

# A Molecular Timeline for the Origin of Photosynthetic Eukaryotes

Hwan Su Yoon,\* Jeremiah D. Hackett,\* Claudia Ciniglia,† Gabriele Pinto,† and Debashish Bhattacharya\*

\*Department of Biological Sciences and Center for Comparative Genomics, University of Iowa; †Dipartimento di Biologia Vegetale, Università “Federico II,” Naples, Italy

The appearance of photosynthetic eukaryotes (algae and plants) dramatically altered the Earth's ecosystem, making possible all vertebrate life on land, including humans. Dating algal origin is, however, frustrated by a meager fossil record. We generated a plastid multi-gene phylogeny with Bayesian inference and then used maximum likelihood molecular clock methods to estimate algal divergence times. The plastid tree was used as a surrogate for algal host evolution because of recent phylogenetic evidence supporting the vertical ancestry of the plastid in the red, green, and glaucophyte algae. Nodes in the plastid tree were constrained with six reliable fossil dates and a maximum age of 3,500 MYA based on the earliest known eubacterial fossil. Our analyses support an ancient (late Paleoproterozoic) origin of photosynthetic eukaryotes with the primary endosymbiosis that gave rise to the first alga having occurred after the split of the Plantae (i.e., red, green, and glaucophyte algae plus land plants) from the opisthokonts sometime before 1,558 MYA. The split of the red and green algae is calculated to have occurred about 1,500 MYA, and the putative single red algal secondary endosymbiosis that gave rise to the plastid in the cryptophyte, haptophyte, and stramenopile algae (chromists) occurred about 1,300 MYA. These dates, which are consistent with fossil evidence for putative marine algae (i.e., acritarchs) from the early Mesoproterozoic (1,500 MYA) and with a major eukaryotic diversification in the very late Mesoproterozoic and Neoproterozoic, provide a molecular timeline for understanding algal evolution.

## Introduction

The photosynthetic eukaryotes (i.e., algae and plants) define a vast assemblage of autotrophs (Graham and Wilcox 2000). The emergence dates of these taxa have proven difficult to establish solely on the basis of fossil or biomarker evidence (Knoll 1992). Recent phylogenetic data suggest that the different algal groups diverged near the base of the eukaryotic tree (Baldauf et al. 2000; Baldauf 2003; Nozaki et al. 2003). This observation makes endosymbiosis, the process that creates plastids (Bhattacharya and Medlin 1995), one of the fundamental forces in the Earth's history. Molecular clock methods that incorporate information from plastid genomes offer a potentially powerful approach to date splits in the algal tree of life. These methods are, however, not without pitfalls, and they require that four general conditions be met: (1) a well-supported and accurate tree that resolves all the important nodes in the phylogeny (this normally entails the use of large multi-gene data sets), (2) reliable fossil calibrations on the tree that provide upper and lower bounds for the nodes of interest, (3) molecular clock methods that account for DNA mutation rate heterogeneity within and across lineages, and (4) a broad taxon sampling that includes the known diversity in lineages (Soltis et al. 2002). Given that one or more of these criteria have not been addressed, it is not surprising that molecular clock estimates are often inconsistent with the fossil record (Benton and Ayala 2003; Heckman et al. 2001). This is especially true for the estimation of ancient divergence times for which there is limited fossil evidence, and modeling DNA sequence evolution is the most error-prone because of the accumulation of superimposed mutations (Whelan, Liò, and Goldman 2001).

In contrast, the fossil data have two significant shortcomings. The first is that fossil dates are always underestimates because the first emergence of a lineage is not likely to be discovered because of the rare and sporadic nature of the fossil record. Second, for unarmored unicellular or filamentous eukaryotes, apart from size (prokaryotes >1 mm in size are unknown), it is very difficult to discriminate them from bacteria (Benton and Ayala 2003; Knoll 2003). The multitude of intracellular features that discriminate living eukaryotic and prokaryotic cells are absent in fossils. In spite of these concerns, molecular and fossil data provide independent and potentially valuable perspectives on biological evolution.

With this in mind, we set out to use a multi-gene approach and reliable fossil constraints to address an outstanding issue in biological evolution, the timing of the cyanobacterial primary endosymbiosis that gave rise to the first photosynthetic eukaryote and the subsequent splits in the algal tree of life. To do this, we erected a six-gene (and five-protein) plastid phylogeny that includes red, green, glaucophyte, and chromist (the chlorophyll-*c*-containing cryptophytes, haptophytes, and stramenopiles [Cavalier-Smith 1986]) algae. Maximum likelihood methods that take into account divergence rate variation were used to calculate emergence dates using trees identified with Bayesian inference. These data establish a molecular timeline for the origin of photosynthetic eukaryotes that is in agreement with the available fossil record.

## Materials and Methods

### Taxon Sampling and Sequencing

Forty-six species were used to infer the plastid phylogeny including 32 red algae including the chromists, 12 green algae and land plants, the glaucophyte *Cyanophora paradoxa*, and a cyanobacterium (*Nostoc* sp. PCC7120) as the outgroup (for strain identifications and GenBank accession numbers, see table 1 in the Supplementary Material online). A total of 42 new plastid

Key words: algal origin, fossil record, molecular clock, divergence time estimates, plastid.

E-mail: dbhatac@blue.weeg.uiowa.edu.

*Mol. Biol. Evol.* 21(5):809–818. 2004

DOI:10.1093/molbev/msh075

Advance Access publication February 12, 2004

sequences were determined in this study. Our sequencing strategy was to focus on red algae and chromists that span the known diversity of these lineages. In particular, we included a broad diversity of extremophilic Cyanidiales, including two mesophilic taxa that we have recently discovered (*Cyanidium* sp. Sybil, *Cyanidium* sp. Monte Rotaro), and members of the other genera in this early-diverging red algal order. Our data set included, therefore, key early-diverging red and green (e.g., *Mesostigma viride*) algae and land plants (e.g., *Anthoceros formosae*), a glaucophyte, and a cyanobacterium.

To prepare DNA, the algal cultures were frozen in liquid nitrogen and ground with glass beads using a glass rod and/or Mini-BeadBeater (Biospec Products, Inc., Bartlesville, Okla.). Total genomic DNA was extracted with the DNeasy Plant Mini Kit (Qiagen, Santa Clarita, Calif.). Polymerase chain reactions (PCR) were done using specific primers for each of the plastid genes (see Yoon, Hackett, and Bhattacharya 2002; Yoon et al. 2002). Four degenerate primers were used to amplify and sequence the photosystem I P700 chlorophyll *a* apoprotein A2 (*psaB*) gene: *psaB*500F; 5'-TCWTGGTTYAAAAATAAYGA-3', *psaB*1000F; 5'-CAAYTAGGHTTAGCTTTAGC-3', *psaB*1050R; 5'-GGYAWWGCATACATATGYTG-3', *psaB*-1760R; 5'-CCRATYGTATTWAGCATCCA-3'. Because introns were found in the plastid elongation factor Tu (*tufA*) and photosystem I P700 chlorophyll *a* apoprotein A1 (*psaA*) genes of some red algae (most likely indicating gene transfer to the nucleus [H. S. Y., D. B. unpublished data]), the reverse transcriptase (RT)-PCR method was used to isolate cDNA. For the RT-PCR, total RNA was extracted with the RNeasy Mini Kit (Qiagen, Santa Clarita, Calif.). To synthesize cDNA from total RNA, M-MLV Reverse Transcriptase (GIBCO BRL, Gaithersburg, Md.) was used according to the manufacturer's protocol. The PCR products were purified with the QIAquick PCR Purification kit (Qiagen), and were used for direct sequencing with the BigDye Terminator Cycle Sequencing Kit (PE-Applied Biosystems, Norwalk, Conn.) and an ABI-3100 at the Center for Comparative Genomics at the University of Iowa. Some PCR products were cloned into pGEM-T vector (Promega, Madison, Wis.) prior to sequencing.

### Phylogenetic Analyses

Sequences were manually aligned with SeqPup (Gilbert 1995). The alignment used in the phylogenetic analyses is available on request from D. B. We prepared a concatenated data set of 16S rRNA (1,309 nt), *psaA* (1,395 nt), *psaB* (1,266 nt), photosystem II reaction center protein D1 (*psbA*) (957 nt), ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbcL*; 1,215 nt), and *tufA* (969 nt) coding regions (a total of 7,111 nt) from photosynthetic eukaryotes and the cyanobacterium *Nostoc* sp. PCC7120 as the outgroup. Because the *rbcL* gene of the green and glaucophyte algae are of cyanobacterial origin, whereas those in the red algae and red-algal-derived plastids are of proteobacterial origin (e.g., Valentin and Zetsche 1990), the evolutionarily distantly related green and glaucophyte *rbcL* sequences were coded as missing data in the phylogenetic analyses. The highly divergent and likely

nonfunctional *tufA* sequence in *Chaetosphaeridium globosum* (Baldauf, Manhart, and Palmer 1990) and the nuclear-encoded land plant *tufA* genes (Baldauf and Palmer 1990) were also excluded from the analysis.

Trees were inferred with Bayesian inference and the minimum evolution (ME) and maximum parsimony (MP) methods. To address the possible misleading effects of nucleotide bias or mutational saturation at third codon positions in the DNA data set (e.g., for *rbcL*, see Pinto et al. 2003), we excluded third codon positions from the phylogenetic analyses (leaving a total of 5,177 nt). In the Bayesian inference of the DNA data (MrBayes, version 3.0b4; Huelsenbeck and Ronquist 2001), we used the general time reversible (GTR) +  $\Gamma$  model with separate model parameter estimates for the three data partitions (16S rRNA, first, and second codon positions in the protein-coding genes). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) from a random starting tree was initiated in the Bayesian inference and run for 2 million generations. Trees were sampled each 1,000 cycles. Four chains were run simultaneously of which three were heated and one was cold, with the initial 200,000 cycles (200 trees) being discarded as the "burn-in." Stationarity of the log likelihoods was monitored to verify convergence by 200,000 cycles (results not shown). A consensus tree was made with the remaining 1,800 phylogenies to determine the posterior probabilities at the different nodes. In the ME analyses, we generated distances using the GTR + I +  $\Gamma$  model (identified with Modeltest version 3.06, [Posada and Crandall 1998] as the best-fit model for our data) with the PAUP\*4.0b8 software (Swofford 2002). Ten heuristic searches with random-addition-sequence starting trees and tree bisection-reconnection (TBR) branch rearrangements were done to find the optimal ME trees. Best scoring trees were held at each step. In addition, we attempted to correct for mutational saturation and base composition heterogeneity in the DNA data by recoding first and third codon positions as purines (R) and pyrimidines (Y [see Phillips and Penny 2003; Delsuc, Phillips, and Penny 2003]). The 16S rDNA and second codon position data were maintained as the original nucleotides in this analysis. A starting tree was generated with the RY-recoded data set using the ME method and the HKY-85 evolutionary model. This tree was used as input in PAUP\* to calculate the parameters for the GTR + I +  $\Gamma$  model. These parameters were then used in a ME-bootstrap analysis (2,000 replications) with the settings described above.

Unweighted MP analysis was also done with the DNA data, using heuristic searches and TBR branch-swapping to find the shortest trees. The number of random-addition replicates was set to 10 for each tree search. To test the stability of monophyletic groups in the ME and MP trees, we analyzed 2,000 bootstrap replicates (Felsenstein 1985) of the DNA data set. We also did a Bayesian analysis in which all three codon positions were included in the data set (7,111 nt). The settings implemented in this inference were the same as described above (i.e., ssgamma), except for the use of a four-partition evolutionary model (i.e., 16S rRNA, first, second, and third codon positions).

In addition to the DNA analyses, we also inferred trees using the five proteins in our data set (i.e., excluding

16S rRNA). An ME tree was inferred with the “Fitch” program (PHYLIP version 3.6; Felsenstein 2002) using the WAG +  $\Gamma$  evolutionary model with 10 random sequence additions and global rearrangements to find the optimal trees. PUZZLEBOOT version 1.03 (<http://hades.biochem.dal.ca/Rogerlab/Software/software.html>) and Tree-Puzzle V5.1 (Schmidt et al. 2002) were used to generate the distance matrix. The gamma value was calculated using Tree-Puzzle. Protein bootstrap analyses using the ME method were done using the settings described above and 500 replicates. A quartet-puzzling–maximum likelihood analysis of the five-protein data set was done with Tree-Puzzle and the WAG +  $\Gamma$  model (50,000 puzzling steps).

### Molecular Clock Analyses

We used the maximum likelihood method to infer the divergence times of different plastid lineages. Seven different constraints were used in this analysis (see fig. 1A and table 2 in the Supplementary Material online). To date divergences in the best Bayesian tree and in the pool of credible Bayesian trees (see fig. 1 in the Supplementary Material online), we used the r8s program (Sanderson 2003) and the Langley-Fitch (LF) method with a “local molecular clock” and the Nonparametric rate smoothing (NPRS, Sanderson [1997]) method, both with the Powell search algorithm. In the LF method, local rates were calculated for 12 different clades (e.g., for each of the chromist plastid lineages, six for non-Cyanidiales red algae, one for the Cyanidiales, one for the Streptophyta [charophytes and land plants], and one for the chlorophyte green algae). Ninety-five percent confidence intervals on divergence dates were calculated using a drop of two ( $s = 2$ ) in the log likelihood units around the estimates (Cutler 2000). Three different starting points were used in each molecular clock analysis to avoid local optima. We chose methods that relax the assumption of a constant molecular clock across the tree because the likelihood ratio test showed significant departure, in our data set, from clock-like behavior ( $P < 0.005$ ).

## Results and Discussion

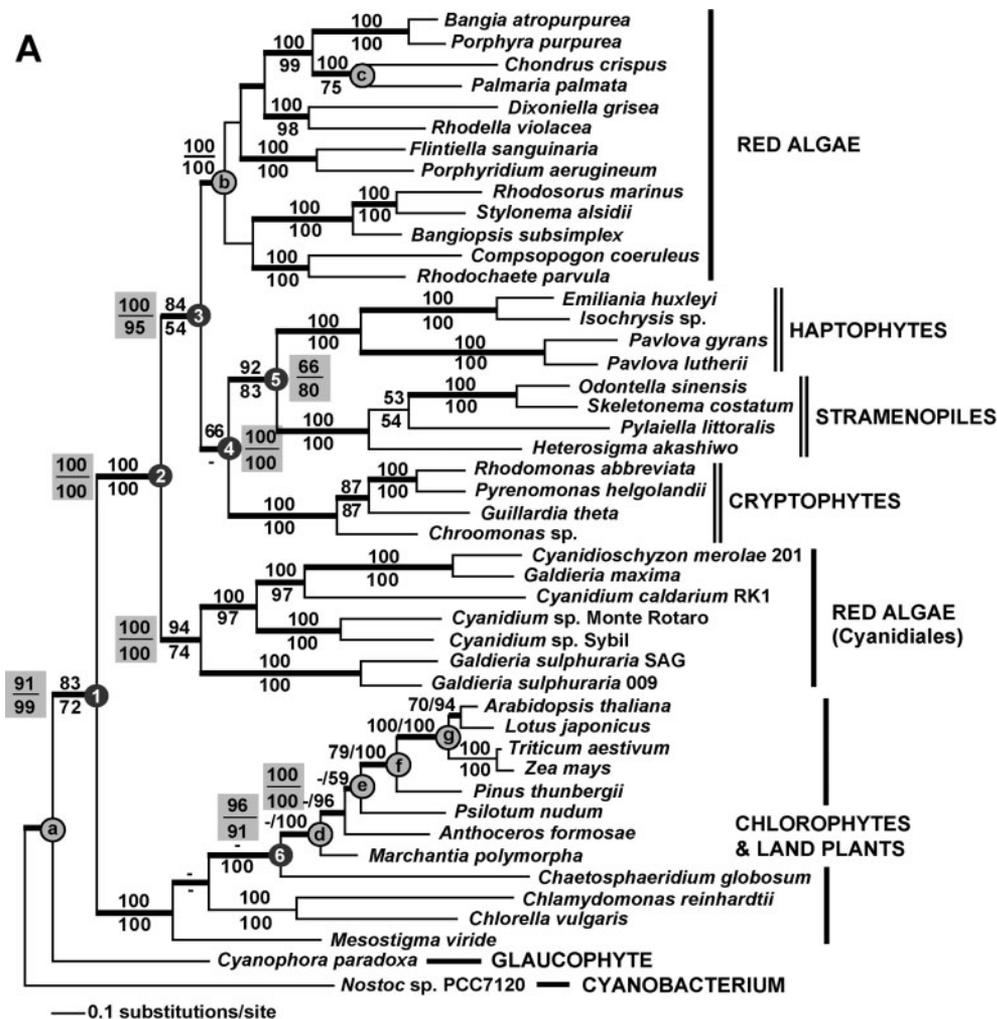
### Phylogenetic Relationships

The Bayesian tree of highest likelihood (excluding the third codon positions in the data), which was identified using the GTR evolutionary model with gamma-distributed rates across sites for three partitions, is shown in figure 1A. This phylogenetic hypothesis has relatively broad taxonomic sampling, including early diverging red (Cyanidiales) and green algal (*Mesostigma viride*) and land plant (e.g., *Marchantia polymorpha*) lineages, and it is consistent with present understanding of algal and plant relationships (Cavalier-Smith 1986; Fast et al. 2001; Karol et al. 2001; Yoon et al. 2002). Most nodes in the phylogeny, except that defining chromist monophyly (the haptophytes and stramenopiles were, however, strongly supported as sister groups), the near-simultaneous radi-

tion of the non-Cyanidiales red algae, and the early divergences in the chlorophyte/land plant lineage (see fig. 1A), have a significant ( $\geq 95\%$ ) posterior probability and strong bootstrap support (ME and MP methods). When we added the third codon positions (see fig. 2 in the Supplementary Material online) and reanalyzed the data using the four-partition model, the resulting Bayesian tree was essentially identical with the tree shown in figure 1A, however, with stronger bootstrap support for many nodes (see the shaded bootstrap values in figure 1A). Bootstrap analysis of the RY-recoded data set using the ME method (see fig. 3 in the Supplementary Material online) resulted in a consensus tree that was consistent with the results described above, with strong support for chromist plastid (94%) monophyly. The order of divergence of the non-Cyanidiales red algae and the early splits among land plants remained unresolved in this analysis (as in fig. 1A).

The ME tree of the five-protein data set is shown in figure 2. This phylogeny mirrors the DNA-based trees, except for the order of divergence of some green algal and land plant lineages (e.g., the position of *Mesostigma*, *Anthoceros*, and *Psilotum*). There was, however, only weak bootstrap support (64%) for chromist monophyly in the protein tree, leading us to question the strong support for this group based on the DNA data. Intriguingly, in all of our analyses the haptophytes and stramenopiles were always found as sister groups with moderate to strong bootstrap support (fig. 1A and fig. 2; see also figs. 2 and 3 in the Supplementary Material online), whereas the inclusion of the cryptophytes as the early divergence in the Chromista was more poorly supported. Third codon positions, which could exhibit nucleotide bias, were critical in the placement of the cryptophytes with the other chromists, with the bootstrap support increasing from 66% to 100% in the ME-GTR analyses when these sites were included in the DNA analysis. Given these results, we suggest that chromist monophyly remains a working hypothesis to explain plastid origin in these taxa, and that this idea remains to be established with the addition of more genes to our data set or through plastid genome comparisons that incorporate a broad taxon sampling. The cryptophytes are candidates for an independent origin of their red algal-derived plastid, whereas the monophyly of haptophytes and stramenopiles is well supported in all of our trees. Existing plastid genome trees using larger combined data sets of plastid proteins (41 [Martin et al. 2002], 39 [Maul et al. 2002], and 41 proteins [Ohta et al. 2003]) suggest polyphyly of the Chromista; however, these analyses all lack a representative of the haptophytes and sample poorly the red plastid lineage and algae containing red algal secondary endosymbionts. In spite of this unresolved issue, we chose to use the protein tree to date the basal splits in algal evolution. This choice was important because it allowed us to address potential error in our DNA-based estimates that could result, for example, from nucleotide composition bias.

Taken together, our analyses provide a generally consistent view of plastid relationships (with the caveat regarding chromist plastid origin), which is summarized in figure 1A. This tree is also interpretable as a “host” phylogeny for the red and green algae and for the



**B**

Node	DNA (1st + 2nd position) Tree		Protein Tree
	7 constraints (conf.) [BCI]	6 constraints (conf.)	6 constraints (conf.)
1	1474 (1449-1513) [1438-1576]	1452 (1401-1519)	1668 (1591-1757)
2	1370 (1350-1416) [1298-1415]	1349 (1301-1407)	1452 (1396-1519)
3	1274 (1261-1305) [1225-1309]	1255 (1204-1302)	1276 (NA)
4	1189 (1172-1219) [1106-1231]	1171 (1126-1216)	1224 (1177-1272)
5	1047 (1025-1077) [958-1102]	1032 (992-1076)	1096 (1038-1152)
6	792 (768-815) [707-835]	787 (762-814)	646 (596-703)
Ⓐ Max. age	3500	3500	3500
	1558 (1531-1602) [1526-1703]	1535 (1480-1600)	1719 (1636-1821)
Ⓑ Cons.	1174-1222	1156 (1116-1199)	Cons. 1174-1222

FIG. 1.—Evolutionary relationships of algal plastids. *A*, Phylogeny of the major algal groups inferred from a Bayesian analysis of the combined plastid DNA sequences of 16S rRNA, *psaA*, *psaB*, *psbA*, *rbcL*, and *tufA*, excluding third codon positions in the protein-coding regions. This is the tree of highest likelihood identified in the Bayesian tree pool using the three-partition analysis and the GTR model ( $-\ln$  likelihood = 60760.73). Results of a minimum evolution (ME)-GTR bootstrap analysis are shown above the branches, whereas the bootstrap values from an unweighted maximum parsimony (MP) analysis are shown below the branches. The bootstrap values in the gray squares were inferred from the full data set including third codon position (see, figure 2 in the Supplementary Material online). The thick nodes represent  $>95\%$  Bayesian posterior probability. The letters within the gray circles indicate nodes that were constrained for the molecular clock analyses. The nodes that were estimated are indicated by the numbers in the filled circles. Dashes indicate nodes that were not recovered in the ME-GTR or MP bootstrap consensus trees. *B*, The divergence time estimates and 95% confidence intervals (in parentheses) for the major phylogenetic splits calculated using the best Bayesian tree and the LF method from the DNA and protein data sets. The values when all seven constraints or when the *Bangiomorpha* (node b) constraint was released are shown. The Bayesian 95% confidence intervals (BCI) for these distributions are also shown for the LF analysis of 696/1800 phylogenies in the credible tree set that were identified with Bayesian inference.

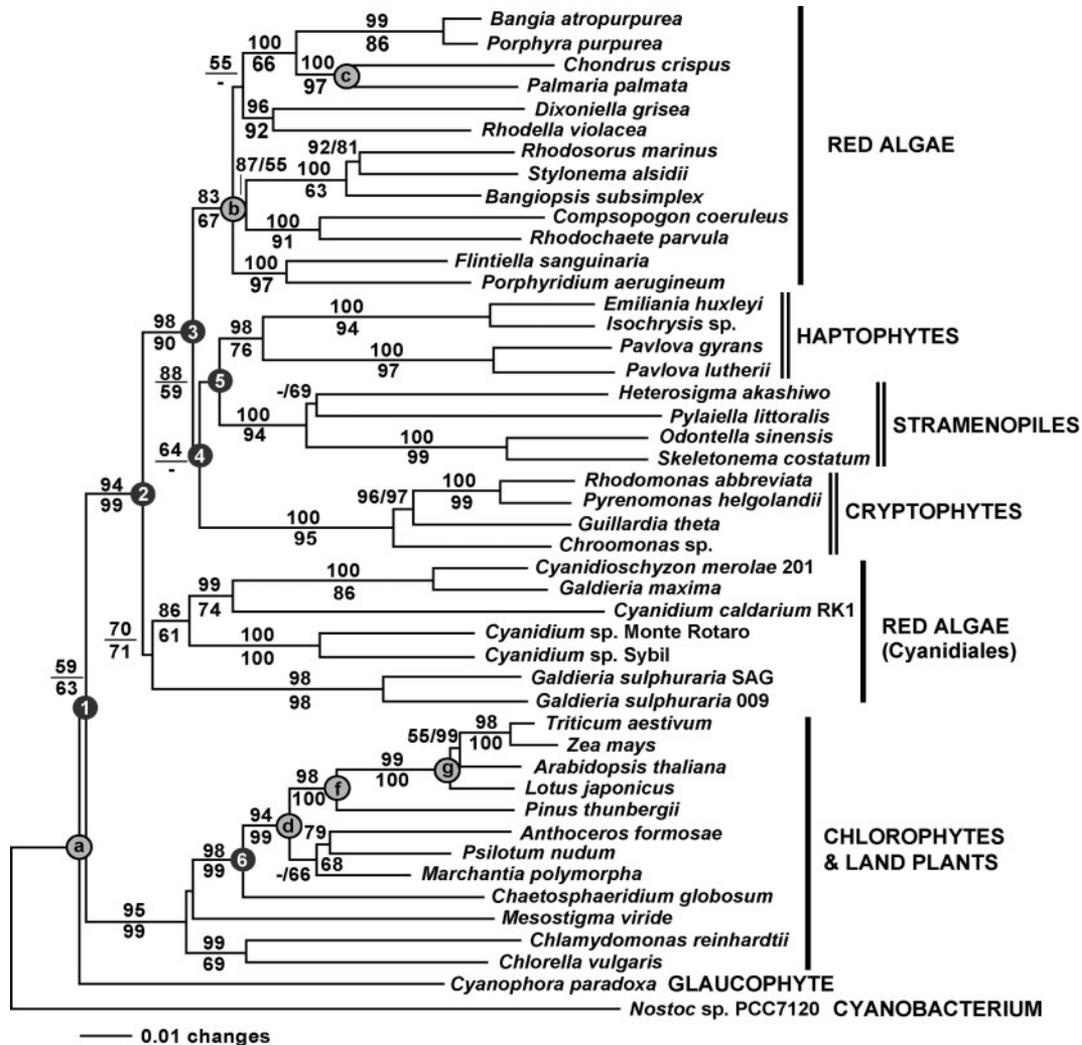


FIG. 2.—Evolutionary relationships of algal plastids using the five-protein data set. The phylogeny was inferred using the ME method, and distance matrices were calculated using the WAG +  $\Gamma$  evolutionary model. The results of a protein ME bootstrap analysis are shown above the branches, whereas puzzle values from a quartet puzzling-maximum likelihood analysis are shown below the branches (WAG +  $\Gamma$  model).

photosynthetic chromists that emerge as a monophyletic clade within the red lineage. The predicted congruence of plastid and host trees is based on phylogenetic evidence from nuclear and mitochondrial loci for the monophyly of red and green algae, with the glaucophytes (together, the Plantae [Cavalier-Smith 1998]) as a weakly supported sister group to this clade (Bhattacharya and Weber 1997; Gray et al. 1998; Moreira, Le Guyader, and Phillippe 2000). Plastid genes in the reds, greens, and glaucophytes are, therefore, surrogate host markers because they have been vertically inherited since the single origin of these taxa. Furthermore, given a single origin of the chromist plastid, then, under the most parsimonious scenario, the Chromista hosts would also be monophyletic (Yoon et al. 2002). Under the model presented here, the lack of a plastid in the early-diverging cryptophytes, in *Goniomonas* spp., and in aplastidial stramenopiles such as oomycetes is regarded in each case as an example of plastid loss (see below [Andersson and Roger 2002]).

#### Divergence Time Estimations

We used the LF method with a “local molecular clock” and the NPRS method using the Powell search algorithm (Sanderson 2003) to calculate divergence dates on the best Bayesian tree using the data set that excluded the third codon positions (i.e., fig. 1A). In addition, 696 of the 1,800 trees that were retained after chain convergence in the Bayesian MCMCMC sampling procedure had a topology identical to the best Bayesian tree. These 696 trees were also used for dating using the LF method, thereby incorporating uncertainty about the evolutionary model parameter estimates and the resulting branch lengths in this procedure. To calibrate the nodes in these trees, we chose six reliable fossil dates that correspond to the radiation of the major algal/plant lineages and a maximum age (i.e., upper bound) for all other divergence date estimates (fig. 1A). We could, however, estimate this node in our analyses. The maximum age constraint  $a$  was a date of 3,500 MYA that marks the presence of the first

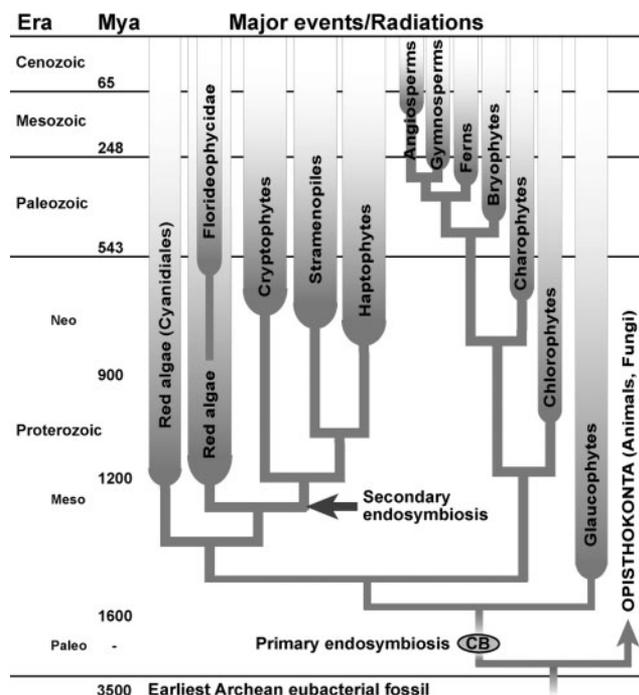


FIG. 3.—Schematic representation of the evolutionary relationships and divergence times for the red, green, glaucophyte, and chromist algae. These photosynthetic groups are outgroup-rooted with the Opisthokonta which putatively ancestrally lacked a plastid. The branches on which the cyanobacterial (CB) primary and red algal chromist secondary endosymbioses occurred are shown.

fossils in the Archean (Schopf et al. 2002; Westall et al. 2001 [but see Brasier et al. 2002 and Garcia-Ruiz et al. 2003]). To address the possibility of pre-Archean life (>3,500 MYA), we also constrained node a with a date of 4,400 MYA that corresponds to be the earliest evidence for a continental crust and oceans on Earth (Wilde et al. 2001). Because both 3,500 MYA and 4,400 MYA constraints gave essentially the same results (e.g., 1,719 vs. 1,720 MYA [node a] and 1,452 vs. 1,453 MYA [node 2] for the 3,500 and 4,400 MYA constraints, respectively), we used the former age in the results presented below. The second node b was constrained at 1,174–1,222 MYA based on the well-preserved fossil of a multicellular *Bangia*-type red alga (*Bangiomorpha*) from rocks dated to this time (Butterfield 2001). Third, we fixed node c at a date of 595–603 MYA based on the Doushantuo Florideophycidae red algal fossils from this time that have reproductive structures (i.e., carposporangia and spermatangia) typical for advanced members of this lineage (Barfod et al. 2002; Xiao, Zhang, and Knoll 1998). We set the four nodes, d–g, in the green lineage with a date of 432–476 MYA for the first appearance of land plants (Kenrick and Crane 1997), 355–370 MYA for seed plant origin (Gillespie, Rothwell, and Scheckler 1981), 290–320 MYA for the split of gymnosperms and the stem lineage leading to extant angiosperms in the Carboniferous (Goremykin, Hansmann, and Martin 1997; Doyle 1998; Bower, Coat, and dePamphilis 2000), and 90–130 MYA for the monocot and eudicot divergence (Crane, Friis, and Pedersen 1995), respectively.

Under these seven constraints and using the LF method, we estimated the split of the red and green algae to have occurred 1,474 MYA on the best Bayesian tree (marked with 1 in fig. 1A; see fig. 1B for the 95% confidence interval). The split of *Cyanophora paradoxa* from the red–green lineage is dated at 1,558 MYA. These results suggest that the primary endosymbiosis in which a nonphotosynthetic eukaryote engulfed a cyanobacterial-like prokaryote and retained it as a cellular organelle (Bhattacharya and Medlin 1995; Delwiche and Palmer 1997), occurred sometime before 1,558 MYA. Our estimate for the date of the split of the glaucophyte from the red and green algae is consistent with a previous molecular clock study that used nuclear multi-gene data to estimate a date of  $1,576 \pm 88$  MYA for the unresolved three-way split of plants, animals, and fungi (see fig. 3 in Wang, Kumar, and Hedges 1999). This age is, however, considerably older than other estimates such as 1,200 MYA and 1,342–1,392 MYA for the split of plants and animals (Feng, Cho, and Doolittle 1997 and Nei, Xu, and Glazko 2001, respectively). Nei, Xu, and Glazko (2001) also estimated an age of 1,578–1,717 MYA for the split of protists (mostly *Plasmodium* data) from the plant-animal-fungal clade. Although it would be very useful to directly compare our estimate to those cited above, the vast differences in the taxon sampling (i.e., our study and other more recent trees are far more species-rich) and phylogenetic hypotheses between these studies make this comparison difficult (see below).

Recent phylogenetic studies with broader taxon sampling suggest that the Plantae are either sister to the chromalveolates (i.e., Chromista and Alveolata [Cavalier-Smith 1999; Fast et al. 2001; Yoon et al. 2002; Harper and Keeling 2003; Bhattacharya, Yoon, and Hackett 2004]) plus Discicristata (i.e., Euglenozoa, Kinetoplastida, and Heterolobosea [Baldauf et al. 2000; Baldauf 2003]) or alternatively, they are paraphyletic, with the greens being most closely related to the chromalveolates and the Discicristata (Nozaki et al. 2003). The second scenario posits primary plastid loss in the common ancestors of the chromalveolates and the Discicristata with subsequent secondary plastid gains in some members of these lineages. The finding of a cyanobacterial-type 6-phosphogluconate dehydrogenase gene (*gnd*) in the non-photosynthetic Heterolobosea (Andersson and Roger 2002) is consistent with this model. The phylogenetic positions of the potentially early-diverging diplomonads and the parabasalids, however, remain to be determined. Regardless of which scenario is correct, these analyses both place the cyanobacterial primary endosymbiosis near the root of the eukaryotic tree, with this event occurring shortly after the split of the Plantae (sensu Nozaki et al. 2003) from the animals and fungi (Opisthokonta [Baldauf et al. 2000; Baldauf 2003; Nozaki et al. 2003]). The primary endosymbiosis must, therefore, have occurred after the split of the Plantae from the opisthokonta and prior to the divergence of the Glaucophyta (see fig. 3). Our molecular clock estimate of 1,558 MYA as the split of the glaucophyte from the red and green algae therefore supports a “late Paleoproterozoic” origin for the primary plastid endosymbiont in the eukaryotic tree of life

(see figure 3). This endosymbiotic event therefore appears to have occurred relatively soon after eukaryotic origin.

Our results also show that the earliest possible date for the putative single secondary endosymbiosis in the Chromista (fig. 1, node 3), in which a non-photosynthetic protist captured a red algal plastid is 1,274 MYA, after the split of the Cyanidiales from the other red algae 1,370 MYA (fig. 1, node 2). This date is consistent with a more limited molecular clock analysis that placed the chromist endosymbiotic event at  $1,261 \pm 28$  MYA (Yoon et al. 2002). The monophyly of chromalveolate plastids (Cavalier-Smith 1999) is supported by recent studies (Fast et al. 2001; Yoon et al. 2002; Harper and Keeling 2003); therefore, it is likely that the alveolates diverged sometime after 1,274 MYA, before the split of the cryptophytes in the Chromista. The stramenopiles and haptophytes split 1,047 MYA (fig. 1, node 5) after the cryptophyte divergence (1,189 MYA; fig. 1, node 4). Each of the chromist lineages in our analyses radiated early in the Neoproterozoic (e.g., 805 MYA for haptophytes, 754 MYA for stramenopiles, and 704 MYA for cryptophytes; fig. 3). These estimates are younger bounds because of the absence of plastid-less forms such as oomycetes and bicoseocids (stramenopiles) in our tree; therefore, the radiation of chromist taxa could potentially go further back into the Neoproterozoic. We estimate the divergence of the charophyte, *Chaetosphaeridium globosum* (Coleochaetales), to have occurred 793 MYA (node 6). Taken together, our data suggest that the split of the glaucophytes from the red and green algae occurred early in the Mesoproterozoic, whereas the latter two groups diverged from each other in the Mesoproterozoic and radiated in the Neoproterozoic.

To test the LF divergence time estimates in which we specified 12 “local rates” in the tree, we also used the NPRS method to accommodate rate inconstancy (Sander-son 1997). The estimated divergence dates using NPRS are older than those using the LF method; however, these differences are relatively minor—e.g., 1,354 MYA for the chromist plastid split (node 3) and 1,255 MYA for the cryptophyte plastid split (node 4; see table 2 in the Supplementary Material online). We also assessed the precision of our divergence time estimates using the credible tree set identified by Bayesian inference. The average divergence times (using the LF method) and the 95% confidence intervals of the distributions are very similar to the results using the best Bayesian tree (see figure 1B). This suggests that there is only minor variation in the branch length estimates in the pool of credible trees used in this analysis (see fig. 1 in the Supplementary Material online); finally, the divergence time estimates (fig. 1B) that were inferred from the protein tree (fig. 2) were generally consistent with the results of the DNA-based analyses (fig. 1B; see also fig. 2B in the Supplementary Material online). We used six or five constraints in the protein analyses because node e, which was not consistent between the DNA and protein trees, had to be excluded from these calculations. Two estimates that were markedly different between the DNA- and protein-based approaches were the estimates of node a for the split of the glaucophyte (1,719 MYA [protein] vs. 1,558 MYA [DNA]) from the red and

green algae, and of node 1 for the split of the red and green algae (1,668 MYA [protein] versus 1,474 MYA [DNA]). These results reflect variation in the branch lengths that unite the glaucophyte to the cyanobacterial outgroup and to the remaining algal plastids (see fig. 2). This discordance may be resolved with increased sampling of glaucophytes or the addition of more data to the protein analysis.

#### Agreement with the Fossil Record and Assessment of Alternative Hypotheses

Given that our divergence time estimates are reasonably accurate, then how consistent are these values with the early eukaryotic fossil record? The first convincing eukaryotic fossils are of single-celled, presumably phototrophic eukaryotes (acritarchs attributed to *Tappania* [see TEM analysis of Javaux, [http://gsa.confex.com/gsa/2002AM/finalprogram/abstract\\_41302.html](http://gsa.confex.com/gsa/2002AM/finalprogram/abstract_41302.html)] from the early Mesoproterozoic (1,500 MYA; Javaux, Knoll, and Walter 2001). Thereafter, the *Bangiomorpha* fossil that was found in rocks dated at  $1,198 \pm 24$  MYA provides compelling evidence (but see Cavalier-Smith 2002) for the presence of multicellular, sexual red algae by this time (Butterfield 2001). Because the red algae are not the most anciently diverged photosynthetic eukaryotes (fig. 1), the primary endosymbiosis that gave rise to the first alga must have occurred before 1,200 MYA and probably before 1,500 MYA (i.e., if acritarchs are the remains of marine algae). These fossil dates agree with our molecular clock estimate of about 1,600 MYA (i.e., late Paleoproterozoic) for the origin of the primary plastid in eukaryotes, thereby placing eukaryote origin before this time. Martin et al. (2003) reached a very similar conclusion in their analysis of the fossil and geological record. Our results also agree with the fossil findings of a putative eukaryotic diversification in the very late Mesoproterozoic and Neoproterozoic (Knoll 1992; 2003). An alternative view of eukaryotic origin is provided by the Neoproterozoic snowball Earth hypothesis (Cavalier-Smith 2002; Hoffman et al. 1998) that was proposed because many unambiguously eukaryotic fossils date from about 850 MYA.

We wanted to address two alternative scenarios that are a consequence of the Neoproterozoic hypothesis. The first is that *Bangiomorpha* is not a red alga (because they did not yet exist) but rather an *Oscillatoria*-like cyanobacterium (Cavalier-Smith 2002). Usage of this constraint would, therefore, lead to false, elevated age estimates for the first origin of algae. To address this issue, we released only the *Bangiomorpha* constraint ( $1,198 \pm 24$  MYA; fig. 1A, node b) and recalculated the dates. Without this constraint, the red–green algal split was estimated at 1,452 MYA (LF method) with a confidence interval of 1,401–1,519 MYA, and the chromist endosymbiosis was 1,255 MYA (12,048–1,302 MYA). Recalculating the date for node b using the six remaining constraints showed a date of 1,156 MYA (1,116–1,199 MYA). These calculations indicate that the *Bangiomorpha* fossil date (regardless of whether the organism is a red alga or a prokaryote) does not have a seriously misleading influence on our

estimation procedure; rather, our clock calculations recover a date for node b that is close to this constraint (1,198 vs. 1,156 MYA) when it is removed from the analysis. The second scenario we addressed is the hypothetical origin of eukaryotes 850 MYA (Cavalier-Smith 2002; Hoffman et al. 1998). Here, we forced node a in figure 1A to be constrained at a maximum age of 850 MYA (instead of 3,500 MYA), excluded the 1,198 MYA *Bangiomorpha* constraint, and recalculated specific divergence times. Under these conditions, when we also released the Florideophycidae constraint (node c) and calculated this date, the age was found to be 342 MYA (327–359 MYA) rather than the reliable fossil date of  $599 \pm 4$  MYA (see table 2 in the Supplementary Material online). These results suggest that forcing the snowball Earth hypothesis onto our phylogeny results in underestimates of divergence times.

Our estimate for the split of the haptophytes and stramenopiles 1,047 MYA (fig. 1) contrasts with a previous analysis done by Medlin et al. (1997), who assumed (based on available data) that the origin of photosynthesis in these groups all occurred via independent red algal secondary endosymbioses (see also Oliveira and Bhattacharya 2000). Their calculations supported plastid origins in haptophytes and stramenopiles at or before the Permian-Triassic boundary 250 MYA (Medlin et al. 1997). A critical difference in our approach is that we assumed, based primarily on multi-gene phylogenetic evidence and a unique GAPDH gene duplication that is shared by chromalveolates, a monophyletic origin of chromist plastids (Cavalier-Smith 1986; Fast et al. 2001; Yoon et al. 2002; Harper and Keeling 2003; fig. 1A). This implies that the common ancestor of the Chromista (not just the later-diverging photosynthetic members) contained the red algal secondary plastid. Consistent with this view, a recent study has shown that the *gnd* gene in *Phytophthora* (Oomycota) is closely related to the homolog of cyanobacterial origin in photosynthetic stramenopiles, supporting the presence of the red algal secondary endosymbiont in *Phytophthora* and *gnd* origin through gene transfer (Andersson and Roger 2002). In contrast, Medlin et al. (1997) rooted their stramenopile nuclear SSU rDNA tree using the nonphotosynthetic oomycetes as the outgroup. The origin of the photosynthetic stramenopiles in their analysis would therefore represent a more recent within-group divergence and not the timing of plastid origin. Interestingly, the haptophyte divergence in the linearized host nuclear SSU rDNA tree used by Medlin et al. (1997) was found to be between 850–ca. 1,750 MYA. Given a photosynthetic ancestor of the haptophytes, these values bracket our date of 1,047 MYA for the haptophyte-stramenopile split in the plastid multi-gene tree.

#### The Long Pause in Algal Radiation

Assuming that our results (and the Paleoproterozoic model) are correct, we are left with an important problem, explaining the presence of algae significantly earlier than the eukaryotic diversification documented in Neoproterozoic fossils (Anbar and Knoll 2002). We believe that this discordance likely reflects a combination of factors. First,

as mentioned above, the first appearance of a fossil is almost always an underestimate of the actual age of the lineage because of the incompleteness of the record (Knoll 1992). Second, if early-diverging forms do not contain a mineralized exoskeleton (e.g., coccoliths in haptophytes [Graham and Wilcox 2000]), then they may not be fossilized, also resulting in an underestimate of the age of the lineage. Third, the first origin and diversification of algal groups may not have been coincident. Early red and green algae may have been unable to radiate 1,500 MYA because of physical factors such as nutrient conditions or tropic competition. Anbar and Knoll (2002) suggested that low nitrogen availability (which is critical for algal growth) that resulted from anoxic and sulfidic oceans may have limited algal diversification in the mid-Proterozoic. Alternatively, Martin et al. (2003) have suggested that low anoxia and high sulfide may themselves have been the major factors limiting the diversification of the first eukaryotes. In either case, these conditions were ameliorated by extensive weathering around 1,250 MYA, potentially laying the foundation for the Neoproterozoic algal radiation seen in the fossil record and in our molecular clock analyses (fig. 3).

#### Supplementary Material

The GenBank accession numbers for the 42 new plastid sequences generated in this study are listed in table 1 of the Supplementary Material online. The six-gene alignment used in the phylogenetic analyses is available on request from D.B.

#### Acknowledgments

This work was supported by grants from the National Science Foundation awarded to D.B. (DEB 01-07754, MCB 02-36631). We thank Kori Osborne for technical assistance and J. Frankel, J. Comeron, and two anonymous reviewers for critical reading of the manuscript.

#### Literature Cited

- Anbar, A. D., and A. H. Knoll. 2002. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? *Science* **297**:1137–1142.
- Andersson, J. O., and A. J. Roger. 2002. A cyanobacterial gene in nonphotosynthetic protists—an early chloroplast acquisition in eukaryotes? *Curr. Biol.* **12**:115–119.
- Baldauf, S. L. 2003. The deep roots of eukaryotes. *Science* **300**:1703–1706.
- Baldauf, S. L., and J. D. Palmer. 1990. Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. *Nature* **344**:262–265.
- Baldauf, S. L., J. R. Manhart, and J. D. Palmer. 1990. Different fates of the chloroplast *tufA* gene following its transfer to the nucleus in green algae. *Proc. Natl. Acad. Sci. USA* **87**:5317–5321.
- Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. *Science* **290**:972–977.
- Barfod, G. H., F. Albareda, A. H. Knoll, S. Xiao, P. Telouk, R. Frei, and J. Baker. 2002. New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils. *Earth Planet. Sci. Lett.* **201**:203–212.

- Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. *Science* **300**:1698–1700.
- Bhattacharya, D., and L. Medlin. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. *J. Phycol.* **31**:489–498.
- Bhattacharya, D., and K. Weber. 1997. The actin gene of the Glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. *Curr. Genet.* **31**:439–446.
- Bhattacharya, D., H. S. Yoon, and J. D. Hackett. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. *BioEssays*: **26**:50–60.
- Bowe, L. M., G. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA* **97**:4092–4097.
- Brasier, M. D., O. R. Green, A. P. Jephcoat, A. K. Kleppe, M. J. Van Kranendonk, J. F. Lindsay, A. Steele, and N. V. Grassineau. 2002. Questioning the evidence for Earth's oldest fossils. *Nature* **416**:76–81.
- Butterfield, N. J. 2001. Paleobiology of the late Mesoproterozoic (ca. 1200 Ma) hunting formation, Somerset Island, Arctic Canada. *Precam. Res.* **111**:235–256.
- Cavalier-Smith, T. 1986. The kingdom Chromista: origin and systematics. Pp. 309–347 in F. E. Round and D. J. Chapman, eds., *Progress in phycolgical research*. Biopress, Bristol, U.K.
- . 1998. A revised six-kingdom system of life. *Biol. Rev. Camb. Philos. Soc.* **73**:203–266.
- . 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.
- . 2002. The neomuran origin of archaeobacteria, the negibacterial root of the universal tree and bacterial megaclassification. *Int. J. Syst. Evol. Microbiol.* **52**:7–76.
- Crane, P. R., E. M. Friis, and K. R. Pedersen. 1995. The origin and early diversification of angiosperms. *Nature* **374**:27–33.
- Cutler, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. *Mol. Biol. Evol.* **17**:1647–1660.
- Delsuc, F., M. J. Phillips, and D. Penny. 2003. Comment on “Hexapod origins: monophyletic or paraphyletic?” *Science* **301**:1482.
- Delwiche, C. F., and J. D. Palmer. 1997. The origin of plastids and their spread via secondary symbiosis. Pp. 53–86 in D. Bhattacharya, ed., *Origins of algae and their plastids*. Springer-Verlag, Vienna, Austria.
- Doyle, J. A. 1998. Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. *Mol. Phylogenet. Evol.* **9**:448–462.
- Fast, N. M., J. C. Kissinger, D. S. Roos, and P. J. Keeling. 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 the bootstrap. *Evolution* **39**:783–791.
- . 2002. PHYLIP (Phylogeny Inference Package) 3.6. Department of Genetics, University of Washington, Seattle, Wash.
- Feng, D. F., G. Cho, and R. F. Doolittle. 1997. Determining divergence times with a protein clock: update and reevaluation. *Proc. Natl. Acad. Sci. USA* **94**:13028–13033.
- Garcia-Ruiz, J.M., S. T. Hyde, A. M. Carerup, A. G. Christy, M. J. Van Kranendonk, and N. J. Welham. 2003. Self-assembled silica-carbonate structures and detection of ancient microfossils. *Science* **302**:1194–1197.
- Gilbert, D. G. 1995. SeqPup, A biological sequence editor and analysis program for Macintosh computer. Indiana University, Bloomington.
- Gillespie, W. H., G. W. Rothwell, S. E. Scheckler. 1981. The earliest seeds. *Nature* **293**:462–464.
- Goremykin, V. V., S. Hansmann, and W. F. Martin. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. *Plant Syst. Evol.* **206**:337–351.
- Graham, L. D., and L. W. Wilcox. 2000. *Algae*. Prentice-Hall, Upper Saddle River, N.J.
- Gray, M. W., B. F. Lang, and R. Cedergren et al. (15 co-authors). 1998. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res.* **26**:865–878.
- Harper, J. T., and P. J. Keeling. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAPDH) indicates a single origin for chromalveolate plastids. *Mol. Biol. Evol.* **20**:1730–1735.
- Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**:1129–1133.
- Hoffman, P. F., A. J. Kaufman, G. P. Halverson, and D. P. Schrag. 1998. A Neoproterozoic snowball earth. *Science* **281**:1342–1346.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**:754–755.
- Javaux, E. J., A. H. Knoll, and M. R. Walter. 2001. Morphological and ecological complexity in early eukaryotic ecosystems. *Nature* **412**:66–69.
- Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche. 2001. The closest living relatives of land plants. *Science* **294**:2351–2353.
- Kenrick, P., and P. R. Crane. 1997. The origin and early evolution of plants on land. *Nature* **389**:33–39.
- Knoll, A. H. 1992. The early evolution of eukaryotes: a geological perspective. *Science* **256**:622–627.
- . 2003. *Life on a young planet*. Princeton University Press, Princeton, N.J.
- Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny. 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.
- Martin, W., C. Rotte, M. Hoffmeister, U. Theissen, G. Gelius-Dietrich, S. Ahr, and K. Henze. 2003. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. *IUBMB Life* **55**:193–204.
- Maul, J. E., J. W. Lilly, L. Cui, C. W. dePamphilis, W. Miller, E. H. Harris, and D. B. Stern. 2002. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. *Plant Cell* **14**:2659–2679.
- Medlin, L. K., W. H. C. F. Kooistra, D. Potter, G. W. Saunders, and R. A. Andersson. 1997. Phylogenetic relationships of the “golden algae” (haptophytes, heterokont chromophytes) and their plastids. Pp. 187–219 in D. Bhattacharya, ed., *Origins of algae and their plastids*. Springer-Verlag, Vienna, Austria.
- Moreira, D., H. Le Guyader, and H. Phillippe. 2000. The origin of red algae and the evolution of chloroplasts. *Nature* **405**:69–72.
- Nei, M., P. Xu, and G. Glazko. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. *Proc. Natl. Acad. Sci. USA* **98**:2497–2502.

- Nozaki, H., M. Matsuzaki, M. Takahara, O. Misumi, H. Kuroiwa, M. Hasegawa, I. T. Shin, Y. Kohara, N. Ogasawara, and T. Kuroiwa. 2003. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. *J. Mol. Evol.* **56**:485–497.
- Ohta, N., M. Matsuzaki, O. Misumi, S. Y. Miyagishima, H. Nozaki, K. Tanaka, T. Shin-I, Y. Kohara, and T. Kuroiwa. 2003. Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. *DNA Res.* **10**:67–77.
- Oliveira, M. C., and D. Bhattacharya. 2000. Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. *Am. J. Bot.* **87**:482–492.
- Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. *Mol. Phylogenet. Evol.* **28**:171–185.
- Pinto, G., P. Albertano, C. Ciniglia, S. Cozzolino, A. Pollio, H. S. Yoon, and D. Bhattacharya. 2003. Comparative approaches to the taxonomy of the genus *Galdieria merola* (Cyanidiales, Rhodophyta). *Cryptogamie Algol.* **24**:13–32.
- Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- Sanderson, M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**:1218–1231.
- . 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**:301–302.
- Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**:502–504.
- Schopf, J. W., A. B. Kudryavtsev, D. G. Agresti, T. J. Wdowiak, and A. D. Czaja. 2002. Laser-Raman imagery of Earth's earliest fossils. *Nature* **416**:73–76.
- Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. USA* **99**:4430–4435.
- Swofford, D. L. 2002. PAUP\*: Phylogenetic analysis using parsimony (\* and other methods) 4.0b8. Sinauer Associates, Sunderland, Mass.
- Valentin, K., and K. Zetsche. 1990. Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. *Plant Mol. Biol.* **15**:575–584.
- Wang, D. Y., S. Kumar, and S. B. Hedges. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. R. Soc. Lond. Ser. B. Biol. Sci.* **266**:163–171.
- Westall, F., M. J. De Witb, J. Dann, S. Van Der Gaast, C. E. J. De Ronded, and D. Gemeke. 2001. Early Archean fossil bacteria and biofilms in hydrothermally-influenced sediments from the Barberton greenstone belt, South Africa. *Precam. Res.* **106**:93–116.
- Whelan, S., P. Liò, and N. Goldman. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. *Trends Genet.* **17**:262–272.
- Wilde, S. A., J. W. Valley, W. H. Peck, and C. M. Graham. 2001. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**:175–178.
- Xiao, S., Y. Zhang, and A. H. Knoll. 1998. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. *Nature* **391**:553–558.
- Yoon, H. S., J. D. Hackett, and D. Bhattacharya. 2002. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. *Proc. Natl. Acad. Sci. USA* **99**:11724–11729.
- Yoon, H. S., J. D. Hackett, G. Pinto, and D. Bhattacharya. 2002. The single, ancient origin of chromist plastids. *Proc. Natl. Acad. Sci. USA* **99**:15507–15512.

Dan Graur, Associate Editor

Accepted December 9, 2003