# Phylogenetic Analyses of the *rbc*L Sequences from Haptophytes and Heterokont Algae Suggest Their Chloroplasts are Unrelated

Niels Daugbjerg\* and Robert A. Andersen†

\*Department of Phycology, University of Copenhagen, Denmark; and †Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine

Using the large subunit of RuBisCo (rbcL) sequences from cyanobacteria, proteobacteria, and diverse groups of algae and green plants, we evaluated the plastid relationship between haptophytes and heterokont algae. The rbcL sequences were determined from three taxa of heterokont algae (Bumilleriopsis filiformis, Pelagomonas calceolata, and Pseudopedinella elastica) and added to 25 published sequences to obtain a data set comprising 1,434 unambiguously aligned sites (~98% of the total rbcL gene). Higher levels of mutational saturation in third codon positions were observed by plotting the pairwise substitutions with and without corrections for multiple substitutions at the same site for first and second codon positions only and for third positions only. In accordance with this finding phylogeny reconstructions were completed by omitting third codon positions, thus using 956 bp in weightedparsimony and maximum-likelihood analyses. The midpoint-rooted phylogenies showed two major clusters, one containing cyanobacteria, glaucocystophytes, a phototrophic euglenoid, chlorophytes, and embryophytes (the green lineage), the other containing proteobacteria, haptophytes, red algae, a cryptophyte, and heterokont algae (the nongreen lineage). In the nongreen lineage, the haptophytes formed a sister group to the clade containing heterokont algae, red algae, and the cryptophyte Guillardia theta. This branching pattern was well supported in terms of bootstrap values in weighted-parsimony and maximum-likelihood analyses (100% and 92%, respectively). However, the phylogenetic relationship among red algae, heterokonts, and a cryptophyte taxon was not especially well resolved. A four-cluster analysis was performed to further explore the statistical significance of the relationship between proteobacteria, red algae (including and excluding Guillardia theta), haptophytes, and heterokont algae. This test strongly favored the hypothesis that the heterokonts and red algae are more closely related to each other than either is to proteobacteria or haptophytes. Hence, this molecular study based on a plastid-encoded gene provides additional evidence for a distant relationship between haptophytes and the heterokont algae. It suggests an evolutionary scenario in which the ancestor of the haptophyte lineage engulfed a phototrophic eukaryote and, more recently, the heterokont lineage became phototrophic by engulfing a red alga.

#### Introduction

Members of the algal class Haptophyceae were originally classified in the Chrysophyceae (Pascher 1913, pp. 43, 48–49; Bourrelly 1957, pp. 232–234), but as ultrastructural data accumulated, they were placed in their own class (Christensen 1962, pp. 72–74). Since receiving their independent status, additional ultrastructural data (Hibberd 1976) and molecular data (e.g., Bhattacharya et al. 1992; Leipe et al. 1994) have brought into question any evolutionary relationship between the Haptophyceae and the Chrysophyceae. The Chrysophyceae have a well-supported evolutionary relationship with other heterokont algae (=chromophyte algae), but the closest relative for the Haptophyceae remains unresolved. The morphological and biochemical characters that are unique to the Haptophyceae or are shared with other protistan groups are described elsewhere (e.g., Hibberd 1976; Andersen 1991; Leipe et al. 1994; Saunders et al. 1995), but despite this knowledge, no consensus classification or phylogenetic hypothesis for the Haptophyceae has been reached. Because many of the shared characters (synapomorphies) are either chloroplast features or associated with chloroplasts, we compared gene sequences encoding the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rbcL)

Key words: chromophytes, haptophytes, heterokont algae, phylogeny, plastid evolution, *rbcL*, red algae, RuBisCo.

Address for correspondence and reprints: Niels Daugbjerg, Department of Phycology, University of Copenhagen, Øster Farimagsgade 2D, DK-1353 Copenhagen K, Denmark. E-mail: nielsd@bot.ku.dk.

Mol. Biol. Evol. 14(12):1242–1251. 1997 © 1997 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038 from various algal lineages in an attempt to resolve the phylogeny of the Haptophyceae and, in a broader sense, to resolve the relationships of algal groups.

As already shown (e.g., Morden et al. 1992; Ueda and Shibuya 1993; McFadden, Gilson, and Waller 1995), rbcL gene comparisons suggest a biphyletic oxigin of phototrophic eukaryotes. The rbcL gene of the green algae/plant lineage, glaucocystophytes, and phototrophic euglenoids is derived from cyanobacteria and form one rbcL lineage; the second rbcL gene lineage consists of the nongreen algae and is derived from proteobacteria. This scenario for the evolution of the *rbc*L gene is well supported in terms of bootstrap values but is opposed by molecular phylogenies based on other chloroplast-encoded genes like psbA, tufA (Morden et al. 1992; Delwiche, Kuhsel, and Palmer 1995), atpB (Douglas and Murphy 1994), GAPDH (Martin et al. 1993), ClpC (Clarke and Eriksson 1996), and SSU rDNA (Bhattacharya and Medlin 1995; Helmchen, Bhattacharya, and Melkonian 1995). Phylogenies based on these genes suggest that there was a single cyangbacterial ancestor of plastids. A number of hypotheses have been suggested to explain the apparently contradictory results. For example, (1) a lateral gene transfer of the rbcLS genes may have occurred from a protesbacteria into the ancestor that gave rise to the nongreen plants; (2) a lateral transfer of the rbcLS operon may have occurred into the cyanobacterial ancestor that gave rise to the nongreen plants; or (3) two rbcLS operons may have been present in a cyanobacterial-like ancestor (that gave rise to plastids) and different copies were re-

Table 1 Oligonucleotide Primer Sequences Used to Amplify and Sequence Heterokont Algae

Primer Code for Forward Primers	Primer Sequence 5'-3'	Primer Code for Reverse Primers	Primer Sequence 5'-3'			
DPrbcL1 (12-6)	AAGGAGGAADHHATGTCT	DPrbcL7 (23–3; rbcS)	AAASHDCCTTGTGTWAGTYTC			
ND <i>rbc</i> L2 (34–53)	AAAAGTGACCGTTATGAATC	NDrbcL8 (1232–1212)	CCAATAGTACCACCACCAAAT			
NDrbcL3 (43–58)	CGTTACGAATCTGGTG	NDrbcL9 (1226–1212)	GTACCACCACCAAAT			
NDrbcL4 (342–356)	AGGTTCACTAGCTAA	NDrbcL10 (983–969)	TGGTCAACACCAGCC			
NDrbcL5 (635-650)	CACAACCATTCATGCG	NDrbcL11 (835–820)	CAGTGTAACCAATTAC			
NDrbcL6 (953–967)	GTAAATGGATGCGTA	NDrbcL12 (527–514)	GCACCTAATAGTGG			

NOTE.—Numbers in parentheses refer to the position of the rbcL gene of the brown alga Pilayella littoralis. Abbreviations (IUPAC code): S (C/G): H (A/T). C); D (A/T/G); W (A/T); Y (C/T).

tained in the green versus nongreen lineages (Palmer 1993). For more details on the origin of plastids as revealed by the rbcL gene, see Assali, Mache, and Loiseaux-de Goër (1990); Assali et al. (1991); Morden et al. (1992); Loiseaux-de Goër (1994).

The haptophytes and heterokont algae both have chloroplasts surrounded by two membranes of endoplasmic reticulum which are continuous with the outer membrane of the nucleus; the plastids have lamellae composed of three appressed thylakoids, and the chloroplast pigments are typically chlorophyll a and c as well as carotenoids (e.g., fucoxanthin, diatoxanthin, diadinoxanthin, etc.) (see Hibberd 1976; Bjørnland and Liaaen-Jensen 1989; Jeffrey 1989; Andersen 1991). Despite these similarities, SSU rDNA sequence comparisons suggest that the haptophytes are distantly related to the heterokont algae (e.g., Bhattacharya et al. 1992; Leipe et al. 1994). After considering the unresolved phylogenetic position of the Haptophyceae in nuclear gene trees, the possibilities for lateral gene transfer of rbcL genes, and the possibilities for additional endosymbiotic origins of nongreen algal plastids, we proposed two hypotheses—hypothesis I: the plastid similarities of haptophytes and heterokont algae resulted from a common ancestral plastid that gave rise to two distinct lineages (single secondary endosymbiotic event); and hypothesis II: the plastid similarities of haptophytes and heterokont algae resulted from convergent evolution of the chloroplast features (two secondary endosymbiotic events). If hypothesis I is correct, then the rbcL genes in haptophytes and heterokont algae should be more similar to each other than they are to those of any other algal lineage. If hypothesis II is correct, then chloroplast genes of the haptophyte and heterokont algae should be dissimilar, and each should be more closely related to another lineage. To test these hypotheses, we analyzed the rbcL gene sequences using weighted-parsimony and maximum-likelihood analyses, and we then tested the results of four monophyletic assemblages using a fourcluster analysis program PHYLTEST (Kumar 1995).

#### Materials and Methods

Cultures

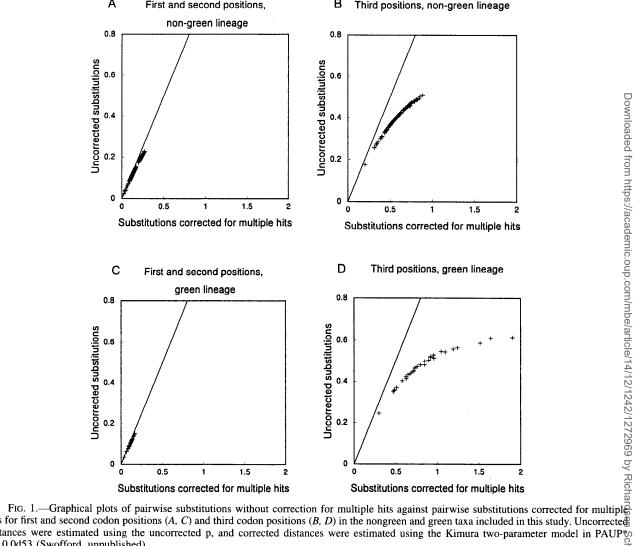
The heterokont algae were obtained from the Provasoli-Guillard National Center for Culture of Marine Phytoplankton (Pelagomonas calceolata CCMP 1214 and Pseudopedinella elastica CCMP 716) and from the Culture Collection of Algae at the University of Texas (Bumilleriopsis filiformis UTEX 309).

DNA Extraction, Amplification and Sequencing

Cells were concentrated by centrifugation and in cubated in 500 µl preheated 2 × CTAB (2% hexade cyltrimethylammonium bromide) isolation buffer and 1% β-mercapto-ethanol for 1–2 h at 60°C (Doyle and € Doyle 1987). Total genomic DNA was extracted using 500 µl of 24:1 chloroform: isoamyl alcohol, cleaned with the GENECLEAN II® kit as recommended by the manufacturer (BIO 101 Inc.). Double-stranded DNA was amplified in 100-μl reaction volumes containing 105 × PCR Buffer II (10 mM Tris-HCl, pH 8.3 (at 25°C). 50 mM KCl), 200 μM dNTP, 0.2 or 1.0 μM of each primer, 2.5 U of Ampli Taq® DNA polymerase, and 2.5 mM MgCl<sub>2</sub>. Amplification conditions were one initial cycle of denaturation at 94°C for 3 min, followed by 30°C cycles each consisting of denaturation at 94°C for 1 ming annealing at 50°C for 1 min, and extension at 72°C for 2 min. PCR products were visualized in a 0.8% agarose gel containing 0.67 μg/ml ethidium bromide in a TAE buffer (40 mM Tris-acetate, 1 mM EDTA). The resulting bands were transferred to Eppendorf tubes and melted? in 700-900 µl NaI. From this step on precipitations were done with the GENECLEAN II® kit. Nucleotide sequences of PCR products were determined using the AmpliCycle<sup>™</sup> (Perkin Elmer) sequencing kit following the recommendations of the manufacturer. Sequencing primers were biotinylated at the 5' end. The sequence reactions were run in a 6% Long Ranger gel and transferred to an Immobilon-S membrane. The band patterns were detected using the NEBlot® Phototope® kit (New\_ England Biolabs) as recommended by the manufacturer and finally visualized by exposure on X-ray film. Oli gonucleotide primers used to amplify and determine the sequence of the rbcL gene in the heterokont algae are shown in table 1.

#### Nucleotide Sequences Used

rbcL sequences from cyanobacteria, proteobacteria and diverse algal and embryophyte lineages were aligned manually and edited using ESEE V1.09d (Cabo) and Beckenbach 1989) and Compare Sequences V3.0 (Siegismund, unpublished). The rbcL sequences used for phylogenetic inference are listed below (with GenBank accession numbers): Alcaligenes eutrophus (M17744), Rhodobacter sphaeroides (M64624), Xan-



hits for first and second codon positions (A, C) and third codon positions (B, D) in the nongreen and green taxa included in this study. Uncorrecte distances were estimated using the uncorrected p, and corrected distances were estimated using the Kimura two-parameter model in PAUP to V4.0.0d53 (Swofford, unpublished).

thobacter flavus (X17252), Anabaena sp. (L02520), Synechococcus sp. (=Anacystis nidulans, D13539), Cyanophora paradoxa (X53045), Antithamnion sp. (X54532), Gelidium floridanum (U00106), Porphyridium aerugineum (X17597), Porphyra purpurea (U38804), Delisea pulchra (U26812), Guillardia theta (formerly known as Cryptomonas  $\Phi$ , X62349), Pleurochrysis carterae (D11140), Chrysochromulina hirta (D45846), Calyptrosphaera sphaeroidea (D45842), Umbilicosphaera sibogae (D45843), Emiliania huxleyi (D45845), Pilayella littoralis (X55372), Odontella sinensis (Z67753), Chlorella ellipsoidea (D10997), Chlamydomonas moewusii (M15842), Chlamydomonas reinhardtii (J01399), Euglena gracilis (M12109), Marchantia polymorpha (X04465), and Zea mays (X86563). rbcL sequences from the three heterokont algae determined in this study have been deposited in GenBank and given the following accession numbers: Pelagomonas calceolata, U89898; Pseudopedinella elastica, U89899; and Bumilleriopsis filiformis, U89900.

#### Phylogenetic Analysis

We examined mutational saturation of third codon positions in the two rbcL lineages by plotting all pairwise substitutions uncorrected for multiple substitutions against those corrected for multiple substitutions ("un corrected p" and Kimura two-parameter model, PAUP V4.0.0.d53; Swofford, unpublished). We calculated these for third positions only and for first and second positions only (fig. 1). Because substitution rates were approximately twice as frequent for third-position nue cleotides (see Results) and because of strong evidence for mutational saturation at this position, we excluded this codon position in weighted-parsimony analyses Parsimony was performed with PAUP V3.1.1 (Swofford) 1993) using the heuristic search option with random ad ⊆ dition of sequences (100 replicates; see Maddison 1991) and a branch-swapping algorithm (tree bisection-recons nection, TBR). Nucleotide sites were weighted (rescaled) consistency index over an interval of 1–1,000) and ther used as input for a bootstrap analysis with 100 replica tions. Bootstrap analyses were conducted in order to find the relative support for the branching pattern (Felsenstein 1985). Maximum-likelihood analyses based on first and second codon positions only (956 base pairs) were performed with fastDNAml V1.1.1 (Olsen et al. 1994)

Table 2 Base Composition of Thymine (T), Cytosine (C), Adenine (A), and Guanine (G) in Percentage of rbcL Sequences at All, First, Second, and Third Codon Positions, Respectively

	ALL	Codor	Posit	IONS	First	Codo	n Posi	TIONS	SECON	D COD	on Pos	SITIONS	THIRD	Codo	v Posr	TIONS
Taxon	T	С	A	G	T	С	A	G	T	С	A	G	T	С	Α	G
Alcaligenes eutrophus	15.5	33.5	17.9	33.1	15.0	23.8	22.6	38.6	27.6	24.3	28.1	20.0	4.0	52.3	3.2	40.5
Rhodobacter sphaeroides	15.9	32.6	18.1	33.4	14.6	23.0	23.8	38.6	28.1	23.4	28.9	19.6	5.1	51.3	1.7	42.0
Xanthobacter flavus	15.6	35.7	18.0	30.7	14.4	22.8	22.2	40.6	27.3	23.3	29.6	19.9	5.3	60.9	2.3	31.≸
Calyptrosphaera sphaeroidea	28.3	21.8	27.5	22.4	15.3	22.4	24.8	37.5	28.8	24.6	29.4	17.2	40.7	18.3	28.3	12.6
Chrysochromulina hirta	29.8	20.6	27.5	22.2	17.2	21.1	24.4	37.3	28.8	23.8	29.6	17.9	43.4	17.0	28.3	11.3
Emiliania huxleyi	29.9	20.8	26.7	22.6	16.1	21.8	24.2	37.9	29.0	23.8	29.2	18.1	44.4	17.0	26.8	11.8
Pleurochrysis carterae	28.5	20.9	27.3	23.4	16.9	21.7	24.5	36.9	28.5	24.1	29.3	18.1	40.1	16.9	28.1	15.Œ
Umbilicosphaera sibogae	29.6	20.7	27.5	22.2	15.9	22.2	24.4	37.5	28.5	24.6	29.4	17.4	44.4	15.3	28.5	11.8
Antithamnion sp	30.7	16.7	32.2	20.4	19.4	15.6	27.6	37.3	29.8	23.6	29.5	17.1	42.8	11.0	39.5	6.8
Gelidium floridanum	29.6	16.6	30.9	21.2	19.2	16.5	26.8	37.6	29.8	22.8	30.4	17.1	44.9	10.6	35.4	9.
Delisea pulchra	32.6	15.8	31.1	20.5	20.8	15.0	26.6	37.6	29.8	23.6	29.6	17.0	47.3	8.6	37.2	6.9
Porphyra purpurea	31.2	17.0	30.1	21.8	18.1	15.6	27.9	38.4	29.5	23.6	29.3	17.5	45.8	11.6	33.1	9.5
Porphyridium aerugineum	31.0	18.8	29.4	20.8	18.8	18.6	25.5	37.1	28.9	23.4	29.5	18.1	45.4	14.4	33.1	7.2
Guillardia theta		18.7	30.5	21.7	18.1	16.9	26.0	39.0	28.7	22.4	30.0	19.0	40.7	16.9	35.4	7.₫
Bumilleriopsis filiformis	30.4	18.6	29.6	21.4	21.7	15.7	24.9	37.8	28.6	22.7	29.3	19.4	41.0	17.4	34.8	6.9
Odontella sinensis	31.6	18.4	29.1	20.9	23.1	14.9	24.0	38.0	28.2	24.6	29.2	18.1	43.5	15.6	34.2	6.₹
Pelagomonas calceolata	30.7	19.9	27.0	22.4	23.3	13.8	25.6	37.3	28.9	24.4	28.7	18.0	39.9	21.5	26.8	11.8
Pseudopedinella elastica	29.9	20.1	27.7	22.3	19.1	18.8	24.5	37.7	28.1	25.1	28.1	18.6	42.6	16.4	30.5	10.€
Pilayella littoralis	32.3	16.4	29.3	22.0	21.5	14.6	24.9	39.0	28.9	22.8	28.9	19.4	46.4	11.8	34.2	7.6
Anabaena sp	25.1	25.1	26.0	23.9	18.3	20.7	22.6	38.4	26.5	23.3	31.3	19.0	30.4	31.3	24.1	14.2
Synechococcus sp	21.7	29.0	22.6	26.6	16.6	22.7	22.9	37.8	26.6	23.3	30.9	19.2	22.0	41.0	14.0	22.9
Cyanophora paradoxa	31.0	19.0	28.7	21.4	21.8	16.6	24.4	37.3	25.9	24.4	29.3	20.5	45.3	16.0	32.3	6.5
Chlamydomonas reinhardtii	30.0	21.0	26.8	22.2	18.5	20.5	21.3	39.7	26.7	22.2	29.3	21.8	44.8	20.3	29.7	5.2€
Chlamydomonas moewusii	30.0	19.0	28.7	22.3	20.9	18.1	21.8	39.2	26.1	22.6	29.7	21.6	42.9	16.2	34.7	6.3
Chlorella ellipsoidea	31.2	18.4	28.2	22.3	20.3	19.2	22.0	38.6	25.4	23.7	29.7	21.1	47.8	12.3	32.8	7. 🗓
Euglena gracilis	34.1	15.9	26.7	23.4	22.0	17.5	22.4	38.2	25.0	23.3	29.3	22.4	55.2	6.9	28.2	9.7
Marchantia polymorpha	33.4	15.5	29.2	21.9	21.1	16.8	24.1	37.9	26.3	22.4	30.6	20.7	52.8	7.1	33.0	7.∦
Zea mays	28.5	19.3	27.7	24.5	20.5	17.7	23.7	38.2	26.1	23.3	30.0	20.7	39.0	16.8	29.5	14.7

and PHYLIP V3.572c (Felsenstein 1993). In order to find the tree with the optimal maximum-likelihood score, searches were repeated by varying the transition to transversion (Ts: Tv) ratio.

### Four-Cluster Analysis

The four-cluster analysis program PHYLTEST V2.0 (Kumar 1995; Kumar and Rzhetsky 1996) was used to evaluate the statistical confidence of the three unrooted phylogenetic hypotheses comparing four monophyletic groups (proteobacteria, red algae [with or without the cryptophyte Guillardia theta], haptophytes, and heterokont algae). Based on the nucleotide sequence data, the topology showing the smallest sum of branch lengths is selected to represent the evolutionary relationships (the minimum-evolution principle; Rzhetsky and Nei 1992, 1993; Rzhetsky et al. 1995).

#### Results

Table 2 shows the nucleotide distribution (in percentage) of thymine, cytosine, adenine, and guanine for all codon positions combined and for first, second, and third codon positions individually. The proteobacteria possess a higher level of guanine and cytosine especially in the third codon positions. The base composition in cyanobacteria is more equally distributed when all positions are combined, but individual variations occur at specific positions. For eukaryotes, the frequency of bases for all positions combined shows a higher level of thymine and adenine, and this is due to extremely high levels of these two bases in the third codon positions.  $\frac{1}{2}$ 

#### Mutational Saturation of Third Codon Positions

Of the total number of substitutions (1,434 nucle otide positions), a little more than half (56.1%) occurred in third codon positions. The frequency of substitutions in second codon positions was lower (17.7%) than that in first codon positions (26.3%). To explore the extent of saturation of the rapidly evolving third codon positions, we plotted nucleotide substitutions uncorrected for multiple substitutions against the differences, with corrections for multiple substitutions (fig. 1). The deflection from linearity when including only third codon positions (fig. 1B and D) suggests that multiple substitutions at this site are increasing more rapidly, probably for dis tantly related taxa. The almost linear relationship be tween uncorrected and corrected substitutions when distance estimates are based on first and second positions only indicates that these positions have not yet reached mutational saturation (fig. 1A and C).

#### Plastid Phylogeny

The inferred phylogenies resulting from weighted parsimony and maximum-likelihood analyses of rbcE sequences are illustrated in figures 2 and 3. Both reconstructions show the green lineage and the nongreen lineage as monophyletic groups. Within the nongreen lineage, the haptophytes have a sister group relationship, highly supported by bootstrap values, to the crypto-

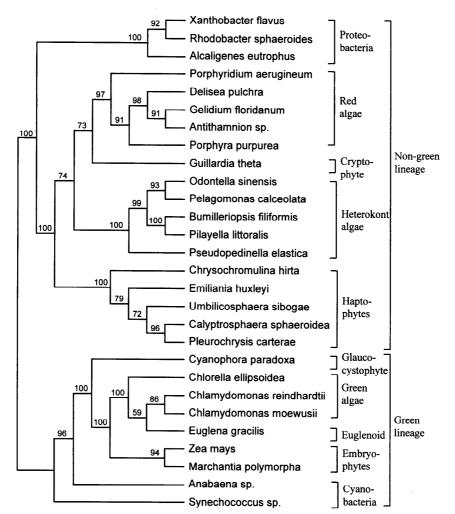


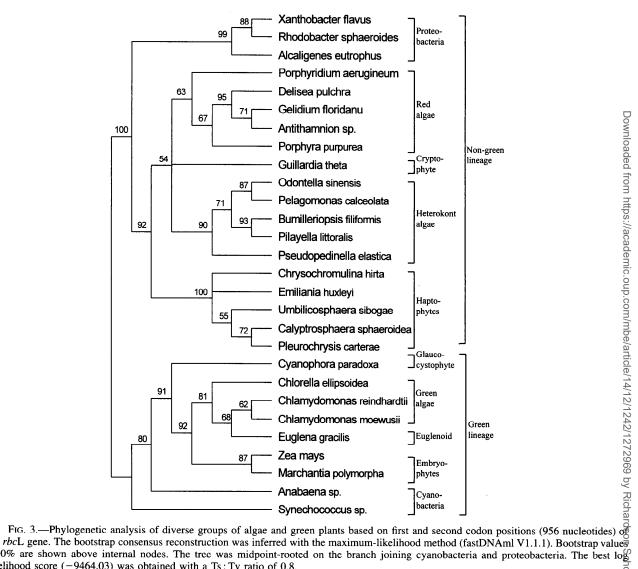
Fig. 2.—Phylogenetic analysis of diverse groups of algae and green plants based on first and second codon positions (956 nucleotides) of the *rbcL* gene. The bootstrap consensus reconstruction was inferred with a weighted-parsimony (rescaled consistency index over the interval 1,000) method. The heuristic search option in PAUP V3.1.1 was used. Bootstrap values ≥50% are shown above internal nodes. The tree was midpoint-rooted on the branch joining cyanobacteria and proteobacteria. The single most parsimonious tree using the rescaled consistency index had a consistency index of 0.74 and a retention index of 0.9.

phyte/red algal/heterokont clade (figs. 2 and 3), indicating a distant relationship between the rbcL gene of the haptophytes and heterokonts. In weighted-parsimony analyses, the haptophytes, red, and heterokont algae all form monophyletic groups which are well supported by bootstrap values (100%, 97%, and 100%, respectively). However, in the maximum-likelihood bootstrap analysis, the monophyletic status of the red algae is not well supported and the cryptophyte Guillardia theta takes an unresolved position (compare figs. 2 and 3). In the parsimony analysis, the cryptophyte is related to the red algae but this relationship is only moderately supported by bootstrap values (73%). The relationship between the brown algae Pilayella littoralis and the xanthophyte Bumilleriopsis filiformis is highly supported in both analyses, and these two taxa form a well-supported sister group to the clade containing the marine centric diatom Odontella sinensis and the pelagophyte Pelagomonas calceolata. The dictyochophyte Pseudopedinella elastica takes the most divergent position within the heterokont algae. The phylogeny between the five haptophyte taxa included is fairly well resolved in terms of bootstrap values, but the relationship does not reflect the classification into the orders recognized by Parke and Dixon (1976; see also Fujiwara et al. 1994).

In the green lineage (here including the glaucocystophyte Cyanophora paradoxa, which possesses cyanelles instead of plastids) the topology among the major groups is generally well supported by high bootstrapy values. The branching patterns obtained by the two physlogeny algorithms applied were identical. Cyanophoral forms a sister group to the green lineage, and the euglenoid Euglena gracilis is a sister taxon to two species of Chlamydomonas and is thus positioned within the chlorophyll-a+b-containing organisms.

#### Testing Alternative Tree Topologies

Using the four-cluster analysis program PHYL $^{\circ}_{\circ}$  TEST, we tested three phylogenetic scenarios (fig. 4) and found that the red algae and heterokont algae were more closely related to each other (with respect to the *rbcL* gene) than either was to haptophytes or to proteobacteria (confidence probability [CP] > 96%). When constraining the red algae and the cryptophyte as a monophyletic



the rbcL gene. The bootstrap consensus reconstruction was inferred with the maximum-likelihood method (fastDNAml V1.1.1). Bootstrap value ≥50% are shown above internal nodes. The tree was midpoint-rooted on the branch joining cyanobacteria and proteobacteria. The best log likelihood score (-9464.03) was obtained with a Ts: Tv ratio of 0.8.

assemblage, this cluster was also more closely related to the heterokont algae than either was to the other groups (table 3).

## Discussion

Phylogeny

The results of the weighted-parsimony and maximum-likelihood analyses show that the rbcL gene of the heterokont algae is more closely related to that of the red algae and cryptophytes than to that of the haptophytes, suggesting that the haptophytes are distantly related to the heterokont algae (figs. 2 and 3). The close relationship of the red algae and cryptophytes to the heterokont algae is also supported by the four-cluster test (table 3). The unresolved position of the cryptophyte Guillardia theta in the maximum-likelihood analysis and the moderate bootstrap support for its position in the parsimony analysis may be due to poor taxon sampling, and if more rbcL gene sequences from cryptophytes become available, then branch support may improve. For example, the addition of cryptophyte taxa to

nuclear-encoded SSU rDNA gene sequence analyses has shown that the cryptophytes form a monophyletic group. within the red algae, not as a sister group (Cavalier Smith et al. 1996).

In other studies using the rbcL gene, Chesnick Morden, and Schmieg (1996) found that the haptophytes and heterokont algae were sister taxa (see their fig. 3) however, this relationship is poorly supported in terms of bootstrap values (=53%). Also, in a study which  $di\bar{d}$ not include haptophytes, Delwiche and Palmer (1996) reported a close evolutionary relationship between the rbcL genes of red algae/cryptophytes and those of het<sup>∞</sup> erokont algae. Thus, our study shows for the first time a distant relationship between the rbcL genes of haps tophytes and heterokonts for which branches are well resolved and branch support is strong.

Results from nuclear-encoded SSU rDNA gene sequence analyses also show no close relationship between the haptophytes and heterokont algae (e.g., Bhattacharya et al. 1992; Leipe et al. 1994; Van de Peer et al. 1996), but because of weak branch support following

Α

Haptophytes

red algae (5 taxa), and heterokont algae (5 taxa).

bootstrap values, the exact relationship of the haptophytes is not known. Results using the SSU rDNA gene sequences from the chloroplast genome suggest that the chloroplast of heterokonts is similar to those of Cyanidium and Galdieria, while the haptophytes form a clade between these algae and the typical red algae or cryptophytes (bootstrap support is weak; Bhattacharva and Medlin 1995; Medlin et al. 1995).

Presumably, organisms like the haptophytes, ciliates, dinoflagellates, and heterokonts (=Stramenopiles), which have mitochondria with tubular cristae, form a monophyletic group (Taylor 1976; Stewart and Mattox 1980). Interestingly, the haptophytes have a periplastidal endoplasmic reticulum which lies just below the plasma membrane (Hibberd 1976), and this bears a resemblance to the alveoli of ciliates (Lynn and Small 1990), the amphisiesmal vesicles of dinoflagellates (Morrill and Loeblich 1983), and the inner membrane complex of apicomplexans (Vivier and Desportes 1990). These ultrastructural features are not found in the heterokonts, thus adding further support for a distant relationship between haptophytes and heterokonts. It may be speculated that the haptophytes have an evolutionary relationship with the alveolates.

Based on our results, we accept hypothesis II, i.e., that the plastids of the haptophytes and the heterokont algae are not closely related. We conclude that the haptophytes and the heterokont algae obtained chloroplasts independently by two endosymbiotic events and that similarities in their plastids have arisen due to convergent evolution. Specifically, our data do not support the hypothesis that heterokonts obtained their chloroplasts by engulfing a haptophyte alga. Accepting hypothesis II is not without complications. The fact that the rbcL gene

sequence analyses place the red algae and cryptophytes between the haptophytes and beta-live to same rbcL lineage implies that the chloroplast ultra-2 structure and pigmentation arose independently, i.e., from different symbionts which probably had phycobilipigments, rather than carotenoids, as light-harvesting ac cessory pigments. If the haptophytes obtained a plastid by engulfing a red alga, then the red algae should di verge earlier in the tree than the haptophytes; this was not observed. However, it is possible that a colorless haptophyte engulfed a red alga very early in the evo lution of red algae, and due to substitution rate differences in the rbcL gene, the haptophyte lineage may branch earlier in the phylogenetic tree (see Leitner et al. 1996). Similarly, if the heterokont plastid either descended directly from a haptophyte or was obtained by engulfing a haptophyte, then the red algae and cryptophyte rbcL genes should not diverge between them. Ac \(\frac{1}{2}\) cepting hypothesis II also implies that the replacement of the (ancestral) phycobilipigment-type light-harvesting complex with a carotenoid-type light-harvesting complex has occurred twice independently. The reason(s) for a change in the photosynthetic apparatus are unknown but one might postulate that (1) there is a selective ad $_{\perp}$ vantage for carotenoid light-harvesting or (2) a deletion removed the phycobisome/phycobiliprotein complex and there was strong selection for the development of an alternative light-harvesting complex. There appears to be no evidence of phycobilipigment genes in the chlo roplast genome of the heterokont algae Odontella and Pilayella (Loiseaux-de Goër 1994; Kowallik et al 1995), and the heterokont chloroplast genome sizes of Synura (91.5 kb) and Chrysodidymus (102 kb) are some of the smallest known (Graham, Graham, and Wujek

Table 3 Results of Four-Cluster Analyses for Determining the Most Likely Topology Between Proteobacteria (P), Red Algae (R), a Cryptophyte (C), Haptophytes (HP), and Heterokont Algae (HA)

Best Tree	Statistical Confidence		202
[(R, HA), (P, HP)]	Better than [(P, R), (HA, HP)]	: CP=99.8	<u></u>
	Better than[(R, HP), (P, HA)]	: CP=96.9	
$[(R+C, HA), (P, HP)] \dots \dots$	Better than [(P, R+C), (HA, HP)]	: CP=99.9	
	Better than [(R+C, HP), (P, HA)]	: CP=97.1	

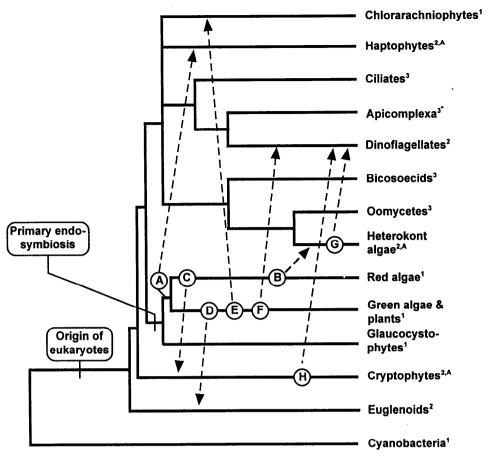


Fig. 5.—Schematic scenario showing the phylogeny of mostly algal groups based on nuclear-encoded small-subunit ribosomal DNA and the probable origin of plastids in the different algal lineages (host relationships redrawn mainly from Bhattacharya and Medlin 1995). The timescale of divergence events is not exact. Numbers in circles refer to a secondary endosymbiotic event (i.e., a heterotrophic eukaryote engulfed a phototrophic eukaryotic alga). The haptophytes probably obtained their chloroplasts by engulfing a taxon similar to the ancestor to the redgreen algal lineage (A), whereas the heterokont algae probably obtained their chloroplasts by engulfing a red algae (B), as did the cryptophytes (C). The plastids in euglenoids (D), chlorarachniophytes (E), and some dinoflagellates (F) most likely originated from the green algal lineage (O) of cryptophytes (H). I autotrophic lineage; a uto-/heterostrophic lineage; host taxa autotrophic; Plasmodium was recently shown to contain a reduced plastid.

1993; Wee, Chesnick, and Cattolico 1993). Furthermore, by sequencing the total chloroplast genome (119,704 bp) of the marine centric diatom Odontella sinensis (Kowallik et al. 1995) and comparing it to the plastid genome of the red algae *Porphyra purpurea* (191,028 bp), it appears that the arrangements of individual genes are almost identical (Stoebe, Freier, and Kowallik 1996). The colinear gene clusters in the two taxa strongly suggest that the diatom chloroplast has been derived by a secondary endosymbiosis involving a red alga. The reasons for change in the photosynthetic apparatus of the haptophytes and the heterokont algae may be different. Molecular studies examining the other plastid genes as well as the complete haptophyte plastid genome may shed light on these topics. Furthermore, the genes for carotenoid synthesis are a specific means for testing the possible independent origin of the carotenoid-type of light-harvesting complex. A carotenoid such as β-carotene is well suited because it also occurs in red algae and in at least some cryptophytes (Rowan 1989, p. 121). Evolution of Plastids in Haptophytes and Heterokonts

The use of chloroplast-encoded genes to infer phy

The use of chloroplast-encoded genes to infer phylogenies for algae is hampered by the likelihood that

plastids have arisen independently in eukaryotic line eages several times as a result of endosymbiotic events i.e., nonphotosynthetic eukaryotes may have engulfed photosynthetic eukaryotes (e.g., Gibbs 1978; Palmer 1993). To date, phylogenies produced using nuclear-encoded genes from eukaryotic algae that include hapto= phytes suffer from either limited taxon sampling or poor resolution (e.g., Bhattacharya et al. 1992; Bhattacharya and Medlin 1995; Saunders et al. 1995). After examin ing these trees, we constructed a generalized tree that includes only well-supported branches (see above references), and we added hypothetical endosymbiotie sources of plastids based on our rbcL analyses (fig. 5) We suggest that the haptophytes (or their ancestor) ob tained a plastid independently and that the source of this plastid was probably either the ancestor to the red alga green algal divergence or a very early ancestor in the lineage which led to the red algae (fig. 5, event A). Like others (e.g., Delwiche and Palmer 1996), we also suggest that a colorless heterokont obtained a plastid by engulfing a red alga (fig. 5, event B). Similarly, the cryptophytes, which appear to be an older lineage, became photosynthetic much more recently by also engulfing a red alga (fig. 5, event *C*) (e.g., Douglas et al. 1991). As has been hypothesized earlier (e.g., Gibbs 1978; Watanabe et al. 1990; McFadden, Gilson, and Waller 1995), those nongreen algae with chlorophyll b (euglenoids, chlorarachniophytes, and at least one dinoflagellate) obtained plastids by engulfing a green alga (fig. 5, events *D*, *E*, and *F*). Finally, the dinoflagellates also appear to have obtained plastids at least twice more by engulfing a heterokont alga (fig. 5, event *G*) (e.g., Tomas and Cox 1973; Chesnick, Morden, and Schmieg 1996) and a pigmented cryptophyte (fig. 5, event *H*) (e.g., Wilcox and Wedemayer 1985; Schnepf and Elbrächter 1988).

#### Codon Bias and Mutational Saturation

It is evident from table 2 that a different codon bias is present when comparing the major groups of organisms included in this study, particularly in third codon positions. Thus, the rbcL gene displays a different codon usage pattern with specific restrictions on the changes of certain bases. Omitting sites with extreme codon biases (e.g., third codon positions in the rbcL gene) is particularly appropriate when the phylogeny estimation is based on parsimony, as this method is more likely to be consistent if character changes are low (Swofford et al. 1996). The method we employed to identify mutational saturation can be implemented by plotting uncorrected substitutions against substitutions corrected for multiple hits using the uncorrected p and, e.g., the Kimura two-parameter model in the PAUP\* (Swofford, unpublished). It represents a simple way to detect mutational saturation of particular sites in protein coding genes.

#### Acknowledgments

A postdoctoral research fellowship from the Carlsberg Foundation to N.D. is acknowledged. This work was also supported by National Science Foundation Grants BRS-94-19498 and EHR-9108766 and Office of Naval Research Grant N00014-92-J-1717 to R.A.A. This is Bigelow Laboratory scientific contribution no. 97004.

#### LITERATURE CITED

- ANDERSEN, R. A. 1991. The cytoskeleton of chromophyte algae. Protoplasma **164**:143–159.
- ASSALI, N., R. MACHE, and S. LOISEAUX-DE GOËR. 1990. Evidence for a composite phylogenetic origin of the plastid genome of the brown alga *Pylaiella littoralis* (L.) Kjellm. Plant Mol. Biol. 15:307–315.
- ASSALI, N., W. F. MARTIN, C. C. SOMERVILLE, and S. LO-ISEAUX-DE GOËR. 1991. 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.
- BHATTACHARYA, D., and L. MEDLIN. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J. Phycol. 31:489–498.
- BHATTACHARYA, D., L. MEDLIN, P. O. WAINRIGHT, E. V. AR-IZTIA, C. BIBEAU, S. K. STICKEL, and M. L. SOGIN. 1992. Algae containing chlorophylls a+c are paraphyletic: molecular evolutionary analysis of the chromophyta. Evolution 46:1801–1817.

- BJØRNLAND, T., and S. LIAAEN-JENSEN. 1989. Distribution patterns of carotenoids in relation to chromophyte phylogeny and systematics. Pp. 37–61 in J. C. GREEN, B. S. C. LEADBEATER, and W. L. DIVER, eds. The chromophyte algae: problems and perspectives. Systematics Association Special Volume, Clarendon Press, Oxford.
- BOURRELLY, P. 1957. Recherches sur les Chrysophycées. Morphologie, Phylogénie, Systématique. Revue Algologique Mémoire Hors-Série no. 1.
- CABOT, E. L., and A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. 5:233-234.
- CAVALIER-SMITH, T., J. A. COUCH, K. E. THORSTEINSEN, P. GILSON, J. A. DEANE, D. R. A. HILL, and G. I. McFADDEN, 1996. Cryptomonad nuclear and nucleomorph 18S rRNA phylogeny. Eur. J. Phycol. 31:315–328.
- CHESNICK, J. M., C. W. MORDEN, and A. M. SCHMIEG. 1996 Identity of the endosymbiont of *Peridinium foliaceum* (Pyrrophyta): analysis of the *rbcLS* operon. J. Phycol. 32:850 857.
- CHRISTENSEN, T. 1962. Alger. Pp. 1–178 in T. W. BÖCHER, M. C. LANGE, and T. SØRENSEN, eds. Botanik, Bd. 2. Systematisk Botanik Nr. 2. Munksgaard, Copenhagen.
- CLARKE, A. K., and M.-J. ERIKSSON. 1996. The cyanobacterium Synechococcus sp. PCC 7942 possesses a close homologue to the chloroplast ClpC protein in higher plants. Plant Mol. Biol. 31:721-730.
- DELWICHE, C., M. KUHSEL, and J. D. PALMER. 1995. Phylogenetic analysis of *tufA* sequences indicates a cyanobacter ial origin of all plastids. Mol. Phylogenet. Evol. 4:110–1285
- DELWICHE, C., and J. D. PALMER. 1996. Rampant horizontal transfer and duplication of rubisco genes in eubacteria and plastids. Mol. Biol. Evol. 13:873–882.
- DOUGLAS, S. E., and C. A. MURPHY. 1994. Structural, trangering scriptional, and phylogenetic analysis of the atpB gene cluster from the plastid of Cryptomonas phi (Cryptophyceae). J. Phycol. 30:329-340.
- DOUGLAS, S. E., C. A. MURPHY, D. F. SPENCER, and M. W. GRAY. 1991. Cryptomonad algae are evolutionary chimaes ras of two phylogenetically distinct unicellular eukaryotes. Nature 350:148-151.
- DOYLE, J. J., and J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: approach using the boostrap. Evolution 39:783-791.
- sion 3.572c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- FUJIWARA, S., M. SAWADA, J. SOMEYA, N. MINAKA, M. KASWACHI, and I. INOUYE. 1994. Molecular phylogenetic analysis of *rbc*L in the Prymnesiophyta. J. Phycol. 30:863–87.
- GIBBS, S. P. 1978. The chloroplast of *Euglena* may have evolved from symbiotic green algae. Can. J. Bot. **56**:2883 2889.
- GRAHAM, L. E., J. M. GRAHAM, and D. E. WUJEK. 1993. UE trastructure of *Chrysodidymus synuroideus* (Synurophysceae). J. Phycol. **29**:330–341.
- HELMCHEN, T. A., D. BHATTACHARYA, and M. MELKONIAN 1995. Analyses of ribosomal RNA sequences from glaucocystophyte cyanelles provide new insights into the evolutionary relationships of plastids. J. Mol. Evol. 41:203–210.
- HIBBERD, D. J. 1976. The ultrastructure and taxonomy of the Chrysophyceae and Prymnesiophyceae (Haptophyceae): a survey with some new observations on the ultrastructure of the Chrysophyceae. Bot. J. Linn. Soc. 72:55–80.

- JEFFREY, S. W. 1989. Chlorophyll c pigments and their distribution in the chromophyte algae. Pp. 13-36 in J. C. GREEN, B. S. C. LEADBEATER, and W. L. DIVER, eds. The chromophyte algae: problems and perspectives. Systematics Association Special Volume, Clarendon Press, Oxford.
- KOWALLIK, K. V., B. STOEBE, I. SCHAFFRAN, P. KROTH-PANIC, and U. Freier. 1995. The chloroplast genome of a chlorophyll a+c containing alga, Odontella sinensis. Plant Mol. Biol. Rep. 13:336-342.
- KUMAR, S. 1995. PHYLTEST: a program for testing phylogenetic hypotheses. Pennsylvania State University, University
- KUMAR, S., and A. RZHETSKY. 1996. Evolutionary relationships of eukaryotic kingdoms. J. Mol. Evol. 42:183-193.
- LEIPE, D. D., P. O. WAINRIGHT, H. J. GUNDERSON, D. PORTER, D. J. PATTERSON, F. VALOIS, S. HIMMERICH, and M. L. So-GIN. 1994. The stramenopiles from a molecular perspective: 16S-like rRNA sequences from Labyrinthuloides minuta and Cafeteria roenbergensis. Phycologia 33:369-377.
- LEITNER, T., D. ESCANILLA, C. FRANZÉN, M. UHLÉN, and J. ALBERT. 1996. Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. Proc. Natl. Acad. Sci. USA 93:10864-10869.
- LOISEAUX-DE GOËR, S. 1994. Plastid lineages. Pp. 137-177 in F. E. ROUND and D. J. CHAPMAN, eds. Progress in phycological research 10. Biopress, Bristol.
- LYNN, D. H., and E. B. SMALL. 1990. Phylum Ciliophora. Pp. 498-523 in L. MARGULIS, J. O. CORLISS, M. MELKONIAN, and D. J. CHAPTMAN, eds. Handbook of Protoctista. Jones and Bartlett, Boston.
- McFadden, G. I., P. R. Gilson, and R. F. Waller. 1995. Molecular phylogeny of chlorarachniophytes based on plastid rRNA and rbcL sequences. Arch. Protistenkd. 145:231-
- MADDISON, D. R. 1991. The discovery and importance of multiple islands of most-parsimonious trees. Syst. Zool. 40:
- MARTIN, W., H. BRINKMANN, C. SAVONNA, and R. CERFF. 1993. Evidence for a chimeric nature of nuclear genomes: eubacterial origin of eukaryotic glyceraldehyde-3-phosphate dehydrogenase genes. Proc. Natl. Acad. Sci. USA 90:8692-8696.
- MEDLIN, L. K., A. COOPER, C. HILL, S. WSIEDEN, and U. WELLBROCK. 1995. Phylogenetic position of the Chromista plastids based on small subunit rRNA coding regions. Curr. Genet. 28:560-565.
- MORDEN, C. W., C. F. DELWICHE, M. KUHSEL, and J. D. PALM-ER. 1992. Gene phylogenies and the endosymbiotic origin of plastids. BioSystems 28:75-90.
- MORRILL, L. C., and A. R. LOEBLICH III. 1983. Ultrastructure of the dinoflagellate amphiesma. Int. Rev. Cytol. 82: 151-
- OLSEN, G. J., H. MATSUDA, R. HAGSTROM, and R. OVERBEEK. 1994. FastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. CA-BIOS 10:41-48.
- PALMER, J. D. 1993. A genetic rainbow of plastids. Nature 364: 762-763.
- PARKE, M., and P. S. DIXON. 1976. Check-list of British marine algae—third revision. J. Mar. Biol. Assoc. UK 56:527–594.
- PASCHER, A. 1913. Chrysomonadinae. In Süsswasserflora Deutschlands, Österreichs und der Schweiz, 2. Verlag von Gustav Fischer, Jena.
- ROWAN, K. S. 1989. Photosynthetic pigments of algae. Cambridge University Press, Cambridge.

- RZHETSKY, A., S. KUMAR, and M. NEI. 1995. Four-cluster analysis: a simple method to test phylogenetic hypotheses. Mol. Biol. Evol. 12:163-167.
- RZHETSKY, A., and M. NEI. 1992. A simple method for estimating and testing minimum-evolution trees. Mol. Biol. Evol. 9:945-967.
- -. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol. Biol. Evol. 10 1073-1095.
- SAUNDERS, G. W., D. POTTER, M. P. PASKIND, and R. A. AN DERSEN. 1995. Cladistic analyses of combined traditional and molecular data sets reveal an algal lineage. Proc. Natl Acad. Sci. USA 92:244-248.
- SCHNEPF, E., and M. ELBRÄCHTER. 1988. Cryptophycean-like double membrane-bound chloroplast in the dinoflagellate Dinophysis Ehrenb.: evolutionary, phylogenetic and toxis cological implications. Bot. Acta 101:196–203.
- STEWART, K. D., and K. MATTOX. 1980. Phylogeny of phy toflagellates. Pp. 433-462 in R. E. Cox, ed. Phytoflagel lates. Elsevier/North-Holland, New York.
- STOEBE, B., U. FREIER, and K. V. KOWALLIK. 1996. The chloroplast genome of a diatom, Odontella sinensis: insight into the origin of chromophyte plastids. Abstract no. 379 from the First European Phycological Congress, Cologne.
- SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Laboratory of Molecular System atics, Smithsonian Institution, Washington, D.C.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. ME HILLIS. 1996. Phylogenetic inference. Pp. 407-514 in D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular Systolic tematics. Sinauer, Sunderland, Mass.
- TAYLOR, F. J. R. 1976. Flagellate phylogeny: a study in const flicts. J. Protozool. 23:28–40.
- TOMAS, R. N., and E. R. Cox. 1973. Observations on the symp biosis of *Peridinium balticum* and its intercellular alga. Ultrastructure. J. Phycol. 9:304-323.
- UEDA, K., and H. SHIBUYA. 1993. Molecular phylogeny of rbcL and its bearing on the origin of plastids: bacteria of cyanobacteria? Pp. 369-376 in S. SATO, M. ISHIDA, and Ho ISHIKAWA, eds. Endocytobiology V. Tübingen University Press, Tübingen, Germany.
- Van de Peer, Y., S. A. Rensing, U.-G. Maier, and R. de WACHTER. 1996. Substitution rate calibration of small sub unit ribosomal RNA identifies chlorarachniophyte endo symbionts as remnants of green algae. Proc. Natl. Acad. Sc. USA **93**:7732–7736.
- VIVIER, E., and I. DESPORTES. 1990. Phylum Apicomplexa. Pp. 549-573 in L. MARGULIS, J. O. CORLISS, M. MELKONIAN, and D. J. CHAPTMAN, eds. Handbook of Protoctista. Jones and Bartlett, Boston.
- WATANABE, M. M., S. SUDA, I. INOUYE, T. SAWAGUCHI, and M. CHIHARA. 1990. Lepidodinium viride gen. et sp. nov (Gymnodiniales, Dinophyta), a green dinoflagellate with 2 chlorophyll a- and b-containing endosymbiont. J. Phycological **26**:741-751.
- Wee, J. L., J. Chesnick, and R. A. Cattolico. 1993. Partia characterization of the chloroplast genome from the chromophytic alga Synura petersenii (Synurophyceae). 📡 Phycol. 29:96-99.
- WILCOX, L. W., and G. J. WEDEMAYER. 1985. Dinoflagellate. with blue-green chloroplasts derived from an endosymbion eukaryote. Science **227**:192–194.
- BARBARA A. SCHAAL, reviewing editor
- Accepted September 5, 1997