

Cyanobacterial Genes Transmitted to the Nucleus Before Divergence of Red Algae in the Chromista

Hisayoshi Nozaki,¹ Motomichi Matsuzaki,² Osami Misumi,^{3,4} Haruko Kuroiwa,^{3,4} Masami Hasegawa,^{1,5} Tetsuya Higashiyama,¹ Tadasu Shin-i,⁶ Yuji Kohara,⁶ Naotake Ogasawara,⁷ Tsuneyoshi Kuroiwa³

- ¹ Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
² Department of Biomedical Chemistry, Graduate School of Medicine, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
³ Department of Life Science, College of Science, Rikkyo (St. Paul's) University, Nishiikebukuro, Toshima-ku, Tokyo 171-8501, Japan
⁴ Bio-oriented Technology Research Advancement Institution (BRAIN), Toranomon, Minato-ku, Tokyo 105-0001, Japan
⁵ The Institute of Statistical Mathematics, 4-6-7, Minami-Azabu, Minato-ku, Tokyo 106-8569, Japan
⁶ Center for Genetic Resource Information, National Institute of Genetics, 1111 Yata, Mishima, Shizuoka 411-8540, Japan
⁷ Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama Ikoma, Nara 630-0101, Japan

Received: 20 October 2003 / Accepted: 28 January 2004

Abstract. The plastids of red algae, green plants, and glaucophytes may have originated directly from a cyanobacterium-like prokaryote via primary endosymbiosis. In 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 eukaryotic host cell. Although phylogenetic analyses of multiple plastid genes from a wide range of eukaryotic lineages have been carried out, the phylogenetic positions of the secondary plastids of the Chromista (Heterokontophyta, Haptophyta and Cryptophyta) are ambiguous in a range of different analyses. This ambiguity may be the result of unusual substitutions or bias in the plastid genes established by the secondary endosymbiosis. In this study, we carried out phylogenetic analyses of five nuclear genes of cyanobacterial origin (6-phosphogluconate dehydrogenase [*gnd*], oxygen-evolving-enhancer [*psbO*], phosphoglycerate kinase [*pgk*], delta-aminolevulinic acid dehydratase [*aladh*], and ATP synthase gamma [*atpC*] genes), using the genome sequence data from the primitive red alga *Cyanidioschyzon merolae* 10D. The sequence data robustly resolved the origin of the cyanobacterial genes in the nuclei of the Chromista

(Heterokontophyta and Haptophyta) and Dinophyta, before the divergence of the extant red algae (including *Porphyra* [Rhodophyceae] and *Cyanidioschyzon* [Cydidiophyceae]). Although it is likely that *gnd* genes in the Chromista were transmitted from the cyanobacterium-like ancestor of plastids in the primary endosymbiosis, other genes might have been transferred from nuclei of a red algal ancestor in the secondary endosymbiosis. Therefore, the results indicate that the Chromista might have originated from the ancient secondary endosymbiosis before the divergence of extant red algae.

Key words: Chromista — Cyanobacterial nuclear genes — Phylogeny — Plastids — Primary endosymbiosis — Secondary endosymbiosis

Introduction

The origin and diversity of plastids (chloroplasts) in eukaryotic cells can be attributed to two types of endosymbiotic events: primary endosymbiosis, by which the primary plastids (with two bounding membranes) originated directly from a cyanobacterium-like prokaryote, and secondary or tertiary

Table 1. List of additional sequences (Andersson and Roger 2002; Archibald et al. 2003) included in the wide range of eukaryotic phylogeny (Figs. 1, 2, 3, 4, 5) and DDBJ/EMBL/GenBank accession numbers of the five nuclear genes

Taxon	Origin or DDBJ/EMBL/GenBank accession number				
	<i>gnd</i>	<i>psbO</i>	<i>pgk</i>	<i>aladh</i>	<i>atpC</i>
<i>Cyanidioschyzon</i>	AB159595 AB159596	AB159597	AB159598	AB159599	AB159600
<i>Porphyra</i>				EsT ^{a,b}	EST ^a

^aObtained from EST data of *Porphyra yezoensis* (Nikaido et al. 2000).

^bShort sequence (< 50%).

endosymbiosis, involving a phototrophic eukaryote and a eukaryotic host cell giving rise to secondary or tertiary plastids (surrounded by three or four bounding membranes) (e.g., Bhattacharya and Medlin 1995; Delwiche 1999; McFadden 2001; Cavalier-Smith 2002; Yoon et al. 2002a). Based on phylogenetic analyses of plastid-coding genes, and the similarity of pigment compositions and plastid genome organization, the secondary plastids have been assigned to either the red lineage with the red algal plastids or the green lineage with green plant plastids (e.g., Bhattacharya and Medlin 1995; Delwiche 1999; McFadden 2001; Cavalier-Smith 2002; Nozaki et al. 2003b). The Chromista, which are comprised of the Heterokontophyta, Haptophyta, and Cryptophyta, have secondary plastids of the red lineage (e.g., Delwiche 1999; McFadden 2001; Cavalier-Smith 2002, 2003). Although phylogenetic analyses using multiple plastid genes from a wide range of eukaryotic lineages suggest that the secondary plastids in the Chromista originated after the divergence of extant red algae (Rhodophyceae and Cyanidiophyceae), the phylogenetic positions of secondary plastids have been found to be ambiguous in a range of different phylogenetic analyses (Martin et al. 2002; Maul et al. 2002; Yoon et al. 2002a, b; Ohta et al. 2003). This ambiguity might be derived from unusual substitutions or bias, related to differences between the primary and the secondary plastids during evolution.

Cyanobacterial genes are encoded in the nuclear genomes in a wide range of primary photosynthetic organisms, and these genes are thought to have been transmitted from a cyanobacterium-like ancestor of plastids by primary endosymbiosis. These genes are usually located in the nuclear genomes of the secondary photosynthetic organisms and are, therefore, useful as phylogenetic markers for primary and secondary endosymbiosis that has occurred during the evolution of eukaryotes. Andersson and Roger (2002) demonstrated that 6-phosphogluconate dehydrogenase (*gnd*) genes with cyanobacterial affinity are present in the nuclear genomes of primary and secondary photosynthetic organisms, including Heterokontophyta, green plants, and the red alga *Porphyra*.

Ishida and Green (2002) investigated the phylogenetic relationships between the plastids from Chromista and Dinophyta based on sequences of the nuclear-encoded plastid gene encoding oxygen-evolving-enhancer (*psbO*). More recently, Archibald et al. (2003) analyzed several cyanobacterial nuclear genes from various eukaryotes, including the chlorarachniophyte *Bigelowiella*. However, since most of these phylogenetic analyses included only a single red algal operational taxonomic unit (OTU) (*Porphyra* from the Rhodophyceae), the phylogenetic position of secondary plastids in the Chromista could not be resolved with respect to divergence of the extant red algae (Andersson and Roger 2002; Ishida and Green 2002; Archibald et al. 2003). Therefore, the study of cyanobacterial nuclear genes from other red algae, especially those of the Cyanidiophyceae, is needed to resolve this problem.

We recently launched a genome project to investigate the primitive red alga *Cyanidioschyzon merolae* (Cyanidiophyceae). This project has resulted in the complete nuclear genome sequence for this taxon (Matsuzaki et al. unpublished). In the present study, we aimed to determine the plastid primary and secondary endosymbiosis based on sequences from nuclear genes of cyanobacterial origin. The analysis of these genes has given robust support to the origin of the cyanobacterial genes in the Chromista nuclei, before the divergence of the extant red algae.

Materials and Methods

Phylogenetic Analyses of Protein Sequences from Each Gene

The data matrices of the five nuclear genes of cyanobacterial origin were found to be essentially the same as those reported in Andersson and Roger (2002) and Archibald et al. (2003), except that they included the sequences listed in Table 1 and excluded amitochondrial eukaryotic organisms, which are known to have unusual gene substitutions due to parasitism or the absence of mitochondria (see Stiller et al. 1998, 2001). The four nuclear genes (*gnd*, *psbO*, phosphoglycerate kinase [*pgk*], delta-aminolevulinic acid dehydratase [*aladh*], and ATP synthase gamma [*atpC*]) used by Archibald et al. (2003) were selected for the present study, as OTUs

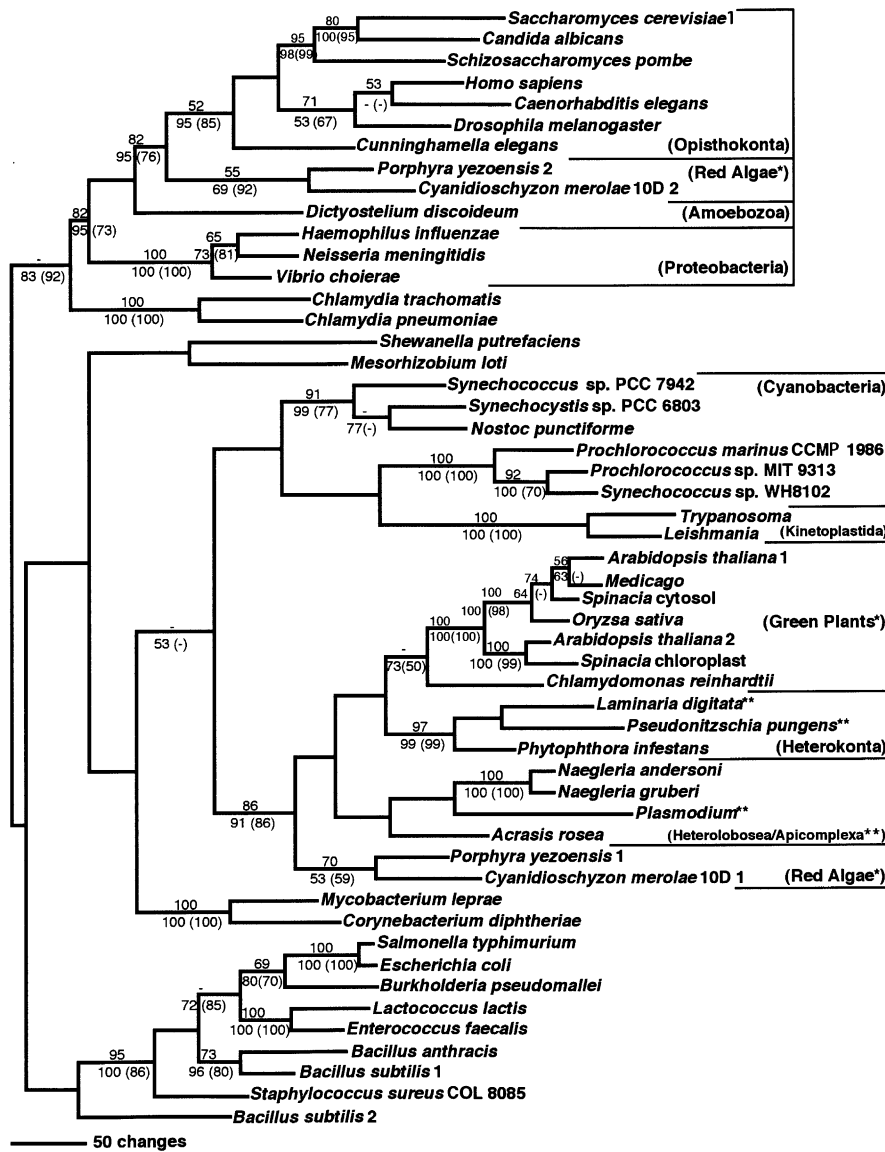


Fig. 1. One of the seven most parsimonious trees (with tree length of 4930 and consistency index of 0.4069) based on 443 amino acid sequences from *gnd* genes of 52 OTUs representing a wide range of prokaryotic and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by AUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and/or maximum likelihood calculation (based on JTT model with rate heterogeneity) are also shown by numbers under the branches without or with parent and double asterisks, respectively. Single and double asterisks indicate taxa with primary and secondary plastids, respectively.

from the Chromista and red algae were included in the data matrices in that study, and two representatives of the red algae (*Cyanidioschyzon* of the Cyanidiphyceae and *Porphyra/Gracilaria* of the Rhodophyceae) were analyzed using the sequences provided in Table 1. Since the *pgk* and *aladh* genes of the chlorarachniophyte *Bigelowiella* can be considered to be results of the lateral gene transfers from the red algae, these genes were excluded from the present phylogenetic analyses. The sequences were aligned using CLUSTAL X (Thompson et al. 1997) with default options. Ambiguous portions with large gaps in the alignment were excluded from the data matrices (available from HN on request).

Maximum parsimony (MP) trees were constructed using a heuristic search (with the tree bisection-reconnection [TBR] branch-swapping algorithm), using PAUP 4.0 b10 (Swofford 2002). Neighbor-joining (NJ) trees (Saitou and Nei 1987) were constructed using CLUSTAL X, and genetic distances were calculated according to the method of Kimura (1983). Maximum likelihood (ML) analyses were carried out using Proml of PHYLIP 3.6 (Felsenstein 2002) based on the JTT model (Jones et al. 1992) and the discrete gamma model for site heterogeneity (using five discrete rate categories) with alpha parameters that were calculated by TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996). The calculations

of bootstrap values (Felsenstein 1985) for the MP and NJ analyses were based on 1000 replications. For the ML method, bootstrap values were calculated based on 100 replications using Seqboot and Proml of PHYLIP 3.6 (Felsenstein 2002).

Phylogenetic Analyses of Limited Eukaryotic OTUs Using Concatenated Protein Sequences from Four Nuclear Genes

Sequence data for 10 representative OTUs (*Spinacea*, monocotyledons [*Oryza*, *Triticum*, and *Hordeum*], *Arabidopsis*, *Chlamydomonas*, Rhodophyceae [*Porphyra* and *Gracilaria*], *Cyanidioschyzon* [Cyanidiphyceae], diatoms [*Phaeodactylum* and *Odontella*], *Anabaena* sp. PCC 7120, *Synechocystis* sp. PCC 6803, and *Prochlorococcus marinus* CCMP 1986) were extracted from the data matrices of the four genes (*psbO*, *pgk*, *aladh*, and *atpC*). In these four nuclear genes, plastid secondary endosymbiosis possibly resulted in the transference of genes to the nuclei of the Chromista (see below). These four amino acid sequences were concatenated to form a single data matrix consisting of 1217 amino acids (available from H.N. on request), which was analyzed phylogenetically as

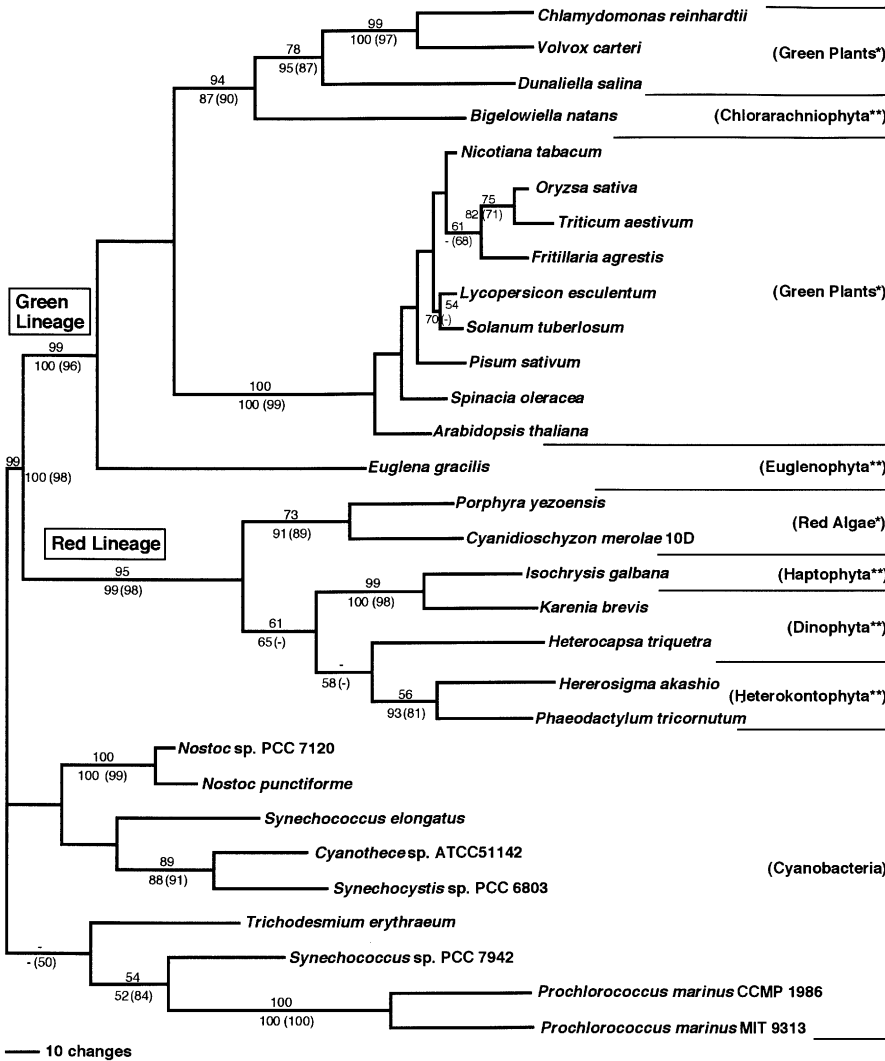


Fig. 2. One of the two most parsimonious trees (with tree length of 1406 and consistency index of 0.6031) based on 228 amino acid sequences from *psbO* genes of 30 OTUs representing cyanobacterial and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by PAUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and/or maximum likelihood calculation (based on JTT model with rate heterogeneity) are also shown by numbers without or with parentheses, respectively. Single and double asterisks indicate taxa with primary and secondary/tertiary plastids, respectively.

described above except that bootstrap values for ML analysis were calculated based on 1000 replications. In addition, all 15 possible trees for green plants (*Spinacea*, monocotyledons, *Arabidopsis*, and *Chlamydomonas*), Rhodophyceae, *Cyanidioschyzon*, diatoms, and cyanobacteria (*Anabaena sp. PCC 7120*, *Synechocystis sp. PCC 6803*, and *Prochlorococcus marinus CCMP 1986*; designated as the outgroup) were examined by PROTML of MOLPHY 2.3 (Adachi and Hasegawa 1996) and AAML of PAML (Yang 1997) without and with the discrete gamma model for site heterogeneity (the relationships within green plants and cyanobacteria were fixed as shown in Fig. 6). Bootstrap values were calculated by the REL method (Kishino et al. 1990) with 10,000 replications. The Kishino and Hasegawa (KH; 1989), weighted Shimodaira and Hasegawa (WSH; 1999), and approximately unbiased (AU) tests were performed using the program CONSEL (Shimodaira and Hasegawa 2001).

Results

Phylogeny of *gnd* Genes

The *gnd* genes of the eukaryotic organisms analyzed in this study formed two distinct lineages, one closely related to the cyanobacterial lineage and the other

with higher affinity to the proteobacteria (Fig. 1). The former lineage was composed of primary and secondary plastid-containing organisms (green plants, red algae, heterokontophytes, and apicomplexans), as well as plastid-lacking organisms (heterokont *Phytophthora* and Heterolobosea) belonging to the Bikonta (Cavalier-Smith 2003) or Plantae (Nozaki et al. 2003a). Although monophyly of the red algae, green plants, and Heterokonta was resolved, the phylogenetic positions of these three groups were ambiguous within the former eukaryotic lineage (Fig. 1). The latter eukaryotic lineage was composed of the Opisthokonta (Metazoa and Fungi), red algae, and cellular slime mold *Dictyostelium*, whereas red algae were resolved as a robust monophyletic sister group to the Opisthokonta, with *Dictyostelium* occupying the most basal position (Fig. 1).

Phylogenetic Analyses of Other Nuclear Genes

The *psbO* protein sequences analyzed in this study demonstrated that the red and green lineages are re-

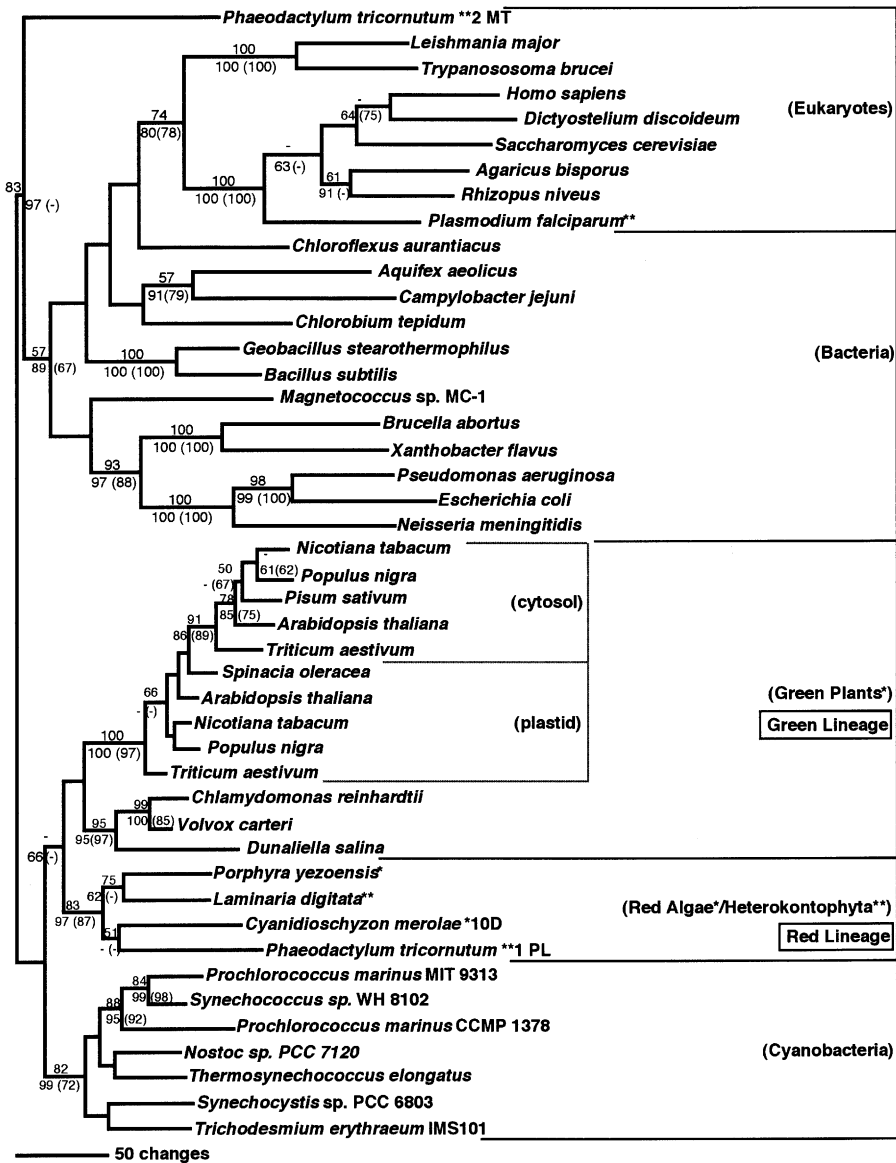


Fig. 3. Single most parsimonious tree (with tree length of 3232 and consistency index of 0.4709) based on 372 amino acid sequences from *pgk* genes of 46 OTUs representing various prokaryotic and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by PAUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and/or maximum likelihood calculation (based on JTT model with rate heterogeneity) are also shown by numbers under the branches without or with parentheses, respectively. Single and double asterisks indicate taxa with primary and secondary plastids, respectively.

solved as two monophyletic groups with high bootstrap support (95–100%) using all three phylogenetic methods (Fig. 2). In the red lineage, the red algae (*Porphyra* and *Cyanidioschyzon*) formed a robust monophyletic group (with 73–91% bootstrap values) that was sister to the lineage composed of Chromista (Haptophyta and Heterokontophyta) and Dinophyta (Fig. 2). Other phylogenetic results of *psbO* gene phylogeny were essentially the same as those reported by Ishida and Green (2002) and Archibald et al. (2003).

In the phylogenetic analyses of *pgk* and *atpC*, the red lineage was resolved as a robust monophyletic group in all of the three phylogenetic methods (with more than 80% bootstrap values) (Figs. 3 and 4). However, only the NJ and ML analysis resolved the red lineage with more than 50% bootstrap values in the phylogeny of *aladh* gene sequences (Fig. 5). Al-

though the analysis of *atpC* gene sequences suggested the robust monophyly of the red algae (*Porphyra* and *Cyanidioschyzon*), with 67–93% bootstrap values (Fig. 4), phylogenetic results for *pgk* suggested the moderate nonmonophyly of the red algae (Fig. 3) and monophyly of the red algae was supported by only the ML method with 50% or more bootstrap values in *aladh* gene phylogeny (Fig. 5).

Phylogenetic analyses of limited eukaryotic OTUs using concatenated protein sequences from four nuclear genes (*psbO*, *pgk*, *aladh*, and *atpC*) demonstrated the robust monophyly of the red (Rhodophyceae, *Cyanidioschyzon*, and Heterokontophyta) and green lineages, with 100% bootstrap values by all three phylogenetic methods. In the red lineage, the red algae were resolved as a robust monophyletic group (with 84–99% bootstrap values) to which the Heterokontophyta was sister.

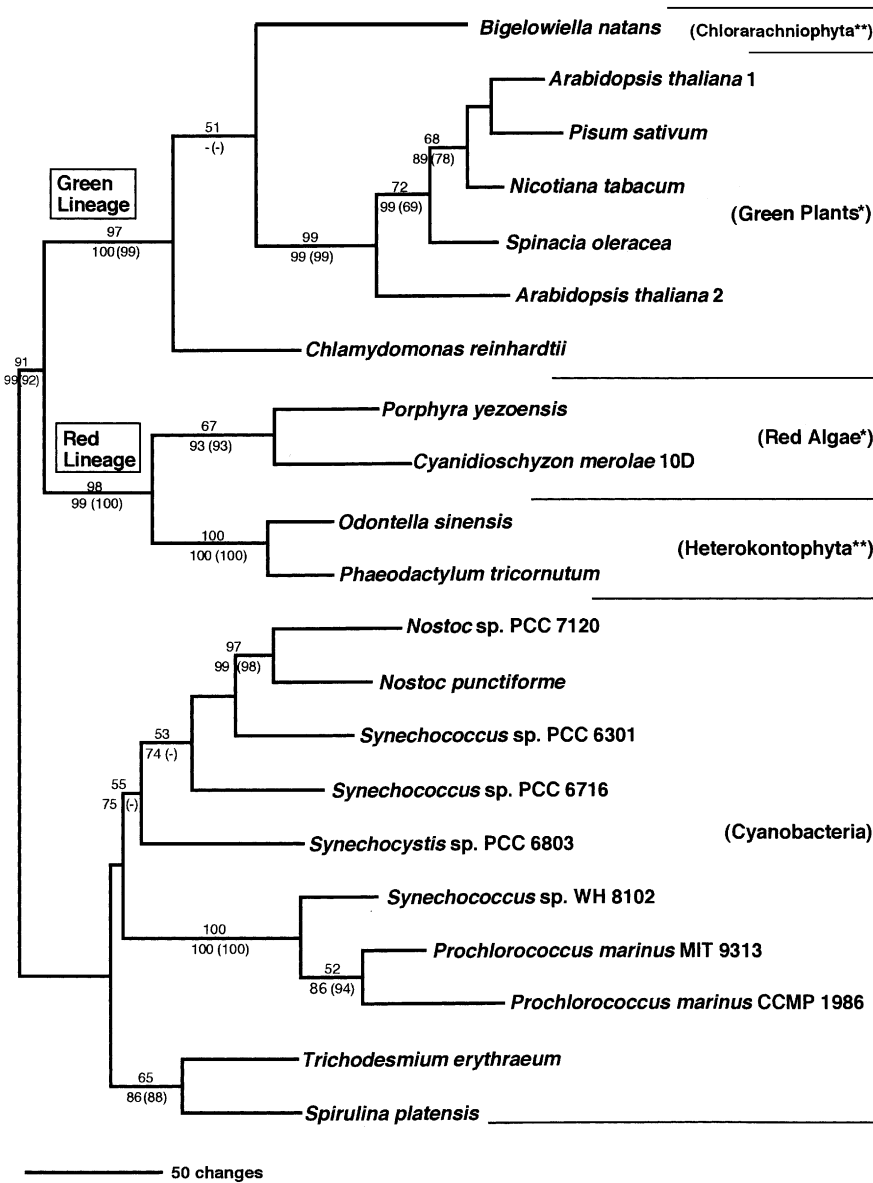


Fig. 4. Single most parsimonious tree (with tree length of 1247 and consistency index of 0.6576) based on 307 amino acid sequences from *atpC* genes of 21 OTUs representing various prokaryotic and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by PAUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and/or maximum likelihood calculation (based on JTT model with rate heterogeneity) are also shown by numbers under the branches without or with parentheses, respectively. Single and double asterisks indicate taxa with primary and secondary plastids, respectively.

From both the PROTML and the AAML analyses, the highest likelihood tree turned out to be Tree 1: (green plants, ((Rhodophyceae, *Cyanidioschyzon*), diatoms), Cyanobacteria) (Table 2). A clade consisting of Rhodophyceae and *Cyanidioschyzon* was supported with 97.8 and 92.7% bootstrap values by the PROTML and AAML analyses, respectively. By contrast, the groupings of diatoms and Rhodophyceae/*Cyanidioschyzon* (Trees 2 and 3; Table 2) were rejected with only 0.2–7.2% bootstrap support by the PROTML and AAML analyses. These groupings were also rejected at the 5% level by the KH and AU tests for the PROTML analysis, although they could not be rejected by the conservative WSH test for the PROTML analysis (highest *P* value was 0.089) and by the three tests for the AAML analysis (highest *P* values were 0.078–0.239) (Table 2).

Discussion

Phylogeny of *gnd* Genes

Andersson and Roger (2002) showed that *gnd* genes with a high cyanobacterial affinity are present in plastid-lacking organisms belonging to the Heterokonta and Heterolobosea (Discicristata), and these genes form a monophyletic group with those from the plastid-containing Apicomplexa, Heterokonta, green plants, and the red alga *Porphyra*. Therefore, two alternative hypotheses have been suggested: either (1) primary plastids might have been introduced into the eukaryotic cells of a common ancestor of the three primary photosynthetic eukaryotic groups as well as Discicristata, Heterokonta, and Alveolata, or (2) secondary plastids might have mediated the *gnd* gene transfer (Andersson and

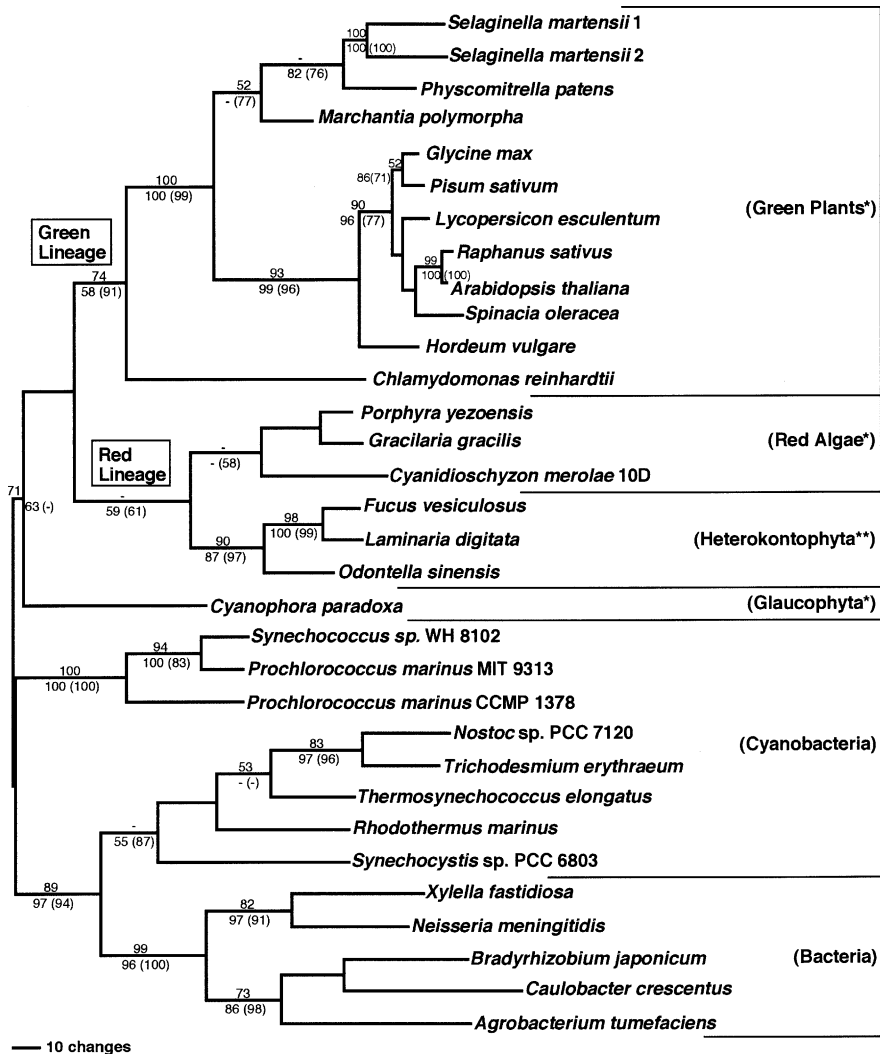


Fig. 5. One of the three most parsimonious trees (with tree length of 1605 and consistency index of 0.5707) based on 310 amino acid sequences from *aladh* genes of 32 OTUs representing various prokaryotic and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by PAUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and/or maximum likelihood calculation (based on JTT model with rate heterogeneity) are also shown by numbers under the branches without or with parentheses, respectively. Single and double asterisks indicate taxa with primary and secondary plastids, respectively.

Roger 2002; Cavalier-Smith 2002). The results of our phylogenetic analyses suggest that monophyly of the extant red algae, as represented by *Porphyra* (Rhodophyceae) and *Cyanidioschyzon* (Cyanidiphyceae), might have resulted from *gnd* gene transfer before the divergence of the extant red algae (Fig. 1). This is consistent with the hypothesis recently proposed by Nozaki et al. (2003a), that primary endosymbiosis might have occurred in a common ancestor of the three primary photosynthetic eukaryotes as well as the organisms lacking primary plastids (Discicristata, Heterokonta, and Alveolata). More recently, Nozaki et al. (2003b) proposed that the *gnd* genes in Heterolobosea (Discicristata) could have resulted from gene transfer in an ancient primary endosymbiosis because their cladistic analysis of plastid gene loss suggested the relatively recent acquisition of the secondary plastid of *Euglena* (after the divergence of Chlorophyceae and Trebouxiophyceae).

In this study, we demonstrated that the *gnd* genes of *Cyanidioschyzon* and *Porphyra*, with a high pro-

teobacterial affinity, form a monophyletic group that is sister to the Opisthokonta in the *gnd* gene phylogeny (Fig. 1). However, such genes have not been reported in other plastid-containing or plastid-lacking organisms (green plants, Heterokonta, Discicristata, and Apicomplexa) belonging to Bikonta (Cavalier-Smith 2003) or Plantae (Nozaki et al. 2003a) (Fig. 1). Therefore, the *gnd* genes with a high proteobacterial affinity might have been lost in these organisms, and this loss might have taken place only once, as suggested by the results of Nozaki et al. (2003a), resolving the red algae as a basal lineage to the monophyletic group composed of other organisms (green plants, Heterokonta, Discicristata, etc.) in Bikonta (Cavalier-Smith 2003) or Plantae (Nozaki et al. 2003a). Alternatively, such genes might have been transmitted to the red algae by the horizontal gene transfer from the common ancestor of the Opisthokonta (Metazoa and Fungi) before the divergence of the extant red algae.

Although the question of which organism is the most basal eukaryote has been the subject of a long-

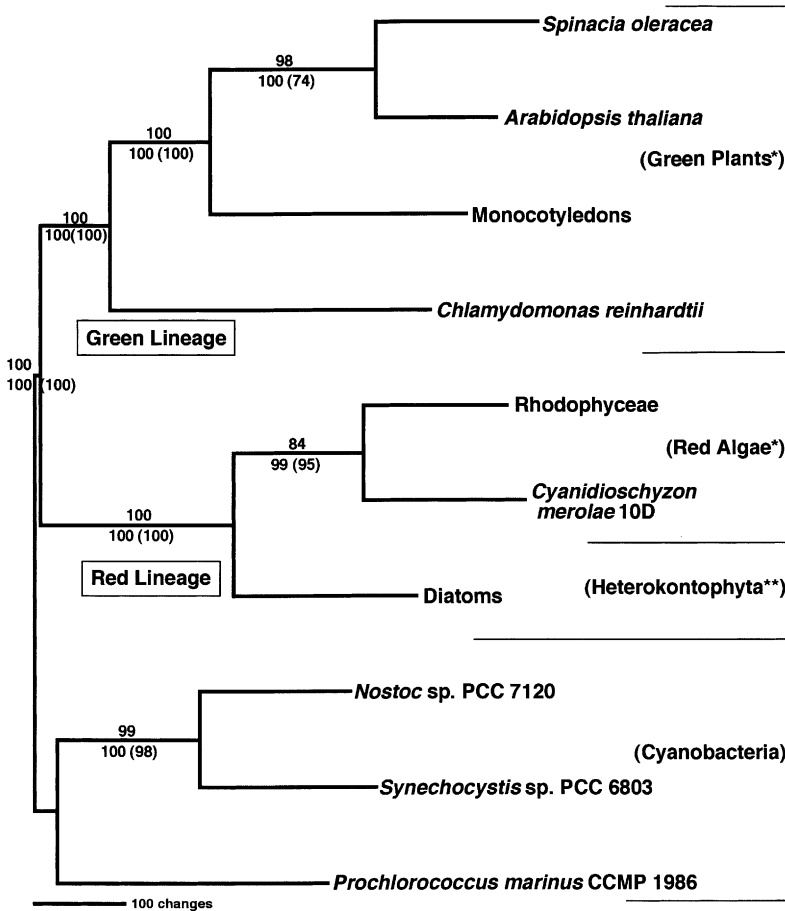


Fig. 6. Single most parsimonious tree (with tree length of 2578 and consistency index of 0.8181) based on concatenated 1217 amino acid sequences from four genes (*psbO*, *pgk*, *aladh*, and *atpC*) of 10 OTUs representing cyanobacterial and eukaryotic taxa. The tree was constructed based on the heuristic search using stepwise addition of 50 random replications (with the TBR branch-swapping algorithm) by PAUP 4.0b10 (Swofford 2002). Branch lengths are proportional to numbers of amino acid substitutions, which are indicated by the scale bar below the tree. Numbers above branches are the bootstrap values (50% or more) based on 1000 replications of the full heuristic analysis (with simple addition sequence). Branches resolved with 50% or more bootstrap values by the neighbor-joining method (based on the Kimura distances) and maximum likelihood analysis (based on JTT model with rate heterogeneity) are also shown by numbers under the branches without and with parentheses, respectively. Single and double asterisks indicate taxa with primary and secondary plastids, respectively.

standing debate, recent phylogenetic studies have suggested two candidates: excavates (including amitochondrial eukaryotes, diplomonads, and parabasalids) and Amoebozoa. Previous phylogenetic studies of eukaryotes based on nuclear protein-coding genes usually included the amitochondrial eukaryotic organisms such as Diplomonadida (e.g., *Giardia lamblia*) and Parabasala (e.g., *Trichomonas*), with the archaeobacterial homologous genes designated as the outgroup (e.g., Hasegawa et al. 1993; Baldauf and Doolittle 1997; Stiller et al. 1998; Moreira et al. 2000; Baptiste et al. 2002). In these studies, amitochondrial organisms are consistently positioned as the most basal group within the eukaryotes (Hasegawa et al. 1993; Baldauf and Doolittle 1997; Stiller et al. 1998; Moreira et al. 2000; Baptiste et al. 2002). However, the basal phylogenetic position of amitochondrial eukaryotes can possibly be considered an artifact, due to the unusual gene substitutions of amitochondrial eukaryotes and the large genetic distance between the Archaeobacteria and Eukaryotes (see Nozaki et al. 2003a). Recently, Cavalier-Smith (2002) and Stechmann and Cavalier-Smith (2002) suggested that the Amoebozoa represent the most basal eukaryotic lineage, based on the diversification of the microtubular cytoskeleton and/or presence/absence of dihydrofo-

late reductase-thymidylate synthase gene fusion. In addition, the present phylogenetic results and the analysis of the *gnd* genes with a high proteobacterial affinity demonstrate the basal position of the cellular slime mold (*Dictyostelium*) to the red algae and Opisthokonta (Fig. 1). The basal position of *Dictyostelium* has also been suggested in some phylogenetic trees based on the presence of recently (after eukaryotic origin) duplicated genes in the nuclear genome. Rivero et al. (2001) showed that 15 Rho (Ras-like small GTPases)-related proteins from *Dictyostelium* hold basal positions to the Metazoa and Fungi, as well as to the green plant *Arabidopsis*. Nozaki et al. (2003a) resolved the cellular slime molds as the most basal mitochondria-containing eukaryotes, based on a paralogous comparison of the concatenated alpha- and beta-tubulin genes of mitochondria-containing organisms.

Phylogenetic Analyses of Other Nuclear Genes

Recent phylogenetic analyses of the plastid multigene sequences have demonstrated that secondary plastids are positioned within the red algal lineages, suggesting secondary endosymbiosis after the divergence of extant red algae (Rhodophyceae and Cyanidiophy-

Table 2. Comparison of alternative trees inferred from maximum likelihood (ML) analyses of concatenated 1217 amino acid sequences from four genes (*psbO*, *pgk*, *aladh* and *atpC*) of cyanobacterial and eukaryotic taxa (see Fig. 6)^a

Tree topology ^b	$\Delta \ln L \pm SE^c$	BT ^d	<i>p</i> value ^e			
			KH	WSH	AU	
	Without Γ^f					
1.	((Green,((Rhodo,Cyani),Diat)),Cyanob)	< -16525.9 >	0.978	0.975	1.000	0.988
2.	((Green,(Rhodo,(Cyani,Diat))),Cyanob)	-22.9 \pm 9.3	0.002	0.009	0.033	0.002
3.	((Green,((Rhodo,Diat),Cyani)),Cyanob)	-19.8 \pm 10.1	0.020	0.025	0.089	0.022
4.	((Green,Diat),(Rhodo,Cyani),Cyanob)	-95.3 \pm 21.5	0	0	0	5e-07
5.	((Green,(Rhodo,Cyani)),Diat),Cyanob)	-101.5 \pm 20.3	0	0	0	6e-08
6.	((Green,(Rhodo,Diat),Cyani),Cyanob)	-127.2 \pm 24.7	0	0	0	1e-41
7.	((Green,Cyani),(Rhodo,Diat),Cyanob)	-127.5 \pm 24.6	0	0	0	4e-49
8.	((Green,Rhodo),(Cyani,Diat),Cyanob)	-139.2 \pm 23.6	0	0	0	2e-06
9.	((Green,(Cyani,Diat),Rhodo),Cyanob)	-139.5 \pm 23.4	0	0	0	2e-07
10.	((Green,Rhodo),Diat),Cyani),Cyanob)	-147.9 \pm 25.5	0	0	0	2e-04
11.	((Green,Diat),Rhodo),Cyani),Cyanob)	-135.8 \pm 26.0	0	0	0	4e-09
12.	((Green,Cyani),Diat),Rhodo),Cyanob)	-147.1 \pm 25.3	0	0	0	4e-44
13.	((Green,Diat),Cyani),Rhodo),Cyanob)	-139.7 \pm 25.3	0	0	0	6e-10
14.	((Green,Rhodo),Cyani),Diat),Cyanob)	-150.1 \pm 24.5	0	0	0	2e-09
15.	((Green,Cyani),Rhodo),Diat),Cyanob)	-143.5 \pm 25.3	0	0	0	7e-81
	With Γ^g					
1.	((Green,((Rhodo,Cyani),Diat)),Cyanob)	< -16032.2 >	0.927	0.922	0.999	0.933
2.	((Green,(Rhodo,(Cyani,Diat))),Cyanob)	-15.2 \pm 7.1	0.002	0.020	0.071	0.013
3.	((Green,((Rhodo,Diat),Cyani)),Cyanob)	-11.7 \pm 8.2	0.072	0.078	0.239	0.084
4.	((Green,Diat),(Rhodo,Cyani),Cyanob)	-74.6 \pm 17.9	0	0	0	2e-32
5.	((Green,(Rhodo,Cyani)),Diat),Cyanob)	-76.3 \pm 17.3	0	0	3e-04	1e-31
6.	((Green,(Rhodo,Diat),Cyani),Cyanob)	-97.7 \pm 21.1	0	0	0	3e-14
7.	((Green,Cyani),(Rhodo,Diat),Cyanob)	-98.3 \pm 20.9	0	0	1e-04	2e-42
8.	((Green,Rhodo),(Cyani,Diat),Cyanob)	-105.0 \pm 19.8	0	0	0	1e-39
9.	((Green,(Cyani,Diat),Rhodo),Cyanob)	-105.0 \pm 19.8	0	0	0	1e-39
10.	((Green,Rhodo),Diat),Cyani),Cyanob)	-110.2 \pm 21.2	0	0	0	1e-32
11.	((Green,Diat),Rhodo),Cyani),Cyanob)	-106.3 \pm 21.5	0	0	0	1e-06
12.	((Green,Cyani),Diat),Rhodo),Cyanob)	-111.2 \pm 20.9	0	0	0	2e-70
13.	((Green,Diat),Cyani),Rhodo),Cyanob)	-108.0 \pm 21.1	0	0	9e-05	3e-54
14.	((Green,Rhodo),Cyani),Diat),Cyanob)	-112.5 \pm 20.5	0	0	3e-05	1e-61
15.	((Green,Cyani),Rhodo),Diat),Cyanob)	-110.8 \pm 20.9	0	0	0	7e-71

^aAll 15 of possible tree topologies of five OTUs are shown. The relationships within green plants and cyanobacteria were fixed as shown in Fig. 6. For details also see text.

^bGreen, green plants; Rhodo, Rhodophyceae; Cyani, *Cyanidioscenzon* (Cyanidiophyceae); Diat, diatoms (Chromista); Cyanob, cyanobacteria (outgroup).

^cDifferences in the log-likelihoods of alternative trees from that of the ML tree (Tree 1) ($\Delta \ln L$) are given \pm 1 SE (estimated by Kishino and Hasegawa's [1989] formula).

^dBootstrap values calculated by the RELL method (Kishino et al. 1990) with 10,000 replications.

^e*p* values of the KH, WSH, and AU tests were estimated by the CONSEL program (Shimodaira and Hasegawa 2001).

^fML analyses without discrete gamma model of site heterogeneity by PROTML of MOLPHY 2.3 (Adachi and Hasegawa 1996).

^gML analyses with discrete Γ model of site heterogeneity by AAML of PAML (Yang 1997). The optimized alpha parameter of the Γ model was optimized for each tree (0.73–0.75).

ceae) (Marin et al. 2002; Maul et al. 2002; Yoon et al. 2002a, b; Ohta et al. 2003). In contrast, the present phylogenetic analyses using nuclear cyanobacterial genes suggest that secondary endosymbiosis in red algae could have taken place before the divergence of the extant red algae, based on the robust monophyly of the extant red algae (Figs. 2 and 4–6, Table 2) and the rejection of Trees 2 and 3 at the 5% level by KH and AU tests for PROTML analysis (Table 2). Itoh et al. (2002) compared the complete genome sequences of two different endosymbionts and demonstrated that the rate of amino acid substitution is twice as high in symbionts as in their relatives, suggesting that the elevated evolutionary rate is mainly

due to an enhanced mutation rate in endosymbionts. The available evidence suggests that secondary plastid genes may carry accelerated and/or unusual gene substitutions compared with those of primary plastids, because the secondary plastids can be considered to be the plastids of reduced eukaryotic endosymbionts (e.g., Delwiche 1999; McFadden 2001). Thus, phylogenetic results of nuclear genes of a high cyanobacterial affinity should be useful for representing natural phylogenetic relationships because such genes in secondary photosynthetic eukaryotes have experienced the same evolutionary environment as those in the primary photosynthetic eukaryotes, even following secondary endosymbiosis, when the

genes should have been transferred from the engulfed photosynthetic eukaryote to the host nucleus. However, the multigene phylogeny presented in this study (Fig. 6) analyzed a limited number of taxa, especially in the red lineages. Analysis of additional OTUs/sequences (especially other members of the Chromista) might better resolve the phylogenetic relationships between plastids within the red lineage.

In contrast to the eukaryotic *gnd* genes, which may have originated directly from cyanobacterium-like ancestors as a result of primary endosymbiosis (see above), transmission of four other genes to the nuclei of the Chromista possibly occurred during secondary endosymbiosis, because the red lineage was found to be monophyletic (Archibald et al. 2003) (Figs. 2–4, 5). Such genes are not recognized in plastid-lacking organisms (Ishida and Green 2002; Archibald et al. 2003). In addition, analysis of these genes demonstrated that secondary endosymbiosis in the red lineage preceded the divergence of the extant red algae (Figs. 2 and 4–6, Table 2). Such ancient secondary endosymbiosis has also been suggested by Fast et al. (2001) and Harper and Keeling (2003). These authors provided evidence that the plastids of Alveolata (Apicomplexa and Dinophyta) and Chromista (Heterokontophyta, Cryptophyta, and Haptophyta) appear to be the result of a single secondary endosymbiotic event on the basis of the distribution and phylogeny of the plastid-targeted glyceraldehyde-3-phosphate dehydrogenase gene. Yoon et al. (2002b), on the basis of the broadly sampled data set and molecular dating of the chloroplast multiple genes, also suggested the single, ancient origin of secondary plastids in the red lineage. In contrast, secondary endosymbiosis in the green lineage (Chlorarachniophyta and Euglenophyta) seems to be a relatively recent event (after the divergence of Chlorophyta and Streptophyta [Nozaki et al. 2003b]) based on phylogenetic analyses of the nuclear-encoded cyanobacterial genes (Archibald et al. 2003) and cladistic analysis of plastid gene losses (Nozaki et al. 2003b). In the *psbO* gene phylogeny, the chlorarachniophyte *Bigeloviella* is robustly resolved as a sister lineage to the Volvocales (*Volvox*, *Chlamydomonas*, and *Dunaleiella*) within the Chlorophyta (Archibald et al. 2003) (Fig. 2). Although MP and NJ analyses of the chloroplast multiple genes robustly resolved the most basal position of the *Euglena* plastid within the green lineage (e.g., Martin et al. 2002; Ohta et al. 2003), the ML method of the same data matrix suggested the close relationship between *Euglena* and *Chlorella* plastids (Martin et al. 2002) and the recent cladistic analysis of plastid gene losses robustly demonstrated the secondary endosymbiosis for the *Euglena* plastid after the divergence of *Chlamydomonas* (Chlorophyceae) and *Chlorella*

(Trebouxiophyceae) within the Chlorophyta (Nozaki et al. 2003b).

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 T.K.), and by the Program for the Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN; to T.K., T.H., and H.N.).

References

- Adachi J, Hasegawa M (1996) Computer Science Monographs, No. 28, MOLPHY Version 2.3: Programs for Molecular Phylogenetics Based on Maximum Likelihood. Institute of Statistical Mathematics, Tokyo
- Archibald JM, Rogers MB, Toop M, Ishida K, Keeling PJ (2003) Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigeloviella natans*. *Proc Natl Acad Sci USA* 100:7678–7683
- 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, Doolittle WF (1997) Origin and evolution of the slime molds (Mycetozoa). *Proc Natl Acad Sci USA* 94:12007–12012
- Bapteste E, Brinkmann H, Lee JA, Moore DV, Sensen CW, Gordon P, Duruflé L, Gaasterland T, Lopez P, Müller M, Philippe H (2002) The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium*, *Entamoeba*, and *Mastigamoeba*. *Proc Natl Acad Sci USA* 99:1414–1419
- 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
- Cavalier-Smith T (2002) Chloroplast evolution: Secondary symbiogenesis and multiple losses. *Curr Biol* 12:R62–R64
- Cavalier-Smith T (2003) Genomic reduction and evolution of novel genetic membranes and protein-targeting machinery in eukaryote-eukaryote chimaeras (meta-algae). *Phil Trans R Soc Lond B Biol Sci* 358:109–133
- Delwiche CF (1999) Tracing the thread of plastid diversity through the tapestry of life. *Am Nat* 154:164–177
- Fast NM, Kissinger JC, Roos DS, Keeling PJ (2001) Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. *Mol Biol Evol* 18:418–426
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using bootstrap. *Evolution* 38:16–24
- Felsenstein J (2002) PHYLIP (Phylogeny Inference Package) version 3.6a3. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle
- Harper JT, Keeling PJ (2003) Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol Biol Evol* 20:1730–1735
- Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993) Early branchings in the evolution of eukaryotes: Ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. *J Mol Evol* 36:380–388
- Ishida K, Green BR (2002) Second- and third-hand chloroplasts in dinoflagellates: phylogeny of oxygen-evolving enhancer 1 (PsbO) protein reveals replacement of a nuclear-encoded plastid gene by that of a haptophyte tertiary endosymbiont. *Proc Natl Acad Sci USA* 99:9294–9299

- 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
- Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *CABIOS* 8:275–282
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press, Cambridge
- Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J Mol Evol* 29:170–179
- Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. *J Mol Evol* 31:151–160
- Martin W, Rujan T, Richly E, Hansen A, Cornelsen S, Lins T, Leister D, Stoebe B, Hasegawa M, Penny D (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
- Moreira D, Le Guyader H, Philippe H (2000) The origin of red algae and the evolution of chloroplasts. *Nature* 405:69–72
- Nikaïdo I, Asamizu E, Nakajima M, Nakamura Y, Saga N, Tabata S (2000) Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, *Porphyra yezoensis*. *DNA Res* 7:223–227
- Nozaki H, Matsuzaki M, Takahara M, Misumil O, Kuroiwa H, Hasegawa M, Shin-i T, Kohara Y, Ogasawara N, Kuroiwa T (2003a) 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
- Nozaki H, Ohta N, Matsuzaki M, Misumi O, Kuroiwa T (2003b) Phylogeny of plastids based on cladistic analysis of gene loss inferred from complete plastid genome sequences. *J Mol Evol* 57:377–382
- Ohta N, Matsuzaki M, Misumi O, Miyagishima S, Nozaki H, Tanaka K, Shin-i T, Kohara Y, Kuroiwa T (2003) Analyses of the complete nucleotide sequence of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*: Genome structure, gene organization, and gene content. *DNA Res* 10:67–77
- Rivero F, Dislich H, Glockner G, Noegel AA (2001) The *Dictyostelium discoideum* family of Rho-related proteins. *Nucleic Acids Res* 29:1068–1079
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114–1116
- Shimodaira H, Hasegawa M (2001) CONSEL: For assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17:1246–1247
- Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297:89–91
- Stiller JW, Duffield ECS, Hall BD (1998) Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. *Proc Natl Acad Sci USA* 95:11769–11774
- Stiller JW, Riley J, Hall BD (2001) Are red algae plants? A critical evaluation of three key molecular data sets. *J Mol Evol* 52:527–539
- Strimmer K, Von Haeseler A (1996) Quartet puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. *Mol Biol Evol* 13:964–969
- Swofford DL (2002) PAUP* 4.0: Phylogenetic analysis using parsimony, version 4.0b10. Computer program distributed by Sinauer Associates, Fitchburg
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Yang Z (1997) PAML: A program package for phylogenetic analysis by maximum likelihood. *CABIOS* 13:555–556
- 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