ACADEMIC PRESS

Molecular Phylogenetics and Evolution 24 (2002) 35–57

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.academicpress.com

Relationships within Cornales and circumscription of Cornaceae—*matK* and *rbcL* sequence data and effects of outgroups and long branches

(Jenny) Qiu-Yun Xiang,[a,*] Michael L. Moody,[b] Douglas E. Soltis,[c] Chaun zhu Fan,[a] and Pamela S. Soltis[d]

[a] *Department of Botany, North Carolina State University, Raleigh, NC 27695-7612, USA*
[b] *Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-4236, USA*
[c] *Department of Botany and the Genetics Institute, University of Florida, Gainesville, FL 32611-5826, USA*
[d] *Florida Museum of Natural History and the Genetics Institute, University of Florida, Gainesville, FL 32611, USA*

Received 9 April 2001; received in revised form 1 March 2002

Abstract

Phylogenetic relationships in Cornales were assessed using sequences *rbcL* and *matK*. Various combinations of outgroups were assessed for their suitability and the effects of long branches and outgroups on tree topology were examined using RASA 2.4 prior to conducting phylogenetic analyses. RASA identified several potentially problematic taxa having long branches in individual data sets that may have obscured phylogenetic signal, but when data sets were combined RASA no longer detected long branch problems. t_{RASA} provides a more conservative measurement for phylogenetic signal than the PTP and skewness tests. The separate *matK* and *rbcL* sequence data sets were measured as the chloroplast DNA containing phylogenetic signal by RASA, but PTP and skewness tests suggested the reverse. Nonetheless, the *matK* and *rbcL* sequence data sets suggested relationships within Cornales largely congruent with those suggested by the combined *matK–rbcL* sequence data set that contains significant phylogenetic signal as measured by t_{RASA}, PTP, and skewness tests. Our analyses also showed that a taxon having a long branch on the tree may not be identified as a "long-branched" taxon by RASA. The long branches identified by RASA had little effect on the arrangement of other taxa in the tree, but the placements of the long-branched taxa themselves were often problematic. Removing the long-branched taxa from analyses generally increased bootstrap support, often substantially. Use of non-optimal outgroups (as identified by RASA) decreased phylogenetic resolution in parsimony analyses and suggested different relationships in maximum likelihood analyses, although usually weakly supported clades (less than 50% support) were impacted. Our results do not recommend using t_{RASA} as a sole criterion to discard data or taxa in phylogenetic analyses, but t_{RASA} and the taxon variance ratio obtained from RASA may be useful as a guide for improved phylogenetic analyses. Results of parsimony and ML analyses of the sequence data using optimal outgroups suggested by RASA revealed four major clades within Cornales: (1) *Curtisia–Grubbia*, (2) *Cornus–Alangium*, (3) *Nyssa–Camptotheca–Davidia–Mastixia–Diplopanax*, and (4) Hydrangeaceae–Loasaceae, with clades (2) and (3) forming a monophyletic group sister to clade (4) and clade (1) sister to the remainder of Cornales. However, there was not strong bootstrap support for relationships among the major clades. The placement of *Hydrostachys* could not be reliably determined, although most analyses place the genus within Hydrangeaceae; ML analyses, for example, placed the genus as the sister of *Hydrangeeae*. Our results supported a Cornales including the systematically problematic *Hydrostachys*, a Cornaceae consisting of *Cornus* and *Alangium*, a Nyssaceae consisting of *Nyssa* and *Camptotheca*, a monogeneric Davidiaceae, a Mastixiaceae consisting of *Mastixia* and *Diplopanax*, and an expanded Grubbiaceae consisting of *Grubbia* and *Curtisia*, and two larger families, Hydrangeaceae and Loasaceae.

Keywords: Cornales; Cornaceae; Hydrostachyaceae; Grubbiaceae; Molecular phylogeny; *rbcL*; *matK*; Long-branch attraction; RASA

* Corresponding author. Fax: +919-515-3436.
E-mail addresses: jenny_xiang@ncsu.edu (J.Q.-Y. Xiang), cfan3@unity.ncsu.edu (C.-Z. Fan), Michael.moody@uconn.edu (M.L. Moody), dsoltis@botany.ufl.edu (D.E. Soltis), psoltis@flmnh.ufl.edu (P.S. Soltis).

1055-7903/02/$ - see front matter
PII: S1055-7903(02)00267-1

1. Introduction

Cornales traditionally represented one of the systematically most problematic groups of flowering plants. The systematic debates largely centered on the circumscription and relationships of the core member, Cornaceae (for reviews see Eyde, 1988; Xiang and Soltis, 1998; Xiang et al., 1993). The family has been defined to include only the genus *Cornus* (e.g., Takhtajan, 1987) to as many as 14 or 15 diverse genera (e.g. Harms, 1898—containing *Garrya*, *Nyssa*, *Camptotheca*, *Davidia*, *Alangium*, *Mastixia*, *Curtisia*, *Aucuba*, *Cornus*, *Corokia*, *Griselinia*, *Helwingia*, *Kaliphora*, *Melanophylla*, and *Toricellia*; Cronquist, 1981—containing all of Harms' genera except *Alangium* and *Garrya*, with the addition of *Aralidium*). As reviewed in earlier papers (see Xiang and Soltis, 1998; Xiang et al., 1993), a total of 17 diverse genera have been considered members of Cornaceae; with 11 of these recognized as monogeneric families under various classifications. During the past 50 years, a diverse array of comparative experimental studies in phytochemistry, serology, wood anatomy, palynology, embryology, and cytology have been conducted to clarify relationships among the putative genera of Cornaceae (Breuer et al., 1987; Chopra and Kaur, 1965; Bate-Smith et al., 1975; Eyde, 1967; Fairbrothers and Johnson, 1964; Ferguson, 1977; Ferguson and Hideux, 1980; Goldblatt, 1978; Li and Chao, 1954; Noshiro and Baas, 1998). Despite this diversity of data, consensus had not been reached regarding the circumscription of the family and the closest relatives of *Cornus* (see Xiang et al., 1993).

Only recently have relationships among the putative cornaceous genera been more clearly understood. Eyde (1988) found a set of morphological, anatomical, and chemical characters that are shared by *Cornus*, *Mastixia*, *Nyssa*, *Camptotheca*, and *Davidia*. He defined Cornaceae to consist of only *Nyssa*, *Camptotheca*, *Davidia*, *Mastixia*, and *Cornus*. Two years later, Eyde discovered that the monotypic aralioid genus *Diplopanax* was a living fossil mastixioid most closely linked to the cornoid fossil *Mastixicarpum* (Eyde and Xiang, 1990). Thus, with the addition of *Diplopanax*, Cornaceae contained six living genera. This concept of Cornaceae was followed by some authors (e.g., Manchester, 1999; Thorne, 1992; Wen and Stuessy, 1993).

Recently, a molecular phylogenetic approach was employed to clarify relationships among the putative cornaceous genera (Xiang and Soltis, 1998; Xiang et al., 1993). Phylogenetic analyses of *rbcL* sequences revealed that nine genera (*Aralidium*, *Aucuba*, *Corokia*, *Garrya*, *Griselinia*, *Helwingia*, *Kaliphora*, *Melanophylla*, and *Toricellia*) previously placed in Cornaceae by different authors were allied with various asterids and distantly related to *Cornus*; eight genera (*Cornus*, *Alangium*, *Camptotheca*, *Curtisia*, *Davidia*, *Diplopanax*, *Mastixia*, and *Nyssa*) and the families, Hydrangeaceae and Loasaceae, formed a monophyletic Cornales clade that was different from all previously proposed concepts of Cornaceae or Cornales (Xiang and Soltis, 1998; Xiang et al., 1993). This circumscription of Cornales was also identified and strongly supported in broader phylogenetic analyses using molecular data (e.g., Chase et al., 1993; Olmstead et al., 1993, 2000; Savolainen et al., 2000a,b; Soltis et al., 2000), although not all the cornalean taxa were included in these broad analyses.

Within Cornales, four major lineages, (1) *Cornus–Alangium*, (2) *Nyssa–Davidia–Camptotheca–Mastixia–Diplopanax*, (3) *Curtisia*, and (4) Hydrangeaceae –Loasaceae, were identified in analyses of *rbcL* sequences (Xiang and Soltis, 1998; Xiang et al., 1993). However, the relationships among these four clades remained unresolved. To resolve relationships among these clades, Xiang et al. (1998a) further employed DNA sequences of the more rapidly evolving chloroplast gene *matK* (but with only a few representatives of Hydrangeaceae and Loasaceae). Parsimony analyses of the combined *rbcL–matK* sequences suggested that *Cornus*, *Alangium*, *Nyssa*, *Davidia*, *Camptotheca*, *Mastixia*, *Diplopanax*, and *Curtisia* formed a monophyletic group sister to Hydrangeaceae and Loasaceae. This clade closely corresponded to the Cornaceae of Eyde (1988) except that he did not include *Alangium* and *Curtisia* (see Xiang et al., 1998a). However, this cornoid group was weakly supported by bootstrap analyses and the monophyly of the group remained uncertain (see Xiang et al., 1998a).

Two phylogenetic analyses of *rbcL* sequences (Hempel et al., 1995—for Loasaceae and Morton et al., 1996—for Ebenales) suggested that two enigmatic African genera, *Grubbia* and *Hydrostachys*, representing two monogeneric families (Grubbiaceae and Hydrostachyaceae) may also belong to Cornales. Analysis of *rbcL* sequences of Xiang (1999) supported *Grubbia* and *Hydrostachys* as members of Cornales and further suggested that *Hydrostachys* may be part of Hydrangeaceae, closely allied with *Kirengeshoma* and *Deutzia*; while *Grubbia* may be either sister to *Curtisia* or the first branching lineage within Cornales. The results of Xiang (1999) also suggested the non-monophyly of *Cornus*, *Alangium*, *Nyssa*, *Davidia*, *Camptotheca*, *Mastixia*, *Diplopanax*, and *Curtisia*, a finding in conflict with earlier analyses (Xiang et al., 1998a). Xiang (1999) called for more data to test these new hypotheses because none of these relationships were supported by high bootstrap values.

The phylogenetic affinity of *Hydrostachys* and *Grubbia* to Cornales was also revealed in recent broad analyses [e.g., *rbcL* sequences of eudicots—Savolainen et al., 2000b; *ndhF* sequences of Asteridae s. l.—Olmstead et al., 2000 and combined sequences of 18S rDNA, *rbcL*, and *atpB* of angiosperms—Soltis et al., 2000 (*Grubbia*

was not included in these two studies); combined sequences of *ndhF*, *atpB*, and *rbcL* of Asteridae—Albach et al., 2001a,b; also see also AGP, 1998]. However, the exact placements of *Hydrostachys* and *Grubbia* within Cornales could not be determined in these analyses due to incomplete sampling of Cornales. Although most phylogenetic analyses involving *Hydrostachys* have suggested that *Hydrostachys* is a member of Hydrangeaceae (Albach et al., 2001a,b; Hempel et al., 1995; Soltis et al., 2000; Xiang, 1999), the placement of *Hydrostachys* within Hydrangeaceae varies among analyses, and in all analyses, *Hydrostachys* had a branch at least three times longer than any other branches of Cornales. Therefore, further phylogenetic analyses with complete sampling (all genera) of Hydrangeaceae, sufficient sampling of Loasaceae, and all other genera of Cornales are necessary to determine the relationships of *Hydrostachys* and *Grubbia* within Cornales.

In addition to *rbcL*, sequences of the chloroplast gene *matK* are now available for most Cornales. We constructed a data set of combined *rbcL–matK* sequences with a more extensive sampling of Cornales than any previous studies, including all genera of Hydrangeaceae, most genera of Loasaceae, and all of the other taxa of Cornales (*Cornus*, *Alangium*, *Nyssa*, *Davidia*, *Camptotheca*, *Mastixia*, *Diplopanax*, *Curtisia*, *Grubbia*, and *Hydrostachys*). We first explored the *matK* and *rbcL* sequence data for the presence of potentially long-branched taxa, assessed suitability of outgroups using a tree-independent method, relative apparent synapomorphy analysis (RASA 2.4; Lyons-Weiler, 2000), and then conducted detailed phylogenetic analyses of the sequence data. We experimented with RASA because of a concern regarding the topological impact of long-branched taxa (see Section 2.4). We sought to determine relationships within Cornales using the most extensive *rbcL* and *matK* data so far assembled for the clade while exploring possible effects of outgroups and long branches on phylogenetic analyses of the group.

2. Materials and methods

2.1. Sequence data

Although most DNA sequences used in the present analyses were generated in earlier studies (Hempel et al., 1995; Hufford et al., 2001; Moody et al., 2001; Morton et al., 1996; Xiang and Soltis, 1998; Xiang et al., 1993; Xiang et al., 1998a), they had not previously been assembled into a single large data set for analysis. The *matK* sequences of *Grubbia* and *Diplopanax* were generated for this study. The DNA of *Grubbia* was provided by M. Chase at Royal Botanical Garden, Kew (No. 495, Goldblatt 9591, MO) and the leaf material of *Diplopanax* was provided by J.-H. Li (Chen, 2001, Ruyuan, Guangdong, China). Amplification of *matK* via PCR for *Grubbia* and *Diplopanax* followed Xiang et al. (1998b). Double-stranded PCR products were cleaned using PEG and then used as the templates for cycle sequencing on a PTC-100 programmable thermal controller following the standard protocol for sequencing kit from Applied Biosystem. The cycle-sequencing products were cleaned by ethanol/sodium acetate precipitation and analyzed on an ABI-377 Automated Sequencer (Applied Biosystems, Foster City, CA 94404, USA).

2.2. Data matrices

Three data matrices were constructed for phylogenetic analyses: (1) *matK* sequences for 52 members of Cornales taxa each of 1353 base pairs (bp), with 18 short alignment gaps, starting from the first site of the 5′ end coding region. This matrix included *Cornus*, *Alangium*, *Nyssa*, *Davidia*, *Camptotheca*, *Diplopanax*, *Mastixia*, and *Curtisia*, all genera of Hydrangeaceae, 12 of the 15 genera of Loasaceae, *Hydrostachys*, and *Grubbia* (see Table 1). The 22 *matK* sequences generated by Xiang et al. (1998a) (GenBank Accession Nos. U96* in Table 1) were missing the last $\sim$ 190 bp of the aligned sequences and the *matK* sequence of *Hydrostachys* was missing the first 62 bp. (2) A narrowly combined *rbcL–matK* sequence data set containing 2857 characters (1504 bp from *rbcL* and 1353 bp from *matK*) of the 44 Cornales taxa each with both *rbcL* and *matK* sequences was available. This matrix included *Cornus*, *Alangium*, *Nyssa*, *Davidia*, *Camptotheca*, *Diplopanax*, *Mastixia*, and *Curtisia*, all genera of Hydrangeaceae, six genera of Loasaceae, *Hydrostachys*, and *Grubbia*. (3) A broadly combined *rbcL–matK* sequence data set containing 66 Cornales taxa, in which all of the taxa for which either *rbcL* or *matK* sequences were available was included. In this combined matrix, a broader sampling (14 of the 15 genera of Loasaceae) than the first combined data matrix was included; five of the taxa were missing the *matK* sequences and eight taxa were missing the *rbcL* sequences (see Table 1). In addition, three sequences were combined from two related species. The *rbcL* of *Cornus kousa* was combined with the *matK* of *Cornus capitata* to represent the Asian flowering dogwoods (Subgen. *Syncarpea*); the *rbcL* of *Alangium chinense* was combined with the *matK* of *Alangium platanifolium* to represent *Alangium*, and the *rbcL* of *Eucnide lobata* was combined with the *matK* of *Eucnide bartonioides* to represent *Eucnide* in Loasaceae. To compare results from analyses of these data and the *rbcL* sequences, the *rbcL* sequence data of Xiang (1999) were reanalyzed in the present study using the same outgroups as those used for *matK* and combined *rbcL–matK* sequence data.

Table 1
Source of sequences used in the study

Species	*rbcL*	*matK*
Cornales		
Cornus L.		
Subgen. *Yinquania* (Zhu) Murrell		
Cornus oblonga Wall.	L11218	U96899
Subgen. *Mesomora* Raf.		
Cornus alternifolia L.	L11212	U96889
Subgen. *Kraniopsis* Raf.		
Cornus obliqua Raf.	L11217	U96898
Subgen. *Cornus*		
Cornus chinensis Wangerin	L11214	U96892
Cornus mas L.	L11216	U96896
Subgen. *Arctocrania* Endl. ex Reichenbach		
Cornus canadensis L.	L01898	U96890
Subgen. *Cynoxylon* (Raf.) Raf.		
Cornus florida L.	L11215	U96894
Cornus kousa Hance	L14395	
Cornus capitata		U96891
Other core members of Cornales		
Alangium chinense (Lour.) Harms	L11209	–
Alangium platanifolium (Sieb. and Zucc.) Harms	–	U96880
Camptotheca acuminata Decne.	L11223	U96888
Curtisia dentata (Burn.) G.A. Sm.	L11222	U96901
Davidia involucrata Baill.	L11223	U96885
Diplopanax stachyanthus Hand.-Mazz.	L11224	AF468494
Mastixia caudatilimba C.Y. Wu ex Soong	L11227	U96887
Nyssa ogeche Marsh	L11228	U96886
Hydrangeaceae		
Broussaisia arguta Gaud.	AF323188	Hufford et al. (2001)
Carpenteria alternifolia Sieb. and Zucc.	AF323191	Hufford et al. (2001)
Carpenteria californica Torr.	L11177	Hufford et al. (2001)
Decumaria sp.	Morgan et al., 1993	Hufford et al. (2001)
Deutzia gracilis Sieb. and Zucc.	L11181	Hufford et al. (2001)
Deutzia rubens Rehder	AF323196	U96884
Dichroa febrifuga Lour.	AF323187	Hufford et al. (2001)
Deinanthe bifida Maxim.	AF323192	Hufford et al. (2001)
Fendlera rupicola Engelm. and Gray	AF323200	Hufford et al. (2001)
Fendlerella utahensis (Wats.) Heller	AF323198	Hufford et al. (2001)
Hydrangea arborescens L.	AF323486	Hufford et al. (2001)
Hydrangea macrophylla Ser.	L11187	Hufford et al. (2001)
Hydrangea anomala D. Don	AF323202	Hufford et al. (2001)
Hydrangea quercifolia Bartram	AF323203	U96882
Jamesia americana T. and G.	AF323201	Hufford et al. (2001)
Kirengeshoma palmata Yatabe	AF323197	Hufford et al. (2001)
Philadelphus caucausicus Koehne	AF323194	Hufford et al. (2001)
Philadelphus hirsutus Nutt.	AF323193	U96881
Pileostegia viburnoides	AF323185	Hufford et al. (2001)
Platycrater arguta Sieb. and Zucc.	AF323190	Hufford et al. (2001)
Schizophragma hydrangeoides Sieb. and Zucc.	AF323189	U96883
Whipplea modesta Torr.	AF323199	Hufford et al. (2001)
Loasaceae		
Blumenbachia latifolia Cambess		Moody et al. (2001)
Cajophora lateritia Klotzsch	U17872	Moody et al. (2001)
Cevallia sinuata Lag.	U17873	–
Eucnide lobata (Hook) A. Gray	U17874	–
Eucnide bartonioides Zucc.	–	Moody et al. (2001)
Eucnide aurea (Gray) Thompson and Ernst	–	Moody et al. (2001)
Eucnide urens (A. Gray) Parry	–	U96902
Fuertesia domingensis Urb.	–	Moody et al. (2001)
Gronovia scandens L.	U17875	–
Kissenia capensis R. Br.	–	Moody et al. (2001)
Klaprothia fasciculate (K. Presl.) Poston	–	Moody et al. (2001)

Table 1 (*continued*)

Species	*rbcL*	*matK*
Loasa loxensis H.B.K.	U17876	Moody et al. (2001)
Mentzelia decapetala (Pursh) Urban and Gilg	U17877	Moody et al. (2001)
Petalonyx nitidus S. Wats.	AF299086	U96904
Petalonyx linearis Greene	–	Moody et al. (2001)
Plakothira parviflora J. Florence	–	Moody et al. (2001)
Schismocarpus matudai Steyerm	U17878	Moody et al. (2001)
Scyphanthus elegans Don	–	Moody et al. (2001)
Hydrostachyaceae		
Hydrostachys multifida A. Juss.	L17879	Moody et al. (2001)
Hydrostachys angustisecta Engl.	Y10708	
Grubbiaceae		
Grubbia tomentosa (Thunb.) Harms	Z83141	AF323184
Outgroups		
(Fouquieriaceae) *Fouquieria splendens* Engelm.	L11675	U96893
(Roridulaceae) *Roridula gorgonias* Planchon	L01590	–
Roridula sp.	–	U96905
(Sarraceniaceae) *Sarracenia flava* L.	L01952	–
Sarracenia purpurea L.	–	U96906
(Longaniaceae) *Plocosperma buxifolium* Benth.		Z70192
(Bignoniaceae) *Paulownia tomentosa* (Thunb.) Steudel	L36447.1	AF051997
(Bignoniaceae) *Kigelia africana* (Lam.) Benth.	AF102648	AF05198
(Araliaceae) *Panax quinquefolius* L.		U58615
(Araliaceae) *Pseudopanax arboreus* (Murr.) Phillipson		U58617
(Scrophulariaceae) *Antirrhinum majus* L.		AF051978
(Orobanchaceae) *Boschniakia hookeri* Walp.		AF051979
(Hamamelidaceae) *Maingaya malayana* Oliver, L.C.		AF 025393
(Polemoniaceae) *Phlox gracili* (Douglas ex Hooker) Greene		L34203
(Polemoniaceae) *Phlox longifolia* Nutt.	AF156732	
(Actinidiaceae) *Actinidia chinensis* Planch.	L01882.2	U61324.1
(Ericaceae) *Rhododendron tashiroi*		AB012749
(Ericaceae) *Rhododendron hippophaeoidea* Balf. F. and W.W. Sm	L01949.2	

2.3. Outgroups

Outgroups were chosen from other groups of the asterid clade (or Asteridae s. l.) suggested to be closely related to the Cornales clade by various broad phylogenetic analyses (Chase et al., 1993; Olmstead et al., 1993; Savolainen et al., 2000a,b; Soltis et al., 2000). *Sarracenia*, *Fouquieria*, and *Roridula* were typically used as outgroups in earlier phylogenetic analyses of *rbcL* and *matK* sequences for Cornales (Xiang et al., 1998a). These taxa were again employed as the outgroups in the present study. To examine if changing outgroups would affect the estimation of ingroup relationships, two other different combinations of taxa were also used as the outgroups to replace *Sarracenia*, *Fouquieria*, and *Roridula* (as SFR hereafter) in the analyses. One suite was *Plocosperma*, *Panax*, and *Paulownia* (PPP hereafter), chosen among the higher asterids. The other suite was *Phlox*, *Actinidia*, and *Rhododendron* (PAR hereafter), which, like SFR, were chosen from Ericales, a clade closely related to Cornales according to the three gene-based angiosperm phylogeny (Soltis et al., 2000). The suitability of these three suites of outgroups was also assessed using RASA.

2.4. Detecting long branches and assessing suitability of outgroups using RASA

Presence of long-branched taxa in a data matrix or use of inappropriate outgroups in an analysis may lead to erroneous phylogenetic estimation (Felsenstein, 1978; Huelsenbeck, 1997, 1998; Lyons-Weiler, 2000; Lyons-Weiler and Hoelzer, 1997). A potentially simple and useful method for assessing the suitability of outgroups and detecting long branches in a data set is the recently developed relative apparent synapomorphy analysis (RASA 2.4) (Lyons-Weiler, 2000; Lyons-Weiler and Hoelzer, 1997). Lyons-Weiler (2000) suggests that

RASA assesses the suitability of outgroups and detects potentially long-branched taxa by examining phylogenetic signal in a data set with discrete characters. In this method, a "cladistic" measure (RAS) is calculated for each pair of taxa in the matrix, which is the amount of "apparent synapomorphy" (character states shared to the exclusion of other taxa) summed for a particular taxon across comparisons with all other taxa in the matrix, for all characters (Lyons-Weiler and Hoelzer, 1997; Lyons-Weiler et al., 1996). This "cladistic" measure (RAS) is plotted against a "phenetic" measure of similarity (E; the overall similarity of a taxon pair; that is, the number of variable characters for which a pair of taxa share character states). According to the authors of the methods, the slope of the regression of RAS on E is usually positive as a function of the amount of character covariation in the data. The observed slope (or observed rate of increase in pairwise RAS per unit pairwise E) is compared to a null slope that represents the relationship between E and RAS for the matrix derived under the assumption that character states do not covary more than expected from a random distribution of phylogenetic signals across pairs of taxa. The homogeneity of the two slopes is assessed using Student's t test with a degree of freedom of $[(N \times (N-1)/2)] - N - 3$. The data exhibit a significant degree of hierarchy if the observed slope is significantly greater than the null (indicated by a significant value of the test statistic, t_{RASA}) (Lyons-Weiler, 2000; Lyons-Weiler et al., 1996).

According to Lyons-Weiler (2000), if problematic taxon pairs (i.e., taxa likely to indicate spurious relationships in phylogeny estimation) are present in the data matrix, they will exhibit a greater amount of apparent cladistic support than that predicted by their "phenetic" similarity. Thus, in the RASA plot, points representing these putative "long-branched taxon" will appear as outlier points in the vertical direction (i.e., clustered near the origin and above the observed regression line). A problematic taxon also makes a much greater contribution to the variance of the cladistic measure (RAS) than to the phenetic variance (Lyons-Weiler, 2000). Therefore, the long-branch taxa can also be identified by examining the variance ratio output on RASA and are indicated by having greater variance ratio values than other taxa in the matrix (Lyons-Weiler, 2000).

RASA evaluates alternative outgroups by examining t_{RASA} of the ingroup data rooted by different outgroup taxa. The combination of outgroup taxa that results in the greatest significant t_{RASA} score in the analysis is considered as the optimal outgroup. According to Lyons-Weiler (2000), the t_{RASA} from a rooted analysis gives an estimate of the plesiomorphy content of putative outgroup taxa. As the character states found within outgroup taxa may be plesiomorphic for the ingroup, but may also be convergent upon autapomorphic and synapomorphic states among the ingroup taxa, the combination of outgroup taxa having the highest plesiomorphy content is considered the best justified outgroup for that ingroup. A significant decrease in t_{RASA} from the rooted analysis suggests that the outgroups used may be inappropriate (Lyons-Weiler, 2000).

An advantage of RASA as stated by Lyons-Weiler and Hoelzer (1997) is that the method is a tree-independent approach of estimating phylogenetic signal for discrete characters (Lyons-Weiler, 2000). Most methods of measuring phylogenetic signal to date have relied on topologies previously selected by some criterion, which therefore are linked to the assumptions and errors of the tree-building algorithm. These measures either do not provide an actual measure of signal: noise per se with probabilistic support (e.g., the ratio measures of homoplasy) or are sensitive to taxon sampling (more taxa more homoplasy) or can be misled by alternative sources of "signal" including noise resulting in branch attraction (e.g., the skewness test) (see Lyons-Weiler and Hoelzer, 1997). Thus, these methods may not reliably distinguish between taxa exhibiting phylogenetic information and those that are likely to indicate spurious relationships in phylogeny estimation.

We experimented with RASA and compared the results with those from permutation tail probability (PTP) and skewness tests (Archie, 1989; Faith and Cranston, 1991; Hillis and Huelsenbeck, 1992) because PTP and skewness tests could be readily performed using PAUP 4.0b4a (Swofford, 2000). PTP test compares the length of the most parsimonious tree for the observed data set to that found for random data sets (Faith and Cranston, 1991). The proportion of times that a tree can be found as short or shorter in the random data sets than the original tree is the PTP value. If PTP is lower than 0.05, significant cladistic covariation of characters exists in the original data set (i.e., the data contain significant phylogenetic signal) (Archie, 1989; Faith and Cranston, 1991). Skewness test assesses phylogenetic signal in a data set by examining the distribution of tree lengths of a random sample of all tree topologies (Hillis and Huelsenbeck, 1992). Data matrices with phylogenetic signal produce tree-length distributions that are strongly skewed to the left, in contrast to the nearly symmetrical distribution produced by data sets composed of random noise (Hillis and Huelsenbeck, 1992).

For RASA, we performed the signal content (SC) analyses on RASA 2.4 for the *rbcL*, *matK*, and the narrowly combined *rbcL–matK* sequence data sets. We examined the values of t_{RASA} and variance ratios to detect potentially long-branched taxa present in these data sets. To evaluate the suitability of outgroup taxa used for phylogenetic analyses, we performed SC analyses of the ingroup alone and of the ingroup rooted by different suites of outgroup taxa. An increase of t_{RASA} from the rooted analysis was used as the indication of suitability of outgroups according to Lyons-Weiler

(2000). For PTP tests, we ran heuristic searches with random taxon addition of 10 replicates for 1000 random data sets. For skewness tests, we evaluated the lengths of 10,000 random trees excluding constant characters from each data set.

2.5. Searching for optimal phylogenetic trees

All phylogenetic analyses were conducted with PAUP 4.0 b4a. PPC (Swofford, 2000). For all data matrices, both parsimony and maximum likelihood (ML) analyses were performed. Parsimony analyses were conducted using the heuristic search option with random taxon addition of 100 replicates, MULPARS on, TBR branch swapping, characters equally weighted, and character states unordered. Analyses were conducted including the entire aligned sequences and also excluding the portion of *matK* sequences (the last ~ 190 bp) missing in some species. Gaps in *matK* sequences were treated as missing data or excluded from the aligned sequences, but coded as binary data in the analyses to see if different coding strategies for gaps affect the topology of the phylogenetic trees.

Maximum likelihood analyses were performed using heuristic searches and the HKY two-parameter model, assuming a proportion of sites to be invariable with gamma-distributed rates at variable sites ($I + \Gamma$). This $HKY + I + \Gamma$ model incorporates both unequal base frequencies and a transition bias and allows for among-site variation of substitution rates. The base frequencies, t_i/t_v ratio, proportion of invariable sites, and α were first estimated from the most parsimonious trees. These values were used for an initial ML analysis. The resulting ML tree was then used to estimate these parameters again and the estimated values were used for a second ML analysis. As the likelihood score of the tree resulting from the second analysis remained the same as that from the first ML analysis, we did not perform further analyses. The ML trees resulting from this second ML analysis were thus used as our final tree from ML analysis.

Bootstrap analyses (Felsenstein, 1985) of 1000 replicates using FAST heuristic searches were performed to estimate support for clades identified in the most parsimonious trees. For ML trees, bootstrap analysis using ML method was not feasible due to the enormous amount of time required for a PowerMac 8600. To obtain estimates of support for the branches on the ML trees, we conducted bootstrap analyses using neighbor-joining method under a maximum likelihood model with parameter settings used for searching for the best ML trees (see Figs. 2 and 5).

Phylogenetic analyses including all Cornales in the matrix or excluding potentially problematic taxa identified by RASA were performed to examine the effect of these taxa on tree topologies.

3. Results

3.1. Results of RASA

Long-branch detection. Results of SC analyses revealed non-significant phylogenetic signal for both the *rbcL* and *matK* sequence data ($t_{RASA} = -3.238643$, $df = 1221$ for *matK*, $t_{RASA} = -1.302802$, $df = 986$ for *rbcL* ; both non-significant), suggesting that there might be potentially problematic taxa (long branches) in the data that obscured the phylogenetic signal. Output of taxon variance ratio suggested that four genera, *Hydrostachys* and three genera of Loasaceae, *Gronovia*, *Eucnide*, and *Schismocarpus*, were potential long-branched taxa in the *rbcL* sequence data and two genera, *Hydrostachys* and *Schismocarpus*, were potential long branched taxa in the *matK* sequence data. Removing these taxa did significantly increase the phylogenetic signal of the data (t_{RASA} increased to significant). In both data sets, *Hydrostachys* had a taxon variance ratio much greater than any other taxa, suggesting that it had introduced more homoplasy than any other potentially long-branched taxa. SC analysis for the narrowly combined *rbcL–matK* sequences revealed significant phylogenetic signal in the data ($t_{RASA} = 10.32417$, $df = 857$, significant), although *Hydrostachys* still had a much greater taxon variance ratio than the remainder of the taxa in the matrix. If RASA is a reliable method, this suggested that the "long-branched" *Hydrostachys* did not disrupt the hierarchical character covariation (or "phylogenetic signal") in the combined data set, although it did in the separate data sets.

Outgroup evaluation. Results of RASA suggested that SFR was an appropriate suite of outgroup taxa for the *rbcL* sequence data (t_{RASA} rooted by SFR was significantly greater than t_{RASA} for the unrooted ingroup taxa; 46.76324 vs. −1.302802; $df = 986$), but may not be a good choice for the *matK* sequence data (t_{RASA} rooted by SFR was smaller than the t_{RASA} unrooted ingroup: −16.60314 vs. −3.238643 and $df = 1221$ including *Hydrostachys* in the matrix; −1.304452 vs. 8.766922 and $df = 1172$ excluding *Hydrostachys* from the matrix). RASA suggested that PPP and PAR were better suites of outgroups than SFR for the *matK* sequence data and PPP were the best among the three suites of outgroups tested (t_{RASA} rooted by $PPP = 26.36227$, t_{RASA} rooted by $PAR = 19.5244$, t_{RASA} unrooted ingroup $= -3.238643$). For the narrowly combined *rbcL–matK* sequence data, SFR was considered to be better outgroups than PPP and PAR (t_{RASA} rooted by $SFR = 11.39249$, t_{RASA} rooted by $PAR = 4.576742$, and t_{RASA} rooted by $PPP = -2.448116$; t_{RASA} unrooted ingroup $= 10.32417$, $df = 857$). When the combined data set was rooted by SFR in the SC analysis, the taxon variance ratios were nearly all equal among taxa (including *Hydrostachys*), but if the analysis was rooted by PAR or PPP, nearly 2/3

of the taxa showed a much greater variance ratio (>10 times greater) than the remainder.

3.2. Results of PTP and skewness tests

The P values from PTP tests for *matK*, *rbcL*, and narrowly combined *rbcL–matK* sequence data sets with SFR as the outgroup were all less than or equal to 0.01 ($P_{matK} = 0.001$, $P_{rbcL} = 0.01$, $P_{matK-rbcL} = 0.001$). PTP tests of 100 replicates were also performed for *matK* sequence data using the other two suites of outgroups (PAR and PPP) and the P values for these tests were both 0.01. Skewness tests indicated that the tree length distributions of all data sets are strongly skewed to the left with a g_1 more negative than the critical values (Table 2). These results from PTP and skewness tests suggested that all data sets contained significant phylogenetic signal (or were phylogenetically structured).

3.3. Results of phylogenetic analyses

RbcL. Parsimony and ML analyses of *rbcL* sequences using SFR as outgroups suggested relationships within Cornales similar to those found in Xiang (1999) (Fig. 1). The same major lineages, *Cornus–Alangium*, *Nyssa–Camptotheca–Davidia–Mastixia–Diplopanax*, Hydrangeaceae–Loasaceae, *Curtisia*, and *Grubbia*, were recognized, but relationships among them were not clearly resolved. For example, parsimony analyses suggested that the *Nyssa–Camptotheca–Davidia–Mastixia–Diplopanax* clade was the sister of the Hydrangeaceae–Loasaceae clade; whereas ML analyses suggested the monophyly of *Cornus–Alangium–Nyssa–Camptotheca–Davidia–Mastixia–Diplopanax–Curtisia–Grubbia*. *Hydrostachys* was placed as the sister of *Deutzia* and *Kirengeshoma* in Hydrangeaceae with low bootstrap support in both parsimony and ML analyses. Parsimony analyses excluding the long-branched taxa identified by RASA one at a time or in different combinations did not change the arrangements of other taxa on the tree.

Table 2
Results of PTP and skewness tests

Date set/Outgroup	No. taxa/No. characters	p_{PTP}	$g_{skewness}$
MatK-ingroup only	52/586	0.001*	−0.510134*
matK excl. HS	50/516	0.001*	−0.510734*
matK/SFR	55/636	0.001*	−0.496487*
matK/PPP	55/630	0.01*	−0.542971*
matK/PAR	55/635	0.01*	−0.545971*
rbcL-ingroup only	46/339	0.001*	−0.612239*
rbcL excl. HGES	42/327	0.01*	−0.623923*
rbcL/SFRD	50/374	0.01*	−0.615507*
matK–rbcL/SFR	47/976	0.001*	−0.495696*

P values for the PTP tests indicates the proportion of a tree can be found as short or shorter than the orginal tree. A value of 0.05 or smaller indicates that cladistic covariation of the observed data is significant. Critical values of g_1 for a data matrix with 250–500 characters and 25 taxa are $P(0.01) = -0.09$ (Hillis and Huelsenbeck, 1992). These values provide conservative estimates for data matrices with more than 25 taxa and more than 500 characters (variable characters, not the total length of sequences). Data sets produce g_1 more negative than the critical values are significantly more structured than are the random data. A "*" indicates that the test result is significant. Abbreviations for outgroups are referred to "outgroups" under Section 2. HS, *Hydrostachys* and *Schismocarpus*. HGES, *Hydrostachys*, *Gronovia*, *Eucnide*, and *Schismocarpus*.

MatK—Parsimony analyses. Analyses of *matK* sequences of the Cornales clade using SFR (non-suitable outgroups based on RASA) as the outgroups, with gaps (see Table 3) coded as missing characters found 45 shortest trees of 1431 steps each with a CI of 0.5189 (excluding uninformative characters) and an RI of 0.6153. The strict consensus of these trees showed a tetrachotomy among four major clades, *Curtisia–Grubbia* (78% bootstrap value), *Cornus–Alangium* (81%), (the nyssoids–mastixioids; 53%), and Hydrangeaceae–Loasaceae (56%), with *Hydrostachys* allied with *Schismocarpus* (23%) (Fig. 2; dashed lines indicate branches not present in the strict consensus tree). Within Hydrangeaceae, *Fendlera* and *Jamesia* formed a clade (96%) and was sister to the remainder of the family. The remainder of the family formed two subclades that corresponded to the *Hydrangeeae* (94%) and *Philadelpheae* (63%) of Hydrangeoideae of Hufford et al. (2001) (see Fig. 2). This branching pattern within Hydrangeaceae was congruent with the earlier parsimony analyses of *rbcL* sequence data and combined *rbcL–matK* sequence data of the family (Hufford et al., 2001; Soltis et al., 1995). Within Loasaceae, clades corresponding to two of the three traditional subfamilies, Gronovioideae and Loasoideae, were recovered, congruent with earlier analyses of *matK* sequence data for the family (Moody et al., 2001) (see Fig. 2).

These relationships within Cornales described above were not affected by different treatments of gaps (i.e., coded as missing or excluded from the aligned sequences and coded as binary data) and missing data (i.e. including or excluding the 3′ end of approximately 190 bp portion of the aligned sequences missing in 22 taxa), except the placement of the particularly long-branched taxon *Hydrostachys*. When the 3′ ends (190 bp) were excluded from parsimony analysis, *Hydrostachys* appeared as the sister to the rest of Cornales.

Parsimony analyses of *matK* sequences using PPP as the outgroup (the best among the three suites according to RASA) found fewer shortest trees (18; CI: 0.5166 excluding uninformative characters; RI: 0.6642) compared to SFR as the outgroup (45 shortest trees; CI: 0.5189; RI: 0.6753). The strict consensus of the 18 trees showed complete resolution among the major clades of Cornales (see Fig. 3A). *Curtisia–Grubbia* was sister to the remainder of Cornales; *Cornus–Alangium* and nyssoids–mastixioids were sisters, forming a clade sister to the monophyletic Hydrangeaceae and Loasaceae. *Hydrostachys*

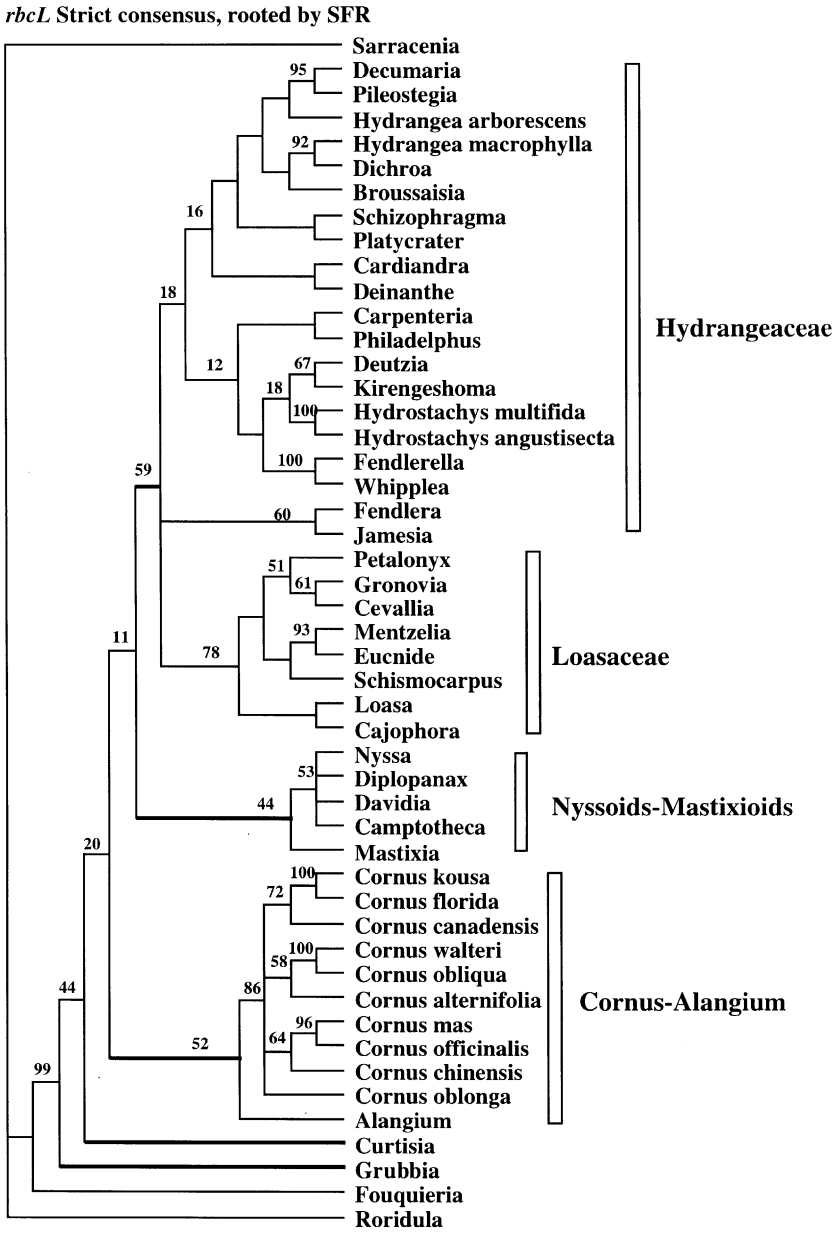

Fig. 1. Strict consensus tree of the 36 most parsimonious trees resulting from parsimony analysis of *rbcL* sequences rooted by SFR (*Sarracenia*, *Fouquieria*, and *Roridula*). Major lineages are indicated by thick branches and numbers are bootstrap values for lower nodes and some terminal nodes with bootstrap values >50%.

was again sister to *Schismocarpus* in Loasaceae, as suggested by parsimony analysis rooted by SFR.

Parsimony analyses using PAR as the outgroup (less optimal outgroup) (found 54 shortest trees of 1630 steps (CI = 0.5114 excluding uninformative characters; RI = 0.6573) showing relationships among major clades similar to those using PPP as the outgroup. The only differences were that the trees rooted by PAR placed *Hydrostachys* sister to the remainder of Cornales and *Curtisia–Grubbia* are placed with Hydrangeaceae–Loasaceae and nyssoids–mastixioids in a weakly supported monophyletic group (Fig. 3B).

Excluding the particularly long-branched taxon *Hydrostachys* from the parsimony analyses resulting in little or no changes in tree topologies, but CI and bootstrap support for the trees generally increased. For example, analyses without *Hydrostachys* rooted by SFR resulted in only one change in the tree topology (the sister relationship between *Schismocarpus* and *Eucnide* was no longer present). The CI, RI, and bootstrap support for the trees without *Hydrostachys* slightly increased compared to those with *Hydrostachys* included (CI: 0.5451 vs. 0.5181; RI: 0.7205 vs. 0.6753; bootstrap support for the monophyly of Hydrangeaceae: 58% vs. 29%; for

Table 3
Insertions and deletions (Indels A–R) in *matK* sequences of Cornales inferred using outgroups

Indels	Taxa	Reference no.	Sequence involved in the indel
A	*Plakothira*	110	–**G**GTTAAAT
	Klaprothia	110	–**G**GTTAAAT
B	*Grubbia*	115	+**T**AAATC
	Curtisia	115	+**T**AAAAA
	Hydrostachys	115	+**T**AAATAAAAAAAATC
	Panax	115	+**T**AAATA
C	*Schismocarpus*	132	–**A**TC
	Fuertesia	132	–**A**TCTAT
D	*Sarracenia*	162	+**T**ATGAC
E	*Sarracenia*	210	+**T**TTAATTACTCA
F	Loasa	266	+**T**TCTAA
G	*Deutzia*	276	+**C**AA
	Philadelphus	276	+**C**AA
	Carpenteria	276	+**C**AA
	Fendlerella	276	+**C**AA
	Whipplea	276	+**C**AA
	Kirengeshoma	276	+**C**AA
H	*Grubbia*	314	–**A**AA
I	*Hydrostachys*	376	+**T**CTTGTCTA
J	*Hydrostachys*	596	–**A**AT
K	*C. florida*	596	+**T**CGTAA
	C. obliqua	596	+**T**CGTAA
	C. canadensis	596	+**T**CATAA
	C. alternifolia	596	+**T**CATAA
	C. oblonga	596	+**T**CATAA
L	*Curtisia*	640	–**A**TCCAGTTC
M	*Mentzelia*	712	–**T**TCGTC
N	*Cardiandra*	723	–**C**GTAACCAATCTTCTC ATTTACGATCA
O	*Schismocarpus*	813	–**T**AGTGTAGA
P	*C. canadensis*	836	+**T**TA
Q	*Roridula*	842	–**A**TCTAT
R	*Broussaisia*	1011	+**A**AA

Reference number indicates the position of the first nucleotide of the indel sequence (in bold face) in the aligned matrix. A minus symbol before the indel sequence indicates that the sequence is absent in the listed taxa, but present in all other unlisted taxa in the matrix. A plus symbol indicates that the indel sequence is present in only the listed taxa, but absent in the unlisted taxa. Indels A, B, C, G, and K are potentially phylogenetically informative.

Loasaceae: 42% vs. 33%; for Hydrangeaceae–Loasaceae: 85% vs. 56%). When rooted by PPP, excluding *Hydrostachys* did not result in changes in tree topology, but bootstrap support increased for some major clades (e.g., the monophyly of Cornales, Hydrangeaceae–Loasaceae clade, Hydrangeaceae, and Loasaceae) and decreased for the monophyly of *Cornus–Alangium–Nyssa–Camptotheca–Davidia–Mastixia*–Loasaceae–Hydrangeaceae (see Fig. 3A, bootstrap values below branches).

MatK—ML analyses. ML analyses of *matK* sequence data using SFR as the outgroup resulted in a tree with the same four major clades within Cornales as those identified in the parsimony analyses rooted by SFR (Fig. 2), but showed complete resolution among them. In the ML tree, *Cornus*, *Alangium*, nyssoids, mastixioids, *Curtisia*, and *Grubbia* formed a monophyletic group sister to Hydrangeaceae–Loasaceae. *Curtisia–Grubbia* was the sister of the nyssoids–mastixioids clade (Fig. 4). However, these relationships were weakly supported with very low bootstrap percentages and short branch lengths (Fig. 4). In this ML tree, *Hydrostachys* appeared as the sister of the *Hydrangeeae* clade (see Fig. 4).

ML analysis using PPP as the outgroup resulted in a tree with a topology nearly identical to the ML tree rooted by SFR (compare Figs. 4 and 5), except that *Curtisia–Grubbia* no longer unites with the nyssoids–mastixioids, rather it appears as a lineage separate from Hydrangeaceae–Loasaceae, *Cornus–Alangium*–nyssoids–mastixioids (Fig. 5A). The ML analyses of *matK* sequences rooted by PAR found a tree with a topology identical to the ML tree rooted by SFR (Fig. 5B). ML analyses excluding *Hydrostachys* using different suites of outgroups did not result in changes in tree topology.

Combined rbcL–matK—Parsimony analyses. Parsimony analysis of the small combined *rbcL–matK* data set (Matrix 2) using SFR as outgroup (optimal outgroup) found 162 shortest trees. The strict consensus of these trees identified the same major clades within Cornales as those recognized by *matK* and *rbcL* sequences alone. Similar to results from *rbcL* and *matK* sequence analyses alone, relationships among *Curtisia–Grubbia*, *Cornus–Alangium*, nyssoids–mastixioids, and Loasaceae–Hydrangeaceae were not resolved in the strict consensus tree (Fig. 6). The monophyly of each of Loasaceae, Hydrangeaceae, and *Philadelpheae* was not supported. *Hydrostachys* appeared as a distinct lineage in a polychotomous clade containing *Hydrangeeae* and taxa of *Philadelpheae*.

Exclusion of *Hydrostachys* from this analysis increased the CI, RI, and bootstrap values within the Hydrangeaceae–Loasaceae clade (CI : 0.4934 vs. 0.4746; RI: 0.6659 vs. 0.6258; bootstrap values for Hydrangeaceae: 71% vs. 38%; Loasaceae: 93% vs. 74%; Hydrangeaceae–Loasaceae: 98% vs. 70%; see Fig. 6).

When using PAR or PPP as the outgroups (nonsuitable outgroups according to RASA), relationships among the four major clades were resolved with low support and resolution within the large Hydrangeaceae–Loasaceae significantly decreased. For example, the monophyly of each of Loasaceae, Hydrangeaceae, and *Philadelpheae* was not recognized) and *Hydrostachys* appeared in the Hydrangeaceae–Loasaceae clade with its sister relationship unresolved.

Combined rbcL–matK—ML analyses. ML analyses of the small combined *rbcL–matK* data set rooted by SFR resulted in a tree with a well-resolved topology similar to the *matK* ML trees (Figs. 4, 5, 7A); similar to *matK* parsimony trees rooted by PPP, this ML tree recognized *Curtisia Grubbia* as the sister of the rest of Cornales. Changing outgroup to PAR or PPP in the analyses for the small combined data set resulted in an ML tree with

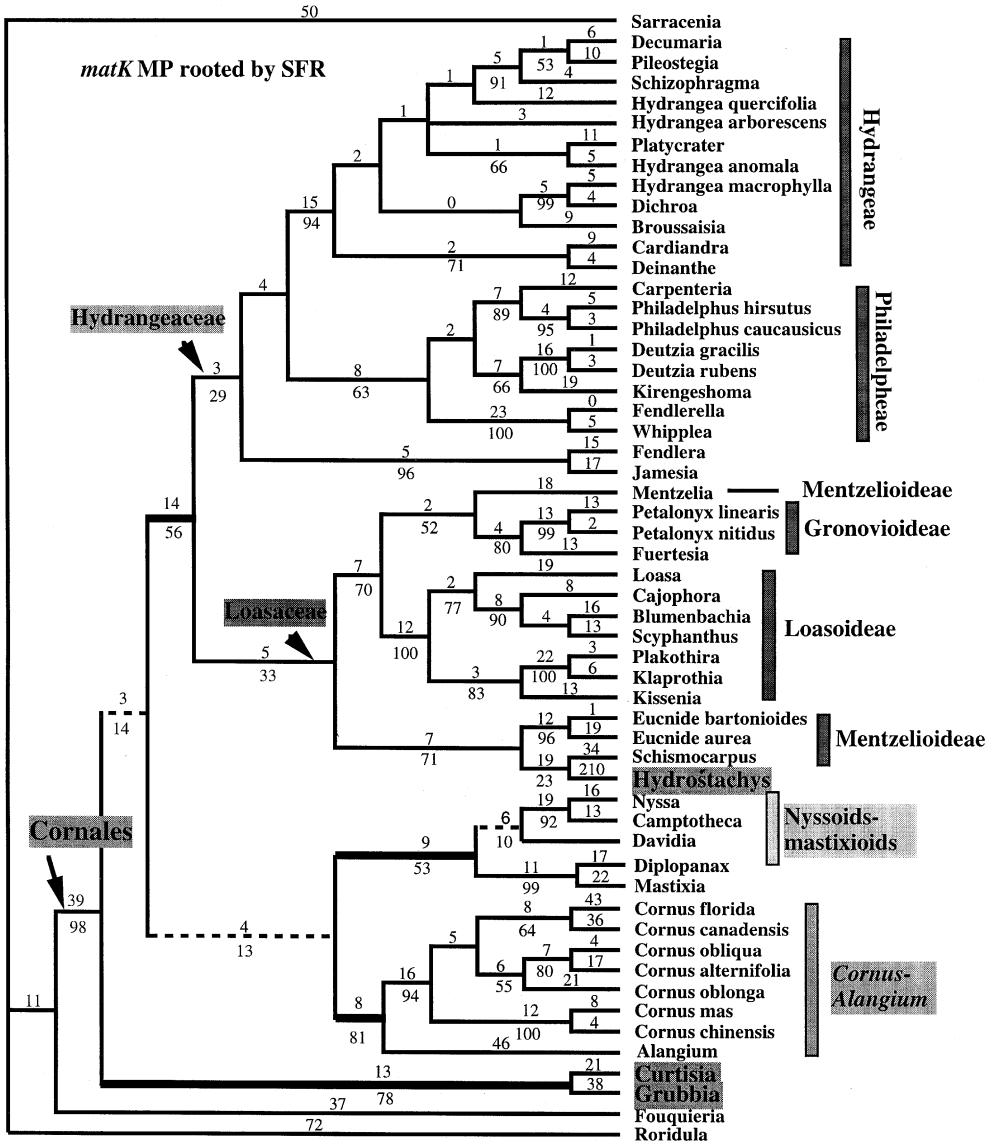

Fig. 2. One of the 45 shortest trees resulting from parsimony analysis of *matK* sequences rooted by *Sarracenia*, *Fouquieria*, and *Roridula*. Numbers above branches are nucleotide changes. Numbers below branches are bootstrap values. Dashed lines indicate the nodes that are not recognized in the strict consensus tree. Bolded lines indicate major lineages.

a topology identical to the *matK* ML trees rooted by PAR or PPP, respectively (Fig. 5, 7B).

Analysis of the large combined *rbcL–matK* sequence data (Matrix 3) required a greater amount of time to complete and found many more trees than did the small combined matrix (8640 vs. 162 in parsimony analysis). Concomitantly, the resolution of the strict consensus of all the shortest trees was reduced. However, the four major clades were still recognized in this tree (Fig. 8) and the reduced resolution in the strict consensus tree largely affected relationships within each major clade. In addition, this strict consensus tree placed *Hydrostachys* as the sister to the remainder of the Cornales. The ML analysis of this matrix was not performed to completion due to the considerable amount of time required as a result of a significant amount of missing data (13 taxa with ~50% of characters missing; see Section 2). The ML tree resulting from this incomplete analysis similarly shows *Hydrostachys* within the Hydrangeaceae and reduced resolution within each clade compared to the smaller matrix.

4. Discussion

4.1. Effects of long branches

The problems caused by long-branched taxa in phylogenetic analysis have been widely recognized. However, few studies have made efforts to identify

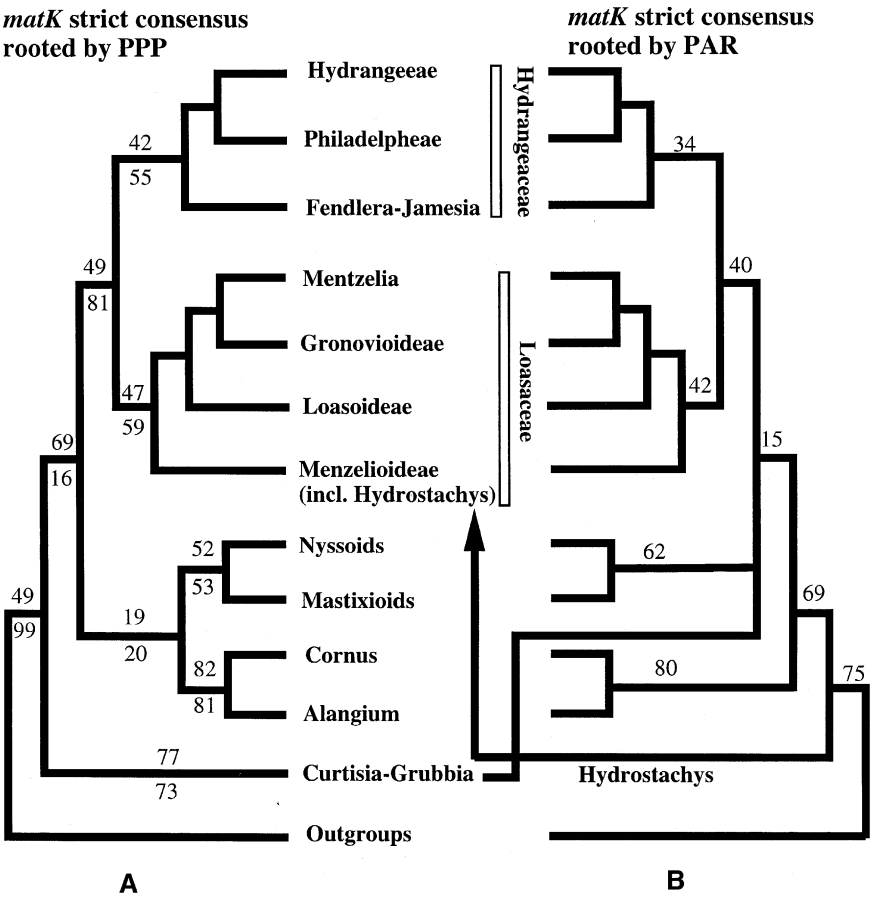

Fig. 3. Simplified strict consensus trees resulting from parsimony analyses of *matK* sequences rooted by PPP (*Plocosperma*, *Paulownia*, and *Panax*; (A) and by PAR (*Phlox*, *Actinidia*, and *Rhododendron*; (B), respectively. Numbers above branches are bootstrap values for major clades and numbers below branches are bootstrap values excluding *Hydrostachys* from the analysis. The trees show relationships of clades labeled in Fig. 2. Branch arrangements within each clade are identical to those shown in Fig. 2.

potentially long-branched taxa and to explore the potential effect of long branches on tree topology with empirical data. We experimented with RASA to identify potentially long-branched taxa in our *matK* and *rbcL* sequence data and examined the effect of these taxa on tree topology. Our results indicated that potential long-branched taxa (e.g., *Schismocarpus*, *Gronovia*, *Eucnide*, *Schismocarpus*, and *Hydrostachys* in the *rbcL* sequence data and *Hydrostachys* and*Schismocarpus* in the *matK* sequence data) significantly obscured phylogenetic signal (as measured by RASA) in both the *rbcL* and *matK* data sets. Removal of these taxa from the *rbcL* and *matK* data sets increased t_{RASA} from non-significant to significant; see Section 3). Interestingly, the effect of these taxa in obscuring phylogenetic signal in the *rbcL* and *matK* sequence data disappeared when the two data sets were combined (indicated by a significant t_{RASA} for the combined data, but non-significant t_{RASA} for the *rbcL* and *matK* data sets). This suggested that increasing the number of characters resulted in increased signal, probably due to greater numbers of informative and compatible characters in the larger data set (Lyons-Weiler et al., 1996). It also suggested that increasing number of characters may overcome some of the problems introduced by long-branched taxa in single gene data sets.

A noteworthy finding of our results from RASA was that data sets with or without phylogenetic signal as measured by t_{RASA} (i.e., the combined *matK–rbcL* data vs. separate *matK* and *rbcL* data sets) all provide similar useful information for grouping taxa within Cornales. Tree topologies derived from *matK*, *rbcL*, and combined *matK–rbcL* sequences were very similar and were different only at weakly supported nodes (comparing Figs. 1–8). PTP and skewness tests both indicated that all of these data sets contain significant phylogenetic signal, in contrast to t_{RASA} which suggested no phylogenetic signal in the *matK* and *rbcL* sequence data. This suggests that t_{RASA} can be misleading if used to detect phylogenetic signal (at least for data sets as small as our *rbcL* and *matK* sequence data—less than 2 kb). It may reject data sets actually containing useful phylogenetic information, such as our *rbcL* and *matK* sequence data sets. Thus, caution must be taken when using t_{RASA} as a measure of

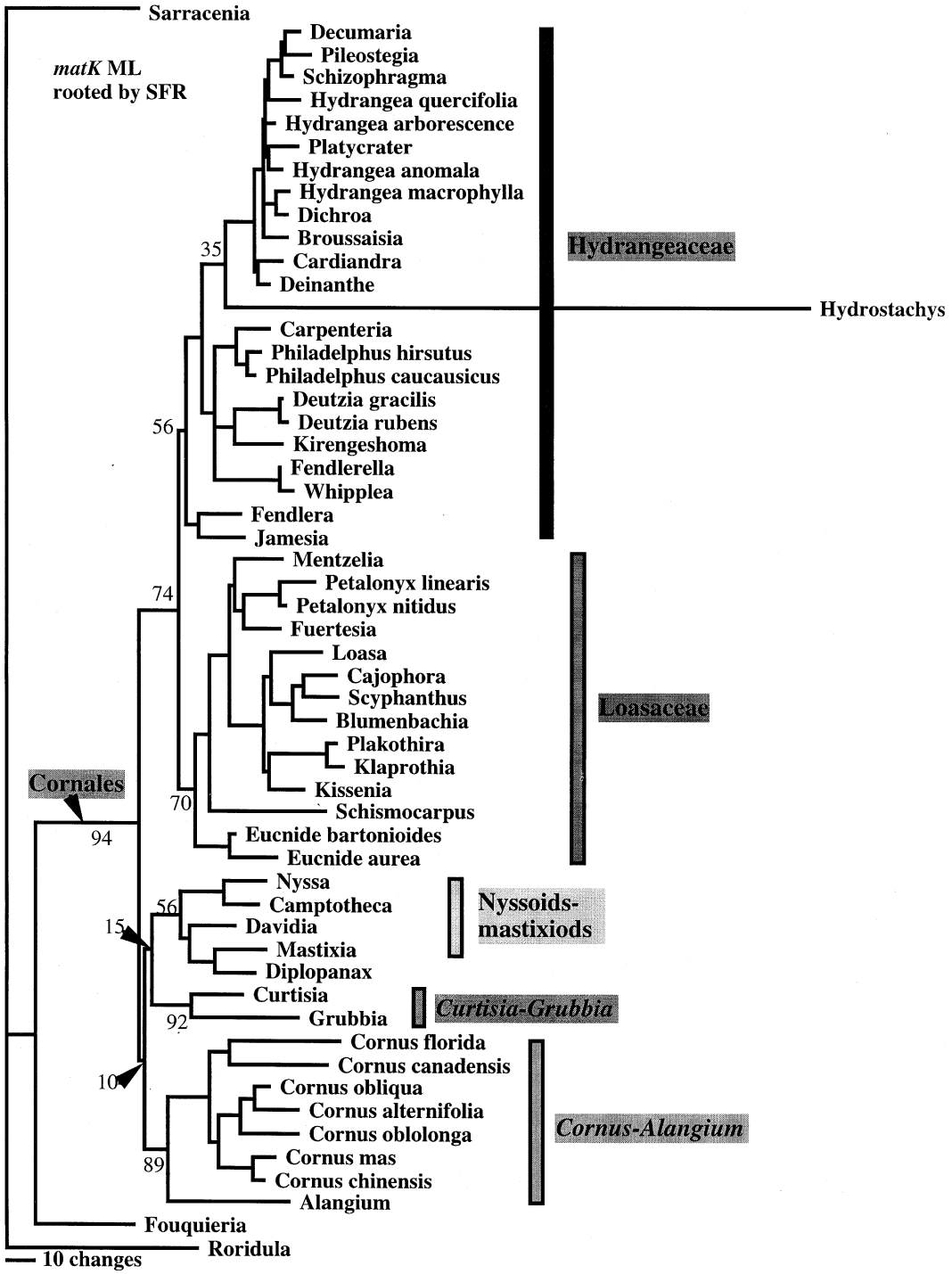

Fig. 4. Phylogram resulting from maximum likelihood analysis of *matK* sequence data rooted by *Sarracenia*, *Fouquieria*, and *Roridula* using the HKY + I + Γ with the following parameter settings: $A = 0.313693$, $C = 0.189293$, $G = 0.160299$, $T = 0.336715$, $t_i/t_v = 1.08055$, proportion of invariable characters = 0.046785, and $\alpha = 1.10947$. − ln = 10148.66.

phylogenetic signal. RASA should not be used alone as a criterion to discard entire sequences or taxa from phylogenetic analyses. Instead, a non-significant or negative t_{RASA} from RASA should be used as a guide to explore further the data for potential causes of "lack" of phylogenetic signal. Compared to PTP and skewness tests, RASA is less sensitive in detecting phylogenetic signal. If used with caution, this property can be advantageous. For example, a negative t_{RASA} will lead the researcher to more thoroughly explore the data (i.e., to identify problematic taxa with conflicting characters in the data).

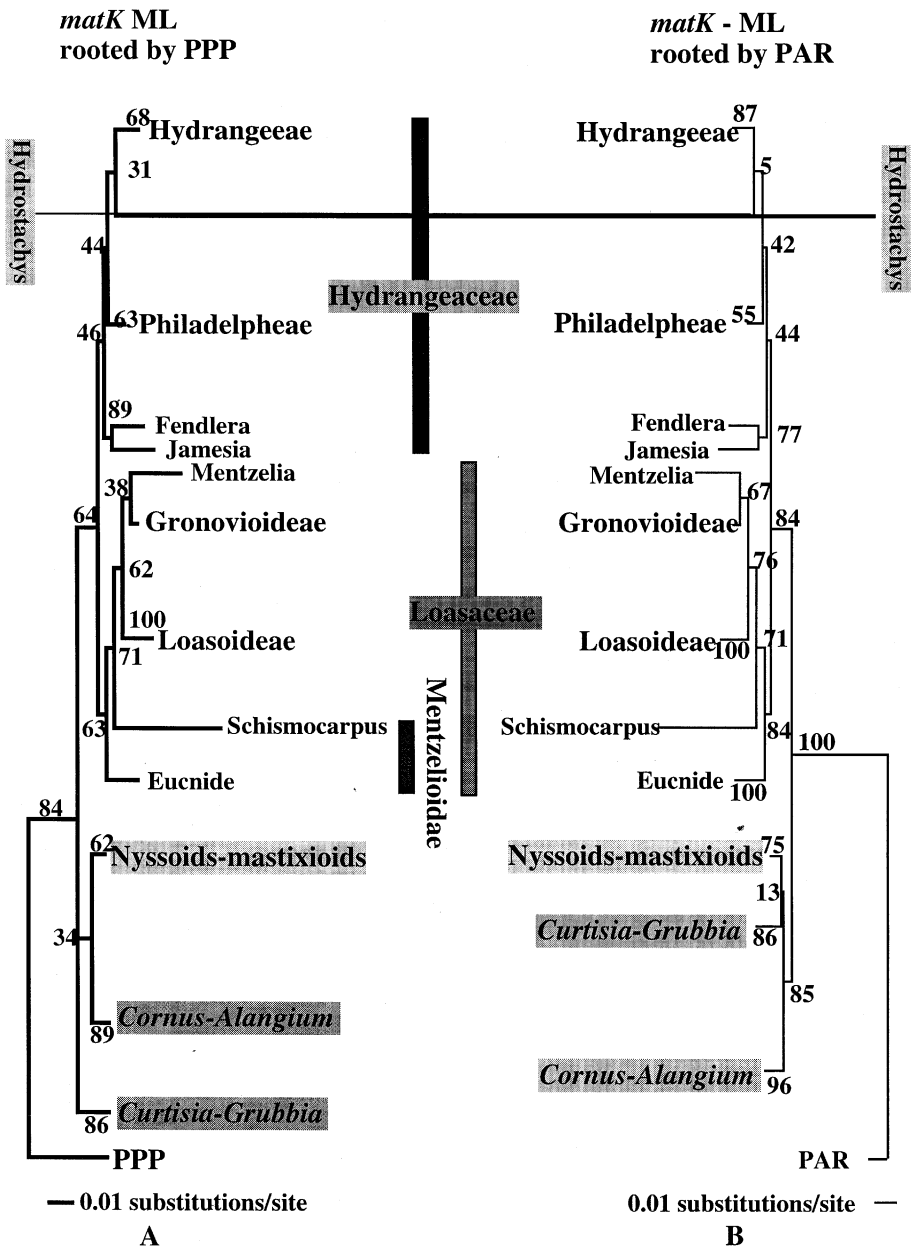

Fig. 5. (A) Simplified phylogram resulting from ML analysis of *matK* sequences rooted by PPP (*Plocosperma*, *Paulownia*, and *Panax*) using the HKY + I + Γ model with the following parameter settings: $A = 0.31786$, $C = 0.192774$, $G = 0.163907$, $T = 0.325459$, $t_i/t_v = 1.062951$, proportion of invariable characters = 0.09643, and $\alpha = 1.037437$. − ln = 10419.533. (B) Simplified phylogram resulting from ML analysis of *matK* sequences rooted by PAR (*Phlox*, *Actinidia*, and *Rhododendron*) using HKY + I + Γ model with the following parameter settings: $A = 0.317673$, $C = 0.189253$, $G = 0.162592$, $T = 0.330482$, $t_i/t_v = 1.101693$, proportion of invariable characters = 0.082110, and $\alpha = 1.048055$. − ln = 10477.20663. Numbers are bootstrap values. Component genera and species for *Hydrangeeae*, *Philadelpheae*, Gronovioideae, Losaoideae, Nyssoids, and Mastixioids are referred to Fig. 2 and the branch arrangements within each of these clades are identical to those shown in Fig. 4.

Another finding of our experiments with RASA for our *matK* and *rbcL* sequence data was that a taxon having a long terminal branch in a phylogenetic tree may not necessarily be identified by RASA as a potentially problematic long-branched taxon (i.e., this taxon may not have a significantly greater taxon variance ratio than other taxa). For example, in trees resulting from both parsimony and ML analyses of the combined *rbcL–matK* sequences rooted by the optimal outgroup SFR, *Hydrostachys* had long branches (e.g., 309 changes in the parsimony tree and ~5 times as long as the next longest branch in the ML tree), but results from SC analysis of the data rooted by the same outgroup did not show a significantly greater taxon variance for *Hydrostachys* compared to other taxa. In other words, it was not recognized as a long-branched taxon in the combined data set by RASA (see Section 3), suggesting that *Hydrostachys* was not a problematic taxon for the

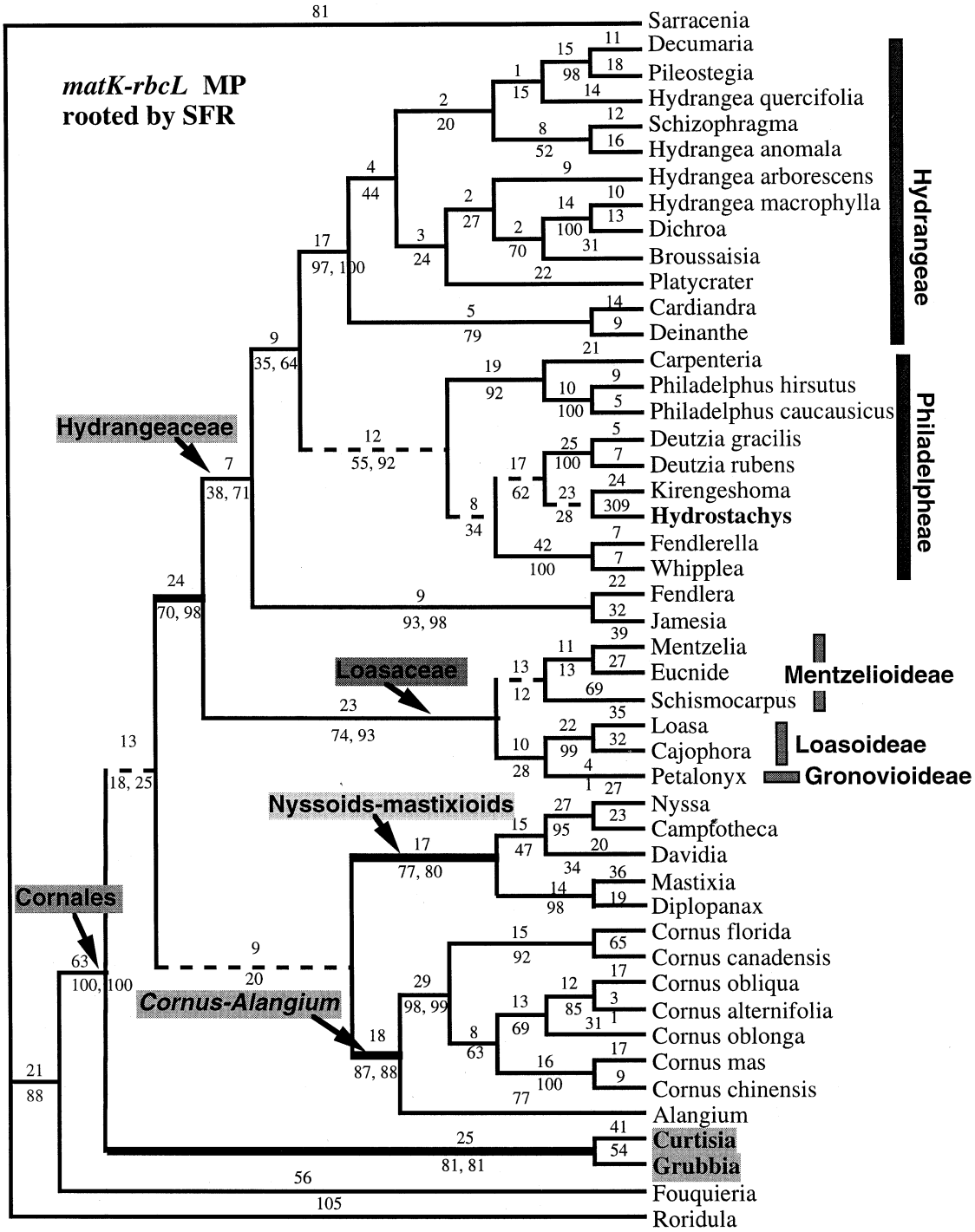

Fig. 6. One of the 162 shortest trees resulting from parsimony analysis of small combined *matK–rbcL* sequences (i.e., all taxa with both gene sequences available) rooted by *Sarracenia*, *Fouquieria*, and *Roridula*. Numbers above branches are nucleotide changes. First numbers below branches are bootstrap values. Second numbers below branches are bootstrap values excluding *Hydrostachys* from the analysis. Dashed lines indicate the nodes that are not recognized in the strict consensus tree. Major clades are indicated by thicker branches.

combined data set. This conclusion is further supported by a significant t_{RASA} for the combined *rbcL–matK* data set from SC analysis and reflects the positive effect of increased number of characters in a data set. These results demonstrated that long branches observed on a phylogenetic tree may not always disrupt the phylogenetic hierarchy of signal in a data set. It may depend on whether there are enough informative and compatible characters to buffer off the effect of the long-branched taxa in the data matrix. Alternatively with more characters false signal caused by convergent evolution may be construed by RASA as phylogenetically informative

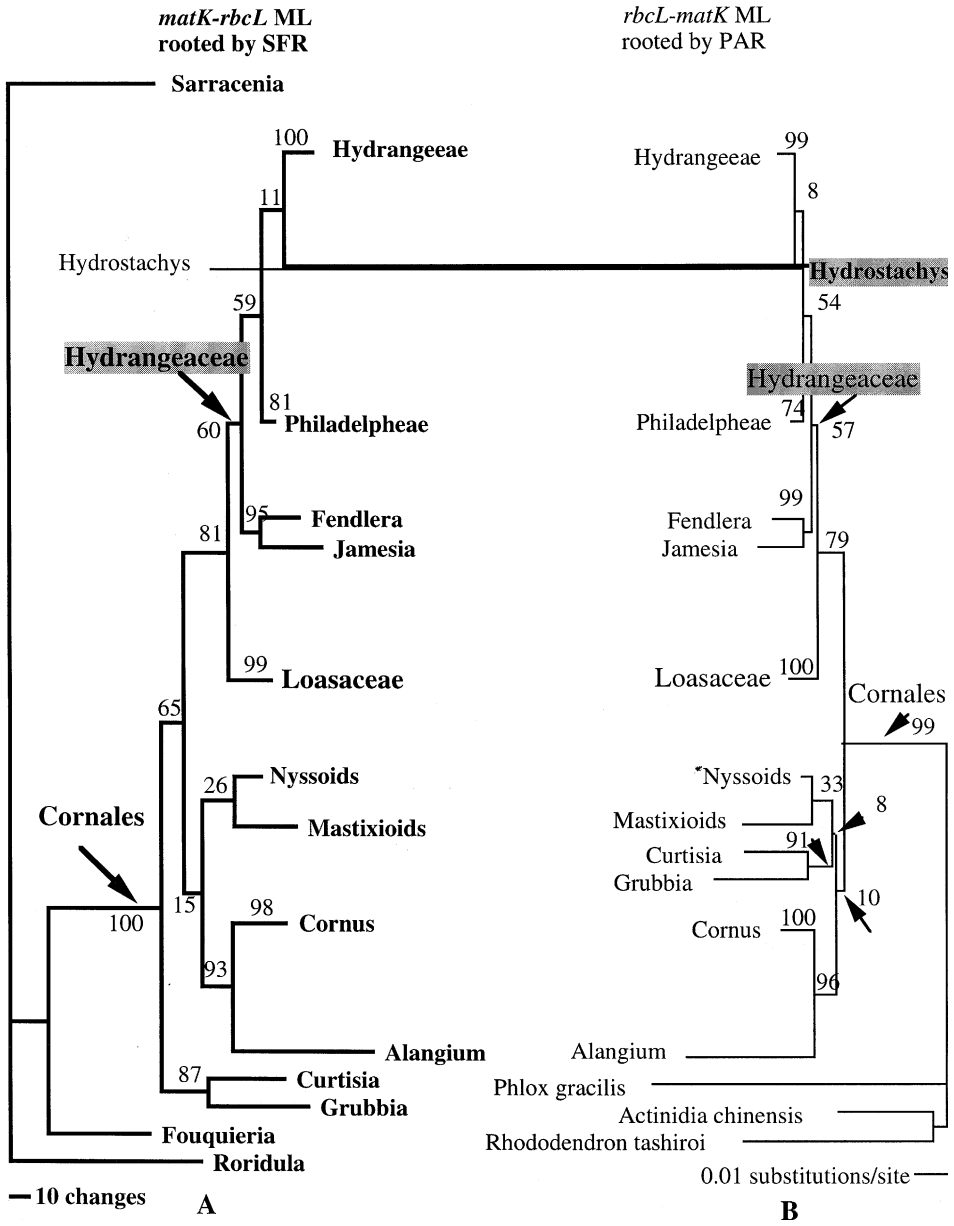

Fig. 7. (A) Simplified phylogram resulting from ML analysis of the small combined *rbcL–matK* sequences (i.e., all taxa with both gene sequences available) using SFR (*Sarracenia*, *Fouquieria*, and *Roridula*) as the outgroup and the HKY + I + Γ with the following parameters values: $A = 0.293202$, $C = 0.201891$, $G = 0.194256$, $T = 0.310651$, $t_i/t_v = 1.229177$, proportion of invariable characters = 0.339276, and $\alpha = 0.92572$. $-\ln = 16730.779$. (B) Simplified phylogram resulting from ML analysis of the small combined *rbcL–matK* sequences using PAR (*Phlox*, *Actinidia*, and *Rhododendron*) as the outgroup and the HKY + I + Γ with the following parameters values: $A = 0.296313$, $C = 0.201535$, $G = 0.192654$, $T = 0.309499$, $t_i/t_v = 1.190606$, proportion of invariable characters = 0.341426, and $\alpha = 0.902010$. $-\ln = 17180.566$. Numbers on both (A) and (B) indicate bootstrap values. Branch arrangements within Hydrangeaceae, Philadelpheae, Loasaceae, Nyssoids, and Mastixioids are referred to Fig. 6.

data, much in the way more characters in a data set often increase branch support for trees containing long branched taxa. In any case the examination of RASA's ability to detect long branches with more characters should be investigated further.

Our results further indicated that even if a taxon is identified as a long branch in the data set by RASA, it is not always problematic in phylogenetic reconstruction. The presence of such potentially long-branched taxa in our data sets identified by RASA seemed to have little effect on the tree topology regarding relationships among those taxa not having long branches. Phylogenetic analyses including and excluding these putatively problematic taxa resulted in trees with nearly identical topologies, demonstrating that these taxa have little impact on the placements of other taxa on the phylogenetic trees. According to our results, the major problem of the long-branched taxa in phylogenetic

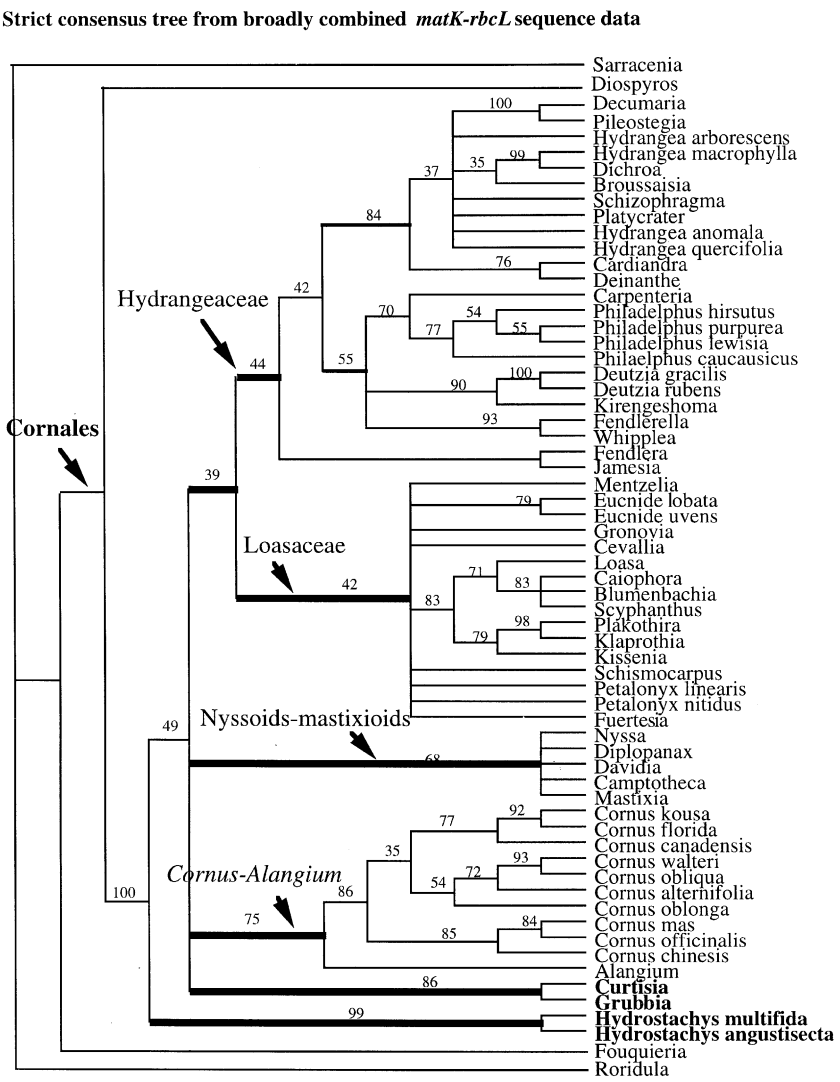

Fig. 8. Strict consensus tree resulting from parsimony analysis of the broadly combined *matK–rbcL* sequence data (i.e., all species with either *matK* or *rbcL* sequence available; see Section 2 for details). Major clades are labeled and indicated by thick branches. Values are bootstrap support.

analyses seemed to be the placement of these taxa themselves on the tree and overall branch support of the tree. Extremely long-branched taxa, such as *Hydrostachys*, moved around in trees resulting from different analyses. For example, ML analyses of the *matK* sequence data rooted by different outgroups placed the long-branched taxon *Hydrostachys* as the sister of *Hydrangeeae* in Hydrangeaceae, whereas parsimony analyses of the data placed the genus as the sister of *Schismocarpus* in Loasaceae or outside of Cornales depending on the outgroups used (Figs. 1–5; see Section 3). When *Schismocarpus* was removed from parsimony analyses, *Hydrostachys* appeared attracted to the outgroup and the other parts of the tree topology remained unchanged. However, the placement of *Schismocarpus*, another potential long-branch in both the *rbcL* and *matK* sequence data sets (i.e., with high taxon variance ratios and removing it from the data significantly increased t_{RASA}), varied little in phylogenetic analyses using different methods. Although *Schismocarpus* was suggested as a potential long branch, its branch as shown on the tree does not exceed the longest branch of the ingroup taxa (Fig. 2) and is still much shorter than that of *Hydrostachys*. All of these results reemphasize that caution must be taken when making decisions to exclude "problematic" taxa from the analyses using RASA as a guide and when interpreting the phylogenetic affinities of long-branched taxa. Multiple analyses should be performed and methods that incorporate unequal rates of evolution, such as maximum likelihood, should be used.

Although our results demonstrated that the long-branched taxa had little effect on placements of other taxa on the phylogenetic trees, removing Hydrostachys from the data matrix did increase bootstrap support for most clades, sometimes dramatically, as a result of the

increased phylogenetic signal in the data (see Figs. 3 and 6). Thus, applying RASA to detect potentially problematic taxa may help improve the results of phylogenetic analyses and perhaps pin-point specific taxa (long-branched taxa) for which caution in the interpretation of phylogenetic results may be needed.

4.2. Effects of outgroups in phylogenetic analyses

RASA suggested that *Sarracenia*, *Fouquieria*, and *Roridula* were the best among the three suites of outgroups for the *rbcL* and combined *matK–rbcL* sequence data sets and that *Plocosperma*, *Paulownia*, and *Panax* were the best outgroups for the *matK* sequence data. The use of different suites of outgroups substantially affects the results of phylogenetic analyses. In general, using less optimal or non-suitable outgroups (as measured by RASA) decreased resolution for relationships in parsimony analyses, but suggested different relationships in ML analyses. For example, parsimony analyses of *matK* sequences using SFR as the outgroup resulted in a tetrachotomy among the four major clades (Fig. 2). In contrast, the parsimony analyses rooted by outgroup PPP or PAR deemed more appropriate by RASA analysis resulted in complete resolution among these clades (Fig. 3). Parsimony analyses of the combined *rbcL–matK* sequences using the unfavored outgroups PPP resulted in reduced resolution within *Hydrangeeae*. ML analyses of *matK* sequences rooted by the unfavored SFR and PAR suggested monophyly of *Curtisia*, *Grubbia*, Nyssoids, and Mastixioids, whereas the ML analysis rooted by the favored PPP suggested the monophyly of *Cornus*, *Alangium*, Nyssoids, and Mastixioids (Figs. 4 and 5). The same results were found for ML analyses of the combined *matK–rbcL* sequences using the unfavored outgroup PAR and favored outgroup SFR (compare Figs. 7A and B). However, it is noteworthy that these effects of outgroups mainly involve weakly supported nodes of the trees.

It must be noted, however, that our results from RASA may have been influenced by some of the limitations of the method. According to Lyons-Weiler et al. (1996), major limitations of RASA in measuring phylogenetic signal include: (1) slightly susceptible to Type II error; (2) gives conservative estimate for data sets with small taxon sampling (i.e., data sets with few taxa appear to be devoid of signal); (3) fails to report multiple contrasting signals if such signals do exist in a data set; a lower amount of signal will be reported if internal heterogeneity exists in a given data set. Given that our data sets were fairly large (with 44–66 ingroup taxa), and the amount of signal estimated for the three data sets was not positively related to the number of taxa in the data sets, the second limitation may have little effect on our results. It is not possible to evaluate whether the third limitation has affected the results without in-depth character examination of the data; it is unknown if more than one set of covarying, but conflicting characters, exist in the *matK* and *rbcL* sequence data, although a lower amount of signal was reported for the two data sets by RASA (compared to PTP and skewness tests).

4.3. Phylogenetic relationships within the Cornales clade

Because employment of different outgroups substantially influenced the topology, although only the weakly supported branches were involved, our following discussion on relationships within Cornales is based on the results from analyses rooted by outgroups favored by RASA (i.e., Figs. 1, 3A, 5A, 6, 7A, 8).

Affinity of Hydrostachyaceae. Hydrostachyaceae contains only the genus *Hydrostachys* with 20–22 species restricted to Madagascar and tropical and southern Africa (Cronquist, 1981; Mabberley, 1997). The phylogenetic affinity of the family has long been a puzzle to plant systematists due to its unusual morphology associated with its aquatic habit (e.g., a tuberous-thickened stem, a basal hold-fast, fibrous roots, a cluster of basal, often pinnatifid or pinnate leaves, inaperturate pollen tetrads, and the lack of stomates, vessels, and many common secondary compounds; Cronquist, 1981; Scogin, 1992; Straka, 1988; Verdcourt, 1986; Watson and Dallwitz, 2000). The morphology, habit, and ecology of *Hydrostachys* suggest a close relationship of the genus to Podostemaceae. Evidence from embryology, palynology, and floral morphology of the *Hydrostachys* in contrast suggested a distant relationship of the genus to Podostemaceae, but a close relationship to Plantaginaceae or Solanaceae (Rauh and Jager-Zurn, 1967; Straka, 1988; also see Verdcourt, 1986). Chemical profiles of *Hydrostachys*, however, do not support either of these proposed relationships (Scogin, 1992).

Nonetheless, Hydrostachyaceae has been allied with Lamiales and Scrophulariales in Asteridae (Dahlgren, 1980; Takhtajan, 1987; Wagenitz, 1992) and placed in Bruniales of Rosanae (Thorne, 1992) and placed in Calitrichales of Asteridae (Cronquist, 1981). Phylogenetic analyses of *rbcL* sequences by Les et al. (1997) allied *Hydrostachys* with various taxa, including Podostemaceae, Cornales, Crassulaceae and Haloragaceae. However, these relationships were questionable given the selective sampling of angiosperms in the analysis. In contrast, several phylogenetic analyses of four different chloroplast gene sequences and 185 rDNA suggested or strongly supported that *Hydrostachys* was a member of Cornales allied within Hydrangeaceae (Albach et al., 2001a,b; Hempel et al., 1995; Olmstead et al., 2000; Soltis et al., 2000; Xiang, 1999; the present study) although many of these too had selective sampling (i.e, did not include Podostomaceae, included only selective groups of asterid taxa, etc.).

However, the exact affinity of *Hydrostachys* within Hydrangeaceae has differed among analyses. For example, analyses of *rbcL* sequences (Xiang, 1999) allied it with *Kirengeshoma* and *Deutzia* and analysis of *rbcL–atpB–ndhF* (Albach et al., 2001a,b) allied it with *Whipplea* and *Fendlerella* (both in the *Philadelpheae* clade). Our ML analyses also suggested that *Hydrostachys* was a member of Hydrangeaceae, but most closely allied with the *Hydrangeeae* clade (see Section 2.4; Figs. 2,4,5,7,8). This alignment was revealed in ML analyses of both the *matK* and combined *rbcL–matK* sequences. In contrast, our parsimony analyses suggested other placements of *Hydrostachys*. For example, analysis of *matK* sequences allied the genus with *Schismocarpus* in Loasaceae (Figs. 1,3) and analysis of the combined *rbcL–matK* sequences did not resolve relationships among *Hydrostachys*, *Hydrangeeae*, and several clades of *Philadelpheae* in Hydrangeaceae (Fig. 6). We believe that the placement of *Hydrostachys* in Loasaceae suggested by parsimony analyses of *matK* sequences was a result of long-branch attraction given that (1) both *Hydrostachys* and *Schismocarpus* had long branches in the parsimony trees, (2) both *Hydrostachys* and *Schismocarpus* were identified as potentially problematic taxa in the *matK* sequence data matrix by RASA, (3) analysis of maximum likelihood (a method known to be less sensitive to long-branch attraction) placed *Hydrostachys* and *Schismocarpus* in different clades, and (4) *Hydrostachys* was united with other taxa in parsimony analyses excluding *Schismocarpus* and/or *Kirengeshoma* from the *matK* and combined *rbcL–matK* data sets (e.g., it was united with the outgroups and sister to *Panax* in the analysis of *matK* sequences without *Schismocarpus*, but united with *Kirengeshoma* in the analysis of combined *rbcL–matK* sequences without *Schismocarpus*, and united with the outgroups again if *Kirengeshoma* was also removed from *rbcL–matK* data set). In contrast, ML analyses excluding *Schismocarpus* did not change the sister relationship between *Hydrostachys* and *Hydrangeeae*. These facts meet all of the following criteria proving long-branch attraction, except criterion (2) (see Siddall and Whiting, 1999): (1) the branches leading to the putatively attracted groups are very long, (2) the support for the attractors must be high, (3) the branches in question are sufficiently long enough to actually attract, (4) some method that is less sensitive to the long-branch attraction problem results in a tree with the branches separated, and (5) the absence of one of the branches should allow the remaining branch to place elsewhere in the pruned tree.

The alignment of *Hydrostachys* outside of Hydrangeaceae was also suggested by parsimony analyses of *matK* and combined *rbcL–matK* sequences for Hydrangeaceae (Hufford et al., 2001). To determine if ML analyses of these data for Hydrangeaceae also disagree with the parsimony analyses, we constructed an *matK* and a combined *rbcL–matK* sequence data sets containing the same taxa included in Hufford et al. (2001) (except that one of the outgroup taxa *Apium* in Hufford et al. was replaced by *Panax* in our data) and conducted analyses using both parsimony and ML (with both the $HKY + I + \Gamma$ model and $GTR + I + \Gamma$ model). The results from our ML analyses again placed *Hydrostachys* in Hydrangeaceae (allied with *Carpenteria* within *Philadelpheae* in *matK* sequence analyses and as the sister of *Philadelpheae* in the combined *rbcL–matK* sequence analyses), although our parsimony analyses showed the same results as those found in Hufford et al. (2001) (i.e., *Hydrostachys* was sister to *Plocosperma* outside of Cornales). This incongruence between ML and parsimony analyses of various data sets regarding the placement of *Hydrostachys* could be the result of differences of the two methods in handling long branches.

Thus, although there is molecular evidence supporting an affinity of *Hydrostachys* within the Cornales clade, the exact placement of *Hydrostachys* in Cornales remained unresolved. An alignment within Hydrangeaceae has been most frequently suggested by analyses of various chloroplast gene sequences with relative broad sampling of Cornales (e.g., Albach et al., 2001a,b; Hempel et al., 1995; Xiang, 1999; the present study), but this relationship was not strongly supported by bootstrap analyses. The difficulty in reliably placing *Hydrostachys* is probably largely due to its high level of molecular divergence from the remainder of the Cornales clade (shown as a very long branch in all the trees and the pairwise sequence divergence values between *Hydrostachys* and other taxa are significantly greater than all other pairwise comparisons of all other taxa in Cornales and outgroups). The molecular evolutionary pattern displayed in *Hydrostachys* may be dramatically different from the remainder of the Cornales clade. An ML model that could take into consideration a different model of evolution for the single branch of *Hydrostachys* would be necessary to potentially reliably solve the affinity of the genus. Data from nuclear DNA sequences may also be helpful to test the various hypotheses.

Affinity of Grubbiaceae. In all parsimony and ML analyses of *rbcL*, *matK*, and combined *rbcL–matK* sequences, *Grubbia* consistently appeared to be the sister of *Curtisia*. Although the two genera seem to have relatively long branches, raising the possibility of long branch attraction between them, the branches to these two genera were similar in length to some other taxa on the trees. Results of RASA did not show them as potentially long-branched taxa in any of the data sets. The close relationship between *Grubbia* and *Curtisia* suggested by the *rbcL* and *matK* sequence data has not been proposed by any previous workers. *Grubbia* has been traditionally placed in Ericales by Cronquist (1981) and Takhtajan (1987) and in Bruniales by Dahlgren (1980) and Thorne (1992), whereas *Curtisia* has been generally

allied with *Cornus* or Cornaceae. Both *Grubbia* and *Curtisia* have only a few species restricted to South Africa. The two are similar in many aspects of morphology and have a similar geographic distribution. These include a woody habit, leathery, simple, and oppositely arranged exstipulate leaves, scalariform vessel end-walls, hermaphrodite reproductive system, three aperturate and colporate pollen grains, a cymose terminal inflorescence unit, minute 4-merous flowers with an epigynous disk, inferior ovary, one pendulous, anatropous, unitegmic, and tenuinucellate ovule per locule, and seeds with oily and copious endosperm. However, all of these features are also found in one or more of the other members of the Cornales clade, particularly, in the cornoid genera. Thus, no apparent morphological synapomorphies can be found for the two genera at present. In the combined *rbcL–matK*-based parsimony tree (Fig. 6), *Curtisia* and *Grubbia* were united by 25 nucleotide changes (a number similar to that supports the Hdyrangeaceae–Loasaceae clade, but higher than those for nyssoids–mastixioids and *Cornus–Alangium* clades) with a bootstrap value of 81% for their sister relationship. A short insertion (Indel B in Table 3) in the *matK* sequences uniquely shared by *Curtisia* and *Grubbia* adds further support for a close relationship between these two genera.

The two genera also exhibit a relatively high level of molecular divergence and several morphological differences between them. For example, the genetic distance between the two sister genera is 0.03954 for the combined *rbcL* and *matK* sequences. This number is higher than those for any other sister genera on the tree except for *Cornus–Alangium*. In *Curtisia*, the leaves are coarsely dentate, the inflorescences are terminal, the androecium has four stamens in one whorl, anthers are not inverted, the gynoecium are 4-carpelled, and fruits are 4-seeded fleshy drupes. Whereas in *Grubbia*, the leaves are entire, the inflorescences are born axillarily, the androecium has eight stamens in two whorls, anthers are inverted, the gynoecium is 2-carpelled, and fruits are non-fleshy and multiple. This evidence of great molecular and morphological differences suggests an ancient divergence of the two genera.

Relationships among major clades and a phylogeny-based classification scheme. Our phylogenetic analyses of *matK* sequences using parsimony with PPP or SFT as out-groups suggested that *Curtisia–Grubbia* is the first branching lineage within Cornales; the *Cornus–Alangium* clade is sister to the *Nyssa–Camptotheca–Davidia–Mastixia*, which is, in turn, sister to the large Hydrangeaceae–Loasaceae clade (Fig. 3). These relationships were also recovered by both parsimony and ML analyses of the combined *rbcL–matK* sequences (Figs. 6,7) (although not in all the parsimony trees) and not in conflict with those suggested by ML analysis of *matK* sequences (Fig. 5A). However, these relationships were not very strongly supported by bootstrap analyses (Figs. 3,6,7). This lack of strong support for basal nodes of the phylogenetic trees (or relationships among major lineages) suggests either conflict in characters or insufficient information in the sequence data to solve deep relationships within the Cornales. Alternatively, the phenomenon may suggest a rapid early radiation of Cornales soon after its origin, a process that results in a polytomy near the root. In this case, increasing the number of characters for phylogenetic analyses will unlikely help resolve these relationships.

Although *Cornus*, *Nyssa*, *Camptotheca*, *Davidia*, *Diplopanax*, and *Mastixia* have been delimitated as Cornaceae by Eyde (1988) and Eyde and Xiang (1990), the monophyly of these genera (with the inclusion of *Alangium*) was supported with low bootstrap values (20% in Fig. 3, 34% in Fig. 5, 15% in Fig. 7) and remains equivocal. Thus this circumscription of Cornaceae is not strongly supported by the *matK* and *rbcL* sequence data but the putative ''synapomorphies'' used to unite these taxa as Cornaceae (e.g., lack of central bundles in their gynoecial vasculature, a germination valve in the fruit stone, H-shaped thinning of pollen aperture) (Eyde, 1988). Given the results of the present phylogenetic analyses, a circumscription of Cornaceae including only *Cornus* or *Cornus* and *Alangium* is strongly supported (see all figures). Further sampling of nuclear genes may add to our current interpretation of Cornaceae.

The monotypic *Diplopanax* had long been placed in Araliaceae before it was discovered to be a living fossil related to the extinct, *Mastixicarpum*, of Mastixiaceae (Eyde and Xiang, 1990). The new alignment of *Diplopanax* closest to *Mastixia* was for the first time clearly confirmed in the present study by the *matK* and *rbcL* sequence data. Although *Diplopanax* was included in previous analyses, only *rbcL* sequence data were available for the taxon. The genus was never united with *Mastixia* and its placement within the nyssoids-mastixioids clade was not strongly supported in previously analyses (e.g., Xiang et al., 1993, 1998a). With the *matK* sequence data available for *Diplopanax* in the present study, the genus is robustly placed as the sister of *Mastixia*, adding strong support to the finding of Eyde and Xiang (1990) for a close relationship between *Mastixia* and *Diplopanax*.

Nyssa, *Davidia*, and *Camptotheca* have been generally regarded as Nyssaceae although sometimes *Davidia* has been separated out as a monogeneric family Davidiaceae (see Xiang et al., 1993). The monophyly of these nyssoids is not recognized in all of the analyses of the present study and has never been robust. In some analyses, *Davidia* was united with *Mastixia–Diplopanax*, rather than with *Nyssa* and *Camptotheca.* However, *Nyssa* and *Camptotheca* have always been recognized as sisters and strongly supported by bootstrap analyses. These relationships among the nyssoids and mastixiods

support a Nyssaceae including *Nyssa* and *Camptotheca*, a monogeneric Davidiaceae, and a Mastixiaceae including *Mastixia* and *Diplopanax*.

Relationships revealed within the Hydrangeaceae in the present study are identical to those found in previous phylogenetic analyses focusing on only the family by Soltis et al. (1995) and Hufford et al. (2001). Two subfamilies, a Jamesioideae including *Jamesia* and *Fendlera* and a Hydrangeoideae consisting of two tribes, *Hydrangeeae* and *Philadelpheae* (Hufford et al., 2001), are supported (see all figures). The taxonomy of Loasaceae cannot be robustly evaluated with the present results because the sampling of Loasaceae was not complete in the analyses. Although the broadly combined *matK–rbcL* sequence data included 15 of the 17 genera of the family, relationships within the family in the analysis of the broadly combined data set were not well resolved as a result of large amount of missing data for some genera (e.g., missing *matK* or *rbcL* sequences) (Fig. 8).

Given the cpDNA data presented here, a circumscription of Cornales including Cornaceae (*Cornus*, *Alangium*), Nyssaceae (*Nyssa*, *Camptotheca*), Davidiaceae (*Davidia*), Mastixiaceae (*Mastixia*, *Diplopanax*), Grubbiaceae (*Grubbia*, *Curtisia*), Hydrostachyaceae (*Hydrostachys*), Loasaceae (15 genera), and Hydrangeaceae (17 genera) is supported.

5. Conclusion

Exploration of the *rbcL*, *matK*, and combined *rbcL–matK* sequence data using RASA suggested the presence of several potential problematic taxa that substantially obscured phylogenetic signal in the *rbcL* and *matK* data sets, but combining genes generally reduced the problems introduced by the long-branched taxa. These long-branched taxa identified by RASA did not have a major impact on phylogenetic relationships of other taxa in these data matrices, but may affect the placements of the long-branched taxa on the tree and the bootstrap support of the tree topology. Thus, it is necessary to detect potentially long-branched taxa in a data matrix prior to phylogenetic analysis for improved results and to place caution in the interpretation of phylogenetic affinities of the identified long-branched taxa.

Comparing to PTP and skewness tests, RASA is less sensitive in detecting phylogenetic signals and is sensitive to the number of characters in the matrix. A data set rejected by RASA may be indicated as containing significant phylogenetic signal in the PTP and skewness tests and still provide useful phylogenetic information. According to our results of RASA, one suite of outgroup taxa suitable for one data set may not necessarily be the best outgroup for other data sets (e.g., PPP was suggested to be best for *matK* sequence data, but not for *rbcL* and the combined *matK–rbcL* data). This would not be a surprise if heterogeneous rates of molecular evolution exist among taxa for the same molecule and among molecules within the same taxon. Use of different combinations of outgroups chosen logically (i.e., from the sister clade of the ingroup) based on previous broad based analyses (Soltis et al., 2000) had substantial impact on tree topologies in these analyses, but mainly only weakly supported branches were affected. We recommend performing exploration of data for phylogenetic signal, presence of potential long branches, and suitability of outgroups before conducting phylogenetic analyses. However, we do not recommend using t_{RASA} as a sole criterion to discard data or taxa from the analyses.

Our parsimony and ML analyses of *matK* and combined *rbcL–matK* sequences for Cornales suggested that the two woody enigmatic genera, *Curtisia* and *Grubbia* from South Africa, are sisters and form a clade probably sister to all other Cornales. Thus, we suggest grouping the two genera into a single family, Grubbiaceae. Other major lineages identified from the analyses include Cornaceae, Nyssaceae–Davidiaceae–Mastixiaceae, and Hydrangeaceae–Loasaceae. Relationships among the major clades remained unclear due to low bootstrap support for the basal nodes. The unusual aquatic genus *Hydrostachys* from South Africa appeared as the sister of the *Hydrangeeae* clade within Hydrangeaceae with low bootstrap support in the ML analyses, but parsimony analyses did not support this placement. The extremely long branch, decay of phylogenetic signal, and ragite placement of *Hydrostachys* in the cornates warrants caution in interpretation of these results considering this taxon. Better ML models capable of incorporating drastically different patterns of molecular evolution may be useful to reliably determine the sister taxa of *Hydrostachys*. Additional data from the nuclear genome may help to resolve relationships among the major clades and to understand better the diversification of this basal lineage of Asterids s. l.

Acknowledgments

We thank the Erin O'Leary-Jepsen and the Molecular Research Core Facility at Idaho State University, the DNA Sequencing and Mapping Center at North Carolina State University for running the sequencing gels and two anonymous reviewers for their thoughtful comments on the manuscripts. The project is partially supported by an NSF Grant DEB-0129069.

Notes. As the *matK* sequence for *Diplopanax* was not available until after the review of the manuscript. Thus, all tree values involving *matK* sequences reported in the paper were from analyses without *Diplopanax*. Addition of *matK* sequence of *Diplopanax* to the *matK* and

combined *matK–rbcL* data did not change tree topologies from all the analyses; the genus always appeared as the sister of *Mastixia* in the mastixioids–nyssoids clade with high bootstrap support. Thus, *Diplopanax* was subsequently added to the trees presented. Results reported for Long-branch Detection and Outgroup evaluation also did not include *Diplopanax* in the *matK* data set.

References

AGP, 1998. An ordinal classification for the families of flowering plants. Ann. Missouri Bot. Gard. 85, 531–553.

Albach, D.C., Soltis, D.E., Chase, M.W., Soltis, P.S., 2001a. Phylogenetic placement of the enigmatic angiosperm *Hydrostachys*. Taxon 50, 781–805.

Albach, D.C., Soltis, D.E., Olmstead, R.G., 2001b. Phylogenetic analysis of the asterids based on sequences of four sequences. Ann. Missouri Bot. Gard. 88, 163–210.

Archie, J.W., 1989. A randomization test for phylogenetic information in systematic data. Syst. Zool. 38, 219–252.

Breuer, B., Stuhlfauth, T., Fock, H., Huber, H., 1987. Fatty acids of some Cornaceae, Hydrangeaceae, Aquifoliaceae, Hamamelidaceae and Styracaceae. Phytochemistry 26, 1441–1445.

Chase, M.W. et al. (41 other authors), 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Missouri Bot. Gard. 80, 528–580.

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

Chopra, R.N., Kaur, H., 1965. Some aspects of the embryology of *Cornus*. Phytomorphology 15, 353–359.

Bate-Smith, E.C., Ferguson, I.K., Hutson, K., Jensen, S.R., Nielsen, B.J., Swain, T., 1975. Phytochemical interrelationships in the Cornaceae. Biochem. Syst. Ecol. 3, 79–89.

Dahlgren, R.M.T., 1980. A revised system of classification of the angiosperms. Bot. J. Linn. Soc. 80, 91–124.

Eyde, R.H., 1967. The peculiar gynoecial vasculature of Cornaceae and its systematic significance. Phytomorphology 17, 172–182.

Eyde, R.H., 1988. Comprehending *Cornus*; puzzles and progress in the systematics of the dogwoods. Bot. Rev. 54, 1–233.

Eyde, R.H., Xiang, Q.-Y., 1990. Fossil mastixioid (Cornaceae) alive in eastern Asia. Am. J. Bot. 77, 689–692.

Fairbrothers, D.E., Johnson, M.A., 1964. Comparative serological studies within the families Cornaceae (dogwood) and Nyssaceae (sur gum). In: Leone, C.A. (Ed.), Taxonomic Biochemistry and Serology. Ronald Press, New York, pp. 305–518.

Faith, D.P., Cranston, P.S., 1991. Could a cladogram this short have arisen by chance alone on permutation tests for cladistic structure. Cladistics 7, 1–28.

Felsenstein, J., 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Ferguson, I.K., 1977. World Pollen and Spore Flora 6: Cornaceae Dum. Almqvist and Wiksell, Stockholm.

Ferguson, I.K., Hideux, M.J., 1980. Some aspects of the pollen morphology and its taxonomic significance in Cornaceae *sens. Lat*. Proc. IV Int. Palynol. Conf., Lucknow 1, 240–249.

Goldblatt, P., 1978. A contribution to cytology in Cornales. Ann. Missouri Bot. Gard. 65, 650–655.

Harms, H., 1898. Cornaceae. In: Engler, A., Prantl, K. (Eds.), Die naturlichen Pflanzenfamilien, Teil III. Abteilung 8. W. Engelmann, Leipzig, pp. 250–270.

Hempel, A.L., Reeves, P.A., Olmstead, R.G., Jansen, R.K., 1995. Implications of *rbcL* sequence data for higher order relationships of the Loasaceae and the anomalous aquatic plant *Hydrostachys* (Hydrostachyaceae). Pl. Syst. Evol. 194, 25–37.

Hillis, D.M., Huelsenbeck, J.P., 1992. Signal, noise, and reliability in molecular phylogenetic analyses. J. Heredity 83, 189–195.

Huelsenbeck, J.P., 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46, 69–74.

Huelsenbeck, J.P., 1998. Systematic bias in phylogenetic analysis, is the *Strepsiptera* problem solved? Syst. Biol. 46, 519–537.

Hufford, L., Moody, M.L., Soltis, D.E., 2001. A phylogenetic analysis of Hydrangeaceae based on sequences of the plastid gene *matK* and their combination with *rbcL* and morphological data. Inter. J. Pl. Sci. 162, 835–846.

Les, D.H., Philbrick, C.T., Novelo, A.R., 1997. The phylogenetic position of river-weeds (Podostemaceae): insights from *rbcL* sequence data. Aquatic Bot. 57, 5–27.

Li, H.-L., Chao, C.-Y., 1954. Comparative anatomy of the woods of the Cornaceae and allies. Quart. J. Taiwan Mus. 7, 119–136.

Lyons-Weiler, J., 2000. RASA. 2.4 for the Mac. Available from: http://test1.bio.psu.edu/LW/list.htm.

Lyons-Weiler, J., Hoelzer, G.A., 1997. Escaping from the Felsenstein Zone by detecting long branches in phylogenetic data. Mol. Phyl. Evol. 8, 375–384.

Lyons-Weiler, J., Hoelzer, G.A., Tausch, R.J., 1996. Relative apparent synapomorphy analysis (RASA) I: the statistical measurement of phylogenetic signal. Mol. Biol. Evol. 13, 749–757.

Mabberley, D.J., 1997. The Plant Book—A Portable Dictionary of the Vascular Plants, second ed. Cambridge University Press, Cambridge.

Manchester, S.R., 1999. Biogeographical relationships of North American Tertiary floras. Ann. Missouri Bot. Gard. 86, 472–522.

Moody, M.L., Hufford, L., Soltis, D.E., Soltis, P.S., 2001. Phylogenetic relationships of Loasaceae subfamily Gronovioideae inferred from *matK* and ITS sequence data. Am. J. Bot. 88 (2).

Morton, C.M., Chase, M.W., Kron, K.A., Swensen, S.M., 1996. A molecular evaluation of the monophyly of the order Ebenales based upon *rbcL* sequence data. Syst. Bot. 21, 567–586.

Noshiro, S., Baas, P., 1998. Systematic wood anatomy of Cornaceae and allies. IAWA J. 19, 43–97.

Olmstead, R.G., Bremer, B., Scott, K.M., Palmer, J.D., 1993. A parsimony analysis of the Asteridae *sensu lato* based on *rbcL* sequences. Ann. Missouri Bot. Gard. 80, 700–722.

Olmstead, R.G., Kim, K-J., Jansen, R.K., Wagstaff, S.J., 2000. The phylogeny of the Asteridae sensu lato based on chloroplast *ndhF* gene sequences. Mol. Phyl. Evol. 16, 96–112.

Rauh, W., Jager-Zurn, I., 1967. Le probleme de la position systematique de Hydrostachyacees. Adansonia 6, 515–523.

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-G., Powell, M., Sheahan, M.C., Soltis, D.E., Soltis, P.S., Weston, P., Whitten, W.M., Wurdack, K.J., Chase, M.W., 2000a. Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcL* gene sequences. Kew Bull. 55, 257–309.

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., 2000b. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcL* gene sequences. Syst. Biol. 49, 306–362.

Scogin, R., 1992. Phylogenetic profile of *Hydrostachys insignis* (Hydrostachyaceae). Aliso 13, 471–474.

Siddall, M.E., Whiting, M.F., 1999. Long-branch abstrations. Cladistics 15, 9–24.

Soltis, D.E., Xiang, Q.-Y., Hufford, L., 1995. Relationships and evolution of Hydrangeaceae based on *rbcL* sequence data. Am. J. Bot. 82, 504–514.

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., Axtell, M., Swensen, S.M., Prince, L.M., Kress, W.J., Nixon, K.C., Farris, J.S., 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. Bot. J. Linn. Soc. 133, 381–461.

Straka, V.H., 1988. Die Pollen morphologie von *Hydrostachys*, Hydrostachyaceae, und ihre bedeutung fur die systematische einreihung der familie. Beitr. Biol. Pflanzen 63, 413–419.

Swofford, D.L., 2000. PAUP: phylogenetic analysis using parsimony, version 4.0b2. Computer Program distributed by Sinauer Assoc., Sunderland, MA, USA.

Takhtajan, A.L., 1987. Sistema Magnoliofitov. Izdatelístvo ìNauka, î Leningrad.

Thorne, R.F., 1992. An updated phylogenetic classification of the flowering plants. Aliso 13, 365–389.

Verdcourt, B., 1986. Hydrostachyaceae. In: Polhill, R.M. (Ed.), Flora of tropical East Africa. Royal Botanical Garden, Kew, Richmond, England, pp. 1–6.

Wagenitz, G., 1992. The Asteridae: evolution of a concept and its present status. Ann. Missouri Bot. Gard. 79, 209–217.

Watson, L., Dallwitz, M.J., 2000. The families of flowering plants: descriptions, illustrations, identification, and information retrieval. Version: 14th December 2000. Available from: http://biodiversity.uno.edu/delta/'.

Wen, J., Stuessy, T.F., 1993. Phylogeny and biogeography of Nyssa (Cornaceae). Syst. Bot. 18, 68–79.

Xiang, Q.-Y., 1999. Systematic affinities of Grubbiaceae and Hydrostachyaceae within Cornales—insights from *rbcL* sequences. Harv. Pap. Bot. 4, 527–542.

Xiang, Q.-Y., Soltis, D.E., 1998. *RbcL* sequence data define a cornaceous clade and clarify relationships of Cornaceae *sensu lato*. In: Boufford, D.E., Ohba, H. (Eds.), Sino-Japanese Flora—Its Characteristics and Diversification. University of Tokyo, Tokyo, pp. 123–137.

Xiang, Q.-Y., Soltis, D.E., Morgan, D.R., Soltis, P.S., 1993. Phylogenetic relationships of *Cornus* L. *sensu lato* and putative relatives inferred from *rbcL* sequence data. Ann. Missouri Bot. Gard. 80, 723–734.

Xiang, Q.-Y., Soltis, D.E., Soltis, P.S., 1998a. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. Am. J. Bot. 85, 285–297.

Xiang, Q.-Y., Crawford, D.J., Wolfe, A.D., Tang, Y.-C., 1998b. Origin and biogeography of *Aesculus* L. (Hippocastanaceae): a molecular phylogenetic perspective. Evolution 52, 988–997.