Current Biology 20, 2217–2222, December 21, 2010 ©2010 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2010.11.035

Report

Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants

Cédric Finet,[1,*] Ruth E. Timme,[2,4] Charles F. Delwiche,[2,5] and Ferdinand Marlétaz[3,6]
[1]Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, WI 53706, USA
[2]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA
[3]Station Marine d'Endoume, UMR 6540, DIMAR, CNRS, Université de la Méditerranée, Chemin de la Batterie des Lions, 13007 Marseille, France

Summary

The Viridiplantae (green plants) include land plants as well as the two distinct lineages of green algae, chlorophytes and charophytes. Despite their critical importance for identifying the closest living relatives of land plants, phylogenetic studies of charophytes have provided equivocal results [1–5]. In addition, many relationships remain unresolved among the land plants, such as the position of mosses, liverworts, and the enigmatic Gnetales. Phylogenomics has proven to be an insightful approach for resolving challenging phylogenetic issues, particularly concerning deep nodes [6–8]. Here we extend this approach to the green lineage by assembling a multilocus data set of 77 nuclear genes (12,149 unambiguously aligned amino acid positions) from 77 taxa of plants. We therefore provide the first multigene phylogenetic evidence that Coleochaetales represent the closest living relatives of land plants. Moreover, our data reinforce the early divergence of liverworts and the close relationship between Gnetales and Pinaceae. These results provide a new phylogenetic framework and represent a key step in the evolutionary interpretation of developmental and genomic characters in green plants.

Results and Discussion

We address here for the first time the question of the origin and early relationships of land plants using the full set of nuclear ribosomal proteins, which have shown to be valuable phylogenetic markers [7, 9]. We also carefully inspect a set of discrete genomic and morphological characters, which previously brought insightful evidence to deep plant phylogeny [10–13]. We took advantage of the increasing number of available expressed sequence tags (ESTs) that allowed us to sample diverse genes from a large number of taxa [14]. In addition, we generated new transcriptomic data by applying pyrosequencing to five selected species of charophyte algae (see Experimental Procedures). These freshwater algae are pivotal in understanding the origin of land plants, because they are thought to be the closest relatives of embryophytes based on morphological similarities [15], molecular phylogeny [13, 16], and rare genomic characters [10, 12]. However, the question as to which charophyte lineage represents the sister to land plants is far from settled [17–19].

We assembled a data set of 77 nuclear genes from 77 taxa, of which 20 are built upon a composite approach designed to take advantage of available taxonomic diversity (see Experimental Procedures and Tables S1 and S2 available online). Phylogenetic analyses of this multigene data set support Coleochaetales as the closest extant relatives of land plants (Figure 1; Bayesian posterior probabilities [PP] = 1, bootstrap replicates [BP] = 91), and not Charales (e.g., *Nitella*) as previously reported [13]. Notably, we observed extensive differences between rates of molecular evolution inferred in land plants and their algal relatives (Figures S1 and S2). This prompted us to investigate the possible impact of long-branch attraction on our reconstructions. First, we used a site-heterogeneous model of evolution (CAT), which has been reported to handle nonphylogenetic signal more efficiently [20, 21], and compared results with those obtained by using a more classical site-homogeneous model (WAG). Statistical comparisons using a cross-validation approach show that the CAT model fits our data better (likelihood score difference = 1697.81 ± 96.39; see Experimental Procedures). We then analyzed four distinct data sets in which deepest roots and fastest evolving taxa were progressively removed (77-, 66-, 64-, and 61-taxon data sets; Figures S1 and S2). *Coleochaete* branching as the sister group to land plants is remarkably robust in this analysis with regard to inference models and taxonomic sampling.

Coleochaetales share numerous characteristics with land plants. These include morphological traits such as complex three-dimensional organization, with parenchyma-like tissues and placental transfer cell wall ingrowths in some species, and zygote retention [15]. Ultrastructural studies have revealed that cytokinesis in *Coleochaete* cells use a phragmoplast very similar to that of embryophytes [22]. Additionally, *Coleochaete* cells have a land-plant-like peroxisome [22]. Several of these characters are found in the Charales and Zygnematales as well, but the morphological features of the Coleochaetales are certainly compatible with the phylogeny found here. The placement of Coleochaetales as the sister group to land plants also seems possible based on the fossil record. For example, the late Silurian–early Devonian fossil *Parka* has been compared to extant *Coleochaete* on the basis of both structure and ecology [23], suggesting that Coleochaetales may be an ancient group, although *Parka* was also substantially larger than any known member of the Coleochaetales. In addition, our phylogeny is consistent with the distribution of a set of molecular signatures and morphological characters under the parsimony criterion (Figure 2). For example, the intron in the *nad5* mitochondrial gene was reported at the same position in *Coleochaete orbicularis*, *Sphagnum* (moss), and *Marchantia* (liverwort) but was found missing in other charophyte and embryophyte taxa [13],

*Correspondence: cedric.finet@ens-lyon.org
[4]Present address: National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892-6510, USA
[5]Present address: Maryland Agricultural Experiment Station, University of Maryland, College Park, MD 20742, USA
[6]Present address: Developmental Biology Unit, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany

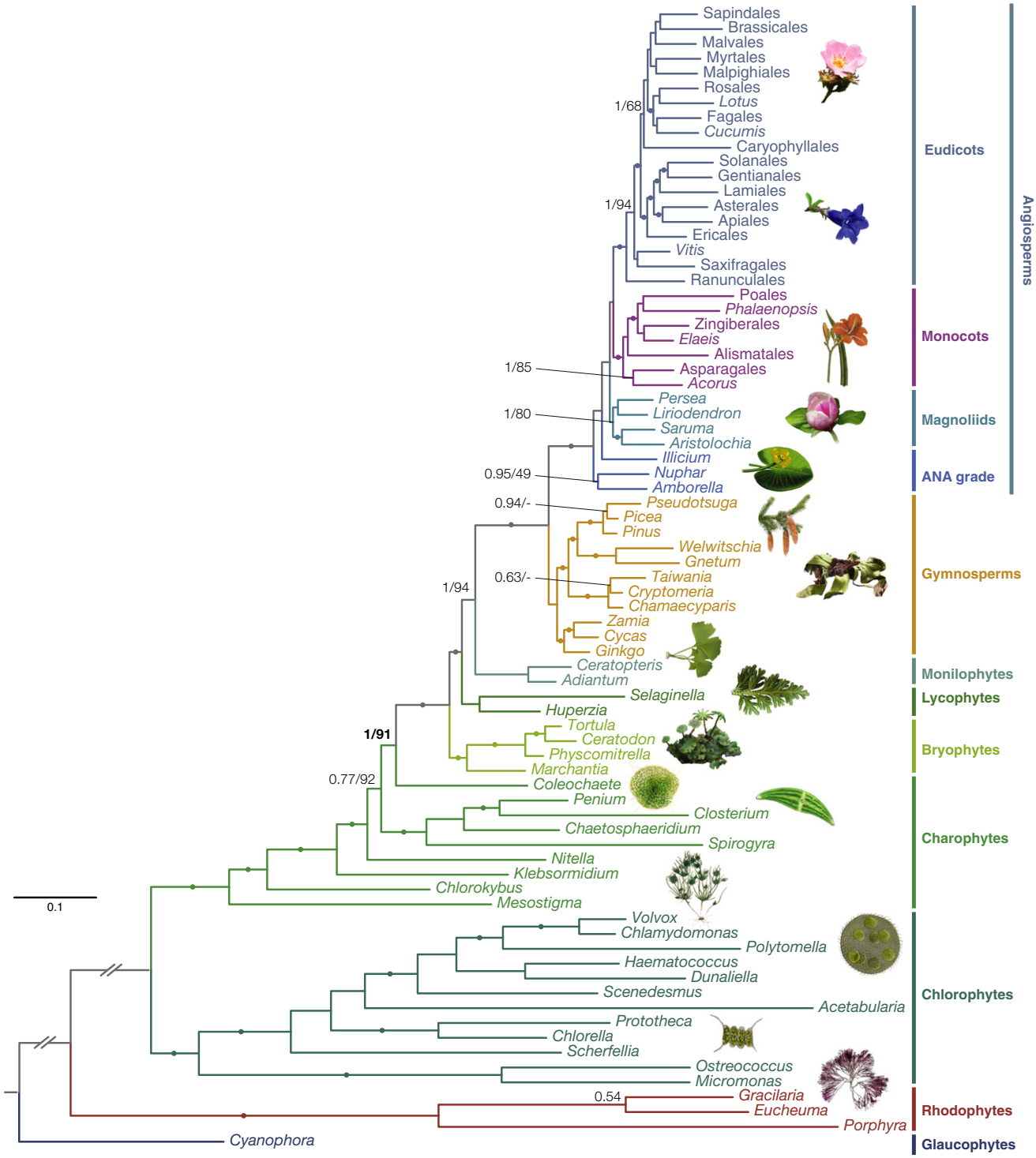

Figure 1. Phylogram of the 77-Taxon Analyses

RAxML maximum-likelihood analyses and PhyloBayes Bayesian analyses were conducted under the PROTMIXWAG model and the CAT model, respectively. The overall matrix represents 12,149 amino acids and exhibits 19% missing data (Table S1). Support values obtained after 100 bootstrap replicates (BP) and Bayesian posterior probabilities (PP) are shown for selected branches (all of the support values are shown in Figures S1 and S2). A dot indicates support values of PP = 1 and BP > 95. Scale bar indicates number of changes per site. According to the most recent phylogenies of eukaryotes, the branch leading to the glaucophytes was used to root this tree.

which may be compatible with a gain in a putative common ancestor of land plants and Coleochaetales, followed by secondary losses (Figure 2). It is noteworthy that this molecular signature further supports the placement of hornworts as the sister to vascular plants. However, the absence of the *nad5* intron in other examined species of *Coleochaete* suggests a much more complex evolution of this character [13].

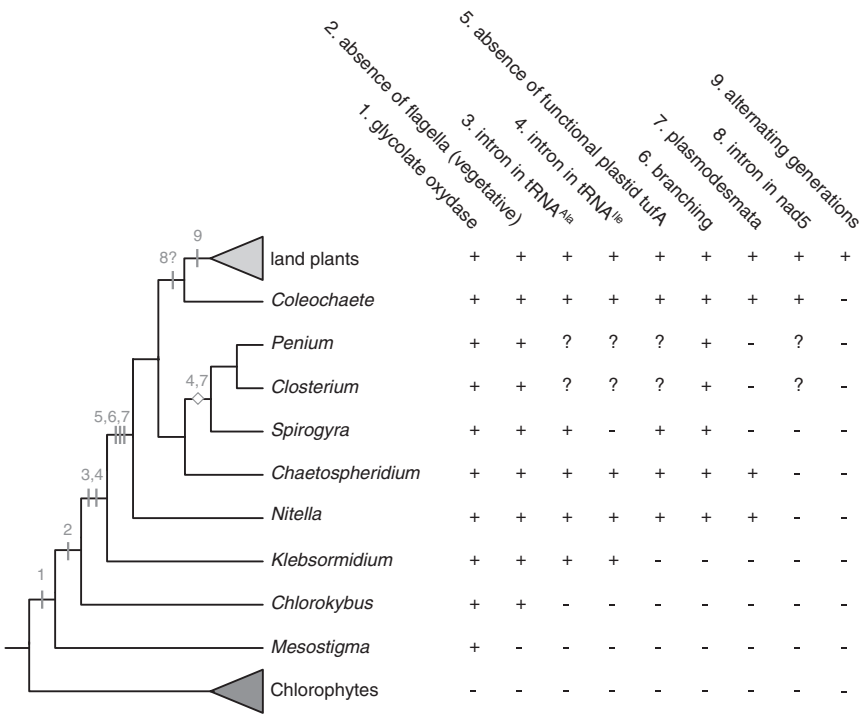

Figure 2. Main Morphological and Molecular Characters of Charophytes Discussed in the New Phylogenetic Framework

Key to character changes: 1, acquisition of the glycolate oxidase, a key enzyme involved in the degradation of glycolate in peroxisomes; 2, loss of the flagella in vegetative phase (reproductive cells remain motile); 3–5, structural markers in chloroplast genomes (see [19] for a more complete study of genomic characters); 6, ability to develop complex, filamentous thalli with branching; 7, acquisition of intercellular communication via plasmodesmata, which are extensions of the plasma membrane connecting the cytoplasm of each cell with that of its neighbors; 8, presence of an intron in the mitochondrial gene *nad5*; 9, acquisition of a life cycle with alternation of diploid and haploid multicellular phases. Characters were mapped based upon the parsimony criterion. In particular, the mapping of the acquisition of the *nad5* intron is one of several possible scenarios. | indicates acquisition; ◇ indicates loss.

Our phylogeny provides insights into the internal structure of the charophytes (Figure 1). None of our results indicate that the charophyte algae represent a monophyletic lineage. The Chlorokybales and the Mesostigmatales are found to be the deepest branches of charophytes. These two orders do not make up a clade in most of our analyses, challenging a recent chloroplast genome-based phylogeny [24]. The placement of Mesostigmatales as an outgroup to the remaining charophytes is supported by the presence of vegetative flagella in Mesostigmatales and chlorophytes, but not in other charophytes (Figure 2). However, the grouping of Chlorokybales and Mesostigmatales has been recovered in some of our trees, suggesting that further analyses will be necessary to settle the relative position of these two groups. The Klebsormidiales are subsequently found to diverge next with strong branch support (PP = 1, BP = 100) (Figures S1 and S2). Hence, our topology identifies the Mesostigmatales, Chlorokybales, and Klebsormidiales as the earliest-diverging charophyte lineages. This finding is congruent with the nuclear (versus chloroplastic) localization of the *tufA* gene in late-diverging charophyte lineages and land plants [2, 11, 25, 26] (Figure 2). More unexpected is the grouping of Zygnematales with *Chaetosphaeridium*, which was formerly allied with Coleochaetales. If corroborated by further analyses, this would have major implications for charophyte systematics. More extensive sampling of Coleochaetales and Charales would likely settle this issue in the future.

Among the flowering plants, relationships of relatively few groups are still in dispute. Although the ANA grade (*A*morborellaceae, *N*ymphaeales, and *A*ustrobaileyales) has been identified as the earliest-diverging branches, the relative positions of Amborellaceae and Nymphaeales remain controversial [27]. Our phylogeny identifies a clade including *Amborella* and Nymphaeales as the sister to all other angiosperms. Such a relationship needs to be corroborated by further analyses, especially by improving the sampling within the ANA grade. Another noteworthy point is the placement of Acorales as the sister group to Asparagales (Figure 1; PP = 1, BP = 85). This result, which has not been previously postulated, challenges the view that Acorales represent the deepest branch of the monocots. In addition, we observe a distinct branching for magnoliids in site-homogeneous and site-heterogeneous analyses, with a poorly supported eudicot sister-group relationship in the first case (BP = 36) and an early angiosperm position in the second (PP = 0.99; Figure 1; Figures S1 and S2). This could indicate residual long-branch effects in this part of the tree and suggests that further attention should be paid to relative relationships of these taxa.

Among gymnosperms, we consistently place Gnetales within conifers as sister group to Pinaceae (Figure 1; PP = 1, BP = 100). Gnetales represent a puzzling order in that they have vessel elements and chlamys-surrounded ovules and perform double fertilization [28] as do flowering plants. These similarities led some authors to formulate the so-called "anthophyte" hypothesis that proposed Gnetales as the sister group of flowering plants [29, 30]. Conversely, our topology rejects the monophyly of conifers and is consistent with the "gnepine" hypothesis [31–34]. This evolutionary scenario implies that Gnetales lost several synapomorphies of conifers (e.g., resin canal, tiered proembryos, ovulate cone scale) [35]. Interestingly, this topology is in agreement with a recent structural genomic study showing that all plastid *ndh* genes are absent across Gnetales and Pinaceae, but not in any other group of gymnosperms [36]. In addition, phylogenetic analyses of the different data sets all provide strong evidence that cycads are allied with *Ginkgo* (Figure 1; Figures S1 and S2).

The approach used here represents a technical advance in plant phylogeny by identifying nuclear ribosomal proteins as valuable phylogenetic markers. These markers are easily retrievable from EST or transcriptomic data because they are abundantly and ubiquitously expressed [37]. Additionally, the full set of nuclear ribosomal protein genes only rarely includes

duplicated copies. This minimizes orthology assessment issues. Among the 136 taxa used in the present study, only the genus *Ostreococcus* challenges this statement with an unusual reduced number of 23 nuclear ribosomal protein genes. The genome of the green alga *Ostreococcus tauri* exhibits highly derived features, such as an extreme degree of compaction and gene loss [38], so this discrepancy is not altogether surprising, but study of the complete genomes of additional early-diverging land plants and green algae will be needed to validate inferences of orthology for data derived from EST surveys. Application of this multigene approach to the green lineage provided several interesting results, such as the branching of *Coleochaete* as the sister group of land plants. However, phylogenomics has proven highly sensitive to systematic error, such as long-branch attraction, which can be resolved by methodological improvements and increased taxonomic sampling [39]. Concerning the grouping of land plants and *Coleochaete*, a definitive conclusion will thus only be achieved through both the collection of more extensive genomic data in charophytes and bryophytes sensu lato and the evaluation of possible systematic error still affecting current models of molecular evolution.

The newly proposed phylogeny strengthens the view that molecular phylogeny can reveal patterns of morphological diversity across very deep lineages in the Tree of Life. For instance, we point out that, during the evolution of charophyte algae, no clear trend is displayed toward increasing complexity. Indeed, land plants share a common ancestor with Coleochaetales, a lineage that shows a great diversity in tissue organization, ranging from filamentous (*C. pulvinata*) to parenchymatous (*C. orbicularis*). Conversely, charophytes found in this analysis to be more distantly related to land plants, like the genus *Nitella*, exhibit whorled branches, suggesting that this grade of organismal complexity may have evolved independently [40]. This could prompt a reappraisal of the fossil record, shedding light on ecological and morphological characters that allowed the very significant colonization of land.

In summary, *Coleochaete* represents a pivotal model for studying the origin of land plants. Developmental genetics approaches and genome sequencing could untangle evolutionary origins of prominent embryophyte features. First, this could help to determine how a shift in tissue organization took place, from filamentous to parenchymatous land plants [41, 42]. Second, enclosed zygotes of *Coleochaete* are reminiscent of land plant embryos whose evolutionary appearance is still mysterious. It would be interesting to determine whether key genes in embryo development are present in *Coleochaete*, revealing a possible ancestral toolkit in the common ancestor of Coleochaetales and land plants.

Experimental Procedures

Generation of Charophyte Transcriptomic Data

A mixture of classical Sanger sequencing of cDNA (EST sequence) and next-generation 454 pyrosequencing yielded high-grade assembly of unigene transcripts for a representative collection of five charophyte species: *Chaetosphaeridium globosum*, *Chlorokybus atmophyticus*, *Klebsormidium flaccidum*, *Nitella hyalina*, and *Penium margaritaceum* (Table S1). EST libraries were built following the same experimental procedure as described previously [43] reporting the collection of *Coleochaete orbicularis* and *Spirogyra pratensis*, also included in the present study.

Multigene Data Set and Composite Taxa

Ribosomal proteins are valuable phylogenetic markers because of conservation among eukaryotes [9], membership of nonmultigenic superfamilies (no orthology assignment issues), and relative abundance in EST databases. Moreover, ribosomal proteins have been shown to be reliable for reconstructing deep phylogenies [7]. In angiosperms, we chose to apply a composite taxon strategy because it has been proven to limit branch lengths and to handle a limited number of taxa that allows the employment of parameter-rich models of evolution. This strategy starts from the definition of a set of unambiguously monophyletic species and goes through the collection of least diverging sequence or species for each marker, so as to ultimately yield a unique composite taxon by concatenation. For example, this has led to strong shortening of the nematode branch by surveying the multiple EST collections available for this clade [7]. We surveyed available EST collections for taxa of interest at NCBI dbEST (Table S1). These data were downloaded, stored locally, and processed using a taxon-building pipeline operated by using Perl scripts. For each taxon, a first BLAST [44] search collected all sequences similar to a canonical set of ribosomal proteins (cutoff score 50). The ribosomal transcripts were assembled using Phrap [45]. A second BLAST search was carried out on assembled transcripts of each taxon and was allowed to select the least dissimilar sequence for each marker and to detect the coding frame for translation. Alignments for individual marker genes were generated using MUSCLE [46] and manually checked using MacVector (MacVector, Inc.). Before marker concatenation, removal of ambiguously aligned regions was performed with Gblocks on individual genes, using the least stringent parameters [47]. The overall 77-taxon matrix represents 12,149 amino acids and exhibits 19% missing data (Table S1), which is far less than amounts observed in recent phylogenomic studies [8, 48] and is well below the "reasonable" amount of missing data allowed for accurate phylogenetic reconstruction [9].

Data Set Validation

The integrity of the data set and especially the possible contamination status were verified by inferring independent trees for independent marker genes using PhyML and the WAG+Γ_4 model. We carefully checked each of the trees for cases exhibiting a well-supported branch (bootstrap percentage > 70%) incongruent with our concatenated analysis. Molecular evolution parameters collected through these analyses are summarized in Table S2.

Because many proximal nodes in plant phylogeny remain ambiguous, especially within relatively recently diverging angiosperms, we attempted to verify that our composite strategy did not utilize taxa with questionable monophyly. We then set up an alternative alignment that included all species involved in composite taxa, in order to check their monophyly (Table S1). This was carried out using the same procedure, except that each original EST collection represented only one species. The resulting tree from 136 taxa, obtained using RAxML (see below), confirmed the monophyly of all composite taxa.

Phylogenetic Reconstruction

The use of a concatenated approach for phylogenetic reconstruction at high taxonomic level has already been demonstrated as a useful tool for resolving longstanding phylogenetic issues [7, 8, 49]. However, such data sets constitute large amounts of information in term of positions and taxa that may be a challenge for tree inference. In particular, it has been shown that phylogenomic data sets, when analyzed using improper methods and models of evolution, can lead to deeply misleading topologies [39, 50, 51]. Recently, development of Bayesian inference has made it possible to handle and estimate the multiple parameters of realistic models of evolution to limit reconstruction artifacts. Particularly, the CAT model introduces multiple profiles of amino acid substitutions that are distinguished by their equilibrium frequencies and are thus able to cope with the heterogeneity of protein data encountered in large phylogenomic alignments [20].

We first used the PhyloBayes 3.2 program, which implements a site-heterogeneous CAT model [52]. For the different data sets, we ran two chains for at least 20,000 cycles (~2,200,000 generations) and removed the first 5,000 cycles as burn-in. Maximal posterior differences recovered were 0.12 for the 77-taxon data set, 0.20 for the 66-taxon data set, 0.27 for the 64-taxon data set, and 0.17 for the 61-taxon data set (Figure S1), values that are all consistent with an accurate estimation of posterior consensus. Run parameters of the chains were also plotted together against time to check appropriate convergence and chain mixing. In addition, we employed RAxML 7.0.4, which allows efficient maximum-likelihood analyses of large data sets [53]. All searches were completed with the PROTMIXWAG setting (an efficient approximation of the WAG+Γ model), and 100 bootstrap replicates were conducted for support estimation.

We statistically compared CAT and WAG models using cross-validation tests [54]. Ten replicates (9 of 10 for the learning set and 1 of 10 for the testing set) were run assuming each model under evaluation for 1,500 cycles (500 being discarded as burn-in). We determined that the CAT model had a much better statistical fit than WAG likelihood difference (CAT versus WAG: 1697.81 ± 96.3913).

The different data sets and obtained topologies are available on TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S10983).

Supplemental Information

Supplemental Information includes two figures and two tables and can be found with this article online at doi:10.1016/j.cub.2010.11.035.

Received: August 9, 2010
Revised: September 25, 2010
Accepted: November 11, 2010
Published online: December 9, 2010

References

1. Marin, B., and Melkonian, M. (1999). Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. Protist *150*, 399–417.
2. Turmel, M., Otis, C., and Lemieux, C. (2005). The complete chloroplast DNA sequences of the charophycean green algae Staurastrum and Zygnema reveal that the chloroplast genome underwent extensive changes during the evolution of the Zygnematales. BMC Biol. *3*, 22.
3. Turmel, M., Otis, C., and Lemieux, C. (2006). The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. Mol. Biol. Evol. *23*, 1324–1338.
4. Rodríguez-Ezpeleta, N., Philippe, H., Brinkmann, H., Becker, B., and Melkonian, M. (2007). Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of Mesostigma in the Streptophyta. Mol. Biol. Evol. *24*, 723–731.
5. Hall, J.D., Karol, K.G., McCourt, R.M., and Delwiche, C.F. (2008). Phylogeny of the conjugating green algae based on chloroplast and mitochondrial nucleotide sequence data. J. Phycol. *44*, 467–477.
6. Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965–968.
7. Marlétaz, F., Martin, E., Perez, Y., Papillon, D., Caubit, X., Lowe, C.J., Freeman, B., Fasano, L., Dossat, C., Wincker, P., et al. (2006). Chaetognath phylogenomics: A protostome with deuterostome-like development. Curr. Biol. *16*, R577–R578.
8. Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D., et al. (2008). Broad phylogenomic sampling improves resolution of the animal tree of life. Nature *452*, 745–749.
9. 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.
10. Frederick, S.E., Gruber, P.J., and Tolbert, N.E. (1973). The occurrence of glycolate dehydrogenase and glycolate oxidase in green plants: an evolutionary survey. Plant Physiol. *52*, 318–323.
11. Baldauf, S.L., and Palmer, J.D. (1990). Evolutionary transfer of the chloroplast tufA gene to the nucleus. Nature *344*, 262–265.
12. Manhart, J.R., and Palmer, J.D. (1990). The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. Nature *345*, 268–270.
13. Karol, K.G., McCourt, R.M., Cimino, M.T., and Delwiche, C.F. (2001). The closest living relatives of land plants. Science *294*, 2351–2353.
14. Philippe, H., and Telford, M.J. (2006). Large-scale sequencing and the new animal phylogeny. Trends Ecol. Evol. (Amst.) *21*, 614–620.
15. Graham, L.E. (1993). Origin of Land Plants (New York: John Wiley & Sons).
16. Qiu, Y.L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA *103*, 15511–15516.
17. McCourt, R.M., Delwiche, C.F., and Karol, K.G. (2004). Charophyte algae and land plant origins. Trends Ecol. Evol. (Amst.) *19*, 661–666.
18. Becker, B., and Marin, B. (2009). Streptophyte algae and the origin of embryophytes. Ann. Bot. (Lond.) *103*, 999–1004.
19. Turmel, M., Pombert, J.F., Charlebois, P., Otis, C., and Lemieux, C. (2007). The green algal ancestry of land plants as revealed by the chloroplast genome. Int. J. Plant Sci. *168*, 679–689.
20. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–1109.
21. Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. *7* (*Suppl 1*), S4.
22. Marchant, H.J., and Pickett-Heaps, J.D. (1973). Mitosis and cytokinesis in Coleochaete scutata. J. Phycol. *9*, 461–471.
23. Niklas, K.J. (1976). Morphological and ontogenetic reconstruction of Parka decipiens Fleming and Pachytheca Hooker from the Lower Old Red Sandstone, Scotland. Trans. R. Soc. Edinb. *69*, 483–499.
24. Lemieux, C., Otis, C., and Turmel, M. (2007). A clade uniting the green algae Mesostigma viride and Chlorokybus atmophyticus represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. BMC Biol. *5*, 2.
25. Baldauf, S.L., Manhart, J.R., and Palmer, J.D. (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.
26. Turmel, M., Otis, C., and Lemieux, C. (2002). The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeridium globosum: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc. Natl. Acad. Sci. USA *99*, 11275–11280.
27. Graham, S.W., and Iles, W.J.D. (2009). Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. Am. J. Bot. *96*, 216–227.
28. Friedman, W.E., and Carmichael, J.S. (1996). Double fertilization in Gnetales: Implications for understanding reproductive diversification among seed plants. Int. J. Plant Sci. *157*, S77–S94.
29. Crane, P.R. (1985). Phylogenetic analysis of seed plants and the origin of angiosperms. Ann. Mo. Bot. Gard. *72*, 716–793.
30. Doyle, J.A., and Donoghue, M.J. (1986). Seed plant phylogeny and the origin of angiosperms: An experimental cladistic approach. Bot. Rev. *52*, 321–431.
31. Bowe, L.M., Coat, G., and dePamphilis, C.W. (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.
32. Chaw, S.M., Parkinson, C.L., Cheng, Y., Vincent, T.M., and Palmer, J.D. (2000). Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. Proc. Natl. Acad. Sci. USA *97*, 4086–4091.
33. Gugerli, F., Sperisen, C., Büchler, U., Brunner, I., Brodbeck, S., Palmer, J.D., and Qiu, Y.L. (2001). The evolutionary split of Pinaceae from other conifers: Evidence from an intron loss and a multigene phylogeny. Mol. Phylogenet. Evol. *21*, 167–175.
34. Qiu, Y.L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R.W., Hendry, T.A., et al. (2007). A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. Int. J. Plant Sci. *168*, 691–708.
35. Mathews, S. (2009). Phylogenetic relationships among seed plants: Persistent questions and the limits of molecular data. Am. J. Bot. *96*, 228–236.
36. Braukmann, T.W., Kuzmina, M., and Stefanović, S. (2009). Loss of all plastid ndh genes in Gnetales and conifers: Extent and evolutionary significance for the seed plant phylogeny. Curr. Genet. *55*, 323–337.
37. Marlétaz, F., Gilles, A., Caubit, X., Perez, Y., Dossat, C., Samain, S., Gyapay, G., Wincker, P., and Le Parco, Y. (2008). Chaetognath transcriptome reveals ancestral and unique features among bilaterians. Genome Biol. *9*, R94.
38. Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., Partensky, F., Degroeve, S., Echeynié, S., Cooke, R., et al. (2006). Genome analysis of the smallest free-living eukaryote Ostreococcus tauri unveils many unique features. Proc. Natl. Acad. Sci. USA *103*, 11647–11652.
39. Jeffroy, O., Brinkmann, H., Delsuc, F., and Philippe, H. (2006). Phylogenomics: The beginning of incongruence? Trends Genet. *22*, 225–231.
40. Cook, C.D.K. (1978). The *Hippuris* syndrome. In Essays in Plant Taxonomy, H.E. Street, ed. (London: Academic Press), pp. 164–176.

41. Graham, L.E. (1984). Coleochaete and the origin of land plants. Am. J. Bot. *71*, 603–608.
42. Dupuy, L., Mackenzie, J., and Haseloff, J. (2010). Coordination of plant cell division and expansion in a simple morphogenetic system. Proc. Natl. Acad. Sci. USA *107*, 2711–2716.
43. Timme, R.E., and Delwiche, C.F. (2010). Uncovering the evolutionary origin of plant molecular processes: Comparison of *Coleochaete* (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes. BMC Plant Biol. *10*, 96.
44. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.
45. Green, P. (1996). Phrap documentation (http://www.phrap.org/phredphrap/phrap.html).
46. Edgar, R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
47. Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. *17*, 540–552.
48. Hejnol, A., Obst, M., Stamatakis, A., Ott, M., Rouse, G.W., Edgecombe, G.D., Martinez, P., Baguñà, J., Bailly, X., Jondelius, U., et al. (2009). Assessing the root of bilaterian animals with scalable phylogenomic methods. Proc. Biol. Sci. *276*, 4261–4270.
49. Bourlat, S.J., Juliusdottir, T., Lowe, C.J., Freeman, R., Aronowicz, J., Kirschner, M., Lander, E.S., Thorndyke, M., Nakano, H., Kohn, A.B., et al. (2006). Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature *444*, 85–88.
50. Baurain, D., Brinkmann, H., and Philippe, H. (2007). Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? Mol. Biol. Evol. *24*, 6–9.
51. Holder, M., and Lewis, P.O. (2003). Phylogeny estimation: Traditional and Bayesian approaches. Nat. Rev. Genet. *4*, 275–284.
52. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.
53. Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.
54. Philippe, H., Derelle, R., Lopez, P., Pick, K., Borchiellini, C., Boury-Esnault, N., Vacelet, J., Renard, E., Houliston, E., Quéinnec, E., et al. (2009). Phylogenomics revives traditional views on deep animal relationships. Curr. Biol. *19*, 706–712.