Phylogeny of Primary Photosynthetic Eukaryotes as Deduced from Slowly Evolving Nuclear Genes

Hisayoshi Nozaki, Mineo Iseki,† Masami Hasegawa,‡[1] Kazuharu Misawa,* Takashi Nakada,* Narie Sasaki,§[2] and Masakatsu Watanabe‖*

*Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan; †Hayama Center for Advanced Studies, Graduate University for Advanced Studies, Kanagawa, Japan; ‡The Institute of Statistical Mathematics, Tokyo, Japan; §Division of Life Science, Graduate School of Humanities and Sciences, Ochanomizu University, Tokyo, Japan; and ‖Department of Evolutionary Studies of Biosystems, School of Advanced Sciences, Graduate University for Advanced Studies, Kanagawa, Japan

Introduction

The biodiversity of photosynthetic eukaryotes, traditionally recognized as nine algal divisions or phyla, is attributed to two kinds of endosymbiotic events involving plastids: primary endosymbiosis and secondary endosymbiosis. Therefore, the phylogenetic positions of primary photosynthetic eukaryotes are fundamental for understanding the evolution of eukaryotic cells and establishing higher taxonomic concepts of eukaryotes. Recently, Rodríguez-Ezpeleta et al. (2005) demonstrated the strong monophyly of the three groups of primary photosynthetic eukaryotes (green plants, glaucophytes, and red algae) based on 143 nuclear genes. However, they analyzed only two divisions of the secondary phototrophs belonging to the Stramenopiles–Alveolata (SA) lineage, and their 143 genes included rapidly evolving genes. Here, we reexamine the phylogeny of the primary phototrophs based on slowly evolving nuclear genes selected mainly from the data matrix of Rodríguez-Ezpeleta et al. (2005), using additional operational taxonomic units (OTUs) of a free-living, secondary phototrophic group (Haptophyta) and of Excavata (Heterolobosea and *Reclinomonas*) that do not belong to the SA lineage. Our phylogenetic results demonstrate the robust non-monophyly of the primary phototrophs and the basal position of red algae within the bikonts, suggesting the loss of plastids in certain eukaryotic lineages under the assumption of the single plastid primary endosymbiosis.

Results and Discussion

Maximum parsimony (MP) (with 84% bootstrap values [BT]) and Bayesian inference (BI) using the WAG+I+Γmodel (with 0.99 posterior probabilities [PP]) (see Supplemental Methods) based on the 5216 × 31 matrix (see *Methods*) robustly resolved the red algae as the most basal lineage within the bikonts sensu Cavalier-Smith (2003) or Plantae sensu Nozaki et al. (2003) (three groups of primary photosynthetic eukaryotes, SA lineage, and the Haptophyta; fig. 1A). In the maximum likelihood (ML) analyses (using PhyML and Proml of PHYLIP; see Supplemental Methods), however, the basal position of the red algae had only weak BT (54%–55%). This weak support may have resulted from the large amount of missing data on the glaucophyte OTUs (21%–32% of 5,216 amino acid positions versus 8.3% in total [5216 × 33 matrix]) because the ML analysis excluding these two glaucophyte species (5216 × 31-2GL matrix) showed increased support (80%–90% BT) for the most basal position of the red algae within the bikonts (fig. 1B), and addition of *Glaucocystis* (with 32% data missing) markedly reduced the support, especially in the ML analyses (Table S2 in Supplemental Material). In the second data matrix (5216 × 33 matrix; *see Methods*) that includes two OTUs of Excavata, the MP analyses and BI with the WAG+I+Γmodel, with relatively high supports (with 82% BT and 0.99 PP, respectively), resolved the most basal position of the red algae plus Excavata within the bikonts (fig. 2A), whereas the ML analyses did not resolve their basal position with 50% or more BT. However, the ML calculations excluding these two glaucophyte OTUs (5216 × 33-2GL matrix) showed increased support (51%–87% BT) for the most basal position of the red algae plus Excavata within the bikonts (fig. 2B). Bayesian inference based on the CAT+Γmodel also supports the most basal position of the red algae or red algae plus Excavata within the bikonts, with 1.00 PP (5216 × 31-2GL and 5216 × 33-2GL matrices), 0.90 PP (5216 × 31 matrix), or 0.65 PP (5215 × 33 matrix).

The highest likelihood trees in the exhaustive ML analyses of the 5216 × 31-2GL and 5216 × 33-2GL matrices favored polyphyletic relationships for primary photosynthetic eukaryotes (Tables S3 and S4 in Supplemental Material). The most basal group within the bikonts was composed of the red algae or the red algae plus Excavata, supported with 95% or 88% BT, using the 5216 × 31-2GL or 5216 × 33-2GL matrices, respectively (figs. 1B, 2B). In the 5216 × 31-2GL matrix, the grouping of green plants with red algae was not rejected at the 5% level by the AU, KH, or WSH test (Table S3). However, this grouping was rejected at the 5% or 1% level , respectively, by the AU or KH test in the 5216 × 33-2GL matrix (table S4). In addition, all seven trees that were not rejected by both the AU and the KH test at the 5% level (Trees 1–5, 7, and 8; Table S4) resolved that the red algae or red algae plus Excavata constitute the most basal lineage within the bikonts.

Based on the very conserved nuclear genes (actin, elongation factor one alpha [EF-1α], α-tubulin, and β-tubulin), the basal phylogenetic position of the red algae within the bikonts was resolved robustly (Nozaki et al. 2003; Nozaki 2005). This phylogenetic result may have arisen from the possible relaxation of the unusually high substitution rates

[1] Present address: School of Life Sciences, Fudan University, Shanghai 200433, China.

[2] Present address: Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, Aichi 464-8602, Japan.

Key words: eukaryote evolution, long branch attraction, phylogeny, plastid endosymbiosis, primary photosynthetic eukaryotes, taxon sampling.

E-mail: nozaki@biol.s.u-tokyo.ac.jp.

Mol. Biol. Evol. 24(8):1592–1595. 2007
doi:10.1093/molbev/msm091
Advance Access publication May 7, 2007

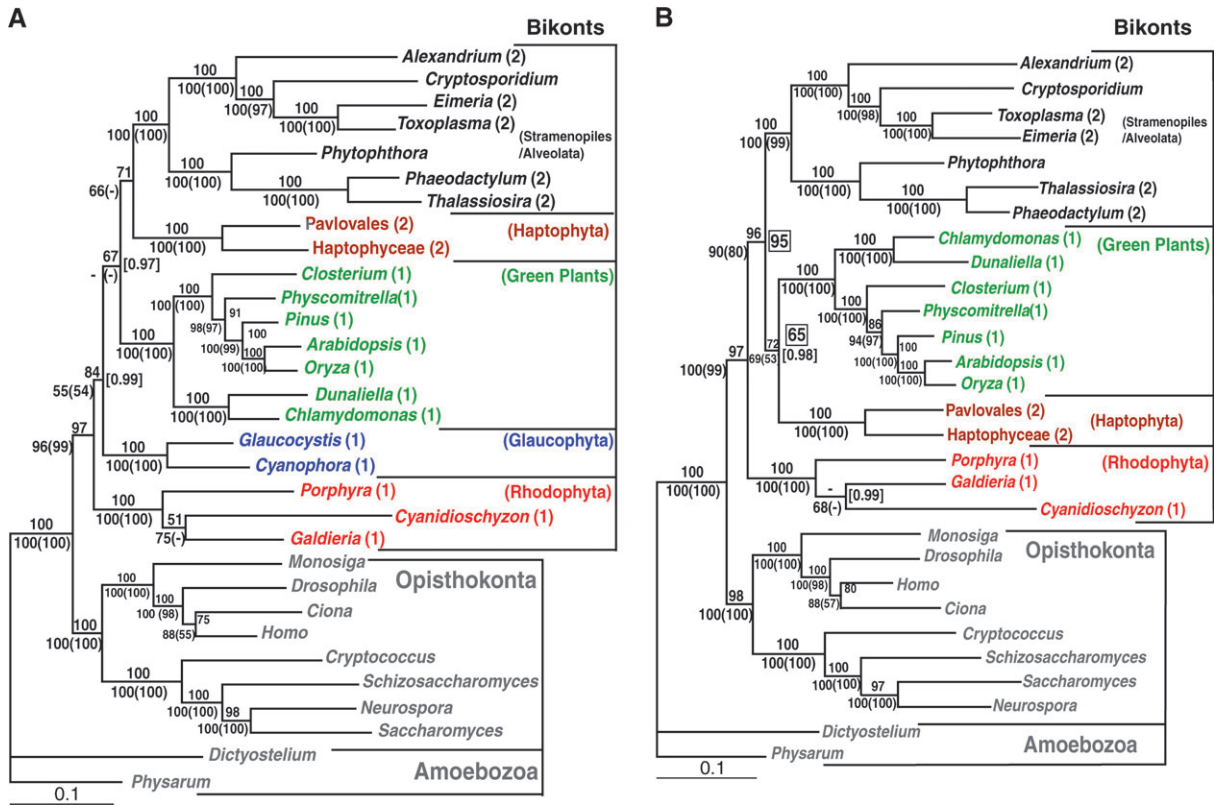

FIG. 1.—Phylogeny based on nuclear encoded protein sequences including Haptophyta, using a 5216 × 31 matrix (31 OTUs) (*A*) and a 5216 ×31-2GL matrix (29 OTUs; excluding glaucophytes) (*B*). The analysis is based on the concatenated data set of slowly evolving nuclear proteins (19 proteins; 5,216 amino acid positions). The tree has been inferred with Bayesian inference with the WAG+I+Γmodel. Posterior probabilities (PP) for all branches are 1.00 except for branches with PP <1.00 (within brackets). Numbers above the branches represent bootstrap values (BT; ≥50%) by maximum parsimony analysis (1,000 replicates). Numbers without or within parentheses below the branches represent BT ≥50% obtained with 1,000 replicates of maximum likelihood (ML) analysis with PhyML (WAG+I+Γmodel) or Proml of PHYLIP (JTT+I+Γ model with global rearrangements), respectively. Numbers in squares show BT (10,000 replicates) calculated by the RELL method in the exhaustive ML analysis (JTT-F+Γ model). Numbers in parentheses just after the species names show possession of primary (1) or secondary (2) plastids. For details, see Supplemental Material.

of the α- and β-tubulin genes in eukaryotes lacking flagellae (e.g., red algae, *Dictyostelium*). However, excluding these two genes, our slowly evolving gene sequences still robustly resolved the basal position of the red algae or the red algae plus Excavata within the bikonts. In addition, the present data matrix including Excavata sequences strongly rejected the monophyly between green plants and red algae in the AU and KH tests. Therefore, the strong monophyly of the three groups of primary phototrophs (Rodríguez-Ezpeleta et al. 2005) may have been due to long branch attraction (LBA) between the Opisthokonta/Amoebozoa and the SA lineage based on the fast evolving genes within the 143 genes. The SA lineage consists mainly of parasites (apicomplexans) and a ciliate (Rodríguez-Ezpeleta et al. 2005), which might have high amino acid substitutions or saturation, especially in fast-evolving genes, as a result of parasitism (Musto et al. 1999; Castro, Austin, and Dowton 2002) and atypical transcription/translation (Brunk 1986; Lozupone, Knight, and Landweber 2001).

Under the assumption of a single event of plastid primary endosymbiosis (Matsuzaki et al. 2004; Rodríguez-Ezpeleta et al. 2005; for an alternative viewpoint, see Stiller, Reel, and Johnson 2003), the nonmonophyly of the primary phototrophs suggested here may be explained by the ancient primary endosymbiosis and the subsequent loss of the primary plastids in the primary plastid-lacking organisms within the bikonts (Nozaki et al. 2003; Nozaki 2005). This hypothesis may also be suggested based on the presence of cyanobacterial or plant-like genes in the nuclei of the plastid-lacking bikonts (Andersson and Roger 2002; Nozaki et al. 2003; Nozaki 2005). In any case, further phylogenetic analyses including other lineages of secondary photosynthetic eukaryotes and related nonparasitic eukaryotes are needed to resolve the correct and reliable evolutionary history of the primary plastids.

Methods

As multigene analyses are expected to be increasingly sensitive to LBA, improved taxon sampling and the selection of positions or genes that evolve more slowly have been suggested for resolving deep branching in phylogenies (Philippe and Laurent 1998; Philippe, Lartillot, and Brinkmann 2005). In addition, we avoided a single OTU in each of the major lineages within the phylogenetic tree. Therefore, we analyzed only 19 slowly evolving genes, and used six additional OTUs from Haptophyta (Haptophyceae and *Pavlova*), Excavata (Heterolobosea [*Naegleria* and

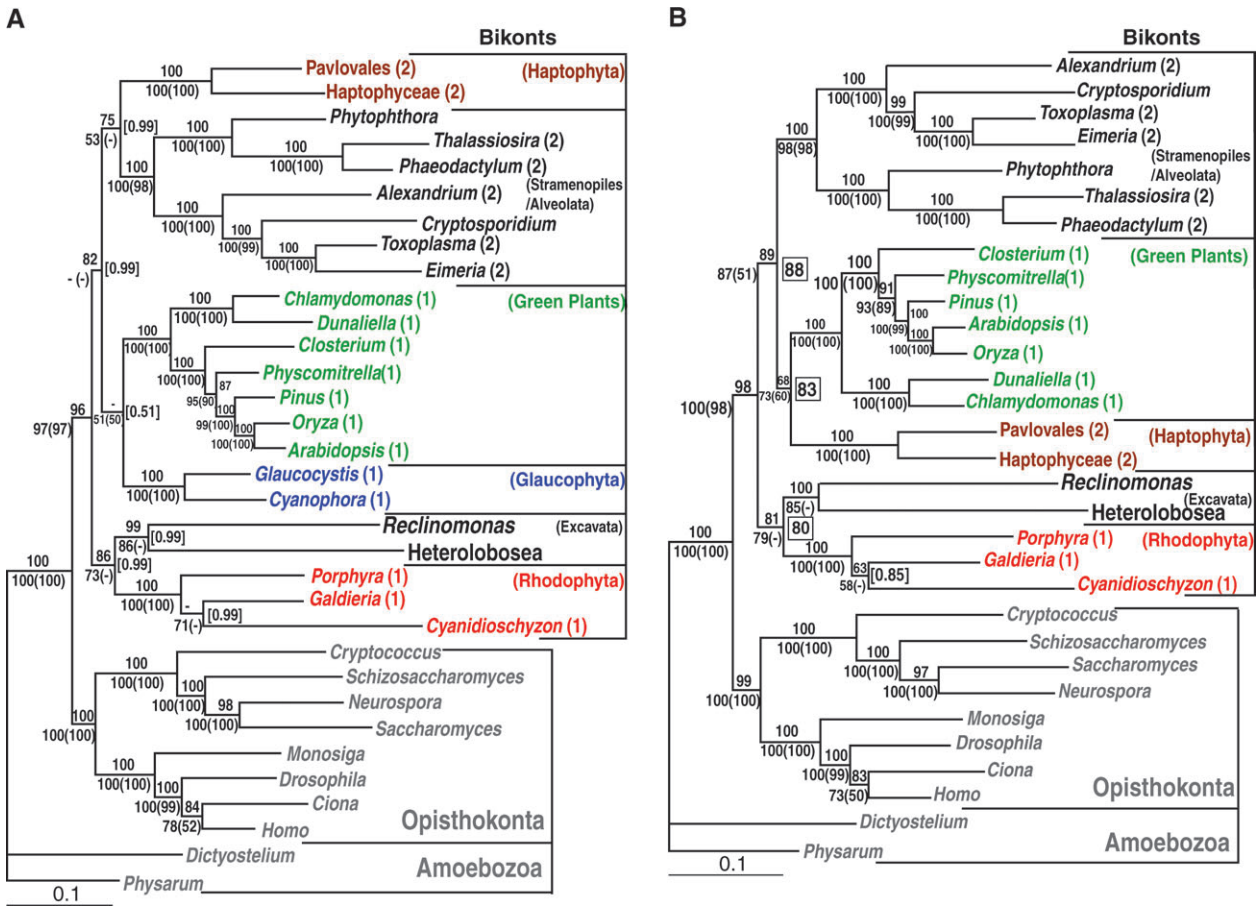

FIG. 2.—Phylogeny based on nuclear encoded protein sequences including Haptophyta and Excavata, using the 5216 ×33 matrix (33 operational taxonomic units [OTUs]) (*A*) and the 5216 ×33-2GL matrix (31 OTUs; excluding glaucophytes) (*B*). For details, see fig. 1.

Sawyeria] and *Reclinomonas*), the red alga *Galdieria*, and the amoeba *Physarum*. The 19 genes used in this study lack complete deletion of a gene in *Physarum* and both two-glaucophyte OTUs, and their *p*-distances do not exceed 0.4 in pairwise distances (based on saturation curves of the distance-correction methods [Philippe and Laurent 1998]) for any combination of OTU for each gene, except for a combination (99 amino acids) between *Dictyostelium* and *Toxoplasma rps17* genes (*p*-distance = 0.40404), and a short alignment (39 amino acids) between the *Cyanophora* and *Physarum nsf1-I* genes (*p*-distance = 0.46154), as well as 16 combinations (*p*-distance $\leq$ 0.43443) in *rpl2*, *rpl27* and *pls3* genes related to the Excavata. Thus, two data matrices without and with the Excavata OTUs were analyzed in this study: the "5216 × 31 matrix" consisting of 5,216 amino-acid sequences (19 genes) from 31 OTUs (excluding Excavata) and the "5216 × 33 matrix" including two OTUs of Excavata. Eighteen genes were selected from the 143 genes of Rodríguez-Ezpeleta et al. (2005) (see Supplementary Material), and the remaining gene was *hsp*90, which has been widely used to determine the macrophylogeny of eukaryotes in other studies (e.g., Harper, Waanders, and Keeling 2005). Because α- and β-tubulin sequences might be relaxed in eukaryotes lacking flagella (e.g., red algae, *Dictyostelium*), and because EF-2 protein sequences might contain unusual phylogenetic information (Stiller, Riley, and Hall 2001), we did not use these three genes. The OTUs analyzed here were the same as those of Rodríguez-Ezpeleta et al. (2005), except for the six additional OTUs (see above), and the exclusion of seven OTUs: *Tetrahymena* having atypical transcription and translation in gene expression (Brunk 1986; Lozupone, Knight, and Landweber 2001), and six OTUs (*Babesia*, *Hydra*, *Phaenerochaete*, *Plasmodium*, *Theileria*, and *Ustilago*) from the Opisthokonta and Alveolata based on their deletion in sequences and/or high substitutions. Because the glaucophyte sequences contained the large amount of missing data (21%–32% of 5,216 amino acid positions), and because such gaps seemed to reduce the phylogenetic resolution (Table S2), two data matrices excluding the Glaucophyta (5216 × 31-2GL matrix [5216 × 31 matrix excluding glaucophytes] and the 5216 × 33-2GL matrix [5216 × 33 matrix excluding glaucophytes]) were also analyzed in this study.

Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are grateful to Dr. N. Rodríguez-Ezpeleta (Université de Montréal, Canada), who kindly provided the alignments of the 143 nuclear proteins. Computation time was provided by the Super Computer System, Human Genome Center, Institute of Medical Science, University of Tokyo. This work was supported by a Grant-in-Aid for Creative Scientific Research (No. 16GS0304) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Literature Cited

Andersson JO, Roger AJ. 2002. A cyanobacterial gene in nonphotosynthetic protists—an early chloroplast acquisition in eukaryotes? Curr Biol. 12:115–119.

Brunk CF. 1986. Genome reorganization in *Tetrahymena*. Int Rev Cytol. 99:49–83.

Castro LR, Austin AD, Dowton M. 2002. Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. Mol Biol Evol. 19:1100–1113.

Cavalier-Smith T. 2003. The excavate protozoan phyla Metamonada Grassé emend. (Anaeromonadea, Parabasalia, *Carpediemonas*, Eopharyngia) and Loukozoa emend. (Jakobea, *Malawimonas*): their evolutionary affinities and new higher taxa. Int J Syst Evol Microbiol. 53:1741–1758.

Harper JT, Waanders E, Keeling PJ. 2005. On the monophyly of chromalveolates using a six-protein phylogeny of eukaryotes. Int J Syst Evol Microbiol. 55:487–496.

Lozupone CA, Knight RD, Landweber LF. 2001. The molecular basis of nuclear genetic code change in ciliates. Curr Biol. 11:65–74.

Matsuzaki M, Misumi O, Shin-i T, et al. (42 co-authors). 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature. 428:653–657.

Musto H, Romero H, Zavala A, Jabbari K, Bernardi G. 1999. Synonymous codon choices in the extremely GC-poor genome of *Plasmodium falciparum*: compositional constraints and translational selection. J Mol Evol. 49:27–35.

Nozaki H. 2005. A new scenario of plastid evolution: plastid primary endosymbiosis before the divergence of the "Plantae," emended. J Plant Res. 118:247–255.

Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, Hasegawa M, Shin-i T, Kohara Y, Ogasawara N, Kuroiwa T. 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.

Philippe H, Lartillot N, Brinkmann H. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol Biol Evol. 22:1246–1253.

Philippe H, Laurent J. 1998. How good are deep phylogenetic trees? Curr Opin Genet Dev. 8:616–623.

Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Loffelhardt W, Bohnert HJ, Philippe H, Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Curr Biol. 15:1325–1330.

Stiller JW, Reel DC, Johnson JC. 2003. A single origin of plastids revisited: convergent evolution in organellar genome content. J. Phycol. 39:95–105.

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.

Takashi Nakada, Associate Editor

Accepted May 2, 2007