1-1-1993

Phylogenetic Hypotheses for the Monocotyledons Constructed from *rbc*L Sequence Data

James F. Smith
Boise State University

PHYLOGENETIC HYPOTHESES FOR THE MONOCOTYLEDONS CONSTRUCTED FROM *rbc*L SEQUENCE DATA[1]

Melvin R. Duvall,[2] Michael T. Clegg,[2] Mark W. Chase,[3,4] W. Dennis Clark,[5] W. John Kress,[6] Harold G. Hills,[3] Luis E. Eguiarte,[7] James F. Smith,[8,9] Brandon S. Gaut,[2] Elizabeth A. Zimmer,[8] and Gerald H. Learn, Jr.[2]

Abstract

DNA sequences for the plastid locus that encodes the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (*rbc*L) were determined for 18 species of monocotyledons in 15 families. These data were analyzed together with sequences for 60 other monocot species in a total of 52 families by the maximum likelihood method producing one, presumably optimal, topology. An additional 26 species were added (104 total monocot species) and analyzed by the parsimony method with an outgroup of 18 dicot species producing 109 trees of 3,932 steps. The *rbc*L data show at least moderate support for seven lineages corresponding to the following orders, superorders, or combinations: Arecanae; Asparagales (excluding Hypoxidaceae) plus Iridaceae; Cyclanthanae plus Pandananae; Dioscoreales; Orchidales; Typhales; and Zingiberanae. Six clades corresponding to families or genera are well supported, including: Agavaceae, Asphodelaceae, Bromeliaceae, Hypoxidaceae, Poaceae, and *Tradescantia*. The two, earliest diverging multispecies clades in our *rbc*L phylogenies, Alismatanae and Aranae, are only weakly supported, and Bromelianae, Commelinanae, and Lilianae are paraphyletic. In all analyses *Acorus calamus* is phylogenetically isolated as the sister species to the remaining species of monocotyledons.

Innovations for the manipulation of nucleic acids and advances in computer technology now permit phylogenetic analysis of homologous sequences of DNA from large numbers of organisms. These base-to-base comparisons of nucleotides afford the highest possible resolution of inherited mutations in DNA molecules and can be applied to questions of higher-order plant systematics. A locus that has been selected for such studies by molecular systematists is the plastid gene that encodes the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase (*rbc*L). Phylogenetic reconstructions for lineages of monocotyledons based on *rbc*L sequence data have been attempted previously for: (1) Arecaceae (Wilson et al., 1990); (2) Bromeliaceae (Clark et al., in prep.); (3) Poaceae (Doebley et al., 1990); and (4) Zingiberales (Smith et al., 1993). These initial attempts met with a disappointing lack of resolution, in some cases because of the low substitution rate of *rbc*L, but strongly suggested that these data would have greater phylogenetic utility when applied to more divergent lineages. Our goal here was to reconstruct higher-order phylogenetic relationships among the mono-

[1] For computer analyses we thank G. J. Olsen (University of Illinois, Urbana), R. Overbeek (Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illlinois), the Concurrent Supercomputing Consortium, R. Whitkus, P. Lowe (both, University of California, Riverside) and M. Berres (University of Minnesota, St. Paul). We also thank D. Jarrell (University of California, Riverside), R. Olmstead (University of Colorado, Boulder), and others for unpublished *rbc*L sequences; J. French (Rutgers University, Piscataway, New Jersey) who provided DNA extracts for Araceae; and W. Armstrong (Palomar College, San Marcos, California) for identification of one species. This work was supported by grants from National Science Foundation to M. Duvall (BSR-9002321), M. Chase (BSR-8906496), and W. D. Clark (BSR-8904637); National Institutes of Health (GM45144) to M. Clegg; Scholarly Studies Program, Smithsonian Institution to W. J. Kress and E. Zimmer; and American Orchid Society to M. W. Chase.

[2] Department of Botany and Plant Sciences, University of California, Riverside, California 92521, U.S.A.
[3] Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.
[4] Present address: Royal Botanic Gardens, Kew, Richmond, Surrey TW9 3AB, U.K.
[5] Department of Botany, Arizona State University, Tempe, Arizona 85287, U.S.A.
[6] Department of Botany, NHB-166, National Museum of Natural History, Smithsonian Institution, Washington, D.C. 20560, U.S.A.
[7] Centro de Ecología, Universidad Nacional Autonóma de México, Apdo. Postal 70-275, 04510 México, D.F., México.
[8] Laboratory of Molecular Systematics, Smithsonian Institution, Suitland, Maryland 20746, U.S.A.
[9] Present address: Department of Biology, Boise State University, Boise, Idaho 83725, U.S.A.

Ann. Missouri Bot. Gard. 80: 607–619. 1993.

cotyledons using DNA sequence data from the *rbc*L locus for a set of species that represented the breadth of taxonomic diversity in the group.

Materials and Methods

DNA sequences of *rbc*L from 104 species in 52 of 104 monocot families were analyzed (see Appendix in this issue for species, locations of vouchers, sequencing methods, extent of sequences, and authors). Eighteen of those sequences (Table 1) were produced specifically for this study by the following methods. Total DNA was extracted from fresh leaf tissue by the method of Doyle & Doyle (1987) for 16 species, the method of Palmer (1986) for *Cyperus alternifolius*, and the method of Jarrell et al. (1992) for *Lemna minuscula*. Approximately 1 μg of each DNA preparation was used to provide template for *Taq*-mediated amplification of the *rbc*L gene using the protocol provided with *Taq* DNA polymerase by the supplier (Promega Corp.). The primers for the reaction were two highly conserved sequences from the *rbc*L coding region of *Spinacia oleracea*. The 5′ primer consisted of the first 30 base pairs of that sequence and the 3′ (reverse) primer corresponded to positions 1351 to 1380 on the complementary strand. Amplifications utilizing these primers produced 1,320 base pairs of the coding sequence of *rbc*L. Primers that are external to the coding sequence occasionally resulted in successful amplifications of more of the *rbc*L coding sequence. For *Sparganium americanum*, a 27 base pair sequence at the promoter region for *rbc*L derived from positions 53742–53768 in the plastid genome of *Oryza sativa* was used as the 5′ primer. For *Gymnostachys anceps*, *Tradescantia* aff. *pallida*, and *Typha latifolia*, a 24 base pair sequence at a ribosome control site downstream of the coding sequence derived from positions 59146–59123 (complement) in the plastid genome of *Nicotiana tabacum* was used as the 3′ primer. Up to 1,428 base pairs of *rbc*L were produced, depending on the primers that were used in the amplification reactions. For the *rbc*L sequences from the monocotyledons analyzed here, 5.7% of the sequence data was missing largely because of the use of internal primers in the amplification reactions.

Single-stranded DNA was produced in a second round of amplification using the double-stranded product as template and the two primers individually. These single-stranded products were precipitated (7.5% polyethylene glycol-8000, 0.94 M NaCl), the DNA pellets were washed twice with ethanol, and sequencing was accomplished by the dideoxynucleotide chain-termination method using a set of conserved, synthetic, internal *rbc*L primers (obtained from G. Zurawski, DNAX Corp.). Both the *rbc*L coding strand and the complementary strand were sequenced for all species. The 18 *rbc*L sequences produced for this study were entered into "GenBank" under the accession numbers listed in Table 1.

The maximum likelihood method of phylogenetic analysis was selected for use here because it is less biased by the heterogeneous nucleotide substitution rates that have been observed for *rbc*L among different lineages of monocotyledons (Gaut et al., 1992) than are other methods (Felsenstein, 1981). Furthermore, under this method, likelihood scores may be calculated for alternate topologies and subjected to comparative tests of statistical significance. However, only a subset of the entire *rbc*L database could be analyzed since the method is computer-intensive. [To establish an estimate of the computer burden, a maximum likelihood analysis of 79 *rbc*L sequences using the program, DNAML in PHYLIP 3.42 (Felsenstein, 1991), was executed on a CRAY "Y-MP8/864" supercomputer. The analysis spanned 11 days with a run time of about 51 CPU hr. At this speed, ten executions of the program with different input orders (typically recommended) of the 79 *rbc*L sequences would have required over three months of real time utilizing more than 20 CPU *days* on this machine.] Consequently, 78 of the original 104 monocot species were selected for maximum likelihood analysis with *Saururus cernuus* as the outgroup (79 species total). Each of the following taxa, Agavaceae, Arecaceae, Bromeliaceae, Nolinaceae, Poaceae, and Zingiberales, invariably constituted a monophyletic lineage in preliminary phylogenetic analyses that included multiple species in each. The subset was thus selected to contain representatives from each of the 52 families while excluding some of the multiple species in these six taxa, preserving the potential for investigating higher order relationships. The species that were retained in the subset are: Agavaceae—*Manfreda maculosa* and *Yucca recurvifolia*; Arecaceae—*Caryota mitis*, *Phoenix reclinata*, and *Serenoa repens*; Bromeliaceae—*Aechmea chantinii* and *Tillandsia elizabethae*; Nolinaceae—*Nolina* (*Beaucarnea*) *recurvata*; Poaceae—*Oryza sativa* and *Zea mays*; and Zingiberales—*Calathea loeseneri*, *Costus barbatus*, *Globba curtisii*, *Hedychium gardnerianum*, *Heliconia latispatha*, *Maranta leuconeura*, *Musa cavendishii*, *Orchidantha fimbriata*, *Phenakospermum guyannense*, *Ravenala madagascariensis*, and *Tapeinocheilos ananassae*.

TABLE 1. Eighteen species for which *rbc*L was sequenced for this study. Vouchers (herbaria), accession numbers for the *rbc*L sequences submitted to GenBank, superordinal, ordinal, and familial alignments are given. Identities for the remaining 104 species that were analyzed may be found in the Appendix for this issue.

Species	Vouchers (herbaria)	Family	GenBank accession
Aranae			
Arales			
Acorus calamus L.	*French 232* (CH)	Acoraceae	M91625
Gymnostachys anceps R. Br.	*Howard 4325* (FTG)	Araceae	M91629
Pistia stratiotes L.	*French 233* (CH)	Araceae	M96963
Lemna minuscula Herter	*Duvall 19920501* (UCR)	Lemnaceae	M91630
Bromelianae			
Typhales			
Typha latifolia L.	*Bradley 24974* (GMUF)	Typhaceae	M91634
Sparganium americanum Nutt.	*Chase 257* (NCU)	Sparganiaceae	M91633
Commelinanae			
Commelinales			
Tradescantia aff. *pallida*	*Bradley 24980* (GMUF)	Commelinaceae	L05041
Tradescantia zebrina hort. ex Bosse	*Bradley 24980* (GMUF)	Commelinaceae	L05042
Cyperales			
Cyperus alternifolius L.	*Duvall 19920602* (UCR)	Cyperaceae	M91627
Lilianae			
Asparagales			
Aloe vera (L.) Burm.	*Bradley 24977* (GMUF)	Asphodelaceae	L05029
Haworthia subfasciata Baker	*Bradley 24978* (GMUF)	Asphodelaceae	L05035
Chlorophytum comosum (Thunb.) Jacques	*Bradley 7331* (GMUF)	Anthericaceae	L05031
Clivia miniata Regel	*Bradley 24976* (GMUF)	Amaryllidaceae	L05032
Nolina (*Beaucarnea*) *recurvata* (Lem.) Hemsl.	*Peterson 12606* (US)	Nolinaceae	L05030
Liliales			
Iris ×*germanicum* L.	*Bradley 25976* (GMUF)	Iridaceae	L05037
Medeola virginiana (L.) Merrill	*Bradley 24972* (GMUF)	Uvulariaceae	M91613
Zingiberanae			
Zingiberales			
Maranta leuconeura E. Morris	*Bradley 24979* (GMUF)	Marantaceae	L05040
Hedychium gardnerianum Ker Gawl.	*Bradley 24975* (GMUF)	Zingiberaceae	M91628

The analysis was executed on the fastest existing computer, the Touchstone Delta Parallel Processing Supercomputer, using "fastDNAml version 1.0.3" (Olsen et al., 1992). Thirty-three replications of the analysis were performed with different randomly determined orders of input. The "categories option," which permits specification of different substitution rates by codon position, was invoked using relative rates of 1.00, 0.85, and 5.80 for first, second, and third codon positions, respectively. These values are based on reported substitution rates at each codon position (Clegg et al., 1986; Ritland & Clegg, 1987). Regional and local branch swapping were employed. This portion of the analysis, which included data from 74 of the 79 species, consumed 125 hours of computer time.

After the initial analysis was performed, the remaining five species (*Aletris farinosa*, *Burmannia biflora*, *Sparganium americanum*, *Stegolepis allenii*, and *Typha latifolia*; total species: 79) with equivocal phylogenetic positions in preliminary analyses were added to the optimal topology from the previous step with 50 (of 120 possible) random

TABLE 2. Distribution of polymorphic sites, constant sites, and sites shared by two or more species (i.e., informative sites), among codon positions of *rbc*L. Sites are calculated for 1,428 base pairs and 79 species. Base substitutions were optimized on the topology of Figures 4 and 5.

	Codon position			Total sites
	First	Second	Third	
Polymorphic sites	221	210	430	861
Constant sites	255	266	46	567
Informative sites	137	124	394	655

input orders. Rearrangements of the optimal topology from this step were performed to arrive at a final result. Kishino & Hasegawa (1989) tests of the likelihood scores associated with topologies produced by each step of the analysis (available from M. Duvall on request) were performed.

Bootstrap and decay analyses were implemented with PAUP Version 3.0s (Swofford, 1991) on a Macintosh IIfx for the *rbc*L data set from 78 species of monocots. A set of 18 trees of length 3,117 was determined with more than ten replications of the input order. Trees one step longer (486 trees) and two steps longer (6,237 trees) were also determined. (An attempt to determine all trees up to four steps longer using a Macintosh Quadra 700 was aborted after 13.8 days of continuous execution. The analysis produced 14,873 trees occupying over 6 Megabytes of memory and was estimated to be 10% complete. Available computer memory was the limiting factor so that the time to perform read-write operations of treefiles from external memory became prohibitive.) Bootstrap analysis with 200 subsamples of the original data matrix was performed with local (i.e., "nearest neighbor interchange" or "NNI") branch swapping.

Three tests to ascertain the effect of constraining topologies so that selected species occupied monophyletic clades were conducted. The selected species groups were: (1) Commelinanae (13 spp.); (2) Alismatanae (4 spp.), and Aranae (4 spp.), together with *Acorus calamus*; and (3) each of nine superorders sensu Dahlgren et al. (1985). These constraints were imposed on parsimony analyses with 10 replications of the input order and NNI branch swapping.

For pragmatic reasons the parsimony method was selected for larger-scale analysis of *rbc*L sequences (104 monocot species). Eighteen dicot species were selected as an outgroup as suggested by a more inclusive analysis of angiosperms (Chase et al., 1993). Ten replications of the input order were executed with the "steepest descent" option invoked and global ("tree-bisection and reconnection" or "TBR") branch swapping employed. All equally most-parsimonious trees were saved.

RESULTS

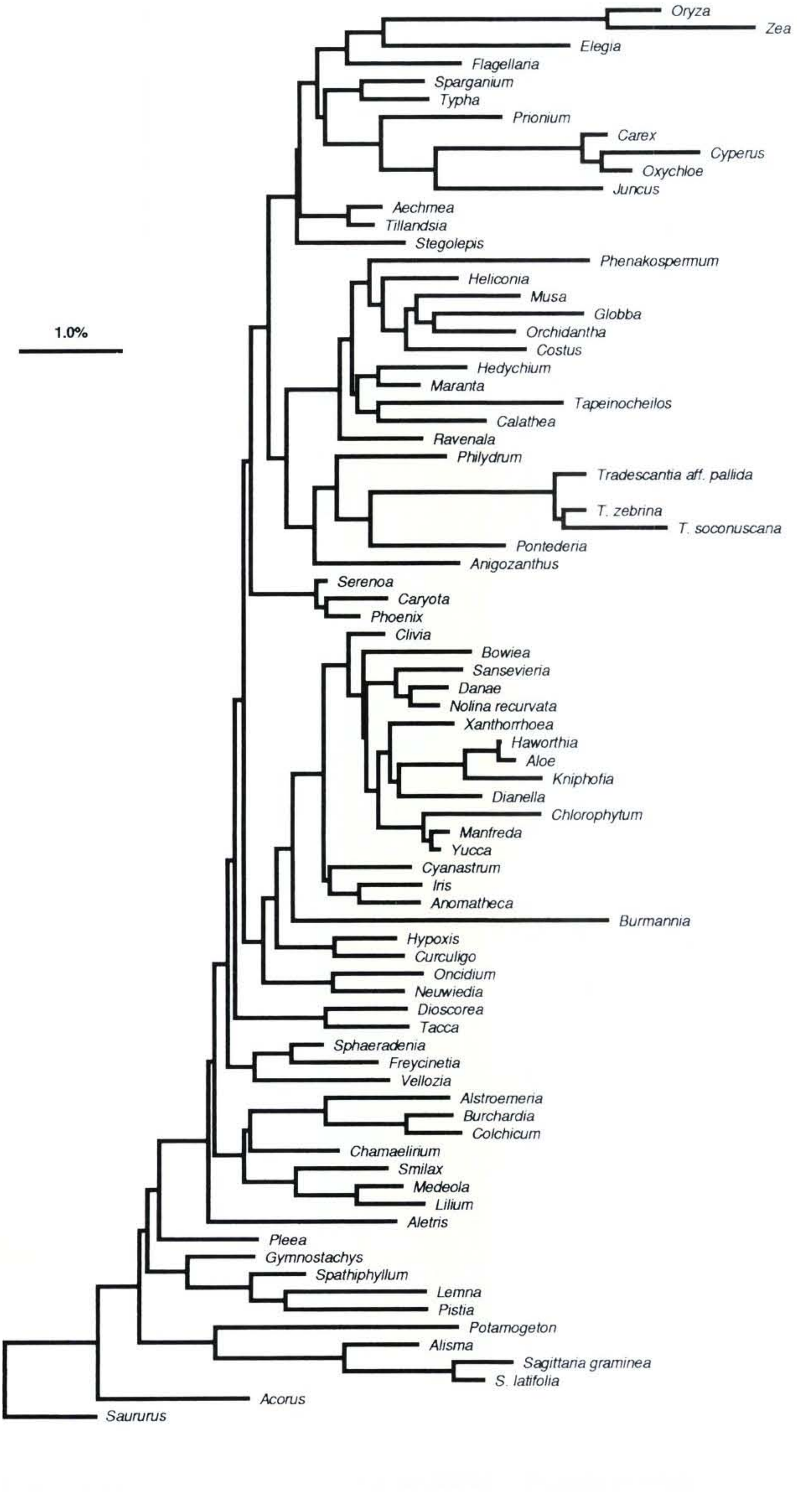
Of the 1,428 bases analyzed for 79 species, 861 were polymorphic and 50% of these were at third codon positions (Table 2). Of the 861 polymorphic sites, 655 were shared by two or more species and of these, 60% were at third codon positions.

The results of our phylogenetic analyses show greater congruence with the taxonomic system of Dahlgren et al. (1985) than with other contemporary systems (reviewed in Goldberg, 1989). References to taxa in this report thus follow that system.

The topology resulting from maximum likelihood analysis (Fig. 1) had an associated likelihood score of −18,878.21. Note that other topologies exist which have likelihood scores that are not significantly different from that of Figure 1. Bootstrap values and decay indices for analyses of 79 species are given (Fig. 2, Table 3). Recall that only trees up to two steps longer than the shortest trees were determined so that clades supported with decay indices of two may also be supported at higher,

→

FIGURE 1. Topology for 78 monocotyledons and the outgroup dicotyledon *Saururus cernuus* (79 spp. total) produced by the maximum likelihood method with a log likelihood of −18,878.21. The analysis was executed using the computer program fastDNAml 1.0.3 (Olsen et al., 1992), on the Touchstone Delta Parallel Processing Supercomputer, the fastest existing computer. The initial phase of the analysis included 74 of these species, employed regional and local branch swapping, specified substitution rates of 1.00, 0.85, and 5.80 for first, second, and third codon position substitutions, respectively, under the "categories" option, was performed with 33 replications of the input order, and consumed more than 125 hours of supercomputer time. Five additional species, *Aletris farinosa*, *Burmannia biflora*, *Sparganium americanum*, *Stegolepis allenii*, and *Typha latifolia*, were added more than 50 times to the initial topology with the largest log likelihood score. Additional rearrangements of the tree with the largest likelihood score produced the tree shown here. Scale bar indicates 1.0% maximum likelihood distance units as calculated under the assumptions specified above. Species are indicated by genera only. For complete binomials see Appendix in this issue.


1.0%
Oryza
Zea
Elegia
Flagellaria
Sparganium
Typha
Prionium
Carex
Cyperus
Oxychloe
Juncus
Aechmea
Tillandsia
Stegolepis
Phenakospermum
Heliconia
Musa
Globba
Orchidantha
Costus
Hedychium
Maranta
Tapeinocheilos
Calathea
Ravenala
Philydrum
Tradescantia aff. pallida
T. zebrina
T. soconuscana
Pontederia
Anigozanthus
Serenoa
Caryota
Phoenix
Clivia
Bowiea
Sansevieria
Danae
Nolina recurvata
Xanthorrhoea
Haworthia
Aloe
Kniphofia
Dianella
Chlorophytum
Manfreda
Yucca
Cyanastrum
Iris
Anomatheca
Burmannia
Hypoxis
Curculigo
Oncidium
Neuwiedia
Dioscorea
Tacca
Sphaeradenia
Freycinetia
Vellozia
Alstroemeria
Burchardia
Colchicum
Chamaelirium
Smilax
Medeola
Lilium
Aletris
Pleea
Gymnostachys
Spathiphyllum
Lemna
Pistia
Potamogeton
Alisma
Sagittaria graminea
S. latifolia
Acorus
Saururus

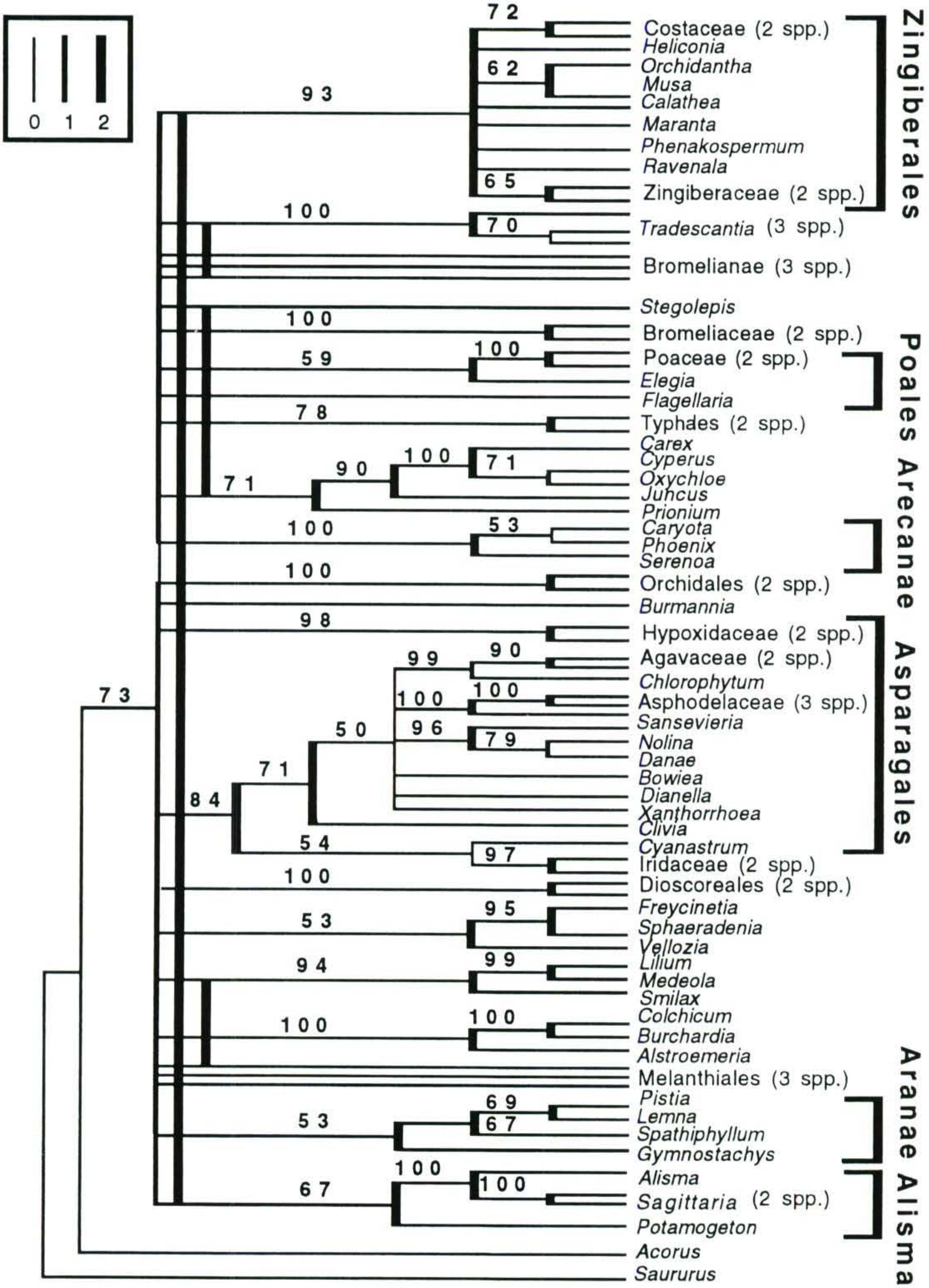

0 1 2
72
62
93
65
100
70
Costaceae (2 spp.)
Heliconia
Orchidantha
Musa
Calathea
Maranta
Phenakospermum
Ravenala
Zingiberaceae (2 spp.)
Tradescantia (3 spp.)
Bromelianae (3 spp.)
Zingiberales
Stegolepis
100
Bromeliaceae (2 spp.)
100
59
Poaceae (2 spp.)
Elegia
Flagellaria
78
Typhales (2 spp.)
Carex
100
71
Cyperus
90
Oxychloe
71
Juncus
Prionium
100
53
Caryota
Phoenix
Serenoa
100
Orchidales (2 spp.)
Burmannia
Poales Arecanae
98
Hypoxidaceae (2 spp.)
99
90
Agavaceae (2 spp.)
Chlorophytum
73
100
100
Asphodelaceae (3 spp.)
Sansevieria
50
96
79
Nolina
Danae
71
Bowiea
84
Dianella
Xanthorrhoea
Clivia
54
97
Cyanastrum
Iridaceae (2 spp.)
100
Dioscoreales (2 spp.)
Asparagales
95
Freycinetia
53
Sphaeradenia
Vellozia
94
99
Lilium
Medeola
Smilax
100
100
Colchicum
Burchardia
Alstroemeria
Melanthiales (3 spp.)
69
Pistia
53
67
Lemna
Spathiphyllum
Gymnostachys
100
Alisma
67
100
Sagittaria (2 spp.)
Potamogeton
Acorus
Saururus
Aranae Alismatanae

TABLE 3. Support for clades corresponding to taxa sensu Dahlgren et al. (1985) by *rbc*L data for 78 species of monocotyledons. Parenthetical numbers immediately following taxa indicate the number of species analyzed. Bootstrap values (200 replicates) and decay indices (up to two steps longer than the most parsimonious topologies) are given. Clades supported with decay indices of two may also be supported at a higher, undetermined, decay index. Also given are the numbers of substitution events supporting each lineage and the number of nonhomoplastic synapomorphies (appearing parenthetically following branch lengths) in the maximum likelihood topology (Fig. 1).

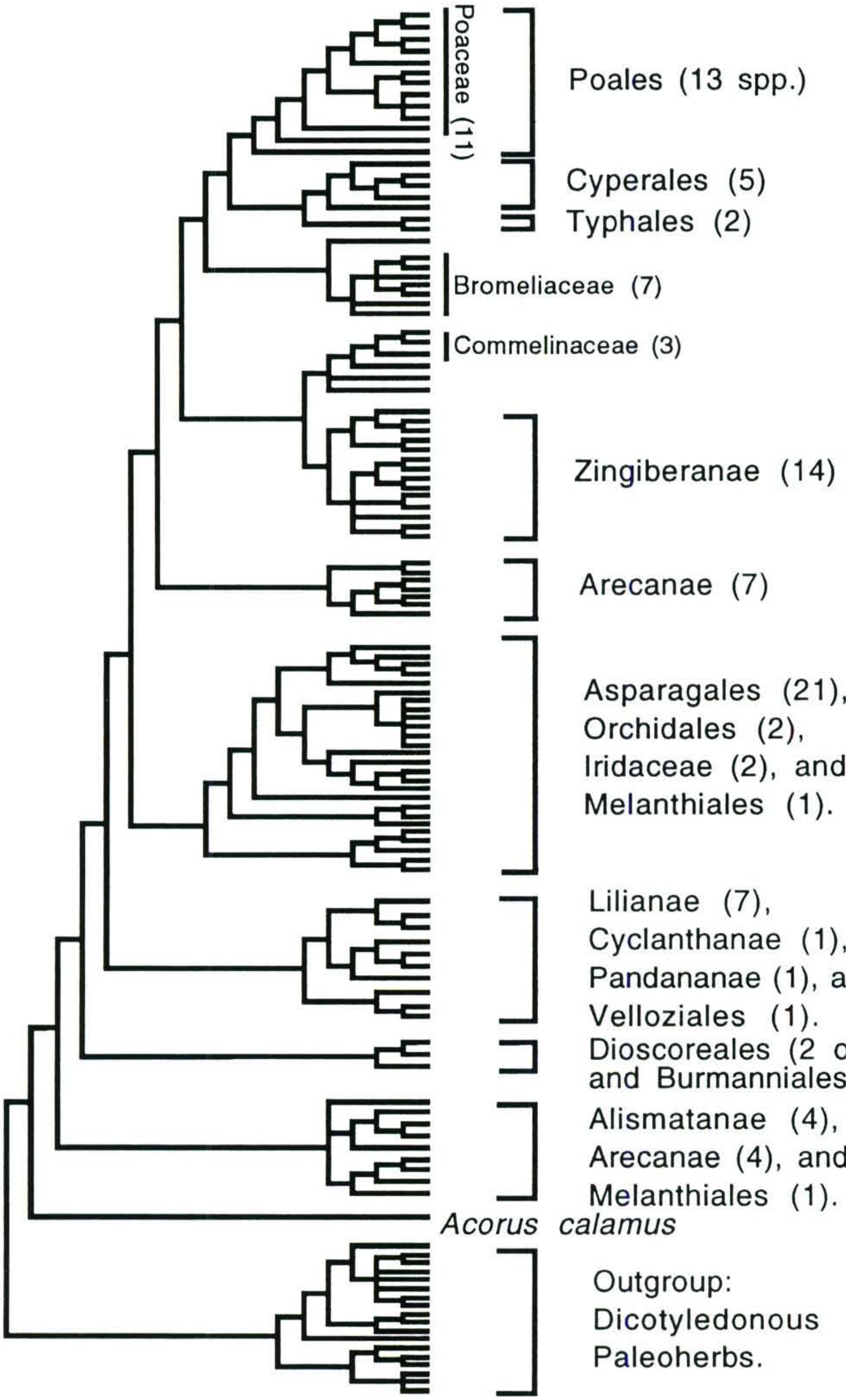
Taxa (number of species)	Bootstrap values (%)	Decay indices	Lengths
Alismatanae (4)	67	≥ 2	37 (3)
Aranae (4)	53	≥ 2	14 (1)
Arecanae (3)	100	≥ 2	31 (4)
Bromelianae (8)	< 50	0	—
Bromeliaceae (2)	100	≥ 2	14 (0)
Typhales (2)	78	≥ 2	21 (11)
Commelinanae (13)	< 50	0	—
Cyperales (5)	71	≥ 2	26 (1)
Poales (3 of 4)	59	≥ 2	21 (2)
Poaceae (2)	100	≥ 2	66 (7)
Tradescantia (3)	100	≥ 2	48 (3)
Cyclanthanae (1) plus Pandananae (1)	95	≥ 2	13 (1)
Lilianae (32)	< 50	0	—
Asparagales (14 of 16) plus Iridaceae (2)	84	≥ 2	13 (1)
Agavaceae (2)	90	≥ 2	3 (0)
Asphodelaceae (3)	100	≥ 2	21 (1)
Hypoxidaceae (2)	100	≥ 2	18 (1)
Dioscoreales (2 of 3)	98	≥ 2	27 (2)
Orchidales (2)	100	≥ 2	23 (1)
Zingiberanae (11)	93	≥ 2	20 (1)
Costaceae (2)	72	≥ 2	—
Zingiberaceae (2)	65	≥ 2	—
Monocotyledons excluding *Acorus calamus* (77 spp.)	73	≥ 2	13 (1)

undetermined, values. Seven lineages found in the maximum likelihood tree that have associated bootstrap values at or above 78% and decay indices greater than or equal to two correspond to orders, superorders, or combinations of these. These are: (1) Arecanae (3 spp.); (2) Asparagales (excluding Hypoxidaceae: 14 spp.) plus Iridaceae (2 spp.); (3) Cyclanthanae plus Pandananae (1 spp. each); (4) Dioscoreales (2 of 3 spp.); (5) Orchidales (2 spp.); (6) Typhales (2 spp.); and (7) Zingiberanae (11 spp.). Six clades of confamilial or congeneric species are supported by the *rbc*L data at a bootstrap value of at least 90% and a decay index of at least two including: Agavaceae, Asphodelaceae, Bromeliaceae, Hypoxidaceae, Poaceae, and *Tradescantia*.

Parsimony analysis of 122 species produced a set of 109 equally parsimonious trees of 3,932 steps over the 1,428 characters (see overview of strict consensus tree, Fig. 3). These trees have consistency indices (excluding uninformative characters) of 0.267 (the low value reflecting the large number of species) and retention indices of 0.633. One of these 109 trees was arbitrarily selected and is given in detail (Figs. 4, 5) to enumerate genera and illustrate comparative branch lengths.

←

FIGURE 2. Majority rule (50%) consensus tree for 79 species. The bootstrap analysis was conducted using PAUP 3.0s with 200 bootstrap subsamples of the data matrix. Percentage values for those branches occurring in at least 50% of the bootstrap topologies are shown. Selected taxa are identified. Branch lengths are arbitrary. Decay values up to two steps longer are indicated as vertical bars of varying thickness overlaid on the bootstrap topology. The thickest lines indicate clades supported in trees at least two steps longer. The thinnest lines indicate clades supported only in maximum parsimony trees.


Poaceae (11)
Poales (13 spp.)
Cyperales (5)
Typhales (2)
Bromeliaceae (7)
Commelinaceae (3)
Zingiberanae (14)
Arecanae (7)
Asparagales (21),
Orchidales (2),
Iridaceae (2), and
Melanthiales (1).
Lilianae (7),
Cyclanthanae (1),
Pandananae (1), and
Velloziales (1).
Dioscoreales (2 of 3),
and Burmanniales (1).
Alismatanae (4),
Arecanae (4), and
Melanthiales (1).
Acorus calamus
Outgroup:
Dicotyledonous
Paleoherbs.

The maximum likelihood tree (79 spp., Fig. 1) and the resolved portions of the consensus tree produced by parsimony (122 spp., Fig. 3) are largely congruent with respect to the constituent species of the seven lineages listed above and the order of divergence of those lineages. Exceptions are: (1) Dioscoreales (excluding *Smilax glauca*) diverge earlier in the parsimony trees; (2) one species each of Cyclanthanae, Pandananae, and Velloziales make up an isolated clade in the maximum likelihood tree, which is found embedded within a clade of seven species of Lilianae in the parsimony tree; (3) *Burmannia biflora* is found in Asparagales in the maximum likelihood tree but with two species of Dioscoreales in the parsimony trees; and (4) *Aletris farinosa* occupies an isolated clade near the base of the maximum likelihood tree but is embedded within the clade of Asparagales in the parsimony trees. In both the maximum likelihood and parsimony trees, two species of Hypoxidaceae are included in Asparagales consistent with Dahlgren et al. (1985).

Discussion

The order of divergence of the seven major lineages in our phylogenetic analyses is in general agreement with widely accepted views on the evolution of the monocotyledons. Recognition of the early divergences of Alismatanae (Cronquist, 1981; Dahlgren et al., 1985) and Dioscoreales (Dahlgren et al., 1985), and the affinities between the Alismatanae and the Aranae (Dahlgren et al., 1985; Grayum, 1991) correlate with the position of species from these superorders near the base of our molecular phylogenies. Postulated later divergences of Arecanae (Doyle, 1973), Bromelianae, Commelinanae, and Zingiberanae (Cronquist, 1981) are also consistent with our analyses.

In general, deep branches in the tree (Figs. 1, 4, and 5) are shorter than the terminal branches, indicating fewer nucleotide changes along the former. This relationship suggests either that the substitution rate during evolutionary radiations of the monocotyledons was unusually slow, that sampling bias occurred in part because of extinction events, or that the original radiations occurred rapidly. The fossil record, and especially that of fossil pollen, for angiosperms in general (Doyle & Hickey, 1976) and probably for the monocotyledons as well (Doyle, 1973) is certainly consistent with the hypothesis of rapid radiation.

Acorus calamus occupies a unique, basal position in all the trees generated for this study. Although *Acorus* was traditionally classified in Araceae because of superficial morphological similarities with the Australian aroid *Gymnostachys anceps*, aroid authorities have recently acknowledged long-recognized difficulties with this classification and, based on a substantial body of evidence (reviewed in Grayum, 1987), proposed removal of *Acorus* to a monogeneric family. A tree (not shown) constrained to include *Acorus* and four species of Arales as monophyletic that was analyzed over the subset of 79 species is 25 steps longer than the shortest trees. Our analysis thus indicates that removal of *Acorus* from the Araceae is consistent with a more parsimonious phylogenetic hypothesis and further offers an explanation for the failure to identify synapomorphies of this genus with other monocot species, if, as we suggest, *Acorus* is distinguished as the most basal extant lineage of monocotyledons.

Several other aspects of the tree of Figure 1, not already considered in the accompanying reports in this issue, are of interest. We have included *rbc*L sequences from 21 species in 12 of 30 families in the large order Asparagales (estimated 5,000 spp. or 10% of monocotyledons) sensu Dahlgren et al. (1985). These species are all found in a single monophyletic clade in both maximum likelihood and parsimony trees, together with only six other species: (1) *Aletris farinosa*, (Melanthiaceae); (2 and 3) two species of Iridaceae: *Anomatheca laxa* and *Iris* × *germanicum*; (4 and 5) two species of Orchidales: *Neuwiedia veratrifolia* (Apostasiaceae) and *Oncidium excavatum* (Orchidaceae); and (6) *Burmannia biflora*. Further, the clade of Asparagales (excluding Hypoxidaceae) is at least moderately supported by boot-

←

Figure 3. Overview of the strict consensus of 109 equally parsimonious trees for 104 monocots and 18 dicots. These trees are of length 3,932, have consistency indices excluding uninformative characters of 0.267, and retention indices of 0.633. Note that the unresolved portions of this topology are within the terminal clades corresponding to: (1) Bromeliaceae; (2) that containing Commelinaceae and three species of Bromelianae; (3) Zingiberanae; (4) Asparagales; and (5) that consisting of Alismatanae, Arecanae, and one species of Melanthiales. The terminal species are identified in the Appendix in this issue and as genera in Figures 4 and 5. The number of species analyzed from each taxon is given parenthetically.

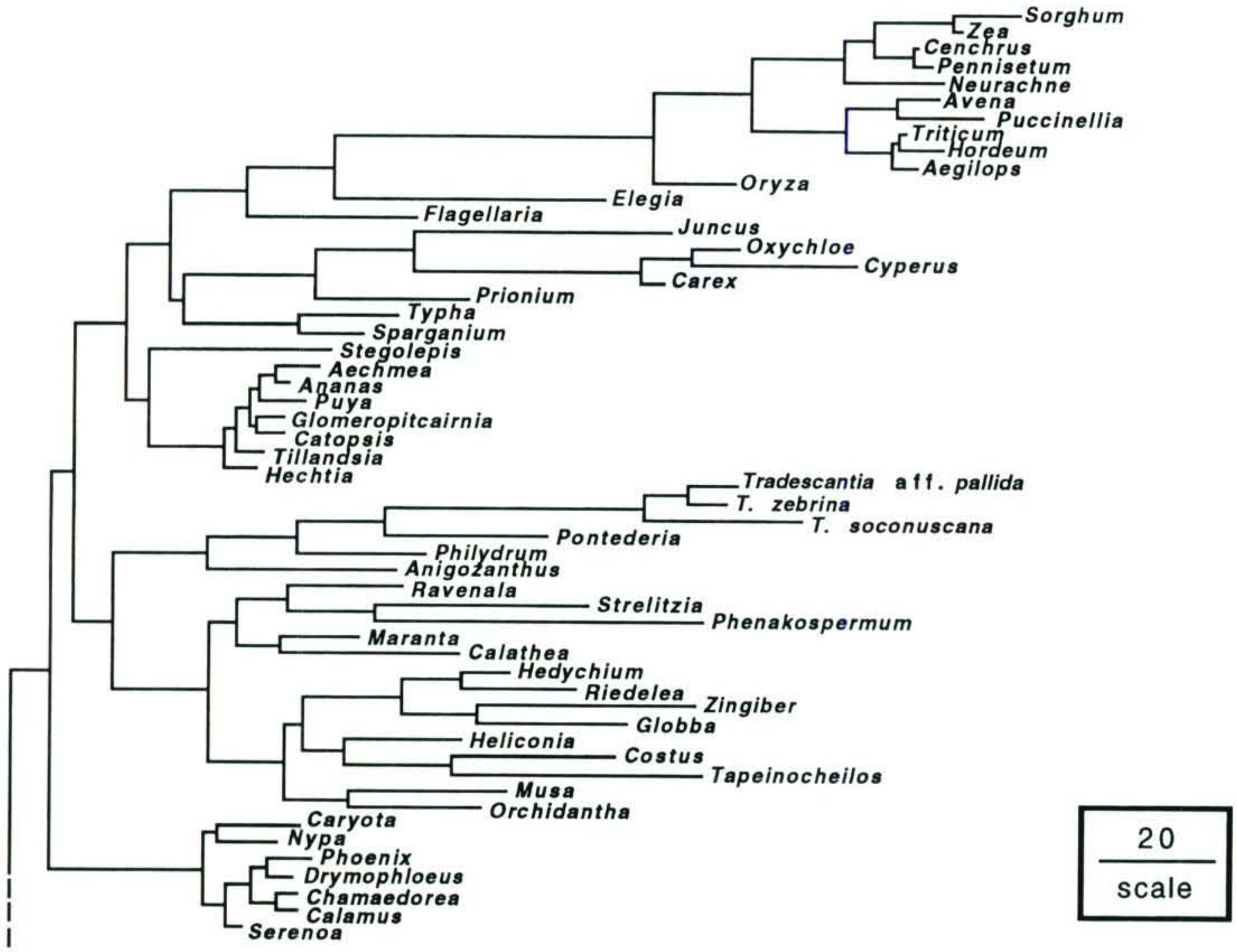


FIGURE 4. Terminal portion of one arbitrarily selected tree from the set of 109 trees (see consensus tree, Fig. 3) of length 3,932, consistency index of 0.267, and retention index of 0.633. This portion of the tree includes species of Commelinanae, Bromelianae, Zingiberanae, and Arecanae (see Fig. 5 for the remainder of the tree). Branch lengths correspond to the number of substitutions optimized along the branches. The scale bar is proportional to a branch length of 20 steps.

strap (84%) and decay (at least two steps longer) analyses. Duvall et al. (in review) have analyzed *rbc*L sequence data from four more species in three additional families of Asparagales that further support the common ancestry of this large order, and they note that the alliance between Iridaceae and Asparagales has morphological and anatomical support.

The historical treatment of Arecanae, Cyclanthanae, and Pandananae as at least marginally related taxa has been contradicted by subsequent taxonomic schemes that treat each as an unrelated superorder (Thorne, 1983) or as separate superorders with a loose alliance between Arecanae and Cyclanthanae (Dahlgren et al., 1985). Phylogenetic hypotheses based on the *rbc*L data support *Freycinetia* (Pandananae) and *Sphaeradenia* (Cyclanthanae) as sister species only distantly related to seven species of Arecanae. This arrangement suggests that Pandananae and Cyclanthanae are more ancient groups than Arecanae. Inclusion of *rbc*L data for a second species of Pandananae, *Pandanus veitchii* does not alter this result (Duvall et al., in review).

The *rbc*L trees suggest that *Smilax glauca* is allied with species of Liliales and isolated from the two other species of Dioscoreales (*Dioscorea polygonoides* and *Tacca integrifolia*). Among the three species a closer relationship has been proposed between the latter two, and *Smilax* has been placed outside of Dioscoreales because of a lack of similarities in secondary chemistry (Dahlgren et al., 1981). The more recent decision to place *Smilax* in Dioscoreales was based on "leaf morphology and floral appearance," although *Smilax* is hypothesized to form a "bridge" between Dioscoreales and species of Lilianae (Dahlgren et al., 1985).

Included in our analyses are 11 species of Poales and five species of the related Cyperales. Phylogenetic analyses of *rbc*L sequences of Poaceae

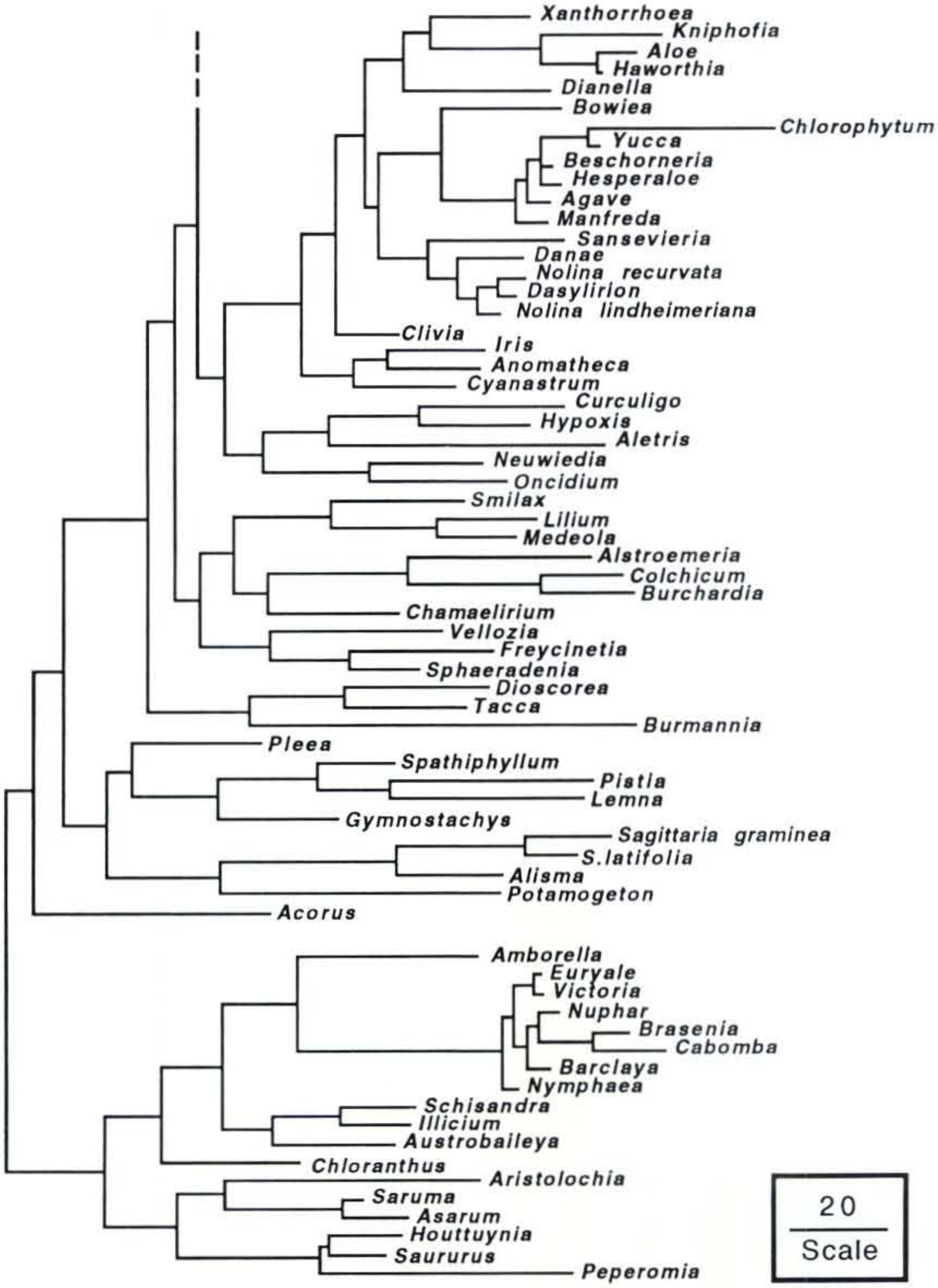


FIGURE 5. Remainder of the tree partially shown in Figure 4. This portion of the tree includes species of Aranae, Alismatanae, Lilianae, outgroup dicotyledons, and *Acorus calamus* (see Fig. 4 for the remainder of the tree). Branch lengths correspond to the number of substitutions optimized along the branches. The scale bar is proportional to a branch length of 20 steps.

have been thoroughly discussed (Doebley et al., 1990). We here note that pooid and panicoid grasses are segregated into sister clades in Figure 2, and that *Oryza sativa* occupies the most basal position of Poaceae. In the clade consisting of Poales (Fig. 3), a species of Restionaceae is most closely related to Poaceae, and Cyperales are found as a component of a sister clade which also contains Typhales contra Dahlgren et al. (1985). This topology is in agreement with the distribution of three inversions in the plastid genomes of these taxa (Doyle et al., 1992) that further predicts that Joinvilleaceae are immediately basal to Poaceae. This prediction has been confirmed by analysis of *rbc*L data as well (Duvall et al., in review).

In our analyses Commelinaceae (three species) cluster with three species of Bromelianae: *Pontederia sagittaria*, *Philydrum lanuginosum*, and *Anigozanthus flavidus* in a clade that is sister to Zingiberanae and separated from 18 other species

of Commelinanae (Fig. 4). A parsimony tree over the 79 species subset constrained to include 13 spp. of Commelinanae as monophyletic (not shown) is 15 steps longer than the tree without constraints. Parsimony analysis of *rbc*L thus does not support the recognition of a monophyletic Commelinanae.

Pistia has been tentatively aligned with *Lemna* based on similarities of seedling structure (Grayum, 1991). Analysis of *rbc*L sequences supports this alignment at a bootstrap value of 69% and a decay index value of at least two.

Three species of Melanthiaceae are found in three different primary lineages. These species exhibit a great deal of variation and when combined into a single family are considered a "difficult" treatment (Dahlgren et al., 1985). Our results support the suggestion (Dahlgren et al., 1985) that the Melanthiaceae should be divided into several families, and we further suggest that those families may not be closely related to each other.

The phylogenetic analysis presented here offers support for the recognition of seven primary lineages of monocotyledons that diverged over a relatively short period of geologic time. As noted, this result is in agreement with the taxonomic treatment of Dahlgren et al. (1985). However, a parsimony tree for the 79 species subset constrained to include nine of ten superorders sensu Dahlgren et al. (1985) as monophyletic groups (not shown) was considerably less parsimonious (52 steps longer). (Note that the tenth superorder, Triuridanae, is composed exclusively of achlorophyllous species unlikely to possess a phylogenetically meaningful copy of the *rbc*L sequence.) With regard to the paraphyletic arrangements of Bromelianae, Commelinanae, and Lilianae, the phylogenetic hypotheses presented here are at odds with those based on morphological, anatomical, chemical, and other characters. These discrepancies may reflect a new understanding of the affinities among these taxa. However, insufficient sampling of the *rbc*L data for these groups of species may also be a factor. For example, we have here included *rbc*L data from only half of the 52 families of Lilianae. Additional sampling of molecular characters may resolve these discrepancies or offer further insight into the phylogenetics of the monocotyledons.

The focus of this project was the phylogenetic framework supported by the *rbc*L data set for the monocotyledons. Another valuable feature of the data set, now generally available in GenBank, will be to further develop our understanding of the mechanisms and underlying probabilities of nucleotide substitution, particularly as influenced by structural and functional constraints of the molecules. These further studies will undoubtedly refine methods of the phylogenetic analysis of molecular data.

Literature Cited

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E. Conti, J. D. Palmer, J. R. Manhart, K. J. Sytsma, H. J. Michaels, W. J. Kress, K. G. Karol, W. D. Clark, M, Hedren, B. S. Gaut, R. K. Jansen, K.-J. Kim, C. F. Wimpee, J. F. Smith, G. R. Furnier, S. H. Strauss, Q.-Y. Xiang, G. M. Plunkett, P. S. Soltis, S. Swensen, S. E. Williams, P. A. Gadek, C. J. Quinn, L. E. Eguiarte, E. Golenberg, G. H. Learn, Jr., S. W. Graham, S. C. H. Barrett, S. Dayanandan & V. A. Albert. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Missouri Bot. Gard. 80: 528–580.

Clegg, M., K. Ritland & G. Zurawski. 1986. Processes of chloroplast DNA evolution. Pp. 275–294 *in* S. Karlin & E. Nevo (editors), Evolutionary Processes and Theory. Academic Press, New York.

Cronquist, A. 1981. An Integrated System of Classification of Flowering Plants. Columbia Univ. Press, New York.

Dahlgren, R., H. Clifford & P. Yeo. 1985. The Families of the Monocotyledons. Springer-Verlag, New York.

———, S. Rosendal-Jensen & B. Nielsen. 1981. A revised classification of the angiosperms with comments on correlations between chemical and other characters. Pp. 149–204 *in* D. Young & D. Seigler (editors), Phytochemistry and Angiosperm Phylogeny. Praeger Scientific, New York.

Doebley, J., M. Durbin, E. Golenberg, M. Clegg & D. Ma. 1990. Evolutionary analysis of the large subunit of carboxylase (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution 44: 1097–1108.

Doyle, J. A. 1973. The monocotyledons: Their evolution and comparative biology. Quart. Rev. Biol. 48: 399–413.

——— & L. Hickey. 1976. Pollen and leaves from the mid-Cretaceous Potomac Group and their bearing on early angiosperm evolution. Pp. 139–206 *in* C. Beck (editor), Origin and Early Evolution of Angiosperms. Columbia Univ. Press, New York.

Doyle, J. J. & J. L. Doyle. 1987. A rapid DNA isolation procedure from small quantities of fresh leaf tissue. Phytochem. Bull. 19: 11–15.

———, J. I. Davis, R. J. Soreng, D. Garvin & M. J. Anderson. 1992. Chloroplast DNA inversions and the origin of the grass family (Poaceae). Proc. Natl. Acad. Sci. U.S.A. 89: 7722–7726.

Duvall, M., M. Chase, D. Soltis & M. Clegg. A phylogeny of seed plants resulting from analysis of DNA sequence variation among the *rbc*L loci of 475 species with particular emphasis on alliances among monocotyledons. *In* P. Hoch (editor), Experimental and Molecular Approaches to Plant Biosystematics. Monogr. Syst. Bot. Missouri Bot. Gard. (in review).

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Molec. Evol. 17: 368–376.

———. 1991. PHYLIP (Phylogeny Inference Package) Version 3.4. University of Washington, Seattle.

GAUT, B., S. MUSE, W. D. CLARK & M. CLEGG. 1992. Relative rates of nucleotide substitution at the *rbc*L locus of monocotyledonous plants. J. Molec. Evol. 35: 292–303.

GOLDBERG, A. 1989. Classification, evolution, and phylogeny of the families of monocotyledons. Smithsonian Contr. Bot. 71. Smithsonian Institution Press, Washington, D.C.

GRAYUM, M. 1987. A summary of evidence and arguments supporting the removal of *Acorus* from the Araceae. Taxon 36: 723–729.

———. 1991. Systematic embryology of the Araceae. Bot. Rev. 57: 167–203.

JARRELL, D., M. ROOSE, S. TRAUGH & R. KUPPER. 1992. A genetic map of citrus based on segregation of isozymes and RFLPs in an intergeneric cross. Theor. Appl. Genet. 84: 49–56.

KISHINO, H. & M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Molec. Evol. 29: 170–179.

OLSEN, G., H. NATSUDA, R. HAGSTROM & R. OVERBEEK. 1992. FastDNAml Version 1.0.3. University of Illinois, Urbana and Argonne National Laboratory, Argonne, Illinois.

PALMER, J. 1986. Isolation and structural analysis of chloroplast DNA. Meth. Enzymology 118: 167–186.

RITLAND, K. & M. CLEGG. 1987. Evolutionary analysis of plant DNA sequences. Amer. Naturalist 130, supplement: S74-S100.

SMITH, J., W. J. KRESS & E. A. ZIMMER. 1993. Phylogenetic analysis of the Zingiberales based on *rbc*L sequences. Ann. Missouri Bot. Gard. 80: 620–630.

SWOFFORD, D. 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.

THORNE, R. 1983. Proposed new realignments in the angiosperms. Nordic J. Bot. 3: 85–117.

WILSON, M., B. GAUT & M. CLEGG. 1990. Chloroplast DNA evolves slowly in the palm family (Arecaceae). Molec. Biol. Evol. 7: 303–314.