Current Biology, Vol. 15, 1325–1330, July 26, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.cub.2005.06.040

Monophyly of Primary Photosynthetic Eukaryotes: Green Plants, Red Algae, and Glaucophytes

Naiara Rodríguez-Ezpeleta,[1] Henner Brinkmann,[1] Suzanne C. Burey,[2] Béatrice Roure,[1] Gertraud Burger,[1] Wolfgang Löffelhardt,[2] Hans J. Bohnert,[3] Hervé Philippe,[1,*] and B. Franz Lang[1,*]

[1]Canadian Institute for Advanced Research
Centre Robert Cedergren
Département de Biochimie
Université de Montréal
2900 Boulevard Edouard-Montpetit
Montréal, Québec, H3T 1J4
Canada

[2]Max L. Perutz Laboratories
University Departments at the Vienna Biocenter
Department of Biochemistry and Ludwig Boltzmann Research Unit for Biochemistry
9 Dr. Bohr-Gasse
Vienna, A-1030
Austria

[3]Department of Plant Biology and Crop Sciences
University of Illinois Urbana/Champaign
1201 W. Gregory Drive
Urbana, Illinois 61801

Summary

Between 1 and 1.5 billion years ago [1, 2], eukaryotic organisms acquired the ability to convert light into chemical energy through endosymbiosis with a Cyanobacterium (e.g., [3–5]). This event gave rise to "primary" plastids, which are present in green plants, red algae, and glaucophytes ("Plantae" sensu Cavalier-Smith [6]). The widely accepted view that primary plastids arose only once [5] implies two predictions: (1) all plastids form a monophyletic group, as do (2) primary photosynthetic eukaryotes. Nonetheless, unequivocal support for both predictions is lacking (e.g., [7–12]). In this report, we present two phylogenomic analyses, with 50 genes from 16 plastid and 15 cyanobacterial genomes and with 143 nuclear genes from 34 eukaryotic species, respectively. The nuclear dataset includes new sequences from glaucophytes, the less-studied group of primary photosynthetic eukaryotes. We find significant support for both predictions. Taken together, our analyses provide the first strong support for a single endosymbiotic event that gave rise to primary photosynthetic eukaryotes, the Plantae. Because our dataset does not cover the entire eukaryotic diversity (but only four of six major groups in [13]), further testing of the monophyly of Plantae should include representatives from eukaryotic lineages for which currently insufficient sequence information is available.

*Correspondence: franz.lang@umontreal.ca (B.F.L.); herve.philippe@umontreal.ca (H.P.)

Results and Discussion

Plastid Genes Significantly Support Plastid Monophyly

The monophyly of plastids is supported by several common features, such as a similar gene content of plastid genomes, the presence of plastid-specific gene clusters that are distinct from those in Cyanobacteria, the conservation of the plastid-protein import machinery and protein-targeting signals, and phylogenies based on plastid and cyanobacterial gene sequences (see [5] and references therein). Yet, some authors have challenged each of these evidences as either weak or inconclusive [14]. In particular, the molecular phylogenies are often based on single or a few genes and are not robust (e.g., [9, 10]). One published multigene phylogeny recovers plastid monophyly, but it includes only a few cyanobacterial taxa [15].

Our analyses are the first to include in a phylogenomic framework data from a broad diversity of Cyanobacteria and other related Bacteria for testing plastid monophyly. Our dataset contains 50 proteins (10,334 amino acid positions) from 16 plastids and 15 Cyanobacteria; 13 additional bacteria (10 Gram-positive bacteria, *Deinococcus*, *Thermus*, and *Chloroflexus*) were added to this dataset for a reduced number of proteins (26 proteins totaling 4,998 amino acid positions) in order to determine the root. Four different phylogenetic inference methods were employed: maximum likelihood (with a concatenate [cML] and a separate [sML] model), Bayesian inference (BI), maximum-likelihood-based distance (Dml), and maximum parsimony (MP). As shown in Figure 1, plastids form a strongly supported monophyletic group (100% bootstrap value [BV]). Within plastids, the relationships among green plants, red algae, and glaucophytes remain unresolved at standard confidence levels. It has been proposed that systematic errors such as long-branch attraction (LBA), compositional bias, and covarion structures are responsible for the recovery of plastid monophyly [16, 17]. We therefore performed analyses with LogDet distances [18], a covarion model [19, 20], and by including only the slowest evolving plastids—*Porphyra* (red alga), *Mesostigma* (green plant), and *Cyanophora* (glaucophyte). All tests of possible artifacts as suggested by Lockhart and coworkers did not affect the strong support for plastid monophyly. Horizontal gene transfer (HGT) is obviously another major concern in cyanobacterial phylogeny [21], but does not seem to affect our results (see the Supplemental Experimental Procedures in the Supplemental Data available with this article online).

Interestingly, with a dataset including 13 additional bacteria as an outgroup (Figure S1), *Gloeobacter* is the deepest branch within Cyanobacteria, consistent with seemingly "primitive" features of the photosynthetic apparatus (see [9] and references therein). However, it cannot be excluded that this is due to LBA and that *Gloeobacter* is highly derived and not early diverging. Similarly, because plastids are fast evolving relative to

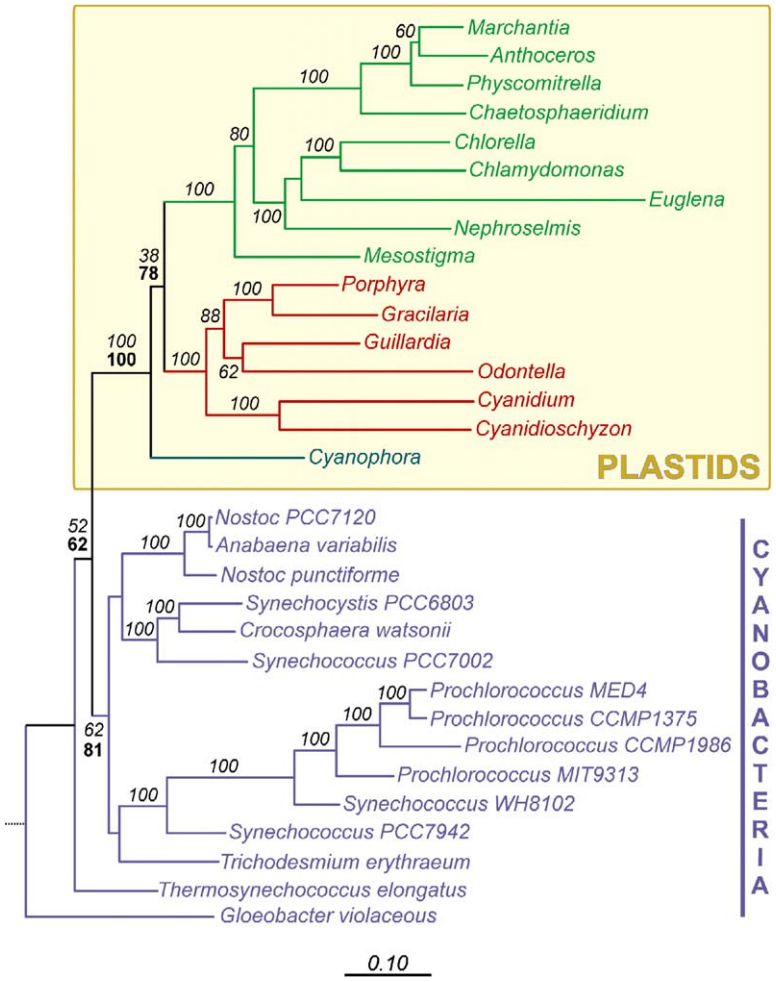

Figure 1. Phylogeny based on plastid and bacterial protein sequences

The analysis is based on the concatenated dataset of plastid-encoded proteins (50 proteins; 10,334 amino acid positions). The tree has been inferred with BI with the WAG+F+Γ model. Numbers in italics represent support values obtained with 100 bootstrap replicates on the concatenated dataset with PhyML (WAG+F+Γ model), and numbers below (in bold) represent bootstrap values based on 10,000 RELL replicates of the sML analysis (see Experimental Procedures for details). The presence of a single value indicates that this branch was constrained in the separate analysis (except for the position of *Euglena*, which was also constrained). The scale bar denotes the estimated number of amino acid substitutions per site. Bootstrap values lower than 50% obtained in both approaches are not shown. The dotted line indicates the position of the root, which was inferred with a dataset of 26 plastid proteins including 13 additional bacteria (see Figure S1). Species names in certain colors denote the following: in green, green plants plus the secondary-plastid-containing *Euglena*; in red, red algae and secondary-plastid-containing *Odontella* and *Guillardia*; and in blue, glaucophytes. Taxon designations are as follows: *Anthoceros formosae*, *Marchantia polymorpha*, *Physcomitrella patens*, *Chaetosphaeridium globosum*, *Euglena gracillis*, *Chlamydomonas reinhardtii*, *Chlorella vulgaris*, *Nephroselmis olivacea*, *Mesostigma viride*, *Cyanidioschyzon merolae*, *Cyanidium caldarium*, *Odontella sinensis*, *Guillardia theta*, *Porphyra purpurea*, *Gracilaria tenuistipitata*, and *Cyanophora paradoxa*. *Prochlorococcus* stands for *Prochlorococcus marinus*. Note that we have not included *Synechococcus* PCC6301 because it is closely related if not identical with *Synechococcus PCC7942*.

most Cyanobacteria, they might be attracted toward the root of the tree by LBA. Genome projects on potentially basally diverging Cyanobacteria (e.g., *Pseudanabaena*; [9]) and improved tree-inference methods are required to resolve these questions with confidence.

Nuclear Genes Significantly Support the Monophyly of Plantae

The monophyly of Plantae has been tested with phylogenies that use nuclear and mitochondrial sequences, but support is weak (e.g., [7, 8, 11, 12]). Strong support for the sister-group relationship of green plants and red algae has been obtained in multiprotein phylogenies, one with 13 nuclear proteins [22] and the other with four mitochondrial proteins [23]. However, the nuclear tree has nonsignificant support for the monophyly of Plantae when the then-available six glaucophyte-protein sequences are included, whereas the mitochondrial phylogeny does not include glaucophytes. In addition, the exclusion of only one protein (elongation factor 2) from the nuclear dataset or the use of alternative mitochondrial datasets (*nad* versus *cob*/*cox* genes) drastically reduces support for the sister-group relationship of green plants and red algae [7, 24]. The use of a limited number of genes is a possible explanation for the lack of significant support for or against the monophyly of Plantae. Indeed, it is well documented that single gene sequences often do not contain sufficient phylogenetic signal to resolve short internal branches, even at moderately deep divergence.

We have therefore performed phylogenetic analyses based on a dataset of 143 orthologous nuclear proteins (30,113 amino acid positions) from 39 species. To overcome the lack of data from glaucophytes, the less-studied group of primary photosynthetic eukaryotes, we have sequenced 4,628 and 8,696 expressed sequence tags (ESTs) from *Cyanophora paradoxa* and *Glaucocystis nostochinearum*, respectively. Our dataset represents all major eukaryotic groups for which sufficient sequence information is available, i.e., not including members of two major, potentially polyphyletic groups Rhizaria (Cercozoa, Radiolaria, Foraminifera, etc.) and excavates (jakobids, malawimonads, Heterolobosea, etc). The monophyly of Plantae remains to be tested with respect to these missing groups. Analyses including diplomonads, parabasalids, and kinetoplastids, the only excavates for which enough data are available, demonstrate that these excavates are fast

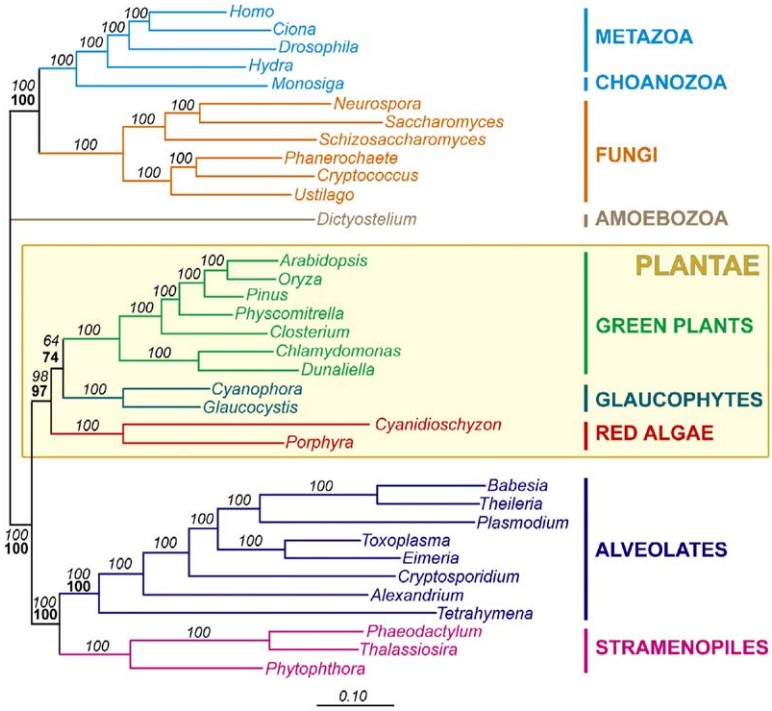

Figure 2. Phylogeny Based on Nuclear-Encoded Protein Sequences

The analysis is based on the concatenated dataset of nuclear encoded proteins (143 proteins; 30,113 amino acid positions). The posterior probabilities for all branches are 1.0. For further details, see Figure 1. The tree is rooted between opisthokonts and other eukaryotes (excluding *Dictyostelium*), a proposal based on the presence/absence of a gene fusion [42, 43]. The position of *Dictyostelium* cannot be deduced with this gene-fusion event because it has lost the corresponding homologous genes. Therefore, a basal trifurcation with *Dictyostelium*, opisthokonts and other eukaryotes is shown. Taxon designations are as follows: *Homo sapiens*, *Ciona intestinalis*, *Drosophila melanogaster*, *Hydra magnipapillata*, *Monosiga brevicollis*, *Candida albicans*, *Saccharomyces cerevisiae*, *Neurospora crassa*, *Phanerochaete chrysosporium*, *Cryptococcus neoformans*, *Ustilago maydis*, *Arabidopsis thaliana*, *Oryza sativa*, *Pinus taeda*, *Physcomitrella patens*, *Closterium peracerosum-strigosum-littorale complex*, *Chlamydomonas reinhardtii*, *Dunaliella salina*, *Cyanidioschyzon merolae*, *Porphyra yezoenesis*, *Cyanophora paradoxa*, *Glaucocystis nostochinearum*, *Babesia bovis*, *Theileria annulata*, *Plasmodium falciparum*, *Eimeria tenella*, *Toxoplasma gondii*, *Cryptosporidium parvum*, *Alexandrium tamarense*, *Phaeodactylum tricornutum*, *Thalassiosira pseudonana*, and *Phytophthora sojae*.

evolving (Figure S4). Because the inclusion of fast-evolving taxa causes phylogenetic artifacts [25, 26], only analyses without these lineages are shown in the following (note, however, that there is no difference in tree topology; Figure 2, Figure S4, and Table S1).

Analyses with the remaining 34 species (cML, sML, and BI, Figure 2; and MP and Dml, not shown) significantly support the monophyly of Plantae and of all other relationships except one (see below), indicating the presence of a strong signal in our dataset. In the cML analysis, the monophyly of all major lineages (e.g., animals, fungi, green plants, alveolates, and stramenopiles [13]) are confirmed with bootstrap values of 100%, which corroborates numerous previous analyses. In addition, the superkingdom Opisthokonta, including Fungi and Holozoa (Metazoa and the choanoflagellate *Monosiga brevicollis*), is recovered at 100%, as are the superensemble Alveolata uniting Apicomplexa, Ciliophora, and Dinoflagellata (100%) and a clade uniting stramenopiles and alveolates (100%). Finally, in the cML analysis, the support value for the monophyly of Plantae is significant (98%). Our inferences with sML, which fits the data best (Table S2) also recover the monophyly of Plantae with high confidence (97% BV; Figure 2). However, the relationships among the three groups of Plantae remain unsupported (64% and 74% BV for the sister-group of green plants and glaucophytes, with the cML and sML approaches, respectively).

To assess the confidence level for the monophyly of Plantae, we retained the best 25 topologies from the exhaustive sML analysis (see Experimental Procedures) and performed several statistical tests. All tests gave essentially the same results; the least-biased and most-rigorous test available to date, the "approximately unbiased" (AU) test, is shown [27, 28] for seven of the most relevant topologies tested (Table 1). The AU test rejects all scenarios in which Plantae are not monophyletic (significance level = 0.05). The relationships within the Plantae remain unresolved, although the sisterhood of glaucophytes and green plants has the highest probability, in agreement with the results of the bootstrap analyses shown in Figure 2. Interestingly, removal of the fast-evolving *Cyanidioschyzon* renders the three alternative arrangements among the lineages of Plantae almost identical (see column AU-33 in Table 1). A detailed discussion of the impact of taxon sampling in phylogenomics will be presented in a separate study (N.R.-E. et al., unpublished data).

How Many Genes Does It Take to Resolve the Monophyly of Plantae?

The above-presented analyses are the first that strongly support the monophyly of Plantae. To verify whether a large number of genes is indeed required to obtain this result, we calculated for each internal branch in Figure 2 the bootstrap values as a function of the number of amino acid positions used (Figure 3). With ~8000 amino acid positions, all internal branches but two are recovered with a BV > 90%. The monophyly of Plantae is supported with only 70% BV, with the same number of amino acid positions (Figure 3, thick

Table 1. Likelihood Tests of Alternative Tree Topologies

Rank	Tree topology	Δ ln *L*[a]	AU[b]	AU-33[c]
1	Best tree; glaucos with greens	–27.5	0.892	0.575
2	Glaucos with reds	27.5	0.297	0.567
3	Glaucos basal to (reds + greens)	42.8	0.147	0.412
4	Reds basal to (alveos + strams)	84.5	**0.044**	**0.018**
5	Glaucos basal to (dicts + opis)	137.6	**0.006**	**0.007**
11	Greens basal to (alveos + strams)	235.2	**3e-07**	**2e-04**
12	Three Plantae lineages unrelated	238.6	**3e-61**	**8e-30**

Comparison of alternative trees with CONSEL [40], inferred from separate maximum-likelihood analyses of 143 proteins and 34 species, with the same model as the analysis in Figure 2. The 25 best topologies from the sML analysis were retained. In the three best topologies, Plantae are monophyletic. All other 22 topologies are rejected at a significance level of 0.05. Topologies 4, 5, and 11 are the best in which only two Plantae lineages are sister groups, and topology 12 is the best in which the three Plantae lineages are unrelated. The following abbreviations are used: glaucos, glaucophytes; reds, red algae; greens, green plants; alveos, alveolates; strams, stramenopiles; dicts, *Dictyostelium*; and opis, opisthokonts.

[a] Log likelihood difference.

[b] Approximate Unbiased test.

[c] When removing *Cyanidioschyzon* from the dataset (AU-33 column), AU values for the best and the second-best tree become almost identical (0.575 and 0.567), eliminating the marginal support for the sister-group relationship of glaucophytes and green plants.

black line). In fact, the support value of this branch increases slowly but regularly with the addition of positions, finally reaching 90% BV at >20,000 amino acid positions. The sisterhood of green plants and glaucophytes (Figure 3; thick dotted line) also increases slowly with the addition of more data, but it reaches only 74% BV with the complete dataset (Figure 2).

In summary, our results show that 30,000 amino acid positions are necessary to recover the monophyly of Plantae with significant support. This explains why other studies, which all used much fewer sequence positions, did not obtain statistically significant support for this clade (e.g., [8, 22, 24]).

Conclusions

Our phylogenomic analyses support the idea that Plantae are monophyletic and that plastids form a monophyletic group to the exclusion of Cyanobacteria, providing compelling evidence for a single origin of primary photosynthesis in eukaryotes. Still, our large datasets are insufficient to resolve the branching order within Plantae and are thus unable to support or reject the common assumption that glaucophytes emerged prior to the divergence of green plants and red algae. Addressing this issue requires analyses that include more taxa and/or more genes of the three Plantae lineages, in particular red algae and glaucophytes. Furthermore, the monophyly of Plantae remains to be tested after addition of several major eukaryotic groups not included here because of the lack of gene sequences or of their fast rate of evolution. The eukaryotic groups from which data are most urgently required, preferentially from slow-evolving species, are the Rhizaria (including Cercozoa, Radiolaria, Foraminifera, etc.), Amoebozoa (Lobosa and Conosa), and the potentially nonmonophyletic Excavata (Euglenozoa, Heterolobosea, jakobids, malawimonads, diplomonads, parabasalids, retortamonads, etc.).

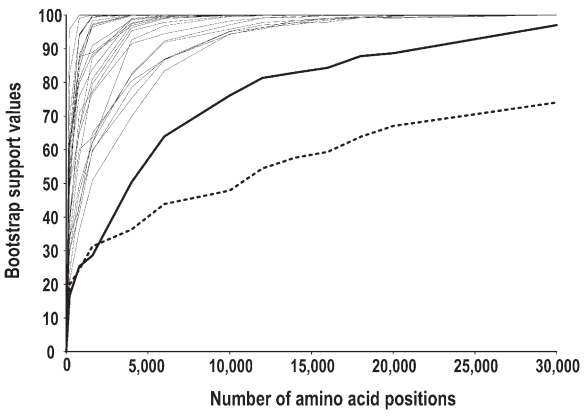

Figure 3. More than 100 Genes Are Required to Recover the Monophyly of Plantae

Evolution of the bootstrap support values (BV) for each internal branch, as a function of the number of amino acid positions. Y and X axes refer to bootstrap values (in %) and number of amino acid positions, respectively. Thick line represents monophyly of Plantae; the dotted line represents green plants + glaucophytes; and the thin lines represent other internal branches in Figure 2.

As we show here, the high support for the monophyly of Plantae critically relies on the use of a large collection of protein sequences. Whereas few relationships (e.g., the sisterhood of animals and fungi) can be convincingly demonstrated already with a small number of sequences (e.g., 13 mitochondrial proteins), we posit that the resolution of other ancient events in the history of eukaryotes will require massive datasets in the order of 100 or more genes [29].

An efficient way to obtain data from many organisms is EST sequencing, which requires limited amounts of cell material and is therefore useful for the exploration of underrepresented eukaryotes, many of which are difficult to grow and unavailable in axenic culture. On the basis of phylogenetic analyses with these data, key species can then be selected for genome projects. Glaucophytes clearly belong to the taxa of prime interest because they contain minimally derived plastids and appear as the slowest-evolving eukaryotes in our phylogenomic analyses (Figure 2). Complete glaucophyte genome sequences would allow for a better understanding of the origin of eukaryotic photosynthesis

while providing deeper insight into the biogenesis of plastids.

Experimental Procedures

Construction of cDNA Libraries, Sequencing, Selection of Orthologous Proteins, and Data Extraction from Multiple Alignments

A detailed description of cDNA-library constructions and sequencing is available with the Supplemental Data. Sequences are available at http://amoebidia.bcm.umontreal.ca/public/pepdb/welcome.php. The nuclear dataset is based on an available alignment [12]. Data from *Cyanophora*, *Glaucocystis* (this study), and additional sequences retrieved from GenBank (http://www.ncbi.nlm.nih.gov) and other sources (see Supplemental Data) were added to the alignment as described [12]. Species evolving at highly accelerated rates (Microsporidia, Euglenozoa, Parabasalida, and Diplomonadida) were not included (but see Figure S4), and only representative (preferentially slowly evolving) members of fungi, animals, and embryophyte plants were used. Unambiguously aligned sequence blocks were extracted with Gblocks [30]; after manual verification, potential paralogs were identified and removed as described [31]. When all orthologous proteins that are available from at least 23 out of the 34 used species are included, the dataset contains 143 proteins (see Supplemental Data for a detailed list), totaling 30,113 amino acid positions. On average, 19% of the amino acids are missing.

The plastid dataset consists of 50 proteins (a total of 10,334 amino acid positions; see Supplemental Data for a detailed list) from 16 plastids and 15 cyanobacteria that were publicly available. The number of land-plant plastids in this data collection was restricted to three slowly evolving species. An alternative plastid dataset including 26 proteins (a total of 4,998 amino acid positions) from 13 additional bacteria was used to root the tree. Sequences were aligned with CLUSTALW [32] and refined manually with MUST [33], and ambiguously aligned positions were removed with Gblocks [30]. The two resulting datasets are available upon request.

Phylogenetic Analyses

The concatenated datasets of nuclear and plastid/cyanobacterial sequences were analyzed by maximum likelihood (ML) with PhyML 2.4 [34], maximum parsimony (MP) with PAUP* 4.0 b10 [35], bayesian inference (130,000 and 120,000 generations for nuclear and plastid dataset respectively, repeated three times with identical results) with MrBayes 3.0 b4 [19], and distance methods with TREE-PUZZLE 5.2 [36] and BIONJ [37]. The reliability of each internal branch was evaluated on the basis of 100 (ML) or 1000 (MP and distance approach) bootstrap replicates. Subsequently, separate ML analyses (sML) were conducted as described [31]. In brief, relationships that are undisputed and supported by 100% bootstrap values (e.g., the monophyly of animals, fungi, and green plants) were constrained, and all resulting tree topologies were exhaustively analyzed independently for each protein to identify the tree topology with the best overall likelihood value (for more details, see Supplemental Data). Site-wise likelihood values were calculated by PAML [38]. The support for each internal branch was evaluated by the RELL method [39], with 10,000 replicates. For likelihood tests of competing tree topologies, p values were calculated with CONSEL [40].

Number of Amino Acid Positions and Bootstrap Support

For sML analysis, the relationship between the number of sequence positions and the bootstrap value was calculated for various internal branches as described [41]. In order to do so, the constraints were adapted to permit testing of groups within Apicomplexa, Fungi, Holozoa, green plants, and stramenopiles. In brief, variable fractions of amino acid positions of the complete dataset (e.g., 1,000; 2,000; 3,000; …; 30,000) were randomly drawn from the dataset, each 100 times. RELL bootstrap analysis was then performed on each of the 100 samples for each size fraction. The average of the bootstrap values for each size fraction was plotted against its size.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, four figures, and two tables and are available with this article online at http://www.current-biology.com/cgi/content/full/15/14/1325/DC1/.

Acknowledgments

We wish to thank Frédéric Delsuc, Nicolas Rodrigue, and three anonymous reviewers for helpful comments on the manuscript; David To for his assistance in configuring the parallel version of MrBayes; and the PEP sequencing team at the Université de Montréal for the supply of *Glaucocystis* data. This work has been supported by operating and equipment funds from Genome Quebec/Canada. B.F.L., G.B., and H.P. are members of the Program in Evolutionary Biology of the Canadian Institute for Advanced Research (CIAR), whom we thank for salary and interaction support. The authors H.P. and B.F.L. are grateful to the Canada Research Chairs Program and the Canadian Foundation for Innovation (CFI) for salary and equipment support. H.J.B. acknowledges support from the National Science Foundation, USA, DBI-9813360 and DBI-0223905, and thanks the W.M. Keck Center for Comparative and Functional Genomics at the University of Illinois, Urbana-Champaign, for DNA sequencing services (ESTs of *Cyanophora*). We are also grateful to the Austrian Research Fund for financing grant P15438 (to W.L.). N.R.E. has been supported by "Programa de Formación de Investigadores del Departamento de Educación, Universidades e Investigación" (Government of Basque Country), and B.R. is a CIHR strategic training fellow in Bioinformatics.

Received: February 13, 2005
Revised: June 7, 2005
Accepted: June 9, 2005
Published: July 26, 2005

References

1. Douzery, E.J., Snell, E.A., Bapteste, E., Delsuc, F., and Philippe, H. (2004). The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc. Natl. Acad. Sci. USA *101*, 15386–15391.
2. Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G., and Bhattacharya, D. (2004). A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. *21*, 809–818.
3. Mereschkowsky, C. (1905). Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. Biologisches Centralblatt *25*, 593–604.
4. McFadden, G.I. (2001). Primary and secondary endosymbiosis and the origin of plastids. J. Phycol. *37*, 951–959.
5. Palmer, J.D. (2003). The symbiotic birth and spread of plastids: how many times and whodunit? J. Phycol. *39*, 4–11.
6. Cavalier-Smith, T. (1981). Eukaryote kingdoms: seven or nine? Biosystems *14*, 461–481.
7. Stiller, J.W., Riley, J., and Hall, B.D. (2001). Are red algae plants? A critical evaluation of three key molecular data sets. J. Mol. Evol. *52*, 527–539.
8. Baldauf, S.L., Roger, A.J., Wenk-Siefert, I., and Doolittle, W.F. (2000). A kingdom-level phylogeny of eukaryotes based on combined protein data. Science *290*, 972–977.
9. Turner, S., Pryer, K.M., Miao, V.P., and Palmer, J.D. (1999). Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J. Eukaryot. Microbiol. *46*, 327–338.
10. Douglas, S.E., and Turner, S. (1991). Molecular evidence for the origin of plastids from a cyanobacterium-like ancestor. J. Mol. Evol. *33*, 267–273.
11. Bhattacharya, D., Helmchen, T., Bibeau, C., and Melkonian, M. (1995). Comparisons of nuclear-encoded small-subunit ribosomal RNAs reveal the evolutionary position of the Glaucocystophyta. Mol. Biol. Evol. *12*, 415–420.
12. Philippe, H., Snell, E.A., Bapteste, E., Lopez, P., Holland, P.W., and Casane, D. (2004). Phylogenomics of eukaryotes: impact

of missing data on large alignments. Mol. Biol. Evol. *21*, 1740–1752.
13. Simpson, A.G., and Roger, A.J. (2004). The real 'kingdoms' of eukaryotes. Curr. Biol. *14*, R693–R696.
14. Stiller, J.W., Reel, D.C., and Johnson, J.C. (2003). A single origin of plastids revisited: convergent evolution in organellar genome content. J. Phycol. *39*, 95–105.
15. Cai, X., Fuller, A.L., McDougald, L.R., and Zhu, G. (2003). Apicoplast genome of the coccidian *Eimeria tenella*. Gene *321*, 39–46.
16. Lockhart, P.J., Howe, C.J., Bryant, D.A., Beanland, T.J., and Larkum, A.W. (1992). Substitutional bias confounds inference of cyanelle origins from sequence data. J. Mol. Evol. *34*, 153–162.
17. Lockhart, P.J., Steel, M.A., Barbrook, A.C., Huson, D.H., Charleston, M.A., and Howe, C.J. (1998). A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. Mol. Biol. Evol. *15*, 1183–1188.
18. Lockhart, P.J., Steel, M.A., Hendy, M.D., and Penny, D. (1994). Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. *11*, 605–612.
19. Huelsenbeck, J.P., and Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics *17*, 754–755.
20. Fitch, W.M. (1976). The molecular evolution of cytochrome c in eukaryotes. J. Mol. Evol. *8*, 13–40.
21. Zhaxybayeva, O., Lapierre, P., and Gogarten, J.P. (2004). Genome mosaicism and organismal lineages. Trends Genet. *20*, 254–260.
22. Moreira, D., Le Guyader, H., and Philippe, H. (2000). The origin of red algae and the evolution of chloroplasts. Nature *405*, 69–72.
23. Burger, G., Saint-Louis, D., Gray, M.W., and Lang, B.F. (1999). Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*. Cyanobacterial introns and shared ancestry of red and green algae. Plant Cell *11*, 1675–1694.
24. Nozaki, H., Matsuzaki, M., Takahara, M., Misumi, O., Kuroiwa, H., Hasegawa, M., Shin-i, T., Kohara, Y., Ogasawara, N., and Kuroiwa, T. (2003). The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. J. Mol. Evol. *56*, 485–497.
25. Sanderson, J.S., and Shaffer, H.B. (2002). Troubleshooting molecular phylogenetic analyses. Annu. Rev. Ecol. Syst. *33*, 49–72.
26. Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. *6*, 361–375.
27. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.
28. Whelan, S., Lio, P., and Goldman, N. (2001). Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet. *17*, 262–272.
29. Bapteste, E., Brinkmann, H., Lee, J.A., Moore, D.V., Sensen, C.W., Gordon, P., Durufle, L., Gaasterland, T., Lopez, P., Muller, M., et al. (2002). The analysis of 100 genes supports the grouping of three highly divergent amoebae: *Dictyostelium, Entamoeba*, and *Mastigamoeba*. Proc. Natl. Acad. Sci. USA *99*, 1414–1419.
30. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.
31. Philippe, H., Lartillot, N., and Brinkmann, H. (2005). Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa and Protostomia. Mol. Biol. Evol. *22*, 1246–1253.
32. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673–4680.
33. Philippe, H. (1993). MUST, a computer package of Management Utilities for Sequences and Trees. Nucleic Acids Res. *21*, 5264–5272.
34. Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. *52*, 696–704.
35. Swofford, D.L. (2002). PAUP: Phylogenetic Analysis using Parsimony (and Other Methods). Version 4 (Sunderland, Massachusetts: Sinauer Associates).
36. Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502–504.
37. Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Mol. Biol. Evol. *14*, 685–695.
38. Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. *13*, 555–556.
39. Kishino, H., Miyata, T., and Hasegawa, M. (1990). Maximum likelihood inference of protein phylogeny, and the origin of chloroplasts. J. Mol. Evol. *31*, 151–160.
40. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.
41. Lecointre, G., Philippe, H., Van Le, H.L., and Le Guyader, H. (1994). How many nucleotides are required to resolve a phylogenetic problem? The use of a new statistical method applicable to available sequences. Mol. Phylogenet. Evol. *3*, 292–309.
42. Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Muller, M., and Le Guyader, H. (2000). Early-branching or fast-evolving eukaryotes? An answer based on slowly evolving positions. Proc. R. Soc. Lond. B. Biol. Sci. *267*, 1213–1221.
43. Stechmann, A., and Cavalier-Smith, T. (2002). Rooting the eukaryote tree by using a derived gene fusion. Science *297*, 89–91.