Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 4641–4644, May 1993
Evolution

Phylogenetic analysis of *rbcL* sequences identifies *Acorus calamus* as the primal extant monocotyledon

(molecular systematics/phylogenetics/ribulose-bisphosphate carboxylase)

MELVIN R. DUVALL*, GERALD H. LEARN, JR., LUIS E. EGUIARTE†, AND MICHAEL T. CLEGG

Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124

Contributed by Michael T. Clegg, March 12, 1993

ABSTRACT **The identity of the oldest lineage of monocotyledons is a subject of debate. Alternative interpretations of morphological homologies are variously consistent with proposals that species of Alismatanae, Dioscoreales, or Melanthiales were the earliest descendants of the first monocotyledons. We present phylogenetic analyses based on DNA sequences of the plastid locus *rbcL* in which *Acorus calamus*, an herb with unspecialized floral features and of uncertain affinities, is supported as a member of the oldest extant lineage of monocotyledons. This conclusion is consistent with a substantial body of morphological, anatomical, and embryological evidence and offers an explanation for the failure to identify any close relationship between *Acorus* and other genera.**

Primal monocotyledons are thought to have diverged from dicotyledons 100–150 million years ago by one of several proposed evolutionary scenarios (1). Proponents of the once widely accepted "phyllode theory" held that the nature of the typically linear monocotyledonous leaf indicated homology to an elaborated dicotyledonous petiole (ref. 2; but see refs. 3 and 4 for countering arguments). This presumed homology supported an ancestry from aquatic dicotyledons with reduced leaves, leading to the conclusion that the most ancient monocotyledons were to be found among aquatic Alismatanae (water plantains and allied species). Alternatively, supporters of what has been designated, somewhat loosely, as the "magnolianan hypothesis" (so-called because of a presumed ancestry from pre-Magnolianae) propose that the atypical, petiolate, reticulately veined leaves of dioscoreoid monocotyledons (yams and related species) are homologous to those of dicotyledons (1). This scenario suggests that the primal monocotyledons, presumably now extinct, first gave rise to the ancestors of Lilianae rather than of Alismatanae. Other lineages, including the melanthioid lilies, have also been proposed to be primal (5). Since the monocotyledons are believed to share a single common ancestry (1, 6–8), only one of these hypotheses can be correct. Clearly there is no consensus on which lineages of monocotyledons are distinguished as the most ancient, nor are the relationships among the major groups well resolved. Morphological homologies are too ambiguous to decide among the competing hypotheses. Until these issues are clarified, the origin and subsequent evolutionary history that produced the 50,000 species of monocotyledons will remain an enigma.

The failure to resolve the higher-order taxonomic relationships among monocotyledons is due to an inability to identify evolutionarily meaningful homologies between long-diverged species. Analysis of molecular characters has the potential to circumvent this difficulty. However, the only inclusive investigation of the molecular systematics of monocotyledons is our broad analysis of plastid DNA sequences sampled across the monocotyledons from 116 species, which suggested a phylogenetic framework for the entire group (7). Here we present an exhaustive analysis of molecular evidence from selected representatives in all major groups (all superorders excluding Triuridanae; see explanation below) to identify the primal extant lineage and to outline higher-order relationships.

One species in this study, *Acorus calamus*, or sweet flag, deserves particular comment. Systematists have sought to establish the affinities of *Acorus*, which exhibits a unique combination of characteristics. Two of the most remarkable of these are (*i*) a cellular type of endosperm development which is otherwise unknown in monocotyledons but common in dicotyledons (9) and (*ii*) a reported absence of double fertilization, unlike virtually all angiosperms (10). The flowers of *Acorus*, which are bisexual, perigoniate, and trimerous, resemble those of primitive Aranae, the sister superorder to the Alismatanae, but also those of primitive monocotyledons in general. Nonetheless, these characteristics, together with a superficial resemblance in leaf morphology to the Australian aroid *Gymnostachys anceps* have been used to justify the traditional classification of *Acorus* in Aranae (1, 2, 5, 11, 12).

Extensive investigations of the numerous unique embryological, anatomical, and floral characteristics of *Acorus* argue against a close relationship with Aranae (9, 13–16). However, no other group of monocotyledons displays more convincing evidence of alliance. The only proposed alternative candidates are species of Typhales with which *Acorus* shares several similarities of anatomy, morphology, and embryology (9) and a host relationship with the same fungal parasite (17). Because of its extraordinary features and problematic systematic history *Acorus calamus* was included in our analysis, together with putatively related species (Aranae, 10 species; Typhales, 2 species).

The plastid locus *rbcL* encodes the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). RuBisCO, the most abundant soluble leaf protein, is an indispensable component of photosynthetic carbon metabolism and is thus ubiquitous among green plants. (Note that one superorder of monocotyledons, Triuridanae, is composed exclusively of nonphotosynthetic species that are unlikely to possess a functional, and therefore phylogenetically meaningful, copy of the *rbcL* sequence.) Furthermore, because the locus is highly conserved and lacks introns, the alignment of sequences of *rbcL* from long-diverged species is a straightforward matter. Finally, the utility of *rbcL* data for phylogenetic reconstructions has been amply demonstrated (6–8, 18–23).

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.

Abbreviation: MLE, maximum likelihood estimation.

*Present address: Nucleic Acid Facility, 1184 Molecular Biology Building, Iowa State University, Ames, IA 50011-3260.
†Present address: Centro de Ecologia, Universidad Nacional Autonoma de Mexico, Apdo. Postal 70-275, 04510 Mexico, D.F., Mexico.

MATERIALS AND METHODS

DNA sequences of a 1327-base-pair portion of the *rbcL* gene were determined for 10 species: *Anchomanes difformis, Ariopsis peltata, Lasia spinosa, Montrichardia arborescens, Symplocarpus foetidus, Xanthosoma sagittifolium* (Araceae), *Eriocaulon microcephalum* (Eriocaulaceae), *Hosta rectifolia* (Funkiaceae), *Ludovia integrifolia* (Cyclanthaceae), and *Orthrosanthus polystachyus* (Iridaceae). Total genomic DNA was isolated from dried shoots (24) of *Eriocaulon microcephalum* and from fresh leaves (35) of the other 9 species. Approximately 1 μg of each DNA preparation was used to provide template for *Taq*-mediated amplification of the *rbcL* gene using the protocol provided with *Taq* DNA polymerase by the supplier (Promega) and synthetic primers homologous to conserved regions between *Zea mays* and *Spinacia oleracea*. The forward primer is homologous to the first 27 base pairs of *rbcL* and the reverse primer corresponds to positions 1355–1378 on the complementary strand. Single-stranded DNA for direct sequencing was prepared in subsequent amplifications. Primers for sequencing were designed and made available by G. Zurawski (DNAX, Palo Alto, CA) based on oligonucleotide sequences conserved between the *rbcL* sequences of *Zea mays* and *Spinacia oleracea*. By these methods, 1327 base pairs of *rbcL* were determined for 10 species representing four superorders of monocotyledons. The *rbcL* sequences for these 10 species have been deposited in the GenBank data base under accession numbers L10246–L10255.

Also analyzed were 21 previously reported *rbcL* sequences (6–8, 18, 25): 17 species sampled across the monocotyledons, including one species each of Alismatanae, Dioscoreales, and Melanthiales; and 4 species of dicotyledonous Nymphaeanae (water lilies). Sequences were analyzed from positions 31–1350 to eliminate missing sequence data at the extreme ends of some sequences.

A maximum likelihood estimation (MLE) of the phylogeny of these species was made using the DNAML computer program version 3.42 of PHYLIP (26) on a Sun SparcStation IPX. MLE is less biased than other methods by the heterogeneity of substitution rates between different lineages (27) that has been observed for *rbcL* among monocotyledons (28). The program was repeatedly executed with different input orders, a transition/transversion ratio of 2.0, and local branch swapping. In the initial analytical phase, 10 replicated executions were performed. Further replications were then executed until the topology with the largest log-likelihood score from the initial phase was produced from two more data input orders (7 more replications, for a total of 17).

DNA sequence data were also analyzed by the parsimony method implemented by PAUP 3.0s (29) on a Macintosh IIfx microcomputer. This method was selected for further analyses because it is less computer-intensive than the DNAML program. The program was initially executed with 250 randomly determined replications of the input order by using the "tree-bisection reconnection" (TBR) branch-swapping strategy (29) with comparisons of over 16.4×10^6 rearrangements from all most-parsimonious trees. A second analysis was performed in which all shortest trees and those one step longer were determined by using TBR branch swapping and two replications of the input order. A third analysis was executed with TBR swapping on all of the trees from the first two analyses, saving trees up to two steps longer than the shortest trees. Note that these analyses did not recover any shorter trees than the first analysis. Finally, a bootstrap analysis was performed with 1000 subsamplings of the data matrix by "nearest-neighbor interchange" branch swapping.

RESULTS AND DISCUSSION

For the 27 monocotyledonous species, of the 1320 nucleotide sites analyzed, 437 (33%) were polymorphic, with 71, 49, and 317 polymorphisms observed at first, second, and third codon positions, respectively.

The tree with the largest log-likelihood score (−8733.036) from the MLE analysis is given (Fig. 1). Tests (30) of the log-likelihood scores of all topologies produced by the MLE method indicated no significant differences in scores. A strict consensus of these trees is given (Fig. 2). Parsimony analysis produced 40 equally parsimonious trees of length 1237 (retention index = 0.547), the strict consensus of which is given together with bootstrap values (Fig. 3). The single difference between the resolved portions of the two consensus trees (Figs. 2 and 3) is the position of *Sagittaria graminea* (Alismatanae), which is embedded in Aranae in the parsimony analysis (Fig. 3). The phylogenetic position of this species in the parsimony analysis is clearly an artifact of sampling too few taxa, as was shown by analyses including *rbcL* data from other species of Alismatanae (not shown).

In the optimized topologies produced by both maximum-likelihood and parsimony methods there is a basal split between *A. calamus* and the remaining 26 monocotyledons,

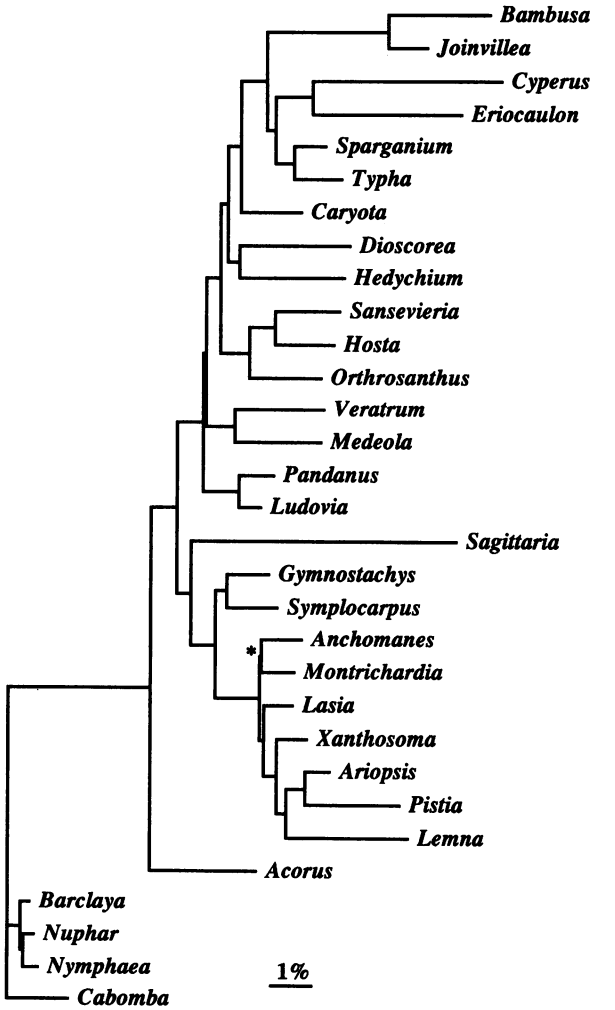

FIG. 1. The maximum likelihood topology with the largest (least negative) log-likelihood score of 17 replicated maximum likelihood analyses of *rbcL* sequences from 27 species of monocotyledons and 4 outgroup species of Nymphaeanae (water lilies). Branch lengths correspond to evolutionary distances. Scale bar indicates 1% sequence divergence. All branches are significantly positive at the $P = 0.05$ level with the exception of that marked with an asterisk.

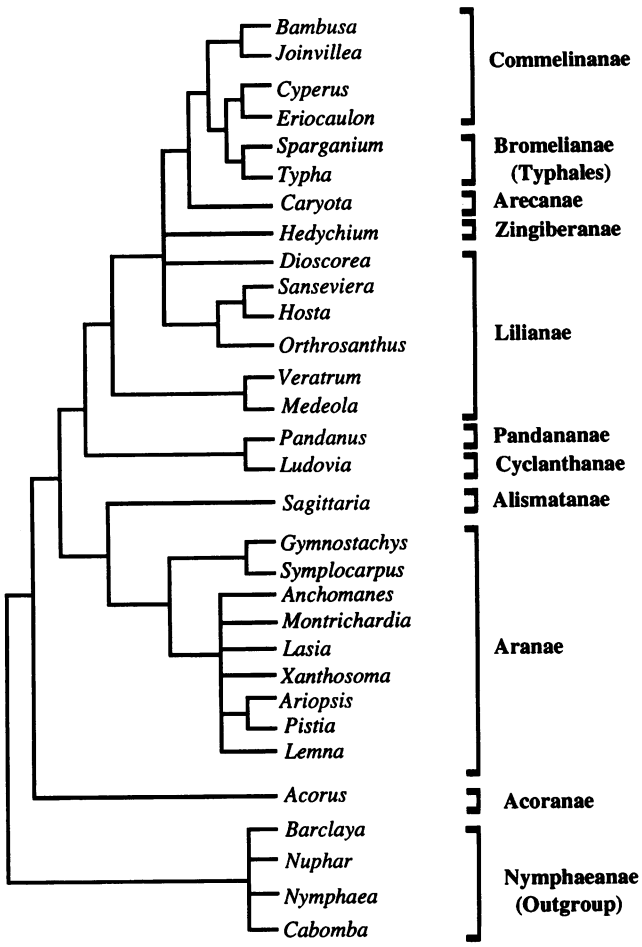

FIG. 2. Strict consensus of topologies produced in 17 replicates of MLE for the 31 species analyzed in Fig. 1. Log-likelihood scores of these 17 topologies were not significantly different from each other. Branch lengths are arbitrary.

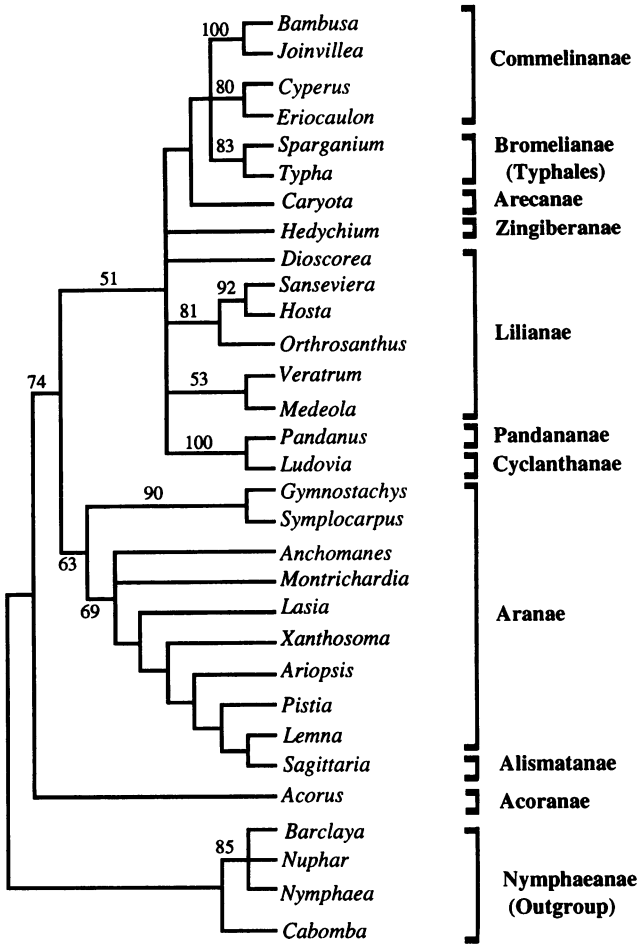

FIG. 3. Strict consensus tree of 40 equally parsimonious trees of length 1237 found by parsimony analysis of the species analyzed in Fig. 1. Bootstrap values are shown for nodes supported in >50% of 1000 bootstrap subsamplings of the data matrix. Branch lengths are arbitrary.

including 9 species of Aranae and two species of Typhales. This split suggests that species of *Acorus* are the most archaic of extant monocotyledons.

The branch segregating *A. calamus* from other monocotyledons is supported in 530 trees one step longer than the most-parsimonious trees and also in 2979 trees that are two steps longer. This branch is found in 740 of the 1000 bootstrap trees. The next best supported position for *Acorus*, in a clade with *Sagittaria graminea* (Alismatanae), is found in only 118 of the 1000 bootstrap trees. The magnitude of this difference is a strong indication of the statistical support for the basal split between *Acorus* and the other monocotyledons. Fifteen nucleotide substitutions segregate the *rbcL* sequences of the other 26 species of monocotyledons from that of *A. calamus* and the four species of Nymphaeales (Table 1). Four of these substitutions (positions 213, 930, 951, and 1149, all silent substitutions), can be inferred to be unambiguous events. An additional two substitutions (positions 537 and 767) each exhibit a single autapomorphic reversal (i.e., a reversal exhibited in only one species) on the tree of Fig. 1. The previously proposed alliances of *A. calamus* with Aranae and Typhales are not supported by the *rbcL* data. Constrained topologies in which *A. calamus* is forced into a position immediately basal to Aranae or Typhales add 7 and 28 steps to the length of the tree, respectively.

The *rbcL* phylogeny reconstructions for the remaining 26 species of monocotyledons (Fig. 1) suggest (*i*) an early divergence and close relationship between Alismatanae and Aranae; (*ii*) later divergences of dioscoreoid and melanthioid Lilianae and paraphyly of this superorder; (*iii*) a well-supported alliance between Pandananae (screw pines) and Cyclanthanae (Panama hat palms); and (*iv*) most recent

Table 1. Positions of nucleotide substitutions in *rbcL* supporting the basal split between *A. calamus* and other monocotyledons

Nucleotide position	Codon position	Substitution
84	3	A → C
213*	3	C → T
228	3	C → T
265	1	C → G
412	1	C → T
474	3	G → A
537†	3	G → A
564	3	A → G
582	3	T → C
612	3	G → A
767†	2	G → T
930*	3	C → T
951*	3	G → A
1020	3	A → G
1149*	3	C → T

Substitutions are shown only through position 1209, the last position of sequence data for the previously published sequences from Nymphaeanae.

*Unambiguous (nonhomoplastic) mutation.

†Substitutions that exhibit only one autapomorphic reversal.

divergences of species of Commelinanae, Bromelianae, and Zingiberanae, as is widely recognized (1, 2). Interestingly, with the exception of the node segregating Pandananae and Cyclanthanae as sister taxa, none of the relationships between superordinal groups is as well supported by the *rbcL* data as is the result for *A. calamus*. This may reflect sparse sampling among the later-diverged clades or rapid radiation of the major lineages of monocotyledons with the effect of fewer substitutions in the deep branches of the tree (Fig. 1).

The dicotyledons selected in our analysis define the ancestral characters for the monocotyledons and are thus critical. The herbaceous Magnolianae and, in particular, Nymphaeanae, have long been proposed to be the descendants of the dicotyledons from which the monocotyledons arose (1, 2). Recent cladistic analyses also support this hypothesis and identify several dicotyledonous taxa as members of the sister clade to the monocotyledons (8, 31). Alternative phylogenetic analyses were thus conducted employing other Nymphaeanae, Piperales, and Chloranthaceae. These selections gave identical results (not shown) with respect to the position of *A. calamus*.

Absolute verification of the phylogenetic position of *A. calamus* in an *rbcL* phylogeny would require analysis of sequence data from every living species. Realistically, however, the inclusion of data from other putatively primitive species should indicate the robustness of our result. Consequently, species from the lineages recognized as archaic were incorporated into the analysis (data not shown), including Alismatanae (three species), Melanthiales (three species), and Dioscoreales (two species). In these analyses *A. calamus* still occupied a singular basal branch (6–8).

None of the competing phylogenetic hypotheses identifying species of Alismatanae, Dioscoreales, or Melanthiales as direct descendants of the first monocotyledons are supported by the *rbcL* data. Alteration of the topology shown in Fig. 3 so that the basal monocotyledon is *Sagittaria graminea* (Alismatanae), *Dioscorea polygonoides*, or *Veratrum parviflorum* (Melanthiales) increases the length of the tree by 9, 22, and 24 steps, respectively.

We propose an alternative hypothesis, which we call the Acoranan hypothesis, implicating the species of *Acorus* as an isolated group that represents the most ancient surviving lineage of the ancestral monocotyledons. This hypothesis is a reasonable, if unanticipated, result, consistent with a failure to identify other monocotyledons with which *Acorus* shares a recent common ancestry. Furthermore, a suite of morphological and anatomical similarities that have been noted as numerous puzzling convergences between *Acorus* and the dicotyledonous Piperales (9) are more logically and simply explained under the Acoranan hypothesis as retained ancestral characters.

Under the Acoranan hypothesis, certain predictions can be made. (*i*) Discovery of fossils attributable to *Acorus* might be expected—in spite of the paludal habitat of the genus, which is not conducive to fossilization—perhaps dating back to the divergence of the monocotyledons. Although allies of *Acorus* are represented in the fossil record of the past 55 million years (32), older fossils marking the time of divergence of the monocotyledons have not been recognized. Perhaps a directed search might confirm this prediction. (*ii*) Phylogenetic analyses of DNA sequences from other loci might be expected to demonstrate the same pattern of relationships reported here. (*iii*) Some aspect of the biology of *Acorus* might be presumed responsible for its persistence. A notable feature of the genus is the presence of ethereal oil cells which contain a complex mixture of volatile compounds with documented insecticidal and bacteriocidal (33) and allelopathic (34) action. Whether these compounds are directly responsible for the long-term success of the genus has not been ascertained. However, given the fresh insight afforded by the Acoranan hypothesis, these and other predictions can be investigated to clarify our understanding of the origin and evolution of the monocotyledons.

We thank J. C. French (Rutgers University, Piscataway, NJ) for his generous donation of DNA extracts from species of Araceae. For plant material we thank A. W. H. Damman (University of Connecticut, Storrs), P. Goldblatt (Missouri Botanical Garden, St. Louis), and A. C. Sanders (University of California Botanic Gardens, Riverside). This work was supported by grants from the National Science Foundation (BSR-9002321 to M.R.D.) and the National Institutes of Health (GM45144 to M.T.C.).

1. Dahlgren, R., Clifford, H. & Yeo, P. (1985) *The Families of the Monocotyledons* (Springer, New York).
2. Cronquist, A. (1981) *An Integrated System of Classification of Flowering Plants* (Columbia Univ. Press, New York).
3. Kaplan, D. R. (1970) *Am. J. Bot.* **57**, 331–361.
4. Kaplan, D. R. (1973) *Q. Rev. Biol.* **48**, 437–457.
5. Thorne, R. F. (1992) *Aliso* **13**, 365–389.
6. Duvall, M., Clegg, M., Chase, M., Clark, W. D., Kress, W. J., Hills, H., Eguiarte, L., Smith, J., Gaut, B., Zimmer, E. & Learn, G., Jr. (1993) *Ann. Mo. Bot. Gard.* **80**, in press.
7. Duvall, M., Chase, M., Soltis, D. & Clegg, M. (1993) in *Experimental and Molecular Approaches to Plant Biosystematics*, ed. Hoch P. (Missouri Bot. Gard., St. Louis), in press.
8. Chase, M., Soltis, D., Olmstead, R., Morgan, D., Les, D., Mishler, B., Duvall, M., Price, R., Hills, H., Qiu, Y.-L., Kron, K., Rettig, J., Conti, E., Palmer, J., Manhart, J., Sytsma, K., Michaels, H., Kress, W. J., Donoghue, M., Clark, W. D., Hedren, M., Gaut, B., Jansen, R., Kim, K.-J., Wimpee, C., Smith, J., Furnier, G., Straus, S., Xiang, Q.-Y., Plunkett, G., Soltis, P., Swensen, S., Eguiarte, L., Learn, G., Jr., Barrett, S., Graham, S., Dayanandan, S. & Albert, V. (1993) *Ann. Mo. Bot. Gard.* **80**, in press.
9. Grayum, M. (1987) *Taxon* **36**, 723–729.
10. Buell, M. (1938) *Bot. Gaz. (Crawfordsville)* **99**, 556–568.
11. Takhtajan, A. (1980) *Bot. Rev.* **46**, 225–359.
12. Thorne, R. (1983) *Nord. J. Bot.* **3**, 85–117.
13. Grayum, M. (1984) Thesis (University of Massachusetts, Amherst).
14. Eyde, R., Nicolson, D. & Sherwin, P. (1967) *Am. J. Bot.* **54**, 478–497.
15. French, J. (1985) *Am. J. Bot.* **72**, 472–486.
16. French, J. (1987) *Bot. Gaz.* **148**, 198–208.
17. Savile, D. (1979) *Bot. Rev.* **45**, 377–503.
18. Les, D., Garvin, D. & Wimpee, C. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10119–10123.
19. Doebley, J., Durbin, M., Golenberg, E., Clegg, M. & Ma, D. (1990) *Evolution* **44**, 1097–1108.
20. Golenberg, E., Giannasi, D., Clegg, M., Smiley, C., Durbin, M., Henderson, D. & Zurawski, G. (1990) *Nature (London)* **344**, 656–658.
21. Albert, V., Williams, S. & Chase, M. (1992) *Science* **257**, 1491–1495.
22. Donoghue, M., Olmstead, R., Smith, J. & Palmer, J. (1992) *Ann. Mo. Bot. Gard.* **79**, 333–345.
23. Giannasi, D., Zurawski, G., Learn, G., Jr., & Clegg, M. (1992) *Syst. Bot.* **17**, 1–15.
24. Rogers, S. & Bendich, A. (1985) *Plant Mol. Biol.* **5**, 69–76.
25. Wilson, M., Gaut, B. & Clegg, M. (1990) *Mol. Biol. Evol.* **7**, 303–314.
26. Felsenstein, J. (1991) *Phylogeny Inference Package* (Univ. Wash. Press, Seattle).
27. Felsenstein, J. (1981) *J. Mol. Evol.* **17**, 368–376.
28. Gaut, B., Muse, S., Clark, W. D. & Clegg, M. (1992) *J. Mol. Evol.* **35**, 292–303.
29. Swofford, D. (1990) *Phylogenetic Analysis Using Parsimony Version 3.0s* (Illinois Natl. Hist. Surv., Champaign, IL).
30. Kishino, H. & Hasegawa, M. (1989) *J. Mol. Evol.* **29**, 170–179.
31. Donoghue, M. & Doyle, J. (1989) in *The Hierarchy of Life*, eds. Fernholm, R., Bremer, K. & Jornvall, H. (Elsevier, Amsterdam), pp. 181–193.
32. Crepet, W. (1977) *Rev. Palaeobot. Palynol.* **25**, 241–252.
33. Brown, D. (1988) *Aroids* (Timber Press, Portland, OR).
34. Dellagreca, M., Monaco, P., Previtera, L., Aliotta, G., Pinto, G. & Pollio, A. (1989) *Phytochemistry* **28**, 2319–2321.
35. Doyle, J. & Doyle, J. (1987) *Phytochem. Bull.* **19**, 11–15.