Analysis of the *Amborella trichopoda* Chloroplast Genome Sequence Suggests That *Amborella* Is Not a Basal Angiosperm

Vadim V. Goremykin, Karen I. Hirsch-Ernst,† Stefan Wölfl,‡ and Frank H. Hellwig**

*Institut für Spezielle Botanik, Universität Jena, Jena, Germany; †Zentrum Pharmakologie und Toxikologie, Universität Göttingen, Göttingen, Germany; ‡Klinik für Innere Medizin, Universität Jena, Jena, Germany

Phylogenetic analyses based on comparison of a limited number of genes recently suggested that *Amborella trichopoda* is the most ancient angiosperm. Here we present the complete sequence of the chloroplast genome of this plant. It does not display any of the genes characteristic of chloroplast DNA of the gymnosperm *Pinus thunbergii* (*chl*B, *chl*L, *chl*N, *psa*M, and *ycf*12). The majority of phylogenetic analyses of protein-coding genes of this chloroplast DNA suggests that *Amborella* is not the basal angiosperm and not even the most basal among dicots.

Introduction

The question of the origin and early diversification of angiosperms was one of the most intriguing issues for the botanists in the past century. The lack of reliable paleobotanic evidence left a wide space for speculations on angiosperm origins. Virtually all groups of land plants have been suggested to fulfill the role of the ancestors of angiosperms (Meyen 1986). From these earlier theories only two have remained that are still discussed in the literature—the euanthial theory by Arber and Parkin (1907, 1908) and the pseudanthial theory put forward by von Wettstein (1924). The first theory implies that all the flowers were derived from the large compound bisexual strobiloid flowers, like those of *Magnolia*, while the second maintains that the flowers evolved from the unisexual strobili, and, therefore, the simple small unisexual flowers, similar to those of Hamamelids, should be considered as the starting point for floral evolution. At the end of the twentieth century this controversy still remained unresolved. Molecular studies suggested that the modern flowering plants evolved from herbaceous forms, so called ''paleoherbs.'' Within paleoherbs, the picture of angiosperm diversification remained unclear, with the root of angiosperms assigned either to *Ceratophyllum*, with small unisexual flowers, or to Nymphaeales, which have large complex bisexual flowers.

Most recently, some papers employing large multigene data sets (Parkinson, Adams, and Palmer 1999; Qiu et al. 1999; Soltis, Soltis, and Chase 1999) suggested that the tropical monotypic family Amborellaceae, with only one species—*Amborella trichopoda* Baill., might belong to the most archaic lineage of the angiosperms. Some later molecular investigations (Barkman et al. 2000; Graham and Olmstead 2000), however, yielded unstable topologies either with *Amborella* as a sister group to the rest of the angiosperms or with *Amborella* as a first group to split off the most basal angiosperm clade, including Nymphaeales.

Here, we present the complete sequence of the *Amborella trichopoda* chloroplast genome, sequenced with overall 10× coverage to achieve a total quality of one mistake among 23,341 bases (6.97 possible mistakes in the whole plastome sequence), as determined using PHRED (Ewing et al. 1998) confidence values for the genomic consensus sequence.

Key words: *Amborella*, chloroplast genomes, angiosperms, gymnosperms, molecular evolution, substitution rates.

E-mail: vadim.goremykin@uni-jena.de.

Mol. Biol. Evol. 20(9):1499–1505. 2003
DOI: 10.1093/molbev/msg159
Molecular Biology and Evolution, Vol. 20, No. 9,

Methods

Genomic Sequencing

Leaves from *Amborella trichopoda* were received from the National Tropical Botanical Garden of Hawaii. A voucher specimen is deposited at the Herbarium Haussknecht (University of Jena, Jena, Germany). Total DNA was extracted using the CTAB-based method (Murray and Thompson 1980) and purified with Qiagen columns (Qiagen, Valencia, Calif.), according to the manufacturer's protocol. The plastome sequence was amplified using a long-range PCR technique. In brief, PCR primers were developed from the alignment of known chloroplast genomic sequences. Using these primers, we covered the entire chloroplast genome of *Amborella trichopoda* with PCR products ranging in size from 4 to 20 kb. The inverted repeat regions were amplified separately, each with two PCR products overlapping in the middle of the repeat and extending to the flanking single-copy regions. PCR products were purified by electrophoresis in low-melting agarose gel (Biozym, Hessisch Oldendorf, Germany). After treatment with agarase (MBI Fermentas, Vilnius, Lithuania), PCR products were sheared by nebulization, yielding fragments of 0.5 to 1.5 kb in length, which were subcloned into the 4Blunt-TOPO vector using the TOPO Shotgun Subcloning kit (Invitrogen, Groningen, The Netherlands), according to the manufacturer's protocol. After transformation of electrocompetent *E. coli* TOP10 cells (Invitrogen), recombinant plasmids were isolated from clones employing the Montage Plasmid Miniprep kit (Millipore, Eschborn, Germany). Automated sequencing was performed on ABI 3100, ABI 377 (ABI), and MegaBACE 1000 (Amersham/Pharmacia Biotech, Uppsala, Sweden) sequencers utilizing the Big Dye Terminator sequencing kit (ABI, Foster City, Calif.).

The *Amborella trichopoda* chloroplast genome sequence reported in this paper has been deposited in the EMBL database (accession number AJ506156). The alignments used in these analyses and primer sequences are available upon request.

Sequence Assembly

All reads were base-called with the PHRED program. Masking of the vector and primer sequences and assembly

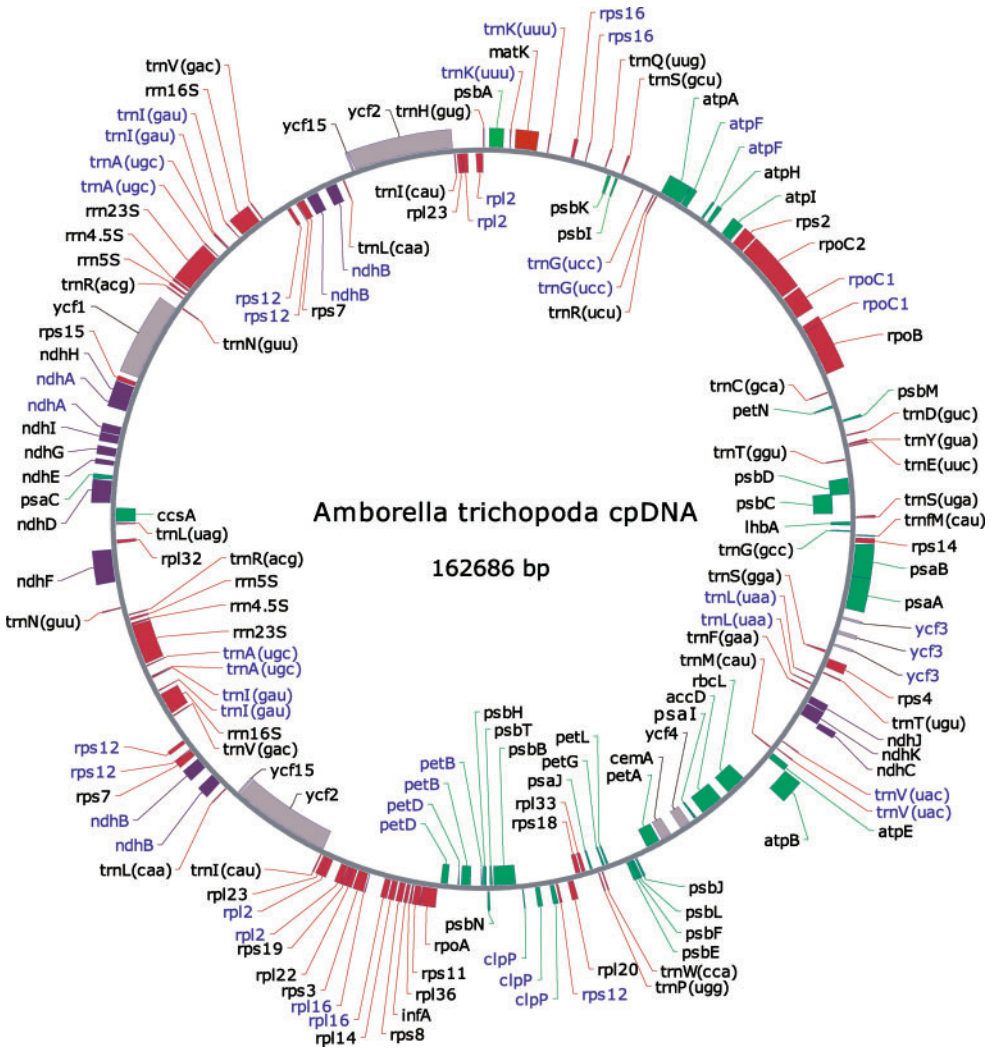

FIG. 1.—*Amborella trichopoda* cp DNA. The topmost part of the map corresponds to the start and the end of the EMBL sequence entry AJ506156. The genes shown inside the circle are transcribed clockwise, those outside the circle are transcribed counterclockwise. The genes of the genetic apparatus are shown in red, photosynthesis genes are shown in green, and genes of NADH dehydrogenase are shown in violet. Grey represents the ORFs, *ycf*s, and genes of unknown function. Intron-containing genes are represented by their exons. Their names are given in blue. In cases in which two genes overlap, one of them is shifted off the map to show its position.

were performed using the STADEN package (Staden, Beal, and Bonfield 2000) on a Linux Pentium III PC. The sequencing data were accumulated with 10× coverage for every PCR product. The remaining gaps were closed by PCR.

Results

General Plastome Properties

We found the 162,686-bases-long plastome of *Amborella trichopoda* (fig. 1) to be the largest ever registered among land plants. Yet, the increased length is not due to the presence of new genes previously undetected in other angiosperm plastomes. The cpDNA of *Amborella* does not contain *chl*B, *chl*L, *chl*N, *psa*M, or *ycf*12 genes characteristic for *Pinus thunbergii* cpDNA. Each gene encoded by the *Amborella trichopoda* chloroplast genome has already been identified in other completely sequenced chloroplast genomes of angiosperms. A total of 132 predicted genes (114 individual gene species and 18 genes duplicated in the inverted repeats), including 87 potential protein-coding genes (80 gene species), eight ribosomal RNA genes (four gene species), and 37 tRNA genes (30 gene species) representing 20 amino acids were identified on the cpDNA under study based on similarity to the entries in the nonredundant EMBL database and on similarity to the sequences in our local database of the chloroplast genes. This gene content is typical for the plastomes of the dicotyledonous angiosperms, being most similar to that of *Calycanthus fertilis* (Goremykin et al. 2003 [EMBL database accession number AJ428413]) when compared with those taxa sequenced. The only difference in gene species content between these two chloroplast genomes is that the cpDNA of *Amborella* does not contain the hypothetical *ACRS* (ACR-toxin sensitivity gene) open reading frame (ORF) in the intron within the tRNA-Ala(ugc) gene, which we found in *Calycanthus* cpDNA. The *ACRS* gene was

previously described in an intron within the mitochondrial tRNA-Ala gene of *Citrus jambhiri* (Ohtani, Yamamoto, and Akimitsu 2002). The cpDNA of *Amborella* is colinear to previously published plastomes of *Nicotiana tabacum* (Shinozaki et al. 1986), *Arabidopsis thaliana* (Sato et al. 1999), and *Spinacia oleracea* (Schmitz-Linneweber et al. 2001) and to that of *Calycanthus fertilis* in respect to gene order and overall homology. It has nine conserved ORFs (six gene species) with no known function (*ycf*s)—ycfs 1, 2, 3, 4, 15, and 72. We did not annotate the last *ycf* in the genomic sequence, since 94% of the 366-bases-long *ycf*72 coding sequence overlap with the second exon of the ribosomal protein gene *rpl*2, encoded on the complementary DNA strand. Such almost completely overlapping genes are not known to exist in chloroplasts. We believe that the lack of stop codons in *ycf*72 is fortuitous and that this *ycf* does not code for any product.

The chloroplast genes and *ycf*s encoded on the cpDNA of *Amborella* are not significantly larger than their respective homologs in other flowering plants. The lengths of the homologous gene subsets of *Amborella* and *Calycanthus* plastomes (90,352 and 90,157 bases, respectively) do not significantly differ. By contrast, the introns and spacers located on the cpDNA of *Amborella* are together 8,674 bases longer than those of *Calycanthus*. The larger size of the cpDNA molecule of *Amborella* is mostly due to unusually large intergenic spacers.

Phylogenetic Considerations and Analyses

Before phylogenetic analyses, divergence at synonymous and nonsynonymous sites was estimated in the codon-based nucleotide alignments of 61 chloroplast genes common to 13 completely sequenced chloroplast genomes of land plants (Ohyama et al. 1986; Shinozaki et al. 1986; Hiratsuka et al. 1989; Wakasugi et al. 1994; Maier et al. 1995; Sato et al. 1999; Hupfer et al. 2000; Kato et al. 2000; Schmitz-Linneweber et al. 2001; Ogihara et al. 2002) by the method of Yang (1997*a*). A comparison of synonymous and nonsynonymous distances between the genes encoded on the *Pinus* cpDNA and their angiosperm homologs (fig. 2) shows that most of the chloroplast coding sequences are very divergent at their synonymous sites for this taxonomical range. One can therefore conclude that using complete sequences of protein-coding genes from cpDNA for investigating affinity between different representatives of angiosperms and gymnosperms could lead to too high variance of phylogenetic reconstruction and to erroneous correction for multiple substitutions per site with substitution models based on total distance measurement.

Different compositional biases in different lineages could also pose problems in phylogeny reconstruction. Critical investigation (Lockhart et al. 1999) of the data set of the translated chloroplast coding sequences sampled from across the whole plant kingdom by Martin et al. (1998) demonstrated that among 48 proteins analyzed, three—*rpo*C1, *rpo*C2, and *rpo*B—showed such bias. Yet this problem does not emerge on the level of the land plants. A 5% chi-square test did not register any significant

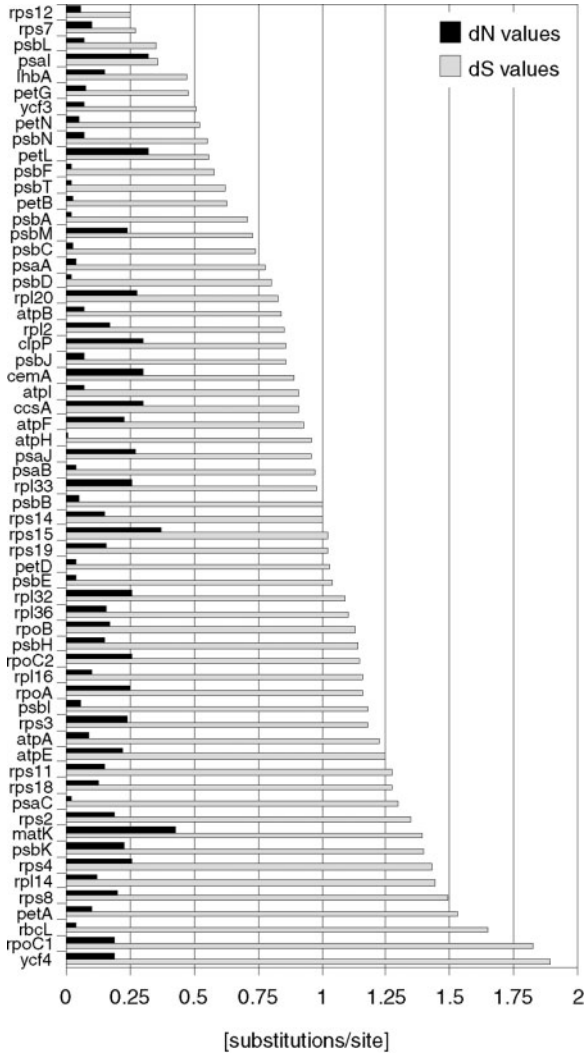

FIG. 2.—Mean rates of synonymous and nonsynonymous substitutions between *Pinus thunbergii* and 10 angiosperms under analysis determined with the yn00 program. The 61 protein-coding genes shown on the diagram are common to 13 completely sequenced chloroplast genomes of the land plants. The black bars represent the distances based on the nonsynonymous substitutions in these proteins. The grey bars represent the distances based on the synonymous substitutions.

deviation in the amino acid composition among the land plant *rpo*B and *rpo*C1 genes we analyzed. In the alignment of *rpo*C2 sequences, the liverwort sequence did not pass the test. This, however, would not be expected to influence the phylogeny inference among angiosperms, since we included two tracheophyte outgroups in our data set, which did not exhibit compositional bias.

We concatenated 61 individual alignments of the first and the second codon positions from the coding genes common to 13 known chloroplast genomes of the land plants into one data set. Another data set was produced with the translated sequences of these genes. There were no missing sequences of complete genes in our data sets. After manual editing we obtained 30,017-positions-long nucleotide and 14,655-positions-long amino acid alignments of good quality.

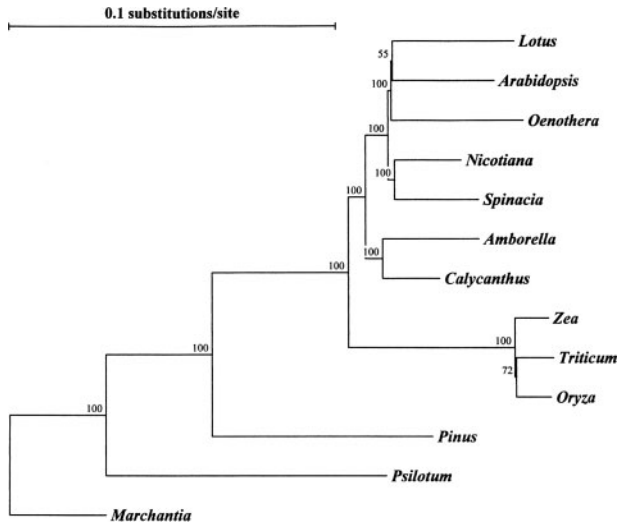

FIG. 3.—Neighbor-Joining tree built from Kimura two-parameter distances derived from analysis of the alignment of the first and the second codon positions from 61 protein-coding genes common to the plastomes of land plants.

Analyses of Nucleotide Sequences

A Neighbor-Joining (NJ) tree from Kimura two-parameter distances (as implemented in the TREECON package [Van de Peer and De Wachter 1994]) based on the alignment of the first and the second codon positions from 61 chloroplast protein-coding genes is presented in figure 3. On this tree, *Amborella trichopoda* does not form the most basal angiosperm clade. Instead, it appears on a common branch with *Calycanthus fertilis*, a member of Laurales, at the base of the eudicot cluster. Both this branch and the monophyly of all dicotyledonous plants under study received the highest bootstrap support (100/100). Applying Jukes-Cantor; Felsenstein F81; Felsenstein F84; Kimura three-parameter; Hasegawa, Kishino, and Yano; Tajima-Nei; Tamura-Nei; and General time-reversible models as implemented in the PAUP package (Swofford 2002) instead of the Kimura two-parameter model did not lead to any changes in tree topology or in bootstrap support values for these two branches.

We repeated all nine above analyses employing gamma correction (PAUP). The alpha shape parameter, estimated from the data (0.28) showed that most sites in the data set were constant or had evolved very slowly, but a few sites exhibited comparatively higher rates. Yet all nine trees we obtained yielded topologies identical to figure 3. The bootstrap support values for the branch uniting all dicotyledonous plants under analysis and for the branch bearing out *Amborella* and *Calycanthus* underwent no changes (100/100 BP).

Log determinant analysis (PAUP) produced the same topology with the same bootstrap proportion support for these two branches, which suggests that the topology on figure 3 is not likely to be an artifact of the compositional biases in different lineages.

These results are supported by maximum parsimony. Maximum parsimony with heuristic algorithm (PAUP and PHYLIP [Felsenstein 1989] implementations) recovered the same topology as shown on figure 3. The branch uniting *Amborella* and *Calycanthus* and the one bearing out all the dicotyledonous plants under study were found in all bootstrap samples (100/100 BP). The same result was obtained using branch and bound algorithm (PAUP). The regression analysis performed with the AUTODECAY program showed that the branch *Calycanthus-Amborella* was reproduced in the first 27 shortest trees, whereas the branch uniting all dicotyledonous plants under study disappeared starting from the 88th shortest tree. The constrained tree with *Amborella* as a sister group to other angiosperms was 87 steps longer than the most parsimonious tree, on which it formed a common clade with *Calycanthus* at the base of the dicotyledonous cluster.

Maximum-likelihood analyses were performed with the Tree-Puzzle program using the fast quartet-puzzling tree search algorithm (Strimmer and von Haeseler 1996). The program provides a method-specific statistical support for the internal tree branches—a percentage of topology bipartitions occurring in intermediate trees it builds in all puzzling steps. Branches showing quartet puzzling support (QPS) above 90% are considered as strongly supported. Employing Tamura-Nei and Hasegawa, Kishino, and Yano substitution models with default settings and the root set to *Marchantia*, we found the same topology as shown in figure 3 with, respectively, 93/94 QPS for the branch uniting all dicotyledonous plants and 99/99 QPS for the branch *Calycanthus-Amborella*. The competing topologies (eudicots basal or *Calycanthus-Amborella* branch basal) did not include one with *Amborella* basal to all angiosperms. We did not register a single bipartition out of 1,000 possible supporting such placement of *Amborella* in the above analyses.

Applying the substitution model fitted with the ModelTest program (Posada and Crandall 1998) to our nucleotide data set with all positions including gaps deleted, we received the ''standard'' topology, with *Amborella* forming the most basal branch and Calycanthus branching second. Changing the alpha shape parameter from 0.72 to 0.28 (the value we previously estimated from the data) with the fitted model resulted in a drastic change of the topology: Eudicots became paraphyletic with *Nicotiana* separated from the rest of them by the branch bearing out the grasses, whereas *Calycanthus* became the most basal and *Amborella* branched off second. A general drop of QPS values across the tree was observed.

Analyses of Protein Sequences

NJ trees built from Kimura, Tajima-Nei (TREECON implementation), and Dayhoff (PHYLIP implementation) protein distances on the basis of the 14,655–amino acids–long alignment yielded topologies identical to the topology presented on figure 3. Bootstrap support for the monophyly of the dicotyledonous plants and the *Calycanthus-Amborella* branch remained on the 100/100 BP level in these analyses.

The topology presented in figure 3 was also recovered in maximum-parsimony (PAUP) analyses of the above data set with 100/100 BP support for the branch bearing out all the dicots under analysis and 96/100 BP support for

the sister group relationship between *Calycanthus* and *Amborella*. We did not register an alternative topology with *Amborella* forming the most basal angiosperm clade in these analyses.

Maximum-likelihood analyses of this data set with Dayhoff, Schwartz, and Orcutt (1978), Henikoff and Henikoff (1992) (Blosum62), and Whelan and Goldman (2001) (WAG) amino acid substitution models as implemented in the Tree-Puzzle program yielded 99 QPS support for the *Amborella-Calycanthus* clade and 100 QPS for the sister group relationship between it and the clade uniting all eudicots (Rosopsida) under analysis. With Müller and Vingron (2000) and Jones, Taylor, and Thornton (1992) models, both these branches were recovered with 99 QPS.

Discussion

The majority of the analyses performed on the basis of joint analysis of the chloroplast coding sequences did not confirm the previously reported basal position of *Amborella trichopoda*. Instead, they showed that *Amborella* originated after splitting of flowering plants into dicots and monocots, being a member of the dicot cluster. In all cases in which the topology presented in figure 3 was recovered (31 out of 33 analyses), it received high bootstrap support (meaning that it was robust but not necessarily correct). More importantly, it was recovered using both nucleotide and protein data, diminishing the possibility that it might be a product of a systematic error.

In the two cases in which this topology could not be confirmed, complex fitted models were used. It is important to note that a small change in the fitted substitution model causes a big difference to the tree obtained. One of the recovered fitted topologies is obviously wrong. Multiple cases have previously been reported in which such fitted models failed to yield the correct tree (Yang 1997*b*; Posada and Crandall 2001), whereas unfitted models recovered correct topologies.

One can also note that, in contrast to the placement of *Amborella* as a sister group to all other angiosperms, the position recovered with unfitted models is in better compliance with the taxonomical views, which were maintained for more than a century. *Amborella trichopoda* was first described by Baillon (1869), who placed it in Monimiaceae (Laurales) because of the similarity of its male flowers with of those of *Hedycarya*. It was further treated accordingly by Bentham and Hooker (1880), Pax (1889), Perkins and Gilg (1901), and Perkins (1925). The female flowers of *Amborella* remained unknown to science until 1948, when they were described by Bailey and Swamy (1948) who also found that *Amborella* had vesselless xylem. In accordance with their suggestions, Pichon (1948) proposed a separate family for the plant, Amborellaceae, within Laurales. The separation of *Amborella* into a separate family was later supported by Money, Bailey, and Swamy (1952) in their detailed investigation of the morphology of Monimiaceae. Following suggestions by Bailey and Swamy (1948) they narrowed the definition of this family, also separating *Trimenia* and *Piptocalyx* into another new family, Trimeniaceae. Yet Money, Bailey, and Swamy (1952) regarded *Amborella* close to Monimiaceae because of their sharing of hippocrepiform sclereids and the similar forms and structures of the pollen.

Takhtajan (1966) viewed the vesselless condition of xylem as an archaic character in angiosperms. Writing on the origins of Laurales, he argued that they developed from certain ancient vesselless members of Magnoliales, the oldest order of angiosperms in his system. Magnoliales themselves, in his opinion, radiated from plants with the vesselless Winteraceae type of xylem. As a consequence, vesselless *Amborella* was placed by Takhtajan (1966) in a basal position within Laurales. This placement of *Amborella* was further supported by Cronquist (1981). Describing the taxonomic position of this plant, he wrote: ''It is clearly a member of Laurales, in which its primitively vesselless wood, alternate leaves, essentially hypogynous flowers, several carpels, abundant endosperm and stamens dehiscent by longitudinal slits mark it as an archaic type.''

On the other hand, association of *Amborella* with Nymphaeales is problematic from the morphological standpoint: in contrast to the water lilies, *Amborella* has small unisexual flowers and abundant endosperm versus endosperm almost lacking in Nymphaeales, which instead develop perisperm. There are significant differences in early endosperm development of *Amborella* and Nymphaeales (Floyd and Friedman 2001). Like Eudicots and Eumagnoliids, *Amborella* has a seven-celled female gametophyte and triploid endosperm, as opposed to the four-celled female gametophyte and diploid endosperm characteristic of Nymphaeales (Williams and Friedman 2002).

Amborella was placed at the root of all other angiosperms when molecular studies by Parkinson, Adams, and Palmer (1999), Qiu et al. (1999), and Soltis, Soltis, and Chase (1999) emerged. Yet, these studies were based on a limited number of characters derived from only a few genes. Furthermore, using unmasked sequences of chloroplast genes with high substitution rates at their synonymous sites (see fig. 2) in maximum-parsimony analyses aimed at finding affinity between angiosperms and spermatophyte outgroups can be misleading.

The uniformity and high resolution of the results provided by the data set of the coding genes from chloroplast DNA in the overwhelming majority of analyses give reasons to believe that further accumulation of the genomic cpDNA data of basal angiosperms should provide means to resolve the first stages of flowering plant evolution. However, further studies on the properties of these data and possible underlying causes influencing the position of *Amborella* obtained here using various models are needed. Although not systematically investigated here, it is possible that the use of different closely related outgroups may influence the position of *Amborella* in the tree. Further studies of this aspect using chloroplast genome sequences from further gymnosperms are required as well.

Acknowledgments

We would like to thank David Lorence for sending us plant material of *Amborella* and Pete Lockhart for

interesting discussion. This project was supported by a grant of the Deutsche Forschungsgemeinschaft (DFG).

Literature Cited

Arber, E. A. N., and J. Parkin. 1907. On the origin of angiosperms. J. Linnean Soc. **38**:29–80.

———. 1908. Studies on the evolution of angiosperms: the relationship of the angiosperms to Gnetales. Ann. Bot. (London) **22**:489–515.

Bailey, I. W., and B. G. L. Swamy. 1948. Amborella trichopoda Baill : a new morphological type of vesselless dicotyledons. J. Arnold Arb. **29**:245–254.

Baillon, H. 1869. Histoire des plantes. Vol. I. L. Hachette & Cie. Paris, London, Leipzig.

Barkman, T. J., G. Chenery, J. R. McNeal, J. Lyons-Weiler, W. J. Ellisens, G. Moore, A. D. Wolfe, and C. W. dePamphilis. 2000. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc. Natl. Acad. Sci. USA **97**: 13166–13171.

Bentham, G., and J. D. Hooker, 1880. Genera plantarum, III(1). Reeve & Company, London.

Cronquist, A. 1981. An integrated system of classification of flowering plants. Columbia University Press, New York.

Dayhoff, M. O., R. M. Schwartz, and B. C. Orcutt. 1978. A model of evolutionary change in proteins. Pp. 345–352 *in* M. O. Dayhoff, ed. Atlas of protein sequence structure, Vol. 5, Suppl. 3. National Biomedical Research Foundation, Washington, D.C.

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

Felsenstein, J. 1989. PHYLIP—phylogeny inference package. Version 3.2. Cladistics **5**:164–166.

Floyd, S. K., and W. E. Friedman. 2001. Developmental evolution of endosperm in basal angiosperms: evidence from Amborella (Amborellaceae), Nuphar (Nymphaeaceae), and Illicium (Illiciaceae). Plant Syst. Evol. **228**:153–169.

Goremykin, V., K. I. Hirsch-Ernst, S. Wölfl, and F. H. Hellwig. 2003. The chloroplast genome of the ''basal'' angiosperm *Calycanthus fertilis*—structural and phylogenetic analyses. Plant Syst. Evol. (in press).

Graham, S. W., and R. G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am. J. Bot. **87**:1712–1730.

Henikoff, S., and J. G. Henikoff. 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA **89**:10915–10919.

Hiratsuka, J., H. Shimada, R. Whittier et al. (16 co-authors). 1989. The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. **217**:185–194.

Hupfer, H., M. Swiatek, S. Hornung, R. G. Hermann, R. M. Maier, W-L. Chiu, and B. Sears. 2000. Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable Euoenothera plastomes. Mol. Gen. Genet. **263**:581–585.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–282.

Kato, T., T. Kaneko, S. Sato, Y. Nakamura, and S. Tabata. 2000. Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. DNA Res. **7**:323–330.

Lockhart, P. J., C. J. Howe, A. C. Barbrook, A. W. D. Larkum, and D. Penny. 1999. Spectral analysis, systematic bias, and the evolution of chloroplasts. Mol. Biol. Evol. **16**:573–576.

Maier, R. M., K. Neckermann, G. L. Igloi, and H. Kossel. 1995. Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Mol. Biol. **251**:614–628.

Martin, W., B. Stoebe, V. Goremykin, S. Hansmann, M. Hasegawa, and K. V. Kowallik. 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393**:162–165.

Meyen, S. V. 1986. Hypothesis of the origin of the angiosperms from Bennetitales by gamoheterotrophy: transition of characters from one sex to another. Zh. Obshzh. Biol. **47**:291–309 [in Russian].

Money, L. L., I. W. Bailey, and B. G. L. Swamy. 1952. The morphology and relationships of the Monimiaceae. J. Arnold Arb. **31**:372–404.

Müller, T., and M. Vingron. 2000. Modeling amino acid replacement. J. Comput. Biol. **7**:761–776.

Murray, M. G., and W. F. Thompson. 1980 Rapid isolation of high molecular weight DNA. Nucleic Acids Res. **8**:4321–4325.

Ogihara, Y., K. Isono, T. Kojim et al. (19 co-authors). 2002. Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Mol. Genet. Genomics. **266**:740–746.

Ohtani, K., H. Yamamoto, and K. Akimitsu. 2002. Sentitivity to *Alternaria alternata* toxin in citrus because of altered mitochondrial RNA processing. Proc. Natl. Acad. Sci. USA **99**:2439–2444.

Ohyama, K., H. Fukuzawa, T. Kohchi et al. (13 co-authors). 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature **322**:572–574.

Parkinson, C. L., K. L. Adams, and J. D. Palmer. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. Curr. Biol. **9**:1485–1488.

Pax, F. 1889. Monimiaceae. Pp. 94–105 *in* A. Engler and K. Prantl, eds. Die Natürlichen Pflanzenfamilien, III(2). W. Engelmann, Leipzig.

Perkins, J., and E. Gilg. 1901. Monimiaceae. Pp 1–122 *in* A. Engler, ed. Das Pflanzenreich, V(101). W. Engelmann, Leipzig.

Perkins, J. 1925. Übersicht über die Gattungen der Monimiaceae. W. Engelmann, Leipzig.

Pichon, P. 1948. Les Monimiacées, famille hétérogène. Bull. Mus. Nat. Hist. Nat. Paris **20**:383–384.

Posada, D., and K. A. Crandall. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics **14**:817–818.

———. 2001. Simple (wrong) models for complex trees: a case from retroviridae. Mol. Biol. Evol. **18**:271–275.

Qiu, Y-L., J. Lee, F. Bernasconi-Quadroni, D. E. Soltis, P. S. Soltis, M. Zanis, E. A. Zimmer, Z. Chen, V. Savolainen, and M. W. Chase. 1999. The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature **402**: 404–407.

Sato, S., Y. Nakamura, T. Kaneko, E. Asamizu, and S. Tabata. 1999. Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. **6**:283–290.

Schmitz-Linneweber, C., R. M. Maier, J. P. Alcaraz, A. Cottet, R. G. Herrmann, and R. Mache. 2001. The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Plant Mol. Biol. **45**:307–315.

Shinozaki, K., M. Ohme, M. Tanaka et al. (23 co-authors). 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. **5**:2043–2049.

Soltis, P. S., D. E. Soltis, and M. W. Chase. 1999. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. Nature **402**:402–403.

Staden, R., K. F. Beal, and J. K. Bonfield. 2000. The Staden package, 1998. Methods Mol. Biol. **132**:115–130.

Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

Takhtajan, A. 1966. Systema et phylogenia Magnoliophytorum. Nauka, Moscow, Leningrad.

Van de Peer, Y., and R. De Wachter. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. Comput. Appl. Biosci. **10**:569–570.

von Wettstein, R. 1924. Handbuch der systematischen Botanik. 3rd edition. Franz Deutige, Leipzig.

Wakasugi, T., J. Tsudzuki, S. Ito, K. Nakashima, T. Tsudzuki, and M. Sugiura. 1994. Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA **91**: 9794–9798.

Whelan, S., and N. Goldman. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. **18**:691–699.

Williams, J. H., and W. E. Friedman. 2002. Identification of diploid endosperm in an early angiosperm lineage. Nature. **415**:522–526.

Yang, Z. 1997*a*. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**: 555–556.

———. 1997*b*. How often do wrong models produce better phylogenies? Mol. Biol. Evol. **14**:105–108.

William Martin, Associate Editor

Accepted April 29, 2003