Plant Molecular Biology **17**: 853–863, 1991.
© 1991 *Kluwer Academic Publishers. Printed in Belgium.*

Evolution of the Rubisco operon from prokaryotes to algae: Structure and analysis of the *rbcS* gene of the brown alga *Pylaiella littoralis*

Nour-Eddine Assali[1,4], William F. Martin[2], Charles C. Sommerville[3] and Susan Loiseaux-de Goër[3]*
[1] *Laboratoire de Biologie Moléculaire Végétale, CNRS, URA 57, Université Joseph Fourier, Grenoble;*
[2] *Institut für Genetik, Technische Universität Braunschweig, Postfach 3329, D-3300 Braunschweig, FRG;*
[3] *Station Biologique de Roscoff, CNRS UPR 4601, BP 74, 29680 Roscoff, France (*author for correspondence);* [4] *present address: Laboratoire de Biochimie, I.A.V. Hassan II, BP 6202, Rabat, Morocco*

Received 25 September 1990; accepted in revised form 21 June 1991

Key words: algae, endosymbiosis, evolution, plastidial DNA, Rubisco operon

Abstract

The *rbcS* gene coding for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) of the brown alga *Pylaiella littoralis* is located within the plastid genome and is transcribed as a single polycistronic mRNA with the gene for the large subunit of Rubisco, *rbcL*. The structure of the Rubisco operon from *P. littoralis* was determined. Molecular phylogenies for *rbcS* and *rbcL* with a wide range of prokaryotes and eukaryotes were constructed which are congruent with recent evidence for polyphyletic plastid origins. Both *rbcL* and *rbcS* of the β-purple bacterium *Alcaligenes eutrophus* clearly cluster with the rhodophyte and chromophyte proteins. The data suggest that the Rubisco operons of red algal and chromophytic plastids derive from β-purple eubacterial antecedents, rather than the cyanobacterial lineage of eubacteria from which other of their genes derive. This implies a lateral transfer of Rubisco genes from β-purple eubacterial ancestors to the cyanobacterial ancestor of rhodophyte and chromophyte plastids.

Introduction

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) is one of the most widely studied plant or bacterial enzymes. It catalyses both the carboxylation of ribulose bisphosphate in CO_2 fixation and oxygenation of the same substrate during photorespiration [26]. Rubisco has a quaternary structure of eight large and eight small subunits (L_8S_8) except in some bacteria, where it can assume a stoichiometry of L_6S_6, or be composed of large subunits only, i.e. L_8, L_6, L_{4-6}, L_2 [14]. Though a wealth of literature has been compiled for Rubisco of chlorophytes, which possess chlorophyll *a* and *b*, little is yet known about Rubisco of chromophytes or rhodophytes, which

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number X55372.

perform photosynthesis with the aid of chlorophyll *c* or phycobiliproteins, respectively. In green algae, *Euglena* and land plants, the small subunit of Rubisco (*rbcS*) is nuclear-encoded, whereas the large subunit *rbcL*) is a component of the plastid DNA. Conversely, in all chromophytes and rhodophytes studied to date, both *rbcL* and *rbcS* genes are located on the plastid genome and form an operon [9, 10, 22, 30, 41–43]. A complete Rubisco operon is also found in the cyanoplasts (formerly known as cyanelles) of *Cyanophora paradoxa* [37].

It is now generally accepted that plastids initially arose through endosymbiosis between a photosynthetic prokaryote and a non-photosynthetic host cell. Molecular data from chloroplasts (plastids of green algae and land plants) suggest that these derive from a cyanobacterium-like endosymbiont [16, 28]. The plastids of non-green algae are structurally quite distinct from green chloroplasts [12] and could have a distinct phylogeny. Recent studies indicate that the *rbcS* gene of *Olisthodiscus luteus* [4], the *rbcL* gene of *Pylaiella littoralis* [2], and both the *rbcL* and *rbcS* of *Ectocarpus siliculosis* [42] are quite similar to their analogues from *Alcaligenes eutrophus*, a β-purple eubacterium as designated by Woese [47]. Here we report the sequence of the plastid encoded *rbcS* gene of the brown alga *Pylaiella littoralis* and the structure of the Rubisco operon. Phylogenetic analyses made from the deduced amino acid sequences of *rbcS* and *rbcL* genes of prokaryotes and eukaryotes are presented and discussed.

Materials and methods

DNA sequencing

The fragment E18, and *Eco* RI fragment of the *P. littoralis* plastid genome which contains the *rbcS* gene, was cloned into M13 mp18 and pUC18 as described previously [2]. The dideoxy chain termination method was used to sequence the fragments using the Amersham sequencing kit (Amersham France SA) for single-strand DNA, and the Pharmacia T7 sequencing kit (Pharmacia LKB) for double-strand DNA. Nucleotide sequences were analysed by the CITI2 programs (Paris) and the alignments were performed with the LINEUP program of the WISGEN Incorporation package [7].

Data handling and dendrograms for rbcL *and* rbcS

Comparisons of 18 *rbcL* sequences were based upon the alignment previously published [2]. Terminal and internal regions of length heterogeneity which could result in distortion of amino acid identities were removed from the alignment such that only positions occupied in all members of the set were compared (468 positions in each *rbcL* sequence). Amino acid identities at these positions were calculated (153 pairwise comparisons). Identities were converted into fractions of substituted residues (divergence), which were corrected for multiple substitutions at identical sites [8]. The resulting matrix of corrected values was used to generate the dendrogram shown in Fig. 4 by the Neighbour Joining (NJ) method [32]. The NJ method was chosen since it has been shown to be more efficient than many other commonly used matrix [36] or molecular sequence [17, 31] methods in recovery of correct topology under a variety of data parameters.

For the *rbcS*, 15 amino acid sequences were aligned with the LINEUP program of the GCG sequence analysis software package [7]. The alignment is shown in Fig. 1. The removal of amino acid sites not occupied in all members of the sequence set yielded 103 positions for comparison in each sequence. Identity matrices, fractions of substituted residues, corrected divergence and trees for *rbcS* were generated as for *rbcL*.

Hybridization analysis

Plastidial DNA was purified from isolated plastids and nuclear DNA from whole cells as previously described [6]. DNAs were digested to completion with *Eco* RI, transferred to nylon

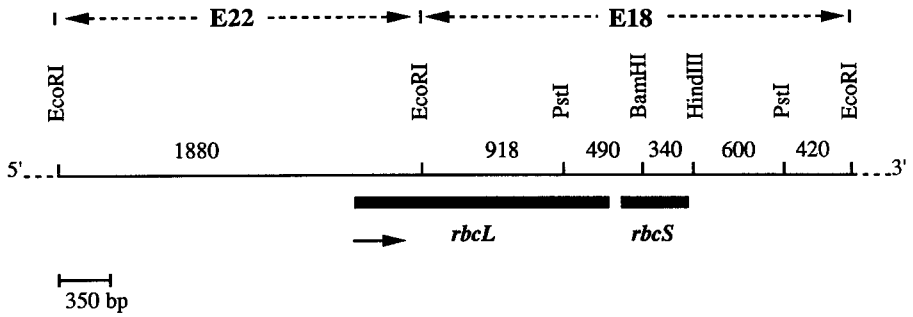


Fig. 1. Physical map of the Rubisco operon found on the large molecule of the *Pylaiella littoralis* plastidial genome. E22 and E18 refer to *Eco* RI fragments described by Gunderson *et al.* [18]. Numbers above the fine line indicate restriction fragment length in bp. Heavy lines indicate the position and size of the *rbcL* and *rbcS* genes. The arrow shows the direction of transcription.

membranes (GeneScreen Plus, Dupont-NEN), and hybridized under low-stringency conditions (rapid hybridization system, Amersham, at 42 °C) with *rbcS* probes from *Anabaena* (gift from R. Haselkorn), from tobacco (gift from J. Flech) and from *Pylaiella*. The probes were labelled using a random-primed labelling kit (Boehringer Mannheim, France) and [α-^{32}P]-dCTP (3000 Ci/mmol, Amersham).

RNA extraction and northern analysis

RNA was purified from isolated plastids by several extractions with phenol-chloroform (plastids in 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 2% SDS, 2% sodium sarcosinate, 10 mM EDTA, 1% β-mercaptoethanol). The nucleic acids in the supernatant were then precipitated by ethanol, resuspended in sterile distilled water and adjusted to 2 M lithium chloride in order to selectively precipitate the RNA. The lithium chloride precipitation was then repeated a second time. RNAs were resolved on an agarose gel in the presence of methyl mercury, transferred to GeneScreen Plus membranes, and hybridized according to the manufacturer's specifications. Two DNA fragments of *P. littoralis*, one internal to the *rbcL* gene and the other to the *rbcS* gene, were purified twice by electrophoresis, labelled by random priming with [α-^{32}P]-dCTP, and used as probes.

Results

Localization of the rbcS *gene*

Although the *rbcL* of *P. littoralis* was localized by hybridization with DNA probes from green plants, attempts to localize the *rbcS* gene of *P. littoralis*, either on plastidial or on nuclear genomes, by hybridization with heterologous probes from *Anabaena* and tobacco were unsuccessful. The same result was obtained for the chromophytes *O. luteus* [30] and *Cryptomonas* Φ [8], and the rhodophyte *Porphyridium aerugineum* [39], due to the fact that the *rbcS* genes of these species are significantly less homologous to those of cyanobacteria or land plants than are their *rbcL* genes. The *rbcS* gene of *P. littoralis* was localized by sequencing the 3′ flanking region of the *rbcL* gene and comparing the open reading frame, starting 195 bp downstream, to the sequence of the *rbcS* gene found in other species. The restriction map of this region is shown in Fig. 1. The nucleotide sequence of the *rbcS* gene, the flanking regions and the predicted amino acid sequence are represented in Fig. 2. Subsequent hybridizations on plastidial and nuclear DNA of *P. littoralis* with the *Pylaiella rbcS* probe detected only one copy of the *rbcS* gene in this alga.

Fig. 2. Nucleotide sequence of the *P. littoralis rbcS* gene and its flanking regions. The deduced amino acid sequence of the 3′ end of the *rbcL* gene and the *rbcS* gene are included. The inverse repeat downstream of the coding region is indicated by arrows.

The 5′ and 3′ flanking regions

The spacer region between the *rbcL* gene and the *rbcS* gene of *P. littoralis* is 195 bp long. Unlike those described for *C. vinosum* [45], *O. luteus* [4], *E. siliculosus* [42] and *Cryptomonas* Φ [10], no palindromic or repeated sequences were found in this spacer. Neither were any promoter-like sequences found. The spacer region of *Pylaiella* is 67% homologous to that of *E. siliculosus* suggesting a close relationship between the two algae. No significant homology was found with any of the other Rubisco operon spacer regions sequenced to date. It is to be noted that in the three red algae studied so far [22, 41, 43] translation of the *rbcS* genes starts at a GTG codon instead of an ATG codon, and an in-frame GTG immediately precedes the ATG putative start codons of *Pylaiella* and *Ectocarpus*.

A long inverted repeat sequence of 69 bp is found at the end of the *rbcS* gene of *P. littoralis*, beginning 10 nucleotides after the stop codon. The calculated ΔG of the stem-loop is −420 kJ. The capability to form a stem-loop has been correlated with the ability to protect upstream mRNA segments from exonucleolytic degradation by interacting with specific proteins which could have a role in the processing and stability of plastid mRNAs [39].

Comparison of different rbcS *genes*

The *rbcS* gene of *P. littoralis* is 417 bp long (ATG to TAA inclusive), and its deduced amino acid sequence is 98.5% homologous to that of *E. siliculosus*, another brown alga. Amino acid sequences from several species have been aligned for comparison (Fig. 3). The amino terminus of the small subunit of Rubisco from the brown alga

Fucus serratus [21] has been included in this comparison. It is clear from this alignment that these 15 sequences segregate into two homogeneous groups based on regions of sequence homology, particularly in the terminal domains. One group is related to the Rubisco gene of the β-purple bacterium *Alcaligenes eutrophus* and the other to that of the cyanobacteria. Rubisco genes from red algae, cryptophytes and chromophytes are among those which cluster with the β-purple bacterium. The Rubisco small subunits of one group share little homology with those of the other group. Only 10 amino acids are conserved in all species, and only two small regions are very similar: those at positions 22 to 28 and 79 to 89 in Fig. 3.

The Rubisco small subunits belonging to the 'β-purple eubacterium-like' lineage are much more conserved between species than are those belonging to the 'cyanobacterium-like' lineage, while the large subunits are equally conserved in both lineages. Most of the *rbcS* genes belonging to the 'cyanobacterium-like' lineage are nuclear-encoded, and these all occur in chlorophyll *b*-containing plants. A characteristic feature of all nuclear-encoded *rbcS* proteins is the presence of a 12 to 18 amino acid insertion starting at position 57 in our alignment (Fig. 3), which are not present in the plastidial genes. On the basis of conserved residues and intron-exon structure [48], this fragment may share a common origin in those species in which it is found. The *rbcS* genes belonging to the 'β-purple eubacterial lineage' are characterized by three conserved regions which can be considered as lineage signatures. These correspond to the first six amino acids MRxTQG, to an amino acid sequence T(D)D(P)HRN starting at position 52 (Fig. 3), and to the long carboxy-terminal region. The sequence at amino acids 52–59 corresponds to the variable region surrounding the 'additional' region in the green plant proteins. The carboxy-terminal region is 31 to 36 amino acids longer (one quarter of the molecule) in this lineage than in the 'cyanobacterial-like lineage'. As the highly conserved regions in each group are absent in the other, their significance to the function of the small subunit in the holoenzyme is questionable.

Co-transcription of the rbcL *and* rbcS *genes*

Northern analysis using intragenic probes from the *rbcL* and the *rbcS* genes of *P. Littoralis* showed that these genes are co-transcribed on a single mRNA of 2.4 kb (not shown) as in other chromophyte and rhodophyte algae. No smaller mRNA products were detected indicating that no independently transcribed *rbcS* gene is present in *Pylaiella*.

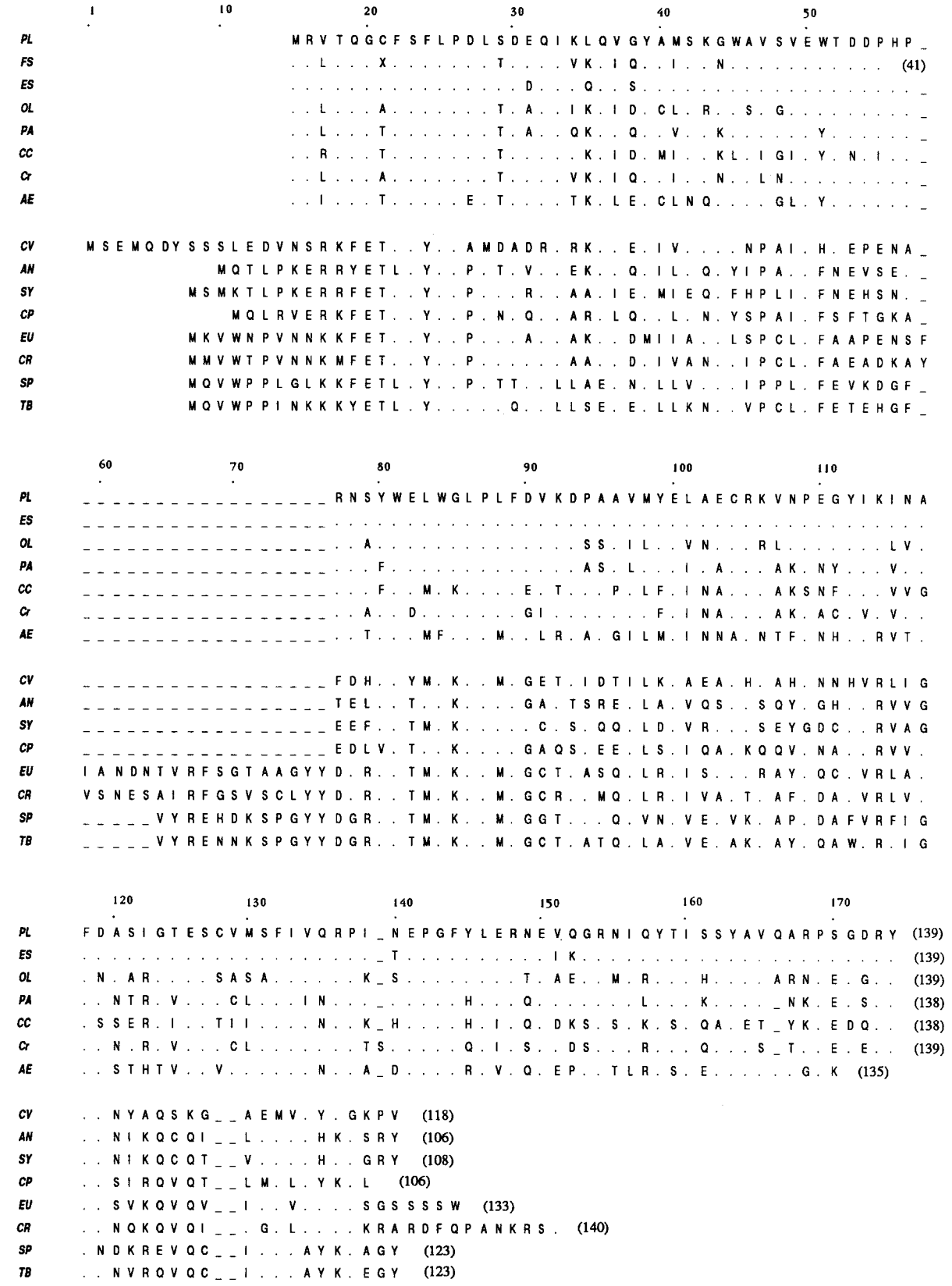
Dendrograms

The resulting neighbour-joining (NJ) trees from amino acid divergence values corrected for multiple substitutions at identical sites are given in Fig. 4. For *rbcL* and *rbcS*, NJ topologies were additionally generated from uncorrected amino acid divergence values. These were identical in branching order to the trees shown in Fig. 4 although, as expected, branches were shorter, particularly towards the base of the tree.

For those species represented in both the *rbcL* and *rbcS* data, branching orders in Fig. 4 are identical with the exceptions of *Synechococcus* and *Chromatium*. The *rbcL* data set is considerably larger (468 vs. 103 sites) and, combined with the overall lower rate of amino acid substitution for *rbcL*, should thus yield a more reliable topology [20] than the *rbcS* data in which several corrected divergence values exceed one substitution per site. For both the *rbcL* and *rbcS* genes of the non-green algae considered in the present contribution, branching orders identical to those in Fig. 4 were also obtained with PROTPARS of the PHYLIP [11] package (data not shown).

Discussion

Perhaps surprisingly, the *rbcS* and *rbcL* genes of *P. littoralis*, as well as those of other chromophytes, cryptophytes and red algae, are more similar to their homologues from the β-purple eubacterium *Alcaligenes eutrophus* than they are to the corresponding genes of green plants (Fig. 4).


1 10 20 30 40 50
PL MRVTQGCFSFLPDLSDEQIKLQVGYAMSKGWAVSVEWTDDPHP_
FS ..L...X.....T....VK.IQ..I..N..... (41)
ES .....D...Q..S..... _
OL ..L...A.....T.A..IK.ID.CL.R..S.G..... _
PA ..L...T.....T.A..QK..Q..V..K.....Y..... _
CC ..R...T.....T.....K.ID.MI..KL.IGI.Y.N.I... _
Cr ..L...A.....T....VK.IQ..I..N..LN..... _
AE ..I...T.....E.T....TK.LE.CLNQ....GL.Y..... _
CV MSEMQDYSSSLEDVNSRKFET..Y..AMDADR.RK..E.IV....NPAI.H.EPENA_
AN MQTLPKERRYETL.Y..P.T.V..EK..Q.IL.Q.YIPA..FNEVSE._
SY MSMKTLPKERRFET..Y..P...R..AA.IE.MIEQ.FHPLI.FNEHSN._
CP MQLRVERKFET..Y..P.N.Q..AR.LQ..L.N.YSPAI.FSFTGKA_
EU MKVWNPVNNKKFET..Y..P...A..AK..DMIIA..LSPCL.FAAPENSF
CR MMVWTPVNNKMFET..Y..P.....AA..D.IVAN..IPCL.FAEADKAY
SP MQVWPPLGLKKFETL.Y..P.TT..LLAE.N.LLV...IPPL.FEVKDGF_
TB MQVWPPINKKKYETL.Y.....Q..LLSE.E.LLKN..VPCL.FETEHGF_
60 70 80 90 100 110
PL _____RNSYWELWGLPLFDVKDPAAVMYELAECRKVNPEGYIKINA
ES _____.....
OL _____A.....SS.IL..VN...RL.....LV.
PA _____F.....AS.L...I.A...AK.NY...V..
CC _____F..M.K....E.T...P.LF.INA...AKSNF...VVG
Cr _____A..D.....GI.....F.INA...AK.AC.V.V..
AE _____T...MF...M..LR.A.GILM.INNA.NTF.NH..RVT.
CV _____FDH..YM.K..M.GET.IDTILK.AEA.H.AH.NNHVRLIG
AN _____TEL..T..K....GA.TSRE.LA.VQS..SQY.GH..RVVG
SY _____EEF..TM.K.....C.S.QQ.LD.VR...SEYGDC..RVAG
CP _____EDLV.T..K....GAQS.EE.LS.IQA.KQQV.NA..RVV.
EU IANDNTVRFSGTAAGYYD.R..TM.K..M.GCT.ASQ.LR.IS...RAY.QC.VRLA.
CR VSNESAIRFGSVSCLYYD.R..TM.K..M.GCR..MQ.LR.IVA.T.AF.DA.VRLV.
SP _____VYREHDKSPGYYDGR..TM.K..M.GGT...Q.VN.VE.VK.AP.DAFVRFIG
TB _____VYRENNKSPGYYDGR..TM.K..M.GCT.ATQ.LA.VE.AK.AY.QAW.R.IG
120 130 140 150 160 170
PL FDASIGTESCVMSFIVQRPI_NEPGFYLERNEVQGRNIQYTISSYAVQARPSGDRY (139)
ES .....T.....IK..... (139)
OL .N.AR....SASA.....K_S.....T.AE..M.R...H....ARN.E.G.. (139)
PA ..NTR.V...CL...IN....._.....H...Q.....L...K....._NK.E.S.. (138)
CC .SSER.I..TII....N..K_H....H.I.Q.DKS.S.K.S.QA.ET_YK.EDQ.. (138)
Cr ..N.R.V...CL.....TS.....Q.I.S..DS...R...Q...S_T..E.E.. (139)
AE ..STHTV..V.....N..A_D....R.V.Q.EP..TLR.S.E.....G.K (135)
CV ..NYAQSKG__AEMV.Y.GKPV (118)
AN ..NIKQCQI__L....HK.SRY (106)
SY ..NIKQCQT__V....H..GRY (108)
CP ..SIRQVQT__LM.L.YK.L (106)
EU ..SVKQVQV__I..V....SGSSSSW (133)
CR ..NQKQVQI__.G.L....KRARDFQPANKRS. (140)
SP .NDKREVQC__I...AYK.AGY (123)
TB ..NVRQVQC__I...AYK.EGY (123)

Purple bacteria are recognized by Woese as a phylum containing most of the well-known Gram-negative bacteria; they constitute an assembly of both photosynthetic and non-photosynthetic species. The β-subdivision is distinguished from the α-, γ- and δ-subdivisions by their cytochromes, 16S RNAs and photoreaction centres [45]. It is important to note that the Rubisco genes of *A. eutrophus* are carried on both chromosomal DNA and plasmids in two *A. eutrophus* strains [1]. The chromosomal *rbcS* gene of the *A. eutrophus* ATCC strain 17707 [1], used to make the phylogenetic tree in Fig. 4, is 54% identical to the plastid gene of *P. littoralis*. Two other *rbcL* and *rbcS* gene sequences, those of the chromosomal and plasmid-encoded genes from *A. eutrophus* strain H16, have been determined but are not yet published (B. Bowien, personal communication). Of these three sequences the most homologous to the *rbcS* of *P. littoralis* (57.5%) is the plasmid-encoded *rbcS* of strain H16. The plasmid-encoded *rbcL* gene also has greater homology to the *rbcL* of *P. littoralis* than either of the chromosomally encoded genes.

The similarity of *A. eutrophus* Rubisco genes to those of non-green algae previously noted [1, 2, 42] is subject to different interpretations. Although the antecedents of all extant β-purple bacteria were (purple) photosynthetic [47], the cyanobacterial nature of rhodophytic (and consequently chromophytic; see below) photosynthesis would not support the view that rhodoplasts are descendants of β-purple bacteria [42]. The present data would, however, be compatible with the view that a cyanobacteria-like antecedent of red algal plastids horizontally received β-purple Rubisco genes via plasmid-mediated transfer. This working hypothesis would envisage the endosymbiont as a chimaera and carries the testable predictions that (1) Rubisco (and perhaps other) gene sequences from further β-purple bacteria, as they become known, will cluster close to rhodophyte, chromophyte and cryptophyte analogous genes and that (2) the majority of rhodo-, chromo- and crytophytic plastid genes will, however, be of cyanobacterial affinity.

During evolution, the Rubisco genes separated into three main lineages, the α-purple eubacterial lineage, the β-purple eubacterial lineage and the γ-purple eubacterial-cyanobacterial lineage [2]. The main dichotomy gave rise to the extant α-purple eubacterial Rubisco genes on one hand, and to all others on the other. No earlier ancestors of these genes are known. In the α-purple eubacterial lineage only *rbcL* genes found in the L_2 or L_{4-6} complexes have been sequenced, to our knowledge, although several α-purple eubacteria, including *Rhodobacter sphaeroides*, do have L_8S_8 Rubisco complexes. It would be interesting to compare sequences of these subunits and especially to know if the α-purple Rubisco small subunits have amino acid sequences corresponding to the very conserved regions of the 'β-purple eubacterial-like' proteins.

The branching order of algal species in the 'β-purple eubacterial-like' lineage, is identical in both *rbcL* and *rbcS* trees, and in other published phylogenetic trees based on Rubisco genes [42]. In this lineage, the first branch giving rise to a plastidial Rubisco gene is that which leads to *Cyanidium caldarium*. If, as suggested by these trees, the *C. caldarium* line has separated much earlier than other red algal lines from a common ancestor, it explains the difficulties in assigning this alga a taxonomic position [23].

Phycologists have postulated that plastids of chromophytes and cryptophytes, which are surrounded by two layers of chloroplastic endoplas-

Fig. 3. Comparison of the *rbcS* amino acid sequence of *P. littoralis* (PL) with those of different prokaryotes and eukaryotes. Non-green algae are: the brown algae *Fucus serratus* (FS, partial sequence) [20] and *Ectocarpus siliculosus* [42], the chromophyte *Olisthodiscus luteus* (OL) [4], the red algae *Porphyridium aerugineum* (PA) [41], *Cyanidium caldarium* (CC) [43], the Cryptophyte *Cryptomonas Φ* (Cr) [9]. The prokaryotes are: the γ-purple eubacterium *Chromatium vinosum* (CV) [45], the β-purple eubacterium *Alcaligenes eutrophus* ATCC 17707 (AE) [1], the cyanobacteria *Anabaena* 7120 (AN) [27] and *Synechococcus* (SY) [32]. The other plastidial genes are those of *Cyanophora paradoxa* (CP) [37], *Euglena gracilis* (EU) [5], the green alga *Chlamydomonas reinhardtii* (CR) [15] and the land plants *Spinacea oleracea* (SP) and *Nicotiana tabacum* (TB) [25].

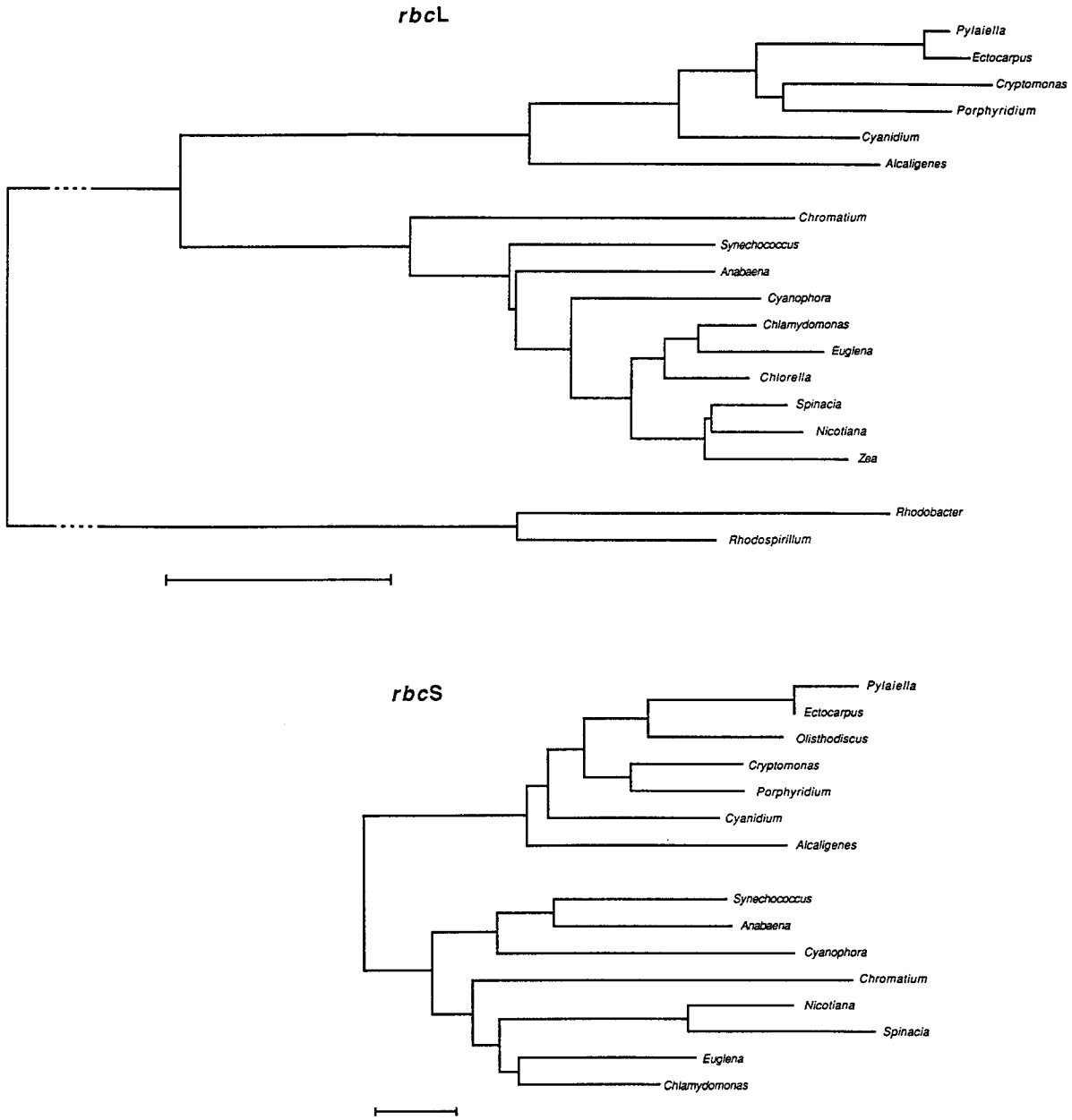


Fig. 4. Dendrograms generated by the neighbour-joining method for *rbcS* and *rbcL*. The bar to the lower right of each tree represents 0.1 amino acid substitution per site (corrected). For *rbcL*, the upper portion of the tree is rooted to the *Rhodospirillum* and *Rhodobacter* sequences, yet the full length of this root branch is not drawn to scale as indicated by dashed lines. The *rbcS* sequence of *Rhodobacter sphaeroides* is not yet available and *Rhodospirillum rubrum* has no *rbcS* gene [13, 40], therefore formal rooting could not be performed for *rbcS*. The deepest branch of the unrooted *rbcS* tree was arbitrarily 'bent' to permit display of relevant species in a form which permits direct comparison of the *rbcS* and *rbcL* topologies.

mic reticulum (CER), have not evolved directly from a photosynthetic prokaryote as those of red algae, but from a secondary endosymbiosis between an eukaryotic host cell and a pre-existing eukaryotic algal cell [for review see 12]. The pre-existing algal cells are thought to have been uni-

cellular red algae. If this theory is correct, the trees shown here favour two different ancestors of the extant red algae as the endosymbionts, one which gave rise to the chromophyte plastids, and another to the cryptophyte plastids.

The third lineage comprises Rubisco genes from the γ-purple eubacterium *Chromatium vinosum*, two cyanobacteria, the cyanoplast of *C. paradoxa* and various chlorophyll *b*-containing organisms. The Rubisco small subunits of chlorophyll *b*-containing organisms which have been studied are all nuclear-encoded. In this lineage the branching order is clearly not the same in both trees for *Chromatium* and *Cyanophora*.

The *rbcL* and *rbcS* sequences of another (unnamed) γ-purple eubacterium have recently been published [38]. This bacterium is a sulphur-oxidizing chemoautotrophic endosymbiont, quite different from the photosynthetic, free-living *C. vinosum*. Although we had already constructed the trees before this sequence appeared, 89% of the amino acids are conserved between both species in the large subunit and 80% in the small subunit, supporting the position of γ-purple bacteria in the *rbcL* and *rbcS* trees. Nevertheless the position of these bacteria in the *rbcS* tree, closely linked to green plastid *rbcS* genes, is probably artefactual, due to stochastic similarities and the low total number of *rbcS* sites compared. In fact, their amino acid sequences have none of the characteristic signatures of the Rubisco small subunits of chlorophyll b-containing plants, i.e. VWxPxxxK in the amino terminal region and the additional fragment starting at position 58 (Fig. 3) immediately followed by DxR. The Rubisco genes of *Cyanophora* are closely related to those of cyanobacteria, the same is true for other *Cyanophora* chloroplastic genes such as *tufA* [24, 29], indicating that a cyanobacterial ancestor different from that which evolved into green plastids could have given rise to the cyanoplasts.

Phylogenetic trees based on entire sequences of the nuclear 18S rDNA genes show that the host cells giving rise to extant algae are widely divergent, also supporting the theory of multiple endosymbiotic events [3, 18, 34, 35]. Endosymbiosis may have been established soon after the divergence from the eukaryotic trees, or long after, as appears to be the case for *Euglena* and the Dinoflagellates.

Phycologists have often thought that the 'Chromophytes', which actually include different eukaryotic lineages, form an artificial group. The results discussed herein support that contention. The only link between dinoflagellates, cryptophytes and what could be called chromophytes *sensu stricto* (Rhaphidophyceae, Crysophyceae, Haptophyceae, diatoms and brown algae) is that their plastids have probably evolved from red alga-like ancestors. No molecular information is yet available about the cytoplasmic origin of the Xanthophyceae or Eustigmatophyceae. Both the organelle and host cell phylogenies indicate deep branching of the different algal lines which in turn indicates a more complex history of algae than can be accounted for by a single host-symbiont event.

Acknowledgements

We wish to thank the Gesellschaft für Biotechnologische Forschung, Braunschweig, FRG, for the use of their computer facilities and Dr R. Haselkorn and Dr J. Flech for hybridization probes.

References

1. Anderson K, Caton J: Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose bisphosphate carboxylase large and small subunit genes and their gene product. J Bact 169: 4547–4558 (1987).
2. Assali NE, Mache R, Loiseaux-de Goër S: Evidence for a composite phylogenetic origin of the plastid genome of the brown alga *Pylaiella littoralis* (L.) Kjellm. Plant Mol Biol 15: 307–315 (1990).
3. Bhattacharya D, Elwood HJ, Goff L, Sogin ML: Phylogeny of *Gracilaria lemaneiformis* (Rhodophyta). J Phycol 26: 181–186 (1990).
4. Boczar BA, Delaney TP, Cattolico RA: Gene for the ribulose-1,5-bisphosphate carboxylase small subunit protein of the marine alga *Olisthodiscus luteus* is similar to that of a chemoautotrophic bacterium. Proc Natl Acad Sci USA 86: 4996–4999 (1989).
5. Chan RL, Keller M, Canaday S, Weil JH, Imbault P:

Eight small subunits of *Euglena* ribulose 1.5 bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. Embo J 9: 333–338 (1990).
6. Dalmon J, Loiseaux S, Bazetoux S: Heterogeneity of plastid DNA of two species of brown algae. Plant Sci Lett 29: 243–253 (1983).
7. Devereux J, Heaberli P, Smithies O: A comprehensive set of sequence analysis programs for the VAX. Nucl Acids Res 12: 387–395 (1984).
8. Dickerson R: The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1: 26–45 (1971).
9. Douglas SE, Durnford DG: The small subunit of ribulose-1,5-bisphosphate carboxylase is plastid-encoded in the chlorophyll *c*-containing alga *Cryptomonas* F. Plant Mol Biol 13: 13–20 (1989).
10. Douglas SE, Durnford DG, Morden CW: Nucleotide sequence of the gene for the large subunit of ribulose-1,5 bisphosphate carboxylase/oxygenase from *Cryptomonas* F: evidence supporting the polyphyletic origin of plastids. J Phycol 26: 500–508 (1990).
11. Felsenstein J: Evolutionary trees from DNA sequences: A maximum likelihood approach. J Mol Evol 17: 368–376 (1981).
12. Gibbs SP: The evolution of algal chloroplasts. In: Weissner W, Robinson DG, Starr RC (eds) Experimental Phycology I, pp. 145–157. Springer-Verlag, Berlin (1990).
13. Gibson JL, Tabita FR: Different molecular forms of ribulose-1,5-bisphosphate from *Rhodopseudomonas sphaeroides*. J Biol Chem 252: 943–949 (1977).
14. Glover HE: Ribulose bisphosphate carboxylase/oxygenase in marine organisms. Int Rev Cytol 115: 67–138 (1988).
15. Goldschmidt--Clermont M, Rahire M: 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 (1986).
16. Gray MW: The bacterial ancestry of plastids and mitochondria. Bioscience 33: 693–699 (1983).
17. Guoy M, Li WH: Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Nature 339: 145–147 (1989).
18. Gunderson H, Elwood H, Ingold A, Kindle K, Sogin ML: Phylogenetic relationships between chlorophytes, chrysophytes and oomycetes. Proc Natl Acad Sci USA 84: 5823–5827 (1987).
19. Hori H, Osawa S: Origin and evolution of organisms as deduced from 5S Ribosomal RNA sequences. Mol Biol Evol 4: 445–472 (1987).
20. Jin L, Nei M: Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7: 82–102 (1990).
21. Keen JN, Pappin DJC, Evans LV: Amino acid sequence analysis of the small subunit of ribulose bisphosphate carboxylase from *Fucus* (Phaeophyceae). J Phycol 24: 324–327 (1988).
22. Kostrzewa M, Valentin K, Maid U, Radetzky R, Zetsche K: Structure of the rubisco-operon from the multicellular red alga *Antithamnion* spec. Curr Genet 18: 465–469 (1990).
23. Lewin RA, Gibbs SP: Algae of uncertain taxonomic position: introduction and bibliography. In: Rosowski JR, Parker BC (eds) Selected Papers in Phycology II, pp. 659–662. PSA, USA (1990).
24. Ludwig W, Weizenegger M, Betzl D, Leidel E, Lenz T, Ludvigsen A, Möllenhoff D, Wenzig P, Schleifer KH: Complete nucleotide sequences of seven eubacterial genes coding for the elongation factor Tu: functional structural and phylogenetic evaluations. Arch Microbiol 153: 241–247 (1990).
25. Mazur BJ, Chui CF: Sequence of a genomic DNA clone for the small subunit of ribulose bisphosphate carboxylase/oxygenase from tobacco. Nucl Acids Res 13: 2373–2386 (1985).
26. Miziorko HM, Lorimer GH: Ribulose-1,5-bisphosphate carboxylase/oxygenase. Ann Rev Biochem 52: 507–535 (1983).
27. Nierzwicki-Bauer SA, Curtis SE, Haselkorn R: 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 (1984).
28. Palmer JD: Evolution of chloroplast and mitochondrial DNA in plants and algae. In: McIntyre RJ (ed) Monographs in Evolutionary Biology: Molecular Evolutionary Genetics, pp. 131–240. Plenum Press, New York.
29. Palmer JD, Baldauf SL, Calie PJ, de Pamphilis CW: Chloroplast gene instability and transfer to the nucleus. In: Clegg M, O'Brien S (eds) Molecular Evolution, vol 122, pp. 1–10. Alan R. Liss, New York.
30. Reith M, Cattolico RA: Inverted repeat of *Olisthodiscus luteus* chloroplast DNA contains genes for both subunits of ribulose-1,5-bisphosphate carboxylase and the 32000 dalton QB protein: Phylogenetic implications. Proc Natl Acad Sci USA 83: 8599–8603 (1986).
31. Saitou N, Imanishi T: Relative efficiencies of the Fitch-Margoliash, Maximum-Parsimony, maximum-likelihood, minimum-evolution and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol Biol Evol 6: 514–525 (1989).
32. Saitou N, Nei M: The neighbor joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406–425 (1987).
33. Shinozaki K, Sugiura M: The gene for the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase is located close to the gene for the large subunit in cyanobacterium *Anacystis nidulans* 6301. Nucl Acids Res 11: 6956–6964 (1983).
34. Sogin ML, Gunderson JH: Structural diversity of Eukaryotic small subunit ribosomal RNAs. Evolutionary implications. Ann NY Acad Sci 503: 125–139 (1987).
35. Sogin ML, Gunderson JH, Elwood HJ, Alonso RA,

Peattie DA: Phylogenetic meaning of the kindom concept, an unusual ribosomal RNA from *Giardia lamblia*. Science 243: 75–77 (1989).
36. Sourdis J, Nei M: Relative efficiencies of the maximum parsimony and distance matrix methods in obtaining the correct phylogenetic tree. Mol Biol Evol 5: 298–311 (1988).
37. Starnes SM, Lambert DH, Maxwell ES, Stevens SE jr, Portis RD, Shively JM: Cotranscription of the large and small subunit genes of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Cyanophora paradoxa*. FEMS Microbial Lett 28: 165–169 (1985).
38. Stein JL, Haygood M, Felbeck H: Nucleotide sequence and expression of a deep-sea ribulose-1,5-bisphosphate carboxylase gene cloned from a chemoautotrophic bacterial endosymbiont. Proc Natl Acad Sci USA 87: 8850–8854 (1990).
39. Stern DB, Jones H, Gruissem W: Function of plastid mRNA 3′ inverted repeats. RNA stabilization and gene-specific protein binding. J Biol Chem 264: 18742–18750 (1989).
40. Tabita FR, McFadden BA: D-Ribulose-1,5-diphosphate carboxylase from *Rhodospirullum rubrum* II: Quaternary structure, composition, catalytic and immunological properties. J Mol Chem 249: 3459–3464 (1974).
41. Valentin K, Zetshe K: 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 (1989).
42. Valentin K, Zetsche K: Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. Plant Mol Biol 15: 575–584 (1990).
43. Valentin K, Zetsche K: Structure of the Rubisco operon from the unicellular red alga *Cyanidium caldarium:* Evidence for a polyphyletic origin of the plastids. Mol Gen Genet 222: 425–430 (1990).
44. Valentin K, Zetsche K: Nucleotide sequence of the gene for the large subunit of Rubisco from *Cyanophora paradoxa*. Phylogenetic implications. Curr Genet 18: 199–202 (1990).
45. Viale AM, Kobayashi H, Akazawa T: 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 Bact 171: 2391–2400 (1989).
46. Whatley JM, Whatley FR: Chloroplast evolution. New Phytol 87: 233–247 (1981).
47. Woese CR: Bacterial evolution. Microbiol Rev 51: 221–270 (1987).
48. Wolter FP, Fritz CC, Willmitzer L, Schell J, Schreier P: *rbcS* genes in *Solanum tuberosum*: Conservation of transit peptide and exon shuffling during evolution. Proc Natl Acad Sci USA 85: 846–850 (1988).