothrix* northern blots showed that an *rbcL* transcript of ca. 2.5 kb is present (data not shown). The sum of the lengths of *Prochlorothrix rbcL* and *rbcS* genes plus the spacer sequence is 1998 bp, suggesting that these genes are transcribed as a dicistronic message. S1 nuclease protection identified an adenine at position −444 relative to the ATG of *rbcL* as the transcription start site of the message (Fig. 2). Primer extension results confirmed the result within two bases, identifying the guanine residue at position −442. Although sequences preceding this transcription start site do not bear a strong resemblance to *E. coli* promoter sites, the best match, having a homology score of 47.9% (M.E. Mulligan, et al. 1984), is highlighted in Fig. 2. Putative −10 and −35 elements cover positions −455 to −450 and −476 to −471, respectively (Fig. 2).

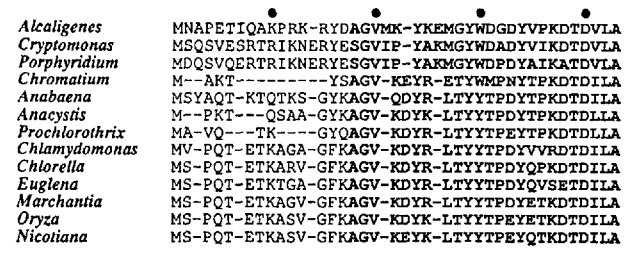


**Fig. 3.** Alignment of rubisco LSU amino acid sequences near the amino terminus. Only the first 43 aligned residues are shown. Species are ordered to reflect similarity among the sequences. Gaps in the sequence are indicated by a dash, and bold dots are placed above every tenth residue. Residues after position 17 are shown in bold.

## *LSU Sequence Comparisons*

The *Prochlorothrix rbcL* gene is 1413 bp in length, which is shortest among the gene sequences published to date. The *rbcL* sequences range from this to 1464 bp in the β-purple bacterium *Alcaligenes eutrophus* (Andersen and Caton 1987) and 1467 bp in the non-chl *b* algae *Cryptomonas φ* (Douglas et al. 1990) and *Porphyridium aerugineum* (Valentin and Zetsche 1989). The *rbcL* genes from cyanobacteria, green algae, and land plants range from 1419 to 1434 bp. Most of the variation in sequence length is accounted for by short gaps at the 5′ end of the gene, a region that is also relatively poorly conserved in sequence (Fig. 3), and by an extension of 21–24 bp at the 3′ end in *Alcaligenes, Cryptomonas,* and *Porphyridium.* Only two other gaps are present in the gene interior of the aligned sequences, found at amino acid positions 274 and 466 [see Douglas et al. (1990) for complete alignment].

Nucleotide and amino acid sequence identities

were calculated for all combinations of sequences representing different photosynthetic types and are presented in Table 2. Gaps were ignored in calculating sequence identity. *Prochlorothrix* shares the highest sequence identity with *Anacystis* at both the nucleotide (81%) and amino acid (88%) levels. The next most similar taxon at the nucleotide level is *Anabaena,* followed by *Chromatium* and the plastid-containing taxa. However, amino acid comparisons differ in that green algae have higher sequence identity to *Prochlorothrix* than do *Anabaena* and higher plant plastid genes. Least similar to *Prochlorothrix* are *Alcaligenes, Cryptomonas,* and *Porphyridium.*

### *SSU Sequence Comparisons*

The SSU from *Prochlorothrix* is more similar in length and sequence similarity to cyanobacterial SSU than to those from any other group of organisms (Fig. 4). The protein length in *Prochlorothrix* is the same as that from *Anabaena* and only two amino acids shorter than that of *Anacystis. Cyanophora* SSU also shares characteristics of cyanobacterial SSUs. Length and sequence in other organisms differ markedly from those of prochlorophyte and cyanobacterial SSU and several structural changes have occurred that have phylogenetic implications. On the basis of gaps present in the sequences it is possible to distinguish most of these lineages. Nuclear-encoded green algal and land plant SSUs share a 12- or 17-amino acid insertion relative to the operon-encoded SSUs of bacteria and non-chl *b* algae at positions 55–71 of the aligned sequence. Length differences divide the organisms into two major groups. Cyanophyte, *Chromatium,* and the cyanobacterial/prochlorophyte SSUs are distinguished from the β-purple bacterium *Alcaligenes* and non-chl *b* algae by a much shorter sequence length, a gap in the latter group of sequences at positions 8–13, and a gap in the former group of sequences at positions 113 and 114. Green algae can be distinguished from all others by an apparent one-amino acid insertion at position 45 and from angiosperms by a five-amino acid insertion at positions 62–66. Dicotyledons are also distinguishable from monocotyledons and gymnosperms by an apparent gap in the monocot/gymnosperm lineage at position 49. Despite these obvious differences among sequences, certain SSU domains have maintained a high level of conservation (boxed amino acids in Fig. 4).

### *Codon Usage and GC Content*

Codon usage of *rbcL* and *rbcS* was determined for *Prochlorothrix* and several other species representing different divisions of photosynthetic organisms for which both gene sequences were available (Table 3). Eight of the 59 synonymously variable codons are not used in either *rbcL* or *rbcS* of *Prochlorothrix.* Similar situations are evident in other species for different codons. In most instances, the codon usage in *rbcL* is reflected in *rbcS.* The principal exception to this is in green algae and land plants where *rbcL* is encoded in the chloroplast and *rbcS* in the nucleus. This agrees with the observation of Grantham et al. (1980) that codon usage patterns tend to be more similar among the various genes of a single genome than in a given gene studied in different species. Among most species analyzed there are four amino acids with one codon more prevalent than other codons. These are asparagine (AAC), glutamic acid (GAA), glycine (GGT), and phenylalanine (TTC). In general, the codons specifying the amino acids appear biased such that the base composition in the wobble position of the codon reflects the overall GC content of the organism, as was shown by Muto and Osawa (1987). This hypothesis was tested by calculating the percentage GC content in third frame wobble positions. The single codon amino acids, methionine and tryptophan, and the termination codons were omitted from the analysis. Table 4 shows the percent GC content in *rbcL* and *rbcS* of each species listed in Table 3. *Prochlorothrix* shows a bias toward a high GC content (65.6% and 72.1% in *rbcL* and *rbcS,* respectively), which is reflected in most other bacterial genes. The purple bacteria *Alcaligenes* and *Chromatium* have the highest GC content and the cyanobacterium *Anacystis* has a GC content very similar to that of *Prochlorothrix.* In contrast, the cyanobacterium *Anabaena* 7120 and the plastid-encoded genes have a much lower GC content (14.9–44.2%). A striking contrast in GC content is evident between the chloroplast-encoded *rbcL* gene and nuclear-encoded *rbcS* gene in *Chlamydomonas* and tobacco (Table 4), such that the plastid genes are biased toward high AT content and nuclear genes are biased toward high GC content. This trend was also evident among other plant sequences examined (data not shown).

Codon usage for some amino acids in *Prochlorothrix rbcL* and *rbcS* indicates influences other than simple GC richness. Particularly striking for some amino acids are the favored use of GGT for glycine and GAA for glutamic acid. Alanine-GCT is also surprisingly frequent and several amino acids (leucine, valine, and proline) show a distinct preference for C over G in the wobble position or vice versa. These biases may indicate selection for codon usage matching tRNA abundances as seen in other bacteria (such as *E. coli,* Ikemura 1985). Codon usage in *Prochlorothrix psbA* (Morden and Golden 1989a) and to a lesser extent in the *rbcL* genes of other prokaryotes (Table 3) also shows these preferences.

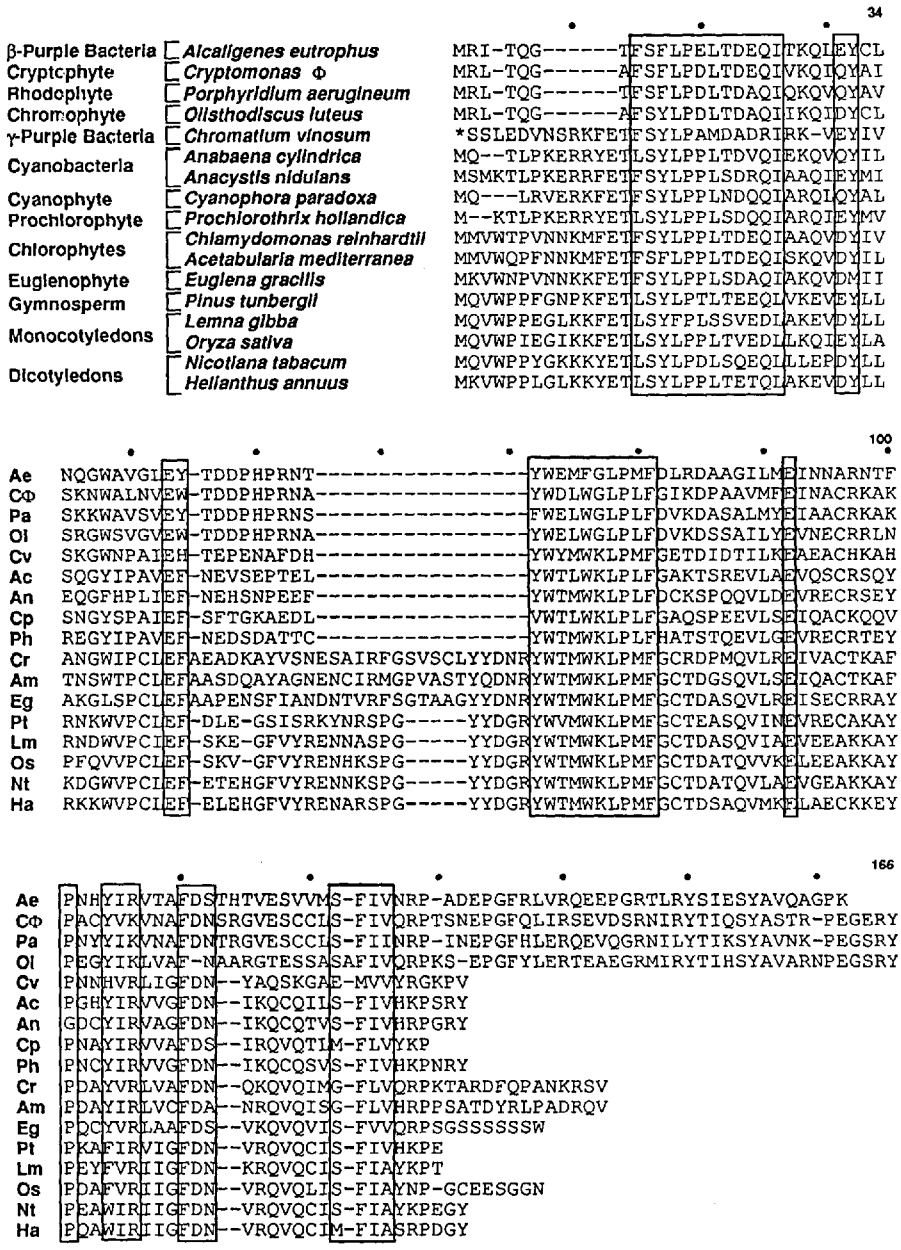


**Fig. 4.** Aligned rubisco SSU sequences. Species are ordered to reflect similarity among the sequences. Gaps in sequences are indicated by a dash, and bold dots are placed above every tenth residue. Boxed residues indicate conserved domains. The asterisk (*) at the amino terminus of the *Chromatium vinosum* sequence represents eight residues, M S E M Q D Y S, unique to this taxon.

## *LSU Phylogeny*

The 14 aligned rubisco LSU sequences of species listed in Table 1 were analyzed for phylogenetic relationships using Wagner parsimony and bootstrap analyses. Positions of gaps were treated as missing characters. As stated previously, the mature rubisco of the α-purple bacterium *R. rubrum* is composed of only two large subunits ($L_2$) and is believed to be ancestral to the other *rbcL* forms; hence it was used as an outgroup in this analysis. Wagner parsimony yielded two equally most parsimonious trees of 1031 steps and a consistency index of 0.787 (Fig. 5A and B). Two major lineages are evident in both trees, one leading to *Alcaligenes* (β-purple bacteria) and the non-chl *b* algae, and one to *Chromatium* (γ-purple bacteria), the cyanobacteria, and ultimately the chlorophyte line. This division is strongly supported by the bootstrap analysis. The difference between the two trees was the branching position of *Prochlorothrix* and *Anacystis*. In both instances the branch nodes of these species is between that of *Chromatium* and *Anabaena*. However in one, *Prochlorothrix* branches between *Anabaena* and *Anacystis* (Fig. 5A) and, in the other, *Prochlorothrix* and *Anacystis* form a clade (Fig. 5B). Bootstrap values for the branches leading to *Prochlorothrix* and the cyanobacteria are relatively low (less than 50%) and cannot be taken with a high degree of confidence.

## *SSU Phylogeny*

Fifteen rubisco SSU sequences were used in the phylogenetic analysis representing each of the major

**Table 3.** Codon usage of *rbcL* and *rbcS* in eubacteria and eukaryotes (continued on page 388)

| Amino acid | Codon | Ph | Ae | Cv | An | Ac | CΦ | Pa | Cr | Nt |
|---|---|---|---|---|---|---|---|---|---|---|
| Ala | GCA | 2 (0) | 3 (0) | 0 (0) | 7 (1) | 13 (1) | 26 (2) | 16 (3) | 8 (0) | 13 (1) |
| Ala | GCT | 13 (1) | 3 (0) | 3 (0) | 19 (2) | 25 (1) | 21 (7) | 27 (7) | 35 (1) | 22 (1) |
| Ala | GCG | 2 (1) | 17 (2) | 12 (1) | 10 (0) | 6 (0) | 2 (2) | 3 (0) | 1 (0) | 4 (1) |
| Ala | GCC | 22 (2) | 26 (6) | 33 (7) | 8 (0) | 4 (1) | 1 (0) | 0 (0) | 1 (12) | 6 (3) |
| Arg | AGA | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 2 (1) | 3 (1) | 6 (3) | 1 (0) | 7 (1) |
| Arg | AGG | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 2 (0) | 0 (0) | 0 (0) | 0 (0) |
| Arg | CGA | 0 (0) | 1 (0) | 0 (0) | 0 (0) | 0 (0) | 1 (0) | 0 (0) | 0 (0) | 6 (0) |
| Arg | CGT | 9 (2) | 0 (1) | 9 (1) | 13 (5) | 18 (5) | 18 (8) | 20 (5) | 30 (0) | 11 (3) |
| Arg | CGG | 4 (0) | 6 (1) | 1 (0) | 3 (0) | 1 (0) | 1 (0) | 1 (0) | 0 (0) | 1 (0) |
| Arg | CGC | 17 (6) | 27 (8) | 17 (5) | 11 (3) | 9 (0) | 0 (1) | 0 (0) | 0 (8) | 5 (0) |
| Asn | AAT | 0 (0) | 0 (1) | 1 (1) | 1 (1) | 0 (1) | 3 (2) | 4 (3) | 0 (0) | 9 (1) |
| Asn | AAC | 13 (4) | 18 (6) | 14 (5) | 13 (2) | 18 (1) | 18 (6) | 18 (5) | 14 (7) | 6 (4) |
| Asp | GAT | 9 (2) | 8 (2) | 5 (3) | 6 (3) | 12 (0) | 16 (7) | 20 (6) | 7 (1) | 23 (3) |
| Asp | GAC | 19 (2) | 25 (5) | 27 (5) | 25 (2) | 16 (2) | 13 (1) | 10 (1) | 23 (7) | 4 (1) |
| Cys | TGT | 5 (1) | 0 (0) | 1 (0) | 2 (0) | 4 (0) | 9 (1) | 5 (3) | 12 (0) | 5 (0) |
| Cys | TGC | 7 (3) | 5 (1) | 8 (1) | 5 (4) | 3 (2) | 0 (3) | 1 (0) | 0 (4) | 4 (3) |
| Gln | CAA | 2 (2) | 2 (1) | 0 (0) | 8 (5) | 12 (7) | 16 (7) | 17 (7) | 8 (0) | 9 (3) |
| Gln | CAG | 11 (4) | 13 (5) | 8 (3) | 4 (2) | 3 (2) | 0 (0) | 0 (0) | 2 (8) | 4 (4) |
| Glu | GAA | 25 (7) | 11 (4) | 9 (4) | 24 (3) | 23 (7) | 23 (7) | 23 (6) | 30 (0) | 24 (5) |
| Glu | GAG | 11 (3) | 10 (7) | 21 (7) | 5 (8) | 8 (2) | 2 (1) | 2 (2) | 0 (6) | 9 (7) |
| Gly | GGA | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 3 (1) | 4 (1) | 2 (0) | 1 (0) | 13 (5) |
| Gly | GGT | 35 (1) | 1 (0) | 13 (0) | 23 (1) | 31 (2) | 41 (4) | 38 (5) | 46 (0) | 23 (0) |
| Gly | GGG | 2 (0) | 7 (1) | 0 (0) | 4 (0) | 2 (0) | 1 (0) | 1 (0) | 0 (0) | 8 (0) |
| Gly | GGC | 8 (2) | 37 (7) | 30 (5) | 18 (3) | 6 (1) | 0 (1) | 2 (1) | 2 (4) | 2 (2) |
| His | CAT | 1 (0) | 2 (1) | 2 (0) | 11 (1) | 1 (0) | 1 (0) | 4 (2) | 0 (0) | 9 (0) |
| His | CAC | 14 (2) | 9 (2) | 13 (5) | 15 (2) | 12 (2) | 8 (1) | 6 (0) | 13 (0) | 5 (1) |
| Ile | ATA | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 1 (0) | 1 (0) | 0 (0) | 2 (0) |
| Ile | ATT | 8 (3) | 0 (0) | 0 (0) | 1 (0) | 6 (3) | 13 (5) | 18 (6) | 15 (0) | 9 (3) |
| Ile | ATC | 15 (3) | 22 (7) | 25 (6) | 21 (7) | 19 (4) | 12 (4) | 8 (2) | 6 (6) | 10 (3) |
| Leu | TTA | 1 (0) | 0 (0) | 0 (0) | 0 (0) | 4 (2) | 18 (8) | 18 (5) | 15 (0) | 9 (0) |
| Leu | TTG | 5 (2) | 0 (0) | 1 (0) | 10 (1) | 15 (2) | 0 (0) | 0 (0) | 0 (0) | 10 (5) |
| Leu | CTA | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 11 (1) | 6 (0) | 15 (2) | 6 (0) | 6 (1) |
| Leu | CTT | 2 (1) | 0 (0) | 0 (0) | 1 (0) | 1 (2) | 15 (1) | 7 (3) | 17 (0) | 10 (3) |
| Leu | CTG | 25 (4) | 31 (8) | 24 (3) | 22 (3) | 8 (3) | 0 (0) | 0 (0) | 0 (8) | 6 (0) |
| Leu | CTC | 10 (0) | 4 (2) | 9 (2) | 10 (3) | 2 (1) | 0 (0) | 0 (0) | 0 (0) | 0 (1) |
| Lys | AAA | 8 (2) | 1 (1) | 23 (8) | 17 (2) | 18 (3) | 22 (6) | 19 (9) | 23 (0) | 21 (1) |
| Lys | AAG | 13 (3) | 20 (1) | 3 (0) | 9 (3) | 7 (3) | 3 (0) | 4 (0) | 0 (7) | 4 (8) |
| Met | ATG | 11 (3) | 17 (5) | 13 (6) | 12 (4) | 8 (1) | 18 (2) | 18 (1) | 13 (7) | 8 (3) |
| Phe | TTT | 8 (0) | 0 (1) | 2 (0) | 5 (1) | 8 (2) | 4 (5) | 4 (4) | 2 (0) | 12 (1) |
| Phe | TTC | 15 (4) | 21 (7) | 19 (5) | 19 (7) | 12 (2) | 16 (2) | 15 (3) | 17 (9) | 9 (4) |
| Pro | CCA | 1 (0) | 0 (0) | 0 (0) | 1 (0) | 4 (2) | 8 (4) | 7 (6) | 13 (0) | 5 (5) |
| Pro | CCT | 4 (0) | 1 (0) | 0 (0) | 5 (2) | 11 (3) | 7 (4) | 14 (1) | 7 (2) | 11 (3) |
| Pro | CCG | 1 (0) | 15 (4) | 14 (5) | 9 (1) | 0 (0) | 0 (1) | 0 (1) | 1 (1) | 3 (0) |
| Pro | CCC | 16 (7) | 6 (5) | 6 (1) | 7 (5) | 5 (3) | 0 (0) | 0 (0) | 0 (6) | 2 (0) |
| Ser | AGT | 1 (0) | 0 (0) | 0 (1) | 0 (0) | 0 (0) | 3 (2) | 1 (1) | 1 (0) | 2 (2) |
| Ser | AGC | 0 (2) | 3 (1) | 2 (3) | 2 (5) | 0 (3) | 2 (3) | 3 (1) | 1 (1) | 3 (1) |
| Ser | TCA | 1 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 6 (2) | 7 (4) | 8 (0) | 3 (2) |
| Ser | TCT | 4 (1) | 0 (0) | 0 (0) | 2 (0) | 9 (4) | 8 (1) | 10 (3) | 5 (1) | 7 (0) |
| Ser | TCG | 0 (0) | 17 (4) | 9 (0) | 10 (2) | 0 (0) | 0 (2) | 1 (0) | 0 (1) | 0 (0) |
| Ser | TCC | 7 (3) | 0 (1) | 5 (4) | 5 (0) | 4 (1) | 1 (0) | 1 (0) | 0 (3) | 2 (0) |
| Thr | ACA | 1 (0) | 0 (0) | 1 (0) | 0 (0) | 10 (2) | 18 (6) | 10 (3) | 12 (0) | 5 (0) |
| Thr | ACT | 2 (2) | 1 (2) | 1 (0) | 1 (2) | 4 (0) | 13 (0) | 17 (3) | 17 (1) | 16 (3) |
| Thr | ACG | 1 (0) | 9 (5) | 4 (0) | 4 (1) | 2 (0) | 0 (0) | 0 (0) | 0 (0) | 1 (0) |
| Thr | ACC | 27 (6) | 22 (4) | 24 (4) | 24 (1) | 17 (4) | 0 (0) | 0 (0) | 0 (5) | 7 (2) |
| Trp | TGG | 8 (2) | 9 (2) | 10 (3) | 9 (2) | 9 (2) | 7 (4) | 8 (3) | 8 (4) | 8 (5) |
| Tyr | TAT | 5 (1) | 4 (0) | 6 (2) | 2 (0) | 4 (2) | 7 (2) | 15 (5) | 2 (0) | 10 (1) |
| Tyr | TAC | 9 (7) | 5 (6) | 12 (5) | 12 (6) | 16 (6) | 13 (4) | 5 (3) | 17 (7) | 8 (9) |
| Val | GTA | 4 (1) | 0 (0) | 0 (0) | 0 (0) | 15 (2) | 22 (6) | 21 (2) | 19 (0) | 17 (0) |
| Val | GTT | 4 (2) | 0 (0) | 1 (0) | 4 (1) | 11 (6) | 15 (2) | 15 (7) | 16 (0) | 16 (2) |
| Val | GTG | 15 (5) | 26 (5) | 12 (2) | 10 (2) | 1 (0) | 0 (0) | 0 (1) | 0 (7) | 2 (5) |

Table 3. Continued from page 387

| Amino acid | Codon | Ph | Ae | Cv | An | Ac | CΦ | Pa | Cr | Nt |
|---|---|---|---|---|---|---|---|---|---|---|
| Val | GTC | 7 (0) | 12 (3) | 19 (5) | 13 (2) | 3 (1) | 0 (0) | 0 (0) | 0 (6) | 1 (1) |
| Ochre | TAA | 0 (1) | 0 (0) | 0 (0) | 1 (1) | 0 (1) | 1 (1) | 1 (1) | 1 (1) | 1 (1) |
| Amber | TAG | 1 (0) | 0 (1) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |
| Opal | TGA | 0 (0) | 1 (0) | 1 (1) | 0 (0) | 1 (0) | 0 (0) | 0 (0) | 0 (0) | 0 (0) |

Amino acids are identified by three letter abbreviations. Numbers outside of the parentheses show the frequency with which each amino acid is encoded by the specified codon for *rbcL*; those inside the parentheses give the values for *rbcS*. Species indicated are *Prochlorothrix hollandica* (Ph), *Alcaligenes eutrophus* (Ae), *Chromatium vinosum* (Cv), *Anacystis nidulans* (An), *Anabaena* 7120 (Ac), *Cryptomonas* Φ (CΦ), *Porphyridium aerugineum* (Pa), *Chlamydomonas reinhardtii* (Cr), and *Nicotiana tabacum* (Nt)

Table 4. Percent GC content of organisms for wobble position of codons in *rbcL* and *rbcS*

| Species | *rbcL* | *rbcS* |
|---|---|---|
| *Prochlorothrix hollandica* | 65.6 | 72.1 |
| *Alcaligenes eutrophus* | 91.6 | 89.1 |
| *Chromatium vinosum* | 82.8 | 81.6 |
| *Anacystis nidulans* | 66.8 | 71.4 |
| *Anabaena* 7120 | 44.2 | 42.2 |
| *Cryptomonas* Φ | 20.5 | 24.0 |
| *Porphyridium aerugineum* | 17.5 | 14.9 |
| *Chlamydomonas reinhardtii* | 21.7 | 95.3[a] |
| *Nicotiana tabacum* | 27.5 | 56.5[a] |

Only the third position of each codon was considered in the analysis; first position wobble was not included. The single codon amino acids methionine and tryptophan were not included in the analysis

[a] *rbcS* in green algae and land plants is nuclear encoded

photosynthetic groups for which SSU sequence information was available. Notable additions to this analysis include the cyanophyte *Cyanophora paradoxa*, the chromophyte *Olisthodiscus luteus*, and the gymnosperm *Pinus tunbergii*. Gaps in amino acid sequences were scored as missing data. Additional characters were added to the end of the data matrix to represent each of the nine gaps shared by two or more taxa. Gaps were then scored as present or absent and given a weight equal to a single character state change. The tree constructed was unrooted. Two equally most parsimonious trees of 564 steps and a consistency index of 0.764 were generated. Figure 6 shows the bootstrap majority-rule consensus tree, which is identical to one of the most parsimonious trees; branch lengths associated with that most parsimonious tree are given. Differences between the two most parsimonious trees were due to lack of resolution among the non-chl *b* algae. This close relationship is also borne out by the low bootstrap values associated with these branches. As with LSU, two main divisions in the tree are evident that separate *Alcaligenes* and the non-chl *b*-containing algae from *Chromatium* and the lineage leading to chlorophytes. This separation was supported in 100% of the bootstrap analyses. *Prochlorothrix* was again identified with the cyanobacteria (including *Cyanophora*) and most closely associated with *Anacystis*. The bootstrap value for the relationship of *Prochlorothrix* to the cyanobacteria is high (78%) as is the value supporting the association of the green algae and land plants (65%).

## Discussion

### *Structural Conservation of Large and Small Subunits*

The rubisco LSU has been highly conserved over its greater than 1.3 billion years of evolutionary history (time estimated by Ochman and Wilson 1987). As expected, the deduced amino acid sequence of *Prochlorothrix* does contain the conserved amino acids that have been predicted to be involved in the active site (Lys-170, His-293, and Lys-329) and that play a role in carbamylation (Lys-196) during rubisco activation (Andersen and Caton 1987). The remainder of the polypeptide is also highly conserved among the different groups with *Prochlorothrix*, maintaining a percent identity with all other species between 56 and 87%. The most variable portion of the polypeptide was the amino terminus. It has been shown with barley (Poulsen et al. 1979) and spinach (Zurawski et al. 1981) that the amino terminus is posttranslationally modified by the cleavage of 14 amino acids from the polypeptide, leaving alanine as the first amino acid of the functional protein. It is expected that higher sequence divergence would occur in this domain as it may be under relaxed selection. Numerous insertion/deletion events and nonconservative changes appear to have occurred, whereas the level of conservation of amino acids increases after the alanine at position 18 (serine in *Cryptomonas* and *Porphyridium*) of the aligned sequence. The amino acid at position 17 is lysine in all sequences except *Prochlorothrix*, *Alcaligenes*, *Chromatium*, *Cryptomonas*, and *Porphyridium*,

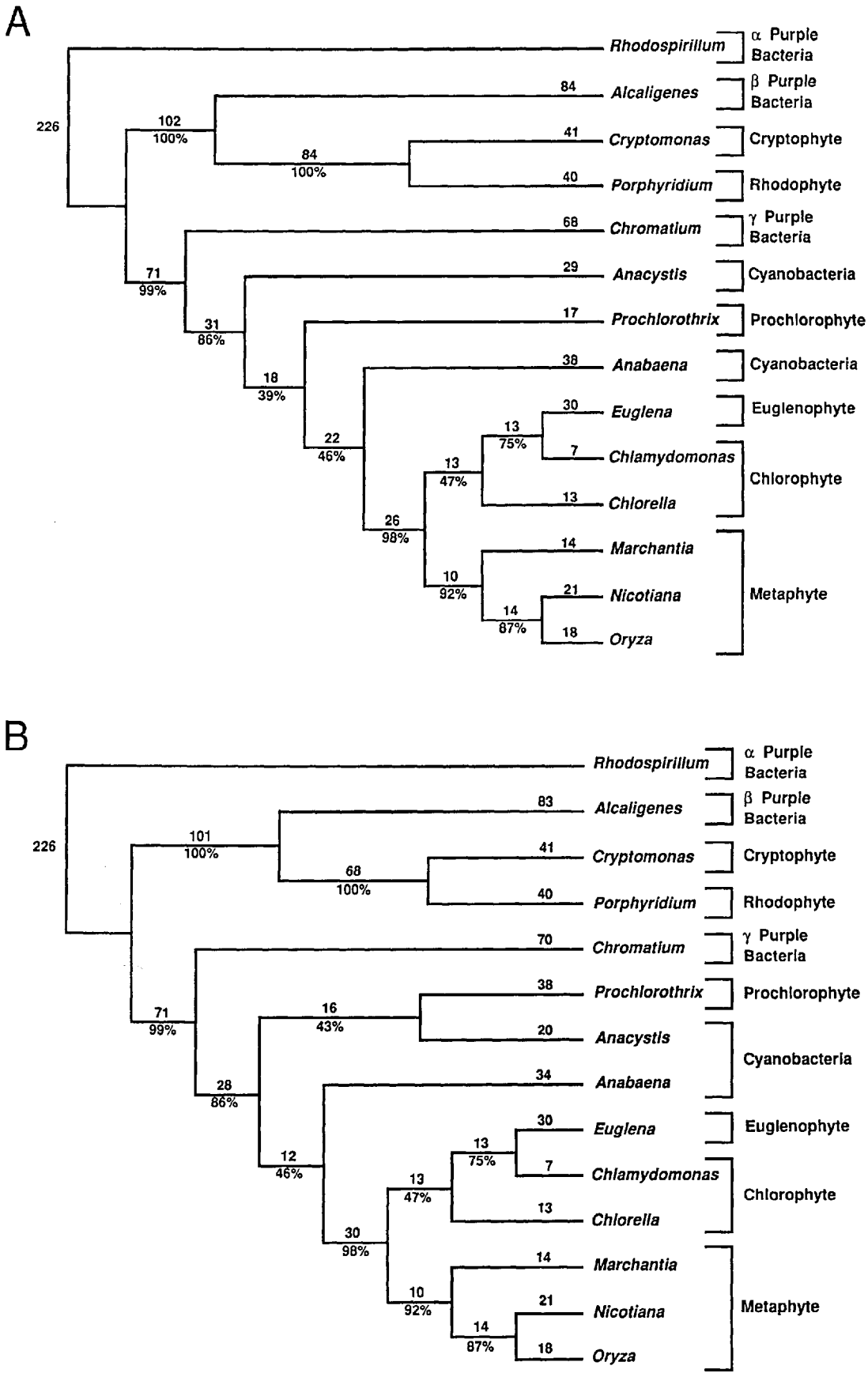


**Fig. 5.** Phylogenetic analysis using Wagner parsimony based on rubisco LSU sequences. Two equally most parsimonious trees were produced with 1031 steps and a consistency index of 0.787. Numbers above the line represent the number of character state changes attributed to that branch. Confidence values for each branch determined by bootstrap analysis are displayed as percentages below that line. Figure 5A, tree 1; Fig. 5B, tree 2.

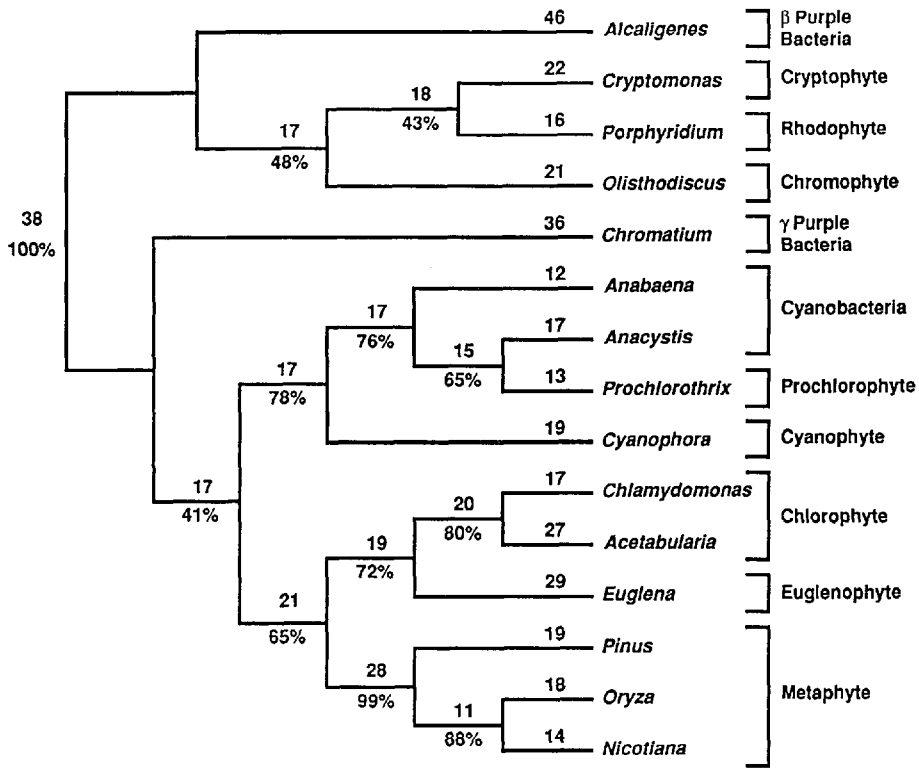


**Fig. 6.** Bootstrap majority-rule consensus tree based on rubisco SSU sequences. Wagner parsimony of rubisco SSU sequences was used to construct two equally most parsimonious trees having 564 steps and a consistency index of 0.764. Numbers above the line represent the number of character state changes attributed to that branch based on the most parsimonious tree that had an identical topology to the consensus tree. Confidence values for each branch determined by bootstrap analysis are displayed as percentages below each line.

which may have bearing on the processing of the protein in addition to its phylogenetic implications.

Certain regions of conservation are present among all rubisco small subunit sequences, suggesting that these may have important functions. Results of x-ray crystallography have shown that each small subunit makes contact with three large subunits and the conserved domains in the small subunits form critical contact zones (Knight et al. 1990; Schneider et al. 1990). Site-directed mutagenesis of specific residues in these zones prevented the accumulation of stable holoenzyme or the ability of subunits to assemble correctly (Fitchen et al. 1990). The large deletion in the SSU of bacteria and nongreen algae is a domain that forms an extensive loop stabilizing hydrogen bonds between adjacent $L_2$ dimers in green algae and plants (Knight et al. 1989, 1990). Because interactions between the large and small subunits are absent among enzymes lacking this domain, no constraints to conserve corresponding residues in the LSU are present, and differences between the LSU sequences of *Anabaena* and spinach have been noted (Knight et al. 1989).

### *Large and Small Subunit Phylogeny*

Phylogenetic analysis of the rubisco LSU indicates that *Prochlorothrix* clusters with the cyanobacteria and, among these, is most closely related to *Anacystis*. This was evident from both nucleic acid/amino acid identities and Wagner parsimony analysis. This was also the result of analyses of 16S rRNA sequences (Turner et al. 1989) and of *psbA* gene sequences (Morden and Golden 1989a,b). In the data presented here, *Anabaena* is more closely associated to the chlorophytes and land plants than are *Prochlorothrix* and *Anacystis* [similar to the results of Turner et al. (1989) and contrasting those of Morden and Golden (1989a)]. However, the confidence limits of this branching arrangement by bootstrap analysis suggest that there may be homoplasy associated with the sequence evolution of these species. Whether this group or some other cyanobacterium is more similar to green plastids is not clear. A strong feature that linked *Prochlorothrix* to green plastids in the *psbA* analysis was a shared seven-amino acid deletion at the carboxy terminus relative to cyanobacterial genes (Morden and Golden 1989a,b). The species here may be at the limit of sequence similarity for phylogenetic determination and only the clustering of major groups is possible without considerably greater amounts of sequence for comparison. Also, only two cyanobacterial *rbcL* sequences were available for comparison; additional sequences might be beneficial in resolving this apparent incongruity.

Relationships inferred from LSU and SSU parsimony analysis and LSU sequence identities are well supported by bootstrap analysis. The clustering of organisms having green plastids (chlorophytes, euglenophytes, and metaphytes) is in agreement with

conventional views (Whatley and Whatley 1981; Cavalier-Smith 1982; Gray 1989). Also, the association of *Porphyridium, Cryptomonas,* and *Olisthodiscus* is not unexpected as it has been previously hypothesized that both cryptophytes and chromophytes are derived from rhodophytes (Whatley and Whatley 1981). However, our results indicate that this lineage is related to the β-purple bacteria (i.e., *Alcaligenes*) rather than to cyanobacteria as was previously suggested (Whatley and Whatley 1981; Cavalier-Smith 1982; Gray 1989), and that cyanobacteria and the chlorophyte lineage are derived from the γ-purple bacteria (i.e., *Chromatium*).

*Plastid Endosymbiosis*

There is strong evidence based on SSU structure and phylogenetic analysis of LSU and SSU which indicate that there have been multiple endosymbiotic origins of plastids. As stated earlier, *Cyanophora paradoxa* shares many features common to both cyanobacteria and plastids and because of this is believed to be the progenitor of all plastid forms (Cavalier-Smith 1982) or at least to the rhodophytes and in turn the other non-chl *b* algae (Whatley and Whatley 1981). Phylogenetic analysis of rubisco SSU (presented here) and 16S rRNA (Giovannoni et al. 1988; Turner et al. 1989) support the cyanelle of *Cyanophora* being closely related to cyanobacteria. There is no evidence based on SSU data to indicate it is ancestral to other classes of organisms as previously thought. However, it is apparent that *Cyanophora* is not associated with the lineage leading to rhodophytes. The *rbcL* data for *Cyanophora* are not yet available.

The *rbcL* analysis suggests the origin of the non-chl *b* algal plastid to have been an organism similar to the β-purple bacterium *A. eutrophus* rather than a cyanophyte. These organisms all share many characteristics of rubisco large and small subunit gene structure and form a strongly supported clade in parsimony analysis. Cryptophytes (containing phycobilins and chl *c*) are believed to have been the result of the endosymbiosis of a rhodophyte (containing phycobilins), and subsequent synthesis of chl *c*, resulting in a plastid with four surrounding membranes and a nucleomorph between the outer and inner two membranes (Whatley and Whatley 1981; Cavalier-Smith 1982). Chromophytes (containing chl *c*) possess four membranes surrounding the plastid with no nucleomorph and are believed to be derived from cryptophytes with concurrent loss of phycobilins and the nucleomorph (Whatley and Whatley 1981; Cavalier-Smith 1982). The results of our study support this hypothesis. *Cryptomonas* and *Olisthodiscus* are closely associated to each other and *Porphyridium* is more distantly related. Sequence data showing that red algal plastids arose from an endosymbiosis separate from green plastids previously have been presented using the proteins ferredoxin and cytochrome *c* (Schwartz and Dayhoff 1981); however purple bacteria were not included in these studies and it was assumed that rhodophyte plastids were derived from cyanobacteria. The endosymbiosis of a purple bacterium giving rise to the plastid of red algae raises questions concerning the origin of phycobilins and oxygenic photosynthesis among rhodophytes. This could have occurred by the independent origin of these complexes in red algae and cyanobacteria or by lateral transfer of genes from a cyanobacterium to a red alga (movement of a gene or gene clusters from one organism to another). Alternatively, the plastids of red algae may have been derived from the endosymbiosis of a cyanobacterium, which at some time may have acquired the *rbcLS* operon from a β-purple bacterium via lateral transfer or by a second endosymbiosis of a β-purple bacterium and lateral transfer among endosymbionts.

Chloroplasts of green algae and land plants have a common origin derived from the cyanobacteria and are apparently well separated phylogenetically from other plastid types. In addition, cyanobacteria and chloroplasts form a lineage derived from the γ-purple bacteria. The taxa within the chloroplast lineage represent a monophyletic group with the exception of the link between *Euglena* and green algae. It previously has been shown that the plastid genome of *Euglena* is distinctly prokaryotic in nature based on plastid $tRNA^{Phe}$ (Chang et al. 1981) and rRNA analysis (Wolters and Erdmann 1988), yet the nuclear DNA is more closely allied to that of trypanosomatid protozoa rather than algal lineages (Chang et al. 1981; Delihas et al. 1981). It was suggested that the plastid of *Euglena* is the result of a separate endosymbiotic event and examination of the three membranes surrounding the plastid indicated that the probable progenitor was a green alga (Gibbs 1978, 1981). Our results show that *Euglena* has a close affinity to *Chlamydomonas,* thus supporting this hypothesis.

*AT vs GC Content*

It previously has been shown that in the course of molecular evolution the GC content of mitochondrial DNAs (mtDNAs) in insects and mammals has diverged considerably (Osawa and Jukes 1989). Even among mammals, human mtDNA has a higher GC content than other species (Osawa and Jukes 1989). A change in GC content has also occurred during the evolution of the plastid. Most purple bacteria and cyanobacteria have a high GC content whereas plastid DNA has a low GC content. Several suggestions have been made regarding potential contributors to determining codon usage and codon bias

including tRNA availability, codon–anticodon H-bond strength, contextual constraints, dinucleotide preference, and overall GC content (Ticher and Graur 1989). However, these do not address the question as to why there has been a switch in GC content between plastid-encoded DNA and its bacterial relatives. It might be proposed that plastids are derived from an AT-rich bacterial endosymbiont. However, evidence we present that *Cryptomonas* and *Porphyridium* (AT rich) are derived from a β-purple bacterium such as *Alcaligenes* (GC rich) suggests this may not be the case.

AT richness seems to be a common feature in the organelle genomes of plants, animals, and fungi (Aota et al. 1988). Two possible explanations may be used to account for a higher AT content in plastids compared to their bacterial progenitors. First, the cost efficiency during replication and transcription of DNA may favor a high AT content because of the number of H-bonds to break during DNA polymerization. Plastid DNA replicates much more rapidly than nuclear DNA at certain stages of plant ontogeny (Lamppa and Bendich 1979; Kuroiwa et al. 1981). Relative chloroplast DNA content increases from approximately 1% to 10% of the total tissue DNA (greater in some species) in the early stages of plant development (Lamppa and Bendich 1979; Kuroiwa et al. 1981). As such, it may be more energetically efficient during replication if the genome has a low, rather than high, GC pressure.

A second factor that could account for the higher AT content (low GC pressure) among plastid genomes is the mutation of C-to-T resulting from the spontaneous deamination of 5-methylcytosine resulting in a thymine residue (Coulondre et al. 1978; Watson et al. 1987) and/or the activity of the deamination-repair system (Muto and Osawa 1987). Although plastid DNA is not typically methylated, there have been recent reports that show plastid DNAs to be methylated at early stages of plant development (Ngernprasirtsiri et al. 1988a,b; Gauly and Kössel 1989) and in ripening tomato fruit (Kobayashi et al. 1990). However, simply having a mechanism to accomplish this is not an indication that it should occur; otherwise all genomes with methylated DNA would be expected to be AT rich. Plant nuclear genomes contain methylated DNAs yet are GC rich as we have indicated. Likewise, animal nuclear genomes are GC rich and there is evidence of increasing GC content in some lineages of animal mitochondrial genomes (Osawa and Jukes 1989).

## *Concluding Remarks*

The phylogenetic relationship of *Prochlorothrix* (and also *Prochloron*) to cyanobacteria and green chloroplasts has been, and continues to be, uncertain. Experiments on physiology and ultrastructure continually point toward *Prochlorothrix* sharing characters of both groups, yet results from two of three studies utilizing molecular data suggest it is not intermediate between them (this study; Turner et al. 1989). The one common factor among all studies based on molecular data is that *Synechococcus* (*Anacystis nidulans*) is most closely related to *Prochlorothrix* of the few cyanobacteria analyzed. Miller and Jacob (1989) pointed out that several concurrent changes must have occurred if chl *b* were to arise in prochlorophytes by some convergent event separate from those of green plastids. These changes include (but are not limited to) thylakoid membrane architecture, the synthesis of chl *b* and chl *a/b* binding proteins, thylakoid stacking, and the deletion of 21 nucleotides at the 3′ end of *psbA* genes. The latter is a strong character that links *Prochlorothrix* to the green plastids. Interestingly, there are 3- and 4-bp direct repeats in the *psbAI* genes of *Synechococcus* (Golden et al. 1986) and *Synechocystis* (Osiewacz and McIntosh 1987), respectively, flanking the region deleted in *psbA* genes of *Prochlorothrix*, green algae, and land plants. If homologous recombination of the direct repeats were to occur the deletion would exactly correspond to the position of the 21-nucleotide (seven-amino acid) gap at the carboxy terminus of green plastid and *Prochlorothrix* genes. However, recombination between 7- and 8-bp direct repeats in the same region of the *psbA* genes of *Anabaena* (Vrba and Curtis 1989) and *Fremyella* (B. Mulligan et al. 1984) would result in deletion of only six of these amino acid residues.

Future studies to elucidate the phylogenetic relationship of *Prochlorothrix* to cyanobacteria and green plastids should be directed at specific genes and gene complexes that show differences between green chloroplasts and cyanobacteria. Investigations in other laboratories of the chl *a/b* binding proteins and characteristics of the water oxidation complex may be helpful in determining the phylogenetic position of this interesting organism.

*Acknowledgments.* We thank Karen Greer, Rixin Li, and Monica Smith for assistance in cloning the *rbcLS* operon, Mark Nalty for helping to map the transcript, and Martin Mulligan for assistance in promoter scoring. We also thank Maria Kuhsel, John Logsdon, Dick Olmstead, Jeff Palmer, and Ken Wolfe for discussions concerning various aspects of this project and manuscript. This work was supported by funds from the College of Science at Texas A&M University and by National Science Foundation grant DMB-8958089 and equipment grant BBS-8703784.

*Note added in proof:* The nucleotide sequence of the *Prochlorothrix hollandica rbcLS* operon has been submitted to the EMBL data bank with the accession number X57359.

Since preparation of this manuscript the *rbcL* sequence from *Cyanophora paradoxa* has been published (Valentin K, Zetsche K (1990) Nucleotide sequence of the large subunit of rubisco from *Cyanophora paradoxa*—phylogenetic implications. Curr Genet 18:199–202). Addition of this sequence to the data set

used to generate Fig. 5 resulted in a tree in which *C. paradoxa* formed a branch between *Anabaena* and the green algae. All other branches were as shown in Fig. 5B.

## References


Aitken A, Stanier R (1979) Characterization of peptidoglycan from the cyanelles of *Cyanophora paradoxa*. J Gen Microbiol 112:219–223

Andersen K, Caton J (1987) Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose bisphosphate carboxylase large and small subunit genes and their products. J Bacteriol 169:4547–4558

Aota S, Gojobori T, Ishibashi F, Maruyama T, Ikemura T (1988) Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res 16:r315–r402

Bayer MG, Maier TL, Gebhart UB, Schenk HEA (1990) Cyanellar ferrodoxin-$NADP^+$–oxidoreductase of *Cyanophora paradoxa* is encoded by the nuclear genome and synthesized on cytoplasmatic 80S ribosomes. Curr Genet 17:265–267

Boczar BA, Delaney TP, Cattolico RA (1989) Gene for the ribulose-1,5-bisphosphate carboxylase small subunit protein of the marine chromophyte *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium. Proc Natl Acad Sci USA 86:4996–4999

Breiteneder K, Seiser C, Löffelhardt W, Michalowski C, Bohnert HJ (1988) Physical map and protein gene map of cyanelle DNA from the second known isolate of *Cyanophora paradoxa* (Kies-strain). Curr Genet 13:199–206

Bremer B, Bremer K (1989) Cladistic analysis of blue-green procaryote interrelationships and chloroplast origin based on 16S rRNA oligonucleotide catalogues. J Evol Biol 2:13–30

Bryant DA, de Lorimier R, Lambert DH, Dubbs JM, Stirewalt VL, Stevens SE Jr, Porter RD, Tam J, Jay E (1985) Molecular cloning and nucleotide sequence of the $\alpha$ and $\beta$ subunits of allophycocyanin from the cyanelle genome of *Cyanophora paradoxa*. Proc Natl Acad Sci USA 82:3242–3246

Bullerjahn GS, Matthijs HCP, Mur LR, Sherman LA (1987) Chlorophyll-protein composition of the thylakoid membrane from *Prochlorothrix hollandica*, a prokaryote containing chlorophyll *b*. Eur J Biochem 168:295–300

Burger-Wiersma T, Post AF (1989) Functional analysis of the photosynthetic apparatus of *Prochlorothrix hollandica* (Prochlorales), a chlorophyll *b* containing procaryote. Plant Physiol 91:770–774

Burger-Wiersma T, Veenhuis M, Korthals HJ, Van de Wiel CCM, Mur LR (1986) A new prokaryote containing chlorophylls *a* and *b*. Nature 320:262–264

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

Cavalier-Smith T (1982) The origins of plastids. Biol J Linn Soc 17:289–306

Chan RL, Keller M, Canaday J, Weil J-H, Imbault P (1990) Eight small subunits of *Euglena* ribulose 1-5 bisphosphate carboxylase are translated from a large mRNA as a polyprotein. EMBO J 9:333–338

Chang SH, Hecker LI, Brum CK, Schnabel JJ, Heckman JE, Silberklang M, RajBhandary UL, Barnett EE (1981) The nucleotide sequence of *Euglena* cytoplasmic phenylalanine transfer RNA. Evidence for possible classification of *Euglena* among the animal rather than the plant kingdom. Nucleic Acids Res 9:3199–3204

Coulondre C, Miller JH, Farabaugh PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 274:775–780

Curtis SE, Haselkorn R (1983) Isolation and sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase from the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 80:1835–1839

Curtis SE, Haselkorn R (1984) Isolation, sequence and expression of two members of the 32 kd thylakoid membrane protein gene family from the cyanobacterium *Anabaena* 7120. Plant Mol Biol 3:249–258

Delihas N, Andersen J, Andresini W, Kaufman L. Lyman H (1981) The 5S ribosomal RNA of *Euglena gracilis* cytoplasmic ribosomes is closely homologous to the 5S RNA of the trypanosomatid protozoa. Nucleic Acids Res 9:6627–6633

Devereux JR (1989) Sequence analysis software package of the genetics computer group, version 6.0. University of Wisconsin Biotechnology Center, Madison WI

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Douglas S, Durnford DG (1989) The small subunit of ribulose-1,5-bisphosphate carboxylase is plastid-encoded in the chlorophyll c-containing alga *Cryptomonas* Φ. Plant Mol Biol 13: 13–20

Douglas S, Durnford DG, Morden CW (1990) Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from the chlorophyll *c*-containing alga, *Cryptomonas* Φ: evidence supporting the polyphyletic origin of plastids from purple bacteria. J Phycol 26:500–508

Dron M, Rahire M, Rochaix J-D (1982) Sequences of the chloroplast DNA region of *Chlamydomonas reinhardii* containing the gene of the large subunit of ribulose bisphosphate carboxylase and parts of its flanking genes. J Mol Biol 162:775–793

Ellis RJ (1985) Synthesis, processing, and assembly of polypeptide subunits of ribulose-1,5-bisphosphate carboxylase/oxygenase. In: Steinback KE, Bonitz S, Arntzen CJ, Bogorad L (eds) Molecular biology of the photosynthetic apparatus. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp 339–347

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

Fitchen JH, Knight S, Andersson I, Brändén C-I, McIntosh L (1990) Residues in three conserved regions of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase are required for quaternary structure. Proc Natl Acad Sci USA 87:5768–5772

Gauly A, Kössel H (1989) Evidence for tissue-specific cytosine-methylation of plastid DNA from *Zea mays*. Curr Genet 15: 371–376

Gibbs SP (1978) The chloroplasts of *Euglena* may have evolved from symbiotic green algae. Can J Bot 56:2883–2889

Gibbs SP (1981) Chloroplasts of some groups may have evolved from endosymbiotic eukaryotic algae. Ann NY Acad Sci 361: 193–207

Gibson JL, Tabita FR (1977) Different forms of D-ribulose-1,5-bisphosphate carboxylase from *Rhodopseudomonas sphaeroides*. J Biol Chem 252:943–949

Gingrich JC, Hallick RB (1985a) The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene I. Complete DNA sequence and analysis of the nine intervening sequences. J Biol Chem 260:16156–16161

Gingrich JC, Hallick RB (1985b) The *Euglena gracilis* chloroplast ribulose-1,5-bisphosphate carboxylase gene II. The spliced mRNA and its product. J Biol Chem 260:16162–16168

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

Golden SS, Brusslan J, Haselkorn R (1986) Expression of a family of *psbA* genes encoding a photosystem II polypeptide in the cyanobacterium *Anacystis nidulans* R2. EMBO J 5: 2789–2798

Golden SS, Brusslan J, Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods Enzymol 153: 215–231

Goldschmidt-Clermont M, Rahire M (1986) Sequence, evolution and differential expression of the two genes encoding variant small subunits of ribulose bisphosphate carboxylase/oxygenase in *Chlamydomonas reinhardtii*. J Mol Biol 191: 421–432

Grantham R, Gautier C, Gouy M (1980) Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res 8:1893–1912

Gray MW (1989) The evolutionary origins of organelles. Trends Genet 5:294–299

Hallenbeck PL, Kaplan S (1988) Structural gene regions of *Rhodobacter sphaeroides* involved in $CO_2$ fixation. Photosynth Res 19:63–71

Hwang S-R, Tabita FR (1989) Cloning and expression of the chloroplast-encoded *rbcL* and *rbcS* genes from the marine diatom *Cylindrotheca* sp. strain N1. Plant Mol Biol 13:69–79

Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 2:13–34

Kassavetis GA, Geiduschek EP (1982) Bacteriophage T4 late promoters: mapping 5′ ends of T4 gene 23 mRNAs. EMBO J 1:107–114

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

Knight S, Andersson I, Brändén C-I (1989) Reexamination of the three-dimensional structure of the small subunit of RuBisCo from higher plants. Science 244:702–705

Knight S, Andersson I, Brändén C-I (1990) Crystallographic analysis of ribulose 1,5-bisphosphate carboxylase from spinach at 2.4 Å resolution. J Mol Biol 215:113–160

Kobayashi H, Ngernprasirtsiri J, Akazawa T (1990) Transcriptional regulation and DNA methylation in plastids during transitional conversion of chloroplasts to chromoplasts. EMBO J 9:307–313

Kuroiwa T, Suzuki T, Ogawa K, Kawano S (1981) The chloroplast nucleus: distribution, number, size, and shape, and a model for the multiplication of the chloroplast genome during chloroplast development. Plant Cell Physiol 22:381–396

Lambert DH, Bryant DA, Stirewalt VL, Dubbs JM, Stevens SE Jr, Porter RD (1985) Gene map for the *Cyanophora paradoxa* cyanelle genome. J Bacteriol 164:659–664

Lamppa GK, Bendich AJ (1979) Changes in chloroplast DNA levels during development of pea (*Pisum sativum*). Plant Physiol 64:126–130

Lemaux PG, Grossman AR (1985) Major light-harvesting polypeptides encoded in polycistronic transcripts in a eukaryotic alga. EMBO J 4:1911–1919

Lewin RA (1975a) Associations of microscopic algae with didemnid ascidians. Phycologia 14:149–152

Lewin RA (1975b) Extraordinary pigment composition of a prokaryotic alga. Nature 256:735–737

Lewin RA (1981) *Prochloron* and the theory of symbiogenesis. Ann NY Acad Sci 361:325–328

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor NY

Martin PG, Dowd JM, Stone SJL (1983) The study of plant phylogeny using amino acid sequences of ribulose-1,5-bisphosphate carboxylase. II The analysis of small subunit data to form phylogenetic trees. Aust J Bot 31:411–419

Martin PG, Boulter D, Penny D (1985) Angiosperm phylogeny studied using sequences of five macromolecules. Taxon 34: 393–400

Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65: 499–559

Maxwell ES, Liu J, Shively JM (1986) Nucleotide sequences of *Cyanophora paradoxa* cellular and cyanelle-associated 5S ribosomal RNAs: the cyanelle as a potential intermediate in plastid evolution. J Mol Evol 23:300–304

Meagher RB, Berry-Lowe S, Rice K (1989) Molecular evolution of the small subunit of ribulose bisphosphate carboxylase: nucleotide substitution and gene conversion. Genetics 123: 845–863

Miller KR, Jacob JS (1989) On *Prochlorothrix*. Nature 338: 303–304

Miller KR, Jacob JS, Burger-Wiersma T, Matthijs HCP (1988) Supramolecular structure of the thylakoid membrane of *Prochlorothrix hollandica*, a chlorophyll *b*-containing prokaryote. J Cell Sci 91:577–586

Moon E, Kao T-H, Wu R (1987) Rice chloroplast DNA molecules are heterogeneous as revealed by DNA sequences of a cluster of genes. Nucleic Acids Res 15:611–630

Morden CW, Golden SS (1989a) *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts. Nature 337: 382–385

Morden CW, Golden SS (1989b) *psbA* genes indicate common ancestry of prochlorophytes and chloroplasts, corrigendum. Nature 339:400

Müller K-D, Salnikow J, Vater J (1983) Amino acid sequence of the small subunit of D-ribulosebisphosphate carboxylase/oxygenase from *Nicotiana tabacum*. Biochim Biophys Acta 742:78–83

Mulligan B, Schultes N, Chen L, Bogorad L (1984) Nucleotide sequence of a multiple-copy gene for the B protein of photosystem II of a cyanobacterium. Proc Natl Acad Sci USA 81:2693–2697

Mulligan ME, Hawley DK, Entriken R, McClure WR (1984) *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase activity. Nucleic Acids Res 12:789–800

Muto A, Osawa S (1987) The guanine and cytosine content of genomic DNA and bacterial evolution. Proc Natl Acad Sci USA 84:166–169

Nargang FL, McIntosh L, Somerville C (1984) Nucleotide sequence of the ribulosebisphosphate carboxylase gene from *Rhodospirillum rubrum*. Mol Gen Genet 193:220–224

Ngernprasirtsiri J, Kobayashi H, Akazawa T (1988a) DNA methylation occurred around lowly expressed genes of plastid DNA during tomato fruit development. Plant Physiol 88: 16–20

Ngernprasirtsiri J, Kobayashi H, Akazawa T (1988b) DNA methylation as a mechanism of transcriptional regulation in nonphotosynthetic plastids in plant cells. Proc Natl Acad Sci USA 85:4750–4754

Nierzwicki-Bauer SA, Curtis SE, Haselkorn R (1984) Cotranscription of genes encoding the small and large subunits of ribulose-1,5-bisphosphate carboxylase in the cyanobacterium *Anabaena* 7120. Proc Natl Acad Sci USA 81:5961–5965

Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. J Mol Evol 26:74–86

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. Nature 322:572–574

Osawa S, Jukes TH (1989) Codon reassignment (codon capture) in evolution. J Mol Evol 28:271–278

Osiewacz JD, McIntosh L (1987) Nucleotide sequence of a

member of the *psbA* multigene family from the unicellular cyanobacterium *Synechocystis* 6803. Nucleic Acids Res 15: 10585

Palmer JD, Jansen RK, Michaels HJ, Chase MW, Manhart JR (1988) Chloroplast DNA variation and plant phylogeny. Ann Mo Bot Gard 75:1180–1206

Poulsen C, Martin B, Svendsen IB (1979) Partial amino acid sequence of the large subunit of ribulosebisphosphate carboxylase from barley. Carlsberg Res Commun 44:191–199

Ritland K, Clegg MT (1987) Evolutionary analysis of plant DNA sequences. Am Nat 130:S74–S100

Sagan L (1967) On the origin of mitosing cells. J Theor Biol 14:225–274

Schaefer MR, Golden SS (1989) Differential expression of members of a cyanobacterial *psbA* gene family in response to light. J Bacteriol 171:3973–3981

Schimper AFW (1883) Über die Entwicklung der Chlorophyll Körner und Farb-Körner. Bot Zeitung 41:105–114

Schneider G, Knight S, Andersson I, Brändén C-I, Lindqvist Y, Lundqvist T (1990) Comparison of the crystal structures of $L_2$ and $L_8S_8$ rubisco suggests a functional role for the small subunit. EMBO J 9:2045–2050

Schneider SU, Leible MB, Yang X-P (1989) Strong homology between the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase of two species of *Acetabularia* and the occurrence of unusual codon usage. Mol Gen Genet 218:445–452

Schwartz RM, Dayhoff MO (1981) Chloroplast origins: inferences from protein and nucleic acid sequences. Ann NY Acad Sci 361:260–269

Shinozaki K, Sugiura M (1982) The nucleotide sequence of the tobacco chloroplast gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Gene 20:91–102

Shinozaki K, Yamada C, Takahata N, Sugiura M (1983) Molecular cloning and sequence analysis of the cyanobacterial gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase. Proc Natl Acad Sci USA 80:4050–4054

Spratt BG, Hedge PJ, te Heesen S, Edelman A, Broome-Smith JK (1986) Kanamycin-resistant vectors that are analogues of plasmids pUC8, pUC9, pEMBL8 and pEMBL9. Gene 41: 337–342

Starnes SM, Lambert DH, Maxwell ES, Stevens SE Jr, Porter RD, Shively JM (1985) Cotranscription of the large and small subunit genes of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Cyanophora paradoxa*. FEMS Lett 28:165–169

Stiekema WJ, Wimpee CF, Tobin EM (1983) Nucleotide sequence encoding the precursor of the small subunit of ribulose 1,5-bisphosphate carboxylase from *Lemna gibba* L.G-3. Nucleic Acids Res 11:8051–8061

Swofford DL (1989) Paup version 3.0. Illinois Natural History Survey, Champaign IL.

Tabita FR, McFadden BA (1974) D-Ribulose 1,5-diphosphate carboxylase from *Rhodospirillum rubrum*, II. Quaternary structure, composition, catalytic and immunological properties. J Biol Chem 249:3459–3464

Ticher A, Graur D (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J Mol Evol 28:286–298

Turner NE, Robinson SJ, Haselkorn R (1983) Different promoters for the *Anabaena* glutamine synthetase gene during growth using molecular or fixed nitrogen. Nature 306:337–342

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

Valentin K, Zetsche K (1989) The genes of both subunits of ribulose-1,5-bisphosphate carboxylase constitute an operon on the plastome of a red alga. Curr Genet 16:203–209

Viale AM, Kobayashi J, Akazawa T (1989) Expressed genes for plant-type ribulose 1,5-bisphosphate carboxylase/oxygenase in the photosynthetic bacterium *Chromatium vinosum*, which possesses two complete sets of the genes. J Bacteriol 171:2391–2400

Volkman JK, Burger-Wiersma T, Nichols PD, Summons RE (1988) Lipids and chemotaxonomy of *Prochlorothrix hollandica*, a planktonic prokaryote containing chlorophylls *a* and *b*. J Phycol 24:554–559

Voordouw G, DeVries PA, Van den Berg WAM, DeClerck EPJ (1987) Site-directed mutagenesis of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Anacystis nidulans*. Eur J Biochem 163:591–598

Vrba JM, Curtis SE (1989) Characterization of a four-member *psbA* gene family from the cyanobacterium *Anabaena* PCC 7120. Plant Mol Biol 14:81–92

Waksman G, Freyssinet G (1987) Nucleotide sequence of a cDNA encoding the ribulose-1,5-bisphosphate carboxylase/oxygenase from sunflower (*Helianthus annuus*). Nucleic Acids Res 15:1328

Watson JK, Hopkins NH, Roberts JW, Steitz JA, Weiner AM (1987) Molecular biology of the gene. Benjamin/Cummings, Menlo Park, CA

Whatley JM (1981) Chloroplast evolution-ancient and modern. Ann NY Acad Sci 361:154–164

Whatley JM, Whatley FR (1981) Chloroplast evolution. New Phytol 87:233–247

Woese CR (1987) Bacterial evolution. Microbiol Rev 51:221–271

Wolter FP, Fritz CC, Willmitzer L, Schell J, Schreier PH (1988) *rbcS* genes in *Solanum tuberosum*: conservation of transit peptide and exon shuffling during evolution. Proc Natl Acad Sci USA 85:846–850

Wolters J, Erdmann VA (1988) Cladistic analysis of ribosomal RNAs—the phylogeny of eukaryotes with respect to the endosymbiotic theory. BioSystems 21:209–214

Xie Y, Wu R (1988) Nucleotide sequence of a ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit gene (*rbcS*) in rice. Nucleic Acids Res 16:7749

Yamamoto N, Kano-Murakami Y, Matsuoka M, Ohashi Y, Tanaka Y (1988) Nucleotide sequence of a full length cDNA clone of ribulose bisphosphate carboxylase small subunit gene from green dark-grown pine (*Pinus tunbergii*) seedling. Nucleic Acids Res 16:11830

Yanisch-Perron C, Vieira J, Messing J (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC vectors. Gene 33:103–119

Yoshinaga K, Ohta T, Suzuki Y, Sugiura M (1988) *Chlorella* chloroplast DNA sequence containing a gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase and a part of a possible gene for $\beta'$ subunit of RNA polymerase. Plant Mol Biol 10:245–250

Zurawski G, Perrot B, Bottomley W, Whitfeld PR (1981) The structure of the gene for the large subunit of ribulose 1,5-bisphosphate carboxylase from spinach chloroplast DNA. Nucleic Acids Res 9:3251–3270

Received August 8, 1990/Revised November 24, 1990/Accepted December 3, 1990