

## Elusive Relationships Within Order Fabales: Phylogenetic Analyses Using *matK* and *rbcL* Sequence Data<sup>1</sup>

M. A. Bello,<sup>2,3,5</sup> A. Bruneau,<sup>4</sup> F. Forest,<sup>3</sup> and J. A. Hawkins<sup>2</sup>

<sup>2</sup> School of Biological Sciences, Lyle Building, The University of Reading, Reading, Berkshire U.K. RG1 6AS

<sup>3</sup> Jodrell Laboratory, The Royal Botanic Gardens Kew, Richmond, Surrey U.K. TW9 3DS

<sup>4</sup> Institut de recherche en biologie végétale, Université de Montréal, 4101 Sherbrooke est, Montréal, Québec, Canada H1X 2B2

<sup>5</sup> Author by correspondence (mangelicabegu@lycos.com)

Communicating Editor: Matt Lavin

**Abstract**—The order Fabales, including Leguminosae, Polygalaceae, Quillajaceae and Surianaceae, represents a novel hypothesis emerging from angiosperm molecular phylogenies. Despite good support for the order, molecular studies to date have suggested contradictory, poorly supported interfamilial relationships. Our reappraisal of relationships within Fabales addresses past taxon sampling deficiencies, and employs parsimony and Bayesian approaches using sequences from the plastid regions *rbcL* (166 spp.) and *matK* (78 spp.). Five alternative hypotheses for interfamilial relationships within Fabales were recovered. The Shimodaira-Hasegawa test found the likelihood of a resolved topology significantly higher than the one calculated for a polytomy, but did not favour any of the alternative hypotheses of relationship within Fabales. In the light of the morphological evidence available and the comparative behavior of *rbcL* and *matK*, the topology recovering Polygalaceae as sister to the rest of the order Fabales with Leguminosae more closely related to Quillajaceae + Surianaceae, is considered the most likely hypothesis of interfamilial relationships of the order. Dating of selected crown clades in the Fabales phylogeny using penalized likelihood suggests rapid radiation of the Leguminosae, Polygalaceae, and (Quillajaceae + Surianaceae) crown clades.

**Keywords**—Leguminosae, macroevolution, Polygalaceae, Quillajaceae, radiation, Surianaceae.

The order Fabales Bromhead, including Leguminosae (= Fabaceae Lindl.), Polygalaceae Hoffmanns. & Link, Quillajaceae D. Don, and Surianaceae Arn., is one of the novel hypotheses of relationships that emerged from angiosperm phylogenies based on DNA sequence data (e.g. Chase et al. 1993; APG 2003). Fabales is part of eurosids I and together with Fagales, Cucurbitales, and Rosales forms the “nitrogen-fixing clade” (Soltis et al. 1995). Polygalaceae and Leguminosae are widely distributed and comprise about 1,000 and 19,325 species respectively (Persson 2001; Lewis et al. 2005a). In contrast, Quillajaceae comprises just two species distributed in Chile and Brazil (Fuks 1983; Lersten and Horner 2005), and Surianaceae, with eight species, includes taxa restricted to Australia and Mexico together with the pantropical *Suriana maritima* L. (Crayn et al. 1995).

Traditional systematic treatments suggested different affinities for the families now included in Fabales, but never linked all of them together (Table 1). The morphological evidence supporting a sister relationship between Leguminosae, Polygalaceae, Quillajaceae, and Surianaceae is scarce, as is the case for many subclades within rosids derived from sequence analyses (Soltis and Soltis 2004; Soltis et al. 2005). Although the presence of vessel elements with single perforations, elagic acid, and a large green embryo have been proposed as potential morphological synapomorphies for the order (Judd et al. 2002), few comparative analyses of morphological characters have been carried out explicitly for Fabales and their closest relatives.

To date, molecular phylogenetic studies including families now referred to the Fabales have focussed either above the level of the order, or on one of the families. Neither the higher-level nor the family-level studies have explicitly optimised sampling to address the question of intrafamilial relationships within the Fabales. Higher-level studies using few exemplars of each family, whether based on single (Chase et al. 1993; Fernando et al. 1993; Crayn et al. 1995; Soltis et al. 1997; Doyle et al. 2000; Savolainen et al. 2000b; Kajita et al. 2001; Wojciechowski et al. 2004) or combined DNA regions (Savolainen et al. 2000a; Soltis et al. 2000), suggested contradictory and poorly-supported sister group relationships within the order. Similarly, in analyses

focussed on one family with few exemplars of related lineages, the interfamilial relationships are not recovered with confidence. For instance, Lavin et al.’s (2005) extensive study of Leguminosae included one member of each of the other families in Fabales, and the phylogeny of Polygalaceae (Persson 2001) sampled one representative of Quillajaceae, three species of Leguminosae, and two of Surianaceae.

Intrafamilial phylogenetic relationships within Polygalaceae, Surianaceae, and Leguminosae have been the subject of several studies. The phylogeny of Polygalaceae has been explored using morphological (Eriksen 1993a) and molecular (Persson 2001; Forest et al. 2007) data. These analyses tested the traditional tribal relationships (Chodat 1896), already challenged by previous studies of ovule, seed anatomy, and floral vascular anatomy (Verkerke 1985; Eriksen 1993b). Chodat’s main divisions of the family [i.e. tribes Polygalaeae Fr., Moutabeae Chodat, and Xanthophylleae Baill., plus Carpolobieae recently proposed by Eriksen (1993a)] are monophyletic according to these analyses. Studies in Surianaceae using *rbcL* data (Fernando et al. 1993; Crayn et al. 1995) revealed relatively well-supported phylogenetic relationships within the family: [((*Recchia*, *Cadellia*) (*Suriana* (*Stylobasium*, *Guilfoylia*)))]. Surianaceae was previously characterised by the pericarp anatomy (which differs in *Guilfoylia* F. Muell.), the apocarpous and gynobasic gynoecium, and the basal-marginal placentation (Fernando and Quinn 1992; Fernando et al. 1993), features that allowed one to predict the placement of the genus *Recchia* Moc. & Sessé within this family before its inclusion in the *rbcL* analysis of Crayn et al. (1995). The legumes are among the most widely explored angiosperm families in phylogenetic terms since numerous reconstructions have been made in recent years using morphological, chemical, and molecular data (e.g. Tucker and Douglas 1994; Chappill 1995; Doyle 1995; Käss and Wink 1996; Doyle et al. 1997, 2000; Bruneau et al. 2000, 2001; Kajita et al. 2001; Herendeen et al. 2003; Wojciechowski 2003, Wojciechowski et al. 2004; Lewis et al. 2005a). Although topological conflicts are observed amongst these reconstructions, most studies recovered a monophyletic status for Leguminosae and subfamily Papilionoideae, and the paraphyletic condition

TABLE 1. Summary of previous relationships proposed for Leguminosae, Polygalaceae, Quillajaceae and Surianaceae. Current Fabales members (in bold) are cited with their most closely related taxa (in italics) in the different systems when specified.

System/ Family	Dahlgren(1980)	Heywood (1993)	Cronquist (1981; 1988)	Thorne (2000)	Takhtajan (1997)
Polygalaceae	Rutiflorae Polygalales <i>Krameriaceae</i> <b>Polygalaceae</b> <i>Trigoniaceae</i> <i>Vochysiaceae</i>	Rosidae Polygalales <i>Krameriaceae</i> <i>Malpighiaceae</i> <b>Polygalaceae</b> <i>Tremandraceae</i> <i>Trigoniaceae</i> <i>Vochysiaceae</i>	Rosidae Polygalales <b>Polygalaceae</b> <i>Tremandraceae</i> <i>Trigoniaceae</i> <i>Vochysiaceae</i> <i>Xanthophyllaceae</i> <i>Emblingiaceae</i>	Geraniae Polygalales <i>Euphroniaceae</i> <i>Krameriaceae</i> <i>Malpighiaceae</i> <b>Polygalaceae</b> <i>Trigoniaceae</i>	Rosidae Geraniae Polygalales <i>Emblingiaceae</i> <b>Polygalaceae</b> <i>Xanthophyllaceae</i>
Leguminosae	Fabiflorae Fabales <b>Leguminosae</b>	Fabales <b>Leguminosae</b>	Fabales <i>Connaraceae</i> <b>Leguminosae</b>	Rutanae Rutales Fabineae <i>Connaraceae</i> <b>Leguminosae</b> <b>Quillajaceae</b> <b>Surianaceae</b>	Rosidae Fabanae Fabales <b>Leguminosae</b>
Surianaceae	Rutales <i>Simaroubaceae</i> <b>Surianaceae</b> Sapindales <b>Stylobasium</b> <i>Connaraceae</i> <i>Emblingiaceae</i>	Sapindales <i>Anacardiaceae</i> <b>Stylobasium</b>	Sapindales <i>Simaroubaceae</i> <i>Recchia</i> Rosales <b>Surianaceae</b>	Rutanae Rutales Fabineae <i>Connaraceae</i> <b>Leguminosae</b> <b>Quillajaceae</b> <b>Surianaceae</b>	Rutanae Rutales <i>Simaroubaceae</i> <b>Surianaceae</b>
Quillajaceae		Rosales Rosaceae Maloideae <i>Malus</i> Mill. <i>Pyrus</i> L. <b>Quillaja</b>	Rosaceae Spiraeoideae <b>Quillaja</b> <i>Kageneckia</i>	Rutanae Rutales Fabineae <i>Connaraceae</i> <b>Leguminosae</b> <b>Quillajaceae</b> <b>Surianaceae</b>	Rosanae Rosales Rosaceae Quillajoideae Rosaeae <b>Quillaja</b>
Not specified	<i>Recchia</i> <b>Quillaja</b>	<i>Cadellia</i> <i>Guilfoylia</i> <i>Recchia</i> <i>Suriana</i>	—	—	—

of subfamily Caesalpinioideae. Recently, it has been suggested that subfamily Mimosoideae is nonmonophyletic (e.g. Luckow et al. 2000; Wojciechowski et al. 2004). So far, the species *Quillaja saponaria* Molina is the only representative of the two species of family Quillajaceae integrated in any phylogenetic analyses (Crayn et al. 1995; Doyle et al. 2000; Savolainen et al. 2000b; Kajita et al. 2001; this study).

In the light of previous studies, and given the current interest in the phylogenetic relationships of legumes and allies, the aim of this study is to reevaluate the interfamilial relationships within Fabales. To this end, sequences from the plastid regions *rbcl* and *matK* gene were selected because the sampling availability from previous analyses and their potential for solving deep hierarchical relationships (e.g. Johnson and Soltis 1995; Hilu and Liang 1997; Miller and Bayer 2001; Fan and Xiang 2003; Hilu et al. 2003; Soltis et al. 2003; Wojciechowski et al. 2004). In comparison to previous studies, additional taxa from Caesalpinioideae and Polygalaceae were included together with almost all species of Surianaceae and *Quillaja saponaria*.

#### MATERIALS AND METHODS

**Taxon Sampling**—This study includes a total of 152 species of Fabales, with representatives of Leguminosae (52), Polygalaceae (93), Quillajaceae (1), and Surianaceae (6). Members of the diverse infrafamilial groups currently recognised within Polygalaceae (Paiva 1998; Bernardi 2000) were included. For Leguminosae, four representatives of Papilionoideae, eight of Mimosoideae and several caesalpinoid members of Detarieae

(11), Cercidae (9), Caesalpinieae (11), and Cassieae (9) clades were preferentially sampled (Appendix 1) due their position within early diverging clades in the legume phylogeny (Bruneau et al. 2001; Wojciechowski et al. 2004). All Surianaceae were included except two of the three species of *Recchia*. As in other studies, Quillajaceae is represented only by *Q. saponaria*. Selected taxa representing closely related families according to recent angiosperm phylogenetic studies (APG 2003; Hilu et al. 2003) were included as outgroup taxa (*rbcl* 14 spp.; *matK* 8 spp). The summary of the taxon sampling, taxonomic ranges covered, voucher information, and sequence Genbank/EMBL accession numbers is provided in Appendix 1.

**DNA Extraction, PCR Amplification, and Sequencing**—Genomic DNA from silica-dried leaves was isolated using the CTAB method (Doyle and Doyle 1987) with modifications following Harris (1995), or using the DNeasy Plant Mini kit (Qiagen Ltd., Crawley, West Sussex U.K.). For the Polymerase Chain Reaction (PCR), primers designed to amplify the *matK* gene in legumes were used (Wojciechowski et al. 2004). Of these, the primers *matK* 1932Ra and *matK* 832R did not function satisfactorily for most of the Polygalaceae sampled. For those taxa additional primers were designed (Table 2). Universal primers for angiosperms were used to amplify *rbcl* (N, 724/SS722, and R; Käss and Wink 1996). PCR amplification was performed using 5 µl reaction buffer [10×], 1.5 µl MgCl<sub>2</sub> [50 mM], 2 µl dNTP [5 mM], 1.75 µl each primer [10 mM], 0.75 µl *Taq* Polymerase [5u µl<sup>-1</sup>] (Biotaq, Bioline Ltd, London), 1 µl of genomic DNA and 36.25 µl of nanopure water per reaction. The PCR reactions were performed on a 2700 Gene Amp thermocycler from Applied Biosystems, following the conditions of Johnson and Soltis (1995) for *matK*. For *rbcl*, the program settings were 2 min at 94°C, 35 cycles of 50 sec at 94°C, annealing at 49°C for 50 sec, 75 sec at 72°C, and a final extension of 7 min at 72°C. PCR products were cleaned using the QIAquick PCR Purification Kit (Qiagen Ltd.). The sequences were generated on an ABI prism 3100 capillary sequencer (Applied Biosystems, Foster City, California).

**Sequence Alignment and Phylogenetic Analysis**—Complementary strands were assembled and edited using Lasergene software (DNASar

TABLE 2. Primers designed for PCR amplification and sequencing of *matK* in Polygalaceae.

Primer designation	Sequence (5' - 3')
66Forward	ATATTTTCCGAAGTACTAGATAAAT
405Reverse	AATCGTAGGGAAAATGGAAT
443Forward	GAAGGGTCGGAAAATTTTAA
786Reverse	AATCTTGGATTAGTTTTTGA
1190Forward	GTGTGCGGTTAAATCCTTCG
1535Reverse	CGCGTACAGCACTTTTGTGT
1651Forward	TAGGGTTCGGGTTTGGTATT
2000Reverse	CGTGCTTGCATTTTTCATTG

Inc, Madison, Wisconsin). An initial alignment was performed with Clustal W, as implemented by Megalign 4.05 (DNASTar, Inc.) with the default settings and visually edited in MacClade 4.0 (Maddison and Maddison 2002). Four data sets were assembled (TreeBASE study number 2226): *rbcL*, *matK*, and two combined *rbcL* + *matK* sets. Of the combined data sets, one comprised only taxa with both *rbcL* and *matK* sequences available (the "reduced taxon combined" data set), and the second included all taxa, even those with only *rbcL* or *matK* sequences (the "all taxon combined" data set). Sequences of selected outgroups were included (*rbcL*, 14 spp.; *matK*, 8 spp).

Phylogenetic analyses implementing maximum parsimony were performed using Winclada (Nixon 2002) running NONA (Goloboff 1993) as a complementary process. The searches were unconstrained and the strategy multiple TBR+TBR was selected. These heuristic searches were performed excluding uninformative characters. For the smaller matrices (*matK* and the "reduced taxon combined" data sets) the analyses consisted of 10,000 replications, keeping a maximum of 300,000 trees and holding 30 trees per replicate. For *matK*, an additional matrix included gaps coded as binary characters (following the "simple indel coding" method from Simmons and Ochoterena 2000) was included. In the case of the larger *rbcL* and "all taxon combined" data sets, the analyses were limited to 1,000 replications, keeping 10,000 trees and 10 trees per replicate. After all analyses, the nelsen strict consensus trees were calculated by NONA in Winclada. Support values were assessed using bootstrap (Felsenstein 1985) and jackknife (Farris et al. 1996) methods on all data sets. These analyses were implemented in Winclada using 500 replicates, 100 search replicates, and one starting tree per replicate in addition to the TBR branch swapping algorithm.

Using Winclada and NONA, the incongruence between *matK* and *rbcL* data sets was explored using the incongruence length difference (ILD) test (Farris et al. 1994) implementing 1,000 replicates, holding 10 trees per replicate, and leaving the operational default. Although the test has been empirically challenged (e.g. Yoder et al. 2001), it has been used to inspect the interaction among data partitions (e.g. Fishbein et al. 2001) and in some cases could represent a starting point for the discussion of combinability.

All data sets were also analysed using Bayesian inference (MrBayes v3.0b4, Huelsenbeck and Ronquist 2001). MrModeltest version 2 (Nylander 2004) was used to choose the model of DNA evolution that best fits the individual partitions. Based on the Akaike Information Criterion (AIC), the general time reversible model with gamma and proportion of invariable sites was selected (GTR + I + G) for both the *rbcL* (-lnL = 14518.45) and *matK* (-lnL = 17946.66) partitions.

Using PAUP\* version 4.0b10 (Swofford 2002), the GTR + SS model was tested against the GTR + I + G model. A *matK* data set of 1,500 base pairs (excluding the *trnK* introns), and a *rbcL* matrix of 1,401 bp were implemented for this comparison. For *matK* the model GTR + I + G (-lnL = 17946.66383) presented a significantly better likelihood in comparison with the GTR + SS model (-lnL = 18290.60679). Similarly for *rbcL* the model GTR + I + G (-lnL = 14518.45011) was significantly better than the GTR+SS model (-lnL = 15748.03458). The level of significance was addressed with the AIC for both markers.

The Bayesian analyses were run for  $5 \times 10^6$  generations with four chains, sampling every 100 generations. Likelihood scores were plotted against generation numbers to determine when the analyses reached stationarity. Trees saved prior to stationarity were excluded ("burn in") and the remaining trees were used to determine the posterior probabilities by constructing a majority rule consensus tree in PAUP 4.0b.

**Topology Comparisons Using the Shimodaira-Hasegawa Test (SH)**—The SH test was implemented in order to compare the likelihood of contrasting topologies. This test is frequently used to explore confidence limits of contrasting phylogenetic hypothesis (e.g. Fishbein and Soltis 2004; Beilstein et al. 2006). Nine different comparisons were performed.

First, resolved vs collapsed topologies for the interfamilial relationships were compared to determine whether resolved trees are more likely than a hard polytomy in the case of Fabales. The topologies compared were derived from the Bayesian analyses of the *matK*, *rbcL*, and reduced combined data sets. In each of these analyses, a post burn-in tree was randomly selected to represent the resolved hypothesis and the unresolved tree was created in MacClade 4.0 by collapsing the interfamilial branches of the selected tree. The rest of the comparisons were between alternative intrafamilial resolutions derived from parsimony and Bayesian inference. The topologies derived from the *rbcL* parsimony analyses matrix were not compared because of their lack of resolution. The parsimony trees randomly selected for each comparison present the same topology as the consensus trees. Similarly, the Bayesian trees used were randomly selected from the post burn-in population of trees with the same topology as the majority rule consensus tree.

**Dating of the Main Lineages Within Fabales**—The short time frame between the divergence of the major lineages of Leguminosae has been used as a criterion to suggest radiation (Lavin et al. 2005). In order to test whether the crown clades of the other families of Fabales radiated within a similarly short time frame, ages for these clades were inferred using penalised likelihood implemented in r8s version 1.7 (Sanderson 2003). The program r8s can infer absolute rates of molecular evolution and divergence times allowing multiple calibration points and assignment of maximum and minimum ages (Sanderson 2002). The calibration points used here were a maximum age of 84 Ma for the Fabales crown node (estimated age for the Fagales and Cucurbitales divergence, Wikström et al. 2001) and 59 Ma for the Leguminosae crown node (estimated age for Leguminosae in Lavin et al. 2005). The 84 Ma estimate for Fabales does not disagree with additional fossil or molecular clock age estimates (Friis and Crane 1989; Crepet et al. 2004). The trees used for dating were subsets of the *matK* post burn-in trees, which include representatives of eurosids I (Begoniaceae, Fagaceae, Rosaceae, Violaceae, Zygophyllaceae) as outgroups as well as *Oxalis stricta*, the first specified outgroup which was pruned prior to r8s analyses. One hundred trees were randomly sampled to estimate the ages of the crown nodes of Polygalaceae, Surianaceae and Quillajaceae + Surianaceae.

## RESULTS

**Data Sets and Parsimony Analysis**—The *rbcL* matrix included 166 sequences, 1,405 characters, and 1.24% missing data (Data set 1). From this data set 911 characters were constant, 132 parsimony-uninformative, and 362 informative. The aligned *matK* data set included 78 sequences, 1,753 characters, and 8.4% of the characters coded as missing data (Data set 2). It comprised 650 constant and 338 variable parsimony uninformative characters, together with 765 informative characters. The combined *matK* + *rbcL* reduced and all taxon sets included 3,158 positions with 75 and 169 taxa respectively (Data sets 3 and 4). In the "reduced taxon combined" set 1,722 sites were constant, 430 parsimony-uninformative and 1,006 informative, and in the "all taxon combined" set 1,568 characters were invariant, 466 parsimony-uninformative, and 1,124 informative. Table 3 summarises the parsimony tree scores obtained for each data set. The ILD test showed significant incongruence between *rbcL* (length = 1,269) and *matK* (length = 2,987) partitions, with a tree length for the combined matrix of 4,290 steps ( $\alpha = 0.05$ ,  $p = 0.001$ ).

In all analyses Fabales is monophyletic with moderate to high bootstrap and jackknife support values (Table 4; Fig. 1). Similarly, Polygalaceae and Surianaceae always appear as monophyletic with high support. With the exception of the topology based on *rbcL*, the support for Leguminosae as monophyletic is moderate to high (Table 4). Congruent interfamilial relationships were recovered in the *matK* and the "all taxon combined" analyses: Quillajaceae and Surianaceae form a clade sister to Leguminosae with Polygalaceae sister to the rest of the order. The relationship (Leguminosae (Quillajaceae + Surianaceae)) is poorly supported in the



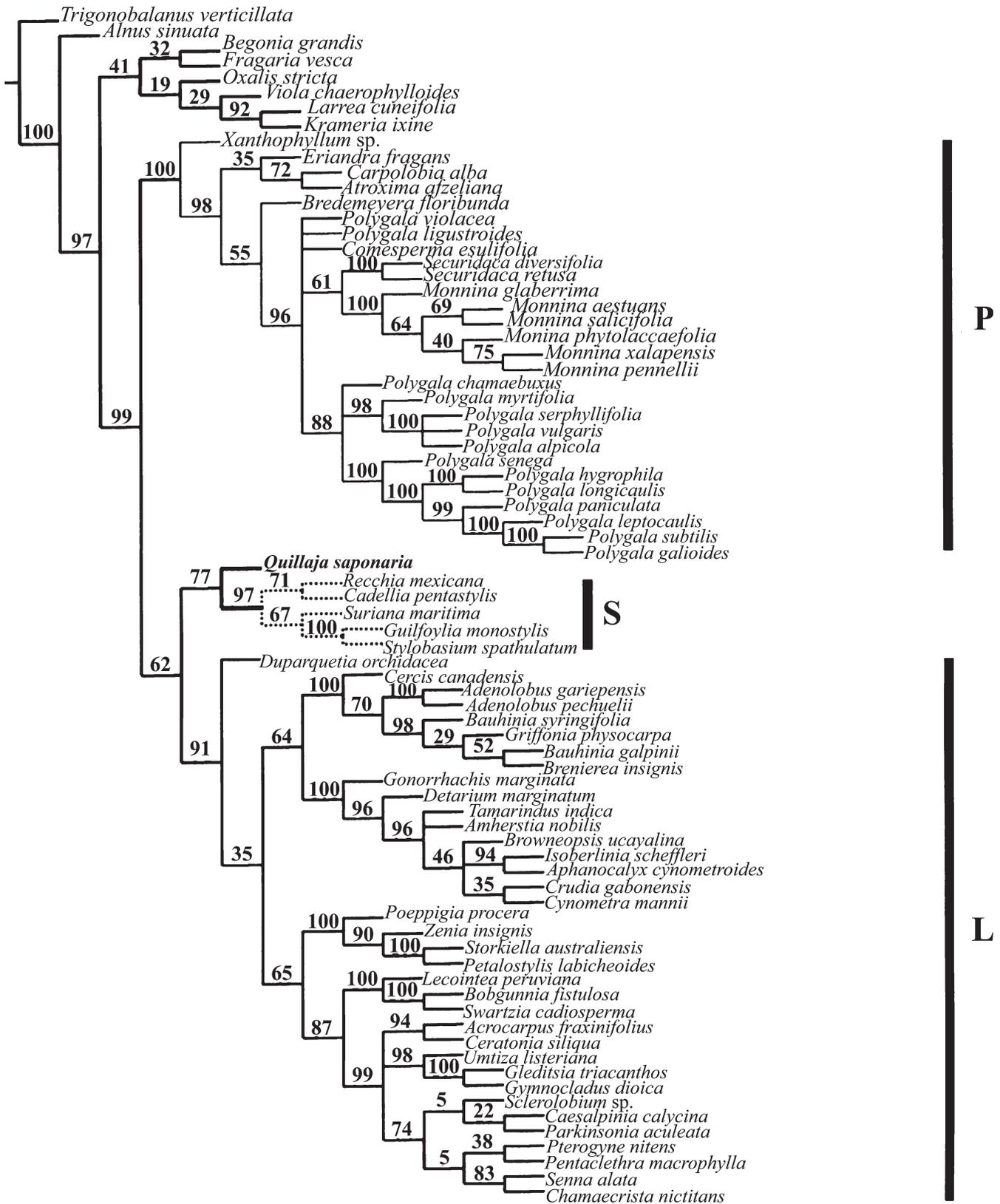


FIG. 1. Nelsen strict consensus tree derived from the *matK* data set parsimony analysis. Values indicate jackknife support. L = Leguminosae, P = Polygalaceae, S = Surianaceae.

to recover consistent and supported interfamilial relationships is found in other orders within the rosids. Some authors have argued that rosids represent the largest remaining problematic group of angiosperms (Soltis and Soltis 2004) together

with other "taxa difficult to relate" such as the closest relatives of monocots and the core eudicots. The pattern of weakly supported or conflicting familial internal branches is a challenge for phylogenetic reconstruction, but also a phenomenon in

TABLE 5. Results from the topology likelihood comparison with the SH test ( $\alpha = 0.05$ ). In cases of significant difference,  $p$  values are underlined and the selected topology is highlighted in bold.

Topologies compared	-lnL	Difference	P values
<b><i>matK</i> Bayesian vs polytomy</b>	21063.259 21078.159	14.900	<u>0.024</u>
<i>rbcl</i> Bayesian vs polytomy	16504.477 16521.734	17.256	0.282
<b>"Reduced taxon combined"</b>	27789.453	14.990	<u>0.029</u>
<b>Bayesian vs polytomy</b>	27804.443		
<i>matK</i> Bayesian vs	21050.360	9.918	0.256
<i>matK</i> parsimony	21060.278		
<b>"Reduced taxon combined"</b>	27789.453	204.793	<u>0.000</u>
<b>Bayesian vs <i>matK</i> parsimony</b>	27994.247		
"Reduced taxon combined"	29670.526	36.166	0.168
parsimony vs <i>matK</i> parsimony	29706.693		
<b><i>matK</i> Bayesian vs "reduced taxon combined" parsimony</b>	21050.360 21125.665	75.305	<u>0.007</u>
<b><i>matK</i> Bayesian vs "reduced taxon combined" Bayesian</b>	21050.360 21447.139	396.779	<u>0.000</u>
<i>rbcl</i> Bayesian vs "all taxa combined" parsimony	16504.477 16553.889	49.412	0.099

need of explanation. It is argued that different processes (i.e. macroevolution vs. microevolution, Kellogg 2002) play an important role in determining the pattern of deep and shallow nodes in angiosperms. Other authors have interpreted lack of resolution among familial clades in several orders of flowering plants as a sign of relatively rapid diversification or radiation (Judd and Olmstead 2004), where episodic lineage radiation produces short branches and persistence of those lineages explains the long branches subtending well supported families (Soltis et al. 2005). Examples where long branches support individual lineages and short branches "weakly" connect them are frequent in the literature [e.g. Cornales Dumort. (Xiang et al. 1998), Cucurbitales Dumort. (Swensen et al. 1998), Ericales Dumort. (Anderberg et al. 2002), Geraniales Dumort. (Savolainen et al. 2000b; Soltis et al. 2000), Lamiales Bromhead (Wortley et al. 2005), Malpighiales Mart. (Soltis et al. 2000), and Saxifragales Dumort. (Fishbein et al. 2001; Fishbein and Soltis 2004)].

The origin of the order need not be interpreted as a hard polytomy. In this study the SH test found a resolved topology to be significantly more likely than a polytomy, (but see discussion for *rbcl* below), suggesting a nonsimultaneous origin of families (soft polytomy) in the Fabales. At deep levels, factors such as extinction, long-branch attraction, and difficult homology assessment (sequence alignment) can impede phylogenetic reconstruction (for further discussion see Mishler 2000). In addition to these potential problems in deep node reconstruction and those generally invoked to explain polytomies [e.g. rate variation across characters, taxon or character compositional bias (see Wortley et al. 2005 for further examples)], the slowly evolving DNA regions usually selected

TABLE 6. Estimated ages with penalized likelihood of selected crown clades within Fabales. Calibration points were implemented for the crown nodes of Fabales (84 Ma) and Leguminosae (59 Ma). The values indicate ages in million years ago (Ma).

Crown clades	Mean	Standard deviation	Minimum	Maximum
Polygalaceae	57.41	8.11	31.77	77.67
Surianaceae	38.69	11.75	19.68	71.85
Quillajaceae + Surianaceae	54.53	11.77	33.04	82.30

for phylogenetic analyses can present selective constraints, increasing the risk of convergence (Mishler 2000). For this particular study, in order to avoid long-branch attraction effects, a Bayesian analysis was performed assuming that likelihood-based approaches are not as vulnerable as parsimony analyses to the long-branch attraction effect (but see Felsenstein 1978; Schulmeister 2004; Sober 2004). The Bayesian analyses recovered monophyletic families but interfamilial relationships remain poorly to moderately supported (Table 4).

**Performance of *matK* and *rbcl***—The plastid gene regions selected for this study have both been frequently used at deep levels, but have presented some difficulties such as high substitution rate of the third codon position of *rbcl*, which could obscure its phylogenetic information (Johnson and Soltis 1994; Steele and Vilgalys 1994; Hilu et al. 2003; Lavin et al. 2005). On the other hand, although *matK*'s success in resolving deep-node relationships has been challenged in phylogenetic analyses of Cornales (Xiang et al. 1998), this plastid region has been useful for several phylogenetic analyses (Johnson and Soltis 1994; Hilu and Liang 1997; Hilu et al. 2003; Wojciechowski et al. 2004). Comparison of *matK* and *rbcl* in phylogenetic reconstructions of Leguminosae shows that *matK* exhibits a higher level of resolution at many hierarchical levels within the family, due perhaps to its faster substitution rates that are more uniform among codon positions (Lavin et al. 2005). Our parsimony analyses using *matK* data recover phylogenies with higher support values for the interfamilial relationships than the *rbcl* analyses, and the *matK* topology is recovered when analysing the "all taxon combined" data set (Table 4). The *matK* and *rbcl* topologies exhibit the same retention index (0.77 in all trees obtained), but *matK* shows less homoplasy (estimated with consistency index) than *rbcl* and the "all taxon combined" data (Table 3).

Additionally, other aspects of the behaviour of *rbcl* in this study weaken our confidence in the *rbcl* topologies. For example, the *rbcl* topologies derived from parsimony and Bayesian analyses are contradictory and were not recovered from any other data set; whereas the Bayesian topology has a moderate support (= 0.93) for the clade (Leguminosae(Quillajaceae (Polygalaceae + Surianaceae))), the parsimony tree is unresolved (Table 4). Contrasting the *rbcl* Bayesian topology with a tree derived from a parsimony analysis of the "all taxon combined" data set, the SH test does not reject this second topology despite its lower likelihood value (Table 5). It has been reported that Bayesian analyses can provide highly supported arbitrary resolutions in cases of hard or nearly hard polytomies (Lewis et al. 2005b), which suggests a need to be cautious regarding the *rbcl*-based Bayesian topology, at least with respect to the interfamilial relationships. Interestingly, and in agreement with this suggestion, the SH test did not discriminate between the *rbcl* Bayesian topology and a hard polytomy scenario (Table 5).

Aside from the results of the SH test, the performance of the *matK* and *rbcl* data sets during the analyses was different; the Bayesian trees from the *rbcl* data set required more generations to reach stationarity in comparison to *matK* (4,000,000 vs 2,633,200 generations). A similar situation was reported by Lavin et al. (2005) in their analyses of *rbcl* and *matK* sequences from Leguminosae, which again could be attributed to *rbcl*'s lack of informativeness. The parsimony analyses also progressed differently for *matK* and *rbcl*, with the *matK* partition recovering only six trees whereas with *rbcl* the memory settings were always surpassed (300,000 trees).

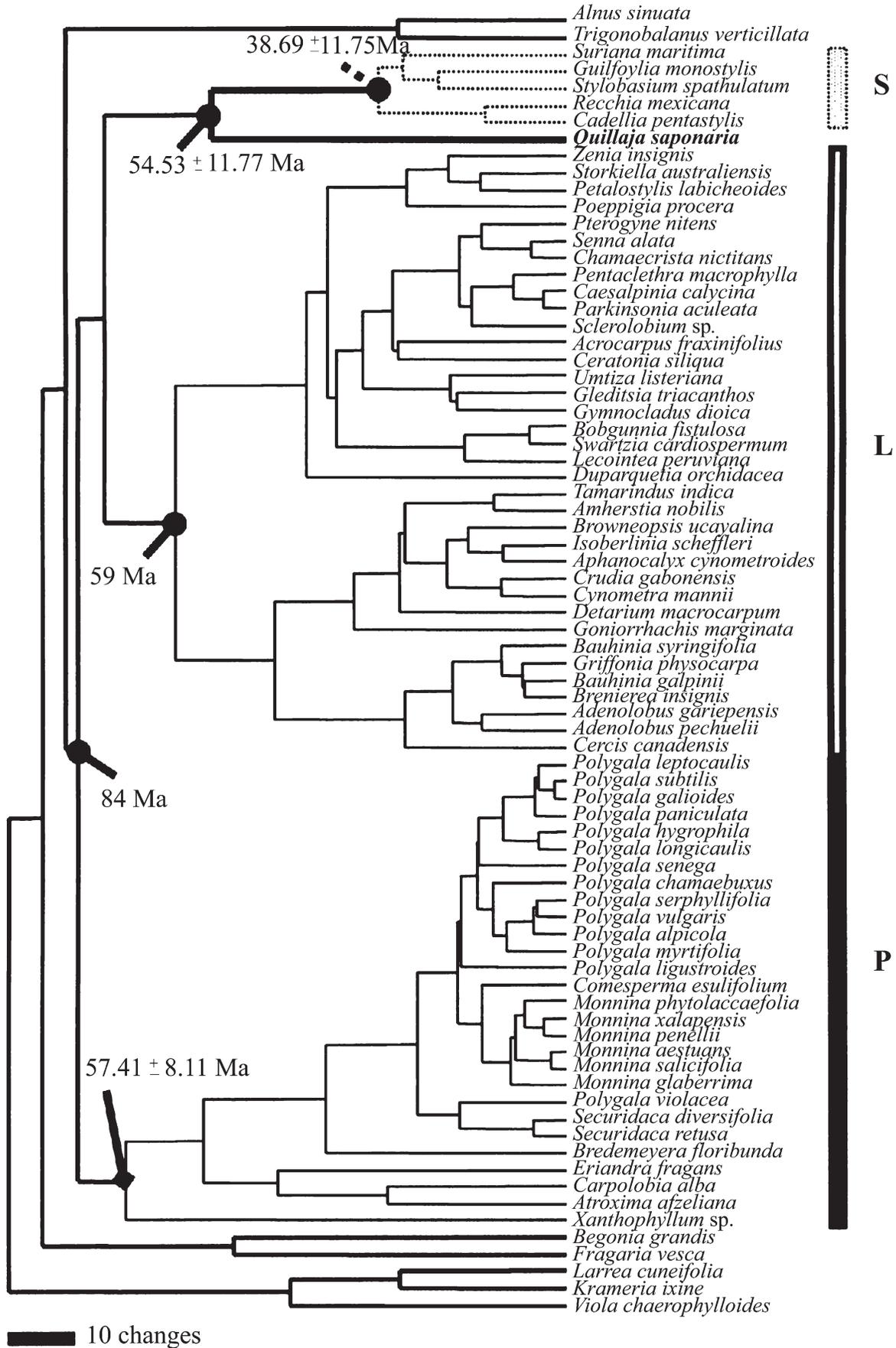


FIG. 2. Chronogram based on one of the post burn-in *matK* trees displaying ages of the Polygalaceae (P), Surianaceae (S), and (Quillajaceae + Surianaceae) crown clades. The constraints used (84 ma maximum age for Fabales and 59 ma fixed age for Leguminosae) are also indicated.

Although the ILD test suggested incongruence between the two regions, combined analyses were investigated. The combined analyses of *matK* and *rbcL* recovered topologies that were congruent with topologies recovered by *matK* alone; if combined analysis measures strength of evidence supporting disparate results (Nixon and Carpenter 1996), the *matK* derived topology is favored as it is recovered in the "all taxon combined" trees; however the low support and the inconsistency between Bayesian and parsimony results are noted. The Bayesian and parsimony analyses of *matK* recover Quillajaceae and Surianaceae as sisters, but unlike the parsimony analysis, the Bayesian analysis finds the clade (Leguminosae + Polygalaceae), although with poor support (Table 4). Although for some comparisons the SH test favours the *matK* Bayesian topology, in which Polygalaceae and Leguminosae are sisters, there is no significant difference between the likelihoods of the *matK* parsimony and Bayesian topologies.

**Does Fabales Represent a Legitimate Case of Radiation?**—Evidence of the history of Fabales from the fossil record and molecular analyses suggests that Leguminosae, Polygalaceae and Surianaceae were all present in the Tertiary and that the origin of the first two families could be traced in the Paleocene [cf. Wikström et al. (2001) who calculated a Campanian age for legumes (74–79 million years, Ma) and a Maastrichtian age for Polygalaceae (66–68 Ma)]. Leguminosae, with an age of ca. 51 Ma according to the oldest identifiable fossil record, represents one of the lineages that appear relatively late in the history of the eudicots displaying high speciation and low extinction rates (Herendeen and Crane 1992; Herendeen et al. 1992; Magallón et al. 1999; Crepet et al. 2004). Lavin et al. (2005) suggest that after a sudden appearance in the Late Paleocene and a short time span between origin and diversification of extant lineages, legume evolution probably involved just the last 60 Ma [cf. the early Paleocene age (64 Ma) of Leguminosae calculated by Davies et al. (2004)]. The minimum age calculated for the stem lineage of the clade that includes Rosaceae and Fagales, members of the nitrogen fixing clade, is estimated at 94 Ma (Friis and Crane 1989; Crepet et al. 2004), whereas other studies estimate the age of that clade around 80–85 Ma (Davies et al. 2004). These proposals are consistent with the idea of a relatively recent origin of families within the Fabales in comparison to some of their closest relatives. We find the Leguminosae crown node is oldest, although very similar in age and to the Polygalaceae and the crown node comprising Quillajaceae and Surianaceae. The diversification of Surianaceae seems to be much more recent (Table 6, Fig. 2).

Fossil evidence from Polygalaceae and Surianaceae is roughly coincident with the average ages here estimated for the crown clade of these families. The earliest putative fossil of Polygalaceae belongs to samaras of *Securidaca* dating from the late Paleocene ca. 58.5 Ma (Crane et al. 1990; Magallón et al. 1999; Pigg et al. 2004, 2008), followed by records from the Eocene (Martin 1982; Zhi-Chen et al. 2004) and Miocene (Huang 1980; Muller 1981; Magallón et al. 1999; see Lavin et al. 2005 for further references). On the other hand, affinity between *Suriana* L. and *Paranyssapollenites changeensis*, a fossil from the Eocene of China (Zhi-Chen et al. 2004), has been claimed. Recent age estimates using Bayesian dating (Forest et al. 2007) suggests that the diversification of Polygalaceae started in the Early Paleocene  $60.4 \pm 8.0$  Ma, very similar to our estimate. Forest et al. (2007) estimated the Surianaceae crown node age to be Early Eocene  $52.5 \pm 10.3$  Ma, much earlier than suggested in our study.

Here we define a deep radiation as the establishment of major lineages (and diverse morphology) in a relatively short time frame. Thus, according to the divergence time analyses performed here, the diversification of Leguminosae (59 Ma), Polygalaceae ( $57.41 \text{ Ma} \pm 8.11 \text{ Ma}$ ) and Quillajaceae+Surianaceae ( $54.53 \pm 11.77 \text{ Ma}$ ) represents a rapid radiation that could explain the lack of reliable resolution at the family level within Fabales.

**Perspectives for Molecular Data**—Whether the addition of more characters, taxa or both offers greatest improvements to phylogenetic resolution and accuracy has been an issue widely debated (e.g. Hendy and Penny 1989; Hillis 1998; Poe and Swofford 1999; Rosenberg and Kumar 2001; Rokas and Carroll 2005). Although some studies suggested that phylogenetic accuracy seems to improve when taxon sampling is increased (Graybeal 1998), others claimed that the addition of more characters is required in order to reconstruct phylogenetic histories (see Rosenberg and Kumar 2001; see Rokas and Carroll 2005). In Saxifragales it is suggested that approximately  $1.8 \times 10^7$  bp would be required to resolve infraordinal relationships (Fishbein et al. 2001). Estimation of the sequence data necessary to resolve interfamilial relationships of Lamiales suggest that at least 2000 parsimony informative nucleotides are necessary to reconstruct deep nodes relationships in this order (Wortley et al. 2005). Ideally, an increase in both characters and taxa is suggested for angiosperm phylogenetics (Soltis et al. 1998; Soltis et al. 2005). However, empirical studies conclude that if the purpose is to identify major subgroups, it is better to analyse more characters than taxa (Bremer et al. 1999; Li et al. 2004). Although numerous lineages of Leguminosae and Polygalaceae were compared with Surianaceae and Quillajaceae, it was not possible to recover well supported interfamilial phylogenetic relationships within the Fabales. As Wortley et al. (2005) argue for the Lamiales, further increases in taxon sampling are unlikely to resolve problematic interfamilial relationships. More taxa may be placed within the well-supported monophyletic families, but the number of characters grouping families is not increased. In Leguminosae even intrafamilial relationships involving the main divergent lineages are challenging despite the increase in taxa sampling; in a recent phylogenetic study where sequences of *matK* and the *trnL* and *3'-trnK* introns from 153 genera of Caesalpinioideae were implemented, relationships among the first branching lineages of the legumes are still not well supported, with Cercideae, Detarieae and the genus *Duparquetia* alternatively resolved as sister group to all of the legumes (Bruneau et al. 2008). In several phylogenetic studies, low support values for major clades have been detected even after the inclusion of an extensive sampling of taxa (e.g. Källersjö et al. 1998; Johnson 2001). Conversely, recent phylogenetic analyses in eurosids I based on complete plastome sequences (i.e. few taxa available) suggests a well supported sister group relationships of Rosales and Fabales (Ravi et al. 2007). Although empirical studies of sequence data of flowering plants suggest potential improvement of phylogenetic resolution after the increase in characters, a vast sequencing effort seems to be required, and even so factors such as rate substitution and combination of different data sets could occasionally decrease an optimum phylogenetic signal (Wortley et al. 2005).

**Perspectives for Morphological Studies**—Recently the use of morphological data for phylogenetic reconstruction has been viewed as problematic (Scotland et al. 2003), and

particularly problematic in the case of rapid radiations, since phenotypic characters could be overwritten by recent radiations, leaving few characters to identify obvious synapomorphies of the orders (Chase et al. 2000). However, if characters that evolved fast during a radiation but experiencing a slower evolutionary rate since then are required to resolve rapid and ancient radiations (Fishbein et al. 2001), then morphological characters seem to be worth exploring when cladogenetic events may be associated with their selection or constraint (see Donoghue and Sanderson, 1992 for further discussion). Phylogenetic analyses of Leguminosae where morphological and molecular data sets are combined, have exhibited better resolution and support values than analyses based only on DNA sequences only (Herendeen et al. 2003; Wieringa and Gervais 2003). However, a recent review suggests that the general picture where morphology and molecular data have been combined is one of increased resolution but not increased support when compared to molecular data alone (Wortley and Scotland 2006).

Fernando et al. (1993) have claimed that the highly derived morphological features of Fabales could not provide evidence of a close relationship between families. Indeed in a survey of wood anatomy features of APG orders (Baas et al. 2000), Fabales is one of the clades not supported by any consistent trait due the heterogeneity displayed in the order. However, there are some characters which are suggestive of certain interfamilial relationships. Notably, Claxton et al. (2005) stated that the pollen morphology of *Quillaja* and Surianaceae is more similar between these two taxa and the early diverging lineages of Leguminosae than to Polygalaceae. Additionally, comparative a floral ontogeny study highlight similarities between Leguminosae and Quillajaceae morphologies (but also notes important differences between Quillajaceae and Surianaceae; Bello et al. 2007). In the light of the relative merits of *matK* and *rbcL* partitions, and taking into account previous morphological studies (Claxton et al. 2005; Bello et al. 2007), we consider the topology suggesting Polygalaceae as sister to the clade (Leguminosae (Quillajaceae+Surianaceae)) to be the most satisfactory hypothesis of relationships within Fabales based on current data (Table 4, Fig. 1; parsimony analyses of *matK* and combined data for all taxa). However, this hypothesis needs to be tested further. Whether morphological data in Fabales are able to arbitrate between alternative interfamilial relationships may only be determined by wider and more careful consideration of phenotypical characters and their inclusion in a combined analysis (Bello et al. unpublished data). One of the reasons advocated for the lack of confidence in rosid relationships is that nonmolecular features have not been explored to identify synapomorphies for groups of families (Soltis and Soltis 2004; Soltis et al. 2005; Endress and Friis 2006; Schönenberger and Balthazar 2006). Since morphological characters are the features that ultimately support the distinctiveness of real biological entities and therefore maximise utility of a phylogeny, studies addressing an integral morphological and molecular survey will be key in the discrimination of elusive relationships at the family level.

**ACKNOWLEDGMENTS.** M.A.B., supported by the programme Alfan, the European Union Programme of High Level Scholarships for Latin America (E04D040438CO), would like to acknowledge the Systematic Association (UK) and J. Betancur, A. Cogollo, D. Giraldo, and F. González from the Universidad Nacional de Colombia for their valuable cooperation. Molecular work in the Bruneau laboratory was supported by funds from the Natural Sciences and Engineering Research Council of Canada.

The authors thank R. Bateman, H. Ochoterena, and P. Rudall who contributed substantially in the improvement of the paper with their comments.

#### LITERATURE CITED

- Albert, V., S.E. Williams, and M.W. Chase. 1992. Carnivorous plants: phylogeny and structural evolution. *Science* 257: 1491–1495.
- Anderberg, A.A., C. Rydin, and M. Källersjö. 2002. Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. *American Journal of Botany* 89: 677–687.
- APG. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.
- Baas, P., E. Wheeler, and M.W. Chase. 2000. Dicotyledonous wood anatomy and the APG system of angiosperm classification. *Botanical Journal of the Linnean Society* 134: 3–17.
- Beilstein, M.A., I.A. Al-Shehbaz, and E.A. Kellogg. 2006. Brassicaceae phylogeny and trichome evolution. *American Journal of Botany* 93: 607–619.
- Bello, M.A., J.A. Hawkins, and P.J. Rudall. 2007. Floral morphology and development in Quillajaceae and Surianaceae (Fabales), the species poor-relatives of Leguminosae and Polygalaceae. *Annals of Botany* 100: 1491–1505.
- Bernardi, L. 2000. Consideraciones taxonómicas y fitogeográficas acerca de 101 Polygalae americanas. *Cavanillesia Altera* 1: 1–456.
- Bremer, B., R.K. Jansen, B. Oxelman, M. Backlund, H. Lantz, and K.-J. Kim. 1999. More characters or more taxa for a robust phylogeny: case study from the coffee family (Rubiaceae). *Systematic Biology* 48: 413–435.
- Bruneau, A., F.J. Breteler, J.J. Wieringa, G.Y.F. Gervais, and F. Forest. 2000. Phylogenetic relationships in tribes Macrolobieae and Detarieae as inferred from chloroplast *trnL* intron sequences. Pp. 121–149 in *Advances in Legume Systematics* vol. 9, eds. P.S. Herendeen and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Bruneau, A., F. Forest, P.S. Herendeen, B. Klitgaard, and G. Lewis. 2001. Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trnL* intron sequences. *Systematic Botany* 26: 487–514.
- Bruneau, A., M. Mercure, G.P. Lewis, and P.S. Herendeen. 2008. Phylogenetic patterns and diversification in the caesalpinioideae legumes. *Canadian Journal of Botany* 86: 697–718.
- Chappill, J.A. 1995. Cladistic analysis of Leguminosae: the development of an explicit hypothesis. Pp. 1–9 in *Advances in legume systematics, phylogeny* vol. 7, eds. M. Crisp and J. Doyle. Richmond: Royal Botanic Gardens, Kew.
- 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. Systma, 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.-Y. Xiang, G.M. Plunkett, P.S. Soltis, S.M. Swensen, S. E. Williams, P.A. Gadek, C.J. Quinn, L.E. Eguiarte, E. Golenberg, G.H. Learn, S.W. Graham, S.C.H. Barrett, S. Dayanandan, and V. Albert. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. *Annals of the Missouri Botanical Garden* 80: 528–580.
- Chase, M.W., M.F. Fay, and V. Savolainen. 2000. Higher-level classifications in the angiosperms: new insights from the perspective of DNA sequence data. *Taxon* 49: 685–704.
- Chodat, R. 1896. Polygalaceae. Pp. 323–345 in *Die Natürlichen Pflanzenfamilien*, eds. A. Engler and K. Prantl. Leipzig: Engelmann.
- Claxton, F., H. Banks, B.B. Klitgaard, and P.R. Crane. 2005. Pollen morphology of families Quillajaceae and Surianaceae (Fabales). *Review of Palaeobotany and Palynology* 133: 221–233.
- Clement, W.L., M.C. Tebbitt, L.L. Forrest, J.E. Blair, L. Brouillet, T. Eriksson, and S.M. Swensen. 2004. Phylogenetic position and biogeography of *Hillebrandia sandwicensis* (Begoniaceae): a rare Hawaiian relict. *American Journal of Botany* 91: 905–917.
- Crane, P., S.R. Manchester, and D.L. Dilcher. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte formation (Paleocene) near Almont, North Dakota. *Fieldiana, Geology* n.s. 20:1–63.
- Crayn, D.M., E.S. Fernando, P.A. Gadeck, and C.J. Quinn. 1995. A reassessment of the familial affinity of the Mexican genus *Recchia* Mocino & Sessé ex DC. *Brittonia* 47: 397–402.

- Crepet, W.L., K.C. Nixon, and M.A. Gandolfo. 2004. Fossil evidence and phylogeny: the age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* 91: 1666–1682.
- Cronquist, A. 1981. An integrated system of classification of flowering plants. *New York: Columbia University Press*.
- Cronquist, A. 1988. *The evolution and classification of flowering plants*. New York: The New York Botanical Garden.
- Dahlgren, R.M.T. 1980. A revised system of classification of the angiosperms. *Botanical Journal of the Linnean Society* 80: 91–124.
- Davies, J., T.G. Barraclough, M.W. Chase, P.S. Soltis, D.E. Soltis, and V. Savolainen. 2004. Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences USA* 101: 1904–1909.
- Donoghue, M.J. and M.J. Sanderson. 1992. The suitability of molecular and morphological evidence in reconstructing plant phylogeny. Pp. 340–368 in *Molecular systematics of plants*, eds. P.S. Soltis D.E. Soltis, and J.J. Doyle. New York: Chapman and Hall.
- Doyle, J.J. 1995. DNA data and legume phylogeny: a progress report. Pp. 11–30 in *Advances in legume systematics, phylogeny* vol. 7, eds. M. Crisp and J. Doyle. Richmond: Royal Botanic Gardens, Kew.
- Doyle, J.J. and J. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.
- Doyle, J.J., J.L. Doyle, J.A. Ballenger, E.E. Dickson, T. Kajita, and H. Ohashi. 1997. A phylogeny of the chloroplast gene *rbcL* in the Leguminosae: taxonomic correlations and insights into the evolution of nodulation. *American Journal of Botany* 84: 541–554.
- Doyle, J.J., J.A. Chappill, D.C. Bailey, and T. Kajita. 2000. Towards a comprehensive phylogeny of legumes: evidence from *rbcL* sequences and non-molecular data. Pp. 1–20 in *Advances in legume systematics* vol. 9, eds. P.S. Herendeen and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Endress, P.K. and E.M. Friis. 2006. Rosids - Reproductive structures, fossil and extant, and their bearing on deep relationships: Introduction. *Plant Systematics and Evolution* 260: 83–85.
- Eriksen, B. 1993a. Phylogeny of the Polygalaceae and its taxonomic implications. *Plant Systematics and Evolution* 186: 33–55.
- Eriksen, B. 1993b. Floral anatomy and morphology in the Polygalaceae. *Plant Systematics and Evolution* 186: 17–32.
- Fan, C. and Q.-Y. Xiang. 2003. Phylogenetic analyses of Cornales based on 26S rRNA and combined 26 rDNA-*matK-rbcL* sequence data. *American Journal of Botany* 88: 1131–1138.
- Farris, J.S., M. Källersjö, A.G. Kluge, and C. Bult. 1994. Testing significance of incongruence. *Cladistics* 10: 315–319.
- Farris, J.S., V.A. Albert, M. Källersjö, D. Lipscomb, and A.G. Kluge. 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* 12: 99–124.
- Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.
- Fernando, E.S. and C.J. Quinn. 1992. Pericarp anatomy and systematics of Simaroubaceae sensu lato. *Australian Journal of Botany* 40: 263–289.
- Fernando, E.S., P.A. Gadeck, D.M. Crayn, and C.J. Quinn. 1993. Rosid affinities of Surianaceae: molecular evidence. *Molecular Phylogenetics and Evolution* 2: 344–350.
- Fishbein, M. and D.E. Soltis. 2004. Further resolution of the rapid radiation of Saxifragales (angiosperms, eudicots) supported by mixed model Bayesian-analysis. *Systematic Botany* 29: 883–891.
- Fishbein, M., C. Hibsich-Jetter, D.E. Soltis, and L. Hufford. 2001. Phylogeny of Saxifragales (Angiosperms, eudicots): analysis of a rapid, ancient radiation. *Systematic Biology* 50: 817–847.
- Forest, F. and J.C. Manning. 2006. Evidence for inclusion of South African endemic *Nylandtia* in *Muraltia* (Polygalaceae). *Systematic Botany* 31: 525–532.
- Forest, F., M.W. Chase, C. Persson, P.R. Crane, and J.A. Hawkins. 2007. The role of biotic and abiotic factors in the evolution of ant-dispersal in the milkwort family (Polygalaceae). *Evolution* 61: 1675–1694.
- Friis, E.M. and P.R. Crane. 1989. Reproductive structures of Cretaceous Hammamelidae. Pp. 155–174 in *Evolution, systematics, and fossil history of the Hammamelidae*, eds. P.R. Crane and S. Blackmore. Oxford: Clarendon Press.
- Fuhs, R. 1983. O genero *Quillaja* Molina (Rosaceae) no Brasil. *Arquivos do Jardim Botânico do Rio de Janeiro* 26: 61–67.
- Goloboff, P.A. 1993. Nona, version 16, Tucuman, Argentina: Published by the author.
- Graybeal, A. 1998. Is better to add taxa or characters to a difficult phylogenetic problems? *Systematic Biology* 47: 9–17.
- Harris, S. 1995. Systematics and randomly amplified polymorphic DNA in the genus *Leucaena* (Leguminosae, Mimosoideae). *Plant Systematics and Evolution* 197: 195–208.
- Hendy, M.D. and D. Penny. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.
- Herendeen, P.S. and P.R. Crane. 1992. Early caesalpinoid fruits from the Palaeogene of southern England. Pp. 57–68 in *Advances in legume systematics, the fossil record* vol. 4, eds. P.S. Herendeen and D.L. Dilcher. Richmond: Royal Botanic Gardens, Kew.
- Herendeen, P.S., W.L. Crepet, and D.L. Dilcher. 1992. The fossil history of the Leguminosae: phylogenetic and biogeographic implications. Pp. 303–315 in *Advances in legume systematics, the fossil record* vol. 4, eds. P.S. Herendeen and D.L. Dilcher. Richmond: Royal Botanic Gardens, Kew.
- Herendeen, P.S., A. Bruneau, and G. Lewis. 2003. Phylogenetic relationships in caesalpinoid legumes: a preliminary analysis based on morphological and molecular data. Pp. 37–62 in *Advances in legume systematics, higher level systematics* vol. 10, eds. B. Klitgaard and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Heywood, V.H. 1993. *Flowering plants of the world*. Oxford: Oxford University Press.
- Hillis, D. 1998. Taxon sampling, phylogenetic accuracy, and investigator bias. *Systematic Biology* 47: 3–8.
- Hilu, K.W. and H. Liang. 1997. The *matK* gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.
- Hilu, K.W., T. Borsch, K. Muller, D.E. Soltis, P.S. Soltis, V. Savolainen, M.W. Chase, M. Powell, L.A. Alice, R. Evans, H. Sauquet, C. Neinhuis, T.A.B. Slotta, J.G. Rohwer, C.S. Campbell, and L.W. Chatrou. 2003. Angiosperm phylogeny based on *matK*. *American Journal of Botany* 90: 1758–1776.
- Huang, T. 1980. Miocene palynomorphs of Taiwan, V. Angiosperm grains. *Taiwania* 25: 57–103.
- Huelsenbeck, J. and F. Ronquist. 2001. MR BAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* 17: 754–755.
- Johnson, K. P. 2001. Taxon sampling and the phylogenetic position of the Passeriformes: evidence from 916 avian cytochrome b sequences. *Systematic Biology* 128: 128–136.
- Johnson, L.A. and D.E. Soltis. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s.str. *Systematic Botany* 19: 143–156.
- Johnson, L.A. and D.E. Soltis. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. *Annals of the Missouri Botanical Garden* 82: 149–175.
- Judd, W.S., C.S. Campbell, E.A. Kellogg, P.F. Stevens, and M.J. Donoghue. 2002. *Plant systematics: a phylogenetic approach*. Sunderland: Sinauer Associates, Inc.
- Judd, W.S. and R.G. Olmstead. 2004. A survey of tricolpate (eudicot) phylogenetic relationships. *American Journal of Botany* 91: 1627–1644.
- Kajita, T., H. Ohashi, Y. Tateishi, C.D. Bailey, and J.J. Doyle. 2001. *rbcL* and legume phylogeny, with particular reference to Phaseolae, Milletiae and allies. *Systematic Botany* 26: 515–536.
- 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 *rbcL* DNA sequences reveals support for major clades of green plants, land plants, seed plants and flowering plants. *Plant Systematics and Evolution* 213: 259–287.
- Kamiya, K., K. Harada, M.M. Clyde, and A.L. Mohamed. 2002. Genetic variation of *Trigonobalanus verticillata*, a primitive species of Fagaceae, from Malaysia revealed by chloroplast sequences and AFLP markers. *Genes & Genetic Systems* 77: 177–186.
- Käss, E. and M. Wink. 1996. Molecular evolution of the leguminosae: phylogeny of the three subfamilies based on *rbcL* sequences. *Biochemical Systematics and Ecology* 24: 365–378.
- Kellogg, E.A. 2002. Are macroevolution and microevolution qualitatively different? Evidence from Poaceae and other families. Pp. 70–84 in *Developmental genetics and plant evolution*, eds. Q.C.B. Cronk, R.M. Bateman, and J.A. Hawkins. London: Taylor & Francis.
- Kita, Y. and M. Kato. 2001. Intrafamilial phylogeny of the aquatic angiosperm Podostemaceae inferred from nucleotide sequences of the *matK* gene. *Plant Biology* 3: 156–163.
- Lavin, M., P.S. Herendeen, and M.F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of the major family lineages immediately following an early Tertiary emergence. *Systematic Biology* 53: 653–661.
- Lersten, N.R. and H.T. Horner. 2005. Macropattern of styloid and druse crystals in *Quillaja* (Quillajaceae) bark and leaves. *International Journal of Plant Sciences* 166: 705–711.

- Lewis, G., B. Schrire, B. Mackinder, and M. Lock. 2005a. *Legumes of the world*. Richmond: Royal Botanic Gardens, Kew.
- Lewis, P.O., M.T. Holder, and K.E. Holsinger. 2005b. Polytomes and bayesian phylogenetic inference. *Systematic Biology* 54: 241–253.
- Li, R.-Q., Z.-D. Chen, A.-M. Lu, D.E. Soltis, P.S. Soltis, and P.S. Manos. 2004. Phylogenetic relationships in Fabales based on DNA sequences from three genomes. *International Journal of Plant Sciences* 165: 311–324.
- Lia, V.V., V.A. Confalonieri, C.I. Comas, and J.H. Hunziker. 2001. Molecular phylogeny of *Larrea* and its allies (Zygophyllaceae): reticulate evolution and the probable time of creosote bush arrival to North America. *Molecular Phylogenetics and Evolution* 21: 309–320.
- Luckow, M., P.J. White, and A. Bruneau. 2000. Relationships among the basal genera of mimosoid legumes. Pp. 165–180 in *Advances in Legume Systematics* vol. 9, eds. P. Herendeen and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Luckow, M., J.T. Miller, D.J. Murphy, and T. Livshultz. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. Pp. 197–220 in *Advances in legume systematics, higher level systematics* vol. 10, eds. B. Klitgaard and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Maddison, D. and W. Maddison. 2002. MacClade: analysis of phylogeny and character evolution. Sunderland: Sinauer Associates, Inc.
- Magallón, S., P.R. Crane, and P.S. Herendeen. 1999. Phylogenetic patterns, diversity and diversification of eudicots. *Annals of the Missouri Botanical Garden* 86: 297–372.
- Martin, H.A. 1982. Changing Cenozoic barriers and the Australian palaeobotanical record. *Annals of the Missouri Botanical Garden* 69: 625–667.
- Miller, J.T. and R.J. Bayer. 2001. Molecular phylogenetics of *Acacia* (Fabaceae:Mimosaceae) based on the chloroplast *matK* coding sequence and flanking *trnK* intron spacer regions. *American Journal of Botany* 88: 697–705.
- Mishler, B.D. 2000. Deep phylogenetic relationships among “plants” and their implications for classification. *Taxon* 49: 661–683.
- Morgan, D., D.E. Soltis, and K.R. Robertson. 1994. Systematic and evolutionary implications of the *rbcl* sequence variation in Rosaceae. *American Journal of Botany* 81: 890–903.
- Muller, J. 1981. Fossil pollen records of extant angiosperms. *Botanical Review* 47: 1–142.
- Nixon, K.C. 2002. WinClada. Ithica, NY: Published by the author.
- Nixon, K.C. and J.M. Carpenter. 1996. On simultaneous analysis. *Cladistics* 12: 221–241.
- Nylander, J. 2004. MrModeltest v2. Uppsala: Evolutionary Biology Centre, Uppsala University.
- Paiva, J.R. 1998. Polygalarum Africanarum et Madagascariensium prodromus atque gerontogaei generis Heterosamara Kuntze, a genere *Polygala* L. segragati et a nobis denuo recepti, synopsis monographica. *Fontqueria* 48: 1–346.
- Persson, C. 2001. Phylogenetic relationships in Polygalaceae based on plastid DNA sequences from the *trnL-F* region. *Taxon* 50: 763–779.
- Pigg, K.B., M.F. Wojciechowski, and M.L. DeVore. 2004. Samaras from the Late Paleocene Almont and Beicegel Creek floras of north Dakota, U.S.A., with potential affinities to *Securidaca* (Polygalaceae). Abstract submitted to Botany 2006, available at <http://www.2004.botanyconference.org/engine/search/index.php?func=detail&aid=594>.
- Pigg, K.B., M.L. DeVore, and M.F. Wojciechowski. 2008. *Paleosecuridaca curtisii* gen. et sp.nov., *Securidaca*-like samaras (Polygalaceae) from the late Paleocene of North Dakota and their significance to the divergence of families within the Fabales. *International Journal of Plant Sciences* 169: 1304–1313.
- Poe, S. and D.L. Swofford. 1999. Taxon sampling revisited. *Nature* 398: 299–300.
- Potter, D., F. Gao, P.E. Bortiri, S.-H. Oh, and S. Baggett. 2002. Phylogenetic relationships in Rosaceae inferred from chloroplast *matK* and *trnL-trnF* nucleotide sequence data. *Plant Systematics and Evolution* 231: 77–89.
- Ravi, V., J.P. Khurana, A.K. Tyagi, and P. Khurana. 2007. Rosales sister to Fabales: towards resolving the rosoid puzzle. *Molecular Phylogenetics and Evolution* 44: 488–493.
- Rokas, A. and S.B. Carroll. 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* 22: 1337–1344.
- Rosenberg, M.S. and S. Kumar. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. *Proceedings of the National Academy of Sciences USA* 98: 10751–10756.
- Sanderson, M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.
- Sanderson, M.J. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics (Oxford, England)* 22: 301–302.
- Savolainen, V., M.W. Chase, S.B. Hoot, C.M. Morton, D.E. Soltis, C. Bayer, M.F. Fay, A.Y.D. Bruijn, S. Sullivan, and Y.-L. Qiu. 2000a. Phylogenetics of flowering plants based upon a combined analysis of plastid *atpB* and *rbcl* gene sequences. *Systematic Biology* 49: 306–362.
- Savolainen, V., M.F. Fay, D.C. Albach, A. Blacklund, M.V.d. Bank, K.M. Cameron, S.A. Johnson, M.D. Lledo, J.-C. Pintaud, M. Powell, M.C. Sheahan, D.E. Soltis, P.S. Soltis, P. Weston, W.M. Whitton, M.C. Wurdack, and M.W. Chase. 2000b. Phylogeny of eudicots: a nearly complete familial analysis based on *rbcl* gene sequences. *Kew Bulletin* 55: 257–309.
- Schönenberger, J. and M. v. Balthazar. 2006. Reproductive structures and phylogenetic framework of the rosids - progress and prospects. *Plant Systematics and Evolution* 260: 87–106.
- Schulmeister, S. 2004. Inconsistency of maximum parsimony revised. *Systematic Biology* 53: 521–528.
- Scotland, R.W., R.G. Olmstead, and J.R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. *Systematic Biology* 52: 539–548.
- Sheenan, M.C. and M.W. Chase. 1996. A phylogenetic analysis of Zygophyllaceae R. Br. based on morphological, anatomical and *rbcl* sequence data. *Botanical Journal of the Linnean Society* 122: 279–300.
- Simmons, M. and H. Ochoterena. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.
- Sober, E. 2004. The contest between parsimony and likelihood. *Systematic Biology* 53: 644–653.
- Soltis, D.E., D. Morgan, A. Grable, P.S. Soltis, and P. Kuzoff. 1993. Molecular systematics of Saxifragaceae sensu stricto. *American Journal of Botany* 80: 1056–1081.
- Soltis, D.E., P.S. Soltis, D.R. Morgan, S.M. Swensen, B.C. Mullin, J.D. Dowd, and P.G. Martin. 1995. Chloroplast gene sequence data suggest a single origin of the predisposition for symbiotic nitrogen fixation in angiosperms. *Proceedings of the National Academy of Sciences USA* 92: 2647–2651.
- Soltis, D.E., P.S. Soltis, D. Nickrent, L.A. Johnson, W.H. Hahn, S.B. Hoot, J.A. Sweere, R.B. Kuzoff, K.A. Kron, M.W. Chase, S.M. Swensen, E.A. Zimmer, S.-M. Chaw, L.J. Gillespie, W.J. Kress, and K.J. Systma. 1997. Angiosperm phylogeny inferred from 18S ribosomal DNA sequences. *Annals of the Missouri Botanical Garden* 84: 1–49.
- Soltis, D.E., P.S. Soltis, M.E. Mort, M.W. Chase, V. Savolainen, S.B. Hoot, and C.M. Morton. 1998. Inferring complex phylogenies using parsimony: an empirical approach using three large DNA data sets for angiosperms. *Systematic Biology* 47: 32–42.
- Soltis, D.E., P.S. Soltis, M.W. Chase, M.E. Mort, D.C. Albach, M. Zanis, V. Savolainen, W.H. Hahn, S.B. Hoot, M.F. Fay, S.M. Axtell, S.M. Swensen, L.M. Prince, W.J. Kress, K.C. Nixon, and J.S. Farris. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcl*, and *atpB* sequences. *Botanical Journal of the Linnean Society* 133: 381–461.
- Soltis, D.E., A. Sinters, M. Zanis, S. Kim, J. Thompson, P. Soltis, L.D. Craene, P. Endress, and J. Farris. 2003. Gunnerales are sister to other core eudicots: implications for the evolution of pentamery. *American Journal of Botany* 90: 461–470.
- Soltis, D.E., P.S. Soltis, P.K. Endress, and M.W. Chase. 2005. *Phylogeny and evolution of angiosperms*. Sunderland: Sinauer Associates, Inc.
- Soltis, P.S. and D.E. Soltis. 2004. The origin and diversification of angiosperms. *American Journal of Botany* 91: 1614–1626.
- Steele, K.P. and R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. *Systematic Botany* 19: 126–142.
- Swensen, S.M., J.N. Luthi, and L.H. Rieseberg. 1998. Datisceaeae revised: monophyly and the sequence of breeding system evolution. *Systematic Botany* 23: 157–169.
- Swofford, D. 2002. PAUP\*. Phylogenetic analysis using parsimony (\*and other methods). Sunderland: Sinauer Associates, Inc.
- Takhtajan, A. 1997. *Diversity and classification of flowering plants*. New York: Columbia University Press.
- Thorne, R.F. 2000. The classification and geography of flowering plants: Dicotyledons of the Class Angiospermae. *Botanical Review* 66: 441–647.
- Thulin, M.B., B. Bremer, J. Richardson, J. Niklasson, M.F. Fay, and M.W. Chase. 1998. Family relationships of the enigmatic rosoid genera *Barbeya* and *Dirachma* from the Horn of Africa region. *Plant Systematics and Evolution* 213: 103–119.
- Tucker, S.C. and A.W. Douglas. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. Pp. 11–32 in

- Advances in legume systematics, structural botany* vol. 6, eds. I. K. Ferguson and S. Tucker. Richmond: Royal Botanic Gardens, Kew.
- Verkerke, W. 1985. Ovules and seeds of the Polygalaceae. *Journal of the Arnold Arboretum* 66: 353–394.
- Wieringa, J. J. and G. Y. F. Gervais. 2003. Phylogenetic analyses of combined morphological and molecular data sets on the *Aphanocalyx-Bykinia-Tetraberlinia* group (Leguminosae, Caesalpinioideae, Detarieae s.l.). Pp. 181–196 in *Advances in legume systematics, higher level systematics* vol. 10, eds. B. Klitgaard and A. Bruneau. Richmond: Royal Botanic Gardens, Kew.
- Wikström, N., V. Savolainen, and M. W. Chase. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 268: 2211–2220.
- Wojciechowski, M. F. 2003. Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective. Pp. 5–35 in *Advances in legume systematics, higher level systematics* vol. 10, eds. B. Klitgaard and A. Bruneau. The Royal Botanic Gardens, Kew.
- Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* 91: 1846–1862.
- Wortley, A. H. and R. W. Scotland. 2006. The effect of combining molecular and morphological data in published phylogenetic analyses. *Systematic Biology* 55: 677–685.
- Wortley, A. H., P. J. Rudall, D. J. Harris, and R. W. Scotland. 2005. How much data are needed to resolve a difficult phylogeny? Case study in Lamiales. *Systematic Biology* 54: 697–709.
- Xiang, Q.-Y., D. E. Soltis, and P. S. Soltis. 1998. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. *American Journal of Botany* 85: 285–297.
- Yoder, A. D., J. A. Irwin, and B. A. Payseur. 2001. Failure of the ILD to determine data combinability for Slow Loris phylogeny. *Systematic Biology* 50: 408–424.
- Yokoyama, J., M. Suzuki, K. Iwatsuki, and M. Hasebe. 2000. Molecular phylogeny of Coriaria, with special emphasis on the disjunct distribution. *Molecular Phylogenetics and Evolution* 14: 11–19.
- Zhi-Chen, S., W. Wei-Ming, and H. Fei. 2004. Fossil pollen records of extant angiosperms in China. *Botanical Review* 70: 425–458.
- APPENDIX 1. Taxon sampling for the phylogenetic analyses of order Fabales based on plastid *matK* and *rbcL*. A dash indicates the region was not sampled. The number in parentheses following the GenBank accession code indicates the paper where these sequences have been previously published: [(1) Doyle et al. 1997; (2) Wojciechowski et al. 2004; (3) Käss and Wink 1996; (4) Kajita et al. 2001; (5) Fernando et al. 1993; (6) Morgan et al. 1994; (7) Soltis et al. 1993; (8) Thulin et al. 1998; (9) Kamiya et al. 2002; (10) Swensen et al. 1998; (11) Sheenan and Chase 1996; (12) Albert et al. 1992; (13) Hilu et al. 2003; (14) Luckow et al. 2003; (15) Yokoyama et al. 2000; (16) Li et al. 2004; (17) Potter et al. 2002; (18) Kita and Kato 2001; (19) Persson 2001; (20) Savolainen et al. 2000a; (21) Lia et al. 2001; (22) Clement et al. 2004; (23) Forest and Manning 2006; (24) Forest et al. 2007; (25) Bruneau et al. 2008]. Information is presented in the following order: **taxon name** GenBank accessions: *rbcL*, *matK*; *voucher specimen of the samples worked in this study* (SOURCE).
- Leguminosae.** *Acrocarpus fraxinifolius* Wight & Arn. AF308699(4), EU361843 (25); *Manos 1416* (DUKE). *Adenolobus garipensis* (E.Mey.) Torre & Hillc. AM234268 (24), EU361844 (25); *Leistner* et al. 246 (K). **A. pechuelii** (Kuntze) Torre & Hillc. AM234264 (24), EU361845 (25); *Oliver* et al. 6527 (K). *Amherstia nobilis* Wall. AM234234 (24), AF542601(13); *Baker 490* (KEP). *Aphanocalyx cynometroides* Oliver AM234241 (24), EU361855 (25); *Wieringa 2355* (WAG). *Archidendron hirsutum* I.C.Nielsen AM234253 (24), —; *Douglas 625* (MEL). *Baphia nitida* Afzel. ex Lodd. AM234261 (24), —; *Bruneau* sn. (LBG). *Bauhinia galpinii* N.E. Br. AM234262 (24), EU361875 (25); *Forest 347* (NBG). **B. gillvum** (Bailey) Pedley AM234266 (24), EU361876 (25); *Weston 2446* (NSW). **B. syringifolia** (F.Muell.) Wunderlin AM234267 (24), EU361878 (25); *Weston 2449* (NSW). **Bobgunnia fistuloides** (Harms) J.H.Kirkbr. & Wiersema AM234258 (24), EU361885 (25); *Breteler 14870* (WAG). *Brenierea insignis* Humbert AM234269 (24), EU361889 (25); *Dupuy M430* (K). *Browneopsis ucayalina* Huber AM234233 (24), EU361894 (25); *Klitgaard 684* (K). *Caesalpinia calycina* Benth AM234236 (24), EU361899 (25); *Lewis 1885* (K). *Calpocalyx dinklagei* Harms AM234257 (24), —; *Breteler 15461* WAG. *Calliandra pittieri* Standl. AM234255 (24), —; *Klitgaard 649* (K). **C. trinervia** Benth. AM234255 (24), —; *Klitgaard 622* (K). *Cassia grandis* L.f. AM234244 (24), —; *Smith 2061* (MT). *Cedrelinga cate-naeformis* (Ducke) Ducke AM234256 (24), —; *Klitgaard 698* (K). *Ceratonina siliqua* L. U74203(1), AY386852 (2). **Cercis canadensis** L. U74188 (1), AY386908 (2). **C. chinensis** Bunge AM234263 (24), —; *Bruneau 1182* (MT). *Chamaecrista nictitans* (L.) Moench var. **jaliscensis** (Greenman) Irwin & Barnaby AM234248 (24), EU361914 (25); *Klitgaard 654* (K). **Crudia gabonensis** Pierre ex Harás AM234230 (24), EU361922 (25); *Wieringa 2585* (WAG.) **Cynometra manni** Oliver AM234231 (24), EU361925 (25); *Wieringa 2498* (WAG). *Delonix elata* (L.) Gamble AM234235 (24), —; *Herendeen 20XII97-1US*. **Detarium macrocarpum** Harms AM234239 (24), EU361929 (25); *Breteler 12528* (WAG). **Dialium guianensis** (Aubl.) Sandw. AM234245 (24), —; *Klitgaard 686* (K). **Duparquetia orchidacea** Baill. —, EU361937 (25); *Bruneau 1098* K. **Erythrophleum ivorense** A.Chev U74205 (1), —. **Gleditsia triacanthos** L. Z70129(3), AY386849(2). **Goniorrhachis marginata** Taub. AM234232 (24), EU361959 (25); *Lewis 1909* (K). **Griffonia physocarpa** Baill. AM234265 (24), EU361961 (25); *Cheek 8013* (K). **Gymnocladus dioica** (L.) Koch U74193(1), AY386929(2). **Inga** sp. AM234254 (24), —; *Klitgaard 677* (K). **Isoberlinia scheffleri** (Harms) Greenway AM234240 (24), EU361983 (25); *Herendeen 16-XII-97-2* (US). **Lecointea peruviana** Standl. AM234260 (24), EU361990 (25); *Klitgaard 679* (K). **Parkia multijuga** Benth. AM234251 (24), —; *Klitgaard 697* (K). **Parkinsonia aculeata** L. Z70157(3), AY386917(2). **Peltophorum pterocarpum** (DC.) Backer ex K.Heyne AM234243 (24), —; *Goyder 3719* (K). **Pentaclethra macrophylla** Benth. AM234250 (24), AF521853(14); *DeWilde 11496* (WAG). **Petalostyllis labicheoides** R.Br. AF308719(4), AY386895(2). **Poeppegia procera** Presl. AM234246 (24), AY386907(2); *Howard 5162* (MT). **Pterogyne nitens** Tul. AM234247 (24), EU362031 (25); *Herendeen 13-XII-97-1* (US). **Saraca palem-banica** Baker. AM234238 (24), —; *Breteler 13509* (WAG). **Senna alata** (L.) Roxb. U74250(1), EU362042 (25); *Bruneau 1076* (K). **Sclerolobium** sp. AM234242 (24), EU362040 (25); *Klitgaard 687* (K). **Storckia australiensis** J.H.Ross & B.Hyland AM234249 (24), EU362052 (25); *Hill 2096* (K). **Swartzia cadiospermum** Spruce ex. Benth. AM234259 (24), EU362053 (25); *Klitgaard 664* (K). **Tamarindus indica** L. Z70160(3), EU362056 (25); *JBM 2138-76* (MT). **Umtiza listeriana** Sim. AM234237 (24), EU362062 (25); *Schrire 2602* (K). **Zenia insignis** Chun. AF308722(4), EU362065 (25); *Manos 1418* (DUKE). **Polygalaceae.** **Atroxima afzeliana** (Oliv.ex Chod.) Staff AM234175 (24), EU604049; *Jongkind 4281* (WAG). **A. liberica** Stapf AM234174 (24), —; *Reitsma* et al. 1115 (NY). **Barnhartia floribunda** Gleason AM234168 (24), —; *Boom* et al. 8594 (K). **Bredemeyera colletioides** (Phil.) Chodat AM234171 (24), —; *Guaglianone* et al. 1587 (NY). **B. floribunda** (Willd.) Chodat EU644699, EU596520; *Bello 742* (COL). **B. microphylla** Hieron AM234173 (24), —; *P. J. M. Maas 8181* (U). **Carpobolba alba** G.Don AM234176 (24), EU604053; *Cable 747* (K). **C. goetzii** Guerde AM234177 (24), —; *Goyder* et al. 3722 (K). **Comesperma ericinum** DC. L29492(5), —; *Gadek* sn (NSW). **C. esulifolium** (Gand.) Prain AM234179 (24), EU596516; *Telford 12350* (CANB). **C. hispidulum** Pedley AM234178 (24), —; *T. White 7078* (NY). **Eriandra fragans** Royen & Steenis. AM234170 (24), EU604051; *Pullen 7234* (K). **Heterosamara tatarinowii** (Regel) Paiva AM234208 (24), —; *Togasi 1223* (GB). **Monnina aestuans** (L.f.) DC. EU644698, EU604037; *Bello 658* (COL). **M. dictyocarpa** Griseb. AM234183, —; *Wall* sn GB. **M. glaberrima** Chodat. EU644697, EU604039; *Betancur 10882* (COL). **M. hirta** (Bonpl.) B.Eriksen. AM234181 (24), —; *Sagastegui 14374* (NY). **M. leptostachya** Benth. AM234182, —; *Harling & Ståhl 26293* (GB). **M. malmeana** Chodat. AM234180 (24), —; *Eiten & Eiten 4249* (NSW). **M. cf. pennellii** Ferreyra. EU644696, EU604036; *Betancur 10731* (COL). **M. phillyrioides** (Bonpl.) B.Eriksen. AM234185 (24), —; *Harling & Ståhl 26293* (GB). **M. phytolaccaefolia** Kunth. EU644695, EU596519; *Betancur 10773* (COL). **M. pterocarpa** Ruiz & Pavon. AM234186, —; *Sagastegui 14908* (GB). **M. salicifolia** Ruiz & Pavón EU644694, EU604038; *González 4131* (COL). **M. xalapensis** Kunth. AM234184 (24), EU604047; *Chase 963* (K). **Monnina** sp. AM234187 (24), —; *Kloch* sn (K). **Monnina** sp.1. EU644693, —; *González 4191* (COL). **Monnina** sp.2. EU644692, —; *González 4190* (COL). **Moutabea aculeata** (Ruiz & Pavon) Poepp. & Endl. AM234169 (24), —; *Smith 1522* (US). **Muraltia alba** Levyns. AJ829696 (23), —; *Goldblatt 9515* (MO). **M. angulosa** Turcz. AJ829685 (23), —; *Forest 183* (NBG, K). **M. ericoides** (Burm.f.) Steud. AJ829695 (23), —; *Forest 315* (NBG, K). **M. heisteria** (L.) DC. AJ829698 (23), —; *Forest 159* (NBG, K). **M. mixta** DC. AM234188 (24), —; *Chase 6567* (K). **M. pauciflora** (Thunb.) DC. AJ829687 (23), —; *Forest 235* (NBG, K). **Nylandtia scoparia** (E. & Z.) Goldblatt & Manning AJ829701 (23), —; *Forest & Manning 249* (NBG, K). **N. spinosa** (L.) Dumort AJ829700, —; *Chase 281* (K). **Polygala acuminata** Willd. AM234195 (24), —; *Wurdack 1818* (NY). **P. alpica** Rupr. AM234191 (24), EU604041; *Chase 11747* (K). **P. amara** L. Z70175(3), —. **P. arillata** D.Don AM234210 (24), —; *Bartholomew et al.* sn (US). **P. bracteolata** L. AJ829702(23), —; *Forest 152* (NBG, K). **P. calcarea** “Lillet” Schultz AM234194 (24), —; *Chase 11322* (K). **P. chamaebuxus** var. **grandiflora** L. AM234198 (24), EU596517; *Chase 11323* (K). **P. comosa** Schkuhr AM234211 (24), —; *Wieringa 3462* (WAG). **P. cowellii** (Britton) S.F.Blake AM234199 (24), —; *Axelrod & Chavez 5936* (NY). **P. cruciata** L. L01945 (12), —. **P. erioptera** DC. Chodat AM234219 (24), —; *Simonsen et al.* sn. **P. fuertesii** (Urb.) S.F.Blake AM234200 (24), —;

*Thomsson 10465* (NY). **P. fruticosa** P.J.Bergius AM234215 (24),—; *Forest 254* (NBS, K). **P. galapeia** Hook.f AM234218 (24),—; *Adersen 1732* (C). **P. galioides** Poir. EU644691, EU604048; *Bello 722* (COL). **P. garcinii** DC. AM234212 (24),—; *Forest 161* (NBS, K). **P. guayaquilensis** Eriksen & Stahl AM234190 (24),—; *Cornejo & Bonifaz 6028* (GB). **P. hebeclada** DC. AM234196 (24),—; *H. S. Irwin et al. 13559* (NY). **P. hispida** Burch ex DC. AM234224 (24),—; *Nanni & Forest 302* (NBS, K). **P. hygrophila** Kunth EU644689, EU604040; *Bello 728* (COL). **P. jamaicensis** Chodat EU644690, —; *Cogollo sn* (COL). **P. klotzschii** Chodat AM234209 (24),—; *Silva & Hatschbach 638* (C). **P. lindheimeri** A.Gray AM234206 (24),—; *Barneby 17972* (NY). **P. longicaulis** Kunth EU644688, EU604042; *Bello 720* (COL). **P. microphylla** L. AM234203 (24),—; *Charpin et al. 22671* (NY). **P. myrtifolia** L. AJ829699 (24), EU604043; *Forest 162* (NBS, K). **P. obscura** Benth AM234207 (24), —; *Holmgren & Lowrey 8022* (NY). **P. oleaefolia** A. St-Hil. & Moq. AM234204 (24), —; *Anderson et al. 35551* (K). **P. oxyphylla** DC. AM234202 (24), EU604054; *Harley et al. 20751* (K). **P. paniculata** L. EU644684, EU596518; *Betancur 10752* (COL). **P. penaea** L. AM234201 (24),—; *Ståhl & Lindström 212* (GB). **P. peduncularis** Burch ex DC. AM234214 (24),—; *Forest 202* (NBS, K). **P. rehmannii** Chodat AM234222 (24),—; *Nanni sn* (NBS). **P. rupestris** Pourr. AM234220 (24),—; *Fay sn.* (K). **P. rhinostigma** Chodat AM234223 (24),—; *Forest & Nanni 295* (NBS, K). **P. sancti-georgii** Riley AM234217 (24),—; *Adersen 1897* (C). **P. senega** L. AM234189 (24), EU604034; *Brouillet 99-11* (MT). **P. serpyllifolia** Fischer-Oost. EU644685, EU604052; *Rudall sn.* (K). **P. subtilis** Kunth EU644683, EU604033; *Bello 719* (COL). **P. tenella** Willd. EU644687, EU604030; *González 4099<sup>a</sup>* (COL). **P. teretifolia** L.f. AM234213 (24),—; *Forest 200* (NBS, K). **P. uncinata** E.Mey. AM234192 (24),—; *Chase 6535* (K). **P. violacea** Aubl. EU644686, EU406035; *Bello 685* (COL). **P. violioides** A.St-Hil. & Moq. AM234205 (24),—; *Ratter et al. 6425* (K). **P. virgata** Vell. AM234216 (24),—; *Forest & Nanni 294* (NBS, K). **P. vulgaris** L. AM234193 (24), EU604046; *Fay*

*316* (K). **Securidaca bialata** Benth. EU644682, —; *Bello 726* (COL). **S. diversifolia** (L.) S.F.Blake. AM234225 (24), EU604055; *Chase 2998* (MICH). **S. retusa** Benth. EU644681, EU604029; *Bello 691* (COL). **S. virgata** Sw. AM234226 (24),—; *Ståhl et al. 2311* (GB). **S. welwitschii** Oliver AM234227 (24),—; *Strid 2887* (GB). **Securidaca** sp. EU644680, —; *Bello 723* (COL). **Xanthophyllum affine** Korth ex Miq. AM234228 (24),—; *Church et al. 1293* (NY). **X. octandrum** (F.Muell.) Domin. AM234229 (24),—; *Forster 9554* (NY). **Xanthophyllum** sp. AJ235799 (20), EU604044; *Coode 7760* (K). **Quillajaaceae**. **Quillaja saponaria** Molina U06822(6), AY386843(2). **Surianaceae**. **Cadellia pentastylis** F.Muell. L29491(5), EU604056; *Fernando & Quinn 1383* (UNSW). **Guilfoylia monostylis** (Benth.) F.Muell. L29494(5), EU604031; *Fernando & Wannan 1262* (UNSW). **Recchia mexicana** Moc. & Sessé AM234270 (24), EU604045. **Suriana maritima** L. U07680(5), AY386950(2). **Stylobasium australe** (Hook.) Prance U07679(5), —. **S. spathulatum** Desf. U06828(6), EU604032; *Latz 13213* (K).

**Outgroups**. **Begoniaceae**. **Begonia glabra** Aubl. AF008962(10), —. **B. grandis** Otto ex A.DC. AF534733(22), AB016466(15). **Betulaceae**. **Alnus glutinosa** (L.) Gaern. EU644678, —; *Chase 10822* (K). **A. sinuata** Rydb. AY263926 (16), AY263914 (16). **Cucurbitaceae**. **Abobra tenuifolia** Naud. AF008961(10), —. **Elaeagnaceae**. **Shepherdia argentea** (Pursh.) Nutt. AJ225787(8), —. **Fagaceae**. **Trigonobalanus verticillata** Forman AB084769(9), AB084771(9). **Krameriaceae**. **Krameria ixine** Löfling EU644679, EU604050; *Fernández 22529* (COL). **K. lanceolata** Torr. Y15032(11), —. **Oxalidaceae**. **Oxalis stricta** L. —, AF542605(13). **Rosaceae**. **Fragaria x annanasa** Duchesne U06805(7), —. **Fragaria vesca** L. —, AF288102(17). **Violaceae**. **Viola chaerophylloides** Makino —, AB038188(18). **Zygophyllaceae**. **Bulnesia arborea** Engl. EU644676, —; *Villalobos 164* (COL). **Kallstroemia** sp. EU644677, —; *Villalobos 134* (COL). **Larrea cuneifolia** Cav. AF200471(21), AF542602(13). **Zygophyllum xanthophyllum** Engl. AJ133872(11), —.