Phylogenomic Analysis Supports the Monophyly of Cryptophytes and Haptophytes and the Association of Rhizaria with Chromalveolates

Jeremiah D. Hackett,†[1,2] Hwan Su Yoon,*[2,3] Shenglan Li,* Adrian Reyes-Prieto,* Susanne E. Rümmele,* and Debashish Bhattacharya**

*Department of Biological Sciences and Roy J. Carver Center for Comparative Genomics, University of Iowa; and †Woods Hole Oceanographic Institution, Woods Hole, Massachusetts

Here we use phylogenomics with expressed sequence tag (EST) data from the ecologically important coccolithophore-forming alga *Emiliania huxleyi* and the plastid-lacking cryptophyte *Goniomonas* cf. *pacifica* to establish their phylogenetic positions in the eukaryotic tree. Haptophytes and cryptophytes are members of the putative eukaryotic supergroup Chromalveolata (chromists [cryptophytes, haptophytes, stramenopiles] and alveolates [apicomplexans, ciliates, and dinoflagellates]). The chromalveolates are postulated to be monophyletic on the basis of plastid pigmentation in photosynthetic members, plastid gene and genome relationships, nuclear "host" phylogenies of some chromalveolate lineages, unique gene duplication and replacements shared by these taxa, and the evolutionary history of components of the plastid import and translocation systems. However the phylogenetic position of cryptophytes and haptophytes and the monophyly of chromalveolates as a whole remain to be substantiated. Here we assess chromalveolate monophyly using a multigene dataset of nuclear genes that includes members of all 6 eukaryotic supergroups. An automated phylogenomics pipeline followed by targeted database searches was used to assemble a 16-protein dataset (6,735 aa) from 46 taxa for tree inference. Maximum likelihood and Bayesian analyses of these data support the monophyly of haptophytes and cryptophytes. This relationship is consistent with a gene replacement via horizontal gene transfer of plastid-encoded *rpl36* that is uniquely shared by these taxa. The haptophytes + cryptophytes are sister to a clade that includes all other chromalveolates and, surprisingly, two members of the Rhizaria, *Reticulomyxa filosa* and *Bigelowiella natans*. The association of the two Rhizaria with chromalveolates is supported by the approximately unbiased (AU)-test and when the fastest evolving amino acid sites are removed from the 16-protein alignment.

Introduction

Multigene phylogenetics and the availability of genome data from protist lineages have provided a major impetus to resolving the eukaryotic tree of life (e.g., Simpson and Roger 2004; Bhattacharya and Katz 2005; Parfrey et al. 2006), leading recently to the hypothesis that eukaryotes comprise 6 putative supergroups (reviewed by Adl et al. 2005; Keeling et al. 2005; Parfrey et al. 2006). These are the protistan roots of all multicellular eukaryotes and comprise the Opisthokonta (e.g., animals, fungi, choanoflagellates), Amoebozoa (e.g., lobose amoebae, slime molds), Archaeplastida or Plantae (red, green [including land plants], and glaucophyte algae), Chromalveolata (e.g., apicomplexans, ciliates, giant kelps), Rhizaria (e.g., cercomonads, foraminifera), and Excavata (e.g., diplomonads, parabasalids). Although within supergroup phylogeny and membership are presently unsettled, in particular for excavates and chromalveolates (Parfrey et al. 2006; Simpson, Inagaki, and Roger 2006), the 6 lineages are a useful tool for testing hypotheses about eukaryote phylogeny and biodiversity. The future addition to phylogenies of uncultured environmental samples or poorly studied organisms such as amoebae or heterotrophic flagellates may expand the number of eukaryotic supergroups or alter their interrelationships. In this study, we generated extensive (6,488 unigenes) expressed sequence tag (EST) data from *Emiliania huxleyi* (Haptophyta) and 496 unique ESTs from the early diverging plastid-lacking *Goniomonas* cf. *pacifica* (Cryptophyta) (Hoef-Emden, Marin, and Melkonian 2002), both of which are chromalveolate protists. *E. huxleyi* is of tremendous importance worldwide because of its dominance in oceanic open waters (Brown and Yoder 1994), its role in cloud production through dimethylsulfide (DMS) release, the effect of *E. huxleyi* "algal blooms" on the optical quality and temperature of oceanic waters, and its role as a major carbon sink in our planet (Buitenhuis et al. 1996).

Studying *E. huxleyi* and *G. pacifica* also advances our understanding of the chromalveolates, a controversial supergroup whose constituent members are yet to be shown to comprise a monophyletic lineage in host (i.e., nuclear) phylogenies. The chromalveolate hypothesis was proposed by Cavalier-Smith (1999) to unite all chlorophyll *c*–containing algae and their nonphotosynthetic sister groups. The chromalveolates originally comprised the chromist (cryptophytes, haptophytes, and stramenopiles; Cavalier-Smith 1986) and alveolate (dinoflagellates, ciliates, and apicomplexans; e.g., Van de Peer and De Wachter 1997) protists. Recent phylogenetic analyses, however, identify two more groups with chromalveolate affinities. The first is the plastid-lacking katablepharid protists that appear as sister to cryptophytes in an analysis of small subunit rDNA (Okamoto and Inouye 2005). The second are telonemids. Analysis of a concatenated dataset of Hsp90 and small subunit rDNA positioned with significant Bayesian support *Telonemia* spp. as sister to cryptophytes in a clade that also included haptophytes (Shalchian-Tabrizi et al. 2006a). Telonemids are widely distributed marine heterotrophic protists that are morphologically distinct from katablepharids. These taxa share characters with different chromalveolates

[1] Present address: Department of Ecology and Evolutionary Biology, The University of Arizona, Tucson, AZ 85721, USA.

[2] These authors contributed equally to the manuscript.

[3] Present address: Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, ME 04575, USA.

Key words: chromalveolates, endosymbiosis, *Emiliania huxleyi*, endosymbiotic gene transfer, *Goniomonas* cf. *pacifica*, haptophyte.

E-mail: debashi-bhattacharya@uiowa.edu.

Mol. Biol. Evol. 24(8):1702–1713. 2007
doi:10.1093/molbev/msm089
Advance Access publication May 7, 2007

© The Author 2007. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

including tubular mitochondrial cristae common with alveolates and some chromists and flagellar hairs common with haptophytes and stramenopiles. However cortical alveoli, which are shared not only with alveolates and some stramenopiles but also with glaucophytes in the Plantae, are also found in *T. antarcticum* (Shalchian-Tabrizi et al. 2006a).

The key feature for originally merging chromalveolates is their photosynthetic organelle (plastid [when present]) that is believed to have arisen from a single secondary (red algal) endosymbiosis. Under this hypothesis, some time after its split from other members of the Plantae a red alga was engulfed by the nonphotosynthetic ancestor of the chromalveolates (Bhattacharya, Yoon, and Hackett 2004) and reduced to a 4-membrane-bound plastid that uniquely contained chlorophyll *c*. In chromists, the 4-membrane-bound plastid was retained in all groups except presumably katablepharids and telonemids (Okamoto and Inouye 2005; Shalchian-Tabrizi et al. 2006a). In the alveolates, the dinoflagellates lost one plastid membrane, the apicomplexans lost photosynthesis and retain a remnant organelle (apicoplast) used for other plastid functions such as fatty acid and heme synthesis (Foth and McFadden 2003), and ciliates have apparently lost the plastid. Support for the chromalveolate hypothesis comes from nuclear gene trees that unite stramenopiles and alveolates (Gajadhar et al. 1991; Baldauf et al. 2000; McEwan and Keeling 2004; Nozaki et al. 2004), plastid characters (e.g., storage products) and gene and genome data from this organelle (Yoon et al. 2002, 2004, 2005, 2006a; Bachvaroff, Sanchez Puerta, and Delwiche 2005), the evolutionary history of host-derived plastid-targeted translocators (Weber, Linka, and Bhattacharya 2006), and unique gene duplication and replacement events shared by these taxa (Fast et al. 2001; Harper and Keeling 2003; Patron, Rogers, and Keeling 2004). All chromalveolates except cryptophytes contain a mitochondrion with tubular cristae. Cryptophytes were thought to be ancestral within the chromists, due to the presence of two distinguishing characters, flattened mitochondrial cristae and the remnant red algal nucleus (the nucleomorph; Greenwood 1974) that is found between the second and third plastid membranes Cavalier-Smith 1999). The sister-group relationship of katablepharids and/or telonemids to cryptophytes leaves unclear the ancestral set of chromist characters.

Despite many attempts, the phylogenetic position of cryptophytes and haptophytes remains unresolved. Previous analyses of nuclear genes (often with extensive taxon sampling) show haptophytes to be unaffiliated with any other eukaryotic lineage (e.g., rDNA [Bhattacharya et al. 1995; Van de Peer and De Wachter 1997; Van de Peer et al. 2000], actin [Stibitz, Keeling, and Bhattacharya 2000], HSP90 [Stechmann and Cavalier-Smith 2003]). In contrast, rDNA trees provide moderate support for the grouping of cryptophyte with glaucophyte algae (e.g., Bhattacharya et al. 1995; Van de Peer and De Wachter 1997). The sister-group relationship of cryptophytes and glaucophytes is supported by the presence of flattened cristae in both groups. However a recent analysis using a 143-protein data set supported robustly the monophyly of Plantae and a specific relationship between glaucophytes and green algae/land plants (Rodriguez-Ezpeleta et al. 2005; see also Petersen et al. 2006; Hackett et al. 2007), although haptophytes and cryptophytes were missing from this data set. These results place glaucophytes clearly within Plantae but do not address the position of cryptophytes in the eukaryotic tree. Another phylogenetic treatment (Harper, Waanders, and Keeling 2005) used a HSP90 data set and a concatenated alignment of 6 proteins (actin, alpha- and beta-tubulin, eEF-1alpha, HSP70, and HSP90) to test chromalveolate monophyly. The single-protein HSP90 and 4-protein (actin, alpha- and beta-tubulin, and HSP90) data sets provided weak to moderate bootstrap support for the monophyly of cryptophytes and haptophytes using a variety of phylogenetic approaches. However, bootstrap support for this clade was markedly lowered with the use of the 6-protein data set, and none of the analyses provided evidence for the union of all chromalveolates, suggesting polyphyly of this group (Harper, Waanders, and Keeling 2005). And finally, a recent analysis of plastid genes aimed at determining the extent of horizontal gene transfer (HGT) within the plastid genome uncovered a ribosomal protein gene (*rpl36*) of foreign (likely eubacterial) origin that is uniquely shared by cryptophytes and haptophytes, providing strong evidence for the monophyly of these taxa (Rice and Palmer 2006). Here we assessed the phylogenetic position of haptophytes and cryptophytes using phylogenomics with the EST sequences from *E. huxleyi* and *G. pacifica*. These analyses included complete genome or EST data from other eukaryotes and prokaryotes available in public databases. Phylogenetic analyses of a concatenated 16-protein data set support the monophyly of cryptophytes and haptophytes and their sister-group relationship to a well supported assemblage that includes other chromalveolates and two members of the Rhizaria.

Materials and Methods

Construction of cDNA Libraries and Annotation of ESTs

Construction of the *Emilania huxleyi* (CCMP 371) cDNA library is described in Li et al. (2006). For this paper, we added EST data from the sequencing of a subtracted library from this species that was not presented in Li et al. (2006). The EST sequences from this species are deposited in the dbEST database of GenBank. Total RNA from *Goniomonas* cf. *pacifica* (CCMP 1869) was prepared as in Li et al. (2006), and the cDNA library was constructed using the Stratagene pBluescript II XR cDNA library construction kit (Stratagene), which contains *Xho*I and *Eco*RI digested pBluescript II SK (+) vector. Because the nonphotosynthetic *G. pacifica* was mass cultured (8 L) with bacteria and rice, algal growth (despite the large culture volume) was extremely slow and likely compromised by bacterial competition. For this reason, we initiated library construction with several nanograms of poly(A+) mRNA rather than the 5 μg prescribed by the Stratagene protocol. The low yield of RNA also made difficult efficient size selection for the libraries. However using the available starting material, a starter library was constructed and normalized as described above. The ESTs were sequenced on an ABI 3730 96-channel capillary DNA sequencer (The Roy J. Carver Center for Comparative Genomics, University of Iowa). All ESTs were processed using Phred

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

(Ewing and Green 1998; Ewing et al. 1998) and Crossmatch. A unigene set of 496 cDNAs was identified using UIcluster 3.05 from a total of 1,233 ESTs. These EST have been released to GenBank.

The unique sequences were annotated using the AutoFACT program and the uniref100, nr, cog, KEGG, pfam, and Smart databases with a bit-score cutoff of 50 (Koski et al. 2005). The ESTs were also assigned to KOG categories using BLAST searches and an e-value cutoff of $9.0E^{-5}$ or less against the KOG database. The unique sequences were assigned to GO categories using the Blast2GO program and presented using the goslim_plant hierarchy (Conesa et al. 2005; http://www.blast2go.de/).

Phylogenomics

The *E. huxleyi* and *G. pacifica* EST data were used as input for the phylogenomics approach (for details, see Li et al. 2006; Reyes-Prieto, Yoon, and Bhattacharya 2006) using the PhyloGenie package of computer programs (Frickey and Lupas 2004). PhyloGenie is used for high-throughput phylogenetic reconstruction with an automated pipeline in which BLAST searches, extraction of homologous sequences from the BLAST results, generation of alignments, phylogenetic tree reconstruction, and neighbor joining bootstrap support values for individual phylogenies are calculated. The local database was assembled as described in Li et al. (2006), and all candidate proteins for phylogenetic analysis were used to build the final alignments by focusing on Plantae and chromalveolate data including *Galdieria sulphuraria* (Barbier et al. 2005); Michigan State University Galdieria Database http://genomics.msu.edu/galdieria/sequence_data.html), *Porphyra yezoensis* (Nikaido et al. 2000; http://www.kazusa.or.jp/en/plant/porphyra/EST), and *Phaeodactylum tricornutum* (Scala et al. 2002; http://www.ncbi.nlm.nih.gov/). EST data were translated over 6 frames using the ExPASy Translate tool (http://www.expasy.ch/tools/dna.html). To generate the final alignments we built a web-based local BLAST tool (Dragonblast V2.1; S.E.R. unpublished data) that included all genome data that was not in the phylogenomics local database, including publicly available data from GenBank (http://www.ncbi.nlm.nih.gov/) and the Protist EST Program (PEP; http://www.bch.umontreal.ca/pepdb/pep_main.html). Each candidate protein identified via phylogenomics was used in TBLASTN searches with Dragonblast to generate the final data sets. The protein alignments were generated using the ClustalW server (http://www.ebi.ac.uk/clustalw/) and then manually refined. Only regions that were unambiguously aligned were retained for phylogenetic analysis. These alignments are available upon request from D.B.

Phylogenetic Analysis

For each of the 16 single-protein data sets that were assembled using phylogenomics and available genome data, a single-gene phylogeny was reconstructed using a bootstrap (100 replications) maximum likelihood approach with tree optimization (PHYML V2.4.3; Guindon and Gascuel 2003) with the WAG + I + Γ evolutionary model. We incorporated a diversity of eukaryotic sequences in these trees through exhaustive database searches to address potential alternative positions for chromalveolate genes resulting from endosymbiotic gene transfer (EGT) associated with the plastid or HGT from a foreign source. Most of the 16 proteins have been used previously (often extensively, e.g., actin, alpha-tubulin, HSP90) in phylogenetic analyses and with one possible exception (see below) did not provide bootstrap support for EGT or HGT. Once this prescreen was complete, we assembled a 46-taxon data set from the 16 proteins (6,735 aa) for phylogenetic analysis. Members of all 6 eukaryotic supergroups were represented in this alignment, including excavates and two Rhizaria, the chlorarachniophyte amoeba *Bigelowiella natans,* and the foraminiferan *Reticulomyxa filosa* (Burki and Pawlowski 2006). To increase the calculation speed, we sampled broadly protistan supergroup members but included only 7 members of the Opisthokonta, a supergroup that has been substantiated in previous phylogenetic analyses (for review, see Parfrey et al. 2006). ProtTest V1.3 (Abascal, Zardoya, and Posada 2005) was used to identify the best-fit model (i.e., RtREV + Γ + F) for the 16-protein alignment. This model was then used in two RAxML (RAxML-VI-HPC, v2.2.1; Stamatakis, Ludwig, and Meier 2005) maximum likelihood analyses, with the first including all 46 taxa and the second excluding the long-branched *Giardia lamblia* (diplomonad) and *Trichomonas vaginalis* (parabasalid) sequences (e.g., Arisue, Hasegawa, and Hashimoto 2005). RAxML was chosen for these final multiprotein analyses because of its speed and evidence that this program outperforms PHYML for real data sets (see Stamatakis, Ludwig, and Meier 2005). These analyses used a random starting tree (one round of taxon addition) and the rapid hill-climbing algorithm (i.e, option -f d in RAxML). To generate bootstrap values for the 46- and 44-taxon trees, we used the RAxML topologies (and best-fit model parameter values) as starting trees in maximum likelihood analyses (100 replicates) under PHYML with tree optimization.

We used Bayesian inference (MrBayes V3.0b4; Huelsenbeck and Ronquist 2001) with the 46- and 44-taxon data set using the RtREV + Γ + F model to calculate posterior probabilities for nodes in the RAxML tree. Metropolis-coupled Markov chain Monte Carlo from a random starting tree was used in this analysis with two independent runs (i.e., nrun = 2 command) and 1 cold and 3 heated chains. The Bayesian analyses were run for 1 million generations each with trees sampled every 100th generation. To increase the probability of chain convergence, we sampled trees after the standard deviation values of the two runs were <0.01 to calculate the posterior probabilities (i.e., after 432,200 generations for the 46-taxon Bayesian analysis and 302,200 generations for the 44-taxon analysis). The remaining phylogenies were discarded as burn-in. In all of these phylogenetic analyses, the branch leading to the Opisthokonta was used to root the tree for visualization purposes (e.g., Arisue, Hasegawa, and Hashimoto 2005).

And finally, to test whether the results of our maximum likelihood were potentially being misled by the class of fastest evolving sites in the data set, we assigned each amino acid position in the 16-protein data set to 1 of 4 ML gamma rate categories using TREE-PUZZLE (V5.2,

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

Schmidt et al. 2002). The RaxML 46-taxon tree was used as input with the WAG + Γ + F model, and the rate category contributing most to each site was recorded, and the sites were sorted based on its rate for each data set. The sites in rate category 4 (i.e., fastest evolving) were removed from the alignment (leaving 5467 aa), and a RaxML tree was inferred using these data with PHYML bootstrap analyses done as described above. This procedure was also done with the 44-taxon data set that excluded *G. lamblia* and *T. vaginalis*.

Tree Topology Tests

To assess the positions of Rhizaria, cryptophytes, haptophytes, and cryptophytes + haptophytes in the 46-taxon RAxML tree, we generated 4 different backbone phylogenies that were identical to the "best" topology but excluded these taxa. Each clade was then added (using MacClade V4.05; Maddison and Maddison 2005) to every possible branch in the respective backbone tree to assess its alternative positions. The site-by-site likelihoods for the trees in the 4 analyses were calculated using the 16-protein data set and TREEPUZZLE (V5.2, Schmidt et al. 2002) with the WAG + Γ + F evolutionary model (the alpha value for the gamma distribution was identified using RAxML) and the default settings. The approximately unbiased (AU-) test was implemented using CONSEL V0.1i (Shimodaira and Hasegawa 2001) to identify the pool of probable trees in each test and to assign their probabilities.

Results and Discussion

Analysis of the *Emiliania huxleyi and Goniomonas* cf. *pacifica* Unigene Sets

A total of 13,981 3′ EST sequences were generated from the starter, normalized, and subtracted cDNA libraries of *E. huxleyi* and assembled into 6,488 clusters (46% gene novelty rate). The ESTs had an average length of 627 bases and an overall G+C content of 64.1%. A total of 1,233 ESTs from the *G. pacifica* library were sequenced and assembled into 496 clusters. These sequences have an average length of 427 bases and a G+C content of 56.4%. Using AutoFACT EST, we putatively annotated 2,288 (35%) *E. huxleyi* ESTs and 275 (55%) *G. pacifica* ESTs, and 1,892 and 186 ESTs, respectively, could be assigned to the KOG classification. The ESTs were also annotated using Blast2GO to assign the ESTs to GO categories. A total of 1,957 *E. huxleyi* (see fig. S1 in supplementary data) and 182 *G. pacifica* ESTs (not shown) could be assigned to GO categories.

Phylogenomic and Single-Protein Analyses

We used an existing phylogenomic pipeline in our lab (Li et al. 2006; Reyes-Prieto, Yoon, and Bhattacharya 2006) to automatically generate alignments and neighbor joining trees with bootstrap support values from genome data. The *E. huxleyi* unigene set was used as the query against a local database that included *G. pacifica* and other chromalveolate, algal, plant, animal, fungal, protist, and bacterial complete genome or EST sequences (for details, see Li et al. 2006). Initially we identified 42 trees (out of 234 that were considered [ca. 1/3 of the 785 BLAST hits identified above]) that contained at least one other chromalveolate (usually *T. pseudonana*) and a member of another eukaryotic lineage with no prokaryotes branching within these taxa. This list was reduced to 16 protein trees that fulfilled the following criteria for acceptance (e.g., Ciccarelli et al. 2006): absence of ancient paralogy, lack of bootstrap support for EGT or HGT from eukaryotes or prokaryotes (see below), ease of alignment, presence in all completed genomes, not of organelle function, and >200 amino acids in length. Not surprisingly, many of these proteins we identified in our search have been used extensively in previous phylogenetic analyses (e.g., actin, HSP90), and the list is dominated by highly expressed translational genes. The final list is 14-3-3 protein, 20S core proteasome subunit beta 4, 26S proteasome regulatory subunit T4, actin, alpha-tubulin, beta-tubulin, eukaryotic translation elongation factor 1 alpha (eEF-1alpha), eukaryotic translation elongation factor 2 (EF-2), eukaryotic translation initiation factor eIF-2 gamma subunit, eukaryotic translation initiation factor eIF-5A, heat shock protein HSP70, heat shock protein HSP90, phosphomannomutase, RuvB-like DNA/RNA helicase reptin, transitional endoplasmic reticulum ATPase, and V-type ATPase V1 subunit B.

We found one clear example of taxon misplacement in the single gene PHYML analyses. This data set (EF2) was then used in a RAxML tree inference with a bootstrap analysis. Here, the EST-derived EF2 homolog in *Karenia brevis* (dinoflagellate; Nosenko et al. 2006) had a strongly supported sister-group relationship (RAxML = 100%) with the excavates *Leishmania major* and *Trypanosoma* spp. (Euglenozoa, see fig. S2). The other dinoflagellates in our tree (*Alexandrium tamarense, Karlodinium micrum*) were in the expected position as sister to apicomplexans (i.e., *Theileria parva, Plasmodium falciparum*). The *K. brevis* EF2 gene is therefore likely to be an example of HGT from the excavate to the dinoflagellate, and these data were removed from the alignment prior to the 16-protein phylogenetic analyses. Another misplacement in the EF2 tree involves the haptophytes. In this case, the EST-based sequences from *Isochrysis galbana* (ISL00004505) and *Pavlova lutheri* (PLL00000385) in the PEP database share a moderately well-supported (RAxML bootstrap = 82%) relationship with the green algae/land plants and red algae. Previous analyses (e.g., Fast et al. 2001; Harper and Keeling 2003; Hackett et al. 2004; Li et al. 2006; Nosenko et al. 2006) have demonstrated that genes of red and green algal origin that encode plastid-targeted proteins are found in the nucleus of chromalveolates. Many of these sequences have putatively originated via EGT from the red algal secondary endosymbiont that is shared by plastid-bearing chromalveolates and from either a second endosymbiosis or multiple HGT events involving a green alga(e) (for details, see Nosenko et al. 2006). Therefore it is conceivable that genes encoding cytosolic proteins such as EF2 may also have undergone EGT to the nucleus of chromalveolates, as suggested by our analysis. Given these results, we removed the *I. galbana* and *P. lutheri* EF2 sequences from our alignment and entered them as missing

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

data. It is also possible that the misplaced EST sequences we identified are the result of culture contamination during cDNA library construction and/or sequencing and are not genuine cases of gene transfer. In either case, it appears that that single-gene trees are required to identify taxon misplacements when dealing with large multigene data sets.

Missing Data

Analysis of the 16-protein alignment showed that overall there was an average of 22.75% missing data (see table S1 in supplementary data). These sites were however not uniformly distributed across taxa (see table S1). Species with completed genomes had very little missing data (e.g., *Arabidopsis thaliana,* 0.0%; *Cyanidioschyzon merolae,* 0.1%), whereas the EST-derived data was highly variable in this respect. For example, although we sampled genes that were present in the *E. huxleyi* unigene set, many gene sequences were incomplete due to the single-pass 3′ cDNA sequencing approach, resulting in 56.1% missing positions. Similar values were found for EST data from other protists (table S1). In spite of the large amount of missing data from the haptophyte and other protists, a large number of positions (i.e., from a total of 6,735 aa) remained in the analysis. This latter aspect, rather than the overall percent missing data is likely a better indicator of the phylogenetic power of our data set (this issue is discussed in Wiens 2003, 2006; Driskell et al. 2004; Philippe et al. 2004; McMahon and Sanderson 2006) and was sufficient to resolve most nodes in the eukaryotic tree (see below). The addition of more chromalveolate and other protist data is required to test this hypothesis.

Multiprotein Phylogeny

The 46- and 44-taxon (excluding *G. lamblia* and *T. vaginalis*) RAxML trees of eukaryotes using the 16-protein data set are shown in Figs. 1 and S3 and provide strong bootstrap and Bayesian support for the monophyly of cryptophytes (PHYML46, PHYML44 = 100; Bayesian posterior probability, BPP = 1.0) and haptophytes (PHYML46, PHYML44 = 100%, BPP = 1.0) and moderate support for the sister-group relationship of these chromalveolate lineages (PHYML46 = 90%, PHYML44 = 87%, BPP = 1.0). We also found strong bootstrap support for the monophyly of 4/6 supergroups using either data set; e.g., Plantae (PHYML46 = 87%, PHYML44 = 98%), Opisthokonta (PHYML46 = 100%, PHYML44 = 100%), Amoebozoa (PHYML46 = 96%, PHYML44 = 99%), Rhizaria (PHYML46 = 100%, PHYML44 = 100%). Of particular interest was the well-supported sister-group relationship of Rhizaria to the stramenopile + alveolate clade (PHYML46 = 100%, PHYML44 = 100%). The latter group was also well supported in our analysis (PHYML46 = 92%, PHYML44 = 84%, see below for details).

Previous single- and multigene phylogenetic analyses have supported stramenopile + alveolate monophyly (Van de Peer and De Wachter 1997; Baldauf et al. 2000; Harper, Waanders, and Keeling 2005; Rodriguez-Ezpeleta et al. 2005). The Plantae has been shown to be monophyletic using plastid (e.g., Rodriguez-Ezpeleta et al. 2005; Yoon et al. 2005; Li et al. 2006) and nuclear gene (Rodriguez-Ezpeleta et al. 2005; Hackett et al. 2007) trees, as well as comparative analyses of the plastid import system (Matsuzaki et al. 2004; McFadden and van Dooren 2004), and plastid targeted translocators of fixed carbon (for red and green algae; Weber, Linka, and Bhattacharya 2006). Our tree is however the first that is consistent with Plantae monophyly when members of all 6 supergroups (albeit with restricted taxon sampling) are included in the analysis. The monophyly of the Amoebozoa taxa in our analysis (e.g., Baldauf et al. 2000; Fahrni et al. 2003; Walker, Dacks, and Martin Embley 2006) and of the Opisthokonta (e.g., Baldauf et al. 2000; Parfrey et al. 2006) is consistent with previous studies, although some aspects (e.g., monophyly of *Dictyostelium discoideum* and *Acanthamoeba castellanii*) differ with respect to previous analyses (e.g., Nikolaev et al. 2006). This latter result likely reflects the limited taxon sampling in our tree of the Amoebozoa. Excavate phylogeny is controversial. In our trees, exclusion of the long-branched *G. lamblia* and *T. vaginalis* from the data set did not change the interrelationships of excavate taxa. We found strong support for a single clade of excavates excluding the malawimonads (PHYML46 = 97%, [minus *G. lamblia* and *T. vaginalis*] PHYML44 = 94%). Our analyses also provided strong bootstrap support for the monophyly of each of these individual groups, Euglenozoa, jakobids, and malawimonads (PHYML46, 44 = 100% in each case). The paraphyly of Excavata evident in our phylogenies is, however, a provisional result because our sample of taxa is not broad enough to assess the overall monophyly of this supergroup. Important excavates that are missing from our tree include oxymonads, retortamonads, *Carpediemonas* , and *Trimastix* spp. (see Simpson 2003; Simpson, Inagaki, and Roger 2006).

Cryptophyte and Haptophyte Monophyly

A close evolutionary relationship between cryptophytes and haptophytes was previously reported by Harper, Waanders, and Keeling (2005) using less sequence data but with a broader taxon sampling. Shalchian-Tabrizi et al. (2006a) also found a specific relationship between cryptophytes and haptophytes in a clade that included telonemid protists. Perhaps most importantly, Rice and Walker (2006) recently identified a rare HGT event involving cryptophyte and haptophyte plastid genomes in which the vertically inherited *rpl36* gene of cyanobacterial origin in these taxa was replaced by one of foreign (eubacterial) provenance. Taken together, these data support the monophyly of cryptophytes and haptophytes and suggest that the red algal nucleomorph genome that is retained in plastid-containing cryptophytes was lost at least twice in evolution; once in the haptophyte ancestor and putatively once in the ancestor of the remaining chromalveolates and Rhizaria. Our data also strongly support a photosynthetic ancestry for *Goniomonas* spp. which now lacks a plastid. Finally, the sister-group relationship of cryptophytes + hatophytes is inconsistent with the Chromista hypothesis that posits monophyly of these taxa with stramenopiles. Under this

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

model, the nucleomorph-bearing cryptophytes are the early divergence with a single loss of this genome prior to the split of haptophytes and stramenopiles (Cavalier-Smith 1986). In our trees that contain members of all major chromalveolate lineages (except katablepharids and telonemids), the stramenopiles share a specific relationship with alveolates (see also Baldauf et al. 2000; Harper, Waanders, and Keeling 2005; Simpson, Inagaki, and Roger 2006) to the exclusion of other chromists (i.e., cryptophytes and haptophytes).

Monophyly of Rhizaria and Stramenopiles + Alveolates

The position of the Rhizaria as sister to the stramenopiles + alveolates is unexpected. In previous phylogenetic analyses, members of the Rhizaria have formed a moderately to well supported monophyletic group [e.g., actin, Keeling (2001); small subunit rDNA + actin, Nikolaev et al. (2004); RPB1, Longet et al. (2003)] that, on the basis of bootstrap analyses, does not share a close relationship with any other eukaryotic supergroup. This idea was recently addressed by Burki and Pawlowski (2006) using a dataset of 85 concatenated nuclear protein sequences (13,258 amino acids). Their phylogenomic analysis of EST data from *B. natans* and *R. filosa* robustly confirms the monophyly of these Rhizaria, with the AU-test suggesting two potential affiliations with other eukaryotes. The first, more highly supported topology ($P = 0.802$) indicated the monophyly of Rhizaria and an excavate clade defined by *G. lamblia* + *T. vaginalis* + Euglenozoa, whereas the second favored a sister-group relationship between Rhizaria and stramenopiles ($P = 0.412$). A close relationship between Rhizaria and other chromalveolates was reported (albeit without bootstrap support) in a 6-protein phylogeny in which *B. natans* was sister to a haptophyte + cryptophyte clade (Harper, Waanders, and Keeling 2005) and in the rDNA analysis of Nikolaev et al. (2006). In our analyses, we find robust bootstrap support for the monophyly of *B. natans* and *R. filosa* with the stramenopile + alveolate clade. If true, this suggests that Chromalveolata now includes Rhizaria within a significantly expanded supergroup.

The monophyly of cryptophytes + haptophytes with the other chromalveolate lineages (and Rhizaria) is moderately supported by our analyses (PHYML46 = 80%; PHYML44 = 86%; BPP = 1.0). This result is consistent with chromalveolate plastid protein trees (e.g., Yoon et al. 2002; Bachvaroff, Sanchez Puerta, and Delwiche 2005) and the phylogeny of chromalveolate nuclear-encoded plastid-targeted proteins (e.g., Fast et al. 2001; Li et al. 2006; Nosenko et al. 2006) that support the presence of a shared red algal secondary endosymbiont among these taxa. These analyses did not, however, exhaustively assess the phylogeny of chlorarachniophyte plastid-targeted proteins. The surprising relationship of Rhizaria to the chromalveolates suggests that, under the most parsimonious scenario, the putative red algal secondary endosymbiont shared by these taxa was lost in the Rhizaria ancestor and in ciliates. Within Rhizaria, a plastid was regained by chlorarachniophytes from a green algal secondary endosymbiosis (Rogers et al. 2007) and by the filose amoeba *Paulinella chromatophora* from a cyanobacterial primary endosymbiosis (Marin, Nowack, and Melkonian 2005; Yoon et al. 2006b).

We tested these results by removing the class of fastest evolving sites from the 16-protein alignment and inferred 46- and 44-taxon trees using RAxML. These trees (fig. 2) were nearly identical to those inferred from the analysis of all sites (figs. 1, S3) but provided equal or higher PHYML bootstrap support for key nodes of interest in the phylogenies. The position of the Rhizaria within chromalveolates (PHYML46 = 99%; PHYML44 = 98%) and the monophyly of the Plantae (PHYML46 = 100%; PHYML44 = 100%) and the cryptophyte + haptophyte clade (PHYML46 = 90%; PHYML44 = 93%) receive similar or higher bootstrap support and within the Plantae, the sister-group relationship of red and green algae is now well supported (PHYML46 = 92%; PHYML44 = 90%). Other nodes in the tree continue to receive significant bootstrap support after exclusion of the fastest evolving sites. This suggests that the phylogenetic signal regarding the Rhizaria and the cryptophytes + haptophytes does not derive from the sites most likely to carry significant homoplasy.

Tree Topology Tests

We used the AU-test to assess the likelihoods of tree topologies that addressed all alternate positions of the Rhizaria or either or both cryptophytes and haptophytes in the 46-taxon tree. Our analysis shows that only two competing positions for Rhizaria are retained in the pool of candidate trees. The first is that shown in figure1 (i.e., the best tree; $P = 0.962$) and the second, less favored position is as sister to alveolates ($P = 0.108$; see table 1). All other positions within figure 1 are significantly rejected, including the association of Rhizaria with excavate lineages ($P < 0.01$), as sister to stramenopiles ($P = 0.048$), or as sister to all chromalveolates ($P = 0.007$).

We assessed the possibility that the significance values generated by the AU-test for the Rhizaria position may be misleading because our pool of candidate trees included many that are highly unlikely (e.g., sister to different animals or fungi), perhaps making it more likely for the test to reject alternative "reasonable" topologies. The AU-test is, however, generally believed to not be susceptible to changes in the number of candidate trees (Shimodaira 2002). Nevertheless, we reran the AU-test using a more limited set of 49 candidate trees that tested all local rearrangements of the Rhizaria within chromalveolates or Plantae, as sister to either, or on the branch uniting these supergroups. The results of this analysis were very similar to the original, and all alternative positions of the Rhizaria were rejected except that shown in figure 1 ($P = 0.959$) and as sister to alveolates ($P = 0.159$). A potential sister-group relationship with stramenopiles was again marginally rejected ($P = 0.047$). This suggests that the AU-test results are not significantly biased by the candidate tree pool size. Given the strength of these results, it is unclear why our findings differ from those of Burki and Pawlowski (2006). These analyses have, however, two major differences. Our data set contains about one-half less sequence data than Burki and Pawlowski (2006), but in contrast is

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

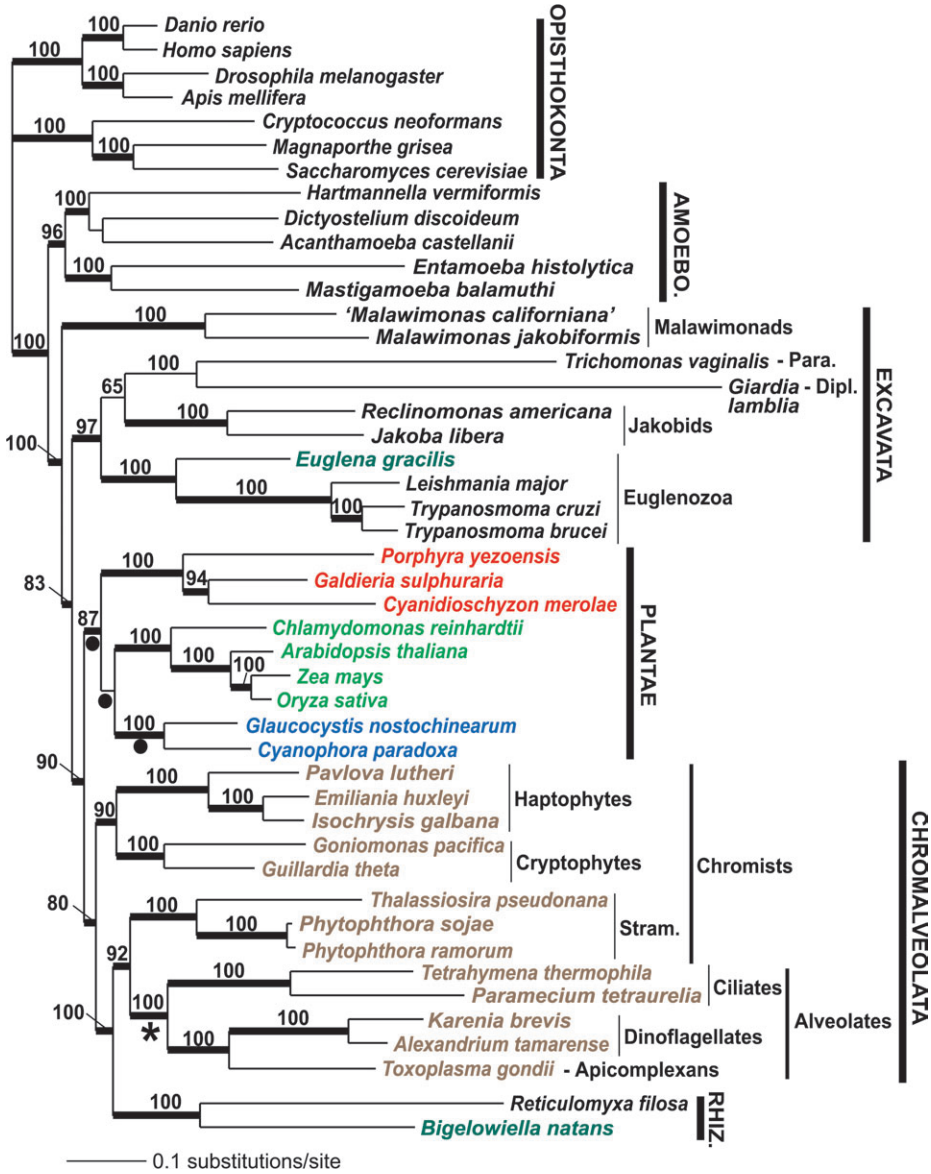

FIG. 1.—Maximum likelihood (RAxML) tree of eukaryotes using the 46-taxon data set and 16 proteins. The results of bootstrap analyses using PHYML are shown above the branches. Only bootstrap values >60% are recorded. The results of a Bayesian inference are shown as thick branches for posterior probabilities = 1.0. The branch lengths in this tree are proportional to the number of substitutions per site (see scale in figure). AMOEBO. is Amoebozoa; RHIZ. is Rhizaria; Stram. is stramenopiles; Para. is parabasalids; and Dipl. is diplomonads. Chromalveolates are shown in brown text, and the Plantae in red (red algae), blue (glaucophyte algae), and green (green algae and land plants) text. *Euglena gracilis* and *Bigelowiella natans* are shown in dark green text because both of these taxa contain a green algal plastid derived from independent secondary endosymbioses (Rogers et al. 2007). Results of the AU-test are shown for the alternate nonrejected positions of Rhizaria (asterisk) and the cryptophyte + haptophyte clade (filled circle).

characterized by a richer taxon sample of both excavates and chromalveolates. That we find both bootstrap and AU-test support for the Rhizaria + chromalveolate relationship suggests that significant phylogenetic signal exists in our data set for this surprising relationship. As stated above, if true, this topology suggests a significantly more complex history of endosymbiosis for chromalveolates and Rhizaria than previously thought, with potentially independent red algal endosymbioses in the cryptophye + haptophyte clade and in the stramenopiles + alveolates, or alternatively, a single plastid gain in the common ancestor of all of these taxa, with plastid loss in the branch uniting the Rhizaria (see also Burki and Pawlowski 2006).

Next we tested the position of the cryptophye + haptophyte clade and found that these taxa had two alternate positions in the tree. The position of highest probability was as shown in the best tree (fig. 1, $P = 0.863$), and the other trees that were not significantly rejected by the AU-test were as sister to all Plantae ($P = 0.389$), to only the glaucophytes ($P = 0.051$), or to the greens + glaucophytes ($P = 0.167$). This same pattern played out when the cryptophytes or haptophytes were moved as individual clades to all other branches in the best tree (see table 1 for details). Interestingly, movement of cryptophytes or haptophytes to the branches uniting taxa other than Plantae (e.g., Amoebozoa, malawimonads) resulted generally in

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

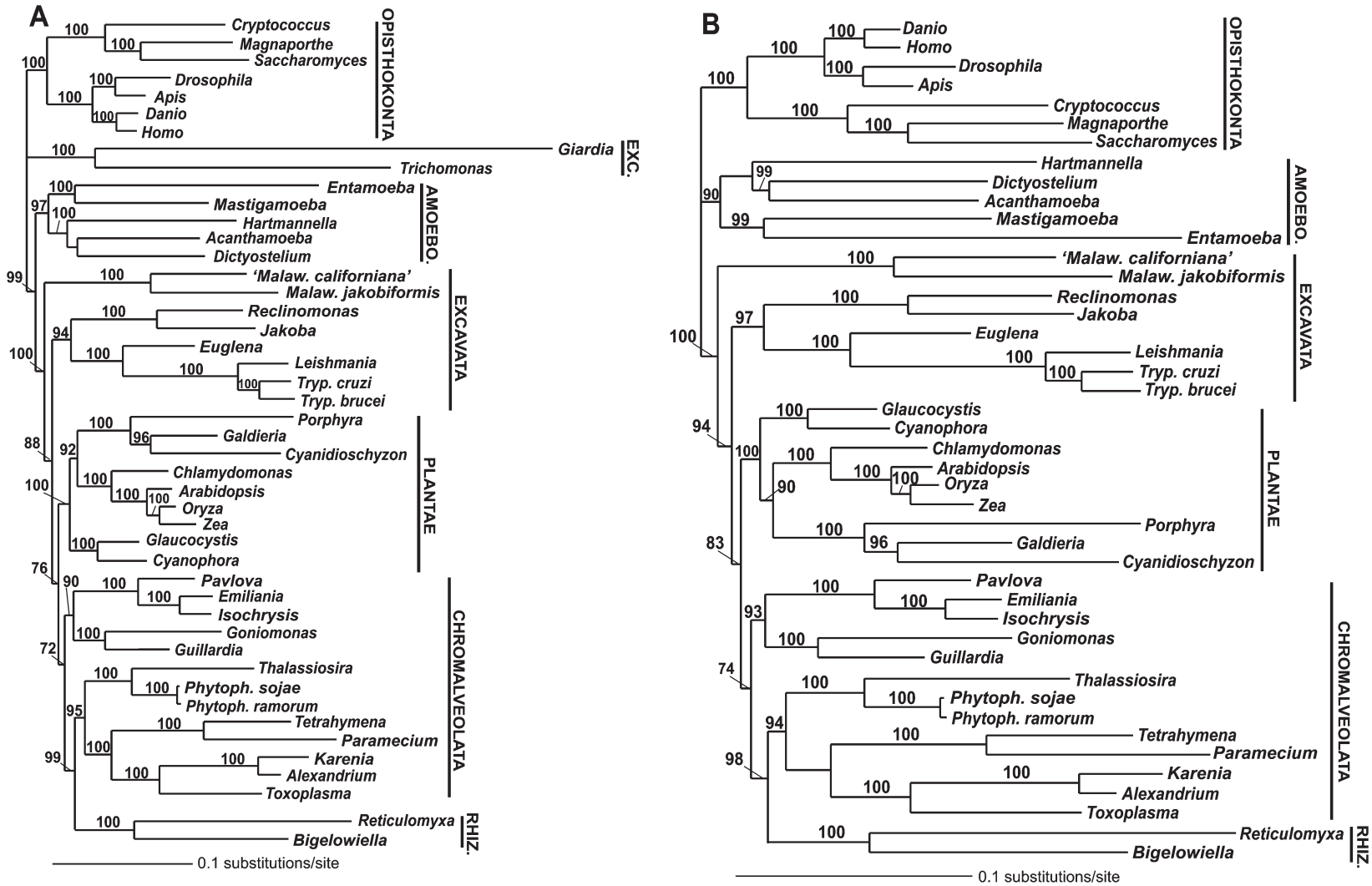

FIG. 2.—Maximum likelihood (RAxML) tree of eukaryotes after the removal of the most divergent class of amino acid sites from the 16-protein data set. (*A*) Phylogeny inferred using the 46-taxon data. (*B*) Phylogeny inferred using the 44-taxon data set after exclusion of the long-branched *G. lamblia* and *T. vaginalis* sequences. The results of bootstrap analyses using PHYML are shown above the branches in these trees. Only bootstrap values >60% are recorded. The branch lengths in the trees are proportional to the number of substitutions per site (see scales in figure). AMOEBO. is Amoebozoa; RHIZ. is Rhizaria; and EXC. Is Excavata.

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

Table 1
Results of the AU-Test for Alternative Tree Topologies Using the 46-Taxon Phylogeny Shown in Fig. 1.

Clade Moved			
Rhizaria	Rank	ΔlnL	AU Probability
Best tree	*1*	*-14.0*	*0.962*
Base alveolates	2	14.0	0.108
Base stramenopiles	3	18.3	0.048*
Base chromalveolates	5	26.0	0.007**
Base eugl.+ diplo.+parabas.+jakobids	11	63.3	0.002**
Base Euglenozoa	25	92.3	6e-007**
Base diplo.+parabas.+jakobids	21	86.8	2e-004**
Base malawimonads	19	83.5	7e-005**
Cryptophytes + Haptophytes	Rank	ΔlnL	AU Probability
Best tree	*1*	*-9.9*	*0.863*
Base Plantae	2	9.9	0.389
Base green+glaucophytes	6	27.6	0.167
Base glaucophytes	8	45.5	0.051
Base reds	7	45.0	6e-005**
Base Plantae+chromalveolates	5	25.8	0.013*
Base Amoebozoa	27	185.2	2e-004**
Base Euglenozoa	20	131.5	1e-077**
Base Opisthokonta	24	171.1	2e-005**
Cryptophytes	Rank	ΔlnL	AU Probability
Best tree	*1*	*-8.1*	*0.892*
Base chromalveolates+Rhizaria	2	8.1	0.383
Base Plantae	4	24.2	0.118
Base Plantae+chromalveolates	5	25.2	0.067
Base reds	10	46.6	0.006**
Base Amoebozoa	27	122.6	1e-032**
Base Euglenozoa	20	89.6	8e-009**
Base Opisthokonta	24	117.2	2e-075**
Haptophytes	Rank	ΔlnL	AU Probability
Best tree	*1*	*-8.1*	*0.888*
Base alveolates+Stramen.+ Rhizaria	2	8.1	0.375
Base Plantae	4	24.5	0.106
Base greens+glaucophytes	7	29.4	0.166
Base greens	8	30.8	0.219
Base glaucophytes	10	37.6	0.079
Base reds	11	50.7	2e-005**
Base Plantae+chromalveolates	9	37.5	0.014*
Base Amoebozoa	30	159.3	4e-004**
Base Euglenozoa	24	123.3	4e-009**
Base Opisthokonta	28	146.8	2e-006**

NOTE.—Significantly rejected trees are marked with * when $P < 0.05$ and with ** when $P < 0.01$. Diplo. is diplomonad, eugl. is Euglenozoa, and parabas. is parabasalids.

trees of significantly lower probability. This pattern may be interpreted in at least 3 ways. Under the first scenario, it is possible that the Plantae association is specific to our analysis. Although potentially true, it is a surprising result, given that virtually all other nodes in our tree are highly supported, and that the only alternate nonrejected associations detected by the AU-test are with one clade, the Plantae. Under a second scenario, our results may reflect undetected EGT (either complete or partial, ancient gene replacements) from a Plantae endosymbiont or HGT from these algae that is not evident in any single-protein PHYML bootstrap analysis (except for EF2, which is a long protein [783 aa in our final alignment] with significant phylogenetic signal). A surprising result in this regard is the nature of the Plantae association. Under the chromalveolate hypothesis, the source of EGT is expected to be from the red algal secondary endosymbiont. However, the AU-test strongly rejects the sister-group relationship of cryptophytes, haptophytes, or cryptophytes + haptophytes with red algae ($P < 0.01$). This suggests that, as evident in the EF2 tree (fig. S1) and analyses of nuclear-encoded plastid-targeted proteins in chromalveolates (e.g., Nosenko et al. 2006), nonred sources may also have provided genes to the nucleus of some or all chromalveolates. If this latter hypothesis is correct, then it may explain the great difficulties that are evident when reconstructing chromalveolate interrelationships, in spite of extensive multigene data or a broad taxon sampling (e.g., Harper Waanders, and Keeling 2005; Burki and Pawlowski 2006). Chromalveolate genomes may carry two distinct, conflicting phylogenetic histories, one of their hosts and one of their history of endosymbiotic gene replacements and HGT events, providing a potentially large impediment to multigene analyses of this group. A final possibility is that the cryptophytes and/or haptophyte host cells may be related to the Plantae, as has been suggested by rDNA trees that often support cryptophyte + glaucophyte monophyly (Bhattacharya et al. 1995; Van de Peer and De Wachter 1997; Okamoto and Inouye 2005), which is supported by the presence of flattened mitochondrial cristae in these taxa.

Conclusions

Although it is obvious that much more careful work needs to be done with chromalveolates and Rhizaria to elucidate their phylogeny, our analyses do provide some important insights. First, it is essential that in multigene analyses of chromalveolates each gene is analyzed individually with a broad taxon sample to identify potential cases of EGT or HGT. In the case of EGT, an obvious alternate position is expected with the Plantae. Second, despite the conflict evident within our data set, the overall phylogenetic signal favors two main findings: (1) the association of Rhizaria with chromalveolates, and (2) the monophyly of cryptophytes + haptophytes (see also Patron, Inagaki, and Keeling 2007). This latter result is supported by the independent plastid genome HGT data of Rice and Palmer (2006). One further consideration with regard to the position of the cryptophytes is that the AU-test does not significantly reject the ancestral position of this lineage relative to all other chromalveolates and Rhizaria ($P = 0.383$). This supports the notion of a single loss of the nucleomorph after the cryptophyte divergence, as has been suggested by plastid gene data (e.g., Yoon et al. 2002), but complicates interpretation of the *rpl36* gene replacement. If the cryptophytes are indeed the early divergence in the chromalveolate + Rhizaria clade, then it is possible two copies of *rpl36* may have existed in their common ancestor with independent, differential losses in the cryptophytes and haptophytes (cyanobacterial gene loss) versus other chromalveolates (eubacterial gene loss). Alternatively, and less parsimoniously, there were two

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

independent *rpl36* gene replacements in cryptophytes and haptophytes. Given these uncertainties, we stress that our results are a working hypothesis that is constrained by the inherent difficulties in inferring "deep" phylogeny due to differences in divergence rates among lineages, covarion substitution processes (e.g., Philippe et al. 2004; Rodriguez-Ezpeleta et al. 2005; Shalchian-Tabrizi et al. 2006b), and potentially confounding EGT. However these results can be tested through the analysis of more sequence data from existing cultured taxa than found in our tree (see EuTree; http://www.biology.uiowa.edu/eu_tree/) and the addition of novel taxa such as katablepharids and telonemids when these turn up in environmental surveys (e.g., picobiliphytes; Not et al. 2007).

Supplementary Material

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

Acknowledgments

This work was supported by grants from the National Science Foundation and the National Aeronautics and Space Administration awarded to D.B. (EF 04-31117, NNG04GM17G). J.D.H. was initially supported at the University of Iowa by an Institutional NRSA (T 32 GM98629) from the National Institutes of Health and is currently supported by a postdoctoral scholarship from the Cooperative Institute for Climate and Ocean Research of the National Oceanic and Atmospheric Administration and the Woods Hole Oceanographic Institution. We are grateful for the constructive comments of two anonymous reviewers.

Literature Cited

Abascal F, Zardoya R, Posada D. 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 21:2104–2105.

Adl SM, Simpson AG, Farmer MA. (28 co-authors). 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. J Eukaryot Microbiol. 52: 399–451.

Arisue N, Hasegawa M, Hashimoto T. 2005. Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. Mol Biol Evol. 22:409–420.

Bachvaroff TR, Sanchez-Puerta MV, Delwiche CF. 2005. Chlorophyll c-containing plastid relationships based on analyses of a multi-gene dataset with all four chromalveolate lineages. Mol Biol Evol. 22:1772–1782.

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.

Barbier G, Oesterhelt C, Larson MD, Halgren RG, Wilkerson C, Garavito RM, Benning C, Weber AP. 2005. Comparative genomics of two closely related unicellular thermoacidophilic red algae, *Galdieria sulphuraria* and *Cyanidioschyzon merolae*, reveals the molecular basis of the metabolic flexibility of *Galdieria sulphuraria* and significant differences in carbohydrate metabolism of both algae. Plant Physiol. 137:460–474.

Bhattacharya D, Helmchen T, Bibeau C, Melkonian M. 1995. Comparisons of nuclear-encoded small-subunit ribosomal RNAs reveal the evolutionary position of the Glaucocystophyta. Mol Biol Evol. 12:415–420.

Bhattacharya D, Katz LA. 2005. Frontiers in genomics: Insights into protist evolutionary biology. J Eukaryot Microbiol. University of Iowa, J Eukaryot Microbiol. 52:170–172. May 19–21, 2004

Bhattacharya D, Yoon HS, Hackett JD. 2004. Photosynthetic eukaryotes unite: Endosymbiosis connects the dots. BioEssays. 26:50–60.

Brown CW, Yoder JA. 1994. Coccolithophorid blooms in the global ocean. J. Geophys Res. 99:7467–7482.

Buitenhuis E, van Bleijswijk J, Bakker D, Veldhuis M. 1996. Trends in inorganic and organic carbon in a bloom of *Emiliania huxleyi* in the North Sea. Mar Ecol Prog Ser. 143:271–282.

Burki F, Pawlowski J. 2006. Monophyly of Rhizaria and multigene phylogeny of unicellular bikonts. Mol Biol Evol. 23:1922–1930.

Cavalier-Smith T. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan plastid origins and the Eukaryote family tree. J Eukaryot Microbiol. 46:347–366.

Cavalier-Smith T. 1986. The kingdom Chromista: Origin and systematics. In: Round FE, Chapman DJ, eds. Progress in Phycological Research. Bristol, U.K.: Biopress. p. 309–347.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. 2006. Toward automatic reconstruction of a highly resolved tree of life. Science. 311:1283–1287.

Conesa A, Gotz S, Garcia-Gomez JM, Terol T, Talon M, Robles M. 2005. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674–3676.

Driskell AC, Ane C, Burleigh JG, McMahon MM, O'Meara BC, Sanderson MJ. 2004. Prospects for building the tree of life from large sequence databases. Science. 306:1172–1174.

Ewing B, Green P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. Genome Res. 8:186–194.

Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8:175–185.

Fahrni JF, Bolivar I, Berney C, Nassonova E, Smirnov A, Pawlowski J. 2003. Phylogeny of lobose amoebae based on actin and small-subunit ribosomal RNA genes. Mol Biol Evol. 20:1881–1886.

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.

Foth BJ, McFadden GI. 2003. The apicoplast: A plastid in *Plasmodium falciparum* and other Apicomplexan parasites. Int Rev Cytol. 224:57–110.

Frickey T, Lupas AN. 2004. PhyloGenie: Automated phylome generation and analysis. Nucleic Acids Res. 32:5231–5238.

Gajadhar AA, Marquardt WC, Hall R, Gunderson J, Ariztia-Carmona EV, Sogin ML. 1991. Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. Mol Biochem Parasitol. 45:147–154.

Greenwood AD. 1974. The Cryptophyta in relation to phylogeny and photosynthesis. In: Saunders JV, Goodchild DJ, eds.

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

Electron microscopy 1974. Canberra: Australian Academy of Sciences. p. 566–567.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 52:696–704.

Hackett JD, Yoon HS, Butterfield NJ, Sanderson MJ, Bhattacharya D. 2007. Plastid endosymbiosis: Sources and timing of the major events. In: Falkowski P, Knoll A, eds. Evolution of primary producers in the sea. Forthcoming: Elsevier.

Hackett JD, Yoon HS, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Nosenko T, Bhattacharya D. 2004. Migration of the plastid genome to the nucleus in a peridinin dinoflagellate. Curr Biol. 14:213–218.

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.

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.

Hoef-Emden K, Marin B, Melkonian M. 2002. Nuclear and nucleomorph SSU rDNA phylogeny in the Cryptophyta and the evolution of cryptophyte diversity. J Mol Evol. 55: 161–179.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17:754–755.

Keeling PJ. 2001. Foraminifera and Cercozoa are related in actin phylogeny: Two orphans find a home? Mol Biol Evol. 18:1551–1557.

Keeling PJ, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE, Roger AJ, Gray MW. 2005. The tree of eukaryotes. Trends Ecol Evol. 20:670–676.

Koski LB, Gray MW, Lang BF, Burger G. 2005. AutoFACT: An automatic functional annotation and classification tool. BMC Bioinformatics. 6:151.

Li S, Nosenko T, Hackett JD, Bhattacharya D. 2006. Phylogenomic analysis identifies red algal genes of endosymbiotic origin in the chromalveolates. Mol Biol Evol. 23:663–674.

Longet D, Archibald JM, Keeling PJ, Pawlowski J. 2003. Foraminifera and Cercozoa share a common origin according to RNA polymerase II phylogenies. Int J Syst Evol Microbiol. 53:1735–1739.

Maddison DR, Maddison WP. 2005. MacClade. Sunderland (MA): Sinauer Associates.

Marin B, Nowack EC, Melkonian M. 2005. A plastid in the making: Evidence for a second primary endosymbiosis. Protist. 156:425–432.

Matsuzaki M, Misumi O, Shin IT. (41 co-authors). 2004. Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. Nature. 428:653–657.

McEwan ML, Keeling PJ. 2004. HSP90, tubulin and actin are retained in the tertiary endosymbiont genome of Kryptoperidinium foliaceum. J Eukaryot Microbiol. 51:651–659.

McFadden GI, van Dooren GG. 2004. Evolution: Red algal genome affirms a common origin of all plastids. Curr Biol. 14:R514–516.

McMahon MM, Sanderson MJ. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. Syst Biol. 55:818–836.

Nikaido 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.

Nikolaev SI, Berney C, Fahrni JF, Bolivar I, Polet S, Mylnikov AP, Aleshin VV, Petrov NB, Pawlowski J. 2004. The twilight of Heliozoa and rise of Rhizaria, an emerging supergroup of amoeboid eukaryotes. Proc Natl Acad Sci USA. 101:8066–8071.

Nikolaev SI, Berney C, Petrov NB, Mylnikov AP, Fahrni JF, Pawlowski J. 2006. Phylogenetic position of *Multicilia marina* and the evolution of Amoebozoa. Int J Syst Evol Microbiol. 56:1449–1458.

Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng JF, Bhattacharya D. 2006. Chimeric plastid proteome in the Florida "red tide" dinoflagellate *Karenia brevis*. Mol Biol Evol. 23:2026–2038.

Not F, Valentin K, Romari K, Lovejoy C, Massana T, Tobe K, Vaulot D, Medlin LK. 2007. Picobiliphytes: A marine picoplanktonic algal group with unknown affinities to other eukaryotes. Science. 315:253–255.

Nozaki H, Matsuzaki M, Misumi O, Kuroiwa H, Hasegawa M, Higashiyama T, Shin IT, Kohara Y, Ogasawara N, Kuroiwa T. 2004. Cyanobacterial genes transmitted to the nucleus before divergence of red algae in the Chromista. J Mol Evol. 59:103–113.

Okamoto N, Inouye I. 2005. The katablepharids are a distant sister group of the Cryptophyta: A proposal for Katablepharidophyta Divisio Nova/Kathablepharida Phylum Novum based on SSU rDNA and beta-tubulin phylogeny. Protist. 156:163–179.

Parfrey LW, Barbero E, Lasser E, Dunthorn M, Bhattacharya D, Patterson DJ, Katz LA. 2006. Evaluating support for the current classification of eukaryotic diversity. PLoS Genet. 2:e220.

Patron NJ, Rogers MB, Keeling PJ. 2004. Gene replacement of fructose-1,6-bisphosphate aldolase supports the hypothesis of a single photosynthetic ancestor of chromalveolates. Eukaryot Cell. 3:1169–1175.

Patron NJ, Inagaki Y, Keeling PJ. 2007. Multiple gene phylogenies support the monophyly of cryptomonad and haptophyte host lineages. Curr Biol. 17:1–5.

Petersen J, Teich R, Becker B, Cerff R, Brinkmann H. 2006. The GapA/B gene duplication marks the origin of Streptophyta (charophytes and land plants). Mol Biol Evol. 23:1109–1118.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PW, Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Mol Biol Evol. 21:1740–1752.

Reyes-Prieto A, Yoon HS, Bhattacharya D. 2006. Phylogenomics and its growing impact on algal phylogeny and evolution. Algae. 21:1–10.

Rice DW, Palmer JD. 2006. An exceptional horizontal gene transfer in plastids: Gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. BMC Biol. 4:31.

Rodriguez-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.

Rogers MB, Gilson PR, Su V, McFadden GI, Keeling PJ. 2007. The Complete chloroplast genome of the chlorarachniophyte *Bigelowiella natans*: Evidence for independent origins of chlorarachniophyte and euglenid secondary endosymbionts. Mol Biol Evol. 24:54–62.

Scala S, Carels N, Falciatore A, Chiusano ML, Bowler C. 2002. Genome properties of the diatom *Phaeodactylum tricornutum*. Plant Physiol. 129:993–1002.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18:502–504.

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012

Shalchian-Tabrizi K, Eikrem W, Klaveness D. (12 co-authors). 2006a. Telonemia, a new protist phylum with affinity to chromist lineages. Proc Biol Sci. 273:1833–1842.

Shalchian-Tabrizi K, Skanseng M, Ronquist F, Klaveness D, Bachvaroff TR, Delwiche CF, Botnen A, Tengs T, Jakobsen KS. 2006b. Heterotachy processes in rhodophyte-derived secondhand plastid genes: Implications for addressing the origin and evolution of dinoflagellate plastids. Mol Biol Evol. 23:1504–1515.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst Biol. 51:492–508.

Shimodaira H, Hasegawa M. 2001. CONSEL: For assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.

Simpson AG. 2003. Cytoskeletal organization, phylogenetic affinities and systematics in the contentious taxon Excavata (Eukaryota). Int J Syst Evol Microbiol. 53:1759–1777.

Simpson AG, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol. 23:615–625.

Simpson AG, Roger AJ. 2004. The real 'kingdoms' of eukaryotes. Curr Biol. 14:R693–696.

Stamatakis A, Ludwig T, Meier H. 2005. RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics. 21:456–463.

Stechmann A, Cavalier-Smith T. 2003. Phylogenetic analysis of eukaryotes using heat-shock protein Hsp90. J Mol Evol. 57:408–419.

Stibitz TB, Keeling PJ, Bhattacharya D. 2000. Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. Mol Biol Evol. 17:1731–1738.

Van de Peer Y, Baldauf SL, Doolittle WF, Meyer A. 2000. An updated and comprehensive rRNA phylogeny of (crown) eukaryotes based on rate-calibrated evolutionary distances. J Mol Evol. 51:565–576.

Van de Peer Y, De Wachter R. 1997. Evolutionary relationships among the eukaryotic crown taxa taking into account site-to-site rate variation in 18S rRNA. J Mol Evol. 45:619–630.

Walker G, Dacks JB, Martin Embley T. 2006. Ultrastructural description of *Breviata anathema*, n. gen., n. sp., the organism previously studied as "*Mastigamoeba invertens*". J Eukaryot Microbiol. 53:65–78.

Weber AP, Linka M, Bhattacharya D. 2006. Single, ancient origin of a plastid metabolite translocator family in Plantae from an endomembrane-derived ancestor. Eukaryot Cell. 5:609–612.

Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. Syst Biol. 52:528–538.

Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. J Biomed Inform. 39:34–42.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol. 21:809–818.

Yoon HS, Hackett JD, Pinto G, Bhattacharya D. 2002. The single, ancient origin of chromist plastids. Proc Natl Acad Sci USA. 99:15507–15512.

Yoon HS, Hackett JD, Van Dolah FM, Nosenko T, Lidie KL, Bhattacharya D. 2005. Tertiary endosymbiosis driven genome evolution in dinoflagellate algae. Mol Biol Evol. 22:1299–1308.

Yoon HS, Muller KM, Sheath RG, Ott FD, Bhattacharya D. 2006a. Defining the major lineages of red algae (Rhodophyta). J Phycol. 42:482–492.

Yoon HS, Reyes-Prieto A, Melkonian M, Bhattacharya D. 2006b. Minimal plastid genome evolution in the *Paulinella* endosymbiont. Curr Biol. 16:R670–672.

Martin Embley, Associate Editor.

Accepted April 25, 2007

Downloaded from http://mbe.oxfordjournals.org/ by guest on September 19, 2012