images, however, favor models with higher viscosities and thicker weak zones. The predicted deformation pattern (Fig. 1C) can match the observations (Fig. 1A) better if a weak zone of 13 km and a viscosity of 1.6 x 10^{18} Pa·s are used. Moreover, we find that the match between the InSAR images and model calculations can be improved if a nonuniform weak layer, (*24*), is used. The weak layer thickness at depth is determined from the geographical variations of the Moho depth (*23*). The calculated amplitude of uplift to the west of the Johnson Valley fault is closer to what is shown on the inteferogram (*25*).

A viscosity on the order of 10^{18} Pa·s in the lower crust is consistent with a maximum viscosity of the lower crust of 10^{19} Pa·s that is inferred from the uplift and tilting of Quaternary lake sediments on the Halloran Hills in the eastern Mojave desert (*26*). The weakness of the lower crust could be related to the thermal structure of the Basin and Range province, so it can help understand the physical mechanism responsible for the extension of the general area.

Our study on postseismic rebound does not resolve the mechanism responsible for interseismic deformation associated with major strike-slip faults (*27*). The postseismic deformation involves a sudden coseismic stress concentration close to the rupture zone, while the interseismic deformation only involves gradual strain concentration. So it is possible that a mechanism other than viscoelastic flow, such as stable sliding, is also related to the interseismic process (*27*).

Because viscosity governs the evolution of the stress field and thus the loading and unloading processes of major earthquake-generating faults, our estimate of the viscosity beneath the Landers earthquake region will help to assess earthquake hazards in southern California and further characterize the behavior of earhquake-related processes.

References and Notes

1. K. Sieh *et al.*, *Science* **260**, 171 (1993).
2. G. Peltzer *et al.*, *ibid.* **273**, 1202 (1996).
3. W. J. Bosl and A. Nur, *Eos* **78**, F491 (1997).
4. J. C. Savage and J. L. Svarc, *J. Geophys. Res.* **102**, 7565 (1997).
5. D. Massonnet, W. Thatcher, H. Vadon, *Nature* **382**, 612 (1996).
6. D. D. Jackson *et al.*, *Science* **277**, 1621 (1997).
7. Y. Bock *et al.*, *J. Geophys. Res.* **102**, 18013 (1997).
8. Z. K. Shen *et al.*, *Bull. Seismol. Soc. Am.* **84**, 780 (1994).
9. K. Heki, S. Miyazaki, H. Tsuji, *Nature* **386**, 595 (1997).
10. A. Nur and G. Mavko, *Science* **183**, 204 (1974).
11. J. B. Rundle, *J. Geophys. Res.* **83**, 5937 (1978); W. Thatcher and J. Rundle, *ibid.* **89**, 7631 (1984).
12. H. J. Melosh and A. Raefsky, *ibid.* **88**, 515 (1983); S. C. Cohen, *ibid.* **89**, 4538 (1984).
13. F. F. Pollitz, *ibid.* **102**, 17921 (1997).
14. J. C. Savage, *ibid.* **95**, 4873 (1990).
15. C. J. Marone, C. H. Scholtz, R. Bilham, *ibid.* **96**, 8441 (1991).
16. A. M. Dziewonski, G. Ekstrom, M. P. Salganik, *Phys. Earth Planet. Inter.* **77**, 151 (1993).
17. C. J. Ammon, A. A. Velasco, T. Lay, *Geophys. Res. Lett.* **20**, 97 (1993).
18. We note that the first-motion mechanism presented by Hauksson *et al.* (*20*) has a rake angle of 170° on a fault striking in N10°W. The mechanism solution of the second point source obtained by Qu *et al.* (*28*) (shown in Fig. 1), which is located near the southern end of the Emerson fault, has a rake angle of 153°, which involves even a larger dip-slip component than Hauksson *et al.*'s (*20*). Also several aftershocks near the Emerson fault show a significant thrust component (*20*). The aftershock distribution along the Emerson fault shown by figures 10a and 10b of (*20*) suggests a northeast dipping fault, which is qualitatively consistent with the solution of Qu *et al.* (*28*). F. Cotton and M. Campillo [*Geophys. Res. Lett.* **22**, 1921 (1995)] showed that the Landers earthquake is almost pure strike-slip, but the rake angle varies considerably around 180° on the Emerson and Camp Rock faults. Their average rake angle is about 176°. Surface rupture measurements [For example, J. R. Arrowsmith and D. D. Rhodes, *Bull. Seism. Soc. Am.* **84**, 511 (1994)] also showed some direct evidence of up to 2.5 m of vertical motion along a short segment of the Emerson fault. The various rake angles suggest that 7 to 45% of the seismic moment is released through dip-slip motion along the Emerson Fault segment, corresponding to 0.3 to 1.8 m of dip-slip displacement. In our calculations, we assume 0.7 m of dip-slip motion.
19. Z. K. Shen *et al.*, *Eos Trans. AGU* **78**, 477 (1997).
20. E. Hauksson, *et al.*, *J. Geophys. Res.* **98**, 19835 (1993).
21. E. Hauksson, *Bull. Seismol. Soc. Am.* **84**, 917 (1994).
22. D. J. Wald and T. H. Heaton, *ibid.*, p. 668.
23. K. B. Richards-Dinger and P. M. Shearer, *J. Geophys. Res.* **102**, 15211 (1997).
24. FEVER (Finite Element code for Visco-Elastic Rheology) is an object-oriented finite element software program developed by one of us (Deng). The code was implemented in C++ and systematically tested against many analytic solutions of linear and non-linear viscoelastic problems. The error of the numerical solution for almost all of the tested cases is within 1% of the corresponding analytic result.
25. We fixed the upper boundary of the weak zone to be 15 km deep, and assume that the lower boundary of the weak layer is the Moho surface, constrained from the Moho-reflected PmP arrivals (*23*). A thicker weak zone to the west of the Johnson Valley fault (Moho is about 32 km deep) (*23*) leads to a larger uplift rate in that region, compared to models with uniform Moho depth (28 km).
26. P. S. Kaufman and L. H. Royden, *J. Geophys. Res.* **99**, 15723 (1994).
27. L. E. Gilbert, C. H. Scholz, J. Beavan, *ibid.*, p. 23975.
28. J. Qu, T. L. Teng, J. Wang, *Bull. Seismol. Soc. Am.* **84**, 596 (1994).
29. Y. Okada, *ibid.* **82**, 1018 (1992); L. Erikson, User's manual for DIS3D: A three-dimensional dislocation program with applications to faulting in the earth, thesis, Stanford University (1986).

30. We thank M. Simons for a critical review of the manuscript and G. Peltzer, J. Savage, C. Scholz, K. Sieh, M. Spiegelman, L. Sykes, T. L. Teng, W. Thatcher, and many people in the seismo lab for discussion. This research was supported by SCEC. SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. This is SCEC contribution 446 and contribution number 8574 of the Division of Geological and Planetary Sciences, California Institute of Technology.

17 August 1998; accepted 2 November 1998

In Search of the First Flower: A Jurassic Angiosperm, *Archaefructus*, from Northeast China

Ge Sun,* David L. Dilcher,* Shaoling Zheng, Zhekun Zhou

Angiosperm fruiting axes were discovered from the Upper Jurassic of China. Angiosperms are defined by carpels enclosing ovules, a character demonstrated in this fossil. This feature is lacking in other fossils reported to be earliest angiosperms. The fruits are small follicles formed from conduplicate carpels helically arranged. Adaxial elongate stigmatic crests are conspicuous on each carpel. The basal one-third of the axes bore deciduous organs of uncertain affinities. No scars of subtending floral organs are present to define the individual fertile parts as floral units, but the leaf-like structures subtending each axis define them as flowers. These fruiting axes have primitive characters and characters not considered primitive.

It has been thought that angiosperms first appeared about 130 million years ago in the Lower Cretaceous (*1, 2*). There are several recent reports of Triassic, Jurassic, and lowermost Cretaceous-aged fossils identified as angiosperms (*3–7*), but none of these reports can be accepted as conclusive evidence for the presence of angiosperms. Many reports of early angiosperms are based on pollen, leaves, and wood with vessels, none of which are definitive characters of angiosperms. Some are based on flowers and fruits that are too poorly preserved to demonstrate ovules or seeds enclosed in the carpels. The unique character of angiosperms is that the ovules

G. Sun, Nanjing Institute of Geology and Palaeontology, Academia Sinica, Nanjing 210008, China. D. L. Dilcher, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. S. Zheng, Shenyang Institute of Geology and Mineral Resources, Shenyang 110032, China. Z. Zhou, Kunming Institute of Botany, Academia Sinica, Kunming 650204, China.

*To whom correspondence should be addressed. E-mail: gsun@jlonline.comi; dilcher@flmnh.ufl.edu

are completely enclosed in a carpel. Here, we describe such early angiosperm fruits collected from the Upper Jurassic "Jianshangou Bed" in the lower part of the Yixian Formation of Huangbanjiegou village near Shangyuan Town of Beipiao City, western Liaoning Province, northeast China (Fig. 1).

The Yixian Formation (*8–12*) consists of layers of volcanic rocks sandwiched between sedimentary rocks. The sedimentary rocks contain abundant freshwater and terrestrial fossils, including plants, bivalves, fish, conchostracans, ostracods, gastropods, insects, turtles, lizards, shrimps, dinosaurs, birds, and mammals that constitute the Jehol biota (*13, 14*). The Yixian Formation is about 2000 to 2500 m thick and is considered to be latest Jurassic in age (*8–15*). We classify our discovery as follows:

Division Magnoliophyta
Class Magnoliopsida
Subclass Archaemagnoliidae
Genus *Archaefructus* Sun, Dilcher, Zheng et Zhou, gen. nov.

Type-species: *Archaefructus liaoningensis* Sun, Dilcher, Zheng et Zhou, sp. nov.

Generic diagnosis: Reproductive axes branched or unbranched, bearing helically arranged fruits (follicles) on short pedicels. Fruits mature distally, occupying the distal two-thirds of an axis; carpels or stamens deciduous, leaving short peg-like pedicel bases on the proximal one-third of an axis. Fruits derived from conduplicate carpels commonly bearing three (two to four) ovules. Fertile axes are subtended by leaf-like structures (*16*). Details of the diagnosis of *Archaefructus liaoningensis* are presented in (*17*).

The fruits presented here are recognized as angiosperms on the basis of the ability to remove seeds completely enclosed within them. The occurrence of this angiosperm-defining character in *Archaefructus* is important because it demonstrates that *Archaefructus* has angiosperm affinities, and it establishes a benchmark in time for when the closed carpel is first found. This character occurs in combination with other reproductive characters, resulting in a new mixture of characters. This unique set of characters should change our understanding of the nature of the early angiosperm flower.

The pollen-bearing organs of *Archaefructus* are unknown. They were not present with carpels in the fossil material examined. The proximal one-third of each fertile axis has what appears to be pedicel bases that may have borne deciduous fruits or other organs such as stamens. *Archaefructus* may have been either unisexual (monoecious or dioecious) or bisexual. No pollen was found attached to any surface of the fruits or axes, and no angiosperm pollen has been isolated from the matrix. The only sterile organs associated with *Archaefructus* are two poorly preserved leaf-like structures (Fig. 2A). The lateral axis is borne in the axil of a leaf-like structure occurring on the main axis. Examination by epi-illumination and fluorescence epi-illumination of the surfaces of this fertile complex revealed some cellular detail of the epidermal cells covering the fruits and axes, but no pollen or evidence of scars of any deciduous organs were found except for the peg-like pedicels basal in each fertile axis. Therefore, there may have been deciduous floral organs of an unknown nature associated with these fruiting axes when they were young. These fossils are fruiting axes bearing individual conduplicate carpels (Fig. 2, A and B), and each axis should be regarded as originating from a floral unit (Fig. 2A). The elongate nature of the axes may have been more extended in the fruiting stage than at pollination. The crowded carpels at the apices suggest this (Fig. 2B). Also, in the young carpels the stigmatic tissue occupies proportionately more area, and the apical prominence appears to continue to enlarge as the carpel matures.

The carpels of *Archaefructus* are closed in a conduplicate fashion, contain more than one ovule, and are clustered together. Subtending each "flower" is a leaf-like structure consisting of a petiole that terminates in a branched pattern of possibly three major veins extending into a crumpled leaf lamina (Fig. 2A). A few of the basal reproductive organs were deciduous at maturity while the subtending leaves and distal carpels remained attached. It is possible that the crumpled leaf-like organs subtending each fertile axis were colored or patterned in some way to attract the attention of insect pollinators. The stigmatic surface may have produced an exudate on which the dipterians, known from the same sediments, may have fed (*12*). It is also possible that the extended tips on the stigmatic crests of each carpel functioned similarly to those of *Ascarina* of the Chloranthaceae, which is wind pollinated (*18*). Thus, there is no single pattern of pollination biology present, as is found for specific fossil angiosperm taxa occurring during the latter Cretaceous (*19*), but both insects and wind may have been involved. Insect pollination offers a biological environment that would have contributed to an early and rapid diversification of the angiosperms.

Archaefructus has helically arranged carpels, and the placement of the two leaf-like organs suggests this pattern continued in the foliage. In contrast, many members of Gnetales (*20*), found in the Mesozoic, are characterized by oppositely placed leaves, branches, and reproductive organs. Probable fossils of Gnetales that co-occur with *Archaefructus* in the Yixian Formation include *Chaoyangia liangii* (*21*) and *Eragrosites changii* (*22*). These were both described recently as the earliest record of angiosperms. *Chaoyangia liangii* has ribbed stems with conspicuous nodes, each bearing two oppositely arranged leaves. The stems branch oppositely to produce a cyme-like pattern on which winged fruits or seeds are borne. These winged fruits or seeds are similar to those previously described as *Gurvanella* (*23, 24*)

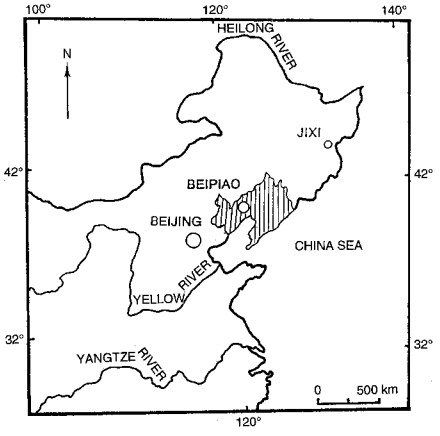


Fig. 1. Map showing the geographic location of the angiosperm fruiting axes *Archaefructus liaoningensis* gen. et sp. nov. Vertical lines represent Liaoning Province. Fossil localities are southwest of Beipiao.

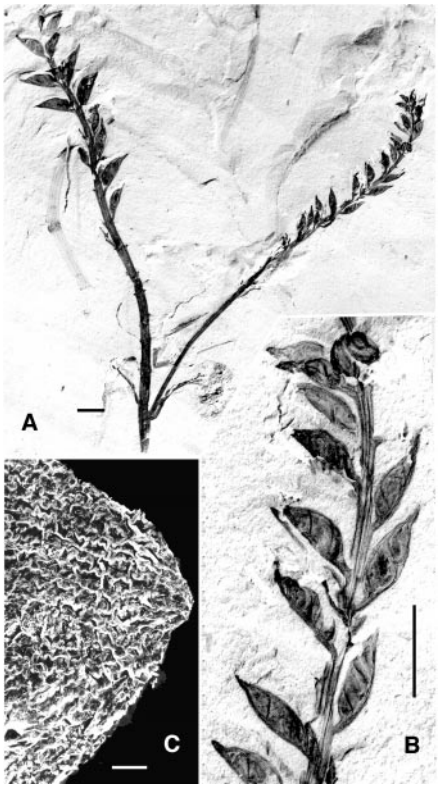


Fig. 2. *Archaefructus liaoningensis* Sun, Dilcher, Zheng et Zhou gen. et sp. nov. (**A**) Holotype, SZ0916; fruiting axes and remains of two subtending leaves. Scale bar, 5 mm. (**B**) Enlarged view of the carpels showing remains of the adaxial crest, abaxial venation, seeds in each carpel, and finger-like prominences. Scale bar, 5 mm. (**C**) Portion of a seed removed from a carpel, as viewed by scanning electron microscopy. Scale bar, 25 μm.

and have a distinct resemblance to the winged seeds of *Welwitchia mirabilis*. *Chaoyangia liangii* is an interesting fossil plant, but the ribbed stems, opposite branching, and winged fruits or seeds suggest that it has affinities with Gnetales rather than the angiosperms. It is unlike any living Gnetales, and careful analysis of the described specimen and additional material needs to be carried out. Before it can be accepted unequivocally as an angiosperm, the nature of the winged fruits or seeds must be clearly understood, and we conclude at this time that it most probably is an extinct genus of Gnetales.

Eragrosites changii is a name given to fossils interpreted as grass-like remains (*22*). These fossils also have reproductive organs borne on ribbed axes with distinct nodes that are oppositely branched, characters typical of Gnetales. In addition, the tightly crowded grass-like seed heads have oppositely arranged bracts that are reminiscent of the seed-bearing organs of *Ephedra* and *Welwitchia*. We consider this fossil grass to represent fossil remains of an extinct Gnetales, and it definitely is not an angiosperm.

Archaefructus presents a new set of characters not previously known in angiosperms. Typically the division Magnoliophyta (*25*) is used for angiosperms or flowering plants, and the class Magnoliopsida is used for the dicotyledons and Liliopsida for the monocotyledons. We suggest that a new subclass, Archaemagnoliidae, be constructed for angiosperms that do not conform to the character sets of any of the existing subclasses of the Magnoliophyta. This new subclass is characterized by flowers subtended by only a single leaf or leaf-like organ. Flowers consist of elongate receptacles bearing conduplicate carpels helically. The nature of the male floral organs is unknown at this time. Flowers appear to terminate axes and predate the evolution of any floral patterns. The subclass does not fit the concepts of "paleoherb" or of "eoangiosperm," as both represent collections of angiosperm taxa already more specialized and modified (*26*) than *Archaefructus* of the subclass Archaemagnoliidae.

Although *Archaefructus* fits the general plan of the "fundamental axis" for the primitive angiosperm (*26*), there are no subtending bracts present; the carpels, leaves, and branching are helical; and the development of carpels is conduplicate (plicate) (*27*) rather than ascidiate (*27, 28*). Some cladograms (*28, 29*) suggest that ascidiate carpels with one or two ovules are most primitive, on the basis of the occurrence of these characters in extant angiosperms such as Chloranthaceae. This family is schematically derived through a gnetalian ancestry based on these characters (*28*). *Archaefructus* does not support this proposed evolutionary scheme.

In extant angiosperms, ovules are formed on the inner surface of the carpel, which histologically is different from the outer surface. After fusion of the carpel, the ovules are enclosed and isolated from external environmental factors. Endress (*27*) has maintained that the so-called "open carpels" of some angiosperms are a myth because secretions produced by the inner lining of the carpels fill any gap. The pollen grain and the pollen tube are required to interact with the biochemical barrier as well as, in most carpels, the physical barrier to the male gametophyte presented by the closed carpel (*30, 31*). This important step in angiosperm reproduction is clearly well developed in the Upper Jurassic in *Archaefructus*. It allowed for incompatibility to develop between the male gametophyte and the carpel very early in angiosperm evolution.

Overall, *Archaefructus* looks more like a seed fern–type plant than like bennettitalian or gnetalian plants, which have received support as ancestral groups (*26, 32, 33*). The leaf-like nature of the fertile shoots, the helical disposition of the carpels, the conduplicate nature of the carpels with multiple ovules, and the subtending leafy structures are characters that would support the possible seed fern ancestry of *Archaefructus*. Gnetales are considered a sister group of the angiosperms, just as they might be thought of as a sister group of some of the Mesozoic seed ferns. The Mesozoic seed ferns are poorly understood and probably do not represent a natural group of plants. Many seed ferns became extinct during the Triassic or the Jurassic, and all became extinct by the mid-Cretaceous (*34*). Perhaps some lineages of Mesozoic seed ferns are the ancestors of the Mesozoic radiation of the angiosperms, explaining why Gnetales and angiosperms are often found to be sister clades.

Archaefructus is more than 85 mm long and consists of two fertile axes, which give rise to nearly 60 carpels and two leaves (Fig. 2, A and B). This compression-impression plant material was recovered by cleaving apart sedimentary layers of rock. The fossil is unlike the charcoalified remains, recovered by sieving, that have added much to our knowledge of early angiosperm reproduction (*35, 36*). Those flowers and fruits are minute relative to the material of *Archaefructus*. Thus, *Archaefructus* is a clear indicator that large reproductive axes of angiosperms existed early in angiosperm evolution, even if only a few have been recovered. This may suggest that the small angiosperm flowers and fruits of early angiosperms are derived and reduced to small sizes from an ancestor with large flowers.

By the mid-Cretaceous and into the lower Upper Cretaceous, a tremendous increase in angiosperm diversity appears in the fossil record (*37–41*). Nearly all of these fossils represent lines of evolution progressing toward extant taxonomic clades of angiosperms at the family or generic level (*24*). The evolution of modern angiosperm taxonomic groups thus seems to have transpired relatively quickly during the Lower Cretaceous.

For nearly a century, many paleobotanists and botanists have considered the angiosperms to have originated in the tropical regions of the world (*42–45*). The presence of *Archaefructus* and early angiosperms from the Lower Cretaceous of Jixi, in northeast China (*46, 47*), suggest that there were early angiosperms in China and that this was one of the areas where early diversification of the angiosperms was taking place. Angiosperms and angiosperm-like plants have also been reported from the early Cretaceous of Mongolia and Lake Baikal in eastern Russia (*23, 24, 48*). These fossils are similar to the Yixian flora and are associated with similar fossil fauna (that is, the Jehol fauna characterized by the *Lycoptera-Eoestheria-Ephemeropsis* assemblage). Therefore, angiosperms may have originated in Asia (*42*).

References and Notes

1. J. A. Doyle and L. J. Hickey, in *Origin and Early Evolution of Angiosperms*, C. B. Beck, Ed. (Columbia Univ. Press, New York, 1976), pp. 139–206.
2. P. R. Crane, *Nature* **366**, 631 (1993).
3. B. Cornet and D. Habib, *Rev. Palaeobot. Palynol.* **71**, 269 (1992).
4. B. Cornet, *Palaeontogr. Abt. B* **213**, 37 (1989).
5. _____, *Evol. Theory* **7**, 231 (1986).
6. _____, *Mod. Geol.* **19**, 81 (1993).
7. C. R. Hill, *Cretaceous Res.* **17**, 27 (1996).
8. Q. Ji and S. A. Ji, *Chin. Geol.* **233**, 30 (1996).
9. L. H. Hou, Z. Zhou, L. D. Martin, A. Feduccia, *Nature* **377**, 616 (1995).
10. D. M. Unwin, *Nature* **391**, 119 (1998).
11. C. C. Labandeira, *Science* **280**, 57 (1998).
12. D. Ren, *ibid.*, p. 85.
13. P. J. Chen, Z. M. Dong, S. N. Zhen, *Nature* **391**, 147 (1998).
14. P. J. Chen *et al.*, *Bull. Nanjing Inst. Geol. Palaeont. Acad. Sin.* **1**, 22 (1980).
15. P. E. Smith *et al.* [*Can. J. Earth Sci.* **32**, 1426 (1995)] suggested that the age of these rocks is Lower Cretaceous (122 million years). However, Chen *et al.* (*13*) maintain that the samples analyzed came from intrusive volcanic rocks many millions of years younger, rather than from the fossil-bearing sediments (which they presume to be about 145 million years ago) deposited in the Upper Jurassic. Associated with the angiosperm fruiting axes reported here are abundant remains of non-angiospermous plants, which indicate a Late Jurassic to Early Cretaceous age. We have identified these plants as *Thallites* sp., *Equisetites* spp., *Onychiopsis elongata* (Geyl.) Yok., *Coniopteris* spp., *Ruffordia goeppertii* (Dunk.) Sew., *Cladophlebis* spp., *Pterophyllum* sp., *Nilssonia* sp., *Otozamites* sp., *Neozamites* sp., *Baiera hastata* Krassilov, *Sphenobaiera* sp., *Czekanowskia rigida* Heer, *Solenites murrayana* L. et H., *Phoenicopsis* sp., *Schizolepis jeholensis* Yabe et Endo, *Pityophyllum* sp., *Pityocladus yabei* (Toyama et Oishi) Chang, *Elatocladus manchurica* (Yok.) Yabe, *Brachyphyllum* sp., *Sphenolepis* sp., *Gurvanella* sp. Krassilov, and *Potamageton? jeholensis* Yabe et Endo. The theropod dinosaur, *Sinosauropteryx*, from the lower part of this formation (*13*) is comparable to *Compsognathus* from the Upper Jurassic (Tithonian) Solnhofen Beds in Germany [P. Wellnhofer, *Archaeopteryx* **11**, 1 (1993)]. In addition, the mouth parts of flies from the Yixian Formation (*12*) are similar to those known from the Upper Jurassic of Kazakhstan (*11*). Therefore, we consider the lower Yixian Formation to be Late Jurassic in age.
16. Etymology: *Archae-* for ancient; *fructus* for fruiting. The species *liaoningensis* is named for the province in China from which the specimens were collected. Specific diagnosis: Same as for the genus. Holotype: Deposited in the Nanjing Institute of Geology and Palaeontology, Academia Sinica, number SZ0916

(Fig. 2). Age and stratigraphy: Upper Jurassic, Yixian Formation.
17. Description: Fertile axes vary in size (Fig. 2A). Main axis is 85 mm long from leaf axis and 3 mm wide basally, tapering to 1 mm wide distally. The lateral fertile axis originates from a leaf axil, 86 mm long and 1 mm wide basally, tapering to 0.3 mm wide distally. Fruits are attached by pedicels that are 0.75 to 1.5 mm long by 0.25 to 0.6 mm wide. Fruits are larger basally, 7 to 9 mm long by 2 to 3 mm wide, each containing three (two to four) seeds. Finger-like prominences extend about 1 mm from apex of fruits (Fig. 2B). The fruits are positioned at acute angles to the axis. The main axis has 18 fruits and 11 peg-like bases of pedicels (that bore deciduous reproductive organs) about 0.5 mm long. The lateral axis has 30 fruits and four peg-like remnants of the pedicels that bore deciduous reproductive organs, and smaller fruits, 5 to 6 mm long by 1.5 to 2 mm wide. The fruits are crowded at the axis apex where fruit size decreases. Fruits near the apex are 3 mm long by 2 mm wide, each with two seeds. One other fruiting axis (SZ0917, not figured) contains 17 fruits crowded into 35 mm of the fragmentary axis. Seeds fill fruits and have an oblique orientation. They appear to be attached to the adaxial side of the fruit. Seeds may overlap within fruits or may be distinctly separated by oblique bands of tissue. Cuticles of the seed coats are thin. Epidermal cells are rectangular-polygonal, about 25 to 45 μm by 12 to 20 μm. Anticlinal cell walls are sinuous and cutinized, about 2.5 to 3.5 μm thick (Fig. 2C). Periclinal cell walls are somewhat unevenly cutinized.
18. P. K. Endress, *Plant Syst. Evol.* **152**, 1 (1986).
19. E. M Friis and W. L. Crepet, in *The Origins of Angiosperms and Their Biological Consequences*, E. M. Friis, W. G. Chaloner, P. R. Crane, Eds. (Cambridge Univ. Press, Cambridge, 1987), pp. 145–179.
20. P. R. Crane, *Int. J. Plant Sci.* **157** (suppl.), S50 (1996).
21. S. Y. Duan, *Sci. China Ser. D* **41**, 14 (1998).
22. Z. Y. Cao, S. Q. Wu, P. A. Zhang, J. R. Li, *Chin. Sci. Bull.* **43**, 230 (1998).
23. V. A. Krassilov, *Palaeontogr. Abt. B* **181**, 1 (1982).
24. _____, *Angiosperm Origins: Morphological and Ecological Aspects* (PENSOFT, Sofia, Bulgaria, 1997).
25. A. Cronguist, *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York, 1981).
26. L. J. Hickey and D. W. Taylor, in *Flowering Plant Origin, Evolution and Phylogeny*, D. W. Taylor and L. J. Hickey, Eds. (Chapman & Hall, New York, 1996), pp. 176–231.
27. P. K. Endress, *Diversity and Evolutionary Biology of Tropical Flowers* (Cambridge Univ. Press, Cambridge, 1994).
28. D. W. Taylor and G. Kirchner, in (*26*), pp. 116–140.
29. D. W. Taylor and L. J. Hickey, *Plant Syst. Evol.* **180**, 137 (1992).
30. D. L. Dilcher, in *Monogr. Syst. Bot. Mo. Bot. Gard.* **53**, 187 (1995).
31. _____, *La importanica del origen de las angiospermas y como formaron el mundo alrededor de ellas*, VI Congreso Latinoamericano de Botanica, Mar Del Plata, Argentina (Royal Botanic Gardens, Kew, UK, 1996), pp. 29–48.
32. P. R. Crane, *Ann. Mo. Bot. Gard.* **72**, 716 (1985).
33. J. A. Doyle and M. Donoghue, *Bot. Rev.* **52**, 1 (1986).
34. T. N. Taylor and E. L. Taylor, *The Biology and Evolution of Fossil Plants* (Prentice-Hall, Englewood Cliffs, NJ, 1993).
35. E. M Friis and P. K. Endress, *Prog. Bot.* **57**, 253 (1996).
36. E. M. Friis, P. R. Crane, K. R. Pedersen, in *Evolution and Diversification of Land Plants*, K. Iwatsuki and P. H. Raven, Eds. (Springer-Verlag, Tokyo, 1997), pp. 121–156.
37. W. L. Crepet, K. C. Nixon, E. M. Friis, J. V. Freudenstein, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8986 (1992).
38. K. C. Nixon and W. L. Crepet, *Am. J. Bot.* **80**, 616 (1993).
39. P. S. Herendeen, W. L. Crepet, K. C. Nixon, *Plant Syst. Evol.* **189**, 29 (1994).
40. J. A. Keller, P. S. Herendeen, P. R. Crane, *Am. J. Bot.* **83**, 528 (1996).
41. M. A. Gandolfo, K. C. Nixon, W. L. Crepet, *ibid.* **85**, 964 (1998).
42. A. Takhtajan, in *Biogeographical Evolution of the Malay Archipelago*, T. C. Whitmore, Ed. (Oxford Univ. Press, New York, 1987), pp. 26–31.
43. D. I. Axelrod, *Science* **130**, 203 (1959).
44. D. Burger, *Rev. Palaeobot. Palynol.* **65**, 153 (1990).
45. J. A. Doyle, in *Proceedings of the 27th International Geological Congress, Palaeontology 2, USSR Academy of Sciences, Moscow, August 4–14* (VNU Science Press, Utrecht, Netherlands, 1984), pp. 23–33.
46. G. Sun, S. X. Guo, S. L. Zheng, T. Y. Piao, X. K. Sun, *Sci. China Ser. B* **36**, 253 (1993).
47. G. Sun and D. L. Dilcher, *Palaeobotanist* **45**, 393 (1996).
48. V. A. Vakhrameev, *Jurassic and Cretaceous Floras and Climates of the Earth* (Cambridge Univ. Press, Cambridge, 1991).
49. Supported by the National Natural Science Foundation of China (projects 39370055 and 39770059), the Academy of Sciences, China (project KZ952-S1-426), and the Becker/Dilcher Endowment for Paleobotany of the University of Florida Foundation. We thank Z. P. Zhang for his help in fossil collecting, and S. W. Mei, Y. Q. Mao, Z. Y. Song, S. W. Zhao, J. Y. Chen, C. Jaramillo, and T. Lott for their assistance with the preparation of the fossil material and this manuscript.

6 August 1998; accepted 22 September 1998

Tracking the Long-Term Decline and Recovery of an Isolated Population

Ronald L. Westemeier,* Jeffrey D. Brawn,† Scott A. Simpson, Terry L. Esker, Roger W. Jansen, Jeffery W. Walk, Eric L. Kershner, Juan L. Bouzat, Ken N. Paige

Effects of small population size and reduced genetic variation on the viability of wild animal populations remain controversial. During a 35-year study of a remnant population of greater prairie chickens, population size decreased from 2000 individuals in 1962 to fewer than 50 by 1994. Concurrently, both fitness, as measured by fertility and hatching rates of eggs, and genetic diversity declined significantly. Conservation measures initiated in 1992 with translocations of birds from large, genetically diverse populations restored egg viability. Thus, sufficient genetic resources appear to be critical for maintaining populations of greater prairie chickens.

The conservation implications of small population size are controversial (*1–4*). A significant loss in genetic variation may decrease fitness or limit the long-term capacity of a population to respond to environmental challenges (*5*). Alternatively, chance environmental and demographic events may pose a more immediate threat to small populations (*1, 2*). Conservation strategies can be different depending on the relative importance of these factors (*1, 3, 6*), but fundamental questions persist because there are few data on long-term changes in the demography and genetics of wild populations.

Here we report the results of a long-term study on a remnant population of greater prairie chickens (*Tympanuchus cupido pinnatus*) in southeastern Illinois (*7*). Over the 35-year period of this study, we documented concurrent declines in population size and fitness as well as an overall reduction in genetic diversity. In addition, we report on a conservation strategy initiated in 1992, whereby translocations of individuals from large, genetically diverse populations increased fitness.

Greater prairie chickens are grassland-dependent birds still found in areas of suitable habitat ranging from northwestern Minnesota south to northeastern Oklahoma, and from southeastern Illinois west to northeastern Colorado (*8*). Leks (or booming grounds) are used as arenas for territorial display and breeding by two or more males (*9*). Loss of habitat suitable for successful nesting and brood rearing is the single most important factor leading to declines, isolation, and extirpations throughout the species' range in the midwestern United States (*10*). The eastern subspecies *Tympanuchus cupido cupido*, also known as the heath hen, has been extinct since 1931 (*11*) and Attwater's prairie chicken *Tympanuchus cupido attwateri*, which is restricted to Texas, is near extinction (*12, 13*).

In Illinois, native prairie habitat for prairie chickens originally covered >60% of the state (Fig. 1), but fewer than 931 ha (<0.01%) of the original 8.5×10^6 ha of high-grade prairie remain (*14*). There were possibly several million prairie chickens statewide in the mid-19th century (*15*); by 1962 an estimated 2000 birds

R. L. Westemeier, Illinois Natural History Survey, Effingham, IL 62401, USA. J. D. Brawn, Illinois Natural History Survey, Champaign, IL 61820, USA. S. A. Simpson and T. L. Esker, Illinois Department of Natural Resources, Newton, IL 62448, USA. R. W. Jansen, Douglas-Hart Nature Center, Mattoon, IL 61938, USA. J. W. Walk and E. L. Kershner, University of Illinois, Department of Natural Resources and Environmental Sciences, Urbana, IL 61801, USA. J. L. Bouzat and K. N. Paige, University of Illinois, Department of Ecology, Ethology, and Evolution, Urbana, IL 61801, USA.

*To whom correspondence should be addressed.
†To whom e-mail should be addressed at j-brawn@uiuc.edu

ministration (NOAA)/Climate Prediction Center (CPC), Washington, DC (2000)]. Updates are available on the Web at www.cpc.ncep.noaa.gov.
5. K. E. Trenberth, J. G. Olson, *J. Clim.* **2**, 1196 (1989).
6. W. J. Randel, F. Wu, *J. Clim.* **12**, 1467 (1999).
7. J. W. Hurrell, H. van Loon, *Tellus* **46A**, 325 (1994).
8. D. W. Waugh, W. J. Randel, S. Pawson, P. A. Newman, E. R. Nash, *J. Geophys. Res.* **104**, 27191 (1999).
9. S. Zhou, M. E. Gelman, A. J. Miller, J. P. McCormack, *Geophys. Res. Lett.* **27**, 1123 (2000).
10. G. J. Marshall, V. Lagun, T. A. Lachlan-Cope. *Int. J. Climatol.*, in press.
11. D. G. Vaughan, G. J. Marshall, W. M. Connolley, J. C. King, R. Mulvaney, *Science* **293**, 1777 (2001) and references therein has a summary of recent temperature trends over Antarctica.
12. D. G. Vaughan, S. M. Doake, *Nature* **379**, 328 (1996).
13. S. S. Jacobs, J. C. Comiso, *J. Clim.* **10**, 697 (1997).
14. E. Hanna, *Weather* **54**, 71 (1999).
15. S. E. Stammerjohn, R. C. Smith, *Clim. Change* **37**, 617 (1997).
16. C. L. Parkinson, *Ann. Glaciol.*, in press.
17. Monthly mean radiosonde data for the stations listed in Table 1 were obtained from the NCEP with the assistance of J. Caron and K. Trenberth at the NCAR. The stations were chosen on the basis of the temporal continuity of their records. For each station, anomalies exceeding 3 SD about the long-term mean were removed from the data. Gridded monthly mean surface air–temperature anomalies were obtained from the Climate Research Unit at the University of East Anglia (*49*). Monthly and daily means of the NCEP/NCAR reanalysis (*50*) were obtained through the NOAA Climate Diagnostics Center. Halley station ozone data were provided by the British Antarctic Survey.
18. Data for the radiosonde stations listed in Table 1 and gridded monthly mean surface temperature data over Antarctica are available starting from 1958. However, because the temporal coverage of these data improved markedly in the late 1960s, our analysis is restricted to the period after 1969. Results based on the NCEP/NCAR reanalysis were found to be in excellent agreement with those derived from the radiosonde data after ~1979, which corresponds to the introduction of satellite data into the reanalysis assimilation scheme. For example, the monthly mean time series of 500-hPa geopotential height averaged over the radiosonde stations listed in Table 1 is correlated with 500-hPa geopotential height anomalies from the reanalysis averaged over the SH polar cap (poleward of 60°S) at a level of $r = 0.89$ for the period 1979–1998, and the amplitude and seasonality of trends calculated from these two time series are virtually identical. Results based on the NCEP/NCAR reanalysis were found to diverge from those derived from the radiosonde data before 1979, particularly in the lower stratosphere. For a thorough comparison of the NCEP/NCAR reanalysis and radiosonde data over Antarctica, see (*51*).
19. Water vapor and well-mixed greenhouse gas trends are estimated to contribute about 25 and 10%, respectively, to the observed cooling trends in the lowermost Antarctic stratosphere, with the remainder due to the massive depletion of ozone [see (*52*)].
20. D. T. Shindell, D. Rind, P. Lonergan, *Nature* **392**, 589 (1998).
21. The robustness of the seasonality of the trends was assessed by recalculating the trends for odd-numbered years only. In both the troposphere and stratosphere, the trends exhibit a seasonality qualitatively identical to that noted in Fig. 1 and Table 2.
22. M. Shiotani, *J. Meteorol. Soc. Jpn.* **68**, 461 (1990).
23. J. W. Kidson, *J. Clim.* **1**, 183 (1988).
24. D. J. Karoly, *Tellus* **42A**, 41 (1990).
25. D. L. Hartmann, F. Lo, *J. Atmos. Sci.* **55**, 1303 (1998).
26. D. Gong, S. Wang, *Geophys. Res. Lett.* **26**, 459 (2000).
27. D. W. J. Thompson, J. M. Wallace, *J. Clim.* **13**, 1000 (2000).
28. The NH counterpart to the SAM is alternatively referred to as the North Atlantic Oscillation (*36*, *53*), the Arctic Oscillation (*37*), and the NAM (*27*).
29. J. G. Charney, P. G. Drazin, *J. Geophys. Res.* **66**, 83 (1961).
30. Time series of the SAM are typically based on the difference in geopotential height between polar and middle latitudes. Because radiosonde data are scarce in the SH middle latitudes, the SAM index used in this study is based solely on polar data. In practice, the time series of the SAM derived from 500-hPa height anomalies averaged over the radiosonde stations in Table 1 is highly correlated with an index of the SAM based on empirical orthogonal function analysis of gridded SH geopotential height anomalies, as used in (*24*, *26*, *27*). The 500-hPa time series shown in Fig. 2 is correlated with the corresponding leading principal component time series of SH (20° to 90°S) 500-hPa geopotential height anomalies from the NCEP/NCAR reanalysis at $r = 0.84$, based on all calendar months of the year, 1979–1998.
31. The significance of the correlations in Table 3 was estimated from the *t* statistic, assuming only one degree of freedom for every 2 years. The significance of the linkages in Fig. 3 was estimated assuming one degree of freedom for every year (the time series used in Fig. 3 consist of 6 months per year).
32. Because the trend in the SAM accounts for a relatively small fraction (9%) of the total month-to-month variance during December-May from 1969–1998, the fractions of the trends that are linearly congruent with the SAM index in Fig. 3 are not strongly sensitive to shared trends in the time series. For the shortest period of record considered (December-May monthly means from 1979–2000), the total fraction was found to vary by ~5% when the indices were detrended before the calculation of the regression coefficients.
33. The correlation between the time series of 500- and 30-hPa height anomalies averaged over the polar cap for the April-May season is $r = 0.62$. The corresponding correlation between April-May 30-hPa height anomalies and total column ozone at Halley station during the previous November is $r = 0.23$.
34. H. K. Roscoe, A. E. Jones, A. M. Lee, *Science* **278**, 93 (1997).
35. H. L. Miller, R. W. Sanders, S. Solomon, *J. Geophys. Res.* **104**, 18769 (1999).
36. J. W. Hurrell, *Science* **269**, 676 (1995).
37. D. W. J. Thompson, J. M. Wallace, *Geophys. Res. Lett.* **25**, 1297 (1998).
38. _____, G. C. Hegerl, *J. Clim.* **13**, 1018 (2000).
39. M. P. Hoerling, J. W. Hurrell, T. Xu, *Science* **292**, 90 (2001).
40. D. T. Shindell, R. L. Miller, G. Schmidt, L. Pandolfo, *Nature* **399**, 452 (1999).
41. D. T. Shindell, G. A. Schmidt, R. L. Miller, D. Rind, *J. Geophys. Res.* **106**, 7193 (2001).
42. J. C. Fyfe, G. J. Boer, G. M. Flato, *Geophys. Res. Lett.* **26**, 1601 (1999).
43. P. J. Kushner, I. M. Held, T. L. Delworth, *J. Clim.* **14**, 2238 (2001).
44. D. M. H. Sexton, *Geophys. Res. Lett.* **28**, 3697 (2001).
45. M. P. Baldwin, T. J. Dunkerton, *J. Geophys. Res.* **104**, 30937 (1999).
46. _____, *Science* **294**, 581 (2001).
47. D. W. J. Thompson, M. P. Baldwin, J. M. Wallace, *J. Climate*, in press.
48. Intergovernmental Panel on Climate Change, *Climate Change 2001: The Science of Climate Change* (Cambridge Univ. Press, Cambridge, UK, 2001).
49. P. D. Jones, *J. Clim.* **7**, 1794 (1994).
50. M. E. Kalnay *et al.*, *Bull. Am. Meteorol. Soc.* **77**, 437 (1996).
51. G. J. Marshall, *J. Clim.* **15**, 659 (2002).
52. P. M. de F. Forster, K. P. Shine, *Geophys. Res. Lett.* **26**, 3309 (1999).
53. H. van Loon, J. Rogers, *Mon. Weather Rev.* **106**, 296 (1978).

54. We thank K. E. Trenberth, J. Caron, and W. J. Randel (NCAR) for assistance with the Antarctic radiosonde data; T. Gill (South African Meteorological Service) and G. Bodeker (National Institute for Water and Atmospheric Research) for providing additional station data; P. D. Jones for providing surface temperatures over Antarctica; J. M. Wallace, K. E. Trenberth, M. P. Baldwin, and G. Marshall for helpful comments at various stages of this research; and H. K. Kim and W. Higgins at the NOAA CPC for the inspiration for Fig. 4. D.W.J.T. is supported by the NSF under grant CAREER: ATM-0132190.

21 December 2001; accepted 26 March 2002

Archaefructaceae, a New Basal Angiosperm Family

Ge Sun,[1]* Qiang Ji,[2] David L. Dilcher,[3]* Shaolin Zheng,[4] Kevin C. Nixon,[5] Xinfu Wang[6]

Archaefructaceae is proposed as a new basal angiosperm family of herbaceous aquatic plants. This family consists of the fossils *Archaefructus liaoningensis* and *A. sinensis* sp. nov. Complete plants from roots to fertile shoots are known. Their age is a minimum of 124.6 million years from the Yixian Formation, Liaoning, China. They are a sister clade to all angiosperms when their characters are included in a combined three-gene molecular and morphological analysis. Their reproductive axes lack petals and sepals and bear stamens in pairs below conduplicate carpels.

The fossil record provides information about the evolution of major groups of organisms living on Earth today as well as those that have become extinct. The earliest history of flowering plants is poorly documented. Some of the sparse data from fossils have been accommodated into current phylogenetic models. Current phylogenetic studies (*1, 2*) and recent paleobotanic finds (*3*) support the nature of the basal angiosperms (*Amborella* and Nymphaeales) consistent with combined multiple gene and morphologic analyses (*4–6*). Newly discovered fossils reveal a combination of unique characters. These fossils consist of new material of *Archaefructus liaoningensis* (*7*) and *A. sinensis* sp. nov. (*8*), a new species preserved as nearly whole plants in various stages of reproductive maturity. The fossils were recovered from the lower part of the Upper Jurassic/Lower Cretaceous Yixian Formation (*9*) in Beipiao and Lingyuan of western Liaoning, China (41°12′N, 119°22′E). The formation is at least 124.6 million years old (*10*) and may be as old as uppermost Upper Jurassic (*11*). All aspects of these plants are known, including

their roots, leaves, and reproductive organs, as complete plants with all organs attached. When all characters of *A. liaoningensis* (*7*) and *A. sinensis* (*8*) are evaluated in a phylogenetic context, they require a new extinct family of flowering plants, Archaefructaceae (*12*). An analysis of the characters of this family demonstrates that it is best considered a sister taxon to extant angiosperms (Fig. 1).

Morphologic characters, especially those of the reproductive (flower) organs, have been the traditional basis for organizing the phylogeny of the angiosperms (*13*). We need to integrate a detailed morphologic character database with the molecular database in order to place fossils into the whole record of angiosperm phylogeny. The early angiosperm fossils have various levels of preservation of morphologic characters but lack any molecular characters. With sufficient data, some fossils can be intercalated into existing taxa in the current molecular-based angiosperm phylogenies (*3*). However, when novel character combinations are present that do not clearly align a fossil with particular extant angiosperm families, and when no molecular data are available, it becomes necessary to use methods that combine morphologic and molecular characters in a "total evidence" cladistic analysis. In this context, we performed numerous phylogenetic analyses of *Archaefructus* with modern angiosperms, using a combined matrix of morphology and molecular data. Figure 1 presents the results of one such analysis, in which we reduced the number of morphologic characters to only those relevant to the fossil. In all analyses, *Archaefructus* was maintained in a position as a sister taxon to the extant angiosperms.

The "flower" of *Archaefructus* is a unique collection of female and male reproductive organs (Fig. 2, A to C, E to G, J, and K). The carpels mature last, after the pollen has been dispersed and the anthers have been lost on the same axis. The shoot apex terminates in the carpel production. The immature carpels are clustered close together and then become spaced out as the axis elongates and they mature; most are arranged helically. Each carpel is attached to the axis by a pedicle that has no visible bract scars or evidence that other organs were ever attached near them. The same is the case for the stalks upon which the pairs of stamens are borne. In Archaefructaceae, the carpels are terminal, pseudo-whorled in threes or subopposite to helical in arrangement, and subtended by helical stamen-bearing stalks.

The stamens were produced in pairs and remained attached to the stalks only while the carpels were young (Fig. 2, A to C, E, J, and K), as suggested by their small size and close spacing. As the carpels matured, the stamens abscised, leaving the short stalks that remain on the mature shoots. Two stamens commonly arise from the terminus of each stalk. The stamens consist of short slender filaments and long anthers. The anthers are basifixed and consist of two distinct parallel thecae, each probably containing two longitudinal pollen sacs. This is the typical organization of modern angiosperm anthers (*14–17*). Each theca opened by a longitudinal slit extending the full length of the anther. Once opened, the anthers probably remained open.

The anthers often show apical extensions (Fig. 2, B, C, and E) that may have served as pollinator attractants (*14–17*). These stamens demonstrate a distinct differentiation between the short filament and the nonlaminar anthers. This finding supports the hypothesis that there is no homologous relationship between the stamen and the carpel (*18*). The pre-Cenomanian record of stamens is sparse, but new discoveries in Lower Cretaceous sediments hold promise that more will be found (*3, 19, 20*). Each stamen record provides useful information for the phylogenetic analysis of angiosperm characters (*21–25*). For example, the presence of nonlaminar stamens early in angiosperm history supports the view that stalked anthers are primitive.

Stamen bundles are formed in a variety of living angiosperms when there is a secondary subdivision of the androecial primordium (*26*). As a result of this secondary primordial activity, a single primordium may produce several stamens. The stamens produced this way are basally fused. The stalks found in *Archaefructus* may represent the remains of stamen filaments that are fused together. The paired stamens of *Archaefructus* may be collateral pairs of stamens that result in doubling of organs, as has been observed in the Magnoliidae and in the Alismatidae (*26*). Such stamen pairs resulting from paired initial primordia might reflect an ancient history of this character found in the stamen bundles of *Archaefructus.* This type of primordia in the androecium has been presented as a possible primitive character (*27, 28*). An alternative hypothesis is that the two stamens attached to the stalks are the last remnant of larger branching systems that contained male flowers or terminal isolated stamens. In support of this hypothesis, the paired stamens attached to the stalks might represent the last remnants of stamens attached to a reduced branching system. We think the paired stamens are best construed as dichotomous remnants of an earlier, more extensive dichotomous branching system.

Pollen has been obtained in situ from the anthers. The pollen is monosulcate (Fig. 2, F and G). It is of moderate size (17 to 36 μm long) with an exine pattern that is vermiform (Fig. 2F) or fossulate, similar to large monosulcate pollen described from the Lower Cretaceous (*29*). A granular texture is evident in high magnification on a scanning electron microscope (SEM) (Fig. 2G). Under epifluo-

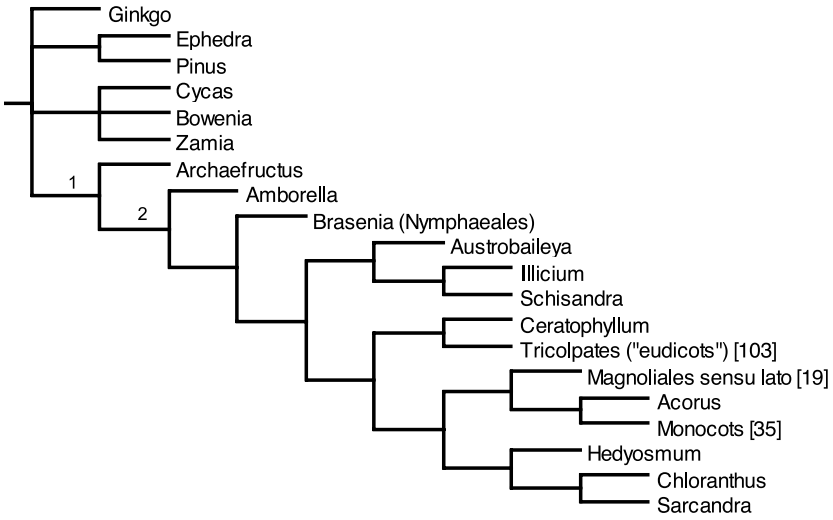


Fig. 1. Consensus cladogram of most parsimonious trees for analysis of 173 living taxa of seed plants, plus the fossil *Archaefructus*. Various analyses included 1628 molecular characters and 17 to 108 morphological characters (*47*). Taxa with numbers in brackets after some names indicate the number of species in that clade that were analyzed as separate terminals and monophyletic in all trees but are not shown here to save space. The Bremer support (or "decay index") for the branches subtending the angiosperms and *Archaefructus* is indicated above the branches. For taxa and character matrix, see (*48*).

[1]Research Center of Palaeontology, Jilin University, Changchun 130026, China. [2]Geological Institute of Chinese Academy of Geosciences, Beijing 100037, China. [3]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA. [4]Shenyang Institute of Geology and Mineral Resources, Shenyang 110032, China. [5]L. H. Bailey Hortorium, Cornell University, Ithaca, NY 14853, USA. [6]Nanjing Institute of Geology and Palaeontology, Academia Sinica, Nanjing 210008, China.

*To whom correspondence should be addressed. E-mail: sunge@jlu.edu.cn, dilcher@flmnh.ufl.edu

rescent microscopy, we observed some isolated pollen on the stigmatic crests of the carpels, particularly on the extended tips of the young carpels (*7*).

Relatively long pedicles and apical extensions characterize the young carpels associated with the shoots bearing stamens (Fig. 2J). The apical extension may be an elongation of the adaxial stigmatic crest that, in combination with the elongated pedicle, could help accommodate wind or insect pollination. As the carpels matured, they probably bent outward, forming a wider angle with the shoot. At the same time, the relative proportions of the parts of the carpel changed, so that the extension of the carpel tip and the pedicle are not as obvious in the fossils of mature fruits.

Even though the stamens matured while the carpels were young, it is impossible to know whether *Archaefructus* was protandrous. The functional nature of the pollen and stigmas cannot be determined from the fossils, but it is possible for pollen to mature and be dispersed before the carpels associated on the same axis were receptive. This type of dichogamy would increase the fitness of *Archaefructus* by establishing a self-isolating mechanism to ensure outbreeding (*16*). It is possible that *Archaefructus* possessed the po-

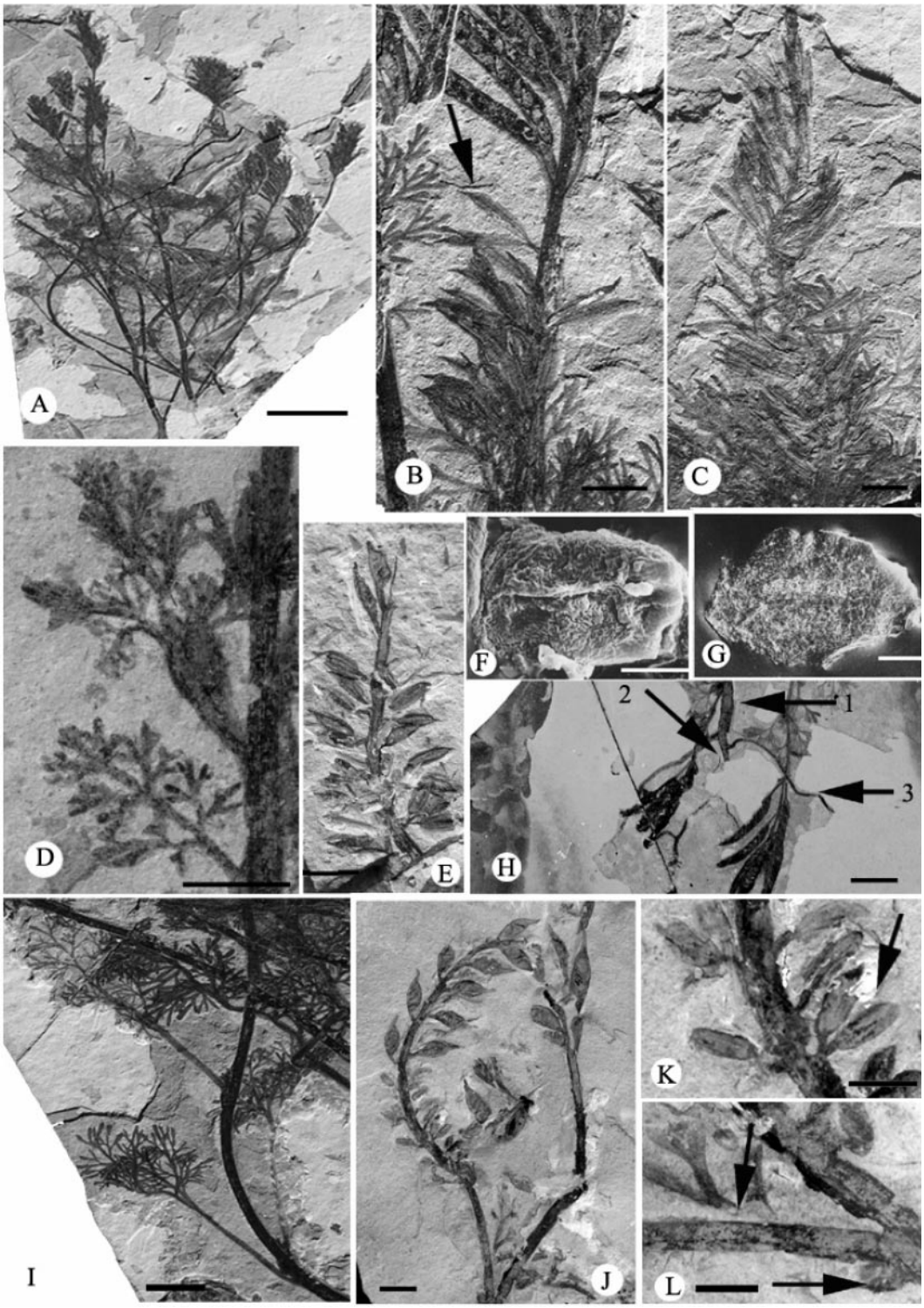


Fig. 2. (**A** to **D**, **H**, and **I**) *Archaefructus sinensis* Sun, Dilcher, Ji et Nixon: samples J-0721 (A, B, C, I), NMD-001 (D), and NMD-002 (H). (**E** to **G**, **J** to **L**) *Archaefructus liaoningensis* Sun, Dilcher, Zheng et Zhou: samples B-2000 (J to L), PB18943 (E to G). (A) Whole specimen (holotype). Scale bar, 5 cm. (B and C) Multiseeded stalked carpels and paired stamens (indicated by arrow). Scale bars, 5 mm. (D) Swollen leaf base of upper dissected leaf. Scale bar, 5 mm. (E) Two young fruits with stamens below, from specimen PB18943. Scale bar, 3 mm. (F and G) SEM images of pollen from specimen PB18943. Pollen grains show monosulcate aperture (F) and rugulate exine (G). Scale bars, 10 μm. (H) Base of stem (1) of a fruiting plant folded in half upon itself; root (2) bears a few simple lateral roots (3). Scale bar, 1 cm. (I) Lower dissected leaf. Scale bar, 5 mm. (J) Paratype showing mature fruits and peglike projections on the main shoot; to the left side is a lateral shoot with stamens and young fruits. Scale bar, 5 mm. (K) Enlargement of portion of (J). Shoot shows paired stamens (indicated by arrow). Scale bar, 2 mm. (L) Same as (J). A leaf base with missing petiole that extended across the lateral shoot with the distal portion of the dissected leaf preserved. Arrows indicate leaf base and leaf blade. Scale bar, 2 mm.

tential in its reproductive biology to avoid self-pollination. The small size of the immature carpels and the ovules contained in them suggests that unless pollination and fertilization occurred, they would never have developed into mature fruits and seeds. Early in angiosperm history, fruit development was probably tied to successful pollination. In this way, angiosperms avoided investing energy in the production of sterile fruit and seed tissue.

The carpels and stamens are borne together on the same flowering shoots. The carpels are terminal and the stamens subtend them. There are no petals, sepals, or other organs associated with the carpels and stamens. The fertile shoots are produced in a leaf axis or a leaf may subtend terminal fertile shoots. The leaf petioles remain attached to these fertile shoots. The leaf blades are seldom preserved in specimens of *A. liaoningensis*, whereas they are more often preserved attached in *A. sinensis* (Fig. 2, A, D, and I). In one specimen of *A. liaoningensis*, the partial remains of a preserved leaf are attached (Fig. 2, J and L). More complete isolated dispersed leaves, three to four times pinnately compound, are found associated from the same sediments.

The Archaefructaceae probably were aquatic plants. The herbaceous nature of the plants is obvious by the thin stems that extend for some distance, which would require water for support. The finely dissected compound leaves suggest an aquatic habitat. The basal leaves have long petioles and are larger than the more distal leaves. The basal leaves are more dissected than the distal leaves. All leaves have a swollen petiole base. It is especially enlarged in those leaves that are most distal from the base but nearest to the reproductive organs and probably the surface of the water. These may have given some buoyancy to the plant. The roots are poorly developed in *A. sinensis* (Fig. 2H). Numerous fish (*Lycoptera*) are preserved and mixed in with the fossil plants or found in association with both species. The reproductive organs of *Archaefructus* probably were exposed above the water during pollination and may have remained so for seed dispersal.

A Lower Cretaceous (125 to 115 million years old) fossil with affinities to the Nymphaeales (*3*) is more specialized in its floral morphology and much smaller than *Archaefructus*. Although *Archaefructus* has some features similar to Cabombaceae in the Nymphaeales, it does not fit the characters of that order or any extant order. We consider *Archaefructus* distinct from the Nymphaeales and *Amborella*. Figure 1 suggests that *Archaefructus* is a sister taxon to all known angiosperms. The characters that are unique to the Archaefructaceae are illustrated in the reconstruction of *A. sinensis* (Fig. 3).

Although the reproductive structures of *Archaefructus* superficially resemble those of *Caytonia* and other seed ferns of similar or older age, they are different when examined closely. *Archaefructus* has female structures clearly interpretable as angiospermous carpels that are closed along an adaxial stigmatic crest, the stamens are angiospermous with bilateral symmetry, and pollen is nonsaccate and monosulcate. In contrast, the outer seed-enclosing structure of *Caytonia* is not conduplicate, and in recent phylogenetic analyses is usually interpreted as homologous with the outer seed integument (*25*). In the male structures usually reconstructed as belonging with *Caytonia* (*Caytonanthus*), the stamen symmetry is radial and the pollen is bisaccate, as in various modern conifers (*30*). All these features place *Caytonia* (and other seed ferns) outside of the clade formed by *Archaefructus* + extant angiosperms. Because of the fragmentary nature of Mesozoic seed ferns such as *Caytonia*, and the markedly different and complex interpretations of their reproductive structures, it was not possible or advisable to include these in our cladistic analyses. For example, in addition to uncertainty in interpretation of homology of the female structures in *Caytonia*, the female and male structures are not organically connected and were likely borne on separate axes. In the current analysis, these taxa would be unstable but would clearly not be closer to angiosperms than *Archaefructus*, which possesses strictly angiospermous features of carpels borne above stamens on bisexual axes. It can be confidently stated that *Archaefructus* is the closest phylogenetically to angiosperms of any available fossil, but has features that exclude placing it within the angiosperm "crown group" or extant clade.

Fig. 3. Reconstruction of *A. sinensis*. These are terminal shoots. The main shoot is more mature so the stamens are deciduous, leaving short pegs. The latter shoot is younger, the carpels are smaller, and the stamens are borne in pairs on short pegs. For interpretation and reconstruction, see (*48*). [Diagram by K. Simons and D. Dilcher]

The lack of similarity between *Archaefructus* and other known fossil plants during the Upper Jurassic/Lower Cretaceous (*31*) provides us with more information about the primitive angiosperm (as defined by the characteristic of seeds enclosed in carpels) than it does about related pre-angiospermous seed plants. However, it should be noted that, like other angiosperms (both fossil and living), *Archaefructus* does not represent the original angiosperm and likely had its own derived features. The complex of features seen in *Archaefructus* provides an important point of extrapolation to the original angiosperm, suggesting the possibility that it lacked petals and sepals (previous phylogenies without *Archaefructus* favor an ancestor with a perianth) and may have been a submerged aquatic (like some Nymphaeales). *Archaefructus* is, rather, part of a complex basal group in angiosperm evolution.

Detailed examination of the reproductive shoots in *Archaefructus* by epifluorescence microscopy reveals a continuous covering of epidermal cells preserved on these shoots extending between the attached organs.

When examined cell by cell, no interruptions in the cell pattern were found that would indicate the presence of scars left by deciduous bracts or other organs, so each entire shoot represents a flower, not an inflorescence. In *Archaefructus* the leaves help to define the morphologic floral units, beginning at the point of its origin in a leaf axis and extending to the reproductive tip of the shoot (*11*). Thus, the *Archaefructus* flower terminates a lateral branch system. Paired stamens arise from the short stalks on these lateral axillary shoots. Each of these stamen units could be considered reduced complex male branches existing in *Archaefructus* as small individual male flowers. In that case, the *Archaefructus* flower would be an inflorescence derived from a complex branched ancestor. On the other hand, the stalk with the stamen pairs could be viewed as a normal androecium in a single flower.

If we consider that the term "flower" is related to an organizational plan (*26*), then it is reasonable to consider that each set of multiple carpels and stamens helically arranged along individual elongated shoots and subtended by leaves could be considered a flower. However, *Archaefructus* does not lend itself to such easy interpretation. The "flower" of *Archaefructus* may represent a stage in evolution in which its reproduction was angiospermous (ovules enclosed in carpels) while the organization of the traditional floral unit(s) was still poorly defined. Perhaps, just as the evolutionary history of the modern ovulate pine cone can be understood only by knowing its complex branching ancestors, the evolutionary history of the flower may also involve complex branched ancestral axes.

The origin of the organization of reproductive organs seen in *Archaefructus* has a bearing on how we view the potential ancestor of flowering plants. Many theories and hypotheses of angiosperm origin have been proposed; among these are the euanthium (euthial) (*32*) and the pseudanthium (*33–35*) theories. The euanthium theory indicates that the angiosperm flower organization evolved from a bisexual strobilus with numerous, helically arranged ovules and pollen-bearing organs, as found in *Cycadoidea* or other Mesozoic bennettialean fossil plants. The ovule and pollen-bearing organs were already differentiated and associated with perianth organs that were suggested to be conspicuous and attractive to insect pollinators (*36, 37*). The pseudanthium theory proposed that the ancestral plants of angiosperms had separate branching systems containing ovules and pollen organs. As the hypothetical floral units that formed these branching systems were clustered together, they eventually condensed and modified into a shoot with terminal carpels subtended by stamens. Petals and sepals probably evolved from subtending modified leaves and became part of the whole unit subtending these newly organized reproductive shoots. Ancestors with male and female organs on separate branches are found in the Mesozoic seed ferns. *Archaefructus* appears to support the pseudanthium theory, with the stalks bearing paired stamens perhaps being remnants of an earlier branching system while the petals and sepals have not yet evolved from associated subtending leaves.

References and Notes

1. M. C. Zane, D. E. Soltis, P. S. Soltis, S. Mathews, M. J. Donoghue, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
2. D. E. Soltis *et al*., *Bot. J. Linn. Soc.* **133**, 381 (2000).
3. E. M. Friis, K. R. Pedersen, P. R. Crane, *Nature* **410**, 357 (2001).
4. J. A. Doyle, *Mol. Phylogenet. Evol*. **9**, 448 (1998).
5. O. I. Nandi, M. W. Chase, P. E. Endress, *Ann. Mo. Bot. Gard*. **85**, 137 (1998).
6. J. A. Doyle, P. K. Endress, *Int. J. Plant Sci.* **161** (suppl. 6), S121 (2000).
7. *Archaefructus liaoningensis* Sun, Dilcher, Zheng et Zhou (Fig. 2, E to G, J to L). Emended description: Main fertile shoots with lateral fertile shoots common (Fig. 2, J and L). Lateral shoots borne in the axes of leaves, and leaves often subtending the main fertile shoot. Main shoots up to 85 mm long and 3 mm wide basally, tapering to 1 mm distally. Lateral fertile shoots up to 86 mm long, tapering from 1 mm wide basally to 0.5 mm distally. Fruits attached by pedicles (0.75 to 1.5 mm long, 0.25 to 0.6 mm wide). Fruits larger basally (7 to 10 mm long, 2 to 3 mm wide), each containing two to four seeds. Finger-like prominences extend about 1 mm past the apex of the fruits (Fig. 2, E and J). Young fruits positioned at acute angles to the shoot while mature fruits spread out at wider angles (Fig. 2, E and J). The fruits are follicles derived from conduplicate carpels. The main shoot has numerous fruits (up to 18 were observed on one specimen) and the lateral shoots also terminate in numerous fruits (up to 30 were observed). The fruits are crowded at the shoot apex and decrease in size distally. Fruits near the apex are about 3 mm long by 2 mm wide, each with two seeds. The seeds fill the fruits and have an oblique orientation. They appear to be attached to the adaxial side of the fruits. Seeds may overlap within the fruits or may be distinctly separated by oblique bands of tissue. Cuticle of the seed coats are thin. Epidermal cells are rectangular-polygonal, about 25 to 45 μm by 12 to 20 μm. Anticlinal cell walls are sinuous and cutinized, about 2.5 to 3.5 μm thick. Periclinal cell walls are somewhat unevenly cutinized. Often 10 to 12 short (0.3 to 0.5 mm long) stalks are located about 2.5 to 5 mm below the fruits in a zone that extends for about 15 mm along the shoot. These stalks may commonly have two or rarely one (perhaps very rarely three) stamens (Fig. 2, E, J, and K). Between 15 and 26 stamens have been observed. Stamens are deciduous as the shoot matures. Stamens consist of a short filament (0.25 to 0.5 mm long) basifixed to an anther (2 to 3 mm long by 1 to 1.2 mm wide) and may have an extended tip (0.5 to 1 mm long) (Fig. 2E). The narrow tip of the anther extends past the thecae and may be a narrow attenuate connective tip or an extension of the filament. The anthers appear to have two distinct thecae parallel in arrangement and perhaps each containing two longitudinal pollen sacs. The pollen isolated from the anthers is more or less elliptic (17 to 36 μm long, 15 to 20 μm wide). The pollen appear monosulcate and have a verriform or fossulate/rugulate exine pattern. Exine appears granular under the SEM (Fig. 2, F and G). The leaves, attached (Fig. 2L) and associated, are small, pinnately dissected three or four times. The petiole base may be slightly swollen. Multiple vascular strands (up to five) are observed in the petiole. Leaves frequently subtend a central fertile shoot or bear a lateral fertile shoot in their axis. The presence of nearly complete isolated leaves suggests that they were deciduous. Petiole about 10 mm, pinnate leaflet branches opposite to alternate, further dissected ending in rounded lobes, 0.5 to 1 mm wide.
8. *Archaefructus sinensis* Sun, Dilcher, Ji et Nixon sp. nov. (Fig. 2, A to D, H, and I). Description: Plants herbaceous, reproductive axes subtended by vegetative shoots 30.1 cm long by 17 cm wide (Fig. 2A). Main axes basally 3 mm wide, narrowing gradually upward, apically 1 mm wide. Roots poorly developed, consisting of a primary and few short secondary axes (Fig. 2H). Leaves dissected, leaf blade dissected two to five times, petiole length variable (0.5 to 4.0 cm). Basal leaves with long petioles with slightly swollen bases (Fig. 2I). Leaves near reproductive organs have short petioles and swollen bases (Fig. 2D). Ultimate leaf segments are about 2 mm long by 0.3 mm wide with rounded tips. Freely forming lateral branches (1 to 1.5 mm wide) in leaf axils diverge from the main stem at 30° to 35°. Each lateral branch terminates in a fertile shoot. Fertile axes terminate with numerous (12 to 20) small (10 to 18 mm long, 1.5 to 2.0 mm wide) carpels. These are subtended by several (8 to 18) short, blunt, helical stalks, each bearing two stamens. Carpels small when anthers mature; carpels helical, whorled, or opposite. Carpels matured into elongate follicles containing multiple seeds (8 to 12). Stamens consist of short fine filaments (about 1 mm long) attached to broad long anthers (4 to 5 mm long, 0.5 to 0.8 mm wide) ending with a prominent tip (0.5 mm long). Petals, sepals, or bracts absent. Holotype: J-0721. Specimen deposited in Geological Institute of Chinese Academy of Geoscience, Beijing, China. Etymology: The specific name refers to where the fossil was found. Number of specimens examined: 5 (two have 6 each, one has 12 reproductive axes attached together in each plant or in various stages of maturity) (Fig. 2A).
9. The age of the Yixian Formation of western Liaoning, China, is uncertain. We allow for a range from 145 to 125 million years, which corresponds with the Upper Jurassic to Lower Cretaceous (*12, 31, 38–45*).
10. C. C. Swisher *et al*., *Nature* **400**, 58 (1999).
11. G. Sun, D. L. Dilcher, S. Zheng, Z. Zhou, *Science* **282**, 1692 (1998).
12. Family Archaefructaceae Dilcher, Sun et Nixon fam. nov. Description: Herbaceous, aquatic plants with branching stems. Branches originate as axillary shoots terminating in reproductive organs. Poorly developed sparsely branched roots. Leaves alternate, petiole bases enlarged, petioles of various lengths. Blades pinnately dissected two to five times, stipules absent. Flowers medium size (5 cm long) and terminal on axillary branches, hypogynous. Carpels numerous, helical to whorled, stalked, conduplicate. Stamens borne in pairs on short helically arranged stalks, filaments short, large basifixed anthers with terminal apical extensions. Pollen monosulcate, exine reticulate to verriform. Fruits multiseeded folicules, persistent stamens deciduous with stalks persisted on floral axis. The family Archaefructaceae consists of a single genus, *Archaefructus*, with two species known, *A. liaoningensis* and *A. sinensis*. These species are currently known to occur in western Liaoning, northeastern China, from Upper Jurassic/Lower Cretaceous sediments (*9*). *Archaefructus liaoningensis* and *A. sinensis* are the current species recognized in the family.
13. A. Cronquist, *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York, 1981).
14. W. G. D'Arcy, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 1–24.
15. P. K. Endress, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 92–110.
16. P. Bernhardt, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 192–220.
17. L. Hufford, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 58–91.
18. W. Burger, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 111–117.

19. W. L. Crepet, K. C. Nixon, in *The Anther, Form, Function and Phylogeny*, W. G. D'Arcy, R. C. Keating, Eds. (Cambridge Univ. Press, New York, 1996), pp. 25–57.
20. E. M. Friis, K. R. Pedersen, P. R. Crane, *Int. J. Plant Sci.* **161** (suppl. 6), S169 (2000).
21. K. C. Nixon, W. L. Crepet, D. A. Stevenson, E. M. Friis, *Ann. Mo. Bot. Gard.* **81**, 484 (1994).
22. M. J. Donoghue, J. A. Doyle, in *The Hierarchy of Life,* B. Fernholm, K. Bremer, H. Jornvall, Eds. (Elsevier, New York, 1989), pp. 181–193.
23. H. Loconte, D. W. Stevenson, *Brittonia* **42**, 197 (1990).
24. _____, *Cladistics* **7**, 267 (1991).
25. J. A. Doyle, M. J. Donoghue, *Bot. Rev.* **52**, 321 (1986).
26. P. K. Endress, *Diversity and Evolutionary Biology of Tropical Flowers* (Cambridge Univ. Press, Cambridge, 1994).
27. W. Hagemann, in *Contemporary Problems in Plant Anatomy,* R. A. White, W. C. Dickinson, Eds. (Academic Press, Orlando, FL, 1984), pp. 301–350.
28. P. Leins, C. Erbar, *Bot. Acta* **104**, 336 (1991).
29. J. V. Ward, J. A. Doyle, C. L. Hotton, *Pollen Spores* **31**, 113 (1989).
30. T. M. Harris, *The Yorkshire Jurassic Flora. II. Caytoniales, Cycadales & Pteridosperms* (British Museum [Natural History], Oxford, 1964).
31. G. Sun, S. L. Zheng, D. L. Dilcher, Y. D. Wang, S. W. Mei, *Early Angiosperms and Their Associated Plants from Western Liaoning, China* (Shanghai Scientific and Technological Education Publishing House, Shanghai, 2001).
32. E. A. N. Arber, J. Parkin, *Ann. Bot. (London)* **22**, 489 (1908).
33. R. R. von Wettstein, *Handbuch der Systematischen Botanik* (Franz Deuticke, Leipzig, ed. 2, 1907).
34. V. A. Krassilov, *Angiosperm Origins: Morphological and Ecological Aspects* (Pensoft, Sofia, Bulgaria, 1997).
35. P. R. Crane, *Ann. Mo. Bot. Gard.* **72**, 716 (1985).
36. T. Delevoryas, *Palaeontographica Abt. B* **121**, 122 (1968).
37. W. L. Crepet, *Palaeontographica Abt. B* **148**, 144 (1974).
38. M. P. Doludenko, E. R. Orlovskaya, *Jurassic Flora of the Karatau* (Science Press, Moscow, 1976).
39. S. E. Wang *et al*., *Jurassic System. In Stratigraphy of China* (Geological Press, Beijing, 2000).
40. Q. Ji *et al*., *Prof. Pap. Stratigr. Palaeont.* **27**, 74 (1999).
41. S. A. Ji, Q. Ji, K. Padian, *Nature* **398**, 573 (1999).
42. J. F. Zhang, *Paleoworld* **11**, 150 (1999).
43. C. H. Lo *et al*., *Geochemica* **28**, 406 (1999).
44. D. F. Wang, *Bull. Chin. Acad. Sci.* **7**, 65 (1983).
45. P. E. Smith *et al*., *Can J. Earth Sci.* **32**, 1426 (1995).
46. K. C. Nixon, *Cladistics* **15**, 407 (1999).
47. Consensus cladogram of most parsimonious trees for analysis of 173 living taxa of seed plants, plus the fossil *Archaefructus*. Various analyses included 1628 molecular characters and 17 to 108 morphological characters. The molecular characters are based on the three-gene matrix (*rbcL*, *atpB*, *18s*) that was recently published for 567 species (*2*). Taxa were selected to provide a good representation of variation throughout the angiosperms, including a dense sampling of the so-called basal angiosperms. The tree shown was generated with a matrix of 1645 characters (17 morphological characters, including only those relevant characters that could be scored for the fossil). Parsimony analysis was undertaken using the parsimony ratchet of Nixon (*46*), with numerous runs of 200 replications for each analysis. In all analyses, *Archaefructus* is a sister taxon to the angiosperms as shown in this tree. Depending on the data set used, the overall length of the tree varied, with an overall consistency index of ~0.18 (consistent with the original three-gene analysis). The taxa *Cycas*, *Bowenia*, *Zamia*, *Ginkgo*, *Ephedra*, and *Pinus* represent the modern gymnosperms; the other taxa in the analysis are angiosperms (flowering plants). Note that this data set does not address the question of whether the gymnosperms are monophyletic, because no taxa outside of the seed plants were included. The tree has been drawn to be neutral on this point, and it supports either hypothesis equally. The tree differs from the original three-gene analysis only in the position of *Ephedra*, which in these trees is more consistent with analyses of other genes that place gnetopsids with Pinaceae, suggesting that the morphology may play a positive role in resolving discrepancies.
48. See supplemental material on *Science* Online (www.sciencemag.org/cgi/content/full/296/5569/899/DC1).
49. Supported by the National Natural Science Foundation of China (projects 39970050 and 39770059), Key Project of Ministry of Land and Resources, China, Special Project Stz. 98-1-02 (Chinese Academy of Sciences), and by NSF grants INT-0074295 and DEB 0090283 and the Becker/Dilcher Endowment for Paleobotany of the University of Florida Foundation. We thank T. A. Lott for assistance with photography, illustration, and manuscript preparation; K. Simons for diagramming the reconstruction of *Archaefructus*; and J. Y. Liu and C. L. Gao of the Natural Museum of Dalian for their help in observation of some specimens of *Archaefructus*. This paper is University of Florida Contribution to Paleobiology publication number 525.

28 December 2001; accepted 20 March 2002

Mammal Population Losses and the Extinction Crisis

Gerardo Ceballos[1] and Paul R. Ehrlich[2]

The disappearance of populations is a prelude to species extinction. No geographically explicit estimates have been made of current population losses of major indicator taxa. Here we compare historic and present distributions of 173 declining mammal species from six continents. These species have collectively lost over 50% of their historic range area, mostly where human activities are intensive. This implies a serious loss of ecosystem services and goods. It also signals a substantial threat to species diversity.

Population extinctions are a more sensitive indicator of the loss of biological capital than species extinctions. This is because many of the species that have lost a substantial portion of their populations [thus altering ecosystems and perhaps reducing the ability of those systems to deliver services (*1*)] are unlikely to go globally extinct and enter the species extinction statistics in the foreseeable future (*2*). Most analyses of the current loss of biodiversity emphasize species extinctions (*3–5*) and patterns of species decline (*6–8*) and do not convey the true extent of the depletion of humanity's natural capital. To measure that depletion, we need to analyze extinctions of both populations and species. Here we give a rough minimum estimate of the global loss of continental mammal populations. We believe that mammals, because of their great taxonomic diversity and the wide range of ecological niches they exploit, can serve as an indicator of what is occurring in the rest of Earth's biota.

Our data consist of historic (i.e., mostly 19th century) and present-day distributional ranges of all of the terrestrial mammals of Australia and subsets of the terrestrial mammal faunas of Africa, South East Asia, Europe, and North and South America (Table 1 and table S1). These subsets consist of all mammal species whose ranges are known to be shrinking for which we had access to data. They comprise roughly 4% of the ~4650 known species. We assume that loss of range area is due to the extinction of populations, but we do not attempt to equate a given areal loss with a precise number of population extinctions due to the complexities of defining and delimiting populations (*9*). Data were gathered from the specialized literature (Web references). In general, because they are better known, most of our range data are from medium- and large-sized species. Whether globally these are more or less liable to population extinction than medium to small species is a matter of conjecture (*10–12*), but at present there is little reason to assume an important directional bias in our samples. There was no correlation between body mass and range shrinkage in our data ($P > 0.05$, $r^2 = 0.22$). There does remain a possible source of bias in the relative lack of very small species in the total sample (*12*).

The ranges were digitized and the historic and present range areas were calculated. For each species, we estimated both total area occupied historically and percent historic range area now occupied. Using ArcView 3.1, the ranges were superimposed to produce synthetic maps summarizing the losses of species populations in 2 degree by 2 degree quadrats (i.e., the number of species that have disappeared from each quadrat because all of their populations previously located in that quadrat have disappeared). The area of these quadrats, of course, varies with latitude, but the average of such quadrats over land is about 30,000 km^2.

Declining species of mammals in our sample had lost from 3 to 100% of their

[1]Instituto de Ecologia, Universidad Nacional Autonoma de Mexico, Apdo, Postal 70-275, Mexico D.F. 04510, Mexico. [2]Department of Biological Sciences, Stanford University, Stanford, CA 94305–5020, USA. E-mail: gceballo@miranda.ecologia.unam.mx (G.C.); pre@Stanford.edu (P.R.E.)

Molecular Phylogenetics and Evolution
Vol. 18, No. 1, January, pp. 117–126, 2001
doi:10.1006/mpev.2000.0868, available online at http://www.idealibrary.com on IDEAL®

The Mitochondrial nad2 Gene as a Novel Marker Locus for Phylogenetic Analysis of Early Land Plants: A Comparative Analysis in Mosses

Susanne Beckert,[1] Hermann Muhle,* Dagmar Pruchner, and Volker Knoop

*Molekulare Botanik and *Systematische Botanik und Ökologie, Universität Ulm, D-89069 Ulm, Germany*

Received February 29, 2000; revised September 12, 2000; published online December 13, 2000

The mitochondrial nad2 gene is established as a novel marker locus for phylogenetic analyses among early land plants. The potential of this gene for phylogenetic resolution was checked with a broad taxon sampling of 42 mosses (Bryopsida, including the enigmatic genus *Takakia*) to allow both a comparative analysis with the recently explored nad5 gene and the fusion of independent data sets. The mitochondrial gene sequences provide valuable phylogenetic information on the relationships of classically defined orders and their respective monophylies. The more rapidly diverging sequences of a group I intron in nad5 and of a group II intron in nad2 add information for fine resolution. Although both genes provide phylogenetic information in the same taxonomic range (above family level), the combined sequence alignment results in an approximate doubling in the number of nodes with significant bootstrap support (>90). According to our data, Buxbaumiales are a paraphyletic taxon in a key position between the earliest branching taxa (Sphagnales, Takakiales, Andreaeales, Polytrichales, and Tetraphidales) and all other orders, possibly to be placed in the subclass Bryidae. A dichotomy in the latter recalls two previously suggested superorders Hypnanae and Dicranannae. Both genes independently question the monophyly of the orders Dicranales and Neckerales and reject the inclusion of the genera *Schistostega*, *Timmia*, and *Encalypta* among Eubryales. © 2000 Academic Press

Key Words: **mosses; bryophytes; evolution; molecular phylogeny; mitochondrial DNA; group II introns.**

INTRODUCTION

Although an abundant fossil record for early vascular plants, most notably lycopods, exists, fossil documents of old land plants with bryophyte characters are rare and controversial (Edwards *et al.*, 1995). Recent phylogenetic analyses, however, clearly support the idea of bryophytes as the earliest land plants (Kenrick and Crane, 1997; Qiu and Palmer, 1999). Systematically, bryophytes have been split into three classes: liverworts, hornworts, and mosses. However, the relationships of these three groups relative to the presumed algal ancestors on the one hand and the vascular plants on the other have remained unclear. Moreover, the monophyly of at least one of these classes, the liverworts, has been questioned (Capesius and Bopp, 1997). Based on the presence of three mitochondrial introns in all land plant groups except the liverworts, Qiu *et al.* (1998) recently deduced liverworts as the earliest branching land plants, confirming suggestions of a bryophyte paraphyly (Mishler and Churchill, 1984).

Rather than short mitochondrial protein-coding sequences of the cox3 gene (Malek *et al.*, 1996), we recently used larger sequence regions from the faster-evolving nad5 gene to derive phylogenetic trees for cryptogamous plants (Beckert *et al.*, 1999; Vangerow *et al.*, 1999). In this contribution we investigate the potential of a novel mitochondrial protein gene, nad2, for phylogenetic analyses among early land plants. This mitochondrial gene is unique among land plants in carrying the only group II intron that is positionally conserved over the large evolutionary distance (presumably exceeding 400 Myr) between the liverwort *Marchantia polymorpha* (Oda *et al.*, 1992) and the seed plant mitochondrial DNAs, such as that of *Arabidopsis thaliana* (Unseld *et al.*, 1997).

Given the observation that at least one group of liverworts, the Marchantiidae, has a strikingly low sequence divergence in mitochondrial DNA (Beckert *et al.*, 1999), we have chosen to investigate the mosses as an alternative early branching and divergent group of land plants. This choice allows a simultaneous exhaustive comparative analysis with the earlier established nad5 data set. Both genes, nad2 and nad5, encode protein subunits of the NADH dehydrogenase (complex I of the mitochondrial respiratory chain).

[1] Current address: Angewandte Mikrobiologie und Hygiene, Universität Ulm, D-89069 Ulm, Germany.

1055-7903/00 $35.00

The aims of this study were (i) a comparison of the phylogenetic information content in two plant mitochondrial loci and within group I and group II intron sequences, (ii) a test for congruency of phylogenetic implications from the two independent data sets and for a differential power of resolution on different taxonomic levels, and (iii) the reconstruction of a phylogeny of mosses with a wide taxon sampling covering all extant orders.

The trees based on mitochondrial DNA extend studies based on other widely used molecular markers, such as the nuclear 18S rRNA (e.g., Capesius and Stech, 1997), the chloroplast rbcL gene (e.g., Goffinet *et al.*, 1998), chloroplast ITS sequences (Samigullin *et al.*, 1998), or combinations of nuclear and chloroplast sequences (Cox and Hedderson, 1999; Hyvönen *et al.*, 1998). The phylogenetic implications of the novel mitochondrial data from the nad2 and nad5 genes are discussed in relation to those deduced from the other studies.

MATERIALS AND METHODS

Plant material and nucleic acid preparation. Plant material was generally collected in the field and vouchered (Table 1), with parallel attempts at cultivation in the greenhouse with varying success. DNA of some species was derived from sterile moss cultures on agar kindly provided by Dr. I. Capesius, Heidelberg. *Takakia lepidozioides* DNA was kindly provided by Dr. Y.-L. Qiu, Zürich. Total nucleic acids were extracted from green plant material in the presence of cetyltrimethylammonium bromide (CTAB) or alternatively with the Plant DNeasy kit (Qiagen). The CTAB extraction method (Doyle and Doyle, 1990) was modified by adding 1% polyvinyl–pyrrolidon (PVP 40), incubating at RT for 15 min, and extracting once with phenol–chloroform. DNA and RNA were differentially precipitated in the presence of 2 M lithium acetate.

PCR amplification, cloning, and sequencing. The novel primers NAD2UP (5′-gga gtt gtn ttt agt acc tct aa-3′) and NAD2DO (5′-agt agt aac gay ttn tca cga tcc at-3′) were designed for amplification of the nad2 gene (see Fig. 1). The upstream primer K (5′-ata tgt ctg agg atc cgc ata g-3′) and the downstream primer L (5′-atc ttt ggc caa gga tcc tac aaa-3′) were routinely used for amplification of the nad5 gene region (Beckert *et al.*, 1999). PCR amplification assays contained 1 μl template DNA (approximately 10 ng–1 μg), 10 μl 10× PCR buffer (100 mM Tris/HCl, pH 8.85, 250 mM KCl, 50 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$), 250 mM each dNTP, 0.25 μg of each primer, 2.5 U DNA polymerase, and double-distilled water to 100 μl. Different commercially available thermostable DNA polymerases were used, e.g., a mixture (90:1) of *Taq* DNA Pol (Gibco BRL) and *Pwo* DNA Pol (Boehringer Mannheim). A typical amplification assay included an initial denaturation (5 min, 94°C) followed by 35 cycles with 1 min denaturation at 94°C, 1 min annealing at 50–55°C, and 2 min 30 s synthesis at 72°C, and a final step of synthesis for 6 min at 72°C. PCR fragments were blunt-end ligated into pBlueskript II SK^+ (Stratagene). Positive clones were sequenced with a Thermosequenase kit (Amersham) using Cy5-fluorescence-labeled oligonucleotides and run on an Alf Express sequencer (Pharmacia). Sequencing primers were universal and reverse primers of the polylinker sequence and additional primers matching internal sequences of the cloned gene fragments.

Sequence analysis. Sequence handling was done with the UWGCG (Genetics Computer Group, University of Wisconsin) software package 9.1 for UNIX (Devereux *et al.*, 1984) including the construction of alignments with the program PILEUP as described in the text.

Phylogenetic tree construction was done with the commercial version (4.02b for Power Macintosh) of the PAUP* software (Swofford, 1999). Factory default settings were used for phylogenetic analyses of the alignments in heuristic searches, notably TBR (tree bisection–reconnection) branch swapping and the minimum-evolution objective function for distance matrix (DM) analyses. Starting trees were obtained via neighbor-joining (NJ) for DM and via stepwise addition for maximum-parsimony (MP) and maximum-likelihood (ML) analyses. Gaps were treated as missing data. Bootstrap values were derived from heuristic searches (generally 500 replicates with TBR, 5000 replicates without TBR for trees in Figs. 2A and 2C). All experimental alterations of parameters (e.g., character inclusion, transition/transversion (ti/tv) ratios, distance measures) are detailed under Results and Discussion.

RESULTS AND DISCUSSION

The nad2 Gene Structure

The mitochondrial nad2 gene (Fig. 1) carries four group II introns that are well conserved among angiosperm plants. We selected it as a novel marker gene for phylogenetic analyses in plants, because one of these intervening sequences (nad2i3) is the single intron that is positionally conserved in the chondriomes of seed plants (e.g., *Arabidopsis thaliana;* Unseld *et al.*, 1997) and the liverwort *Marchantia polymorpha* (Oda *et al.*, 1992). The nad2 gene is interrupted by the *trans*-splicing intron nad2i2 in seed plants (Binder *et al.*, 1992) but this intron is *cis*-arranged in ferns (Malek and Knoop, 1998). Amplification of nearly the entire nad2 coding region encompassing all known intron insertion sites turned out to be feasible in bryophytes (Fig. 1), where intron nad2i2 has so far been found to be generally absent. To our surprise we found that,

TABLE 1

Plant Material of This Study

Classification[a]		Database accessions[b]		
Order	Species	nad5	nad2	Voucher[c]
Sphagnales	*Sphagnum fallax*	*AJ001225*	A3299524	ULM:Muhle180597-2
Takakiales	*Takakia lepidozioides*	A3291553	A3299525	Qiu97126
Andreaeales	*Andreaea nivalis*	*AJ001226*	A3299526	ULM:Muhle140897-3
Polytrichales	*Atrichum undulatum*	*AJ001229*	A3299527	ULM:Muhle170897-2
	Pogonatum urnigerum	A3291554	A3299528	ULM:Muhle170997-15
Tetraphidales	*Tetraphis pellucida*	*AJ224855*	A3299529	ULM:Muhle151197-1
Buxbaumiales	*Diphyscium sessile*	*Z98972*	A3299530	ULM:Muhle191097-2
	Buxbaumia aphylla	A3291555	A3299531	ULM:Muhle070398-1
Bryidae				
Timmiales	*Timmia bavarica*	*Z98963*	A3299532	ULM:Muhle161197-1
Encalyptales	*Encalypta streptocarpa*	A3291556	A3299533	ULM:Muhle151197-5
Funariales	*Funaria hygrometrica*	*Z98959*	A3299534	ULM:Muhle291197-1
	Physcomitrella patens	*Z98960*	A3299535	SC-IC
Dicrananae				
Dicranales	*Dichodontium pellucidum*	A3291557	A3299536	ULM:Muhle191097-6
	Orthodicranum montanum	A3291558	A3299537	ULM:Muhle070198-1
Ditrichaceae	*Ceratodon purpureus*	*Z98955*	A3299538	ULM:Muhle191097-4
	Ditrichum cylindricum	A3291559	A3299539	ULM:Muhle281097-2
Leucobryaceae	*Leucobryum glaucum*	A3291560	A3299540	ULM:Muhle281097-6
Fissidentales	*Fissidens cristatus*	*Z98954*	A3299541	ULM:Muhle200497-3
Grimmiales	*Racomitrium lanuginosum*	A3291561	A3299542	ULM:Muhle070997-10
Pottiales	*Pottia truncata*	*Z98957*	A3299543	ULM:Muhle191097-3
	Tortula latifolia	A3291562	A3299544	ULM:Muhle070198-2
Cinclidotaceae	*Cinclidotus riparius*	A3291563	A3299545	ULM:Muhle200497-5
Schistostegales	*Schistostega pennata*	*AJ224856*	A3299546	ULM:Muhle221097-10
Hypnanae				
Bartramiales	*Bartramia halleriana*	*Z98961*	A3299547	ULM:Muhle140897-6
	Plagiopus oederi	*Z98962*	A3299548	ULM:Muhle140897-10
Aulacomniaceae	*Aulacomnium androgynum*	A3291564	A3299549	ULM:Muhle090897-1
(Eu)-Bryales	*Pohlia nutans*	A3291565	A3299550	ULM:Muhle090897-4
	Orthodontium lineare	A3291566	A3299551	ULM:Muhle090897-3
Mniaiceae	*Mnium hornum*	A3291567	A3299552	ULM:Muhle090897-2
Orthotrichales	*Ulota crispa*	A3291568	A3299553	ULM:Muhle200497-6
Hedwigiaceae	*Hedwigia ciliata*	*Z98966*	A3299554	ULM:Muhle070997-9
	Rhacocarpus purpurascens	*Z98967*	A3299555	IC
Isobryales (Leucodontales)				
Leucodontaceae	*Pterogonium gracile*	*Z98968*	A3299556	ULM:Muhle150597-1
Neckeraceae	*Homalia trichomanoides*	A3291569	A3299557	ULM:Muhle291197-3
Fontinalaceae	*Fontinalis antipyretica*	A3291570	A3299558	ULM:Muhle291197-5
Thamniaceae	*Thamnobryum alopecurum*	A3291571	A3299559	ULM:Muhle291197-4
Hypnales				
Brachytheciaceae	*Tomenthypnum nitens*	A3291572	A3299560	ULM:Muhle070198-4
Plagiotheciaceae	*Sharpiella seligeri*	A3291573	A3299561	ULM:Muhle191097-5
Amblystegiaceae	*Hygrohypnum ochraceum*	A3291574	A3299562	ULM:Muhle191197-4
	Scorpidium scorpioides	A3291575	A3299563	ULM:Muhle070997-6
Leskeaceae	*Leskea polycarpa*	A3291576	A3299564	ULM:Muhle231197-1
Lembophyllaceae	*Isothecium alopecurum*	A3291577	A3299565	ULM:Muhle291197-6

[a] Taxonomic designations are subject to discussion and controversy as discussed in the text; for a recent reference see Vitt *et al.*, 1998; for a comparison of taxonomic treatments among the Hypnanae see e.g., Hedenäs, 1994. Names of families are indicated only for those not naming the respective order.

[b] All accessions are new data obtained for this study, except for some of the nad5 data set established earlier (Beckert *et al.*, 1999), that are given in italics.

[c] Vouchers of materials collected by H. Muhle are held at the herbarium of the University of Ulm. SC-IC indicates sterile moss cultures maintained on agar; IC indicates DNA kindly provided by Dr. I. Capesius, Heidelberg. *Takakia lepidozioides* DNA was kindly provided by Dr. Y.-L. Qiu, Zürich.

instead of nad2i3, another one of the four nad2 introns known from seed plants, nad2i1, is positionally conserved as the single intervening sequence in all moss species so far investigated (Table 1), including the enigmatic genus *Takakia*. The conserved intron nad2i1 contributes sequence information of variable length

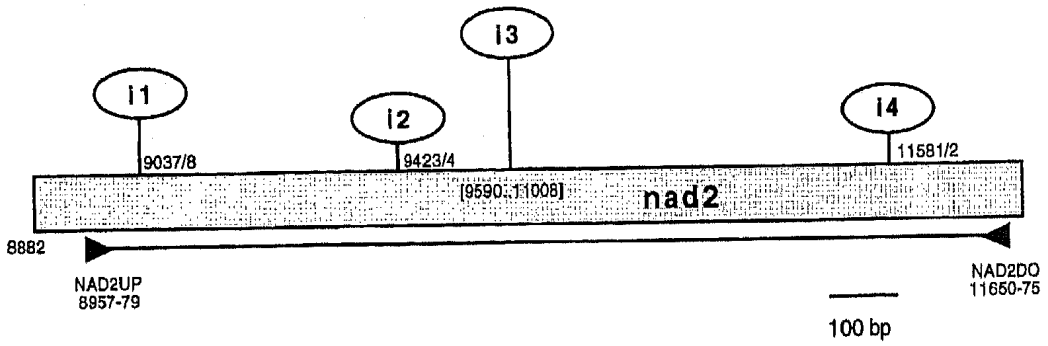


FIG. 1. The nad2 gene reading frame is interrupted by four group II introns well conserved in angiosperm species (i1, i2, i3, and i4). One of these, intron nad2i2, is *trans*-splicing (Binder *et al.*, 1992), causing a noncontinuous gene arrangement and precluding PCR amplification of the entire gene in seed plants. *Cis*-arranged counterparts to this intron have been identified in ferns and *Isoetes* (Malek and Knoop, 1998). Intron i3 of the nad2 gene is the single intron position conserved in the chondriomes of angiosperms and the liverwort *Marchantia polymorpha* and the only intervening sequence in nad2 of the liverwort. A coding region of 1252 bp + respective intron sequences is PCR-amplified (line with inverted arrowheads) from diverse mosses (Table 1) and intron i1 is conserved in mosses and seed plants. Numbering is according to the *Marchantia* chondriome database entry M68929.

(from 936 bp in *Diphyscium* to 1004 bp in *Sphagnum*), adding to the well-conserved amplified coding region of the nad2 gene (417 codons, 1252 bp). By differing in the group II intron distribution, the nad2 gene stands in contrast to the single group I intron of the nad5 gene, which was found conserved in mosses and liverworts (Beckert *et al.*, 1999). These observations on the one hand underline the positional stability of the mitochondrial introns on the class level, but may on the other hand also be taken as a recommendation for caution in the choice of introns whose sole presence or absence would be taken as an indicator for the phylogeny of the major land plant groups (Qiu *et al.*, 1998).

RNA Editing and Intron Indels

Best starting alignments of the full mitochondrial nad2 and nad5 sequences including introns were obtained with PILEUP of the GCG program package when gap and gap extension penalties were set to 2 and 0, respectively, and required only very minor manual modification. No length variations exist for the coding regions. Insertions unique to single species and small regions of ambiguous alignment in the introns were excluded for the phylogenetic analyses. Two shared indels of *Sphagnum* and *Takakia* in the nad5 intron were binarily recoded in the alignment matrix. RNA editing events, exchanging cytidine to uridine, to reconstitute conserved codon identities can be predicted from the moss nad2 alignment compared to the known seed plant cDNA sequences or the *M. polymorpha* sequence, as was similarly observed for the nad5 sequences of bryophytes (Steinhauser *et al.*, 1999). How RNA editing positions should be treated in phylogenetic analyses is not entirely clear, but it was suggested that DNA contains slightly more phylogenetic information than cDNA (Bowe and dePamphilis, 1996). Based on analysis of the nad5 gene among pteridophytes, we had recently suggested that editing sites most likely contribute some homoplasy to the data set (Vangerow *et al.*, 1999). However, as shown below and also observed in the analyses of nad5 alignments of pteridophytes (Vangerow *et al.*, 1999), neither the variation of alignment parameters nor the inclusion of all positions (RNA editing sites, ambiguous intron indels) affect the tree topologies significantly. The respective potential character exclusions can be found in the assumptions block of the corresponding NEXUS file of the alignment (available from the authors upon request) and their influence in phylogenetic analysis is discussed below.

Phylogenetic Analysis: Congruent Phylogenetic Information in nad2 and nad5

Phylogenetic trees based on the maximum-likelihood approach (for the default HKY model and ti/tv ratio of 2) are shown in Fig. 2 for the nad2 gene (A), the nad5 gene (C), and the fused data set (B). Total alignment length and numbers of excluded (either affected by RNA editing or within dubiously aligned intron regions) and parsimony-informative (pi) positions are given below each tree. Bootstrap values derived from heuristic parsimony searches are indicated where they exceed 70. Both genes have the same ratio of parsimony-informative positions vs total included characters (0.18), but the nad2 alignment contributes slightly more characters. By and large, both genes determine the same groupings with confidence and thus appear to supply information on similar taxonomic levels. Notably, no single node determined with confidence by one gene is contradicted by an alternative topology determined by the other. When the number of reliably determined nodes with bootstrap confidence of varying levels (>70/>80/>90) is listed, an increase in the number of nodes with bootstrap support at all three levels is observed upon combining the data sets (Fig. 2). Most significant is the near doubling of the number of nodes

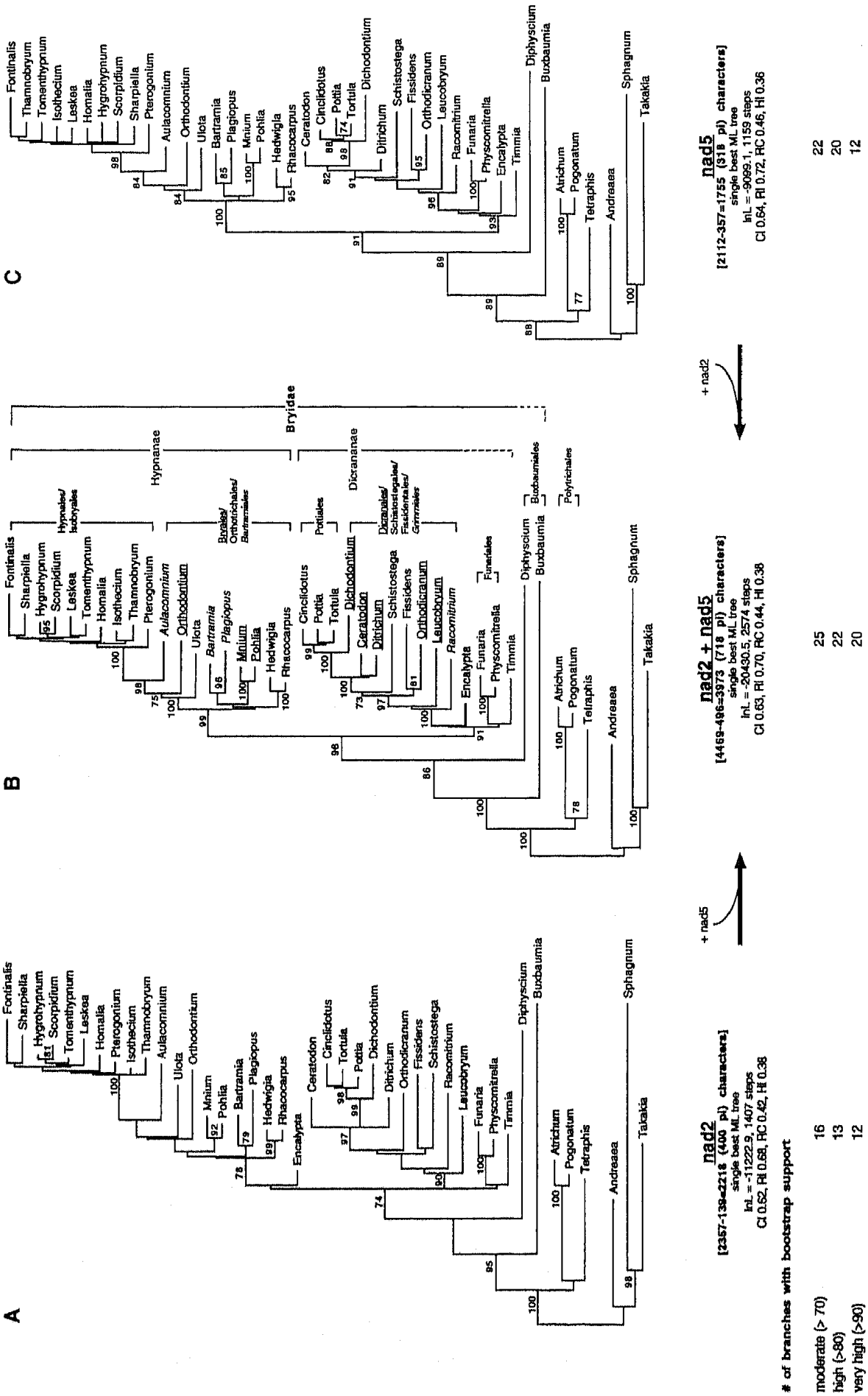


FIG. 2. Phylogenetic trees based on the nad2 (A), the nad5 (C), and the joined nad2 + nad5 sequence alignment (B) obtained with the maximum-likelihood (ML) approach. The numbers of characters (complete alignment – excluded characters = included characters, pi, parsimony-informative sites) are indicated. Bootstrap value percentages exceeding 70 derived from parsimony analyses (500 replicates with TBR branch-swapping in B, 5000 replicates without branch-swapping in A and C) are given next to the respective nodes in the ML topologies. The numbers of nodes determined with bootstrap reliabilities exceeding 70, 80, and 90, respectively, are given below each tree for comparison. A systematic classification is given in B with typographical distinction (*italics*, underlining) of some ordinal assignments to indicate lacking monophyly in Dicranales, Bryales, and Bartramiales. For lack of appropriate outgroups simultaneously containing introns nad5i2 and nad2i1, the trees are unrooted.

with very high bootstrap support (>90) in comparison to the single-gene approaches (20 vs 12). The combined alignment thus results in 50% (20/40) of all nodes in the tree determined with very high bootstrap reliability, a number that increases to 61% (19/31) when the terminal Hypnales/Isobryales cluster is ignored, which remains largely unresolved by the well-conserved mitochondrial sequences.

Phylogenetic Analysis: Congruent Results from Different Methods

The results of different tree construction methods (maximum-parsimony, maximum-likelihood, and distance matrix approaches) for the joint nad2 + nad5 alignment reduced for ambiguous positions are shown side-by-side in Fig. 3. The respective tree statistics are given below each tree and bootstrap values from 500 replicates exceeding 70 are indicated for the MP and DM analyses (Figs. 3A and 3C), both for the alignment reduced for ambiguous positions and for the complete nonreduced alignment (separated by slashes).

The ML tree calculated for a ti/tv ratio of 3 is shown in Fig. 3B for comparison to the one in Fig. 2B (ML searches with empirical ti/tv ratios were aborted after 50 h, at which time trees showing only minor topological variations in the terminal Hypnales/Isobryales cluster were identified).

Four equally parsimonious trees of 2571 steps, likewise with minor topological differences in the terminal Hypnales/Isobryales group only, were obtained in the heuristic search, and the tree with highest likelihood is shown in Fig. 3A. All distance measures available in PAUP were alternatively used for DM analyses. The shortest trees (2579 steps) were obtained with the Tajima–Nei/Tamura–Nei distances (Fig. 3C), and trees of 2582 steps were obtained with other distance measures.

Topological differences between all tree types were rare and in all cases restricted to branches without significant bootstrap support (Fig. 3). The DM analysis shows support for two groupings not identified (but also not rejected with statistically supported alternative topologies) by the two other tree-building approaches: an Encalypta/Timmia/Funariales group and a Hedwigiaceae/Bartramiaceae group.

To test for the influence of ambiguous positions (RNA editing sites and highly variable intron regions), the full alignment (no characters excluded) was also used for tree construction. Differences are essentially restricted to minor variations of bootstrap support. The single topological change of a node supported with a bootstrap value exceeding 70 upon inclusion of all characters is the loss of a *Tetraphis*/Polytrichales linkage in the parsimony analysis with *Tetraphis* then appearing basal to Bryidae (however, without bootstrap support). A *Tomenthypnum*/Amblystegiaceae linkage is suggested by MP and DM bootstrap analyses and a *Homalia/Isothecium/Thamnobryum* group receives moderate bootstrap support in the DM analysis only after including all characters.

Phylogenetic Implications for a Cladistic System of Mosses: Basal Lineages

Earlier classifications of *Takakia* had suggested its inclusion among liverworts, e.g., in the order Calobryales (Schuster, 1984), due to gametophytic similarities, such as the absence of rhizoids and food-conducting cells, lateral origin of branches, or the development of water-conducting cells. The late discovery of its moss-like sporophytes (Smith and Davison, 1993) had suggested its reclassification as a moss. In an elaborate cladistic evaluation of a large morphological data set (Garbary and Renzaglia, 1998), the gametophytic similarities now appear as plesiomorphies. Molecular data such as 18S rRNA sequences (Hedderson *et al.*, 1998) or the presence of mitochondrial introns generally absent in liverworts (Y. L. Qiu, V. Knoop, and B. Crandall-Stotler, unpublished observations) likewise cannot support an inclusion of *Takakia* among the liverworts. The presence of i1 as the single intron in the nad2 gene of *Takakia* is a character exclusively shared with all mosses investigated and supports the suggested reclassification of *Takakia* into the mosses.

Placing the root in the moss trees with confidence, however, is hindered by the mutual absence of either the nad2i1 or the nad5 group I intron in potential outgroups. *Sphagnum* appears basal-most in a double-outgroup approach with the liverwort *Marchantia* and the lycopod *Lycopodium*, which allows inclusion of both intron sequences for phylogenetic information (the nad5 group I intron shared with the liverwort and nad2i1 shared with the lycopod, not shown). However, bootstrap support for *Takakia* branching after *Sphagnum* is as yet lacking, and the inclusion of additional phylogenetic information from other loci is needed to settle the question about placing the root on either branch.

The orders Andreaeales, Polytrichales, and Tetraphidales branch before a monophyletic group of derived (arthrodontous) mosses, possibly to be designated a subclass Bryidae (Fig. 2B). This observation supports systems that place at least Sphagnales and Andreaeales in subclasses (Vitt, 1984) or even classes (Fukarek *et al.*, 1992) of their own and likewise justifies the definition of subclasses Tetraphididae and Polytrichidae set apart from the Bryidae (Walther, 1983). A junction of the latter two (Fig. 2B) receives high bootstrap support only from the DM analysis of the reduced alignment (Fig. 3). A subclass Bryidae may be defined with or without the Buxbaumiales, which are clearly placed as an intervening (possibly paraphyletic) taxon at the root of this subclass, with *Diphyscium* being more closely related to (other) Bryidae than *Buxbaumia*. A novel ordinal classification of mosses by Vitt

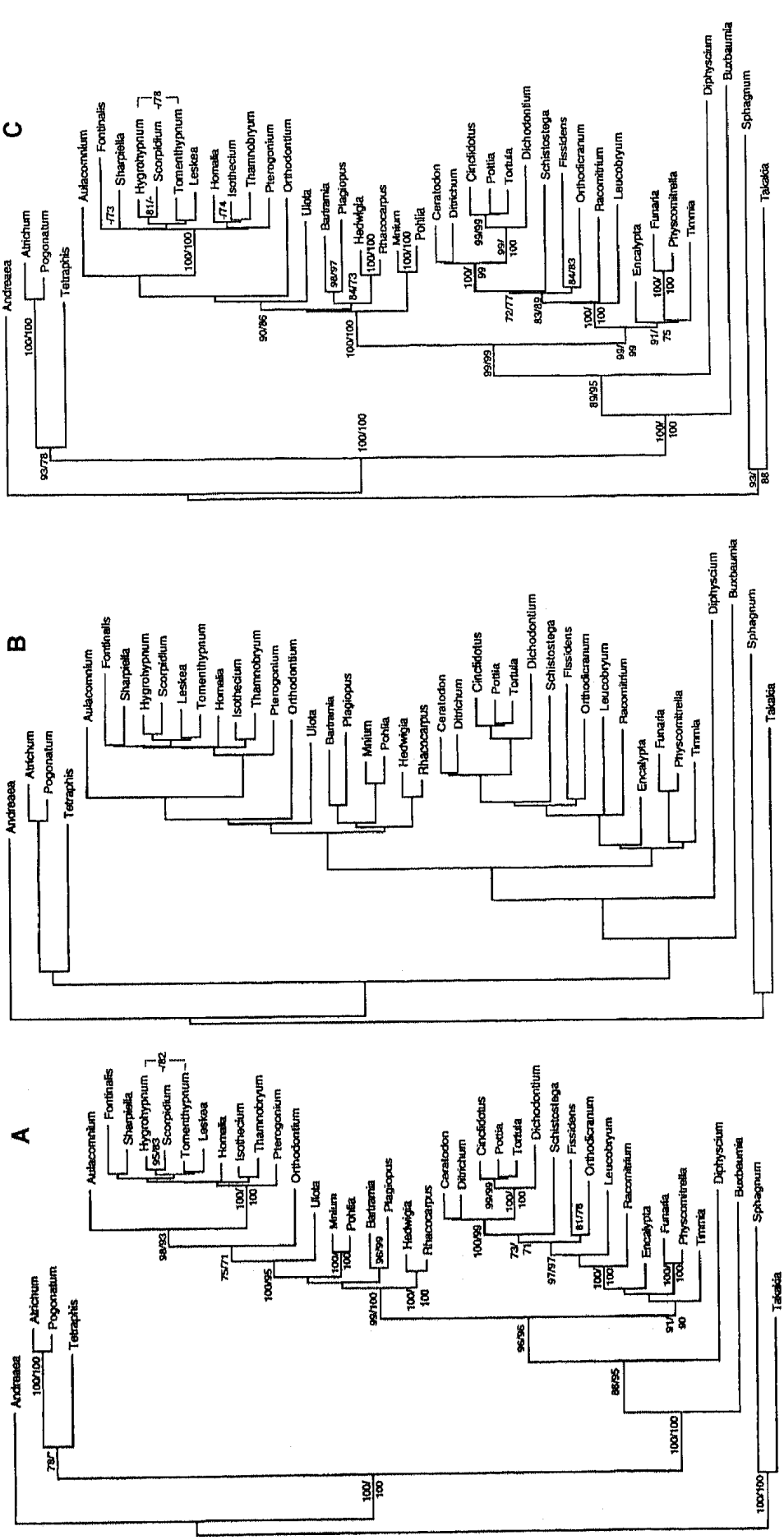


FIG. 3. Results of tree building approaches are shown for the maximum-parsimony (A), maximum-likelihood (B), and distance matrix (C) methods for comparison. The trees shown are based on the same joint nad2 + nad5 alignment (4469 positions reduced by 496 ambiguous positions) as in Fig. 2B. Bootstrap values exceeding 70 obtained from 500 replicates in heuristic searches with TBR branch-swapping are indicated in A and C both for the reduced data set (before the slash) and for the complete nonreduced alignment (behind the slash) for comparison. Tree statistics are given below each tree, likewise both for the reduced and for the complete alignment, respectively (highest likelihood tree only for the parsimony analyses). The single topological change of a node supported with a bootstrap value exceeding 70 with one data set (reduced alignment) not identified with the other (full alignment) is marked with an asterisk in the parsimony tree (A).

et al. (1998) was based on a joint cladistic evaluation of peristome types and molecular data and suggested the definition of a class Polytrichopsida including Tetraphidales, Buxbaumiales, and Diphysciales next to Bryopsida (i.e., Bryidae in the sense used here). Although clearly not a monophyletic group, Polytrichopsida in that sense is recognizable as a paraphyletic assembly between the earliest lineages (*Sphagnum, Andreaea,* and *Takakia*) and the arthrodontous mosses in the Bryidae.

Two monophyletic subgroups become immediately apparent in the Bryidae, and they remind one of the concepts "Hypnanae" and "Dicranane" *sensu* Frahm and Frey (1992) or a similar dichotomy of suborders according to Vitt (1984), although modifications would be required for both concepts (Fig. 2B) to match the current observations. The Hypnanae includes only species with diplolepidous-alternate peristomes, the Dicrananae group contains species with haplolepidous and diplolepidous-opposite peristomes.

The "Dicrananae" Cluster

The orders Timmiales, Encalyptales, and Funariales, although having diplolepidous peristomes, show slightly stronger affinity to the Dicrananae group, with reasonable bootstrap support coming mainly from nad5 (Fig. 2). The inclusion of Encalyptales among Dicrananae is in accordance with this concept *sensu* Frahm and Frey (1992), but neither the junction of Timmiales and Bartramiales in a superorder Bartramianae (Frahm and Frey, 1992) nor the inclusion of Timmiaceae among (Eu-)Bryales (or Bryinae acording to Vitt, 1984) receives support from the mitochondrial analyses presented here, with the latter groups clearly to be included in the alternative superorder Hypnanae. Interestingly, a very basal placement of *Timmia* is also found with an alternative molecular approach (Cox and Hedderson, 1999) and the diverging taxonomic treatments of *Timmia* may reflect the difficulties of reasonable classification based on morphological characters alone. The apparent former misplacement of *Timmia,* which is now clearly considered a much more basal branching genus, is addressed by Vitt *et al.* (1998), suggesting the reevaluation of morphological markers.

The diplolepidous flanged peristomes of Buxbaumiales and Encalyptales may have been the starting point for the evolution of the Bryidae peristomes into the haplolepidous forms of the Dicrananae, possibly via the diplolepidous-opposite (=*Funaria*-type) on the one hand and into the diplolepidous-alternate (possibly via diplolepidous orthotrichaceous) peristomes of the Hypnanae on the other (Vitt, 1984, see also Vitt *et al.*, 1998 and Cox and Hedderson, 1999). However, the definite positions of *Timmia, Encalypta,* and the Funariales need to be ultimately resolved.

The orders Grimmiales, Fissidentales, Dicranales, Pottiales, and Schistostegales are joined in the Dicrananae group (Fig. 2B). This is in accordance with most systematic assessments (e.g., Vitt *et al.*, 1998; Walther, 1983), but not with the inclusion of *Schistostega* (together with *Timmia*) in the Bryinae (=Eubryales) in an older concept, *sensu* Vitt (1984).

Both the concept of the genera-rich family Dicranaceae as monophyletic and the ordinal definition of Dicranales including Ditrichaceae and Leucobryaceae requires reconsideration. *Dichodontium* is placed in a monophyletic group together with the (itself monophyletic) Pottiales and then joined by the sister family Ditrichaceae, but to the exclusion of *Orthodicranum* and *Leucobryum.* The branching pattern of *Orthodicranum* relative to *Schistostega* and *Fissidens* is not fully resolved but the entire group of haplolepidous mosses with *Racomitrium* and *Leucobryum* at their base receives maximal bootstrap support. The inclusion of Fissidentales with Dicranales according to Vitt *et al.* (1998) is thus in accordance with our trees, but this inclusion should likewise contain the Pottiales.

The "Hypnanae" Cluster

Among the "Hypnanae" a group of 10 most derived species of the orders Hypnales and Isobryales is set apart very confidently. This group (encompassing Hypninae (=Hypnobryales) and Leucodontinae (=Isobryales) *sensu* Vitt) may coincide with the occurrence of a pleurocarpous growth habit in the restricted sense of LaFarge-England (1996), as also commented on by Cox and Hedderson (1999); however, a larger taxon sampling is needed to address this point. The family level Amblystegiaceae is recognized with the close linkage of *Hygrohypnum* and *Scorpidium;* however, further branching details in this group remain unresolved by both mitochondrial genes. This Hypnales/Isobryales cluster includes the families Neckeraceae, Fontinalaceae, and Leucodontaceae and thus rejects their inclusion in an order "Neckerales" *sensu* Frahm and Frey (1992) together with species of Orthotrichales. *Ulota* (Orthotrichaceae), likewise a species of the concept "Neckerales" *sensu* Frahm and Frey (1992), is branching earlier, nested among members of the (Eu-)Bryales (including Bartramiaceae). Clearly, the more widely accepted concept of distinguishing Isobryales (Leucodontales) and Orthotrichales is much better reflected in the mitochondrial phylogenies than the fusion of the two groups in a potential order Neckerales. The transitory status of these taxa may be morphologically mirrored by the transitions from acrocarpous to pleurocarpous growth forms in these groups (Hedenäs, 1994). The family Bryaceae *sensu stricto* (*Orthodontium* and *Pohlia*) is not identified as monophyletic in the nad gene trees. Interestingly, a *Pohlia/Mnium* linkage to the exclusion of *Orthodontium* was likewise identified when sequences from the nucleus and chloroplast genomes were investigated for a larger taxon set focusing on ciliate arthrodontous mosses (Cox and Hedderson,

1999). The concept of a subclass Bryidae comprising Bryales, Leucodontales, and Hypnales as suggested recently (Vitt *et al.*, 1998) finds no support. Hypnales and Leucodontales (Isobryales) are clearly linked to the exclusion of Bryales, which appear nonmonophyletic in an as yet incompletely resolved branching order, together with Orthotrichales (and Bartramiales) basal to the Hypnales–Isobryales group.

CONCLUDING REMARKS

The mitochondrial nad2 gene is established as a novel and informative phylogenetic marker locus in early land plants. Focusing on an extensive taxon sampling of mosses, we find congruent and complementing results in our phylogenetic trees based on nad2 and the nad5 locus. No single contradiction is observed in the independent phylogenetic trees and combination of the two data sets results in increased resolution and statistical confidence supported by both types of organellar intron sequences, group I and group II, in the two genes. Both genes appear most suitable for determining ordinal relationships within a class of phylogenetically old plants. Although both genes (with exceptions) will fail to resolve relationships at or below family level, they appear helpful in testing monophyletic concepts for some families (as, e.g., shown for the genera-rich Dicranaceae). At least on these taxonomic levels, mitochondrial sequences (Beckert *et al.*, 1999 and this study) thus appear to be a rich source of information, most notably compared to other molecular markers such as the chloroplast rbcL, rps4, and trn loci or nuclear 18S rRNA sequences.

Recent studies based on chloroplast ITS sequences (Samigullin *et al.*, 1998) do not as yet give a complete picture for comparison due to some taxon undersampling, but may eventually be shown to be a good data set to complement the mitochondrial data for establishing phylogenetic relationships of orders and above. As an example, a Dicrananae–Hypnanae dichotomy in the sense of this study is recognizable with a smaller taxon set in the tree based on chloroplast ITS sequences, although the placement of Buxbaumiales remained unclear in that study. None of the reliably determined branches among the mosses in the ITS tree (i.e., those associated with bootstrap values exceeding 80) contradict any of the statements of this study. Significantly, *Schistostega* is also placed in the "Dicrananae" complex in the sense reported here. The placement of Schistostegaceae as a family of Eubryales (=Bryinae in Vitt, 1984) is thus questioned by two independent molecular approaches.

From an overall perspective, the cladogram suggested by Vitt (1984) is well reflected in the nad gene-based phylogenies. The placement of Buxbaumiales at the root of arthrodontous mosses *in toto* (Bryidae) instead of inclusion in the Hypnanae cluster is now adequately considered in a novel system (Vitt *et al.*, 1998) and the accompanying separation of *Diphyscium* in an order of its own is likewise comprehensible from the novel mitochondrial trees.

The combination of particularly high sequence conservation combined with sufficient sequence variability in the positionally stable organellar introns in plant chondriomes appears of promise for ongoing studies to elucidate phylogenetic relationships of ancestral lineages. We will thus continue to explore novel molecular markers from plant mitochondrial genomes, hoping to resolve additional issues of early land plant evolution.

ACKNOWLEDGMENTS

We are very grateful to The Deutsche Forschungsgemeinschaft for Grant Kn411-2 to V.K., Profs. Drs. Ingrid Capesius, Heidelberg and Yin-Long Qiu, Zürich for very kind exchange of materials and information, Dr. D. Swofford, Suitland, Maryland, for making beta test versions of PAUP* available, Prof. Dr. Axel Brennicke for continuous support, Enya and Blue Tonic World for entertainment during manuscript preparation, and Franz Fritsche for greenhouse work.

REFERENCES

Beckert, S., Steinhauser, S., Muhle, H., and Knoop, V. (1999). A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial nad5 gene. *Pl. Syst. Evol.* **218**: 179–192.

Binder, S., Marchfelder, A., Brennicke, A., and Wissinger, B. (1992). RNA editing in trans-splicing intron sequences of nad2 mRNAs in *Oenothera* mitochondria. *J. Biol. Chem.* **267**: 7615–7623.

Bowe, L. M., and dePamphilis, C. W. (1996). Effects of RNA editing and gene processing on phylogenetic reconstruction. *Mol. Biol. Evol.* **13**: 1159–1166.

Capesius, I., and Bopp, M. (1997). New classification of liverworts based on molecular and morphological data. *Pl. Syst. Evol.* **207**: 87–97.

Capesius, I., and Stech, M. (1997). Molecular relationships within mosses based on 18S rRNA gene sequences. *Nova Hedwigia* **64**: 525–533.

Cox, C. J., and Hedderson, T. A. J. (1999). Phylogenetic relationships among the ciliate arthrodontous mosses: Evidence from chloroplast and nuclear DNA sequences. *Pl. Syst. Evol.* **215**: 119–139.

Devereux, R., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**: 387–395.

Doyle, J. J., and Doyle, J. L. (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13–15.

Edwards, D., Duckett, J. G., and Richardson, J. B. (1995). Hepatic characters in the earliest land plants. *Nature* **374**: 635–636.

Frahm, J.-P., and Frey, W. (1992). "Moosflora," UTB, Stuttgart.

Fukarek, F., Schultze-Motel, J., and Siegel, M. (1992). "Moose, Farne, Nacktsamer," Urania, Leipzig.

Garbary, D. J., and Renzaglia, K. S. (1998). Bryophyte phylogeny and the evolution of land plants: Evidence from development and ultrastructure. *In* "Bryology for the Twenty-first Century" (J. W. Bates, N. W. Ashton, and J. D. Duckett, Eds.), pp. 45–63. Maney and Br. Bryol. Soc., Leeds.

Goffinet, B., Bayer, R. J., and Vitt, D. H. (1998). Circumscription and phylogeny of the Orthotrichales (Bryopsida) inferred from rbcL sequence analyses. *Am. J. Bot.* **85**: 1324–1337.

Hedderson, T. A., Chapman, R., and Cox, C. J. (1998). Bryophytes

and the origins and diversification of land plants: New evidence from molecules. *In* "Bryology for the Twenty-first Century" (J. W. Bates, N. W. Ashton, and J. D. Duckett, Eds.), pp. 65–77. Maney and Br. Bryol. Soc., Leeds.

Hedenäs, L. (1994). The basal pleurocarpous diplolepidous mosses—A cladistic approach. *Bryologist* **97**: 225–243.

Hyvönen, J., Hedderson, T. A., Smith Merrill, G. L., Gibbings, J. G., and Koskinen, S. (1998). On phylogeny of the Polytrichales. *Bryologist* **101**: 489–504.

Kenrick, P., and Crane, P. R. (1997). The origin and early evolution of plants on land. *Nature* **389**: 33–39.

La Farge-England, C. (1996). Growth form, branching pattern and perichaetial position in mosses: Cladocarpy and pleurocarpy redefined. *Bryologist* **99**: 170–186.

Malek, O., Lättig, K., Hiesel, R., Brennicke, A., and Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO J.* **15**: 1403–1411.

Malek, O., and Knoop, V. (1998). Trans-splicing group II introns in plant mitochondria: The complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA* **4**: 1599–1609.

Mishler, B. D., and Churchill, S. P. (1984). A cladistic approach to the phylogeny of the 'bryophytes'. *Brittonia* **36**: 406–424.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., Kohchi, T., and Ohyama, K. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *J. Mol. Biol.* **223**: 1–7.

Qiu, Y. L., Cho, Y., Cox, J. C., and Palmer, J. D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394**: 671–674.

Qiu, Y. L., and Palmer, J. D. (1999). Phylogeny of early land plants: Insights from genes and genomes. *Trends Plant Sci.* **4**: 26–30.

Samigullin, T. H., Valiejo-Roman, K. M., Troitsky, A. V., Bobrova, V. K., Filin, V. R., Martin, W., and Antonov, A. S. (1998). Sequences of rDNA internal transcribed spacers from the chloroplast DNA of 26 bryophytes: Properties and phylogenetic utility. *FEBS Lett.* **422**: 47–51.

Schuster, R. M. (1984). Evolution, phylogeny and classification of the Hepaticae. *In* "New Manual of Bryology" (R. M. Schuster, Ed.), Vol. 2, pp. 892–1070. Hattori Botanical Laboratory, Nichinan.

Smith, D. K., and Davison, P. G. (1993). Antheridia and sporophytes in Takakia ceratophylla (Mitt.) Grolle: Evidence for reclassification among the mosses. *J. Hatt. Bot. Lab.* **73**: 263–271.

Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing—Extreme in hornworts and dividing the liverworts? *J. Mol. Evol.* **48**: 303–312.

Swofford, D. L. (1999). PAUP*, Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer, Sunderland, MA.

Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* **15**: 57–61.

Vangerow, S., Teerkorn, T., and Knoop, V. (1999). Phylogenetic information in the mitochondrial nad5 gene of pteridophytes: RNA editing and intron sequences. *Plant Biol.* **1**: 235–243.

Vitt, D. H. (1984). Classification of the Bryopsida. *In* "New Manual of Bryology" (R. M. Schuster, Ed.), Vol. 2, pp. 696–759. Hattori Botanical Laboratory, Nichinan.

Vitt, D. H., Goffinet, B., and Hedderson, T. (1998). The ordinal classification of the mosses: Questions and answers for the 1990s. *In* "Bryology for the Twenty-first Century" (J. W. Bates, N. W. Ashton, and J. D. Duckett, Eds.), pp. 113–123. Maney and Br. Bryol. Soc., Leeds.

Walther, K. (1983). "Syllabus der Pflanzenfamilien." Kapitel V.2, "Bryophytina. Laubmoose," Borntraeger, Berlin.

Noncoding plastid *trnT-trnF* sequences reveal a well resolved phylogeny of basal angiosperms

T. BORSCH,* K. W. HILU,† D. QUANDT,* V. WILDE,‡ C. NEINHUIS* & W. BARTHLOTT*

**Botanisches Institut und Botanischer Garten, Friedrich-Wilhelms-Universität Bonn, Bonn, Germany*
† Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
‡ Paläobotanik, Forschungsinstitut Senckenberg, Frankfurt, Germany

Keywords:

angiosperms;
chloroplast genome;
group I intron;
molecular evolution;
noncoding DNA;
phylogeny;
secondary structure;
trnT-trnF.

Abstract

Recent contributions from DNA sequences have revolutionized our concept of systematic relationships in angiosperms. However, parts of the angiosperm tree remain unclear. Previous studies have been based on coding or rDNA regions of relatively conserved genes. A phylogeny for basal angiosperms based on noncoding, fast-evolving sequences of the chloroplast genome region *trnT-trnF* is presented. The recognition of simple direct repeats allowed a robust alignment. Mutational hot spots appear to be confined to certain sectors, as in two stem-loop regions of the *trnL* intron secondary structure. Our highly resolved and well-supported phylogeny depicts the New Caledonian *Amborella* as the sister to all other angiosperms, followed by Nymphaeaceae and an *Austrobaileya–Illicium–Schisandra* clade. *Ceratophyllum* is substantiated as a close relative of monocots, as is a monophyletic eumagnoliid clade consisting of Piperales plus Winterales sister to Laurales plus Magnoliales. Possible reasons for the striking congruence between the *trnT-trnF* based phylogeny and phylogenies generated from combined multi-gene, multi-genome data are discussed.

Introduction

Flowering plants (angiosperms) are the largest and most diverse group in the plant kingdom. They have undergone an extensive radiation since the Cretaceous, and at present comprise approximately 270 000 species of remarkably diverse biological forms, spanning and dominating most habitats on earth and providing the vast majority of our food crops. Connected to this immense diversity and importance has been the need for understanding their origin and evolution. Recent contributions based on DNA sequences from genes belonging to the three plant genomes (nuclear, chloroplast and mitochondrial) analysed individually and in combinations have provided new insights into flowering plant phylogeny and radically changed our concept of their systematic relationships (Chase *et al.*, 1993; Soltis *et al.*, 1997, 1999a, 2000; Mathews & Donoghue, 1999, 2000; Qiu *et al.*, 1999, 2000; Barkman *et al.*, 2000; Graham & Olmstead, 2000; Savolainen *et al.*, 2000). Although many new lineages have recently been identified, there are still disputable clades in the global angiosperm tree because of incongruence among phylogenies, poor branch resolution or lack of convincing statistical support. As a consequence, additional areas of evidence from new genomic regions or other sources, like the fossil record, remain crucial.

The recent surge in applying molecular techniques in systematic biology has also raised important issues relevant to understanding patterns of molecular evolution of genes and genomes and their implications for organismal phylogenies. The issue of incongruence among phylogenies inferred from different genes underscores a central problem in phylogenetic studies, namely that of segregating gene trees that reflect gene phylogenies from organismal trees that depict the evolutionary history of the organisms (e.g. Doyle, 1992; Moritz &

Correspondence: Thomas Borsch, Botanisches Institut und Botanischer Garten, Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany.
Tel.: +49-228-73-2681; fax: +49-228-73-3120;
e-mail: borsch@uni-bonn.de

Hillis, 1996). Differences between gene trees and organismal trees can be caused either by intrinsic biases of the genes, such as functional constraints resulting in heterogeneity in rates and modes of substitution, or by extrinsic factors such as deep coalescence, gene duplication and horizontal gene transfer (Doyle, 1992; Swofford *et al.*, 1996). Combining data sets have helped to resolve most problems, which arose from single gene analyses of angiosperms (Qiu *et al.*, 1999, 2000; Soltis *et al.*, 1999a, 2000). Nevertheless, in some cases, like the analysis of combined *rbcL* and *atpB* data sets (Savolainen *et al.*, 2000), potential dominance of information from one gene could generate evolutionary noise that obscures to varying degrees the true organismal phylogeny. For example, parsimony analysis of *atpB* sequences alone resolves *Ceratophyllum* as sister to *Acorus*, and the two as sister to all other monocots, whereas the combined analysis of *atpB* and *rbcL* shows *Ceratophyllum* as sister to all other angiosperms, reflecting its position in the *rbcL* analysis alone. Combining data from different genes may also cause a decrease in resolution in parts of the phylogeny and create weak support of some clades when there is incongruence between original data sets. These shortcomings may only be overcome by sampling high numbers of independently varying characters (e.g. Graham & Olmstead, 2000; Qiu *et al.*, 2000).

Coding regions of rather conserved genes are typically used in reconstructing deep-level phylogenies, such as relationships among major angiosperm lineages. This practice is based on the premise that the low rates of substitution characteristic of those genes reduce incidents of multiple hits that could obscure historical signal, keeping levels of homoplasy at a minimum. In addition, relative ease of sequence alignment makes homology assessment within so-called conserved genes very straightforward. In contrast, noncoding regions have been deemed unsuitable for resolving such phylogenies because of high mutational rates. Noncoding regions, on the other hand, being functionally less constrained than coding regions (e.g. Morton & Clegg, 1993; Clegg *et al.*, 1994) may render fixation of a greater number of substitutions during cladogenesis closer to a stochastic process (i.e. selectively closer to neutral mutations; Jukes & King, 1971; Kimura, 1983). Consequently, the mutations would not to a larger extent be biased by and reflect the functional evolution of the gene.

Our application of *trnT-trnF* sequences to a phylogenetic analysis of the waterlily genus *Nymphaea* (Borsch, 2000) demonstrated that alignment of outgroup sequences beyond the Nymphaeaceae sensu APG (1998; corresponds to Nymphaeales as comprising the genera *Brasenia*, *Cabomba*, *Nuphar*, *Barclaya*, *Ondinea*, *Victoria*, *Euryale*, *Nymphaea*) is possible and led us to employ the region in investigating relationships among basal angiosperms. The *trnT-trnF* region is located in the large single-copy region of the chloroplast genome, approximately 8 kb downstream of *rbcL*. Three highly conserved transfer RNA genes [tRNA genes for threonine (UGU), leucine (UAA) and phenylalanine (GAA)] are found in tandem, separated by spacers of several hundred base pairs (bp) (Fig. 1). The high variability of the two spacers and the intron in *trnL* have led to the wide use of *trnT-trnF* sequences in addressing relationships at the species and genus levels (e.g. Taberlet *et al.*, 1991; Van Ham *et al.*, 1994; Sang *et al.*, 1997; Small *et al.*, 1998; Bakker *et al.*, 2000). Moreover, the region was quite informative in phylogenetic studies of families like Asteraceae (Bayer & Starr, 1998), Arecaceae (Asmussen & Chase, 2001) and Rhamnaceae (Richardson *et al.*, 2000) and orders like Laurales (Renner, 1999) and Magnoliales (Sauquet *et al.*, in press).

In the present study, the entire *trnT-trnF* region was sequenced from 32 families representing most lineages of basal angiosperms. The confinement of the extreme variability to certain mutational hot spots and the presence of a majority of length mutational events in simple sequence repeats (SSRs) of 3–5 bp facilitated the alignment. Mutationally flexible stretches of sequence in the *trnL* intron correspond to two stem-loop regions in P8 of the proposed RNA secondary structure. This study presents a phylogenetic tree for basal angiosperms based on *trnT-trnF* sequence data that is largely congruent with multi-gene, multi-genome studies and demonstrates that fast-evolving, noncoding sequences do not

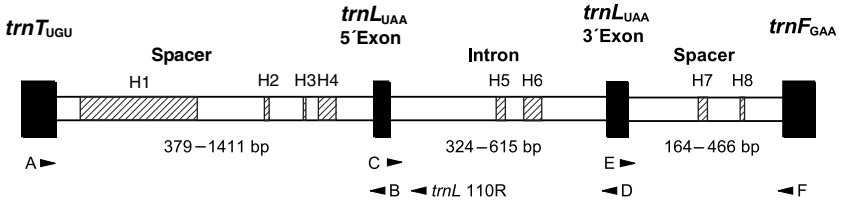


Fig. 1 Structure of the *trnT-trnF* region in basal angiosperms and gymnospermous outgroups based on the data set used in the present study. tRNA genes (*trnT* and *trnF* are each 73 bp long) and exons (*trnL*-5′ is 35 bp and 3′ is 50 bp) are represented by black boxes. The spacers and the intron are illustrated by an empty bar with mutational hot spots in grey. Proportions reflect average sequence length of the sequenced taxa. Mean sequence lengths (and standard deviations, SD) in bp are 298 in H1 (SD = 222), 12 in H2 (SD = 2), four in H3 (SD = 2), 40 in H4 (SD = 7), 12 in H5 (SD = 7), 30 in H6 (SD = 31), 19 in H7 (SD = 21), and 7 in H8 (SD = 3). Minimum and maximum sizes of the spacers and the intron among the taxa sequenced are indicated below the bar. Positions of primers are marked by arrows.

necessarily show total saturation when applied to deep-level phylogenetic questions in angiosperms, but on the contrary, yield a phylogeny with many of the nodes receiving statistical support. This empirical analysis therefore is in line with expectations drawn from recent simulation studies (Hillis, 1998) in that higher evolutionary rates may be beneficial for reconstructing correct phylogenies.

Materials and methods

Material

Sequences from the *trnT-trnF* region were obtained for 38 angiosperms representing 28 families and three gymnosperms. The species, their respective families and the sources of material are listed in Table 1. The *Pinus*

Table 1 Taxa used in the study, their respective families, source of material, location of voucher specimens, and GenBank numbers of deposited sequences.

Genus/species	Family	Garden/field origin	Voucher	GenBank number
Acorus gramineus L.	Acoraceae	Bonn Bot. Gard.	Borsch 3458 (BONN)	AY145336
Aextoxicon punctatum Ruiz & Pav.	Aextoxicaceae	Bonn Bot. Gard.	Borsch 3459 (BONN)	AY145362
Amborella trichopoda Baill.	Amborellaceae	University of California, Sta. Catarina Bot. Gard.	Borsch 3480 (VPI)	AY145324
Annona muricata L.	Annonaceae	Bonn Bot. Gard.	Borsch 3460 (BONN)	AY145352
Asimina triloba Dun.	Annonaceae	Bonn Bot. Gard.	Borsch 3461 (BONN)	AY145353
Orontium aquaticum L.	Araceae	Bonn Bot. Gard.	Borsch 3457 (BONN)	AY145338
Araucaria araucana C. Koch	Araucariaceae	Bonn Bot. Gard.	Borsch 3462 (BONN)	AY145321
Nypa fruticans Wurmb.	Arecaceae	Bonn Bot. Gard.	Borsch 3463 (BONN)	AY145339
Aristolochia pistolochia L.	Aristolochiaceae	France, Herault	Borsch 3257 (FR)	AY145341
Saruma henryi Oliv.	Aristolochiaceae	Bonn Bot. Gard.	Borsch 3456 (BONN)	AY145340
Austrobaileya scandens C. White	Austrobaileyaceae	Bonn Bot. Gard.	Borsch 3464 (BONN)	AY145326
Buxus sempervirens L.	Buxaceae	Bonn Bot. Gard.	Borsch 3465 (BONN)	AY145357
Brasenia schreberi Gmelin	Cabombaceae	USA, Virginia	Borsch & Wieboldt 3298 (VPI. FR)	AY145329
Cabomba caroliniana Grey	Cabombaceae	USA, Virginia	Ludwig, J.C. s.n. (VPI)	AY145328
Calycanthus floridus L. var. *laevigatus* (Willd.) T. & G.	Calycanthaceae	Bonn Bot. Gard.	Borsch 3455 (BONN)	AY145349
Canella winterana Gaertn.	Canellaceae	Bonn Bot. Gard.	Borsch 3466 (BONN)	AY145348
Ceratophyllum demersum L.	Ceratophyllaceae	USA, Virginia	Wieboldt 16073 (VPI)	AY145335
Chloranthus brachystachys Bl.	Chloranthaceae	Bonn Bot. Gard.	Borsch 3467 (BONN)	AY145334
Dicentra eximia (Ker Gawl.) Torr.	Fumariaceae	Bonn Bot. Gard.	Borsch 3468 (BONN)	AY145361
Ginkgo biloba L.	Ginkgoaceae	Virginia Tech Bot Gard.	Borsch 3469 (VPI)	AY145323
Gnetum gnemon L.	Gnetaceae	Bonn Bot. Gard.	Borsch 3470 (BONN)	AY304546
Illicium floridanum Ellis	Illiciaceae	USA, Florida	Borsch & Wilde 3104 (VPI, FR)	AY145325
Lactoris fernandeziana Phil.	Lactoridaceae	DNA from Tod Stuessy	Stuessy s.n.	AY145324
Umbellularia californica Nutt.	Lauraceae	Bonn Bot. Gard.	Borsch 3471 (BONN)	AY145350
Liriodendron tulipifera L.	Magnoliaceae	USA, Virginia	Slotta s.n. (VPI)	AY145356
Magnolia virginiana L.	Magnoliaceae	USA, Maryland	Borsch & Neinhuis 3280 (VPI, FR)	AY145354
Michelia champaca L.	Magnoliaceae	Bonn Bot. Gard.	Borsch 3472 (BONN)	AY145355
Myristica fragrans Houtt.	Myristicaceae	Bonn Agr. Bot. Gard.	Borsch 3473 (BONN)	AY145351
Nelumbo nucifera subsp. *lutea* (Willd.) Borsch & Barthlott	Nelumbonaceae	USA, Missouri	Borsch & Summers 3220 (FR)	AY145359
Nuphar advena (Aiton) W.T. Aiton	Nymphaeaceae	USA, Florida	Borsch & Wilde 3093 (FR)	AY145351
Nuphar lutea (L.) Sibth. & Sm.	Nymphaeaceae	Germany, Hesse	Borsch 3337 (FR)	AY145330
Nymphaea odorata Ait. subsp. *odorata*	Nymphaeaceae	USA, Georgia	Borsch & Wilde 3132 (VPI, BONN)	AY145333
Victoria cruciana Orbign.	Nymphaeaceae	Bonn Bot. Gard.	Borsch 3474 (BONN)	AY145332
Piper angustum Rudge	Piperaceae	Missouri Bot. Gard.	Acc. 910150	AY145345
Piper spec.	Piperaceae	Bonn Bot. Gard.	Borsch 3475 (BONN)	AY145346
Platanus occidentalis L.	Platanaceae	USA, Virginia	Slotta s.n. (VPI)	AY145358
Houttuynia cordata Thunb.	Saururaceae	Bonn Bot. Gard.	Borsch 3481 (BONN)	AY145344
Saururus cernuus L.	Saururaceae	USA, Florida	Borsch & Wilde 3108 (VPI, FR)	AY145343
Schisandra rubriflora	Schisandraceae	Bonn Bot. Gard.	Borsch 3477 (BONN)	AY145327
Tofieldia glutinosa (Michx.) Pers.	Tofieldiaceae	USA,	Borsch, Hellquist, Wiersema 3393 (VPI, BONN)	AY145337
Trochodendron aralioides P.F. Siebold & J.G. Zuccarini	Trochodendraceae	Bonn Bot. Gard.	Borsch 3478 (BONN)	AY145360
Drimys winteri J.R. Forster & G. Forster	Winteraceae	Bonn Bot Gard.	Borsch 3479 (BONN)	AY145347

trnT-trnF sequence was obtained from the complete sequence of the chloroplast genome (Tsudzuki *et al.*, 1992; GenBank number NC001631). Classification is in accordance with the APG (1998) system. However, for the Chloranthales (comprising Chloranthaceae) and Winterales (comprising Canellaceae and Winteraceae), an ordinal rank is accepted because (1) published ordinal names exist, (2) these groups are now identified as clearly monophyletic lineages, and (3) they do not belong to the basal angiosperm grade comprising Amborellaceae, Nymphaeaceae, Austrobaileyaceae, Illiciaceae, Schisandraceae and Trimeniaceae.

DNA isolation, amplification and sequencing

Total genomic DNA was isolated from frozen (stored at −80 °C) or silica-gel-dried leaf tissue using a modified (2% cetyltrimethylammoniumbromide, 1% polyvinylpyrrolidone, 100 mM Tris (pH 8), 20 mM EDTA, 1.4 M NaCl) (CTAB) method. The isolation procedure was modified in the present study by introducing triple CTAB extractions to yield optimal quantities of high-quality DNA from tissues with considerable amounts of secondary compounds that occur in many basal angiosperms. This protocol is a modification of a miniprep procedure described in Liang & Hilu (1996). About 100 mg of dry tissue (equaling approximately 300 mg of fresh tissue) were ground in liquid N_2 and incubated at 65 °C for 30 min with 700 μL of CTAB. After centrifuging and transferring the supernatant into a clean tube, the same tissue was reincubated twice with CTAB solution. All three preparations were kept separate. The CTAB solutions were then extracted with chloroform twice, and the DNA was subsequently precipitated with ethanol. After separately resuspending the pellets from all extraction steps in TE, two cleaning steps were carried out: the first by adding one-half volume 7.5 M ammonium acetate and precipitating with 100% ethanol, and the second by adding one-half volume 3 M sodium acetate and precipitating with ethanol. Genomic DNA from the second and third extractions was usually clean enough to be directly used for polymerase chain reaction (PCR) amplification.

The region was PCR-amplified in two overlapping fragments with universal primers (Taberlet *et al.*, 1991) annealing to the tRNA genes. Primers a and d or rps4-5F (5′-AGGCCCTCGGTAACGSG-3′, designed in this study) and d were used to amplify the *trnT-L* spacer together with the *trnL* intron, and primers c and f were used to amplify the *trnL* intron and the *trnL-F* spacer. Amplification conditions were: 34 cycles of 94 °C (1 min) denaturation, 52 °C (1 min) annealing, 72 °C (2 min) extension and 72 °C (15 min) final extension. The PCR products were then purified using a QiaQuick gel extraction kit (QIAGEN, Inc., Valencia, CA, USA) and directly sequenced with an ABI Prism™ BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer, Norwalk, CT, USA) on ABI 310 and 377 automated sequencers. In addition to the above mentioned primers, trnL110R (5′-GAT TTG GCT CAG GAT TGC CC-3′) was designed as an additional universal sequencing primer for angiosperms.

Sequence alignment

Sequence divergence in noncoding regions is caused by a variety of small structural changes in addition to substitution events. We concur with the opinion expressed by Thorne *et al.* (1992), Gu & Li (1995), Benson (1997), Kelchner (2000), and others that the nature of the underlying molecular processes leading to structural changes has to be used as the basis for alignment. Therefore, the processes creating length mutations need to be considered as of particular importance for homology assessment. In *trnT-trnF* sequences, most of the structural changes are known to be SSRs of 4 bp and more (e.g. Van Ham *et al.*, 1994; Bayer & Starr, 1998). Small indels (1–3 bp) are rare and usually confined to poly-A/T strings. Several algorithms for multiple sequence alignment have been developed (e.g. McClure *et al.*, 1994). However, currently available algorithms do not always allow a safe recognition of structural motifs of unpredictable kind, length and complexity (e.g. Graham *et al.*, 2000; Kelchner, 2000), such as SSRs occurring in tandem, shorter indels occurring within larger, clearly delimited indels, or small inversions. These difficulties are caused by defining nucleotides as discrete and independent characters throughout all alignment positions (Kelchner, 2000), regardless of the possibility that a single length mutational event might have involved several nucleotides at once or not. This also explains why the application of fixed gap costs in current alignment algorithms can result in an alignment that deviates from the optimum (i.e. if length mutations are considered putative single events). Therefore, alignment was carried out by eye based on direct sequence comparison using QUICKALIGN 1.5.5 (Müller, 2000), a program designed for optimal manual sequence adjustment. For stringency, rules for manual alignment are required that consider known mechanisms of sequence evolution as well as other, similarity-based criteria for homology assessment, as proposed by Golenberg *et al.* (1993), Kelchner & Clark (1997), Hoot & Douglas (1998), Graham *et al.* (2000), Simmons & Ochoterena (2000) and others, which have been accepted in many studies including the present study. Similarity is a valid criterion to hypothesize homology not only in morphological but also in molecular characters (Doyle & Davis, 1998). Indels are called ‘entire’ (i.e. positional extension is identical in all taxa in which an indel occurs; Graham *et al.*, 2000) or ‘overlapping’ (i.e. positional extensions differ in different taxa). Overlapping indels may be explained by two or more subsequent length mutational events in one taxon, or by different, overlapping length mutational events in

different taxa. Inversions are not discussed as they were not found in the present data set.

The rules employed for the *trnT-trnF* alignment are presented below.

1 Gap insertion. For the insertion of gaps, attention was given to both the potentially inserted sequence *and* its neighbouring sequences. A gap was inserted only when it prevented the inclusion of *more than two* substitutions among closely adjacent nucleotides in the alignment. This decision is based on empirical data from analyses of *trnT-trnF* sequences in basal angiosperms (Borsch, 2000) where length mutations were found to occur approximately half as frequently as substitutions.

2 Placement of gaps. For the placement of gaps, the recognition of sequence motifs was given priority following Kelchner & Clark (1997), which in this data set are only direct SSRs. Golenberg *et al.* (1993), who first proposed alignment rules for length-variable DNA sequences, called multinucleotide repeats 'Type 1b gaps'. Giving priority to a motif can result in insertions that are correctly aligned as nonhomologous (i.e. with different positional extensions) although sequence similarity would warrant their inaccurate placement under the same column (e.g. 6B, see Kelchner, 2000).

3 Homonucleotide strings. Individual positions in homonucleotide strings of different lengths (poly-As or -Ts) are considered to be of uncertain homology (Golenberg *et al.*, 1993; Hoot & Douglas, 1998; Kelchner, 2000) and are therefore excluded. Slipped strand mispairing (Levinson & Gutman, 1987) is likely to have led to numerous length mutational events involving one to several nucleotides. As only nucleotides of the same kind are involved, accurate motif recognition is not possible.

4 Determination of entire indels. Entire indels of the same positional extension and of complete sequence similarity were very easily assessed as primary homologous sensu De Pinna (1991) and consequently placed in the same column(s) of the alignment. During primary homology assessment, no inference had to be made regardless of whether the length mutational event occurred in a common ancestor of all taxa sharing it or in parallel in different lineages. This is analogous to the fact that the synapomorphic status of a substitution in a particular position is not inferred in the alignment process. Recognition of a repeat motif was regarded as further evidence for correctly recognizing a length mutational event.

5 Substitutions in indels. If entire indels of the same positional extension differed by individual substitutions, two principally different cases were distinguished. (A) Direct repeats with exact duplication of a sequence template that has already acquired a substitution (compared with other taxa in the alignment). The presence of autapomorphic or synapomorphic substitutions in the template sequence in this case implies that the repeat event happened after the substitution event. Compared with taxa without substitutions, those motifs provide evidence for unravelling the parallel nature of an insertion event before its potential synapomorphic status could have been tested in a phylogenetic context. As cases without substitutions do not allow such inference, and levels of homoplasy in length mutational events should be assessed equally over all alignment parts, positional extension of indels is regarded as a decisive criterion. A side-effect is that such substitutions in indels can receive double weight in phylogenetic analysis, but the signal would still be in favour of the correct topology. (B) Repeats with substitutions not found in their template sequence: this implies that substitutions either occurred in the template or inserted sequence during or after the replication process. As there is no way of distinguishing which of the nucleotides were the template and which were inserted, correct assignment of these variable positions is not possible. Consequently, variable positions of case B were excluded from the analysis for objective homology assessment following Kelchner (2000) and Asmussen & Chase (2001). We followed this more conservative approach, although Graham *et al.* (2000) did not see the need for exclusion.

6 Overlapping indels. Those indels can be explained by two or more length mutational events and are also called 'progressive step indels' (Kelchner, 2000). For their alignment, a parsimony principle is employed where the least number of steps required is assumed as most probable. The least number of steps can only be inferred using a global perspective. For detecting alternative explanations, all sequences that are length variable in the respective region were placed in close proximity. Gaps were then placed so that only a minimum number of rectangles are required to describe the gaps globally. When this criterion is applied, two different cases need to be distinguished. (1) Overlapping indels with complete sequence similarity can be easily considered primary homologous following Kelchner (2000; example 5). This assumption is valid regardless of whether or not repeat motifs can be identified or the origin of an inserted sequence can be determined. (2) In case of overlapping indels differing by individual substitutions, homology assessment can follow sequence similarity criteria to place overlapping sequence parts (i.e. nucleotides present in taxa with shorter than the largest gaps). Where SSRs were involved, the above-mentioned rules had to be applied. If different placements of overlapping sequence parts (including different arrangement of variable sites) requiring the same minimal number of length mutational events were possible, homology was considered uncertain. Other authors (Gatesy *et al.*, 1993; Davis *et al.*, 1998; Simmons & Ochoterena, 2000) do not think that the latter have to be excluded because these alternative positions are considered to be neutral in parsimony searches. We followed the more conservative approach.

7 Regions of uncertain homology. Those regions (referred to here as hot spots) were excluded from phylogenetic analysis following Swofford & Olsen (1990) and Swofford

et al. (1996). As these hot spots are confined to a few blocks, their removal does not constitute a subjective exclusion of information. The core structure of *trnT-trnF* sequences is represented across the data set. Furthermore, these excluded blocks are comparatively small. Depending on the species, they represent approximately 7.7% (*Amborella*) to an average of about 20% of the entire region for the ingroup taxa and below the average for the outgroups (*Araucaria* = 13.8%; *Pinus* = 14.7%; *Ginkgo* = 12.9%; see Table 2). Accurate information on the location of hot spots in the sequences of all species is provided in Table 2. A similar approach was adopted in broad scale analyses of 18S rDNA in angiosperms and land plants (e.g. Soltis *et al*., 1997, 1999b).

Determination of secondary structure

Based on the Michel–Westhof model of the catalytic core (Michel & Westhof, 1990), Cech *et al.* (1994) proposed a convention to draw secondary structures of group I introns, which is followed here. Cech *et al.* (1994) pointed out that introns might vary considerably in size and number of helical elements, especially at P8. Consequently, the different helical elements (stem-loop regions P1 and 2, P4 and 5, P6, P8 and P9) as well as the cloverleaf structure of the tRNA-leucine have been predicted using free energy minimization (Jaeger *et al*., 1989; Zucker, 1994; Zucker *et al*., 1999). In order to characterize the borders of highly variable parts with the P8 stem-loop region of the *trnL* intron, we chose an integrated approach of comparative sequence analysis and free energy minimization as proposed by Jaeger *et al.* (1990). Predictions of secondary structures based on free energy minimization were computed with RNA structure 3.6 (Mathews *et al*., 2001) and with the *mfold* server (www.bioinfo.math.edu/~mfold) that allowed a more adequate selection of parameters.

Phylogenetic analysis

Analyses were based on nucleotide substitutions, and gaps were treated as missing characters. This approach also allowed us to compare the results with those based on coding genomic regions. Phylogenetic trees were constructed with PAUP*4.0b6 (Swofford, 2001) employing maximum parsimony (MP) with heuristic searches consisting of 100 and 1000 replicates of random stepwise addition with MULPARS in effect and tree bisection reconnection (TBR) branch swapping. Characters were optimized with ACCTRAN. Measures of support for individual clades are based on bootstrap analysis of 500 replicates and decay analysis as implemented in PAUP* and AutoDecay (Erikson & Wikstrom, 1996). Numbers of steps per site were calculated using the CHART option of MacClade 3.07 (Maddison & Maddison, 1997). The data were also analysed with maximum likelihood (ML) implemented in PAUP*. A general time reversible model was employed as an approach for direct estimation of substitution rate matrix parameters and nucleotide frequencies via ML. We are aware that under these settings calculation time might be higher compared with less complex models. Four heterogeneous rate categories across sites were specified after an approximation of the gamma distribution. Heuristic search was performed with starting trees obtained by 'as-is' stepwise addition, and TBR was used as branch swapping algorithm with MulTrees in effect.

Results

Variability of the *trnT-trnF* region in basal angiosperms

In the angiosperm taxa studied, the overall length of the *trnT-trnF* region (excluding the tRNA genes; Fig. 1, Table 2) ranges from 1309 to 2255 bp; the *trnT-L* spacer accounts for 467–1411 bp, the *trnL* intron for 324–615, and the *trnL-F* spacer for 164–441. The *trnT-trnF* region is similar in length within the four gymnosperm taxa sequenced as outgroups, except for *Gnetum* in which the two spacers are considerably shorter (280 and 138 bp) and the intron is only somewhat reduced in size (346 bp). The alignment (see 'Supplementary material') was performed through the gymnosperms with the exception of *Gnetum*. The latter had accumulated numerous autapomorphies of, sometimes, unclear homology in the spacers, and thus was excluded from the analysis. The overall sequence alignment is 4622 bp long (without tRNA genes; including hot spots). High variability commonly detected at lower taxonomic levels turned out to be confined to certain mutational hot spots (Figs 1 and 2; Table 2). This pattern allowed us to exclude from the analyses all positions of uncertain homology.

The striking difference between the absolute length of the region and the alignment is caused by length mutations, occurring at about half the frequency of nucleotide substitutions. Indels inserted in the alignment range in length from 1 to 200 bp, and most of the insertions identified as simple repeat motifs were 4–6 bp long. Inversions were absent. Several of the length mutations are synapomorphic, defining specific clades, some of which are cited in the following. Although a detailed account of microstructural changes in *trnT-trnF* in basal angiosperms goes beyond the scope of the present paper, the following are examples of major synapomorphies: the Nymphaeaceae (represented by *Brasenia*, *Cabomba*, *Nuphar*, *Nymphaea* and *Victoria* and corresponding to Nymphaeales) share a 'TTATG' – insertion in alignment positions 1341–1345 in the *trnT-L*-spacer and an 'AAATG' – SSR in positions 4603–4607 of the *trnL-F*-spacer; the lineage of Piperaceae and Saururaceae within Piperales a 'CTTT' – SSR in positions

Table 2 Actual length of the *trnT-trnF* region in basal angiosperms and gymnospermous outgroups and positions of mutational hot spots in the respective sequences (counts start with position 1 from the 5′ end of each spacer and the intron). Note that length variation of the spacers and the intron is mainly caused by the insertion of nucleotides within the hot spot regions, which differ depending on the species. Where there are no insertions in a hot spot region in individual taxa, the hot spot is considered as not present, 'n. p.' For mean sizes of hot spots see Fig. 1.

Taxon	Length of *trnT-L*-spacer (bp)	Length of *trnL*-intron (bp)	Length of *trnL-F*-spacer (bp)	Position of H1	Position of H2	Position of H3	Position of H4	Position of H5	Position of H6	Position of H7	Position of H8
Araucaria	412	495	466	60–94	215–218	n.p.	289–325	241–248	290–292	192–294	n.p.
Pinus	423	489	377	49–88	214–217	278	294–336	252–263	296–298	137–219	267–270
Ginkgo	379	500	362	51–88	201–204	270	271–295	241–248	286–304	137–189	258–269
Amborella	474	475	374	74–88	206–218	321–323	337–388	233–240	285–295	n.p.	270–275
Illicium	617	518	243	62–242	377–390	469–472	486–523	257–264	302–314	n.p.	127–133
Austrobaileya	684	476	389	67–309	444–456	535–537	551–589	234–241	279–291	150–183	279–285
Schisandra	554	484	396	56–197	327–339	418–420	434–464	242–249	287–293	145–173	279–285
Cabomba	484	508	396	76–140	253–265	351–353	376–402	242–245	283–325	176–220	288–293
Brasenia	479	522	359	77–128	241–253	337–339	362–387	240–243	281–338	162–194	257–263
Nuphar lut	483	588	365	61–124	241–253	338–340	354–388	243–246	284–405	158–190	256–261
Nuphar ad	478	607	365	61–124	241–253	338–340	354–388	244–247	285–424	158–190	256–261
Victoria	467	577	373	62–113	225–237	322–324	338–370	241–244	282–392	145–183	246–255
Nymphaea	476	521	379	62–113	225–237	327–329	343–384	241–244	282–336	174–206	269–275
Chloranthus	797	495	351	55–455	584–596	672	694–717	250–257	295–311	166–171	242–248
Ceratophyllum	838	530	441	64–476	595–607	687–688	706–749	243–256	307–351	190–206	303–312
Acorus	726	522	376	56–359	489–501	580–582	598–636	243–272	318–331	174–180	260–268
Tofieldia	1385	521	239	60–1000	1126–38	1217–19	1236–78	244–262	300–333	24–34	125–131
Orontium	768	615	164	69–364	508–520	599–601	618–671	269–293	331–413	n.p.	n.p.
Nypa	794	522	345	61–424	556–568	655–659	677–711	244–268	306–331	170–180	n.p.
Saruma	750	505	356	53–360	493–505	583–586	608–641	245–256	298–328	171–180	247–253
Aristolochia	716	512	371	55–323	453–465	558–560	582–618	261–278	321–345	167–180	257–263
Lactoris	795	498	373	79–412	548–560	638–640	654–695	255–265	300–321	171–184	262–269
Saururus	877	495	350	56–461	595–607	700–709	726–768	258–265	302–319	176–191	261–270
Houttuynia	1411	491	353	56–979	1117–29	1222–31	1248–91	257–264	301–315	174–189	262–271
Piper ang.	844	490	375	56–426	560–572	665–674	691–734	255–262	299–314	172–193	271–280
Piper spec.	802	491	381	556–379	513–525	618–627	644–687	255–262	299–315	172–199	277–286
Drimys	717	497	359	60–346	480–492	569–571	588–629	237–254	292–316	174–179	249–255
Canella	793	479	258	56–412	546–558	637–639	656–699	239–252	290–309	73–78	148–154
Calycanthus	653	324	332	77–274	400–412	494–496	513–555	n.p.	n.p.	151–156	216–228
Umbellularia	545	484	362	33–160	286–293	374–376	393–452	242–254	292–310	176–181	252–258
Myristica	881	503	300	55–505	642–653	732–734	751–797	235–258	296–324	175–181	252–258
Annona	–	349	378	–	–	–	–	241–258	n.p.	181–191	257–263
Asimina	–	496	390	–	–	–	–	242–259	297–315	188–198	264–275
Magnolia	766	492	355	56–409	539–551	630–632	649–687	241–258	296–313	168–174	245–251
Michelia	772	492	356	56–415	545–557	636–638	655–693	241–258	296–313	169–175	246–252
Liriodendron	783	491	361	57–411	542–554	633–635	652–698	241–258	296–312	170–180	251–257
Buxus	685	507	377	62–298	419–431	510–516	533–580	252–269	312–328	177–182	256–262
Platanus	1011	525	365	56–630	752–764	843–850	867–910	258–276	319–346	151–156	235–241
Nelumbo	1035	527	400	56–663	784–796	875–877	894–937	257–276	319–342	187–192	285–291
Trochodendron	1077	441	368	57–719	825–837	916–918	935–978	n.p.	248–262	191–196	269–275
Dicentra	713	476	359	56–319	433–445	526–528	545–585	238–249	289–307	158–161	232–238
Aextoxicon	857	511	354	87–459	593–605	691–693	710–752	261–278	321–337	166–171	239–245

3643–3646 in the *trnL-F*-spacer; a clade of *Magnolia*, *Michelia* and *Liriodendron* within Magnoliaceae which, based on substitutions only receives 68% bootstrap support (BS) with MP, shares a 'GAATC' – SSR in positions 2622–2626 in the *trnL*-intron; and the two species of *Nuphar* share a 'GATTT' – SSR in positions 1373–1377 in the *trnT-L*-spacer. It appears that synapomorphic indels occur at various taxonomic levels, and vary considerably in their distribution where some branches of the basal angiosperm tree (e.g. the one leading to the Nymphaeaceae) are marked by many indels, whereas others have few or none. Thorough analyses of their type and distribution with broader taxon sampling will help to assess their phylogenetic utility at various taxonomic levels including the genus level. Long indels were rather rare and were restricted to individual taxa (autapomorphic). Further, long insertions (>20 bp) generally do not occur as repeated motifs. Most prominent examples are the 176-bp insertion in the *trnT-L* spacer and the 200-bp deletion in the *trnL-F* spacer of *Austrobaileya* and *Illicium*, respectively. Interestingly, both genera are members of the same small clade (Figs 3 and 4).

Proposed secondary structure

The proposed secondary structures for the tRNA-Leucine and the *trnL* intron in *Nymphaea odorata* are given in Fig. 2. The P6 and P8 stem-loop regions account for most of the sequence length variation in the intron. Minimum free energy configurations reveal several helical elements, labelled using roman numerals (I–III; Fig. 2) within an extensive P8 region. Within helical element I, repetitive elements are inserted in a number of basal angiosperm taxa (hot spot H5; Table 2). This is not the case in *Nymphaea*. Therefore, the respective position is marked by a single arrow in Fig. 2C. A second mutational hot spot (H6; Table 2) that was also excluded from phylogenetic analysis falls into helical element II. Two arrows border the respective AT-rich string of repetitive elements. It is important to note that P8 is conserved for most of its primary sequence in angiosperms. Only two positions are prone to larger inserts, which can be of independent origin and may vary considerably in length among taxa.

Phylogeny of basal angiosperms

The two spacers and the intron provided a set of 3112 characters excluding hot spot regions and exons. The positions of excluded parts with respect to the alignment with a total length of 4707 bp are (Fig. 1): 256–1276 (H1), 1538–1550 (H2), 1729–1750 (H3), 1795–1927 (H4), 2194–2228 (trnL-5′ exon), 2720–2749 (H5), 2837–2990 (H6), 3330–3379 (trnL-3′ exon), 4025–4145 (H7), 4403–4418 (H8). Of these 3112 characters, 928 characters were variable in angiosperms (1070 in whole data set), of which 608 are parsimony-informative in angiosperms (738 in the whole data set). The relative contributions of the three *trnT-trnF* sections are summarized in Table 2. The MP analysis of *trnT-trnF* sequences resulted in two shortest trees of 3198 steps (consistency index = 0.565, retention index = 0.592), differing only in the position of *Dicentra* (Ranunculales) being either basal in eudicots (Fig. 3) or sister to a eudicot clade consisting of *Buxus*, *Aextoxicon* and *Trochodendron* (tree not shown). Increasing the number of replicates during random stepwise addition from 100 to 1000 found the same two trees, increasing the possibility that the most parsimonious trees were recovered. The ML analysis resulted in one tree with a score of $-\ln L = 18720.06573$ (not shown). The ML phylogeny differed from the MP in placing the Chloranthales as sister to the eumagnoliids (defined here to comprise Laurales, Magnoliales, Piperales, Winterales) instead of being sister to the eudicots, as in the MP trees.

The *trnT-trnF* MP strict consensus (Fig. 2) clearly depicts [99% BS, decay value (DE) of 13] the New Caledonian woody shrub *Amborella* as sister to all remaining angiosperms. Diverging next is the herbaceous water lily lineage Nymphaeaceae (94% BS, six DE), followed by the *Schisandra–Illicium–Austrobaileya* clade (100% BS, 18 DE) that comprises small trees and woody lianas. The Chloranthaceae, a tropical family with very reduced flowers, emerges as sister to the eudicots, but with bootstrap support <50%. Piperales, comprising Piperaceae, Saururaceae, Aristolochiaceae and Lactoridaceae (members of what have been known as paleoherbs), are highly supported (97% BS, nine DE) in a clade sister to woody Canellaceae and Winteraceae (Winterales). Support for the latter sister relationship is weak (64% BS, one DE). Magnoliales and Laurales, the first of which gains moderate (73% BS, two DE) and the latter strong (100% BS, eight DE) support, appear in a clade sister to the Piperales–Winterales clade. The clade of these four lineages (eumagnoliids) is weakly supported (60% BS and two DE). The eudicots, encompassing the dicot families with tricolpate or tricolpate-derived pollen, constitute a monophyletic lineage (100% BS, 13 DE) in which *Dicentra* (Ranunculales) appears in a polytomy with a well-supported *Nelumbo–Platanus* clade (Proteales, 98% BS, 10 DE) and a clade containing the other eudicots (represented by *Trochodendron*, *Aextoxicon* and *Buxus*). A phylogram (MP) of one of the two shortest trees is shown in Fig. 3 to illustrate branch lengths and ML branch lengths leading to some important nodes are presented in the following. The branch leading from the first angiosperm node to the subtrees where Nymphaeaceae and Austrobaileyales are basal has 53 and 26 steps (Fig. 4). The branch leading from the root node to *Amborella* is 73 steps long, whereas the branches of water lilies and Austrobaileyales are 156 and 158 steps on average (mean of all terminal taxa belonging to respective clades). The

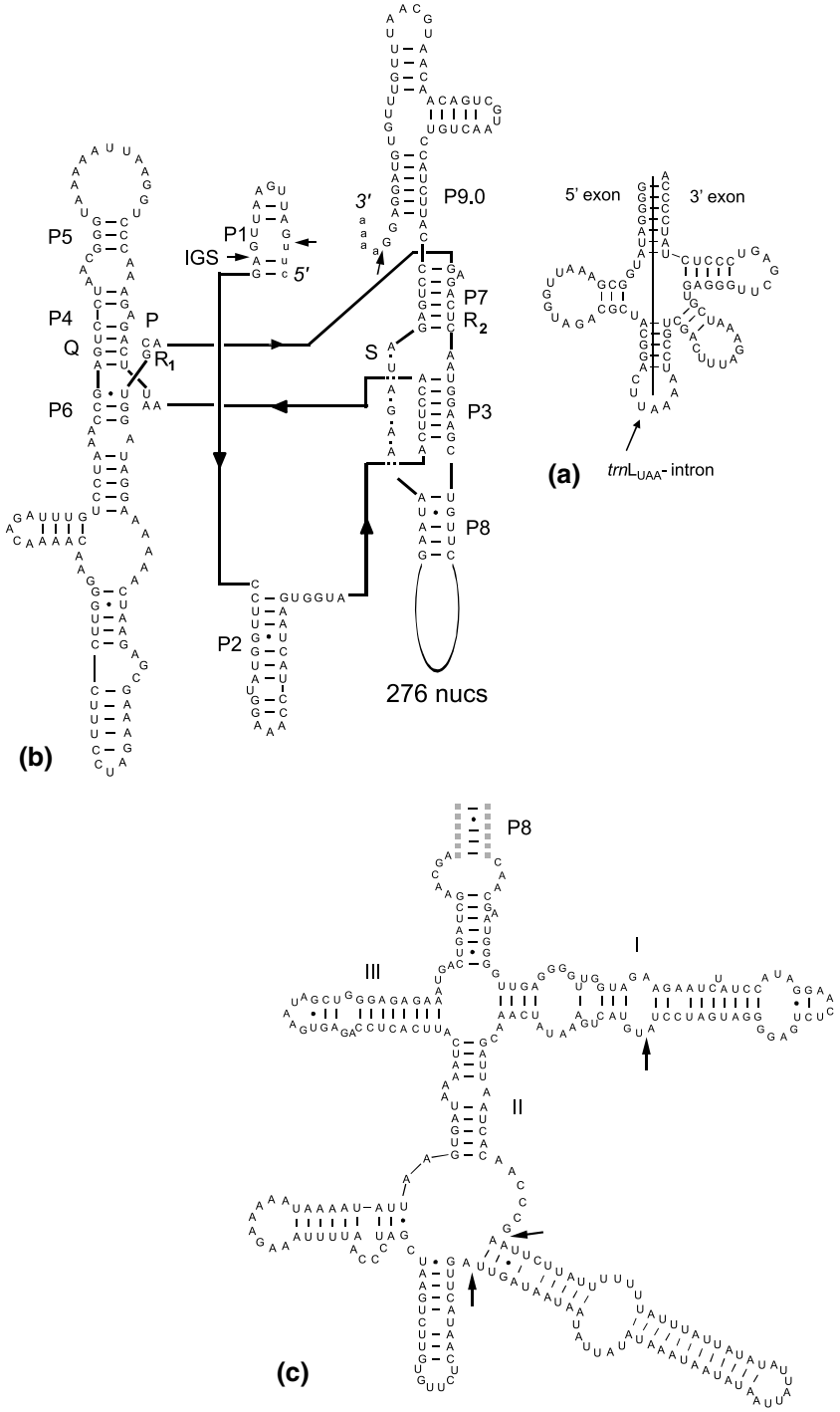


Fig. 2 Proposed secondary structure of the tRNA for leucine (UAA) encoded by the *trnL* gene in *Nymphaea odorata*. (a) Cloverleaf structure corresponding to the two exons, (b) proposed secondary structure for the intron and (c) P8 stem-loop region. Three main helical elements are labelled using roman numerals I–III. The single arrow in helical element I indicate the position of a repeat region, which is missing in *Nymphaea*. The two arrows in helical element II border an AT-rich string of repetitive elements that cannot be aligned across angiosperms and was therefore excluded from phylogenetic analysis.

branch leading from the basal grade to the subtree where the monocot–*Ceratophyllum* clade is basal is 33 steps. The branch length with ML are 0.029 and 0.010 from the first angiosperm node to the subtrees with Nymphaeaceae and Austrobaileyales at the base; 0.087 from the root node to *Amborella*, and 0.101 and 0.068, respectively, to taxa of Nymphaeaceae and Austrobaileyaceae. The branch leading from the basal grade to the subtree with the monocot–*Ceratophyllum* clade basal is 0.041.

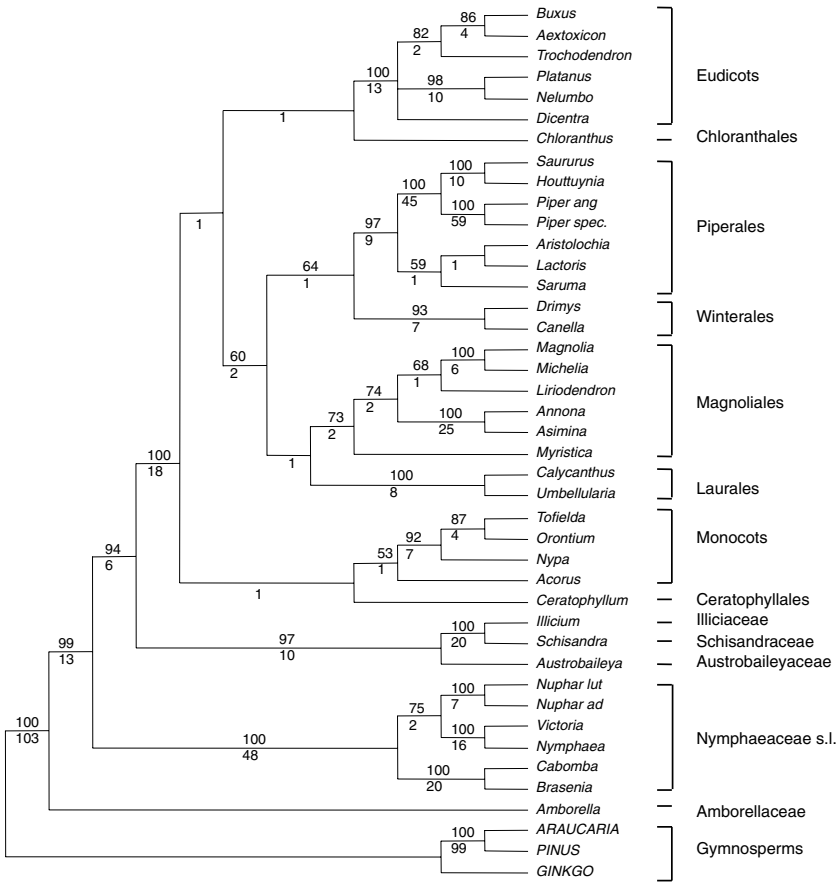


Fig. 3 Strict consensus of the two most parsimonious trees (3198 steps, CI = 0.565, RI = 0.592) showing phylogenetic relationships among basal angiosperms based on noncoding sequences from the plastid region *trnT-trnF*. BS values >50% are given above and DE values below branches.

Discussion

Molecular evolution of *trnT-trnF* and implications for phylogenetic utility

Sequences of *trnT-trnF* presented here from across seed plants allow us to examine early evolution in flowering plants with fast-evolving, noncoding DNA and to compare the evolution of the two spacers and the intron over a broad evolutionary scale. The region has been widely used in systematic studies below the family level, and often only the *trnL* intron and *trnL-F* spacer are employed (e.g. Gielly *et al.*, 1996; Sang *et al.*, 1997; Small *et al.*, 1998). It is rather striking to see this noncoding region providing strong historical signal and high resolution deep in angiosperm phylogeny. The numbers of variable and informative sites correlate with the number of aligned characters for each of the three parts (Table 3). However, looking at actual sequence lengths it appears that the *trnL* intron sequences contain only 63% variable sites compared with 83 and 98% in the *trnT-L* and *trnL-F* spacers, respectively. This seems to be caused by fewer small length mutational changes in the intron compared with the spacers.

Estimates of variability for noncoding DNA cannot be carried out very easily because of multiple-nucleotide mutational events (i.e. length mutations). Aligning sequences with insertions results in an accelerated character number and thus underestimation of variability. We therefore consider a 'corrected' percentage value based on average actual sequence length (Table 3) to present a more accurate approximation of variability. The differences between the corrected and uncorrected values could be substantial, and in this data set they range from 2.0 to 3.6-fold. The correction measure used here should be considered as a rough approximation. It is obvious that differences in sequence lengths among the sampled taxa may bias the mean sequence length, but we consider the adoption of such a correction measure provides a more realistic picture than using uncorrected values.

Compared with the *trnT-L* spacer, it seems that the *trnL* intron and the downstream spacer, *trnL-F*, evolve in concert. Length variability is much higher in *trnT-L* than in the other two regions, as indicated by the SE of average length of 200 vs. 50 (Table 3). In addition, 4–6 bp simple direct repeat motifs are about 30% less frequent in the intron and *trnL-F* than *trnT-L* (Borsch,

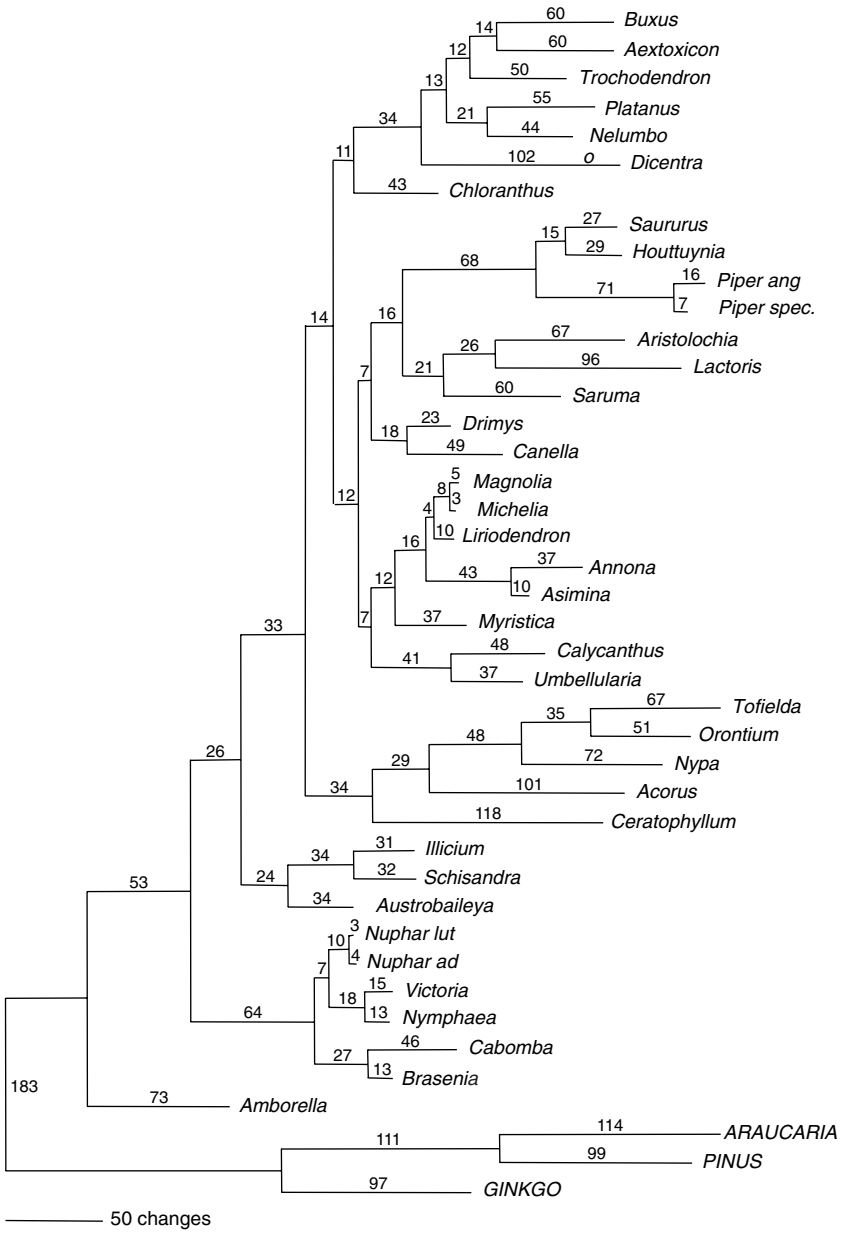


Fig. 4 One of the two shortest trees (3198 steps) found in parsimony analyses of the *trnT-trnF* data set. Branch lengths (ACCTRAN optimization) are indicated above branches. The second tree only differed by the position of *Dicentra* among basal eudicots.

2000). Moreover, the mutational hot spot sectors are much smaller in the *trnL* intron and *trnL-F* spacer compared with *trnT-L* (Table 2). It is also worth mentioning that the respective tRNA genes of the *trnL* intron and *trnL-F* spacer are transcribed in the same direction (Shinozaki *et al.*, 1986; Kanno & Hirai, 1993). The relatively conserved length of the *trnL* intron may relate to the role this group I intron plays in splicing during mRNA processing (Kuhsel *et al.*, 1990).

These evolutionary patterns are also reflected in the gymnosperm species examined here, including the divergent *Gnetum*. The absolute size and degree of sequence divergence are proportionally less pronounced in the intron (346 bp) than in the spacers (280 and 138 bp; in angiosperms the average length of the intron and spacers are 500, 739 and 355 bp, respectively; Table 2). The extreme divergence found in the sequence and absolute size of *trnT-trnF* of *Gnetum* is important in the light of anthophyte hypothesis (e.g. Crane, 1985; Doyle & Donoghue, 1986) that proposes the Gnetales and extinct Bennettitales as closest relatives for angiosperms. Information from this and other phylogenetic studies (e.g. Goremykin *et al.*, 1996; Qiu *et al.*, 1999, 2000; Bowe *et al.*, 2000; Chaw *et al.*, 2000; Donoghue & Doyle, 2000), as well as the analysis of genes controlling floral development (Winter *et al.*, 1999) point to the rejection

Table 3 Variation and relative contribution (excluding mutational hot spots) of the three parts of the *trnT-trnF* region in angiosperms and gymnospermous outgroups. High numbers of insertions characteristic of noncoding regions expands the alignment, causing underestimation of variability; for a better approximation, the amount of variability is also calculated on the basis of average actual length of sequences (corrected). Note that character numbers are always based on the alignment.

	trnT-L spacer	*trnL* intron	*trnL-F* spacer	*trnT-trnF*
Average sequence length (bp)	730	500	355	1590
Standard deviation	235	51	49	223
Average sequence length excluding hot spots	376	459	331	1167
Standard deviation	24	33	45	53
Number of characters	1005	919	1188	3112
Variable characters	313	289	326	928
% variable characters (corrected)	31 (83)	31 (63)	27 (98)	30 (79)
Parsimony informative characters	211	179	218	608
% informative characters (corrected)	21 (56)	20 (39)	18 (66)	20 (52)

of the anthophyte hypothesis and acceptance of the monophyly of extant gymnosperms (e.g. Chaw *et al.*, 2000). Divergence of *Gnetum* in the *trnT-trnF* region could be caused by either an accelerated mutational rate or by a very long separation from other lineages including angiosperms. In our main analysis, *Gnetum* was not included because large portions of *trnT-trnF* sequence cannot be aligned. However, a negative effect from not including this gymnosperm lineage in our basal angiosperm analysis is not to be expected, as *Gnetum* most likely is not the immediate sister of angiosperms, and its inclusion would probably have only resulted in additional long-branch attraction effects. The secondary structure of the *trnL* intron (Fig. 2) reveals that the highly length-variable sectors that cannot be aligned across angiosperms are confined to smaller stem-loops within P8. Insertion–deletion events are a characteristic mode of divergence in noncoding regions and tend to be site-dependent (Clegg *et al.*, 1994). The situation in *trnT-trnF* is also comparable with angiosperm 18S rDNA, in which four highly variable regions have been identified by Soltis *et al.* (1997, 1999b). These regions are also located in loops of the proposed ribosomal RNA secondary structure and were excluded from the phylogenetic analysis because of difficulties in alignment. The two hot spots in the *trnL-F* spacer are small areas in which tandem duplications and repetitive elements accumulate, whereas the large mutational hot spot in *trnT-L* seems to be of a different nature. Length variability in *trnT-L* is caused by the addition of sequence in a certain area that seems to occur independently in different lineages and may involve insertions of larger fragments of so far unknown origin, particularly in monocots (Table 2).

The *trnT-trnF* region is known to be fast-evolving; depending on the taxonomic group, it evolves up to three times faster than *rbcL* (e.g. Bayer & Starr, 1998, for Asteraceae; Reeves *et al.*, 2001, for Iridaceae). *TrnT-trnF*, and many other noncoding parts of the large single copy (LSC) region of the chloroplast genome differ considerably in their rates of evolution from noncoding DNA in the inverted repeat (IR; Olmstead & Palmer, 1994; Soltis & Soltis, 1998). This fast rate of evolution in LSC noncoding regions may have led to the notion that the majority of their sites would be saturated when they are used in phylogeny reconstruction at higher taxonomic levels. On the contrary, Fig. 5 indicates that the largest number of variable sites changed only once in the present data set.

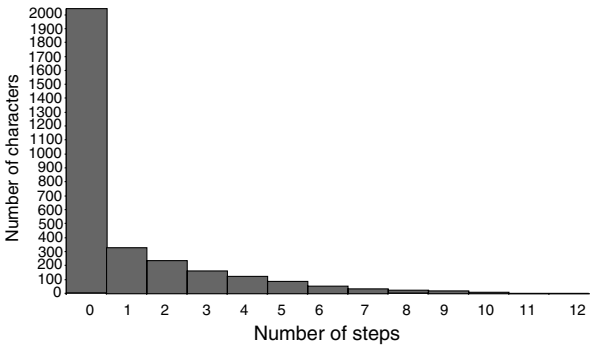


Fig. 5 Amount of variability among characters in the *trnT-trnF* data set containing 3112 characters calculated over tree one of 3198 steps; the *x*-axis indicates the level of variability (i.e. number of steps for a character) and the *y*-axis shows the number of characters for each level of variability. Most of the variable characters have changed only one time.

Reconstructing deep-level phylogeny in the angiosperms with *trnT-trnF* sequences has produced results that are highly congruent with trees inferred from multigene, multi-genome data sets as discussed in detail below. Reasons for this unexpected strong performance of the *trnT-trnF* data set at deep levels may come from the ability of the majority of the sequence positions to evolve rather freely. Based on the secondary structure of the *trnL* intron of *Nymphaea*, 68% of the characters in the present analysis are contributed by the P8 and P6 stem-loops. In contrast, the functionally highly constrained and evolutionarily conserved P, Q, R and S regions,

which act as a core in RNA catalysis (Cech, 1988; Michel & Westhof, 1990; Besendahl *et al.*, 2000), account for only about 9% of the *trnL* intron sequence length (based on the secondary structure model for *Nymphaea*). These regions are not length variable in basal angiosperms and contain only one informative and three additional variable positions that are autapomorphic. Consequently, the effect of the P, Q, R and S regions on phylogeny reconstruction is minimal. Compensatory mutations could present a problem in a phylogenetic analysis because two different character states change dependently, resulting in double weighting of the respective changes. In the *trnL* secondary structure of *Nymphaea*, 95 positions appear in stems which corresponds to the maximum number of potentially co-evolving sites. In the present data set only 20% of these stem positions are variable, so that the overall proportion of possibly co-evolving sites across basal angiosperms is very low. This seems to be in contrast to the maximum possible 73% of compensatory mutations reported from 18S rDNA (Soltis *et al.*, 1997).

Congruence of *trnT-trnF* with multi-gene multi-genome phylogenies

The overall angiosperm phylogeny resolved in this *trnT-trnF* study is highly congruent with that based on multi-gene, multi-genome data sets (Qiu *et al.*, 1999, 2000; Soltis *et al.*, 1999a, 2000). The emergence of a grade of *Amborella*, Nymphaeaceae and *Schisandra–Illicium–Austrobaileya* as the three most basal branches (the basal angiosperm grade) is in agreement with phylogenies based on combined data sets (except *rbcL* plus *atpB*). This grade has not been observed in analyses of individual genes with the exception of *atpB* (Savolainen *et al.*, 2000). In the 18S rDNA sequence analysis, *Austrobaileya*, *Illicium* and *Schisandra* either appear in a clade second to *Amborella* or as the most basal lineage, depending on sampling and outgroup choice (Soltis *et al.*, 1997). Analysis of the *trnT-trnF* region, like the five-gene analysis of Qiu *et al.* (1999, 2000) and the six-gene analysis by Zanis *et al.* (2002), stands out in its strong statistical support for the basal grade (nearly 100% BS for the relevant nodes in both studies). This additional evidence from a genomic region with a basically different evolutionary mode and tempo is of particular importance as rooting the angiosperms with *Amborella* has been discussed in terms of possible long-branch attraction. The five-gene data set shows *Amborella* to have the longest branch (357 steps; Qiu *et al.*, 2000). Taxon deletion analyses in the same study found a likelihood measure in favour of an *Amborella* plus Nymphaeaceae clade rather than *Amborella* alone as the first branch. Similar results were obtained with likelihood analyses and noise reduction experiments of a data set consisting of sequences from the three plant genomes (Barkman *et al.*, 2000). In the present analysis, the node uniting all other angiosperms above *Amborella* does not only receive higher bootstrap support (99% compared with 91% in six-gene, 88% in five-gene and 65% in three-gene analyses) but also a decay value of 13. Recent extensive analyses on the root of the angiosperms using MP, ML and Bayesian methods of phylogeny reconstruction with an 11-gene data set favoured *Amborella* as sister to all other angiosperms, with less evidence for an *Amborella* plus Nymphaeaceae clade and almost no evidence for Nymphaeaceae alone as respective sister lineages to all other angiosperms (Zanis *et al.*, 2002). Additional support for this basal grade in flowering plants comes from this *trnT-trnF* data set. Congruence between MP and ML analyses of *trnT-trnF* is not only in topology but also in branch lengths. Unrooted MP subtrees of the angiosperms in this study (Fig. 4) reveal that the branches leading to *Amborella* (126 steps) and the water lilies (103 steps on average) are quite similar in length. Branch lengths determined with ML are 0.116 vs. 0.101 and correlate well with those found using MP. Consequently, long-branch attraction to the outgroup does not seem to be a very probable factor to have influenced the basal-most position of *Amborella* in this study, given that Nymphaeaceae and *Amborella* are more or less equally divergent. Therefore, the *trnT-trnF* data support *Amborella* instead of Nymphaeaceae plus *Amborella* as the most basal angiosperm lineage.

The recognition of the basal grade in flowering plants points to strong shifts in habit and habitat quite early in their evolutionary history as exemplified by the divergence of the herbaceous aquatic Nymphaeaceae as the second extant lineage.

The position of monocots varies among trees based on different data sets although none of these alternative placements is well supported. Combined analysis of *rbcL*, *atpB* and 18S rDNA (Soltis *et al.*, 1999a, 2000) shows monocots unresolved with Winterales, Laurales, Magnoliales, Chloranthales and Piperales; *rbcL* data alone (Chase *et al.*, 1993) depicted them unresolved with Laurales and Piperales; *atpB* (Savolainen *et al.*, 2000) analysis revealed monocots sister to eudicots and paraphyletic to *Ceratophyllum* (Ceratophyllales); phytochrome genes *PHYA* and *PHYC* (Mathews & Donoghue, 1999) place monocots as sister to *Chloranthus* (Chloranthales) in a position basal to eudicots; and three of four 18S rDNA data sets show them sister to *Ceratophyllum* within basal eudicots, with *Acorus* resolved independently from the rest of the monocots (Soltis *et al.*, 1997). The present study places monocot-*Ceratophyllum* as the next-branching clade after the basal grade but without bootstrap support, in line with the six-gene analysis of Zanis *et al.* (2002). The five-gene analyses of Qiu *et al.* (1999, 2000) showed the clade sister to Chloranthales in the same phylogenetic position; these nodes, however, collapse in their strict consensus (Qiu *et al.*, 2000). Underlying these inconsistencies in the position of monocots is weak support in all these studies.

In contrast to the difficulties in defining the monocot position among angiosperms, the relationship between monocots and the dicot *Ceratophyllum* is gaining support. The relationship between monocots and the aquatic *Ceratophyllum* inferred in this study is congruent with those based on a large number of slowly evolving chloroplast genes (Graham & Olmstead, 2000) or genes from all three genomes (Qiu *et al.*, 2000; Zanis *et al.*, 2002). The congruence between ML and MP trees implies that the position of *Ceratophyllum* in this study is not influenced by heterogeneity in rates of substitution among taxa and concomitant long-branch attraction. Such a phenomenon could affect MP more strongly than ML analysis (e.g. Huelsenbeck, 1995). The relationship of *Ceratophyllum* to monocots is supported by loss of primary roots. In addition, *Ceratophyllum* shares with the Alismatales, one of the most basal lineages in monocots, the presence of achene fruits (Les & Schneider, 1995). The fossil record is in line with this molecular-based hypothesis because earliest records for both *Ceratophyllum* and monocots have been found almost contemporaneously in the Cretaceous. *Ceratophyllum*-like fruits have been recognized in the Aptian of Australia (Krassilov, 1997), and the earliest fossils that can be assigned to the monocots are triuridaceous flowers from the Turonian (early Upper Cretaceous) of the USA (Gandolfo *et al.*, 2000) and aroid fruits from the Albian (Herendeen & Crane, 1995). Recent calculations by Bremer (2000) postulate that the major monocot lineages may have diverged from each other during the Early Cretaceous. As many basal monocots are aquatic or nearly so (Chase *et al.*, 2000), a possible aquatic ancestor for the *Ceratophyllum*–monocot group ought to be considered.

Among other important angiosperm relationships supported by the *trnT-trnF* data is the association between the largely herbaceous Piperales and woody Winterales. This association is in contrast with the traditional classification that places the woody Canellaceae and Winteraceae into a more broadly circumscribed order Magnoliales (Cronquist, 1981). The *trnT-trnF*-based position of the Winterales with the Piperales is also depicted in the phytochrome gene (Mathews & Donoghue, 1999, 2000; Graham & Olmstead, 2000), five-gene (Qiu *et al.*, 1999, 2000) and six-gene (Zanis *et al.*, 2002) trees analyses. Thus, hypothesized relationships based on different molecular data sets now seem to converge, whereas morphology favours a sister-group relationship of Winterales with Laurales (Doyle & Endress, 2000). The close affinity of the two families Canellaceae and Winteraceae has been suggested earlier based on *rbcL* (Chase *et al.*, 1993; Qiu *et al.*, 1993) and phytochemical data (Gottlieb *et al.*, 1989). The classification of Piperales to contain the four families Piperaceae, Saururaceae, Lactoridaceae and Aristolochiaceae (APG, 1998) can be clearly defended based on the strong statistical support for this clade. However, the position of Lactoridaceae sister to *Aristolochia* within Aristolochiaceae with no support (Fig. 2) or with weak support in the five-gene analysis (Qiu *et al.*, 1999, 2000) might be spurious. Analyses of a larger *trnT-trnF* data set of Aristolochiaceae and Piperales (Neinhuis *et al.*, 1999) displayed the Lactoridaceae sister to the Aristolochiaceae. *Lactoris* has the longest branch among Piperales in this and other data sets (e.g. Graham & Olmstead, 2000), and thus its position might be influenced by long-branch attraction. Morphological information (González & Rudall, 2001) supports relationships of *Lactoris* to Saururaceae and Aristolochiaceae.

In line with previous molecular studies, both Laurales and Magnoliales are resolved as monophyletic lineages with monophyly of Laurales gaining very strong statistical support. However, Magnoliales gain weak support here, which is most likely the result of a low rate of substitutions in the *trnT-trnF* region in this lineage and the subsequent low number of synapomorphic mutations uniting them. The eudicot lineage is very well supported (100% BS, 13 DE), but the branch leading to *Dicentra* (Ranunculales) is very long (Fig. 3). The long branch may have caused the conflict between the shortest trees at the base of the eudicots, which may not be maintained when the eudicots are more densely sampled.

Contributions of noncoding *trnT-trnF* sequence data to understanding basal angiosperm relationships

Recent molecular approaches based on single and combined gene data sets have provided immense insight into the evolution of flowering plants. These contributions are redefining angiosperm classification. Hypotheses from the precladistic era recognized the Magnoliales (Takhtajan, 1980; Cronquist, 1981) with their large showy flowers and a high number of spirally arranged carpels, to be the most ancestral flowering plants (the so called 'Magnolialean Hypothesis'; see Qiu *et al.*, 2000, for overview of basal angiosperm relationships). Analyses of an 18S rDNA data set by Hamby & Zimmer (1992) resolved Nymphaeaceae *s.str.* as the sister group to all other angiosperms. However, results of the first large-scale molecular phylogenetic analyses based on *rbcL* depicted the aquatic *Ceratophyllum* as the first-branching angiosperm (Chase *et al.*, 1993; Qiu *et al.*, 1993). Subsequent intense efforts of sequencing multiple genes from different genomes culminated into a first general hypothesis of what could be the root of the angiosperms (e.g. Mathews & Donoghue, 1999; Qiu *et al.*, 1999; Soltis *et al.*, 1999). The picture has changed not only by revealing *Amborella* as sister to all other angiosperms but also by providing strong corroborative evidence from various genomic regions, including *trnT-trnF*, in support of an *Amborella*, Nymphaeaceae and *Illicium–Schisandra–Austrobaileya* grade. Moreover, *Trimenia* (not sampled here) has been shown to be a member of the *Illicium–Schisandra–Austrobaileya* clade (e.g. Qiu *et al.*, 1999; Renner, 1999; Zanis *et al.*, 2002). In addition, receiving increased

evidence from this noncoding DNA and from multiple gene studies (Qiu *et al.*, 1999, 2000; Graham & Olmstead, 2000) is a core eumagnoliid clade encompassing Winterales plus Piperales and Laurales plus Magnoliales. This finding provides a phylogenetic framework for one of the most species-rich groups of basal angiosperms. Instead of the broad circumscription of the eumagnoliids (Soltis *et al.*, 2000) as comprising all angiosperms with monosulcate or monosucate-derived pollen except the basal grade and *Ceratophyllum*, this term might better be confined to the above-mentioned clade of four orders.

Remaining in flux are the positions of the Chloranthales, the eudicots (which comprise approximately 75% of angiosperm diversity) and the *Ceratophyllum*–monocot clade, even when combining six genes with a quite dense taxon sampling (Zanis *et al.*, 2002). Chloranthaceae and members of the basal grade are the only angiosperms that lack post-genital carpel fusion (Endress & Igersheim, 2000). This evidence, and the extensive occurrence of unambiguously identified chloranthaceous fossils in the lower Cretaceous (Crane *et al.*, 1995; Friis *et al.*, 1999), point to a position more basal than currently inferred from molecular data sets. Perhaps, Chloranthaceae have diverged right after the separation of the *Illicium–Schisandra–Trimenia–Austrobaileya* clade. In order to reveal possible parallelisms or reversals in structural characters and to improve robustness of the molecular-derived phylogenies, the addition of genomic regions that evolve under different functional constraints as well as the integration of information from morphology, palaeobotany, and developmental genetics are needed. Better understanding of sampling effects and of patterns of molecular evolution in conjunction with the development of algorithms that more effectively reflect the evolutionary modes of the different genomic regions used in molecular systematics will perhaps allow further progress in this area.

Conclusions

Most striking is the congruence between angiosperm phylogenies based on sequences from the noncoding *trnT-trnF*, the five (Qiu *et al.*, 2000; mitochondrial *atp1* and *matR*, chloroplast *atpB* and *rbcL*, and nuclear 18S) and six combined genes (Zanis *et al.*, 2002), and generally the three-gene (Soltis *et al.*, 1999a; Soltis *et al.*, 2000; chloroplast *atpB* and *rbcL*, and nuclear 18S RNA) and 17-gene (Graham & Olmstead, 2000) analyses. Soltis *et al.* (1999a, 2000), and Qiu *et al.* (1999, 2000) suggested that phylogenies based on combined data representing different genomes are more reliable than phylogenies based on individual genes because gene- or genome-specific bias can be largely ruled out. The strong congruence in topology as well as statistical support of major nodes between trees based on *trnT-trnF* and multiple gene/genome sequence data underscore the effectiveness of this fast-evolving, noncoding region in reconstructing phylogenies at high taxonomic levels. Low constraints (i.e. freedom for a greater number of sequence positions to vary) could result in a more equitable distribution of phylogenetic information across the region in contrast with only a few, localized, potentially variable positions; such a pattern is expected to reduce the average level of homoplasy in a genomic region. Emphasis on the utility of neutral nucleotide substitutions as phylogenetic markers is underscored in several studies at the generic and subfamilial levels (e.g. Bakker *et al.*, 2000). Further support for this concept comes from the strong agreement between *trnT-trnF* and *matK* phylogenies (Hilu *et al.*, in press). The *matK* gene also appears to be under far less selectional constraint than other genes used in phylogeny reconstruction, as is evident from the considerably higher rate of nonsynonymous substitution that is up to 26 times that of other genes and about seven-fold that of *rbcL* (Olmstead & Palmer, 1994; Hilu & Liang, 1997).

It is also important to note that the widely accepted concept of distinguishing 'slow' and 'fast' evolving genes often rather considers average amounts of variability in the genomic region under study than picturing rates at individual sites. The latter are the actual source of variability, and a slow evolving gene may have its few variable positions evolving at rates similar to the rates at most individual positions in a fast evolving gene. Consequently, expected levels of multiple hits per variable site might not necessarily differ in a genomic region in which the sites evolve slowly on average. In other words, assuming the benefits of slow evolving genes as described in the introduction may also be problematic, simply because the currently used concepts of 'slow' vs. 'fast' are an oversimplification. In fact the benefits expected for a slow evolving region (reduced incidents of multiple hits, low levels of homoplasy) might not always apply. These conceptual issues have to be further investigated, but this requires a comparative characterization of different genomic regions, which we are not attempting in this paper.

The assumption of high efficiency of functionally constrained/evolutionarily conserved DNA, in contrast with less constrained and fast-evolving DNA, for resolving deep-level phylogenies was often also based on the notion that the less constrained third codon positions are highly saturated, high in homoplasy and uninformative because of excess multiple hits (e.g. Swofford *et al.*, 1996). Nonetheless, third codon positions of *rbcL* were shown to provide most of the historical signal in analyses of bryophytes and land plants (Lewis *et al.*, 1997; Kallersjö *et al.*, 1998). A similar observation was also made by Savolainen *et al.* (2000) in the combined analysis of *atpB* and *rbcL* data for angiosperms. In a recent simulation study, Hillis (1998) increased evolutionary rates originally observed in a set of 228 angiosperm sequences by a factor of 10. He found that the amount of the tree that was inferred correctly was achieved with far fewer characters

compared with the sequences evolving at lower rates. Although the average character changed 23.6 times under the latter conditions (Hillis, 1998), signal was not obscured by homoplasy. Based on the relative numbers of supported nodes in different data partitions, Kallersjö *et al.* (1999) asserted that homoplasy can result in better recognition of groups and thus can increase phylogenetic structure. Soltis *et al.* (1999b), recognizing this phenomenon, gave higher weight to the more freely evolving loop characters in the18S rDNA in their land plant study. However, the effect of differential weighting on the phylogeny was not significant, possibly because of profound differences in rate of evolution between18S loop and stem regions (Soltis & Soltis, 1998; Soltis *et al.*, 1999b). Rapidly evolving sites may increase the chance of generating synapomorphies for particular groups without being obscured by multiple hits simply by expanding the information basis. Further, the number of sites that are co-evolving or that have a greater likelihood to change to particular nucleotides because certain amino acids are favoured by the protein secondary and tertiary structures may be lower. The latter have been identified as a reason to explain parallelisms and reversals in RuBisCo (Kellogg & Juliano, 1997).

Other noncoding regions have also been used in phylogeney reconstruction at higher taxonomic levels. However, those studies have so far only used noncoding sequences from the chloroplast inverted repeat (IR), an extremely slowly evolving region, with average rates of substitutions that are even lower than the more conserved protein-coding genes of the LSC region (Graham *et al.*, 2000). The IR regions used in such studies are the internal transcribed spacer (ITS) (Goremykin *et al.*, 1996), the *rpl2* intron, 3′-*rps12* intron, *ndhB*-intron, and spacers between 3′-*rps12* and *rps7*, and the spacer between *rps7* and *ndhB* (Graham & Olmstead, 2000; Graham *et al.*, 2000). Graham et al. (2000) found slightly higher CIs but similar RIs comparing IR protein coding and noncoding data sets. In contrast, CIs and RIs from noncoding IR regions were substantially lower than from protein coding genes outside the IR. This points to different mutational dynamics between the IR and the single copy regions of the chloroplast genome. Therefore, except for the presence of length mutations, noncoding regions in the IR may not be directly comparable with other noncoding parts such as *trnT-trnF*. A more conclusive view on the effects of noncoding vs. coding and 'fast' vs. 'slow' evolving genomic regions will require comparative studies of a larger number of functionally different genomic regions based on identical sampling schemes.

In our data set, the relaxed selection pressure across the noncoding parts of *trnT-trnF* appears to have provided an ideal situation that allowed the recovery of a robust phylogenetic structure. Detailed study of the molecular evolution of the *trnT-trnF* region is currently underway. The effectiveness of the *trnT-trnF* sequences in phylogeny reconstruction is even more evident when we consider that the average actual length of *trnT-trnF* (Table 2) is less than about 20% of the length of the five- and six-gene data sets (Qiu *et al.*, 2000; Zanis *et al.* 2002). The present data set recovered the same relationships with equal or greater support, from substantially fewer nucleotides. Therefore, the approach is considerably cheaper in terms of sequencing effort but requires much more prudence in alignment. The *trnT-trnF* data set constitutes new and strong evidence for understanding relationships among basal angiosperms, particularly within major clades. Furthermore, the results provide strong argument for the application of noncoding regions in molecular systematics at deeper levels. Utility of these genomic regions, however, should be individually tested.

The present study underscores the importance of recognizing patterns and mechanisms of molecular evolution of genomic regions used in molecular systematics to augment the probabilities of employing historic signals in phylogeny reconstruction and recovering correct organismal phylogenies. Moreover, analysis of noncoding regions is not subject to problems of differential weighting of codon positions or synonymous vs. nonsynonymous mutations that, when applied, might influence data-decisiveness in phylogenetic analyses (Davis *et al.*, 1998; Savolainen *et al.*, 2000). Thus, gene trees inferred from noncoding regions should theoretically depict a rather close approximation of the evolutionary history of the group. Our findings demonstrate that alignable noncoding regions like *trnT-trnF* can be particularly promising in phylogeny reconstruction deeper than the species and generic levels.

Acknowledgments

We thank D. L. Dilcher, L. A. Alice, S. E. Scheckler and K. Müller for comments on the manuscript, Tod Stuessy and Daniel Crawford for providing *Lactoris* DNA, and the Missouri Botanical Garden and the Arboretum of the University of California, Santa Cruz, for contributing plant material for *Piper* and *Amborella*, respectively. Supported in part by a scholarship from Studienstiftung des deutschen Volkes to T.B. Helpful comments by two anonymous reviewers are acknowledged.

Supplementary material

The following material is available from: http://www.blackwellpublishing.com/products/journals/suppmat/JEB/JEB577/JEB577sm.htm

Appendix S1

Overall sequence alignment 4707 bp in length, including the *trnT-L*-spacer, the *trnL* gene, and the *trnL-F*-spacer. Positions of the *trnL* 5′-exon are 2194–2228, and of the *trnL* 3′-exon 3330–3379. Positions of hot spots are: 256–1276 (H1), 1538–1550 (H2), 1729–1750 (H3), 1795–1927

(H4), 2720–2749 (H5), 2837–2990 (H6), 4025–4145 (H7) and 4401–4416 (H8).

Appendix S2

Character matrix of the *trnT-trnF*-region used in phylogenetic analysis of basal angiosperms (3112 characters). Hot spots H1–H8 and *trnL* exons are excluded.

References

APG. 1998. An ordinal classification for the families of flowering plants. *Ann. Missouri Botanic. Garden* **85**: 531–553.

Asmussen, C.B. & Chase, M.W. 2001. Coding and noncoding plastid DNA in palm systematics. *Am. J. Bot.* **88**: 1103–1117.

Bakker, F.T., Culham, A., Gomez-Martinez, R., Carvalho, J., Compton, J., Dawtrey, R. & Gibby, M. 2000. Patterns of nucleotide substitution in angiosperm cpDNA *trnL* (UAA)-*trnF* (GAA) regions. *Mol. Biol. Evol.* **17**: 1146–1155.

Barkman, T.J., Chenery, G., Mcneal, J.R., Lyons-Weiler, J., Ellisens, W., Moore, G., Wolfe, A.D. & Depamphilis, C.W. 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.

Bayer, R.J. & Starr, J.R. 1998. Tribal phylogeny of the Asteraceae based on two non-coding chloroplast sequences, the *trnL* intron and the *trnL/trnF* intergenic spacer. *Ann. Missouri Botanic. Garden* **85**: 242–256.

Benson, G. 1997. Sequence alignment with tandem duplication. *J. Comput. Biol.* **4**: 351–367.

Besendahl, A., Qiu, Y.-L., Lee, J., Palmer, J.D. & Bhattacharya, D. 2000. The cyanobacterial origin and vertical transmission of the plastid tRNA Leu group I intron. *Current Genetics* **37**: 12–23.

Borsch, T. 2000. *Phylogeny and Evolution of the Genus Nymphaea (Nymphaeaceae)*. PhD Thesis, University of Bonn, Bonn.

Bowe, L.M., Coat, G. & DePamphilis, C.W. 2000. Phylogeny of seed plants on all three genomic compartments: extant gymnosperms are monophyletic and gnetales' closest relatives are conifers. *Proc. Natl. Acad. Sci. USA* **97**: 4092–4097.

Bremer, K. 2000. Early Cretaceous lineages of monocot flowering plants. *Proc. Natl. Acad. Sci. USA* **97**: 4707–4711.

Cech, T.R. 1988. Conserved sequences and structures of group I introns: building an active site for RNA catalysis – review. *Gene* **73**: 259–271.

Cech, T.R., Damberger, S.H. & Gutell, R. 1994. Representation of the secondary and tertiary structure of group I introns. *Nat. Struct. Biol.* **1**: 273–280.

Chase, M.W., Soltis, D.E., Olmstead, R.G., Morgan, D., Les, D.H., Mishler, B.D., Duvall, M.R., Price, R.A., Hills, H.G., Qiu, Y.-L., *et al.* 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann. Missouri Botanic. Garden* **80**: 528–580.

Chase, M.W., Soltis, D.E., Soltis, P.S, Rudall, P.J., Fay, M.F., Hahn, W.H., Sullivan, S., Joseph, J., Molvray, M., Kores, P.J., *et al.* 2000. High level systematics of the monocotyledons: an assessment of current knowledge and a new classification. In: *Monocots: Systematics and Evolution* (K. L. Wilson & D. A. Morrison, eds), pp. 3–16. CSIRO Publishing, Collingwood, Australia.

Chaw, S.M., Parkinson, C.L., Cheng, Y., Vincent T.M. & 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.

Clegg, M.T., Gaut, B.S., Learn, G.H. & Morton, B.R. 1994. Rates and patterns of chloroplast DNA evolution. *Proc. Natl. Acad. Sci. USA* **91**: 6795–6801.

Crane, P.R. 1985. Phylogenetic analysis of seed plants and the origin of angiosperms. *Ann. Missouri Botanic. Garden* **72**: 716–793.

Crane, P.R., Friis, E.M. & Pedersen, K.R. 1995. The origin and early diversification of angiosperms. *Nature* **374**: 27–33.

Cronquist, A. 1981. *An Integrated System of Classification of Flowering Plants*. New York Botanical Garden, New York.

Davis, J.I., Simmons, M.P., Stevenson, D.W. & Wendel, J.W. 1998. Data decisiveness, data quality, and incongruence in phylogenetic analysis: an example from the monocotyledons using mitochondrial *atpA* sequences. *Syst. Biol.* **47**: 282–310.

De Pinna, M.C.C. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**: 367–394.

Donoghue, M.J. & Doyle, J.A. 2000. Seed plant phylogeny: demise of the anthophyte hypothesis? *Current Biol.* **10**: R106–R109.

Doyle, J.J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Bot.* **17**: 144–163.

Doyle, J.J. & Davis, J.I. 1998. Homology in molecular phylogenetics: a parsimony perspective. In: *Molecular Systematics of Plants II* (D. L. Soltis, P. S. Soltis & J. J. Doyle, eds), pp. 101–131. Kluwer, Boston.

Doyle, J.A. & Donoghue, M.J. 1986. Seed plant phylogeny and the origin of angiosperms: an experimental cladistic approach. *Botanic. Rev.* **52**: 321–431.

Doyle, J.A. & Endress, P.K. 2000. Morphological phylogenetic analyses of basal angiosperms:comparison and combination with molecular data. *Int. J. Plant Sci.* **161**(Suppl.): 121–153.

Endress, P.K. & Igersheim, A. 2000. Gynoecium structure and evolution in basal angiosperms. *Int. J. Plant Sci.* **161**(Suppl.): 211–213.

Erikson, T. & Wikstrom, N. 1996. *AutoDecay 3.0*. University of Stockholm, Stockholm, Sweden.

Friis, E.M., Pedersen, K.R. & Crane, P.R. 1999. Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in early cretaceous floras from Portugal. *Ann. Missouri Botanic. Garden* **86**: 259–296.

Gandolfo, M.A., Nixon, K.C. & Crepet, W.L. 2000. Monocotyledons: a review of their early cretaceous record. In: *Monocots: Systematics and Evolution* (K. L. Wilson & D. A. Morrison, eds), pp. 44–51. CSIRO Publishing, Collingwood, Australia.

Gatesy J., De Salle, R. & Wheeler, W. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylogenetics Evol.* **2**: 152–157.

Gielly, L., Yuan, Y.M., Küpfer, P. & Taberlet, P. 1996. Phylogenetic use of noncoding regions in the genus Gentiana L.: chloroplast *trnL* (UAA) intron versus nuclear ribosomal internal transcribed spacer sequences. *Mol. Phylogenetics Evol.* **5**: 460–466.

Golenberg, E.M., Clegg, M.T., Durbin, M.L., Doebley, J. & Ma, D.P. 1993. Evolution of a non-coding region of the chloroplast genome. *Mol. Phylogenetics Evol.* **2**: 52–64.

González, F. & Rudall, P. 2001. The questionable affinities of *Lactoris*: evidence from branching pattern, inflorescence

morphology, and stipule development. *Am. J. Bot.* **88**: 2143–2150.

Goremykin, V., Bobrova, V., Pahnke, J., Tritsky, A., Antonov, A. & Martin, W. 1996. Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support Gnetalean affinities of angiosperms. *Mol. Biol. Evol.* **13**: 383–396.

Gottlieb, O.R., Kaplan, M.A.C., Kubitzki, K. & Toledo Barros, J.R. 1989. Chemical dichotomies in the magnolialean complex. *Nordic J. Bot.* **8**: 437–444.

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

Graham, S.W., Reeves, P.A., Burns, A.C.E. & Olmstead, R.G. 2000. Microstructural changes in noncoding chloroplast DNA: Interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *Int. J. Plant Sci.* **161**(Suppl.): 83–96.

Gu, X. & Li, W.H. 1995. The size distribution of insertions and deletions in human and rodent pseudogenes suggests the logarithmic gap penalty for sequence alignment. *J. Mol. Evol.* **40**: 464–473.

Hamby, R.K. & Zimmer, E.A. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: *Molecular Systematics of Plants* (P. S. Soltis, D. S. Solis & J. J. Doyle, eds), pp. 50–91. Chapman & Hall, New York.

Herendeen, P.S. & Crane, P.R. 1995. The fossil history of monocotyledons. In: *Monocotelydons: Systematics and Evolution* (P. J. Rudall, P. J. Cribb, D. F. Cutler & C. J. Humphries, eds), pp. 1–21. Royal Botanic Gardens, Kew, England.

Hillis, D.M. 1998. Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* **47**: 3–8.

Hilu, K.W. & Liang, H. 1997. The *matK* gene: sequence variation and application in plant systematics. *Am. J. Bot.* **84**: 830–839.

Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainer, V., Chase, M.W., Powell, M.P., Alice, L.A., Evans, R., Sauquet, H., Neinhuis, C., Slota, T.A.B., Rohwer, J.G., Campbell, C.S. & Chatrou, L. In press. Angiosperm phylogeny based on *matK* sequence information. *Am. J. Bot.*

Hoot, S.B. & Douglas, A.W. 1998. Phylogeny of the Proteaceae based on *atpB* and *atpB-rbcL* intergenic spacer region sequences. *Aust. J. Syst. Bot.* **11**: 301–320.

Huelsenbeck, J.P. 1995. Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**: 17–48.

Jaeger, J.A., Turner, D.H. & Zucker, M. 1989. Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* **86**: 7706–7710.

Jaeger, J.A., Turner, D.H. & Zucker, M. 1990. Predicting optimal and suboptimal structure for RNA. *Meth. Enzymol.* **183**: 281–306.

Jukes, T.H. & King, J.L. 1971. Deleterious mutations and neutral substitutions. *Nature* **231**: 114–115.

Kallersjö, M., Farris, J.S., Chase, M.W., Bremer, B., Fay, M.F., Humphries, C.J., Petersen, G., Seberg, O. & Bremer, K. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants, and flowering plants. *Plant. Syst. Evol.* **213**: 259–287.

Kallersjö, M., Albert, V.A. & Farris, J.S. 1999. Homoplasy increases phylogenetic structure. *Cladistics* **15**: 91–93.

Kanno, A. & Hirai, A. 1993. A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Current Genetics* **23**: 166–174.

Kelchner, S.A. 2000. The evolution of noncoding chloroplast DNA and its application in plant systematics. *Ann. Missouri Botanic. Garden* **87**: 482–498.

Kelchner, S.A. & Clark, L.G. 1997. Molecular evolution and phylogenetic utility of the chloroplast *rpl16* intron in *Chusquea* and the Bambusoideae (Poaceae). *Mol. Phylogenetics Evol.* **8**: 385–397.

Kellogg, E.A. & Juliano, N.D. 1997. The structure and function of RuBisCo and their implications for systematic studies. *Am. J. Bot.* **84**: 413–428.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge.

Krassilov, V.A. 1997. *Angiosperm Origins: Morphological and Ecological Aspects*. Pensoft, Sofia, Bulgaria.

Kuhsel, M.G., Strickland, R. & Palmer, J.D. 1990. An ancient group-I intron shared by bacteria and chloroplasts. *Science* **250**: 1571–1573.

Les, D.H. & Schneider, E.L. 1995. The Nymphaeales, Alismatidae and the theory of an aquatic monocotyledon origin. In: *Monocotelydons: Systematics and Evolution* (P. J. Rudall, P. J. Cribb, D. F. Cutler & C. J. Humphries, eds), pp. 23–42. Royal Botanic Gardens, Kew, England.

Levinson, G. & Gutman, G. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**: 203–221.

Lewis, L.A., Mishler, B.D. & Vilgalys, R. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylogenetics Evol.* **7**: 377–393.

Liang, H. & Hilu, K.W. 1996. Application of the *matK* gene sequences to grass systematics. *Can. J. Bot.* **74**: 125–134.

Maddison, W.P. & Maddison, D.R. 1997. *MacClade 3.07. Analysis of Phylogeny and Character Evolution (computer program)*. Sinauer Associates, Sunderland, Massachusetts.

Mathews, D.H., Zucker, M. & Turner, D.H. 2001. *RNAstructure 3.6 (computer program)*www.bioinfo.math.edu.

Mathews, S. & Donoghue, M.J. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* **286**: 947–950.

Mathews, S. & Donoghue, M.J. 2000. Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. *Int. J. Plant Sci.* **161**(Suppl.): 41–55.

McClure, M.A., Vasi, T.K. & Fitch, W.M. 1994. Comparative analysis of multiple protein sequence alignment methods. *Mol. Biol. Evol.* **11**: 571–592.

Michel, F. & Westhof, E. 1990. Modeling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J. Mol. Biol.* **216**: 585–610.

Moritz, C. & Hillis, D.M. 1996. Molecular systematics: context and controversies. In *Molecular Systematics* (D. M Hillis, C. Moritz & B. K. Mable, eds), 2nd edn, pp. 1–13. Sinauer Associates, Sunderland, Massachusetts.

Morton, B.R. & Clegg, M.T. 1993. A chloroplast DNA mutational hotspots and gene conversion in a non-coding region near *rbcL* in the grass family (Poaceae). *Current Genetics* **24**: 357–365.

Müller, J.J. 2000. *QuickAlign 1.5.5. (computer program)*. Available from www.uni-bonn.de/~uzsfjj/

Neinhuis, C., Borsch, T. & Hilu, K.W. 1999. *Phylogenetic relationships within Aristolochiaceae based on trnT-trnF* sequences. Abstracts XVI. Abstract number 351. International Botanical Congress, St. Louis, Missouri.

Olmstead R.G. & Palmer, J.D. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *Am. J. Bot.* **81**: 1205–1224.

Qiu, Y.-L., Chase, M.W., Les, D.H. & Parks, C.R. 1993. Molecular phylogenetics of the magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. *Ann. Missouri Botanic. Garden* **80**: 587–606.

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

Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D.E., Soltis, P.S., Zanis, M., Zimmer, E.A., Chen, Z., Savolainen, V. & Chase, M.W. 2000. Phylogeny of basal angiosperms: analyses of five genes from three genomes. *Int. J. Plant Sci.* **161**(Suppl.): 3–27.

Reeves, G., Chase, M.W., Goldblatt, P., Rudall, P., Fay, M.F., Cox, A.V., Lejeune, B. & Souza-Chies, T. 2001. Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *Am. J. Bot.* **88**: 2074–2087.

Renner, S.S. 1999. Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. *Am. J. Bot.* **86**: 1301–1315.

Richardson, J.E., Fay, M.F., Cronk, Q.C.B., Bowman, D. & Chase, M.W. 2000. A phylogenetic analysis of Rhamnaceae using *rbcL* and *trnL-F* plastid DNA sequences. *Am. J. Bot.* **87**: 1309–1324.

Sang, T., Crawford, D.J. & Stuessy, T.F. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae). *Am. J. Bot.* **84**: 1120–1136.

Sauquet, H., Doyle, J.A., Scharaschkin, T., Borsch, T., Hilu, K.W., Chatrou, L.W. & Le Thomas, A. In press. Phylogenetic analysis of Magnoliales and Myristicaceae based on Multiple data sets: implications for character evolution. *Botanic. J. Lim. Soc.*

Savolainen, V., Chase, M.W., Hoot, S.B., Morton, C.M., Soltis, D.E., Bayer, C., Fay, M.F., De Bruijn, A.Y., Sullivan, S. & Qiu, Y.-L. 2000. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Syst. Biol.* **49**: 306–362.

Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwomgse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.-Y. Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. & Sugiura, M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome. *Plant Mol. Biol. Reporter* **4**: 110–147.

Simmons, M.P. & Ochoterena, H. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Syst. Biol.* **49**: 369–381.

Small, R.L., Ryburn, J.A., Cronn, R.C., Seelanan, T. & Wendel, J.F. 1998. The tortoise and the hare: choosing between non-coding plastome and *Adh* sequences for phylogeny reconstruction in a recently diverged plant group. *Am. J. Bot.* **85**: 1301–1305.

Soltis, D.E. & Soltis, P.S. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. In: *Molecular Systematics of Plants II* (D.L. Soltis, P.S. Soltis & J.J. Doyle, eds), pp. 1–42. Kluwer, Boston.

Soltis, P.S., Soltis, D.E. & Chase, M.W. 1999a. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402**: 402–404.

Soltis, D.E., Soltis, P.S., Chase, M.W., Mort, M.E., Albach, D.C., Zanis, M., Savolainen, V., Hahn, W., Hoot, S.B., Fay, M.F., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.L., Nixon, K.C. & Farris, J.S. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanic. J. Linn. Soc.* **133**: 381–461.

Soltis, D.E., Soltis, P.S., Nickrent, D.L., Johnson, L.A., Hahn, W.J., Hoot, S.B., Sweere, J.A., Kuzoff, R.K., Kron, K.A., Chase, M.W., *et al.* 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. Missouri Botanic. Garden* **84**: 1–49.

Soltis, P.S., Soltis, D.E., Wolf, P.G., Nickrent, D.L., Chaw, S.-M. & Chapman, R.L. 1999b. The phylogeny of land plants inferred from 18S rDNA sequences: pushing the limits of rDNA signal? *Mol. Biol. Evol.* **16**: 1774–1784.

Swofford, D.L. 2001. *PAUP: Phylogenetic Analysis Using Parsimony, version 4.0b6 (computer program)*. Sinauer, Sunderland.

Swofford, D.L. & Olsen, G.J. 1990. Phylogeny reconstruction. In: *Molecular Systematics* (D. M. Hillis & C. Moritz, eds), pp. 411–501. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D.L., Olsen, G.J., Waddell, P.J. & Hillis, D.M. 1996. Phylogenetic inference. In: *Molecular Systematics* (D. M. Hillis, C. Moritz & B. K. Mable, eds), 2nd edn, pp. 407–514. Sinauer Associates, Sunderland, Massachusetts.

Taberlet, P., Gielly, L., Pautou, G. & Bouvet, J. 1991. Universal primers for amplification of three non-coding chloroplast regions. *Plant Mol. Biol.* **17**: 1105–1109.

Takhtajan, A. 1980. Outline of the classification of flowering plants (Magnoliophyta). *Botanic. Rev.* **46**: 225–359.

Thorne, J.L., Kishino, H. & Felsenstein, J. 1992. Inching towards reality: an improved likelihood model of sequence evolution. *J. Mol. Evol.* **34**: 3–16.

Tsudzuki, J., Nakashima; K., Tsudzuki, T., Hiratsuka, J., Shibata, M., Wakasugi, T. & Sugiura, M. 1992. Chloroplast DNA of black pine retains a residual inverted repeat lacking rRNA genes: nucleotide sequences of *trnQ*, *trnK*, *psbA*, *trnI* and *trnH* and the absence of *rps16*. *Mol. Gen. Genet.* **232**: 206–214.

Van Ham, R.C.H.J., T'hart, H., Mes, T.H.M. & Sandbrink, J.M. 1994. Molecular evolution of non-coding regions of the chloroplast genome in the Crassulaceae and related species. *Current Genet.* **25**: 558–566.

Winter, K.U., Becker, A., Münster, T., Kim, J.T., Saedler, H. & Theissen, G. 1999. MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. *Proc. Natl. Acad. Sci. USA* **96**: 7342–7347.

Zanis, M., Soltis, D.E., Soltis, P.S., Mathews, S. & Donoghue, M.J. 2002. The root of angiosperms revisited. *Proc. Natl. Acad. Sci. USA* **99**: 6848–6853.

Zucker, M. 1994. Prediction of RNA secondary structure by energy minimization. In: *Computer Analysis of Sequence Data* (A. M. Griffin & H. G. Griffin, eds) **25**(II), pp. 267–294. CRC Press, Totowa, NJ.

Zucker, M., Mathews, D.H. & Turner, D.H. 1999. Algorithms and thermodynamics for RNA secondary structure prediction: a practical guide. In: *RNA Biochemistry and Biotechnology* (J. Barciszewski & B. F. C. Clark, eds). NATO ASI Series, Kluwer Academic Publishers.

Received 20 June 2002; revised 6 September 2002; accepted 14 March 2003

Int. J. Plant Sci. 162(5):1111–1140. 2001.

1058-5893/2001/16205-0012$03.00

THE FLOWERS IN EXTANT BASAL ANGIOSPERMS AND INFERENCES ON ANCESTRAL FLOWERS

Peter K. Endress[1]

Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zurich, Switzerland

This is a combination of a review and original data on floral structure, development, and biology of representatives of all families of the ANITA grade and, in addition, Chloranthaceae and Ceratophyllaceae. Since the ANITA grade has been identified as the basalmost grade of extant angiosperms based on molecular studies by a number of authors, it has become possible to search for potential plesiomorphies among flowers of extant basal angiosperms. They may include the following traits: flowers small, pollination by small insects (dipters, thrips, moths); flowers with moderate or low number of floral organs, in spiral (or whorled) arrangement, with a tendency to form organ series in Fibonacci numbers (3, 5, 8); flowers bisexual (but easily becoming unisexual because of low level of synorganization between organs), protogynous; tepals (in spiral flowers) with gradual transitions between bractlike, sepal-like, and petal-like forms; stamens with short filaments, anthers with a connective tip, with more or less bulging disporangiate thecae; thecae opening by a longitudinal slit and not by valves. Carpels free, styleless, extremely ascidiate, with one or only few anatropous ovules, sealed by secretion and not by postgenital fusion; stigmas wet, with multicellular protrusions. Among members of the ANITA grade, there is a trend to form extragynoecial compita. In those taxa with the relatively most complicated gynoecium architecture (including an extragynoecial compitum), there is a concomitant trend to have less strongly ascidate to almost plicate carpels (Nymphaeaceae, Schisandraceae, Illiciaceae).

Keywords: Amborellaceae, ANITA grade, Austrobaileyaceae, basal angiosperms, Cabombaceae, Ceratophyllaceae, Chloranthaceae, floral biology, floral development, floral structure, Illiciaceae, Nymphaeaceae, Schisandraceae, Trimeniaceae.

Introduction

The group formerly known as magnoliids (Magnoliidae) was long thought to represent the basalmost extant angiosperms (e.g., Cronquist 1988). It comprises ca. 35 families (in the circumscription of Takhtajan [1997] and with the inclusion of his Nymphaeidae). However, the magnoliids are extremely diverse in structure, and until recently, it was uncertain which groups of the magnoliids were the basalmost. Concomitantly, the large flowers of *Magnolia*, with a high number of floral organs on a long floral axis (or medium-sized flowers of Winteraceae; cf. Gottsberger 1974), were taken as the prototype for primitive flowers, as shown in textbooks up to the 1970s.

In the 1980s, the situation changed with the recovery of many exquisitely preserved flower fossils from the Lower Cretaceous, the time of early diversification of the angiosperms (Friis et al. 1986). These early flowers were all surprisingly small, even tiny, on the order of 1 mm. One family, the Chloranthaceae, was especially well represented among these fossils. They supported earlier findings on Chloranthaceae-like pollen as the earliest, well-recognizable pollen of flowering plants (Couper 1958; Kemp 1968; Doyle 1969, 1977; Muller 1970; Doyle et al. 1977; Hughes et al. 1979). Extant Chloranthaceae have tiny flowers and very few floral organs.

A new perspective opened up with the landmark work by Chase et al. (1993) on the phylogeny of seed plants based on the *rbc*L gene and 500 species of seed plants, in which *Ceratophyllum* appeared as the basalmost angiosperm. *Ceratophyllum* also has tiny flowers. Furthermore, this study supported earlier structural cladistic studies that the magnoliids (and therefore also the dicotyledons) do not constitute a monophyletic group (Donoghue and Doyle 1989; Doyle and Hotton 1991).

The latest breakthrough came at the Sixteenth International Botanical Congress in St. Louis, at the August 3, 1999, symposium, where four more or less independent groups of authors all reached the same conclusion, that the genus *Amborella* is the basalmost clade among extant angiosperms, followed by Nymphaeales and a clade comprising Austrobaileyacae, Trimeniaceae, and Illiciales; this basal grade was termed the ANITA grade (Mathews and Donoghue 1999, 2000; Qiu et al. 1999, 2000, 2001; Soltis et al. 1999, 2000*a*; Graham and Olmstead 2000; Graham et al. 2000). Additional support came from Parkinson et al. (1999), Renner (1999) (with Chloranthaceae basalmost), Borsch et al. (2000), and Barkman et al. (2000) (with *Amborella* and Nymphaeales basalmost). These studies also corroborate the view that the magnoliids are highly paraphyletic. This unanimous result is especially remarkable because different approaches were used by these groups of authors. Important in all these studies is the use of multiple genes and a larger sampling of taxa than before. The results seem to be better supported than those of all former studies. In some earlier studies, Nymphaeales (Hamby and

[1] E-mail pendress@systbot.unizh.ch.

Manuscript received January 2001; revised manuscript received March 2001.

Zimmer 1992; Doyle et al. 1994) and the ANITA grade (Soltis et al. 1997) also appeared at the base, but these topologies were less well supported and therefore were not discussed so widely at the time, or members of the ANITA grade came out together as a clade but not as earliest branching angiosperms (Chase et al. 1993; Qiu et al. 1993).

In the premolecular era, it was tentatively suggested that a group consisting of Chloranthaceae, Trimeniaceae, and *Amborella*, among other groups, might be especially primitive in angiosperms (Endress 1986). This assumption was based on the early and widespread fossil record of Chloranthaceae (see above) and the floral similarities between the three families (see also Endress 1987*a*). Upchurch (1984, p. 549) argued that "groups such as Amborellaceae, Austrobaileyaceae, Schisandraceae, and certain Chloranthaceae would be more primitive in their stomatal structure than Magnoliales," based on Lower Cretaceous leaf remains.

In a combination of the three gene analysis by Soltis et al. (1999, 2000*a*) and a morphological analysis that is more detailed than previous ones, the ANITA grade also comes to the base of the tree, if *Amborella* is chosen as sister to all other angiosperms (Doyle and Endress 2000). In contrast to the gene trees alone, Chloranthaceae immediately follows the ANITA grade. In a tree based on morphology alone, Chloranthaceae goes even more toward the base between *Amborella* and Trimeniaceae.

This new and much better supported phylogenetic framework can be taken as a basis to do evolutionary studies (Doyle and Endress 2000; Kuzoff and Gasser 2000; Soltis et al. 2000*b*). We can extrapolate better about what early angiosperms were like and how they evolved. However, fossils and all of the methods we had before are now even more important to improve the picture on early angiosperm evolution. It is best to use as much evidence as possible, and it is important to do an evolutionary evaluation not only on the basis of cladograms available but also with consideration of the biological context of traits.

It is striking that six of the seven families of the ANITA grade have only one or two genera, and two families even have only a single species. Most of these families are geographically very scattered in the Tropics. The two monotypic families (Amborellaceae and Austrobaileyaceae) have very small relic areas and are therefore especially threatened by extinction. They are woody or herbaceous, never large trees; a number of them are scrambling or viny plants (Feild et al. 2000). The latter is true for *Amborella*, *Austrobaileya*, part of Trimeniaceae, and Schisandraceae. Ceratophyllaceae also have a single genus.

What is shared by all those clades thought to be basalmost at one time or another—*Amborella*, Chloranthaceae, and *Ceratophyllum*—is that the flowers are extremely small, in the magnitude of a few millimeters. Thus, this aspect of our concept of ancestral flowers has not changed since the 1980s.

These phylogenetic hypotheses on basal angiosperms are supported by ever more fossil finds of these groups from the Lower Cretaceous (leaves reminiscent of taxa of the ANITA grade [Upchurch 1984], Nymphaeaceae-like plants and flowers [Mohr and Friis 2000; Friis et al. 2001], seeds reminiscent of *Illicium* or Nymphaeales [Friis et al. 2000], *Amborella*-like pollen [Hughes and McDougall 1987; Doyle and Endress 2000], and Chloranthaceae-like pollen and flowers [Walker and Walker 1984; Friis et al. 1986, 1999, 2000; Crane et al. 1989; Brenner 1996; Eklund 1999]).

This publication is a combination of original material and a review. Most of the illustrations are new. The flowers of the families considered are not described in detail, but features of special interest for this comparison are pointed out. Such a comparison seems particularly appropriate because the author investigated flowers of most relevant families in earlier studies. Furthermore, comparison with earlier attempts of syntheses shows the progress made in the past 15 yr (Endress 1986, 1990*a*, 1994*c*).

Material and Methods

The following taxa and collections were used for this study (E: collected by P. K. Endress; BGZ: Botanic Garden of the University of Zurich; in cases in which more than one collection was used, each is indicated for the respective figures). For scanning electron microscopy, specimens fixed in FAA or ethanol were critical-point dried and sputter-coated with gold.

Amborellaceae

Amborella trichopoda Baill.; male flowers: E s.n., coll. 1983 (figs. 2*A*–2*C*, 14*C*), grown from fruits received by H. S. MacKee 38408, cult. BGZ; female flowers: cult. Botanical Garden University of California, Santa Cruz (received via H. Tobe) (fig. 2*D*, 2*E*); H. S. MacKee 38909 (fig. 2*F*; fig. 14*A*, 14*B*), New Caledonia.

Austrobaileyaceae

Austrobaileya scandens C. T. White; E 9083, (fig. 1*C*), grown from fruits, collected on Mt. Lewis, Queensland, cult. BGZ; E 4218, (figs. 1*D*, 4*F*), Boonjee, Queensland; B. Gray 2044, Boonjee, Queensland (fig. 4*A*–4*E*).

Cabombaceae

Cabomba furcata Schult. & Schult. f.; E 00-58, cult. BGZ.

Ceratophyllaceae

Ceratophyllum demersum L.; E 4836 (fig. 13*A*), E 9851 (fig. 13*B*, 13*C*, 13*E*, 13*F*), E 5853 (fig. 13*D*), cult. BGZ.

Chloranthaceae

Sarcandra chloranthoides Gardner; E 5291, grown from fruits collected by K. U. Kramer 6582 in Kerala, India, cult. BGZ.

Illiciaceae

Illicium anisatum L.; E 535 (fig. 7*H*), cult. Brione, Switzerland; E 2681 (fig. 7*C*), E 4536 (fig. 7*A*, 7*B*), E 00-60 (fig. 7*D*–7*G*), cult. Isole di Brissago, Switzerland.

Illicium floridanum J. Ellis; E 7519, cult. BGZ.

Nymphaeaceae

Nuphar advena Aiton; E 00-61, cult. BGZ.

Victoria cruziana A.D. Orb.; E 9983 (fig. 11*A*–11*D*), E 96-

Fig. 1 Anthetic flowers of representatives of all families of the ANITA grade. *A*, *Amborella trichopoda*; male inflorescence (cultivated specimen). *B*, *Trimenia neocaledonica*; left, three bisexual flowers with the stamens fallen; middle, flower bud; right, male flower at anthesis (natural habitat). *C*, *D*, *Austrobaileya scandens*. *C*, Flower in female phase (cultivated specimen). *D*, Flower in male phase, with visiting fly having a pollen load on its back (natural habitat). *E*, *Kadsura japonica*; left, female flower; middle, intermediate flower; right, male flower (cultivated specimen). *F*, *Illicium floridanum*; flower in male phase (cultivated specimen). *G*, *Cabomba furcata*; flowers in male phase (cultivated specimen). *H*, *Nuphar advena*; flower in male phase (cultivated specimen).

57 (fig. 10*A*, 10*C*, 10*E*; fig. 11*E*, 11*F*), E 98-150 (fig. 10*B*, 10*D*, 10*F*), cult. BGZ.

Schisandraceae

Kadsura japonica Benth.; E s.n., s. dat., cult. old BGZ.

Schisandra chinensis Baill.; E 99-59 (fig. 5*A*–5*F*; fig. 6*A*, 6*B*), E 99-67 (fig. 5*G*), cult. BGZ.

Trimeniaceae

Piptocalyx moorei Oliv.; E 4367, New South Wales, Australia.

Trimenia neocaledonica Baker f.; E 6316, New Caledonia.

Trimenia papuana Ridl.; E 4066 (fig. 14*D*), E 4087 (fig. 3*A*–3*F*), Papua New Guinea.

Results

Amborellaceae

Amborellaceae are a monotypic family, endemic to moist forests in New Caledonia (Jérémie 1982). *Amborella trichopoda* is a small, scrambling tree, quite inconspicuous. The flowers were little known for a long time, apart from an anatomical study based on herbarium material by Bailey and Swamy (1948). Only recently they were studied with fixed material (Endress and Igersheim 2000*b*; for the gynoecium, see also Endress 1986 and Endress and Igersheim 1997*a*, 2000*a*). The flowers are small, only ca. 3–5 mm in diameter, whitish or cream in color, with a green ovary (fig. 1*A*). They are functionally unisexual and dioecious. Duration of anthesis of an individual flower is unknown. Collett (1999) noted nocturnal scent and moth visits in a cultivated specimen. Preliminary observations in the field by T. Feild (personal communication) suggest that wind plays a role in pollination. Floral phyllotaxis is spiral, which is easily seen in mature flowers. The divergence angles between subsequent organs are regularly ca. 138° and thus represent the most common pattern of spiral phyllotaxis (Fibonacci pattern) (fig. 2*A*–2*D*, 2*F*). Floral organ number is quite variable. The pollination organs are preceded by several tepals (fig. 2*C*, 2*F*). The outermost organs are small scale. A surprising feature is that the floral organs are arranged on a flat, expanded floral base. This floral base is cuplike in bud and then irregularly tears when the flower expands. In this respect, it resembles flowers of some Monimiaceae (Endress 1980*b*; Endress and Igersheim 2000*b*).

In male flowers, there are about nine to 11 tepals (the two prophylls are not counted) and ca. 12–21 stamens (fig. 2*C*). Prophylls are the first two organs of a lateral flower, which are usually not incorporated in the floral architecture; in *Amborella* they are commonly small scales on the peduncle. In some flowers, there is an undifferentiated pyramidal body in the center, which may be an undifferentiated gynoecium or just a remnant of the floral apex (fig. 14*C*). Tepals and stamens are arranged on a floral cup. The open flower presents the inner surface of the floral cup and the stamens, whereas the tepals are reflexed and partly hidden when the flower is viewed from above (Endress and Igersheim 2000*b*). The stamens have a short, broad, flat filament, which elongates during anthesis. The anther is triangular, with a short connective tip, which may be secretory (fig. 2*C*). The two thecae are disporangiate, strongly introrse, and bulging. Each theca opens by a longitudinal slit. Tepals (also in female flowers) and stamens are served by a single vascular trace (Endress and Igersheim 2000*b*).

Female flowers are slightly smaller than male. They contain about seven or eight tepals (prophylls not counted) and mostly five carpels. In addition, most female flowers have one or two staminodes; they look like stamens but they are sterile (fig. 2*F*). Thus, these flowers are structurally bisexual, although functionally unisexual.

The carpels are barrel-shaped, slightly stipitate, and strongly ascidiate (fig. 14*A*). They have a large caplike stigma that surrounds the orifice of the inner space and has irregular multicellular papillae. On carpels from a young floral bud, one can see the urn-shaped structure; there is only a small entrance near the apex (fig. 2*E*). Each carpel has a single, ventral-median, almost orthotropous, pendant ovule with two integuments (Endress and Igersheim 1997*a*, 2000*b*). Tobe et al. (2000) describe the ovule as hemianatropous; however, the curvature is restricted to the base of the ovule and is not expressed in a strong asymmetry of the integuments. In transverse sections of a carpel at anthesis at the level of maximal seclusion of the inner space, the inner space is a narrow slit. It is not postgenitally fused, but secretion can be seen inside in several lacunae (angiospermy type 1; Endress and Igersheim 1997*a*). The surface of the ovary is somewhat furrowed by the appressed inner stamens in bud, and the ridges in between are covered with uni- or bicellular, upward-directed hairs (fig. 14*B*). This is also the case in the central pyramidal body in male flowers (fig. 14C) (see above).

Floral development is surprisingly difficult to study. When the floral organs are initiated, the flowers are still extremely small, but the organs are crowded and hidden in the concave floral base from very early on. We were not able to dissect the youngest flowers without destroying the floral center.

Trimeniaceae

Trimeniaceae comprise one or two genera (*Trimenia*, *Piptocalyx*) and eight species, occurring in tropical and subtropical rain forests, scattered in the Western Pacific region (Wagner and Lorence 1999). They are shrubs, little trees, or vines. Of all families of the ANITA grade, the flowers of Trimeniaceae are most similar to those of *Amborella*. The first comparative morphological account on the family was given by Money et al. (1950) based on herbarium material; this was followed by a study based on field observations and fixed material (Endress and Sampson 1983).

Flowers are bisexual and male in monoecious distribution; female flowers were not found (Endress and Sampson 1983). It is not known whether they are self-compatible or self-incompatible. Bisexual flowers are protogynous (this study). When the bud opens, the stigma becomes exposed before the stamens open (fig. 3*H*). The flowers are ca. 5–7 mm in diameter. As in *Amborella*, the flowers are whitish or cream with a green ovary (fig. 1*B*), and the floral organs are spirally arranged in a Fibonacci pattern (divergence angles of ca. 138°) and variable in number (Endress and Sampson 1983; this study) (fig. 3*A*–3*D*). Organ numbers are tepals 0–21 (the two

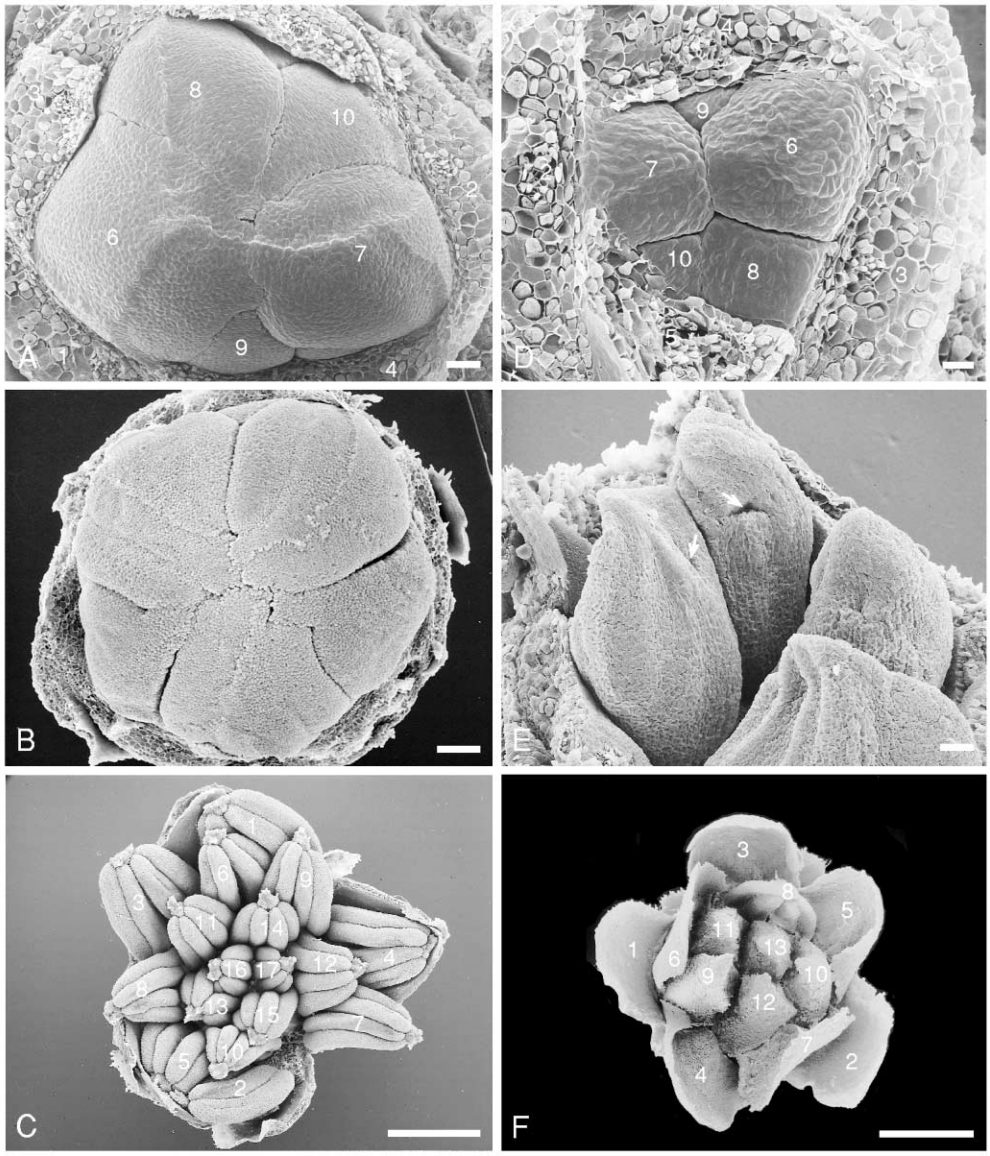


Fig. 2 *Amborella trichopoda* (Amborellaceae). *A–C*, Male flowers. *A*, Bud; outer tepals removed; subsequent organs numbered, beginning with the outermost organ on the figure. *B*, Bud; tepals removed. *C*, Flower at early anthesis; subsequent stamens numbered. *D–F*, Female flowers. *D*, Bud; outer tepals removed; subsequent organs numbered, beginning with the outermost organ on the figure. *E*, Carpels of a bud; arrows point to entrance into the ascidiate carpels. *F*, Flower at early anthesis; subsequent organs numbered, beginning with outermost tepal that is well visible on the figure (*1–7*, inner tepals; *8*, staminode; *9–13*, carpels). Magnification bars: *A*, *D*, *E* = 50 μm; *B* = 0.2 mm; *C*, *F* = 1 mm.

prophylls not counted) and stamens 7–16; but there is only one carpel (Endress and Sampson 1983). Another difference is that there is no floral cup, and tepals and stamens are caducous. Duration of anthesis of an individual flower, pollination biology, and breeding systems are unknown. Lack of floral secretions (except for the stigma), the relatively large size of the stigma, and andromonoecy suggest that wind may at least partly play a role in pollination. In *Trimenia papuana*, the flowers have no noticeable scent; however, in *Piptocalyx moorei* a faint cinnamon-like scent and visits by bees were observed (P. K. Endress, personal observations), although bees may not be the primary pollinators.

The spathulate tepals fall off when the flower opens (hence the name *Piptocalyx*!). Thus, they do not take part in optical attraction of the open flowers. Stamens have an anther with a short connective protrusion, which seems to be nonsecretory, and a filament of about the length of the anther (or shorter). The anther has two extrorse to slightly introrse disporangiate thecae (fig. 3*H*). The thecae are not or only slightly bulging. Each theca opens by a longitudinal slit. Tepals and stamens have a single vascular trace (Endress and Sampson 1983). Transitional forms between tepals and stamens were found (Endress and Sampson 1983).

The carpels are greatly reminiscent of those of *Amborella*,

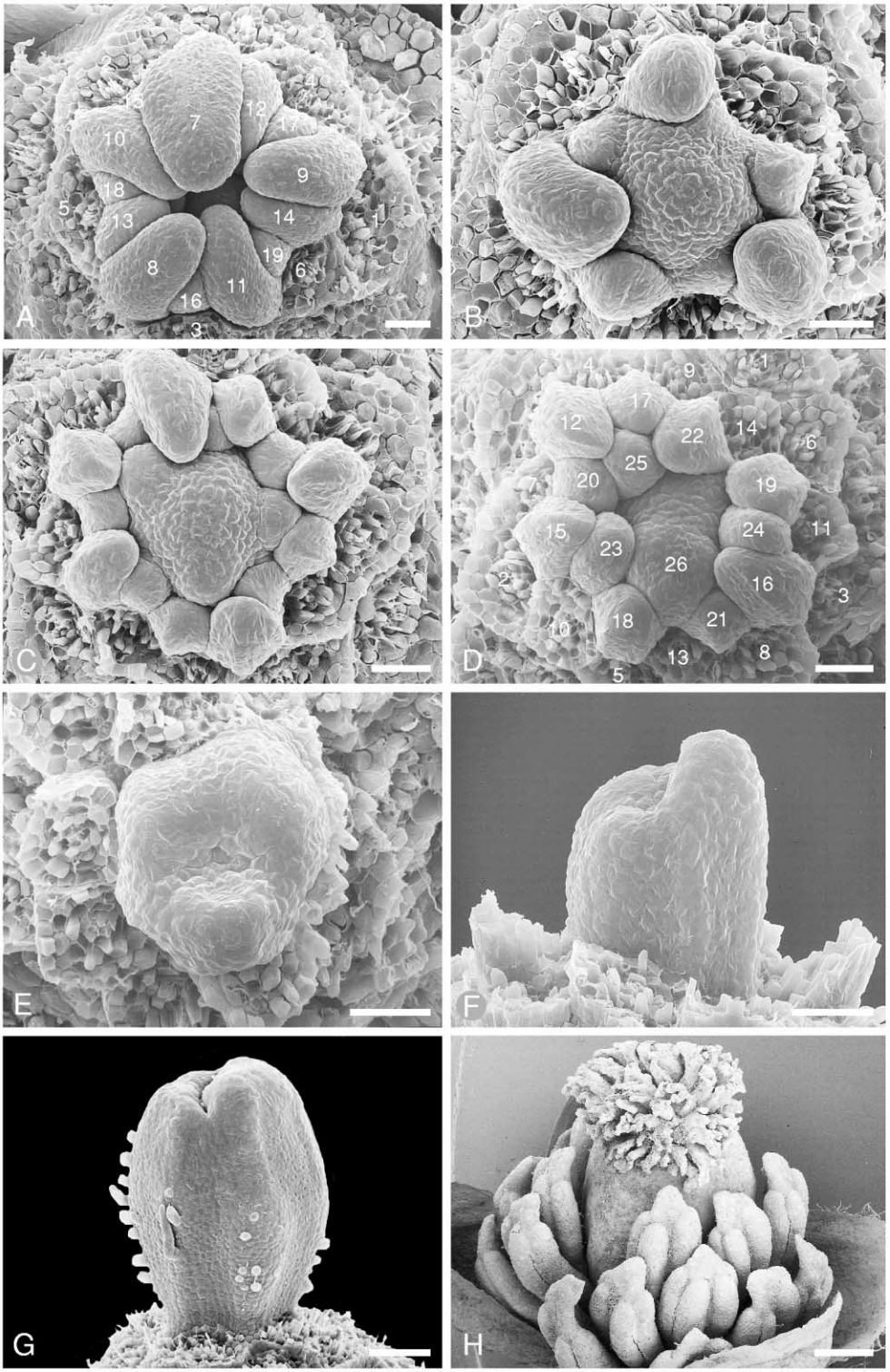


Fig. 3 Trimeniaceae. *A–G*, *Trimenia papuana*; flowers with outer organs partly removed, before anthesis. *A*, Flower with inner tepals. *B*, Flower with innermost tepals and floral apex. C, Flower with stamens and floral apex. *D*, Flower with stamens and incipient carpel (organ 26). *E*, *F*, Young carpel. *E*, From above. *F*, From the side. *G*, Older carpel (from the side). *H*, *Piptocalyx moorei*; flower at female stage of anthesis. *A*, *D*, Subsequent floral organs numbered, beginning with the outermost organ scar on the figure. Magnification bars: *A–F* = 50 μm; *G* = 0.1 mm; *H* = 0.5 mm.

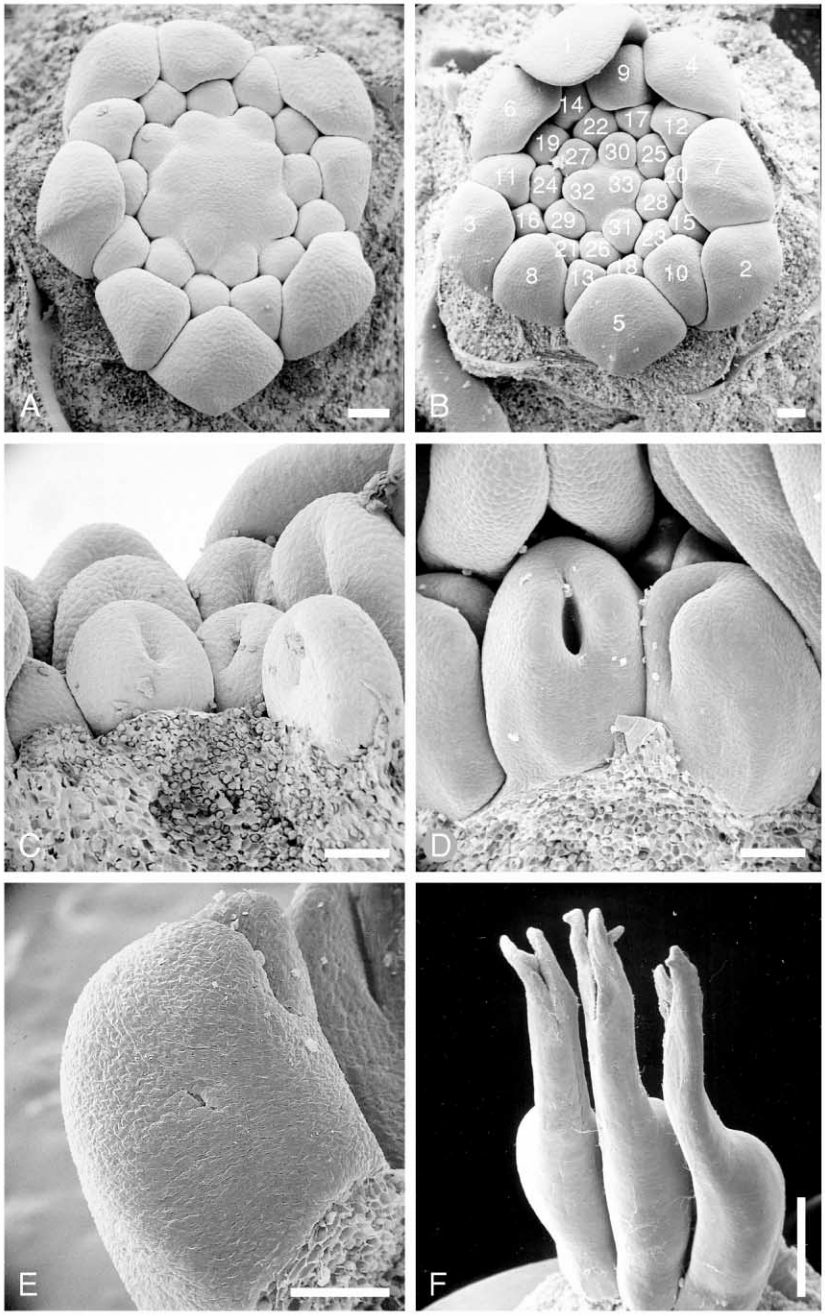


Fig. 4 *Austrobaileya scandens* (Austrobaileyaceae). *A*, Young flower bud with stamens and staminodes. *B*, Young flower bud with all organs formed; subsequent floral organs numbered, beginning with outermost unremoved organ. *C*, Young staminodes and young carpels with incipient ascidiate shape. *D*, Slightly older carpels, with ascidiate part more elongate. *E*, Carpel with transverse tips beginning to form. *F*, Carpels shortly before anthesis. Magnification bars: *A–E* = 0.1 mm; *F* = 1 mm.

barrel shaped, stipitate, and extremely ascidiate, with a similar large stigma that surrounds the orifice of the inner space and has irregular multicellular, multiseriate protrusions, and with a single, ventral-median pendant ovule with two integuments; the ovary surface is somewhat furrowed by the appressed inner stamens in bud, and the ridges are covered with three-cellular strigose hairs, with a long, tanniferous terminal cell (fig. 3*G*, 3*H*; fig. 14*D*). The ovule is, however, anatropous, and not orthotropous (Endress and Sampson 1983). As in *Amborella*, the inner space of the carpel is a narrow, unfused slit with secretion (angiospermy type 1; Endress and Igersheim 1997*a*).

Because the floral apex is convex, early floral development can easily be studied. Although the floral organ primordia are relatively small (narrow) as compared to the floral apex, other

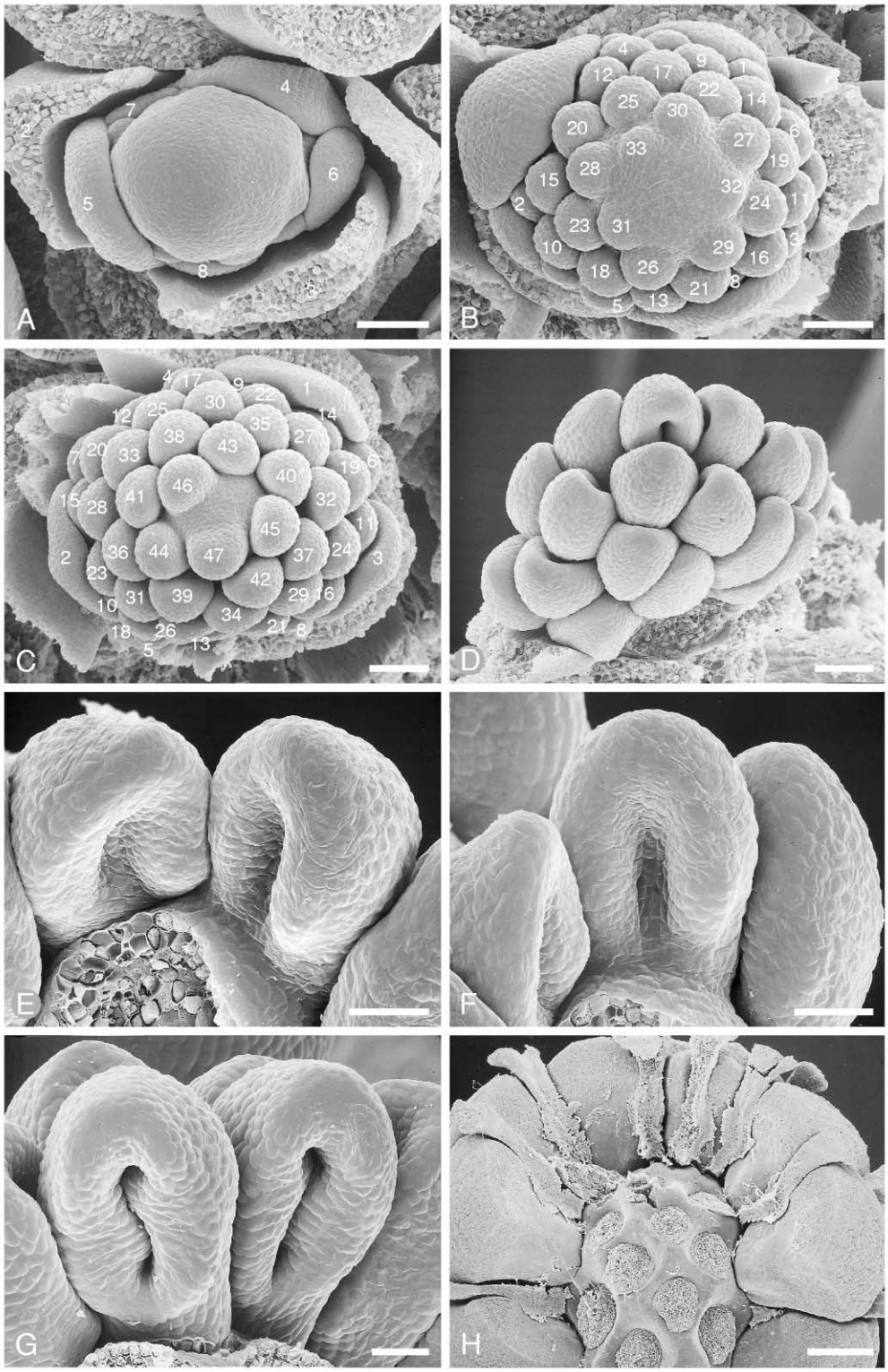


Fig. 5 Schisandraceae. *A–G*, *Schisandra chinensis*; female flowers. *A*, Young flower bud with first carpels initiated. *B*, Young floral bud with most carpels initiated. *C*, Young floral bud with all organs initiated. *D*, Young gynoecium (from the side). *E*, Carpels of young gynoecium; ascidiate base not yet developed. *F*, Same as *E* (one carpel in frontal view). *G*, Somewhat older carpels with ascidiate part present. *A–C*, Subsequent floral organs numbered, beginning with the outermost possible organ. *H*, *Kadsura japonica*; female flower at anthesis, showing gynoecium with carpels removed on one side; stigmas with secretion. Magnification bars: *A–D* = 0.1 mm; *E–G* = 50 μm; *H* = 1 mm.

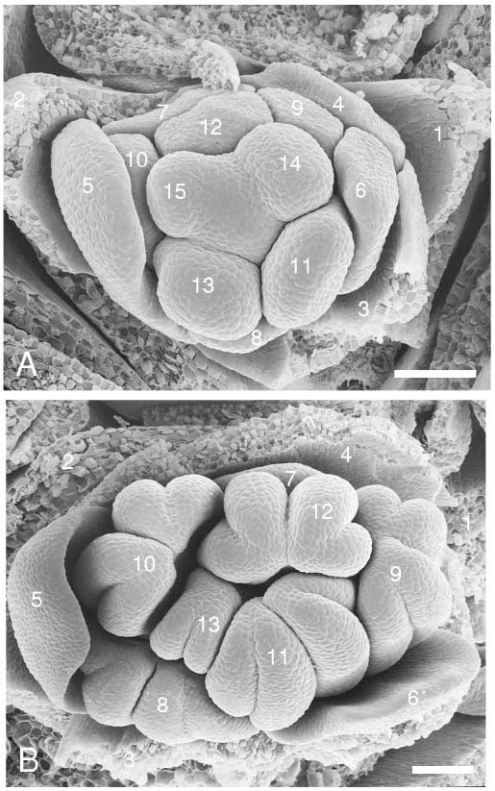


Fig. 6 *Schisandra chinensis* (Schisandraceae); male flowers. *A*, Young bud, just after completion of organ formation. *B*, Older bud with anthers differentiated. *A*, *B*, Subsequent organs numbered, beginning with outermost organ on the figure. Magnification bars = 0.1 mm.

phyllotactic patterns than the Fibonacci pattern were not found (Endress and Sampson 1983; this study). The single carpel is initiated in the same divergence angle from the last formed stamen (fig. 3*D*). The carpel primordium soon becomes chair-like, and then the entrance is lifted up by extensive intercalary elongation of the carpel (fig. 3*E*–3*G*).

Austrobaileyaceae

Austrobaileyaceae are a monotypic family in tropical rain forests of a small area in northeastern Australia (Bailey and Swamy 1949; Endress 1980*c*); *Austrobaileya scandens* plants are vines. The flowers are larger than in the previous families, ca. 5 cm in diameter. They are bisexual and protogynous (Endress 1980*c*) (fig. 1*C*, 1*D*). *Austrobaileya* is self-incompatible (Prakash and Alexander 1984; P. K. Endress, personal observation). Organ numbers are 19–23 tepals, seven to 11 stamens, nine to 16 staminodes, and 10–13 carpels. The flowers are quite spectacular, with the outer bractlike or sepal-like tepals green, the inner petal-like tepals greenish or yellowish with brown/purple spots (these features change gradually from the outermost to the innermost tepals), with broad, yellowish stamens that bear two elongate, disporangiate, strongly bulging, longitudinally dehiscing (introrse) thecae on the upper surface, with dark purple inner staminodes and a bright yellow gynoecium. Duration of anthesis of an individual flower is ca. 5 d (P. K. Endress, preliminary personal observation); it is functionally female on the first day; on the second day the anthers open and the stigma becomes hidden by the staminodes; the flower remains more or less unchanged for a few days (P. K. Endress, preliminary personal observation on cultivated specimens). The flowers smell like decaying fish and are pollinated by flies (Endress 1980*c*) (fig. 1*D*). Thus, these flowers are biologically very specialized. The smell is produced by the stamens and inner staminodes. These organs have a conspicuously papillate surface. The papillate epidermal cells have relatively dense cytoplasm with large nuclei, and presumably secrete the volatile molecules that are responsible for the smell (Endress 1980*c*). The staminodes are complicated not only at the histological but also at the morphological level. They have a furrow on the ventral side, and drosophilid eggs were found in this furrow (Endress 1984*b*). Thus, it seems that the flowers mimic rotting organic matter and bring visiting flies to oviposit on them. It is unknown whether the eggs or larvae survive when the floral organs have fallen to the ground. Tepals and stamens have a single vascular trace (Endress 1980*c*).

The carpels are stipitate and extremely ascidiate (Endress 1980*c*, 1983) (fig. 4*C*–4*F*). Each carpel contains four to 10 anatropous, bitegmic ovules, which are arranged in two longitudinal lines along the ventral side of the ovary (Endress 1980*c*). The carpels are completely free from each other, but at anthesis, all of the unicellular-papillate stigmas are held together by stigmatic secretion (Endress 1980*c*). Pollen tubes may cross between carpels in this zone, which may be called an extragynoecial compitum. The inner space of the carpels is not postgenitally fused but widely gaping and filled with secretion (angiospermy type 1; Igersheim and Endress 1997).

Floral development shows that all organs are arranged in spiral phyllotaxis, with divergence angles of ca. 138°, thus according to the Fibonacci pattern (Endress 1980*c*, 1983) (fig. 4*A*, 4*B*). The carpels originate as oblique bowl-shaped structures, and the entrance is soon lifted up by intercalary elongation (fig. 4*C*–4*F*). Both flanks of the entrance later elongate and form two lateral stigmatic lobes at anthesis (Endress 1980*c*, 1983).

Schisandraceae

Schisandraceae consist of two genera, *Schisandra* and *Kadsura*, and ca. 40 species (Saunders 1997, 1998, 2000). The main distribution is in tropical to temperate forests in eastern Asia; *Schisandra* also has one species in North America. The plants are climbing shrubs.

The flowers are unisexual, monoecious, or dioecious (Saunders 2000). It is not known whether they are self-compatible or self-incompatible. The flowers are whitish cream or dull red; the ovaries are green. Flower diameter is ca. 1–2 cm in many species, but the range has not been studied (cf. figures in Smith 1947; Saunders 1998, 2000) (fig. 1*E*). Floral organ number is variable. Flowers in *Kadsura* have seven to 24 tepals, 15–74 stamens, and 17 to ca. 300 carpels (Saunders 1998); flowers in *Schisandra* have five to 20 tepals, four to 60 stamens, and 12–120 carpels (Saunders 2000). Duration of anthesis of an individual flower, pollination biology, and breeding systems are unknown for the family in its natural habitat (Saunders

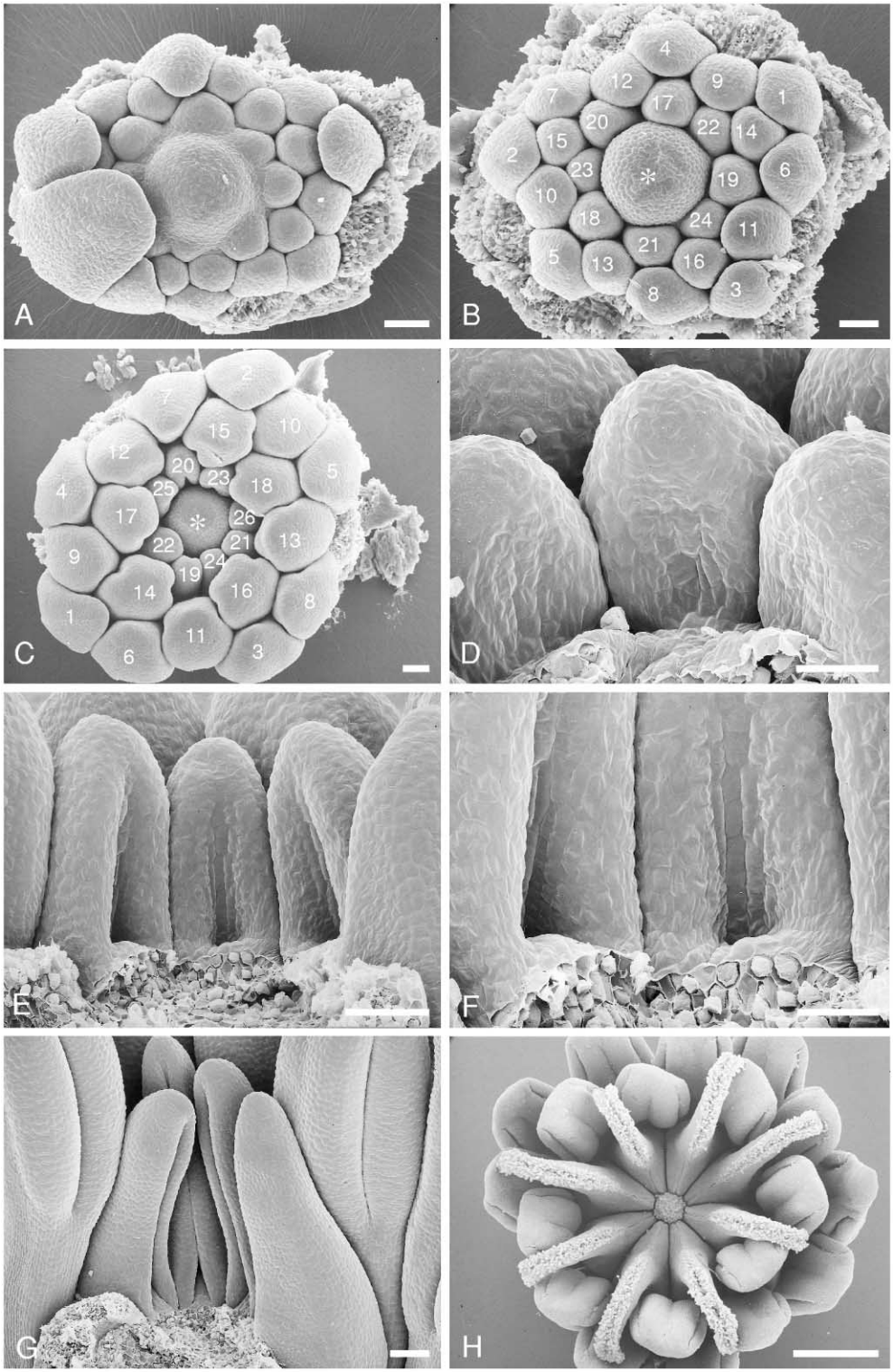


Fig. 7 *Illicium anisatum* (Illiciaceae). *A*, Young flower bud after initiation of all organs. *B*, Slightly older bud with hump (asterisk) in the floral center formed. *C*, Slightly older bud with hump (asterisk) overtopped by stamens and carpels. *D*, Young carpels with ascidiate base not yet developed. *E*, Somewhat older carpels with short ascidiate part present. *F*, Same as *E* in higher magnification. *G*, Carpels with flanks more or less closed, except for the basal part, where the median ovule is formed. *H*, Anthetic flower in female stage. *A*, Most tepals removed; *B*, *C*, *H*, all tepals removed. *B*, *C*, Subsequent stamens and carpels numbered. Magnification bars: *A*–*C*, *E*, *G* = 0.1 mm; *D*, *F* = 50 μm; *H* = 1 mm.

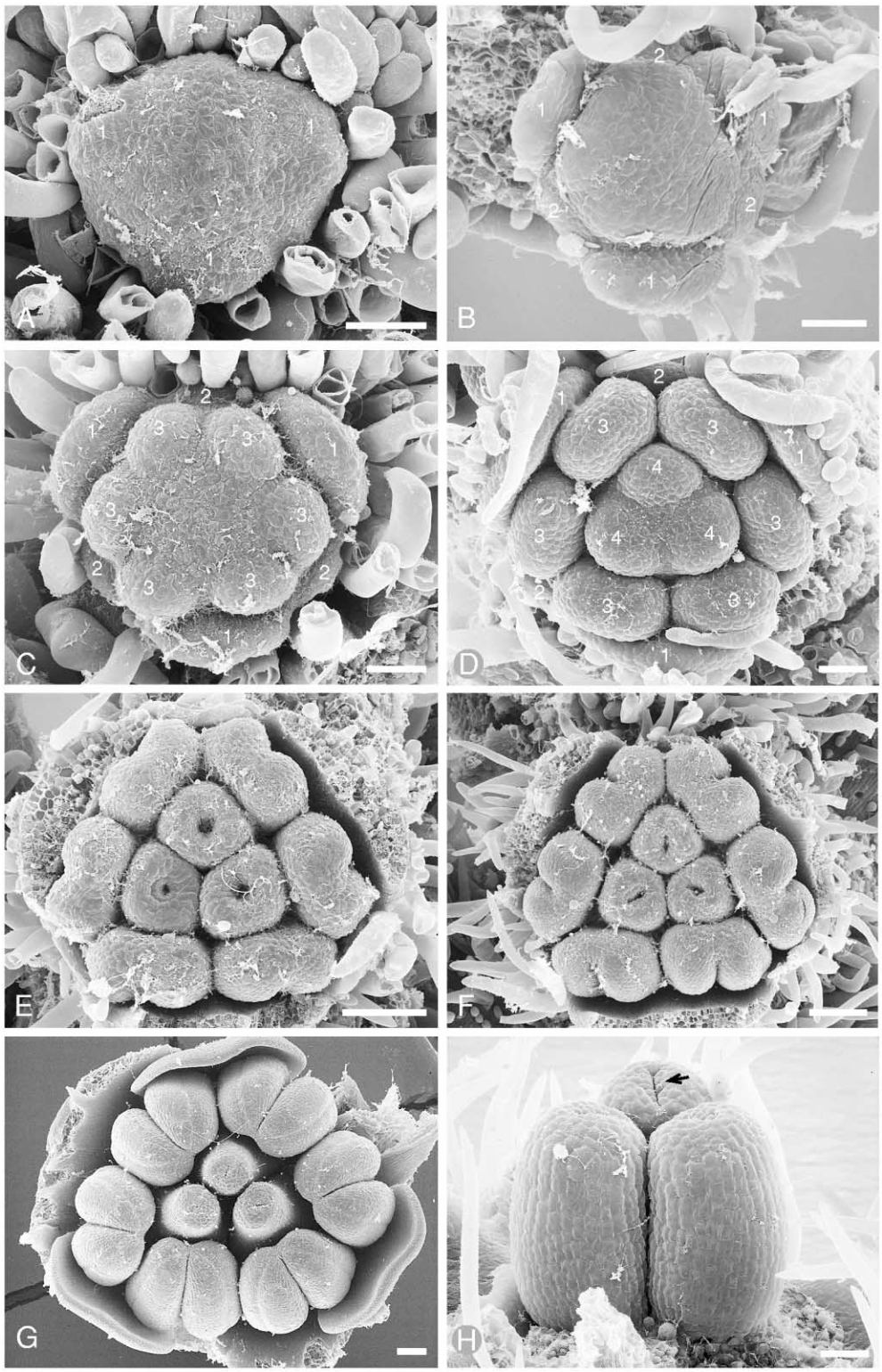


Fig. 8 *Cabomba furcata* (Cabombaceae). *A–G*, Successively older flower buds (from above). *A*, "Sepal" whorl initiated. *B*, "Petal" whorl initiated. *C*, Stamen whorl initiated. *D*, Carpel whorl initiated. *E*, Carpels have become ascidiate. *F*, Carpels forming slitlike openings. *G*, Carpels closed. *H*, Similar stage (carpels from the side). *A–D*, Organs of subsequent whorls numbered. *E–G*, "Sepals" removed. Magnification bars: *A–D*, *H* = 50 μm; *E–G* = 0.1 mm.

1998, 2000). In *Kadsura japonica*, in addition to male and female flowers, I found intermediate flowers with a few stamens below the gynoecium; it is unknown whether these stamens are fertile (fig. 1*E*). The rare occurrence of bisexual flowers was also reported for *Schisandra chinensis* (Ueda 1988). This is reminiscent of female flowers in *Amborella* with staminodes below the carpels. In male flowers of *K. japonica*, the uppermost stamens are reduced and probably sterile. In *Kadsura coccinea*, the male flowers have conspicuously enlarged inner staminodes of unknown function (Saunders 1998).

From the outside to the inside of the flower, the tepals gradually change from smaller, green to larger, colored organs. In some species, the tepals are caducous. The androecium shows pronounced synandry in many species. It would be interesting to know whether the massive synandrium has a particular function in pollination biology. The filaments are commonly shorter than the anthers. The anthers have two disporangiate, introrse or extrorse, and bulging thecae. Each theca opens by a longitudinal slit. The connective is commonly extremely broad and carnose as compared with the small thecae, so that the stamen may have the shape of a hammer. Tepals are served by one to three vascular traces, stamens by a single trace (Ozenda 1949; Tucker and Bourland 1994). Saunders (2000) mentions the occurrence of transitional forms between tepals and stamens.

The slightly stipitate carpels are pronouncedly ascidiate. The stigma surrounds the orifice of the inner space. The carpels are not postgenitally fused but filled with secretion (angiospermy type 1; Igersheim and Endress 1997; Endress and Igersheim 2000*a*). Each carpel contains two to five (to 11) anatropous, bitegmic ovules (in lateral position) (Smith 1947; Leinfellner 1966; Igersheim and Endress 1997). The gynoecium in its entirety is unique. The dorsal parts of the carpels are thick and form a compact shield at the periphery of the gynoecium, especially in *Kadsura*. The uppermost part of the stigma of each carpel emerges at the periphery of this shielded structure. Secretion is also present between the carpels (Igersheim and Endress 1996; Endress and Igersheim 2000*a*) (fig. 5*H*). This is reminiscent of *Austrobaileya*. However, the difference is that the peripheral shields form a secondary inner space so that this secretion is not visible from the outside. Thus, it is also an extragynoecial compitum but more elaborated than in *Austrobaileya*.

Developmental studies show that floral phyllotaxis is spiral according to the Fibonacci pattern, both in male and female flowers (see also Tucker and Bourland 1994; Liu and Lu 1999) (fig. 5*A*–5*C*; fig. 6*A*, 6*B*). The young carpels conspicuously curve toward the floral center so that the carpel opening becomes hidden by the next inner (neighboring) carpels (fig. 5*D*). The carpel opening can only be seen if one-half of the gynoecium is removed. The morphological differentiation of the basal region of the carpel is somewhat retarded. The conspicuous basal ascidiate part of anthetic carpels becomes visible only relatively late (fig. 5*E*–5*G*). Still later the short stipe is formed (fig. 5*G*).

Illiciaceae

Illiciaceae have a single genus *Illicium* with ca. 40 species in forests of eastern Asia and North and Central America (Saunders 1995). The flowers are bisexual and protogynous (fig. 1*F*). *Illicium floridanum* is self-incompatible (Thien et al. 1983). Duration of anthesis of an individual flower is 2–3 d in *Illicium parviflorum* (White and Thien 1985) and 12–14 d in *I. floridanum* (Thien et al. 1983); in both species, the flowers are functionally female on the first day. The flowers of *I. floridanum* have a faint, unpleasant smell. Nectar is produced in small quantities at the base of the petals and stamens (Thien et al. 1983, 2000; White and Thien 1985; Bernhardt 2000). The flowers are creamy white or dull red (*I. floridanum*), and the ovaries are green. Principal pollinators are various flies (Thien et al. 1983). Floral diameter is ca. 1–3 cm; the range is not well known. Organ number is variable: seven to 33 tepals, four to ca. 50 stamens, and five to 21 carpels (Smith 1947; Saunders 1995).

The tepals gradually change from outer, bractlike and sepallike organs to larger, inner showy organs. The stamens have a broad filament that is about the length of the anther. The anther is triangular in *Illicium anisatum*, and thus resembles *Amborella*. It has two disporangiate, introrse, slightly bulging thecae, each opening by a longitudinal slit. According to Keng (1965), in section *Badiana*, the tepals have a single vascular trace, whereas in section *Cymbostemon*, they have five or more. The stamens have a single vascular trace (Hiepko 1965; Keng 1965). Hiepko (1965) mentions the occurrence of transitional forms between tepals and stamens.

The carpels are plicate but the single (anatropous, bitegmic) ovule has a ventral-median position, which indicates a reduced ascidiate (peltate) carpel structure (Leinfellner 1965; Erbar 1983). In contrast to the other ANITA families shown before, the carpel flanks are partially postgenitally fused (angiospermy type 3; Igersheim and Endress 1997; Endress and Igersheim 2000*a*). At anthesis, the carpels are more or less free. They are grouped around a central hump of diverse forms in different species (Keng 1965). This hump is unicellular papillate and slightly secretory like a stigma, although each carpel has a normal stigma in the distal region (fig. 7*H*). Williams et al. (1993) showed that this hump is involved in a peculiar kind of extragynoecial compitum. Pollen tubes may grow from one carpel to another around this hump, like the cars that go around a roundabout. Williams et al. (1993) found different pathways of pollen tubes from the stigma to the ovule.

At anthesis, it looks as if the carpels are positioned in a whorl. However, young stages show that all organs are spirally arranged (Robertson and Tucker 1979; Erbar and Leins 1983; Ronse Decraene and Smets 1993; this study) (fig. 7*A*–7*C*). The divergence angles are ca. 138°, i.e., a Fibonacci pattern. The carpels curve early toward the central hump that is formed by the remaining floral apex. Thus, their opening becomes hidden by the hump. It can only be seen by cutting the flower in half and removing the hump. As in Schisandraceae, the ascidiate base becomes visible relatively late, and in contrast to Schisandraceae, it remains extremely short (fig. 7*D*–7*G*).

Cabombaceae

Cabombaceae comprise two genera, *Brasenia* and *Cabomba*, with six species of water plants distributed mainly in America, with *Brasenia* also in the Old World (Ørgaard 1991; Williamson and Schneider 1993). Nymphaeales (Cabomba-

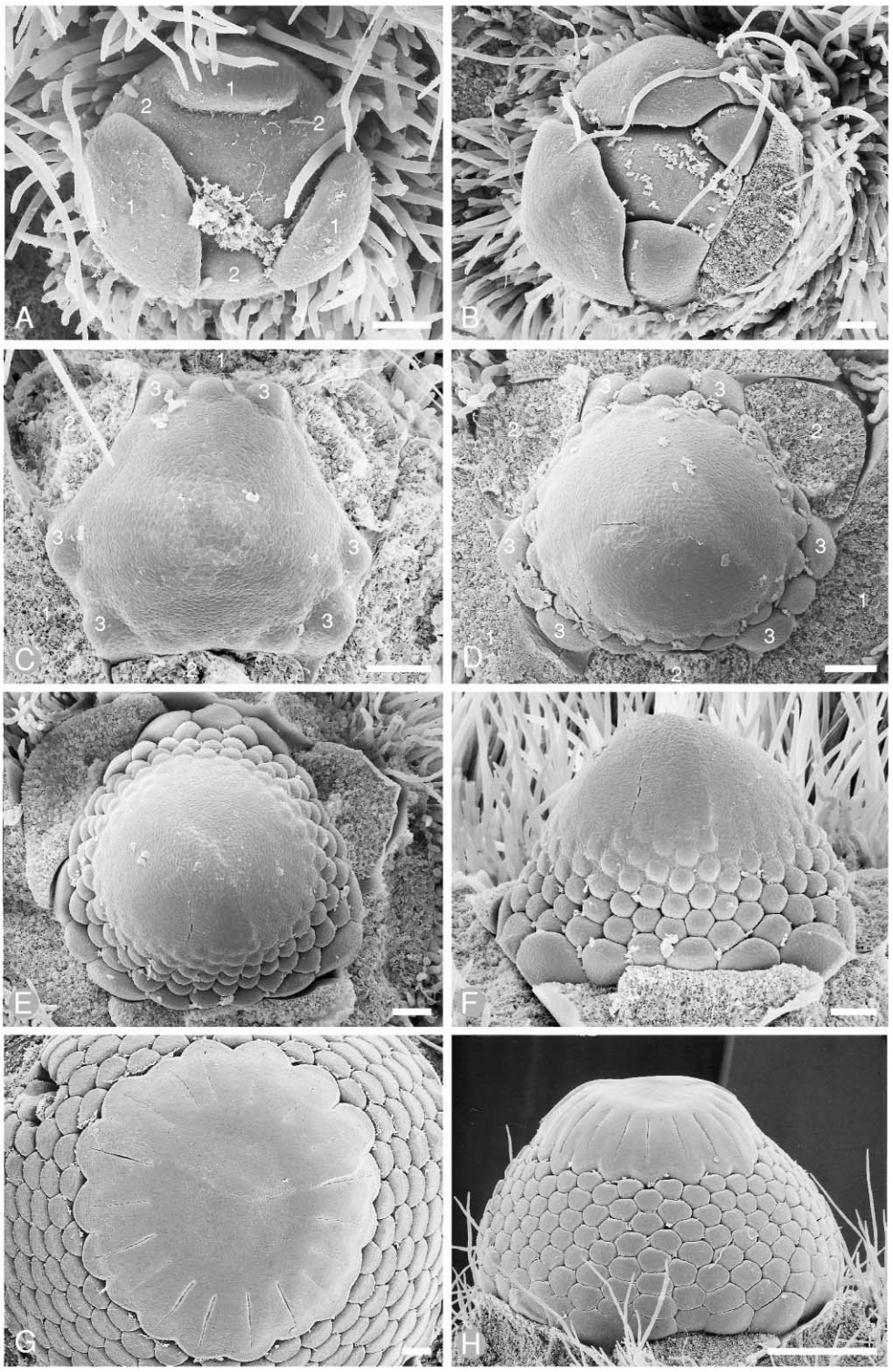


Fig. 9 *Nuphar advena* (Nymphaeaceae). *A–E*, Successively older flower buds (from above). *A*, Second "whorl" of "sepals" initiated. *B*, Slightly older stage. *C*, Petals initiated in double positions. *D*, Slightly older stage. *E*, Androecium being initiated; stamen primordia approximately in orthostichies, alternating with and opposite to series of lowermost visible organs ("petals" and stamens). *F*, Same bud (from the side). *G*, Floral bud with gynoecium formed. *H*, Same stage as *G* (from the side) showing slight irregularities in stamen position. *A*, *C*, *D*, Subsequent "whorls" of "sepals" and "petals" numbered. *C–H*, "Sepals" removed. Magnification bars: *A–G* = 0.1 mm; *H* = 0.5 mm.

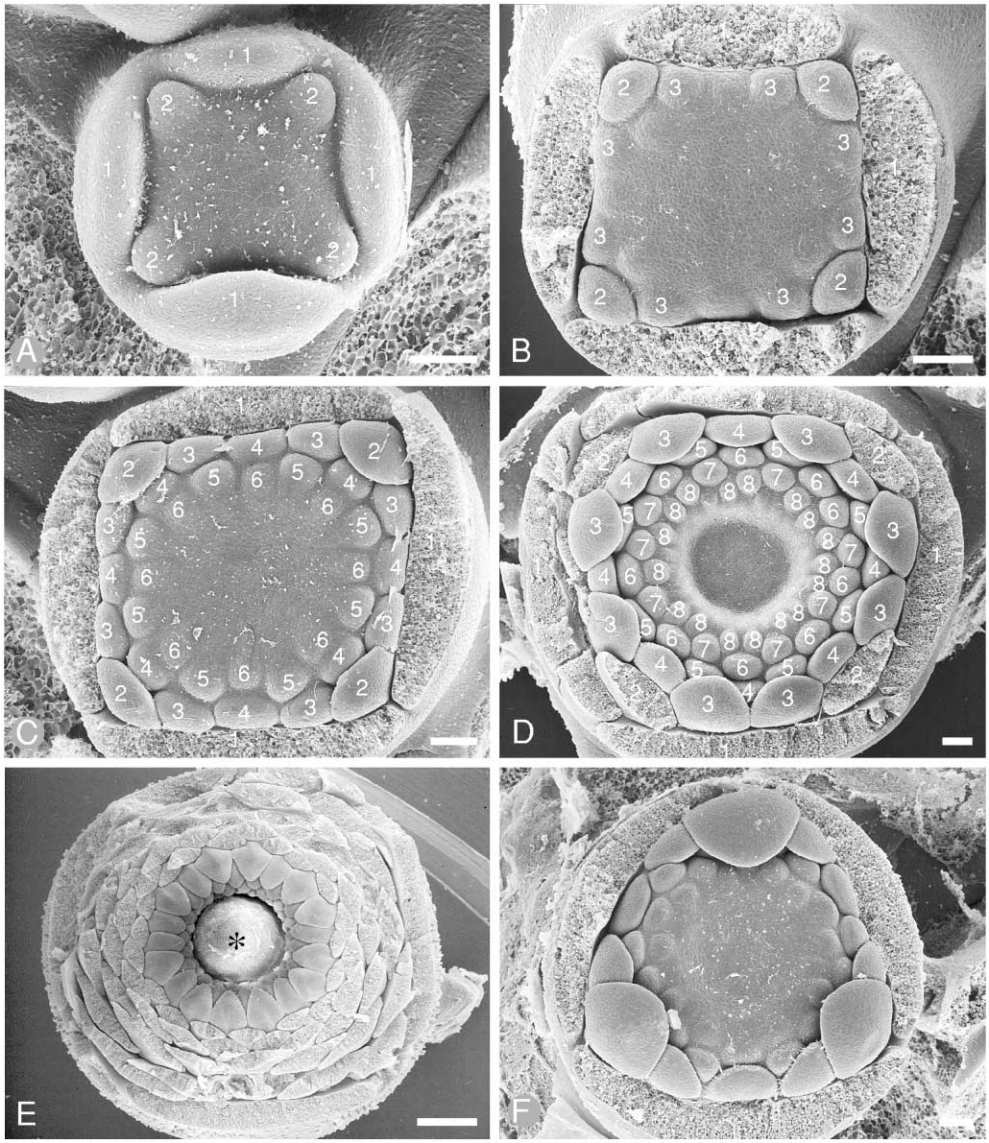


Fig. 10 *Victoria cruziana* (Nymphaeaceae). *A–E*, Successively older floral buds with tetramerous whorls of organs (from above). *A*, "Tepals" and first whorl of "petals" initiated. *B*, Subsequent "petal" whorls initiated in double positions. *C*, "Petals" forming alternating octomerous whorls. *D*, Slightly older stage; inner organs with slightly irregular position. *E*, All organs formed, including hump in the floral center (asterisk). *F*, Floral bud with trimerous whorls (similar stage as *C*). *A–D*, Subsequent whorls of floral organs numbered. *B–F*, Outer organs removed. Magnification bars: *A–D*, *F* = 0.1 mm; *E* = 0.5 mm.

ceae and Nymphaeaceae) are different from the other ANITA clades in many respects, which seems to be at least partly because they are water plants. In contrast to the other ANITA members, the floral organs are not spiral but whorled in the Nymphaeales. *Cabomba* has the simplest flowers. In both *Cabomba* (Schneider and Jeter 1982) and *Brasenia* (Osborn and Schneider 1988), the flowers are bisexual and protogynous (fig. 1*G*). As far as I have seen in the literature, it is unknown whether Cabombaceae are self-compatible or self-incompatible. Duration of anthesis of an individual flower is 2 d. Floral diameter is ca. 0.5–1.5 cm in *Cabomba* (Ørgaard 1991) and ca. 2 cm in *Brasenia* (Richardson 1969; Osborn and Schneider 1988). The flowers of *Cabomba* are white, yellow, purplish pink, or violet (Ørgaard 1991); nectar is produced by two auricles of each petal (Schneider and Jeter 1982; Vogel 1998) (fig. 1*G*); the flowers are pollinated by various small insects, especially flies (Schneider and Jeter 1982). In contrast, the flowers of *Brasenia* are dull purple; they are wind pollinated (Osborn and Schneider 1988; Osborn et al. 1991).

The flowers of *Cabomba* consist of four trimerous (or dimerous) organ whorls (fig. 8*A–8G*). Only the stamens are in double positions, thus there are six stamens in one whorl (fig. 8*C–8G*). *Brasenia* flowers also begin with trimerous whorls but differ in the androecium and gynoecium, with 18–36 stamens with double positions in several whorls and six to 18 carpels in simple whorls (Ito 1986; Ronse Decraene and Smets

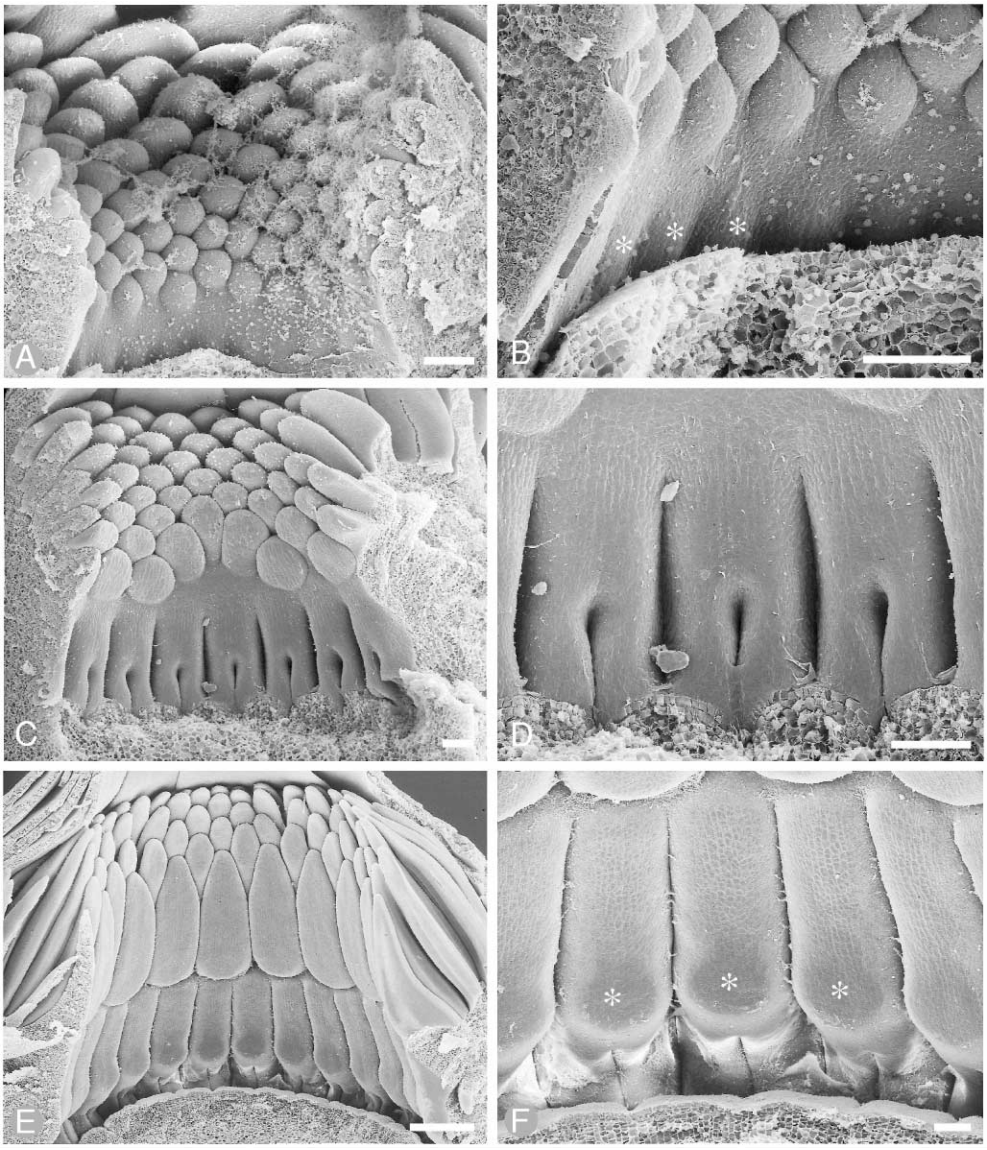


Fig. 11 *Victoria cruziana* (Nymphaeaceae) carpel development; flowers cut in half, and hump in the floral center removed to show innermost stamens and young carpels. *A*, Carpels just being initiated; stamens approximately in orthostichies; carpel primordia alternating with innermost stamens. *B*, Same as *A* in higher magnification; carpel primordia marked with asterisks. *C*, Carpels demarcated from each other by a longitudinal slit; ascidiate part not developed, except for carpel in the middle. *D*, Same as *C* in higher magnification. *E*, Carpel apices beginning to form. *F*, Same as *E* in higher magnification; carpel apices marked with asterisks. Magnification bars: *A–D*, *F* = 0.1 mm; *E* = 0.5 mm.

1993). Richardson (1969) adscribed "very low helices" to the floral phyllotaxis, which is, however, not evident in his illustrations. In contrast to *Cabomba*, in *Brasenia*, the organs of the first and second perianth whorl are similar (Hiepko 1965). All perianth organs have a single vascular trace in both *Brasenia* (Hiepko 1965; Ito 1986) and *Cabomba* (Ito 1986). The anthers are introrse (*Brasenia*) or slightly extrorse (*Cabomba*). They have two disporangiate thecae. The thecae are more or less bulging (Chifflot 1902). Each theca opens by a longitudinal slit. The filament is about the same length as the anther (*Cabomba*) or much longer (*Brasenia*) (Ito 1986; Osborn and Schneider 1988; Endress 1994*d*).

The slightly stipitate carpels are extremely ascidiate, the small, capitate stigma (with bi- or pluricellular papillae) encompassing the upper end of the ascidiate zone. In *Brasenia*, the stigma descends on the ventral side of each carpel; however, the carpels are nevertheless extremely ascidiate (P. K. Endress, personal observation). The carpels are not postgenitally fused but filled with secretion (angiospermy type 1; Igersheim and Endress 1998; Endress and Igersheim 2000*a*). Each carpel contains one to three (to five) anatropous, bitegmic ovules in ventral, lateral, or dorsal position (Richardson 1969; Moseley et al. 1984; Ito 1986; Igersheim and Endress 1998). The gynoecium in *Cabomba* has an indument of strigose, upward-directed, four-cellular hairs, with a long, tanniferous end cell, similar to those in *Trimenia* (fig. 14*E*, 14*F*).

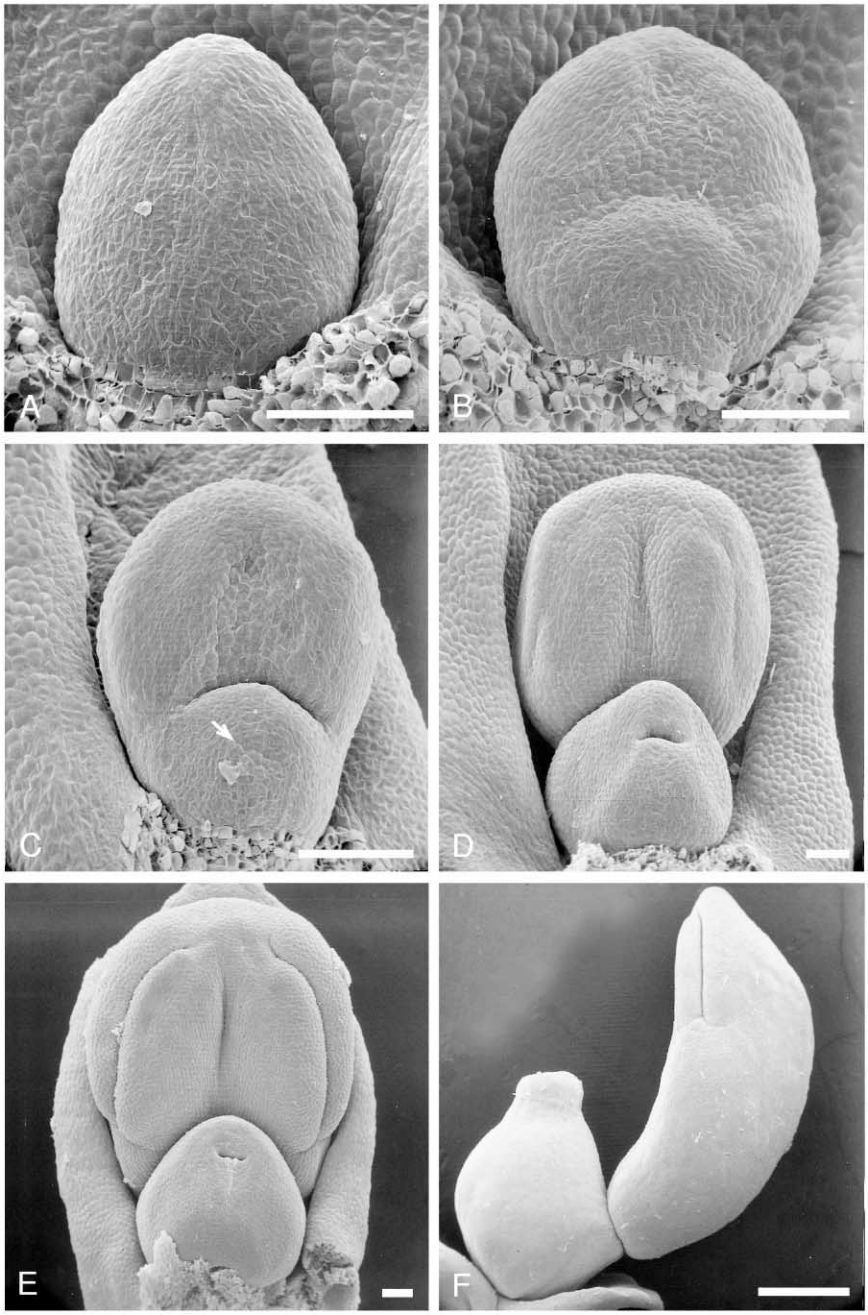


Fig. 12 *Sarcandra chloranthoides* (Chloranthaceae); flower development. *A–E*, Adaxial views. *A*, Flower primordium. *B*, Stamen and carpel primordium distinguishable. *C*, Carpel with shallow concavity (arrow) indicating incipient ascidiate part. *D*, Lower part of carpel has elongated. *E*, Entrance into carpel closed. *F*, Anthetic flower (from the side). Magnification bars: *A–E* = 0.1 mm; *F* = 1 mm.

Floral development in *Cabomba* shows that the organs of the inner perianth whorl are strongly retarded in bud as compared to those of the outer perianth whorl and stamens and expand only shortly before anthesis (fig. 8*C*–8*G*). This was also noticed by Hiepko (1965) and Tucker and Douglas (1996). It is a feature that is typical for petals in many eudicots (cf. Hiepko 1965). In *Brasenia*, in contrast, the inner perianth organs seem not to be retarded as compared to the outer ones (Hiepko 1965). In *Cabomba*, the ascidiate carpels begin development as bowl-like structures (fig. 8*E*, 8*F*). The entrance is soon lifted up by intercalary elongation (fig. 8*G*, 8*H*).

Nymphaeaceae

Nymphaeaceae have six genera and ca. 60 species of water plants, with a worldwide distribution in tropical to temperate regions (Schneider and Williamson 1993). The flowers range from small, ca. 1–2 cm (*Ondinea*) to very large, up to 50 cm

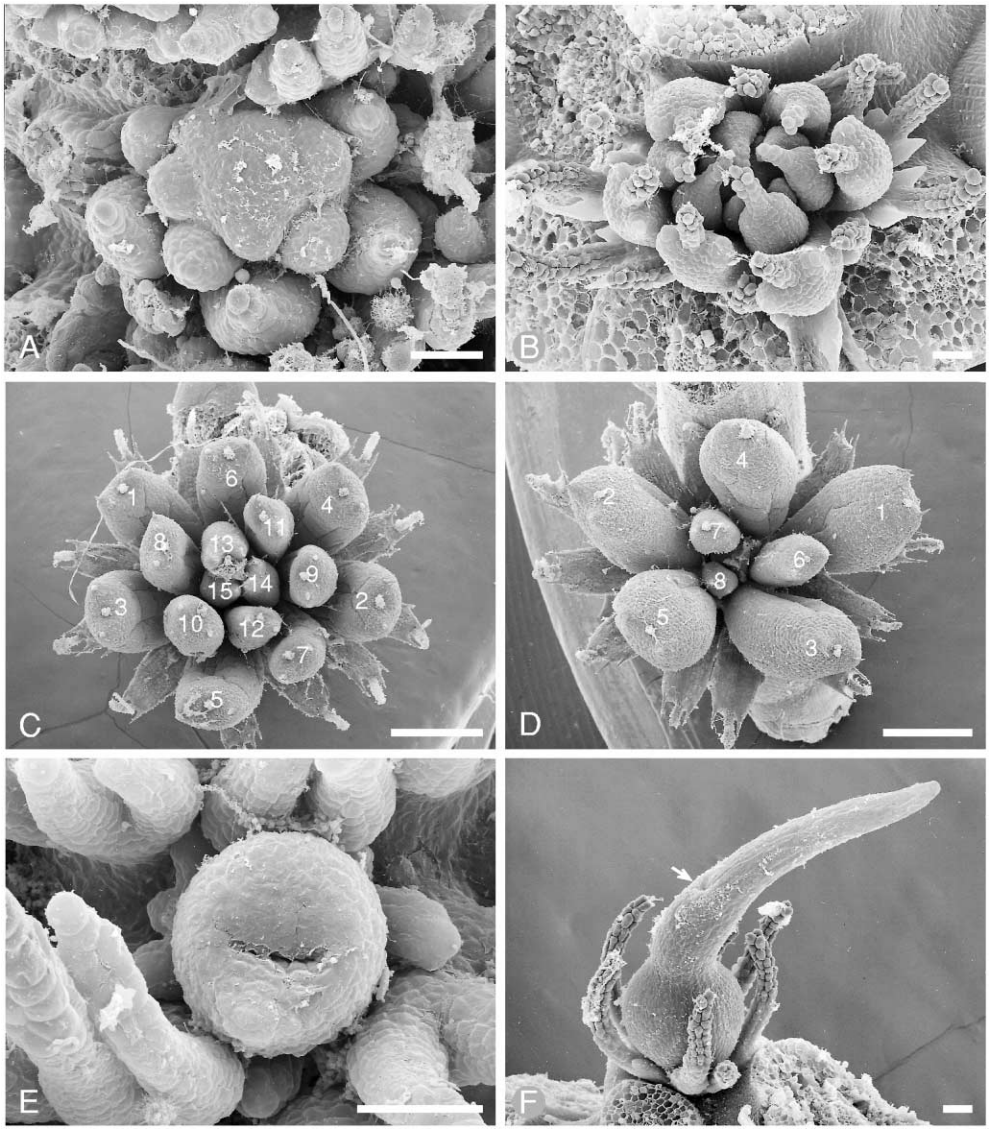


Fig. 13 *Ceratophyllum demersum* (Ceratophyllaceae). *A–D*, Male flower development. *A*, Young flower bud with inner stamens being initiated. *B*, All stamens formed; anthers with secretory connective tips. *C*, Early anthesis; connective tips fallen. *D*, Later anthesis with outer stamens fallen. *E–F*, Female flower development. *E*, Young carpel showing ascidiate structure. *F*, Carpel before anthesis; entrance marked with arrow. *C*, *D*, Subsequent stamens numbered beginning with the first organs visible in the figures. Magnification bars: *A*, *B*, *E*, *F* = 50 μm; *C*, *D* = 0.5 mm.

(*Victoria*; Schneider and Williamson 1993). Organ number varies widely: "sepals" three or four (up to 14) (Beal 1956; Schneider and Williamson 1993), "petals" zero to four (*Ondinea*; den Hartog 1970; Kenneally and Schneider 1983) up to six to 51 (*Nymphaea*; van Royen 1962) and 50–70 (*Victoria*; Schneider 1976), stamens 15–34 (*Ondinea*; den Hartog 1970; Kenneally and Schneider 1983) up to 15–750 (*Nymphaea*; van Royen 1962), carpels three to 14 (*Ondinea*; den Hartog 1970; Schneider 1983) up to five to 47 (*Nymphaea*; Wiersema 1987). The flowers are white, yellow, pink, red, purple, or blue (Wiersema 1988). They are bisexual and protogynous (but protogynous or homogamous in *Nuphar lutea* [Heslop-Harrison 1955*a*] and *Nymphaea alba* [Heslop-Harrison 1955*b*]; homogamous in *Nymphaea jamesiana*, *Nymphaea lingulata*, *Nymphaea ampla* [Wiersema 1988], and *Nymphaea capensis* [Orban and Bouharmont 1995]; however, these authors only say that the outermost anthers open already on the first day of anthesis and not whether they open at the same time the stigma becomes receptive) (fig. 1*H*). Breeding systems are poorly known in Nymphaeaceae. However, all previous studies found self-compatibility (*Nuphar lutea* [Ervik et al. 1995] and *N. capensis* [Orban and Bouharmont 1995]). Partial cleistogamy, with flowering under water, occurs in *Euryale* and *Barclaya* (Kadono and Schneider 1987; Williamson and Schneider 1994), which also indicates self-compatibility. There is no evidence of self-incompatibility in *Nymphaea*

(Wiersema 1988). Duration of anthesis in an individual flower is 4–8 d (*N. lutea* [van der Velde 1986]; *N. alba* [van der Velde 1986]; *N. candida* [van der Velde 1986]; *N. gigantea* [Schneider 1982*b*]), 3 d (*Ondinea* [Schneider 1983]; *N. ampla* [Prance and Anderson 1977]; *Nymphaea elegans* [Schneider 1982*a*]; *N. odorata* [Schneider and Chaney 1981]; *N. capensis* [Orban and Bouharmont 1995]), or 2 d (*Nymphaea mexicana* [Capperino and Schneider 1985]; *Euryale ferox*, if chasmogamous [Kadono and Schneider 1987]). Day-flowering *Nymphaea* species are commonly pollinated by bees and flies (Wiersema 1988). *Victoria* and several tropical species of *Nymphaea* flower at night, and anthesis of a flower extends over two nights. These flowers are scented and thermogenic; they are predominantly pollinated by *Cyclocephala* beetles (*Victoria cruziana* [Valla and Cirino 1972], *Victoria amazonica* [Prance and Arias 1975], *Nymphaea blanda* [Cramer et al. 1975], *Nymphaea rudgeana* [Cramer et al. 1975; Prance and Anderson 1977], and *Victoria amazonum* [Prance 1980]). *Barclaya* is probably pollinated by flies (Williamson and Schneider 1994). *Nuphar* is pollinated predominantly by flies (Lippok and Renner 1997) in North America and also by beetles and bees (Lippok et al. 2000). Nectar is produced on petals of *Nuphar* (Hiepko 1965; Lippok and Renner 1997). In species of *Nymphaea* and *Ondinea*, the floral center secretes a puddle of fluid, which washes the pollen from the bodies of pollinators and again may function as a kind of extragynoecial compitum (Schmucker 1932; Meeuse and Schneider 1980; Endress 1982).

In *Nuphar*, "petals" are retarded in development, compared with "sepals" and stamens (fig. 9*C*–9*H*); however, in *N. alba*, they are only slightly retarded with regard to stamens but not as compared to "sepals" (Hiepko 1965). In both *Nuphar* and *Nymphaea*, the "petals" and "sepals" commonly have only a single vascular trace (Moseley 1961; Hiepko 1965), in *Nymphaea tetragona* one to three (Ito 1983, 1984), and in *Nuphar* and *Ondinea*, "sepals" sometimes more than one (Moseley 1965; Williamson and Moseley 1989). In *Victoria*, staminodes are present between "petals" and stamens (Hiepko 1965). Stamens have an elongate, introrse anther (latrorse in *Ondinea*; den Hartog 1970), with a broadened or tapering apical protrusion. The anther has two disporangiate thecae. The thecae are bulging or not bulging (Chifflot 1902); each theca opens by a longitudinal slit (in *Nuphar* by an H-shaped double valve; Hufford 1996). The large stamens of *Victoria* are served by three vascular traces (Hiepko 1965), the outer stamens by even more (Heinsbroek and van Heel 1969). This is also found in *Nymphaea lotus* (van Heel 1977), whereas in other, smaller-flowered *Nymphaea* species, stamens have only one vascular trace (Moseley 1958), which is also found in *Nuphar* (Moseley 1965). The presence of several vascular traces in these stamens can be seen as an autapomorphy because of floral gigantism. However, in general, floral vascular supplies are complex in all Nymphaeaceae (Moseley et al. 1993) because of their relatively large flowers, as compared to other members of the ANITA grade.

Carpels are ascidiate but they are united over their flanks, and in the floral center, there is a protrusion, reminiscent of that in *Illicium*. At anthesis, the carpels are filled with secretion. However, in contrast to *Cabomba*, each carpel is postgenitally fused at the periphery (angiospermy type 3; Igersheim and Endress 1998). Each carpel contains three to 400 or more ovules (Igersheim and Endress 1998) at both flanks on laminar diffuse placentae. The ovules are anatropous and bitegmic, except for *Barclaya*, which has orthotropous ovules (Schneider 1978; Winter 1993). In some taxa, similar four-cellular hairs as in *Cabomba* occur on the carpels.

Interpretion of floral phyllotaxis of Nymphaeaceae in the literature is confusing. In the specimens of *Nuphar* and *Victoria* studied here, phyllotaxis begins whorled with the outer perianth members. Although in *Nuphar* the outermost five or six organs ("sepals"; cf. also Padgett et al. 1999) are initiated sequentially in a spiral pattern, the position is in two whorls (fig. 9*A*, 9*B*). This seems to be effected by a longer plastochron between the third and fourth organ of the flower. The three outermost organs will later cover the inner ones. In *Victoria*, there are three or, more often, four organs in the outermost whorl ("sepals") (fig. 10*A*–10*F*). The organs of the second whorl are much narrower than those of the first whorl. In both *Victoria* and *Nuphar*, the organs of the third whorl are in double positions, forming a whorl of eight or six organs (similar in *Nymphaea*; Ronse Decraene and Smets 1993) (fig. 9*C*, 9*D*; fig. 10*B*–10*D*, 10*F*). In *Victoria*, more octomerous whorls follow in alternation; after a few whorls, phyllotaxis becomes more irregular (fig. 10*D*, 10*E*). In *Nuphar*, phyllotaxis becomes irregular sooner because the individual organs are very small in comparison with the entire floral apex, which quickly leads to the decay of the initially established whorled pattern (fig. 9*E*–9*H*). In both genera, the carpels seemingly form one whorl (fig. 9*G*, 9*H*; fig. 11*A*–11*F*). What one can see is that the phyllotaxis of these flowers is not spiral. It starts whorled and then becomes irregular. *Nuphar* and *Victoria* also are congruent in that, in both, the third whorl shows organs in double position. Wolf (1991) mentions lability in the androecium of *N. alba*, with whorled, spiral, and irregular patterns. However, the perianth is whorled, at least in the beginning.

In *Nuphar*, the floral apex becomes dome shaped during stamen initiation. The carpels are initiated in a whorl around the summit of the dome-shaped floral apex, more or less alternating with the uppermost stamens. They do not rise much from the surface but appear rather as radial depressions (see also Troll 1933; Moseley 1972) (fig. 9*G*, 9*H*). This leads to the syncarpous condition without an internal compitum at anthesis. Thus, a carpel does not have a clearly defined apex, and therefore lacks the differentiation into a dorsal and ventral side. One could say it has an apical slit. At first, the carpel opening is turned obliquely to the side; at anthesis, by differential thickening of the floral center, it is more or less vertical. In *Victoria*, the floral apex becomes depressed during formation of tepals and stamens. In the center of this depression, a dome-shaped part is formed later, which is devoid of organs. The carpels originate in a whorl at the lowermost part of the depression, which has vertical sides. The young carpels are hidden by the central dome. The carpels are initiated as radially elongate, flat mounds (fig. 11*A*, 11*B*). In contrast to *Nuphar*, radial grooves appear between the carpels. However, a clear-cut carpel apex is lacking as well at this stage (fig. 11*C*, 11*D*). A longitudinal depression is formed in each carpel primordium, which later appears as a slit. The carpel appears slightly ascidiate (fig. 11*D*). Only now a mound forms at the outer end of the slit, which may correspond to a delayed carpel apex (fig. 11*E*, 11*F*). Gynoecium development in *Nymphaea* is sim-

ilar (cf. Troll 1933). At anthesis, these "carpel appendages" function as osmophores together with inner "petals" and are thermogenic (Prance and Arias 1975). At anthesis the entrance into the carpels is less steep than at initiation (cf. Schneider 1976); the same is true for *N. alba* (Payer 1857; Troll 1933).

Chloranthaceae

Chloranthaceae consist of four genera and ca. 75 species (Todzia 1993), with a scattered distribution in wet forests of tropical to warm temperate regions of the Old and New World. Chloranthaceae are not part of the ANITA grade, but they appear immediately above in some analyses (Doyle and Endress 2000). Flowers are in some respects similar to Trimeniaceae and Amborellaceae (Endress 1986, 1987*a*; Endress and Igersheim 1997*a*, 2000*a*).

The flowers are bisexual (*Chloranthus*, *Sarcandra*) or unisexual (*Ascarina*, *Hedyosmum*) (fig. 12*F*); bisexual flowers are protogynous (von Balthazar and Endress 1999). Self-incompatibility was recorded in *Chloranthus spicatus*, self-compatibility in *Sarcandra glabra*, and agamospermy in *Sarcandra chloranthoides* (von Balthazar and Endress 1999). Four other *Chloranthus* species seem to be self-compatible (Ma et al. 1997; Luo and Li 1999; Wang et al. 1999). Duration of anthesis of an individual flower is ca. 10-12 d in *Sarcandra* (two species) (von Balthazar and Endress 1999) and ca. 6 d in *Chloranthus* (two species) (Luo and Li 1999; von Balthazar and Endress 1999). *Chloranthus* and *Sarcandra* are insect-pollinated; thrips were found as the main pollinators in three *Chloranthus* species (Ma et al. 1997; Luo and Li 1999). *Hedyosmum* and perhaps *Ascarina* are wind-pollinated (cf. Endress 1987*a* and literature cited therein).

Flowers are extremely reduced in organ number and very small in size (ca. 1–6 mm diameter). A perianth is lacking, except in *Hedyosmum*, with three short, scalelike organs that are probably tepals. The presence of one single stamen is characteristic for *Hedyosmum*, *Sarcandra*, and part of *Ascarina*. The stamen is not distinctly differentiated into anther and filament. It has two disporangiate, mostly introrse thecae, which open by a longitudinal slit or by a rudimentary H-shaped double valve (*Sarcandra*) (Endress 1987*a*). In *Chloranthus*, the androecium is a three-lobed scale with two thecae or none on the middle lobe but only one theca on each side lobe (on the lateral side). The thecae are not bulging except for some *Chloranthus* species (e.g., *Chloranthus japonica*).

There is only one carpel, which is barrel-shaped, stipitate, and extremely ascidiate (fig. 12*F*). The carpel contains a single ventral-median, orthotropous, bitegmic ovule. The stigma surrounds the orifice of the inner space. In *Hedyosmum*, the dorsal part of the carpel becomes extended into an elongate stigmatic area (Endress 1971). The stigma surface is smooth or unicellular papillate, but in some taxa, there are also multicellular protrusions (Endress 1987*a*). The carpels are not postgenitally fused but filled with secretion (angiospermy type 1; Endress and Igersheim 1997*a*, 2000*a*).

Flower development has been studied in *Sarcandra* and *Chloranthus* (Endress 1987*a*). In both, there are no perianth rudiments. The unicarpellate gynoecium becomes apparent only when the uni- to tristaminate androecium is relatively far into development (fig. 12*A*, 12*B*). The carpel develops as an oblique bowl-like structure, which results in a pronouncedly ascidiate form at anthesis (fig. 12*C*–12*F*).

Ceratophyllaceae

Ceratophyllaceae have one cosmopolitan genus, *Ceratophyllum*, with about six species of submerged water plants (Les 1993). *Ceratophyllum* appeared as basalmost clade of the angiosperms in the first large *rbc*L analysis (Chase et al. 1993). In the recent multigene analyses, it is not at the base, but its position is not stable; it tends to come out as sister of monocots or as sister of eudicots (Graham et al. 2000; Qiu et al. 2000; Soltis et al. 2000*a*).

The flowers are unisexual. Floral structure is highly reduced (Endress 1994*b*), and floral size is minute (ca. 0.5–1.5 mm diameter). Duration of anthesis of an individual flower is unknown. Pollination takes place under water. The stamens detach from the male flowers one by one, beginning at the periphery of the flower. They ascend to the water surface and successively release their pollen, which sinks down, eventually hitting female flowers. Because *Ceratophyllum* lives completely under water, lignified tissue is almost completely lacking. Even the vascular bundles do not form any lignified cell walls (Endress 1994*b*). Most curiously, the only lignified cells in the flowers are the one-celled tips of the stamens that make contact with the water-air boundary when the detached stamens float up to the water surface.

Tepals are lacking. The organs surrounding the stamens or the carpel may rather be interpreted as bracts because occasionally a lateral flower is formed between these organs and the pollination organs (discussion in Les 1993; Endress 1994*b*). The male flowers contain three to 46 stamens (Sehgal and Ram 1981; Wilmot-Dear 1985; Les 1993); the innermost stamens are retarded and sterile and, thus, are staminodes (Shamrov 1981; Endress 1994*b*) (fig. 13*C*, 13*D*). The stamens consist of an almost sessile anther with two disporangiate, extrorse, nonbulging thecae; each theca opens by an irregular longitudinal slit. The female flowers have only one stipitate, extremely ascidiate carpel; the ovary is barrel shaped; there is a long, stylelike part without a distinctly differentiated stigma, surprisingly on the side of ovule attachment, which is the ventral side (Troll 1933; Shamrov 1983; Endress 1994*b*) (fig. 13*F*). The orifice of the inner surface of the carpel is about at midlength of the stylelike extension. The ventral and dorsal parts of the inner surface of the carpel are appressed to each other but apparently not postgenitally fused (angiospermy type 1; Igersheim and Endress 1998). Each carpel has one median, pendant, orthotropous, unitegmic ovule.

Floral development shows that stamen phyllotaxis is diverse, either spiral (Fibonacci or Lucas pattern) or whorled (tri- or tetramerous) (Endress 1994*b*) (fig. 13*A*, 13*C*, 13*D*). The young anthers have a secretory tip (as do the foliage leaves), which breaks off before anthesis (fig. 13*B*, 13*C*). The carpel is initiated as a cuplike organ, which results in the pronouncedly ascidiate form at anthesis (fig. 13*E*, 13*F*).

Discussion

Flower Size

Flower size is interesting in the context of the changing paradigms of primitive flowers. Before the 1970s, large flowers

with numerous organs were seen as primitive in angiosperms, from which then smaller flowers evolved by reduction. The prime importance of the process of reduction in size and number of organs as a major trend in floral evolution was almost dogma for a long time. With the establishment of phylogenetic studies and with knowledge of more fossils, it has become obvious that evolutionary changes have proceeded in very different directions. Since small- and medium-sized flowers with few or only a moderate number of organs are predominant in the ANITA grade, the presence of small flowers in early angiosperms has become more probable. An evolutionary concept that has to be explored for ancestral angiosperm flowers is that paedomorphic evolutionary changes were a potentially important precondition for extensive radiations. It was discussed earlier for angiosperms (Takhtajan 1976) but should be explored in greater depth in light of the present knowledge.

Flowers are small (less than 1 cm in diameter) in *Amborella*, *Cabomba*, and Trimeniaceae (also in Chloranthaceae and *Ceratophyllum*). They are medium sized (1–3 cm) in *Brasenia* (Cabombaceae), *Nuphar* p.p., *Ondinea* (Nymphaeaceae), Schisandraceae, and *Illicium*. They are large (more than 3 cm) in *Austrobaileya* and most Nymphaeaceae. The very large flowers in some Nymphaeaceae are highly specialized, and neither phylogenetic studies of extant plants (Les et al. 1999) nor the fossil record support a basal position for these large-flowered taxa (Mohr and Friis 2000).

Floral Phyllotaxis

Floral phyllotaxis is evolutionarily plastic in basal angiosperms, which appears to be correlated with a lack or only a low degree of synorganization of floral organs (Endress 1987*b*, 1990*b*). It is spiral in *Amborella* (Endress and Igersheim 2000*b*), *Austrobaileya* (Endress 1980*c*), Trimeniaceae (Endress and Sampson 1983), Schisandraceae (Tucker and Bourland 1994; this study). But it is whorled in Nymphaeales: Cabombaceae (Ronse Decraene and Smets 1993; Tucker and Douglas 1996; this study), Nymphaeaceae (Wolf 1991; this study). In *Illicium*, with its peculiar extragynoecial compitum, the carpels appear whorled in anthetic flowers, but they are initiated spirally like the other floral organs (Robertson and Tucker 1979; Erbar and Leins 1983; Ronse Decraene and Smets 1993; this study). The anthetic position results from spacing of the carpels around a central cone. In *Ceratophyllum*, whorled (trimerous and tetramerous) and spiral flowers (Fibonacci and Lucas pattern) occur in the same species (Endress 1994*b*). In *Nuphar*, the first five or six floral organs are initiated in a spiral sequence. However, after the first three organs are initiated, there seems to be a longer plastochron, which results in a whorled arrangement: the three outermost organs approximately form a trimerous whorl, followed by two or three smaller organs, which alternate with the outer organs and thus form a second whorl. Cabombaceae and Nymphaeaceae have a common pattern: The flower begins with two alternating whorls of three (or in Nymphaeaceae more often four) organs. In the third whorl, the organs have double positions, thus six (or eight) organs. This number is then propagated to the subsequent whorls. The innermost whorls may go down to the original number of three (or four). However, there is a difference between the taxa: the third whorl, which shows double organ positions, consists of stamens in Cabombaceae, small petal-like organs in *Nuphar*, and large tepals in *Victoria* and *Nymphaea*. Although in Nymphaeaceae the most common "sepal" number (or organ number of the outer whorl) is four, I also found flowers with three "sepals" in *Victoria* (fig. 10*F*), and with transitions between three and four. Thus, the number and arrangement of floral organs are more similar among genera of Nymphaeaceae and even between Cabombaceae and Nymphaeaceae than expected from the records in the literature. However, the organ categories of the second and subsequent whorls are not the same in every genus.

It may be speculated and inferred from phylogeny (Les et al. 1999) that, in Nymphaeales, the original number is three, not four, because three is common in Cabombaceae and *Nuphar* and because it is a Fibonacci number. Four could be based on a secondary increase concomitant with an increase of floral primordium diameter in the large-flowered genera of Nymphaeaceae. This number could then become more or less fixed in most of the family and still retained in the secondarily small-flowered genus *Ondinea*.

Perianth

The perianth is diverse, and some features are correlated with phyllotaxis. Because floral phyllotaxis is spiral in many taxa of the ANITA grade, perianth organ number is not fixed. The innermost perianth organs tend to be in a series of organs consisting of a Fibonacci number because every time a Fibonacci number is complete, the organ distribution attains a relative peak of regularity. However, the Fibonacci number is flexible, sometimes even within a species (Hirmer 1931; Endress 1987*b*). Thus, the number 3 tends to occur in *Nuphar* (3 is regular in the whorls of *Cabomba*) (fig. 1*G*, 1*H*), 5 in *Amborella*, and among *Austrobaileya* in flowers with relatively few tepals (fig. 1*C*), but 8 in flowers with relatively many tepals (fig. 1*D*). The number seems to be dependent on the breadth of the tepals relative to the circumference of the flower: the broader the tepals, the lower the Fibonacci number in a series. In Nymphaeaceae, with whorled phyllotaxis, 4, a non-Fibonacci number, is predominant, except in *Nuphar*. But trimerous flowers were also found in *Victoria* (fig. 10*F*). Thus, an attempt to assign a specific number of perianth members in a series as plesiomorphic in extant angiosperms may be misleading because the number changes so easily. However, at the level of Nymphaeales, trimery seems to be basic, as it is present not only in Cabombaceae but also in the basal genus *Nuphar* of Nymphaeaceae (in contrast to Les et al. 1999).

In *Amborella*, Trimeniaceae, *Austrobaileya*, *Illicium*, and, to some extent, in Schisandraceae, the outermost organs on the floral axis are small, bractlike structures. However, there is no clear demarcation between these and the larger perianth members but rather a gradual transition series. This was also discussed for other magnoliids with spiral floral phyllotaxis (Endress 1980*a*, 1980*c*; Endress and Sampson 1983). In *Amborella* and Trimeniaceae, the first two organs of lateral flowers of the botryoids were not counted as tepals, as they have a lateral position like prophylls, and there seems to be a longer plastochron between phyllomes 2 and 3. However, the question of where the floral organs begin needs comparative de-

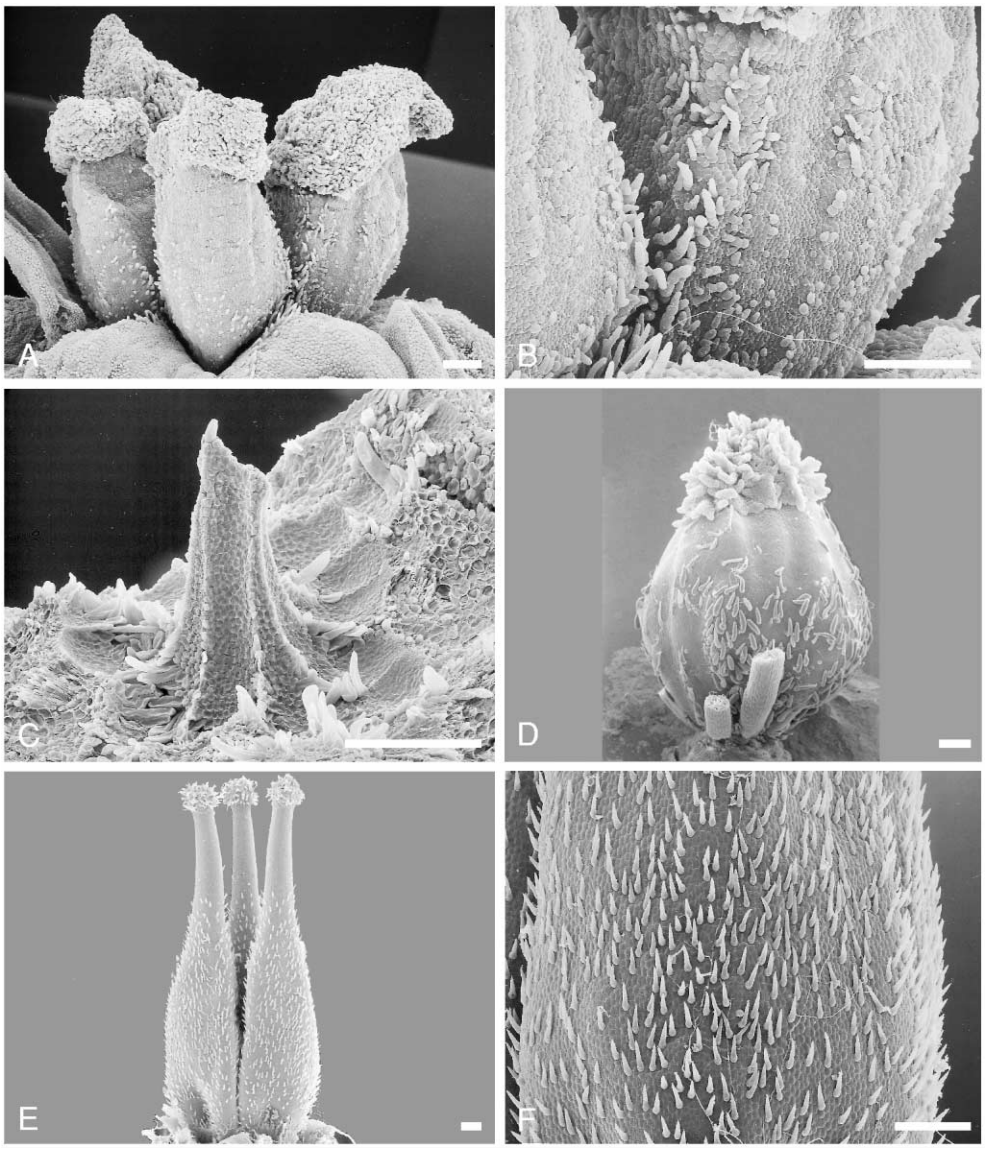


Fig. 14 Hairs on carpels. *A–C*, *Amborella trichopoda*. *A*, Gynoecium at anthesis with barrel-shaped carpels, showing longitudinal grooves and ridges and lines of hairs preferentially on the ridges. *B*, Same as *A* in higher magnification. *C*, Rudimentary structure (gynoecium?) in center of male flower, showing longitudinal grooves and ridges with lines of hairs. *D*, *Trimenia papuana* (Trimeniaceae); gynoecium before anthesis, showing longitudinal grooves and ridges and lines of hairs preferably on the ridges. *E*, *F*, *Cabomba furcata* (Cabombaceae). *E*, Gynoecium at anthesis, covered with hairs. *F*, Same as *E* in higher magnification. Magnification bars: *A–F* = 0.2 mm.

velopmental studies for the other taxa with spiral floral phyllotaxis as well.

Differentiation of the perianth into sepals and petals is most pronounced in eudicots. "Typical" petals are colored, have a single vascular trace, are retarded in development as compared to sepals and stamens, and are ephemeral. Among magnoliids, perianth organs with these attributes are lacking, although there is often a differentiation into outer, green "sepaloid" tepals and inner, colored "petaloid" tepals. It is noteworthy, however, that among magnoliids, it is only in *Cabomba*, one of the basalmost genera in extant angiosperms, that the perianth behaves almost as in eudicots! Also in Nymphaeaceae, esp. *Nuphar*, "petals" with a conspicuously retarded development are present (Hiepko 1965). A perianth is lacking altogether in *Ceratophyllum* and most Chloranthaceae (except for *Hedyosmum*).

Another unusual feature in Nymphaeaceae, in which they resemble eudicots and not magnoliids, is the occurrence of blue perianth organs containing anthocyanins (Fossen et. al. 1998; Fossen and Andersen 1999). Blue pigments are otherwise lacking in basal angiosperms (Gottsberger and Gottlieb 1981). In addition, in *Cabomba* the nectariferous parts of the petals are strongly UV absorbing, whereas the peripheral parts are only weakly absorbing (Burr and Barthlott 1993). Otherwise, there are no conspicuous contrasting patterns within individual perianth parts in the genera studied (*Illicium*, *Kadsura*, *Euryale*,

Victoria [Burr and Barthlott 1993] and *Nymphaea* and *Nuphar* [Giesen and Van der Velde 1983; Burr and Barthlott 1993; Langanger et al. 2000]).

Tepals fall off when the flowers open in Trimeniaceae and after anthesis in *Austrobaileya*, Schisandraceae, and *Illicium*. They have a narrow tepal attachment region at the floral base, in contrast to *Amborella*, Cabombaceae, and Nymphaeaceae, in which the tepals are not caducous.

In general, the morphological distinction between sepals and petals is less simple than it may seem from consideration of typical cases and from the original ABC model of flower development (Coen and Meyerowitz 1991). In contrast to stamens and carpels, which are clearly defined by their male and female structure and function, sepals and petals cannot unambiguously be defined by their structures and functions (Endress 1994*a*). It is to be hoped that future developmental genetic studies, combined with comparative structural studies, may give deeper insight into the complexity of this problem (Baum and Whitlock 1999; Kramer and Irish 2000; Theissen 2001).

Androecium

The presence of relatively more extrorse anthers in magnoliids and basal monocots than in other angiosperms (Endress 1996) may suggest that an extrorse shape is plesiomorphic in angiosperms. However, with the focus on the ANITA grade, this hypothesis is not supported. Anthers are introrse in *Amborella*, Nymphaeaceae, *Austrobaileya* and Illiciaceae, and also Chloranthaceae (Endress and Hufford 1989; Endress 1994*d*). The presence of introrse and extrorse anthers in the same family, as in Cabombaceae, Trimeniaceae, and Schisandraceae, suggests rather that the direction of the thecae is evolutionarily plastic at low systematic levels.

A similar case is the frequent presence of valvate anther dehiscence in magnoliids (Endress and Hufford 1989) and in Cretaceous floral fossils (Friis et al. 1991), which suggested that this is a plesiomorphic feature for angiosperms. In contrast, in the ANITA grade, anthers have simple longitudinal dehiscence. The only exception is *Nuphar* (Nymphaeaceae), in which H-shaped valves were found (Hufford 1996). *Nuphar* also seems to be the only member of the ANITA grade in which the anthers have a thick and broad apex. In general, valvate dehiscence occurs only in anthers with nonprotruding thecae, thus, in thecae with a thick connective, and especially also in anthers with a thick and broad apex (Endress and Hufford 1989; Hufford and Endress 1989; Endress 1994*a*). In the ANITA grade, pollen sacs are conspicuously bulging in *Amborella*, *Austrobaileya*, Schisandraceae, and some *Chloranthus* species, or more or less bulging in *Cabomba*, some Nymphaeaceae, Trimeniaceae, and *Illicium* (the definition of "bulging" may be problematic when the thecae are latrorse). The evolutionary interpretation is that valvate dehiscence is not plesiomorphic in angiosperms, but a trend to form anthers with thick connectives did occur in Magnoliales and Laurales, which then led to the differentiation of valvate dehiscence in several clades. In higher-evolved angiosperms, anthers are more elaborated, with less sterile tissue, i.e., with thin connectives and, thus, with protruding thecae, in which valvate dehiscence is not possible for architectural reasons (Hufford and Endress 1989; Endress 1994*a*).

Inner staminodes occur in *Austrobaileya* and in the male flowers of *Amborella* (exceptionally), some Schisandraceae, and *Ceratophyllum*. In the female flowers of *Amborella*, the presence of staminodes demonstrates the fundamentally bisexual organization of the flowers.

Gynoecium

Angiospermy by secretion, and without postgenital carpel fusion (angiospermy type 1; Endress and Igersheim 1997*b*, 2000*a*), is the most common condition in the ANITA grade and is also present in Chloranthaceae and Ceratophyllaceae. The only exceptions are *Illicium* with angiospermy type 2 (closure by partial postgenital fusion at the periphery but with a complete, secretion-filled canal) and more advanced members of the Nymphaeales with angiospermy type 3 (closure by complete postgenital fusion at the periphery and with an incomplete, secretion-filled canal). More or less correlated with angiospermy type 1 is the pronouncedly ascidiate form of the carpels. Other magnoliids, including groups that were considered most primitive in earlier decades, such as Magnoliaceae, Degeneriaceae, Winteraceae, all have postgenitally fused carpels (angiospermy type 4). Angiospermy types 2–4 are the most common conditions not only in Magnoliales but also in Piperales and most Laurales (Endress and Igersheim 1997*a*, 1997*b*, 2000*a*; Igersheim and Endress 1997, 1998); they are also predominant in basal eudicots (Endress and Igersheim 1999). The situation is more ambiguous in basal monocots (Igersheim et al. 2001). Thus, it looks as though angiospermy came about first by secretion and only secondarily by postgenital fusion.

The carpels are pronouncedly ascidiate in most taxa. In early development, they appear as oblique bowl-like structures. The opening is then lifted up by intercalary elongation of the basal part. This results in a tubular shape (cf. also Leinfellner 1969; Taylor 1991). The stigmatic region surrounds the opening in *Amborella*, *Cabomba*, *Austrobaileya*, and Trimeniaceae (also in Chloranthaceae and *Ceratophyllum*). Most interestingly, all these families have the least elaborated architecture of the entire gynoecium among the basalmost angiosperms. Thus, this strongly supports its plesiomorphic status for angiosperms. In *Amborella*, Nymphaeaceae, Schisandraceae, and *Illicium*, the carpels strongly curve toward the floral center in early stages and the entrance (opening) is then hidden. It remains more or less hidden (or at least its lowermost part) up to anthesis in Nymphaeaceae, Schisandraceae, and *Illicium* (but not in *Amborella*). In groups with more elaborate floral centers, the carpels are less obviously ascidiate, especially in *Illicium* and in most Nymphaeaceae, which both have a central hump around which the carpels are grouped. The carpels turn their opening toward the hump, and together they form a peculiar extragynoecial compitum; to some degree this is also the case in Schisandraceae, but if a central protrusion is present at all, it does not take part in the compitum (Tucker and Bourland 1994; Endress and Igersheim 2000*a*). In these more specialized (synorganized) gynoecia with carpel openings that face the floral center, the carpel wall is retarded on the ventral side in earlier development (Schisandraceae), or it remains more or

less reduced up to anthesis (*Illicium*, Nymphaeaceae). Thus, a trait of several taxa of the ANITA grade is the tendency to form extragynoecial compita that provide a potential communication medium for pollen tubes. It occurs in all three ANITA clades in different ways. It is inconspicuous in *Amborella*, where young female flowers simply have contiguous secretory stigmas, but it is very conspicuous in *Austrobaileya*, Schisandraceae, *Illicium*, and *Nymphaea* (Endress and Igersheim 2000*a*).

The stigma in *Amborella*, Trimeniaceae, and a few Chloranthaceae has irregular, pluricellular, pluriseriate protrusions. In Cabombaceae, they are bi- to pluricellular, uniseriate, and in some Nymphaeaceae, they are pluricellular, uniseriate. In Austrobaileyaceae, Schisandraceae, and Illiciaceae the stigma is unicellular papillate. In some Chloranthaceae and Ceratophyllaceae, it is smooth. Thus, the unusual concentration of stigmas with multicellular protrusions in the ANITA grade is noteworthy.

A massive hump in the center of flowers is present: (1) in bisexual flowers of Nymphaeaceae and *Illicium*, very similar in both, and in an Early Cretaceous fossil nymphaealean flower (Friis et al. 2001); (2) in unisexual (male) flowers of *Amborella* and Schisandraceae. Strigose hairs are formed on the carpels of *Amborella*, Trimeniaceae, and *Cabomba*. Carpels of *Amborella* and Trimeniaceae have longitudinal furrows and ribs caused by pressure of stamens in bud, with rows of hairs between the sites of pressure (fig. 14*A*–14*D*).

Archaefructus, an enigmatic compressed fossil with structures suggestive of conduplicate carpels (Sun et al. 1998), would be interesting for further study. Originally thought to be of Jurassic age (Sun et al. 1998), it is more likely Lower Cretaceous (Barremian or Aptian) (Swisher et al. 1999; Barrett 2000). However as long as only the surface shape and no structural details are known, it cannot even be ascertained whether it is an angiosperm (or closely related angiophyte), and therefore, it is of limited help for evolutionary hypotheses on gynoecium evolution.

Nectaries (and Other Floral Secretory Structures)

Secretion of nectar in flowers has been recorded for some representatives of the ANITA grade. However, in basal angiosperms, there are no highly differentiated nectaries, such as disk nectaries in eudicots or septal nectaries in monocots. In *Cabomba*, each petal has two lateral nectariferous auricles (the tissue without intercellular spaces and the cells with dense cytoplasm, with a concentration of glandular protrusions; Vogel 1998). In *Nuphar*, petals are nectariferous on their dorsal side (Müller 1893; Hiepko 1965; Lippok and Renner 1997). In *Kadsura japonica*, nectar is secreted from the inner surface of the inner tepals (Saunders 1998). In *Illicium*, very small quantities of nectar were found at the base of tepals and stamens (Thien et al. 1983, 2000; White and Thien 1985). In some other magnoliids, nectarioles are present, i.e., several small areas on an organ that secrete nectar at the surface or subepidermally and release it through a stoma; they occur on tepals in *Aristolochia* and *Chimonanthus* and on subtending bracts in *Peperomia* and *Piper* (Vogel 1998). Nectarioles with stomata also seem to occur in species of *Liriodendron* and *Magnolia* (Müller 1893; Daumann 1930). In *Asimina*, the base of the inner tepals secretes nectar (Baillon 1866); it is ridged and has a continuous secretory epidermis without stomata (personal observation). A comparative study of nectar-secreting structures in the basal angiosperms would be of interest.

It is unclear what is secreted by the connective tips of *Amborella* (Endress and Igersheim 2000*b*) and also whether the secretions of the wet stigmas of some representatives of the ANITA grade are consumed by pollinators (e.g., *Amborella, Austrobaileya*, Schisandraceae, *Illicium, Cabomba*, Nymphaeaceae).

Floral Biology, Breeding Systems

Protogyny is omnipresent in those basal angiosperms that have bisexual flowers (in all magnoliids, plus basal monocots and eudicots) (Endress 1984*a*, 1990*a*, 1992; Lloyd and Webb 1986; Bernhardt and Thien 1987; von Balthazar and Endress 1999; Buzgo and Endress 2000; Thien et al. 2000). There are reports of homogamy in some *Nymphaea* and *Nuphar* species (Heslop-Harrison 1955*a*, 1955*b*; Wiersema 1988; Orban and Bouharmont 1995). However, it was not specified by these authors whether anthers open at the same time as stigma receptivity begins or just some time on the first day of anthesis (which could then still be after onset of receptivity).

Breeding systems are unknown in *Amborella*, Trimeniaceae, Schisandraceae, Cabombaceae, and *Ceratophyllum*. The few taxa studied in Nymphaeaceae are all self-compatible (*Nuphar lutea* [Ervik et al. 1995], *Nymphaea capensis* [Orban and Bouharmont 1995], and the partially cleistogamous *Euryale* and *Barclaya* [Kadono and Schneider 1987; Williamson and Schneider 1994]; see also Weller et al. 1995). In a review article, Brewbaker (1959) indicated the occurrence of self-incompatibility in Nymphaeaceae, but without giving a reference. Among the ANITA grade, self-incompatibility is known only from *Austrobaileya* (Prakash and Alexander 1984; P. K. Endress, personal observation) and *Illicium floridanum* (Thien et al. 1983). In Chloranthaceae, self-incompatibility was found in *Chloranthus spicatus* (von Balthazar and Endress 1999), whereas in other *Chloranthus* species and in *Sarcandra glabra*, self-compatibility is present (Ma et al. 1997; Luo and Li 1999; von Balthazar and Endress 1999; Wang et al. 1999). Agamospermy has not been reported from any member of the ANITA grade. However, it was found in *Sarcandra chloranthoides* (von Balthazar and Endress 1999).

Duration of anthesis of single flowers is unknown in some families of the ANITA grade, such as *Amborella*, Trimeniaceae, and Schisandraceae. In *Austrobaileya*, it is ca. 5 d (P. K. Endress, preliminary observations), in Cabombaceae 2 d, in Nymphaeaceae mostly 2 or 3 d, rarely 4 or more, and in *Illicium* 2–3 d or up to 14 d. In *Sarcandra* and *Chloranthus*, it is ca. 6–11 d or more (Luo and Li 1999; von Balthazar and Endress 1999). Anthesis duration is unknown in *Ceratophyllum*.

Pollination biology of the ANITA grade has been very unevenly studied (reviews in Endress 1990*a*; Thien et al. 2000). *Amborella* has only been observed in cultivation and has been found to produce a floral scent at night, which attracts moths (Collett 1999). Small moths (and beetles and bees) were also observed in cultivated *Schisandra chinensis* (Kozo-Poljanski 1946). Flies as major pollinators were reported in *Cabomba*

(Schneider and Jeter 1982), *Nuphar* (Lippok and Renner 1997), *Austrobaileya* (Endress 1980*c*), and *Illicium floridanum* (Thien et al. 1983, 2000). Beetles play a role as pollinators in Nymphaeaceae, especially in large-flowered species of *Nymphaea* and in *Victoria* but less so in smaller-flowered taxa of the ANITA grade; they are also important in some highly specialized flowers of other magnoliids (review in Thien et al. 2000). Bees as predominant pollinators are known only from some Nymphaeaceae among the ANITA grade. Thrips seem to be important pollinators of *Chloranthus* (Luo and Li 1999). It may be expected that they also play a role in Schisandraceae, which have narrow, slitlike gates between the partially united stamens and between the carpels. Thrips as pollinators have also been observed in other magnoliids (e.g., Gottsberger 1977, 1988, 1999; Gottsberger et al. 1980; Thien 1980; Jürgens et al. 2000; Thien et al. 2000; Williams et al. 2001) and some cycads (Mound and Terry 2001). They may play a larger role than previously supposed in pollination of basal angiosperms. Among the ANITA grade, wind pollination was only found for the water plant *Brasenia* (Osborn and Schneider 1988). It is unclear whether in *Amborella* and Trimeniaceae wind may also be at least partly involved in pollination. Preliminary field observations by T. Feild (personal communication) suggest that wind pollination plays a role in *Amborella*. In Chloranthaceae, *Hedyosmum* is clearly wind pollinated, probably also *Ascarina* (cf. Endress 1987*a*).

Flowers of *I. floridanum* were found to be slightly thermogenic (Dieringer et al. 1999; Thien et al. 2000). Thien et al. (2000) hypothesized that even only small differences in flower temperature from the ambient temperature could be attractive for flies and other insects and may be more common than previously supposed. In earlier studies, only highly thermogenic flowers (or inflorescences) were investigated, such as, among magnoliids, *Magnolia* (Azuma 1999; Dieringer et al. 1999), *Annona*, and *Rollinia* (Gottsberger 1989). From the results of Dieringer et al. (1999) in *Magnolia tamaulipana*, it seems that there is selection pressure on the production of larger flowers, as they were significantly more frequently visited by pollinators (beetles) than smaller flowers, probably due to more scent and heat production. Floral scent may have played a prominent role in early flower evolution (Pellmyr and Thien 1986). However, comparative studies of floral scents in basal angiosperms are lacking.

The especially early presence in the fossil record of insects that play a major role in the pollination of basal angiosperms, such as flies, thrips, and primitive moths, and the minor role of younger insect groups, such as bees or syrphids (Ren 1998; Grimaldi 1999; Mound and Terry 2001), indicates that the former insect groups indeed also could have been major pollinators of ancestral angiosperms and could have played a role in the early radiation of angiosperms. However, it should be emphasized once more that pollination biology and breeding systems in some families of the ANITA grade are largely unknown and studies are much needed. It is still an enigma when self-incompatibility evolved in angiosperms or in their ancestors (Weller et al. 1995). In nonangiosperm seed plants, prezygotic self-incompatibility seems to be largely lacking. The only incidence reported seems to be in the conifer *Picea glauca* (Runions and Owens 1998).

Potential Plesiomorphic Features in Extant Basal Angiosperms and Inferences on Early (Ancestral) Angiosperm Flowers

Does the recognition of the ANITA grade change the paradigm for primitive flowers? It cannot be expected that the first angiosperm flowers 130 million years ago or so resembled any extant representative of the ANITA grade in every respect. One may object that all extant plants have the same age, and therefore may all have evolved equally far from the common ancestor. However, we have to take into account that the phylogenetic tree is not symmetrically branched. On the contrary, it appears to be highly asymmetrical. Some clades have undergone repeated extensive radiations, with the result that their descendants have surely undergone radical changes, whereas others have radiated less and may have remained more conservative in their features. The latter is precisely what is indicated by the topology of the basal ANITA grade. These taxa seem not to have radiated much; they are the last remnants of their clades, but it is not likely that they are the remnants of phylogenetic branches that have undergone repeated extensive radiations. Thus, the extant taxa of the ANITA grade may indeed have kept more ancestral features than any other extant taxa of the angiosperms, even if these "basal" lines have each undergone great changes. But what are these features? We are lucky that not only a single clade but a grade consisting of two or three clades has been recognized as the basalmost part of angiosperms. Since this allows parsimony optimization of basal states on cladograms, it can be assumed that states found in more than one clade of the basal grade are likely to be plesiomorphic.

As diverse as some of these basal angiosperm flowers may look at first sight, we should ask: Are there taxa in the basalmost extant angiosperms (ANITA grade) in which the flowers resemble each other to a high degree—without being sister taxa—or are there unique, shared features that are not present in other "magnoliids"? Such resemblances or such shared features could indicate plesiomorphic traits that might well also characterize flowers of ancestral angiosperms of the Lower Cretaceous. The more unique they are, the less they are likely to be convergences. Such features may not have been noticed before, or they may not have attracted much attention, although they have been known earlier.

Amborella and *Trimenia/Piptocalyx* are such a pair of especially similar taxa within the ANITA grade, without being sister taxa. In addition, *Ascarina* (Chloranthaceae) bears special similarities, especially in the gynoecium, with those. This resemblance is also of interest because Chloranthaceae are pulled down in morphological analyses between *Amborella* and the rest of the ANITA grade, if *Amborella* is chosen as sister to all other angiosperms (Doyle and Endress 2000) and because Chloranthaceae are an important element in the early angiosperm fossil record (Friis et al. 2000). Thus, there is some probability that these genera exhibit special plesiomorphic features in their floral structure not present in other angiosperms, as also suggested earlier (Endress 1986).

All the other groups are more isolated in their floral structure, without any similar counterpart. This is true for Cabombaceae, Nymphaeaceae, *Austrobaileya*, and Illiciales. Schisandraceae and *Illicium* are more similar to each other, which

is not surprising because they form a clade, as recognized long ago.

Amborella and Trimeniaceae flowers are similar in the following features: inflorescences botryoids; flowers tiny, with bisexual organization (if functionally unisexual, with rudiments of organs of the other gender); floral organs spirally arranged; organ number variable and low to medium; perianth simple, of small, whitish tepals; stamens whitish, with short filaments; anthers with short connective protrusion; thecae slightly or strongly protruding, longitudinally dehiscing; carpels barrel shaped, styleless, with longitudinal grooves produced by the appressed stamens in bud, covered with (often more than one cellular) hairs that are directed upward; angiospermy by secretion, and not by postgenital fusion (type 1; cf. Endress and Igersheim 2000*a*), however, surfaces appressed to each other, forming a narrow slit; extremely ascidiate up to the stigma; stigma capitate, with coarse "papillae" formed by irregular multicellular, multiseriate protrusions, with a single ventral-median, pendant, bitegmic, crassinucellar ovule. *Ascarina* has similar but glabrous carpels. However, its flowers are almost always unisexual, devoid of a perianth, and contain only a single stamen or a single carpel. Trimeniaceae, Cabombaceae, and Nymphaeaceae share a special kind of hair on the carpels (fig. 14*A*–14*F*). The hairs are strigose, uniseriate, three- or four-cellular, with a long, tanniferous terminal cell and two or three very short basal cells (Endress and Igersheim 1997*a*; Igersheim and Endress 1998). In *Amborella*, hairs on carpels are one- or two-cellular, and variously tanniferous or not so.

What can we tentatively conclude for ancestral flowers after evaluation of the entire ANITA grade, as compared with other magnoliids (cf. also Doyle and Endress 2000)? Flowers most likely were small, with moderate or low numbers of organs that were spirally arranged (or whorled as in Nymphaeales), protogynous. Carpels were probably free, styleless, extremely ascidiate, with one or only a few ovules; inner space closed by secretion and not by postgenital fusion; wet stigmas with pluricellular protrusions. Pollination was probably by small insects, especially dipters (and thrips, moths?). The presence of small flowers in the ANITA grade is in accordance with the fossil record of Early Cretaceous flowers (Friis and Endress 1990; Crane et al. 1995; Dilcher 2000; Friis et al. 2000). It also fits with the hypothesis of the role of paedomorphic innovations at the beginning of successful new clades (Takhtajan 1976). The evolution of angiosperm carpels and ovules may have resulted by progenesis (abbreviation of development of precursor organs; for term, see Alberch et al. 1979). Potential ancestors of angiosperms, such as glossopterids or Caytoniales, had larger and less compact reproductive organs (Doyle 1998; Frohlich and Parker 2000).

There are some labile or variable traits among the ANITA grade, which makes inferences on basic states more difficult. Flowers are bisexual but easily become unisexual perhaps because of the low degree of synorganization between androecium and gynoecium at this evolutionary level. Floral organ number is variable, even within an individual. Both self-incompatibility and self-compatibility occur in the ANITA grade; agamospermy is known from Chloranthaceae (in addition to self-incompatibility and self-compatibility) (von Balthazar and Endress 1999) but not from any family of the ANITA grade.

Conspicuous trends of specialization in the ANITA grade include the increase of flower size and organ number and the tendency to form extragynoecial compita in various ways (in absence of normal syncarpy). A few genera have become specialized for wind pollination. Beetle pollination has evolved in some large-flowered Nymphaeaceae, combined with a shorter anthesis of only two nights. When these trends arose in the history of angiosperms cannot be answered at present.

Acknowledgments

This comparative study would not have been possible without the great support in the field and the provision of additional specimens by P. Burgess, A. Floyd, B. Gray, B. P. M. Hyland, H. S. MacKee, and G. McPherson; my thanks go to all of them; H. Tobe is thanked for sending material. I am indebted to J. A. Doyle for critical discussions on basal angiosperms and for reading the manuscript, to E. M. Friis for information on fossils, and to T. Feild for communication of field observations. I thank A. Doust, S. C. Tucker, and an anonymous referee for reviewing the manuscript. I thank R. Siegrist for preparing specimens, U. Jauch for support with the SEM, A. Zuppiger for photography, and A. Bernhard for helping with the tables.

Literature Cited

Alberch P, SJ Gould, GF Oster, DB Wake 1979 Size and shape in ontogeny and phylogeny. Paleobiology 5:296–317.

Azuma H, LB Thien, S Kawano 1999 Floral scents, leaf volatiles and thermogenic flowers in Magnoliaceae. Plant Species Biol 14:121–127.

Bailey IW, BGL Swamy 1948 *Amborella trichopoda* Baill., a new morphological type of vesselless dicotyledon. J Arnold Arbor Harv Univ 29:245–254.

——— 1949 The morphology and relationships of *Austrobaileya*. J Arnold Arbor Harv Univ 30:211–226.

Baillon H 1866 Sur des pétales à structure anormale. Adansonia 6: 253–254.

Barkman TJ, G Chenery, JR McNeal, J Lyons-Weiler, WJ Elisens, G Moore, AD Wolfe, CW 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.

Barrett PM 2000 Evolutionary consequences of dating the Yixian formation. Trends Ecol Evol 15:99–104.

Baum DA, BA Whitlock 1999 Genetic clues to petal evolution. Curr Biol 9:525–527.

Beal EO 1956 Taxonomic revision of the genus *Nuphar* Sm. of North America and Europa. J Elisha Mitchell Sci Soc 72:317–346.

Bernhardt P 2000 Convergent evolution and adaptive radiation of beetle-pollinated angiosperms. Plant Syst Evol 222:293–320.

Bernhardt P, LB Thien 1987 Self-isolation and insect pollination in the primitive angiosperms: new evaluations of older hypotheses. Plant Syst Evol 156:159–176.

Borsch T, KW Hilu, V Wilde, C Neinhuis, W Barthlott 2000 Phylogenetic analysis of noncoding chloroplast DNA sequences reveals *Amborella* as basalmost angiosperm. Am J Bot 87 6(suppl): 115–116.

Brenner GJ 1996 Evidence for the earliest stage of angiosperm pollen

evolution: a paleoequatorial section from Israel. Pages 91–115 *in* DW Taylor, LJ Hickey, eds. Flowering plant origin, evolution and phylogeny. Chapman & Hall, New York.

Brewbaker JL 1959 Biology of the angiosperm pollen grain. Indian J Genet Plant Breed 19:121–133.

Burr B, W Barthlott 1993 Untersuchungen zur Ultraviolettreflexion von Angiospermenblüten. II. Magnoliidae, Ranunculidae, Hamamelididae, Caryophyllidae, Rosidae. Trop Subtrop Pflanzenwelt 87: 1–193.

Buzgo M, PK Endress 2000 Floral structure and development of Acoraceae and its systematic relationships with basal angiosperms. Int J Plant Sci 161:23–41.

Capperino ME, EL Schneider 1985 Floral biology of *Nymphaea mexicana* Zucc. (Nymphaeaceae). Aquat Bot 23:83–93.

Chase MW, DE Soltis, RG Olmstead, D Morgan, DH Les, BD Mishler, MR Duvall, et al 1993 Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann Mo Bot Gard 80:526–580.

Chifflot J-B-J 1902 Contributions à l'étude de la classe des Nymphéinées. Ann Univ Lyon, NS, I, Sci, Méd 10:1–294.

Coen ES, EM Meyerowitz 1991 The war of the whorls: genetic interactions controlling flower development. Nature 353:31–37.

Collett R 1999 Arboretum collection makes headline news. Bull UCSC Arbor Bot 46:603–632.

Couper RA 1958 British Mesozoic microspores and pollen grains: a systematic and stratigraphic study. Palaeontogr Abt B 103:75–179.

Cramer JM, ADJ Meeuse, PA Teunissen 1975 A note on the pollination of nocturnally flowering species of *Nymphaea*. Acta Bot Neerl 24:489–490.

Crane PR, EM Friis, KR Pedersen 1989 Reproductive structure and function in Cretaceous Chloranthaceae. Plant Syst Evol 165:211–226.

——— 1995 The origin and early diversification of angiosperms. Nature 374:27–33.

Cronquist A 1988 The evolution and classification of flowering plants. 2d ed. New York Botanical Garden, Bronx, N.Y.

Daumann E 1930 Das Blütennektarium von *Magnolia* und die Futterkörper in der Blüte von *Calycanthus*. Planta 11:108–116.

den Hartog C 1970 *Ondinea*, a new genus of Nymphaeaceae. Blumea 18:413–417.

Dieringer G, RL Cabrera, M Lara, L Loya, P Resyes-Castillo 1999 Beetle pollination and floral thermogenicity in *Magnolia tamaulipana* (Magnoliaceae). Int J Plant Sci 160:64–71.

Dilcher DL 2000 Toward a new synthesis: major evolutionary trends in the angiosperm fossil record. Proc Natl Acad Sci USA 97: 7030–7036.

Donoghue MJ, JA Doyle 1989 Phylogenetic analysis of angiosperms and the relationships of Hamamelidae. Pages 17–45 *in* PR Crane, S Blackmore, eds. Evolution, systematics, and fossil history of the Hamamelidae. Vol 1. Clarendon, Oxford.

Doyle JA 1969 Cretaceous angiosperm pollen of the Atlantic coastal plain and its evolutionary significance. J Arnold Arbor Harv Univ 50:1–35.

——— 1977 Patterns of evolution in early angiosperms. Pages 501–546 *in* A Hallam, ed. Patterns of evolution. Elsevier, Amsterdam.

——— 1998 Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. Mol Phylogenet Evol 9:448–462.

Doyle JA, P Biens, A Doerenkamp, S Jardiné 1977 Angiosperm pollen from the pre-Albian Lower Cretaceous of equatorial Africa. Bull Cent Rech Explor-Prod Elf-Aquitaine 1:451–473.

Doyle JA, MJ Donoghue, EA Zimmer 1994 Integration of morphological and ribosomal RNA data on the origins of angiosperms. Ann Mo Bot Gard 81:405–418.

Doyle JA, PK Endress 2000 Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. Int J Plant Sci 161(suppl):S121–S153.

Doyle JA, CL Hotton 1991 Diversification of early angiosperm pollen in a cladistic context. Pages 169–195 in S Blackmore, SH Barnes, eds. Pollen and spores: patterns of diversification. Clarendon, Oxford.

Eklund H 1999 Big survivors with small flowers: fossil history and evolution of Laurales and Chloranthaceae. PhD diss. Uppsala University. Acta Universitatis Upsaliensis, Uppsala.

Endress PK 1971 Bau der weiblichen Blüten von *Hedyosmum mexicanum* Cordemoy (Chloranthaceae). Bot Jahrb Syst 91:39–60.

——— 1980*a* Floral structure and relationships of *Hortonia* (Monimiaceae). Plant Syst Evol 133:199–221.

——— 1980*b* Ontogeny, function and evolution of extreme floral construction in Monimiaceae. Plant Syst Evol 134:79–120.

——— 1980*c* The reproductive structures and systematic position of the Austrobaileyaceae. Bot Jahrb Syst 101:393–433.

——— 1982 Syncarpy and alternative modes of escaping disadvantages of apocarpy in primitive angiosperms. Taxon 31:48–52.

——— 1983 The early floral development of *Austrobaileya*. Bot Jahrb Syst 103:481–497.

——— 1984*a* The flowering process in the Eupomatiaceae (Magnoliales). Bot Jahrb Syst 104:297–319.

——— 1984*b* The role of inner staminodes in the floral display of some relic Magnoliales. Plant Syst Evol 146:269–282.

——— 1986 Reproductive structures and phylogenetic significance of extant primitive angiosperms. Plant Syst Evol 152:1–28.

——— 1987*a* The Chloranthaceae: reproductive structures and phylogenetic position. Bot Jahrb Syst 109:153–226.

——— 1987*b* Floral phyllotaxis and floral evolution. Bot Jahrb Syst 108:417–438.

——— 1990*a* Evolution of reproductive structures and functions in primitive angiosperms (Magnoliidae). Mem N Y Bot Gard 55:5–34.

——— 1990*b* Patterns of flower construction in ontogeny and phylogeny. Biol J Linn Soc 39:153–175.

——— 1992 Protogynous flowers in Monimiaceae. Plant Syst Evol 181:227–232.

——— 1994*a* Diversity and evolutionary biology of tropical flowers. Cambridge University Press, Cambridge.

——— 1994*b* Evolutionary aspects of the floral structure in *Ceratophyllum*. Plant Syst Evol Suppl 8:175–183.

——— 1994*c* Floral structure and evolution of primitive angiosperms: recent advances. Plant Syst Evol 192:79–97.

——— 1994*d* Shapes, sizes and evolutionary trends in stamens of Magnoliidae. Bot Jahrb Syst 115:429–460.

——— 1996 Diversity and evolutionary trends in angiosperm anthers. Pages 92–110 *in* WG D'Arcy, RC Keating, eds. The anther: form, function and phylogeny. Cambridge University Press, Cambridge.

Endress PK, L Hufford 1989 The diversity of stamen structures and dehiscence patterns among Magnoliidae. Bot J Linn Soc 100:45–85.

Endress PK, A Igersheim 1997*a* Gynoecium diversity and systematics of the Laurales. Bot J Linn Soc 125:93–168.

——— 1997*b* Patterns of angiospermy in basal angiosperms. Am J Bot 84(suppl):190.

——— 1999 Gynoecium diversity and systematics of the basal eudicots. Bot J Linn Soc 130:305–393.

——— 2000*a* Gynoecium structure and evolution in basal angiosperms. Int J Plant Sci 161(suppl):S211–S223.

——— 2000*b* The reproductive structures of the basal angiosperm *Amborella trichopoda* (Amborellaceae). Int J Plant Sci 161(suppl): S237–S248.

Endress PK, FB Sampson 1983 Floral structure and relationships of the Trimeniaceae (Laurales). J Arnold Arbor Harv Univ 64:447–473.

Erbar C 1983 Zum Karpellbau einiger Magnoliiden. Bot Jahrb Syst 104:3–31.

Erbar C, P Leins 1983 Zur Sequenz von Blütenorganen bei einigen Magnoliiden. Bot Jahrb Syst 103:433–449.

Ervik F, SS Renner, KA Johanson 1995 Breeding system and polli-

nation of *Nuphar luteum* (L.) Smith (Nymphaeaceae) in Norway. Flora 190:109–113.

Feild TS, MA Zweiniecki, T Brodribb, T Jaffré, MJ Donoghue, NM Holbrook 2000 Structure and function of tracheary elements in *Amborella trichopoda*. Int J Plant Sci 161:705–712.

Fossen T, OM Andersen 1999 Delphinidin 3-galloyl-galactosides from blue flowers of *Nymphaea caerulea*. Phytochemistry 50:1185–1188.

Fossen T, A Larsen, OM Andersen 1998 Anthocyanins from flowers and leaves of *Nymphaea* × *marliacea* cultivars. Phytochemistry 48: 823–827.

Friis EM, PR Crane, KR Pedersen 1986 Floral evidence for Cretaceous chloranthoid angiosperms. Nature 320:163–164.

——— 1991 Stamen diversity and *in situ* pollen of Cretaceous angiosperms. Pages 197–224 *in* S Blackmore, SH Barnes, eds. Pollen and spores, patterns of diversification. Clarendon, Oxford.

Friis EM, PK Endress 1990 Origin and evolution of angiosperm flowers. Adv Bot Res 17:99–162.

Friis EM, KR Pedersen, PR Crane 1999 Early angiosperm diversification: the diversity of pollen associated with angiosperm reproductive structures in Early Cretaceous floras from Portugal. Ann Mo Bot Gard 86:259–296.

——— 2000 Reproductive structure and organization of basal angiosperms from the Early Cretaceous (Barremian or Aptian) of Portugal. Int J Plant Sci 161(suppl):S169–S182.

——— 2001 Fossil evidence of water lilies in the Early Cretaceous. Nature 410:357–360.

Frohlich MW, Parker DS 2000 The mostly male theory of flower evolutionary origins: from genes to fossils. Syst Bot 25:155–170.

Giesen TG, G Van der Velde 1983 Ultraviolet reflectance and absorption patterns in flowers of *Nymphaea alba* L., *Nymphaea candida* Presl, and *Nuphar lutea* (L.) Sm. (Nymphaeaceae). Aquat Bot 16:369–376.

Gottsberger G 1974 The structure and function of the primitive angiosperm flower—a discussion. Acta Bot Neerl 23:461–471.

——— 1977 Some aspects of beetle pollination in the evolution of flowering plants. Plant Syst Evol Suppl 1:211–226.

——— 1988 The reproductive biology of primitive angiosperms. Taxon 37:630–643.

——— 1989 Comments on flower evolution and beetle pollination in the genera *Annona* and *Rollinia* (Annonaceae). Plant Syst Evol 167:189–194.

——— 1999 Pollination and evolution in Neotropical Annonaceae. Plant Species Biol 14:143–152.

Gottsberger G, OR Gottlieb 1981 Blue flower pigmentation and evolutionary advancement. Biochem Syst Ecol 9:13–18.

Gottsberger G, I Silberbauer-Gottsberger, F Ehrendorfer 1980 Reproductive biology in the primitive relic angiosperm *Drimys brasiliensis* (Winteraceae). Plant Syst Evol 135:11–39.

Graham SW, RG Olmstead 2000 Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot 87: 1712–1730.

Graham SW, PA Reeves, ACE Burns, RG Olmstead 2000 Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. Int J Plant Sci 161(suppl):S83–S96.

Grimaldi D 1999 The co-radiation of pollinating insects and angiosperms in the Cretaceous. Ann Mo Bot Gard 86:373–406.

Hamby RK, EA Zimmer 1992 Ribosomal RNA as a phylogenetic tool in plant systematics. Pages 50–91 *in* PS Soltis, DE Soltis, JJ Doyle, eds. Molecular systematics in plants. Chapman & Hall, New York.

Heinsbroek PG, WA van Heel 1969 Note on the bearing of the pattern of vascular bundles on the morphology of the stamens of *Victoria amazonica* (Poepp.) Sowerby. Proc K Ned Akad Wet Ser C Biol Med Sci 72:431–444.

Heslop-Harrison Y 1955*a* Biological Flora of the British Isles: *Nuphar*. J Ecol 43:342–364.

——— 1955*b* Biological Flora of the British Isles: *Nymphaea* L. em. Sm. (nom. conserv.) J Ecol 43:719–734.

Hiepko P 1965 Vergleichend-morphologische und entwicklungsgeschichtliche Untersuchungen über das Perianth bei den Polycarpicae. Bot Jahrb Syst 84:359–508.

Hirmer M 1931 Zur Kenntnis der Schraubenstellungen im Pflanzenreich. Planta 14:132–206.

Hufford L 1996 The origin and early evolution of angiosperm stamens. Pages 58–91 *in* WG D'Arcy, RC Keating, eds. The anther: form, function and phylogeny. Cambridge University Press, Cambridge.

Hufford LD, PK Endress 1989 The diversity of anther structures and dehiscence patterns among Hamamelididae. Bot J Linn Soc 99: 301–346.

Hughes NF, GE Drewry, JF Laing 1979 Barremian earliest angiosperm pollen. Palaeontology 22:513–535.

Hughes NF, AB McDougall 1987 Records of angiospermid pollen entry into the English Early Cretaceous succession. Rev Paleobot Palynol 50:255–272.

Igersheim A, M Buzgo, PK Endress 2001 Gynoecium diversity and systematics of basal monocots. Bot J Linn Soc 136:1–65.

Igersheim A, PK Endress 1996 Functional syncarpy in the two primitive apocarpous angiosperms *Kadsura japonica* (Schisandraceae) and *Galbulimima baccata* (Himantandraceae). Abstract. Fourteenth International Congress of Sexual Plant Reproduction, February 1996, Melbourne.

——— 1997 Gynoecium diversity and systematics of the Magnoliales and winteroids. Bot J Linn Soc 124:213–271.

——— 1998 Gynoecium diversity and systematics of the paleoherbs. Bot J Linn Soc 127:289–370.

Ito M 1983 Studies in the floral morphology and anatomy of Nymphaeales. I. The morphology of vascular bundles in the flower of *Nymphaea tetragona* George. Acta Phytotax Geobot 34:18–26.

——— 1984 Studies in the floral morphology and anatomy of the Nymphaeales. II. Floral anatomy of *Nymphaea tetragona* George. Acta Phytotax Geobot 35:94–102.

——— 1986 Studies in the floral morphology and anatomy of Nymphaeales. III. Floral anatomy of *Brasenia schreberi* Gmel. and *Cabomba caroliniana* A. Gray. Bot Mag Tokyo 99:169–184.

Jérémie J 1982 Amborellaceae. Pages 157–160 *in* Flore de la Nouvelle Calédonie et dépendances. 11. Museum National d'Histoire Naturelle, Phanérogamie, Paris.

Jürgens A, AC Webber, G Gottsberger 2000 Floral scent compounds of Amazonian Annonaceae species pollinated by small beetles and thrips. Phytochemistry 55:551–558.

Kadono Y, EL Schneider 1987 The life history of *Euryale ferox* Salisb. in southwestern Japan with special reference to reproductive ecology. Plant Species Biol 2:109–115.

Kemp EM 1968 Probable angiosperm pollen from British Barremian to Albian strata. Palaeontology 11:421–434.

Keng H 1965 Observations on the flowers of *Illicium*. Bot Bull Acad Sinica 6:61–73.

Kenneally KF, EL Schneider 1983 The genus *Ondinea* (Nymphaeaceae) including a new subspecies from the Kimberley region, Western Australia. Nuytsia 4:359–365.

Kozo-Poljanski BM 1946 The floral mechanism of woo-we-dzy *Schisandra chinensis* (Turcz.) Baill. C R Acad Sci USSR 53:749–751.

Kramer EM, VF Irish 2000 Evolution of the petal and stamen developmental programs: evidence from comparative studies of the lower eudicots and basal angiosperms. Int J Plant Sci 161(suppl): S29–S40.

Kuzoff RK, CS Gasser 2000 Recent progress in reconstructing angiosperm phylogeny. Trends Plant Sci 5:330–336.

Langanger M, S Jokl, M Musso 2000 UV-reflectance in flowers of *Nymphaea alba* L. and *Nuphar lutea* (L.) Sm. (Nymphaeaceae). Aquat Bot 67:13–21.

Leinfellner W 1965 Über die Karpelle verschiedener Magnoliales. I. Oesterr Bot Z 113:383–389.
——— 1966 Über die Karpelle verschiedener Magnoliales. III. *Schisandra* (Schisandraceae). Oesterr Bot Z 113:563–569.
——— 1969 Über die Karpelle verschiedener Magnoliales. VIII. Überblick über alle Familien der Ordnung. Oesterr Bot Z 117: 107–127.
Les DH 1993 Ceratophyllaceae. Pages 246–250 *in* K Kubitzki, JG Rohwer, V Bittrich, eds. The families and genera of vascular plants. Vol 2. Springer, Berlin.
Les DH, EL Schneider, DJ Padgett, PS Soltis, DE Soltis, M Zanis 1999 Phylogeny, classification and floral evolution of water lilies (Nymphaeaceae; Nymphaeales): a synthesis of non-molecular, *rbcL*, *matK*, and 18S rDNA data. Syst Bot 24:28–46.
Lippok B, AA Gardine, PS Williamson, SS Renner 2000 Pollination by flies, bees, and beetles of *Nuphar ozarkana* and *N. advena* (Nymphaeaceae). Am J Bot 87:898–902.
Lippok B, SS Renner 1997 Pollination of *Nuphar* (Nymphaeaceae) in Europe: flies and bees rather than *Donacia* beetles. Plant Syst Evol 207:273–283.
Liu Z, A-M Lu 1999 Ontogeny of staminate and carpellate flowers of *Schisandra sphenanthera* (Schisandraceae). Acta Bot Sin 41: 1255–1258.
Lloyd DG, CJ Webb 1986 The avoidance of interference between the presentation of pollen and stigmas in angiosperms. I. Dichogamy. N Z J Bot 24:135–162.
Luo Y-B, Z-Y Li 1999 Pollination ecology of *Chloranthus serratus* (Thunb.) Roem. et Schult. and *Ch. fortunei* (A. Gray) Solms-Laub. (Chloranthaceae). Ann Bot 83:489–499.
Ma S-B, Y-H Wang, M-K Cui 1997 A contribution to the reproductive biology of *Chloranthus holostegius* (Chloranthaceae) in Mile population. Acta Bot Yunnanica 19:415–422.
Mathews S, MJ Donoghue 1999 The root of angiosperm phylogeny inferred from duplicate phytochrome genes. Science 286:947–950.
——— 2000 Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. Int J Plant Sci 161(suppl):S41–S55.
Meeuse BJD, EL Schneider 1980 *Nymphaea* revisited: a preliminary communication. Isr J Bot 28:65–79.
Mohr BAR, EM Friis 2000 Early angiosperms from the Lower Cretaceous Crato Formation (Brazil), a preliminary report. Int J Plant Sci 161(suppl):S155–S167.
Money LL, IW Bailey, BGL Swamy 1950 The morphology and relationships of the Monimiaceae. J Arnold Arbor Harv Univ 31: 372–404.
Moseley MF Jr 1958 Morphological studies on the Nymphaeaceae. I. The nature of the stamens. Phytomorphology 8:1–29.
——— 1961 Morphological studies of the Nymphaeaceae. II. The flower of *Nymphaea*. Bot Gaz 122:233–259.
——— 1965 Morphological studies of the Nymphaeaceae. III. The floral anatomy of *Nuphar*. Phytomorphology 15:54–84.
——— 1972 Morphological studies of Nymphaeaceae. VI. Development of flower of *Nuphar*. Phytomorphology 21:253–283.
Moseley MF Jr, IJ Mehta, PS Williamson, H. Kosakai 1984 Morphological studies of the Nymphaeaceae (*sensu lato*). XIII. Contributions to the vegetative and floral structure of *Cabomba*. Am J Bot 71:902–924.
Moseley MF, EL Schneider, PS Williamson 1993 Phylogenetic interpretations from selected floral vasculature characters in the Nymphaeaceae *sensu lato*. Aquat Bot 44:325–342.
Mound LA, I Terry 2001 Thrips pollination of the central Australian cycad, *Macrozamia macdonnellii* (Cycadales). Int J Plant Sci 162: 147–154.
Muller J 1970 Palynological evidence on early differentiation of angiosperms. Biol Rev 45:417–450.
Müller L 1893 Grundzüge einer vergleichenden Anatomie der Blumenblätter. Nova Acta Acad Leopold Carol 59:1–356.
Orban I, J Bouharmont 1995 Reproductive biology of *Nymphaea capensis* Thunb. var. *zanzibarensis* (Casp.) Verdc. (Nymphaeaceae). Bot J Linn Soc 119:35–43.
Ørgaard M 1991 The genus *Cabomba* (Cabombaceae)—a taxonomic study. Nord J Bot 11:179–203.
Osborn JM, EL Schneider 1988 Morphological studies of the Nymphaeaceae sensu lato. XVI. The floral biology of *Brasenia schreberi*. Ann Mo Bot Gard 75:778–794.
Osborn JM, TN Taylor, EL Schneider 1991 Pollen morphology and ultrastructure of the Cabombaceae: correlations with pollination biology. Am J Bot 78:1367–1378.
Ozenda P 1949 Recherches sur les Dicotylédones apocarpiques. Jouve, Paris.
Padgett DJ, DH Les, GE Crow 1999 Phylogenetic relationships in *Nuphar* (Nymphaeaceae): evidence from morphology, chloroplast DNA, and nuclear ribosomal DNA. Am J Bot 86:1316–1324.
Parkinson CL, KL Adams, JD Palmer 1999 Multigene analyses identify the three earliest lineages of extant flowering plants. Curr Biol 9:1485–1488.
Payer J-B 1857 Traité d'organogénie comparée de la fleur. Masson, Paris.
Pellmyr O, LB Thien 1986 Insect reproduction and floral fragrances: keys to the evolution of the angiosperms? Taxon 35:76–85.
Prakash N, JH Alexander III 1984 Self-incompatibility in *Austrobaileya scandens*. Pages 214–216 *in* EG Williams, RB Knox, eds. Pollination '84. School of Botany, University of Melbourne, Melbourne.
Prance GT 1980 A note on the pollination of *Nymphaea amazonum* Mart. & Zucc. (Nymphaeaceae). Brittonia 32:505–507.
Prance GT, AB Anderson 1977 Studies of the floral biology of Neotropical Nymphaeaceae. 3. Acta Amazonica 6:163–170.
Prance GT, JR Arias 1975 A study of the floral biology of *Victoria amazonica* (Poepp.) Sowerby (Nymphaeaceae). Acta Amazonica 5: 109–139.
Qiu Y-L, MW Chase, DH Les, CR Parks 1993 Molecular phylogenetics of the Magnoliidae: cladistic analyses of nucleotide sequences of the plastid gene *rbcL*. Ann Mo Bot Gard 80:587–606.
Qiu Y-L, J Lee, F Bernasconi-Quadroni, DE Soltis, PS Soltis, M Zanis, EA Zimmer, Z Chen, V Savolainen, MW Chase 1999 The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407.
——— 2000 Phylogeny of basal angiosperms: analyses of five genes from three genomes. Int J Plant Sci 161(suppl):S3–S27.
Qiu Y-L, J Lee, BA Whitlock, F Bernasconi-Quadroni, O Dombrovska 2001 Was the ANITA rooting of the angiosperm phylogeny affected by long-branch attraction? Mol Biol Evol (in press).
Ren D 1998 Flower-associated Brachycera flies as fossil evidence for Jurassic angiosperm origins. Science 280:85-88.
Renner SS 1999 Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. Am J Bot 86: 1301–1315.
Richardson FC 1969 Morphological studies of the Nymphaeaceae. IV. Structure and development of the flower of *Brasenia schreberi* Gmel. Univ Calif Publ Bot 47:1–45.
Robertson RE, SC Tucker 1979 Floral ontogeny of *Illicium floridanum*, with emphasis on stamen and carpel development. Am J Bot 66:605–617.
Ronse Decraene LP, Smets EF 1993 The distribution and systematic relevance of the androecial character polymery. Bot J Linn Soc 113: 285–350.
Runions CJ, JN Owens 1998 Evidence of pre-zygotic self-incompatibility in a conifer. Pages 255–264 *in* SJ Owens, PJ Rudall, eds. Reproductive biology in systematics, conservation and economic botany. Royal Botanic Gardens, Kew.
Saunders RMK 1995 Systematics of the genus *Illicium* L. (Illiciaceae) in Malesia. Bot J Linn Soc 117:333–352.

——— 1997 A taxonomic revision of *Schisandra* section *Sphaerostema* (Schisandraceae). Edinb J Bot 54:265–287.
——— 1998 Monograph of *Kadsura* (Schisandraceae). Syst Bot Monogr 54:1–106.
——— 2000 Monograph of *Schisandra* (Schisandraceae). Syst Bot Monogr 58:1–118.
Schmucker T 1932 Physiologische und ökologische Untersuchungen an Blüten tropischer *Nymphaea*-Arten. Planta 16:376–412.
Schneider EL 1976 The floral anatomy of *Victoria* Schomb. (Nymphaeaceae). Bot J Linn Soc 72:115–148.
——— 1978 Morphological studies on the Nymphaeaceae. IX. The seed of *Barclaya longifolia* Wall. Bot Gaz 139:223–230.
——— 1982*a* Notes on the floral biology of *Nymphaea elegans* (Nymphaeaceae) in Texas. Aquat Bot 12:197–200.
——— 1982*b* Observations on the pollination biology of *Nymphaea gigantea* W. J. Hooker (Nymphaeaceae). West Aust Nat 15:71–72.
——— 1983 Gross morphology and floral biology of *Ondinea purpurea* den Hartog. Aust J Bot 31:371–382.
Schneider EL, T Chaney 1981 The floral biology of *Nymphaea odorata* (Nymphaeaceae). Southwest Nat 26:159–165.
Schneider EL, JM Jeter 1982 Morphological studies of the Nymphaeaceae. XII. The floral biology of *Cabomba caroliniana*. Am J Bot 69:1410–1419.
Schneider EL, PS Williamson 1993 Nymphaeaceae. Pages 486–493 *in* K Kubitzki, JG Rohwer, V Bittrich, eds. The families and genera of vascular plants. Vol 2. Springer, Berlin.
Sehgal A, HYM Ram 1981 Comparative developmental morphology of two populations of *Ceratophyllum* L. (Ceratophyllaceae) and their taxonomy. Bot J Linn Soc 82:343–356.
Shamrov II 1981 Some peculiar features of the development of the anther in *Ceratophyllum demersum* and *C. pentacanthum* (Ceratophyllaceae). Bot Zh (Moscow & Leningrad) 66:1464–1473.
——— 1983 Anthecological investigation of three species of the genus *Ceratophyllum* (Ceratophyllaceae). Bot Zh (Moscow & Leningrad) 68:1357–1365.
Smith AC 1947 The families Illiciaceae and Schisandraceae. Sargentia 7:1–224.
Soltis DE, PS Soltis, MW Chase, ME Mort, DC Albach, M Zanis, V Savolainen, et al 2000*a* Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. Bot J Linn Soc 133:381–461.
Soltis DE, PS Soltis, DL Nickrent, LA Johnson, WJ Hahn, SB Hoot, JA Sweere, et al 1997 Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. Ann Mo Bot Gard 84:1–49.
Soltis PS, DE Soltis, MW Chase 1999 Angiosperm phylogeny inferred from multiple genes as a research tool for comparative biology. Nature 402:402–404.
Soltis PS, DE Soltis, MJ Zanis, S Kim 2000*b* Basal lineages of angiosperms: relationships and implications for floral evolution. Int J Plant Sci 161(suppl):S97–S107.
Sun G, DL Dilcher, S Zheng, Z Zhou 1998 In search of the first flower: a Jurassic angiosperm, *Archaefructus*, from northeast China. Science 282:1692–1695.
Swisher CC, Y-Q Wang, X-L Wang, X Xu, Y Wang 1999 Cretaceous age for the feathered dinosaurs of Liaoning, China. Nature 400: 58–61.
Takhtajan A 1976 Neoteny and the origin of flowering plants. Pages 207–219 *in* CB Beck, ed. Origin and early evolution of angiosperms. Columbia University Press, New York.
——— 1997 Diversity and classification of flowering plants. Columbia University Press, New York.
Taylor DW 1991 Angiosperm ovules and carpels: their characters and polarities, distribution in basal clades, and structural evolution. Postilla 208:1–40.
Theissen G 2001 Development of floral organ identity: stories from the MADS house. Curr Opin Plant Biol 4:75–85.
Thien LB 1980 Patterns of pollination in primitive angiosperms. Biotropica 12:1–13.
Thien LB, H Azuma, S Kawano 2000 New perspectives in the pollination biology of basal angiosperms. Int J Plant Sci 161(suppl): S225–S235.
Thien LB, DA White, LY Yatsu 1983 The reproductive biology of a relict—*Illicium floridanum* Ellis. Am J Bot 70:719–727.
Tobe H, T Jaffré, PH Raven 2000 Embryology of *Amborella* (Amborellaceae): descriptions and polarity of character states. J Plant Res 113:271–280.
Todzia CA 1993 Chloranthaceae. Pages 281–287 *in* K Kubitzki, JG Rohwer, V Bittrich, eds. The families and genera of vascular plants. Vol 2. Springer, Berlin.
Troll W 1933 Beiträge zur Morphologie des Gynaeceums. IV. Über das Gynaeceum der Nymphaeaceen. Planta 21:447–485.
Tucker SC, JA Bourland 1994 Ontogeny of staminate and carpellate flowers of *Schisandra glabra* (Schisandraceae). Plant Syst Evol Suppl 8:137–158.
Tucker SC, AW Douglas 1996 Floral structure, development, and relationships of paleoherbs: *Saruma*, *Cabomba*, *Lactoris*, and selected Piperales. Pages 141–175 *in* DW Taylor, LJ Hickey, eds. Flowering plant origin, evolution and phylogeny. Chapman & Hall, New York.
Ueda K 1988 Sex change in a woody vine species, *Schisandra chinensis*, a preliminary note. J Jpn Bot 63:319–321.
Upchurch GR Jr 1984 Cuticle evolution in Early Cretaceous angiosperms from the Potomac Group of Virginia and Maryland. Ann Mo Bot Gard 71:522–550.
Valla JJ, DR Cirino 1972 Biología floral del irrupé, *Victoria cruziana* D'Orb. (Nymphaeaceae). Darwiniana 17:477–500.
van der Velde G 1986 Developmental stages in the floral biology s.l. of Dutch Nymphaeaceae (*Nymphaea alba* L., *Nymphaea candida* Presl, *Nuphar lutea* (L.) Sm.). Acta Bot Neerl 35:111–113.
van Heel WA 1977 The pattern of vascular bundles in the stamens of *Nymphaea lotus* L. and its bearing on stamen morphology. Blumea 23:345–348.
van Royen P 1962 Sertulum Papuanum. 5. Nymphaeaceae. Nova Guinea Bot 8:103–126.
Vogel S 1998 Remarkable nectaries: structure, ecology, organophyletic perspectives. II. Nectarioles. Flora 193:1–29.
von Balthazar M, PK Endress 1999 Floral bract function, flowering process, and breeding systems of *Sarcandra* and *Chloranthus* (Chloranthaceae). Plant Syst Evol 218:161–178.
Wagner WL, DH Lorence 1999 A revision of *Trimenia* Seem. (Trimeniaceae) in the Marquesas Islands with description of a new species, *Trimenia nukuhivensis*. Adansonia 21:225–230.
Walker JW, AG Walker 1984 Ultrastructure of Lower Cretaceous angiosperm pollen and the origin and early evolution of flowering plants. Ann Mo Bot Gard 71:464–521.
Wang YH, K Yang, SB Ma 1999 Reproductive biology of *Chloranthus henryi* (Chloranthaceae) in northeastern Yunnan. Acta Bot Yunnanica 21:218–224.
Weller S, MJ Donoghue, D Charlesworth 1995 The evolution of self-incompatibility in flowering plants: a phylogenetic approach. Pages 355–382 *in* PC Hoch, AG Stephenson, eds. Experimental and molecular approaches to plant biosystematics. Mo Bot Gard Monogr Syst Bot 53.
White DA, LB Thien 1985 The pollination of *Illicium parviflorum* (Illiciaceae). J Elisha Mitchell Sci Soc 101:15–18.
Wiersema JH 1987 A monograph of *Nymphaea* subgenus Hydrocallis (Nymphaeaceae). Syst Bot Monogr 16:1–112.
——— 1988 Reproductive biology of *Nymphaea* (Nymphaeaceae). Ann Mo Bot Gard 75:795–804.
Williams EG, TL Sage, LB Thien 1993 Functional syncarpy by intercarpellary growth of pollen tubes in a primitive apocarpous angiosperm, *Illicium floridanum* (Illiciaceae). Am J Bot 80:137–142.
Williams G, P Adams, LA Mound 2001 Thrips (Thysanoptera) pol-

lination in Australian subtropical rain forests, with particular reference to pollination of veiny wilkiea *Wilkiea huegeliana* (Tul.) A.DC. (Monimiaceae). J Nat Hist, Vol 35 (in press).

Williamson PS, MF Moseley 1989 Morphological studies of the Nymphaeaceae sensu lato. XVII. Floral anatomy of *Ondinea purpurea* subspecies *purpurea* (Nymphaeaceae). Am J Bot 76: 1779–1794.

Williamson PS, EL Schneider 1993 Cabombaceae. Pages 157–161 *in* K Kubitzi, JG Rohwer, V Bittrich, eds. The families and genera of vascular plants. Vol 2. Springer, Berlin.

——— 1994 Floral aspects of *Barclaya* (Nymphaeaceae): pollination, ontogeny and structure. Plant Syst Evol Suppl 8:159–173.

Wilmot-Dear M 1985 *Ceratophyllum* revisited—a study in fruit and leaf variation. Kew Bull 40:243–271.

Winter AN 1993 Some aspects of the reproductive biology of *Hydrostemma longifolium* (*Barclaya longifolia*) (Barclayaceae). Bot Zh 78:69–83.

Wolf M 1991 Blütenphyllotaxis von Nymphaeaceae: ist das Androecium von *Nymphaea*, *Nuphar* etc. spiralig? Symposium Morphologie, Anatomie und Systematik, University of Göttingen.

American Journal of Botany 90(12): 1758–1776. 2003.

ANGIOSPERM PHYLOGENY BASED ON *matK* SEQUENCE INFORMATION[1]

KHIDIR W. HILU,[2,14] THOMAS BORSCH,[3] KAI MÜLLER,[3]
DOUGLAS E. SOLTIS,[4] PAMELA S. SOLTIS,[5] VINCENT SAVOLAINEN,[6]
MARK W. CHASE,[6] MARTYN P. POWELL,[6] LAWRENCE A. ALICE,[7]
RODGER EVANS,[8] HERVÉ SAUQUET,[9] CHRISTOPH NEINHUIS,[10]
TRACEY A. B. SLOTTA,[2] JENS G. ROHWER,[11]
CHRISTOPHER S. CAMPBELL,[12] AND LARS W. CHATROU[13]

[2]Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061 USA; [3]Botanisches Institut, Friedrich-Wilhelms-Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany; [4]School of Biological Sciences, Washington State University, Pullman, Washington 99164 USA; [5]Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida 32611 USA; [6]Molecular Systematics Section, Royal Botanic Gardens, Kew, Richmond, Surrey, TW9 3DS, UK; [7]Department of Biology, Western Kentucky University, Bowling Green, Kentucky 42101 USA; [8]Biology Department, Acadia University, 24 University Avenue, Wolfville, Nova Scotia, Box 1X0, Canada; [9]Muséum National d'Histoire Naturelle, 16, rue Buffon, 75005 Paris, France; [10]Institut für Botanik, Zellescher Weg 22, 01062 Dresden, Germany; [11]Institut für Allgemeine Botanik, Universität Hamburg, Ohnhorststr. 18, 22609 Hamburg, Germany; [12]Department of Biological Sciences, University of Maine, Orono, Maine 04469-5722 USA; and [13]National Herbarium of the Netherlands, Utrecht University Branch, Heidelberglaan 2, 3584 CS Utrecht, Netherlands

Plastid *matK* gene sequences for 374 genera representing all angiosperm orders and 12 genera of gymnosperms were analyzed using parsimony (MP) and Bayesian inference (BI) approaches. Traditionally, slowly evolving genomic regions have been preferred for deep-level phylogenetic inference in angiosperms. The *matK* gene evolves approximately three times faster than the widely used plastid genes *rbcL* and *atpB*. The MP and BI trees are highly congruent. The robustness of the strict consensus tree supercedes all individual gene analyses and is comparable only to multigene-based phylogenies. Of the 385 nodes resolved, 79% are supported by high jackknife values, averaging 88%. *Amborella* is sister to the remaining angiosperms, followed by a grade of Nymphaeaceae and Austrobaileyales. Bayesian inference resolves *Amborella* + Nymphaeaceae as sister to the rest, but with weak (0.42) posterior probability. The MP analysis shows a trichotomy sister to the Austrobaileyales representing eumagnoliids, monocots + Chloranthales, and *Ceratophyllum* + eudicots. The *matK* gene produces the highest internal support yet for basal eudicots and, within core eudicots, resolves a crown group comprising Berberidopsidaceae/Aextoxicaceae, Santalales, and Caryophyllales + asterids. Moreover, *matK* sequences provide good resolution within many angiosperm orders. Combined analyses of *matK* and other rapidly evolving DNA regions with available multigene data sets have strong potential to enhance resolution and internal support in deep level angiosperm phylogenetics and provide additional insights into angiosperm evolution.

Key words: angiosperms; Bayesian inference; *matK*; phylogeny; systematics.

Phylogenetic analysis of gene sequences has significantly impacted views of angiosperm relationships (Dahlgren, 1980; Takhtajan, 1987; Cronquist, 1988; Thorne, 1992). Consequently, the overall phylogeny of angiosperms has been radically revised at all levels. Some subclasses, such as Dilleniidae and Hamamelidae, have been shown to be polyphyletic with their constituent families now placed (APG, 1998; APG II, 2003) in several distantly related clades. The composition of other groups has also been altered to varying degrees, e.g., Rosidae, Asteridae, Ericales, Cornales, and Saxifragales. Contributions toward this reassessment of angiosperm phylogeny have come primarily from large data sets of individual genes or combined analyses of these data sets (e.g., Chase et al., 1993; Qiu et al., 1998, 1999, 2000; Hoot et al., 1999; Soltis et al., 1999, 2000, 2003; Olmstead et al., 2000; Savolainen et al., 2000a, b; Zanis et al., 2002). In addition, extensive analyses of morphological, anatomical, and phytochemical characters from across angiosperm families (Nandi et al., 1998) have also contributed to modern views of angiosperm relationships. Consequently, a new concept for the overall phylogeny of flowering plants has emerged, depicting a basal grade of Amborellaceae, Nymphaeaceae (sensu APG II, 2003), and Austrobaileyales, followed by eumagnoliids (sensu APG II, 2003, to include Canellales, Laurales, Magnoliales, and Piperales), monocots, Ceratophyllales, Chloranthaceae, and eudicots. However, a number of questions remain unanswered due to variable or unresolved positions and weak support for various lineages. This situation is particularly true for the eudicots, which constitute about 75% of angiosperm species diversity (Drinnan et al., 1994). Among eudicots, the basal grade lacks convincing bootstrap (BS)/jackknife (JK) support (Qiu et al., 1998; Hoot et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003). Moreover, relationships among the major clades of core eudicots (i.e., Berberidopsidaceae/Aextoxicaceae, Saxifraga-

[1] Manuscript received 27 March 2003; revision accepted 3 July 2003.

The authors thank Stefan Wanke of University of Bonn for the sequence of *Lactoris*, Prof. Wilhelm Barthlott for his continuous support of this project and molecular systematics at the Botanical Institute in Bonn, and the Deep Time Research Coordination Network for facilitating interaction among coauthors. The work is supported in part by DFG grants BO 1815-1/1 to T. B. and Ro700/3 to J. G. R., NSF grants DEB 9634231 to K. W. H. and DEB 9806945 to C. S. C., and the Netherlands Organisation for Scientific Research grant S85-324 to L. W. C.

[14] E-mail: hilukw@vt.edu.

les, Caryophyllales, rosids, asterids, and Santalales) remain uncertain (Hoot et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003). Phylogenetic relationships among rosids also remain unclear (Savolainen et al., 2000a; Soltis et al., 2000, 2003). For basal angiosperms, there are still questions concerning the position of eumagnoliids, monocots, Ceratophyllaceae, and Chloranthaceae. In addition, alternative hypotheses, albeit with weak support, for the position of *Amborella* as sister to all other angiosperms have emerged, depicting waterlilies alone or along with *Amborella* in that position (Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002). Resolving relationships among these groups is not only essential for a comprehensive systematic treatment of angiosperms, but also for understanding patterns of species diversification and character evolution.

Angiosperm phylogenetic studies based on individual genes have faced two difficulties: limited resolution and low internal support for major clades and topological incongruence (Olmstead and Sweere, 1994; Soltis et al., 1997, 2000, 2003; Mathews and Donoghue, 1999, 2000; Savolainen et al., 2000a). Combining data sets in multigene analyses improved resolution and internal support (Soltis et al., 2000, 2003; Parkinson et al., 1999; Graham and Olmstead, 2000; Qiu et al., 2000; Savolainen et al., 2000a, b; Zanis et al., 2002; Sauquet et al., 2003). Combined analyses of genes from different subcellular compartments are considered to be a good method to estimate organismal phylogeny (e.g., Donoghue and Sanderson, 1992; Hillis, 1996, 1998; Kim, 1998), a view supported by empirical studies (Qiu et al., 1999, 2000; Soltis et al., 2000, 2003; Zanis et al., 2002). Consequently, the consensus tree based on combined *rbcL, atpB,* and 18S rDNA sequences of Soltis et al. (2000; henceforth referred to as the three-gene analysis) may be considered the most reliable overall angiosperm phylogeny so far available.

In the majority of broad angiosperm phylogenetic studies, authors have emphasized using sequence information from slowly evolving genes based on the notion that the number of multiple hits and levels of homoplasy are expected to be relatively low (Farris, 1977; Swofford et al., 1996; Olmstead et al., 1998; Graham et al., 2000). However, use of slowly evolving genomic regions can result in severe limitations in taxon sampling due to need for sequencing a large number of nucleotides per species to obtain sufficient number of variable characters. Consequently, it restricts the number of taxa that can reasonably be sequenced and analyzed cladistically, introducing a new set of phylogenetic problems as pointed out in several recent studies (e.g., Graybeal, 1998; Rannala et al., 1998; Pollock et al., 2002). For example, Graham and Olmstead (2000) sequenced 13.4 kilobases (kb) of slowly evolving cpDNA genes, but as a result could only include 19 taxa in a study of basal angiosperms. This raises the problem of taxon density, an issue addressed by several authors (e.g., Graybeal, 1998; Hillis, 1998; Bremer et al., 1999; Zwickl and Hillis, 2002). Therefore, genomic regions that can provide sufficient signal in deep level phylogeny reconstruction without compromising taxon representation are essential for accurate assessment of evolutionary histories. The rapidly evolving *matK* gene satisfies these prerequisites.

The *matK* gene is ~1600 base pairs (bp) in most angiosperms, located within the *trnK* intron, and functionally may be involved in splicing group II introns coding for tRNALys (UUU; Neuhaus and Link, 1987; Ems et al., 1995). Believed to code for a maturase based on structural similarities to other such genes (Neuhaus and Link, 1987; Mohr et al., 1993), *matK* is the only maturase of higher plant plastids (Vogel et al., 1997). The *trnK* intron, including the *matK* exon, is transcribed in one piece (Chiba et al., 1996) and is expressed at the protein level in *Solanum* (Du Jardin et al., 1994). The *matK* open reading frame (ORF) is maintained intact except at the 3′ end where frameshift substitutions slightly alter the length with apparently minimal impact on function (Hilu and Alice, 1999). These data and the analysis of the RNA-binding activity of a *trnK*-encoded polypeptide from *Sinapis* (Liere and Link, 1995) further support a *matK* function in splicing group II introns. The presence of *matK* as a free-standing ORF in the plastid genome of the parasitic *Epifagus virginiana* (Ems et al., 1995), which has lost ~65% of its genes (Wolfe et al., 1992), also points to the functional significance of *matK*.

The *matK* gene stands out among genes used in angiosperm systematics in its substantially greater number of: (1) nucleotide substitutions, (2) nonsynonymous mutations, and (3) insertion/deletion events or indels (Johnson and Soltis, 1994, 1995; Olmstead and Palmer, 1994; Hilu and Liang, 1997; Soltis and Soltis, 1998; K. W. Hilu, K. Müller, and T. Borsch, unpublished data). The gene also exhibits a relatively high proportion of transversions, with the transition/transversion ratio (ti/tv) approaching unity (Olmstead and Palmer, 1994; Hilu and Liang, 1997). The percentage amino acid substitution for *matK* between the monocot rice and the eudicot tobacco is up to sixfold higher than for *rbcL* and *atpB* (41% vs. 7–8%; Olmstead and Palmer, 1994). Among-site rate variability for the three codon positions shows that *matK* is not skewed toward the third position as is the case in most protein-coding genes used in angiosperm systematics. Substitution rates in the first and second codon positions in *matK* approach those of the third position (Johnson and Soltis, 1994, 1995; Hilu and Liang, 1997; Hilu et al., 1999), a situation that elevates the rate of nonsynonymous changes. These data point to either a low correlation between structure and function with a rather small core being functionally important (e.g., domain X; Mohr et al., 1993; Hilu and Liang, 1997), or that the enzyme's function as a maturase might require a particular stereochemistry in which the actual amino acid sequence is of reduced importance. Therefore, *matK* has evolutionary patterns and tempo that distinguish it from most genes used in angiosperm phylogeny reconstruction (Olmstead and Palmer, 1994; Hilu and Liang, 1997).

Some of these attributes of *matK* may have discouraged researchers from using *matK* sequences in broad studies such as overall angiosperm relationships. Another reason for infrequent use of *matK* at broad levels may be that taxon-specific primers are usually required. The location of *matK* within the *trnK* intron and its close proximity to *psbA* provide nearly universal primers for its amplification, and the need to design primers for sequencing is counterbalanced by the quality of the data provided. Effective sequencing strategies for *matK* are discussed in Materials and Methods.

This analysis provides an angiosperm tree based on the largest data set so far compiled for *matK*. We compare the topology obtained with this gene to previously published topologies based on single gene and multigene data sets. We also examine patterns of variability in *matK*. A parsimony approach has been chosen for data analysis to allow for direct comparison with the three-gene analysis of Soltis et al. (2000). We evaluated the effect of including Gnetales as an outgroup on the

topology; the angiosperms and Gnetales represent the two most divergent groups of seed plants (e.g., Bowe et al., 2000). In addition, a Bayesian analysis (Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2002) was performed.

MATERIALS AND METHODS

Taxon sampling and plant material—This study includes representatives of 374 angiosperm genera from 240 families and all orders recognized by APG II (2003) and 12 gymnosperm genera (Appendix 1; see Supplemental Data accompanying the online version of this article). Large families are represented wherever possible by more than one genus. A large proportion of the *matK* sequences was generated specifically for this study, and additional sequences were taken from GenBank (Appendix 1).

DNA isolation, polymerase chain reaction (PCR) amplification, and sequencing—Total cellular DNA was isolated from fresh, silica-dried, or herbarium specimens using the hexadecyltrimethylammonium bromide (CTAB) procedure of Doyle and Doyle (1987) or its modification (Borsch et al., 2003). Because a large number of the sequences available cover a region from around position 400 through the stop codon of *matK*, this region (~1200 bp) became the focus of this study to avoid potential problems associated with large amounts of missing data. For gene amplification, either the entire *trnK* intron was amplified using primers trnK3914F and trnK2R (Johnson and Soltis, 1995) or, in most cases, the 3′-two-thirds of the *trnK* intron was amplified with a forward primer located approximately 480 bp into the coding region and trnK2R (for information on primers see Appendix 2 in Supplemental Data accompanying the online version of this article). By amplifying this region, the PCR primers could also be used for sequencing because primer annealing was guaranteed within the otherwise rather variable coding region of *matK*. Primer NYmatK480F (Borsch et al., 2003), originally designed for *Nymphaea*, turned out to be useful for many angiosperms. For highly divergent taxa, such as *Gnetum* and *Welwitschia*, the whole *trnK* intron was amplified and sequenced first with the amplification primers; specific internal sequencing primers were subsequently designed by ''walking'' into the region. For some taxa, internal primers 390F and 1326R were used (Johnson and Soltis, 1994, 1995). Cycle sequencing was performed using a Big Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems, Foster City, California, USA), and extension products were electrophoresed on ABI 310, 373, and 377 automated sequencers (Applied Biosystems).

Sequence alignment and phylogenetic analysis—Sequences were aligned using QuickAlign (Müller, 2003) or ClustalX (Thompson et al., 1997) followed by manual adjustments. All sequences were translated into amino acids and their ORFs checked. Several in-frame gaps were inserted to align the sequences. Frame-shift mutations near the stop codon formed a mutational hot spot 15 bp upstream of the stop codon; this section was excluded from the analysis. Due to differences in amplification procedures used by the collaborators, the sequences of some taxa lack ca. 200 bp at the 3′ end of the *matK* gene.

All aligned positions were given equal weight, and gaps were treated as missing data. Parsimony analyses were conducted using PAUP* *4.0b6* (Swofford, 2001) and PRAT (Müller, 2002). PRAT is a program written for this study; it generates command files that execute parsimony ratchet searches (Nixon, 1999) using PAUP*. Program options include random addition cycles of the ratchet and parsimony jackknifing, applying the ratchet in each replicate. In this study, 10 random addition cycles of 150 ratchet iterations each were used. Each iteration is comprised of two rounds of tree-bisection-reconnection (TBR) swapping, one on a randomly reweighted data set and the other on the original matrix, saving one minimum-length tree. Random upweighting affected 25% of the positions. Because each random addition cycle soon converged on the same tree score, cycles were not extended beyond 150 iterations and further cycles were not added. Shortest trees collected from the different tree islands were subjected to a final TBR swapping with 5000 saved trees, from which a strict consensus tree was computed. To estimate internal support, parsimony jackknifing with 500 cycles was carried out according to the approach and parameters suggested by Farris et al. (1996) for large data sets, with TBR swapping on five saved trees per cycle. The deleted fraction of characters was e^{-1}, which means that bootstrap frequencies agree with jackknife frequencies (Farris et al., 1996). This allows us to compare jackknife support values obtained here with bootstrap values reported in other studies. We also compare jackknife values from studies that used the same deletion percentage, e.g., by employing the program JAC (J. S. Farris, unpublished program). Two searches were performed on the *matK* data set; one on a matrix that included Gnetales among the outgroup taxa (matrix A) and the other on a matrix that excluded this order (matrix B). The second analysis was performed to evaluate the potential effects of Gnetales on the analysis.

Bayesian inference used the program MrBayes (Huelsenbeck and Ronquist, 2001). Calculations of likelihood were based on a general time reversible model of nucleotide substitution, assuming different stationary nucleotide frequencies and site-specific rate categories for each codon position. The posterior probability (PP) was estimated by sampling trees from the PP distribution, using Metropolis-coupled Markov chain Monte Carlo simulations. Four chains were run for 500 000 generations, starting with one of the shortest trees found with the parsimony ratchet, and the temperature of heated chains was set to 0.2. Chains were sampled every 10 generations. Likelihood scores converged on a stable value after generation 100 000 (the ''burn in'' of the chain), and calculations of PP were based upon the trees sampled after this generation.

Number of steps and consistency, retention, and rescaled consistency indices (CI, RI, and RC, respectively;) (Kluge and Farris, 1969; Farris, 1989) for the three codon positions of *matK* were calculated with PAUP*. Lists of steps for each codon position were subjected to the nonparametric Mann-Whitney *U* test to evaluate differences in nucleotide substitutions at these positions. Because the underlying sample distribution is largely unknown, no parametric test was applied.

RESULTS

Sequence variability and substitution patterns—The ~1200-bp sequenced region of *matK* resulted in 1749 aligned characters due to the insertion of gaps. Except for the 15-bp region upstream of the stop codon, all indels occurred in multiples of three nucleotides (up to 9 bp in length). Three genera in Caryophyllales (*Anredera*-Basellaceae, *Halophytum*-Halophytaceae, and *Rhipsalis*-Cactaceae) had an inversion 6–24 bp in length. Due to its location in a palindromic region, the actual size of the inversion could not be determined. Inversions are often associated with such palindromic motifs (see Graham et al., 2000; Kelchner, 2000). Of the aligned characters, 1221 (70%) are variable and 1083 (62%) are potentially parsimony-informative (based on matrix A). The distribution of variable sites among codon positions is 414, 386, and 421 for the first, second, and third codon positions, respectively. The overall nucleotide *p* distance is 0.216 and translates into an amino acid *p* distance of 0.339 using MEGA (Kumar et al., 2001). Thus, amino acid variation in *matK* is higher than nucleotide variability. The *p* distance at synonymous sites is 0.351, which is twice as high as the *p* distance at nonsynonymous sites (0.176). Based on unambiguous transitions and transversions traced on a single tree from matrix A, the ti/tv ratio in *matK* is 1.275. According to Holmquist (1983), a ratio of 0.4 and below is an indication of highly saturated sequences, a ratio that is certainly not reached here. However, saturation is a complex issue, and its magnitude may differ depending on nucleotide and codon positions along a genomic region; a more complete analysis will be presented elsewhere. Nevertheless, the ratio obtained in this analysis does not point towards a high level of saturation in *matK*. Base substitutions are fairly evenly distributed across the length of *matK* (Fig. 1). The low nucleotide variability depicted for the end of the

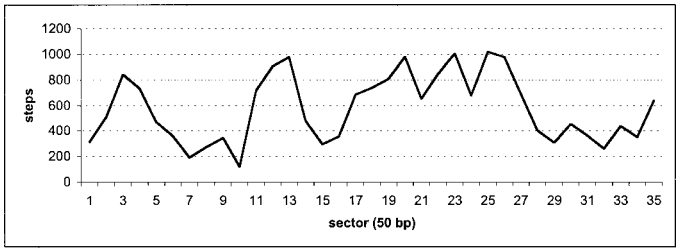


Fig. 1. Distribution of substitutional changes along the coding region, starting with position ~500 in *matK* (begin sector 1) and ending at the stop codon (end sector 36). Sectors represent 50 bp.

sequenced region corresponds in part to the conserved domain *X* (Mohr et al., 1993; Hilu and Liang, 1997). However, this decrease in variability may be accentuated by the amount of missing data at the end (see Material and Methods).

Measured on one of the shortest trees, the number of steps is greater at third positions compared to first and second positions (Table 1). Moreover, in third codon positions, the overall level of homoplasy is higher than in second positions, but lower than in first. The *U* test shows that equality of the distributions for steps at third positions vs. those at first and second positions can be rejected (Table 1, $P < 0.0001$).

Phylogenetic results—Parsimony trees were 20 646 and 20 192 steps in length for matrices A and B, respectively. The CI, RI, and RC values were identical to two decimal places in both searches (CI = 0.14, RI = 0.64, RC = 0.09). In matrix A, the performance of the ratchet using PAUP* and PRAT was compared to the strategy of random addition replicates, saving a limited number of trees per cycle (maxtrees = 1000, nreps = 1000). The latter approach resulted in minimum-length trees of 20 651 steps after several weeks of computation on a 350 Mhz Macintosh G4. In contrast, shorter trees were encountered after only 3 min when PAUP* executed a PRAT command file on the same computer; tree collection from 1500 islands was completed in about 22 h. A summary tree containing the major angiosperm lineages based on the strict consensus of 5000 trees from matrix A is provided in Fig. 2, with detailed strict consensus trees depicted in Figs. 4–12. A total of 305 of 385 nodes (79%) receive jackknife support greater

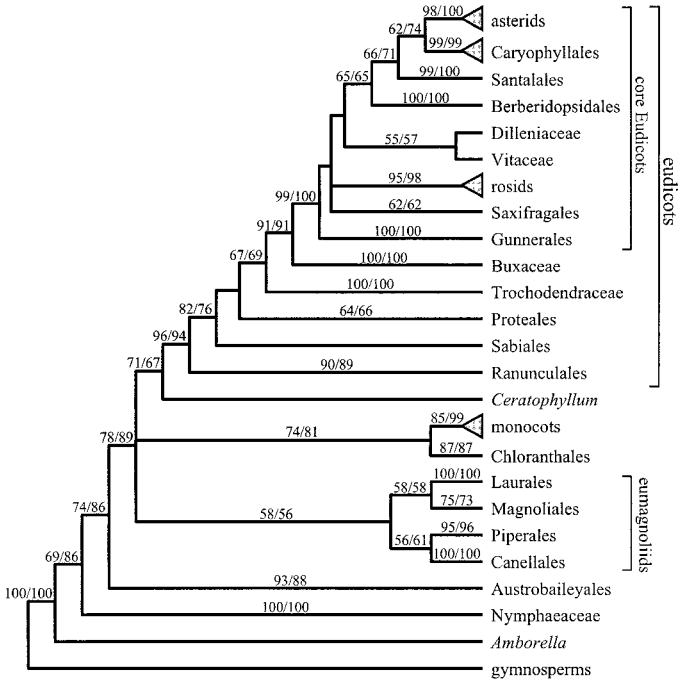


Fig. 2. Summary of angiosperm strict consensus tree based on parsimony analysis of *matK* gene sequences using gymnosperms as the outgroup. Numbers above branches are jackknife values derived from heuristic-based searches on matrices A and B (A/B, Gnetales included or excluded, respectively, from outgroup). Large clades are indicated by triangles.

than 50%, and support levels average 88%. Deletion of Gnetales (matrix B) increased support for some clades (Fig. 2). Homoplasy levels are comparable to those obtained with the analyses of other large matrices. For example, Soltis et al. (2000) reported a CI of 0.12 for their combined three-gene data set (567 taxa) compared to a CI of 0.14 here (374 taxa).

Overall internal support for a tree based on third codon positions is higher than for trees produced from analyzing first or second positions only (Table 1). In the former case, 234 nodes received jackknife support greater than 50%, whereas first and second positions yield 223 and 180 supported nodes, respectively. These data underscore the phylogenetic utility of

TABLE 1. Characteristics of the different codon positions in *matK*. Values are based on the first shortest tree (20 192 steps) found in search B. Jackknife support for phylogenies based on individual codon positions was estimated as described in text, using 100 replicates. Support values in the "Total" column are based on the 500 jackknife replicates of search A. Pi, parsimony informative; CI, consistency index; RI, retention index; RC, rescaled consistency index; *U*, *U* statistic derived from the *U* test, comparing steps at first and second codon positions with those at third codon positions; *P*, corresponding probabilities ($P < 0.0001$ in both tests).

Codon	Position 1	Position 2	Position 1 + 2	Position 3	Total
Characters	583	583	1166	583	1749
Variable characters	414	386	800	421	1221
% Variable	71%	66%	69%	72%	70%
Pi	362	334	696	387	1083
% Pi	62%	57%	60%	66%	62%
CI	0.149	0.176	0.161	0.114	0.140
RI	0.638	0.646	0.642	0.634	0.638
RC	0.095	0.114	0.103	0.072	0.089
Steps	6459	4958	11 417	9384	20 801
U (vs. 3rd)	31 075	25 327	—	—	—
P values	2.204×10^{-5}	7.048×10^{-10}	—	—	—
Supported nodes	223	180	273	234	305
% Average support	81.2	82.6	85.3	83.6	88.4

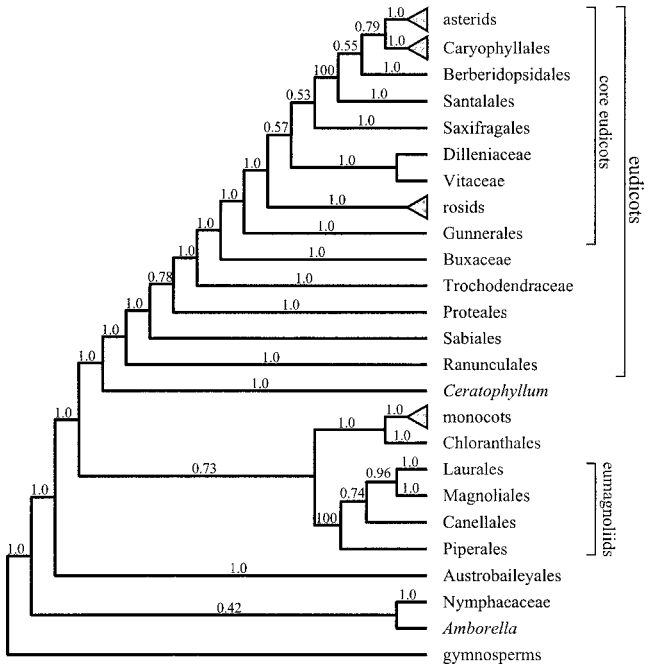


Fig. 3. Summary of angiosperm phylogeny based on Bayesian inference from the *matK* gene sequences using gymnosperms as the outgroup. Numbers above branches are posterior probabilities. Large clades are indicated by triangles.

third codon positions as demonstrated by Chase et al. (1995), Källersjö et al. (1999), Savolainen et al. (2000a), and others and are correlated with the number of informative sites at the different positions (66% in third vs. 62% and 57% in first and second positions). In addition, levels of homoplasy (CI, Table 1) are not on average higher in 3rd codon positions than in first and second positions. Therefore, differences in numbers of steps, variable sites, and phylogenetic structure are not as high among codon positions in *matK* as they are in more conserved genes such as *rbcL* (Källersjö et al., 1999; K. W. Hilu, K. Müller, and T. Borsch, unpublished data).

Analyses of matrices A and B result in consensus trees with identical topologies, although jackknife support for the basal nodes increases when Gnetales are excluded (Fig. 2). The Bayesian analysis produced a tree that is similar in topology to the MP tree (Fig. 3). Noteworthy differences in the Bayesian tree compared to the parsimony tree are the sister group relationship of *Amborella* + Nymphaeaceae to the remaining angiosperms in the former, though support is low (0.42 PP), and the positions of Canellales/Piperales and Saxifragales (Figs. 2, 3). The discussion is based on the MP matrix A; results from both matrix B and Bayesian analysis will be contrasted only when relevant. Bootstrap and/or jackknife support of 50–74% is considered low, 75–84% moderate, and >85% high (Chase et al., 2000). Posterior probabilities for nodes in the Bayesian tree are interpreted as reliable when above 0.95. In cases for which considerable differences exist in JK percentages between matrices A and B, both percentages are reported as A/B.

The MP tree depicts the angiosperms as monophyletic (100%) with the New Caledonian *Amborella trichopoda* (Amborellaceae) as sister to the rest of the flowering plants (69/86%), followed successively by Nymphaeaceae and an *Austrobaileya-Illicium-Schisandra* (Austobaileyales) clade as sister to the remaining angiosperms. These lineages correspond to the ANITA grade sensu Qiu et al. (2000). Monocots are (85/99%) sister to *Chloranthus* (74/81%). A weakly supported eumagnoliid clade (58/56%) consists of Piperales + Canellales (56/61%) and Laurales + Magnoliales (58%). *Ceratophyllum* is sister to eudicots (71%). The eumagnoliids, monocots + Chloranthaceae, and *Ceratophyllum* + eudicots diverge after the Austrobaileyales and their relationships are unresolved in the strict consensus tree. Eudicots are strongly supported (96%) and include a basal grade of Ranunculales, Sabiaceae, Proteales, Trochodendraceae, and Buxaceae (including Didymelaceae; APG II, 2003) that are subsequent sister to the core eudicots. The sister group relationship of Ranunculales to the remaining eudicots and Buxaceae to the core eudicots receive the highest support yet (82% and 91%, respectively). Within the core eudicots, Gunnerales are sister to a trichotomy of Saxifragales, rosids, and a clade comprising Dilleniaceae + Vitaceae, Berberidopsidaceae + Aextoxicaceae, Santalales, and Caryophyllales + asterids. Internal support is low for many core eudicot lineages.

DISCUSSION

Early-diverging angiosperms—Phylogenetic relationships among early-diverging (basal) angiosperms have been well studied due to their importance in understanding character evolution and early diversification in angiosperms (Parkinson et al., 1999; Qiu et al., 1999, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002; Borsch et al., 2003). Most recent phylogenetic analyses of angiosperms, including this study, have converged on a basal assemblage of lineages characterized by monosulcate or monosulcate-derived pollen. This is in contrast to the eudicots, a group that comprises the remaining angiosperms defined by their triaperturate or triaperturate-derived pollen. Amborellaceae, Nymphaeaceae, and Austrobaileyales have emerged with strong support as successive sisters to other angiosperms (Figs. 2–3). In contrast, relationships among eumagnoliids, monocots, Chloranthaceae, and Ceratophyllaceae remain uncertain.

The position of *Ceratophyllum* has varied in previous analyses from sister to all angiosperms, monocots, Chloranthaceae, eudicots, or in an unresolved position (Chase et al., 1993; Qiu et al., 2000; Savolainen et al., 2000a; Zanis et al., 2002). Similarly, Chloranthaceae has appeared as sister to either monocots or eudicots or in an unresolved position with other lineages (Chase et al., 1993; Mathews and Donoghue, 1999; Graham and Olmstead, 2000; Savolainen et al., 2000a; Soltis et al., 2000; Zanis et al., 2002). Increased number of characters has not enhanced our understanding of relationships among these lineages. The eumagnoliids, monocots, *Ceratophyllum*, and Chloranthaceae were unresolved in the five-gene study of Qiu et al. (2000). The combined 11-gene data set (Zanis et al., 2002) showed *Ceratophyllum* + monocots diverging after Illiciales (= Austrobaileyales), followed by Chloranthaceae, but with only 52% BS as sister to a clade comprising eumagnoliids + eudicots. A recent analysis of combined sequence data for basal angiosperms from the relatively fast-evolving noncoding *trnT-trnF* plus *matK* (K. W. Hilu, K. Müller, and T. Borsch, unpublished data) shows 99–100% JK support for the Amborellaceae, Nymphaeaceae, and Austrobaileyales grade, but weak support for the relationships among the rest. Therefore, resolving the relationships among these enigmatic lineages re-

mains one of the major challenges in angiosperm phylogenetics.

Amborellaceae, Nymphaeaceae, and Austrobaileyales—A general consensus exists on the branching pattern of these three most basal nodes (Parkinson et al., 1999; Qiu et al., 1999, 2000; Soltis et al., 1999, 2000; Mathews and Donoghue, 2000; Zanis et al., 2002; Borsch et al., 2003). Nevertheless, alternative relationships depicting Nymphaeaceae alone or *Amborella* + Nymphaeaceae as sister to the rest have been recovered by some methods of analysis (Parkinson et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Qiu et al., 2000; Zanis et al., 2002). The position of *Amborella* as the first branching angiosperm is supported here with 69% JK value in MP (matrix A, Fig. 4). Exclusion of Gnetales from the outgroup (matrix B) increases support for this node to 86%, comparable to analyses based on five genes (Qiu et al., 1999, 2000) and 11 genes (Zanis et al., 2002). Similar JK results were obtained for the Nymphaeaceae as successive sister to the rest (Fig. 2). The three-gene analysis of Soltis et al. (2000) provided only moderate jackknife support for the same topology. The jackknife support achieved by *matK* is substantial considering that the number of variable characters in this *matK* data set is less than one quarter of those analyzed in the three-gene matrix. Among other individual genes, only 18S rDNA (Soltis et al., 2000; <50% JK), *atpB* (Savolainen et al., 2000a; <50% BS), and *trnT-trnF* (Borsch et al., 2003; 94–100% JK) identify *Amborella* and Nymphaeaceae in these positions.

In contrast with the MP topology, BI places *Amborella* + Nymphaeaceae as sister to all other angiosperms, although support for this clade is very low (0.42 PP, Fig. 3). The meaning of such a low posterior probability is particularly dubious given the recent simulations of Suzuki et al. (2002). Kishino-Hasegawa tests (Kishino and Hasegawa, 1989) carried out by Parkinson et al. (1999) and Qiu et al. (2000) on other matrices of combined genes could not reject this position for *Amborella* + Nymphaeaceae. Neighbor-joining analyses of the six- and nine-gene matrices (Barkman et al., 2000) in which ''noisy'' positions were removed with relative apparent synapomorphy analysis (RASA; Lyons-Weiler et al., 1996) provided high bootstrap support value for an *Amborella* + Nymphaeaceae clade. However, recent studies suggest that RASA introduces errors when used in removing noisy sites (Farris, 2002; Simmons et al., 2002). Analyses of an expanded 11-gene data set (Zanis et al., 2002) showed high bootstrap and posterior probabilities for *Amborella* as sister to all other angiosperms, whereas *Amborella* + Nymphaeaceae as sister to all other angiosperms could be rejected in two of the three tests. The *matK* data favor the hypothesis that *Amborella* is sister to all other angiosperms.

Eumagnoliids—As recognized by APG II (2003), eumagnoliids include Canellales, Laurales, Magnoliales, and Piperales. Evidence from *matK* is in line with combined multigene analyses (Graham and Olmstead, 2000; Qiu et al., 2000; Zanis et al., 2002), the strict consensus trees based on analyses of *phyA* + *phyC* (Mathews and Donoghue, 2000), and *trnT-trnF* sequences (Borsch et al., 2003) in supporting this definition of eumagnoliids.

Savolainen et al. (2000a) and Soltis et al. (2000) recognized the magnoliids to include Chloranthaceae and monocots in addition to these four orders but with low support; we will refer to this clade here as eumagnoliids sensu lato (s.l.). Although the MP analysis of *matK* shows monocots + Chloranthaceae as an unresolved lineage with eumagnoliids and *Ceratophyllum* + eudicots (Fig. 4), the Bayesian analysis (Fig. 3) recovers eumagnoliids s.l. with low probability (0.73), encompassing monocots + Chloranthaceae (<0.5 PP) and eumagnoliids (1.0 PP). Eumagnoliids s.l. was also recovered with similar topology in a combined *matK* + *trnT-trnF* data analysis, but with <50% JK support (K. W. Hilu, K. Müller, and T. Borsch, unpublished data). Whereas stronger evidence points to the composition of eumagnoliids to include only the Laurales-Magnoliales-Canellales-Piperales, the hypothesis of an expanded eumagnoliids to include monocots and Chloranthaceae cannot be disregarded. Accepting the hypothesis of a eumagnoliid s.l. clade implies that carpel evolution in Chloranthaceae is secondarily ascidiate as pointed out by Doyle and Endress (2000) and Endress and Igersheim (2000).

Although MP support for the Magnoliales and Laurales sister group relationship was weak (58/56% JK), the Bayesian approach infers the same node but with 0.96 PP. This relationship has been inferred from a number of molecular data sets (Mathews and Donoghue, 1999, 2000; Qiu et al., 1999, 2000; Barkman et al., 2000; Graham and Olmstead, 2000; Graham et al., 2000; Zanis et al., 2002; Borsch et al., 2003). In addition, Sauquet et al. (2003) found particularly good support for this relationship from parsimony analysis of molecular and combined molecular and morphological data (decay index 14 and 17, BS 99%) based on unrooted trees in which all four orders were represented. However, support for this relationship from morphological data alone has been either lacking (Doyle and Endress, 2000) or weak (Sauquet et al., 2003). The internal structure of Magnoliales inferred from *matK* differs from the multigene-based results of Sauquet et al. (2003) in the arrangement of the five families above Myristicaceae; however, this inconsistency is not well supported (JK ≤ 51%) and may be explained by extremely short branches. The internal structure of Laurales based on *matK* is congruent with the minimum evolution tree in Renner and Chanderbali (2000), but not with the trees retrieved using MP in the same study or in Renner (1999).

The MP analysis shows a Piperales + Canellales relationship, but with low support. However, Bayesian analysis (Fig. 3) depicts Piperales as sister to Canellales (74% PP) and Laurales + Magnoliales. The sister group relationship of Piperales to Canellales is consistent with other molecular analyses (Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Qiu et al., 2000; Zanis et al., 2002; Borsch et al., 2003). The internal structure of Piperales in the *matK* tree shows Aristolochiaceae paraphyletic to Lactoridaceae and Saururaceae-Piperaceae in both parsimony (Fig. 4) and Bayesian approaches (tree not shown). A similar situation also occurred in analysis of *atpB* (Savolainen et al., 2000a). Phylogeny reconstruction in Piperales is complicated by the presence of several long branches. Results of higher sampling density using sequences of the complete *trnK* intron and other genomic regions are congruent with those of *matK* in showing a Piperaceae + Saururaceae clade sister to an Aristolochiaceae + Lactoridaceae clade (S. Wanke, University of Bonn, personal communication).

Monocots—The *matK* sequences provide good support (85% JK, Fig. 5) for the monophyly of monocots, with *Acorus* followed by Alismatales as sisters to other monocots. Individual gene analyses, such as *rbcL* and *atpB* (Chase et al., 1993, 1995; Savolainen et al., 2000a) and 18S (Soltis et al., 1997),

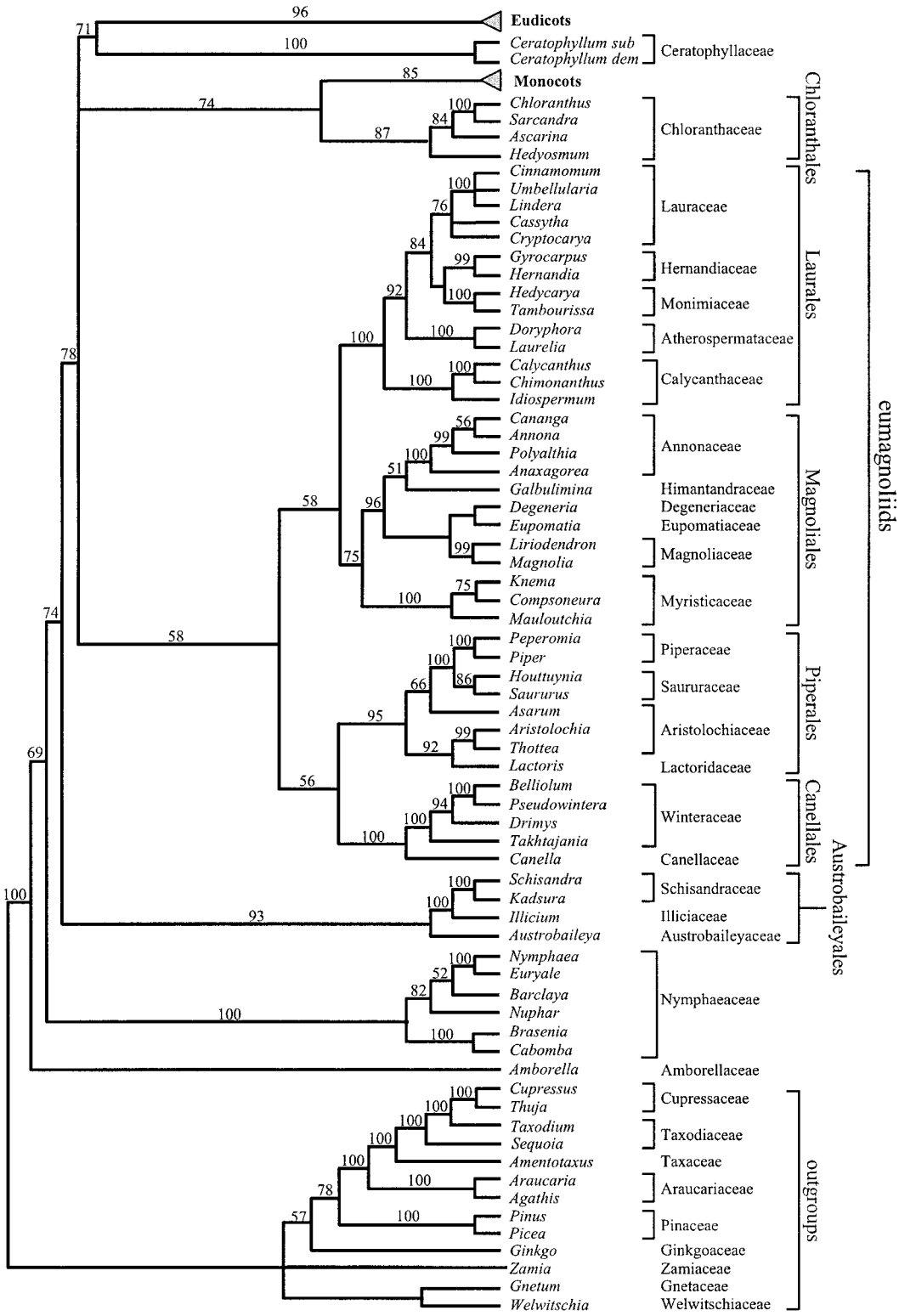


Fig. 4. Strict consensus tree highlighting relationships among basal angiosperm lineages. Numbers above branches are jackknife values derived from heuristic searches of matrix A. Large clades are indicated by triangles.

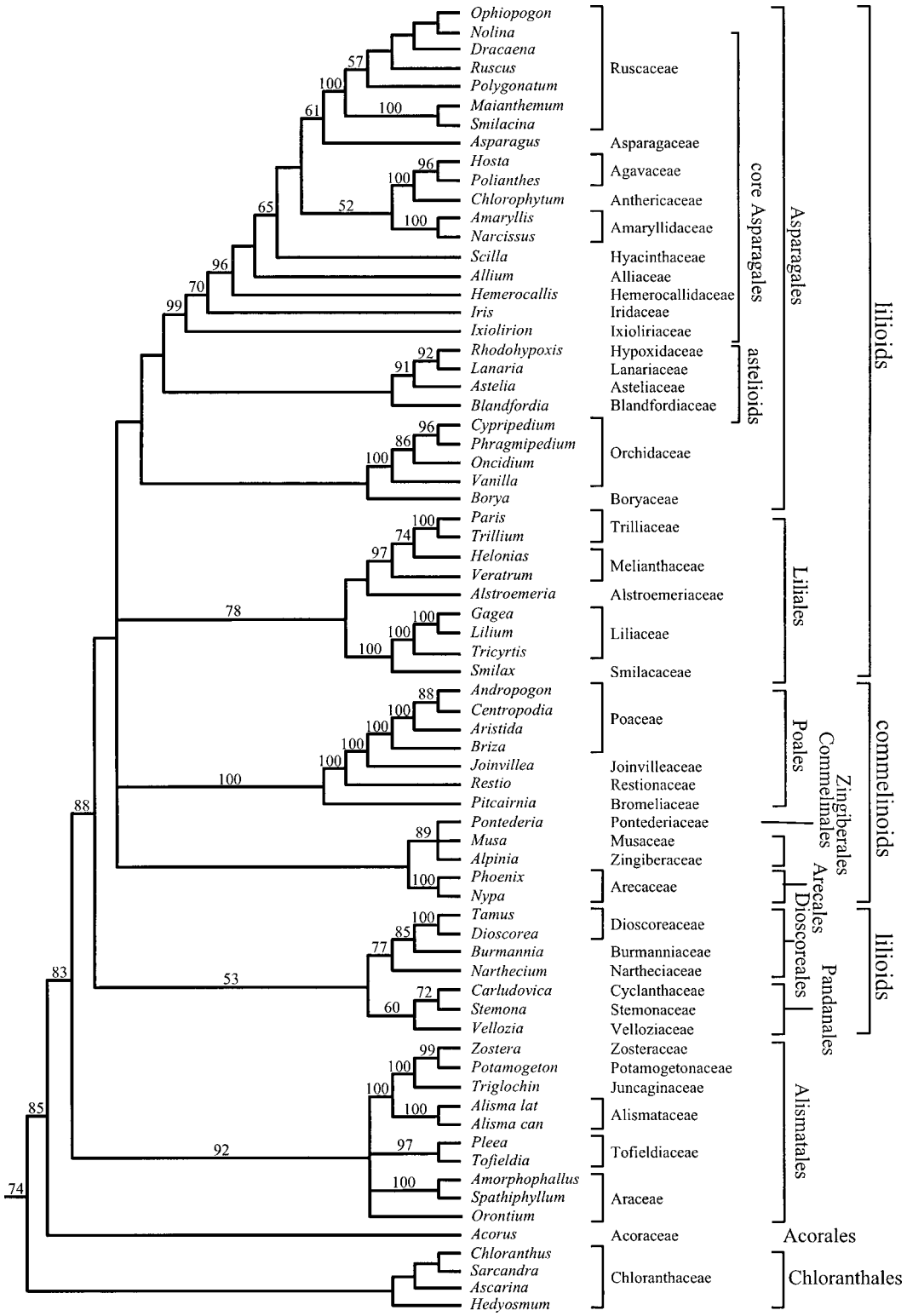


Fig. 5. Strict consensus tree highlighting relationships among monocots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

have indicated such a topology but with low or <50% support. Support for these relationships with *matK* is comparable to data sets that combined three or more genes (Table 2; Chase et al., 2000; Soltis et al., 2000; Zanis et al., 2002). The circumscription and internal structure of most monocot orders, particularly Liliales and Dioscoreales, revealed by *matK* is congruent with those suggested by the three-gene analysis, with similar or higher JK/PP support (Figs. 5–6).

TABLE 2. Comparative bootstrap/jackknife values for various nodes in angiosperm trees based on single and combined genes analyses. Large-scale analyses that covered most major angiosperm lineages are compared. Support value for the monophyly of a clade is reported when its name is cited alone, whereas support values for the others are given for a specified position in the tree. In a grade, the support by which a taxon is excluded from remaining clades (i.e., its sister group) is that at the node representing the common ancestor of the clade from which the taxon is excluded. For example, support for *Amborella* as the first clade in angiosperms refers to the node that separates *Amborella* from remaining angiosperms. ''nr'' (not resolved) denotes unresolved node, whereas ''—'' refers to taxa/clades that were not sampled. Data for *rbcL*, *atpB*, and *rbcL*/*atpB* were obtained from Savolainen et al. (2000a), 18S rDNA from Soltis et al. (1997), *rbcL*/*aptB*/18S from Soltis et al. (2000), and *rbcL*/*atpB*/18S/26S rDNA from Soltis et al. (2003). Savolainen et al. (2000a) used bootstrap, whereas remaining studies employed jackknife. For comparability see Materials and Methods.

Node	*rbcL*	*atpB*	18S	*rbcL*/*atpB*	*rbcL*/*aptB*/18S	*rbcL*/*atpB*/18S/26S	*matK*
Angiosperms	53	60	100	100	100	—	100
Amborella first clade in angiosperms	nr	<50	nr	nr	65	—	69
Nymphaeaceae second clade in angiosperms	nr	<50	nr	nr	72	—	74
Austrobaileyales third clade in angiosperms	nr	<50	nr	nr	71	—	78
Eumagnoliids s.str.	nr	<50	nr	<50	nr	—	58
Piperales sister to Canellales	nr	<50	nr	nr	nr	—	56
Monocots	59	<50	<50	86	95	—	85
Eudicots	72	<50	<50	89	99	100	96
Ranunculales first clade in eudicots	<50	<50	nr	67	59	87	82
Ranunculales	51	<50	<50	94	98	100	90
Sabiales second clade in eudicots	nr	nr	nr	nr	100	nr	<50
Proteales	<50	<50	nr	60	84	73	64
Core eudicots	<50	<50	<50	91	100	100	99
Gunnerales first clade in core eudicots	<50	nr	nr	<50	nr	84	<50
Gunnerales	57	nr	—	80	75	85	100
Rosids	<50	<50	nr	61	99	79	95
Eurosids I	<50	<50	nr	<50	77	<50	52
Eurosids II	<50	<50	nr	<50	95	88	nr
Asterids	<50	66	nr	92	99	99	98
Euasterids I	<50	<50	nr	<50	56	58	<50
Euasterids II	<50	<50	nr	51	88	87	91
Caryophyllales	84	74	<50	97	100	100	99
Saxifragales	<50	<50	68	<50	98	100	62
Santalales	<50	<50	66	86	100	100	99
Dilleniaceae/Vitaceae	nr	nr	nr	nr	nr	nr	55
Berberidopsidales	<50	95	—	97	100	100	100
Cornales	52	74	nr	96	98	100	99
Ericales	<50	<50	nr	97	98	100	98
Supported nodes (>50)	24%	14%	7%	55%	83%	82%	83%

Acorus has been inferred as sister to *Ceratophyllum* (Savolainen et al., 2000a, b, with *atpB* alone), sister to remaining monocots (Soltis et al., 2000), or placed within Alismatales (Qiu et al., 2000); monocot sampling was sparse in both studies and support in both cases was weak or lacking. Compared to three-gene analyses (Chase et al., 2000; Soltis et al., 2000), *matK* provides better support for the Alismatales as sister to the commelinoid/lilioid clades (88% vs. 78% JK) and for the monophyly of the Alismatales (92% vs. 75% JK).

Following Alismatales, the *matK* consensus tree depicts (53% JK; Fig. 5) Dioscoreales and Pandanales (lilioids) in a clade sister to remaining monocots that appear in a polytomy of two commelinoid lineages (Arecales + Zingiberales and Poales) and the two remaining lilioid clades (Liliales and Asparagales). Clarification of the phylogenetic relationships among the lilioid orders Asparagales, Dioscoreales, Liliales, and Pandanales, as well as within the commelinoids (Poales, Commelinales, Zingiberales, and Arecales), has been a major problem because previous molecular studies, including the detailed three-gene analysis of monocots (Chase et al., 2000), did not yield topologies with high internal support. The emergence in the MP analysis of a Dioscoreales + Pandanales clade (53% JK; Fig. 5) as sister to the remaining lilioid and commelinoid lineages is supported by 1.0 PP in the BI analysis (Fig. 6). Therefore, *matK* data also indicate paraphyly of the lilioids to the commelinoids, a hypothesis that requires further evaluation. Significantly, Liliales and most Asparagales uniquely share a particular type of epicuticular wax (parallel platelets; Barthlott et al., 2003). The commelinoids do not form a clade in the MP analysis of the *matK* data. In contrast, this lineage was resolved with 1.0 PP in the Bayesian tree (Fig. 6), in line with the three-gene analyses of Chase et al. (2000) and Soltis et al. (2000) that resolved this group with weak support. The sister group relationship between Commelinales and Zingiberales as inferred earlier by the three-gene studies with moderate support is well supported with *matK* sequences (89% JK, 1.0 PP). Moreover, *matK* places Arecales as sister group to the Commelinales-Zingiberales clade (Figs. 5, 6; <50% MP, 0.88 BI).

Chase et al. (1995, 2000) defined Asparagales in a broad sense to include the astelioids, Boryaceae and Orchidaceae. However, the combined *rbcL*, *atpB*, and 18S rDNA analyses of Soltis et al. (2000) and Chase et al. (2000) only recovered the core Asparagales (sensu Chase et al., 1995) as a monophyletic group with high support. Hypoxidaceae, Asteliaceae, Boryaceae, Blandfordiaceae, and Orchidaceae appeared in their strict consensus tree as unresolved at the base of the Asparagales s.l. clade. The *matK* study recovers core Asparagales (Ixoliriaceae to Convallariaceae, Figs. 5, 6) with better support (99% JK, 1.0 PP). At the base of Asparagales, two

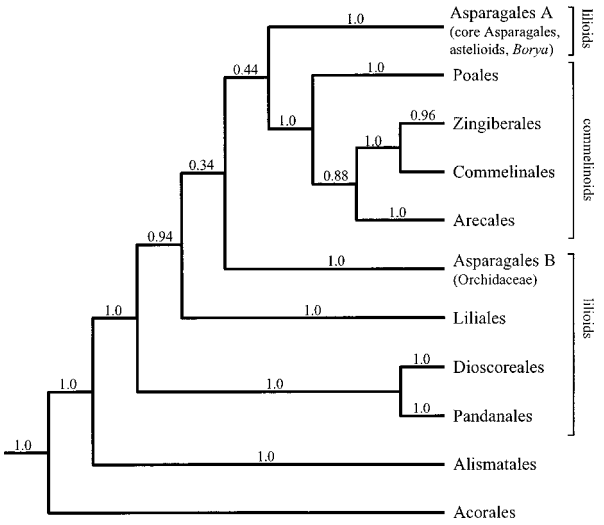


Fig. 6. Summary of monocot relationships inferred from Bayesian analysis. Numbers above branches are posterior probabilities.

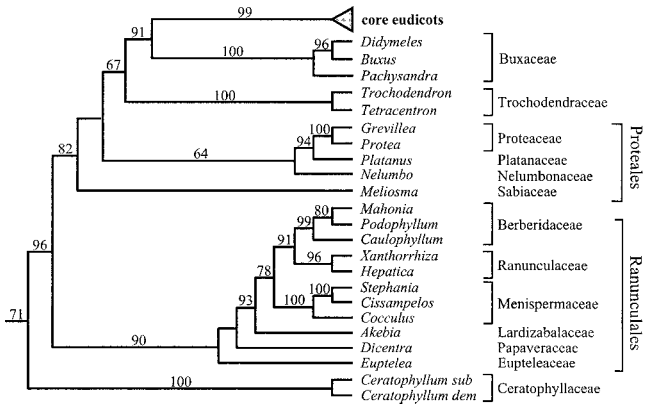


Fig. 7. Strict consensus tree highlighting relationships among basal eudicots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

clades, Boryaceae + Orchidaceae and Blandfordiaceae + Asteliaceae + Lanariaceae + Hypoxidaceae (the latter group was termed the astelioid clade by Fay et al., 2000; Fig. 5) were found successive sisters to the rest in MP but with <50% JK support (annotated as lower Asparagales in Fig. 5). This topology is identical to the one inferred from the second search of the three-gene data set by Chase et al. (2000), although none of the relevant nodes received good support. The *matK* places Boryaceae as sister to Orchidaceae in the MP tree with <50% JK and shows it nested within the astelioid clade in the Bayesian tree with high PP (the sucessive branching order is *Blandfordia*, *Borya*, *Astelia*, and *Rhodohypoxis* plus *Lanaria*; all nodes receive 1.0 PP; details are not shown in Fig. 6). The orchids appear as sister to the commelinoids in the Bayesian tree (Fig. 6). A recent combined *rbcL* + *atpB* + *trnL-F* parsimony analysis of Asparagales (Fay et al., 2000) showed Boryaceae sister to the astelioid clade, which are in turn sister to the core Asparagales. Orchidaceae appear basal in the latter analysis, but none of these basal nodes within lower Asparagales are well supported. As a result, the order of the first branches within the Asparagales and their broader circumscription to include the orchids are still in need of further testing.

Eudicots—The eudicots include over 200 000 species and comprise about 75% of all angiosperm species (Drinnan et al., 1994). This clade has consistently been recovered in all recent single-gene and multiple-gene phylogenetic analyses of angiosperms (e.g., Chase et al., 1993; Hoot et al., 1999; Barkman et al., 2000; Graham and Olmstead, 2000; Mathews and Donoghue, 2000; Qiu et al., 2000; Savolainen et al., 2000a, b; Borsch et al., 2003; Soltis et al., 2003) and in nonmolecular phylogenetic analyses (Drinnan et al., 1994; Nandi et al., 1998). The strong support (100%) for eudicot monophyly achieved by combining genes from several genomes is corroborated by *matK* (Figs. 2, 3). The eudicot clade encompasses a basal grade and a strongly supported core clade that includes the large rosid and asterid clades and the smaller Gunnerales, Caryophyllales, Santalales, Berberidopsidaceae/Aextoxicaceae, and Saxifragales clades. Eudicots have tricolpate or tricolpate-derived pollen (e.g., tricolporate, pantoporate; e.g., Nandi et al., 1998). Morphologically similar pollen having three apertures in Austrobaileyales have been identified as trisyncolpate (Takahashi, 1994), but these are apparently convergent.

Early-diverging eudicots—The same lineages of early-diverging eudicots (Ranunculales, Proteales, Sabiaceae, Trochodendraceae including Tetracentron, and Buxaceae including Didymelaceae) have been consistently recovered (e.g., Chase et al., 1993; Soltis et al., 1997, 2000, 2003; Hoot et al., 1999; Savolainen et al., 2000a), although relationships among them are unclear. In addition, support for the nodes is low in several cases (Table 2). Both MP and BI analyses of our *matK* data yield identical topologies with good support for a grade of Ranunculales/Sabiaceae/Proteales/Trochodendraceae (Fig. 7). The *matK* tree is for the most part in agreement with those obtained from the combined three- and four-gene analyses (Soltis et al., 2000, 2003). However, *matK* data provide considerably higher support for relationships in this group than in previous studies (Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000), particularly for the position of Ranunculales as sister to all other eudicots (82/76% JK, 1.0 PP). Within Ranunculales, the exact position of Eupteleaceae and Papaveraceae at the base is method dependent in *matK*. *Euptelea* is sister to all other Ranuncuales though without support in MP, whereas Papaveraceae were basal in MB, excluded with 75% PP from the rest of Ranunculales. Family relationships among remaining Ranunculales are identical in both approaches (Bayesian not shown) and in agreement with the results from the multigene analyses of Hoot et al. (1999) and Soltis et al. (2000, 2003).

The divergence of Sabiaceae after Ranunculales gains <50% JK support but 1.0 PP. The position of Sabiaceae and its inferred affinity to Proteales has varied (Hoot et al., 1999; Qiu et al., 2000; Savolainen et al., 2000a; Soltis et al., 2000, 2003), but with weak or no support. Although *matK* presents a potential position for Sabiaceae branching after Ranunculales, the hypothesis of a Sabiaceae + Proteales clade (see Soltis et al., 2000) cannot be excluded.

In line with previous studies (Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000), the *matK* tree places *Nelumbo* as the sister to the rest of Proteales. Also evident is the high support for a sister group relationship of Proteaceae and Platanaceae (94% JK, 1.0 PP). Several anatomical and morpho-

logical synapomorphies have been identified for Proteales (see Nandi et al., 1998; Savolainen et al., 2000a).

Buxaceae and Trochodendraceae exchanged positions in the basal eudicot grade in previous studies; these nodes always received low support (Table 2). This *matK* study shows for the first time good support for a Trochodendraceae/Buxaceae/core eudicot including Gunnerales (Figs. 2, 3, 7) successive branching. The support is particularly strong for Buxaceae being sister to core eudicots (91% JK, 1.0 PP).

Core eudicots—Core eudicots, composed of Gunnerales, Caryophyllales, Berberidopsidales, Saxifragales, Santalales, Vitaceae, Dilleniaceae, asterids, and rosids, received 100% JK and 1.0 PP support with *matK*. The core eudicots were recovered with high support in the combined *rbcL*/*atpB* analysis, and 100% support was achieved with the addition of 18S and 26S rDNA sequences (Hoot et al., 1999; Soltis et al., 2000, 2003). Thus, the core eudicots stand now as one of the best-supported major clades of angiosperms (Figs. 2, 3). Soltis et al. (2003) discussed the implications for character evolution of Gunnerales sister to the rest of the core eudicots.

The *matK* data provide more structure for the major branches within core eudicots than previous single or combined gene analyses (Figs. 8, 9; Table 2). A crown group comprising Berberidopsidales/Santalales/Caryophyllales/asterids is inferred here (65% JK and 1.0 PP; Figs. 2, 3, 8). The sister group relationship of asterids and Caryophyllales is weakly supported (62% JK, 0.79 PP). The study based on 18S rDNA sequences (Soltis et al., 1997) showed Caryophyllales nested within asterids (BS $<$ 50%), but combined analysis of 18S rDNA and *rbcL* sequences placed the order within rosids (Soltis and Soltis, 1997); sparse sampling in the latter study might have affected Caryophyllales placement. All previous analyses have indicated affinities of Caryophyllales to Dilleniaceae (see discussion under Dilleniaceae/Vitaceae), but never with high support. Santalales and Berberidopsidaceae/Aextoxicaceae change positions depending on method of analysis. This unstable position is reflected by the low posterior probability (0.55) for the node uniting Berberidopsidaceae/Aextoxicaceae and asterids/Caryophyllales.

Gunnerales—Support for the monophyly of Gunnerales (i.e., the sister group relationship of *Gunnera* and *Myrothamnus*) with *matK* is also high compared to all previous analyses (1.0 PP and 100% JK; Figs. 2, 3, 8), exceeding the 85% JK support achieved by the four-gene analysis (Soltis et al., 2003). The *matK* analysis provides strong support (99/100% JK and 1.0 PP) for the inclusion of Gunnerales in core eudicots (Figs. 2, 8). However, Gunnerales being sister to the remaining core eudicots receives $<$50% JK support in MP, yet is highly probable (1.0 PP) in the Bayesian analysis. This position for Gunnerales in the core eudicot clade is in agreement with studies based on individual and combined sequence data analyses (Chase et al., 1993; Hoot et al., 1999; Savolainen et al., 2000a; Soltis et al., 2000). The solidified relationship of the *Gunnera* + *Myrothamnus* clade is important for future analyses of the remarkable evolutionary diversification of these two genera. *Myrothamnus* comprises two species of small, poikilohydric shrubs on African inselbergs, whereas *Gunnera* (40 species) includes hemicryptophytes, living in symbiosis with the cyanobacterium *Nostoc* for nitrogen fixation and inhabiting more or less wet habitats, mostly in the Southern Hemisphere.

Saxifragales—In Saxifragales, four well-supported clades of interest are resolved with *matK*. One (97%) includes Iteaceae as sister to Grossulariaceae/Saxifragaceae (95%). These three families were part of a large polytomy in trees based on individual and combined analyses of *rbcL, atpB,* and 18S rDNA (Savolainen et al., 2000a; Soltis et al., 2000). The inclusion of the 26S rDNA sequence data improved resolution for Saxifragales with the sister group relationship of Iteaceae (plus Pterostemonaceae) and Saxifragaceae receiving 100% JK. A clade with similar topology and support was recovered in a combined analysis of data from five nuclear and plastid genes (Fishbein et al., 2001). A second clade resolved here includes Cercidiphyllaceae and Daphniphyllaceae (78% JK), which were unresolved in previous large-scale analyses. A third clade includes Haloragaceae, Tetracarpaeaceae, and *Aphanopetalum* (formerly included in Cunoniaceae; 95%). A similar clade was recovered in the five-gene study of Fishbein et al. (2001). A fourth, weakly supported clade (61% JK) encompasses Altingiaceae (*Rhodoleia* and *Altingia*). *Rhodoleia* has been placed in Rhodeliaceae, Hamamelidaceae, or Altingiaceae.

Vitaceae/Dilleniaceae—Dilleniaceae is sister here to Vitaceae, although support for this relationship varies with the method applied (weak using MP, but 1.0 PP in BI). The phylogenetic positions of these families among core eudicots have been difficult to assess and differ in each analysis. Floral and endosperm characters were proposed as potential synapomorphies for the relationship of Dilleniaceae to Caryophyllales and Vitaceae to rosids (Savolainen et al., 2000a). In contrast, several features are shared by Dilleniaceae and Vitaceae, such as calcium oxalate raphides (Metcalfe and Chalk, 1950), an endotesta containing radially elongate cells and a tracheidal endotegmen (Corner, 1976). Similarities of Dilleniaceae to Vitaceae have also been pointed out by Nandi et al. (1998).

Berberidopsidales—The sister group relationship of Berberidopsidaceae and Aextoxicaceae is another case in which *matK* provides the greatest support (100% JK; Fig. 8) among single-gene analyses. The results are in close agreement with trees inferred from multigene data sets (e.g., Savolainen et al., 2000a; Soltis et al., 2000, 2003).

Santalales—The strong support (99% JK; Fig. 8) for the monophyly of Santalales and internal relationships are in agreement with previous molecular studies. This analysis provides the highest support (66/71% JK and 1.0 PP) for the sister group relationship of Santalales to Caryophyllales + asterids. The position of Santalales has varied considerably in previous molecular studies. Soltis et al. (2000) noted that Santalales and Caryophyllales may be related and in turn be sister to the asterids; both relationships receive weak support in this study.

Caryophyllales—The *matK* sequence data strongly support the broadly circumscribed Caryophyllales (99% JK, 1.0 PP) as first indicated by analyses of *rbcL* alone (Savolainen et al., 2000b) and in later studies (Savolainen et al., 2000a; Soltis et al., 2000, 2003; Cuénoud et al., 2002). Two Caryophyllales clades, here named Caryophyllales I and II, are recognized with *matK* (Fig. 8).

In Caryophyllales I, *Rhabdodendron* (Rhabdodendraceae) and *Simmondsia* (Simmondsiaceae) emerge in a polytomy with a clade containing the rest of Caryophyllales I (Fig. 8), which is in agreement with Cuénoud et al. (2002). Core Caryophyl-

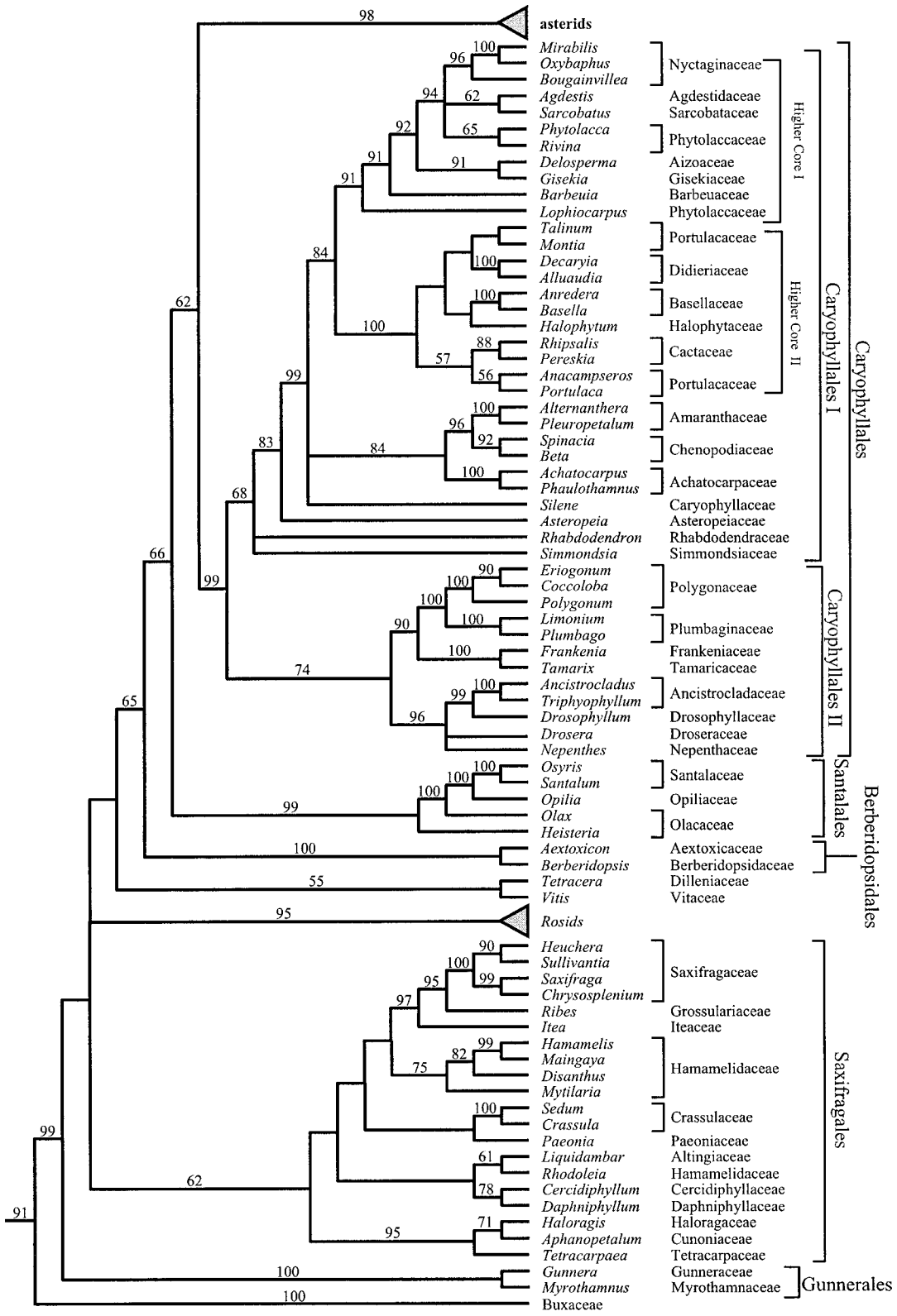


Fig. 8. Strict consensus tree highlighting relationships among core eudicots. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

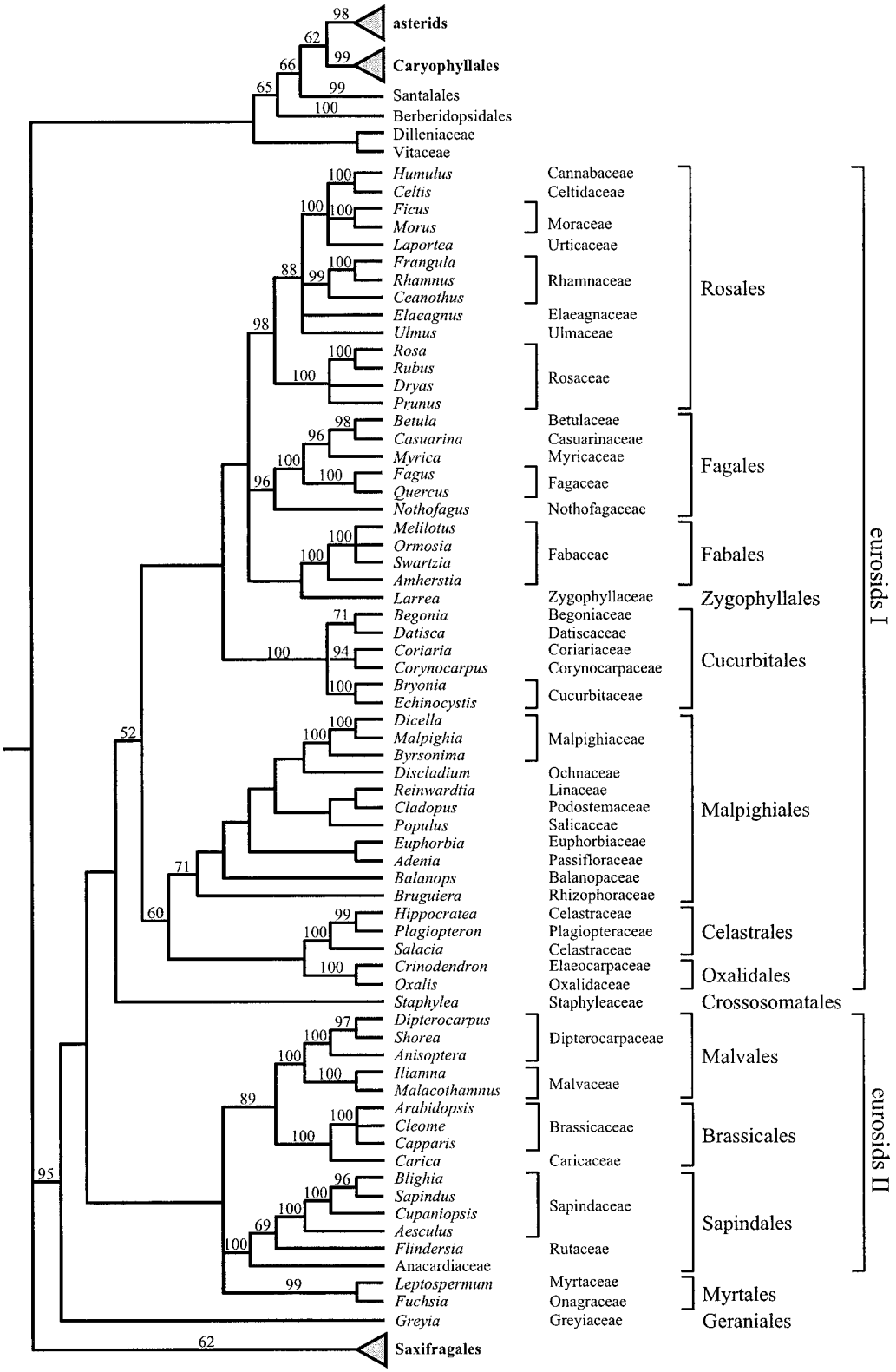


Fig. 9. Strict consensus tree highlighting relationships among rosids. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

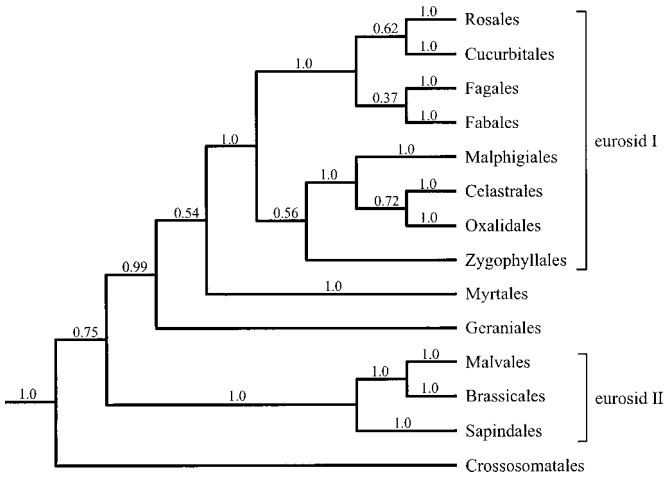


Fig. 10. Summary of relationships among rosids based on Bayesian inference. Numbers above branches are posterior probabilities.

lales (83% BS, 1.0 PP) sensu Cuénoud et al. (2002) includes *Asteropeia* (Asteropeiaceae) as sister to a polytomy comprising Caryophyllaceae, an Amaranthaceae-Chenopodiaceae-Achatocarpaceae clade that was also resolved in Kadereit et al. (in press), and a higher core Caryophyllales clade (Fig. 8). The latter group in turn consists of two major clades, higher core Caryophyllales I and II. Caryophyllaceae appears unresolved in MP, but emerges sister to the rest in BI (0.59 PP), a position also reflected by the *matK*-MP analysis of Cuénoud et al. (2002) and *rbcL* + *atpB* analysis of Savolainen et al. (2000a). Relationships revealed here for Caryophyllales are in general agreement with Cuénoud et al. (2002).

In Caryophyllales II, two subclades are recovered. One encompasses the carnivorous families Ancistrocladaceae, Drosophyllaceae, Droseraceae, and Nepenthaceae in a subclade (96% JK, 1.0 PP) and the Polygonaceae (including Coccolobaceae), Plumbaginaceae, Frankeniaceae, and Tamaricaceae in another (90% JK, 1.0 PP).

Rosids—The rosid clade (not including Vitaceae) corresponds to ''eurosids'' of Soltis et al. (2000). The strong support for rosids with *matK* (95% JK and 1.0 PP) is only comparable to that obtained from the combined *atpB, rbcL,* and 18S rDNA sequence data (99% JK; Soltis et al., 2000). However, the addition of 26S rDNA data to the three-gene data set (Soltis et al., 2003) decreased the JK value for rosids to 79%. Among single data sets, only *rbcL* (Chase et al., 1993; Savolainen et al., 2000a, b) and 18S rDNA (Soltis et al., 1997) demonstrated a rosid clade of similar circumscription but with <50% support. In this circumscription, the rosids include eurosids I and II plus Myrtales, Crossosomatales, and Geraniales (Fig. 9).

Relationships among eurosid I, eurosid II, Crossosomatales, Geraniales, and Myrtales varied among all previous single gene (Chase et al., 1993; Soltis et al., 1997; Savolainen et al., 2000a, b) and multigene (Soltis et al., 2000, 2003) analyses. In this study, topological differences are evident between the MP and Bayesian trees (Figs. 9, 10), reflecting low levels of internal support. The Geraniales/Myrtales plus eurosid II/Crossosomatales/eurosid I grade found in the MP analysis is contrasted with the BI tree that resolves Crossosomatales as sister to the remaining rosids (0.75 PP), followed by eurosid II (excluded from the rest by 0.99 PP), Geraniales, with Myrtales and eurosid I forming the terminal clade (0.54 PP). Low support for such relationships was also evident in previous molecular studies. In contrast to the ambiguous relationships among the major rosid lineages, the rosid orders are generally well supported by *matK* data. All orders receive 1.0 PP (Fig. 10) and 96–100% JK except for Malpighiales (71%). For most orders, this level of confidence was achieved only by two or more genes (Savolainen et al., 2000a; Soltis et al., 2000, 2003).

Eurosid I—Weak support (52%) is obtained for this clade in the MP analysis. In contrast, support in BI is high (1.0 PP). Previous molecular studies mostly yielded <50% support for the eurosid I clade except for the combined three-gene analysis (77% JK; Soltis et al., 2000). However, support declined to <50% when 26S rDNA sequences were added. Within eurosids I, *matK* reveals two major clades. The first comprises Celastrales, Oxalidales, and Malpighiales, which are weakly supported by MP (60%, Fig. 9) but with 1.0 PP. The other eurosid I clade includes Rosales, Fagales, Fabales, and Cucurbitales (the nitrogen-fixing clade; Soltis et al., 1995) as well as Zygophyllales. This clade recently received some support in the three-gene analysis (68% JK; Soltis et al., 2000). The Bayesian tree shows 1.0 PP support for this clade (Fig. 10). The MP topology differs from that of the corresponding Bayesian tree by depicting Zygophyllales as sister to Fabales, albeit with <50% JK support. Anthroquinones are a potential synapomorphy for Zygophyllaceae and the nitrogen-fixing clade (Sheahan and Chase, 2000). Because none of the previous topologies demonstrated good internal support, the position of Zygophyllales remains questionable.

Malpighiales is currently recognized to include at least 30 families, among which relationships have been difficult to establish. The two analyses of *matK* (Figs. 9, 10) agree upon the arrangement at the basal nodes, with Rhizophoraceae and Balanopaceae as successive sisters to the rest (0.60 and 0.87 PP, respectively; <50% JK). In the most extensive sampling of the order thus far (Savolainen et al., 2000b; *rbcL* only), Rhizophoraceae were sister to Erythroxylaceae (not included here) and the pair sister to the rest. No resolution, however, was provided for such relationships with three genes (Soltis et al., 2000). For other clades in Malpighiales, support is only achieved for Passifloraceae + Salicaceae in the BI tree (1.0 PP).

Eurosid II—A eurosid II clade is not recovered in the *matK* MP tree, but instead its components appear in two well-supported clades that form a polytomy with Myrtales (Figs. 9, 10). The three- and four-gene analyses (Soltis et al., 2000, 2003) recovered eurosids II, but relationships among its members lacked internal support. The eurosid II taxa resolved with *matK* appear in two subclades, Sapindales (100% JK) and Brassicales + Malvales (89% JK). In contrast, the Bayesian analysis inferred the three orders in a clade with 1.0 PP support (Fig. 1). The *matK* data provide high internal support for a sister group relationship of Brassicales and Malvales (89% JK, 1.0 PP); such a relationship was obtained in the *atpB* and *rbcL*/*atpB* analyses, although support in those cases was <50% and 62%, respectively (Savolainen et al., 2000a). In contrast, the three- and four-gene analyses (Soltis et al., 2000, 2003) weakly or moderately depicted Sapindales and Malvales

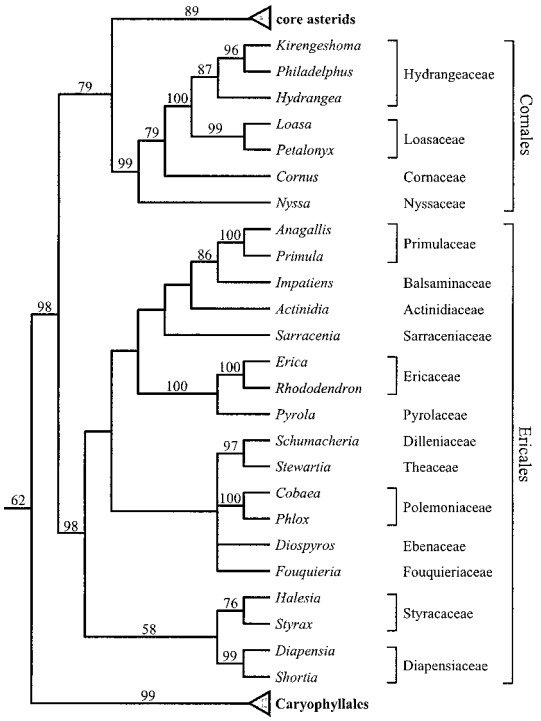


Fig. 11. Strict consensus tree highlighting relationships within Cornales and Ericales of the asterid clade. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

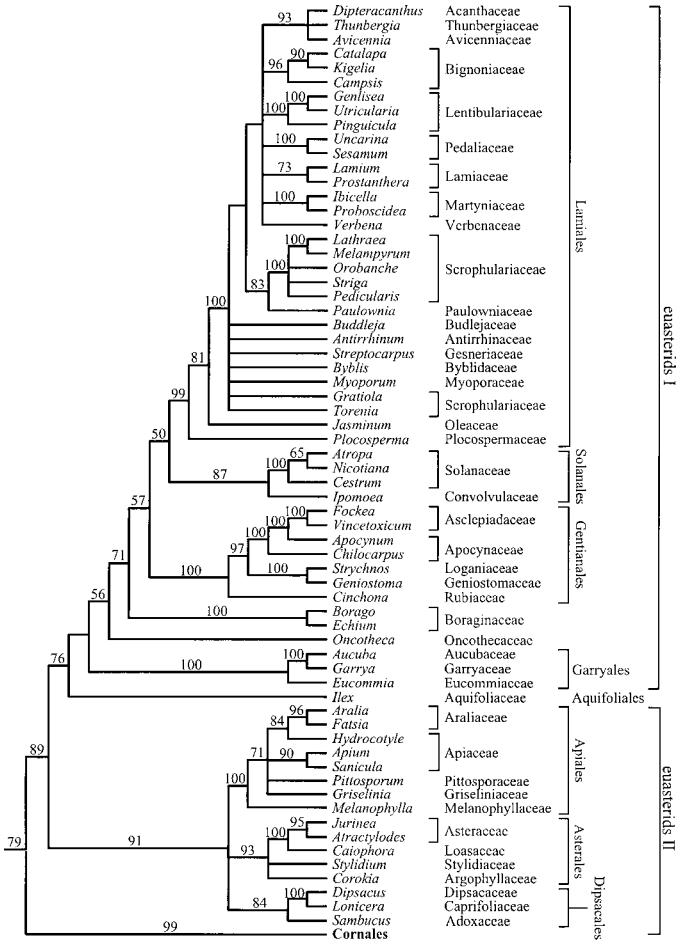


Fig. 12. Strict consensus tree depicting relationships among higher asterids. Numbers above branches are jackknife values derived from heuristic searches of matrix A.

in a clade. No clear morphological synapomorphies for this clade are evident (Judd et al., 1994; Gadek et al., 1996).

Asterids—This study reveals a strongly supported (98% JK, 1.0 PP) asterid clade sensu Olmstead et al. (1992, 1993) that includes four major lineages: Cornales, Ericales, and euasterid I and II (Figs. 2, 11). Bremer et al. (2002) referred to euasterid I and II as Lamids and Campanulids, respectively. Internal support within asterids using *matK* is surpassed only by the analysis of six-genomic region by Bremer et al. (2002) and is similar to what was obtained in the combined three- and four-gene analyses (Soltis et al., 2000, 2003; Albach et al., 2001). Our *matK* data alone show Ericales as sister to remaining asterids (79% JK; Fig. 11). In contrast, most other studies depict Cornales (Albach et al., 2001, *ndhF* tree; Bremer et al., 2002; Soltis et al., 2003) or Cornales + Ericales (Savolainen et al., 2000a; Soltis et al., 2000) in this position.

Ericales/Cornales—The monophyly of Ericales and of Cornales each receives strong support in the *matK* tree (98–99% JK and 1.0 PP; Fig. 11). The backbone of Ericales remains unresolved in this and all other molecular studies including Bremer et al. (2002) and Anderberg et al. (2002). In Cornales, Nyssaceae and Cornaceae are successive sisters to other members of the order, which is in agreement with Xiang et al. (1998), Soltis et al. (2000), and Albach et al. (2001).

Euasterid I (Lamids)—The euasterid I clade (consisting of Garryales, Oncothecaceae, Boraginales, Gentianales, Solanales, and Lamiales including Plocospermataceae) has a 1.0 PP but <50% JK support with MP (Fig. 12). Lack of support in MP is in line with other studies, including multigene data sets (e.g., Soltis et al., 2000, 2003; Albach et al., 2001; Bremer et al., 2002). The *matK* data provide moderate support for a sister position of Aquifoliales to euasterids I in both MP and BI. The sister group relationship of Garryales and Oncothecaceae to remaining euasterid I in the *matK* tree (Fig. 12) is in agreement with Bremer et al. (2002); however, BI shows 1.0 PP support compared to the low MP support (56–>50% JK) in both studies. Support for relationships among the remaining orders of euasterid I (Boraginaceae, Gentianales, Lamiales, and Solanales) is weak, a situation that has been encountered in all previous studies (Olmstead et al., 1992, 1993, 2000; Chase et al., 1993; Cosner et al., 1994; Savolainen et al., 2000a; Soltis et al., 2000, 2003; Albach et al., 2001; Bremer et al., 2002).

Gentianales are well resolved and strongly supported as monophyletic (mostly 100% JK, 1.0 PP) with Rubiaceae being sister to the rest. Support within Gentianales has been either very weak or moderate, except for the *ndhF* study of Olmstead et al. (2000) and the six-gene-region study of Bremer et al. (2002). A sister-group relationship between Lamiales and Solanales is inferred by *matK* (Fig. 12), receiving 0.8 PP (tree not shown), but only 50% JK.

Monophyly of Lamiales is supported by 99% JK, and the first branching position of Plocospermataceae in the Lamiales

receives strong support here (81% JK, 1.0 PP) and in Bremer et al. (2002). This position for Plocospermataceae was suggested by a broad sampling of *rbcL* (Savolainen et al., 2000a), but with only 56% BS. Plocospermataceae is a small family (one genus, three species) from Central America. The sister group relationship of *Paulownia* to a clade comprising parasitic and hemi-parasitic tribes of the former Scrophulariaceae, as well as the holoparasitic former family Orobanchaceae sensu stricto (s.s.), is congruent with the topology found by Olmstead et al. (2001) using *ndhF, rbcL,* and *rps2.* The Bayesian approach provides more resolution in Lamiales (see Müller et al., in press).

Euasterid II (Campanulids)—Within euasterids II, the Apiales, Asterales, and Dipsacales form a strongly supported clade (91% JK, 1.0 PP; Fig. 12) that basically corresponds to euasterids II excluding Aquifoliales. This alliance of Aquifoliales with euasterids II based on *matK* data is not well supported (76% JK, 0.81 PP; Fig. 12). Aquifoliales appeared as the first branching lineage in euasterids II in most previous studies with highest support achieved in combined analyses of 3–6 genomic regions (Soltis et al., 2000, 2003; Bremer et al., 2002). The relationships among the three euasterids II orders remain unclear.

PROSPECTS OF USING *matK* IN ANGIOSPERM PHYLOGENETICS

Sequence information from the *matK* gene produce an angiosperm tree that is considerably more robust than any previous single gene tree. Congruence is high between our *matK* tree and those based on multiple genes representing one, two, or all three genomes (Qiu et al., 1999; Savolainen et al., 2000a, b; Soltis et al., 2000, 2003; Zanis et al., 2002). The analyses of Qiu et al. (1999) and Zanis et al. (2002) were based on 8733 (five genes) and over 15 000 (11 genes) nucleotides, respectively, and thus represent approximately eight and 13 times the number of characters used here. Congruence between our *matK* phylogenies and the various multigene/multigenome phylogenies of angiosperms underscores the utility of *matK* in angiosperm phylogenetics.

When clades from the backbone phylogeny of angiosperms are compared in various molecular phylogenetic studies (Table 2), 83% received jackknife support >50% with *matK* compared with 7–24% for individual analyses of *rbcL, atpB,* and 18S rDNA. Relationships revealed by *matK* data are more robust than those derived from combining *rbcL* and *atpB* sequences (Savolainen et al., 2000a). The number of nodes receiving >50% support with *matK* is in the same range as the combined analyses of 3–4 genes from two genomic compartments (also see Table 2). Examples where *matK* stands out in terms of support are the backbone of the angiosperms, basal eudicots, core eudicots, asterids, eurosid II, and Cornales.

The topology of the angiosperm tree was not influenced by the exclusion of the Gnetales from the outgroup taxa (matrix B), but JK support generally increased at various nodes, implying a higher level of homoplasy introduced by Gnetales relative to other outgroup taxa. In contrast, the Bayesian approach, although it yields results largely congruent with the most parsimonious trees, provides alternative hypotheses for some relationships. However, alternative topologies were confined to areas of the tree at which internal support (JK or BS) has always been low.

Patterns of molecular evolution in *matK* that make it notable among other genes used in studying plant phylogenetics are quantity of information (number of parsimony-informative sites/rate of change at variable positions) and quality of characters (signal vs. noise). The *matK* gene differs from coding genes used in phylogenetic reconstruction in the nearly equitable rates of nucleotide substitution among its three codon positions and the high relative rate of nonsynonymous substitution. This evolutionary mode was previously demonstrated in smaller-scale analyses (Olmstead and Palmer, 1994; Johnson and Soltis, 1995; Hilu and Liang, 1997; Soltis and Soltis, 1998; Cuénoud et al., 2002). Such a pattern would imply relatively relaxed selection on amino acid composition in relation to function as determined by physiochemical and structural properties. In an initial analysis comparing different plastid regions in basal angiosperms, K. W. Hilu et al. (unpublished manuscript) demonstrated that purifying selection was determined to be less significant in *matK* than in other protein coding genes, whereas phylogenetic signal at informative positions was found to be highest.

Although progress has been achieved in understanding angiosperm relationships in this study, several parts of the tree remain unresolved or unsupported. Outstanding among these are the positions of monocots, Chloranthaceae, eudicots, and *Ceratophyllum* among basal angiosperms. Within eudicots, relationships among the major lineages of the core eudicots remain for the most part unclear. Combining *matK* sequences with other gene sequences has strong potential to provide more information for inference of angiosperm phylogeny. Using additional rapidly evolving genomic regions is desirable to provide insight needed to improve our understanding of angiosperm evolution.

LITERATURE CITED

ALBACH, D. C., P. S. SOLTIS, D. E. SOLTIS, AND R. G. OLMSTEAD. 2001. Phylogenetic analysis of asterids based on sequences of four genes. *Annals of the Missouri Botanical Garden* 88: 163–212.

ANDERBERG, A. A., C. RYDIN, AND M. KÄLLERSJÖ. 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany* 89: 677–687.

APG (ANGIOSPERM PHYLOGENY GROUP). 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85: 531–553.

APG II. 2003. An update of APG classification for the orders and families of flowering plants. *Botanical Journal of the Linnean Society* 141: 399–436.

BARKMAN, T. J., G. CHENERY, J. R. MCNEAL, J. LYONS-WEILER, W. 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. *Proceedings of the National Academy of Sciences, USA* 97: 13 166–13 171.

BARTHLOTT, W., I. THEISEN, T. BORSCH, AND C. NEINHUIS. 2003. Epicuticular waxes and vascular plant systematics: integrating micromorphological and chemical data. *In* T. F. Stuessy, V. Mayer, and E. Hörandl [eds.], Deep morphology: toward a renaissance of morphology in plant systematics, 189–206. *Regnum Vegetabile 141.* Ganter Verlag, Rugell, Liechtenstein.

BORSCH, T., K. W. HILU, D. QUANDT, V. WILDE, C. NEINHUIS, AND W. BARTHLOTT. 2003. Non-coding plastid *trnT-trnF* sequences reveal a well supported phylogeny of basal angiosperms. *Journal of Evolutionary Biology* 16: 558–576.

BOWE, L. M., G. COAT, AND C. W. DEPAMPHILIS. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales closest relatives are conifers. *Proceedings of the National Academy of Sciences, USA* 97: 4092–4097.

BREMER, B., K. BREMER, N. HEIDARI, P. ERIXON, R. G. OLMSTEAD, A. A.

ANDERBERG, M. KALLERSJO, AND E. P. BARKHORDARIAN. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxoninc levels. *Molecular Phylogeny and Evolution* 24: 274–301.

BREMER, B., R. K. JANSEN, B. OXELMAN, M. BACKLUND, H. LANTZ, AND K.-J. KIM. 1999. More characters or more taxa for a robust phylogeny—case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.

CHASE, M. W., M. R. DUVALL, H. G. HILLS, J. G. CONRAN, A. V. COX, L. E. EGUIARTE, J. HARTWELL, M. F. FAY, L. R. CADDICK, K. M. CAMERON, AND S. HOOT. 1995. Molecular phylogenetics of Lilianae. *In* P. J. Rudall, P. J. Cribb, D. F. Cutler, and C. J. Humphries [eds.], Monocotyledons: systematics and evolution, 109–137. Royal Botanic Gardens, Kew, UK.

CHASE, M. W., D. E. SOLTIS, P. S. SOLTIS, P. J. RUDALL, M. F. FAY, W. H. HAHN, S. SULLIVAN, J. JOSEPH, T. GIVNISH, K. J. SYTSMA, AND C. PIRES. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. *In* K. K. L. Wilson and D. Morrison [eds.], Proceedings of monocots II: the second international symposium on the comparative biology of the monocotyledons, 3–16. CSIRO Press, Sydney, Australia.

CHASE, M. W., ET AL. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.

CHIBA, T., T. HARADA, S. GOTO, R. ISHIKAWA, AND M. NIIZEKI. 1996. Transcription of tRNA genes from large-scale plastid DNA deletion clearly reveals the action of nuclear-encoded RNA polymerase in the plastid. *Journal of Plant Physiology* 148: 652–656.

CORNER, E. J. H. 1976. The seeds of dicotyledons, vols. 1, 2. Cambridge University Press, Cambridge, UK.

COSNER, M. E., R. K. JANSEN, AND T. G. LAMMERS, 1994. Phylogenetic relationship in the Campanulales based on *rbcL* sequences. *Plant Systematics and Evolution* 190: 79–95.

CRONQUIST, A. 1988. The evolution and classification of flowering plants, 2nd ed. New York Botanical Garden, Bronx, New York, USA.

CUÉNOUD, P., V. SAVOLAINEN, L. W. CHATROU, M. W. POWELL, R. J. GRAYER, AND M. W. CHASE. 2002. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL, atpB,* and *matK* DNA sequences. *American Journal of Botany* 89: 132–144.

DAHLGREN, R. M. T. 1980. A revised classification of the angiosperm. *Botanical Journal of the Linnean Society* 80: 91–124.

DONOGHUE, M. J., AND M. J. SANDERSON. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, 340–368. Chapman and Hall, New York, New York, USA.

DOYLE, J. J., AND J. L. DOYLE. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf material. *Phytochemical Bulletin* 19: 11–15.

DOYLE, J. A., AND P. K. ENDRESS. 2000. Morphological phylogenetic analysis of basal angiosperms: comparison and combination with molecular data. *International Journal of Plant Sciences* 161(Supplement): 121–153.

DRINNAN, A. N., N. R. CRANE, AND S. B. HOOT. 1994. Patterns of floral evolution in the early diversification of non-magnoliid dicotolydons (eudicots). *Plant Systematics and Evolution (Supplement)* 8: 93–122.

DU JARDIN, P., D. PORTETELLE, L. HARVENGT, M. DUMONT, AND B. WATHELET. 1994. Expression of intron-encoded maturase-like polypeptides in potato chloroplasts. *Current Genetics* 25: 158–163.

EMS, S. C., C. W. MORDEN, AND K. W. DIXON. 1995. Transcription, splicing, and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* 29: 621–733.

ENDRESS, P. K., AND A. IGERSHEIM. 2000. Gynoecium structure and evolution in basal angiosperms. *International Journal of Plant Sciences* 161 (Supplement): 211–213.

FARRIS, J. S. 1977. Phylogenetic analysis under Dollo's law. *Systematic Zoology* 26: 77–88.

FARRIS, S. J. 1989. The retention index and the rescaled consistency index. *Cladistics* 5: 417–419.

FARRIS, J. S. 2002. RASA attributes highly significant structure to randomized data. *Cladistics* 18: 334–353.

FARRIS, J. S., V. A. ALBERT, M. KÄLLERSJÖ, D. LIPSCOMB, AND A. G. KLUGE. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.

FAY, M. F., P. J. RUDALL, S. SULLIVAN, K. L. STOBART, A. Y. BRUIJN, G. REEVES, F. QAMARUZ-ZAMAN, W. P. HONG, J. JOSEPH, W. J. HAHN, J. G. CONRAN, AND M. W. CHASE. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. *In* K. K. L. Wilson and D. Morrison [eds.], Proceedings of monocots II: the second international symposium on the comparative biology of the monocotyledons, 347–359. CSIRO Press, Sydney, Australia.

FISHBEIN, M., C. HIBSCH-JETTER, D. E. SOLTIS, AND L. HUFFORD. 2001. Phylogeny of Saxifragales (angiosperms, eudicots): analysis of a rapid, ancient radiation. *Systematic Biology* 50: 817–847.

GADEK, P. A., E. S. FERNANDO, C. J. QUINN, S. B. HOOT, T. TERRAZAS, M. C. SHEAHAN, AND M. W. CHASE. 1996. Sapindales: molecular delimitation and infraordinal groups. *American Journal of Botany* 83: 802–811.

GRAHAM, S. W., AND R. G. OLMSTEAD. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS, AND R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161(Supplement): 83–96.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Systematic Biology* 47: 9–17.

HILLIS, D. M. 1996. Inferring complex phylogenies. *Nature* 383: 130–131.

HILLIS, D. M. 1998. Taxonomic sampling, phylogenetic accuracy and investigator bias. *Systematic Biology* 47: 3–8.

HILU, K. W., AND L. A. ALICE. 1999. Systematic and evolutionary implications of *matK* indels in Poaceae. *American Journal of Botany* 86: 1735–1741.

HILU, K. W., L. A. ALICE, AND H. LIANG. 1999. Phylogeny of Poaceae inferred from *mat*K sequences. *Annals of the Missouri Botanical Garden* 86: 835–851.

HILU, K. W., AND H. LIANG. 1997. The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.

HOLMQUIST, R. 1983. Transitions and transversions in evolutionary descent: an approach to understanding. *Journal of Molecular Evolution* 19: 134–144.

HOOT, S. B., S. MAGALLÓN, AND P. R. CRANE. 1999. Phylogeny of basal eudicots based on three molecular datasets: *atpB, rbcL,* and 18S nuclear ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 86: 1–32.

HUELSENBECK, J. P., B. LARGET, R. E. MILLER, AND F. RONQUIST. 2002. Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* 51: 673–688.

HUELSENBECK, J. P., AND F. RONQUIST. 2001. MrBayes. Bayesian inference of phylogeny. *Bioinformatics* 17: 754–755.

JOHNSON, L. A., AND D. E. SOLTIS. 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* 19: 143–156.

JOHNSON, L. A., AND D. E. SOLTIS. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. *Annals of the Missouri Botanical Garden* 82: 149–175.

JUDD, W. S., R. W. SAUNDERS, AND M. J. DONOGHUE. 1994. Angiosperm family pairs: preliminary phylogenetic analyses. *Harvard Papers in Botany* 5: 1–51.

KADEREIT, G., T. BORSCH, K. WEISING, AND H. FREITAG. In press. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C_4 photosynthesis. *International Journal of Plant Sciences.*

KÄLLERSJÖ, M., V. A. ALBERT, AND J. S. FARRIS. 1999. Homoplasy increases phylogenetic structure. *Cladistics* 15: 91–93.

KELCHNER, S. A. 2000. The evolution of noncoding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482–498.

KIM, J. 1998. Large-scale phylogenies and measuring the performance of phylogenetic estimators. *Systematic Biology* 47: 43–60.

KISHINO, H., AND M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* 29: 170–179.

KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Systematic Zoology* 18: 1–32.

KUMAR, S., K. TAMURA, I. B. JAKOBSEN, AND M. NEI. 2001. MEGA 2.

Molecular evolutionary genetics analysis software. Arizona State University, Tempe, Arizona, USA.

LIERE, K., AND G. LINK. 1995. RNA-binding activity of the *matK* protein encoded by the chloroplast *trnK* intron from mustard (*Sinapis alba* L.). *Nucleic Acids Research* 23: 917–921.

LYONS-WEILER, J., G. A. HOELZER, AND R. J. TAUSCH. 1996. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. *Molecular Biology and Evolution* 13: 749–757.

MATHEWS, S., AND M. J. DONOGHUE. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

MATHEWS, S., AND M. J. DONOGHUE. 2000. Basal angiosperm phylogeny inferred from duplicate phytochromes A and C. *International Journal of Plant Sciences* 161(Supplement): 41–55.

METCALFE, C. R., AND L. CHALK. 1950. Anatomy of the dicotyledons. Leaves, stem, and wood in relation to taxonomy with notes on economic uses, 1st ed., vols. 1–2, Clarendon Press, Oxford, UK.

MOHR, G., P. S. PERLMAN, AND A. M. LAMBOWITZ. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. *Nucleic Acids Research* 21: 4991–4997.

MÜLLER, K. 2002. PRAT: computer program for phylogenetic analyses of large data sets. Program distributed by the author, Botanical Institute, University of Bonn, Bonn, Germany. http://www.botanik.uni-bonn.de/system/download/PRAT

MÜLLER, K. 2003. QuickAlign: a new alignment editor. *Plant Molecular Biology Reporter* 21: 5.

MÜLLER, K., T. BORSCH, L. LEGENDRE, S. POREMBSKI, I. THEISEN, AND W. BARTHLOTT. In press. Evolution of carnivory in Lamiales and Lentibulariaceae. *Plant Biology*.

NANDI, O. I., M. W. CHASE, AND P. L. ENDRESS. 1998. A combined cladistic analysis of angiosperms using *rbcL* and non-molecular data sets. *Annals of the Missouri Botanical Garden* 85: 137–212.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast tRNALys (UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase related polypeptide. *Current Genetics* 11: 251–257.

NIXON, K. 1999. The parsimony ratchet: a rapid means for analyzing large data sets. *Cladistics* 15: 407–414.

OLMSTEAD, R. G., B. BREMER, K. SCOTT, AND J. D. PALMER. 1993. A parsimony analysis of the Asteridae *sensu lato* based on *rbc*L sequences. *Annals of the Missouri Botanical Garden* 80: 700–722.

OLMSTEAD, R., G. C. W. DEPAMPHILIS, A. D. WOLFE, N. D. YOUNG, W. J. ELISONS, AND P. A. REEVES. 2001. Disintegration of the Scrophulariaceae. *American Journal of Botany* 88: 348–361.

OLMSTEAD, R. G., R. K. JANSEN, K. J. KIM, AND S. J. WAGSTAFF. 2000. The phylogeny of the Asteridae s.l. based on chloroplast *ndhF* sequences. *Molecular Phylogenetics and Evolution* 16: 96–112.

OLMSTEAD, R. G., H. J. MICHAELS, K. M. SCOTT, AND J. D. PALMER. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbc*L. *Annals of the Missouri Botanical Garden* 79: 249–265.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

OLMSTEAD, R. G., P. A. REEVES, AND A. C. YEN. 1998. Patterns of sequence evolution and implications for parsimony analysis of chloroplast DNA. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants II, 164–187. Kluwer, Boston, Massachusetts, USA.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular datasets in the Solanaceae. *Systematic Biology* 43: 467–481.

PARKINSON, C. L., K. L. ADAMS, AND J. D. PALMER. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. *Current Biology* 9: 1485–1488.

POLLOCK, D. D., D. J. ZWICKL, J. M. MCGUIRE, AND D. M. HILLIS. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664–671.

QIU, Y.-L., M. W. CHASE, S. B. HOOT, E. CONTI, P. R. CRANE, K. J. SYTSMA, AND C. R. PARKS. 1998. Phylogenetics of the Hamamelidae and their allies: parsimony analyses of nucleotide sequences of the plastid gene *rbcL*. *International Journal of Plant Science* 159: 891–905.

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.

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. 2000. Phylogeny of basal angiosperms: analyses of five genes from three genomes. *International Journal of Plant Sciences* 161: S3–S27.

RANNALA, B., J. P. HUELSENBECK, Z. YANG, AND R. NIELSEN. 1998. Taxon sampling and the accuracy of large phylogenies. *Systematic Biology* 47: 702–710.

RENNER, S. S. 1999. Circumscription and phylogeny of the Laurales: evidence from molecular and morphological data. *American Journal of Botany* 86: 1301–1315.

RENNER, S. S., AND A. S. CHANDERBALI. 2000. What is the relationship among Hernandiaceae, Lauraceae, and Monimiaceae, and why is this question so difficult to answer? *International Journal of Plant Sciences* 161(Supplement): S109–S119.

SAUQUET, H., J. A. DOYLE, T. SCHARASCHKIN, T. BORSCH, K. W. HILU, L. W. CHATROU, AND A. LE THOMAS. 2003. Phylogenetic relationships in the Magnoliales and Myristicaceae: a multi-level approach based on morphology and molecules. *Botanical Journal of the Linnean Society* 142: 125–186.

SAVOLAINEN, V., M. W. CHASE, S. B. HOOT, C. M. MORTON, D. E. SOLTIS, C. BAYER, M. F. FAY, A. Y. DE BRUIJN, S. SULLIVAN, AND Y.-L. QIU. 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Systematic Biology* 49: 306–362.

SAVOLAINEN, V., ET AL. 2000b. Phylogeny of the eudicots: a nearly complete familial analysis of the *rbcL* gene sequences. *Kew Bulletin* 55: 257–309.

SHEAHAN, M. C., AND M. W. CHASE. 2000. Phylogenetic relationships within Zygophyllaceae based on DNA sequences of three plastid regions, with special emphasis on Zygophylloideae. *Systematic Botany* 25: 371–384.

SIMMONS, M. P., C. P. RANDLE, J. V. FREUDENSTEIN, AND J. W. WENZEL. 2002. Limitations of relative apparent synapomorphy analysis (RASA) for measuring phylogenetic signal. *Molecular Biology and Evolution* 19: 14–23.

SOLTIS, D. E., A. E. SENTERS, M. J. ZANIS, S. KIM, J. D. THOMPSON, P. S. SOLTIS, L. R. DE CRAENE, P. K. ENDRESS, AND J. S. FARRIS. 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany* 90: 461–470.

SOLTIS, D. E., AND P. S. SOLTIS. 1997. Phylogenetic relationships among Saxifragaceae sensu lato: a comparison of topologies based on 18S rDNA and *rbcL* sequences. *American Journal of Botany* 84: 504–522.

SOLTIS, D. E., AND P. S. SOLTIS. 1998. Choosing an approach and an appropriate gene for phylogenetic analysis. *In* P. S. Soltis, D. E. Soltis, and J. J. Doyle [eds.], Molecular systematics of plants, vol. II, 1–42. Kluwer, Boston, Massachusetts, USA.

SOLTIS, D. E., P. S. SOLTIS, D. R. MORGAN, S. M. SWENSEN, B. C. MULLIN, J. M. DOWD, AND P. G. MARTIN. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Science, USA* 92: 2647–2651.

SOLTIS, D. E., ET AL. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.

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: 402–404.

SOLTIS, P. S., ET AL. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.

SUZUKI, Y., G. V. GLAZKO, AND M. NEI. 2002. Overcredibility of molecular phylogenetics obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences, USA* 99: 16 138–16 143.

SWOFFORD, D. L. 2001. PAUP: phylogenetic analysis using parsimony, ver. 4.0b6. Sinauer, Sunderland, Massachusetts, USA.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, AND D. HILLIS. 1996. Phylogenetic inference. *In* D. M. Hillis, C. Moritz, and B. K. Mable [eds.], Molecular systematics, 2nd ed., 407–514. Sinauer, Sunderland, Massachusetts, USA.

TAKAHASHI, M. 1994. Exine development in *Illicium religiosum* Sieb. et Zucc. (Illiciaceae). *Grana* 33: 309–312.

TAKHTAJAN, A. 1987. System of Magnoliophyta. Academy of Sciences, USSR, Leningrad, Russia.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, AND D. G. HIGGINS. 1997. The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

THORNE, R. F. 1992. An updated classification of the flowering plants. *Aliso* 13: 365–389.

VOGEL, J., T. HÜBSCHMANN, T. BÖRNER, AND W. R. HESS. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for *matK* as an essential splice factor. *Journal of Molecular Biology* 270: 179–187.

WOLFE, A. D., C. W. MORDEN, AND J. D. PALMER. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Science, USA* 89: 10 648–10 652.

XIANG, Q.-Y., D. E. SOLTIS, AND P. S. SOLTIS. 1998. Phylogenetic relationships of Cornaceae and close relatives inferred from m*atK* and *rbcL* sequences. *American Journal of Botany* 85: 285–297.

ZANIS, M., D. E. SOLTIS, P. S. SOLTIS, S. MATHEWS, AND M. J. DONOGHUE. 2002. The root of angiosperms revisited. *Proceedings of the National Academy of Science, USA* 99: 6848–6853.

ZWICKL, D. J., AND D. M. HILLIS. 2002. Increased taxon sampling greatly reduces phylogenetic error. *Systematic Biology* 51: 588–598.

The origin and early evolution of plants on land

Paul Kenrick & Peter R. Crane

The origin and early evolution of land plants in the mid-Palaeozoic era, between about 480 and 360 million years ago, was an important event in the history of life, with far-reaching consequences for the evolution of terrestrial organisms and global environments. A recent surge of interest, catalysed by palaeobotanical discoveries and advances in the systematics of living plants, provides a revised perspective on the evolution of early land plants and suggests new directions for future research.

The origin and early diversification of land plants marks an interval of unparalleled innovation in the history of plant life. From a simple plant body consisting of only a few cells, land plants (liverworts, hornworts, mosses and vascular plants) evolved an elaborate two-phase life cycle and an extraordinary array of complex organs and tissue systems. Specialized sexual organs (gametangia), stems with an intricate fluid transport mechanism (vascular tissue), structural tissues (such as wood), epidermal structures for respiratory gas exchange (stomates), leaves and roots of various kinds, diverse spore-bearing organs (sporangia), seeds and the tree habit had all evolved by the end of the Devonian period. These and other innovations led to the initial assembly of plant-dominated terrestrial ecosystems, and had a great effect on the global environment.

Early ideas on the origin of land plants were based on living groups, but since the discovery of exceptionally well-preserved fossil plants in the Early Devonian Rhynie Chert, research has focused almost exclusively on the fossil record of vascular plants[1,2]. During the 1970s, syntheses of palaeobotanical and stratigraphic data emphasized the Late Silurian and Devonian periods as the critical interval during which the initial diversification of vascular plants occurred[1,2], and identified a group of simple fossils (rhyniophytes, such as *Cooksonia* and *Rhynia*) as the likely ancestral forms[2]. They also supported earlier hypotheses of two main lines of evolution: one comprising clubmosses (Fig. 1f) and extinct relatives, the other including all other living vascular plants (ferns, horsetails and seed plants; Fig. 1g–j) and related fossils[1,2]. During the past two decades, the discovery of fossil spores from as far back as the mid-Ordovician period[3], improved knowledge of living green algae[4,5], renewed interest in the phylogenetic position of other relevant groups such as mosses and liverworts[5], and advances in molecular systematics[5–14], together with unexpected new data on the structure and biology of Silurian and Devonian fossils[15–25], have provided a broader perspective on the origin of a land flora[26]. These new data indicate that the early diversification of land plants substantially pre-dates the Late Silurian to Early Devonian, and suggest that the main basal lineages originated over a period of more than 100 million years (Myr).

Patterns in the early fossil record

Evidence on the origin and diversification of land plants has come mainly from dispersed spores and megafossils. Gray recognized three new plant-based epochs (Eoembryophytic, Eotracheophytic and Eutracheophytic) spanning the origin and early establishment of land plants: each is characterized by the relative abundance of spore types and megafossils[3]. This synthesis highlights diversification and floral change in the Ordovician and Silurian[3,27,28], and emphasizes a major discrepancy between evidence from spores and megafossils: unequivocal land plant megafossils are first recognized in the fossil record roughly 50 Myr after the appearance of land plant spores.

Eoembryophytic (mid-Ordovician [early Llanvirn: ~476 Myr] to Early Silurian [late Llandovery: ~432 Myr])[3]. Spore tetrads (comprising four membrane-bound spores; Fig. 2d) appear over a broad geographic area in the mid-Ordovician and provide the first good evidence of land plants[3,26,29]. The combination of a decay-resistant wall (implying the presence of sporopollenin) and tetrahedral configuration (implying haploid meiotic products) is diagnostic of land plants. The precise relationships of the spore producers within land plants are controversial, but evidence of tetrads and other spore types (such as dyads) in Late Silurian and Devonian megafossils[16,30], as well as data on spore wall ultrastructure[25] and the structure of fossil cuticles[31], support previous suggestions of a land flora of liverwort-like plants (Fig. 1c)[3]. Some early spores and cuticles may also represent extinct transitional lineages between charophycean algae (Fig. 1a, b) and liverworts (Box 1), but precise understanding of their affinities is hindered by the dearth of associated megafossils.

Eotracheophytic (Early Silurian [latest Llandovery: ~432 Myr] to Early Devonian [mid Lochkovian: ~402 Myr])[3]. The early Silurian (latest Llandovery) marks the beginning of a decline in diversity of tetrads and a rise to dominance of individually dispersed, simple spores, which are found in several basal land plant groups (such as hornworts, some mosses, and early vascular plants)[3]. Although tetrads remain dominant in some Early Devonian localities from northwestern Europe[32], the elaboration of simple spores and turnover of spore 'species'[3] provide evidence of increasing land plant diversity and vegetational change. Although spores have been observed in Silurian megafossils, the affinities of most dispersed forms remain unknown, indicating that substantial land plant diversity is currently undetected in the megafossil record[30].

The earliest unequivocal land plant megafossils are from the mid-Silurian of northern Europe[33], and lowermost Upper Silurian of Bolivia[34] and Australia[35], and the uppermost Silurian of northwestern China[36]. Early assemblages include clubmosses (such as *Baragwanathia*) and related early fossils (such as zosterophylls, some species of *Cooksonia*), and various other plants of uncertain affinity (such as *Salopella* and *Hedeia*; Fig. 3). These data document an influx into land plant communities of diverse but generally small (usually less than 10 cm tall) organisms related to vascular plants (Fig. 3). Exceptions to the generally small size include the clubmoss *Baragwanathia*[37] and the large and much-branched *Pinnatiramosus* from the early Silurian of China[38]. The habit and branching of *Pinnatiramosus* is similar to that of green algae in the Caulerpales, but the presence of tracheid-like tubes is inconsistent with this interpretation[39]. Additional details, including conclusive data on reproductive structures, are needed to clarify the relationships of this enigmatic plant.

Data from northern Europe, Siberia, Podolia (southwestern Ukraine), Libya, Vietnam, Bolivia, Australia and Xinjiang and Yunnan (China) document increasing land plant diversity into

the base of the Devonian[33–36,40]. These fossils, together with the relative chronology implicit in current hypotheses of relationship, imply a minimum mid-Silurian origin for several important vascular plant groups (Box 1; Fig. 4).

Eutracheophytic (Early Devonian [late Lochkovian: ~398 Myr] to mid-Permian [~256 Myr])[3]. In the Early Devonian (late Lochkovian) the diversity of spores and megafossils increased dramatically[29,40–42]. Early assemblages include the classic floras from the Rhynie Chert[20,43,44], the Gaspé Peninsula of eastern Canada[43,44], New York State[43,44], the Rhine Valley of Germany[45], Belgium[46], Australia[35] and Yunnan Province (China)[33], which document a substantial increase in vascular plant diversity, including the appearance and early diversification of many important living groups.

Building a land plant

Phylogenetic studies favour a single origin of land plants from charophycean green algae (Box 1). Based on the ecology of living species, a freshwater origin of land plants seems likely, but direct evidence from the fossil record is inconclusive as mid-Palaeozoic charophytes are found in both freshwater and, more commonly, marine facies[47]. Living charophycean algae (Fig. 1a, b) possess several biosynthetic attributes that are expressed more fully among land plants, including the capacity to produce sporopollenin, cutin, phenolic compounds and the glycolate oxidase pathway[4,48]. However, the absence of well-developed sporophytes, gametophytes with sexual organs of land plant type, cuticle and non-motile, airborne, sporopollenin-walled spores suggests that these innovations evolved during the transition to the land[4,18]. In contrast to animal groups, the entire multicellular diploid phase of the plant life cycle probably evolved in a terrestrial setting.

The transition from an aqueous to a gaseous medium exposed plants to new physical conditions that resulted in key physiological and structural changes. Important metabolic pathways leading to lignins, flavenoids, cutins and plant hormones in vascular plants probably arose from pre-existing elements of primary metabolism in charophycean algae and bryophytes[4]. Although the evolution of these pathways is poorly understood, possible phenolic precursors have been detected in charophycean algae[31], and elements of auxin metabolism have been recognized in mosses and hornworts[49].

Phylogenetic studies predict that early land plants were small and morphologically simple, and this hypothesis is borne out by fossil evidence (Fig. 3). Early fossils bear a strong resemblance to the simple spore-producing phase of living mosses and liverworts (Fig. 1d, e and 5)[16,50–52], and these similarities extend to the anatomical details of the spore-bearing organs and the vascular system[19]. The extinct *Cooksonia* and similar early fossils (such as *Tortilicaulis*, *Uskiella*, *Caia*[15–17,81]) (Fig. 3) suggests that simple early land plants (once grouped as rhyniophytes[1]) are an unnatural assemblage[26]. Some *Cooksonia* species may be among the precursors to vascular plants (protracheophytes), whereas others are vascular plants apparently allied to the clubmoss lineage[26].

Clubmosses emerge from a poorly resolved grade of extinct *Zosterophyllum*-like plants (Fig. 4), although most zosterophylls form a monophyletic group[26]. Within clubmosses, early leafy herbaceous fossils such as *Baragwanathia* and *Asteroxylon* are basal[26,82], and living Lycopodiaceae are resolved as sister group to a calde that comprises the extinct herbaceous Protolepidodendrales, living *Selaginella* and the predominantly arborescent carboniferous lepidodendrids, including living *Isoetes*[26,82] (Fig. 4).

Euphyliophytes make up more than 99% of living vascular plants and exhibit much greater diversity than lycophytes. Relationships among basal euphyllophytes are still poorly understood[26]. Further progress requires a better understanding of the relationships of several fossil groups of uncertain status (such as Trimerophytina, Cladoxylales and Zygopteridales)[26,79].

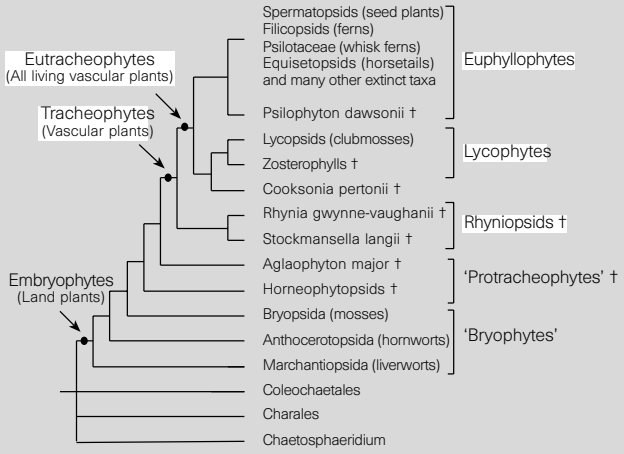
Box 1 Relationships among land plants

Land plants (embryophytes) are most closely related to the Charophyceae, a small group of predominantly freshwater green algae, within which either Coleochaetales (~15 living species; Fig. 1a) or Charales (~400 living species; Fig. 1b), or a group containing both, is sister group to land plants[4,5,10,12,74].

Land-plant monophyly is supported by comparative morphology[4,5,26,75] and gene sequences (18S rRNA, mitochondrial DNA: *cox III*)[12,14]. Relationships among the major basal living groups are uncertain[4,5,26,76,77], but the best-supported hypothesis resolves liverworts (Fig. 1c) as basal and either mosses (Fig. 1e) or hornworts (Fig. 1d) as the living sister group to vascular plants (tracheophytes)[4,5,13,14,26,75]. Less parsimonious hypotheses recognize bryophyte monophyly and either a sister-group relationship with vascular plants[26] or an origin from within basal vascular plants[14,76,78].

Among vascular plants, living ferns (Fig. 1g), horsetails (Fig. 1i) and seed plants (Fig. 1j) (euphyllophytes) are the sister group to clubmosses (Fig. 1f)[13,14,26,75,79]. Euphyllophyte monophyly is strongly supported by comparative morphology[26] and a unique 30-kb inversion in the chloroplast genome[8], as well as sequence data from 18S rRNA[13] and mitochondrial DNA (*cox III*)[14]. These data also provide evidence that the enigmatic Psilotaceae (Fig. 1h) (a group of simple plants once thought to be living relicts of the earliest vascular plants) are more closely related to the fern–seed plant lineage than to basal vascular plants (clubmosses or the extinct rhyniophytes). Within vascular plants, molecular and morphological assessments of phylogeny at the level of orders and below give similar results[11], but at deeper levels (for example, the divergence of major groups of ferns, horsetails and seed plants) phylogenetic resolution is poor. These difficulties highlight the problems of approaches based solely on living species[7,80]. Combined analyses of molecular sequences from multiple loci, and large-scale structural characteristics of the genome (such as introns and inversions), may be more useful in assessing deep phylogenetic patterns.

Megafossils fill some of the substantial morphological 'gaps' among living groups. Phylogenetic analyses[19,26] interpolate two Early Devonian Rhynie Chert plants, *Aglaophyton* and *Horneophyton*, between bryophytes and basal vascular plants as they possess some features unique to vascular plants (a branched, nutritionally independent sporophyte) but also retain bryophyte-like characteristics (terminal sporangia, columella in *Horneophyton*, and the absence of leaves, roots and tracheids with well-defined thickenings). The discovery of previously unrecognized diversity in



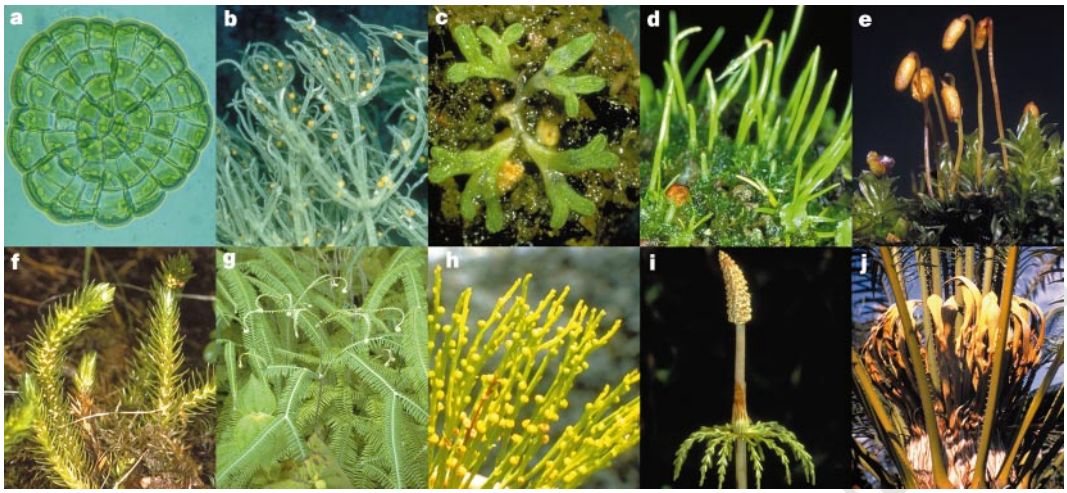


Figure 1 Morphological diversity among basal living land plants and potential land-plant sister groups. **a**, *Coleochaete orbicularis* (Charophyceae) gametophyte; magnification ×75 (photograph courtesy of L. E. Graham). **b**, *Chara* (Charophyceae) gametophyte; magnification ×1.5 (photograph courtesy of M. Feist). **c**, *Riccia* (liverwort) gametophyte showing sporangia (black) embedded in the thallus; magnification ×5 (photograph courtesy of A. N. Drinnan). **d**, *Anthoceros* (hornwort) gametophyte showing unbranched sporophytes; magnification ×2.5 (photograph courtesy of A. N. Drinnan). **e**, *Mnium* (moss) gametophyte showing unbranched sporophytes with terminal sporangia (c… sule); magnification ×4.5 (photograph courtesy of W. Burger). **f**, *Huperzia* (clubmoss) sporophyte with leaves showing sessile yellow sporangia; magnification ×0.8. **g**, *Dicksonia* (fern) sporophyte showing leaves with circinate vernation; magnification ×0.08. **h**, *Psilotum* (whisk fern) sporophyte with reduced leaves and sporangial synangia (three fused sporangia); magnification ×0.4. **i**, *Equisetum* (horsetail) sporophyte with whorled branches, reduced leaves, and a terminal cone; magnification ×0.4. **j**, *Cycas* (seed plant) sporophyte showing leaves and terminal cone with seeds; magnification ×0.05 (photograph courtesy of W. Burger).

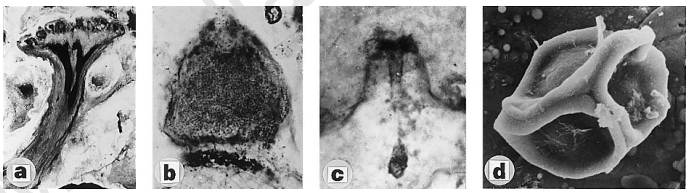


Figure 2 **a**, Longitudinal section (part of a silicified early fossil gametophyte (*Kidstonophyton discoides* from the Rhynie Chert). Antheridia (male sexual organs) are located on the upper surface of the branch; magnification ×3.4. **b**, Longitudinal section of an antheridium of *Lyonophyton rhyniensis* from the Rhynie Chert; magnification ×40. **c**, Longitudinal section of archegonium (female sexual organ) of *Langiophyton mackiei* from the Rhynie Chert; magnification ×80. **a**–**c** are from the Remy Collection (slides 200, 90 and 330), Abteilung Paläobotanik, Westfälische Wilhelms-Universität, Münster, Germany (photographs courtesy of H. Hass and H. Kerp). **d**, Scanning electron micrograph of *Tetrahedraletes medinensis* showing a spore tetrad of possible liverwort affinity from the Late Ordovician (photograph courtesy of W. A. Taylor); magnification ×670.

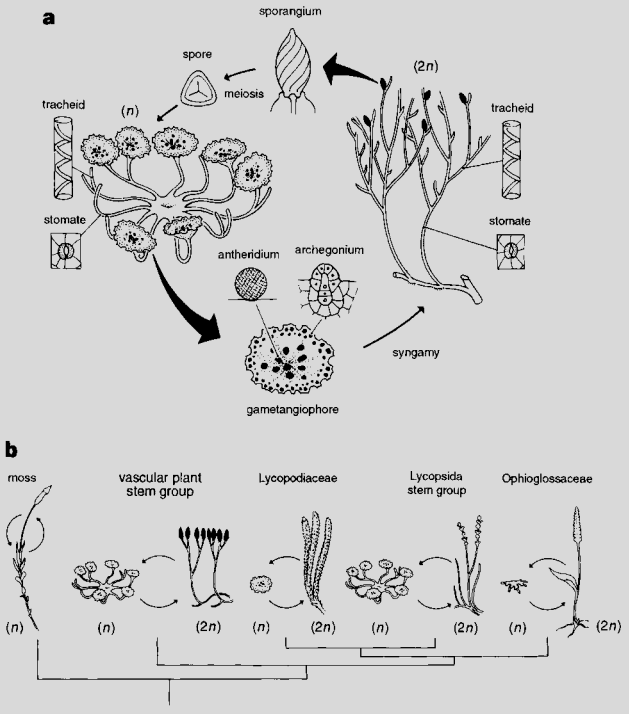
Figure 3 Sporophyte diversity in Early Devonian rhyniophyte fossils. **a**, *Cooksonia pertonii apiculispora*: sporophyte (incomplete proximally) with terminal sporangium[15]; magnification ×15. **b**, *Tortilicaulis offaeus*: sporophyte (incomplete proximally) with terminal sporangium[81]; magnification ×40. **c**. *Tortilicaulis offaeus*: sporophyte (incomplete proximally) with terminal bifurcating sporangium[81]; magnification ×30. **d**, Transverse section of sporangium showing thick wall and central spore mass; magnification ×70. **e**, Details of epidermis at rim of sporangium; magnification ×45. **f**, Stomate with two reniform guard cells (stippled); magnification ×120.

Box 2 Early evolution of the land plant life cycle

Land-plant life cycles are characterized by alternating multicellular sexual (haploid gametophyte, *n*) and asexual phases (diploid sporophyte, 2*n*). Phylogenetic studies indicate that land plants inherited a multicellular gametophyte from their algal ancestors but that the sporophyte evolved during the transition to the land. Most megafossils are sporophytes, and until recently there was no direct early fossil evidence for the gametophyte phase. Recent discoveries of gametophytes in the Rhynie Chert (Early Devonian, 380–408 Myr) have shed new light on the evolution of land-plant life cycles[18,20].

Early gametophytes (**a** in figure) are more complex than in living plants and have branched stems bearing sexual organs on terminal cup- or shield-shaped structures (Fig. 2a). Archegonia (female gametangia) are flask-shaped with a neck canal and egg chamber, and are sunken as in hornworts and most vascular plants (Fig. 2c). Antheridia (male gametangia) are roughly spherical, sessile or with a poorly-defined stalk, and superficial (Fig. 2b). Gametophytes are very similar to associated sporophytes, and shared anatomical features (water-conducting tissues, epidermal patterns, and stomates) have been used to link corresponding elements of the life cycle[18,20]. Our provisional reconstruction of the life cycle of an early vascular plant is based on information from anatomically preserved plants and contemporaneous compression fossils.

The similarities between gametophyte and sporophyte in early fossil vascular plants contrast strongly with the marked dissimilarities typical of living land plants (**b** in figure). The phylogenetic position of fossils suggests that, after the development of a simple, unbranched, 'parasitic' sporophyte among early land colonizers at the bryophyte grade (such as mosses) there was elaboration of both gametophyte and sporophyte in vascular plants. The implications for interpreting life cycles in living vascular plants[18,26] are shown. The small, simple, often subterranean and saprophytic gametophytes of living clubmosses (such as Lycopodiaceae) and ferns (such as Psiloataceae, Stromatopteridaceae, Ophioglossaceae) result from morphological loss. Phylogenetic evidence indicates that gametophyte reduction was independent in clubmosses and the fern–seed plant lineage. These data provide a new interpretation of the gametophyte morphology of living clubmosses (Lycopodiaceae)[18].



fossil record also documents significant differences from living groups, particularly in life cycles and the early evolution of the sexual phase (Box 2).

In common with some animal groups, internalization of vital functions and organs (such as gas exchange surfaces and sexual organs), combined with the development of impermeable exterior surfaces, seem to have been primary responses to life on land. Together, these changes resulted in more highly differentiated plants with stomates, multicellular sexual and spore-bearing organs, water-conducting and other tissue systems[52–54]. Morphological differentiation occurred in both phases of the life cycle (gametophyte and sporophyte), but there was subsequently a dramatic reduction in the gametophyte and a great increase in sporophyte complexity among vascular plants (Box 2). Apical growth and branching coupled with delayed initiation of spore-bearing organs were important innovations of vascular plants that led to a more complex architectural framework on which subsequent morphological diversification was based. The fossil record clearly shows that many vascular-plant organs can be interpreted in terms of modification (especially dichotomization and sterilization) of basic structural units such as the spore-bearing tissues and the stem[26,54]. In ferns and seed plants, much morphological diversity is clearly attributable to modifications of branching systems into a variety of leaf-like organs, whereas the relatively conservative clubmoss bauplan has a dearth of organ systems that can be interpreted as modified branches. In both lineages, however, meristem dormancy and abortion were early innovations, providing evidence of hormonal control and substantial phenotypic flexibility[21,26].

Early terrestrial ecosystems

The advent of land plants had important consequences for energy and nutrient fluxes among terrestrial and freshwater ecosystems[29,55] and hence for the evolution of animal groups that live in these habitats. The vegetational changes of the Silurian and Devonian also had a major impact on the atmosphere and other aspects of the global environment. The evolution of roots is thought to have been an important factor in the reduction of atmospheric CO_2 concentrations through increased weathering of Ca–Mg silicate minerals brought about by mechanical disruption and soil acidification[56,57]. Accelerated weathering has also been linked to the formation of Devonian and Early Carboniferous marine black shales[58], but this requires further investigation in view of similar deposits earlier and later in the geological record. Root-like impressions have been recognized in Late Silurian palaeosols[59], but the earliest unequivocal evidence comes from Early Devonian vascular plants[26], which have modified prostrate stems bearing rhizoids resembling those of living bryophytes. More substantial roots capable of anchoring large trees evolved independently in several groups during the Middle to Late Devonian.

A further series of innovations in vascular plants, including the biosynthesis of lignin and the origin of lateral meristems (cambium), were critical to the development of large plants, and these developments may have been stimulated by competition for light. Trees evolved independently in several major groups, resulting in stratified forest communities by the end of the Middle Devonian and the production of large amounts of highly decay-resistant organic material (in the form of lignified wood). The early evolution of lignin-decomposing fungi (some Ascomycetes, and Basidiomycetes) is still poorly understood[24], but these groups would have been essential for recycling much of the organic carbon.

The earliest land plants probably encountered terrestrial ecosystems that had been occupied by bacteria and protists[60,61], algae[4], lichens[23,62] and fungi[24] since the Late Proterozoic. A variety of enigmatic plants (such as *Protosalvinia*[44,63]) were also present, and some of the largest elements (*Prototaxites* 'trunks' >69 cm in diameter) may have been fungi[64]. Such organisms, or perhaps some rhyniophytes[16], may be the source of the microscopic tubular

Figure 4 Simplified phylogenetic tree showing the minimum stratigraphic ranges of selected groups based on megafossils (thick bars) and the minimum implied range extensions (thin lines). Also illustrated alongside the scale are minimum age estimates for the appearance of certain important land-plant features (from the bottom: spore tetrads, cuticles, single trilete spores, megafossils and stomates). The first unequivocal record of charophycean algae is based on calcified charalean oogonia (female sex organs) from the Late Silurian (Pridoli, ~410 Myr)[83] and distinctive gametophytes from the Early Devonian Rhynie Chert[44]. Proposed similarities between living *Coleochaete* and Early Devonian *Parka* remain to be confirmed[44]. Note that megafossil evidence for vascular plants precedes megafossil evidence of bryophytes and charophycean algae. Confirmation that the Early Devonian *Sporogonites* is a plant at the bryophyte grade could help to reduce this discrepancy. Tre, Tremadoc; Arg, Arenig; Lln, Llanvirn; Llo, Llandeilo; Crd, Caradoc; Ash, Ashgill; Lly, Llandovery; Wen, Wenlock; Lud, Ludlow; Pri, Pridoli; Lok, Lochkovian (Gedinnian); Prg, Pragian (Siegenian); Ems, Emsian; Eif, Eifelian; Giv, Givetian; Frs, Frasnian; Fam, Famennian; Tou, Tournaisian; Vis, Visean; Spk, Serpukhovian; Bsh, Bashkirian; Mos, Moscovian; Kas, Kasimovian; Gze, Gzelian.

Figure 5 Diversity of water-conducting cells (tracheids) in early land plants (median longitudinal section through cells, basal and proximal end walls not shown; cells are ~20–40 μm diameter). **a**, Top, bryophyte hydroid; bottom, details of hydroid wall showing distribution of plasmodesmata-derived micropores (10–50 nm diameter; stipple)[94]. **b**, Top, S-type tracheid (fossil) of Rhyniopsida; bottom, details of S-type cell wall showing distribution of plasmodesmata-derived micropores (stipple) and 'spongy' interior to thickenings[19]. **c**, Top, G-type tracheid (fossil) of basal extinct eutracheophytes, which closely resemble the tracheids of some living vascular plants; bottom, details of G-type cell wall showing pores distributed between thickenings[19]. **d**, Top, scalariform pitted P-type tracheid (fossil) typical of trimerophyte grade plants (euphyllophytes); bottom, details of P-type cell wall showing pit chambers and sheet with pores that extends over pit apertures[26].

fragments commonly extracted from Silurian and Early Devonian sediments[28]. These tubes are often associated with cellular cuticular fragments (*Nematothallus* and *Cosmochlaina*) that may represent fragmented cuticular material from bryophyte-like plants[31]. The discovery of fungal arbusculae in Early Devonian megafossils[22] confirms earlier suggestions that endomycorrhizal associations were an important innovation in the colonization of the land[65].

In contrast to megascopic plants, which appear to have colonized the land only once, many animal groups made the transition to terrestrial existence independently and overcame the problems of water relations in different ways[52,66,67]. Early evidence for terrestrial animals is sparse[29,67–69], but by the Early Devonian exquisitely preserved arthropod faunas are known from several localities in North America, Germany and the United Kingdom[29,66,67]. These faunas document the appearance of diverse arthropod communities including centipedes, millipedes, trigonotarbids and their living relatives spiders, pseudoscorpians, mites (orbatids and endeostigmatids), arthropleurids (extinct arthropods), archaeognathans (primitive wingless insects), collembolans and possibly bristletails. Available evidence indicates that these animals were mainly predators and detritivores and, until the appearance of vertebrate herbivores in the latest Palaeozoic, most energy flow into animal components of early terrestrial ecosystems was probably through the decomposer pathway rather than direct herbivory[29]. Indirect evidence for herbivory comes from wound responses in the tissues of some fossil plants[70,71], and perhaps also from fossil faecal pellets containing abundant spores[70,72].

Future directions

The fossil record of spores, combined with phylogenetic studies, indicates that groups related to living bryophytes were early colonisers of the land, and suggests that several major lineages of vascular plant had already evolved by the mid-Silurian. Megafossils of land plants, however, appear much later. In these assemblages there is a conspicuous bias toward the recognition and perhaps representation of vascular plants. The most important source of data on early megafossils has been the northern European (Laurussian) region, but the appearance of megafossils in this area coincides with facies changes driven by a widespread marine regression[28,73], and the Silurian land-plant megafossils are from marine sediments[3]. It seems likely that the onset of continental conditions in the Devonian of northern Europe allowed megafossils to be preserved at a time when vascular plants were well established but still diversifying. The rapid appearance of vascular plants in this region owes as much to changing geological conditions as to rapid biological diversification[27,28]. Intensified sampling in areas that are remote from these regional events is therefore a high priority.

Palaeobotanical evidence shows that the major groups of living land plants are relicts, even though much modern species diversity *within* these groups may have evolved more recently. Data from the fossil record are therefore especially important for clarifying homologies among major organ systems which may otherwise be difficult to detect as a result of morphological divergence and extinction. Such combined studies of living and fossil plants provide an improved basis for comparative studies of plant development. They indicate, for example, that the ontogeny of leaves and spore-bearing organs in clubmosses are likely to share substantial similarities, but are unlikely to exhibit common features with leaves in seed plants, ferns and horsetails. They also suggest that fundamental features of land plants, such as the spore-bearing organs, stems, stomates and sexual organs, are each under the same kind of developmental control in all main groups. To explore these issues further, data are needed on the molecular basis of plant development from a broader selection of land plants than are currently under study. In the context of a more complete understanding of plant diversity than that provided by living plants alone, such data should be expected to confirm the underlying unity and relative simplicity of developmental processes in land plants. □

Paul Kenrick is at the Swedish Museum of Natural History, Box 50007, S-104 05, Stockholm, Sweden; Peter R. Crane is at The Field Museum, Roosevelt Road at Lake Shore Drive, Chicago, Illinois 60605, and the Department of the Geophysical Sciences, University of Chicago, USA.

1. Banks, H. P. Reclassification of Psilophyta. *Taxon* **24,** 401–413 (1975).
2. Chaloner, W. G. & Sheerin, A. in *The Devonian System* (eds House, M. R., Scrutton, C. T. & Bassett, M. G.) 145–161 (The Palaeontological Association, London, 1979).
3. Gray, J. Major Paleozoic land plant evolutionary bio-events. *Palaeogeog. Palaeoclimatol. Palaeoecol.* **74,** 153–169 (1993).
4. Graham, L. E. *Origin of Land Plants* (Wiley, New York, 1993).
5. Mishler, B. D. *et al.* Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. MO Bot. Gard.* **81,** 451–483 (1994).
6. Manhart, J. R. & Palmer, J. G. The gain of two chloroplast tRNA introns marks the green algal ancestors of land plants. *Nature* **345,** 268–270 (1990).
7. Manhart, J. R. Phylogenetic analysis of green plant *rbc*L sequences. *Mol. Phylogenet. Evol.* **3,** 114–127 (1994).
8. Raubeson, L. A. & Jansen, R. K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255,** 1697–1699 (1992).
9. Chapman, R. L. & Buchheim, M. A. Ribosomal RNA gene sequences: analysis and significance in the phylogeny and taxonomy of green algae. *Rev. Plant Sci.* **10,** 343–368 (1991).
10. McCourt, R. M., Karol, K. G., Guiry, M. D. & Feist, M. Phylogeny of extant genera in the family Characeae (Charales, Charophyceae) based on *rbc*L sequences and morphology. *Am. J. Bot.* **83,** 125–131 (1996).
11. Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbc*L sequences. *Am. Fern J.* **85,** 205–282 (1995).
12. Kranz, H. D. *et al.* Towards the origin of land plants: phylogenetic relationships among charophytes, bryophytes and vascular plants inferred from complete small-subunit ribosomal RNA gene sequences. *J. Mol. Evol.* **41,** 74–84 (1995).
13. Kranz, H. D. & Huss, V. A. R. Molecular evolution of pteridophytes and their relationships to seed plants: evidence from complete 18S rRNA gene sequences. *Plant Syst. Evol.* **202,** 1–11 (1996).
14. Kessel, R., von Haeseler, A. & Brennicke, A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **91,** 634–638 (1994).
15. Edwards, D., Davies, K. L. & Axe, L. A vascular conducting strand in the early land plant *Cooksonia*. *Nature* **357,** 683–685 (1992).
16. Edwards, D., Duckett, J. G. & Richardson, J. B. Hepatic characters in the earliest land plants. *Nature* **374,** 635–636 (1995).
17. Fanning, U., Edwards, D. & Richardson, J. B. A diverse assemblage of early land plants from the Lower Devonian of the Welsh Borderland. *Bot. J. Linn. Soc.* **109,** 161–188 (1992).
18. Kenrick, P. Alternation of generations in land plants: new phylogenetic and morphological evidence. *Biol. Rev.* **69,** 293–330 (1994).
19. Kenrick, P. & Crane, P. R. Water-conducting cells in early fossil land plants: implications for the early evolution of tracheophytes. *Bot. Gaz.* **152,** 335–356 (1991).
20. Remy, W. Gensel, P. G. & Hass, H. The gametophyte generation of some early Devonian land plants. *Int. J. Plant Sci.* **154,** 35–58 (1993).
21. Remy, W. & Hass, H. New information on gametophytes and sporophytes of *Aglaophyton major* and inferences about possible environmental adaptations. *Rev. Palaeobot. Palynol.* **90,** 175–194 (1996).
22. Remy, W., Taylor, T. N., Hass, H. & Kerp, H. Four hundred-million-year-old vesicular arbuscular mycorrhizae. *Proc. Nat Acad. Sci. USA* **91,** 11841–11843 (1994).
23. Stein, W. E., Harmon, G. D. & Hueber, F. M. in *International Workshop on the Biology and Evolutionary Implications on Early Devonian Plants* (Westfälische Wilhelms-Universität Münster, Germany, 1994).
24. Taylor, T. N. & Osborne, J. M. The Importance of fungi in shaping the paleoecosystem. *Rev. Palaeobot. Palynol.* **90,** 249–262 (1996).
25. Taylor, W. A. Ultrastructure of lower Paleozoic dyads from southern Ohio. *Rev. Palaeobot. Palynol.* **92,** 269–280 (1996).
26. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institution Press, Washington DC, 1997).
27. Gray, J. The microfossil record of early land plants: advances in understanding of early terrestrialization, 1970–1984 *Phil. Trans. R. Soc. Lond. B* **309,** 167–195 (1985).
28. Gray, J. & Boucot, A. J. Early vascular land plants: proof and conjecture. *Lethaia* **10,** 145–174 (1977).
29. DiMichele, W. A. *et al.* in *Terrestrial Ecosystems Through Time: Evolutionary Paleoecology of Terrestrial Plants and Animals* (ed. Behrensmeyer, A. K.) 205–325 (Univ. Chicago Press, 1992).
30. Fanning, U., Richardson, J. B. & Edwards, D. in *Pollen and Spores* (eds Blackmore, S. & Barnes, S. H.) 25–47 (Clarendon, Oxford, 1991).
31. Kroken, S. B., Graham, L. E. & Cook, M. E. Occurrence and evolutionary significance of resistant cell walls in charophytes and bryophytes. *Am. J. Bot.* **83,** 1241–1254 (1996).
32. Wellman, C. H. & Richardson, J. B. Sporomorph assemblages from the 'Lower Old Red Sandstone' of Lorne, Scotland. *Special Papers Palaeontol.* **55,** 41–101 (1996).
33. Edwards, D. in *Palaeozoic Palaeogeography and Biogeography* (eds McKerrow, W. S. & Scotese, C. R.) 233–242 (Geological Society, London, 1990).
34. Morel, E., Edwards, D. & Iniquez Rodriguez, M. The first record of *Cooksonia* from South America in the Silurian rocks of Bolivia. *Geol. Mag.* **132,** 449–452 (1995).
35. Tims, J. D. & Chambers, T. C. Rhyniophytina and Trimerophytina from the early land flora of Victoria, Australia. *Palaeontology* **27,** 265–279 (1984).
36. Cai, C.-Y., Dou, Y.-W. & Edwards, D. New observations on a Pridoli plant assemblage from north Xinjiang, northwest China, with comments on its evolutionary and palaeographical significance. *Geol. Mag.* **130,** 155–170 (1993).
37. Hueber, F. M. Thoughts on the early lycopsids and zosterophylls. *Ann. MO Bot. Gard.* **79,** 474–499 (1992).
38. Cai, C. *et al.* An early Silurian vascular plant. *Nature* **379,** 592 (1996).
39. Geng, B.-Y. Anatomy and morphology of *Pinnatiramosus*, a new plant from the Middle Silurian (Wenlockian) of China. *Acta Bot. Sin.* **28,** 664–670 (1986).
40. Raymond, A. & Metz, C. Laurussian land-plant diversity during the Silurian and Devonian: mass extinction, sampling bias, or both? *Paleobiology* **21,** 74–91 (1995).
41. Edwards, D. & Davies, M. S. in *Major evolutionary radiations* (eds Taylor, P. D. & Larwood, G. P.) 351–376 (Clarendon, Oxford, 1990).
42. Knoll, A. H., Niklas, K. J., Gensel, P. G. & Tiffney, B. H. Character diversification and patterns of evolution in early vascular plants. *Paleobiology* **10,** 34–47 (1984).
43. Gensel, P. G. & Andrews, H. N. *Plant Life in the Devonian* (Praeger, New York, 1984).

44. Taylor, T. N. & Taylor, E. L. *The Biology and Evolution of Fossil Plants* (Prentice Hall, New Jersey, 1993).
45. Schweitzer, H.-J. Die Unterdevonflora des Rheinlandes. *Palaeontographica B* **189,** 1–138 (1983).
46. Gerrienne, P. Inventaire des végétaux éodévoniens de Belgique. *Ann. Soc. Géol. Belg.* **116,** 105–117 (1993).
47. Tappan, H. N. *The Paleobiology of Plant Protists* (Freeman, San Francisco, 1980).
48. Raven, J. Plant responses to high O_2 concentrations: relevance to previous high O_2 episodes. *Palaeogreg. Palaeoclimatol. Palaeocol.* **97,** 19–38 (1991).
49. Sztein, A. E., Cohen, J. D., Slovin, J. P. & Cooke, T. J. Auxin metabolism in representative land plants. *Am. J. Bot.* **82,** 1514–1521 (1995).
50. Edwards, D. New insights into early land ecosystems: a glimpse of a Lilliputian world. *Rev. Palaeobot. Palynol.* **90,** 159–174 (1996).
51. Edwards, D., Fanning, U. & Richardson, J. B. Stomata and sterome in early land plants. *Nature* **323,** 438–440 (1986).
52. Raven, J. A. Comparative physiology of plant and arthropod land adaptation. *Phil. Trans. R. Soc. Lond. B* **309,** 273–288 (1985).
53. Raven, J. A. The evolution of vascular plants in relation to quantitative functioning of dead water-conducting cells and stomata. *Biol. Rev.* **68,** 337–363 (1993).
54. Niklas, K. J. *Plant Allometry: The Scaling of Form and Process.* (Univ. Chicago Press, 1994).
55. Beerbower, R. in *Geological Factors and the Evolution of Plants* (ed. Tiffney, B. H.) 47–92 (Yale Univ. Press, New Haven, CT, 1985).
56. Berner, R. A. GEOCARB II: a revised model of atmospheric CO_2 over Phanerozoic time. *Am. J. Sci.* **294,** 56–91 (1994).
57. Mora, C. I., Driese, S. G. & Colarusso, L. A. Middle to Late Paleozoic atmospheric CO_2 levels from soil carbonate and organic matter. *Science* **271,** 1105–1107 (1996).
58. Algeo, T. J., Berner, R., Maynard, J. B. & Scheckler, S. E. Late Devonian oceanic anoxic events and biotic crises: "rooted" in the evolution of vascular land plants? *GSA Today* **5,** 45, 64–66 (1995).
59. Retallack, G. J. in *Paleosols: their Recognition and Interpretation* (ed. Wright, V. P.) (Blackwell, Oxford, 1986).
60. Knoll, A. H. The early evolution of eukaryotes: a geological perspective. *Science* **256,** 622–627 (1992).
61. Bengtson, S. (ed) *Early life on Earth.* (Columbia Univ. Press, New York, 1994).
62. Taylor, T. N., Hass, H., Remy, W. & Kerp, H. The oldest fossil lichen. *Nature* **378,** 244 (1995).
63. Hemsley, A. R. in *Ultrastructure of Fossil Spores and Pollen* (eds Kurmann, M. H. & Doyle, J. A.) 1–21 (Royal Botanic Gardens, Kew, 1994).
64. Hueber, F. M. in *International Workshop on the Biology and Evolutionary Implications of Early Devonian Plants* (Westfälische Wilhelms-Universität, Münster, 1994).
65. Simon, L., Bousquet, J., Léveque, C. & Lalonde, M. Origin and diversification of endomycorrhizal fungi with vascular plants. *Nature* **363,** 67–69 (1993).
66. Selden, P. A. & Edwards, D. in *Evolution and the Fossil Record* (eds Allen, K. C. & Briggs, D. E. G.) 122–152 (Belhaven, London, 1989).
67. Gray, J. & Shear, W. Early life on land. *Am. Sci.* **80,** 444–456 (1992).
68. Gray, J. & Boucot, A. J. Early Silurian nonmarine animal remains and the nature of the early continental ecosystem. *Acta Palaeontol. Pol.* **38,** 303–328 (1994).
69. Retallack, G. J. & Feakes, C. R. Trace fossil evidence for Late Ordovician animals on land. *Science* **235,** 61–63 (1987).
70. Scott, A. C., Stephenson, J. & Chaloner, W. G. Interaction and coevolution of plants and arthropods during the Palaeozoic and Mesozoic. *Phil. Trans. R. Soc. Lond. B* **336,** 129–165 (1992).
71. Banks, H. P. & Colthart, B. J. Plant-animal-fungal interactions in early Devonian trimerophytes from Gaspé, Canada. *Am. J. Bot.* **80,** 992–1001 (1993).
72. Edwards, D., Seldon, P. A., Richardson, J. B. & Axe, L. Coprolites as evidence for plant-animal interaction in Siluro-Devonian terrestrial ecosystems. *Nature* **377,** 329–331 (1995).
73. Allen, J. R. L. Marine to fresh water: the sedimentology of the interrupted environmental transition (Ludlow-Siegenian) in the Anglo-Welsh region. *Phil. Trans. R. Soc. Lond. B* **309,** 85–105 (1985).
74. Melkonian, M. & Surek, B. Phylogeny of the Chlorophyta: congruence between ultrastructural and molecular evidence. *Bull. Soc. Zool. Fr.* **120,** 191–208 (1995).
75. Bremer, K., Humphries, C. J., Mishler, B. D. & Churchill, S. P. On cladistic relationships in green plants. *Taxon* **36,** 339–349 (1987).
76. Garbary, D. J., Renzaglia, K. S. & Duckett, J. G. The phylogeny of land plants: a cladistic analysis based on male gametogenesis. *Plant Syst. Evol.* **188,** 237–269 (1993).
77. Capesius, I. A molecular phylogeny of bryophytes based on nuclear-encoded 18S rRNA genes. *J. Plant Physiol.* **146,** 59–63 (1995).
78. Taylor, T. N. The origin of land plants: some answers, more questions. *Taxon* **37,** 805–833 (1988).
79. Rothwell, G. W. in *Pteridology in Perspective* (eds Cantino, J. A., Gibby, M. & Johns, R. J.) (Royal Botanic Gardens, Kew) (in the press).
80. Albert, V. A. *et al.* Functional constraints and rDNA-based evidence for land plant phylogeny. *Ann. MO Bot. Gard.* **81,** 534–567 (1994).
81. Edwards, D., Fanning, U. & Richardson, J. B. Lower Devonian coalified sporangia from Shropshire: *Salopella* Edwards & Richardson gen. nov. *Horneophyton lignieri* Edwards. *Bot. J. Linn. Soc.* **116,** 89–110 (1994).
82. Bateman, R. M., DiMichele, W. A. & Willard, D. A. Experimental cladistic analysis of anatomically preserved lycopsids from the Carboniferous of Euramerica: an essay on paleobotanical phylogenetics. *Ann. MO Bot. Gard.* **79,** 500–559 (1992).
83. Feist, M. & Grambast-Fessard, N. in *Calcareous Algae and Stromatolites* (ed. Riding, R.) 189–203 (Springer, Berlin, 1991).
84. Hébant, C. in *Bryophyte Systematics* (eds Clarke, G. C. S. & Duckett, J. G.) 365–383 (Academic, London, 1979).

Acknowledgements. We thank W. G. Chaloner, D. Edwards, J. A. Raven, P. S. Herendeen, E. M. Friis, S. Bengtson and especially J. Gray for criticisms of earlier drafts of this manuscript; W. Burger, J. Cattel, A. N. Drinnan, M. Feist, L. E. Graham, H. Haas, H. Kerp, W. A. Taylor and P. Lidmark for assistance with illustrations. This work was supported in part by the Swedish Natural Science Research Council (NFR) and the National Science Foundation.

Correspondence should be addressed to P.K. (e-mail: pb-paul@nrm.se).

joining trees and the amino-acid maximum parsimony phylogenies, and 100 replicates for the nucleotide maximum likelihood tree and the amino-acid distance-based analyses (Dayhoff PAM matrix) (see Supplementary Information for additional trees and summary of bootstrap support). We performed tests of alternative phylogenetic hypotheses using Kishino–Hasegawa[29] (parsimony and likelihood) and Templeton's non-parametric[30] tests.

Received 30 October; accepted 4 December 2000.

1. Eisenberg, J. F. *The Mammalian Radiations* (Chicago Univ. Press, Chicago, 1981).
2. Novacek, M. J. Mammalian phylogeny: shaking the tree. *Nature* **356,** 121–125 (1992).
3. O'Brien, S. J. *et al.* The promise of comparative genomics in mammals. *Science* **286,** 458–481 (1999).
4. Springer, M. S. *et al.* Endemic African mammals shake the phylogenetic tree. *Nature* **388,** 61–64 (1997).
5. Stanhope, M. J. *et al.* Highly congruent molecular support for a diverse clade of endemic African mammals. *Mol. Phylogenet. Evol.* **9,** 501–508 (1998).
6. McKenna, M. C. & Bell, S. K. *Classification of Mammals above the Species Level* (Columbia Univ. Press, New York, 1997).
7. Mouchatty, S. K., Gullberg, A., Janke, A. & Arnason, U. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. *Mol. Biol. Evol.* **17,** 60–67 (2000).
8. Stanhope, M. J. *et al.* Molecular evidence for multiple origins of Insectivora and for a new order of endemic African insectivore mammals. *Proc. Natl Acad. Sci. USA* **95,** 9967–9972 (1998).
9. Simpson, G. G. The principles of classification and a classification of mammals. *Bull. Am. Mus. Nat. Hist.* **85,** 1–350 (1945).
10. McKenna, M. C. in *Phylogeny of the Primates: a Multidisciplinary Approach* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
11. Easteal, S. The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics* **124,** 165–173 (1990).
12. Kumar, S. & Hedges, S. B. A molecular timescale for vertebrate evolution. *Nature* **392,** 917–920 (1998).
13. Reyes, A. *et al.* Where do rodents fit? Evidence from the complete mitochondrial genome of *Sciurus vulgaris*. *Mol. Biol. Evol.* **17,** 979–983 (2000).
14. Gregory, W. K. The orders of mammals. *Bull. Am. Mus. Nat Hist.* **27,** 1–524 (1910).
15. Graur, D., Duret, L. & Gouy, M. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). *Nature* **379,** 333–335 (1996).
16. Huchon, D., Catzeflis, F. M. & Douzery, E. J. P. Molecular evolution of the nuclear von Willebrand factor gene in mammals and the phylogeny of rodents. *Mol. Biol. Evol.* **16,** 577–589 (1999).
17. Kay, R. F., Ross, C. & Williams, B. A. Anthropoid origins. *Science* **275,** 797–804 (1997).
18. Pettigrew, J. D. Flying Primates? Megabats have advanced pathway from eye to midbrain. *Science* **231,** 1304–1306 (1986).
19. Waddell, P. J., Okada, N. & Hasegawa, M. Towards resolving the interordinal relationships of placental mammals. *Syst. Biol.* **48,** 1–5 (1999).
20. Krettek, A., Gullberg, A. & Arnason, U. Sequence analysis of the complete mitochondrial DNA molecule of the hedgehog, *Erinaceus europaeus*, and the phylogenetic position of the Lipotyphla. *J. Mol. Evol.* **41,** 952–957 (1995).
21. Rose, K. D. & Emry, R. J. in *Mammal Phylogeny: Placentals* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 81–102 (Springer, New York, 1993).
22. Miyamoto, M. M. & Goodman, M. Biomolecular systematics of eutherian mammals: phylogenetic patterns and classification. *Syst. Zool.* **35,** 230–240 (1986).
23. Madsen, O. *et al.* Parallel adaptive radiations in two major clades of placental mammals. *Nature* (in the press).
24. Eizirik, E., Murphy, W. J. & O'Brien, S. J. Molecular dating and biogeography of the early placental mammal radiation. *J. Hered.* (in the press).
25. Hedges, S. B., Parker, P. H., Sibley, C. G. & Kumar, S. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381,** 226–229 (1996).
26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25,** 4876–4882 (1997).
27. Swofford, D. L. *PAUP* Phylogenetic Analysis using Parsimony and Other Methods* (Sinauer, Sunderland, Massachusetts, 1998).
28. Strimmer, K. & Von Haeseler, A. Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* **13,** 964–969 (1996).
29. Kishino, H. & Hasegawa, M. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29,** 170–179 (1989).
30. Templeton, A. R. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* **37,** 221–244 (1983).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*. Sequences are deposited in GenBank under accession numbers AY011125–AY012154.

Acknowledgements

We thank D. Hirschmann, M. Houck, R. Montali, R. Baker, G. Harris, K. Helgen, A. L. Roca, M. Roelke-Parker, A. Grafodatsky, O. Serov and T. Oleksyk for help in obtaining samples and technical assistance, and M. Smith and M. Dean for helpful suggestions. We also thank the NCI Frederick Molecular Technology Center for technical support, and the Advanced Biomedical Computer Center for computational assistance. All tissue samples were obtained with appropriate permits (CITES) issued to the National Cancer Institute, National Institutes of Health (principal officer, S.J.O). Y.P.Z. is supported by the Natural Science Foundation of China and Chinese Academy of Sciences. E.E. is supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

Correspondence and requests for materials should be addressed to S.J.O. (e-mail: obrien@mail.ncifcrf.gov).

Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants

Kathleen M. Pryer*, Harald Schneider*, Alan R. Smith†, Raymond Cranfill†, Paul G. Wolf‡, Jeffrey S. Hunt* & Sedonia D. Sipes‡

* *Department of Botany, The Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, Illinois 60605, USA*
† *University Herbarium, University of California, 1001 Valley Life Sciences Building 2465, Berkeley, California 94720, USA*
‡ *Department of Biology, 5305 Old Main Hill, Utah State University, Logan, Utah 84322, USA*

Most of the 470-million-year history of plants on land belongs to bryophytes, pteridophytes and gymnosperms, which eventually yielded to the ecological dominance by angiosperms 90 Myr ago[1–3]. Our knowledge of angiosperm phylogeny, particularly the branching order of the earliest lineages, has recently been increased by the concurrence of multigene sequence analyses[4–6]. However, reconstructing relationships for all the main lineages of vascular plants that diverged since the Devonian period has remained a challenge. Here we report phylogenetic analyses of combined data—from morphology and from four genes—for 35 representatives from all the main lineages of land plants. We show that there are three monophyletic groups of extant vascular plants: (1) lycophytes, (2) seed plants and (3) a clade including equisetophytes (horsetails), psilotophytes (whisk ferns) and all eusporangiate and leptosporangiate ferns. Our maximum-likelihood analysis shows unambiguously that horsetails and ferns together are the closest relatives to seed plants. This refutes the prevailing view that horsetails and ferns are transitional evolutionary grades between bryophytes and seed plants[7], and has important implications for our understanding of the development and evolution of plants[8].

Estimates of a phylogeny for the main groups of land plants, each with highly divergent morphologies, have been many, and all have been contested. Bryophytes (liverworts, hornworts and mosses) are consistently shown to be a basal grade, but their relationships to one another and to vascular plants are controversial[1,2,9–13]. Most phylogenetic analyses of vascular plants consistently reconstruct two main lines of evolution: the Lycophytina (clubmosses and relatives), with 1% of extant diversity, and the Euphyllophytina (all other vascular plants)[1,2,10,11,14–17]. Extant Euphyllophytina[1,2] comprises six major monophyletic lineages: Equisetopsida (horsetails), Polypodiidae (leptosporangiate ferns), Spermatophytata (seed plants), Psilotidae (whisk ferns; simple plants regarded by some to be living relicts of the earliest vascular plants[7,18]), Marattiidae and Ophioglossidae (eusporangiate ferns). Phylogenetic assessments based on single genes[10,11,14,15,19] and/or morphology[1,7,12,17,20] have provided only weak and usually contradictory evidence for the relationships among these euphyllophyte lineages. Resolving these relationships would not only stabilize a pivotal region of vascular plant phylogeny but is also key to identifying the most appropriate outgroup for addressing questions related to the evolution of seed plants.

Recent palaeontological studies[1,2,7] attempted to demonstrate that approaches based solely on living species would have difficulties reconstructing relationships among major lineages of vascular plants. Inadequate taxon sampling, rate heterogeneity across DNA nucleotide sites among lineages, and inappropriate algorithms also have been cited as impediments to resolving ancient branching events[21]. However, as predicted by recent theoretical studies[22], combined analysis of DNA sequences from multiple loci proves to

be very useful in inferring deep phylogenetic patterns[4–6]. With few exceptions[12,20], broad phylogenetic studies rely solely on combined nucleotide sequence data, with authors arguing that morphological character homology assessment among ancient and divergent groups is too challenging. This practice ignores the higher complexity of morphological characters that can conserve character states over time and that have a lower probability of random evolution of similar structures.

We obtained DNA sequences (5,072 aligned base pairs) of four genes from two plant genomes: plastid *atpB*, *rbcL* and *rps4*, and nuclear small-subunit ribosomal DNA. We also assembled a congruent data set of 136 vegetative and reproductive morphological/anatomical characters. We sampled 35 representatives from all major monophyletic lineages of land plants. The selection of taxa reflects our focus on basal vascular plants, and all six Euphyllophytina[1] lineages are represented by two or more members. Five bryophytes

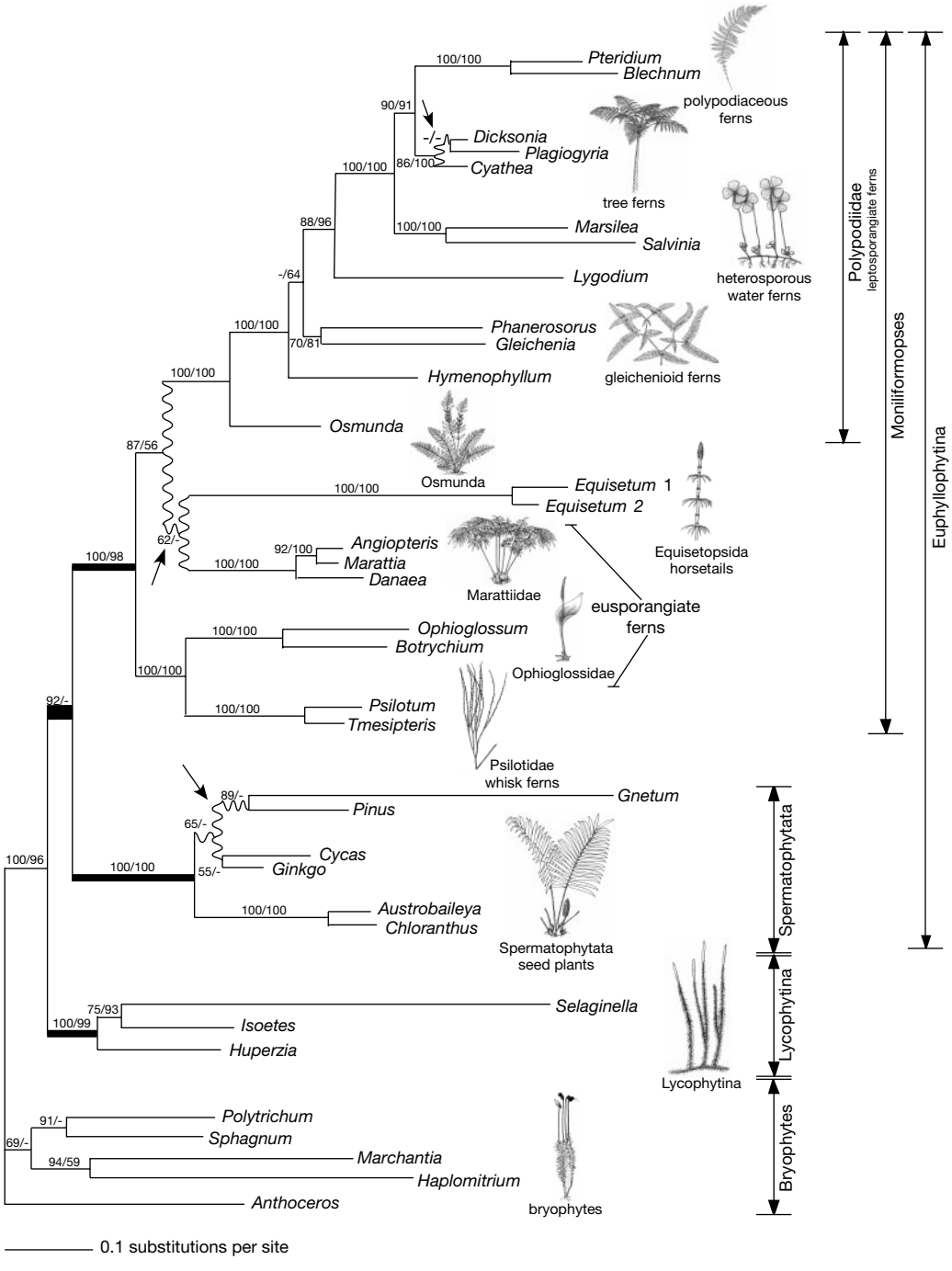


Figure 1 Phylogenetic relationships for all the main lineages of vascular plants inferred from maximum-likelihood (ML) analysis of the combined chloroplast *rbcL*, *atpB*, *rps4* and nuclear small-subunit rDNA data set. Numbers at nodes and before the slash are ML bootstrap values ≥50%; maximum parsimony (MP) bootstrap values ≥50% appear after the slash when these same nodes were supported in the MP unequally weighted analysis of the combined four-genes plus morphology data set (single MP tree = 14165.04 steps). A minus sign indicates a node had less than 50% bootstrap support in one or the other analysis. The topology is rooted by all bryophytes, hence relationships depicted among them are arbitrary. Branches leading to the three monophyletic clades of vascular plants (lycophytes, seed plants and horsetails+ferns) are drawn medium thick. The branch supporting the Euphyllophytina, with horsetails+ferns as sister group to seed plants, is the thickest. Wiggled lines (at straight arrows) indicate three areas of conflict between the ML and MP analyses. Branch lengths are proportional to number of substitutions per site (scale bar). Thumbnail sketches of plant representatives accompany major clades. Taxonomy follows ref. 1.

were specified as outgroups. We analysed the data sets using both maximum-parsimony (MP) and maximum-likelihood (ML) optimization criteria; bootstrap (BS) analyses were conducted to measure the stability of observed phylogenetic patterns.

Using ML on the combined four-gene data set we recovered one most likely tree (-ln likelihood = 36466.6365) for each of the 100 replicates (Fig. 1). We also observed an essentially identical topology using MP on the combined four-gene and morphology data set (three areas that differ are highlighted on Fig. 1). Regardless of the analytical approach (MP or ML), three major lineages emerged as monophyletic clades with exceptional support (100% BS). The first clade comprises the Lycophytina, increasingly recognized as a distinct group of vascular plants only distantly related to other extant pteridophytes and seed plants[1,16]. The second diverging lineage corresponds to seed plants. The third, novel, clade includes all non-seed-producing lineages of Euphyllophytina, including horsetails (Equisetopsida), leptosporangiate ferns (Polypodiidae), eusporangiate ferns (Marattiidae, Ophioglossidae) and whisk ferns (Psilotidae). Seed plants, ferns and horsetails are united as a monophyletic group, to the exclusion of lycopods, in both the ML (92% BS) and MP (<50% BS) analyses.

We observed one unambiguous length discrepancy in *rps4* that can be interpreted as a molecular 'signature' providing additional support for horsetail–fern monophyly. A portion of the *rps4* alignment is shown for base pairs 646–696 (Fig. 2), which includes 27 ambiguously aligned base pairs (658–684) flanked by unambiguously aligned regions. The ambiguously aligned region was excluded entirely from the ML analysis. In the MP analysis, the same region was recoded simply as a single absence/presence character for the observed length increase. This multi-residue length increase in horsetails and ferns is not as likely to be a random convergence as is a single point mutation and provides further evidence for this clade.

Within the horsetail–fern lineage, Psilotidae is most closely related to Ophioglossidae (100% BS). Although this association was only weakly suggested in recent single-gene analyses[11,19,20], the current evidence unambiguously invalidates the traditional morphological and palaeobotanical view that Psilotidae are relatively unaltered descendants of the psilotophytes, among the earliest vascular plant fossils[7,18]. Ophioglossidae and Psilotidae differ so radically in phenotype that this close relationship, implying a shared origin of phenotypic simplification, was never before explicitly considered. All other ferns and horsetails make up its sister clade (87% BS). The relationships of horsetails also have been controversial: sister to seed plants[7], sister to leptosporangiate (Polypodiidae) and eusporangiate (Ophioglossidae and Marattiidae) ferns[1], or as a basal grade euphyllophyte lineage[17]. Our analysis clearly (100% BS) places *Equisetum* within the non-lycophyte pteridophyte clade, although its exact relationships within this clade are not yet well resolved. In the ML analysis, *Equisetum* is sister to Marattiidae (62% BS), whereas in the MP analysis, it is sister to leptosporangiate ferns (<50% BS). This study also confirms a sister relationship between tree ferns and the more derived 'polypodiaceous' leptosporangiate ferns (90% BS), and places the heterosporous water ferns as sister to this clade (100% BS) (Fig. 1). Relationships among these groups were equivocal in earlier studies[17,20].

The only noteworthy disagreement between our ML and MP analyses is localized within seed-plant relationships, a subject of much current controversy[21,23,24]. Our ML analysis resolved gymnosperms as monophyletic (65% BS) and *Gnetum* as sister to *Pinus* (89% BS). Our MP analysis supports *Gnetum* as basal among seed plants (87% BS), and all other gymnosperms as monophyletic (67% BS) and sister to angiosperms.

In the ML analysis of the combined four-gene data set, there is persuasive support for the Euphyllophytina (92% BS), with a basal dichotomy indicating that the horsetail–fern clade (100% BS) is the closest relative to seed plants (100% BS). To the best of our knowledge, this relationship has been proposed only once previously[1], as a tentative hypothesis on the evidence of a single anatomical character (protoxylem distribution). This led to the provisional classification of the horsetail–fern clade as infradivision Moniliformopses (moniliforms); Psilotidae, however, was not included in that study[1]. Although this same deep dichotomy is also robustly resolved in the MP analysis of the combined four-genes plus morphology data set, the Euphyllophytina node is weakly supported (<50% BS). Exceptionally long branches in each of the three main clades (Fig. 1: *Selaginella*, *Gnetum* and *Equisetum*) and the greater sensitivity of MP over ML to long-branch attraction (statistical inconsistency) effects[21,25] probably explain why parsimony bootstrapping failed to recover this clade with high confidence. When these long-branch taxa were removed and the combined four-genes plus morphology data set was re-analysed with MP, this same basal Euphyllophytina node was highly supported (83% BS, results not shown). Each of our separate single-gene analyses, with the exception of *rps4*, did not resolve the horsetail–fern clade, and none was able to determine confidently the closest relatives to seed plants. Only our morphological data set, when analysed alone with MP, provided the same conclusions

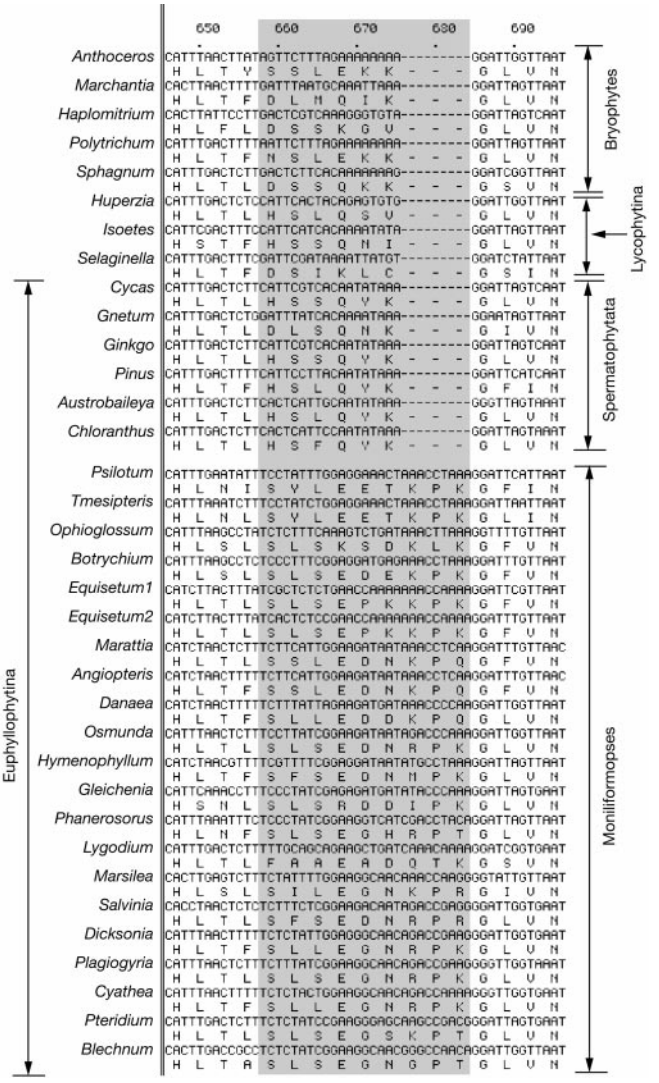


Figure 2 A portion of the chloroplast *rps4* alignment. An ambiguously aligned region (grey box) containing a 9-base-pair length difference distinguishes horsetails and ferns (bottom block) from bryophytes, lycophytes and seed plants (top block). Amino-acid translations are interleaved below each DNA sequence. Dashes indicate gaps.

regarding the Euphyllophytina as when the four genes were analysed simultaneously with ML. A study using mitochondrial small-subunit rDNA sequence data[10] with a smaller selection of taxa suggested support for this hypothesis; however, critical euphyllophyte taxa (Psilotidae and Marattiidae) were not included. A more recent study[26] that combines data from two genes (nuclear and mitochondrial small-subunit rDNA) strongly corroborates a horsetail–fern clade as sister to seed plants, despite a limited sampling of only seven euphyllophyte taxa from all pertinent lineages.

Our report of a basal dichotomy in the Euphyllophytina, a split that occurred in the early–mid Devonian (about 400 Myr ago), necessitates abandonment of the prevailing view that ferns and horsetails represent paraphyletic successive grades of increasing complexity in early vascular plant evolution, which eventually led to the more complex seed plants, and ultimately to angiosperms. A parallel deep reorganization of metazoan phylogeny has recently been proposed[27], with 'simple' bilaterian taxa (for example, platyhelminths and nemerteans) being displaced from the base of the metazoan tree to within the large lophotrochozoan clade.

A corollary of the demise of the paraphyletic interpretation of early vascular plant evolution is that it is now necessary to confront the many recurring models that derive morphological, developmental and physiological conditions in seed plants from an 'intermediate' or 'primitive' pteridophyte ancestor. We predict that this will require a significant revision in the interpretation of the underlying processes of vascular plant evolution. For example, a number of homeotic genes, such as the MADS-box genes that encode transcription factors critical for regulating physiological and developmental processes, especially flower development, have been well studied in angiosperms[28]. Clarifying the origin of these genes has been hampered by the few reports of homologues from non-seed plants, and therefore it is not known to what extent changes in number, regulation and function of these and other homeotic genes may have driven land plant evolution. The study of these genes from across a stable phylogenetic framework is critical. We note that all the main plant model organisms (for example, *Arabidopsis*, *Glycine*, *Lycopersicon*, *Oryza*, *Petunia* and *Zea*) are recently evolved angiosperms. Efforts to promote developmental and genomic research on model systems in the horsetail–fern clade (for example, *Ceratopteris*[29]), will probably lead to an improved understanding of fundamental aspects of vascular plant development and evolution[8]. □

Methods

Taxon sampling and morphological data set

We selected 35 taxa to sample explicitly at least two members of each major monophyletic group of land plants. The various groups were determined from recent broad-scale phylogenetic analyses[1,12,17,20], and we specified the bryophytes *Anthoceros*, *Haplomitrium*, *Marchantia*, *Polytrichum* and *Sphagnum* as outgroups. Our morphological data set comprised 136 parsimony-informative characters (H.S. *et al.*, manuscript in preparation), which we, for the most part, adopted or modifed from recent studies[1,7,12,17,20].

Gene sequencing

We amplified chloroplast *rbcL*, *atpB*, *rps4*, and nuclear small-subunit rDNA genes for all 35 taxa from total cellular DNA by polymerase chain reaction (PCR) and sequenced them using an ABI 377 automated DNA sequencer (PE Applied Biosystems). Details of taxon sampling, DNA isolation, PCR amplification, sequencing, sequence alignment, exclusion and recoding of ambiguously aligned regions, data set combinability testing, and phylogenetic analysis will be published elsewhere (K.M.P. *et al.*, manuscript in preparation). Most *atpB*, *rps4*, nuclear small-subunit rDNA, and some *rbcL* sequences were generated as part of this study. For voucher information, GenBank numbers and the aligned data matrices, see Supplementary Information and http://www.fmnh.org/research_collections/botany/botany_sites/ferns/publications.html; data matrices are also available in TreeBASE, accession number S543, at http://www.herbaria.harvard.edu/treebase/.

Phylogenetic analyses

We conducted heuristic MP (unequal weighting schemes, 1,000 random-addition replicates, tree bisection-reconnection (TBR) branch swapping) and ML (general time-reversible model, accommodating unequal nucleotide frequencies and different-probabilities for each of six substitution types, plus three heterogeneous rate categories across sites following a discrete approximation of the gamma distribution, 100 random-addition replicates) analyses using PAUP* version 4.0b2a[30]. The ML analysis was restricted to the combined four-gene data set because it is not possible to simultaneously implement two models of evolution, one for morphology and one for DNA sequence data, in any currently available computer programs. We further performed both parsimony bootstrap (unequal weighting schemes, 1,000 replicates, each with 10 random-addition replicates and TBR branch swapping) and likelihood bootstrap analyses (212 replicates, using identical parameters to those used to find the most likely tree).

Received 25 July; accepted 27 November 2000.

1. Kenrick, P. & Crane, P. R. *The Origin and Early Diversification of Land Plants: A Cladistic Study* (Smithsonian Institution Press, Washington DC, 1997).
2. Kenrick, P. & Crane, P. R. The origin and early evolution of plants on land. *Nature* **389,** 33–39 (1997).
3. Lupia, R., Lidgard, S. & Crane, P. R. Comparing palynological abundance and diversity: Implications for biotic replacement during the Cretaceous angiosperm radiation. *Paleobiology* **25,** 305–340 (1999).
4. Soltis, P. S., Soltis, D. E. & Chase, M. W. Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* **402,** 402–404 (1999).
5. Qiu, Y.-L. *et al.* The earliest angiosperms: Evidence from mitochondrial, plastid and nuclear genomes. *Nature* **402,** 404–407 (1999).
6. Barkman, T. J. *et al.* 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 (2000).
7. Rothwell, G. W. Fossils and ferns in the resolution of land plant phylogeny. *Bot. Rev.* **65,** 188–218 (1999).
8. Graham, L. E., Cook, M. E. & Busse, J. S. The origin of plants: Body plan changes contributing to a major evolutionary radiation. *Proc. Natl Acad. Sci. USA* **97,** 4535–4540 (2000).
9. Bremer, K. Summary of green plant phylogeny and classification. *Cladistics* **1,** 369–385 (1985).
10. Duff, R. J. & Nickrent, D. L. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *Am. J. Bot.* **86,** 372–386 (1999).
11. Hedderson, T. A., Chapman, R. & Cox, C. J. in *Bryology for the Twenty-first Century* (eds Bates, J. W., Ashton, N. W. & Duckett, J. G.) 65–77 (Maney Publishing and the British Bryological Society, Leeds, 1998).
12. Mishler, B. D. *et al.* Phylogenetic relationships of the "green algae" and "bryophytes". *Ann. Missouri Bot. Gard.* **81,** 451–483 (1994).
13. Qiu, Y.-L., Cho, Y., Cox, J. C. & Palmer, J. D. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* **394,** 671–674 (1998).
14. Hiesel, R., von Haeseler, A. & Brennicke, A. Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proc. Natl Acad. Sci. USA* **91,** 634–638 (1994).
15. Kranz, H. D. & Huss, V. A. R. Molecular evolution of pteridophytes and their relationships to seed plants: Evidence from complete 18S rRNA gene sequences. *Plant Syst. Evol.* **202,** 1–11 (1996).
16. Raubeson, L. A. & Jansen, R. K. Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science* **255,** 1697–1699 (1992).
17. Stevenson, D. W. & Loconte, H. in *Pteridology in Perspective* (eds Camus, J. M., Gibby, M. & Johns, R. J.) 435–467 (Royal Botanic Gardens, Kew, 1996).
18. Wagner, W. H. Jr Systematic implications of the Psilotaceae. *Brittonia* **29,** 54–63 (1977).
19. Manhart, J. R. Phylogenetic analysis of green plant *rbcL* sequences. *Mol. Phylogenet. Evol.* **3,** 114–127 (1994).
20. Pryer, K. M., Smith, A. R. & Skog, J. E. Phylogenetic relationships of extant ferns based on evidence from morphology and *rbcL* sequences. *Am. Fern J.* **85,** 205–282 (1995).
21. Sanderson, M. J., Wojciechowski, M. F., Hu, J. M., Sher Khan, T. & Brady, S. G. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* **17,** 782–797 (2000).
22. Hillis, D. M. Inferring complex phylogenies. *Nature* **383,** 130–131 (1996).
23. Bowe, L. M., Coat, G. & dePamphilis, C. W. 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 (2000).
24. Chaw, S.-M., Parkinson, C. L., Cheng, Y., Vincent, T. M. & Palmer, J. D. 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 (2000).
25. Lewis, P. O. in *Molecular Systematics of Plants II; DNA Sequencing* (eds Soltis, D. E., Soltis, P. S. & Doyle, J. J.) 132–163 (Kluwer Academic, Boston, 1998).
26. Renzaglia, K. S., Duff, R. J., Nickrent, D. L. & Garbary, D. J. Vegetative and reproductive innovations of early land plants: Implications for a unified phylogeny. *Phil. Trans. R. Soc. Lond. B* **355,** 769–793 (2000).
27. Adoutte, A. *et al.* The new animal phylogeny: Reliability and implications. *Proc. Natl Acad. Sci. USA* **97,** 4453–4456 (2000).
28. Theissen, G. *et al.* A short history of MADS-box genes in plants. *Plant Mol. Biol.* **42,** 115–149 (2000).
29. Hickok, L. G., Warne, T. R. & Fribourg, R. S. The biology of the fern *Ceratopteris* and its use as a model system. *Int. J. Plant Sci.* **156,** 332–345 (1995).
30. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other Methods)* (Sinauer, Sunderland, Massachusetts, 1999).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank R. Lupia, F. M. Lutzoni, B. D. Mishler, L. Newstrom-Lloyd and S. Zoller for critical comments on the manuscript; Z. Dabich, J. Bélisle, R. Lupia and D. Kieffer for assistance in rendering Fig. 1; F. M. Lutzoni and V. A. Funk for advice on phylogenetic analyses; I. Capesius, S. Boyles, B. Goffinet, M. Hasebe, M. Kato, M. Kessler, B. D. Mishler, R. Moran, J. Shaw, W. C. Taylor, Y.-L. Qiu, D. Wall, J. Wheeler, and greenhouse managers at Humboldt State University, University of California at Davis, University of California Botanical Garden at Berkeley, New York Botanical Garden for plant material; S. W.

Graham, P. S. Soltis and J. Therrien for sharing unpublished sequence data; and D. Ferguson, E. Grismer, J. Irwin and L. Sappelsa for general assistance in the initial stages of the project. This work was supported by grants from the NSF to K.M.P., A.R.S., P.G.W. and R. C., the Green Plant Phylogeny Research Coordination Group (USDA grant), and by the Pritzker Foundation Fund of The Field Museum.

Correspondence and requests for materials should be addressed to K.M.P. (e-mail: kpryer@fieldmuseum.org).

Export by red blood cells of nitric oxide bioactivity

John R. Pawloski, Douglas T. Hess & Jonathan S. Stamler

Howard Hughes Medical Institute and Department of Medicine, Box 2612, Duke University Medical Center, Durham, North Carolina 27710, USA

Previous studies support a model in which the physiological O_2 gradient is transduced by haemoglobin into the coordinate release from red blood cells of O_2 and nitric oxide (NO)-derived vasoactivity to optimize oxygen delivery in the arterial periphery[1,2]. But whereas both O_2 and NO diffuse into red blood cells, only O_2 can diffuse out[3–5]. Thus, for the dilation of blood vessels by red blood cells, there must be a mechanism to export NO-related vasoactivity, and current models of NO-mediated intercellular communication should be revised. Here we show that in human erythrocytes haemoglobin-derived *S*-nitrosothiol (SNO), generated from imported NO, is associated predominantly with the red blood cell membrane, and principally with cysteine residues in the haemoglobin-binding cytoplasmic domain of the anion exchanger AE1. Interaction with AE1 promotes the deoxygenated structure in SNO–haemoglobin, which subserves NO group transfer to the membrane. Furthermore, we show that vasodilatory activity is released from this membrane precinct by deoxygenation. Thus, the oxygen-regulated cellular mechanism that couples the synthesis and export of haemoglobin-derived NO bioactivity operates, at least in part, through formation of AE1–SNO at the membrane–cytosol interface.

As the first step in analysing the fate of haemoglobin (Hb)-derived NO *in situ*, we determined the disposition of NO groups transfered physiologically from the haems of Hb to β-chain Cys 93 in intact human erythrocytes[3,4]. Red blood cells (RBCs) held at less than 1% O_2 were exposed for 5 min to physiological amounts of NO (100 nM to 1 μM; NO:haem ratios 1:1,000 to 1:100) followed by reoxygenation (21% O_2), and membrane and cytosolic fractions were prepared. Fractions were solubilized with Triton X-100 (TX100), and the NO content of extracts was measured by photolysis/chemiluminescence[3,4]. At the lower NO:haem ratios, which produced intracellular NO concentrations matching those found *in vivo* (100–800 nM), recovery of NO was essentially complete, that is, none was lost to nitrate (Fig. 1a). In this model system, about 15–20% of NO incorporated by RBCs was present as SNO; the remainder was ascribed largely to iron nitrosyl haem (FeNO)[1,3,4,6]. Most iron nitrosyl Hb was recovered with the cytosolic fraction (Fig. 1b). In contrast, SNO was associated predominantly with the membrane fraction (Fig. 1c). These results confirm that, in intact RBCs[7] as with isolated reactants[3,4], Hb will efficiently capture and preserve NO, and form SNO, under physiological conditions. Unexpectedly, however, the formation of SNO is compartmentalized within the RBC.

Haemoglobin associates with the cytoplasmic face of the RBC membrane through specific protein–protein interactions[8–10]. To determine the disposition of Hb-derived membrane SNO, we examined the interaction of SNO–Hb[5,6] with inside-out vesicles (IOVs) prepared by everting RBC membrane ghosts[11]. IOVs incubated with SNO–Hb and washed at pH 8 to remove bound Hb incorporated about 450 pmol NO per mg of TX100-extracted IOV protein (Fig. 1d). All the incorporated NO was present in complex with thiol, that is, as SNO. It is important to note that SNO was not detected in extracts of IOVs exposed to NO in the absence of Hb (data not shown).

To rule out the possibility that apparent NO group transfer to IOVs was an artefact of residual membrane-bound SNO–Hb, we incubated IOVs with SNO–Hb immobilized on Sephadex beads. After centrifugal separation, washes at pH 7 and solubilization in TX100, extracts of IOVs were free of Hb as assessed by spectrophotometric detection of haem. SNO was present in those extracts at somewhat higher levels than in extracts derived from IOVs incubated with free SNO–Hb (suggesting a greater loss of

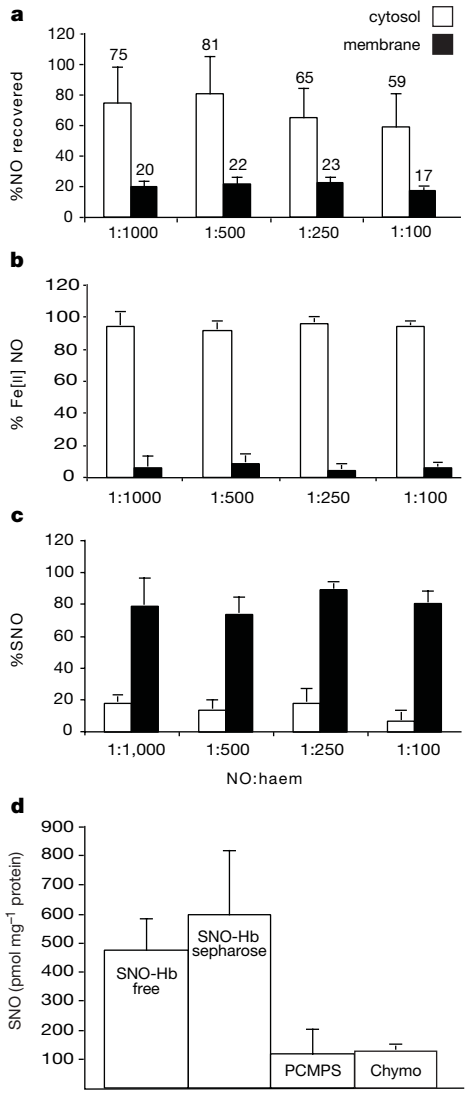


Figure 1 Haemoglobin-derived SNO is associated with cysteine thiols of RBC membrane proteins. **a**–**c**, Distribution in cytosolic and membrane fractions of NO groups after exposure of intact RBCs to NO. Recovery of NO is essentially complete at low, physiological NO:haem ratios (**a**), which yield 100–800 nM intracellular NO; FeNO is predominantly cytosolic (**b**), whereas SNO is largely membrane associated (**c**) ($P < 0.05$ for all pairwise comparisons). **d**, SNO content of IOVs exposed to free or Sepharose-bound SNO–Hb (50 nmol SNO–Hb per mg IOV protein). Transfer of NO groups to the membrane is greatly reduced ($P < 0.05$) after treatment of IOVs with the thiol-modifying reagent PCMPS and after mild digestion of IOVs with chymotrypsin (chymo). ($n = 3$–7 for **a**–**d**.)

J Mol Evol (2002) 55:265–271
DOI: 10.1007/s00239-002-2324-2

JOURNAL OF MOLECULAR EVOLUTION

Divergent Intron Conservation in the Mitochondrial *nad2* Gene: Signatures for the Three Bryophyte Classes (Mosses, Liverworts, and Hornworts) and the Lycophytes

Dagmar Pruchner,[1] Susanne Beckert,[1] Hermann Muhle,[2] Volker Knoop[1]

[1] Molekulare Botanik, Universität Ulm, D-89069 Ulm, Germany
[2] Systematische Botanik & Ökologie, Universität Ulm, D-89069 Ulm, Germany

Received: 27 August 2001 / Accepted: 8 February 2002

Abstract. The slow-evolving mitochondrial DNAs of plants have potentially conserved information on the phylogenetic branching of the earliest land plants. We present the *nad2* gene structures in hornworts and liverworts and in the presumptive earliest-branching vascular land plant clade, the Lycopodiopsida. Taken together with the recently obtained *nad2* data for mosses, each class of bryophytes presents another pattern of angiosperm-type introns conserved in *nad2*: intron nad2i1 in mosses; intron nad2i3 in liverworts; and both introns, nad2i3 and nad2i4, in hornworts. The lycopods *Isoetes* and *Lycopodium* show diverging intron conservation and feature a unique novel intron, termed nad2i3b. Hence, mitochondrial introns in general are positionally stable in the bryophytes and provide significant intraclade phylogenetic information, but the *nad2* introns, in particular, cannot resolve the interclade relationships of the bryophyte classes and to the tracheophytes. The necessity for RNA editing to reconstitute conserved codon entities in *nad2* is obvious for all clades except the marchantiid liverworts. Finally, we find that particularly small group II introns appear as a general feature of the *Isoetes* chondriome. Plant mitochondrial peculiarities such as RNA editing frequency, U-to-C type of RNA editing, and small group II introns appear to be genus-specific rather than gene-specific features.

Key words: Group II introns — Bryophytes — Lycopods — RNA editing

Introduction

Among the three genomes in the plant cell the mitochondrial DNA, the "chondriome," appears to offer some advantages for cladistic analyses of old branches in the land plant phylogeny. First, compared to nuclear and chloroplast DNA, the mitochondrial sequence drift is particularly slow in plants. Second, mitochondrially encoded genes do not present paralogy problems potentially associated with distinguishing members of nuclear gene families, such as the phytochrome (Alba et al. 2000) and PEP carboxylase gene families (Gehrig et al. 2001). Third, the occurrence of mitochondrial introns may help to define monophyletic groups. In contrast to those of the chloroplast genome, the mitochondrial introns are positionally divergent in the chondriomes of angiosperms (Unseld et al. 1997; Kubo et al. 2000) and the liverwort *Marchantia polymorpha* (Oda et al. 1992) but largely appear to be positionally stable, approximately at the level of classes defined by classical systematics (Beckert et al. 1999; Vangerow et al. 1999). When present or absent in more than one

Correspondence to: Dr. Volker Knoop, Botanisches Institut der Universität Bonn, Kirschallee 1, D-53115 Bonn, Germany; *email:* volker.knoop@uni-bonn.de

clearly defined monophyletic group, they could then help to identify higher-order clades to delineate the earliest branches of the land plant phylogeny. Indeed, on the basis of three mitochondrial introns present in all land plants other than the liverworts, it was suggested that the latter are the sister group to all other embryophytes (Qiu et al. 1998). However, the use of intron positions as molecular characters in plant phylogeny has also been questioned (Nickrent et al. 2000). In any case, for introns to be phylogenetically informative, it will be necessary both to gain further insights into the within-clade stability of plant mitochondrial introns (belonging mainly to the group II class of structured organellar introns) and to identify introns that are present or absent in more than one clade.

Extending initial analyses of the *cox3* gene (Malek et al. 1996), we explore the phylogenetic information potential of further mitochondrial protein coding sequences such as *nad5*, *nad2*, *nad4*, and *nad7* (Beckert et al. 1999; Pruchner et al. 2001). The mitochondrial *nad2* gene of plants is characterized by the only group II intron that is conserved in position when the mitochondrial DNA of the liverwort *Marchantia polymorpha* is compared with those of angiosperms. Somewhat surprisingly, however, investigation of the *nad2* gene structure in moss species of all extant orders had revealed that another one of the four known angiosperm introns is universally conserved as the only intervening sequence in *nad2* (Beckert et al. 2001).

With the work reported here we wished to gain information on the within-clade stability of mitochondrial introns by exploring the *nad2* gene in the other bryophyte classes, the liverworts and hornworts. We, furthermore, analyzed *nad2* in members of two orders of the lycopods, the likely sister group to all other vascular plants (euphyllophytes). The *nad2* exon–intron structure is reported for nine liverwort, two hornwort, and two lycopod species. We find perfect clade-specific conservation of intron positions within each bryophyte class but different intron occurrences in the lycopods. Introns in the genus *Isoetes* are particularly small and the *nad2* gene analysis furthermore suggests frequent RNA editing in this genus and the hornworts but none in the marchantiid liverworts.

Materials and methods

Plant Material and Nucleic Acid Preparation. Total nucleic acids were extracted from plant material (Table 1) with the Plant DNeasy kit (Qiagen) according to the protocol supplied by the manufacturer.

PCR Amplification, Cloning, and Sequencing. The PCR amplicon as depicted in Fig. 1 is flanked by terminal primers NAD2UP (5′-ggagttgtntttagtacctctaa-3′) and NAD2DO (5′-aacaayggagagggytatagmaag-3′), respectively. Overlapping PCRs were occasionally performed combining NAD2UP with internal primer NAD2D2 (5′-cctaagatca tagaagcact gc-3′) and NAD2DO with internal primer NAD2U2 (5′-caatctagtggtatctttat ggg-3′) or NAD2A2 (5′-tacgtgtttytatttatrgttyct-3′). PCR amplification assays contained 1 µl of template DNA (approximately 10 ng–1 µg), 10 µl of 10× PCR buffer [100 m*M* Tris/HCl, pH 8.85, 250 m*M* KCl, 50 m*M* $(NH_4)_2SO_4$, 20 m*M* $MgSO_4$], a 250 m*M* concentration of each dNTP, 0.25 µg of each primer, 2.5 U of DNA polymerase, and double-distilled water to 100 µl. Different commercially available thermostable DNA polymerases were used, e.g., a mixture (90:1) of *Taq* DNA Pol (Gibco BRL) and *Pwo* DNA Pol (Boehringer Mannheim). A typical amplification assay included an initial denaturation (5 min, 94°C) followed by 35 cycles of 1 min of denaturation at 94°C, 1 min of annealing at 50–55°C, and 2 min 30 s of synthesis at 72°C and a final step of synthesis for 6 min at 72°C. PCR fragments were blunt-end ligated into pBlueskript II SK+ (Stratagene). Positive clones were sequenced with a Thermosequenase kit (Amersham) using Cy5 fluorescence-labeled oligonucleotides and run on an Alf Express sequencer (Pharmacia). Sequencing primers were universal and reverse primers of the polylinker sequence and additional primers matching internal sequences of the cloned gene fragments.

Table 1. List of species under investigation together with sequence accession numbers and lengths in base pairs of the *nad2* amplicon in Fig. 1

Species	Accession No.	Length (bp)	Intron(s) present
Liverworts			
Marchantiidae			
Bucegia romanica	AJ344059	2664	i3
Corsinia coriandrina	AJ344060	2672	i3
Lunularia cruciata	AJ344061	2667	i3
Ricciocarpos natans	AJ344062	2655	i3
Sphaerocarpos donnelli	AJ409124	2677	i3
Jungermanniidae			
Fossombronia pusilla	AJ409122	2632	i3
Haplomitrium mnioides	AJ409123	2583	i3
Pellia epiphylla	AJ344063	2249	i3
Plagiochila asplenioides	AJ344064	2686	i3
Hornworts			
Anthoceros agrestis	AJ409117	4836	i3, i4
Phaeoceros laevis	AJ409118	5054	i3, i4
Lycopods			
Isoetes durieui	AJ409119	2890	i1, i2, i3, i3b
Lycopodium annotinum	AJ409120	3456	i1, i3b

Results

The plant mitochondrial *nad2* gene structure is shown in Fig. 1. Introns nad2i1, nad2i2, nad2i3, and nad2i4 are conserved among dicot and monocot angiosperms (Binder et al. 1992; Morawala-Patell 1998). Intron nad2i2 is trans-disrupted in flowering plants, precluding PCR amplification of the entire gene but *cis*-arranged versions of nad2i2 have been identified in pteridophytes (Malek et al. 1997). Amplicons of *nad2* were cloned and sequenced for mosses covering all extant orders and this revealed the presence of nad2i1 as the only intervening sequence in this bryophyte class, without exception (Beckert et al. 2001). In

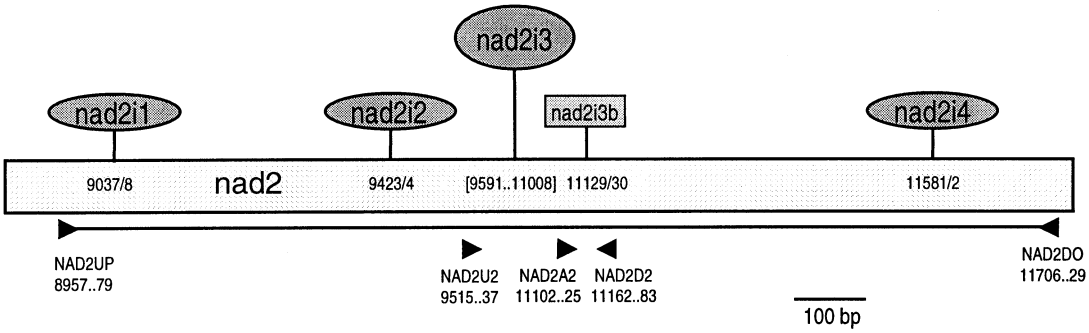


Fig. 1. The *nad2* gene structure. All intron insertions are group II. Introns nad2i1, nad2i2, nad2i3, and nad2i4 are conserved among flowering plants, where intron nad2i2 is split *in trans*. Intron nad2i3b is a lycopod-specific intron discovered in the course of this study. The *line* denotes the extension of a PCR amplicon, encompassing a 417-amino acid coding region plus the respective intervening sequences, which was analyzed for the species listed in Table 1. Internal primers were used to subdivide the amplicon for PCR in some cases. Numbering is according to the *Marchantia polmorpha* chondriome database entry M68929.

contrast, nad2i3 is the only intervening sequence in the *Marchantia polymorpha nad2* gene and, additionally, the only intron in the liverwort chondriome that is conserved in position in angiosperm chondriomes (Unseld et al. 1997; Kubo et al. 2000).

Intron nad2i3 Is Conserved and the Only Intervening Sequence in Liverworts

To determine whether the occurrence of nad2i3 is unique to *Marchantia polymorpha*, a complex thalloid (marchantiid) liverwort, we amplified, cloned, and sequenced *nad2* gene regions from diverse liverwort species. Both the closely related species *Bucegia romanica*, *Corsinia coriandrina*, and *Lunularia cruciata* and the more distant marchantiid liverworts *Ricciocarpos natans* and *Sphaerocarpos donnelli* reveal nad2i3 as the single intron in *nad2*. Members of the sister group of jungermanniid liverworts (*Pellia epiphylla*, *Fossombronia pusilla*, *Haplomitrium mnioides*, and *Plagiochila asplenioides*) were likewise included in our analyses. As in the marchantiid subgroup, nad2i3 was observed as the single *nad2* intron. Two further observations in the liverwort *nad2* data are noteworthy as discussed below.

Lack of RNA Editing and Extreme Sequence Conservation in the Marchantiidae

The necessity for RNA editing, the common phenomenon of cytidine-to-uridine nucleotide exchanges to reconstitute conserved codons in plant mitochondrial RNAs, is observed for the jungermanniid but not for the marchantiid *nad2* sequences (Fig. 2). Indeed, with one event in each 10th codon (42 of 417), the RNA editing frequency appears to be particularly high for *Haplomitrium*, a genus for which we report here the first mitochondrial protein coding sequence. For comparison, only 13 codon changes can be predicted for *Pellia* in the corresponding *nad2* gene region (Fig. 2).

There is an extraordinarily high degree of sequence conservation among the marchantiid compared to the jungermanniid liverworts. In fact, the amino acid translations are identical for *Bucegia*, *Lunularia*, and *Sphaerocarpos nad2* sequences. On the contrary, divergence among jungermanniid species is significantly higher and best exemplified by *Pellia*, which features a 400-bp deletion in intron nad2i3, a one-codon deletion in the upstream exon, and a two-codon insertion in the downstream exon (Fig. 2). Consequently, a phylogenetic tree derived from the *nad2* nucleotide alignment of the liverwort sequences features long branches for the jungermanniid but very short terminal branches for the marchantiid species (Fig. 3). Comparison of bootstrap support values derived from the complete and the exon-only alignment immediately demonstrates how much phylogenetic information for within-clade resolution is contained in the nad2i3 intron sequence: Only two branches receive bootstrap support exceeding 70 from the coding region, whereas inclusion of the intron sequences results in five branches with (increased) bootstrap support (Fig. 3). Unfortunately, however, due to the incongruent intron occurrence among the bryophyte groups (see below), the *nad2* intron sequences cannot help to resolve the interclass relationships. Both observations, the particularly high mitochondrial DNA sequence conservation and the lack of RNA editing among the marchantiid liverworts, are in accord with results previously obtained for the *nad5* gene (Beckert et al. 1999; Steinhauser et al. 1999).

Two Introns Conserved in the Hornworts

Amplification of *nad2* from the hornworts *Anthoceros agrestis* and *Phaeoceros laevis* required overlapping

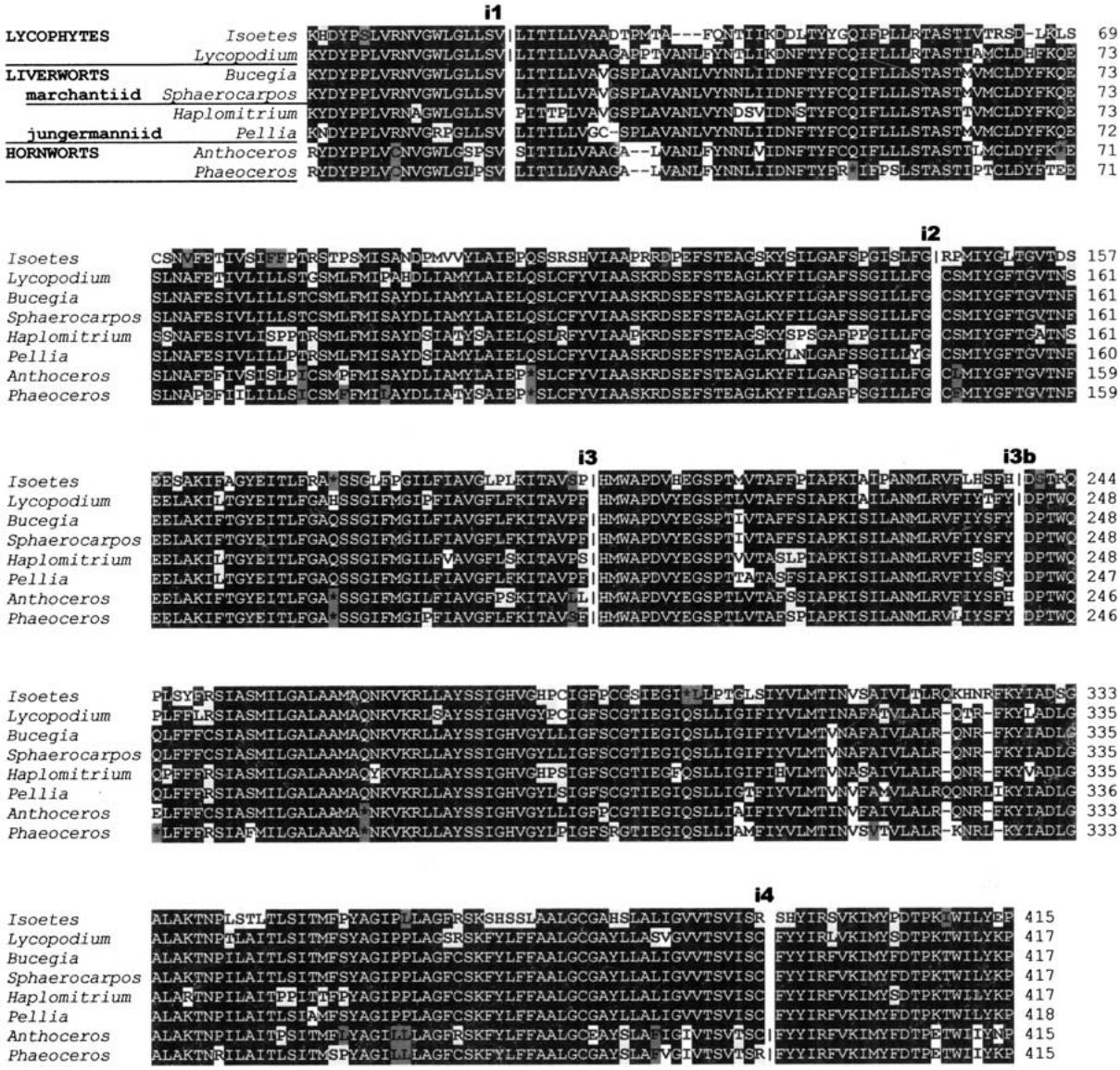


Fig. 2. Conceptual amino acid translations from DNA nucleotide sequences obtained in this study. Taxa were selected to demonstrate divergence in sequence and RNA editing patterns. Missing codons are indicated by *dashes*; *asterisks* denote stop codons in the DNA sequence. Intron insertion sites are shown as gaps in the alignment and the presence of introns in the respective taxa is indicated by *vertical lines*. *Background shading* highlights amino acid identities that are likely subject to change by C-to-U or U-to-C RNA editing.

PCRs due to the simultaneous presence of introns nad2i3 and nad2i4. Neither intron nad2i1 nor intron nad2i2 is present. Thus, each of the three bryophyte classes shows a unique pattern of *nad2* intron conservation: nad2i1 in mosses, nad2i3 in liverworts, and both nad2i3 and nad2i4 in hornworts.

Predictions of RNA editing, again, are in line with observations made before for the *nad5* gene (Steinhauser et al. 1999). Hornworts show frequent RNA editing in both directions of pyrimidine exchange, with the "reverse" (U-to-C) exchange contributing to the reconstitution of conserved codons and to the elimination of stop codons from the reading frame. 24 C to U editings, accompanied by 15 reverse editings, can be predicted for the genomic *Phaeoceros laevis nad2* sequence (Fig. 2). RNA editing is somewhat less frequent in the sister hornwort *Anthoceros agrestis* (18 C-to-U and 12 U-to-C exchanges).

A Novel Intron in the Lycopods

The *nad2* gene was also analyzed for the two lycopods *Lycopodium annotinum* (Lycopodiales) and *Isoetes durieui* (Isoetales). The presence of nad2i2 as a conventional cis-arranged intron had been shown earlier for *Isoetes lacustris* (Malek and Knoop 1998). The complete *nad2* amplicon of *Isoetes durieui* confirms

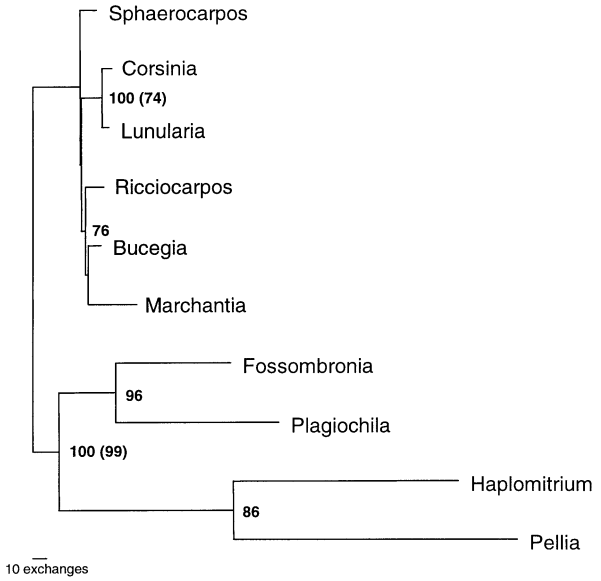


Fig. 3. Most parsimonious phylogenetic tree of 891 steps derived from the complete nucieotide alignment of liverwort *nad2* sequences. Bootstrap percentage (1000 replicates) support (only where exceeding 70) is indicated at the respective branchings obtained from the complete alignment including intron nad2i3 and, in parentheses, from the exon regions only.

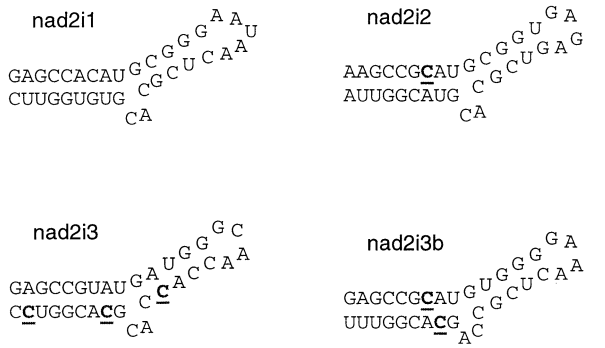


Fig. 4. Signature domain V sequences of the four *Isoetes durieui* group II introns in *nad2*. Potential sites for C-to-U RNA editing to improve RNA helix stability are *underlined*.

this and, additionally, reveals the presence of introns nad2i1 and nad2i3 (but not i4) and a novel, additional intron, termed nad2i3b. Only introns nad2i1 and nad2i3b are conserved in *Lycopodium*. Although the Lycopodiopsida are a divergent class, their monophyly is currently not doubted (Pryer et al. 2001), thus the absence of nad2i2 and nad2i3 in *Lycopodium* clearly suggests independent losses or gains of introns since divergence from the common vascular plant ancestor. The quest for cis-arranged ancestors of the trans-splicing angiosperm mitochondrial introns in *nad* genes had identified particularly small introns in the genus *Isoetes* (Malek and Knoop 1998). This trend is clearly continued with the now identified other homologues of angiosperm *nad2* introns and the novel nad2i3b intron. The sizes are 517, 384, 358, and 386 nucleotides (nt), respectively, for nad2i1, nad2i2, nad2i3, and nad2i3b from *Isoetes durieui*. Intron sizes are 1520 nt for nad2i1 and 684 nt for nad2i3b in *Lycopodium annotinum*, for comparison. The typical domain V secondary structure signatures (Bonen and Vogel 2001) of the *Isoetes* introns unequivocally classify them as belonging to the group II class (Fig. 4). The extraordinary frequency of RNA editing in *Isoetes* (see below) and base mismatches in domains V potentially corrected by pyrimidine exchanges (Fig. 4) make it difficult to derive complete intron secondary structures, which are potentially influenced by RNA editing.

With 17 presumptive C-to-U RNA edits in the *nad2* amplicon sequence, *Lycopodium* shows a moderate frequency of RNA editing. *Isoetes*, on the other hand, shows the necessity for a particularly high frequency of RNA editing to reconstitute a total of 76 codons, including 11 reverse U-to-C exchanges. In one instance, five genomically encoded amino acids in a row, SHSSL, can be expected to be converted into the evolutionary conserved FYLFF motif, *Isoetes* thus confirms its status as the record plant genus of RNA editing (Malek et al. 1996).

Discussion

Although the comparison of the liverwort *Marchantia* (Oda et al. 1992) and angiosperm (Unseld et al. 1997; Kubo et al. 2000) chondriomes may at first glance give another impression, plant mitochondrial group II introns are fairly stable in position. Even for more complexly arranged genes, such as the trans-splicing genes *nad1*, *nad2*, and *nad5* of flowering plants, introns are conserved in comparison to monocot and dicot species (for references see Malek and Knoop 1998). Our earlier studies had shown that mitochondrial introns likewise appear to be stable in the phylogenetically older divisions pteridophytes and bryophytes, at least on the level of classes defined by classical systematics: intron nad5i2 is observed without exception in 15 liverworts and more than 50 mosses (Beckert et al. 1999); intron nad5i2b is observed in 30 genera of ferns and fern allies, with the single exception of the eusporangiate fern *Ophioglossum* (Vangerow et al. 1999); and intron nad2i1 is, without exception, present in 42 mosses (Beckert et al. 2001). Similar observations have been made for the *nad4* and *nad7* genes, however, so far based only on smaller taxon samplings (Pruchner et al. 2001).

The obvious exceptions, however, clearly advocating that independent intron gain or loss can occur, should not go unnoticed. Variations in mitochondrial intron presence in angiosperms are documented for the two downstream introns in *nad4* (Gass et al. 1992), the two introns in the *cox2* gene (Unseld et al. 1997), an intron in *nad7* (Pla et al.

1995), an intron in *nad1* (Bakker et al. 2000), and the intron in the *rps3* gene (Kubo et al., 2000). The lack of intron nad5i2b in the eusporangiate fern *Ophioglossum*, which otherwise is a universally conserved pteridophyte signature intron, is another case in point. The lack of intron nad2i3 in *Lycopodium* reported here is a similar example of intron loss. Likewise, accepting a monophyly of vascular plants and lycopods, the lack of nad2i2 in *Lycopodium* is a further example either of intron loss or of separate intron gain in *Isoetes*. Interestingly these observations do not come as the first peculiarities in the genus *Lycopodium:* the liverwort-type intron cox3i1, which so far appears to be absent in members of all other land plant groups, is present in *Lycopodium* (Malek et al. 1996).

This report was aimed at investigating the intron occurrence pattern in *nad2* of basal land plants, the only mitochondrial gene in which the liverwort *Marchantia* and angiosperms share an intron position. While the intron sequences clearly provide potential for within-clade fine resolution, as shown here for nad2i3 among liverworts, the ultimate goal should be the identification of higher-order clades, e.g., by linking one of the bryophyte classes to the vascular plants. Obviously, the introns of the *nad2* gene can be of little help: each bryophyte class presents different *nad2* introns conserved, none of which establishes a clear link of one group with the tracheophytes.

The occurrence of intron nad2i3 in both liverwort subgroups contradicts a linkage of jungermanniid liverworts alone to the mosses (Capesius and Bopp 1997) and, rather, supports classical views on liverwort monophyly and is in accord with other molecular data (Hedderson et al. 1996; Lewis et al. 1997; Beckert et al. 1999). The *nad2* sequences reported here also clearly confirm two trends that were observed earlier: RNA editing is clearly more genus-specific than gene-specific. The high frequency of RNA editing in *Isoetes* identified previously (Malek et al. 1996) and the apparent absence of editing in marchantiid, but not jungermanniid, liverworts (Steinhauser et al. 1999) are corroborated. Likewise, a particularly high ratio of U-to-C editing in hornworts, as identified in *nad5*, can be predicted from the novel *nad2* sequences.

Acknowledgments. We gratefully acknowledge the gift of *Haplomitrium* DNA from Prof. Dr. Yin-Long Qiu, Amherst, MA. Research in the laboratory of V.K. is supported by grants from the Deutsche Forschungsgemeinschaft, the Land Baden-Württemberg, and the Universität Ulm.

References

Alba R, Kelmenson PM, Cordonnier-Pratt MM, Pratt LH (2000) The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. Mol Biol Evol 17:362–373

Bakker FT, Culham A, Pankhurst CE, Gibby M (2000) Mitochondrial and chloroplast DNA-based phylogeny of Pelargonium (Geraniaceae). Am J Bot 87:727–734

Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst Evol 218:179–192

Beckert S, Muhle H, Pruchner D, Knoop V (2001) The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: A comparative analysis in mosses. Mol Phylogenet Evol 18:117–126

Binder S, Marchfelder A, Brennicke A, Wissinger B (1992) RNA editing in trans-splicing intron sequences of *nad2* mRNAs in Oenothera mitochondria. J Biol Chem 267:7615–7623

Bonen L, Vogel J (2001) The ins and outs of group II introns. Trends Genet 17:322–331

Capesius I, Bopp M (1997) New classification of liverworts based on molecular and morphological data. Plant Syst Evol 207: 87–97

Gass DA, Makaroff CA, Palmer JD (1992) Variable intron content of the NADH dehydrogenase subunit 4 gene of plant mitochondria. Curr Genet 21:423–430

Gehrig H, Heute V, Kluge M (2001) New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. Mol Phylogenet Evol 20:262–274

Hedderson TA, Chapman RL, Rootes WL (1996) Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Plant Syst Evol 200:213–224

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res 28:2571–2576

Lewis LA, Mishler BD, Vilgalys R (1997) Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene rbcL. Mol Phylogenet Evol 7:377–393

Malek O, Knoop V (1998) Trans-splicing group II introns in plant mitochondria: The complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA 4:1599–1609

Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15:1403–1411

Malek O, Brennicke A, Knoop V (1997) Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. Proc Natl Acad Sci USA 94:553–558

Morawala-Patell V, Gualberto JM, Lamattina L, Grienenberger JM, Bonnard G (1998) Cis- and trans-splicing and RNA editing are required for the expression of nad2 in wheat mitochondria. Mol Gen Genet 258:503–511

Nickrent DL, Parkinson CL, Palmer JD, Duff RJ (2000) Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol Biol Evol 17:1885–1895

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

Pla M, Mathieu C, De Paepe R, Chetrit P, Vedel F (1995) Deletion of the last two exons of the mitochondrial *nad7* gene results in lack of the NAD7 polypeptide in a Nicotiana sylvestris CMS mutant. Mol Gen Genet 248:79–88

Pruchner D, Nassal B, Schindler M, Knoop V (2001) Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol Gen Genet 266:608–613

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409:618–622

Qiu YL, Cho Y, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V (1999) Plant mitochondrial RNA editing — extreme in hornworts and dividing the liverworts? J Mol Evol 48:303–312

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of Arabidopsis thaliana contains 57 genes in 366,924 nucleotides. Nature Genet 15:57–61

Vangerow S, Teerkorn T, Knoop V (1999) Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. Plant Biol 1:235–243

Mol Genet Genomics (2001) 266: 608–613
DOI 10.1007/s004380100577

ORIGINAL PAPER

D. Pruchner · B. Nassal · M. Schindler · V. Knoop

Mosses share mitochondrial group II introns with flowering plants, not with liverworts

Received: 18 June 2001 / Accepted: 8 August 2001 / Published online: 13 September 2001

Abstract Extant bryophytes are regarded as the closest living relatives of the first land plants, but relationships among the bryophyte classes (mosses, liverworts and hornworts) and between them and other embryophytes have remained unclear. We have recently found that plant mitochondrial genes with positionally stable introns are well suited for addressing questions of plant phylogeny at a deep level. To explore further data sets we have chosen to investigate the mitochondrial genes *nad4* and *nad7*, which are particularly rich in intron sequences. Surprisingly, we find that in these genes mosses share three group II introns with flowering plants, but none with the liverwort *Marchantia polymorpha* or other liverworts investigated here. In mitochondria of *Marchantia*, *nad7* is a pseudogene containing stop codons, but *nad7* appears as a functional mitochondrial gene in mosses, including the isolated genus *Takakia*. We observe the necessity for strikingly frequent C-to-U RNA editing to reconstitute conserved codons in *Takakia* when compared to other mosses. The findings underline the great evolutionary distances among the bryophytes as the presumptive oldest division of land plants. A scenario involving differential intron gains from fungal sources in what are perhaps the two earliest diverging land plant lineages, liverworts and other embryophytes, is discussed. With their positionally stable introns, *nad4* and *nad7* represent novel marker genes that may permit a detailed phylogenetic resolution of early clades of land plants.

Keywords Plant mitochondrial DNA · Group II introns · RNA editing · Mosses · Liverworts

Communicated by R. Hagemann

D. Pruchner · B. Nassal · M. Schindler · V. Knoop (✉)
Molekulare Botanik, Universität Ulm,
Albert-Einstein-Allee 11, 89081 Ulm, Germany
E-mail: volker.knoop@biologie.uni-ulm.de
Tel.: +49-731-5022615
Fax: +49-731-5022626

Introduction

The use of mitochondrial DNA for evolutionary studies has appeared as a relative latecomer in the molecular phylogenetic analysis of land plants. Short sequences of the mitochondrial *cox3* gene, encoding subunit 3 of cytochrome oxidase, had shown promise for phylogenetic analyses of ancient land plant groups (Hiesel et al. 1994). Further studies were extended to a second region of the same gene (Malek et al. 1996), to the *nad5* gene, encoding subunit 5 of NADH dehydrogenase (Beckert et al. 1999; Vangerow et al. 1999) and the gene for the small mitochondrial rRNA (Duff and Nickrent 1999).

Owing to their large sizes and structural complexity only three complete chondriome sequences of plants have been obtained so far. The complete mitochondrial DNA sequences from the liverwort *Marchantia polymorpha* (Oda et al. 1992) and the model flowering plant *Arabidopsis thaliana* (Unseld et al. 1997) have recently been joined by the complete chondriome sequence of sugar beet (Kubo et al. 2000). While the overall mitochondrial gene complement in the different plants is quite similar, the positions of the group I and group II introns, which are characteristic for organellar genomes in plants and fungi, differ strikingly between the liverwort and the seed plants. Overall, 25 group II introns are present in the mitochondrial DNA of *Marchantia*, and 23 are found in *Arabidopsis*. However, none of these introns is shared by *Marchantia* and flowering plants, with the exception of a single positionally conserved intron in the *nad2* gene.

Studies of *nad5* have shown that a group I intron is conserved between (and is exclusive to) mosses and liverworts, so far without exception (Beckert et al. 1999). Similarly, a group II intron at a different position in *nad5* has qualified as a signature for the pteridophytes, with only rare exceptions (Vangerow et al. 1999). Hence, the occurrence of at least some plant mitochondrial introns appears to correlate with the definitions of systematic classes or divisions, reflecting a high level of positional stability. Among angiosperms, most introns

are positionally conserved between monocot and dicot species. Variation in the incidence of mitochondrial introns between the completely sequenced dicot chondriomes of *Arabidopsis* and *Beta* is restricted to those in the *cox2* gene and the *rps3* gene encoding ribosomal protein 3 of the small subunit. Further variability in the occurrence of mitochondrial introns among angiosperms has been described for the two terminal introns in *nad4* (Gass et al. 1992), and the third intron of *nad7*, which is lacking in *Nicotiana sylvestris* (Pla et al. 1995).

It appears worthwhile to bridge the large evolutionary distance between liverworts and flowering plants, which are separated by more than 400 Myr of evolution, by analysing further loci in the mitochondrial DNA, given that it may contain significant phylogenetic information. We focus on genes that contain introns in both the liverwort and seed plant chondriomes, i.e those encoding other subunits of the NADH dehydrogenase, complex I of the mitochondrial respiratory chain: *nad2*, *nad4*, *nad5* and *nad7*. The analysis of *nad2* has already revealed that mosses do not share the only intron conserved between the liverwort *Marchantia* and angiosperms, but rather carry another group II intron conserved in angiosperms (Beckert et al. 2001).

Here we report new results for the mitochondrial genes *nad4* and *nad7*, and find that mosses share three group II introns with angiosperms, but none with liverworts. The group I intron in *nad5* shared between mosses and liverworts (Beckert et al. 1999) now presents itself as the exception rather than the rule, or it may indicate that group I and group II introns used different modes of invading new loci during the evolution of early land plants.

Materials and methods

Plant material and nucleic acid preparation

Total nucleic acids were extracted from green plant material (Table 1) with the Plant DNeasy kit (Qiagen) according to the protocol supplied by the manufacturer. DNA from *Takakia lepidozioides* and *Bazzania trilobata* was kindly provided by Dr. Y.-L. Qiu (Amherst, Mass.).

Table 1 New sequence accessions produced in the course of this study

Gene	Species	Accession No.	Voucher specimen[a]
nad4	*Bazzania trilobata*	AJ310800	Qiu97085
	Corsinia coriandrina	AJ310801	GC
	Riccia fluitans	AJ310802	GC
	Lunularia cruciata	AJ310803	GC
	Takakia lepidozioides	AJ409092	Qiu97126
	Timmia bavarica	AJ409093	ULM:Muhle161197-1
nad7	*Takakia lepidozioides*	AJ309978	Qiu97126
	Ulota crispa	AJ309977	ULM:Muhle200497-6
	Leucobryum glaucum	AJ309976	ULM:Muhle281097-6
	Dichodontium pellucidum	AJ309975	ULM:Muhle191097-6

[a]GC indicates that the species is maintained in our greenhouse. DNAs of *Bazzania trilobata* and *Takakia lepidozioides* were kind gifts of Dr. Y.-L. Qiu (Amherst, Mass.)

PCR amplification, cloning and sequencing

PCR amplicons depicted in Fig. 1 are flanked by the terminal primers N4UP (5′-ACAGCCAAATTTCARTTTGTGGAA-3′) and N4DO (5′-TYAATSAAATTTTCCATGTTGCAC-3′), and N7UP (5′-GGNCCNCARCAYCCNGCNGC-3′) and N7DO (5′-TCTATCTACCTCTCCAAACACAAT-3′), respectively. PCR amplification assays contained 1 µl of template DNA (approximately 10 ng–1 µg), 10 µl of 10×PCR buffer (100 mM TRIS-HCl pH 8.85, 250 mM KCl, 50 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$), 250 mM of each dNTP, 0.25 µg of each primer, 2.5 U of DNA polymerase and double-distilled water to 100 µl. Different commercially available thermostable DNA polymerases were used, e.g. a mixture (90:1) of *Taq* DNA Pol (Gibco BRL) and *Pwo* DNA Pol (Boehringer Mannheim). A typical amplification assay included an initial denaturation step for 5 min at 94°C, followed by 35 cycles of 1 min at 94°C, 1 min at 50°–55°C and 2 min 30 s at 72°C, with a final extension for 6 min at 72°C. PCR fragments were blunt-end ligated into pBluescript II SK + (Stratagene). Positive clones were sequenced with a Thermosequenase kit (Amersham) using Cy5-labelled fluorescent oligonucleotides and an ALF Express sequencer (Pharmacia). Sequencing primers were universal and reverse primers for the polylinker sequence and additional primers matching internal sequences of the cloned gene fragments.

Results

The exon-intron structures of the plant mitochondrial genes *nad4* and *nad7* are shown in Fig. 1. All introns under consideration belong to the group II class. We here propose a nomenclature for the plant mitochondrial introns, numbering them by order of appearance in the reading frames, taking into account known introns both in the chondriome of the liverwort *M. polymorpha* (Oda et al. 1992) and in angiosperms like *A. thaliana* (Unseld et al. 1997). As an example, *nad7* in *Marchantia* carries only introns nad7i3 and nad7i6, while seed plants have nad7i1, nad7i2, nad7i4 and nad7i5 (Bonen et al. 1994). Primers were designed for PCR amplicons containing all known intron insertion sites of each gene (Fig. 1).

The *nad4* gene

In *nad4*, intron nad4i1 is a characteristic angiosperm intron, while nad4i3 and nad4i4 may or may not be present in any given species of flowering plant; for example, both are lacking in *Lactuca sativa* (Geiss et al. 1994). A single intron, nad4i2, is present at a different location in *nad4* of the liverwort *M. polymorpha* (Fig. 1A). To first determine whether the group II intron nad4i2, as the sole intervening sequence in this gene in *M. polymorpha*, is unique to this liverwort, we investigated *Corsinia coriandrina* and *Lunularia cruciata* (two other marchantiid liverworts), *Riccia fluitans* (a more distantly related species) and *Bazzania trilobata* as a member of the neighbour group of jungermanniid liv-

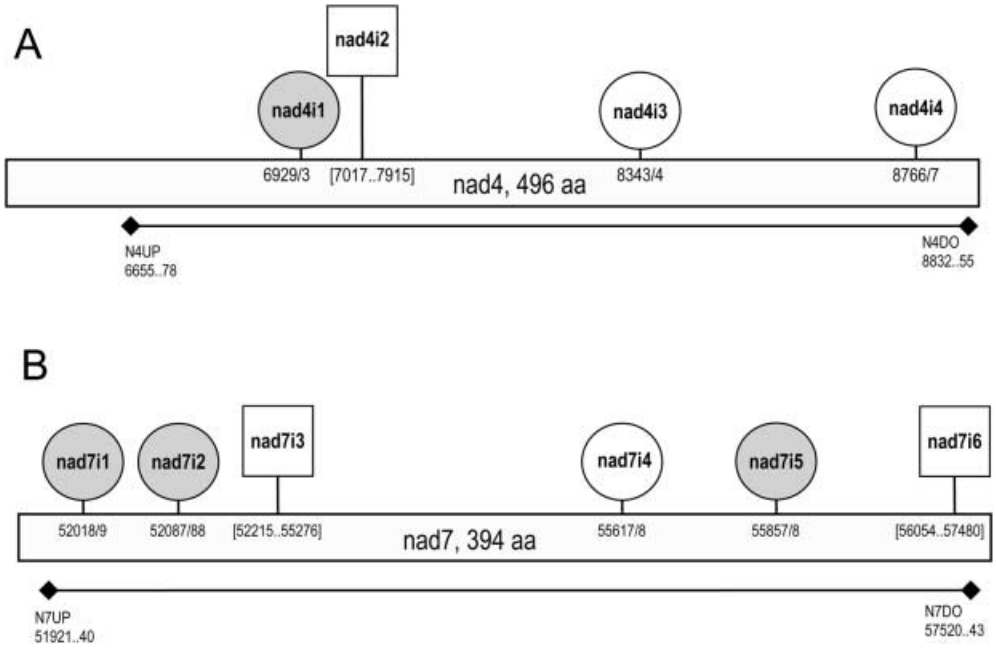


Fig. 1A, B Group II intron positions in the plant mitochondrial genes *nad4* (**A**) and *nad7* (**B**). The *shaded circles* designate introns that are conserved in angiosperms, *open circles* indicate those that are present in only some of the angiosperms examined. The *squares* indicate introns identified in the liverwort *Marchantia polymorpha*. Numbering refers to the coordinates of the complete chondriome sequence from *M. polymorpha* (Accession No. M68929). The extents of the PCR amplicons, encompassing all intron insertion sites, and the primer binding sites are indicated

erworts. The novel DNA sequences were deposited in the database under Accession Nos. AJ310800–AJ310803, respectively. In all cases, we found nad4i2 as the single, "liverwort-type", intron present (Figs. 1A and 2).

The situation is completely different in the mosses. The complete sequence of *nad4* from *Timmia bavarica* is deposited in the database under Accession No. AJ409093. Only nad4i1, the first of the typical angiosperm introns, is present. To investigate whether the singular presence of nad4i1 is unique to *Timmia* we cloned the *nad4* amplicon of *Takakia lepidozioides* and sequenced it (available under Accession No. AJ409092). Evidence has accumulated that the isolated bryophyte genus *Takakia* should be placed in the mosses – probably among the basalmost branching genera (see Newton et al. 2000 and Beckert et al. 2001 and references therein). As in *Timmia,* only nad4i1 is present in the *Takakia nad4* amplicon. Hence, the angiosperm intron nad4i1 is also the typical signature intron in mosses, while the two "optional" introns nad4i3 and nad4i4, which vary in occurrence among the angiosperms, and the liverwort intron nad4i2, are absent. The *nad4* gene clearly qualifies for phylogenetic analyses as a further marker gene, in which clade-specific, rapidly diverging intron sequences can provide a phylogenetic signal that will allow for finer discrimination among liverwort and moss orders, respectively.

The modification of mRNAs by RNA editing to reconstitute conserved codon identities is a hallmark of plant mitochondrial gene expression, except in the subgroup of marchantiid liverworts, where conserved codons are already genomically encoded (Steinhauser et al. 1999). The *nad4* gene is no exception to this rule: 13 RNA editing events to reconstitute evolutionarily conserved codons can be postulated for the jungermanniid liverwort *Bazzania* (Fig. 2). The need for a strikingly high frequency of RNA editing can be deduced from the genomic *nad4* sequence from *Takakia*. With 35 postulated RNA editing events in 418 codons this amounts to altering every 12th codon, a value that significantly exceeds the previous maximum – the 1 in 48 codons that need to be changed in *nad5* – based on a survey of a large number of mosses (Steinhauser et al. 1999).

The *nad7* gene

As in the *nad4* gene, liverwort and flowering plants share none of the intron positions mapped in *nad7* (Fig. 1B). The mitochondrial *M. polymorpha nad7* reading frame contains stop codons, making it a transcribed but unspliced pseudogene (Takemura et al. 1995), while the functional *nad7* gene copy is encoded in the nucleus (Kobayashi et al. 1997).

Investigation of the *nad7* gene in mosses confirms the observations made on the *nad4* gene. Amplification, cloning and sequencing revealed angiosperm-type introns nad7i1 and nad7i2, but none of the others, in the *T. lepidozioides* sequence (deposited in the database under Accession No. AJ309978). To investigate whether the basally diverging genus *Takakia* might in some way be unique in its intron structure, we chose to examine other, derived moss genera. The complete *nad7* amplicon sequences were also determined for *Ulota crispa*

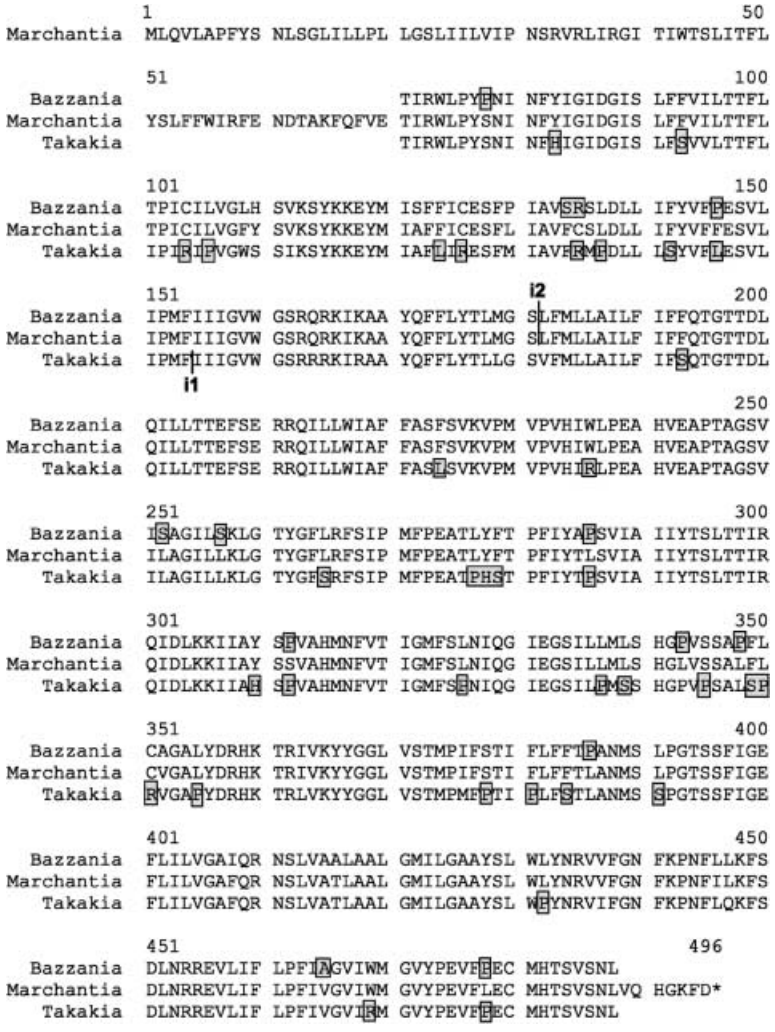


Fig. 2 Protein sequences deduced from the *nad4* gene coding regions of *Bazzania trilobata* and *Takakia lepidozioides* are aligned with the deduced sequence of the NAD4 polypeptide from *Marchantia polymorpha*. Intron insertion sites for nad4i2 in the liverworts and for nad4i1, which is conserved in the mosses, are indicated. The sequences from the jungermanniid liverwort *Bazzania* and the moss *Takakia* both show the need for RNA editing to reconstitute codon identities conserved in other species such as *Marchantia*, as highlighted by the *shaded boxes*

(Orthotrichaceae), *Leucobryum glaucum* (Leucobryaceae) and *Dichodontium pellucidum* (Dicranaceae) and deposited in the database under Accession numbers AJ309975–AJ309977, respectively. In all cases, cangiosperm-type introns nad7i1 and nad7i2 were found to interrupt the coding sequences, but none of the other four introns so far known were present (Figs. 1B and 3). In the meantime, the *nad7* gene sequence from the moss *Physcomitrella patens* (Funariales) has also been deposited in the database (Hashimoto and Sato 2001) and this sequence likewise shows nad7i1 and nad7i2 as the only interrupting sequences. With two neighbouring introns, each approximately 1000 bp long, the *nad7* gene may well present a novel, phylogenetically informative gene, most notably for those interested in further resolving the phylogeny of the mosses.

The NAD7 protein is particularly well conserved in evolution. Comparison of the polypeptide sequences deduced from the *nad7* genes of different species confirms the necessity for particularly frequent RNA editing in *Takakia* (Fig. 3). Whereas only a single RNA editing position has to be postulated for *Leucobryum* and none for *Ulota* or *Dichodontium*, 14 codons can be expected to be reconstituted by C to U RNA editing in the *Takakia* sequence.

Discussion

While it appears reasonable to consider the tracheophytes (vascular plants) a monophyletic group, evidence is accumulating that bryophytes are paraphyletic. The presence of introns at particular positions in the mitochondrial genomes of all land plant groups other than liverworts, as revealed by Southern hybridization, followed by the cloning and sequencing of selected examples, had suggested liverworts as the earliest diverging land plants (Qiu et al. 1998). The hybridization approach allows for screening of a large number of representative plant taxa, but apparent lack of hybridization may lead to misinterpretations; e.g. in the case of the extremely short group II introns in the lycopod genus *Isoetes* (Malek and Knoop 1998) or high degrees of sequence divergence. Conversely, positive hybridization signals may be due to cross-hybridization of homologous members of intron families at different locations

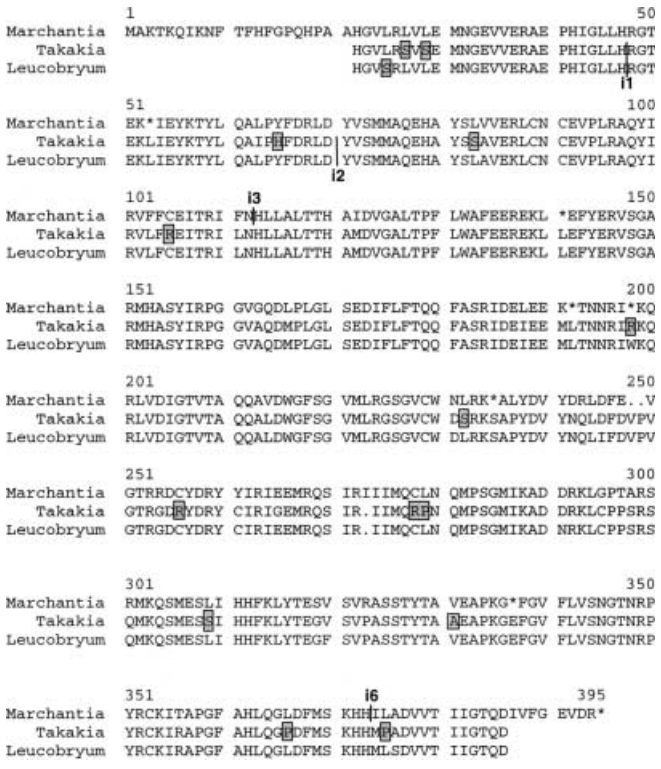


Fig. 3 Protein sequences deduced from the coding regions of the *nad7* genes of *Takakia lepidozioides* and *Leucobryum glaucum* are aligned with the polypeptide sequence deduced from the mitochondrial *Marchantia nad7* pseudogene. The *asterisks* mark stop codons in the *Marchantia* sequence. Candidate RNA editing positions are designated as in Fig. 2. Intron insertion sites for nad7i3 and i6 in the liverwort and for nad7i1 and nad7i2, present in all mosses, are indicated

(Ohyama et al. 1993). To give a striking example, a group II intron in the *nad5* gene of the lycopod *Huperzia* has been duplicated at a position only 850 bp away (Vangerow et al. 1999).

We have used PCR amplification followed by cloning and sequencing, which allows for exact identification of intron positions and has the additional benefit of identifying other sequence characteristics, such as the surprisingly high frequency of RNA editing required in the basal moss *Takakia*. Primer design for PCR amplification can, however, become a matter of guesswork for species like the hornworts, where the pyrimidine composition of a genomic sequence can not be deduced due to very frequent C-U RNA editing in both directions (Steinhauser et al. 1999). Indeed, all our attempts to obtain the *nad4* and *nad7* amplicons (Fig. 1) from hornworts have so far been unsuccessful.

The strikingly divergent occurrence of mitochondrial introns in liverworts and other embryophytes suggests a differential invasion of mitochondrial sequences by introns in the two most ancient land plant lineages. Alternative explanations would require the massive loss of some, followed by a gain of other, mitochondrial introns in at least one of the two lineages. This scenario appears less likely, most notably in light of the absence of homologous introns in completely sequenced green algal chondriomes (Wolff et al. 1994; Turmel et al. 1999; Kück et al. 2000; Nedelcu et al. 2000). This observation apparently extends to algae of the Charophyceae, where we have so far found that introns are generally absent in *nad* genes (unpublished observations).

Which organisms already present on land during Ordovician times could possibly qualify as a biological source for the differential intron gains in the earliest land plants? Intimate physical contact of plants with other organisms soon after the move from water to land may have arisen as symbiotic interactions with fungi. It has been demonstrated that bryophytes are capable of mycorrhiza-like symbiosis (Read et al. 2000; Schüssler 2000) and that mycorrhiza-like fungi (similar to extant Glomales, the zygomycete order that comprises typical arbuscular mycorrhizal fungi) were present during Ordovician times (Redecker et al. 2000). This idea of horizontal intron transfer from fungi by group II intron transposition into conserved sites (Bonen and Vogel 2001) is supported by two circumstantial observations: Firstly, several group I introns occupy identical positions in the mitochondrial genomes of the liverwort *Marchantia* and fungi (Ohta et al. 1993) and a group II intron insertion position shared between fungi and plants has likewise been reported (Schäfer et al. 1998). Secondly, if other organisms, such as eubacteria, were the donor sources of these introns, chloroplast genomes should have experienced a similar differential gain of introns, but the general congruence of intron distribution in the plastomes of liverwort and other embryophytes makes this possibility seem unlikely.

Studies on the mitochondrial DNAs of symbiotic fungi would ideally address whether introns homologous to those identified here are differentially present, e.g. in different species of the Glomales. However, this approach is precluded by our current inability to cultivate the mycorrhizal fungi in isolation from their symbiotic partners.

Acknowledgements Financial support to V.K. was provided by Grant Kn411/2-3 from the Deutsche Forschungsgemeinschaft (DFG) and by an Anfangsförderung from the University of Ulm. We are grateful to Prof. Dr. Y.-L. Qiu (Amherst, Mass.) for collaboration and discussion, and for providing DNAs from *Bazzania trilobata* and *Takakia lepidozioides*.

References

Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst Evol 218:179–192

Beckert S, Muhle H, Pruchner D, Knoop V (2001) The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol Phylogenet Evol 18:117–126

Bonen L, Vogel J (2001) The ins and outs of group II introns. Trends Genet 17:322–331

Bonen L, Williams K, Bird S, Wood C (1994) The NADH dehydrogenase subunit 7 gene is interrupted by four group II introns in the wheat mitochondrial genome. Mol Gen Genet 244:81–89

Duff RJ, Nickrent DL (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am J Bot 86:372–379

Gass DA, Makaroff CA, Palmer JD (1992) Variable intron content of the NADH dehydrogenase subunit 4 gene of plant mitochondria. Curr Genet 21:423–430

Geiss KT, Abbas GM, Makaroff CA (1994) Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. Mol Gen Genet 243:97–105

Hashimoto K, Sato N (2001) Characterization of the mitochondrial *nad7* gene in *Physcomitrella patens*: similarity with angiosperm *nad7* genes. Plant Sci 160:807–815

Hiesel R, von Haeseler A, Brennicke A (1994) Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. Proc Natl Acad Sci USA 91:634–638

Kobayashi Y, Knoop V, Fukuzawa H, Brennicke A, Ohyama K (1997) Interorganellar gene transfer in bryophytes: the functional *nad7* gene is nuclear encoded in *Marchantia polymorpha*. Mol Gen Genet 256:589–592

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res 28:2571–2576

Kück U, Jekosch K, Holzamer P (2000) DNA sequence analysis of the complete mitochondrial genome of the green alga *Scenedesmus obliquus*: evidence for UAG being a leucine and UCA being a non-sense codon. Gene 253:13–18

Malek O, Knoop V (1998) Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA 4:1599–1609

Malek O, Lättig K, Hiesel R, Brennicke A, Knoop V (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15:1403–1411

Nedelcu AM, Lee RW, Lemieux C, Gray MW, Burger G (2000) The complete mitochondrial DNA sequence of *Scenedesmus obliquus* reflects an intermediate stage in the evolution of the green algal mitochondrial genome. Genome Res 10:819–831

Newton A, Cox CJ, Duckett JG, Wheeler JA, Goffinet B, Hedderson TAJ, Mishler BD (2000) Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. Bryologist 103:187–211

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

Ohta E, Oda K, Yamato K, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K, Michel F (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res 21:1297–1305

Ohyama K, Oda K, Ohta E, Takemura M (1993) Gene organization and evolution of introns of a liverwort, *Marchantia polymorpha*, mitochondrial genome. In: Brennicke A, Kück U (eds) Plant mitochondria. VCH, Weinheim pp 115–129

Pla M, Mathieu C, De Paepe R, Chetrit P, Vedel F (1995) Deletion of the last two exons of the mitochondrial *nad7* gene results in lack of the NAD7 polypeptide in a *Nicotiana sylvestris* CMS mutant. Mol Gen Genet 248:79–88

Qiu YL, Cho Y, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674

Read DJ, Ducket JG, Francis R, Ligron R, Russell A (2000) Symbiotic fungal associations in 'lower' land plants. Philos Trans R Soc Lond B Biol Sci 355:815–830

Redecker D, Kodner R, Graham LE (2000) Glomalean fungi from the Ordovician. Science 289:1920–1921

Schäfer B, Kaulich K, Wolf K (1998) Mosaic structure of the *cox2* gene in the petite negative yeast *Schizosaccharomyces pombe*: a group II intron is inserted at the same location as the otherwise unrelated group II introns in the mitochondria of higher plants. Gene 214:101–112

Schüssler A (2000) *Glomus claroideum* forms an arbuscular mycorrhiza-like symbiosis with the hornwort *Anthoceros punctatus*. Mycorrhiza 10:15–21

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V (1999) Plant mitochondrial RNA editing – extreme in hornworts and dividing the liverworts? J Mol Evol 48:303–312

Takemura M, Nozato N, Oda K, Kobayashi Y, Fukuzawa H, Ohyama K (1995) Active transcription of the pseudogene for subunit 7 of the NADH dehydrogenase in *Marchantia polymorpha* mitochondria. Mol Gen Genet 247:565–570

Turmel M, Lemieux C, Burger G, Lang BF, Otis C, Plante I, Gray MW (1999) The complete mitochondrial DNA sequences of *Nephroselmis olivacea* and *Pedinomonas minor*. Two radically different evolutionary patterns within green algae. Plant Cell 11:1717–1730

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. Nat Genet 15:57–61

Vangerow S, Teerkorn T, Knoop V (1999) Phylogenetic information in the mitochondrial *nad5* gene of pteridophytes: RNA editing and intron sequences. Plant Biol 1:235–243

Wolff G, Plante I, Lang BF, Kück U, Burger G (1994) Complete sequence of the mitochondrial DNA of the chlorophyte alga *Prototheca wickerhamii*. Gene content and genome organization. J Mol Biol 237:75–78

A decade of progress in plant molecular phylogenetics

Vincent Savolainen and Mark W. Chase

Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond TW9 3DS, UK

Over the past decade, botanists have produced several thousand phylogenetic analyses based on molecular data, with particular emphasis on sequencing *rbcL*, the plastid gene encoding the large subunit of Rubisco (ribulose bisphosphate carboxylase). Because phylogenetic trees retrieved from the three plant genomes (plastid, nuclear and mitochondrial) have been highly congruent, the 'Angiosperm Phylogeny Group' has used these DNA-based phylogenetic trees to reclassify all families of flowering plants. However, in addition to taxonomy, these major phylogenetic efforts have also helped to define strategies to reconstruct the 'tree of life', and have revealed the size of the ancestral plant genome, uncovered potential candidates for the ancestral flower, identified molecular living fossils, and linked the rate of neutral substitutions with species diversity. With an increased interest in DNA sequencing programmes in non-model organisms, the next decade will hopefully see these phylogenetic findings integrated into new genetic syntheses, from genomes to taxa.

Phylogenetics – the study of the evolutionary history and relationships of biological taxa – has been revolutionized by DNA sequence data. In the early 1980s, plant physiologists characterized a plastid gene, *rbcL*, encoding the large subunit of ribulose bisphosphate carboxylase (Rubisco), the most abundant enzyme on earth [1]. Because *rbcL* is a key photosynthetic gene, Zurawski and his colleagues were interested in comparing *rbcL* gene sequences from as many taxa as possible, thereby possibly increasing our knowledge of photosynthetic pathways and improving attempts to manipulate photosynthesis, for example, in crops. To achieve this goal, they distributed *rbcL* primers free of charge at a time when all phases of sequencing were costly. As a by-product of this initiative, plant systematists collected *rbcL* DNA sequences for a broad sampling of seed plants (499 species), resulting in one of the first collaborative large-scale phylogenetic analyses, just a decade ago [2]. Since then, several thousand molecular-based phylogenetic analyses have been published for all types of organisms [3]. Rather than reviewing phylogenetic methodologies or the details of ten years in plant phylogenetics, we will concentrate here on some of the major and recent advances, from assembling the general 'tree of life' to the evolution of genes, genomes and the origin of biodiversity. Our discussion emphasizes results from *rbcL* analyses, but we have also included several other relevant publications covering our understanding of plant taxonomy, evolution and methodology.

Towards assembling the 'tree of life': size matters

Over time genome sequences evolve – undergoing mutation and fixation in populations. The extent of the substitution differences in homologous sequences often reflects the evolutionary distinctiveness of organisms with respect to each other; thus, this information can be used to reconstruct molecular phylogenetic trees. Although for prokaryotes a complete-genome approach might be necessary due to the large numbers of horizontal transfers that occurred during early stages of life on Earth [4,5], large-scale multigene-based phylogenetic analyses are practical for many eukaryotes and particularly for plants. In addition, nucleotide changes are roughly clocklike, although the speed at which the clock ticks is usually different between lineages; nevertheless, providing that one can correct for this RATE HETEROGENEITY (see Glossary), nucleotide divergence can also be used as a surrogate for time (Box 1).

Several METHODS TO BUILD PHYLOGENETIC TREES have been developed, but building trees remains a hypercomplex

Box 1. Calibrating molecular phylogenetic trees with fossils

To calibrate molecular phylogenetic trees with fossils (or any biogeographical and tectonic event of known age), several options are available. The simplest way is to look at nucleotide divergence between pairs of extant taxa in a tree, which are the products of molecular change (divergence) that has arisen since these taxa evolved from a common ancestor; this date can be inferred from the fossil record and provides a rate of change that can be used to calculate in turn the ages of all the other nodes of the tree. This procedure, however, assumes a constant molecular clock throughout the tree (i.e. equal rates in each branch from the root), unless it is subdivided into subtrees in which different fossils can be used to provide several estimates for the rates of substitutions in the respective parts. An alternative is to correct first for rate heterogeneity across the tree. For example, it can be appropriate to assume that despite the fact that rates can differ among lineages, they are autocorrelated along lineages from parent to daughter branches, that is, rates are at least partly heritable. Several algorithms can then model the evolution of differential rates along these lineages and can apply some corrections, thereby transforming molecular branch lengths into relative time. Then one fossil calibration point can be used to transform relative time into absolute ages as described above. With more complex algorithms, it is also possible to use simultaneously several fossils for calibrations and to fix minimum, maximum or intervals of ages for some nodes in the tree [55,71].

Corresponding author: Vincent Savolainen (v.savolainen@kew.org).

http://tigs.trends.com 0168-9525/$ - see front matter doi:10.1016/j.tig.2003.10.003

Glossary

Angiosperms (flowering plants): plants with flowers and ovules enclosed in an ovary.
Bootstrap: a computational technique in which a percentage of the original data are deleted and randomly resampled to recreate a matrix of the original size, which is used to evaluate support for the groups on the phylogenetic trees.
Convergence: nucleotide changes resulting in identity driven by chance or selection for similar function but not due to common history.
Eudicots: the group of flowering plants with triaperturate pollen.
Functional constraints: the effect of natural selection on DNA to conserve function at the protein level.
Homoplastic changes: any nucleotide changes resulting in identity at a given nucleotide position not due to common history, namely, convergence, parallelism and reversion.
Jackknife: a computational technique in which data points of the original matrix are randomly deleted and the analysis rerun to evaluate clarity of patterns in phylogenetic trees and expressed as percentages of such replicates in which a group of taxa occurs.
Maximum likelihood methods: a computational technique in which phylogenetic trees are built according to models of nucleotide evolution (i.e. incorporating different frequencies of change and nucleotide composition as well as probabilities of change).
Methods to build phylogenetic trees: any of three main categories of computational techniques commonly used to build DNA-based phylogenetic trees: (i) distance methods, in which pairwise genetic distances are used to build trees; (ii) maximum parsimony methods, in which overall nucleotide changes are minimized in the tree-building process (usually with equal probabilities for all changes, but which can also incorporate uneven probabilities much as in maximum likelihood methods); and (iii) maximum likelihood methods (see above). Recently, Bayesian methods have been used in phylogeny inference [71].
Monocots: flowering plants with uniaperturate pollen and parallel leaf venation, comprising palms, grasses, orchids, irises, etc. (Figure 5).
Dicots: a term that referred to the group of plants with two cotyledons (the two specialized leaves that provide nutrients to the growing plantlet) but that phylogenetic studies have shown to be an artificial taxon (Figure 4).
Nonparametric rate smoothing: a computational technique in which rate heterogeneity in DNA sequences is corrected across lineages to make branch lengths proportional to time only.
Rate heterogeneity: the presence of significant difference in the amount of nucleotide changes between lineages or at sites within a DNA region.
Reversion: any nucleotide change that results in restoration of the initial nucleotide (e.g. adenine changing to thymine, and then returning to the original base: that is, thymine back to adenine).
Root: the first split (node) of a phylogenetic tree.
Taxon (pl. taxa): any level in the classification of organisms, for example, species, genus, family and order (Figure 3).
Triaperturate pollen: pollen with three openings, through one of which the pollen tube germinates and transfers the sperm to the ovule.
Ultrametric: referring to a phylogenetic tree in which branch lengths (genetic divergence) are proportional to time only and within which rate heterogeneity among lineages, if any, has been corrected.
Uniaperturate pollen: pollen with a single opening, through which the pollen tube germinates and transfers the sperm to the ovule.
Vascular plants: all plants with tissues specialized for the transport of water, nutrients and minerals.

mathematical problem because the number of solutions (possible trees) that ideally should be evaluated increases exponentially with TAXON number. For example, when using just over 100 taxa, the number of possible trees exceeds the number of particles in the universe. This problem has been brought sharply into focus as a result of large-scale sequencing projects focused on ANGIOSPERM phylogeny and more generally towards assembling the 'tree of life'.

DNA sequences can have rates of substitution at some sites that are so high that the information could be lost due to multiple changes, REVERSIONS and saturation (Figure 1); as a result, sequences from distantly related taxa might be spuriously attracted to each other by one of several forms of 'long branch attraction' [6]. Simulations in four-taxon

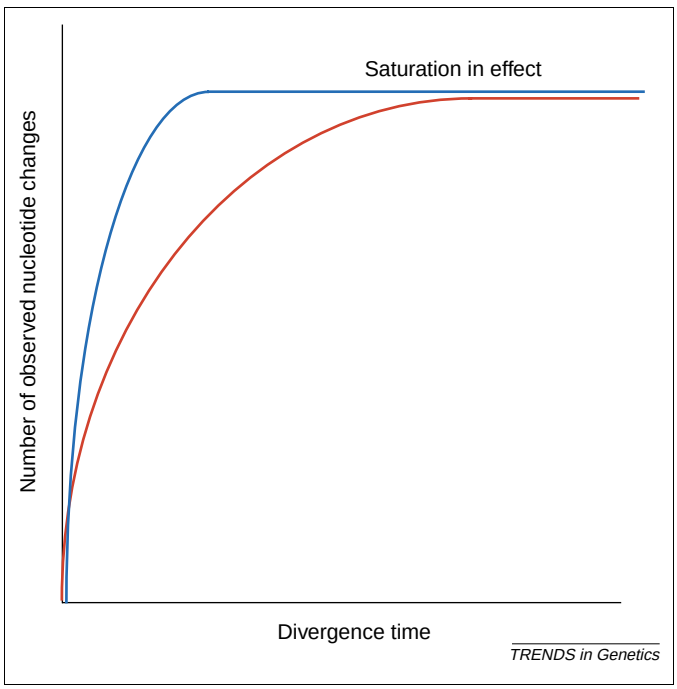


Figure 1. Saturation: when observed nucleotide changes are plotted against time, a plateau is reached when divergence time is great enough for reversions to mask the true number of substitutions; note that DNA sequences with higher substitution rates (blue) reach saturation more quickly than sequences with lower rates (red).

cases have shown that most tree-building methods would be inconsistent (i.e. converge on a wrong solution) in case of saturation unless methods are used that will correct for unobserved changes [7]. MAXIMUM LIKELIHOOD METHODS have been popular in this respect, but these methods are immense consumers of computer time. If a simple four-taxon case cannot be solved readily, even after sequencing several thousand nucleotides, how can a reliable phylogenetic tree with several thousand taxa be built? An answer came from a study of ribosomal DNA (rDNA) sequences in angiosperms: bigger is better – that is, more taxa are at least as beneficial as longer gene sequences. To evaluate how phylogenetic reconstruction is improved when adding more taxa or nucleotides, Hillis performed a simulation experiment with a large tree [8]. He used a 223-plant taxon, nonclocklike tree based on 18S rDNA as a model tree, and simulated on this tree the evolution of DNA sequences of various lengths. Then, using these artificial sequences in phylogenetic analyses, he asked how many variable nucleotide positions are necessary to recover the model tree. Unexpectedly, he found that as few as 5000 variable base positions (i.e. when all sites changed at least once in the tree) were sufficient to recover in every detail the model tree correctly using maximum parsimony [8]. When Hillis then simulated sequence evolution at rates up to ten times faster, the tree was correctly inferred with even fewer nucleotides [9]. Because the four-taxon studies showed that most phylogenetic methods would fail to recover correct trees if nucleotide change does not follow a constant clock [10], these results at first surprised the phylogenetic community. However, it was quickly realized that larger trees reveal more nucleotide changes overall (there are more branches on which nucleotides can change),

and this makes it easier to recover an accurate phylogenetic signal (Figure 2). In particular, although the number of inferred HOMOPLASTIC CHANGES (i.e. base positions that share nucleotides due to CONVERGENCE and reversion) in larger datasets is higher, and these were at first regarded as 'noise', they can be locally informative: they can reflect relationships in restricted parts of the tree in spite of being globally uninformative (Figure 2). For example, although third-codon positions in protein-coding genes accumulate more changes than first or second positions as a result of the redundancy of the genetic code, they are often more informative than other codon positions in plant datasets (sometimes also including bacteria) [11–14], an observation that contrasts with findings in animals [15] (but see Ref. [16]). These findings have been of immense general importance – outside of angiosperm studies – and they have reorientated strategies used to reconstruct the 'tree of life'.

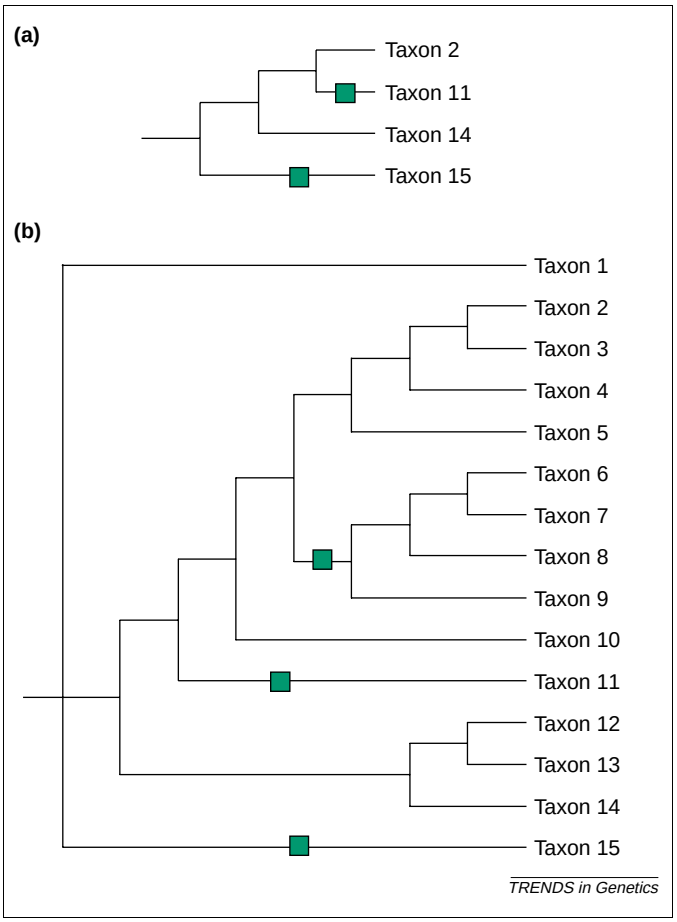


Figure 2. Illustration of how recovery of phylogenetic signal is easier in larger trees. The green square represents a hypothetical substitution (e.g. adenine to cytosine) at one particular site. **(a)** In the smaller tree, this change occurred independently twice, that is, along the branches leading to taxa 11 and 15, and therefore this substitution is a convergence and does not tell anything about evolutionary relationships. **(b)** When additional taxa are added to give the larger tree, this substitution is found on another branch, namely, the ancestor leading to the group of taxa 6 to 9. In this latter case, this change reflects common ancestry despite the fact that overall it is homoplastic. Bigger trees simply have more chance to exhibit such substitutions: that is, substitutions that are 'locally' informative of shared evolutionary history.

Three genomes, one tree

In plant phylogenetics, perhaps one of the most reliable measures of confidence in our trees is the congruence between the information retrieved from the three genomes (plastid, nuclear and mitochondrial). Phylogenetic analyses of angiosperms comprising up to a few thousand taxa (up to 2538 [11]) have been performed with the plastid *rbcL* gene [2,17], *rbcL* combined with plastid *atpB* [13], plastid inverted repeat [18], and various combinations of nuclear rDNA [19–21], nuclear phytochrome genes [22] and mitochondrial *matR* and *atp1* genes [23,24]. Data matrices containing many additional genes have recently been analysed for flowering plants [25]. Although there are sometimes differences of pattern in the published trees, strongly incongruent groupings have rarely been found [26], that is, no contradictory groups depicted in analyses of different genomes received support as measured by the BOOTSTRAP or JACKKNIFE. At the taxonomic level (Figure 3) of families and above, all three genomes appear to be tracking the same evolutionary history. The main factors that could alter detection of historical patterns would be differing structural and FUNCTIONAL CONSTRAINTS (i.e. those caused by strong selection), but combining several genes would be expected to average out such forces operating on individual genes.

There have been reports that DNA sequences from the three genomes evolve at different rates, with those from the nuclear genome being the fastest and those from mitochondrial and plastid DNA the slowest [27]; gene rearrangements are frequent in the mitochondrion, but this does not have an affect on phylogeny reconstruction based on the gene sequences. This situation contrasts with that of animals in which mitochondrial DNA has a higher

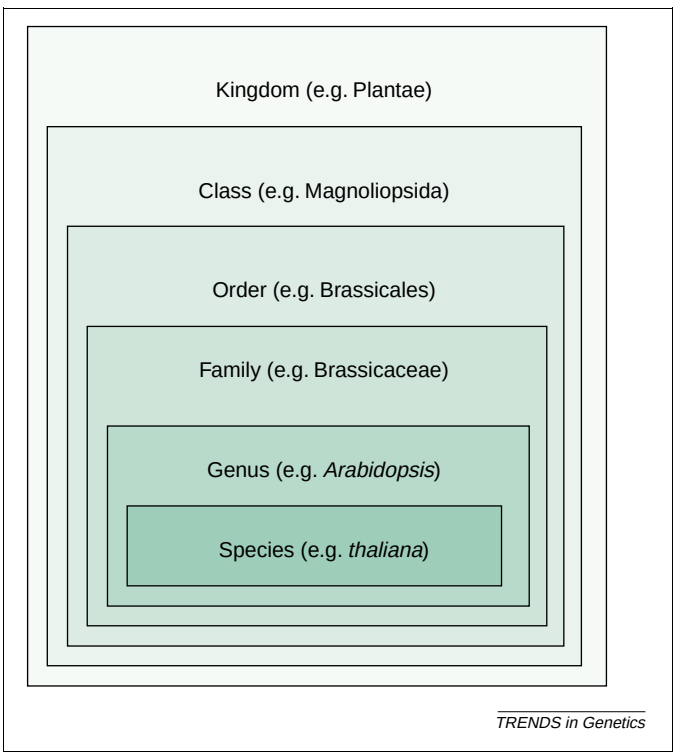


Figure 3. The systematic hierarchical categories of the classification of organisms using the example of *Arabidopsis*.

rate of nucleotide substitution than the nuclear genome and a lower rate of structural evolution than plant mitochondrial and plastid genomes [28]. Furthermore, there has been a great deal of confusion caused by genes being described as rapidly or slowly evolving; for example, 'rapidly evolving' or 'higher rates' could mean higher rates at the same variable sites, more variable sites in some homologous genes or a combination of both [29]. One main issue has been the effect of differential structural and functional constraints, and there have been some concerns about how these might affect phylogenetic inference, especially for the small organellar genomes (with fewer genes) so often used in phylogeny reconstruction and where constraints might be stronger as a result of 'lack of space'. For example, in animals differential functional constraints acting on nonneutral nucleotides of different proteins of the mitochondrial genome have resulted in incorrect evolutionary relationships receiving strong support [28,30]. For anciently diverged plants, concerns have also been raised [31,32], but in angiosperms close examination of plastid genes for their signal content (i.e. nucleotide changes shared due to common history) showed that these genes exhibited evenly distributed phylogenetic information [14] in the different codon positions, amino acids, chemical properties, hydrophobicity and charge, which is the opposite of the animal mitochondrial genome. It is clear that if severe functional constraints were acting on the plastid genome of flowering plants, we would have expected these sites to exhibit changes that not only reflect common history but also convergent changes necessary to preserve function; this was not the case [14]. Therefore, at least for angiosperms, it seems that botanists have made enormous strides in phylogeny inference due to characteristics inherent to the plastid genome (in terms of rates and types of changes at variable sites).

'A rose is still a rose but otherwise everything else in botany has been turned on its head'

Although not as drastic as stated in *The Independent*, 'Botanists reclassify all plants… A rose is still a rose but otherwise everything else in botany has been turned on its head' (pp. 1 and 3, 23 November 1998), botanists have produced the first DNA sequence-based classifications for a major group of organisms. Because angiosperm phylogenetic trees containing several hundreds of taxa were highly congruent although produced by genes in different genomes, botanists decided that it was time to translate the resulting patterns of relationships into a new and comprehensive classification. Rather than a classification reflecting the subjective views of a single author (i.e. based on intuitive ideas of plant evolution), the 'Angiosperm Phylogeny Group' (APG) aimed objectively to interpret published phylogenetic trees and compile them into a hierarchical system at and above the level of family. Their first classification was published in 1998 [33], and an update appeared in early 2003 [34]. The APG classification reflects evolutionary relationships that were newly discovered for ~60% of angiosperm families [33,35]; the main objective of this classification was to maximize information, thus making the system predictive [20].

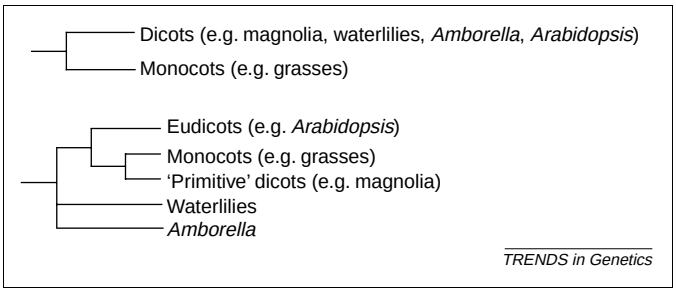


Figure 4. The major splits within angiosperms as they were viewed before the molecular phylogenetics era (top) and more recently as demonstrated by the use of molecular data (below).

The broad picture of angiosperm relationships has changed, with the first split among angiosperms not being that of MONOCOTS versus DICOTS, as stated in most botany textbooks, but instead one in which the 'primitive dicots' are closer to the monocots, a relationship reflected in their UNIAPERTURATE POLLEN grains versus the TRIAPERTURATE POLLEN of other dicots, the latter now being termed 'EUDICOTS' [33] (Figure 4). Perhaps one of the most spectacular changes of ideas concerns the sacred lotus (*Nelumbo*); because of its morphology and habitat preferences, the lotus was always considered a close relative of the water lilies (Nymphaeaceae), a group of 'primitive' dicots, whereas based on DNA sequence it is a eudicot for which the closest relatives are the northern temperate plane tree (*Platanus*) and the southern-hemisphere *Protea* family (Proteaceae) [20].

Rooting the phylogenetic tree of the angiosperms

Discovering new relationships is of course not only relevant to classification. Finding the ROOT of angiosperms, for example, has been the focus of several studies because it provides a direction and temporal scale for plant evolution (mostly calibrated with the fossil record) (Box 1), thereby permitting the production of explicit hypotheses of how traits such as genome size, colinearity of genes on chromosome arms and development have changed during the past 125 million years. Such ideas can then be used to generate research programmes designed to evaluate such predictions. The large flowers of *Magnolia* were long considered the archetype of the angiosperm flower because of their numerous, spirally arranged floral parts, but it has recently become evident that other flower types are equally as 'primitive' as those of *Magnolia*. These include the flowers of unusual plants such as *Amborella* (but see Ref. [36] for an alternative and controversial view) and *Piper* (the source of black pepper) (Figure 5), which were found to be outside the major clades in phylogenetic trees for angiosperms. It must, however, be stressed that knowing how remnants of basal lineages appear today does not necessarily tell us much about the traits of the ancestral angiosperms [37]. The first flowers could have been different from those of every extant group, and we will not know about them until their fossils have been found. The oldest angiosperm fossils are water lilies [38] and another aquatic plant, belonging to the newly described family Archaefructaceae [39], both ~125 million years old. Molecular systematic studies have refined ideas about

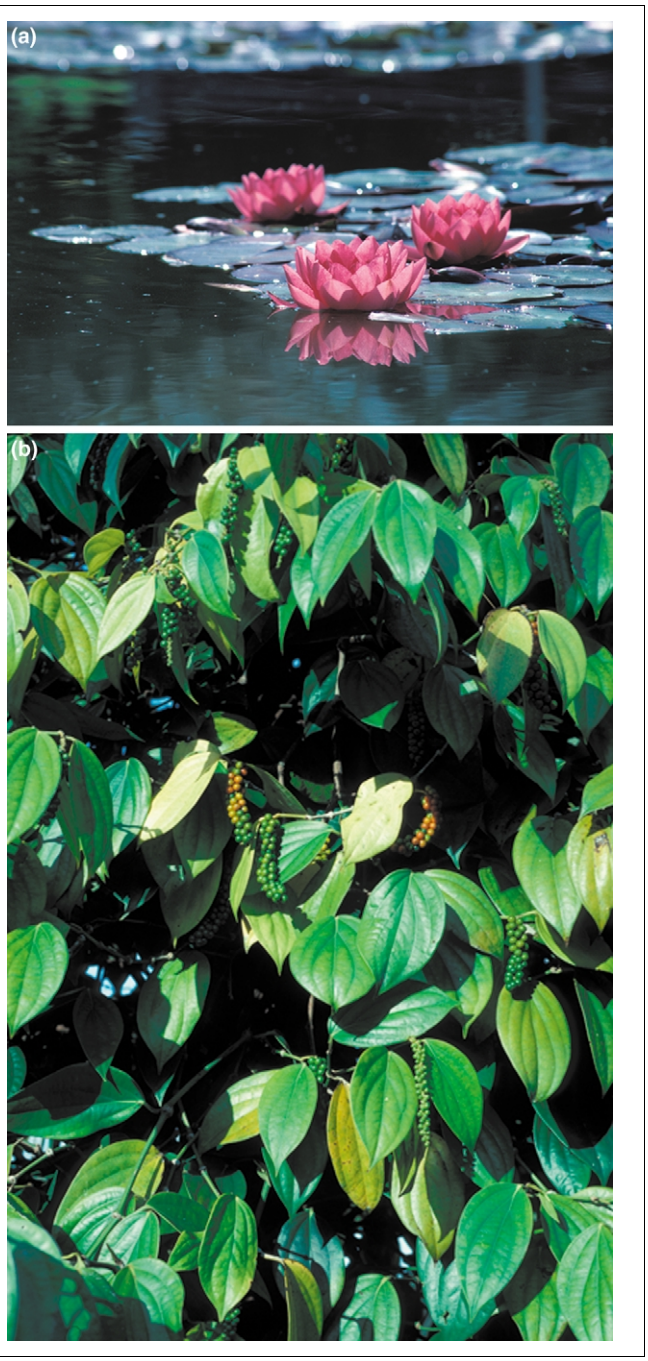


Figure 5. Two potential candidates of the archetype of the angiosperm flower: waterlily (*Nymphaea*, top) and black pepper (*Piper*, below). Photograph, courtesy of P. Gasson, Kew.

which sorts of fossils to look for, but the study of extant lineages alone cannot reveal all that is important for understanding the early angiosperms. For this, studies of fossils are essential.

Genome changes and plant evolution

As described above, it is difficult to infer the floral archetype of the angiosperms solely from knowledge of the phylogenetic relationships of extant species; we can, however, study several other important traits of the early angiosperms in this way, as long as we do not expect them to have been too plastic during the early stages of evolution. For example, by mapping genome sizes gathered from online databases [40] onto the general angiosperm phylogenetic tree, it was possible to infer that the ancestral genome was probably small [41]. How plant genomes increased to the large sizes observed in some modern groups [e.g. >127 pg per unreplicated gametic nucleus (1C value), in *Fritillaria*, a close relative of the lily] remains unexplained, but studies of selfish DNA and other retrotransposable elements could provide key answers [42–47]. At the least, knowing plant relationships can now help pinpoint, which lineages should be of interest, namely, those that have experienced the most drastic expansions or contractions in their genomes, especially because genome change might have provided major contributions to angiosperm radiations. For example, it is known that up to 70% of extant species are descended from taxa in which polyploidization events have occurred [48].

Features of genome evolution have also provided insight into plant phylogeny and vice versa. One example is the striking case of loss of the standard plant telomeric sequences (*Arabidopsis*-type repeats) and their replacement by other categories of repeats. *In situ* hybridization with telomeric probes demonstrated that onion (*Allium*) and aloe (*Aloe*) lack the typical repeats that cap all chromosome arms in the majority of plants [49]. By looking at the DNA-based phylogenetic analysis, it was clear that both species were members of the same order, Asparagales (as redescribed by APG [33], Figure 6) note that in many previous classifications these species were regarded as only distantly related), and therefore most Asparagales genera were examined for absence of the standard telomeric sequences [49]. Apart from a few closely related species of *Ornithogalum* (star of Bethlehem), none of the species between the aloe and onion has the typical plant telomeric repeats. Without phylogenetic information, none of these patterns would have been likely to be investigated in this manner, and clearly 'tree thinking' played a key role in this discovery.

Molecular clocks and molecular living fossils

The estimation of divergence times between species is important because it makes it possible to determine the speed of a variety of evolutionary processes, such as chromosome rearrangements, emergence of new forms of viruses and production of new body plans. When Zuckerlandl and Pauling found that the number of amino acid substitutions in haemoglobin was correlated with fossil-based time divergence estimates in vertebrates, the concept of the 'molecular clock' was born [50]. However, we now know that this clock ticks at varying speeds between lineages of organisms, and fossil-based versus DNA-based age estimates usually disagree, with molecules generally providing much older ages [51]. Using the broadly sampled angiosperm phylogenetic tree (based on plastid *rbcL* and *atpB* and nuclear 18S rDNA and comprising ~75% of all families [20,21]), NONPARAMETRIC RATE SMOOTHING (NPRS) [52] was applied to correct for rate heterogeneity across lineages, thus making the tree ULTRAMETRIC [53]. This chronogram was calibrated with reliable fossil data (the unique structure of the nuts of oaks and their allies; Box 1), and error estimates for the ages of the nodes of that tree were calculated by reapplying the NPRS protocols to

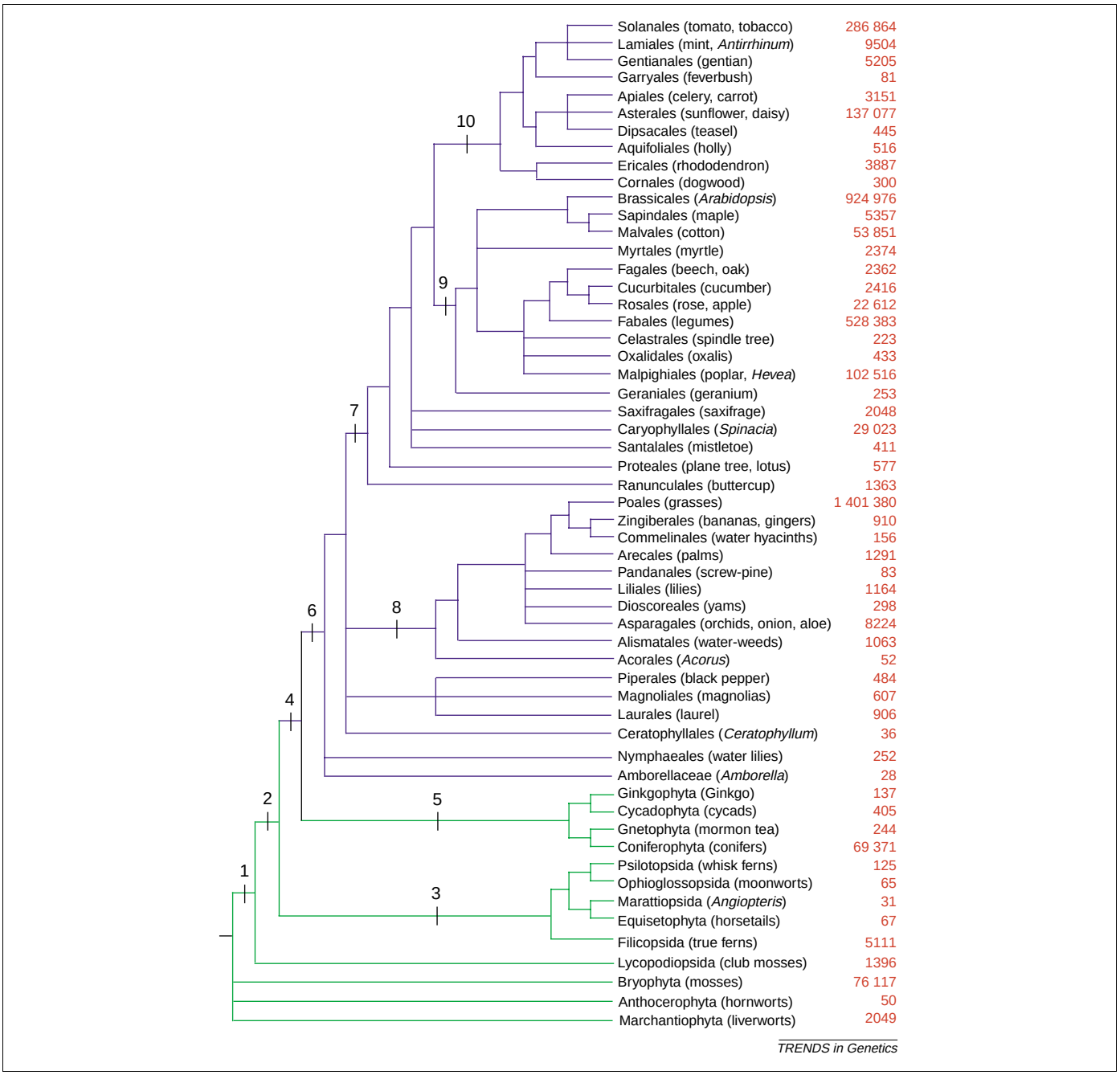


Figure 6. A summary of the terrestrial plant 'tree of life' [20,59] showing vascular plants (all descendants from node 1), which comprise angiosperms (nodes 6–10 depicted in blue) and remaining vascular plants (nodes 1-5 depicted in green). The main groups are leafy plants (node 2), ferns and their allies (node 3), seed plants (node 4), gymnosperms (node 5), flowering plants (node 6), eudicots (node 7), monocots (node 8), rosids (node 9) and asterids (node 10) (time scale not enforced). For flowering plants, most orders are indicated with some of their typical representatives or model organisms. Numbers on the right indicate the number of nucleotides entries held in EBI or GenBank in early November 2002, summing entries from mitochondrial, plastid and nuclear genomes. Several groups have a large number of entries because of the sequencing effort on model organisms. For conifers ~91% of entries belong to *Pinus*; for mosses, *Physcomitrella* (92%); for Malpighiales, *Populus* (92%); for Fabales, *Glycine* (57%) and *Medicago* (33%); for Asterales, *Lactuca* (49%) and *Helianthus* (33%); for Brassicales, *Brassica* (51%) and *Arabidopsis* (48%); and for Poales, *Zea* (26%), *Hordeum* (24%), *Oryza* (23%) and *Triticum* (20%).

bootstrapped DNA matrices. This molecular dating work is the largest published so far in terms of number of taxa (see Refs [54–56] for complementary references). It provides ages for the origin of nearly all angiosperm families, and most of these are in agreement for groups with a good fossil record [57]. However, like most previous studies of molecular clocks, the ages of the deepest nodes were underestimated by the fossil record, whereas the ages of the most recent groups thigh degree of correspondence, for most lineages, between fossil ages and the clock estimates [53] means that the ages of those without a fossil record can now be more reliably estimated than ever before, and this includes the great majority of angiosperm families and orders.

Looking at the VASCULAR PLANTS as a whole, a similar NPRS dating exercise was recently performed [58] using the most comprehensive phylogenetic tree for all lineages of vascular plants based on four genes [59]. Ages were depicted from single genes or combinations, in maximum parsimony and maximum likelihood frameworks, with several fossils of undisputed ages used as calibration

points [58]. Many DNA-based age estimates were in agreement with those from fossils, but it was also discovered that certain lineages have drastically decreased their rates of molecular evolution. This was the case, for example, with tree ferns, which were considered to be 'molecular living fossils' (see also Ref. [60]), paralleling at the genome level the relative morphological stasis they have exhibited for the past 200 million years [58].

Perspectives

There is no doubt that certain angiosperm lineages have been more successful than others in terms of species production, and several authors have documented these major shifts [61–67], although the factors responsible for increased rates of speciation remain unclear. Now that biodiversity is a major concern for society in general and biology in particular, perhaps only shared with human health, understanding the factors involved in its origin is fundamental. An examination of molecular rates in sister families of angiosperms showed that the more species-rich families have, on average, an increased rate of neutral substitutions in both plastid and nuclear genes [65]. In addition, the more diverse families in terms of morphology also have higher rates of DNA substitution [65], but this was not observed for animals [68]. This higher rate of background mutation (perhaps involving deficient DNA repair and exposure to mutagenic radiation) might affect developmental genes, thereby increasing morphological diversity (although alternative explanations are possible, especially regarding the effects of population size and structure on substitutions). This also holds for odd ecological niches with, for example, parasitism and carnivory in plants being associated with higher substitution rates [69].

Finally, it is clear that a decade of plant phylogenetics has resulted in major steps towards understanding the relationships between genes and species diversity. However, out of around 300 000 species of land plants, only 13 species account for over 81% of all plant nucleotides entries in EMBL and/or GenBank (genome data excluded, Figure 6). Large-scale sequencing projects can help explain the origins of phenotypic diversity [70] and, hopefully, intensive DNA sequencing of non-model organisms during the next decade will lead to new genetic syntheses, the phylogenomic era.

Acknowledgements

We thank Arne Mooers, Emmanuel Douzery, Peter Lockhart, several anonymous reviewers and the Editor for their comments on the manuscript.

References

1 Zurawski, G. *et al.* (1981) The structure of the gene for the large subunit of ribulose 1,5 bisphosphate carboxylase from spinach chloroplast DNA. *Nucleic Acids Res.* 9, 3251–3270
2 Chase, M.W. *et al.* (1993) Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene rbcL. *Ann. MO Bot. Gard.* 80, 528–580
3 Pagel, M. (1999) Inferring historical patterns of biological evolution. *Nature* 401, 877–884
4 Kyrpides, N.C. and Olsen, G.L. (1999) Archaeal and bacterial hyperthermophiles: horizontal gene exchange or common ancestry? *Trends Genet.* 15, 298–299
5 Wolf, Y.I. *et al.* (2002) Genome trees and the tree of life. *Trends Genet.* 18, 472–479
6 Hendy, M.D. and Penny, D. (1989) A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 310–321
7 Huelsenbeck, J.P. (1995) Performance of phylogenetic methods in simulation. *Syst. Biol.* 44, 17–48
8 Hillis, D.M. (1996) Inferring complex phylogenies. *Nature* 383, 130–131
9 Hillis, D.M. (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias. *Syst. Biol.* 47, 3–8
10 Huelsenbeck, J. and Hillis, D. (1993) Success of phylogenetic methods in the four-taxon cases. *Syst. Biol.* 42, 247–264
11 Källersjö, M. *et al.* (1998) Simultaneous parsimony jackknife analysis of 2538 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Syst. Evol.* 213, 259–287
12 Källersjö, M. *et al.* (1999) Homoplasy increases phylogenetic structure. *Cladistics* 15, 91–93
13 Savolainen, V. *et al.* (2000) Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. *Syst. Biol.* 49, 306–362
14 Savolainen, V. *et al.* (2002) Phylogeny reconstruction and functional constraints in organellar genomes: plastid *atpB* and *rbcL* sequences versus animal mitochondrion. *Syst. Biol.* 51, 638–647
15 Swofford, D.L. *et al.* (1996) Phylogenetic inference. In *Molecular Systematics* (Hillis, D.M. *et al.*, eds), pp. 407–514, Sinauer Associates Inc.
16 Yoder, A.D. and Yang, Z. (2000) Estimation of primate speciation dates using local molecular clocks. *Mol. Biol. Evol.* 17, 1081–1090
17 Savolainen, V. *et al.* (2000) Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. *Kew Bull.* 55, 257–309
18 Goremykin, V. *et al.* (1996) Noncoding sequences from the slowly evolving chloroplast inverted repeat in addition to *rbcL* data do not support Gnetalean affinities of angiosperms. *Mol. Biol. Evol.* 13, 383–396
19 Soltis, D.E. *et al.* (1997) Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Ann. MO Bot. Gard.* 84, 1–49
20 Soltis, P.S. *et al.* (1999) Angiosperm phylogeny inferred from multiple genes as a tool for comparative biology. *Nature* 402, 402–404
21 Soltis, D.E. *et al.* (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, *rbcL*, and *atpB* sequences. *Bot. J. Linn. Soc.* 133, 381–461
22 Mathews, S. and Donoghue, M.J. (1999) The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286, 947–950
23 Qiu, Y-L. *et al.* (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. *Nature* 402, 404–407
24 Qiu, Y-L. *et al.* (2000) Phylogeny of basal angiosperms: analysis of five genes from three genomes. *Int. J. Plant Sci.* 161, S3–S27
25 Graham, W.W. and Olmstead, R.G. (2000) Utility of 17 chloroplast genes for inferring the phylogeny of basal angiosperms. *Am. J. Bot.* 87, 1712–1730
26 Soltis, D.E. *et al.* (1998) Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Syst. Biol.* 47, 32–42
27 Palmer, J.D. (1985) Evolution of chloroplast and mitochondrial DNA in plants and algae. In *Molecular Evolutionary Genetics* (MacIntyre, R.J., ed.), pp. 131–240, Plenum Press
28 Naylor, G.J.P. and Brown, W.M. (1997) Structural biology and phylogenetic estimation. *Nature* 388, 527–528
29 Steel, M. *et al.* (2000) Invariable sites models and their use in phylogeny reconstruction. *Syst. Biol.* 49, 225–232
30 Naylor, G.J.P. and Brown, W.M. (1998) *Amphioxus* mitochondrial DNA, chordate phylogeny, and the limits of inference based on comparisons of sequences. *Syst. Biol.* 47, 61–76
31 Lockhart, P.J. *et al.* (1998) A covariotide model explains apparent phylogenetic structure of oxygenic photosynthetic lineages. *Mol. Biol. Evol.* 15, 1183–1188
32 Lockhart, P.J. *et al.* (1999) Spectral analysis, systematic bias, and the evolution of chloroplast. *Mol. Biol. Evol.* 16, 573–576
33 Angiosperm Phylogeny Group (APG), (1998) An ordinal classification for the families of flowering plants. *Ann. MO Bot. Gard.* 85, 531–553
34 Angiosperm Phylogeny Group (APG), (2003) An update of the

angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. *Bot. J. Linn. Soc.* 141, 399–436
35 Chase, M.W. *et al.* (2000) Higher-level classification in the angiosperms: new insights from the perspective of DNA sequence data. *Taxon* 49, 685–704
36 Goremykin, V.V. *et al.* (2003) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol. Biol. Evol.* 20, 1499–1505
37 Donoghue, M.J. and Doyle, J.A. (2000) Seed plant phylogeny: demise of the anthophyte hypothesis? *Curr. Biol.* 10, R106–R109
38 Friis, E.M. *et al.* (2001) Fossil evidence of waterlilies (Nymphaeales) in the Early Cretaceous. *Nature* 410, 357–360
39 Sun, G. *et al.* (2002) Archaefructaceae, a new basal angiosperm family. *Science* 296, 899–904
40 Bennett, M.D. (1998) Plant genome values: how much do we know? *Proc. Natl. Acad. Sci. U. S. A.* 95, 2011–2016
41 Leitch, I.J. *et al.* (1998) Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann. Bot. (Lond.)* 82, 85–94
42 Palmer, J.D. and Delwiche, C.R. (1998) The origin and evolution of plastids and their genomes. In *Molecular Systematics of Plants II* (Soltis, D.E. *et al.*, eds), pp. 375–409
43 Cho, Y. *et al.* (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14244–14249
44 Cho, Y.R. and Palmer, J.D. (1999) Multiple acquisitions via horizontal transfer of a group I intron in the mitochondrial cox1 gene during evolution of the Araceae family. *Mol. Biol. Evol.* 16, 1155–1165
45 Palmer, J.D. *et al.* (2000) Dynamic evolution of mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6960–6966
46 Bennetzen, J.L. (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* 115, 29–36
47 Wendel, J.F. *et al.* (2002) Feast and famine in plant genomes. *Genetica* 115, 37–47
48 Grant, V. (1971) *Plant Speciation*, Columbia Press
49 Adams, S.P. *et al.* (2001) Loss and recovery of *Arabidopsis*-type telomere repeat sequences 5′-(TTTAGGG)$_n$-3′ in the evolution of a major radiation of flowering plants. *Proc. R. Soc. London B Biol. Sci.* 268, 1541–1546
50 Zuckerlandl, E. and Pauling, L. (1962) Molecular disease, evolution, and genetic heterogeneity. In *Horizons in Biochemistry* (Kasha, M. and Pullman, B., eds), pp. 189–225, Academic Press
51 Heckman, D.S. *et al.* (2001) Molecular evidence for the early colonization of land by fungi and plants. *Science* 293, 1129–1133
52 Sanderson, M.J. (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* 14, 1218–1231
53 Wikström, N. *et al.* (2001) Evolution of the angiosperms: calibrating the family tree. *Proc. R. Soc. London B Biol. Sci.* 268, 2211–2220
54 Goremykin, V.V. *et al.* (1997) Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. *Plant Syst. Evol.* 206, 337–351
55 Sanderson, M.J. (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109
56 Britton, T. *et al.* (2002) Phylogenetic dating with confidence intervals using mean path lengths. *Mol. Phylog. Evol.* 24, 58–65
57 Magallón, S. *et al.* (1999) Phylogenetic pattern, diversity, and diversification of eudicots. *Ann. MO Bot. Gard.* 86, 297–372
58 Soltis, P.S. *et al.* (2002) Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. *Proc. Natl. Acad. Sci. U. S. A.* 99, 4430–4435
59 Pryer, K.M. *et al.* (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to the seed plants. *Nature* 609, 618–622
60 Sanderson, M.J. and Doyle, J.A. (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. *Am. J. Bot.* 88, 1499–1516
61 Sanderson, M. and Donoghue, M. (1994) Shifts in diversification rate with the origin of angiosperms. *Science* 264, 1590–1593
62 Sanderson, M.J. and Wojciechowski, M.F. (1996) Diversification rates in a temperate legume clade – why are there so many species of *Astragalus* (Fabaceae). *Am. J. Bot.* 83, 1488–1502
63 Barraclough, T.G. *et al.* (1996) Rate of *rbcL* gene sequence evolution and species diversification in flowering plants (angiosperms). *Proc. R. Soc. London B Biol. Sci.* 263, 589–591
64 Savolainen, V. and Goudet, J. (1998) Rate of gene sequence evolution and species diversification in flowering plants: a re-evalutation. *Proc. R. Soc. London B Biol. Sci.* 265, 603–607
65 Barraclough, T.G. and Savolainen, V. (2001) Evolution rates and species diversity in flowering plants. *Evolution* 55, 677–683
66 Magallón, S. and Sanderson, M.J. (2001) Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780
67 Savolainen, V. *et al.* (2002) Is cladogenesis heritable? *Syst. Biol.* 51, 835–843
68 Bromham, L. *et al.* (2002) Testing the relationship between morphological and molecular rates of change along phylogenies. *Evolution* 56, 1921–1930
69 Jobson, R.W. and Albert, V.A. (2002) Molecular rates parallel diversification contrasts between carnivorous plant sister lineages. *Cladistics* 18, 127–136
70 Soltis, D.E. *et al.* (2002) Missing links: the genetic architecture of flower and floral diversification. *Trends Plant Sci.* 7, 22–31
71 Thorne, J.L. and Kishino, H. (2002) Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702

Darwin's abominable mystery: Insights from a supertree of the angiosperms

T. Jonathan Davies*†, Timothy G. Barraclough*†, Mark W. Chase†‡, Pamela S. Soltis§¶, Douglas E. Soltis¶‖, and Vincent Savolainen†‡****

*Department of Biological Sciences and Natural Environment Research Council Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, United Kingdom; †Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew TW9 3DS, United Kingdom; and §Florida Museum of Natural History, ‖Department of Botany, and ¶Genetics Institute, University of Florida, Gainesville, FL 32611

Communicated by Peter R. Crane‡, Royal Botanic Gardens, Surrey, United Kingdom, December 8, 2003 (received for review June 20, 2003)

Angiosperms are among the major terrestrial radiations of life and a model group for studying patterns and processes of diversification. As a tool for future comparative studies, we compiled a supertree of angiosperm families from published phylogenetic studies. Sequence data from the plastid *rbcL* gene were used to estimate relative timing of branching events, calibrated by using robust fossil dates. The frequency of shifts in diversification rate is largely constant among time windows but with an apparent increase in diversification rates within the more recent time frames. Analyses of species numbers among families revealed that diversification rate is a labile attribute of lineages at all levels of the tree. An examination of the top 10 major shifts in diversification rates indicates they cannot easily be attributed to the action of a few key innovations but instead are consistent with a more complex process of diversification, reflecting the interactive effects of biological traits and the environment.

In a letter to J. D. Hooker dated July 22, 1879 (1), Charles Darwin described the rapid rise and early diversification within the angiosperms as "an abominable mystery." Angiosperms are regarded as one of the greatest terrestrial radiations of recent geological times. The major lineages originated 130–90 million years ago (mya) (2, 3), followed by a dramatic rise to ecological dominance 100–70 mya (4). Approximately 250,000 extant species have been recognized (5), although estimates vary, and the final number might be double this (6). Within the group, sister clades can differ in species richness over several orders of magnitude. Darwin attempted to identify a single causal explanation for the rapid diversification of angiosperms but described his own efforts as "wretchedly poor" (1).

Subsequent attempts to understand angiosperm diversification have come from a variety of fields. Studies of the fossil record have explored the origin of angiosperms and the spatiotemporal patterns of their radiation (3, 7–9). A complementary approach has been the use of systematic data of living species to identify major trends in angiosperm evolution and their possible effects on diversification (10). For example, many authors have investigated the importance of biological traits, such as biotic pollination (2, 11, 12), biotic seed dispersal (13–15), and life history flexibility (16, 17), as putative key innovations. Increasingly, such studies rely on knowledge of phylogenetic relationships among higher taxa to estimate net diversification rates and pinpoint independent evolutionary events (18–21), thereby circumventing the problems associated with comparing higher taxa of different ages (22).

Recent advances in molecular phylogenetics have heralded a new era in plant phylogenetics. Since the molecular phylogenetic tree of angiosperms based on plastid *rbcL* sequence data by Chase *et al.* (23), a succession of large-scale angiosperm trees has appeared over the last decade (24–26). Increased sampling of taxa and the use of multiple genes (27–29) have led to increased resolution and confidence in angiosperm relationships (30). These data have become a major resource for comparative biology, but to date no single analysis has included all currently recognized angiosperm families.

Here we use a supertree approach to combine recent phylogenetic data into the first complete family-level phylogenetic tree of the angiosperms, a task that was described as "formidable" and "impossible to meet" just over a decade ago (18). We present this tree, together with dates calibrated by using the fossil record and estimated from molecular branch lengths, as a compilation of current knowledge and a tool for comparative plant biology. In addition, we use the supertree to present the first complete survey of diversification among familial angiosperm lineages. Our aim is to identify at which points on the tree major shifts occurred and use this information to guide the examination of factors that might explain the mystery of angiosperm diversification.

Methods

Supertree Construction. Supertree methods are being used increasingly to combine multiple sources of phylogenetic data into a single analysis. We used matrix representation with parsimony (MRP), which codes branching patterns of individual source trees as a binary matrix and missing taxa as question marks. The matrices for all of the trees are then combined, and a tree search is performed on the combined matrix using parsimony (31, 32). The best practice for supertree analyses is an active area of research (33), but MRP is widely recognized as one of the best current methods and has been successfully applied in a large number of studies (34–36).

Forty-six source trees were selected from published and unpublished work on the basis of either their comprehensive coverage or resolution of previously poorly understood relationships, with the aim of maximizing the number of families represented (a list of source trees is given in Table 2, which is published as supporting information on the PNAS web site). To take into account levels of support for relationships, we used bootstrap percentages for nodes in the source trees as character weights for the MRP binary matrix, following the method of Salamin *et al.* (34) (further details are provided in *Supporting Methods*, which is published as supporting information on the PNAS web site). Family delineations followed the Angiosperm Phylogeny Group (APG) classification (37, 38). For six families, we were unable to find published phylogenetic treatments (listed in Table 3, which is published as supporting information on the PNAS web site).

The SUPERTREE0.8B program [www.tcd.ie/Botany/NS/

Abbreviations: mya, millions of years ago; SG, Slowinski–Guyer measure of imbalance; logN, maximum likelihood estimate of shift in diversification rate; APG, Angiosperm Phylogeny Group; MRP, matrix representation with parsimony.

‡M.W.C. and V.S. are employed by the Jodrell Laboratory, within the Royal Botanic Gardens, of which P.R.C. is director. The Royal Botanic Gardens did not provide any funding specifically for this research.

**To whom correspondence should be addressed. E-mail: v.savolainen@kew.org.

www.pnas.org/cgi/doi/10.1073/pnas.0308127100

SuperTree.html (34)] was used to create a single binary matrix representing all of the relationships in the above trees. The binary matrix was analyzed with PAUP4.0B8 (39) by using weighted parsimony with the following heuristic search: 250 replicates of random taxon addition, subtree pruning–regrafting branch swapping, and holding 10 trees at each replicate. The saved trees were then used as the starting trees in another search using tree bisection–reconnection with a tree limit of 10,000 equally most parsimonious trees.

As estimation of divergence times and consequently diversification rates requires a completely bifurcating topology; all subsequent analysis was performed on one of the most parsimonious supertrees. To examine whether arbitrary resolutions may have biased our results, we repeated each subsequent analysis of diversification rates excluding nodes that collapsed in the strict consensus tree.

The topology of the supertree was compared to that of the three-gene (*atpB*, *rbcL*, and 18S rDNA) bootstrap tree generated from the matrix of Soltis *et al.* (28). Sampling both the plastid and nuclear genome and with broad taxonomic coverage, this tree is regarded as the best estimate of angiosperm phylogeny to date. Therefore, as quality control for the supertree, we checked whether strongly supported relationships in this source tree are also present in the supertree. We used a parsimony equivalent of the Shimodaira–Hasegawa test (40) to compare tree lengths for three-gene source data optimized onto each tree topology in turn by using 500 bootstrap replicates and 10,000 random trees. Second, we compared the number of nodes in common between the two trees by using the program TREECORRECT1.2B [www.tcd.ie/Botany/NS/software.html (41)].

Dating. We estimated the amount of molecular change along branches in the tree by using a matrix of *rbcL* sequences compiled from the source matrices or downloaded from GenBank (www.ncbi.nlm.nih.gov). The *rbcL* gene was chosen because it has been sequenced for most of the taxa in the supertree. Branch lengths were optimized onto the supertree by using maximum likelihood assuming an HKY85 + Γ DNA substitution model in PAUP4.0b8 (39). This model provides a compromise between model complexity and computational time (42). The phylogenetic tree was arranged with Amborellaceae as sister to the rest of the angiosperms (27, 28, 43–45). To correct for variation in substitution rate among lineages, we used nonparametric rate smoothing (46), as implemented in TREEEDIT V1.0 A10 (http://evolve.zoo.ox.ac.uk/software/TreeEdit). A single family on the supertree, Triuridaceae, lacked *rbcL* sequence data and was placed arbitrarily halfway along the branch leading to its sister clade.

The tree was calibrated in units of millions of years by using the split between Fagales and Cucurbitales set to 84 mya [after Wikström *et al.* (47)]. To check consistency of date estimates, we also calibrated the tree, setting the stem lineage subtending the eudicot crown group set to 126 mya (48), and compared the alternative dates.

Measuring Diversification. Species numbers for families were taken from Watson and Dallwitz (refs. 50 and 51 and http://biodiversity.uno.edu/delta/angio). If the generic composition of a family differed from that currently accepted by the APG (37, 38), species richness was adjusted to be in agreement with the APG classification (see *Supporting Methods*).To determine whether there is significant variation in diversification rates among angiosperm lineages, we calculated the overall tree imbalance by using the mean tree imbalance measure of Fusco and Cronk (51) as modified by Purvis *et al.* (52) on the strict consensus tree, because arbitrary resolutions of polytomies have been shown to inflate imbalance (53, 54).

We used two complementary methods to pinpoint where diversification rates changed on the tree. First, we estimated net diversification rates for all clades on the tree using $\log(N)/t$, where N is the number of species within a clade, and t is the time since the clade diverged from its sister clade on the dated tree. Changes in diversification rate on the tree were calculated by subtracting the rate for each clade from the rate of its immediate nesting clade. We refer to this measure as maximum likelihood estimate of shift in diversification rate (logN) rate shifts. Second, we compared the species richness of all sister clades on the tree by using the Slowinski–Guyer measure of imbalance (SG; ref. 56), which assigns a probability of observing an equal or greater difference in species numbers at each node under a general null model that diversification rates in the two daughter clades have been equal. Sister clades are the same age, and therefore this approach accounts for possible effects of different clade ages on current species richness using information on topology alone. Due to the nested nature of phylogenetic comparisons, families with a large or small number of species can influence the degree of imbalance at nodes nesting nearer the root (56). We corrected for this nonindependence by using a heuristic approach, described in *Supporting Methods*.

The distribution of shifts in diversification rate across the tree using the latter two measures were explored by using randomization tests to examine whether the diversification rate is phylogenetically conserved. First, we examined heritability of diversification rates among branches of the tree. Details of randomization test procedures are provided in *Supporting Methods*. Second, we looked for concentration of shifts in diversification rate in either particular time windows or particular angiosperm orders recognized by the APG.

Finally, for both measures of shifts in diversification rate, we identified the top 10 shifts found across the tree. Because the logN measure includes the direction of each shift as well as magnitude, we identified the top 10 increases and decreases in diversification rates separately. We then categorized the affected clades in terms of several factors previously proposed to influence diversification rates in angiosperms, ranging from pollination syndrome to geographic range (taken from Watson and Dallwitz's online database, http://biodiversity.uno.edu/delta/angio). Clades were labeled polymorphic if they exhibit a mixture of possible values. In addition, we used taxonomic descriptions to identify any other general features of the clades. The goal was not to perform a comprehensive test of correlates of diversification in angiosperms but rather to explore whether single factors or simple combinations might be associated with the major shifts in angiosperm radiation. We also recorded the level of support for the nodes: one explanation for large shifts might be phylogenetic error, for example, if a small family were mistakenly placed as sister to a larger clade. The nonindependence of characters within the MRP matrix violates the assumptions of the bootstrap; thus estimates of node support were inferred from the individual source trees.

Results and Discussion

Supertree and Dates. Our MRP analysis generated 10,000 most parsimonious supertrees, one of which is summarized in Fig. 1 and presented in full in Fig. 2, which is published as supporting information on the PNAS web site; nodes collapsing in the strict consensus are indicated with an arrow on the latter. Although undoubtedly more equally most parsimonious trees could have been found with continued branch swapping, it may be reasonable to assume that those nodes liable to collapse in a strict consensus of all most parsimonious trees were identified by using the search implemented.

The final trees include 379 terminal taxa representing monophyletic clades, mostly families but also higher clades in cases for which recognized families are not monophyletic. Because the source trees were predominantly molecular, and all of the matrices included sequence data for *rbcL*, the supertree is

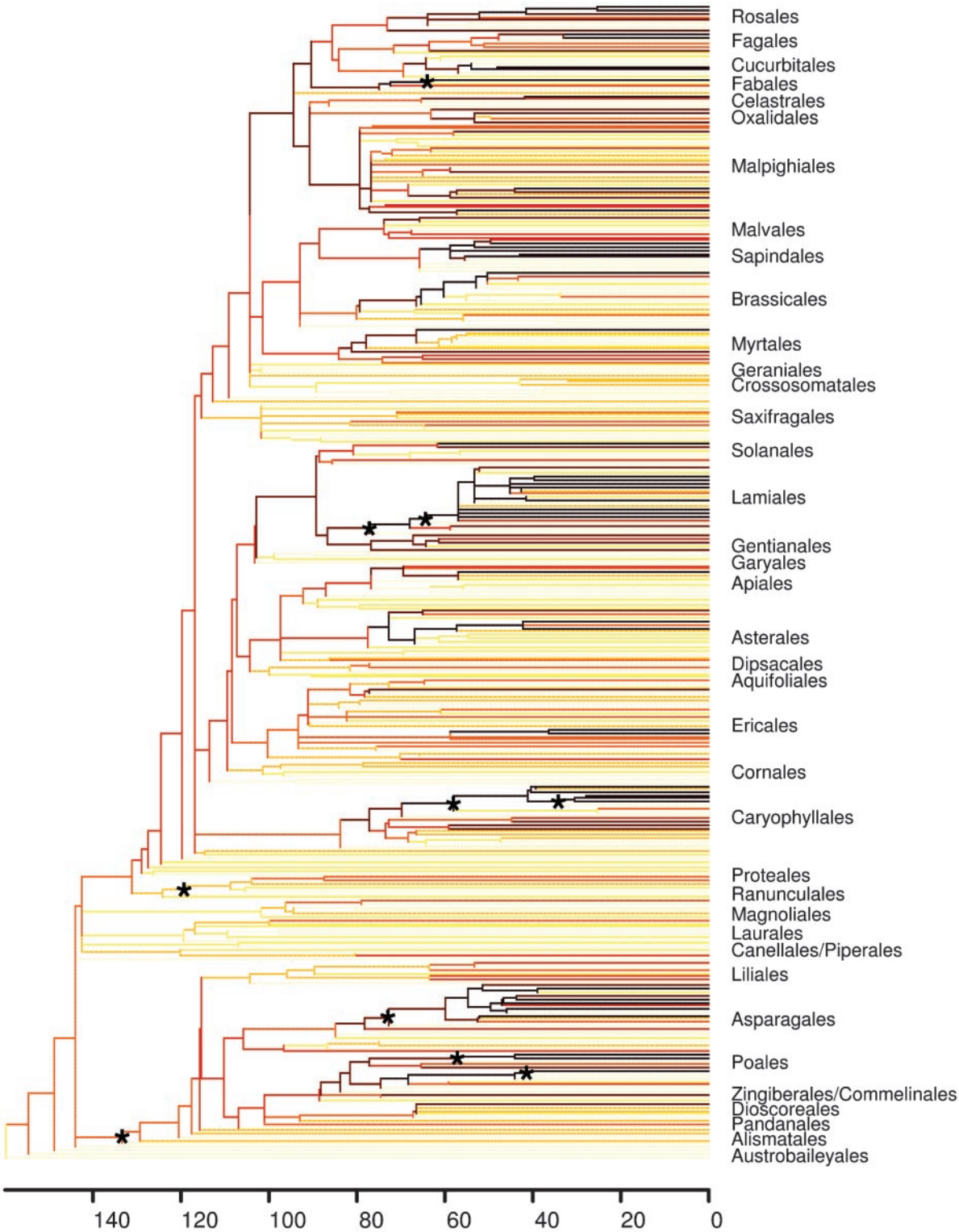


Fig. 1. One of 10,000 most parsimonious supertrees with dates obtained by used nonparametric rate smoothing transformation of maximum likelihood branch lengths from *rbcL* sequence data. The time scale was calibrated by using the split between Fagales and Cucurbitales at 84 mya. The strength of shading reflects diversification rates estimated as log (number of species)/age since split from sister clade. See Fig. 2 for a larger figure showing names of all terminal taxa. Diversification rates vary from low (yellow to orange) to high (red to black). Asterisks indicate the top 10 most imbalanced nodes referred to in Table 1.

inevitably biased toward the *rbcL* gene tree. We retained the input of *rbcL* data because the best estimate of relationships within the source trees is likely to come from combined analysis of all available markers (57).

Because the topology derives from existing phylogenetic hypotheses, we do not present an in-depth discussion of recovered relationships. As noted above, a few families do not appear monophyletic, most noticeable within Caryophyllales, despite

our following the APG classification for circumscription of families. These families are known to be nonmonophyletic, for example Portulacaceae and Phytolaccaceae, but changes in circumscription were judged in the last APG classification to be premature until comprehensive studies are performed. The occurrence of nonmonophyletic families and polytomies within the supertree highlight areas in need of more rigorous analysis and more data.

Differences between the supertree and individual source trees could in principle be caused by hard incongruence among studies or by phylogenetic errors due to relationships with low levels of bootstrap support (58). Most conflicts between the supertree and the three-gene source tree are at nodes with weak support in the three-gene tree. Only 69% of nodes in the source tree are found in the supertree, but all nodes with bootstrap support >70% were present. The supertree was not significantly different from the three-gene source tree in its fit to the three-gene molecular data (modified Shimodaira–Hasegawa test, $P = 0.58$). These results indicate that the weighted MRP analysis accurately reproduced the relationships supported by the best sampled source tree.

There remains active debate over methods for calibrating phylogenetic trees (42, 47, 59), but many alternative methods are not applicable to such large data sets. The alternative calibration point, the origin of the eudicots, produced slightly younger estimates of divergence times, dating the split between Fagales and Cucurbitales ≈10 my younger than that suggested by the fossil record and leading to, on average, 89% younger dates than for the alternative calibration. More generally, there remain examples of inconsistencies in fossil and molecular dates for angiosperm lineages, with a tendency for molecular dates to overestimate deeper nodes, such as the origin of the eudicots, and underestimate more terminal nodes (e.g., Poaceae, Moraceae, and Salicaceae) (47). Discrepancies between molecular and fossil dates are frequent in all groups where comparisons have been made (60). Whether these differences relate to biases in molecular dating procedures, errors in fossil sampling and identification, or both remains to be investigated thoroughly. At present, we have no means to correct for these differences and therefore simply present our results as a comprehensive molecular estimate of branching events for all angiosperm families calibrated by fossil dates assumed to be robust. Because our later analyses rely predominantly on relative age estimates of different families, rather than absolute age, we discuss only results using the Fagales–Cucucurbitales calibration point.

Patterns of Diversification. Analysis of the supertree revealed significant imbalance in net diversification rates among angiosperm lineages compared to the null model that all lineages have an equal diversification rate (weighted mean $I = 0.72$, $P < 0.001$; I, tree imbalance). The comprehensive taxonomic sampling of the supertree allows increased confidence in these findings, which broadly correspond to previous estimates of phylogenetic imbalance within the angiosperms (51) and coincide with the general pattern found across a wide range of taxa (61, 62). Placing the six families not represented in the source trees in the final supertree based on published statements of their likely affinities (see Table 3) did not change our results; we discuss below only those results excluding these families.

The two methods of reconstructing shifts in diversification rate on the tree yielded mostly similar results. Nodes that exhibit a significant SG value tend to have a large logN rate shift. The few exceptions to this trend were cases in which two sister clades with balanced species numbers were joined by a relatively long stem branch. This led to reconstruction of a high rate in both sister clades compared to the rate expected for their nesting clade, a situation not recognizable from topology alone. Overall the measures give the same visual picture of diversification: frequent shifts in diversification rate have occurred across the tree (Fig. 1).

The randomization test found that diversification rates are significantly phylogenetically heritable between related lineages, but only marginally so (logN rate shifts, $P = 0.040$; SG values, $P = 0.031$). Hence, sister families are only marginally more likely to have similar species numbers than two families chosen at random, indicating that diversification rate is a labile attribute.

There was also only weak evidence that particular orders of angiosperms have experienced a greater frequency of shifts than others (randomization test, logN rate shifts, $P > 0.1$; SG values, $P = 0.036$), excluding collapsing nodes from the analysis further reduced significance in both analyses. However, the frequency of reconstructed shifts did vary among time windows, and the exact pattern differed between the SG and logN methods of assigning rate shifts (logN rate shifts, $P = 0.024$; SG values, $P > 0.1$; see Figs. 3 and 4, which are published as supporting information on the PNAS web site). Nodes in more recent time periods tended to display a greater logN rate shift ($P < 0.001$) than expected under the null model, associated with the observation of sister families with long stem branches outlined above. One possible explanation would be if diversification rates have increased uniformly across all lineages within very recent time periods. However, an alternative explanation is that this pattern reflects a bias due to the use of families as terminal taxa: shifts occurring within families can be reconstructed only as occurring in the entire family in our analyses. Reconstructed shifts in diversification rates at nodes deeper in the tree would be unaffected by any such bias; hence our overall results are not affected by the sampling of families as terminal taxa, providing all terminal clades are monophyletic, and we can assign all recognized species of angiosperms to one of the tips in the tree.

The top 10 most imbalanced nodes (SG measure) in the strict consensus supertree are shown in Table 1. Equivalent tables for the logN rate shifts are in Tables 2–8, which are published as supporting information on the PNAS web site. The exact membership of the tables varies with the measure of rate shifts used and whether we correct for nesting of species richness or not, but the general conclusions are unchanged. The top 10 nodes do not reflect poorly supported parts of the tree, rejecting phylogenetic inaccuracy as an explanation for their high imbalance. None of the biological traits stand out as unequivocal key innovations explaining the major shifts in diversification. As can be seen from Table 1 (see also Tables 4 and 5), clades with higher species richness tend to be more polymorphic in the traits considered and cover a wider geographical range, but whether this is a cause or an effect of increased species richness is difficult to evaluate at this level (e.g., see ref. 63). Similarly, major shifts near the root of the tree, such as those leading to the core eudicots and monocots, are characterized by species-rich clades that are polymorphic in all traits considered in this paper and have cosmopolitan distributions. In contrast, the species-poor sister lineages are polymorphic for only approximately a quarter of the traits considered and have typically much more restricted distributions.

Conclusion

As a tool for comparative biology, we have reconstructed a dated supertree of angiosperm families with species numbers presented for the terminals. Our analyses revealed a strikingly labile pattern of diversification rate in the angiosperms. This pattern is not the result solely of phylogenetic inaccuracy and misplaced taxa, because many of the nodes with major shifts are strongly supported in the source trees.

Our results uphold Darwin's suspicions that simple explanations for the mystery of angiosperm diversification are inadequate. Our calibration of the diversification of the major angiosperm lineages does show an early rapid radiation of the basal

Table 1. Top 10 most imbalanced nodes (SG) and their derived clades

Imbalance (SG)	Sister clades	Age, mya	Node support/ (source ref.)	Pollination mode	Dispersal mode	Habit	Strictly dioecious	Chromosome no.	Geographic distribution	Lifestyle
0.0002	**Lamiales I***	77	53/(25)	Biotic/?	Poly	Poly	None	6–96	Cosmopolitan	Poly
	Plocospermataceae			?	Abiotic	Woody	None	?	Central America	Perennial
0.0002	**Poaceae**	41	97/(66)	Abiotic	Poly	Poly	Poly	4–122	Cosmopolitan	Poly
	Ecdeiocoleaceae			Abiotic	?	Herbaceous	None	?	Australia	Perennial
0.0004	**Monocots**	133	87/(26)	Poly	Poly	Poly	Poly	4–180†	Cosmopolitan	Poly
	Acoraceae			?	Poly	Herbaceous	None	24	Old World and North America	Perennial
0.0007	**Asparagales***	72	69/(26)	Biotic/?	Poly	Poly	Poly	6–180	Cosmopolitan	Perennial
	Xeronemataceae			?	Biotic	Herbaceous	None	?	New Zealand and New Caledonia	Annual
0.0010	**Lamiales II***	64	43/(25)	Biotic/?	Poly	Poly	None	6–96	Cosmopolitan	Poly
	Tetrachondraceae			?	Abiotic	Herbaceous	None	?	New Zealand and Patagonia	Perennial
0.0011	**Fabaceae**	64	54/(28)	Biotic	Poly	Poly	None	10–112	Cosmopolitan	Poly
	Surianaceae			?	Biotic	Woody	None	?	Pan subtropical to tropical	Perennial
0.0012	**Caryophyllales I***	85	98/(28)	Poly	Poly	Poly	Poly	12–144	Cosmopolitan	Poly
	Asteropeiaceae/ Physenaceae			Abiotic/?	Abiotic	Woody	Poly	?	Madagascar	Perennial
0.0014	**Caryophyllales II***	74	40/(25)	Poly	Poly	Poly	Poly	12–144	Cosmopolitan	Poly
	Stegnospermaceae			?	Abiotic	Woody	None	?	North & central America	Perennial
0.0015	**Ranunculales***	119	88/(26)	Poly	Poly	Poly	Poly	12–56†	Cosmopolitan	Poly
	Eupteleaceae			Abiotic	Abiotic	Woody	None	28	East Asia	Perennial
0.0015	**Cyperaceae/ Juncaceae**	47	55/(28)	Poly	Abiotic	Poly	Poly	12–112†	Cosmopolitan	Poly
	Thumiaceae			Abiotic	Abiotic	Herbaceous	None	?	North South America	?

Bold indicates the larger clade; the respective nodes are indicated in Fig. 1 by asterisks. Other than where stated, ecological data were derived from Watson and Dallwitz's online database (refs. 49 and 50 and http://biodiversity.uno.edu/delta/angio). Poly, polymorphic; ?, unknown.
*Taxonomic description of clades is given in Table 6.
†Values obtained from Plant DNA C-values Database 2.0 (www.rbgkew.org.uk/cval/homepage.html).

lineages, and this could be taken to account for what Darwin considered to be the "rapid rise and early diversification" of the angiosperms, which was his "abominable mystery" (1). However, numerous other shifts in diversification rates have occurred throughout the history of angiosperms, including several large increases in rates in recent time periods. The pattern is not consistent with a simple model in which diversification is driven by a few major key innovations but rather argues for a more complex process in which propensity to diversify is highly labile: there are "winners" and "losers" at all levels, and shifts occur repeatedly. This conclusion is supported by our tabulation of characteristics of clades affected by the major shifts and previous studies on incomplete phylogenetic trees (21, 64). Traits that may characterize particular species-rich clades are not sufficient to guarantee phylogenetic success, because within all species-rich higher clades we observe several shifts to slower rates of diversification.

Together, these results have implications for future analyses on how the interaction between traits and the environment affects diversification: some traits convey success in some environments but not others. Phylogenetic studies of diversity rely on inferences from current species numbers in terminal clades. Therefore, patterns of diversification reconstructed onto phylogenetic trees depend on the age of lineages, their intrinsic attributes, and also the environments experienced since their origin, particularly recent conditions. Global environments have changed considerably during the history of angiosperm radiation: which lineages are diverse now depends on the match between traits and recent climates, e.g., the rise to dominance of grasses during the late Tertiary is linked to global cooling and drying (65). Ultimately, increasing phylogenetic resolution at the level of genera and below may be needed to produce detailed models of how these interacting effects influence diversification. Our supertree represents a step toward this goal.

We thank Nicolas Salamin and Andy Purvis for help with the supertree. This work was supported by a Natural Environment Research Council studentship (to T.J.D.), a Royal Society University Research Fellowship (to T.G.B.), a U.S.–U.K. Fulbright Distinguished Professorship (to D.E.S. and P.S.S.), and U.S. National Science Foundation Grant DEB-0090283 (to D.E.S., P.S.S., D. L. Dilcher, and P. S. Herendeen).

1. Darwin, F. & Seward, A. C., eds. (1903) *More Letters of Charles Darwin* (John Murray, London), Vol. 2.
2. Labandeira, C. C., Dilcher, D. L., Davis, D. R. & Wagner, D. L. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 12278–12282.
3. Crane, P. R., Friis, E. M. & Pedersen, K. J. (1995) *Nature* **374,** 27–33.
4. Barrett, P. M. & Willis, K. J. (2001) *Biol. Rev.* **76,** 411–447.
5. Wilson, E. O. (1992) *The Diversity of Life* (Belknap, Harvard Univ. Press, Cambridge, MA).
6. Bramwell, D. (2002) *Plant Talk* **28,** 32.
7. Axelrod, D. I. (1952) *Evolution (Lawrence, Kans.)* **6,** 29–60.
8. Doyle, J. A. (1978) *Annu. Rev. Ecol. Syst.* **9,** 365–392.
9. Niklas, K. J. & Tiffney, B. H. (1994) *Philos. Trans. R. Soc. London B* **345,** 35–44.

10. Dilcher, D. L. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7030–7036.
11. Ricklefs, R. E. & Renner, S. S. (1994) *Evolution (Lawrence, Kans.)* **48,** 1619–1636.
12. Crepet, W. L. (1984) *Ann. Mo. Bot. Gard.* **71,** 607–630.
13. Eriksson, O. & Bremer, B. (1992) *Evolution (Lawrence, Kans.)* **46,** 258–266.
14. Herrera, C. M. (1989) *Am. Nat.* **133,** 309–322.
15. Smith, J. F. (2001) *Am. Nat.* **157,** 646–653.
16. Midgely, J. J. & Bond, W. J. (1991) *Philos. Trans. R. Soc. London B* **333,** 209–215.
17. Regal, P. J. (1997) *Science* **196,** 622–629.
18. Sanderson, M. J. & Donoghue, M. J. (1996) *Trends Ecol. Evol.* **11,** 15–20.
19. Dodd, M. E., Silvertown, J. & Chase, M. W. (1999) *Evolution (Lawrence, Kans.)* **53,** 732–744.
20. Barraclough, T. G. & Savolainen, V. (2001) *Evolution (Lawrence, Kans.)* **55,** 677–683.
21. Magallón, S. & Sanderson, M. J. (2001) *Evolution (Lawrence, Kans.)* **55,** 1762–1780.
22. Barraclough, T. G., Nee, S. & Harvey, P. H. (1998) *Evol. Ecol.* **12,** 751–754.
23. Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y. L., *et al.* (1993) *Ann. Mo. Bot. Gard.* **80,** 528–580.
24. Chase, M. W., Duvall, M. R., Hills, H. G., Conran, J. G., Cox, A. V., Caddick, L. R., Cameron, K. M. & Hoot, S. B. (1995) in *Monocotyledons: Systematics and Evolution*, eds. P. J. Rudall, P. J. Cribb, D. F. Cutler & Humphries, C. J. (Royal Botanic Gardens, Kew, U.K.), pp. 109–137.
25. Savolainen, V., Fay, M. F., Albach, D. C., Backlund, A., van der Bank, M., Cameron, K. M., Johnson, S. A., Lledo, M. D., Pintaud, J.-C., Powell, M., *et al.* (2000) *Kew Bull.* **55,** 257–309.
26. Savolainen, V., Chase, M. W., Hoot, S. B., Morton, C. M., Soltis, D. E., Bayer, C., Fay, M. F., De Bruijn, A. Y., Sullivan, S. & Qiu, Y. L. (2000) *Syst. Biol.* **49,** 306–362.
27. Soltis, P. S., Soltis, D. E. & Chase, M. W. (1999) *Nature* **402,** 402–404.
28. Soltis, D. E., Soltis, P. S., Chase, M. W., Mort, M. E., Albach, D. C., Zanis, M., Savolainen, V., Hahn, W. H., Hoot, S. B., Fay, M. F., *et al.* (2000) *Bot. J. Linn. Soc.* **133,** 381–461.
29. Graham, S. W. & Olmstead, R. G. (2000) *Am. J. Bot.* **87,** 1712–1730.
30. Qiu, Y.-L., Lee, J. H., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, P. S., Zanis, M., Zimmer, E. A., Chen, Z. D., Savolainen, V. & Chase, M. W. (1999) *Nature* **402,** 404–407.
31. Baum, B. R. (1992) *Taxon* **41,** 3–10.
32. Ragan, M. A. (1992) *Mol. Phylogenet. Evol.* **1,** 53–58.
33. Bininda-Emonds, O. R. P., Gittleman, J. L. & Steel, M. A. (2002) *Annu. Rev. Ecol. Syst.* **33,** 265–289.
34. Salamin, S., Hodkinson, T. R. & Savolainen, V. (2002) *Syst. Biol.* **51,** 136–150.
35. Purvis, A. (1995) *Philos. Trans. R. Soc. London B* **348,** 405–421.
36. Bininda-Emonds, O. R. P., Gittleman, J. L. & Purvis, A. (1999) *Biol. Rev. Cambridge Philos. Soc.* **74,** 143–175.
37. Angiosperm Phylogeny Group (1998) *Ann. Mo. Bot. Gard.* **85,** 531–553.
38. Angiosperm Phylogeny Group II (2003) *Bot. J. Linn. Soc.* **141,** 399–436.
39. Swofford, D. L. (2001) PAUP4.0b8 (Sinauer, Sunderland, MA).
40. Shimodaira, H. & Hasegawa, M. (1999) *Mol. Biol. Evol.* **16,** 114–116.
41. Savolainen, V., Chase, M. W., Salamin, N., Soltis, D. E., Soltis, P. S., López, A. J., Fédrigo, O. & Naylor, G. J. P. (2002) *Syst. Biol.* **51,** 638–647.
42. Sanderson, M. J. & Doyle, J. A. (2001) *Am. J. Bot.* **88,** 1499–1516.
43. Mathews, S. & Donoghue, M. J. (2000) *Int. J. Plant Sci.* **161,** S41–S55.
44. Zanis, M. J., Soltis, D. E., Soltis, P. S., Mathews, S. & Donoghue, M. J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6848–6853.
45. Qiu, Y.-L., Lee, J., Bernasconi-Quadroni, F., Soltis, D. E., Soltis, S. P., Zanis, M., Zimmers, E. A., Chen, Z., Savolainen, V. & Chase, M. W. (2000) *Int. J. Plant Sci.* **161,** S3–S27.
46. Sanderson, M. J. (1997) *Mol. Biol. Evol.* **14,** 1218–1231.
47. Wikström, N., Savolainen, V. & Chase, M. W. (2001) *Proc. R. Soc. London B* **268,** 2211–2220.
48. Drinnan, A. N., Crane, P. R. & Hoot, S. B. (1994) *Plant Syst. Evol.* **8,** 93–122 suppl.
49. Watson, L. & Dallwitz, M. J. (1991) *Australian Syst. Bot.* **4,** 681–695.
50. Dallwitz, M. J. (1980) *Taxon* **29,** 41–46.
51. Fusco, G. & Cronk, Q. C. B. (1995) *J. Theor. Biol.* **175,** 235–243.
52. Purvis, A., Katzourakis, A. & Agapow, P. M. (2001) *J. Theor. Biol.* **214,** 99–103.
53. Mooers, A. Ø., Page, R. D. M., Purvis, A. & Harvey, P. H. (1995) *Syst. Biol.* **44,** 332–342.
54. Heard, S. B. & Mooers, A. Ø. (1996) *Syst. Biol.* **45,** 115–118.
55. Slowinski, J. B. & Guyer, C. (1993) *Am. Nat.* **142,** 1019–1024.
56. Sanderson, M. J. & Donoghue, M. J. (1994) *Science* **264,** 1590–1593.
57. Huelsenbeck, J. P., Bull, J. J. & Cunningham, W. (1996) *Trends Ecol. Evol.* **11,** 152–158.
58. Soltis, D. E., Soltis, P. S., Mort, M. E., Chase, M. W., Savolainen, V., Hoot, S. B. & Morton, C. M. (1998) *Syst. Biol.* **47,** 32–42.
59. Soltis, P. S., Soltis, D. E., Savolainen, V., Crane, P. R. & Barraclough, T. G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 4430–4435.
60. Benton, M. J. & Ayala, F. J. (2003) *Science* **300,** 1698–1700.
61. Purvis, A. (1996) in *New Uses for New Phylogenies*, eds. Harvey, P. H., Brown, A. J. L., Smith, J. M. & Nee, S. (Oxford Univ. Press, Oxford), pp. 153–168.
62. Savolainen, V., Heard, S. B., Powell, M. P., Davies, T. J. & Mooers, A. Ø. (2002) *Syst. Biol.* **51,** 1–9.
63. Ricklefs, R. E. & Renner, S. S. (2000) *Evolution (Lawrence, Kans.)* **54,** 1061–1065.
64. Sims, H. J. & McConway, K. J. (2003) *Evolution (Lawrence, Kans.)* **57,** 460–479.
65. Chapman, G. P. (1996) *The Biology of Grasses* (CAB International, Oxon, U.K.).
66. Bremer, K. (2002) *Evolution (Lawrence, Kans.)* **56,** 1374–1387.

Rate heterogeneity among lineages of tracheophytes: Integration of molecular and fossil data and evidence for molecular living fossils

Pamela S. Soltis*†, Douglas E. Soltis‡, Vincent Savolainen§, Peter R. Crane¶, and Timothy G. Barraclough‖

*Florida Museum of Natural History and the Genetics Institute, and ‡Department of Botany and the Genetics Institute, University of Florida, Gainesville, FL 32611; §Molecular Systematics Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3DS, United Kingdom; ¶Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, United Kingdom; and ‖Department of Biology and Natural Environment Research Council Centre for Population Biology, Imperial College at Silwood Park, Ascot, Berkshire SL5 7PY, United Kingdom

Contributed by Peter R. Crane, February 12, 2002

Many efforts to date evolutionary divergences by using a molecular clock have yielded age estimates that are grossly inconsistent with the paleontological evidence. Such discrepancies often are attributed to the inadequacy of the fossil record, but many potential sources of error can affect molecular-based estimates. In this study, we minimize the potential error caused by inaccurate topology and uncertain calibration times by using a well-supported tree, multiple genes, and multiple well-substantiated dates to explore the correspondence between the fossil record and molecular-based age estimates for major clades of tracheophytes. Age estimates varied because of gene effects, codon position, lineage effects, method of inferring branch lengths, and whether or not rate constancy was assumed. However, even methods designed to ameliorate the effects of rate heterogeneity among lineages could not accommodate the substantially slower rates observed in *Marattia* + *Angiopteris* and in the tree ferns. Both of these clades of ferns have undergone dramatic decelerations in their rates of molecular evolution and are ''molecular living fossils,'' consistent with their relative morphological stasis for the past 165–200 million years. Similar discrepancies between the fossil record and molecular-based age estimates noted in other studies may also be explained in part by violations of rate constancy among lineages.

For nearly four decades, biologists have attempted to infer divergence dates from molecular data by using the concept of a molecular clock (1, 2). However, these efforts have met with only mixed success, as evidence for rate heterogeneity has accumulated (e.g., refs. 3–7), and as it has become clear that many estimated divergence times are grossly inconsistent with the fossil record (e.g., refs. 8–10). Although ''the clock'' has been known for some time to ''tick'' at different rates in different lineages and different genes, most studies that have used molecular data to estimate divergence times have neither considered potential sources of error or bias, nor provided confidence levels for the estimates reported. Furthermore, although the fossil record is typically regarded as sufficiently reliable to provide dates to calibrate the clock, when dates inferred from molecular data conflict with the fossil record, the latter is often dismissed as inadequate.

Many sources of error and bias can affect molecular-based estimates of divergence times. Obviously, an incorrect topology will yield erroneous estimates, although the magnitude of the problem depends on the extent of the topological error (11). Likewise, inaccurate calibration will bias the resulting estimates for other divergences. Equally seriously, however, heterogeneous rates of evolution among lineages are well known (3–7), and a failure to recognize such heterogeneity can compromise resulting estimates of divergence times. Inadequate sampling of taxa, coupled with rate heterogeneity, can compound the problem. For example, most molecular-based estimates of the age of the angiosperms greatly exceed the date inferred from the fossil record, 125–135 million years ago (mya). However, taxon sampling in these studies is skewed toward herbaceous species, especially grasses, which have elevated rates of molecular evolution relative to woody species (4). Estimates of divergence times may also vary among genes or other data partitions (e.g., 1st and 2nd vs. 3rd codon positions); such effects may be accommodated by different substitution models and should be evaluated in studies that combine multiple genes. A further key potential source of error or bias is the method used to estimate divergence dates. Although nearly any phylogram for any group of organisms clearly portrays violation of a molecular clock, with interspersed long and short branches, few studies that estimate divergence times test for clock-like behavior, and fewer still attempt to accommodate this violated assumption. Alternative methods, designed to accommodate rate inconstancy, have been proposed [e.g., nonparametric rate smoothing (NPRS; refs. 12 and 13); hidden-Markov methods (see ref. 13); likelihood methods (14, 15); Bayesian methods (16, 17); alternatives reviewed by Sanderson and Doyle (11)] but have rarely been tested, and their effectiveness is unknown.

Utilization of the fossil record also confronts many potential errors that could create problems in calibrating a molecular clock or for comparisons with molecular-based dates. Differing degrees of uncertainty in dating fossils is an inherent feature of the study of the geological record. The relevant fossils must also be accurately positioned on a cladogram of extant taxa based on synapomorphies. Further, it is also important that the date for the stem lineage of an extant group, which corresponds to the time a lineage diverged from its extant sister group, is not confused with the date for the crown group, which corresponds to the age of the extant group's most recent common ancestor. Molecular-based dates correspond to the ages of the crown groups, and thus it is critical that the fossils under consideration are also referable to the crown group. Finally, of course, fossils only provide minimum age estimates, and the fossil record inevitably incorporates many biases and real gaps.

Molecular-based estimates of divergence times in plants reveal a vast range of dates: for example, the age of the angiosperms has been estimated as 350–420 mya (18), >319 mya (8, 9), 200 mya (19, 20), 160 mya (7), to 140–190 mya (11). However, although some of these studies examined potential error caused by calibration time, lineage effects, or substitutional noise, only Sanderson and Doyle (11) thoroughly investigated multiple sources of error. Despite the uncertainties and the multiplicity of potential errors associated with molecular-based estimates of divergence times, the presence of large molecular data sets will

Abbreviations: mya, million years ago; NPRS, nonparametric rate smoothing; MP, maximum parsimony; ML, maximum likelihood; LR, likelihood ratio.

†To whom reprint requests should be addressed. E-mail: psoltis@flmnh.ufl.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

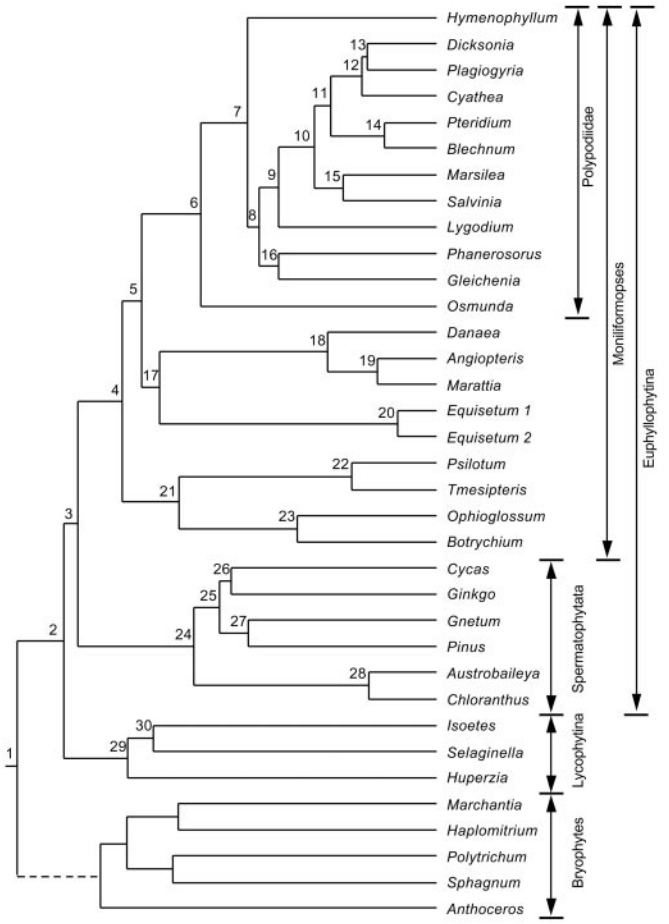


Fig. 1. ML tree from Pryer *et al.* (21) with outgroup monophyletic and nodes numbered.

continue to stimulate attempts to apply a molecular clock. Furthermore, even approximate age estimates for groups that lack a fossil record are better than none. Thus, it is imperative that such analyses are placed on the most secure foundation possible, consider potential sources of error, and determine the best methods to deal with realities such as rate heterogeneity among lineages or genes. The robustness of current methods to violations of their assumptions also needs careful examination.

In this article, we minimize errors of topology and calibration by using a well-supported tree that includes all major clades of tracheophytes and multiple strongly supported dates across a broad span of geologic time, to explore the correspondence between the fossil record and molecular-based age estimates. We also evaluate variation caused by method of estimation and calibration point. We estimate the ages, with confidence intervals, of major clades of tracheophytes, using (*i*) a tree based on four genes and morphology (21), (*ii*) data from four genes singly and combined, and (*iii*) multiple calibration points representing well-substantiated dates in the fossil record. We also test for rate constancy among lineages and heterogeneity of rates among genes and codon positions in protein-coding genes and evaluate sensitivity of age estimates to branch lengths inferred by maximum parsimony (MP) and maximum likelihood (ML), the effectiveness of NPRS relative to an assumption of rate constancy, and the correspondence of molecular-based estimates to the fossil record.

Materials and Methods

Topology and Calibration Points. The phylogenetic tree of tracheophytes used in this study is the ML tree of Pryer *et al.* (21) inferred from analysis of the plastid genes *rbcL*, *atpB*, and *rps4* and nuclear 18S ribosomal DNA (Fig. 1); a nearly identical tree was obtained in MP analyses of these four genes plus 136 morphological characters. This tree shows strong support for three major clades—lycophytes, seed plants, and horsetails + ferns (Moniliformopses)—with equally strong support for most relationships within each of these clades. Pryer *et al.*'s tree shows a basal polytomy with the interrelationships of hornworts, liverworts + mosses, and tracheophytes unresolved. Because dichotomous branching at the base of the tree is required for computation of likelihoods under the assumption of a molecular clock, we made the bryophyte outgroup monophyletic, with hornworts sister to liverworts + mosses, in some analyses. Using MACCLADE version 3.05 (22), we tested the impact of relationships among outgroups on the estimates of divergence times in the ingroup by rearranging the outgroups to conform to the following topologies: hornworts, liverworts, mosses, tracheophytes [abbreviated HLM; consistent with analyses of land plant relationships based on 18S rDNA (e.g., refs. 23 and 24]); liverworts, hornworts, mosses, tracheophytes [abbreviated LHM; consistent with analyses of morphology (25, 26), the distribution of introns (27), and some DNA sequence data sets (e.g., ref. 28)]; basal polytomy, as reported by Pryer *et al.* (21).

The dates for the calibration points used in this paper are based on the time scale of Harland *et al.* (29). Ages of clades are minimum ages estimated conservatively for the crown group by the first appearance of fossils clearly referable to one of the constituent lineages based on morphological synapomorphies. For example, although the time of origin of the angiosperms is unclear, the dates selected as calibration points correspond to fossils that are clearly angiosperms and are thus conservative. Further justification for the dates used is provided in the *Appendix*, which is published as supporting information on the PNAS web site, www.pnas.org. Four different calibration points were used, and in two cases a more conservative and less conservative calibration were used to explore the potential effects.

Tests of Rate Heterogeneity. In all ML analyses, we used an HKY85 model of DNA evolution (30) in which we estimated base frequencies and the transition/transversion ratio from the data; to account for rate heterogeneity among sites, we used a gamma distribution (31) with the alpha shape parameter esti-

Table 1. Parameter values for the HKY85 + Γ model on the tracheophyte tree, with outgroup monophyletic

Gene	Frequency of adenine	Frequency of cytosine	Frequency of guanine	Frequency of thymine	Transit./ Transv. ratio	Alpha, shape parameter
rbcL	0.283	0.219	0.182	0.317	3.946	0.233
atpB	0.305	0.198	0.156	0.341	5.303	0.288
rps4	0.345	0.192	0.201	0.261	3.354	0.779
18S rDNA	0.209	0.251	0.255	0.285	2.586	0.182
Combined	0.277	0.221	0.206	0.296	3.948	0.233

Table 2. LR tests of lineage effects, based on a χ^2 distribution and 33 df

Genes	-ln likelihood without molecular clock enforced	-ln likelihood with molecular clock enforced	Λ	P
rbcL	14,619.9	15,773.2	2,307	<0.0001
atpB	13,533.5	13,826.0	585	<0.0001
rps4	10,775.2	11,045.8	541	<0.0001
18S rDNA	9,053.2	9,249.4	392	<0.0001
Combined	43,136.6	43,752.5	1,232	<0.0001

Table 3. LR tests of gene effects, computed without (−cl) and with (+cl) enforcing a molecular clock, based on a χ^2 distribution and 36 df

Genes	Λ (−cl)	Λ (+cl)
rbcL vs. *atpB*	152	89
rbcL vs. *rps4*	297	203
rbcL vs. 18S rDNA	38,006	38,643
atpB vs. *rps4*	3,756	3,830
atpB vs. 18S rDNA	1,064	846
rps4 vs. 18S rDNA	1,042	840

All values are significant at $P < 0.0001$.

mated from the data (Table 1). Although we did not test the HKY85 + Γ model against alternative models, this model represents a reasonable compromise between generality of the model and computational time required.

For each gene taken separately and all genes combined, rate heterogeneity across lineages was tested by using a likelihood ratio (LR) test (32). Significance was assessed by comparing $\Lambda = -2 \log \mathrm{LR}$, where LR is the difference between the -ln likelihood of the tree, with and without enforcing a molecular clock, with a χ^2 distribution (with $n - 2$ degrees of freedom, where n is the number of taxa).

Rate heterogeneity between pairs of data partitions (genes or codon positions) was also tested by using a LR test: LR = [ln L − (ln L1 + ln L2)], where L1 is the likelihood of the tree with one partition, L2 is the likelihood of the tree with the second partition, and L is the likelihood of the tree with both partitions combined. The test statistic Λ was compared with a χ^2 distribution, with degrees of freedom computed following Sanderson and Doyle (11). Rate heterogeneity among partitions was assessed with and without enforcing a molecular clock. In the tests without a molecular clock enforced, likelihoods were computed on the tree of Pryer *et al.* (21); however, the basal polytomy of this tree precluded the computation of likelihood values on this tree under the assumption of a molecular clock. Therefore, tests of rate heterogeneity among genes with a molecular clock enforced used the tree with outgroups monophyletic (see above).

Estimation of Ages. Because all tests of rate heterogeneity among lineages were highly significant (Table 2), we dated the nodes by using the NPRS method of Sanderson (12). Using PAUP* 4.0 (33), we calculated MP and ML branch lengths when single genes, or all combined, were optimized onto the tree of Pryer *et al.* (21). These trees with branch lengths were then transformed into ultrametric trees by using the NPRS method implemented in the software TREEEDIT (version 1.0 alpha 4–61, August 2000, written by Andrew Rambaut and Mike Charleston and available at http://evolve.zoo.ox.ac.uk/software/TreeEdit/main.html). To transform relative time to absolute ages we calibrated the trees by using dates from the fossil record. To compute error estimates for the ages inferred from single genes or all combined, we reapplied the NPRS procedure to 100 bootstrapped matrices obtained by resampling the data irrespective of codon position by using PHYLIP 3.573c (34).

Results and Discussion

Tests of Rate Heterogeneity. All genes, separate and combined, show significant rate heterogeneity among lineages (Table 2). Furthermore, all pairs of genes evolve at significantly different rates across this tree (Table 3), whether or not a molecular clock is enforced. The relative rates of evolution of the four genes are *rps4* > *atpB* ≈ *rbcL* > 18S rDNA. With the exception of 1st versus 2nd codon positions in *rps4*, computed with a molecular clock, all codon positions evolve at significantly different rates in the three protein-coding genes, whether or not a molecular clock is enforced (Table 7, which is published as supporting information on the PNAS web site).

Comparison of Estimates from Different Partitions. Age estimates varied considerably among genes (Table 4). For example, considering estimates only for node 2 (tracheophytes), when 125 mya was used as a conservative calibration point for node 28 (angiosperms), values for node 2 using MP ranged from 414.3 mya (*rps4*) to 513.2 mya (18S rDNA), and using ML ranged from 490.5 mya (*rps4*) to 680.3 mya (18S rDNA). The plastid gene *rps4* typically yielded the youngest age estimates for a given node, followed in order of increasing age by *atpB* and *rbcL*, with the oldest age estimates consistently provided by the only nuclear gene, 18S rDNA. However, deviations from this general pattern were observed for some nodes (e.g., nodes 22, 24, 25, and 26) for which *rbcL* or *atpB* provided the oldest age estimates (Table 4). The standard deviations for all estimates are also high (Table 4), for individual genes and for the combined matrix. Thus, considerable variance surrounds each age estimate.

Age estimates also varied dramatically by codon position (Table 8, which is published as supporting information on the PNAS web site). For example, when 125 mya was used for node 28 (angiosperms), values for node 2 (age of tracheophytes) using MP ranged from 403.8 (3rd position) to 814.9 (1st) for *atpB*, from 193.6 (2nd) to 506.8 (3rd) for *rbcL*, and from 361.4 (3rd) to 517.5 (2nd) for *rps4*. ML values for 3rd positions were generally older than MP estimates, often nearly twice as old (Table 8), suggesting that multiple substitutions may have occurred at some 3rd positions. Age estimates obtained by using a calibration of 377.4 mya for node 29 (lycophytes) showed similar patterns among codon positions but different ages (data not shown).

Comparison of Estimates from Different Methods. ML and MP age estimates differed greatly (Table 4), with the ML estimates considerably older than those obtained with MP for all data partitions; the MP estimates agree more closely with the fossil record (Table 4). Because ML corrects for multiple substitutions, ML estimates may be expected to be older than MP estimates, but the ML estimates for most nodes are clearly inconsistent with the fossil record.

Because significant lineage effects were detected, we used Sanderson's (12) NPRS method to ameliorate rate differences among clades. This method estimates rates and divergence times by using a criterion that maximizes the autocorrelation of rates within clades. However, the effectiveness of this approach for accommodating rate inconstancy has not been tested, and Sanderson and Doyle's (11) preliminary analyses with angiosperms suggest that NPRS may actually aggravate rather than ameliorate the problem, at least when rates of molecular evolution change abruptly.

To examine the effects of using NPRS, we compared the results obtained with NPRS to those obtained with the widely used approach of defining the relative age of a node in a nonultrametric tree as the maximum branch length from that

Table 4. Ages of nodes with standard deviations from bootstrapped matrices inferred from the optimization of single genes or the combined data set using MP (upper values) or ML (lower values), an estimated age for node 28 of 125 mya as a calibration point, and the tracheophyte tree of Pryer *et al.* (21) with the outgroup specified as monophyletic (see Fig. 1 for node numbers and text for details)

Node	*rbcL*	*atpB*	*rps4*	18S rDNA	Combined	Direct
1	541.6 ± 75.7	506.0 ± 93.4	459.6 ± 92.2	599.4 ± 89.3	546.8 ± 44.0	545.0
	651.0 ± 114.2	662.5 ± 170.4	496.8 ± 165.7	683.3 ± 167.9	716.8 ± 72.0	696.0
2	497.2 ± 69.6	462.0 ± 84.0	414.3 ± 83.5	513.2 ± 80.4	495.9 ± 39.3	493.7
	646.3 ± 114.0	654.8 ± 168.3	490.5 ± 163.2	680.3 ± 167.5	710.1 ± 71.5	688.7
3	460.4 ± 63.0	430.6 ± 79.2	388.6 ± 77.2	481.9 ± 74.9	460.9 ± 36.0	458.0
	612.0 ± 106.3	645.6 ± 165.6	474.8 ± 159.1	621.0 ± 151.5	683.4 ± 67.5	661.6
4	401.9 ± 56.2	379.3 ± 71.5	329.4 ± 66.7	428.9 ± 72.2	398.4 ± 32.9	395.3
	535.1 ± 91.3	568.1 ± 148.6	409.9 ± 140.9	616.9 ± 150.2	600.0 ± 59.9	579.7
5	357.1 ± 50.8	336.9 ± 65.8	297.5 ± 60.6	398.7 ± 66.6	354.4 ± 29.9	350.9
	514.7 ± 91.0	542.1 ± 144.1	374.2 ± 129.0	616.5 ± 150.3	567.8 ± 58.4	543.7
6	321.2 ± 45.6	272.0 ± 54.3	245.6 ± 49.1	340.4 ± 59.7	297.1 ± 26.6	293.6
	442.3 ± 77.6	432.3 ± 118.0	301.3 ± 104.7	568.1 ± 149.0	453.8 ± 49.3	435.0
7	265.2 ± 37.7	217.9 ± 43.7	192.4 ± 38.2	311.5 ± 54.3	238.8 ± 21.5	235.6
	370.5 ± 65.9	329.2 ± 87.1	253.6 ± 91.9	568.1 ± 149.0	366.3 ± 39.7	350.8
8	215.6 ± 30.7	189.4 ± 37.9	157.3 ± 32.9	271.9 ± 50.6	198.2 ± 18.3	195.2
	343.1 ± 59.0	311.2 ± 90.0	240.9 ± 91.3	563.6 ± 150.0	347.1 ± 36.8	328.0
9	172.6 ± 24.2	160.7 ± 32.5	132.4 ± 28.5	247.1 ± 51.4	163.7 ± 15.1	161.4
	298.0 ± 55.2	280.6 ± 73.0	228.0 ± 86.9	523.4 ± 144.6	313.0 ± 35.0	292.1
10	131.3 ± 18.8	122.0 ± 24.5	97.0 ± 22.2	224.6 ± 48.2	123.3 ± 11.8	121.5
	240.3 ± 41.7	214.0 ± 55.7	169.2 ± 66.5	505.0 ± 138.7	240.1 ± 27.3	225.9
11	98.5 ± 15.5	100.0 ± 20.0	82.4 ± 19.5	197.1 ± 46.8	97.6 ± 9.4	95.9
	208.0 ± 38.1	195.3 ± 49.2	147.8 ± 58.0	465.7 ± 123.6	210.5 ± 23.2	195.2
12	63.7 ± 10.1	61.3 ± 13.4	54.2 ± 14.6	155.8 ± 49.3	61.8 ± 6.8	60.8
	150.9 ± 31.6	138.7 ± 42.0	124.3 ± 47.4	465.7 ± 123.6	155.4 ± 21.4	139.9
13	51.0 ± 8.6	54.9 ± 11.4	42.5 ± 11.9	155.8 ± 49.3	51.0 ± 6.3	50.2
	129.4 ± 30.0	134.6 ± 40.7	115.2 ± 43.5	465.7 ± 123.6	142.5 ± 21.2	127.7
14	57.6 ± 9.9	48.9 ± 10.4	43.5 ± 12.0	197.1 ± 46.8	52.6 ± 5.4	51.6
	116.0 ± 26.4	84.2 ± 24.1	70.7 ± 32.7	465.7 ± 123.6	105.4 ± 13.9	98.1
15	96.7 ± 13.8	90.7 ± 20.2	71.4 ± 17.6	175.9 ± 38.3	91.1 ± 9.7	89.4
	169.2 ± 33.6	160.4 ± 46.7	157.3 ± 65.5	435.4 ± 127.8	185.3 ± 24.2	173.2
16	174.3 ± 24.9	151.2 ± 32.0	129.6 ± 27.9	256.9 ± 49.9	160.8 ± 15.0	158.3
	305.3 ± 54.3	289.0 ± 80.1	217.7 ± 85.9	556.5 ± 144.1	312.5 ± 35.2	293.2
17	297.3 ± 43.1	293.8 ± 57.6	248.9 ± 53.2	381.2 ± 62.4	299.8 ± 26.0	297.5
	487.7 ± 91.7	487.0 ± 133.8	368.7 ± 127.9	615.6 ± 150.5	539.0 ± 54.5	511.8
18	128.3 ± 21.0	135.1 ± 34.1	116.7 ± 29.4	269.9 ± 55.8	135.4 ± 15.4	132.4
	181.4 ± 37.6	221.1 ± 65.6	141.0 ± 59.4	425.1 ± 122.2	214.1 ± 30.7	202.1
19	59.3 ± 13.1	81.1 ± 24.0	48.2 ± 16.4	102.8 ± 36.5	64.2 ± 9.2	62.2
	113.9 ± 33.0	216.5 ± 65.7	62.6 ± 40.2	133.5 ± 51.0	122.7 ± 25.7	111.4
20	64.8 ± 12.7	62.1 ± 14.5	48.9 ± 14.3	114.5 ± 67.9	61.5 ± 6.7	60.3
	70.7 ± 18.2	56.3 ± 17.8	49.8 ± 23.7	127.6 ± 80.0	71.9 ± 9.3	73.0
21	334.2 ± 50.6	312.8 ± 58.5	252.5 ± 52.5	369.5 ± 67.6	324.0 ± 27.8	320.6
	444.5 ± 81.3	450.9 ± 120.3	352.8 ± 126.5	509.3 ± 132.1	494.4 ± 53.8	475.2
22	155.7 ± 35.6	122.1 ± 28.6	106.9 ± 29.6	133.7 ± 52.6	131.8 ± 15.3	131.5
	180.0 ± 52.5	132.4 ± 42.0	107.8 ± 47.1	148.1 ± 67.1	158.9 ± 22.2	157.6
23	175.1 ± 27.1	194.2 ± 37.3	146.7 ± 33.2	236.0 ± 58.2	183.6 ± 18.3	182.0
	209.8 ± 40.4	256.5 ± 71.7	239.9 ± 94.8	359.4 ± 113.5	269.2 ± 33.5	258.5
24	352.0 ± 47.6	322.8 ± 58.8	338.2 ± 67.6	304.2 ± 46.1	343.7 ± 25.3	340.2
	459.3 ± 76.0	457.0 ± 117.6	409.8 ± 129.7	386.0 ± 93.3	465.4 ± 44.8	447.8
25	282.8 ± 40.7	259.4 ± 49.1	265.2 ± 53.4	261.5 ± 41.4	277.5 ± 21.2	274.0
	402.1 ± 74.6	418.6 ± 107.1	372.9 ± 122.4	376.8 ± 91.6	424.5 ± 41.3	402.2
26	247.3 ± 38.2	234.4 ± 46.7	229.4 ± 50.4	234.9 ± 37.1	246.9 ± 19.7	242.8
	387.3 ± 72.0	387.6 ± 93.0	360.2 ± 120.3	350.8 ± 82.6	401.7 ± 40.0	378.8
27	227.0 ± 32.3	205.5 ± 40.1	239.2 ± 50.7	190.3 ± 34.3	222.9 ± 18.1	219.7
	323.7 ± 59.2	376.0 ± 99.1	372.9 ± 122.4	298.0 ± 76.6	373.0 ± 36.8	346.8
28	125, 125	125, 125	125, 125	125, 125	125, 125	125, 125
29	365.7 ± 52.0	364.5 ± 70.4	289.1 ± 61.2	422.2 ± 73.3	374.7 ± 32.2	372.7
	563.1 ± 108.3	564.0 ± 150.0	385.1 ± 133.4	605.4 ± 151.7	599.1 ± 67.3	568.0
30	277.4 ± 42.6	310.3 ± 59.9	227.4 ± 50.5	291.2 ± 55.1	294.3 ± 26.3	292.7
	496.5 ± 103.1	538.7 ± 147.9	373.2 ± 134.0	508.7 ± 127.5	557.0 ± 64.0	522.1

''Direct'' age estimates are NPRS estimates computed directly from the tree by using the original combined data set rather than the bootstrapped matrices.

node to any of the tips descended from it (see e.g., refs. 13, 35, and 36). This approach was repeated with both MP and ML branch lengths for all genes combined. With all six calibrations examined, the estimated ages were clearly anomalously old, with all but one estimate for the age of tracheophytes ranging from 1.0 billion to 4.0 billion years and all but one estimate for the age of land plants ranging from 1.1 billion to 4.3 billion years (data not shown). The earliest fossil record of probable embryophytes (a more inclusive clade that comprises tracheophytes plus bryophytes) is from the Middle Ordovician (Llanvirn 476.1–472.7

Table 5. Effect of outgroup topology on ages (in mya) inferred for selected nodes, using MP, the combined data set, and the angiosperm (node 28) calibration point of 125 mya

Outgroup topol.	Node 2	Node 29	Node 3	Node 4	Node 24
Basal polytomy	497	372	460	397	341
Monophyletic	494	373	456	395	340
HLM	496	384	458	398	338
LHM	493	378	458	398	338

Node 2, tracheophytes, Node 29, lycophytes, Node 3, euphyllophytes, Node 4, moniliforms, Node 24, seed plants. See text for details on topologies.

mya). The more ''reasonable'' estimates, of 740 mya for tracheophytes and 794 mya for land plants, both of which are approximately 350 million years older than the fossil record, came from the calibration of lycophytes at 400 mya. However, use of this calibration point resulted in estimates for other lineages, such as the angiosperms (56 mya), *Marattia* + *Angiopteris* (38 mya), and tree ferns (127 mya), much younger than the fossil record (conservatively 125, 166.1, and 166.1 mya, respectively). Estimates obtained by using the ML branch lengths were even more problematic: the older dates were far older, and the younger dates were far younger. The use of NPRS, although not sufficient to account for all rate heterogeneity among lineages, certainly brought at least some estimated ages into line with the fossil record.

Effects of Outgroup Topology on Estimates. The effects of outgroup topology were ascertained through comparisons of divergence estimates by using MP branch lengths and the calibration point of 125 mya for the angiosperms. The most severe effects were at the basal tracheophyte nodes, although the differences were no more than 3 mya or 4 mya, except for the lycophyte dates, which differed by 12 mya (Table 5). Effects at more internal and terminal nodes were minimal, with differences of 1–2 mya (data not shown). Thus, as Sanderson and Doyle (11) found in their analysis of the age of the angiosperms, relationships among the outgroups have surprisingly little effect, even on basal nodes of the ingroup.

Comparison of Estimates from Different Calibration Points. The use of different calibration points had a major impact on the age estimates for nodes (Table 6). For example, when node 28 (angiosperms) was used and a conservative age estimate of 125 mya used, estimates for the ages of node 15 (Marsileales + Salviniales), node 25 (gymnosperms), and node 29 (lycophytes) agree reasonably closely with estimates from the fossil record (89.8 mya vs. 90 mya; 274.5 mya vs. 290 mya; 371.6 mya vs. 400 or 377.4 mya, respectively). However, the estimate for node 6 (*Osmunda* + all other leptosporangiate ferns) is 294.7 mya, which is somewhat older than the first appearance of the crown group Polypodiidae in the Late Permian (255–230 mya), and the estimate for node 2 (tracheophytes) extends back to the Cambrian whereas there is no reliable fossil evidence for the group until the Late Silurian (Ludlovian, *ca.* 415 mya). In contrast, the estimates for node 12 (tree ferns) and node 19 (*Marattia* + *Angiopteris*) are considerably younger than the fossil record indicates (61.0 mya vs. 166.1 mya; 62.4 mya vs. 166.1 mya, respectively). Even under the most conservative interpretation, there is no doubt that Dicksoniaceae, *Angiopteris*, and *Marattia* are all present in the Middle Jurassic flora of Yorkshire, northern England (37, 38).

Table 6. Ages of nodes inferred from the optimization of the combined data set using MP, the tracheophyte tree of Pryer *et al.* (21) with a basal polytomy, and various nodes as calibration points (see Fig. 1 for node numbers and text for details)

Node	Node 28, 125 mya	Node 28, 131.8 mya	Node 12, 166.1 mya	Node 19, 166.1 mya	Node 25, 290 mya	Node 29, 377.4 mya	Node 29, 400 mya
1	581.3, *545.0*	612.9	1,581.3	1,546.5	614.2	590.2	625.6
2	497.3, *493.7*	524.3	1,352.8	1,323.0	525.4	505.0	535.2
3	460.4, *458.0*	485.4	1,252.5	1,225.0	486.5	467.5	495.5
4	397.0, *395.3*	418.6	1,080.2	1,064.4	419.5	403.2	427.4
5	352.3, *350.9*	371.6	958.6	937.4	372.3	357.8	379.2
6	294.7, *293.6*	310.8	801.8	784.1	311.4	299.3	317.2
7	236.5, *235.6*	249.3	643.3	629.2	249.9	240.1	254.5
8	196.0, *195.2*	206.8	533.2	521.5	207.1	199.1	211.0
9	162.0, *161.4*	170.8	440.9	431.1	171.2	164.5	174.4
10	121.9, *121.5*	128.6	331.7	324.4	128.8	123.8	131.2
11	96.3, *95.9*	101.5	262.0	256.3	101.8	97.8	103.7
12	61.0, *60.8*	64.3	165.3	162.4	64.5	62.0	65.7
13	50.4, *50.2*	53.1	137.1	134.1	53.2	51.1	54.2
14	51.8, *51.6*	54.7	141.1	138.0	54.8	52.6	55.8
15	89.8, *89.4*	94.7	244.3	238.9	94.9	91.1	96.6
16	159.0, *158.3*	167.6	432.4	422.9	168.0	161.4	171.1
17	298.6, *297.5*	314.8	812.5	794.6	315.5	303.2	321.4
18	132.9, *132.4*	140.1	361.6	353.6	140.4	134.9	143.0
19	62.4, *62.2*	65.8	169.8	166.1	66.0	63.4	67.2
20	60.5, *60.3*	63.8	164.6	160.9	63.9	61.4	65.1
21	321.8, *320.6*	339.3	875.4	856.0	340.0	326.7	346.3
22	131.9, *131.5*	139.0	358.8	350.9	139.3	133.9	141.9
23	182.6, *182.0*	192.5	496.8	485.8	192.9	185.4	196.5
24	341.0, *340.2*	359.6	927.8	907.4	360.4	346.4	367.1
25	274.5, *274.0*	289.4	746.7	730.2	290	278.7	295.4
26	243.1, *242.8*	256.4	661.4	646.8	256.9	246.9	261.7
27	220.0, *219.7*	232.0	598.4	585.3	232.4	223.4	236.8
28	125, *125*	131.8	330.7	332.6	132.1	126.9	134.5
29	371.6, *372.7*	391.9	1,011.1	988.8	392.7	377.4	400
30	291.2, *292.7*	307.0	792.3	774.8	307.7	295.7	313.4

Node 28, angiosperms; node 12, *Dicksonia*/*Plagiogyria*/*Cyathea*; node 19, *Angiopteris*/*Marattia*; node 25, gymnosperms; node 29 lycopsids. For the angiosperm calibration at 125 mya, values computed with the outgroup specified as monophyletic are in italics.

We obtained very similar results when dates for either node 29 (lycophytes) or node 25 (gymnosperms) were used as calibration points. The estimates for the age of seed plants, gymnosperms, and angiosperms (in the former) and estimates for the age of lycophytes and angiosperms (in the latter) agreed closely with the fossil record, whereas the estimates for the ages of both node 12 (tree ferns) and node 19 (*Marattia* + *Angiopteris*) were again very low compared with the fossil record. In contrast, however, when fossil dates for nodes 12 (tree ferns) and 19 (*Marattia* + *Angiopteris*) were used as calibration points, the estimates for all other nodes became anomalously old (Table 6).

Conclusions

We detected significant rate heterogeneity among lineages of land plants and among genes, even those from the plastid genome. Age estimates based on techniques that assume rate constancy among lineages are highly skewed, with most basal nodes being several hundred million years too old and some internal and terminal nodes being much too young, based on interpretations of the fossil record. NPRS provides estimates that are much more in line with the known history of life on earth. However, NPRS cannot accommodate all of the lineage effects, and age estimates vary substantially depending on the calibration point used.

Estimates of ages for clades of seed plants and lycophytes are reasonably consistent with each other, and with the fossil record, when other seed plant or lycophyte nodes are used for calibration. However, ages for several fern groups inferred from calibrations using seed plants or lycophytes are much too young compared with their unequivocal fossil record. Even when very conservative fern fossil dates are used to estimate the ages of seed plants and lycophytes, the results are strongly at odds not just with paleobotanical data but the whole corpus of geochronological knowledge. Our interpretation is that some clades, notably *Marattia* + *Angiopteris* and the tree ferns, have apparently experienced a dramatic slowdown in their rates of molecular evolution. This pattern cannot be an artifact of insufficient sampling of ferns: all extant members of the (*Marattia* + *Angiopteris*) + *Danaea* clade were included in the Pryer *et al.* (21) tree. Likewise, the tree fern clade is also well sampled, and the overall backbone of the clade of leptosporangiate ferns (Polypodiidae) is also well represented.

Marattia, *Angiopteris*, and the tree ferns are "molecular living fossils," consistent with their relatively stable morphologies through time. Two clades of angiosperms with good fossil records have also been considered molecular living fossils: *Nelumbo* + *Platanus* and *Fagus* + *Carya* (11). The correspondence between relative stasis in morphological features and relative stasis in gene sequences indicates that, in some cases and in broad terms, the genome may evolve as a unit over long periods. At least in angiosperm families, the rate of morphological evolution correlates with the rate of neutral molecular substitutions (39). This pattern stands in stark contrast to that observed for many angiosperm groups that have radiated recently on oceanic islands and exhibit extensive morphological divergence with minimal molecular evolution (40).

We thank Kathleen Pryer and collaborators for sharing their phylogenetic tree and data sets with us before the publication of their paper, Mike Sanderson and Jim Doyle for sharing their unpublished work and for valuable discussion and advice, and Mike Sanderson for helpful comments on the manuscript. This work was supported in part by a U.S.-U.K. Fulbright Distinguished Professorship (to P.S.S. and D.E.S.), National Science Foundation Grant DEB-0090283 (to D.E.S., P.S.S., D. L. Dilcher, and P. S. Herendeen), a Swiss National Science Foundation grant (to V.S.), and a Royal Society University Research Fellowship (to T.G.B.).

1. Zuckerkandl, E. & Pauling, L. (1962) in *Horizons in Biochemistry*, eds. Kasha, M. & Pullman, B. (Academic, New York), pp. 189–225.
2. Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins*, eds. Bryson, V. & Vogel, H. J. (Academic, New York), pp. 97–166.
3. Britten, R. J. (1986) *Science* **231,** 1393–1398.
4. Gaut, B. S., Muse, S. V., Clark, W. D. & Clegg, M. T. (1992) *J. Mol. Evol.* **35,** 292–303.
5. Clegg, M. T., Gaut, B. S., Learn, G. H., Jr. & Morton, B. R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 6795–6801.
6. Li, W.-H. (1997) *Molecular Evolution* (Sinauer, Sunderland, MA).
7. Goremykin, V., Hansmann, S. & Martin, W. (1997) *Plant Syst. Evol.* **206,** 337–351.
8. Martin, W., Gierl, A. & Saedler, H. (1989) *Nature (London)* **339,** 46–48.
9. Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H. & Cerff, R. (1993) *Mol. Biol. Evol.* **10,** 140–162.
10. Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L. & Hedges, S. B. (2001) *Science* **293,** 1129–1133.
11. Sanderson, M. J. & Doyle, J. A. (2001) *Am. J. Bot.* **88,** 1499–1516.
12. Sanderson, M. J. (1997) *Mol. Biol. Evol.* **14,** 1218–1231.
13. Sanderson, M. J. (1998) in *Molecular Systematics of Plants II*, eds. Soltis, D. E., Soltis, P. S. & Doyle, J. J. (Kluwer, Boston), pp. 242–264.
14. Rambaut, A. E. & Bromham, L. D. (1998) *Mol. Biol. Evol.* **15,** 442–448.
15. Yoder, A. D. & Yang, Z. (2000) *Mol. Biol. Evol.* **17,** 1081–1090.
16. Thorne, J. L., Kishino, H. & Painter, I. S. (1998) *Mol. Biol. Evol.* **15,** 1647–1657.
17. Huelsenbeck, J. P., Larget, B. & Swofford, D. L. (2000) *Genetics* **154,** 1879–1892.
18. Ramshaw, J. A. M., Richardson, D. L., Meatyard, B. T., Brown, R. H., Richardson, M., Thompson, E. W. & Boulter, D. (1972) *New Phytol.* **71,** 773–779.
19. Wolfe, K. H., Gouy, M., Yang, Y.-W., Sharp, P. M. & Li, W.-H. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6201–6205.
20. Larouche, J., Li, P. & Bousquet, J. (1995) *Mol. Biol. Evol.* **12,** 1151–1156.
21. Pryer, K. M., Schneider, H., Smith, A. M., Cranfill, R., Wolf, P. G., Hunt, J. S. & Sipes, S. D. (2001) *Nature (London)* **409,** 545–648.
22. Maddison, W. P & Maddison, D. R. (1992) MACCLADE (Sinauer, Sunderland, MA), version 3.05.
23. Nickrent, D. L., Parkinson, C. L., Palmer, J. D. & Duff, R. J. (2000) *Mol. Biol. Evol.* **17,** 1885–1895.
24. Soltis, P. S., Soltis, D. E., Wolf, P. G., Nickrent, D. L., Chaw, S.-M. & Chapman, R. L. (1999) *Mol. Biol. Evol.* **16,** 1774–1784.
25. Mishler, B. D.& Churchill, S. P. (1984) *Brittonia* **36,** 406–424.
26. Kenrick, P. & Crane, P. R. (1997) *The Origin and Diversification of Land Plants: A Cladistic Study* (Smithsonian, Washington, DC).
27. Qiu, Y.-L., Cho, Y., Cox, J. C. & Palmer, J. D. (1998) *Nature (London)* **394,** 671–674.
28. Lewis, L. A., Mishler, B. D. & Vilgalys, R. (1997) *Mol. Phylogenet. Evol.* **7,** 377–393.
29. Harland, W. B., Armstrong, R. L., Cox, A. V., Craig, L. E., Smith, A. G. & Smith, D. G. (1989) *A Geologic Timescale* (Cambridge Univ. Press, Cambridge, U.K.).
30. Hasegawa, M., Kishino, H. & Yano, T. (1985) *J. Mol. Evol.* **21,** 160–174.
31. Yang, Z. (1993) *Mol. Biol. Evol.* **10,** 1396–1401.
32. Felsenstein, J. (1988) *Annu. Rev. Genet.* **22,** 521–565.
33. Swofford, D. L. (1998) PAUP* 4.0: *Phylogenetic Analysis using Parsimony* (* and Other Methods) (Sinauer, Sunderland, MA).
34. Felsenstein, J. (1993) PHYLIP: *Phylogeny Inference Package* (Univ. of Washington, Seattle).
35. Hillis, D. M., Mable, B. K. & Moritz, C. (1996) in *Molecular Systematics*, eds. Hillis D. M., Moritz, C. & Mable, B. K. (Sinauer, Sunderland, MA), 2nd Ed., pp. 515–544.
36. Bremer, K. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 4707–4711.
37. Harris, T. M. (1961) *The Yorkshire Jurassic Flora: I. Thallophyta-Pteridophyta* (British Museum Natural History, London).
38. Hill, C. R. (1987) *Rev. Paleobot. Palynol.* **51,** 65–93.
39. Barraclough, T. G. & Savolainen, V. (2001) *Evolution* **55,** 677–683.
40. Baldwin, B. G., Crawford, D. J., Francisco-Ortega, J., Kim, S.-C., Sang, T. & Stuessy, T. (1998) in *Molecular Systematics of Plants II*, eds. Soltis, D. E., Soltis, P. S. & Doyle, J. J. (Kluwer, Boston), pp. 410–441.

The Role of Phylogenetics in Comparative Genetics[1]

Douglas E. Soltis* and Pamela S. Soltis

Department of Botany and the Genetics Institute, University of Florida, Gainesville, Florida 32611; and Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, Florida 32611

WHY PHYLOGENY MATTERS

Many biologists agree that a phylogenetic tree of relationships should be the central underpinning of research in many areas of biology. Comparisons of plant species or gene sequences in a phylogenetic context can provide the most meaningful insights into biology. This important realization is now apparent to researchers in diverse fields, including ecology, molecular biology, and physiology (see recent papers in *Plant Physiology*, e.g. Hall et al., 2002a; Doyle et al., 2003). Examples of the importance of a phylogenetic framework to diverse areas of plant research abound (for review, see Soltis and Soltis, 2000; Daly et al., 2001). One obvious example is the value of placing model organisms in the appropriate phylogenetic context to obtain a better understanding of both patterns and processes of evolution. The fact that tomato (*Lycopersicon esculentum*) and other species of this small genus actually are embedded within a well-marked subclade *Solanum* (and, hence, are more appropriately referred to as species of *Solanum*; tomato has been renamed as of *Solanum lycopersicon*; e.g. Spooner et al., 1993; Olmstead et al., 1999) is a powerful statement that is important to geneticists, molecular biologists, and plant breeders in that it points to a few close relatives of *S. lycopersicon* (out of a genus of several hundred species) as focal points for comparative genetic/genomic research and for crop improvement. Snapdragon (*Antirrhinum majus*) was historically part of a broadly defined Scrophulariaceae, a family that is now known to be grossly polyphyletic (i.e. not a single clade). Phylogenetic studies indicate that Scrophulariaceae should be broken up into several families (Olmstead et al., 2001), and snapdragon and its closest relatives are part of a clade recognized as the family Plantaginaceae.

A phylogenetic framework has revealed the patterns of evolution of many morphological and chemical characters, including complex pathways such as nitrogen-fixing symbioses, mustard oil production, and chemical defense mechanisms (for review, see Soltis and Soltis, 2000; Daly et al., 2001). However, the importance of phylogeny reconstruction applies not only to the organisms that house genes but also to the evolutionary history of the genes themselves. For example, are the genes under investigation the members of a single well-defined clade, all members of which appear to descend from a recent common ancestor as a direct result of speciation (orthologous genes), or do the sequences represent one or more ancient duplications (paralogous genes; see also Doyle and Gaut, 2000)? Gene families are, of course, the norm in studies of nuclear genes, but investigators are often bewildered by the diversity of genes encountered in a survey of a family of genes from a diverse array of plants. Phylogenetic methodology offers several solutions by permitting inferences of putative orthology among a set of sequences.

Examples of the phylogenetic analysis of gene families abound (e.g. genes encoding: heat shock proteins, Waters and Vierling, 1999; phytochrome, Kolukisaoglu et al., 1995; Mathews and Sharrock, 1997; and actin, McDowell et al., 1996). A noteworthy recent example involves MADS box genes, which encode transcription factors that control diverse developmental processes in plants. Some of the best known examples of MADS box genes include the A, B, and C class floral genes that control the identity of floral organs (for review, see Ma and dePamphilis, 2000). Phylogenetic analyses indicate that a minimum of seven different MADS box gene lineages were already present in the common ancestor of extant seed plants approximately 300 million years ago (mya; Becker et al., 2000). Thus, a diverse tool kit of MADS box genes was available before the origin of the angiosperms.

A phylogenetic perspective also provides the basis for comparative genomics (e.g. Soltis and Soltis, 2000; Walbot, 2000; Daly et al., 2001; Kellogg, 2001; Hall et al., 2002a; Mitchell-Olds and Clauss, 2002; Pryer et al., 2002; Doyle and Luckow, 2003). However, obtaining the appropriate phylogenetic perspective may be difficult: What phylogenetic hypotheses are already available for the group of interest? Are phylogenetic studies underway on a particular group, and is it possible to obtain unpublished trees? Is the phylogenetic underpinning for a lineage of interest sound enough for use in comparative genetic/genomic analyses? Not all phylogenetic trees are of equal quality, and the most fruitful phylogenomic compar-

[1] This work was supported in part by the National Science Foundation (Deep Time Research Coordination Network and the Floral Genome project grants).

* Corresponding author; e-mail dsoltis@botany.ufl.edu; fax 352–846–2154.

http://www.plantphysiol.org/cgi/doi/10.1104/pp.103.022509.

isons will be those based on the strongest phylogenetic inferences.

We cannot address all of the crucial issues relating to the importance of phylogeny in a comprehensive fashion and, therefore, will focus on a few main topics. We provide: (a) phylogenetic summaries and references for major clades of land plants, with an emphasis on angiosperm model systems; (b) a "primer" of phylogenetic methods, including evaluation of parsimony, distance, maximum likelihood (ML), and Bayesian methods, the importance of measures of internal support in phylogenetic inference, and methods of analysis of large data sets; and (c) use of molecular data to estimate divergence times of genes or organisms. A major goal is to foster increased interaction and communication between phylogeneticists and physiologists/molecular geneticists by providing contacts and references for those requiring a phylogenetic backbone for analyses.

SELECTION OF TAXA AND PHYLOGENETIC TREES IN COMPARATIVE STUDIES. A SUMMARY OF LAND PLANT PHYLOGENY

One question that systematists are frequently asked is: Where would I find the most recent phylogenetic tree for group (fill in the blank)? We provide a brief summary of relevant trees below, with a focus on land plants. In addition, selected trees for angiosperms can be found at http:/www.mobot.org/MOBOT/research/APweb//, http://www.flmnh.ufl.edu/deeptime/and http://plantsystematics.org/). Researchers can also consult Tree of Life (http://tolweb.org/tree/phylogeny.html) and TreeBASE (http://www.treebase.org/treebase). Phylogenetic questions can also be posed directly to experts working on various groups of plants; a partial list of phylogenetic consultants is provided in Table I (for a larger list, see also http://www.flmnh.ufl.edu/deeptime/).

Land Plants. Origin and Relationships

Understanding patterns of gene and genome evolution across land plants requires an understanding of the phylogeny of land plants, or embryophytes. Molecular data indicate that the sister group (i.e. the closest relative; two sister groups share a common ancestor not shared with any other group) of land plants is Charales (stoneworts) from the charophycean lineage of green algae (Karol et al., 2001; Fig. 1; see also http://www.flmnh.ufl.edu/deeptime/).

Table I. *Partial list of phylogenetic experts for various clades of land plants.*
For a larger list of experts, see the Deep Time Web site (http://www.flmnh.ufl.edu/deeptime/).

Clade(s)	Contact Person	E-Mail Address
Mosses, liverworts	Jonathan Shaw, Department of Biology, Duke University, Durham, NC 27708	shaw@duke.edu
Ferns	Kathleen Pryer, Department of Biology, Duke University, Durham, NC 27708	pryer@duke.edu
Basal angiosperms	Douglas Soltis, Department of Botany, University of Florida, Gainesville, FL 32611	dsoltis@botany.ufl.edu
	Pamela Soltis, Florida Museum of Natural History, University of Florida, Gainesville, FL 32611	psoltis@flmnh.ufl.edu
Monocots	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk
Poaceae	Elizabeth Kellogg, Department of Biology, University of Missouri, St. Louis 8001 Natural Bridge Rd, St. Louis, MO 63121	kellogg@msx.umsl.edu
Rosids	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk
	Douglas Soltis, Department of Botany, University of Florida, Gainesville, FL 32611	dsoltis@botany.ufl.edu
Fabaceae	Jeff Doyle, Department of Plant Science, Cornell University, Ithaca, NY 14853	jjd5@postoffice.mail.cornell.edu
	Matt Lavin, Department of Plant Sciences, Montana State University, Bozeman, MT 59717	mlavin@montana.edu
Brassicaceae	Ishan Al-Shehbaz, Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166	Ishan.Al-Shehbaz@mobot.org
Asterids	Walter Judd, Department of Botany, University of Florida, Gainesville, FL 32611	wjudd@botany.ufl.edu
	Richard Olmstead, Department of Botany, University of Washington, Seattle, WA 98195	olmstead@u.washington.edu
Caryophyllales	Mark Chase, Molecular Section, Jodrell Laboratory, Royal Botanic Gardens, Kew, Richmond, TW9 3DS UK	M.Chase@rbgkew.org.uk

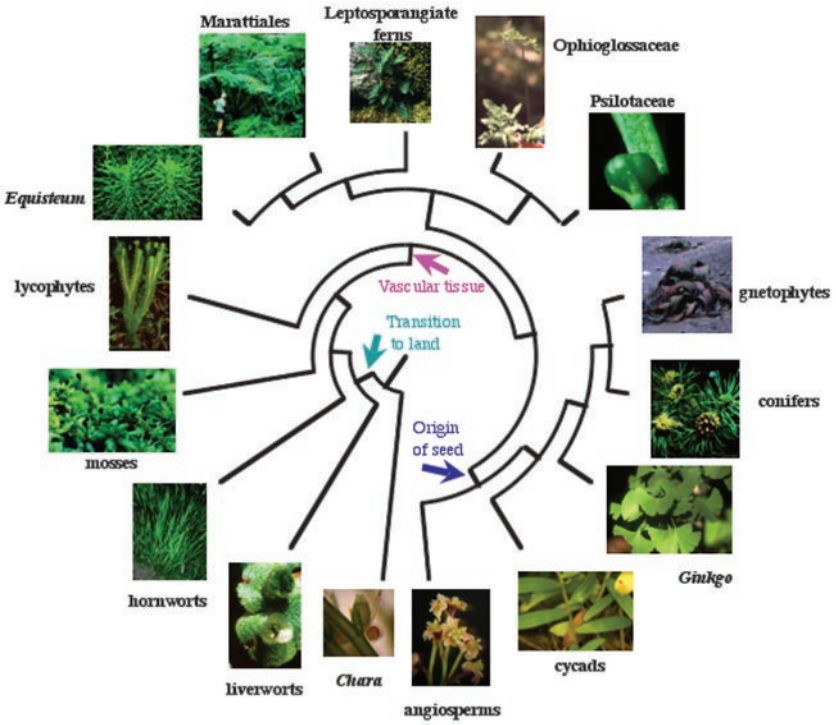


Figure 1. Summary of phylogenetic relationships among major lineages of embryophytes (land plants). Charales are the sister group of the embryophytes. Within the embryophytes, liverworts, hornworts, and mosses are the basal most lineages; however, their precise branching order is uncertain. One of the best supported topologies is depicted with liverworts, hornworts, and mosses as successive sisters to the tracheophytes (vascular plants). Within tracheophytes, there are two clades: monilophytes and spermatophytes (seed plants). Data from Karol et al. (2001), Pryer et al. (2001), and Soltis et al. (2002). Photograph of *Chara* courtesy of R. McCourt; photographs of *Welwitschia* and *Ophioglossum* courtesy of H. Wilson; photographs of *Ginkgo* sp. and *Zamia* courtesy of J. Manhart; photograph of *Polypodium* courtesy of J. Reveal; *Anthoceros* taken from CalPhotos (http://elib.cs.berkeley.edu); other photographs from the online teaching collection of the Botanical Society of America (http://www.botany.org/).

Plants colonized the land approximately 450 mya. Within the land plants, the three lineages long known as the "bryophytes" (liverworts, hornworts, and mosses) do not form a single clade in most analyses but instead form a grade that subtends the tracheophytes (Fig. 1). Furthermore, the precise branching order of the three "bryophyte" lineages remains ambiguous, with different topologies suggested by various data sets. A branching order of liverworts, hornworts, and mosses has emerged as one favored arrangement (e.g. Karol et al., 2001); other data suggest that hornworts, followed by a clade of mosses + liverworts, are the basal branches of the embryophytes (Renzaglia et al., 2000).

Tracheophytes

Vascular plants (tracheophytes) constitute a large and well-defined clade of land plants comprising the lycophytes (e.g. *Lycopodium*, *Selaginella*, and *Isoetes*) as sister to two well-marked clades—monilophytes and seed plants (Pryer et al., 2001; Fig. 1).

Monilophytes (or Moniliforms)

Both molecular and morphological analyses of tracheophytes have recognized a clade of *Equisetum*, Marattiaceae, Psilotaceae, Ophioglossaceae, and leptosporangiate ferns (Kenrick and Crane, 1997; Pryer et al., 2001). Kenrick and Crane (1997) first suggested the presence of this clade (based on one morphological character) and designated these plants Moniliformopses or "moniliforms"; they are now referred to more commonly as monilophytes (Judd et al., 2002). This monilophyte clade unites ancient lineages not previously considered closely related and is sister to a clade of all remaining tracheophytes—the seed plants (Fig. 1).

Seed Plants

Despite repeated efforts, it has been difficult to resolve phylogenetic relationships among extant seed plants, that is, angiosperms and the four lineages of living gymnosperms: cycads, *Ginkgo biloba*, conifers, and Gnetales (for review, see Donoghue and Doyle, 2000; Soltis et al., 2002). Analyses of morphological data generally concur in suggesting that angiosperms and Gnetales are sister groups (the "anthophyte" hypothesis), with extant gymnosperms paraphyletic (that is, not forming a clade but rather a grade; Donoghue and Doyle, 2000).

However, the sister group relationship of Gnetales and angiosperms has not been supported by most molecular analyses. Analyses of combined data sets of multiple genes representing all three plant genomes (plastid, mitochondrion, and nucleus) have found strong support for a clade of extant gymnosperms (Fig. 1; e.g. Bowe et al., 2000; Chaw et al., 2000; Pryer et al., 2001; Soltis et al., 2002). However, some extinct gymnosperms (e.g. Caytoniales and Bennettitales) may be more closely related to angiosperms than to any lineage of living gymnosperm (Donoghue and Doyle, 2000). Cycads and *Ginkgo biloba* are sisters to the remaining living gymnosperms. The relationship between cycads and *Ginkgo biloba*

is unclear; in some analyses, cycads and *Ginkgo biloba* are successive sisters to a clade of conifers and Gnetales, whereas in others, *Ginkgo biloba* and cycads form a clade that is sister to other extant gymnosperms. Some molecular analyses support a surprising placement of Gnetales within conifers as sister to Pinaceae (Bowe et al., 2000; the "gne-pine" hypothesis of Chaw et al., 2000).

The placement of Gnetales within conifers is an excellent example of a molecular phylogenetic result that must be viewed with caution, for several reasons. First, the placement of Gnetales within conifers is supported largely by mitochondrial genes; genes from other genomes do not place Gnetales within conifers. Furthermore, there is conflict between first and second versus third codon positions of cpDNA genes, with different positions supporting different placements of Gnetales. In addition, because most analyses of seed plants have involved small numbers of taxa, the gne-pine hypothesis may be an artifact of inadequate taxon sampling in some analyses. Our current interpretation of relationships among extant seed plants, showing Gnetales as sister to all conifers, is depicted in Figure 1. Analysis of extant gymnosperms exemplifies the complexities inherent in phylogenetic analysis of ancient lineages that have undergone significant extinction.

Angiosperms

The impact of molecular phylogenetic analyses on the angiosperms (flowering plants) has been particularly profound (e.g. Qiu et al., 1999; Graham and Olmstead, 2000; Soltis et al., 2000; Bremer et al., 2002; see below). Because of the wealth of molecular phylogenetic data, angiosperms became the first major group of organisms to be reclassified based largely on molecular data (Angiosperm Phylogeny Group [APG], 1998); data have accumulated so rapidly that this classification was recently revised (APG II, 2003). Readers will find that some family circumscriptions and ordinal groups have changed considerably from traditional classifications (e.g. Cronquist, 1981). Comprehensive trees depicting family level relationships for nearly all of the 300+ angiosperm families (e.g. Soltis et al., 2000; Zanis et al., 2002) and the APG II classification are available at http://www.flmnh.ufl.edu/deeptime/. Although recent classifications (e.g. Cronquist, 1981) may still provide some useful family descriptions, these classifications do not depict current concepts of phylogeny. For interpretations of data in a phylogenetic context and for consistency, authors are urged to follow the APG II (2003) classification.

Think "Eudicots." Abandon "Dicots"

The angiosperms, a clade of 260,000+ species (Takhtajan, 1997), first appeared in the fossil record, conservatively, approximately 130 mya (Hughes, 1994). Standard classifications divided the angiosperms into two large groups, typically recognized at the Linnean rank of class: Magnoliopsida (dicots) and Liliopsida (monocots). Thus, standard comparative studies of physiological pathways and genetic/genomic data have spanned this "monocot-dicot split." However, even preliminary morphology-based studies of angiosperms suggested that this "monocot-dicot split" did not accurately portray relationships. Molecular phylogenetic analyses clearly indicate that the traditional "dicots" are paraphyletic, with the monocots (a clade of $\pm$65,000 species) emerging from among the basal branches of angiosperms (Fig. 2). Following this basal grade of monocots and traditional "primitive dicots" (e.g. Amborellaceae, Nymphaeaceae, Austrobaileyales, and magnoliid clade) is a well-supported clade, the eudicots (Fig. 2). The eudicot clade contains 75% of all angiosperm species, united by the shared feature of triaperturate pollen (pollen with three grooves). The term "monocots" is still useful in that it designates a clade. In contrast, the term "dicots" should be abandoned because it does not correspond to a clade. This change in concept and terminology has already been accepted by many entry level biology and botany textbooks. Comparisons of genes or characters should be based on sister groups, if possible or, minimally, on other monophyletic groups. For example, because the sister group of the monocots remains uncertain, monocots could be compared with members of eudicots or magnoliids. Most of the published molecular comparisons of monocots and dicots have used eudicots as placeholders (e.g. Arabidopsis, *Brassica* spp., and *Antirrhinum* spp.) for the dicots. Thus, many such comparisons are still valid, even if the terminology used ("dicot") was incorrect.

NO SUBCLASSES

Perhaps the best known classification of angiosperms is that of Cronquist (1981), who recognized six subclasses of dicots, Magnoliidae, Hamamelidae, Rosidae, Dilleniidae, Caryophyllidae, and Asteridae, and five subclasses of monocots, although these were followed less frequently. Molecular phylogenies indicate that these subclasses, like the classes Magnoliopsida and Liliopsida, should also be abandoned. The Magnoliidae are paraphyletic, and both the Hamamelidae and Dilleniidae are grossly polyphyletic, with constituent members appearing throughout much of the angiosperm tree. Thus, "Magnoliidae," "Hamamelidae," and "Dilleniidae" do not refer to monophyletic groups, and these names are no longer valid. Cronquist's concepts of Rosidae, Asteridae, and Caryophyllidae must be expanded and revised to correspond to monophyletic groups; these clades are the rosids, asterids, and Caryophyllales sensu

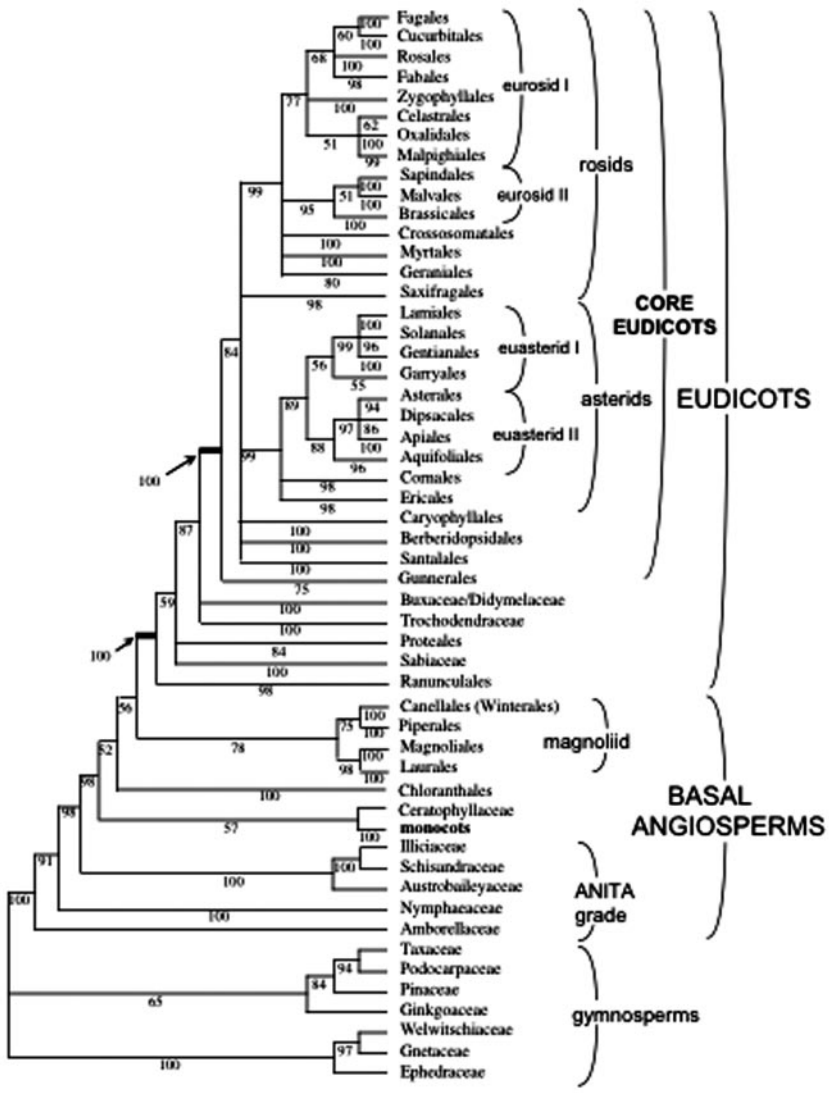


Figure 2. Summary tree of angiosperm relationships based on Soltis et al. (2000, 2003), with basal angiosperm relationships modified following Zanis et al. (2002). Numbers are jackknife values. Values for basal angiosperms are from Zanis et al. (2002); the value for the placement of Gunnerales is from Soltis et al. (2003); other eudicot values are from Soltis et al. (2000).

(APG II, 2003). Although Caryophyllales are recognized at the ordinal level (see APG II, 2003), both rosids and asterids are supraordinal groups that are not assigned a Linnean rank in the APG II classification.

It is important to note that deep-level angiosperm phylogeny is not yet resolved. Relationships among the major clades of eudicots (e.g. rosids, asterids, Caryophyllales, Saxifragales, Santalales, and a few smaller clades) are unresolved (Fig. 2), presenting a limitation for many areas of comparative biology, including comparative genomics.

MODEL GROUPS. OPPORTUNITIES FOR COMPARATIVE GENETICS AND GENOMICS IN THE ANGIOSPERMS

The phylogenetic trees available for many families of angiosperms facilitate interpretation of the evolution of diverse characters (molecular, physiological, and genetic). These trees also aid in the appropriate choice of representative taxa for comparative studies (see also Daly et al., 2001; Hall et al., 2002a); it is often useful to choose representative taxa from across the breadth of a clade and not simply one or two taxa from only a small part of the diversity of that clade.

Because trees depicting organismal phylogenies have accumulated so rapidly, it is often difficult for the nonexpert to know how to obtain a tree for a group of interest. Unfortunately, there is no single source that serves as a compendium of all intrafamilial phylogenetic trees. Judd et al. (2002) provide trees and relevant references for many families of tracheophytes. However, because it is an entry level textbook, many families are not covered. Therefore, we provide a short list of experts (Table I) who can assist with phylogenetic questions for major groups of embryophytes. A larger list is available on the Deep Time website.

Monocots

Molecular analyses have clarified many (but far from all) relationships within monocots (Chase et al., 2000; Soltis et al., 2000), and further analyses are underway (M. Chase and J. Davis, personal communication). The sister group of the monocots remains unclear, but the most comprehensive analyses suggest Ceratophyllaceae (Zanis et al., 2002; Fig. 2).

Poaceae

The Poaceae, or grass family, are an ideal focal point for comparative genetic/genomic research (Kellogg, 2001). The Grass Phylogeny Working Group (2001) has provided the most comprehensive and best supported tree for the grass family. Complete sequencing of the rice (*Oryza sativa*) genome and of entire cpDNA genomes for some genera, as well as extensive genetic/genomic data for crops including wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), and maize (*Zea mays*), make tribe Triticeae of particular interest; a firm phylogenetic framework is available not only for the tribe (Kellogg, 2001) but also for individual genera, such as *Hordeum* (Petersen and Seberg, 2003).

Antirrhinum Spp. (Snapdragon and Relatives)

Snapdragon (Plantaginaceae and Lamiales) is one of the best model systems for the study of floral developmental genetics and offers numerous opportunities for comparative genetic and genomic research. Although *Antirrhinum* spp. have long been placed in the family Scrophulariaceae, molecular phylogenetic studies indicate that the traditionally recognized Scrophulariaceae are not a single clade but actually represent a number of distinct clades: Scrophulariaceae in the strict sense; Plantaginaceae, which includes *Antirrhinum*, *Plantago*, and *Veronica*; Orobanchaceae, which contains all of the parasitic taxa formerly placed in either Orobanchaceae or Scrophulariaceae; the new family Calceolariaceae; an expanded Stilbaceae; and an expanded Phyrmaceae (Olmstead et al., 2001).

Solanaceae

Solanaceae contain a number of model organisms, including tomato and potato (*Solanum tuberosum*), tobacco (*Nicotiana tabacum*), peppers (*Capsicum annuum*), and petunia (*Petunia hybrida*). The family has also served as a model for studies of reproductive incompatibility and organization of the nuclear genome. A molecular phylogenetic framework and a provisional reclassification are now available for the family (Olmstead et al., 1999). Molecular studies have also confirmed that Convolvulaceae represent the sister group of Solanaceae (Soltis et al., 2000). As noted, tomato (formerly *Lycopersicon*) is clearly embedded within the large genus *Solanum*, which also includes potatoes. Thus, potato and tomato share very similar linkage maps (e.g. Tanksley et al., 1988; Doganlar et al., 2002) because they share a recent common ancestor.

Legumes (Fabaceae)

The closest relative of the Fabaceae has long been considered a mystery. Phylogenetic analyses have recently shown the closest relatives of Fabaceae to be Surianaceae and Polygalaceae (Soltis et al., 2000). Considerable progress has been made in recent years in clarifying relationships across the family as a whole and also within subclades within the family (Doyle and Luckow, 2003). Recent analyses have also identified the closest relatives of several important crop genera, including *Medicago*, *Gycine*, and *Pisum* (e.g. Kajita et al., 2001; Hu et al., 2002; for review, see Doyle and Luckow, 2003).

Brassicaceae

Brassicaceae offer important opportunities in comparative genomics by extending out from the complete genome sequence of Arabidopsis (e.g. Hall et al., 2002a; Mitchell-Olds and Clauss, 2002). Initial molecular phylogenetic analyses indicated the presence of a broadly defined Brassicaceae (Brassicaceae sensu lato) that also include Capparaceae. More recently, Hall et al. (2002b) found evidence for three well-supported clades within Brassicaceae sensu lato—Capparaceae subfamily Capparoideae, Capparaceae subfamily Cleomoideae, and Brassicaceae sensu stricto—with the latter two clades as sister groups. Rather than a single broadly defined family Brassicaceae, it may be more appropriate to recognize three families: Capparaceae, Cleomaceae, and Brassicaceae (Hall et al., 2002b). The model plants *Brassica* sp. and Arabidopsis are in Brassicaceae. It may be informative to include members of Capparaceae (e.g. *Capparis* spp.) and Cleomaceae (*Cleome* spp.) in comparative genetic and genomic analyses.

Recent phylogenetic studies of Arabidopsis and relatives (Koch et al., 1999, 2001, 2003; Koch, 2003; O'Kane and Al-Shehbaz, 2003) have provided an initial tree for Brassicaceae sensu stricto and identified an Arabidopsis clade that contains the closest relatives of Arabidopsis. However, a more comprehensive analysis of the family is required and is well underway (M. Beilstein, E. Kellogg, and I. Al-Shehbaz, personal communication).

Brassicales

Brassicaceae are part of a well-supported Brassicales (i.e. "glucosinolate clade"; e.g. Rodman et al., 1998; Soltis et al., 2000), a clade of 15 families that were not considered closely related in recent classifications (e.g. Cronquist, 1981). The order offers the

opportunity to investigate the evolution of a host of features considered characteristic of Brassicaceae. Some aspects of genomic and genic diversification will be better understood by extending out from Brassicaceae to relatives in Brassicales.

PHYLOGENY RECONSTRUCTION. A PRIMER

Alignment ("Garbage in; Garbage out")

Alignment of nucleotide and amino acid sequences is a major consideration, particularly in studies of genes from divergent taxa (e.g. rice and Arabidopsis). It seems obvious to state that the phylogenetic analysis of sequences begins with the appropriate alignment of the data themselves, yet alignment remains one of the most difficult and poorly understood facets of molecular data analysis. Detailed coverage of the topic is beyond the scope of this Update, but excellent overviews are provided by Doyle and Gaut (2000) and Simmons and Ochoterena (2000). We will simply restate, as Doyle and Gaut (2000) stress, that researchers should not accept alignments produced with the default settings of any computer algorithm without a critical evaluation by eye. Furthermore, there may be multiple "good" alignments, and all of these should be subjected to phylogenetic analysis.

Life after Neighbor Joining (NJ)

Inferences of orthology require phylogenetic analysis. Although expression patterns and knowledge of function may provide clues to orthology relationships, orthology, by definition, requires historical analysis to disentangle the products of gene duplication and speciation (for useful review of orthology and paralogy, see Doyle and Gaut, 2000; Jensen, 2001; Koonin, 2001). Thus, molecular biologists and geneticists suddenly need to become phylogeneticists. Although molecular phylogeny reconstruction is a relatively young discipline, it nonetheless has a rich and sometimes contentious background, encompassing diverse philosophies and methodologies that are not necessarily apparent to users of most available computer packages. Several approaches can be used in phylogeny reconstruction of molecular sequences: maximum parsimony (MP), maximum likelihood (ML), distance-based methods such as NJ, and Bayesian inference (BI), a new method of phylogenetic inference (Huelsenbeck et al., 2002). All of these methods have strengths and weaknesses (e.g. Swofford et al., 1996; Lewis, 1998; Doyle and Gaut, 2000; Huelsenbeck et al., 2002; Nei and Kumar, 2000), some of which are summarized in Table II.

Although there is a desire among many investigators for rapid phylogeny reconstruction and "instant tree," it may be prudent to explore several methods (e.g. Swofford et al., 1996; Doyle and Gaut, 2000; Nei and Kumar, 2000). There remains a tendency to place more trust in phylogenetic results supported by multiple approaches (Doyle and Gaut, 2000). Regardless of method of phylogenetic inference, however, some measure of internal support (e.g. bootstrap, jackknife, and posterior probabilities; see below) is essential.

Many non-systematists employ NJ to the exclusion of other methods (Nei and Kumar, 2002). The distance measures used in NJ and other distance methods are typically based on models of nucleotide substitution. The NJ algorithm is fast and readily available in software packages such as MEGA (http://www.megasoftware.net/) and PAUP*. However, it also has important weaknesses. For example, NJ provides only a single tree, precluding comparison with other topologies. In reality, many optimal trees may be found in MP and ML analyses, depending on the data set, and these methods allow all optimal or near-optimal trees to be compared. Furthermore, different trees can be obtained with NJ depending on the entry order of the taxa (Farris et al., 1996; see Table II). One solution is to run multiple NJ analyses with different random entry orders of the taxa, accompanied by bootstrap or jackknife analysis (see below). Finally, because sequence differences are summarized as distance values, it is impossible to identify the specific character changes that support a branch. Although proponents of NJ, Nei and Kumar (2000) nonetheless argue for a pluralistic approach. Other methods of phylogenetic inference should be explored in addition to NJ.

MP is preferred by many phylogeneticists because of its theoretical basis and the diagnosable units it produces. The advantages of parsimony over NJ are several (Table II), an important one being that parsimony seeks to recover all shortest trees. Depending on the data set, a parsimony search may yield one (or a few) to hundreds or thousands of equally short trees. These shortest trees can be summarized in a strict consensus tree, which depicts only the nodes present in all equally short trees. In addition, MP analysis provides diagnoses (i.e. specific sets of characters) for each clade and branch lengths in terms of the number of steps (or changes) on each branch of a tree.

Statistical methods of phylogeny reconstruction, incorporating models of nucleotide (or amino acid) substitution, are preferred by many molecular phylogeneticists (see Lewis, 1998). Both ML and BI rely on such models to reconstruct both topology and branch lengths and, thus, are computationally intensive. ML analysis finds the likelihood of the data, given a tree and a model of molecular evolution. Like ML, BI has had a long tenure in statistics. However, it has only recently been introduced into phylogenetics (see Huelsenbeck et al., 2001, 2002). Although BI uses the same models of evolution as some other methods of phylogenetic analyses (e.g. ML and NJ), it represents a powerful tool and perhaps the wave of the future in phylogenetic inference. BI is based on a quantity referred to as the posterior probability of a

Table II. *Comparison of methods of phylogeny reconstruction*

Method	General	Advantages	Disadvantages
Parsimony	Simplest explanation is the best (Ockham's razor)	By minimizing no. of steps, it also minimizes the no. of additional hypothesis (parallel or reversal nucleotide substitutions)	Different results may be obtained based on the entry order of sequences (therefore, perform multiple searches)
	Select the tree or trees that minimize the amount of change (no. of steps)	Searches identify numerous equally parsimonious (shortest) trees; treats multiple hits as an inevitable source of false similarity (homoplasy)	Relatively slow (compared with NJ) with large data sets
		Basic method can be modified by weighting schemes to compensate for multiple hits	Highly unequal rates of base substitution may cause difficulties (e.g. long branch attraction)
		Readily implemented in PAUP*	
		Can identify individual characters that are informative or problematic	
		Can infer ancestral states	
NJ	Involves estimation of pair-wise distances between nucleotide sequences	Fast	Different results may be obtained based on the entry order of sequences
	Pair-wise distances compensate for multiple hits by transforming observed percent differences into an estimate of the no. of nucleotide substitutions using one of several models of molecular evolution	Provides branch lengths	Only a single tree produced; cannot evaluate other trees
	Minimum evolution is a common distance criterion for picking an optional tree (sum of all branch lengths is the smallest)	Uses molecular evolution model	Branch lengths presented as distances rather than as discrete characters (steps)
	NJ algorithm provides a good approximation of the minimum evolution tree	Readily implemented in PAUP* and MEGA	Cannot identify characters that are either informative or problematic
			Cannot infer ancestral states
Maximium Likelihood	Involves estimating the likelihood of observing a set of aligned sequences given a model of nucleotide substitution and a tree	A statistical test (the likelihood ratio test) can be used to evaluate properties of trees	Computationally very intensive (much slower than other methods
		Nucleotide substitution models are used directly in the estimation process, rather than indirectly (as in parsimony)	Practical with only small nos. (fewer than 50) of sequences
		Flexible, models that can incorporate parameters of base frequencies, substitution rates, and variation in substitution rates and, therefore, are "general"; Jukes-Cantor sets a single substitution rate and is more "restricitive"	
		Easily implemented in PAUP*	
		Uses all of the data (invariable sites and unique mutations are still informative, unlike parsimony analysis)	
Bayesian	Uses a likelihood function and an efficient search strategy	Based on the likelihood function, from which it inherits many of its favorable statistical properties	Very large memory demands
	Based on a quality called the posterior probability of a tree	Uses models as in ML	
	Researcher may specify belief in a prior hypothesis prior to analysis	Can be used to analyze relatively large data sets	
		Provides support values	Posterior probabilities (measure of internal support) can be overestimates

tree, a value that can be interpreted as the probability that a tree is correct, given the data. BI uses a likelihood function to compute the posterior probability. Although BI allows the researcher to specify a prior belief in relationships (Table II; Huelsenbeck et al., 2001, 2002), this option has not been explored extensively to date, and Bayesian analyses typically assign equal prior probability values to all possible trees. Whereas ML is not feasible for large data sets (more than perhaps 50 taxa), BI (as implemented in MrBayes; see Huelsenbeck et al., 2001) incorporates a faster search strategy (using Markov chains) and can be used on data sets of several hundred taxa to find tree, branch lengths, and support (but see Suzuki et al., 2002).

Certainly a frustrating aspect of phylogenetic analysis to those outside of the field is the number of inference methods available. NJ is widely used, in part, because of its speed and ready availability in computer packages such as MEGA. It also is part of alignment packages such as MegAlign (http://www.dnastar.com/cgi-bin/php.cgi?r10.php). However, parsimony can be readily implemented using PAUP* (Swofford, 1998; NJ and ML are also part of the PAUP package). PAUP* is often not employed by molecular biologists, however, because the user friendly version with pull-down menus is made for Macintosh, not Windows, operating systems.

Internal Support for Clades

Some measure of internal support for clades should be provided on all phylogenetic trees. Resampling approaches, such as the bootstrap and the jackknife, are easily computed using PAUP* for parsimony, NJ, and ML analyses, and parsimony jackknifing is performed by *Jac* (Farris et al., 1996). The pros and cons of the jackknife versus bootstrap have been discussed (e.g. Farris et al., 1996; Soltis and Soltis, 2003). A reasonable number of replications should be employed, but "reasonable" varies with the size of the data set, the specifications of the analysis, and the patience of the investigator. It has been argued (Farris et al., 1996) that resampling methods should maximize the number of replicates at the expense of detailed searches in each replicate. Thus, with "fast" methods that conduct little or no branch swapping per replicate, 1,000 or more replicates are quickly obtained. A smaller number of replicates (e.g. 100) may be suitable for bootstrap and jackknife analyses that include detailed searches per replicate.

Interpretations of bootstrap and jackknife values vary (for review, see Soltis and Soltis, 2003), although few view these values in a strict statistical sense. Bootstrap values are conservative, but biased, measures of phylogenetic accuracy (Hillis and Bull, 1993), with values of 70% or greater corresponding to "true" clades in experimental phylogenies (Hillis and Bull, 1993). Thus, some consider values of 70% or more as indicators of strong support, whereas others reserve "strong support" for values of 90 or 95% and above. Although different phylogenetic methods may yield different optimal topologies, the differences generally involve poorly supported clades. Those clades that are strongly supported generally appear in topologies regardless of the method of phylogenetic inference. Additional measures of support include the decay index or Bremer support (Bremer, 1994) for parsimony analyses and the posterior probabilities generated in BI.

Measures of internal support indicate those relationships in which we should, and should not, have confidence. A recently identified clade of MADS-box genes appears as the sister group to the well-known B class floral genes that specify the identity of petals and stamens in Arabidopsis and snapdragon. Becker et al. (2002) termed this new clade B_{sister} and determined that these genes are present in diverse seed plants. Although the monophyly of the B_{sister} clade received 92% bootstrap support, the placement of the B_{sister} clade as sister to the clade of B class genes received only 77% bootstrap support. With this level of support, it is reasonable to question whether the B_{sister} clade is really the sister group of the clade of B class genes. Increased sampling of B_{sister} genes from additional taxa and more rigorous analyses are needed to establish with certainty the placement of the B_{sister} clade within the MADS box genes of plants.

MOLECULAR CLOCKS. RATES AND DATES OF GENE DIVERSIFICATION

Many efforts to date evolutionary divergences using a molecular clock have yielded age estimates that are grossly inconsistent with the fossil record, regardless of method of tree construction. For example, molecular-based estimates of divergence times in plants reveal a vast range of dates. Using molecular data, the age of the angiosperms has been estimated as 350 to 420 mya, greater than 319, 200, to 140 to 190 mya (for review, see Sanderson and Doyle, 2001). However, the oldest unequivocal angiosperm fossils are 125 to 135 mya (for review, see Soltis et al., 2002).

Many sources of error and bias can affect molecular-based estimates of divergence times (see Sanderson and Doyle, 2001; Soltis et al., 2002). Obviously, an incorrect topology will yield erroneous estimates, with the magnitude of the problem depending on the extent of the topological error (Sanderson and Doyle, 2001). Inaccurate calibration will bias the resulting estimates. Also problematic are heterogeneous rates of evolution among lineages (see Sanderson and Doyle, 2001; Soltis et al., 2002). Inadequate taxon sampling can compound the problem. Estimates of divergence times can also vary among genes or other data partitions (e.g. among codon positions). Another potential source of error is the method used to estimate divergence dates. Sanderson and Doyle

(2001) used molecular data to examine angiosperm divergences and found that the age of crown group angiosperms ranges from 68 to 281 mya, depending on data, tree, and assumptions, with most estimates falling between 140 and 190 mya.

Given that rate heterogeneity among lineages is common in most molecular-based trees, can we reliably use molecular data to estimate divergence times? Simple clock-based approaches to estimating divergence times are not likely to yield meaningful estimates. However, several approaches have been proposed when the assumption of rate constancy is violated: linearized trees (Takezaki et al., 1995), nonparametric rate smoothing (Sanderson, 1997, 1998), penalized likelihood (Sanderson, 2002), Bayesian approaches (e.g. Huelsenbeck et al., 2002; Thorne and Kishino, 2002), and "PATH" (Britton et al., 2002; for review of methods and instructions for implementing nonparametric rate smoothing, see http://www.flmnh.ufl.edu/deeptime/). Although methods to accommodate deviations from a steady molecular clock are still under development, it is nonetheless possible to estimate dates of divergence, given: (a) a reliable calibration point or points, (b) adequate sampling of taxa and characters, and (c) a method that is robust to rate heterogeneity. Confidence intervals for the estimated dates and consistency with the fossil record provide means for assessing the reliability of age estimates. Despite attempts to accommodate deviations from constant evolutionary rates, however, confidence intervals are typically large, and divergence times should be interpreted carefully.

SUMMARY AND FUTURE PROSPECTS

An exciting recent development is the merging of phylogenetics and genomics. Phylogenetic hypotheses have become the framework for the choice of organisms in genomic analyses, and more and more molecular biologists are using phylogenetic trees to guide their sampling of taxa for comparative research. This trend will continue. Systematics is moving rapidly; therefore, molecular biologists are encouraged to contact systematics "experts" for help in obtaining the best supported trees for a given clade of interest. We stress the importance of a rigorous phylogenetic analysis of data. It is ironic, for example, that researchers may spend years gathering gene sequence data, but then want an immediate phylogenetic "answer" within seconds or minutes. A thorough phylogenetic analysis, evaluating alternative alignments, exon versus intron boundaries, using different phylogenetic methods, and obtaining estimates of internal support, may take several weeks or more, and this should not be considered an unreasonable investment of time. Our review of issues relating to phylogeny reconstruction also illustrates the need for more "quick courses" in phylogeny reconstruction for molecular biologists interested in constructing gene trees.

ACKNOWLEDGMENTS

We thank Jeff Doyle, Bernie Hauser, Alice Harmon, and two anonymous reviewers for helpful comments on earlier drafts of this paper.

Received February 27, 2003; returned for revision March 30, 2003; accepted May 12, 2003.

LITERATURE CITED

Angiosperm Phylogeny Group (1998) An ordinal classification for the families of flowering plants. Ann Missouri Bot Gard **85:** 531–553

Angiosperm Phylogeny Group II (2003) An updated classification of the angiosperms. Bot J Linn Soc **141:** 399–436

Becker A, Kaufmann K, Freialdenhoven A, Vincent C, Li MA, Saedler H, Theissen G (2002) A novel MADS-box gene subfamily with a sister-group relationship to class B floral homeotic genes. Mol Genet Genomics **266:** 942–950

Becker A, Winter K-U, Meyer B, Saedler H, Theissen G (2000) MADS-box gene diversity in seed plants 300 million years ago. Mol Biol Evol **17:** 1425–1434

Bowe LM, Coat G, DePamphilis CW (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

Bremer K (1994) Branch support and tree stability. Cladistics **10:** 295–304

Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Källersjö M, Anderberg A, Barkhordarian E (2002) Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. Mol Phylogenet Evol **24:** 274–301

Britton T, Oxelman B, Vinnersten A, Bremer K (2002) Phylogenetic dating with confidence intervals using mean path lengths Mol Phylogenet Evol **24:** 58–65

Chase MW, Soltis DE, Soltis PS, Rudall PJ, Fay MF, Hahn WJ, Sullivan S, Joseph J, Molvray M, Kores PJ et al. (2000) Higher-level systematics of the monocutyledons: an assessment of current knowledge and a new classification. *In* KL Wilson, DA Morrison, eds, Monacots: Systematics and Evolution. CSIRO Publishing, Victoria, Australia, pp 3–16

Chaw SM, Parkinson CL, Cheng Y, Vincent TM, Palmer JD (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

Cronquist A (1981) An Integrated System of Classification of Flowering Plants. Columbia University Press, New York

Daly DC, Cameron KM, Stevenson DW (2001) Plant systematics in the age of genomics. Plant Physiol **127:** 1328–1333

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD (2002) A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics **161:** 1697–16711

Donoghue MJ, Doyle JA (2000) Seed plant phylogeny: demise of the anthophyte hypothesis? Curr Biol **10:** R106–R109

Doyle JJ, Gaut B (2000) Evolution of genes and taxa: a primer. Plant Mol Biol **42:** 1–23

Doyle JJ, Luckow MS (2003) The rest of the iceberg: legume diversity and evolution in a phylogenetic context. Plant Physiol (in press)

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. Cladistics **12:** 99–124

Graham SW, Olmstead RG (2000) Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. Am J Bot **87:** 1712–1730

Grass Phylogeny Working Group (2001) Phylogeny and subfamilial classification of the grasses (Poaceae). Ann Missouri Bot Gard **88:** 373–457

Hall AE, Fiebig A, Preuss D (2002a) Beyond the *Arabidopsis* genome: opportunities for comparative genomics. Plant Physiol **129:** 1439–1447

Hall JC, Sytsma KJ, Iltis HH (2002b) Phylogeny of Capparaceae and Brassicaceae based on chloroplast sequence data. Am J Bot **89:** 1826–1842

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst Biol **42:** 182–192

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. Science **294:** 2310–2314

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. Syst Biol **51:** 673–688

Hu J-M, Lavin M, Wojciechowski M, Sanderson MJ (2002) Phylogenetic analysis of nuclear ribosomal ITS/5.8S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe-Cyclolobium*, the core Millettieae, and the *Callerya* group. Syst Bot **27:** 722–733

Hughes NF (1994) The Enigma of Angiosperm Origins. Cambridge University Press, Cambridge, UK

Jensen RA (2001) Orthologs and paralogs: we need to get it right. Genome Biol **2:** 1002.1–1002.3

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ (2002) Plant Systematics: A Phylogenetic Approach. Sinauer Associates, Inc., Sunderland, MA

Kajita T, Ohashi H, Tateishi Y, Bailey CD, Doyle JJ (2001) *rbcL* and legume phylogeny with particular reference to Phaseoleae, Millettieae, and allies. Syst Bot **26:** 515–536

Karol KG, McCourt RM, Cimino MT, Delwiche CF (2001) The closest living relatives of land plants. Science **294:** 2351–2352

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol **125:** 1198–1205

Kenrick P, Crane PR (1997) The Origin and Early Evolution of Land Plants. Smithsonian Institution Press, Washington, DC

Koch M (2003): Molecular Phylogenetics, Evolution and Population Biology in the Brassicaceae. *In* AK Sharma, A Sharma A, eds, Plant Genome: Biodiversity and Evolution, Vol 1: Phanerogams. Science Publishers, Inc., Enfield, NH (in press)

Koch M, Bishop J, Mitchell-Olds T (1999) Molecular systematics and evolution of *Arabidopsis* and *Arabis*. Plant Biol **1:** 529–537

Koch M, Haubold B, Mitchell-Olds T (2001) Molecular systematics of the Brassicaceae: evidence from coding plastid *matK* and nuclear *Chs* sequences. Am J Bot **88:** 534–544

Koch M, Mummenhoff K, Al-Shehbaz IA (2003): Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae): a review of a decade of studies. Ann Missouri Bot Gard (in press)

Kolukisaoglu HM, Marx MS, Weigmann C, Hanelt S, Schneider-Portsch AW (1995) Divergence of the phytochrome gene family predates angiosperm evolution and suggests that *Selaginella* and *Equistem* arose prior to *Psilotum*. J Mol Evol **41:** 329–337

Koonin EV (2001) An apology for orthologs: or brave new memes. Genome Biol **2:** 1005.1–1005.2

Lewis P (1998) Maximum likelihood as an alternative to parsimony for inferring phylogeny using nucleotide sequence data. *In* DE Soltis, PS Soltis, JJ Doyle, eds, Molecular Systematics of Plants II: DNA Sequencing. Kluwer, Boston, pp 132–187

Ma H, dePamphilis C (2000) The ABCs of floral evolution. Cell **101:** 5–8

Mathews S, Sharrock RA (1997) Phytochrome gene diversity. Plant Cell Environ **20:** 666–671

McDowell JM, Huang S, McKinney EC, An Y-Q, Meacher RB (1996) Structure and evolution of the actin gene family in *Arabidopsis thaliana*. Genetics **142:** 587–602

Mitchell-Olds T, Clauss MJ (2002) Plant evolutionary genomics. Curr Opin Plant Biol **5:** 74–79

Nei M, Kumar S (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Oxford

O'Kane SL, Al-Shehbaz IA (2003) Phylogenetic position and generic limits of *Arabidopsis* (Brassicaceae) based on sequences of nuclear ribosomal DNA. Ann Missouri Bot Gard (in press)

Olmstead RG, dePamphilis CW, Wolfe AD, Young ND, Elisons WJ, Reeves A (2001) Disintegration of the Scrophulariaceae. Am J Bot **88:** 348–361

Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer J (1999) Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. *In* M Nee, DE Symon, RN Lester, JP Jessop, eds, Solanaceae IV. Royal Botanic Gardens, Kew, UK, pp 111–117

Petersen G, Seberg O (2003) Phylogenetic analyses of the diploid species of *Hordeum* (Poaceae) and a revised classification of the genus. Syst Bot **28:** 293–306

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf P, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature **409:** 618–622

Pryer KM, Schneider H, Zimmer EA, Banks JA (2002) Deciding among green plants for whole genome studies. Trends Plant Sci **7:** 550–554

Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature **402:** 404–407

Renzaglia KS, Duff RJ, Nickrent DL, Garbary DJ (2000) Vegetative and reproductive innovations if early land plants: implications for a unified phylogeny. Philos Trans R Soc Lond B **355:** 769–793

Rodman JE, Soltis PE, Sytsma KJ, Karol KG (1998) Parallel evolution of glucosinolate biosynthesis inferred from congruent nuclear and plastid gene phylogenies. Am J Bot **85:** 997–1006

Sanderson MJ (1997) A nonparametric approach to estimating divergence times in the absence of rate constancy. Molec Biol Evol **14:** 1218–1231

Sanderson MJ (1998) Estimating rate and time in molecular phylogenies: beyond the molecular clock. *In* DE Soltis, PS Soltis, JJ Doyle, eds, Molecular Systematics of Plants II. Kluwer, Boston, pp 242–264

Sanderson MJ (2002) Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol Biol Evol **19:** 101–109

Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from *rbcL* and 18S rDNA data. Am J Bot **88:** 1499–1516

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol **49:** 369–381

Soltis DE, Soltis PS (2000) Contributions of plant molecular systematics to studies of molecular evolution. Plant Mol Biol **42:** 45–75

Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, Savolainen V, Hahn WH, Hoot SB, Fay MF et al. (2000) Angiosperm phylogeny inferred from a combined data set of 18S rDNA, rbcL and atpB sequences. Bot J Linn Soc **133:** 381–461

Soltis PS, Soltis DE (2003) The bootstrap in phylogeny reconstruction. Am Stat (in press)

Soltis PS, Soltis DE, Savolainen V, Crane PR, Barraclough T (2002) Rate heterogeneity among lineages of land plants: integration of molecular and fossil data and evidence for molecular living fossils. Proc Natl Acad Sci USA **99:** 4430–4435

Spooner DM, Anderson GJ, Jansen RK (1993) Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes and pepinos (Solanaceae). Am J Bot **80:** 676–688

Susuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc Natl Acad Sci USA **99:** 16138–16143

Swofford DL (1998) PAUP* 4.0: Phylogenetic Analysis Using Parsimony (and Other Methods). Sinauer Associates, Sunderland, MA

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. *In* DM Hillis, C Moritz, BK Mable, eds, Molecular Systematics. Sinauer Associates, Sunderland, MA, pp 407–514

Takezaki N, Rzhetsky A, Nei M (1995) Phylogenetic test of the molecular clock and linearized trees. Mol Biol Evol **12:** 823–833

Takhtajan A (1997) Diversity and classification of flowering plants. Columbia University Press, New York

Tanksley SD, Bernatsky R, Lapitan NL, Prince JP (1988) Conservation of gene repertoire but not gene order in pepper and tomato. Proc Natl Acad Sci USA **84:** 6419–6423

Thorne JL, Kishino H (2002) Divergence time and evolutionary rate estimation with multilocus data. Syst Biol **51:** 689–702

Walbot V (2000) A green chapter in the book of life. Nature **408:** 794–795

Waters ER, Vierling E (1999) The diversification of plant cytosolic small heat shock proteins preceded the divergence of mosses. Mol Biol Evol **16:** 127–139

Zanis MJ, Soltis DE, Soltis PS, Mathews S, Donoghue MJ (2002) The root of the angiosperms revisited. Proc Natl Acad Sci USA **99:** 6848–6853

A Molecular Timeline for the Origin of Photosynthetic Eukaryotes

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

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

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

Introduction

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

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

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

Mol. Biol. Evol. 21(5):809–818. 2004
DOI:10.1093/molbev/msh075
Advance Access publication February 12, 2004

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

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

Materials and Methods

Taxon Sampling and Sequencing

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

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

To prepare DNA, the algal cultures were frozen in liquid nitrogen and ground with glass beads using a glass rod and/or Mini-BeadBeater (Biospec Products, Inc., Bartlesville, Okla.). Total genomic DNA was extracted with the DNeasy Plant Mini Kit (Qiagen, Santa Clarita, Calif.). Polymerase chain reactions (PCR) were done using specific primers for each of the plastid genes (see Yoon, Hackett, and Bhattacharya 2002; Yoon et al. 2002). Four degenerate primers were used to amplify and sequence the photosystem I P700 chlorophyll *a* apoprotein A2 (*psa*B) gene: psaB500F; 5′-TCWTGGTTYAAAAATAAYGA-3′, psaB1000F; 5′-CAAYTAGGHTTAGCTTTAGC-3′, psaB1050R; 5′-GGYAWWGCATACATATGYTG-3′, psaB-1760R; 5′-CCRATYGTATTWAGCATCCA-3′. Because introns were found in the plastid elongation factor Tu (*tuf*A) and photosystem I P700 chlorophyll *a* apoprotein A1 (*psa*A) genes of some red algae (most likely indicating gene transfer to the nucleus [H. S. Y., D. B. unpublished data]), the reverse transcriptase (RT)-PCR method was used to isolate cDNA. For the RT-PCR, total RNA was extracted with the RNeasy Mini Kit (Qiagen, Santa Clarita, Calif.). To synthesize cDNA from total RNA, M-MLV Reverse Transcriptase (GIBCO BRL, Gaithersburg, Md.) was used according to the manufacturer's protocol. The PCR products were purified with the QIAquick PCR Purification kit (Qiagen), and were used for direct sequencing with the BigDye Terminator Cycle Sequencing Kit (PE-Applied Biosystems, Norwalk, Conn.) and an ABI-3100 at the Center for Comparative Genomics at the University of Iowa. Some PCR products were cloned into pGEM-T vector (Promega, Madison, Wis.) prior to sequencing.

Phylogenetic Analyses

Sequences were manually aligned with SeqPup (Gilbert 1995). The alignment used in the phylogenetic analyses is available on request from D. B. We prepared a concatenated data set of 16S rRNA (1,309 nt), *psa*A (1,395 nt), *psa*B (1,266 nt), photosystem II reaction center protein D1 (*psb*A) (957 nt), ribulose-1,5-bisphosphate carboxylase/oxygenase (*rbc*L; 1,215 nt), and *tuf*A (969 nt) coding regions (a total of 7,111 nt) from photosynthetic eukaryotes and the cyanobacterium *Nostoc* sp. PCC7120 as the outgroup. Because the *rbc*L gene of the green and glaucophyte algae are of cyanobacterial origin, whereas those in the red algae and red-algal-derived plastids are of proteobacterial origin (e.g., Valentin and Zetsche 1990), the evolutionarily distantly related green and glaucophyte *rbc*L sequences were coded as missing data in the phylogenetic analyses. The highly divergent and likely nonfunctional *tuf*A sequence in *Chaetosphaeridium globosum* (Baldauf, Manhart, and Palmer 1990) and the nuclear-encoded land plant *tuf*A genes (Baldauf and Palmer 1990) were also excluded from the analysis.

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

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

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

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

Molecular Clock Analyses

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

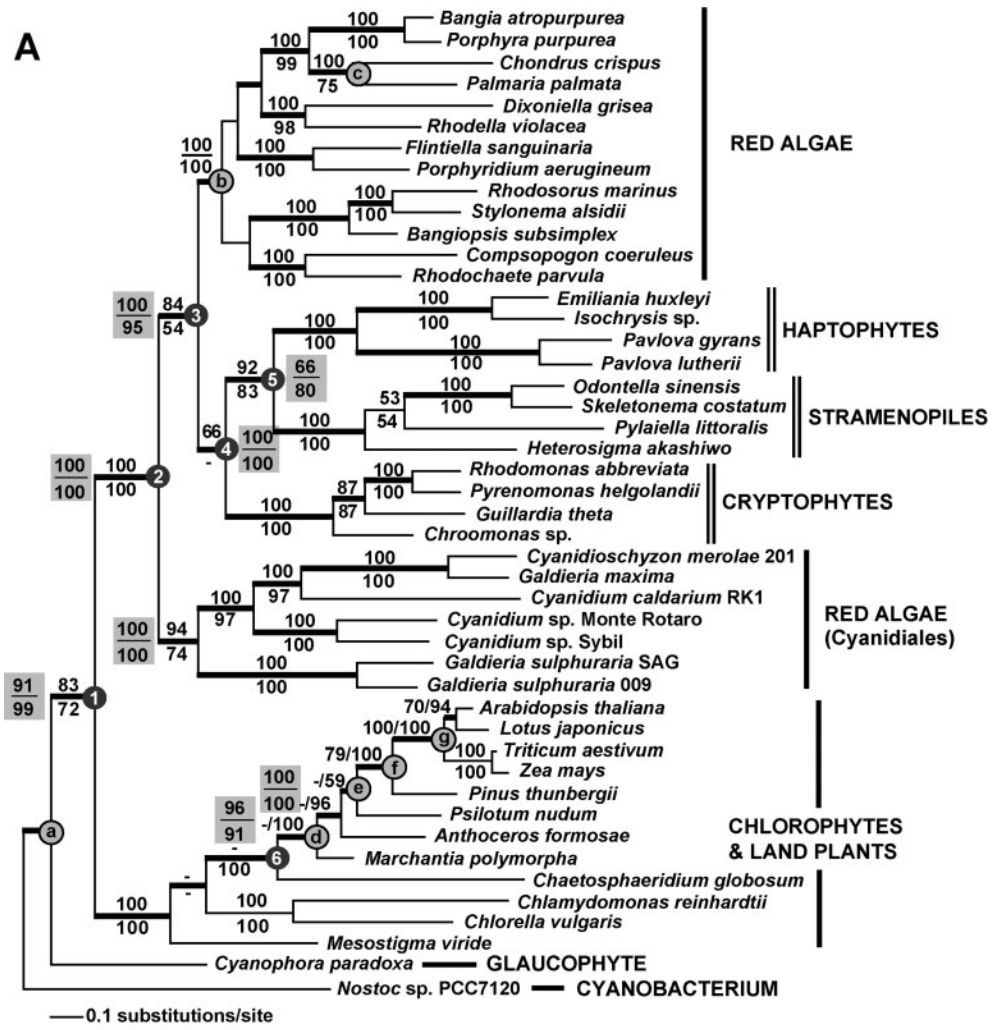
Results and Discussion

Phylogenetic Relationships

The Bayesian tree of highest likelihood (excluding the third codon positions in the data), which was identified using the GTR evolutionary model with gamma-distributed rates across sites for three partitions, is shown in figure 1A. This phylogenetic hypothesis has relatively broad taxonomic sampling, including early diverging red (Cyanidiales) and green algal (*Mesostigma viride*) and land plant (e.g., *Marchantia polymorpha*) lineages, and it is consistent with present understanding of algal and plant relationships (Cavalier-Smith 1986; Fast et al. 2001; Karol et al. 2001; Yoon et al. 2002). Most nodes in the phylogeny, except that defining chromist monophyly (the haptophytes and stramenopiles were, however, strongly supported as sister groups), the near-simultaneous radiation of the non-Cyanidiales red algae, and the early divergences in the chlorophyte/land plant lineage (see fig. 1*A*), have a significant (≥95%) posterior probability and strong bootstrap support (ME and MP methods). When we added the third codon positions (see fig. 2 in the Supplementary Material online) and reanalyzed the data using the four-partition model, the resulting Bayesian tree was essentially identical with the tree shown in figure 1*A*, however, with stronger bootstrap support for many nodes (see the shaded bootstrap values in figure 1*A*). Bootstrap analysis of the RY-recoded data set using the ME method (see fig. 3 in the Supplementary Material online) resulted in a consensus tree that was consistent with the results described above, with strong support for chromist plastid (94%) monophyly. The order of divergence of the non-Cyanidiales red algae and the early splits among land plants remained unresolved in this analysis (as in fig. 1*A*).

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

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



B

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

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

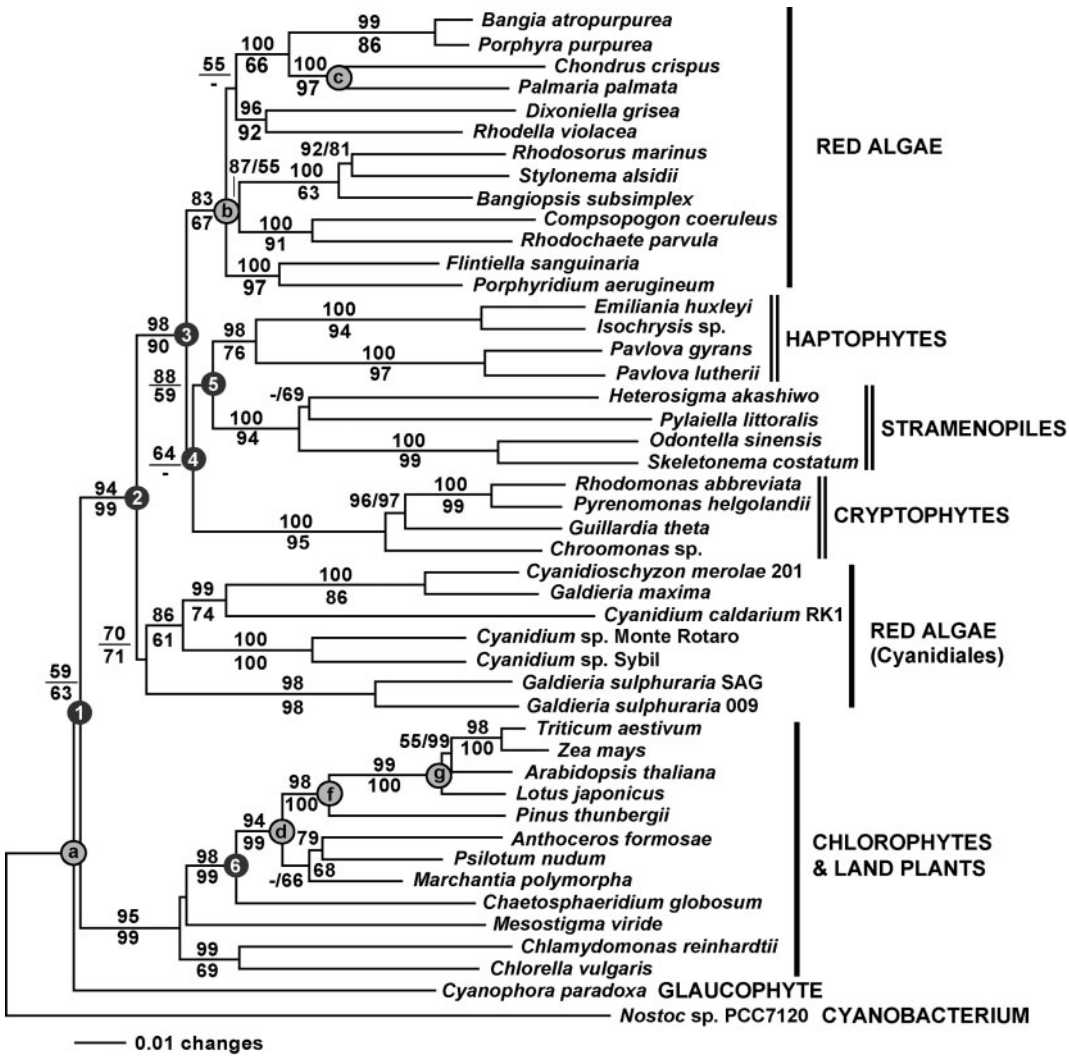


FIG. 2.—Evolutionary relationships of algal plastids using the five-protein data set. The phylogeny was inferred using the ME method, and distance matrices were calculated using the WAG + Γ evolutionary model. The results of a protein ME bootstrap analysis are shown above the branches, whereas puzzle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

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

Divergence Time Estimations

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

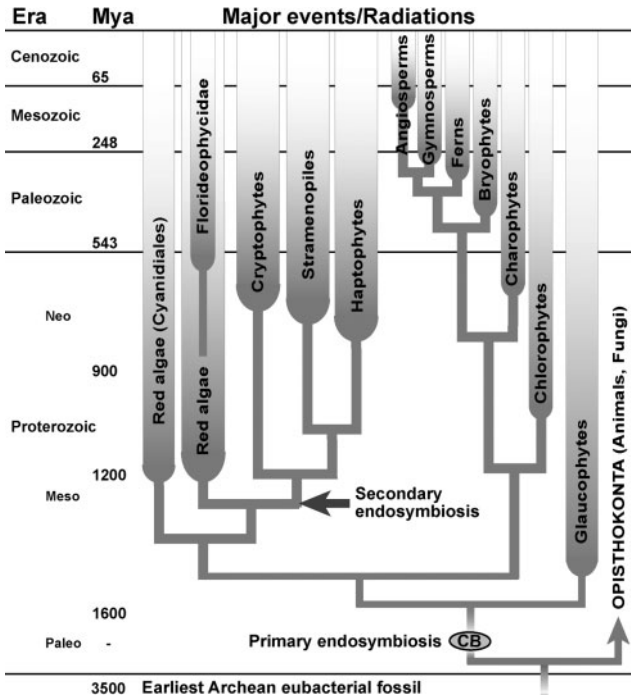


FIG. 3.—Schematic representation of the evolutionary relationships and divergence times for the red, green, glaucophyte, and chromist algae. These photosynthetic groups are outgroup-rooted with the Opisthokonta which putatively ancestrally lacked a plastid. The branches on which the cyanobacterial (CB)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

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

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

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

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

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

To test the LF divergence time estimates in which we specified 12 "local rates" in the tree, we also used the NPRS method to accommodate rate inconstancy (Sanderson 1997). The estimated divergence dates using NPRS are older than those using the LF method; however, these differences are relatively minor—e.g., 1,354 MYA for the chromist plastid split (node 3) and 1,255 MYA for the cryptophyte plastid split (node 4; see table 2 in the Supplementary Material online). We also assessed the precision of our divergence time estimates using the credible tree set identified by Bayesian inference. The average divergence times (using the LF method) and the 95% confidence intervals of the distributions are very similar to the results using the best Bayesian tree (see figure 1*B*). This suggests that there is only minor variation in the branch length estimates in the pool of credible trees used in this analysis (see fig. 1 in the Supplementary Material online); finally, the divergence time estimates (fig. 1*B*) that were inferred from the protein tree (fig. 2) were generally consistent with the results of the DNA-based analyses (fig. 1*B*; see also fig. 2B in the Supplementary Material online). We used six or five constraints in the protein analyses because node e, which was not consistent between the DNA and protein trees, had to be excluded from these calculations. Two estimates that were markedly different between the DNA- and protein-based approaches were the estimates of node a for the split of the glaucophyte (1,719 MYA [protein] vs. 1,558 MYA [DNA]) from the red and green algae, and of node 1 for the split of the red and green algae (1,668 MYA [protein] versus 1,474 MYA [DNA]). These results reflect variation in the branch lengths that unite the glaucophyte to the cyanobacterial outgroup and to the remaining algal plastids (see fig. 2). This discordance may be resolved with increased sampling of glaucophytes or the addition of more data to the protein analysis.

Agreement with the Fossil Record and Assessment of Alternative Hypotheses

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

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

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

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

The Long Pause in Algal Radiation

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

Supplementary Material

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

Acknowledgments

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

Literature Cited

Anbar, A. D., and A. H. Knoll. 2002. Proterozoic ocean chemistry and evolution: a bioinorganic bridge? Science **297**:1137–1142.

Andersson, J. O., and A. J. Roger. 2002. A cyanobacterial gene in nonphotosynthetic protists—an early chloroplast acquisition in eukaryotes? Curr. Biol. **12**:115–119.

Baldauf, S. L. 2003. The deep roots of eukaryotes. Science **300**:1703–1706.

Baldauf, S. L., and J. D. Palmer. 1990. Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. Nature **344**:262–265.

Baldauf, S. L., J. R. Manhart, and J. D. Palmer. 1990. Different fates of the chloroplast *tuf*A gene following its transfer to the nucleus in green algae. Proc. Natl. Acad. Sci. USA **87**: 5317–5321.

Baldauf, S. L., A. J. Roger, I. Wenk-Siefert, and W. F. Doolittle. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

Barfod, G. H., F. Albarede, A. H. Knoll, S. Xiao, P. Telouk, R. Frei, and J. Baker. 2002. New Lu-Hf and Pb-Pb age constraints on the earliest animal fossils. Earth Planet Sci. Lett. **201**:203–212.

Benton, M. J., and F. J. Ayala. 2003. Dating the tree of life. Science **300**:1698–1700.

Bhattacharya, D., and L. Medlin. 1995. The phylogeny of plastids: a review based on comparisons of small-subunit ribosomal RNA coding regions. J. Phycol. **31**:489–498.

Bhattacharya, D., and K. Weber. 1997. The actin gene of the Glaucocystophyte *Cyanophora paradoxa*: analysis of the coding region and introns, and an actin phylogeny of eukaryotes. Curr. Genet. **31**:439–446.

Bhattacharya, D., H. S. Yoon, and J. D. Hackett. 2004. Photosynthetic eukaryotes unite: endosymbiosis connects the dots. BioEssays: **26**:50–60.

Bowe, L. M., G. Coat, and C. W. dePamphilis. 2000. Phylogeny of seed plants based on all three genomic compartments: extant gymnosperms are monophyletic and Gnetales' closest relatives are conifers. Proc. Natl. Acad. Sci. USA **97**:4092–4097.

Brasier, M. D., O. R. Green, A. P. Jephcoat, A. K. Kleppe, M. J. Van Kranendonk, J. F. Lindsay, A. Steele, and N. V. Grassineau. 2002. Questioning the evidence for Earth's oldest fossils. Nature **416**:76–81.

Butterfield, N. J. 2001. Paleobiology of the late Mesoproterozoic (ca. 1200 Ma) hunting formation, Somerset Island, Arctic Canada. Precam. Res. **111**:235–256.

Cavalier-Smith, T. 1986. The kingdon Chromista: origin and systematics. Pp. 309–347 *in* F. E. Round and D. J. Chapman, eds., Progress in phycological research. Biopress, Bristol, U.K.

———. 1998. A revised six-kingdom system of life. Biol. Rev. Camb. Philos. Soc. **73**:203–266.

———. 1999. Principles of protein and lipid targeting in secondary symbiogenesis: Euglenoid, Dinoflagellate, and Sporozoan plastid origins and the eukaryote family tree. J. Eukaryot. Microbiol. **46**:347–366.

———. 2002. The neomuran origin of archaebacteria, the negibacterial root of the universal tree and bacterial megaclassification. Int. J. Syst. Evol. Microbiol. **52**:7–76.

Crane, P. R., E. M. Friis, and K. R. Pedersen. 1995. The origin and early diversification of angiosperms. Nature **374**:27–33.

Cutler, D. J. 2000. Estimating divergence times in the presence of an overdispersed molecular clock. Mol. Biol. Evol. **17**: 1647–1660.

Delsuc, F., M. J. Phillips, and D. Penny. 2003. Comment on "Hexapod orgins: monophyletic or paraphyletic?" Science **301**:1482.

Delwiche, C. F., and J. D. Palmer. 1997. The origin of plastids and their spread via secondary symbiosis. Pp. 53–86 *in* D. Bhattacharya, ed., Origins of algae and their plastids. Springer-Verlag, Vienna, Austria.

Doyle, J. A. 1998. Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. Mol. Phylogenet. Evol. **9**:448–462.

Fast, N. M., J. C. Kissinger, D. S. Roos, and P. J. Keeling. 2001. Nuclear-encoded, plastid-targeted genes suggest a single common origin for apicomplexan and dinoflagellate plastids. Mol. Biol. Evol. **18**:418–426.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

———. 2002. PHYLIP (Phylogeny Inference Package) 3.6. Department of Genetics, University of Washington, Seattle, Wash.

Feng, D. F., G. Cho, and R. F. Doolittle. 1997. Determining divergence times with a protein clock: update and reevaluation. Proc. Natl. Acad. Sci. USA **94**:13028–13033.

Garcia-Ruiz, J.M., S. T. Hyde, A. M. Carnerup, A. G. Christy, M. J. Van Krankendonk, and N. J. Welham. 2003. Self-assembled silica-carbonate structures and detection of ancient microfossils. Science **302**:1194–1197.

Gilbert, D. G. 1995. SeqPup, A biological sequence editor and analysis program for Macintosh computer. Indiana University, Bloomington.

Gillespie, W. H., G. W. Rothwell, S. E. Scheckler. 1981. The earliest seeds. Nature **293**:462–464.

Goremykin, V. V., S. Hansmann, and W. F. Martin. 1997. Evolutionary analysis of 58 proteins encoded in six completely sequenced chloroplast genomes: revised molecular estimates of two seed plant divergence times. Plant Syst. Evol. **206**:337–351.

Graham, L. D., and L. W. Wilcox. 2000. Algae. Prentice-Hall, Upper Saddle River, N.J.

Gray, M. W., B. F. Lang, and R. Cedergren et al. (15 co-authors). 1998. Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. **26**:865–878.

Harper, J. T., and P. J. Keeling. 2003. Nucleus-encoded, plastid-targeted glyceraldehyde-3-phosphate dehydrogenase (GAP-DH) indicates a single origin for chromalveolate plastids. Mol. Biol. Evol. **20**:1730–1735.

Heckman, D. S., D. M. Geiser, B. R. Eidell, R. L. Stauffer, N. L. Kardos, and S. B. Hedges. 2001. Molecular evidence for the early colonization of land by fungi and plants. Science **293**:1129–1133.

Hoffman, P. F., A. J. Kaufman, G. P. Halverson, and D. P. Schrag. 1998. A Neoproterozoic snowball earth. Science **281**:1342–1346.

Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.

Javaux, E. J., A. H. Knoll, and M. R. Walter. 2001. Morphological and ecological complexity in early eukaryotic ecosystems. Nature **412**:66–69.

Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche. 2001. The closest living relatives of land plants. Science **294**:2351–2353.

Kenrick, P., and P. R. Crane. 1997. The origin and early evolution of plants on land. Nature **389**:33–39.

Knoll, A. H. 1992. The early evolution of eukaryotes: a geological perspective. Science **256**:622–627.

———. 2003. Life on a young planet. Princeton University Press, Princeton, N.J.

Martin, W., T. Rujan, E. Richly, A. Hansen, S. Cornelsen, T. Lins, D. Leister, B. Stoebe, M. Hasegawa, and D. Penny. 2002. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. Proc. Natl. Acad. Sci. USA **99**:12246–12251.

Martin, W., C. Rotte, M. Hoffmeister, U. Theissen, G. Gelius-Dietrich, S. Ahr, and K. Henze. 2003. Early cell evolution, eukaryotes, anoxia, sulfide, oxygen, fungi first (?), and a tree of genomes revisited. IUBMB Life **55**:193–204.

Maul, J. E., J. W. Lilly, L. Cui, C. W. dePamphilis, W. Miller, E. H. Harris, and D. B. Stern. 2002. The *Chlamydomonas reinhardtii* plastid chromosome: islands of genes in a sea of repeats. Plant Cell **14**:2659–2679.

Medlin, L. K., W. H. C. F. Kooistra, D. Potter, G. W. Saunders, and R. A. Andersson. 1997. Phylogenetic relationships of the "golden algae" (haptophytes, heterokont chromophytes) and their plastids. Pp. 187–219 *in* D. Bhattacharya, ed., Origins of algae and their plastids. Springer-Verlag, Vienna, Austria

Moreira, D., H. Le Guyader, and H. Phillippe. 2000. The origin of red algae and the evolution of chloroplasts. Nature **405**: 69–72.

Nei, M., P. Xu, and G. Glazko. 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA **98**:2497–2502.

Nozaki, H., M. Matsuzaki, M. Takahara, O. Misumi, H. Kuroiwa, M. Hasegawa, I. T. Shin, Y. Kohara, N. Ogasawara, and T. Kuroiwa. 2003. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. J. Mol. Evol. **56**:485–497.

Ohta, N., M. Matsuzaki, O. Misumi, S. Y. Miyagishima, H. Nozaki, K. Tanaka, T. Shin-I, Y. Kohara, and T. Kuroiwa. 2003. Complete sequence and analysis of the plastid genome of the unicellular red alga *Cyanidioschyzon merolae*. DNA Res. **10**:67–77.

Oliveira, M. C., and D. Bhattacharya. 2000. Phylogeny of the Bangiophycidae (Rhodophyta) and the secondary endosymbiotic origin of algal plastids. Am. J. Bot. **87**:482–492.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. **28**:171–185.

Pinto, G., P. Albertano, C. Ciniglia, S. Cozzolino, A. Pollio, H. S. Yoon, and D. Bhattacharya. 2003. Comparative approaches to the taxonomy of the genus *Galdieria merola* (Cyanidiales, Rhodophyta). Cryptogamie Algol. **24**:13–32.

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

Sanderson, M. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. **14**:1218–1231.

———. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301–302.

Schmidt, H. A., K. Strimmer, M. Vingron, and A. von Haeseler. 2002. Tree-Puzzle: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics **18**:502–504.

Schopf, J. W., A. B. Kudryavtsev, D. G. Agresti, T. J. Wdowiak, and A. D. Czaja. 2002. Laser-raman imagery of Earth's earliest fossils. Nature **416**:73–76.

Soltis, P. S., D. E. Soltis, V. Savolainen, P. R. Crane, and T. G. Barraclough. 2002. Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. Proc. Natl. Acad. Sci. USA **99**:4430–4435.

Swofford, D. L. 2002. PAUP*: Phylogenetic analysis using parsimony (* and other methods) 4.0b8. Sinauer Associates, Sunderland, Mass.

Valentin, K., and K. Zetsche. 1990. Rubisco genes indicate a close phylogenetic relation between the plastids of Chromophyta and Rhodophyta. Plant Mol. Biol. **15**:575–584.

Wang, D. Y., S. Kumar, and S. B. Hedges. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. Proc. R. Soc. Lond. Ser. B. Biol. Sci. **266**:163–171.

Westall, F., M. J. De Witb, J. Dann, S. Van Der Gaast, C. E. J. De Ronded, and D. Gerneke. 2001. Early Archean fossil bacteria and biofilms in hydrothermally-influenced sediments from the Barberton greenstone belt, South Africa. Precam. Res. **106**:93–116.

Whelan, S., P. Liò, and N. Goldman. 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet. **17**:262–272.

Wilde, S. A., J. W. Valley, W. H. Peck, and C. M. Graham. 2001. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. Nature **409**: 175–178.

Xiao, S., Y. Zhang, and A. H. Knoll. 1998. Three-dimensional preservation of algae and animal embryos in a Neoproterozoic phosphorite. Nature **391**:553–558.

Yoon, H. S., J. D. Hackett, and D. Bhattacharya. 2002. A single origin of the peridinin- and fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis. Proc. Natl. Acad. Sci. USA **99**:11724–11729.

Yoon, H. S., J. D. Hackett, G. Pinto, and D. Bhattacharya. 2002. The single, ancient origin of chromist plastids. Proc. Natl. Acad. Sci. USA **99**:15507–15512.

Dan Graur, Associate Editor

Accepted December 9, 2003