

Phylogeny of Plastids Based on Cladistic Analysis of Gene Loss Inferred from Complete Plastid Genome Sequences

Hisayoshi Nozaki,¹ Njij Ohta,² Motomichi Matsuzaki,³ Osami Misumi,^{1,4} Tsuneyoshi Kuroiwa⁵

¹ Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

² Department of Molecular Biology, Faculty of Science, Saitama University, Shimo-Ohkubo, Saitama-shi, Saitama 338-8570, Japan

³ Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

⁴ Bio-oriented Technology Research Advancement Institution (BRAIN), Toranomon, Minato-ku, Tokyo 105-0001, Japan

⁵ Department of Life Science, College of Science, Rikkyo (St.Paul's) University, Nishiikebukuro, Toshima-ku, Tokyo 171-8501, Japan

Received: 13 February 2003 / Accepted: 30 April 2003

Abstract. Based on the recent hypothesis on the origin of eukaryotic phototrophs, red algae, green plants, and glaucophytes constitute the “primary photosynthetic eukaryotes” (whose plastids may have originated directly from a cyanobacterium-like prokaryote via primary endosymbiosis), whereas the plastids of other lineages of eukaryotic phototrophs appear to be the result of secondary or tertiary endosymbiotic events (involving a phototrophic eukaryote and a host cell). Although phylogenetic analyses using multiple plastid genes from a wide range of eukaryotic lineages have been carried out, some of the major phylogenetic relationships of plastids remain ambiguous or conflict between different phylogenetic methods used for nucleotide or amino acid substitutions. Therefore, an alternative methodology to infer the plastid phylogeny is needed. Here, we carried out a cladistic analysis of the “loss of plastid genes” after primary endosymbiosis using complete plastid genome sequences from a wide range of eukaryotic phototrophs. Since it is extremely unlikely that plastid genes are regained during plastid evolution, we used the irreversible Camin-Sokal model for our cladistic analysis of the loss of plastid genes. The cladistic analysis of the 274 plastid protein-coding genes resolved the 20 operational taxonomic units representing a wide range of eukaryotic

lineages (including three secondary plastid-containing groups) into two large monophyletic groups with high bootstrap values: one corresponded to the red lineage and the other consisted of a large clade composed of the green lineage (green plants and *Euglena*) and the basal glaucophyte plastid. Although the sister relationship between the green lineage and the Glaucophyta was not resolved in recent phylogenetic studies using amino acid substitutions from multiple plastid genes, it is consistent with the *rbcL* gene phylogeny and with a recent phylogenetic study using multiple nuclear genes. In addition, our analysis robustly resolved the conflicting/ambiguous phylogenetic positions of secondary plastids in previous phylogenetic studies: the *Euglena* plastid was sister to the chlorophycean (*Chlamydomonas*) lineage, and the secondary plastids from the diatom (*Odontella*) and cryptophyte (*Guillardia*) were monophyletic within the red lineage.

Key words: Evolution — Camin-Sokal model — Cladistic analysis — Gene loss — Phylogeny — Plastids — Plastid genes — Primary plastid — Secondary plastid

Introduction

It is a general tenet of recent evolutionary biology that the red algae, green plants, and glaucophytes

Table 1. List of additional OTUs (Martin et al. 2002) included in the cladistic analyses (Fig. 1) and origin of the plastid genome sequences

OTU	Reference	DDBJ/EMBL/GenBank accession number or URL
<i>Chlamydomonas reinhartii</i>	Maul et al. (2002)	http://bti.cornell.edu/bti2/chlamyweb/
<i>Spinacia oleracea</i>	Schmitz-Linneweber et al. (2001)	NC_002202
<i>Psilotum nudum</i>	Wakasugi et al. (unpublished)	NC_003386
<i>Chaetosphaeridium globosum</i>	Turmel et al. (2002)	NC_004115
<i>Cyanidioschyzon merolae</i> 10D	Ohta et al. (2003)	AB002583

constitute the “primary photosynthetic eukaryotes,” whose plastids have two bounding membranes and likely originated directly from a cyanobacterium-like prokaryote via primary endosymbiosis (e.g., Bhattacharya and Medlin 1995; Delwiche 1999; McFadden 2001; Cavalier-Smith 2002). By contrast, the plastids of other lineages of eukaryotic phototrophs appear to be the result of secondary or tertiary endosymbiotic events (involving a phototrophic eukaryote and a host cell) because they are surrounded by three or four bounding membranes, and the endosymbiotic remnant of the photosynthetic eukaryotic nucleus, called the nucleomorph, may be recognized between the bounding membranes (e.g., Delwiche 1999; McFadden 2001; Cavalier-Smith 2002a; Yoon et al. 2002a). Based on phylogenetic analyses of plastid-coding genes and the similarity of the plastid genome organization (e.g., Morden et al. 1992; Nelissen et al. 1995; Bhattacharya and Medlin 1995), the plastids from all photosynthetic eukaryotes can be considered the product of a single primary endosymbiosis. Recently, complete sequences from various plastids have been determined (e.g., Ohya et al. 1986; Ohta et al. 2003) and phylogenetic analyses using multiple plastid genes from a wide-range of eukaryotic lineages have been carried out in order to resolve robust phylogenetic relationships of plastids (e.g., Martin et al. 2002; Yoon et al. 2002a, b; Maul et al. 2002; Ohta et al. 2003). However, some of the phylogenetic relationships of plastids remain ambiguous or conflict between different phylogenetic methods used for nucleotide and amino acid sequences, especially the phylogenetic positions of the secondary plastids (e.g., Martin et al. 2002; Yoon et al. 2002a, b; Maul et al. 2002; Ohta et al. 2003). Therefore, an alternative methodology to infer the plastid phylogeny is needed.

The most dramatic feature distinguishing plastids from free-living cyanobacteria is the tremendous reduction in the size and gene content of plastid genomes (Martin and Herrmann 1998; Palmer and Delwiche 1998; Delwiche 1999; Martin et al. 2002). In addition, the size of the photosynthetic plastid genomes and number of genes are variable, ranging from 118 to 201 kbp with 62–209 genes (see Ohta et al. 2003). This suggests that plastids have gradually lost their genes during evolution, even after the pri-

mary endosymbiosis, since the reverse evolution of plastid gene loss is generally impossible (Martin et al. 2002). Therefore, the loss of plastid genes seems to be a good indicator of plastid evolution and it should be an alternative to using nucleotide/amino acid substitutions to construct the plastid phylogeny.

This study inferred plastid phylogeny and evolution based on the “loss of plastid genes” deduced from complete plastid genome sequences from a wide range of eukaryotic phototrophs. We used the Camin-Sokal or irreversible model (Camin and Sokal 1965) to conduct a cladistic analysis of the loss of plastid genes. The cladistic analysis of 274 genes from 20 OTUs (operational taxonomic units) produced robust support for the phylogenetic positions of the glaucophyte plastid and the secondary plastids of *Euglena* and the red lineage.

Materials and Methods

The data matrix we used to examine the maintenance or loss of plastid protein-coding genes was that used by Martin et al. (2002, supplementary data). It included the five additional OTUs listed in Table 1 and the *cysW* gene (Ohta et al. 2003), but excluded two nonphotosynthetic organisms (*Plasmodium falciparum* and *Epifagus virsiniensis*) whose plastids are highly reduced. Seven pseudogenes of *Pinus* (*ndhB*, *ndhC*, *ndhD*, *ndhE*, *ndhH*, *ndhI*, and *ndhK*, Wakasugi et al. 1994), which Martin et al. (2002) designated as loss (-), were considered to represent the maintenance of the genes in the present analyses. The annotations of the genes from the additional plastid sequences were based on their descriptions in GenBank, and on their similarity using the search program Blastx at NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). The data matrix consisted of 274 characters from 20 eukaryotic OTUs (Fig. 1), and is available from HN on request.

We carried out a cladistic analysis of the loss of plastid genes after primary endosymbiosis using this data matrix with PAUP 4.0b10 (Swofford 2002). Since the regaining of plastid genes during plastid evolution cannot generally be deduced, we used an irreversible Camin-Sokal model (Camin and Sokal 1965; Wiley et al. 1991) to set character types for the loss or maintenance of 269 plastid genes in PAUP4.0b10. Five other genes (*cuwI*, *matK*, *ycf13*, *ycf68*, and *ycf74*) harbored in the mobile group I and II introns (see Lambowitz and Belfort 1993) were designated as unordered characters. The loss of plastid genes is a derived character. Therefore, the loss and maintenance of a gene are polarized *a priori* in the transformation series. Consequently, the root or ingroup node of the phylogenetic tree was determined without outgroup comparison after the cladistic analyses using PAUP4.0b10 (see Wiley et al. 1991). A cladistic or maximum parsimony analysis of the data matrix was carried out using a branch-and-branch search with PAUP 4.0b10 and the robustness of the resulting lineages was

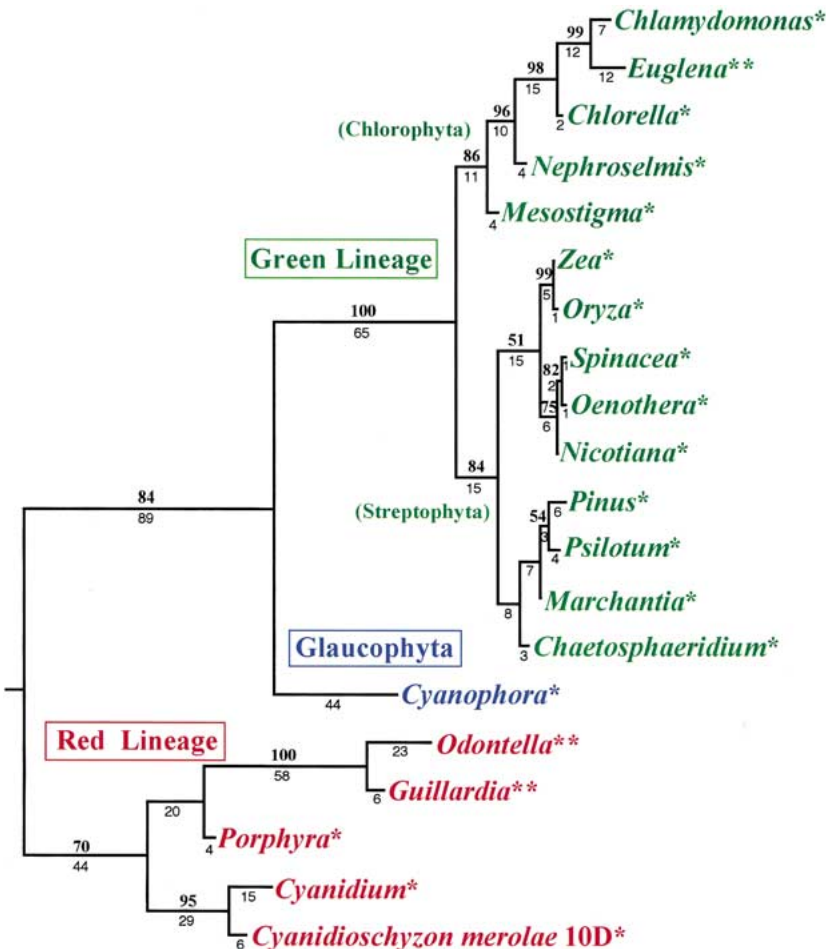


Fig. 1. Single most parsimonious (MP) tree (with tree length of 557 and consistency index of 0.4147) based on the present cladistic analyses of 274 plastid genes from 20 OTUs representing a wide-range of eukaryotic phototrophs. The tree was constructed based on the branch-and-bound search by PAUP 4.0b10 (Swofford 2002). Irreversible Camin-Sokal model (Camin and Sokal 1965; Wiley et al. 1991) was used for setting character types in loss or maintenance of 269 plastid genes by PAUP 4.0b10, whereas five other genes harbored in the mobile group I and II introns were designated unordered. Branch lengths are proportional to numbers of character changes, which are indicated by numbers below branches, based on ACCTRAN by PAUP 4.0b10. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the branch-and-bound search by PAUP 4.0b10. Single and double asterisks indicate the primary and secondary plastids, respectively.

tested with bootstrap analyses (Felsenstein 1985) involving 1000 replications of the branch-and-branch search with PAUP 4.0b10.

Results

Based on our cladistic analysis of 274 genes from 20 OTUs, a single MP tree was constructed and is presented in Fig. 1. The tree was 557 steps long, with a consistency index of 0.4147 and a retention index of 0.8939.

Two major monophyletic groups were resolved with relatively high bootstrap values (70–84%), with the root of the plastids positioned between these two monophyletic groups. One group consisted of the green lineage (green plants plus *Euglena*) and the glaucophyte *Cyanophora*, and the other (the red lineage) contained three red algal plastids (*Porphyra*, *Cyanidium*, and *Cyanidioschyzon*) and secondary plastids from a diatom (*Odontella*) and cryptophyte (*Guillardia*).

Within the former group, *Cyanophora* is basal to the green lineage, which was subdivided into two clades supported with relatively high bootstrap support (84–86%). One of the two clades was composed of two prasinophycean algae (*Mesostigma* and

Nephroselmis) in basal positions and a robust monophyletic group (98% bootstrap support) consisting of *Chlorella*, *Chlamydomonas*, and *Euglena*, with high bootstrap support (99%) for the sister relationship between *Chlamydomonas* and *Euglena*. The other clade consisted of five angiosperms (*Zea*, *Oryza*, *Spinacea*, *Oenothera*, and *Nicotiana*), the gymnosperm *Pinus*, the fern *Psilotum*, the liverwort *Marchantia*, and the charophycean alga *Chaetosphaeridium*, although the basal phylogenetic relationships within the clade were almost completely ambiguous (Fig. 1).

Two robust monophyletic groups were resolved within the red lineage with 95–100% bootstrap values: one composed of the two secondary plastids (*Guillardia* and *Odontella*) and the other consisting of two cyanidiophycean algae (*Cyanidium* and *Cyanidioschyzon*) (Fig. 1). However, the phylogenetic position of the red alga *Porphyra* was ambiguous within the red lineage.

Discussion

Our cladistic analysis of 274 genes from 20 OTUs, including three secondary plastids, resolved two ma-

major lineages with relatively high bootstrap support: the red lineage and a large monophyletic group composed of the green lineage (green plants and *Euglena*) and the basal glaucophyte plastid. These phylogenetic relationships suggest that the in-group node of the plastid lineage is positioned between these two major lineages and a sister relationship of the plastids in the green lineage and the Glaucophyta (Fig. 1). In contrast, recent plastid multigene phylogenies involving more than 8,000 amino acid sequences demonstrated that the glaucophyte *Cyanophora* occupies the basal position within the plastids (Adachi et al. 2000; Martin et al. 2002, Ohta et al. 2003). This discrepancy may result from the small taxon sampling (see Zwickl and Hillis 2002) from the Glaucophyta (only *Cyanophora paradoxa*) for both phylogenetic analyses. Additional OTUs from the Glaucophyta may resolve this problem. However, the sister relationship between the green lineage and the glaucophyte plastid resolved using our cladistic method is consistent with the *rbcL* gene phylogeny (Delwiche and Palmer 1997) and one recent phylogenetic study using multiple nuclear genes (Nozaki et al. 2003). In contrast to the monophyly of most of the plastid genes within the cyanobacterial lineage (e.g., Morden et al. 1992; Nelissen et al. 1995; Bhattacharya and Medlin 1995), the *rbcL* gene tree contains two separate lineages: the red lineage and the green/glaucophyte lineage (Delwiche and Palmer 1997). This polyphyletic status of the plastids is generally attributed to horizontal gene transfer (Morden et al. 1992; Delwiche and Palmer 1997). Very recently, Nozaki et al. (2003) carried out phylogenetic analyses of various lineages of only mitochondria-containing eukaryotic organisms using nuclear multigene sequences, including the complete sequence from the primitive red alga *Cyanidioschyzon merolae*. They resolved the position of the red algae as basal to a large monophyletic group including green plants and the Glaucophyta. If one assumes that primary endosymbiosis occurred only once and the phylogenetic relationships resolved by Nozaki et al. (2003) are reliable, our phylogenetic position of the Glaucophyta (Fig. 1) should be natural.

Since the plastids of the Euglenophyta or Euglenozoa (Discicristata) are surrounded by three bounding membranes and contain chlorophylls a and b, they are considered to have originated from the secondary endosymbiosis of green plants (e.g., Delwiche 1999; McFadden 2001). However, the phylogenetic position of the *Euglena* plastid conflicts between different methods used to deduce multigene phylogenies, even based on the same amino acid sequence data (8,308 a.a. from 16 OTUs) by Martin et al. (2002). The basal position of the *Euglena* plastid within the green lineage was resolved with 86–100% bootstrap support using distance methods, whereas

ML analyses resolved the sister relationship between *Euglena* and *Chlorella* with 64% bootstrap support (Martin et al. 2002). Turmel et al. (1999), on the basis of ML analyses of 7,499 a.a. from 37 plastid genes, showed that the *Euglena* plastid is sister to the clade composed of *Chlorella* (Trebouxiophyceae) and *Chlamydomonas* (Chlorophyceae). Our cladistic analysis demonstrated that *Euglena* and *Chlamydomonas* formed a very robust clade (99% bootstrap support) that is sister to *Chlorella* (Fig. 1). Since recent molecular phylogenetic analyses based on 18S rRNA genes and comparative ultrastructural data suggest the sister relationship between the Chlorophyceae and Trebouxiophyceae (e.g., Friedl 1997; Graham and Wilcox 2000), the ancestor of the euglenophyte plastid should be a chlorophycean alga.

Very recently, several cyanobacterial or plant-like genes were found in plastid-lacking organisms (Heterolobosea and Kinetoplastida) belonging to the Discicristata (Andersson and Roger 2002; Hannaert et al. 2003). Since the Euglenozoa have secondary plastids and are closely related to the Kinetoplastida, Hannaert et al. (2003) discussed that a common ancestor of both Euglenozoa and Kinetoplastida already acquired the secondary plastids that gave rise to the plant-like genes in the Kinetoplastida nuclear genome. However, phylogenetic analyses of various euglenoid taxa suggested that the evolution of phototrophy occurred in the common ancestor of the monophyletic plastid-containing group (Euglenales) distally positioned within the Euglenozoa (Montegut-Felkner and Triemer 1997; Preisfeld et al. 2000; Müllner et al. 2001), and the Heterolobosea are positioned outside the clade composed of Euglenozoa and Kinetoplastida (Baldauf et al. 2000; Nozaki et al. 2003). In addition, the present cladistic analysis suggested the relatively recent acquisition of the secondary plastid of *Euglena* (after the divergence of Chlorophyceae and Trebouxiophyceae) (Fig. 1). Therefore, the secondary plastid endosymbiosis can hardly be considered before the divergence of Euglenozoa and Kinetoplastida or Heterolobosea, and the cyanobacterial or plant-like genes found in Heterolobosea and Kinetoplastida (Andersson and Roger 2002; Hannaert et al. 2003) are possibly derived from the primary plastid endosymbiosis. Nozaki et al. (2003) suggested that the primary plastid endosymbiosis likely occurred once in the common ancestor of three primary plastid-containing lineages as well as in Discicristata, Heterokontophyta, and Alveolata (apicomplexans and Ciliophora), from which the primary plastids were subsequently lost, but some of their genes have remained in the nuclear genomes (Andersson and Roger 2002; Nozaki et al. 2003).

The phylogenetic positions of the secondary plastids within the red lineage conflict in recent phy-

logenetic analyses of multigene sequences (Martin et al. 2002; Maul et al. 2002; Yoon et al. 2002a, b; Ohta et al. 2003). Based on combined *psaA* and *psbA* sequences, the cryptophytes are robustly separated from the clade composed of the Cyanidiophyceae (*Cyanidium*, *Cyanidioschyzon*, etc.) and other secondary plastid-containing algae (Heterokontophyta and Haptophyta) (Yoon et al. 2002a). Non-monophyly of the secondary plastids in the red lineage was also resolved in the phylogenetic studies using more than 8,000 amino acid sequences from limited numbers of OTUs (Martin et al. 2002; Maul et al. 2002; Ohta et al. 2003). In contrast, Yoon et al. (2002b) resolved a large monophyletic group composed of the secondary plastids from Heterokontophyta, Cryptophyta, and Haptophyta in the red lineage, in an analysis of the combined 16S rRNA, *psaA*, *psbA*, *rbcL*, and *tufA* genes DNA sequences (5827 nt from 29 OTUs of the red lineage plus 6 OTUs). They argue that the separate phylogenetic position of the Cryptophyta is derived from the small number of OTUs analyzed by Martin et al. (2002). Our results based on gene loss (Fig. 1) are essentially consistent with Yoon et al. (2002b) in that the secondary plastids of the red lineage have a single ancestor.

The phylogenetic position of *Mesostigma* is conflicting in recent phylogenetic analyses using multiple plastid gene sequences (Lemieux et al. 2000; Martin et al. 2002; Maul et al. 2002; Ohta et al. 2003). This primitive green alga is positioned most basally within the green plants (Lemieux et al. 2000; Maul et al. 2002) or represents the most basal lineage within the Streptophyta (embryophytes and Charophyceae) (Martin et al. 2002; Ohta et al. 2003). In contrast, the present cladistic analysis of loss of plastid genes suggested the basal position of *Mesostigma* within the Chlorophyta (*Nephroselmis*, *Chlamydomonas*, etc.) (Fig. 1). This situation may result from limited taxon sampling from Chlorophyta in the present study (especially in lacking ulvophycean algae). However, pigment composition (Fawley and Lee 1990) and presence of stigma (Graham and Wilcox 2000) in *Mesostigma* suggest its phylogenetic position within the Chlorophyta.

As discussed above, the phylogenetic positions of the secondary plastids in both the red and green lineages appeared ambiguous, even when a large nucleotide or amino acid sequences from multiple plastid genes were analyzed. This is probably based on the unusual bias for gene substitutions related to secondary endosymbiosis, which also caused highly divergent gene sequences of the vestigial nucleus or nucleomorph in the Cryptophyta and Chlorarachniophyta (see Van de Peer et al. 1996). According to Adachi et al. (2000), *Euglena* deviates most strongly from other plastid genomes at the

level of amino acid composition. Recently, Itoh et al. (2002) compared the complete genome sequences of two different endosymbionts, and demonstrated that the rate of amino acid substitution is two times higher in symbionts than in their relatives, suggesting that the elevated evolutionary rate is mainly due to an enhanced mutation rate in endosymbiosis. Therefore, traditional molecular phylogenetic methods using nucleotide or amino acid substitutions should have limited efficiency for resolving the phylogeny of secondary and tertiary plastids especially when the numbers of OTUs are limited. Since the loss of the plastid genes are not directly related to such unusual gene substitutions, the present phylogenetic results regarding the secondary plastids (Fig. 1) can be reliable alternative hypotheses. However, as in the traditional molecular phylogenetic studies using nucleotide or amino acid substitutions, the present cladistic method is potentially associated with the possible clustering of some lineages due to convergent and/or accelerated evolution especially when the taxon sampling is poor. Further information of complete plastid genome sequences should resolve more natural phylogenetic relationships of plastids.

Acknowledgments. This study was supported by Grant-in-Aid for Scientific Research on Priority Areas (c) "Genome Biology" from the Ministry of Education, Culture, Sports, Science and Technology, Japan (No. 1320611 to TK), and by the Program for the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN; to TK and HN).

Addendum

During the submission of this paper (received on 13 February 2003), Grzebyk et al. (J. Phycol. 39: 259-267, April 2003 issue) published a minireview "Mesozoic radiation of eukaryotic algae the portable plastid hypothesis." They showed a phylogenetic analysis of 15 plastids inferred from gene presence and gene loss in 256 plastid protein-coding genes using the Camin-Sokal parsimony method. Their phylogenetic results are essentially the same as those of the present paper in the phylogenetic positions of the Glaucophyta and the monophyly of the red secondary plastids. However, Grzebyk et al. (2003) did not resolve robustly basal phylogenetic relationships within the green lineage, and did not demonstrate the chlorophycean phylogenetic position of the euglenoid secondary plastids due to the lack of *Chlamydomonas*. Although they robustly resolved the sister relationships between the Glaucophyta and the green plastid lineage as well as between *Chlorella* and euglenoid plastids, these results were not accepted based on the "maintenance" of certain plastid genes (Grzebyk et al. 2003).

References

- Adachi J, Waddell PJ, Martin W, Hasegawa M (2000) Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J Mol Evol* 50:348–358
- Andersson JO, Roger AJ (2002) A cyanobacterial gene in non-photosynthetic protists—An early chloroplast acquisition in Eukaryotes? *Curr Biol* 12:115–119
- Baldauf SL, Roger AJ, Wenk-Siefert I, Doolittle WF (2000) A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* 290:972–977
- Bhattacharya D, Medlin L (1995) The phylogeny of plastids: A review based on comparisons of small subunit ribosomal RNA coding regions. *J Phycol* 31:489–498
- Camin JH, Sokal RR (1965) A method for deducing branching sequences in phylogeny. *Evolution* 19:311–326
- Cavalier-Smith T (2002) Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr Biol* 12:R62–R64
- Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 154:164–177
- Delwiche CF, Palmer JD (1997) The origin of plastids and their spread via secondary symbiosis. In: Bhattacharya D (ed) *Origin of algae and their plastids*. Springer-Verlag, Wien, pp 53–86
- Fawley MW, Lee CM (1990) Pigment composition of the scaly green flagellate *Mesostigma viride* (Micromonadophyceae) is similar to that of the siphonous green alga *Bryopsis plumosa* (Ulvophyceae). *J Phycol* 26:666–670
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 38:16–24
- Friedl T (1997) The evolution of the green algae. In: Bhattacharya D (ed) *Origin of algae and their plastids*. Springer-Verlag, Wien, pp 87–101
- Graham LE, Wilcox LW (2000) *Algae*. Prentice Hall, Upper Saddle River, NJ
- Hannaert V, Saavedra E, Duffieux F, Szikora JP, Rigden DJ, Michels PAM, Opperdoes FR (2003) Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc Natl Acad Sci USA* 100:1067–1071
- Itoh T, Martin W, Nei M (2002) Acceleration of genomic evolution caused by enhanced mutation rate in endocellular symbionts. *Proc Natl Acad Sci USA* 99:12944–12948
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 16:111–120
- Lambowitz AM, Belfort M (1993) Introns as mobile genetic elements. *Ann Rev Biochem* 62:587–622
- Martin W, Herrmann RG (1998) Gene transfer from organelles to the nucleus: How much, what happens, and why? *Plant Physiol* 118:9–17
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, et al. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc Natl Acad Sci USA* 99:12246–12251
- Maul JE, Lilly JW, Cui L, dePamphilis CW, Miller W, Harris EH, Stern DB (2002) The *Chlamydomonas reinhardtii* plastid chromosome: Islands of genes in a sea of repeats. *Plant Cell* 14:2659–2679
- McFadden GI (2001) Primary and secondary endosymbiosis and the origin of plastids. *J Phycol* 37:951–959
- Montegut-Felkner AE, Triemer RE (1997) Phylogenetic relationships of selected euglenoid genera based on morphological and molecular data. *J Phycol* 33:512–519
- Morden CW, Delwiche CF, Kuhse M, Palmer JD (1992) Gene phylogenies and the endosymbiotic origin of plastids. *BioSystems* 28:75–90
- Müllner AN, Angeler DG, Samuel R, Linton EW, Triemer RE (2001) Phylogenetic analysis of phagotrophic, phototrophic and osmotrophic euglenoids by using the nuclear 18S rDNA sequence. *Int J Syst Evol Micr* 51:783–791
- Nelissen B, Van de Peer Y, Wilmotte A, De Wachter R (1995) An early origin of plastids within the cyanobacterial divergence is suggested by evolutionary trees based on complete 16S rRNA sequences. *Mol Biol Evol* 12:1166–1173
- Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, Hasegawa M, et al. (2003) The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J Mol Evol* 56:485–497
- Ohta N, Matsuzaki M, Misumi O, Miyagishima S, Nozaki H, Tanaka K, et al. (2003) Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Research* 10:67–77
- Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, et al. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. *Nature* 322:572–574
- Palmer JD, Delwiche CF (1998) The origin and evolution of plastids and their genomes. In: Soltis PS, Soltis DE, Doyle JJ (eds) *Molecular systematics of plants. II. DNA sequencing*. Kluwer Academic, Boston, pp 375–409
- Preisfeld A, Berger S, Busse I, Liller S, Ruppel HG (2000) Phylogenetic analyses of various euglenoid taxa (Euglenozoa) based on 18S rDNA sequence data. *J Phycol* 36:220–226
- Schmitz-Linneweber C, Maier RM, Alcaraz JP, Cottet A, Herrmann RG, Mache R (2001) The plastid chromosome of spinach (*Spinacia oleracea*): Complete nucleotide sequence and gene organization. *Plant Mol Biol* 45:307–315
- Swofford DL (2002) PAUP* 4.0: Phylogenetic Analysis Using Parsimony, version 4.0b10. Sinauer Associates, Inc., Sunderland, MA
- Turmel M, Otis C, Lemieux C (1999) The complete chloroplast DNA sequence of the green alga *Nephroselmis olivacea*: Insights into the architecture of ancestral chloroplast genomes. *Proc Natl Acad Sci USA* 96:10248–10253
- Turmel M, Otis C, Lemieux C (2002) The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proc Natl Acad Sci USA* 99:11275–11280
- Van de Peer Y, Rensing SA, Maier UG, De Wachter R (1996) Substitution rate calibration of small subunit ribosomal RNA identifies chlorarachniophyte endosymbionts as remnants of green algae. *Proc Natl Acad Sci USA* 93:7732–7736
- Wakasugi T, Tsudzuki J, Ito S, Nakashima K, Tsudzuki T, Sugiura M (1994) Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. *Proc Natl Acad Sci USA* 91:9794–9798
- Wiley EO, Siegel-Causey D, Brooks DR, Funk VA (1991) *The complete cladist: A primer of phylogenetic procedures*. Special Publication No 19. Museum of Natural History, University of Kansas, Lawrence, KS
- Yoon HS, Hackett JD, Bhattacharya D (2002a) A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc Natl Acad Sci USA* 99:11724–11729
- Yoon HS, Hackett JD, Pinto G, Bhattacharya D (2002b) The single, ancient origin of chromist plastids. *Proc Natl Acad Sci USA* 99:15507–15512
- Zwickl DJ, Hillis DM (2002) Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 51:588–598