Syst. Biol. 48(3):413–435, 1999

More Characters or More Taxa for a Robust Phylogeny—Case Study from the Coffee Family (Rubiaceae)

BIRGITTA BREMER,[1,4] ROBERT K. JANSEN,[2] BENGT OXELMAN,[1] MARIA BACKLUND,[1] HENRIK LANTZ,[1] AND KI-JOONG KIM[2,3]

[1]*Department of Systematic Botany, Evoluntionary Biology Centre, Uppsala University, Norbyv, 18D, SE-75236 Uppsala, Sweden*
[2]*Section of Integrative Biology and Institute of Cellular and Molecular Biology, University of Texas, Austin Texas 78713, USA*
[3]*Department of Biology, Yeungnam University, Keongsan, Keonbuk, Korea 712-749*

Abstract.—Using different data sets mainly from the plant family Rubiaceae, but in parts also from the Apocynaceae, Asteraceae, Lardizabalaceae, Saxifragaceae, and Solanaceae, we have investigated the effect of number of characters, number of taxa, and kind of data on bootstrap values within phylogenetic trees. The percentage of supported nodes within a tree is positively correlated with the number of characters, and negatively correlated with the number of taxa. The morphological analyses are based on few characters and weakly supported trees are expected. The percentage of supported nodes is also dependent on the kind of data analyzed. In analyses of Rubiaceae based on the same number of characters, RFLP data give trees with higher percentage of supported nodes than *rbcL* and morphological data. We also discuss the support values for particular nodes at the familial and subfamilial levels. Two new data sets of *ndh*F and *rbc*L sequences of Rubiaceae are analyzed and together with earlier studies of the family we can conclude that the monophyly of the Rubiaceae is supported and within the family there are three well supported, but not easily characterized, large subfamilies, Rubioideae, Cinchonoideae s.s. and Ixoroideae s.l. There are also a few genera (*Luculia* and *Coptosapelta*) unclassified to subfamily. [Bootstrap; morphology; *ndh*F; phylogeny; *rbc*L; RFLP; Rubiaceae; subfamilies; support.]

The family is a central concept in flowering plant systematics and has been so for > 200 years, since the publication of Genera Plantarum by Jussieu (1789). Jussieu's families were circumscribed to include a manageable number of related genera (cf. review by Stevens, 1997), but they did not necessarily represent equivalent units, or sister groups, in a cladistic context. For most botanists today, a family represents a natural, supposedly monophyletic group of genera. A family should also be recognizable by several morphological characters. Some well-known families are the grasses, legumes, and carrots—which were recognized as groups long before the family concept was introduced. In larger families there has also been a tradition to divide these into subfamilies, tribes, or both. In the Rubiaceae, as many as eight different subfamilies and ~40 tribes have been recognized, each comprising more or less natural groups of genera.

Although families do not represent equivalent units evolutionarily in many cases, they are still the focus of classification and systematics. In phylogenetic analyses, especially those based on molecular data, it has become very popular to reconstruct family-level phylogenies. Several studies have shown or confirmed that some families are para- or polyphyletic: e.g., Saxifragaceae s.l. (Morgan and Soltis, 1993), Scrophulariaceae (Olmstead and Reeves, 1995), Amaranthaceae (Downie et al., 1997), Apocynaceae s.s. (Sennblad and Bremer, 1996), and Verbenaceae (Wagstaff and Olmstead, 1997); but several investigations have confirmed the monophyly of recognized families: e.g., many families in the *rbc*L analyses of Chase et al. (1993), Rubiaceae (Bremer and Struwe, 1992; Bremer et al., 1995), Asteraceae (Bremer et al., 1992; Kim and Jansen, 1995); Gesneriaceae (Smith et al., 1993; Smith, 1996), Hydrangeaceae (Soltis et al., 1995), Poaceae (Nadot et al., 1995), Polemoniaceae (Johnson et al., 1996), Iridaceae (Souza-Chies et al., 1997), and Leguminoseae (Doyle et al., 1997). In the very large analyses of 2,538 *rbc*L sequences, Källersjö et al. (1998) found

[4]Authors listed in order of contribution. Address correspondence to Birgitta Bremer. E-mail: birgitta.bremer@systbot.uu.se.

that 73% of the angiosperm families (217 families were represented by > 1 sequence) are monophyletic, the monophyly of 14% is uncertain, and only 13% are clearly para- or polyphyletic. Except for some genera that are misplaced in certain families, one can assume that most plant families are monophyletic as currently circumscribed.

The rooting and basal branching of a family phylogeny is essential for infrafamilial classification, but basal clades are not always well supported. Concern about lack of support for basal clades is only part of the more general concern about lack of support for nodes in phylogenetic trees, a subject that has been noticed and discussed by different authors. Several methods or approaches have been proposed to obtain better support, e.g., more characters or taxa, a different sampling strategy, more informative genes (cf. Missouri Botanical Garden 1995), or different kinds of weighting schemes, such as a priori weighting for different substitutions or codon positions (e.g., Albert et al., 1993; Allard and Carpenter, 1996) or homoplasy-based methods such as successive weighting (Farris, 1969).

With simulated data, when the "true phylogeny" is known, the effect on accuracy of adding more taxa or characters has been discussed and explored. Graybeal (1998) used relatively small taxon numbers (up to 30 taxa) but many characters (from 1,000 to 80,000) and started with a 4-taxon tree and added taxa with a sample strategy to break up long branches in the "Felsenstein zone." Phylogenetic accuracy was improved as the number of taxa increased but not when more characters were added. An exception was in the smallest data matrix (~ 8 taxa and 1,000 characters), where a decline in accuracy with increasing taxon numbers was observed. Kim (1998) found that adding taxa either increased or decreased the difference in parsimony in comparison with the model tree when taxa had been randomly added to the simulation. These partly contradictory results may simply illustrate the importance of choosing a clever taxonomic sampling strategy, as stressed by Hillis (1998). In a recent article about taxon sampling, character numbers, and long-branched trees, Poe and Swofford (1999) showed, e.g., "that adding characters can be the more favourable strategy, even for long-branched trees, and that adding slowly evolving taxa to subdivide long branches can reduce accuracy." Cummings et al. (1995) used whole-genome trees as true trees and analyzed data sets with different numbers of randomly selected nucleotide sites to see what effect the number of sites had. They found that the proportion of the resulting trees that were identical to the whole-genome trees distinctly increased with larger numbers of sites. However, with real data and without information from whole-genome trees, or when the true phylogeny is unknown, phylogenetic support is probably the best estimator of phylogenetic robustness. Furthermore, it has not been shown whether support and accuracy show the same connection to taxon and character sampling.

The Rubiaceae is one of the largest of the angiosperm families with ~ 10,000 species. The family is easily recognized, but it has a problematic and much-discussed infrafamilial phylogeny and classification (e.g., Verdcourt, 1958; Bremekamp, 1966; Robbrecht and Puff, 1986; Robbrecht, 1988, 1993; Bremer and Jansen, 1991; Andersson and Persson, 1991; Andersson, 1993; Andreasen and Bremer, 1996; Natali et al., 1996; Manen and Natali, 1996). Most botanists who encounter unidentified Rubiaceae specimens observe that they are difficult to identify or classify. The family is biologically diverse, and one problem with identification is that unique morphological characters are uncommon, whereas most features concerning life forms, flowers, and fruits are homoplastic. The early classification was simple but artificial with two recognized subfamilies, based on ovule number per carpel (e.g., in Schumann, 1891). Much progress and a deeper biological insight into the family were incorporated in the two classification systems from the middle of this century by Verdcourt (1958) and Bremekamp (1966). Verdcourt divided the Rubiaceae into three subfamilies: Rubioideae, Cinchonoideae, and Antirheoideae (as Guettardoideae). Bremekamp accepted the three subfamilies

(although with different circumscriptions) but also five others: Ixoroideae, Urophylloideae, Hillioideae, Gleasonioideae, and Pomazotoideae. The two classification schemes were partly similar but also very different, and no consensus was reached among botanists regarding which system should be used (e.g., the British botanists preferred Verdcourt's system, the Dutch preferred Bremekamp's, and the Americans preferred the old system of Schumann). Later Robbrecht (1988) wrote a very useful book about Rubiaceae, in which he presented another system, with four subfamilies: Rubioideae, Cinchonoideae, Ixoroideae, and Antirheoideae. The main difference between Robbrecht's classification and the earlier systems is that the Antirheoideae are much more widely circumscribed. None of these classifications were based on phylogenetic analyses so it was impossible to evaluate which system provided the most accurate phylogenetic framework for the Rubiaceae.

In recent years, phylogenetic analyses of representatives from all subfamilies and several outgroups based on different data sets (molecular and morphological) have been published (Bremer et al., 1995; Bremer, 1996a) and show that the family is monophyletic but that some of the deepest branches are unresolved or only weakly supported—a situation that makes subfamily classification uncertain and some taxa difficult to classify. With this empirical study we hope to provide some new insights into the following questions: (1) Is it better to use more characters or more taxa to get a more robust (supported) phylogeny (at the base of the family tree and for the rest of the tree)? (2) Are some kinds of data better than others? (3) Will an additional data set, from *ndh*F sequences, provide better resolution or support for the subfamilial classification of Rubiaceae?

Material and Methods

Two new sets of taxa from the Rubiaceae were compiled and analyzed for this study. The first included 43 species (42 genera; Table 1) representing all major lineages of the family; the three subfamilies Cinchonoideae, Ixoroideae, and Rubioideae (Bremer et al., 1995); and several taxa with uncertain taxonomic positions but expected to be basal within the phylogenetic tree. Three different subsets of characters were produced for these 43 taxa: one from *rbc*L sequences (Rubiaceae_43_*rbc*L), one from *ndh*F sequences (Rubiaceae_43_*ndh*F), and one from a matrix of both *rbc*L and *ndh*F sequences (Rubiaceae_43_*rbc*L/*ndh*F). The other new taxon set is a *rbc*L matrix that included the 43 species from above, plus 119 other *rbc*L sequences (Table 1) from the family (Rubiaceae_163_*rbc*L). Eleven outgroups, representing the rest of the Gentianales or the sister group of Rubiaceae (Bremer, 1996b; Backlund et al., 1999), were used (Table 1). We also reanalyzed a set of Rubiaceae data from Bremer (1996a): Rubiaceae_33_morph, Rubiaceae_33_RFLP, Rubiaceae_33_*rbc*L, and Rubiaceae_33-morph/RFLP/*rbc*L.

DNA was extracted, amplified, and sequenced following the protocols of Bremer et al. (1995) for *rbc*L and of Kim and Jansen (1995) or Backlund et al. (1999) for *ndh*F. The *rbc*L data matrix comprised nucleotide positions 27–1,428 (1–26 are excluded because they are the 5′ PCR primer site) of the *rbc*L sequence, and the aligned *ndh*F data matrix comprised aligned nucleotide positions 1–2,303. Both genes encode proteins from the chloroplast genome. Alignments were made manually by using the reading frames of the corresponding amino acid sequence. Gaps (only in *ndh*F) were treated as missing data. New *rbc*L and *ndh*F sequences were generated for 12 and 39 taxa, respectively. Sequences are accessioned in EMBL (Table 1). A few sequences were obtained from GenBank, and the others have been published earlier by the first author alone or in collaboration with others (Olmstead et al., 1993; Bremer et al., 1995; Bremer, 1996a, 1996b; Andreasen and Bremer, 1996; Endress et al., 1996, Andreasen et al., 1998, and Backlund et al., 1999).

Parsimony analyses of the new sets of taxa were conducted with PAUP version 3.1.1 (Swofford, 1993) on a PowerMac 9500/200, with all character changes

TABLE 1. List of investigated species. Earlier unpublished sequences are indicated with a * and voucher information is given. All other taxa, except the two indicated with a literature reference, have been published earlier by the first author alone or in collaboration with coauthors (see *Material and Methods*). Names of herbaria are abbreviated according to Holmgren et al. (1990).

	Accession EMBL/ GenBank	EMBL/ GenBank	Source/voucher information
	*rbc*L	*ndh*F	
Apocynaceae			
Alstonia scholaris R. Br.	X91760	AJ011982	
Kopsia fruticosa DC.	L14402	AJ235824	
Gelsemiaceae			
Gelsemium sempervirens Ait.	L14397	AJ011984	
Mostuea brunonis Didr.	L14404	AJ235828	
Gentianaceae			
Anthocleista grandiflora Gilg	L14389	AJ235829	
Gentiana procera Holm	L14398	L36400	*ndh*F from Olmstead and Reeves, 1995
Loganiaceae			
Antonia ovata Pohl	AJ235817	AJ235832	
Geniostoma rupestre J.R. Forst. & G. Forst.	Z68828	AJ235835	
Logania vaginalis F. Muell.	Z68826	AJ235837	
Spigelia marilandica L.	L14007		
Spigelia anthelmia L.		AJ235840	
Strychnos nux-vomica L.	L14410		
Strychnos potatorum L. f.		AJ235841	
Rubiaceae			
Agathisanthemum bojeri Klotzsch	Z68787		
Aidia micrantha (K. Schum.) Bullock ex F. White	Z68844		
Alberta magna E. Mey.	AJ224843		
Alberta magna E. Mey.	Y18708*	AJ236282*	Cult. Kirstenbosch Bot. Gard., Bremer & Bremer 3773 (UPS)
Alibertia edulis A. Rich.	Z68843		
Alseis "lugonis"	Y18709*	AJ236283*	Ecuador, Bremer et al. 3353 (QCA, QCNE,UPS)
Amphidasya ambigua (Standley) Standley	Y11844		
Anthospermum herbaceum L. f.	X83623	AJ236284*	Tanzania, Bremer 3093 (UPS)
Antirhea lucida (Sw.) Benth. & Hook.	X83624		
Aoranthe penduliflora (K. Schum.) C. Somers	Y11845		
Argostemma hookeri King	Z68788		
Bertiera breviflora Hiern	X83625		
Bouvardia glaberrima Engelm.	X83626		
Burchellia bubalina Sims	Z68833		
Calochone redingii (De Wild.) Keay	Z68845		
Calycophyllum candidissimum DC.	X83627	AJ236285*	Cult. Fairchild Bot. Gard., Sanders 1805 (FTG)
Canthium coromandelicum (Burm.f.) Alston	Z68851		
Capirona decorticans Spruce	Y18710*	AJ236286*	Ecuador, Bremer et al. 3357 (QCA,UPS)
Carphalea glaucescens (Hiern) Verdc.	Z68789	AJ236287*	Herb. Med. Plant Project 215 (UPS)
Casasia clusiifolia Urb.	Z68831		
Catesbaea spinosa L.	X83628		
Cephalanthus natalensis Oliver	Y18711*		Cult. Kirstenbosch Bot. Gard., Bremer & Bremer 3768 (UPS)
Cephalanthus occidentalis L.	X83629	AJ236288*	Cult. Stockholm Univ., Forbes s.n. (SUNIV)
Chassalia parviflora Verdc.	Z68790		
Chazaliella abrupta (Hiern) Petit & Verdc.	Z68791		
Chimarrhis hookeri K. Schum	Y18712*	AJ236289*	Herb. Lao Magin 44 (UPS)
Chiococca alba Hitchc.	L14394	AJ130835*	Cult. Stockholm Univ., Bremer 2703 (UPS)

(continued on next page)

TABLE 1. Continued.

	Accession EMBL/ GenBank	EMBL/ GenBank	Source/voucher information
	*rbc*L	*ndh*F	
Chomelia sp.	Y11846		
Cinchona pubescens (succirubra) Vahl	X83630	AJ235843	
Coccocypselum hirsutum Bartl. ex DC.	X87145		
Coffea arabica L.	X83631	AJ236290*	Cult. Fairchild Bot. Gard., Sanders 1803 (FTG)
Condaminea corymbosa DC.	Y18713*	AJ236291*	Ecuador, Bremer et al. 3387 (QCA, QCNE, UPS)
Conostomium quadrangulare (Rendle) Cufod.	Z68792		
Coprosma pumila Hook. f.	X87146		
Coptosapelta flavescens Korth.	Y18714*	AJ236292*	Thailand, Puff 950720-1/2 (WU)
Coussarea macrophylla Muell. Arg.	Y11847		
Cremaspora triflora (Thonn.) K. Schum.	Z68856		
Cruciata glabra (L.) Ehrend.	X81097		Manen and Natali, 1996
Cubanola domingensis (Britton) A. Aiello	X83632		
Damnacanthus indicus Gaertn. f.	Z68793		
Danais xanthorrhoea (K. Schum.) Bremek.	Z68794	AJ236293*	Tanzania, Bremer 3079 (UPS)
Deppea grandiflora Schlecht.	X83633		
Didymaea alsinoides (Cham. & Schlecht.) Standley	Z68795		
Didymosalpinx norae (Swynnerton) Keay	Z68834		
Emmenopterys henryi Oliver	Y18715*	AJ236294*	Cult. Meise Bot. Gard. Robbrecht s.n. (UPS)
Erithalis fruticosa L,	X83635	AJ236295*	Cult. Fairchild Bot. Gard., Mea 1803 (FTG)
Euclinia longiflora Salisb.	Z68835		
Exostema caribaeum Borhidi & Muniz	X83636	AJ236296*	Cult. Fairchild Bot. Gard., Misitis 2 (FTG)
Faramea multiflora A. Rich.	Z68796		
Feretia aeruginescens Stapf	Z68857		
Gaertnera sp.	Z68797		
Gardenia thunbergia L. f.	X83637	AJ235844	
Genipa americana L.	Z68839		
Geophila repens (L.) I. M. Johnston	Z68798		
Glossostipula concinna (Standley) D.H. Lorence	Z68846		
Gonzalagunia affinis Standley ex Steyerm.	Y11848		
Guettarda uruquensis Cham. & Schlecht.	X83638	AJ236297*	Cult. Fairchild Bot. Gard., Gillis 9575 (FTG)
Haldina cordifolia (Roxb.) C.E. Ridsdale	X83639		
Hallea (*Mitragyna*) *rubrostipulata* (K Schum.) Leroy	X83640		
Hamelia cuprea Griseb.	X83641		
Hedyotis fruticosa L.	Z68799		
Heinsia crinita (Afzel.) G. Tayl.	Y11849		
Hillia triflora (Oersted) C.M. Taylor	X83642	AJ236298*	Cult. Univ. of Colorado, Bremer 3101 (UPS)
Hippotis sp.	Y11850		
Hoffmannia refulgens ghiesbreghtii	X83644		
Hydnophytum formicarum Jack	X83645		
Isertia laevis[A] (Triana) B.M. Boom	Y11852		
Isertia laevis[B] (Triana) B.M. Boom	Y11853		
Isertia pittieri (Standley) Standley	Y11851		
Ixora biflora Fosberg	Z68866		
Ixora coccinea L.	X83646	AJ236299*	Cult. Uppsala, Bot. Gard., Bremer 3104 (UPS)
Ixora hookeri (Oudem.) Bremek.	Z68864		
Kailarsenia ochreata (F. Muell.) C.F. Puttock	Z58847		
Keetia zanzibarica (Klotzsch) Bridson	X83647		
Kohautia caespitosa Schinzl.	Z68800		

(continued on next page)

TABLE 1. Continued.

	Accession EMBL/ GenBank *rbc*L	EMBL/ GenBank *ndh*F	Source/voucher information
Kraussia floribunda Harv.	Z68858		
Ladenbergia pavonii (Lamb.) Standley	Z68801		
Lasianthus pedunculatus E. A. Bruce	Z68802		
Leptactina platyphylla (Hiern) Wernham	Z68867		
Luculia grandifolia Ghose	X83648	AJ011987	
Manettia bicolor Paxt.	Z68803		
Mapouria Cf. umbrosa	Z68804		
Massularia acuminata (G. Don) Bullock ex Hoyle	Z68841		
Meyna tetraphylla (Hiern) Robyns	X83649		
Mitchella repens L.	Z68805		
Mitriostigma axillare Hochst.	X83650		
Morinda citrifolia L.	X83651	AJ236300*	Cult. Uppsala Bot. Gard., Bremer 3106 (UPS)
Mussaenda arcuata Poir.	Y11854	AJ236301*	Gabon, McPehrson 16213 (M0)
Mussaenda erythrophylla Schum. & Thonn.	X83652	AJ130836*	Cult. Fairchild Bot. Gard., Gillis 10838 (FTG)
Mycetia malayana Craib	Z68806		
Myrmecodia platyrea Becc.	X87147		
Nauclea orientalis L.	X83653		
Nertera granadensis Druce	X83654		
Neurocalyx zeylanicus Hook.	Z68807		
Oldenlandia cf. corymbosa	X83655	AJ130837*	Tanzania, Bremer 3075 (UPS)
Oldenlandia goreensis Summerhayes	Z68808		
Opercularia vaginata Labill.	Z68809		
Ophiorrhiza mungos L.	X83656	AJ130838*	Cult. Meisse Bot. Gard., Robbrecht s.n. (UPS)
Oxyanthus cf. zanguebaricus (Hiern) Bridson	Z68838		
Oxyanthus pyriformis (Hochst.) Skeels	Z68836		
Palicourea sp.	Z68810		
Paracoffea melanocarpa (Welw. ex Hiern) Leroy	Z68853		
Parapentas silvatica (K. Schum.) Bremek.	X83657		
Pauridiantha paucinervis (Hiern) Bremek.	Z68811	AJ236302*	Tanzania, Bremer 3090 (UPS)
Pavetta abyssinica Fresen.	Z68863		
Pavetta lanceolata Eckl.	Z68865		
Pentagonia macrophylla Benth.	X83658	AJ236303*	Cult. Duke Univ., McDade 595A (DUKE)
Pentanisia longituba Oliver	Z68812		
Pentanopsis fragrans Rendle	Z68813		
Pentas lanceolata (Forssk.) Deflers	X83659	AJ236304*	Cult. Univ. of Connecticut, Bremer 2702 (S)
Pentodon pentandrus Vatke	X83660		
Phuopsis stylosa Benth. & Hook. f.	X81103		
Phyllis nobla L.	Z68814		
Pinckneya pubens Michx.	X83661	AJ130839*	Cult. Univ. of California, Forbes s.n. (S)
Placopoda virgata Balf. f.	Z68815		
Plocama pendula Ait.	Z68816		
Pogonopus speciosus (Jacq.) Schum.	X83662	AJ236305*	Cult. Fairchild Bot. Gard., Gillis 11168 (FTG)
Porterandia crosbyi (Burkill) A.C. Smith & S.P. Darwin	Z68840		
Posoqueria latifolia Roem. & Schult.	Z68850		
Pouchetia gilletii De Wild.	Z68859		
Pseudomussaenda flava Verdc.	Y11855	AJ236306*	Cult. Copenhagen Bot. Gard., Nissen s.n. (UPS)
Pseudosabicea arborea (K. Schum.) N. Halle	Y11856		
Psilanthus mannii Hook. f.	Z68852		

(continued on next page)

TABLE 1. Continued.

	Accession EMBL/ GenBank *rbc*L	EMBL/ GenBank *ndh*F	Source/voucher information
Psychotria kirkii (*bacteriophila*) Hiern	X83663	AJ236307*	Cult. Uppsala Bot. Gard., Bremer 3866 (UPS)
Psychotria peteri E. Petit	Z68817		
Psychotria poeppigiana Muell. Arg.	Z68818		
Psychotria sp.	Z68819		
Rachicallis americana Hitchcock	X83664		
Ramosmania rodriguesii D.D. Tirvengadum	Z68860		
Randia aculeata L.	Z68832		
Randia (*Atractocarpus*) *fitzalani* F. Muell. ex Benth.	Z68848		
Randia moorei F. Muell. ex Benth.	Z68849		
Richardia pilosa Ruiz & Pav.	Z68820		
Rogiera suffrutescens A. Borhidi	X83665	AJ236308*	Cult. Univ. of Connecticut, Bremer 2712 (S)
Rondeletia odorata Jacq.	Y11857	AJ235845	
Rosenbergiodendron longiflorum (Ruiz & Pav.) Fagerl.	Z68830		
Rothmannia longiflora Salisb.	Z68837		
Rubia tinctorum L.	X83666		
Rudgea cf. *lorentensis* Standley	Z68821		
Rustia splendens Standley,	Y18716*	AJ236310*	Herb. UPS, Delprete 6378 (UPS)
Rutidea orientalis Bridson	Z68862		
Sabicea villosa Roem. & Schult.	Y11858		
Sarcocephalus latifolius (J. E. Smith) E. A. Bruce	X83667		
Schradera subandina Krause	Y11859		
Scyphiphora hydrophyllacea Gaertn. f.	Y18717*	AJ236311*	Sri Lanka, K. Bremer et al. 99 (S)
Serissa foetida Lam	Z68822		
Simira viridiflora (Allem. & Saldanha) Steyermark	Y18718*	AJ236312*	Cult. Rio. de Janeiro, da Silva Neta 100395 (RB)
Spermacoce (*laevis*) *assurgens* Ruiz & Pav.	Z68823	AJ236309*	Tanzania, Bremer 3062 (UPS)
Spermadictyon suaveolens Roxb.	Z68824		
Strumpfia maritima Jacq.	Y18719*	AJ236313*	Herb. UPS, Killip 41057 (UPS)
Sukunia longipes A. C. Smith	Z68842		
Synaptantha tillaeacea Hook. f.	Y18720*		Herb. K, Lazarides & Palmer 272 (K)
Tamridaea capsulifera (Balf.) Thulin & B. Bremer	Y11860		
Tarenna cymosa (Willd. ex Roem. & Schultes) B. Verdc.	X83634		
Tarenna neurophylla (S. Moore) Bremek.	Z68861		
Theligonum cynocramb	X83668		
Timonius sp.	Y18721*		Irian Jaya, Ridsdale 2204 (L)
Tricalysia cryptocalyx Baker	Z68854		
Tricalysia ovalifolia Hiern	Z68855		
Uncaria rhynchophylla Miq.	X83669		
Vangueria madagascarensis Gmelin	X83670	AJ130840*	Cult. Fairchild Bot. Gard., Sanders 1798 (FTG)
Virectaria major (K. Schum.) Verdc.	Y11861		
Warszewiczia cordata Spruce ex K. Schum.	Y18722*	AJ236314*	Ecuador, Bremer et al. 3333 (QCA, QCNE, UPS)

weighted equally. Only phylogenetically informative characters were included. Heuristic searches were performed with 20–100 random stepwise additions and TBR branch swapping with MULPARS. The amount of support for monophyletic groups was evaluated by using bootstrap values (Felsenstein, 1985) with 1,000 replicates and NNI branch swapping. We used the bootstrap method instead of alterna-

tive methods because most other family-level analyses have used this method, which facilitates comparisons with our results.

To investigate the connection between character numbers and percentage of supported nodes (arbitrarily set to 75%), as well as the effect from different kinds of data, we analyzed the different subsets with different numbers of characters from the Rubiaceae_43 and Rubiaceae_33 studies, respectively: Rubiaceae_43_*rbc*L, Rubiaceae_43_*ndh*F, and Rubiaceae_43_*rbc*L/*ndh*F; and Rubiaceae_33_morph, Rubiaceae_33_RFLP, Rubiaceae_33_*rbc*L, and Rubiaceae_33-morph/RFLP/*rbc*L. From each of the larger subsets (Table 2) we repeatedly and randomly jackknifed characters until the number of the characters included in the smaller data sets were reached (e.g., from the Rubiaceae_43_*rbc*L/*ndh*F matrix of 715 characters we randomly sampled 492 and 223 informative characters). All of these matrices were bootstrapped with 1,000 replicates and NNI branch swapping. Confidence intervals (95% CI) were calculated and standard *t*-tests were applied. To correct for multiple tests on the same data, we used sequential Bonferroni correction (Holm, 1979; Rice, 1989).

To investigate the connection between the number of taxa and percentage of supported nodes, we included information from seven "original" *rbc*L analyses (Rubiaceae_163_*rbc*L and Rubiaceae_43_*rbc*L from this paper; Rubiaceae_49_*rbc*L from Bremer et al. [1995]; Rubiaceae_33_*rbc*L from Bremer [1996a]; Rubiaceae_21_*rbc*L from Bremer [1996b]; and 97 Rubiaceae genera from Källersjö et al., [1998]). In all these Rubiaceae analyses, the sampling strategy has been not only to select taxa within suspected monophyletic groups that will represent the overall diversity of the family (strategy 3; Hillis, 1998), but also to pinpoint the position of certain interesting taxa. The purpose in all analyses was not to subdivide long branches. From the data set with the largest number of taxa (Rubiaceae_163_*rbc*L; Table 2), we repeatedly and randomly jackknifed taxa until the number of the taxa included in the smaller data sets were reached (97, 49, 43, 33, and 21). To get a more even distribution of taxon sample sets, we also analyzed 2 taxon sets with 125 and 75 taxa each. All of the randomly selected matrices were bootstrapped with 1,000 replicates and NNI branch swapping.

We specified the support in two ways. In comparisons of general support between different trees, we calculated the percentage of nodes with bootstrap values of $\geq$ 75%; for particular nodes of the Rubiaceae and its subfamilies, we indicated the actual bootstrap values.

The analyses of the Rubiaceae were compared with results from several other selected studies from recently published family-level investigations. The criteria we used in selecting groups for comparison were (1) that the analyses should cover entire plant families, not just certain subfamilies or tribes; (2) that support values should be included in the publications; and (3) that the taxon sampling was the same for different number of characters or that the same character set was analyzed for different numbers of taxa. We found very few studies that fit these criteria; they dealt with only five families, including 18 analyses, and mainly of molecular data: Apocynaceae (Endress et al., 1996), Asteraceae (Jansen et al., 1998), Lardizabalaceae (Hoot et al., 1995), Saxifragaceae (Johnson and Soltis, 1995), and Solanaceae (Olmstead and Sweere, 1994). One of these included morphological data (Apocynaceae; Endress et al., 1996).

Results

Number of Characters, Different Data Sets, and Support

Within the Rubiaceae_43 analyses (Table 2 and Fig. 1a) based on different molecular data, the percentage of supported nodes within trees varied between 36.6% and 58.5%. For the original data the lowest value was from the *rbc*L analysis of 223 informative characters (5.19 char/taxon for the ingroup only, which will be the case in all analyses if not otherwise stated). The permutated data varied between 34.1% to 58.5%, with the lowest value from 1 of the

TABLE 2. Different family analyses, number of informative characters, proportion of characters per taxon, and number and percentage of nodes with bootstrap values ≥75%. Boldface indicates "original" analyses. Under the Rubiaceae_43 and Rubiaceae_33 follow additional analyses of new matrices with the same number of characters randomly jackknifed from the other matrices. For the randomly sampled matrices, the mean values ± SD are calculated for number and percentage of nodes supported with bootstrap values ≥75%.

Family_number of taxa_kind of data	Family node % support	No. of characters	Proportion of characters/taxon	Nodes with support ≥75% No.	Nodes with support ≥75% %
Rubiaceae_43_*rbc*L	**83**	**223**	**5.19**	**15**	**36.6**
Sample 1–8 from 43_*ndh*F		223	5.19	16.1±1.8	39±4.4
Sample 1–8 from 43_*rbc*L/*ndh*F		223	5.19	17.1±2.6	41.8±5.7
Rubiaceae_43_*ndh*F	**100**	**492**	**11.44**	**22**	**53.7**
Sample 1–8 from 43_*rbc*L/*ndh*F		492	11.44	21.1±2.1	51.5±5.1
Rubiaceae_43_*rbc*L/*ndh*F	**100**	**715**	**16.63**	**24**	**58.5**
Rubiaceae_33_morph		**35**	**1.06**	**1**	**3.2**
Sample 1–8 from 33_RFLP		35	1.06	3.6±1.3	11.7±4.2
Sample 1–8 from 33_*rbc*L		35	1.06	1.9±0.8	6.0±2.7
Sample 1–8 from 33_morph/RFLP/*rbc*L		35	1.06	3.0±1.7	9.7±5.5
Rubiaceae_33_RFLP		**161**	**4.88**	**13**	**41.9**
Sample 1–8 from 33_*rbc*L		161	4.88	8.4±1.2	27.0±3.8
Sample 1–8 from 33_morph/RFLP/*rbc*L		161	4.88	11.6±1.8	37.5±6.0
Rubiaceae_33_*rbc*L		**228**	**6.91**	**8**	**25.8**
Sample 1–8 from 33_morph/RFLP/*rbc*L		228	6.91	13.8±1.3	44.4±4.1
Rubiaceae_33_morph/RFLP/*rbc*L		**424**	**12.55**	**15**	**48.4**
Apocynaceae_14_morph	**< 50**	**36**	**2.57**	**3**	**25**
Apocynaceae_14_*rbc*L	**85**	**68**	**4.86**	**6**	**50**
Apocynaceae_14_*mat*K	**98**	**155**	**11.07**	**6**	**50**
Apocynaceae_14_morph/*rbc*L/*mat*K	**100**	**259**	**18.5**	**7**	**58.3**
Asteraceae_23_*rbc*L		**124**	**5.39**	**5**	**2.4**
Asteraceae_23_RFLP		**173**	**7.52**	**11**	**52.4**
Asteraceae_23_*ndh*F		**237**	**10.3**	**10**	**47.6**
Lardizabalaceae_7_18S	**99**	**25**	**3.57**	**2**	**40**
Lardizabalaceae_7_*atp*B	**100**	**42**	**6**	**2**	**40**
Lardizabalaceae_7_*rbc*L	**< 50**	**56**	**8**	**2**	**40**
Lardizabalaceae_7_18S/*atp*B/*rbc*L	**100**	**123**	**17.57**	**4**	**80**
Solanaceae_17_*rbc*L		**63**	**3.71**	**2**	**13.3**
Solanaceae_17_RFLP		**100**	**5.88**	**5**	**33.3**
Solanaceae_17_*ndh*F		**209**	**12.29**	**5**	**33.3**
Solanaceae_17_*rbc*L/RFLP/*ndh*F		**372**	**21.88**	**10**	**66.7**

*ndh*F analyses of 223 informative characters being 5.19 char/taxon. Trees with the highest percentage of supported nodes are from combined analyses of *rbc*L and *ndh*F data: 1 from the original analysis of 715 informative characters (16.63 char/taxon), and the other from a permutated matrix of 492 characters (11.44 char/taxon). The support was significantly higher for the *rbc*L/*ndh*F_492 than for the *rbc*L_223 matrix ($p < 0.001$, one-sample *t*-test) and also for the *rbc*L/ndhF_715 than for the *rbc*L/*ndh*F_492 matrix ($P = 0.018$). The percentage of supported nodes for the original *rbc*L_223 analysis was within the 95% CI of both *ndh*F_223 and *rbc*L/*ndh*F_223 permutated data (Fig. 1a). Likewise, the percentage of supported nodes for the original *ndh*F_492 analysis was within the 95% CI of *rbc*L/*ndh*F_223 permutated data. From the Rubiaceae_43 analyses, we conclude that increasing the number of characters significantly increased the number of supported nodes within the trees. However, we found no significant difference between the different data sets. The *ndh*F and *rbc*L/*ndh*F data did not differ significantly from the *rbc*L data in support per informative character, but the number of informative characters per sequenced nucleotide was higher for *ndh*F than *rbc*L.

Within the Rubiaceae_33 analyses (Table 2 and Fig. 1b) the percentage of supported

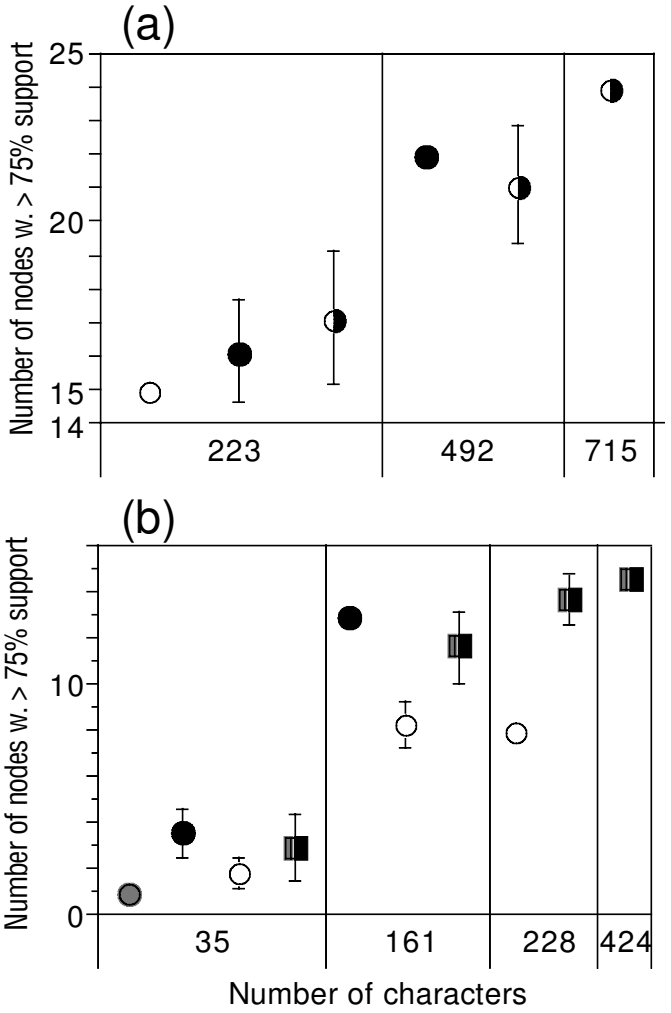

FIGURE 1. (a) Relationship between number of nodes supported by ≥75% bootstrap values, data set (*rbc*L = white, *ndh*F = black, combined *rbc*L/*ndh*F = white/black), and number of informative characters in the Rubiaceae_43 matrix. Data symbols represent mean values, lines 95% CI for the resampled data sets (n = 8). Circles without lines represent the observed values for the original data sets. (b) Same as in (a), but for morphology (grey), RFLP (black), and *rbc*L (white) in the Rubiaceae_33 matrix. Combined data are represented by squares.

nodes in the four different original analyses varied between 3.2% and 48.4%. The lowest value was from the morphological analysis (Rubiaceae_33_morph) of only 35 informative characters (3.2 char/taxon); the highest value was from the combined morphological and DNA data (Rubiaceae_33_morph RFLP/*rbc*L) with 424 characters (12.55 char/taxon). In the analyses of permutated data, the percentage of supported nodes varied between 3.2% and 51.6%, the lowest value being from 1 of the *rbc*L analyses. The highest value was from permutated data of the combined matrix of 228 characters (6.91 char/taxon). The permutated matrices of 161 and 228 characters yielded a significantly higher number of supported nodes than did the matrices of 35 characters ($P < 0.001$). The RFLP_35 matrices provided significantly better support than did the morphological analysis (morph_35, $P = 0.004$) but was not sig-

nificantly different from analyses of the *rbc*L_35 and morph/RFLP/*rbc*L_35 data sets ($P >= 0.05$, Bonferroni corrected *t*-test). The RFLP_161 data were significantly better than *rbc*L_161 (one-sample *t*-test, $P < 0.0001$) but were within the 95% CI of the combined data. The combined morph/RFLP/*rbc*L_228 was significantly better than the *rbc*L_228 data set ($P < 0.001$). As in the Rubiaceae_43 analyses, there was a positive correlation between character number and the number of nodes with bootstrap values ≥ 75%. In the Rubiaceae_33 analyses the RFLP gave better support per character than did the morphological or *rbc*L data. Comparisons of the consistency index (ci) and retention index (ri) show that all sets had about the same ci but that RFLP data had higher ri (RFLP, ci = 0.463, ri = 0.785; *rbc*L, ci = 0.459, ri = 0.672; morph, ci = 0.467 ri = 0.673).

Within the different analyses of Rubiaceae the number and percentage of supported nodes (≥75% bootstrap values) within a tree increased with higher numbers of informative characters and with a higher proportion of characters per taxon.

In other family studies using multiple data sets for the same taxa, there was also increased support when there were more characters or a higher ratio of characters per taxon. Within the Solanaceae (Olmstead and Sweere, 1994) the lowest percentage of supported nodes was 13.3% for the *rbc*L analysis of 63 (3.71 char/taxon) informative characters (including ingroup and one outgroup), and the highest 66.7% for a combined analysis (RFLP/*rbc*L/*ndh*F) of 372 characters (21.88 char/taxon). Within the Apocynaceae (Endress et al., 1996), the number of characters and percentage of supported nodes varied from 25% for the morphological analysis of 36 characters (2.57 char/taxon) to 58.3% for 259 characters (18.5 char/taxon) of the combined analysis (for the ingroup taxa). Within Asteraceae (Jansen et al., 1998) the number of characters and percentage of supported nodes varied from only 2.4% for the 124 *rbc*L characters (5.39 char/taxon) to 52.4% for the 173 (7.52 char/taxon) RFLP characters. Within Lardizabalaceae (Hoot et al., 1995) the percentage of supported nodes was 40% to 80%, the highest values coming from the analyses with the highest number of characters or the highest number of characters per taxon.

The percentage of supported nodes depended not only on the number of characters and taxa, but also on the kind of data examined. In the Rubiaceae_33 analyses, *rbc*L and RFLP data gave better support than morphological data and RFLP was better than *rbc*L. The combined data gave intermediate results. In the Asteraceae, the pattern was the same, with higher support for the RFLP data. In the Apocynaceae, the support for *rbc*L and *mat*K was equal, despite the fact that *mat*K includes many more informative characters. In the Lardizabalaceae, the combined matrix gave distinctly better support than either of the separate analyses (all these have the same support but different numbers of characters, with the greatest number of characters from *rbc*L, followed by *atp*B, and 18S).

Number of Taxa and Support

The analyses of repeatedly and randomly jackknifed taxa, using the same type of characters (*rbc*L) from the Rubiaceae_163_*rbc*L matrix (Table 3), demonstrated that an increase in taxon number decreases the percentage of nodes supported with ≥75% bootstrap values (Fig. 2; Spearman rank correlation $\rho = -0.964$, $P = 0.018$). The highest percentage was for the smallest sample of 21 taxa (52.6% of the nodes supported).

Support for Family and Subfamily Classification of Rubiaceae

When comparing the different Rubiaceae analyses, the highest support for the family and subfamily nodes were found in the analyses that included the largest number of characters or the highest proportion of characters per taxon, the *ndh*F and *rbc*L/*ndh*F data sets, respectively (Table 4). Strict consensus trees resulting from phylogenetic analyses of the 43-taxon sample of Rubiaceae based on *rbc*L, *ndh*F, and combined data for *rbc*L and *ndh*F are presented in Figures 3–5. The trees show the same general pattern of relationships, namely, that Rubiaceae are well-supported by bootstrap values ranging between 83% and 100%;

TABLE 3. Number of taxa and number and percentage of nodes with bootstrap values ≥75% for various Rubiaceae and Saxifragaceae analyses. Boldface indicates "original" analyses. Additional analyses are based on the same number of taxa but randomly jackknifed from the large Rubiaceae_163_*rbc*L matrix. For the randomly sampled matrices, the mean values ± SD are calculated for number and percentage of nodes supported with bootstrap values ≥75%.

Family_number of taxa_kind of data	No. of ingroup taxa	Nodes with support ≥75%	
		No.	%
Rubiaceae_21_*rbc*L	**21**	**10**	**52.6**
Sample 1–7 from 163_*rbc*L	21	9.3± 2.4	48.9± 12.8
Rubiaceae_33_*rbc*L	**33**	**9**	**29**
Sample 1–7 from 163_*rbc*L	33	14.6± 2.2	47.0± 7.2
Rubiaceae_43_*rbc*L	**43**	**15**	**36.6**
Sample 1–7 from 163_*rbc*L	43	15.9± 2.4	38.7± 5.9
Rubiaceae_49_*rbc*L	**49**	**18**	**38.3**
Sample 1–7 from 163_*rbc*L	49	18.7± 3.6	39.8± 7.6
Rubiaceae_75_*rbc*L	**75**	**20**	**27.4**
Sample 1–7 from 163_*rbc*L	75	26.1± 2.8	35.8± 3.8
Rubiaceae_97_*rbc*L	**97**	**26**	**27.4**
Sample 1–7 from 163_*rbc*L	97	33.6± 3.5	35.3± 3.7
Rubiaceae_125_*rbc*L	**125**	**30**	**24.4**
Sample 1–7 from 163_*rbc*L	125	41.0± 4.9	33.3± 4.0
Rubiaceae_163_*rbc*L	**163**	**49**	**30.4**
Saxifragaceae_20_*mat*K/*rbc*L/RFLP	**20**	**14**	**77.8**
Saxifragaceae_22_*mat*K/*rbc*L/RFLP	**22**	**13**	**65**
Saxifragaceae_43_*mat*K/*rbc*L/RFLP	**43**	**20**	**48.8**

that there are large and well-supported groups corresponding to the subfamilies Rubioideae (96–100%), Cinchonoideae s.s. (99–100%), and Ixoroideae s.l. (62–100%), and that some investigated taxa are not included in any of these groups. In all three analyses, the genera *Luculia* and *Coptosapelta* formed single branches at the base of the trees. The amount of resolution was very different within the subfamilies. The Rubioideae

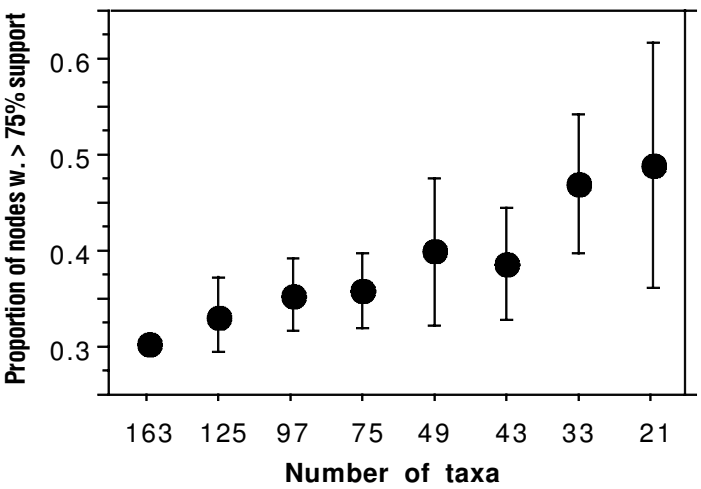

FIGURE 2. Relationship between number of nodes supported by ≥75% bootstrap values and number of taxa. Circles represent mean values, lines 95% CI for the resampled data sets ($n = 8$).

TABLE 4. Number of taxa, number of informative characters, proportion of characters to taxa, percentage of supported nodes (≥ 75% bootstrap value) within the trees, mean values of support for all nodes, and % bootstrap support for particular nodes within different analyses from the family Rubiaceae. Subfamily Rubioideae = RUBI, subfamily Cinchonoideae s.s. = CINC s.s., and subfamily Ixoroideae s.l. = IXOR s.l. A* indicates that the value cannot be calculated because no outgroups were included.

Family_number of taxa_kind of data	No. of taxa	No. of characters	Proportion char/taxon	% nodes with support ≥75%	Mean values of support for all nodes	Family node % support	RUBI	CINC s.s.	IXOR s.l.	No. of unclassified basal taxa
Rubiaceae_21_*rbc*L	21	181	8.62	52.6	73.68	53	91	90	71	1
Rubiaceae_33_*rbc*L	33	213	6.91	29	56.65	*	100	69	34	2
Rubiaceae_43_*rbc*L	43	223	5.19	36.6	63.88	83	96	99	62	2
Rubiaceae_49_*rbc*L	49	264	5.39	38.3	61.85	74	97	78	50	7
Rubiaceae_75_*rbc*L	75	279	3.72	27.4	50.52	*	83	94	47	2
Rubiaceae_97_*rbc*L	97	397	4.09	27.4	49.14	92	100	77	53	2
Rubiaceae_125_*rbc*L	125	355	2.84	24.4	50.67	*	96	90	46	2
Rubiaceae_163_*rbc*L	163	405	2.5	30.4	51.35	90	95	88	52	2
Rubiaceae_43_*ndh*F	43	492	11.44	53.7	63.54	100	100	100	100	2
Rubiaceae_43_*rbc*L/*ndh*F	43	715	16.63	58.5	69.27	100	100	100	100	2

were the most resolved and had more or less the same resolution in all three analyses. The Ixoroideae s.l., were the least resolved, obtaining the best result from the *ndh*F and combined analyses. The Cinchonoideae s.s. had an intermediate resolution and, again, there was better resolution in the *ndh*F and the combined analyses.

The new Rubiaceae_43 analyses include 11 additional genera including *Coptosapelta* of the tribe Coptosapelteae, which occurs as an unresolved basal node in the family outside of the three subfamilies. *Strumpfia*, a morphologically odd genus with fused anthers (Igersheim, 1993), has been treated as a genus of uncertain position, but this study clearly shows that it belongs to the subfamily Cinchonoideae s.s., close to the Chiococceae s.l. The other nine sequenced genera with uncertain systematic positions all belong in the Ixoroideae s.l. For example, *Scyphiphora* clearly groups in the clade including *Ixora* and *Gardenia*. The others, *Alseis, Condaminea, Emmenopterys, Pentagonia, Rustia, Simira, Capirona, Chimarris*, and *Warszewiczia*, which earlier were placed in different tribes of the subfamily Cinchonoideae, clearly belong to the Ixoroideae s.l.

In the *rbc*L analysis of 163 taxa of Rubiaceae and 11 outgroups (Table 1; tree not shown) the family node is well supported (90%). The three subfamily groups Rubioideae, Cinchonoideae s.s., and Ixoroideae s.l. are supported with 95%, 88%, and 52% bootstrap values, respectively. At the base of the family, outside the subfamily groups, we find the same two genera, *Luculia* and *Coptosapelta*, as in the Rubiaceae_43 analyses.

Discussion

Our results indicate that it may be better to add characters, rather than taxa, to obtain improved support for a phylogeny. If the goal is to identify major subgroups, the most cost-effective approach is to restrict the taxon sampling (from each subgroup) and to use more characters to get well-supported groups. However, high support values are not the only criterion for a robust phylogenetic hypothesis. High explanatory power will also be achieved if the analysis includes many taxa. Further, if one suspects long-branch attraction, it is probably better to add more taxa than characters to break up long branches (Graybeal, 1998) unless the taxa to be added are too slowly evolving (Poe and Swofford, 1999). Practicing systematists must by necessity consider using many characters and large numbers of taxa in species-rich groups to generate a robust phylogeny for tracing character evolution and developing classifications.

In the Rubiaceae and in the other plant families compared (Asteraceae, Apocynaceae, Lardizabalaceae, and Solanaceae) increasing the number of characters will increase the support considerably for the entire tree. Sanderson (1989) came to the same conclusion in a survey of published studies covering plants and animals (based on morphological and molecular data). We also expect, in agreement with theoretical arguments (Felsenstein, 1985; Sokal and Shao, 1985), that bootstrap values are strongly influenced by the number of characters. In a tree without homoplasy (perfect Hennigian data) a support value of 95% for a node requires at least three characters (Felsenstein, 1985). The smallest number of characters needed to obtain these support values for all nodes would be three times as many as the number of nodes—with no homoplasy, with an even distribution of characters on the tree, and with binary characters. This should be kept in mind when considering analyzing only a few characters, such as the Rubiaceae_33_morph analysis, which included only 35 characters. Despite these earlier studies, the influence of character numbers has received little attention and has had little effect on how phylogenetic studies of plants have been designed, especially morphological analyses. Olmstead and Sweere (1994) discussed the relationship between character number, bootstrap values, and the "decay" index in the context of combining different molecular data sets to get better support. However, these authors stressed the approach of combining different character sets (genes) more than just increasing the number of characters.

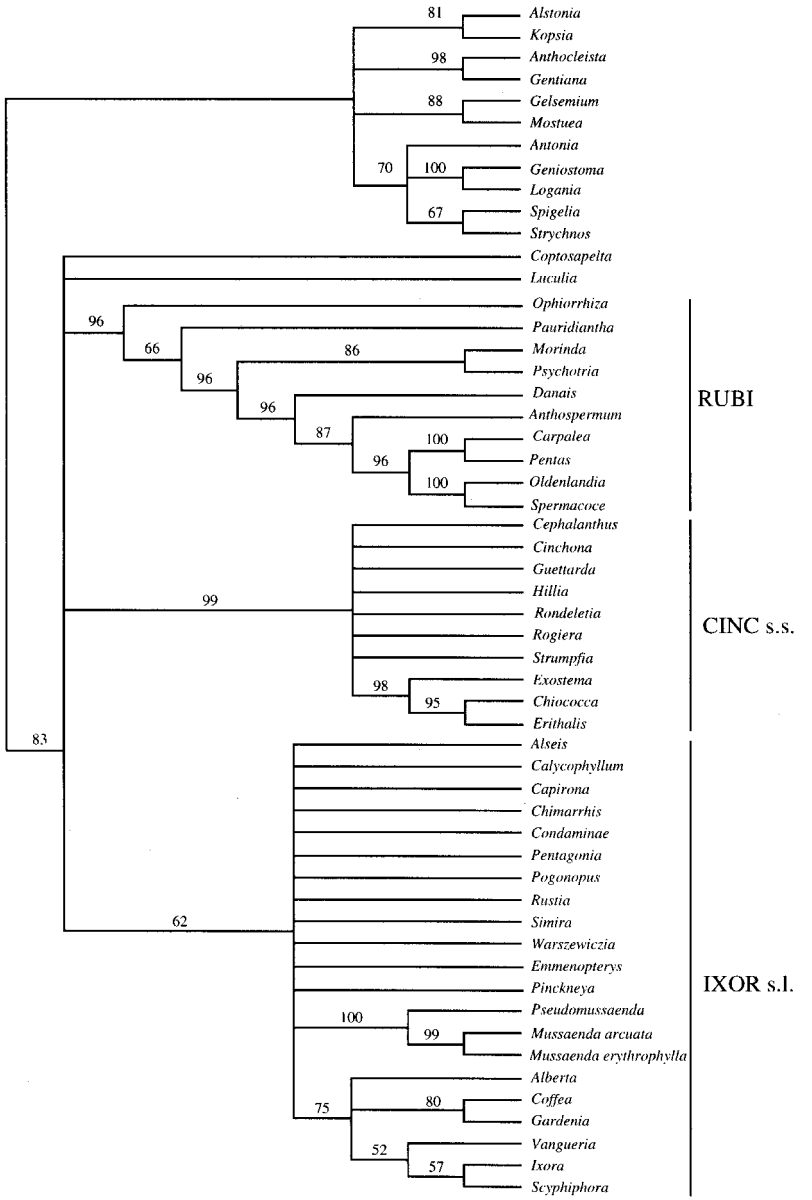

FIGURE 3. Strict consensus tree of 16,583 equally parsimonious trees of Rubiaceae based on *rbc*L sequences, after a heuristic search with 20 replicates and TBR branch swapping (all trees with a ci of 0.387 and a ri of 0.632). Numbers above nodes indicate bootstrap values (≥50%) from 1,000 replicates with NNI branch swapping; all bootstrap values ≥50% are indicated. Vertical bars and corresponding letters represent: CINC s.s. = subfamily Cinchonoideae s.s., IXOR s.l. = subfamily Ixoroideae s.l., RUBI = subfamily Rubioideae.

In the Rubiaceae (Table 3 and Fig. 2), increasing the number of taxa will decrease support within the tree. In the Saxifragaceae (Johnson and Soltis, 1995), there is also a decrease in number of supported nodes. In this example, the support is very dependent on the taxon sampling, which is illustrated by the large differ-

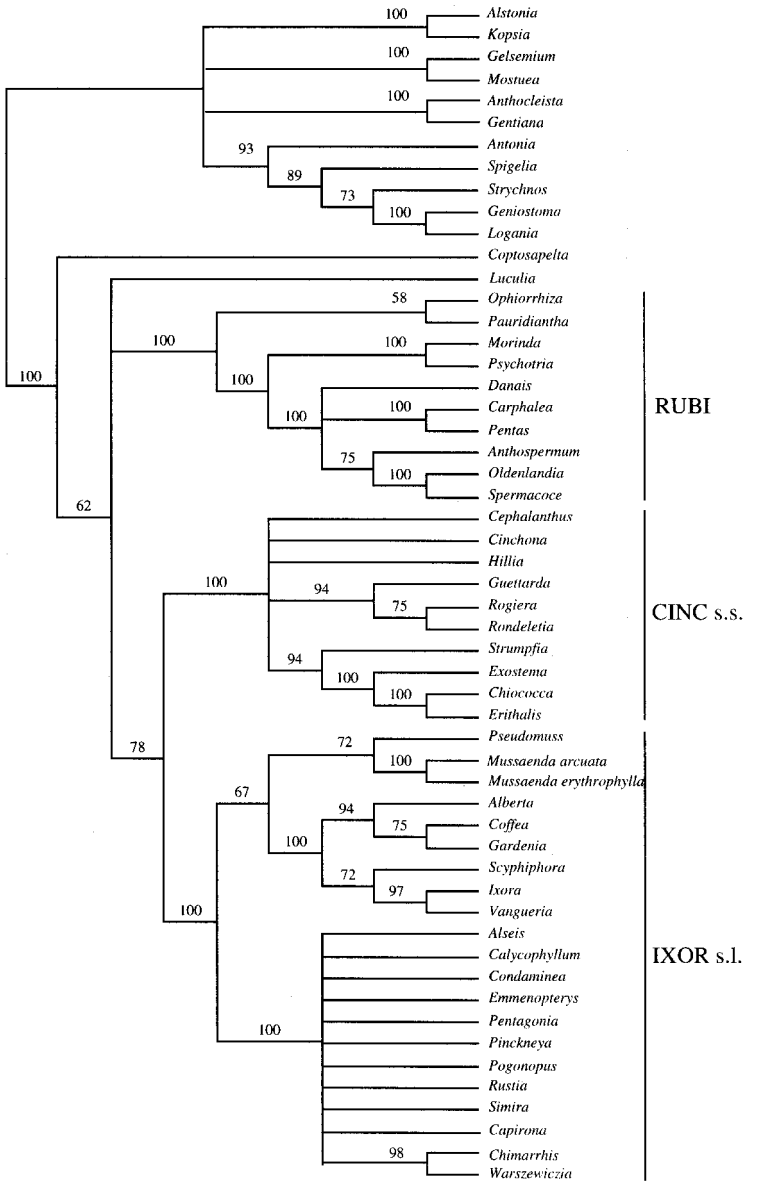

FIGURE 4. Strict consensus tree of 1,088 equally parsimonious trees of Rubiaceae based on *ndh*F sequences, after a heuristic search with 20 replicates and TBR branch swapping (all trees with a ci of 0.465 and a ri of 0.664). Numbers above nodes indicate bootstrap values (≥50%) of 1,000 replicates with NNI branch swapping, all bootstrap values ≥50% are indicated. Vertical bars and corresponding letters represent: CINC s.s. = subfamily Cinchonoideae s.s., IXOR s.l. = subfamily Ixoroideae s.l., RUBI = subfamily Rubioideae.

ence in support between the two smallest samples, which differ only by two taxa (Saxifragaceae_20_*mat*K/*rbc*L/RFLP, 77.8% supported nodes ≥75%; Saxifragaceae_22_*mat*K/*rbc*L/RFLP, 65.0% supported nodes ≥75%, and Saxifragaceae_43_*mat*K/*rbc*L/RFLP, 48.8% supported nodes ≥75%). Furthermore, in the Rubiaceae there is a signifi-

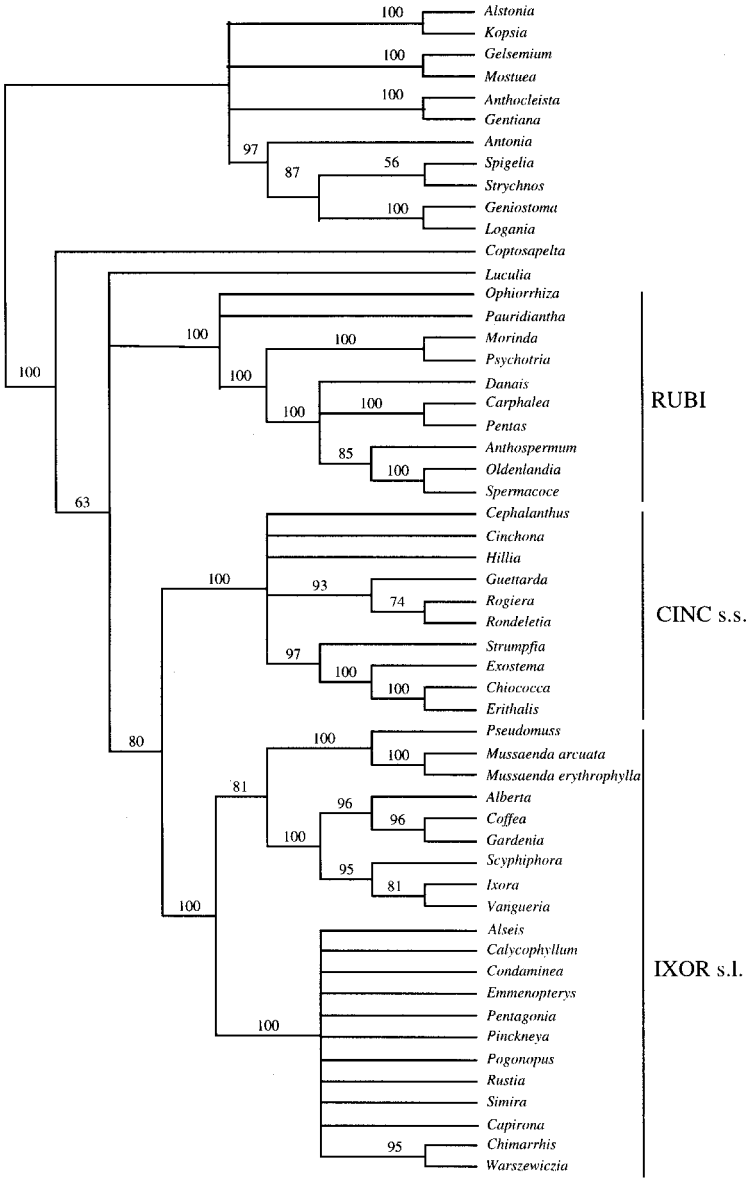

FIGURE 5. Strict consensus tree of 13 equally parsimonious trees of Rubiaceae based on *rbc*L and *ndh*F sequences, after a heuristic search with 100 replicates and TBR branch swapping (all trees with a ci of 0.442 and a ri of 0.652), all bootstrap values ≥50% are indicated. Numbers above nodes indicate bootstrap values (≥50%) of 1,000 replicates with NNI branch swapping. Vertical bars and corresponding letters represent: CINC s.s. = subfamily Cinchonoideae s.s., IXOR s.l. = subfamily Ixoroideae s.l., RUBI = subfamily Rubioideae.

cant connection between the number of taxa and number of informative characters, but with an increased number of taxa the proportion of characters to taxa will decrease significantly for a particular data source (e.g., *rbc*L). The support is positively correlated with the proportion of characters per taxon, which probably is the main reason

for the decreasing support when more taxa are added and thus the proportion of characters to taxa are decreasing (Table 4).

If the tendency to decrease the percentage of supported nodes with increasing number of taxa is general, it would be problematic for analyses of very large data sets. However, in the large analysis of 2,538 *rbc*L sequences of green plants (Källersjö et al., 1998), no such clear effect, at least not for specific groups, was detected in the analysis in comparison with other analyses of the same taxonomic groups. One explanation may be that the support for a particular group is not affected if distant taxa are added to the analysis.

Are Some Kinds of Data Better Than Others?

In the rather short history of cladistic analysis, there have been several intense discussions of which characters should be used and how different character sets should be treated. One debate concerned morphological versus molecular data (reviewed in Hillis, 1987; Patterson, 1987; Fernholm et al., 1989; Givnish and Sytsma, 1997). The consistency of the characters or the level of homoplasy has been shown to be about the same for morphological and molecular data (e.g., Sanderson and Donoghue [1989]; in Rubiaceae, Bremer [1996a]; and in Apocynaceae, Endress et al. [1996]). However, the number of characters is usually much higher in molecular data sets. In the present study (Table 2, Rubiaceae_33_morph and Apocynaceae_14_morph), support in the morphological analyses is very low. This is partly a result of the very low number of characters used. When the same low number of characters are sampled and analyzed from *rbc*L and RFLP data, the support is also very low, but for the RFLP data, the support is significantly higher than for morphology (Fig. 1b). The problem with morphological data is to get enough characters; in the Rubiaceae_33_morph analysis (Bremer, 1996a), the 35 characters are not sufficient to resolve the tree, or to give many supported nodes. For example, the morphological matrix should have at least 93 characters to provide 95% bootstrap values for all nodes—if the characters are binary, without homoplasy, and evenly distributed on the tree.

Another important debate is about whether morphological and molecular sets of characters should be analyzed separately or together. (The debate started with Miyamoto [1983, 1985], and has been continued by e.g., Hillis [1987], Kluge [1989], Sytsma [1990], Barrett et al. [1991, 1993], Bull et al. [1993], de Queiroz [1993], Kluge and Wolf [1993], Nelson [1993], Eernisse and Kluge [1993], Bremer [1996a], Page [1996]. The more general debate on combining data has continued and, interestingly, Miyamoto [Miyamoto and Fitch, 1995], who was one of the first advocates for the combined approach, has changed his opinion.) There are a number of reasons to combine data. One is to increase the number of characters; another is that a combined analysis will result in a more globally parsimonious solution than if the data are analyzed separately and later combined in a consensus tree.

A third discussion, perhaps not so intense, has focused on which molecular markers should be used. Plant molecular systematics began with restriction site mapping of chloroplast DNAs (e.g., Sytsma and Schaal, 1985; Jansen and Palmer, 1987; Chase and Palmer, 1989, Wilson et al., 1990). Although the restriction site approach still provides very valuable phylogenetic information, especially at lower taxonomic levels, there are some limitations of the approach (reviewed in Jansen et al., 1998). Rather soon, these methods became less popular, and sequencing of the chloroplast gene *rbc*L (Doebley et al., 1990; Soltis et al., 1990; Olmstead and Palmer, 1994) became the method á *la mode*. The *rbc*L gene is still the most widely sequenced gene, but the search for other molecular markers has been intense, as evidenced by the 1993 symposium entitled "Alternative genes for phylogenetic reconstruction in plants" (Missouri Botanical Garden, 1995). In the symposium contributions, eight different genes were investigated, both chloroplast and nuclear, and in several of these studies alternative genes provided greater phy-

logenetic potential than *rbc*L. This conclusion is often justified by a higher observed nucleotide difference per site: *mat*K (Johnson and Soltis, 1995); and *ndh*F (Olmstead and Reeves, 1995). Several investigators stress the importance of combining different molecular data sets when highly divergent taxa are studied (cf. Hoot et al., 1995; Nickrent and Soltis, 1995). However, nothing is explicitly said about the need for increasing just the number of informative characters. A good approach would perhaps be to sequence longer informative regions of DNA or to sequence a larger number of genes to get more characters. The approach of sequencing large parts of DNA has been adopted in zoology, where the whole mitochondrial genome is sequenced and a large portion is analyzed (e.g., Cummings et al., 1995; Arnason et al., 1996).

Different character sets (genes) have different information content, depending on the variability of the DNA—as can be seen from the distance values or level of homoplasy, as measured by the ci, for the particular piece of DNA within a specific taxon sample. In the present study of Rubiaceae (Fig. 1b) the kind of data studied affects the levels of support; e.g., in the 33 analyses the RFLP data are significantly better than the morphological and *rbc*L data. However, when comparing results from different families, we find that particular genes do not always give the same level of support or resolution, which perhaps should not be expected, because families are not always evolutionarily comparable units. Often one gene performs much better than other genes in one family than it does in another family, and also better than what is expected from the number of characters. For example, in the Apocynaceae, the support is the same in the *mat*K and *rbc*L analyses, but the number of informative characters is twice as many for *mat*K as for *rbc*L. In the Lardizabalaceae, the support is the same for 18S, *atp*B, and *rbc*L, despite the fact that the number of characters is different (25, 42, and 56, respectively). Restriction site data seem to perform much better than sequence data of both *rbc*L and *ndh*F genes in the examples of Solanaceae and Asteraceae (Table 2). This is a bit ironic because restriction site comparisons have been abandoned and replaced by sequencing in most laboratories—obviously not because of lack of information, but because of the complicated and time-consuming work involved (cf. Givnish and Sytsma, 1997; Jansen et al., 1998).

Will Addition of a New Data Set from ndhF Better Resolve or Support the Subfamilial Classification of Rubiaceae?

None of the phylogenetic trees of the Rubiaceae show complete resolution or high support for relationships of basal nodes. The resolution does not drastically change with the addition of new data from *ndh*F sequences. The percentage of supported nodes increases within the whole tree, and the support for the family and subfamily nodes increases. Most genera of Rubiaceae that have been investigated cladistically are placed with high support in one of the three major subfamilies, but in all analyses there are a few genera (*Luculia* and *Coptosapelta*) left as single branches at the base of the tree. In many cases, as in Rubiaceae, the family is easily recognized and strongly supported, but why are some nodes unresolved or with low support? Different possible explanations could be that (1) we have not yet identified the best characters or markers for these nodes, or, perhaps along the same line, (2) we have not obtained a sufficient number of characters, or (3) the unresolved or weakly supported nodes reflect rapid speciation or slow character evolution (or both) during certain time periods. In many molecular studies, strongly supported branches correspond to groups that are morphologically very distinct and easy to recognize. Because systematics has a long history, many morphologically distinct groups have already been identified and named, and in angiosperm systematics the family rank concept has been used for many of these easily recognized groups. If systematists already have characterized these groups and named them as families, then obviously we should find families that are easily recognized and well-supported. Furthermore, it follows that subfamilies are likely to be more difficult to recognize and weakly supported

because if they were not so, they would already be named as families.

Phylogenetic Implications in Rubiaceae

The first phylogenetic analysis of the Rubiaceae was performed with RFLP data (Bremer and Jansen, 1991). In this analysis, no support values were calculated; more recently, however, analyses using several different molecular data sets and morphology data have been performed, all of which included support values. Seven of these analyses are presented in this study. The topologies are mostly congruent in the different analyses but the support for the various groups differs (Table 4).

The new phylogenetic analyses of Rubiaceae presented here (Rubiaceae_43 and Rubiaceae_163) and the earlier analyses identify three large well-supported clades containing most of the investigated taxa. These major groups are easily circumscribed as the three subfamilies Rubioideae, Cinchonoideae s.s., and Ixoroideae s.l., but with a few unclassified genera at the base of the tree. The subfamily Rubioideae is strongly supported in all analyses. This subfamily has been accepted by most systematists ever since it was established in its modern sense (Verdcourt, 1958), and its circumscription is now only slightly different (Robbrecht, 1988; Bremer, 1996b; Manen and Natali, 1996; Natali et al., 1996). The Rubioideae are characterized by a combination of morphological characters, including herbaceous habit, raphides, valvate aestivation, and articulate hairs. The second subfamily, Cinchonoideae s.s., has strong support in the analyses. The circumscription of this subfamily has been very different in the past, and compared with earlier classifications, all molecular phylogenies indicate that the subfamily should be much more narrowly circumscribed by excluding the *Mussaenda* group (formerly part of tribe Isertieae [Bremer and Thulin, 1998]) and large parts of the tribes Cinchoneae and Condamineeae (cf. Bremer, 1996b; this study). The taxa in the narrowly circumscribed Cinchonoideae are characterized by a few nonmolecular traits: Most members are woody, aestivation is imbricate, and many contain complex indole alkaloids. The level of support for the third subfamily, Ixoroideae s.l., varies in the different analyses, but it is very high in all those with many characters. Before molecular studies, this subfamily included only the taxa with contorted aestivation and a stylar pollen presentation (Ixoroideae s.s.; [Bremekamp, 1966; Robbrecht, 1988]), corresponding to the clade including *Ixora* and *Gardenia* (Figs. 3–5). However, all molecular investigations indicate that many taxa previously included in Cinchonoideae, including most taxa with enlarged calyx lobes (e.g., *Calycophyllum, Emmenopterys, Warszewiczia*, and *Pinckneya*), are more closely related to the Ixoroideae s.s. than to the Cinchonoideae. Because these taxa differ in many morphological characters from the Ixoroideae s.s., (e.g., in aestivation and fruit types), it is difficult to find any unique diagnostic morphological characters for the subfamily. Nevertheless, they are best placed in the Ixoroideae s.l.; otherwise, several new subfamilies would have to be described within this clade.

Luculia and *Coptosapelta* are found as isolated lineages at the base of the molecular phylogenies (Figs. 3–5). *Luculia* has been investigated several times with molecular (e.g., RFLP in Bremer and Jansen [1991]; *rbc*L in Bremer et al. [1995]) and morphological (Bremer and Struwe, 1992) data. It is usually placed in an unresolved basal position. Robbrecht (1988) classified *Luculia* in the same tribe as *Cinchona* of the subfamily Cinchonoideae. In a treatment of the South American tribe Cinchoneae, Andersson and Persson (1991) excluded *Luculia* and placed it in a new emended tribe Coptosapelteae with *Coptosapelta* and several other genera. However, our data do not support a close relationship between *Luculia* and *Coptosapelta*. The tribe Coptosapelteae, as circumscribed by Andersson and Persson (1991), is not monophyletic according to the molecular data. Several genera of their tribe are not even closely related to *Coptosapelta* or *Luculia* (e.g., *Uncaria* and *Mitragyna* both belong to the tribe Naucleeae of the Cinchonoideae [Bremer et al., 1995]).

Conclusions

It is better to analyze more characters than to investigate more taxa if the purpose is to get a strongly supported tree. The percentage of supported nodes within a tree is positively correlated with the number of characters and negatively correlated with the number of taxa. In our study, some kinds of data were better than others. In the analyses of Rubiaceae, RFLP data gave trees with higher percentage of supported nodes than did *rbc*L and morphological data. However, we found no significant difference between the *rbc*L and *ndh*F data. Within Rubiaceae it seems impossible to achieve a completely resolved subfamilial classification, i.e., to include all genera in one of the three large subfamilies. Several data sets now indicate that most genera of the Rubiaceae can be classified in either of the 3 large subfamilies (Rubioideae, Cinchonoideae s.s., and Ixoroideae s.l.) but a few genera (so far only *Luculia* and *Coptosapelta*) do not belong to any of the large subfamilies and are better left unclassified with respect to subfamily.

Acknowledgments

We thank all helpful colleagues and friends; for providing plant material or DNA: M. Chase, L. McDade, T. McDowell, R. Moberg, C. Puff, C. Ridsdale, E. Robbrecht, R. Sanders, and the directors of the botanical gardens and herbaria listed in Table 1; J. Rönnholm and N. Heidari for skillful technical help; Ove Eriksson for discussion of the statistics; and K. Andreasen, K. Bremer, two anonymous reviewers, and the editors for critical reading of the manuscript. This study was supported by Swedish Natural Science Research Council B-BU 1487-322 to B. B. and National Science Foundation DEB-9200 707 to R. K. J.

References

Albert, V. A., M. W. Chase, and B. D. Mishler. 1993. Character-state weighting for cladistic analysis of protein-coding DNA sequences. Ann. Mo. Bot. Gard. 80:752–766.

Allard, M. W., and J. M. Carpenter. 1996. On weighting and congruence. Cladistics 12:183–198.

Andersson, L. 1993. Pollen characteristics of the tribes Calycophylleae, Cinchoneae and Hillieae (Rubiaceae). Nord. J. Bot. 13:405–417.

Andersson, L., and C. Persson. 1991. Circumscription of the tribe Cinchoneae (Rubiaceae)—a cladistic approach. Plant Syst. Evol. 178:65–94.

Andreasen, K., B. G. Baldwin, and B. Bremer. 1999. Phylogenetic utility of the internal transcribed spacer of nuclear ribosomal DNA in the subfamily Ixoroideae (Rubiaceae). Plant Syst. Evol. 217:119–131.

Andreasen, K., and B. Bremer. 1996. Phylogeny of the subfamily Ixoroideae (Rubiaceae). Opera Bot. Belg. 7:119–138.

Annals of the Missouri Botanical Garden. 1995. Alternative genes for phylogenetic reconstruction in plants. Ann. Missouri Bot. Gard. 82(2).

Arnason, U., A. Gullberg, A. Janke, and X. Xu. 1996. Pattern and timing of evolutionary divergences among hominoids based on analyses of complete mtDNAs. J. Mol. Evol. 43:650–661.

Backlund, M., B. Oxelman, and B. Bremer. 1999. Phylogenetic relationships within the Gentianales based on *ndh*F and *rbc*L sequences, with particular reference to the Loganiaceae. Am. J. Bot. (in press).

Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. Syst. Zool. 40:486–493.

Barrett, M., M. J. Donoghue, and E. Sober. 1993. Crusade? A reply to Nelson. Syst. Biol. 42:216–217.

Bremekamp, C. E. B. 1966. Remarks on the position, the delimitation and the subdivision of the Rubiaceae. Acta Bot. Neerl. 15:1–33.

Bremer, B. 1996a. Combined and separate analyses of morphological and molecular data in the plant family Rubiaceae. Cladistics 12:21–40.

Bremer, B. 1996b. Phylogenetic studies within Rubiaceae and relationships to other families based on molecular data. Opera Bot. Belg. 7:33–50.

Bremer, B., K. Andreasen, and D. Olsson. 1995. Subfamilial and tribal relationships in the Rubiaceae based on *rbc*L sequence data. Ann. Mo. Bot. Gard. 82:383–397.

Bremer, B., and R. K. Jansen. 1991. Comparative restriction site mapping of the chloroplast DNA implies new phylogenetic relationships within the Rubiaceae. Am. J. Bot. 78:198–213.

Bremer, K., R. K. Jansen, P. O. Karis, M. Källersjö, and S. C. Keeley. 1992. A review of the phylogeny and classification of the Asteraceae. Nord. J. Bot. 12:141–148.

Bremer, B., and L. Struwe. 1992. Phylogeny of the Rubiaceae and the Loganiaceae: Congruence or conflict between morphological and molecular data? Am. J. Bot. 79:1171–1184.

Bremer, B., and M. Thulin. 1998. Collapse of Isertieae, re-establishment of Mussaendeae, and a new genus of Sabiceeae (Rubiaceae): Phylogenetic relationships based on *rbc*L data. Plant Syst. Evol. 211:71–92.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42:384–397.

Chase, M. W., and J. D. Palmer. 1989. Chloroplast DNA systematics of lilioid monocots: Resources, feasibility, and an example from the Orchidaceae. Am. J. Bot. 76:1720–1730.

Chase, M. W., D. E. Soltis, R. G. Olmstead, D. Morgan, D. H. Les, B. D. Mishler, M. R. Duvall, R. A. Price, H. G. Hills, Y.-L. Qiu, K. A. Kron, J. H. Rettig, E.

CONTI, J. D. PALMER, J. R. MANHART, K. J. SYTSMA, H. J. MICHAELS, W. J. KRESS, K. G. KAROL, W. D. CLARK, M. HEDRÉN, B. S. GAUT, R. K. JANSEN, K. J. KIM, C. F. WIMPEE, J. F. SMITH, G. R. FURNIER, S. H. STRAUSS, Q. XIANG, G. M. PLUNKETT, P. S. SOLTIS, S. SWENSEN, S. E. WILLIAMS, P. A. GADEK, C. J. QUINN, L. E. EGUIARTE, E. GOLDENBERG, G. H. J. LEARN, S. W. GRAHAM, S. C. H. BARRETT, S. DAYANANDAN, AND V. A. ALBERT. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann. Mo. Bot. Gard. 80:528–581.

CUMMINGS, M. P., S. P. OTTO, AND J. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Mol. Biol. Evol. 12:814–822.

DE JUSSIEU, A. L. 1789. Genera plantarum. Viduam Herissant, Paris.

DE QUEIROZ, A. 1993. For consensus (sometimes). Syst. Biol. 42:368–372.

DOEBLEY, J., M. DURBIN, E. GOLDBERG, M., M. T. CLEGG, AND D. P. MA. 1990. Evolutionary analysis of the large subunit of carboxylas (*rbc*L) nucleotide sequence among the grasses (Gramineae). Evolution 44:1097–1108.

DOWNIE, S. R., D. S. KATZ-DOWNIE, AND K. J. CHO. 1997. Relationships in the Caryophyllales as suggested by phylogenetic analyses of partial chloroplast DNA ORF2280 homologue sequences. Am. J. Bot. 84:253–273.

DOYLE, J. J., J. L. DOYLE, J. A. BALLENGER, E. E. DICKSON, T. KAJITA, AND H. OHASHI. 1997. A phylogeny of the chloroplast gene *rbc*L in the Leguminosae: Taxonomic correlations and insights into the evolution of nodulation. Am. J. Bot. 84:541–554.

EERNISSE, D. J., AND A. G. KLUGE. 1993. Taxonomic congruence versus total evidence, and amniote phylogeny inferred from fossils, molecules, and morphology. Mol. Biol. Evol. 10:1170–1195.

ENDRESS, M. E., B. SENNBLAD, S. NILSSON, L. CIVEYREL, M. W. CHASE, S. HUYSMANS, E. GRAFSTRÖM, AND B. BREMER. 1996. A phylogenetic analysis of Apocynaceae s. str. and some related taxa in Gentianales: A multidisciplinary approach. Opera Bot. Belg. 7:59–102.

FARRIS, J. S. 1969. A successive approximations approach to character weighting. Syst. Zool. 18:374–385.

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

FERNHOLM, B., K. BREMER, AND H. JÖRNVALL, eds. 1989. The hierarchy of life—molecules or morphology in phylogenetic analysis. Excerta Medica, Amsterdam.

GRAYBEAL, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem. Syst. Biol. 47:9–17.

GIVNISH, T. J., AND SYTSMA, K. J. 1997. Homoplasy in molecular vs. morphological data: The likelihood of correct phylogenetic inference. Pages 55–101 *in* Molecular evolution and adaptive radiation. (T. J. Givnish, and K. J. Sytsma, eds.) Cambridge University Press, Cambridge, England.

HILLIS, D. M. 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18:23–42.

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

HOLM, S. 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6:65–70.

HOLMGREN, P. K., N. H. HOLMGREN, AND L. C. BARNETT. 1990. Index herbariorum, 8th ed. New York Botanical Garden, New York.

HOOT, S. B., A. CULHAM, AND P. R. CRANE. 1995. The utility of *atp*B gene sequences in resolving phylogenetic relationships: Comparisons with *rbc*L and 18S ribosomal DNA sequences in the Lardizabalaceae. Ann. Mo. Bot. Gard. 82:194–207.

IGERSHEIM, A. 1993. The character states of the caribbean monotypic endemic *Strumpfia* (Rubiaceae). Nord. J. Bot. 13:545–559.

JANSEN, R. K., AND J. D. PALMER. 1987. A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc. Natl. Acad. Sci. USA 84:5818–5822.

JANSEN, R. K., J. L. WEE, AND D. MILLIE. 1998. Comparative utility of chloroplast restriction site and DNA sequence data for phylogenetic studies in plants. Pages 87–100. Molecular systematics of plants, 2nd ed. *In* D. Soltis, P. Soltis, and J. Doyle, (eds.), Chapman and Hall: New York.

JOHNSON, L. A., J. L. SCHULTZ, D. E. SOLTIS, AND P. S. SOLTIS. 1996. Monophyly and generic relationships of Polemoniaceae based on *mat*K sequences. Am. J. Bot. 83:1207–1224.

JOHNSON, L. A. AND D. E. SOLTIS. 1995. Phylogenetic interference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *mat*K sequences. Ann. Mo. Bot. Gard. 82:149–175.

KÄLLERSJÖ, M., J. S. FARRIS, M. W. CHASE, B. BREMER, M. F. FAY, C. J. HUMPHRIES, G. PETERSEN, O. SEBERG, AND K. BREMER. 1998. Simultaneous parsimony jackknife analysis of 2538 *rbc*L DNA sequences reveals support for major clades of green plants, land plants, seed plants, and flowering plants. Plant Syst. Evol. 213:259–287.

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

KIM, K. J., AND R. K. JANSEN. 1995. *ndh*F sequence evolution and the major clades in the sunflower family. Proc. Natl. Acad. Sci. USA 92:10379–10383.

KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among Epicrates (Boidae, Serpentes). Syst. Zool. 38:7–25.

KLUGE, A. G., AND A. J. WOLF. 1993. Cladistics: What's in a word? Cladistics 9:183–199.

MANEN, J.-F., AND A. NATALI. 1996. The chloroplast *atpB-rbc*L spacer in Rubiaceae. Opera Bot. Belg. 7:51–57.

MIYAMOTO, M. M. 1983. Frogs of the *Eleutherodactylus rugulosus* group: A cladistic study of allozyme, morphological, and karyological data. Syst. Zool. 32:109–124.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1:186–189.

MIYAMOTO, M. M., AND W. M. FITCH. 1995. Testing species phylogeny and phylogenetic methods with congruence. Syst. Biol. 44:64–76.

MORGAN, D. R., AND D. E. SOLTIS. 1993. Phylogenetic relationships among members of Saxifragaceae sensu lato based on *rbc*L sequence data. Ann. Mo. Bot. Gard. 80:631–660.

NADOT, S., G. BITTAR, L. CARTER, R. LACROIX, AND B. LEJEUNE. 1995. A phylogenetic analysis of Monocotyledons based on the chloroplast gene *rps*4, using parsimony and a new numerical phenetic method. Mol. Phylogenet. Evol. 4:257–282.

NATALI, A., J.-F. MANEN, AND F. EHRENDORFER. 1996. Tribal, generic and specific relationships in the Rubioideae–Rubieae (Rubiaceae) based on sequence data of the cpDNA intergene spacer. Opera Bot. Belg. 7:193–203.

NELSON, G. 1993. Why crusade against consensus? A reply to Barrett, Donoghue, and Sober. Syst. Biol. 42:215–216.

NICKRENT, D. L., AND D. E. SOLTIS. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbc*L sequences. Ann. Mo. Bot. Gard. 82:208–234.

OLMSTEAD, R. G., B. BREMER, K. M. SCOTT, AND J. D. PALMER. 1993. A parsimony analysis of the Asteridae sensu lato based on *rbc*L sequences. Ann. Mo. Bot. Gard. 80:700–722.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: A review of methods and data analysis. Am. J. Bot. 81:1205–1224.

OLMSTEAD, R. G., AND P. A. REEVES. 1995. Evidence for the polyphyly of the Scrophulariaceae based on chloroplast *rbc*L and *ndh*F sequences. Ann. Mo. Bot. Gard. 82:176–193.

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

PAGE, R. D. M. 1996. On consensus, confidence, and "total evidence." Cladistics 12:83–92.

PATTERSON, C. 1987. Molecules and morphology in evolution: Conflict or compromise. Cambridge University Press, Cambridge.

POE, S., AND SWOFFORD, D. L. 1999. Taxon sampling revisited. Nature 398:299–300.

RICE, E. R. 1989. Analyzing tables of statistical tests. Evolution 43:223–225.

ROBBRECHT, E. 1988. Tropical woody Rubiaceae. Opera Bot. Belg. 1:1–271.

ROBBRECHT, E. 1993. Supplement to the 1988 outline of the classification of the Rubiaceae Index to genera. Opera Bot. Belg. 6:173–196.

ROBBRECHT, E., AND C. PUFF. 1986. A survey of the Gardenieae and related tribes (Rubiaceae). Bot. Jahrb. Syst., Pflanzengesch. Pflanzengeog. 108:63–137.

SANDERSON, M. J. 1989. Confidence limits on phylogenies: The bootstrap revisited. Cladistics 5:113–129.

SANDERSON, M. J., AND M. J. DONOGHUE. 1989. Patterns of variation in level of homoplasy. Evolution 43:1781–1795.

SCHUMANN, K., ed. 891. Rubiaceae. Die natürlichen Pflanzenfamilien 4 (4). Wilhelm Engelmann, Leipzig, Germany.

SENNBLAD, B., AND B. BREMER. 1996. The familial and subfamilial relationships of Apocynaceae and Asclepiadaceae evaluated with *rbc*L data. Plant Syst. Evol. 202:153–175.

SMITH, J. F. 1996. Tribal relationships within Gesneriaceae: A cladistic analysis of morphological data. Syst. Bot. 21:497–513.

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

SOKAL, R. R., AND K. SHAO. 1985. Characters stability in 39 data sets. Syst. Zool. 34:83–89.

SOLTIS, D. E., P. S. SOLTIS, M. T. CLEGG, AND M. DURBIN. 1990. *rbc*L sequence divergence and phylogenetic relationships in Saxifragaceae sensu lato. Proc. Natl. Acad. Sci. USA 87:4640–4644.

SOLTIS, D. E., Q.-Y. XIANG, AND L. HUFFORD. 1995. Relationships and evolution of Hydrangeaceae based on *rbc*L sequence data. Am. J. Bot. 82:504–514.

SOUZA-CHIES, T. T., G. BITTAR, S. NADOT, L. CARTER, E. BESIN, AND B. LEJEUNE. 1997. Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene *rps*4. Plant Syst. Evol. 204:109–123.

STEVENS, P. 1997. How to interpret botanical classifications—suggestions from history. BioScience 47:243–250.

SWOFFORD, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony. Version 3.1. 1. Computer program. Illinois Natural History Survey, Champaign.

SYTSMA, K. J. 1990. DNA and morphology: Inference of plant phylogeny. Trends Ecol. Evol. 5:104–110.

SYTSMA, K. J., AND B. A. SCHAAL. 1985. Phylogenetics of the *Lisianthus skinneri* (Gentianaceae) species complex in Panama using DNA restriction fragment analysis. Evolution 37:594–608.

VERDCOURT, B. 1958. Remarks on the classification of the Rubiaceae. Bull. Jard. Bot. l'État 28:209–281.

WAGSTAFF, S. J., AND R. G. OLMSTEAD. 1997. Phylogeny of Labiateae and Verbenaceae inferred from *rbc*L sequences. Syst. Bot. 22:165–177.

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

Received 20 April 1998; accepted: 8 February 1999
Associate Editor: D. Hibbett