J Mol Evol (1992) 35:385–404

Journal of Molecular Evolution

Molecular Phylogenies of Plastid Origins and Algal Evolution

W. Martin,[1] C.C. Somerville,[2] and S. Loiseaux-de Goër[2]

[1] Institut für Genetik, Technische Universität Braunschweig, Konstantin-Uhde-Str. 5, D-3300 Braunschweig, Germany
[2] Station Biologique de Roscoff, CNRS LP 4601, BP 74, F-29682 Roscoff Cedex, France

Summary. An overview of recent molecular analyses regarding origins of plastids in algal lineages is presented. Since different phylogenetic analyses can yield contradictory views of algal plastid origins, we have examined the effect of two distance measurement methods and two distance matrix tree-building methods upon topologies for the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit nucleotide sequence data set. These results are contrasted to those from bootstrap parsimony analysis of nucleotide sequence data subsets. It is shown that the phylogenetic information contained within nucleotide sequences for the chloroplast-encoded gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, integral to photosynthesis, indicates an independent origin for this plastid gene in different plant taxa. This finding is contrasted to contrary results derived from 16S rRNA sequences. Possible explanations for discrepancies observed for these two different molecules are put forth. Other molecular sequence data which address questions of early plant evolution and the eubacterial origins of algal organelles are discussed.

Key words: Endosymbiosis — Molecular phylogeny — Algal evolution — Plastid origins

Introduction

Eukaryotes are chimeras. One can study the molecular evolution of nuclei and organelles as distinct entities, which themselves share a common ancestor. Implicit in this assertion is the belief that endosymbiosis has played a major role in evolution. Chromophyte algae may be chimeras within chimeras. Chromophytes are characterized by the presence of flagella, chlorophyll *a* and chlorophyll *c*, the absence of chlorophyll *b*, and the presence of three or four membranes surrounding their chloroplasts (Christensen 1989). The number and cellular continuum of membranes surrounding chromophyte plastids have attracted particular attention, since these and other cytological characters have been interpreted as evidence of secondary endosymbiosis (Tomas and Cox 1973) between eukaryotic hosts and eukaryotic endosymbionts in chromophyte evolution (Gibbs 1978, 1981; Whatley 1981, 1989; Cavalier-Smith 1986; see Gibbs 1990, Sitte 1990, and Melkonian 1991 for recent reviews). Strong molecular evidence in support of chromophyte secondary endosymbiosis was recently brought forth (Douglas et al. 1991) through analysis of DNA sequences from the nucleomorph, putatively a vestigial nucleus of the eukaryotic endosymbiont, from the cryptophyte *Cryptomonas* Φ.

The idea that plastids may be endosymbiotic descendants of once free-living organisms is not particularly new. Schimper (1883) is generally given credit for the idea, based upon observations concerning plastid division, that plant cells are derived from the symbiotic unification of a colorless organism with green, autonomous plastid progenitors:

> Sollte es sich definitiv bestätigen, dass die Plastiden in den Eizellen nicht neu gebildet werden, so würde ihre Beziehung zu dem sie enthaltenden Organismus einigermassen an eine Symbiose erinnern. Möglicherweise verdanken die grünen Pflanzen wirklich einer

Offprint requests to: W. Martin

> Vereinigung eines farblosen Organismus mit einem von Chlorophyll gleichmässig tingierten ihren Ursprung.[1]

Mereschkowsky (1905) explicitly suggested that cyanobacteria gave rise to plastids,[2] and that different cyanobacteria gave rise to the plastids of different plant groups. In a series of subsequent publications (1910a–d) he established the framework for a general phylogeny of organisms in which (1) what we now term prokaryotes arose before (albeit independently of) what we now term eukaryotes and (2) the plastids of chlorophytes, rhodophytes, and chromophytes originated through independent endosymbiotic events. The general scheme of eukaryotic evolution proposed by Merschkowsky (1910d) was largely based upon biochemical, hereditary, and metabolic considerations and is reproduced in Fig. 1; the reader may refer to the original literature for an elegant elaboration and justification of the arguments upon which the figure was based.

A number of molecular evolutionary contributions have appeared recently in the literature which relate to questions clearly formulated by Mereschkowsky; these concern the endosymbiotic origins of various algal plastids, the evolution of the symbionts prior to symbiosis, and the origins and early evolution of eukaryotes. Some of the results have been interpreted as molecular evidence for multiple (polyphyletic) origins of plastids, this in light of vehement arguments favoring monophylesis both for plastids and eukaryotes (Cavalier-Smith 1987b). In this article, molecular data regarding the independent phylogenies of plastid and nuclear symbiotic partners before endosymbiosis on the one hand, and subsequent to the establishment of plastids on the other, will be considered. Since molecular evolutionary relationships between organisms or organelles are inferred through calculation, rather than through observation (as in Fig. 1), various parameters involved in reconstruction of phylogenetic trees from molecular sequence data influence the topology or topologies obtained. This in turn colors considerably the *interpretation* of a phylogenetic data set and thereby one's view of evolutionary process. Therefore, a data set particular to plastid evolution, i.e., nucleotide sequences for the large subunit of Rubisco (*rbc*L), will first be considered in order to contrast the topologies generated by two distance measurement methods and two distance matrix methods for tree construction. The true phylogeny of the organisms under study is not known. For comparison, however, one can compare the phylogenetic trees obtained to branching orders estimated through parsimony bootstrap and jackknife (Felsenstein 1985) analyses of *rbc*L nucleotide sequences. The strength of molecular data concerning polyphyly vs monophyly for plastid origins will be addressed. Some of the conflicting results which have been obtained with different genes as molecular markers and, in some cases, potential causes for the differing results will be discussed.

Material and Methods

Comparison of Two Distance Measures. Amino acid sequences were first aligned by eye with the LINEUP program of the WISGEN package (Devereux et al. 1984). Corresponding codon insertions were introduced into the nucleotide sequences; amino- and carboxyterminal regions of length heterogeneity were deleted from the alignment yielding 1413 nucleotide positions (471 codons) in each sequence for comparison. Divergence at nonsynonymous sites between aligned nucleotide sequences (alignment available upon request from the authors) was measured as K_a with the weighted pathway method (Li et al. 1985) and as d_N with the method of Nei and Gojobori (1986). An average *rbc*L pair has about 1060 nonsynonymous sites. Values of K_a and d_N were plotted against each other for each pairwise comparison (Fig. 2).

Construction of Trees. Matrices of distance values were analyzed with the Neighbor-Joining (NJ: Saitou and Nei 1987) or with the Fitch-Margoliash least-squares (FM: Fitch and Margoliash 1967) method (FITCH of the PHYLIP package (Felsenstein 1981)). The NJ method was chosen since it has been shown in computer simulation to be equally or more efficient than other matrix or molecular sequence methods tested in recovery of correct topology under a variety of sequence parameters (Saitou and Imanishi 1987; Sourdis and Krimbas 1987; Saitou and Nei 1987; Jin and Nei 1990). The FM method was chosen for comparison since it is widely used. FITCH was run with the "jumble" and "global" options. References for sequences are *Alcaligenes* (Andersen and Caton 1987); *Anabaena* (Curtis and Haselkorn 1983); *Chlamydomonas* (Dron et al. 1982); *Euglena* (Gingrich and Hallick 1985); *Zea* (McIntosh et al. 1980); *Nicotiana* (Shinozaki and Sugiura 1982); *Chromatium* (Viale et al. 1989); *Chlorella* (Yoshinaga et al. 1988); *Prochlorthix* (Morden and Golden 1991); *Rhodobacter sphaeroides* (Gibson et al. 1991); the unnamed γ-purple eubacterium (Stein et al. 1990); *Cryptomonas* (Douglas et al. 1990); *Cylindrotheca* (Hwang and Tabita 1991); *Pylaiella* (Assali et al. 1991a); *Cyanidium* (Valentin and Zetsche 1990b); *Cyanophora* (Valentin and Zetsche 1990c); *Ectocarpus* (Valentin and Zetsche 1990a); *Antithamnion* (Kostrzewa et al. 1990); *Astasia* (Siemeister and Hachtel 1990); *Cryptomonas* (Douglas et al. 1990); and *Porphyridium* (Valentin and Zetsche 1989). The

[1] "If it can be conclusively confirmed that plastids do not arise de novo in egg cells, the relationship between plastids and the organisms within which they are contained would then be somewhat reminiscent of a symbiosis. Green plants may in fact owe their origin to the unification of a colorless organism with one uniformly tinged with chlorophyll" Schimper (1883) pp. 112–113. Translation by the authors.

[2] "A theory [of endosymbiosis], as suggested here, would gain strength if the existence of any free-living organisms similar to the symbiont could be proven. Lower forms of the Cyanophyceae can be considered as such organisms" Merschkowsky (1905) p. 599. Translation from the original German text by the authors.

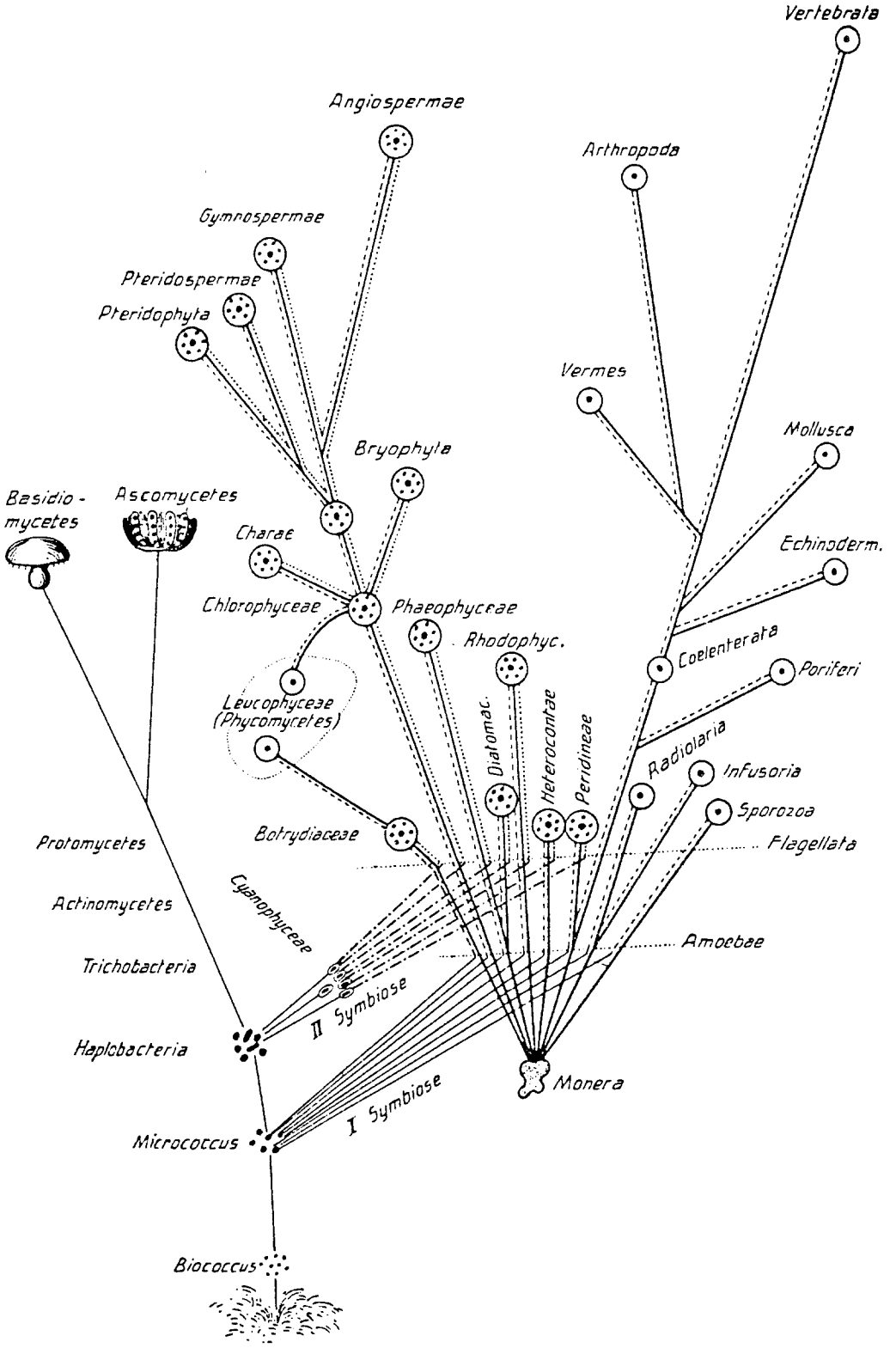


Fig. 1. Phylogenetic tree constructed and presented by Mereschkowsky (1910d), summarizing his views at the time concerning cellular evolution. In his scheme, prokaryotes (*das Mykoplasma: Biococcus*), then known to tolerate extreme conditions, were supposed to have arisen very early in evolution, at a time when the Earth's surface had cooled sufficiently to retain water (<100°C), yet was too warm (>50°C) to support life of the more fragile eukaryotes (*das Amöboplasma: Monera*). The latter arose independently of the former at a time when the Earth's surface had cooled to <50°C; these later were suggested to have obtained their nuclei (not mitochondria) through a first series of endosymbioses. In a second series of endosymbioses, flagellates were proposed to have obtained plastids through endosymbiosis with such diverse cyanobacteria as contained either green, red, or brown pigments. Phycomycetes (included oomycetes) were depicted as a polyphyletic group, distinct from other fungi, which evolved through secondary loss of photosynthesis. The arguments put forth upon which the diagram is based relate to metabolic, biochemical, pigmentation, and growth environment characters.

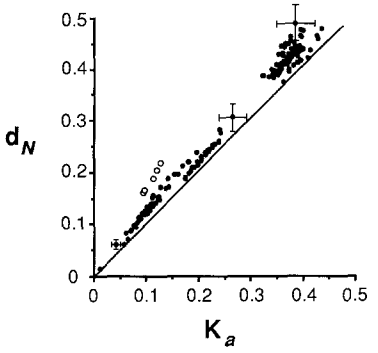


Fig. 2. Comparison of estimates for divergence at nonsynonymous sites between the 20 *rbc*L sequences from species shown in Fig. 3 as values of d_N (Nei and Gojobori 1986) and K_a (Li et al. 1985) for divergence at nonsynonymous sites. Each comparison of d_N vs K_a is represented as a single dot in the graph. The five open circles represent the comparisons between the *Astasia rbc*L sequence and those of chlorophyte plastids. (See text and Fig. 3.) The diagonal is not a regression but rather shows the slope expected for equal estimates of divergence with the two methods. Standard errors for typical mean estimates of K_a and d_N vary with comparison and degree of divergence; for the range of divergence 0.1–0.4 in *rbc*L sequences, the standard errors are approximately 10–15% of the mean, as shown by error bars for three typical individual comparisons.

outgroup used is the form II *rbc*L of *Rhodospirillum rubrum* (Narang et al. 1984); the *Rhobacter* sequence analyzed is of the form I type (Gibson et al. 1991). *Rhodospirillum* also contains a form I type Rubisco (Falcone and Tabita 1991). The *Rhodospirillum* outgroup sequence is quite difficult to align with the remaining *rbc*L sequences throughout most of its length; K_a and d_N values ranged from 2.5 to 3.5 in most comparisons. In order to reduce stochastic error in the distance matrix trees due to the large standard error attached to estimates of divergence in the *Rhodospirillum* comparisons, the *Rhodospirillum* outgroup distance was set to 2.0 for all 20 pairwise comparisons. This procedure produced a "synthetic" root and yielded NJ topologies for the remaining 20 species which were identical to those produced in absence of the outgroup. The *rbc*L nucleotide sequences were analyzed by bootstrap parsimony analysis (DNABOOT of PHYLIP, Felsenstein 1985). Various subsets of the nucleotide sequences were studied. (See Table 2.) Bootstrap replicates for the *Rhodospirillum* outgroup sequence were only performed with second positions.

Results and Discussion

An Overview of the Data and a Consideration of rbc*L Sequences*

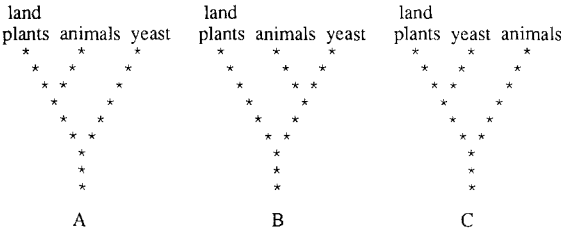
Molecular phylogenies which concern endosymbiosis and the general relationship of various plant groups to other organisms are summarized in Table 1 in order to provide both a general (yet not exhaustive) overview of the scope of current investigation and convenient reference to studies encompassing taxa of interest. Various molecular markers and methods of phylogenetic construction have been employed. Ribosomal RNA sequences are known from a large number of organisms and have thus been used to establish affinities among endosymbionts much more extensively than have protein sequences, though this trend is not altogether adequately reflected in the table. Most plastid-encoded proteins employed to date for evolutionary studies are involved in photosynthesis [e.g., the large (*rbc*L), and in some cases, the small (*rbc*S) subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco), *psb*A, etc.]. Photosynthetic apparatus genes obviously cannot be employed, as the 16S and 18S rRNAs with some limitations can, to *concomitantly* investigate nuclear and organellar phylogenies. Elongation factor Tu, the product of *tuf*A, however, can be used to connect nuclear, plastid, and mitochondrial phylogenies (Palmer et al. 1989). Plastid genomes in chlorophytes appear to evolve quite conservatively with little or no tendency to accept DNA from other compartments (Palmer et al. 1988), as sequence comparison of three complete chloroplast genomes has shown (Ohyama et al. 1986; Shinozaki et al. 1986; Hiratsuka et al. 1989). Chloroplast genes were thus very probably a genome component *at the time of endosymbiosis*, making these genes presumably quite reliable markers for reconstruction of organellar phylogeny. As will be discussed later, it is not known whether this generalization also applies to plastid genomes of rhodophytes and chromophytes, where molecular data are still scarce.

Plant nuclear genomes appear to have acquired a number of protein-coding DNA sequences from endosymbionts through intracellular gene transfer (Brinkmann et al. 1987; Clegg et al. 1991; Baldauf and Palmer 1990; Liaud et al. 1990). Though nuclear, such genes necessarily reflect the evolution of plastids or, in some cases, perhaps mitochondria (Martin et al. 1990), rather than that of the eukaryotic host. To date, 18S rRNAs have been more widely used than other markers to study evolutionary relationships between eukaryotic nuclei, as these genes are experimentally quite accessible and were, in all likelihood, components of the eukaryotic host cell genome before endosymbiosis. Ribosomal RNA phylogenies are not, however, always supported by analyses of protein coding regions (cf. Loomis and Smith 1990).

Discrepancies Between Published Trees

There are *discrepancies* between the trees listed in Table 1 too vast in number to be considered in any detail. Yet as a case in point, one may consider the three possible rooted branching orders for land

plants, metazoa, and yeast; each of these can be found more than once among the 11 contributions listed in Table 1 which examined representatives from all three groups:



Obviously, all three cannot be simultaneously correct. A and B are found four times each and C is found three times. That different authors arrive at different molecular phylogenies for the same taxa and the same gene is disconcerting, though not at all surprising. Various parameters can affect a given tree-building method's result; these have been characterized to some extent in the literature. The amount of data upon which a given molecular phylogeny is based, the alignment generated, the amount of divergence between the sequences, differences in substitution rates between different lineages, the true phylogeny of taxa analyzed, tree-building methods themselves, and other factors may exert an influence upon which topology or topologies are obtained (Felsenstein 1988; Penny and Hendy 1986; Saitou and Imanishi 1989; Sourdis and Krimbas 1987; Gouy and Li 1989b; Swofford and Olsen 1990; Jin and Nei 1990; Nei 1991). As far as the specific example of land plants, metazoa, and yeast is concerned, it can be noted that Gouy and Li (1989a) carefully analyzed sequences for 18S rRNA, 16S rRNA, ten tRNA families, and six conservatively evolving proteins and reported conclusive evidence supporting yeast (yet not by corollary all fungi, see later) as the outgroup for higher plants and metazoa (alternative A shown above), a topology depicted only four times in the 11 aforementioned trees. In light of this, one could question the amount of confidence warranted by the phylogenies for genes of algal nuclei and plastids depicted in the various studies of Table 1.

Bootstrap and Jackknife Analysis of rbc*L Data*

A number of methods for assessing the reliability of phylogenetic trees derived from sequence data are available (for a review see Felsenstein 1988; see also Li 1989). One of these, the bootstrap (Felsenstein 1985) has been employed in several of the studies listed in Table 1. As almost invariably is the case with real data, the true phylogeny of the organisms from which the 19 *rbc*L nucleotide sequences are derived (Table 1) is not known, one therefore cannot compare the efficiency of different methods in obtaining the correct tree, as in computer simulations (cf. Jin and Nei 1990), but one can compare the results obtained with different distance matrix methods in their ability to recover topologies supported by bootstrap parsimony analysis of the *rbc*L sequence data. Bootstrap (BS) and half-delete jackknife (JK) analyses (DNABOOT of the PHYLIP package; Felsenstein 1981, 1985) were therefore performed with several subsets of the *rbc*L nucleotide data for 19 and one subset for 20 (including the *Rhodospirillum* outgroup) species relevant to questions of plant evolution. Since parsimony analysis operates under the assumption that the probability of a base substitution at a given site is small over the lengths of time involved in a branch, synonymous sites should best be removed from the *rbc*L data set, since they do not fulfill this condition. Yet selectively removing synonymous sites is tedious, since the number and location of synonymous sites is a property of codons. One can approximate the removal of synonymous sites by deleting third positions from the sequence alignment (i.e., with a system editor). This eliminates all fourfold degenerate sites and most twofold degenerate sites,[3] thus (1) decreasing stochastic similarity ("background noise") of sequences due to the existence of only four possible character states for any site and (2) lowering considerably the probability of base substitution in a given branch. It should be noted that the nucleotide sequence alignment for the 20 genes consisting of first plus second positions (each 942 bases long) contains almost all of the nondegenerate sites (avg. 910 sites/pair) and thus almost completely contains yet is not identical to the subset of nonsynonymous sites (avg. 1060 sites/pair) from which the two distance measurements [as numbers of nonsynonymous substitutions per site K_a (Li et al. 1985) and d_N (Nei and Gojobori 1986)] were calculated. This circumstance should not introduce serious error into the analysis.[4]

Second codon positions consist only of nondegenerate sites and evolve more slowly than either of

[3] Compare Li et al. (1985) and Nei and Gojobori (1986) for lucid treatments of codon degeneracy in phylogenetic sequence distance measurement.

[4] As an alternative, one could measure divergence at first and second positions with the method of Tajima and Nei (1984) and subject the identical alignment to bootstrap analysis, yet that distance measurement will not discriminate between synonymous and nonsynonymous substitutions.

Table 1. Condensed overview of several recent molecular phylogenies relevant to the context of molecular evolution of plastid and nuclear genes of algae, with emphasis on rhodophytes and cryptophytes*

Taxon:													
rhodophytes	–	1	1	–	–	1	–	1	1	2	1	–	1
chromophytes	–	1	–	–	1	2	–	2	3	2	2	1	4
cryptophyte (nu)	–	–	–	–	–	–	–	1	–	1	–	–	2
cryptophyte (nm)	–	–	–	–	–	–	–	–	–	1	–	–	–
chlorophytes	4	3	1	1[n]	3	8	–	5	5	5	10	4	4
euglenoids	1	1	1	–	1	2	–	1	1	1	1	1	–
cyanophora	1	–	–	–	1	–	–	–	–	–	–	–	–
cyanobacteria	17[a]	6[a]	7	1	6[a]	1	6	–	–	–	–	–	–
α-purple bacteria[t]	–	–	–	1	–	1	5	–	–	–	–	–	–
β-purple bacteria[t]	–	–	–	1	–	2	5	–	–	–	–	–	–
γ-purple bacteria[t]	–	1	–	1	–	1	5	–	–	–	–	–	–
δ-purple bacteria[t]	–	–	–	–	1	–	5	–	–	–	–	–	–
dinoflagellates	–	–	–	–	–	–	–	1	1	1	1	1	1
kinetoplastida	–	–	–	–	–	–	–	3	1	1	3	1	–
ciliates	–	–	–	–	–	–	–	4	2	4	5	2	3
fungi	–	–	–	–	–	–	–	3	5	3	8	3	2
oomycetes	–	–	–	–	–	–	–	1	1	1	1	1	–
amoeba	–	–	–	–	–	–	–	1	1	1	1	1	1
metazoa	–	–	–	–	–	–	–	h	h	h	h	h	h
Compartment:													
plastid	pla	pla	pla		pla	pla							
nucleus								nuc	nuc	nuc	nuc	nuc	nuc
prokaryotic	pro	pro	pro	pro	pro	pro	pro						
Gene/molecule:	16S	16S	16S	16S	16S	16S	16S	18S	18S	18S	18S	18S	28S
Sites analyzed:	667	1130	S_{ab}	818	703	120	g	1240	1530	1001	1263	1530	280
Tree method(s):	FM[b] BS	FM	clus	FM[b]	ST	FM	e	ST	FM[b]	NJ BS	FM[b]	FM[b]	FM PP
Reference:	1	2	3	4	5	6	7	8	9	10	11	12	13

* Numbers in the columns indicate numbers of species analyzed for a given taxon in the corresponding reference numbered at the base of the column. The selection of literature is not exhaustive but is intended to provide rough orientation concerning which genes are currently under study with which types of data analysis methods. Several authors employed additional programs; those listed here concern only trees published as figures. References are:

1. Turner et al. 1989
2. Witt and Stackebrandt 1988
3. Fox et al. 1980
4. Yang et al. 1985
5. Wolters et al. 1989
6. Markowicz et al. 1991
7. Woese 1987
8. Eschbach et al. 1991
9. Bhattacharya et al. 1990
10. Douglas et al. 1991
11. Hendriks et al. 1991
12. Gunderson et al. 1987
13. Christen et al. 1991
14. Valentin and Zetsche 1990a
15. Valentin and Zetsche 1990b
16. This article
17. Morden and Golden 1991
18. Assali et al. 1991a
19. Palmer et al. 1989
20. Ludwig et al. 1990
21. Assali and Loiseaux-de Goër 1992
22. Sogin et al. 1989
23. Morden and Golden 1989
24. Ariztia et al. 1991
25. Michels et al. 1991
26. Perasso et al. 1989
27. Förster et al. 1990
28. Pühler et al. 1989
29. Giovannoni et al. 1988
30. Douglas and Turner 1991

the other two positions; they should be highly suited to parsimony (during bootstrap and jackknife procedures), although analysis of only second codon positions further reduces the size of the original data set. For comparison, separate alignments for first plus second (942 sites), second positions (471 nondegenerate sites) and first positions (471 most of which are nondegenerate) for bootstrap (BS) and jackknife (JK) analysis were generated. The alignments are available upon request from the authors.

The results of these analyses are summarized in Table 2. They demonstrate that on the one hand there is sufficient information contained within the *rbc*L data to clearly resolve some portions of the tree through parsimony analysis (particularly the deepest and uppermost branches). On the other hand, the intermediate branches within the chlorophyte and rhodophyte/chromophyte portions of the tree (cf. Fig. 3) cannot be resolved with the first and second positions of *rbc*L. For the most part, the various data sets (first vs second vs first plus second codon positions) gave similar results, although first codon positions revealed somewhat fewer distinct groups in the 100 parsimony replicates. In general, the data set consisting of only second positions

Table 1. Extended

–	3	3	1	2	–	–	–	–	–	–	–	1	–	–	–	2
–	1	3	–	3	–	–	1	1	–	5	–	7	2	–	–	3
–	–	1	1	1	–	–	–	–	–	–	–	2	–	–	–	1
–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
6	5	4	5	5	4[i]	2	3	2	8	4	5[p]	5	3	2	1	7
1	1	2	1	1	1	1	1	1	–	–	–	–	1	–	–	2
1	1	1	–	1	1	1	–	–	–	–	–	–	–	–	1	1
2	2	2[a]	3[a]	2	1	2	1	1	7[a]	–	–	–	–	–	29	15
–	–	1	1	2	–	1	–	–	–	–	1	–	–	–	–	1
–	–	1	1	1	–	1	–	–	–	–	–	–	–	–	–	–
–	–	2	1	1	1	1	–	1	1	2	1	–	1	–	–	
–	–	–	–	–	–	–	–	1	–	–	2	–	–	–	–	–
–	–	–	–	–	–	–	–	1	–	1	–	–	–	–	–	–
na	na	na	na	na	–	–	na	1	na	–	3	–	–	1	–	–
na	na	na	na	na	–	–	na	1	na	2	–	3	2	–	–	–
na	na	na	na	na	2[k]	3[n]	na	2	na	5	4	4	8	1	–	–
na	na	na	na	na	–	–	na	2	na	3	–	–	3	–	–	–
na	na	na	na	na	–	–	na	–	na	1	–	–	1	–	–	–
na	na	na	na	na	2	2[n]	na	h	na	h	h	h	h	h	–	–
pla	pla	pla	pla	pla	pla	pla	pla		pla		pla[p]			pla	pla	pla
					nuc			nuc		nuc	nuc[p]	nuc	nuc	nuc		
pro		pro	pro	pro	pro	pro	pro	pro	pro		pro			pro	pro	pro
*rbc*L	*rbc*L	*rbc*L	*rbc*L[s]	*rbc*L[s]	*tuf*A	*tuf*A	*psa*B	SSU[f]	*psb*A	18S	*gap*C	28S	18S	*rpo*	16S	16S
450[r]	450[r]	1413	450[r]	450[r]	410[r]	735[r]	450[r]	[g]	352[r]	950	330[r]	450	1700	715[r]	1000	862
CL	CL	NJ	PP	NJ	PP	FM	NJ	FM	PP	FM[b]	FM	FM	FM	FM	FM[b]	FM[b]
		BS	BS			PH		PP		BS						BS
14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Numbers of sites refers to nucleotides unless otherwise indicated. Abbreviations and annotations: FM: Fitch-Margoliash distance matrix method (Fitch and Margoliash 1967), ST: distance matrix method of Sattath and Tversky (1977); CL: CLUSTAL (Higgins and Sharp 1988; employs a UPGMA matrix method (Sneath and Sokal 1973); PP: PAUP (Swafford 1989); PH: various programs of the PHYLIP package (Felsenstein 1985); NJ: Neighbor-Joining distance matrix method (Saitou and Nei 1987); BS: bootstrap assessment of confidence intervals (Felsenstein 1985, 1988); JK: jackknife assessment of confidence intervals (Felsenstein 1985, 1988); *rpo:* DNA-dependent RNA polymerases; *gap*C: glyceraldehyde-3-phosphate dehydrogenase; *tuf*A elongation factor Tu; *psa*B and *psb*A genes of the photosynthetic machinery; *rbc*L: large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase; [a]including prochlorophytes; [b]modified FM (e.g., Elwood et al. 1985) or other least-squares method; [e]various methods; [f]16S and 18S rRNAs pairwise compared; [g]alignment critical, see original reference; [h]one or more metazoa analyzed; [i]chloroplast and nuclear genes of *Arabodopsis* studied; [k]nuclear and mitochondrial genes of *Saccharomyces* studied; [n]only mitochondrial genes analyzed; [na]not applicable since the organism is not known to possess this gene; [p]two different nuclear genes analyzed, one of endosymbiotic origin; [r]amino acid sequences; [s]small subunit of rubisco also analyzed; [t]the α-, β-, γ-, and δ-purple bacteria are *defined* taxonomically to a large extent upon the basis of 16S rRNA data (Woese 1987).

permitted detection of more branches which were consistent with the distance matrix trees (i.e., the branch bearing *Chlamydomonas* and *Chlorella,* or that bearing *Prochlorothrix* and *Synechococcus,* cf. Fig. 3) than did the other two data sets. Yet probably due to the somewhat slower rate of substitutions at second positions, the first and first plus second data sets permitted detection of common branching of some more closely related groups in >95/100 replicates (i.e., *Zea* and *Nicotiana*), whereas second positions did not. The removal of third codon positions of course markedly improves the BS and JK results when distantly species are compared (threshold options were not used in the analyses). For example, a data set was generated consisting of only third positions for the 20 *rbc*L genes; among other curious and meaningless results obtained with the third position data set, the five prokaryotes formed a monophyletic clade (100/100 replicates), simply due to the extremely high GC content of these genes at third positions (data not shown). It is interesting to note that the parsimony trees derived from analysis of *rbc*L nucleotide sequences do not differ dramatically from those obtained through analysis of the amino acid sequences (Morden and Golden 1991), although the results cannot be directly compared since different species were considered. BS and JK approaches to phylogeny can assess the reliability of branches within a tree but do not directly provide a topology which

Table 2. Summary of bootstrap (BS) and jackknife (JK) analyses for 20 *rbc*L sequences at first plus second (1 + 2 pos), second (2 pos), and first (1 pos) codon positions

Chromatium										*	*										γ-purple eubacteria
Unnamed										*	*										γ-purple eubacteria
Prochlorothrix										*							*			*	Prochlorophyte
Synechococcus										*						*	*			*	Cyanobacterium
Cyanophora										*					*	*	*				Cyanophora
Chlamydomonas										*				*	*	*	*	*	*	*	Chlorophyte
Chlorella										*				*	*	*	*		*	*	Chlorophyte
Euglena										*			*	*	*	*	*	*		*	Chlorophyte
Astasia										*			*	*	*	*	*	*		*	Chlorophyte
Nicotiana										*		*		*	*	*	*				Chlorophyte
Zea										*		*		*	*	*	*				Chlorophyte
BS 1 + 2 pos . . .	100	89	69	–	55	64	99	100	100	100	100	100	97	71	99	47	93	54	–	39	–	
JK 1 + 2 pos . . .	100	89	70	–	54	72	100	100	100	100	100	100	98	61	98	49	94	53	–	43	–	
BS 2 pos . . .	100	97	–	54	56	75	98	100	100	100	100	87	61	68	72	–	91	–	44	–	34	
JK 2 pos . . .	100	93	–	53	62	84	100	100	100	100	100	90	71	76	66	–	98	–	52	–	41	
BS 1 pos . . .	100	34	50	–	–	45	97	97	100	100	100	99	88	37	88	–	52	–	–	–	–	
JK 1 pos . . .	100	35	–	–	–	46	99	95	100	100	100	97	95	39	94	–	54	–	–	–	–	
Pylaiella	*	*	*	*	*	*	*		*												Phaeophyte
Ectocarpus	*	*	*	*	*	*	*		*												Phaeophyte
Cylindrotheca		*	*	*	*	*	*		*												Diatom
Cryptomonas			*		*	*	*		*												Cryptophyte
Antithamnion				*	*	*	*		*												Rhodophyte
Porphyridium						*	*		*												Rhodophyte
Cyanidium							*		*												Rhodophyte
Alcaligenes								*	*												β-purple eubacteria
Rhodobacter								*	*												α-purple eubacteria

Numbers indicate the number of times (rounded to integers) out of 100 replicates that those species indicated by vertical columns of asterisks occurred on a common branch in the analysis indicated. A dash (-) indicates that the grouping occurred in less (threshold chosen arbitrarily) than 33 replicates. All groupings which occurred in more than 33 replicates during any analysis are shown. The bootstrap and jackknife analyses for second positions which included the *Rhodospirillum rbc*L sequence depicted this sequence as the outgroup in 100/100 replicates and gave otherwise very similar results to those without *Rhodospirillum* except that *Prochlorothrix* and *Synchococcus* shared a common branch, as in the distance matrix trees, in 34/100 replicates (data not shown). Since non of the species listed below the numerical portion of the table ever grouped with those above same, a 100 appears for each analysis in the columns corresponding to the species indicated by eight and 11 vertical asterisks, respectively; although the second of these two columns is redundant, it has been given for clarity. For further examples to illustrate the table, the two γ-purple eubacteria shared a single common branch in 100/100 replicates of all six analyses. The two pheophytes shared a common branch with the diatom in 89/100 bootstrap replicates for the *rbc*L data set consisting of first plus second positions and 97/100 jackknife replicates for the data set consisting of only second codon positions; these three species then shared a common branch with the cryptophyte in 69/100 bootstrap replicates for first and second positions.

diagrammatically reflects both relatedness and *degree* of relatedness between sequences in the form of a discrete topology with meaningful branch lengths; for this, distance matrix methods were employed.

Two Distance Measures and Two Matrix Methods

Trees were generated with distance matrix methods and compared to the BS and JK results. The analyses focussed upon tree construction with two distance matrix methods: (1) the Neighbor-Joining (NJ) method (Saitou and Nei 1987), which has been compared to other methods in computer simulation and shown to be efficient in recovery of correct topology (Saitou and Nei 1987; Saitou and Imanishi 1989; Jin and Nei 1990) and (2) the Fitch-Margoliash (FM) least-squares (Fitch and Margoliash 1967; FITCH of the PHYLIP package) method, which has been widely employed in studies of rhodophyte and chromophyte evolution (Table 1).

It has been shown that the method of distance measure between sequences is an important factor for obtaining the correct topology (Jin and Nei 1990)

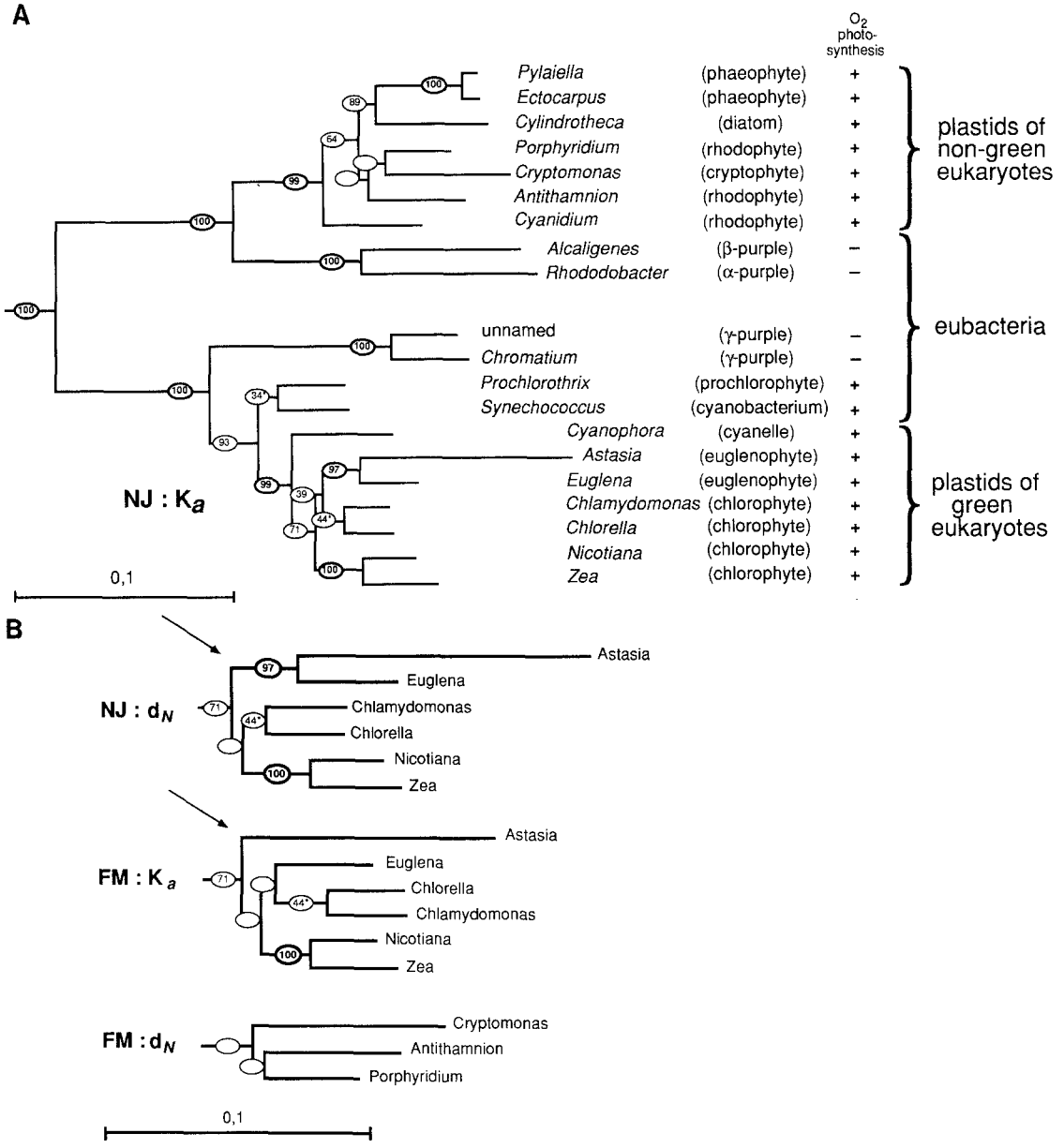


Fig. 3. A Neighbor-Joining tree for values of divergence at nonsynonymous sites between *rbc*L coding regions as measured with the weighted pathway method (Li et al. 1985) showing bootstrap confidence intervals at nodes. (See Material and Methods.) The scale bar indicates a length of K_a = 0.1; branch lengths (horizontal) are drawn to scale. Numbers left of nodes are derived from bootstrap analysis (Felsenstein 1985; DNABOOT of the PHYLIP package) and indicate the number of times out of 100 that those species *drawn* to the right on a given branch *occurred* to the right on that branch in the bootstrap analysis for first plus second codon positions (942 sites), i.e., in 100/100 replicates, neither of the two γ-purple bacteria occurred on a branch within the portion of the tree to the right of and containing the branch labeled "93," but in 7/100 bootstrap replicates, one of the eight "green" organisms or plastids occurred in the γ-purple branch. An asterisk (*) next to a number indicates that the corresponding branch was detected in less than 33 replicates for first plus second positions (942 sites) but was detected the given number of times in the data set consisting only of second positions (471 sites; see also Table 1). Branches that carry 95% confidence intervals for monophyly of groups included on the branch (≥95/100 trees since groups were not selected subsequent to analysis) are indicated with bold ovals; branches with less than 95% confidence intervals are indicated with thin ovals. Ovals without numbers indicate groupings detected in 0–33 bootstrap replicates with either first, second, or first plus second codon positions. The unlabeled outgroup is the form II *rbc*L of *Rhodospirillum rubrum* (Narang et al. 1984). Bootstrap replicates for the *Rhodospirillum* outgroup were only performed with second positions. The capacity of each organism to perform oxygen evolving-photosynthesis is indicated with "+" or "–." **B** Portions of trees constructed for matrices of K_a (Li et al. 1985) or d_N (Nei and Gojobori 1986) with the Fitch-Margoliash least-squares (FM) method (FITCH (Fitch and Margoliash 1967) of PHYLIP (Felsenstein 1981)) and d_N with the NJ method. FITCH was run with the "jumble" and "global" options. Only those portions of trees are shown in which the branching *order* (irrespective of branch length) *differed* from that of Fig. 3A. Numbers and ovals are as in Fig. 3A; all portions of trees *not* shown in Fig. 3B are identical in branching order to that shown in Fig. 3A. The scale bar indicates divergence 0.1 (K_a or d_N) at nonsynonymous sites. Arrows emphasize the different position of *Astasia*'s sequence in the FM:K_a tree.

when distance matrix methods are used. Divergence at synonymous sites between most of the *rbc*L sequences studied was too great to be reliably estimated (data not shown), so only nonsynonymous sites were considered. There are several different methods available for estimating the numbers of nonsynonymous substitutions between protein coding regions (Miyata and Yasunga 1980; Li et al. 1985; Nei and Gojobori 1986) which have been shown (Nei and Gojobori 1986) to all give very similar, yet not identical, estimates of divergence. For the purpose of tree-building, divergence at nonsynonymous sites between the *rbc*L sequences was measured with two different methods through computer programs kindly provided by the respective authors: (1) the weighted pathway method of Li et al. (1985) in which Kimura's two-parameter method model and a number of other factors including alternative pathways of amino acid replacement are considered and (2) the less complicated yet quite robust method of Nei and Gojobori (1986), in which the Jukes-Cantor model is employed. The comparison of values obtained for divergence at nonsynonymous sites, d_N (Nei and Gojobori 1986) and K_a (Li et al. 1985), is given in Fig. 2.

The method of Li et al. considers the relative likelihoods of alternative pathways for codon change, and should, in theory, tend to estimate lower numbers of nonsynonymous substitutions at the expense of higher estimates for numbers of synonymous substitutions. Thus one would expect values of K_a to be somewhat lower than those for d_N. Indeed, as shown in Figure 2 for the *rbc*L data, K_a was in general about 10% lower than d_N. The true divergence between the sequences is not known, so one cannot tell whether (1) K_a is underestimating, (2) d_N overestimating divergence, (3) both methods are giving underestimates, or (4) both are giving overestimates. At any rate, K_a was generally lower than d_N, and considerably lower than d_N (30–40%) in some specific comparisons involving the *Astasia rbc*L sequence. *Astasia longa* is a very close relative of *Euglena* (Walne and Kivic 1990) though nonphotosynthetic. There may therefore be different functional constraints upon the *rbc*L gene product, which our data suggest is evolving much faster than that of other *rbc*L genes. Thus, although K_a and d_N give quite similar estimates in computer simulation under different selection schemes and for globin genes (Nei and Gojobori 1986), with *rbc*L data the methods show appreciable differences in estimates of divergence at nonsynonymous sites.

No examples from the literature are known to us in which the effect of different distance measure methods have been directly compared upon real data in order to assess their effect upon branching order for the same sequences (sequence divergence vs genetic distance have been directly compared: Nelson et al. 1991). Thus, for the *rbc*L data set, NJ and FM trees with both values of d_N and K_a for divergence at nonsynonymous sites were constructed. The trees are summarized in Fig. 3. Comparison of these four results to the BS and JK analyses revealed that the NJ tree for values of K_a found more branches (irrespective of their length) which were supported by the bootstrap parsimony data in Table 2 than the other three trees. The complete NJ topology for values of K_a is thus shown in Fig. 3A. The *portions* of the FM trees for K_a and d_N and the NJ tree for values of d_N which *differ* in branching order from that shown in Fig. 3A, regardless of branch length, are shown in Fig. 3B. It should be noted that two different types of information are contained with the figures. The trees (or portions thereof in Fig. 3B) depict a discrete topology (not necessarily the true one) for the sequences. The amount of divergence between sequences is reflected in the branch lengths and reflects the evolution of codons within the sequences. The numbers given within ovals contain information derived from bootstrap analysis which was performed upon 2/3 or 1/3 codons (cf Table 2).

Both NJ trees (constructed with values of K_a and d_N) detected *all* branching orders which are supported in 95% of bootstrap replicates. The same is, however, not true for the FM method. The FM tree for K_a did not detect common branching of the two euglenoids, *Euglena* and *Astasia*. In our opinion, this finding should warrant caution concerning FM topologies which have note been checked with other methods. These two euglenoids are morphologically extremely similar and unquestionably very closely related through common descent, although *Astasia,* similar to "bleached" *Euglenas,* is nongreen (Siemeister and Hachtel 1990). Their *rbc*L sequences share a common branch in 97/100 bootstrap replicates (98/100 with the half-delete jackknife; see Table 2) for first plus second codon positions (942 sites), yet these sequences do not share a common branch in the FM tree for K_a, whereas they do in the NJ tree using the same K_a matrix as well as in the FM tree using the d_N matrix. This example clearly underscores with real data the findings from computer simulations, that the method of distance measure is a very important parameter for the efficiency of distance matrix methods (Jin and Nei 1990).

Very minor differences in measurement of divergence at nonsynonymous sites between K_a and d_N are thus sufficient to effect change in a portion of the FM topology which is found in more than 95/100 bootstrap replicates. This is a noteworthy result. The instability of *Astasia*'s position may relate to the high substitution rate within the *Astasia rbc*L

sequence. Using either *Zea, Anabaena, Prochlorothrix,* or *Cyanophora* as the outgroup, relative rate tests for substitutions at nondegenerate sites (Li and Tanimura 1987) showed that the substitution rate of *Astasia rbc*L is significantly higher at the 99% level (data not shown) than that of *Euglena*. This finding is relevant since computer simulation has shown that the FM method is less efficient under a variety of parameters than the NJ method in recovering correct topology, particularly when rates of evolution vary drastically in different branches (Saitou and Imanishi 1989). This property would then also seem to apply to real data in the case of *rbc*L. This case in point of *Astasia*'s branching in the FM tree of K_a values should warrant caution concerning interpretation of trees for real data for which no method estimating sampling error has been applied. The problem of how to depict the reliability of a topology within a figure has been solved differently by different authors (i.e., Hori et al. 1985; Gouy and Li 1989a; Martin et al. 1989; Ariztia et al. 1991; Morden and Golden 1991; Douglas et al. 1991). When distance matrix methods are employed to construct trees from rRNA sequences, estimates of K_{nuc} (Kimura 1980) are often used as the distance measure. If the standard errors attached to estimates of K_{nuc} are depicted directly in the tree, as Hori et al. (1985) did, a clearer picture of the data underlying and contained within the tree is provided, even if no methods of specifically addressing the reliability of individual branches are employed. Standard errors for estimates of divergence were not depicted directly in the tree of Fig. 3, yet these are indicated roughly in the graph of Fig. 2, in order to emphasize the uncertainty inherent in the measurement itself. Rather, as in Ariztia et al. (1991) and Morden and Golden (1991), results of bootstrap analysis were depicted within the tree itself.

Thus, the *rbc*L data appear to contain relevant evolutionary information concerning the origins of algal plastids. The *interpretation* of this and other data will be the topic of the following discussion.

Origins of Plant Nuclei

At present, not much is known about the biology of host cells prior to endosymbiosis; Cavalier-Smith has suggested (1987b) that these may have been anaerobic and lacking a cell wall, perhaps in some ways similar to *Giardia lamblia,* a protozoan parasite which possesses nuclei yet lacks both mitochondria and endoplasmatic reticulum (Kabnick and Peattie 1991). Most molecular evolutionary studies which might address this question have, with few exceptions (e.g., Pühler et al. 1989), been performed with 18S rRNAs (Table 1), such that our current view of early eukaryotic evolution encompasses many taxa yet relies heavily upon sequences for this single gene. Although rDNA in eukaryotes belongs to the middle repetitive fraction of the genome and is subject to processes of unequal crossover and gene conversion which may, in some cases, alter these sequences in a manner different from stochastic base-substitution models underlying molecular phylogenetic analyses (Dover 1987, 1989; Hillis et al. 1991), rRNA genes contain useful phylogenetic information. Typical for trees constructed from 18S sequences (e.g., Sogin et al. 1989) is the very deep branching within the eukaryotic portion of the tree for microsporidians and *Giardia lamblia*. The 18S rRNA branch of *Euglena*'s nucleus also emerges quite near the base of the eukaryotic trees on a common branch with trypanosomes and other kinetoplastida (Gunderson et al. 1987; Sogin et al. 1989; Bhattacharya et al. 1990; Eschbach et al. 1991). With the exception of their chlorophylls, euglenoids are cytologically and biochemically quite distinct from chlorophytes and appear related to kinetoplastida (Leedale 1967; Solomon et al. 1991), such that the deep branching of their 18S rRNA, far removed from that of chlorophytes, is not incongruent with other data. It is of interest to note that a second nuclear marker for kinetoplastida (*Trypanosoma*), glyceraldehyde-3-phosphate dehydrogenase (*gap*C), has been studied (Michels et al. 1991) and emerges similarly deep within the eukaryotic *gap*C tree. The fact that a second nuclear marker of the euglenoid–trypanosome lineage so behaves may be an indication that the position of the *Euglena* nuclear branches adequately reflect the evolution of the genes, rather than being an artefact of analysis as shown above for the *Astasia rbc*L sequence in FM trees.

The nuclear rRNA trees listed in Table 1 show common features concerning the clustering of algal nuclear rRNAs with those of certain other eukaryotic groups; the 18S trees are typified by that of Douglas et al. (1991), which is reproduced with permission of the authors in Fig. 4. These eukaryotic groups include (1) brown algae, diatoms, chrysophytes, and oomycetes, (2) dinoflagelletaes and ciliates, and (3) rhodophytes as an independent group; the positions of these groups relative to one another vary from tree to tree, analogous to the plant–animal–yeast problem discussed above. Concerning the first of these clusters, several lines of cytological/ultrastructural evidence (Beakes 1989) and molecular data (Förster et al. 1990; Ariztia et al. 1991) have been presented supporting oomycetes, diatoms, and chrysophytes as a natural group. Corollary to this view is the diphyletic nature of fungi, a theory supported by cell wall and biochemical data

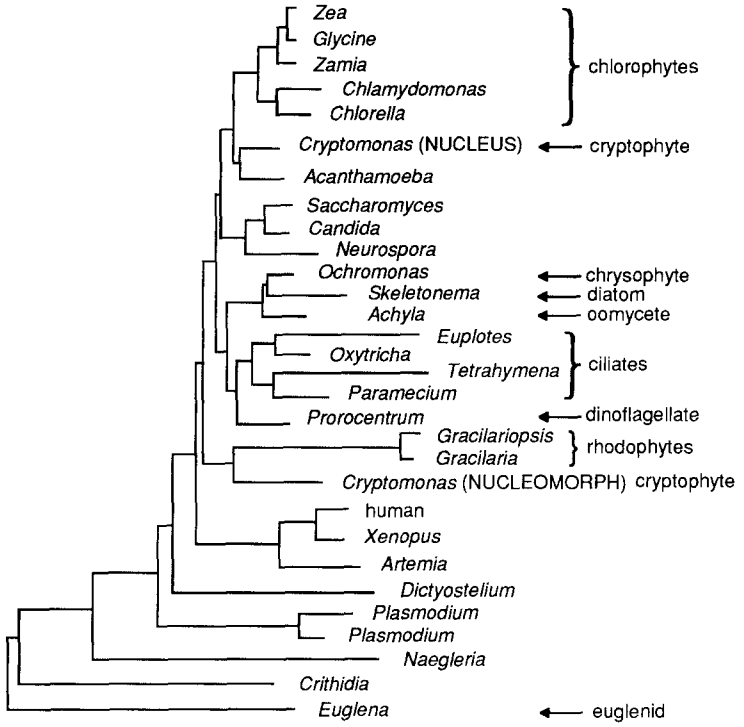


Fig. 4. Phylogenetic tree for nuclear 18S rRNA sequences constructed and presented by Douglas et al. (1991), reproduced here in modified form with the kind permission of the authors. The tree was constructed with the Neighbor-Joining method (Saitou and Nei 1987) for a matrix of K_{nuc} distance values (Kimura 1980). The nuclear and nucleomorph rRNA sequences of *Cryptomonas* are indicated. Of particular interest is the distinct separation of *Cryptomonas* nuclear and nucleomorph rRNA sequences. (See text.) For further details, the original reference should be consulted.

(Bartnicki-Garcia 1987) and perceived by Mereschkowsky (1910d: Fig. 1). Ariztia et al. (1991) have presented very careful analyses of 18S rRNA sequences. It should be noted that in some cases there exist considerable disconcordance between phylogenies derived from rRNA and protein sequences for the same organisms. Loomis and Smith (1990), for example, analyzed sequences for eight different proteins, all of which yielded phylogenies for *Dictyostelium* relative to yeast and animals different from that reflected by the 18S rRNA data in Ariztia et al. (1991). [One may note that 16S rRNA, Rubisco, and *psb*A sequences also testify differently to the relationship of *Prochlorothrix* to cyanobacteria and green plastids (Turner et al. 1989; Morden and Golden 1989).] In general, we would currently tend toward the view that the origins and early evolution of algal nuclear lineages have not been studied with sufficient numbers of different genes so as to provide a final picture of phylesis for the nuclei of these highly diverse organisms, yet the perspectives for future studies of these questions are promising.

Origins of Plastids

Molecular data unequivocally support endosymbiotic origins for plastids. (See references to Table 1.) Several recent molecular contributions on *rbc*L sequences have appeared in the literature (Valentin and Zetsche 1990b; Douglas et al. 1990; Morden and Golden 1991; Assali et al. 1991a) which have indicated that the plastid genes of rhodophytes and chromophytes are only distantly related to the plastids of chlorophytes, and that the latter are much more closely related to cyanobacteria than the former. These data have been interpreted as evidence compatible with polyphyletic plastid origins (Douglas et al. 1990; Valentin and Zetsche 1990b; Assali et al. 1991a), a suggestion forwarded previously on the basis of biochemical evidence (e.g., Mereschkowsky 1910a–d; Raven 1970). Polyphyletic plastid origins are intended here to indicate independent *primary* endosymbioses either (1) between different *prokaryotic* plastid progenitors and any given host cell lacking plastids or (2) between the same prokaryotic symbiont (within the resolution of available methods) and different host cells. Secondary endosymbiosis, in which a photosynthetic *eukaryotic* alga possessing plastids (regardless of their evolutionary history) is engulfed by another eukaryote would not necessitate, but would not exclude, polyphylesis for plastids.

The ancestors of plastids have traditionally been sought among the cyanobacteria (Mereschkowsky 1905) or prochlorophytes (Raven 1970); cyanobacteria, prochlorophytes, chloroplasts, and rhodoplasts are unique in that they evolve oxygen in photosynthesis. Cyanobacteria and rhodoplasts share phycobilins as a unique common trait; prochlorophytes and chloroplasts share chlorophylls *a* and *b* as a unique common trait. Purple bacteria on the

other hand (1) do not evolve oxygen in photosynthesis, (2) do not possess phycobilins, and (3) do not possess chlorophylls *a* and *b*. It is thus somewhat surprising that the *rbc*L sequences of both an α-purple and a β-purple bacterium reveal such close affinity to the *rbc*L genes of rhodo- and chromophyte plastids, whereas the cyanobacterial and prochlorophyte *rbc*L genes appear quite closely related to their counterparts of green plastids and γ-purple bacteria (Fig. 3A).

The *rbc*L data suggest that cyanobacteria are descendants of a well-differentiated "purple" bacterial lineage. This is an intriguing finding which conflicts to some extent with bacterial phylogenies based upon 16S rRNA sequences (Woese 1987). Furthermore, since O_2-evolving photosynthesis occurs in chlorophyte, rhodophyte, and chromophyte plastids, yet not in purple bacteria, the *rbc*L data would suggest that the *common ancestor of purple bacteria and cyanobacteria would have been an O_2-evolving photoautotroph,* and that extant purple bacteria have lost their capacity to use water as an electron donor. This finding is compatible with the considerable body of biochemical evidence suggesting that the common ancestors of purple bacteria and cyanobacteria possessed *both* photosystems I *and* II (reviewed by Olson and Pierson 1987). It was previously suggested on the basis of cytochrome *c* (Dickerson 1980) and 16S rRNA (Woese 1987) data that chlorophyll-based photosynthesis in purple bacteria represents a biochemical symplesiomorphy, the nonphotosynthetic state of many purple bacteria being due to independent secondary loss. The *rbc*L sequence data would go one step further by suggesting that the common ancestor of purple- and cyanobacteria was capable of oxygen-evolving photosynthesis, that only cyanobacteria and plastids retained both photosystems, and that plastids are of polyphyletic origin.

Yet *rbc*L sequences reflect somewhat different affinities between cyanobacteria, purple bacteria, and plastids than do 16S rRNA data. Based upon 16S rRNA sequences, cyanobacteria appear distinct from the purple bacteria (Woese 1987; Stackebrandt et al. 1988) and are depicted to have diverged from these prior to the separation of the α- and δ-purple lineages from the β- and γ-purple lineages (Woese 1987; Zavarzin et al. 1991), though one may keep in mind that the distinctions between α-, β-, δ-, and γ-purple bacteria rest heavily upon a single gene (16S) phylogeny; 16S rRNA sequence phylogenies depict either (1) somewhat closer affinities between rhodoplasts and cyanobacteria than between the latter and chloroplasts or (2) common branching of two main plastid lineages derived of cyanobacteria, one leading to chromophytes, euglenoids and red algae, the other leading to chlorophytes (cf. Fox et al. 1980; Witt and Stackebrandt 1988; Wolters et al. 1989; Markowicz and Loiseaux-de Goër 1991; Douglas and Turner 1991). The 16S sequence of *Alcaligenes* (β-purple) clusters with α- and γ-purple bacteria to the exclusion of both red and green plastids (Markowicz and Loiseaux-de Goër 1991). Discrepancies between 16S rRNA and *rbc*L data concerning plastid origins remain to be explained.

If one assumes that the *rbc*L gene phylogeny (Fig. 3) soundly reflects the evolution of the species and organelles under study, explanations must be sought in the evolution of 16S rRNA genes which may account for the data. Aside from the problems of fluctuating GC content in bacterial evolution (Muto and Osawa 1987), one may recall that ribosomal RNAs evolve in the structural context of roughly 60 ribosomal proteins, about two-thirds of which are encoded by the nucleus in higher plants and green algae (Christopher et al. 1988; Mache 1990; Subramanian et al. 1991). Moreover, plastid ribosomes possess a number of unique features which their counterparts in free-living prokaryotes do not (Zhou and Mache 1989; Subramanian et al. 1991). One thus cannot exclude the possibility that functional constraints relating to molecular coevolution of ribosomal RNAs within the structural context of their evolving cognate ribosomal proteins (Johnson et al. 1990; Martin et al. 1990) subsequent to endosymbiosis may have generated considerable discontinuity in the evolution of 16S rRNAs from prokaryotes to plastids. Selective neutrality of many mutations observed in plastid 16S ribosomal RNA may not hold, thereby influencing 16S rRNA phylogenies.

If, on the other hand, one assumes that 16S rRNA gene phylogenies soundly reflect the evolution of the species and organelles under study, explanations must be sought in the evolution of *rbc*L genes. As one alternative, horizontal gene transfer has been suggested for *rbc*L from a β-purple bacterial donor to a cyanobacterium in the rhodophyte plastid lineage to explain the differences between *rbc*L and other genetic data (Boczar et al. 1989; Morden and Golden 1991; Assali et al. 1991a); i.e., red plastids derive from cyanobacteria which possessed β-purple bacterial *rbc*L genes. As a second alternative, the ancestors to plastids, cyanobacteria, and purple and bacteria may have simply possessed two *rbc*L genes, one of which was secondarily lost in each of the lineages leading to red and green plastids. Some α-purple bacteria do possess two *rbc*L genes in distinct Calvin cycle operons (Gibson et al. 1991), although the form II *rbc*L sequence, clearly an outgroup to the genes given in

Fig. 3, shares no greater similarity to red or green *rbc*L sequences studied (all K_a values >2.0; data not shown) and is thus an unlikely candidate to represent a descendant of any duplication event which would account for the topology observed. Nonetheless, neither early gene duplication, as in the case of the ferredoxins (Schwartz and Dayhoff 1981), nor association between a putative *rbc*L and ferrodoxin gene duplication, can be formally ruled out. As a further alternative, one may consider the possibility that plastid *rbc*L sequences perhaps do not strictly reflect the evolution of plastids, but may also contain sequences once native to mitochondria, believed to be descendants of α-purple bacteria. In this scenario, intracellular *rbc*L gene transfer between organelles (mitochondrium to plastid) in the red lineage rather than horizontal gene transfer between organisms could be invoked to account for the similarity of red plastid and purple bacterial *rbc*L genes; it is currently not known whether rhodo- and chromoplasts are as refractory to DNA uptake during evolution as chloroplasts are (Palmer et al. 1988; Clegg et al. 1991). If plastid genomes ultimately prove to be chimeric (Loiseaux-de Goër et al. 1988; Cattolico and Loiseaux-de Goër 1989), the long-term objective of clarifying plastid phylogeny with molecular methods would appear to require the study of many more (if not all) genes of a given plastid, not just a few such as 16S rRNA and *rbc*L.

Molecular data from rhodophyte and chromophyte plastid genomes are still rather scarce. The analyses of other genes of rhodophyte and chromophyte plastids paint a less complete picture for polyphylesis vs monophylesis of plastids. Ferredoxin phylogenies have been presented (Hase et al. 1983) that would indicate an origin of chromophyte plastids from red algal endosymbionts, though due to an early gene duplication, ferredoxin sequences could support either polyphyletic or monophyletic plastid origins (Schwartz and Dayhoff 1981). Analyses of admittedly short 5S rRNA sequences conflict in that they depict both clustering of rhodoplasts within the green plastid lineage (van den Eynde et al. 1988) as well as distinct separation of chloroplast and rhodoplast lineages (Sommerville et al. 1992). The sequence for *psa*B from a chromophyte (Assali and Loiseaux-de Goër 1992) does not display the divergence from chlorophyte and cyanobacterial counterparts found for the *rbc*L sequence from the same organism. However, for ferredoxin, 5S RNA, and *psa*B, reference sequences from purple bacteria have not been analyzed to such an extent that questions of purple vs cyanobacterial origins for rhodo- and chromophyte plastids could currently be addressed with these sequences. The case for monophyly of rhodo-, chromo-, and chlorophyte plastids has been argued on the basis of overall gross similarities in plastid genome structure (Kowallik 1989). It is, however, quite clear that these plastids are distinct in many respects from those of chlorophytes, i.e., through the presence of phycobiliproteins, through different nuclear/plastid gene localization (a single plastid operon for both subunits of Rubisco; Valentin and Zetsche 1989), and through the presence of a bimolecular plastid genome in some chromophytes (Loiseaux-de Goër et al. 1988; Cattolico and Loiseaux-de Goër 1989).

It would seem clear that the discrepancies between the still-rather-narrowly-based *rbc*L, 16S, and other data sets should warrant investigation of further genes from prokaryotes and rhodo- and chromophyte plastids in order to clarify the evolutionary processes underlying their descent. Again, it is of interest to note that the *rbc*L data are quite compatible with the hypotheses that the common ancestor of purple bacteria and cyanobacteria was capable of oxygen-evolving photosynthesis (Olson and Pierson 1987; van Gorkom 1987).

Origins of Plant Mitochondria

Chlorophyte, animal, and fungal mitochondria, as characterized by 16S (Yang et al. 1985) and cytochrome *c* (Dickerson 1980) sequences, appear to derive from α-purple bacteria. Mitochondrial ancestors were at one time photoautotrophs (Dickerson 1980; Woese 1987) and were almost certainly capable of aerobisis (Margulis 1981). It is of some interest to note in this context that cyanobacteria are also fully capable of aerobic respiration in which an electron transport chain separate from that of photosynthesis is employed (Pescheck 1987). Since purple bacteria and cyanobacteria appear to have possessed relatively recent common ancestors in bacterial evolution (Woese 1987) and in light of the arguments put forth above concerning the oxygen-evolving photosynthetic capacity for the common ancestor of cyano- and purple bacteria, it would seem likely that the energy-producing systems of eukaryotic cells, photosynthesis and mitochondrial respiration, have descended through differential loss of one or the other electron transport chain from a photosynthetic prokaryote which possessed both ATP-yielding systems.

Compared to plastids, relatively few sequences of genes for which adequate reference sequences are known have been determined from plant mitochondria. Chlorophyte mitochondrial DNA studied to date evolves at a rate almost 100 times lower than

that of animal mitochondria (Wolfe et al. 1987); *Chlamydomonas* may be an exception in this respect (Gray et al. 1989). The extreme rate difference poses considerable, yet perhaps not insurmountable, problems in data analysis when plant, animal, and fungal mitochondrial sequences are compared (Gray et al. 1989). To our knowledge, no molecular sequences at all are currently available from mitochondria of other than the chlorophyte lineage of plants. DNA sequences from rhodo- and chromophyte mitochondria should be of interest concerning the phylogeny of these organelles.

Secondary Endosymbiosis

Whereas eukaryotes in general are chimeric, some eukaryotes, the chromophyte algae, may be more highly chimeric than others and require hypotheses more complex than simple endosymbiosis to account for their evolutionary origins. Chromophytes are characterized by the presence of chlorophyll *a* and chlorophyll *c*, the absence of chlorophyll *b*, and the presence of three or four membranes surrounding their chloroplasts (Christensen 1989). The number and cellular continuum of membranes surrounding chromophyte plastids have attracted particular attention, since these and other cytological characters have been interpreted as evidence of secondary endosymbiosis (Tomas and Cox 1973) between eukaryotic hosts and eukaryotic endosymbionts in chromophyte evolution (Gibbs 1978, 1981; Whatley 1981, 1989; Cavalier-Smith 1986, 1987a; see also Gibbs 1990, Sitte 1990, and Melkonian 1991 for recent reviews). The first clear molecular evidence in support of secondary endosymbiosis for the origin of chromophyte plastids was recently brought forth (Douglas et al. 1991) through analysis of DNA sequences from the nucleomorph, putatively a vestigial nucleus of the eukaryotic endosymbiont, from the cryptophyte *Cryptomonas* Φ. A number of modern symbioses involving eukaryotic algae which fall short of endosymbiosis sensu stricto, yet are highly reminiscent of same, are known. The host cells are typically dinoflagellates or ciliates, while the symbionts are cryptophyte, chrysophyte, or green unicellular algae which can assume habits very similar to that of plastids. A few examples are discussed here. (There are numerous other examples.)

The precursor of a true endosymbiosis is perhaps analagous to that which can be observed in the case of the dinoflagellate *Amphidinium poecilochroum*. This species engulfs cryptophyte algae which concomitantly relinquish their cell wall. The remainder of the cellular structures remain intact for some time within the dinoflagellate cell before finally being digested. The cellular content of the cryptophytes are surrounded in the host by one membrane and were once thought to be dinoflagellate plastids, yet *A. poecilochroum*, in contrast to the following examples, cannot be cultivated autotrophically (Larsen 1988).

Gymnodinium acidotum has developed a closer symbiotic relationship with a cryptophyte, *Chroomonas* sp. (Farmer and Roberts 1990; Fields and Rhodes 1991). In this case the cryptophyte is protected by a perialgal vacuole for extended periods of time, but is finally digested, and the symbiosis needs to be reestablished. To some extent, metabolic coordination between the "symbiont" and the host cell has been achieved, as demonstrated by the presence of starch grains in the host cell cytoplasm and by divisions of the cryptophyte within the host cell. *Peridinium balticum* and *Peridinium foliaceum* have established a permanent symbiotic relationship with chrysophytes. Each contains an entire ramified chrysophyte in its cytoplasm which is separated from the host by one membrane. Here again the cells divide synchronously and starch grains can be found in the cytoplasm of the host (Tomas and Cox 1973; Jeffrey and Vesk 1976).

In the dinoflagellate *Lepidodinium viridae*, the endosymbiotic alga is reduced to roughly the same stage as the plastid (algal-remnant) of modern cryptophytes. The endosymbiont is separated from the host by a double membrane, and a vestigial nucleus is contained within the periplastidal space. The plastids are distributed throughout the host cytoplasm and likely represent a green algae, as suggested by pigment content and ultrastructure (Wanatabe et al. 1987, 1990). Finally, the dinoflagellate *Amphidinium wigrense* contains plastids which appear, on the basis of pigmentation and ultrastructure, to derive from cryptophyte algae. These are separated from the host cytoplasm by one extra membrane, as are the plastids of *Euglena*, and no nucleomorph is left (Wilcox and Wedemayer 1985).

Only recently have molecular data been presented in direct support of secondary endosymbiosis; the 18S RNA sequences from the nucleus and nucleomorph of *Cryptomonas* Φ, a cryptophyte, show that these two compartments have undergone a considerable period of separate evolution before reunion within the same cell wall (Douglas et al. 1991; Fig. 4). The nucleomorph sequence is closely related to the nuclear rRNA sequence of red algae, suggesting that these may have given rise to the secondary endosymbiont. The *rbc*L data provide further indirect support for secondary endosymbiosis in that the very close affinities between *rbc*L sequences of red algal plastids and those of chro-

mophyte plastids may be interpreted as a secondary symbiotic origin for chromophyte plastids from rhodophyte ancestors (Valentin and Zetsche 1990a,b; Morden and Golden 1991; Assali et al. 1991a).

The *rbc*L and *rbc*S genes of *Euglena* represent an interesting case and can be interpreted in a relatively straightforward manner. The plastids of *Euglena* are surrounded by three membranes instead of two as in chlorophytes. This circumstance and other plastid ultrastructural features have been forwarded (Gibbs 1978, 1981; Whatley 1981) as evidence in favor of a secondary endosymbiosis from a chlorophyte progenitor for the origin of *Euglena*'s plastids. In this scenario, a chlorophyte would have been phagocytized by the *Euglena* host cell; the third chloroplast membrane is interpreted as a vestigial plasmalemma of the endosymbiont (Gibbs 1978; Whatley 1989). This hypothesis would be consistent both (1) with the deep branching of *Euglena*'s nuclear rRNA and cytochrome *c* (Kemmerer et al. 1991) sequences amongst eukaryotes, quite far removed from the chlorophyte nuclear lineage and (2) with the close relationship of *Euglena*'s *rbc*L and *rbc*S genes to their chlorophyte counterparts (Morden and Golden 1991; Assali et al. 1991a). Also, the *tuf*A (Palmer et al. 1989; Ludwig et al. 1990) and *psa*B (Assali and Loiseaux-de Goër 1992) data reflect a close affinity between some of *Euglena*'s plastid genes and those of chlorophytes. However, the 16S-rRNA genes of *Euglena* appear to be somewhat more closely related to those of rhodophytes than to those of chlorophytes (Markowicz and Loiseaux-de Goër 1991; Douglas and Turner 1991), leaving a discrepancy and thus some open questions concerning *Euglena*'s 16S rRNA genes. *If* the 16S rRNA and protein phylogenies accurately reflect the phylogeny of the genes, this discrepancy could suggest a chimeric nature for the *Euglena* plastome, notwithstanding the other alternatives discussed in earlier sections.

It is interesting to note in this context that *rbc*S is a nuclear gene in *Euglena* and chlorophytes, whereas it is contained within the plastid-encoded *rbc*-operon in rhodo- and chromophytes (Valentin and Zetsche 1989; Assali et al. 1991a). In *Euglena*, the nuclear *rbc*S gene structure is wholly distinct from that found within chlorophytes (Chan et al. 1990). The *Euglena rbc*S gene encodes an octameric polypeptide from which Rubisco small subunit monomers are posttranslationally processed from the octameric precursor *subsequent* to transport of the precursor through the three plastid membranes; the transit peptide of the *Euglena rbc*S precursor has characteristics of both a signal peptide and a transit peptide, as would be expected since the precursor must traverse an extra membrane in routing (Chan et al. 1990). This is in marked contrast to the conservative gene structure of monomeric nuclear *rbc*S genes in chlorophytes (Wolter et al. 1988), the products of which must only pass the chloroplast envelope. If indeed *Euglena*'s plastids are of secondary endosymbiotic origin, interesting questions concerning intracellular gene transfer follow: the lack of a nucleomorph in *Euglena* would suggest either (1) that plastid genes (i.e., *rbc*S) were transferred directly from the plastid genome to the secondary host nucleus, or (2) that *Euglena*'s (no longer existent) secondary endosymbiont nucleus already contained transferred genes (i.e., *rbc*S) from the plastid before secondary endosymbiosis, which were then transferred in a second step to the host cell nucleus. In chromophyte algae with more than three plastid membranes, the nature of precursor transport and transit peptides for nuclear-encoded chloroplast proteins should be of interest concerning the evolutionary origin of both the genes and the protein-sorting systems (Smeekens et al. 1990), which insure that the gene products reach the organelle for which they are destined.

Conclusion

It is quite clear that different molecular markers can provide conflicting views for the course of plastid evolution and that a single-gene phylogeny necessarily reflects the evolutionary history of the gene, rather than that of the entire genome within which it is contained. We have covered some of the current data which address the question and have discussed this in the general context of early eukaryotic and plastid progenitor evolution. We have shown how at least one artefact of data analysis may result in erroneous and perhaps misleading trees. We have demonstrated the importance of distance measure in real data sets for the performance of two distance matrix methods. We have tried to point out areas in which we feel that more analyses need to be performed and more genes need to be studied in order to clarify some aspects of eukaryotic evolution specifically as it applies to plant organelles. We are confident that molecular evolutionary studies of eukaryotic organellar origins will continue to provide stimulating questions for some time to come.

Acknowledgments. We wish to thank H. Brinkmann, H.-J. Bandelt, and R. Cerff for many stimulating discussions and critical remarks on the MS. We thank the Gesellschaft für Biotechnologische Forschung, Braunschweig, for the generous permission to use their computer facilities.

References

Andersen K, Caton J (1987) Sequence analysis of the *Alcaligenes eutrophus* chromosomally encoded ribulose-1,5-bisphos-

phate carboxylase large and small subunit genes and their gene products. J Bateriol 169:4547–4558

Ariztia E, Anderson R, Sogin M (1991) A new phlogeny for chromophyte algae using 16S-like rRNA sequences from *Mallomonas papillosa* (Synurophyceae) and *Tribonema aequale* (Xanthophyceae). J Phycol 27:428–436

Assali N, Martin W, Sommerville C, Loiseaux-de Goër S (1991a) Evolution of the Rubisco operon from prokaryotes to algae: Structure and analysis of the *rbc*S gene of the brown alga *Pylaiella littoralis*. Plant Mol Biol 17:853–863

Assali N, Loiseaux-de Goër S (1992) Sequence and phylogeny of the *psa*B gene of the brown alga *Pylaiella littoralis*. J Phycol 28:209–213

Baldauf S, Palmer J (1990) Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. Nature 344:262–265

Bartnicki-Garcia S (1987) The cell wall: A crucial structure in fungal evolution. In: Rayner A, Brasier CM, Moore D (eds) Evolutionary biology of the fungi. Cambridge University Press, Cambridge, pp 389–403

Beakes G (1989) Oomycete fungi: their phylogeny and relationship to chromophyte algae. In: Green J, Leadbeater BSC, Diver WL (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford, pp 325–342

Bhattacharya D, Elwood HJ, Goff L, Sogin ML (1990) Phylogeny of *Gracilaria lemaneiformis* (Rhodophyta) based on sequence analysis of its small subunit ribosomal RNA coding region. J Phycol 26:181–186

Boczar BA, Delaney TP, Cattolico RA (1989) The 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

Brinkmann H, Martinez P, Quigley F, Martin W, Cerff R (1987) Endosymbiotic origin and codon bias of the nuclear gene for chloroplast glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Evol 26:24–33

Cattolico RA, Loiseaux-de Goër S (1989) Analysis of chloroplast evolution and phylogeny: A molecular approach. In: Green J, Leadbeater BSC, Diver WL (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford, pp 85–100

Cavalier-Smith T (1986) The kingdom chromista: origin and systematics. Prog Phycol Res 4:309–347

Cavalier-Smith T (1987a) The origin of eukaryote and archaebacterial cells. Ann NY Acad Sci 503:17–54

Cavalier-Smith T (1987b) The simultaneous symbiotic origin of mitochondria, chloroplasts and microbodies. Ann NY Acad Sci 503:55–71

Chan R, Keller M, Canaday S, Weil J, Imbault P (1990) 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

Christen R, Ratto A, Baroin A, Perasso R, Grell K, Adouette A (1991) An analysis of the origins of the metazoans, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of tripoblasts. EMBO J 10:499–503

Christensen T (1989) The Chromophyta, past and present. In: Green J et al. (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford

Christensen T (1989) The Chromophyta, past and present. In: Green J et al. (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford

Christopher DA, Cushman JC, Price CA, Hallick RB (1988) Organisation of ribosomal protein genes *rpl*23, *rpl*12, *rps*19, *rpl*22 and *rps*3 on the *Euglena gracilis* chloroplast genome. Curr Genet 14:275–286

Clegg M, Learn G, Golenberg E (1991) Molecular evolution of chloroplast DNA. In: Selander R, Clark A, Whittam TS (eds) Evolution at the molecular level. Sinauer, Sunderland, MA, pp 135–149

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

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

Dickerson R (1980) Evolution and gene transfer in purple photosynthetic bacteria. Nature 283:210–212

Douglas S, Dunford D, Morden C (1990) Nucleotide sequence of the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from *Cryptomonas* Φ: Evidence supporting the polyphyletic origin of plastids. J Phycol 26:500–508

Douglas S, Murphy C, Spencer D, Gray M (1991) Cryptomonad algae are evolutionary chimaeras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148–151

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

Dover GA (1987) DNA Turnover and the molecular clock. J Mol Evol 26:47–58

Dover GA (1989) Linkage disequilibrium and molecular drive in the rDNA multigene family. Genetics 122:249–252

Dron M, Rahire M, Rochaix JD (1982) Sequence of the chloroplast region of *Chlamydomonas reinhardtii* containing the gene for the large subunit of ribulose-1,5-bisphosphate carboxylase and parts of its flanking genes. J Mol Biol 162:775–793

Elwood HJ, Olsen GJ, Sogin M (1985) The small subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytrocha nova* and *Stylonicha pustulata*. Mol Biol Evol 2: 399–410

Eschbach S, Wolters J, Sitte P (1991) Primary and secondary structure of the nuclear small subunit ribosomal RNA of the cryptomonad *Pyrenomonas salina* as inferred from the gene sequence: Evolutionary implications. J Mol Evol 32:247–252

Falcone D, Tabita F (1991) Expression of endogenous and foreign ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) genes in a RuBisCO deletion mutant of *Rhodobacter sphaeroides*. J Bacteriol 173:2099–2108

Farmer MA, Roberts KR (1990) Organelle loss in the endosymbiont of *Gymnodinium acidotum* Dinophyceae. Protoplasma 153:178–185

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum-likelihood approach. J Mol Evol 17:368–376

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

Felsenstein J (1988) Phylogenies from molecular sequences: Inference and reliability. Annu Rev Genet 22:521–565

Fields SD, Rhodes RG (1991) Ingestion and retention of *Chroomonas* spp. (Cryptophyceae) by *Gymnodinium acidotum* (Dinophyceae). J Phycol 27:525–529

Fitch WM, Margoliash E (1967) Construction of phylogenetic trees. Science 155:279–284

Förster H, Coffey M, Elwood H, Sogin M (1990) Sequence analysis of the small subunit ribosomal RNAs of three zoosporic fungi and implications for fungal evolution. Mycologia 82: 306–312

Fox G, Stackebrandt E, Hespell R, Gibson J, Maniloff J, Dyer T, Wolf R, Balch W, Tanner R, Magrum L, Zablen L, Blekemore R, Gupta R, Bonen L, Lewis B, Stahl D, Luehrsen K, Chen K, Woese C (1980) The phylogeny of prokaryotes. Science 209:457–463

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

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

Gibbs S (1990) The evolution of algal chloroplasts. In: Weissner W et al. (eds) Experimental phycology I. Springer, Berlin

Gibson JL, Falcone DL, Tabita FR (1991) Nucleotide sequence, transcriptional analysis and expression of genes encoded within the form I CO_2 fixation operon of *Rhodobacter sphaeroides*. J Biol Chem 266:14646–14653

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

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

Gouy M, Li W-H (1989a) Molecular phylogeny of the kingdoms Animalia, Plantae, and Fungi. Mol Biol Evol 6:109–122

Gouy M, Li W-H (1989b) Phylogenetic analysis based on rRNA sequences supports the archaebacterial rather than the eocyte tree. Nature 339:145–147

Gray M, Cedergren R, Abel Y, Sankoff D (1989) On the evolutionary origin of the plant mitochondrion and its genome. Proc Natl Acad Sci USA 86:2267–2271

Gunderson JH, Elwood H, Ingold A, Kindle K, Sogin M (1987) Phylogenetic relationships between chlorophytes, chrysophytes, and oomycetes. Proc Natl Acad Sci USA 84:5823–5827

Hase T, Inoue K, Hagihara N, Matsubara H, Williams M, Rogers L (1983) Ferredoxin from *Aphanothece haplophitica*, a unicellular blue-green alga: close relationship to ferredoxins from filamentous blue-green algae and phylogenetic implications. J Biochem 94:1457–1464

Hendriks L, Goris A, Neefs J-M, van de Peer Y, Hennebert G, de Wachter R (1989) The nucleotide sequence of the small ribosomal subunit RNA of the yeast *Candida albans* and the evolutionary position of the fungi among the eukaryotes. System Appl Microbiol 12:223–229

Hendriks L, De Baere R, Van de Peer Y, Neefs J, Goris A, De Wachter R (1991) The evolutionary position of the rhodophyte *Porphyra umbilicalis* and the basidiomycete *Leucosporidium scottii* among other eukaryotes as deduced from complete sequences of small ribosomal subunit RNA. J Mol Evol 32:167–177

Higgins DH, Sharp PM (1988) CLUSTAL: A package for performing multiple sequence alignment on a microcomputer. Gene 73:237–244

Hillis DM, Moritz C, Porter C, Baker J (1991) Evidence for biased gene conversion in ribosomal DNA. Science 251:308–310

Hiratsuka J, Shimada H, Whittier R, Ishibashi T, Sakamoto M, Mori M, Kondo C, Honji Y, Sun C-R, Meng B-Y, Li Y-Q, Kanno A, Nishizawa Y, Hirai A, Shinozaki K, Sugiura M (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol Gen Genet 217:185–194

Hori H, Lim B-L, Osawa S (1985) Evolution of green plants as deduced from 5S rRNA sequences. Proc Natl Acad Sci USA 82:820–823

Hwang S-R, Tabita FR (1991) Cotranscription, deduced primary structure, and expression of the chloroplast-encoded *rbc*L and *rbc*S genes of the marine diatom *Cylindrotheca* sp. strain N1. J Biol Chem 266:6271–6279

Jeffrey SW, Vesk M (1976) Further evidence for a membrane-bound endosymbiont within the dinoflagellate *Peridinium foliaceum*. J Phycol 12:450–455

Jin L, Nei M (1990) Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol Biol Evol 7:82–102

Johnson CH, Kruft K, Subramanian AR (1990) Identification of a plastid-specific ribosomal protein in the 30S subunit of chloroplast ribosomes and isolation of the cDNA clone encoding its cytoplasmic precursor. J Biol Chem 265:12790–12795

Kabnick K, Peattie D (1991) *Giardia:* A missing link between prokaryotes and eukaryotes. Am Scientist 79:34–43

Kemmerer EC, Lei M, Wu R (1991) Structure and molecular evolutionary analysis of a plant cytochrome *c* gene: Surprising implications for *Arabidopsis thaliana*. J Mol Evol 32:227–237

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

Kostrzewa M, Valentin K, Maid U, Radetzky R, Zetsche K (1990) Structure of the rubisco operon from the multicellular red alga *Antithamnion spec.* Curr Genet 18:465–460

Kowallik K (1989) Molecular aspects and phylogenetic implications of plastid genomes of certain chromophytes. In: Green J et al. (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford

Larsen J (1988) An ultrastructural study of *Amphidinium poecilochroum* (Dinophyceae), a phagotrophic dinoflagellate feeding on small species of cryptophytes. Phycologia 27:366–377

Leedale G (1967) Euglenoid flagellates. Prentice Hall, Englewood Cliffs, NJ

Li W-H, Wu C-I, Luo C-C (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Li W-H, Tanimura M (1987) The molecular clock runs more slowly in man than in apes and monkeys. Nature 326:93–96

Li W-H (1989) A statistical test of phylogenies estimated from sequence data. Mol Biol Evol 6:424–435

Liaud M-F, Zhang DX, Cerff R (1990) Differential intron loss and endosymbiotic transfer of chloroplast glyceraldehyde-3-phosphate dehydrogenase genes to the nucleus. Proc Natl Acad Sci USA 87:8918–8922

Loiseaux-de Goër S, Markowicz Y, Dalmon J, Audren H (1988) Physical maps of the two circular plastid DNA molecules of the brown alga *Pylaiella littoralis* (L.) Kjellm. Curr Genet 14:155–162

Loomis W, Smith D (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc Natl Acad Sci USA 87:9093–9097

Ludwig W, Weizenegger D, Betzl D, Leidel E, Lenz T, Ludvigsen D, Möllenhoff D, Wenzig P, Schleifer KH (1990) Complete nucleotide sequences of seven eubacterial genes coding for the elongation factor Tu: functional, structural and phylogenetic evaluations. Arch Microbiol 153:241–247

Mache R (1990) Chloroplast ribosomal proteins and their genes. Plant Sci 72:1–12

Margulis L (1981) Symbiosis in cell evolution. Freeman, San Francisco

Markowicz Y, Loiseaux-de Goër S (1991) Plastid genomes of Rhodophyta and Chromophyta share a common origin which differs from that of Chlorophyta and have a composite origin as do those of Euglenophytes. Curr Genet 20:427–430

Martin W, Cerff R (1986) Prokaryotic features of a nucleus encoded enzyme: cDNA sequences for chloroplast and cytosolic glyceraldehyde-3-phosphate dehydrogenases from mustard (*Sinapis alba*). Eur J Biochem 159:323–331

Martin W, Gierl A, Saedler H (1989) Molecular evidence for pre-Cretaceous angiosperm origins. Nature 339:46–48

Martin W, Lagrange T, Li Y-F, Bisanz-Seyer C, Mache R (1990) Hypothesis for the evolutionary origin of the chloroplast ribosomal protein L21 of spinach. Curr Genet 18:553–556

McIntosh L, Poulsen C, Bogorad L (1980) Chloroplast gene sequence for the large subunit of ribulose bisphosphate carboxylase from maize. Nature 288:556–560

Melkonian M (1991) Systematics and evolution of the algae. Prog Botany 52:271–307

Mereschkowsky C (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biol Centralbl 25:593–604

Mereschkowsky C (1910) Theorie der zwei Plasmaarten als Grundlage der Symbiogenesis, einer neuen Lehre von der Entstehung der Organismen. Biol Centralbl 30:278–288

Mereschkowsky C (1910b) ibid. 30:289–303

Mereschkowsky C (1910c) ibid. 30:322–347

Mereschkowsky C (1910d) ibid. 30:353–367

Michels P, Marchand M, Kohl L, Allert S, Vellieux FMD, Wierenga R, Opperdoes F (1991) The cytosolic and glycosomal isozymes of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma brucei* have a distant evolutionary relationship. Eur J Biochem 198:421–428

Miyata T, Yasunga T (1980) Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36

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

Morden CW, Golden SS (1991) 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*. J Mol Evol 32:379–395

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

Narang F, McIntosh L, Sommerville C (1984) Nucleotide sequence of the ribulose bisphosphate carboxylase gene from *Rhodospirillium rubrum*. Mol Gen Genet 193:220–224

Nei M (1991) Relative efficiencies of different tree-making methods for molecular data. In: Miymoto MM, Cracraft JL (eds) Recent advances in phylogenetic studies of DNA sequences. Oxford University Press, Oxford

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous substitutions. Mol Biol Evol 3:418–426

Nelson K, Whittam TS, Selander RK (1991) Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc Natl Acad Sci USA 88:6667–6671

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organisation deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Olson JM, Pierson BK (1987) Evolution of reaction centers in photosynthetic prokaryotes. Int Rev Cytol 108:209–248

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

Palmer J, Baldauf S, Calie P, de Pamphilis C (1989) Chloroplast gene instability and transfer to the nucleus. In: Clegg M, O'Brein S (eds) Molecular evolution. Alan R Liss, New York, pp 97–106

Penny D, Hendy M (1986) Estimating the reliability of evolutionary trees. Mol Biol Evol 3:403–417

Perasso R, Baroin A, Qu L-H, Bachellerie J-P, Adoutte A (1989) Origin of the algae. Nature 339:142–144

Peschek GA (1987) Respiratory electron transport. In: Fay P, van Baalen C (eds) The cyanobacteria. Elsevier, Amsterdam

Pühler G, Leffers H, Gropp F, Palm P, Klenk H-P, Lottspeich F, Garrett R, Zillig W (1989) Archaebacterial DNA dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc Natl Acad Sci USA 86:4569–4573

Raven P (1970) A multiple origin for plastids and mitochondria. Science 169:641–646

Saitou N, Imanishi T (1989) 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

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

Schimper A (1883) Über die Entwicklung der Chlorophyllkörner und Farbkörper. Bot Ztg 41:105–114

Schwartz R, Dayhoff M (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 for the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase gene. Gene 20:91–102

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng B-Y, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organisation and expression. EMBO J 5: 2043–2049

Siemeister G, Hachtel W (1990) Structure and expression of a gene encoding the large subunit of ribulose-1,5-bisphosphate carboxylase (*rbc*L) in the colourless euglenoid flagellate *Astasia longa*. Plant Mol Biol 14:825–833

Sitte P (1990) Phylogenetische Aspekte der Zellevolution. Biol Rundsch 28:1–18

Smeekens S, Weisbeek P, Robinson C (1990) Protein transport into and within chloroplasts. Trends Biochem Sci 15:73–76

Sogin M, Gunderson J, Elwood H, Alonso R, Peattie D (1989) Phylogenetic meaning of the kingdom concept: an unusual Ribosomal RNA from *Girdia lamblia*. Science 243:75–77

Solomon JA, Walne PL, Dawson NS, Willey RL (1991) Structural characterisation of *Eutreptia* (Euglenophyta) II. The flagellar root system and putative vestigal cytopharnyx. Phycologia 30:402–414

Somerville CC, Jouannic S, Loiseaux-de Goër S (1992) Sequence, proposed secondary structure and phylogenetic analysis of the chloroplast 5S rRNA gene of the brown alga *Pylaiella littoralis* (L.) Kjellm. J Mol Evol 34:246–253

Sourdis J, Krimbas C (1987) Accuracy of phylogenetic trees estimated from DNA sequence data. Mol Biol Evol 4:159–166

Stackebrandt E, Murray RGE, Trüper HG (1988) *Proteobacteria* classis nov., a name for the phylogenetic taxon that includes the "purple bacteria and their relatives." Int J Syst Bacteriol 38:321–325

Stein JL, Haygood M, Felbeck H (1990) 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

Subramanian AP, Stahl D, Prombona A (1991) Ribosomal proteins, ribosomes, and translation in plastids. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids. Academic Press, New York, pp 191–215

Swofford DL (1989) PAUP: phylogenetic analysis using parsimony, Version 3. Illinois Natural History Survey, Champaign, IL

Swofford DL, Olsen GJ (1990) Phylogeny reconstruction. In:

Hillis DM, Moritz C (eds) Molecular systematics. Sinauer, Sunderland, MA

Tajima F, Nei M (1984) Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1:269–285

Tomas R, Cox E (1973) Observations on the symbiosis of *Peridinium balticum* and its intracellular alga. J Phycol 9:273–289

Turner S, Burger-Wiersma T, Giovannini S, Murr L, Pace N (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 in the plastome of a red aga. Curr Genet 16:203–209

Valentin K, Zetsche K (1990a) Rubisco genes indicate a close phylogenetic relationship between the plastids of Chromophyta and Rhodophyta. Plant Mol Biol 15:575–584

Valentin K, Zetsche K (1990b) Structure of the Rubisco operon from the unicellular red alga *Cyanidium caldarium:* Evidence for a polyphyletic origin of plastids. Mol Gen Genet 222:425–430

Valentin K, Zetsche K (1990c) Nucleotide sequence for the large subunit of Rubisco from *Cyanophora paradoxa*. Phylogenetic implications. Curr Genet 18:199–202

Van den Eynde H, De Baere R, De Roeck E, Van de Peer Y, Vandenberghe A, Willekens P, De Wachter R (1988) The 5S ribosomal RNA Sequences of a red algal chloroplast and a gymnosperm chloroplast. Implications for the evolution of plastids and cyanobacteria. J Mol Evol 27:126–132

van Gorkom HJ (1987) Evolution of photosynthesis. In: Amesz J (ed) Photosynthesis. Elsevier, Amsterdam, pp 343–350

Viale AM, Kobayashi H, 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 117: 2391–2400

Walne PL, Kivic PA (1990) Phylum Euglenida. In: Margulis et al. (eds) Handbook of Protoctista. Jones and Bartlett, Boston

Wanatabe MM, Tsutomu S, Inouye I, Suda S, Sawaguchi T, Chihara M (1987) A green dinoflagellate with chlorophylls *a* and *b:* Morphology, fine structure of the chloroplast and chlorophyll composition. J Phycol 23:382–389

Wanatabe MM, Suda S, Inouye I, Sawaguchi T, Chihara M (1990) *Lepidodinium viridae* gen. et sp. nov. (Gymnodiniales, Dinophyta): A green dinoflagellate with a chlorophyll *a* and *b*-containing endosymbiont. J Phycol 26:741–751

Whatley J (1981) Chloroplast evolution: Ancient and modern. Annals NY Acad Sci 361:154–164

Whatley J (1989) Chromophyte chloroplasts: a polyphyletic origin? In: Green J et al. (eds) The chromophyte algae: problems and perspectives. Clarendon, Oxford

Wilcox LW, Wedemayer GJ (1985) Dinoflagellates with blue-green chloroplasts derived from an endosymbiotic eukaryote. Science 227:192–194

Witt D, Stackebrandt E (1988) Disproving the hypothesis of a common ancestry for the *Ochromonas danica* chrysoplast and *Heliobacterium chlorum*. Arch Microbiol 150:244–248

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

Wolfe KH, Li W-H, Sharp P (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84:9054–9058

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

Wolters J, Erdmann V, Stackebrandt E (1989) Current status of the molecular phylogeny of plastids. In: Nardon P, Gianinazzi-Pearson V, Grenier AM, Margulis L, Smith DC (eds) Endocytobiology IV. INRA, Lyon, pp 545–552

Yang D, Oyaizu Y, Oyaizu H, Olsen GJ, Woese CR (1985) Mitochondrial origins. Proc Natl Acad Sci USA 82:4443–4447

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 the β′ subunit of RNA polymerase. Plant Mol Biol 10:245–250

Zavarin GA, Stackebrandt E, Murray RGE (1991) A correlation of phylogenetic diversity in the *Proteobacteria* with influences of ecological forces. Can J Microbiol 37:1–6

Zhou D-X, Mache R (1989) Presence in the stroma of chloroplasts of a large pool of a ribosomal protein not structurally related to any *E. coli* ribosomal protein. Mol Gen Genet 219: 204–208

Received February 27, 1992/Revised April 27, 1992