A PHYLOGENY OF LEGUMES (LEGUMINOSAE) BASED ON ANALYSIS OF THE PLASTID \textit{matK} GENE RESOLVES MANY WELL-SUPPORTED SUBCLADES WITHIN THE FAMILY\textsuperscript{1}

\textbf{MARTIN F. WOJCIECHOWSKI,\textsuperscript{2,5} MATT LAVIN,\textsuperscript{3} AND MICHAEL J. SANDERSON\textsuperscript{4}}

\textsuperscript{2}School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501 USA; \textsuperscript{3}Department of Plant Sciences, Montana State University, Bozeman, Montana 59717 USA; and \textsuperscript{4}Section of Evolution and Ecology, University of California, Davis, California 95616 USA

Phylogenetic analysis of 330 plastid \textit{matK} gene sequences, representing 235 genera from 37 of 39 tribes, and four outgroup taxa from euroids \textit{I} supports many well-resolved subclades within the Leguminosae. These results are generally consistent with those derived from other plastid sequence data (\textit{rbcL} and \textit{trnL}), but show greater resolution and clade support overall. In particular, the monophyly of subfamily Papilionoideae and at least seven major subclades are well-supported by bootstrap and Bayesian credibility values. These subclades are informally recognized as the \textit{Cladrastis} clade, genistoid sensu lato, dalbergioid sensu lato, mirbelioid, millettioid, and robbinioid clades, and the inverted-repeat-lacking clade (IRLC). The genistoid clade is expanded to include genera such as \textit{Poeclanthoe}, \textit{Cycloclamidium}, \textit{Bowdichia}, and \textit{Diplotrepis} and thus contains the vast majority of papilionoids known to produce quinolizidine alkaloids. The dalbergioid clade is expanded to include the tribe Amorpheae. The mirbelioids include the tribes Bossiaeeae and Mirbeliaceae, with Hypocryptaeae as its sister group. The millettioids comprise two major subclades that roughly correspond to the tribes Millettieae and Phaseoleae and represent the only major papilionoid clade marked by a macromorphological apomorphy, pseudoracemose inflorescences. The robbinioids are expanded to include \textit{Sesbania} and members of the tribe Loteae. The IRLC, the most species-rich subclade, is sister to the robbinioids. Analysis of the \textit{matK} data consistently resolves but modestly supports a clade comprising papilionoid taxa that accumulate canavanine in the seeds. This suggests a single origin for the biosynthesis of this most commonly produced of the nonprotein amino acids in legumes.

\textbf{Key words:} caesalpinoid legumes; Leguminosae; \textit{matK}; mimosoid legumes; papilionoid legumes; phylogeny.

The legume family is the third largest family of angiosperms (Mabberley, 1997) with approximately 730 genera and over 19,400 species worldwide (Lewis et al., in press). Legumes are second only to Poaceae (the grasses) in agricultural and economic importance. The family includes horticultural varieties and many species harvested as crops and for oils, fiber, fuel, timber, medicines, and chemicals. Ranging in habit from large trees to annual herbs, the family is well represented throughout temperate and tropical regions of the world (Rundel, 1989). The Leguminosae is particularly diverse, however, in tropical forests with a seasonally dry aspect and temperate shrublands tailored by xeric climates. Legumes are noticeably absent to poorly represented in mesic temperate habitats, including many arctic and alpine regions and the understory of cool temperate forests. The predilection of legumes for semi-arid to arid habitats is related to a nitrogen-demanding metabolism, which is thought to be an adaptation to climatically variable or unpredictable habitats whereby leaves can be produced economically and opportunistically (McKey, 1994). Indeed, nitrogen fixation via root-nodulating symbiotic bacteria is just one of several ways (in addition to associations with arbuscular mycorrhizae, ectomycorrhizae, and uptake of inorganic nitrogen compounds) in which legumes obtain high levels of nitrogen to meet the demands of their metabolism (Sprent, 1994, 2001). All legumes play an important role in the terrestrial nitrogen cycle regardless of whether they form root nodules (Sprent, 2001). Considered to be a tropical family with perhaps a late Cretaceous origin (65–70 Mya), the Leguminosae has an abundant and continuous fossil record since the Tertiary (Crepet and Taylor, 1985, 1986; Crepet and Herendeen, 1992; Herendeen et al., 1992). The occurrence of diverse assemblages of taxa representing all three subfamilies at multiple localities dating from the middle to upper Eocene, especially the Mississippi Embayment of southeastern North America, suggests that most major lineages of woody legumes (except for the tribe Cercidaceae) were present and had diversified extensively by this time (Herendeen et al., 1992).

Reconstructing the phylogenetic relationships of the Leguminosae is essential for understanding the origin and diversification of this ecologically and economically important family of angiosperms. Comprehensive phylogenetic analyses of Leguminosae began with the plastid gene \textit{rbcL} (Doyle, 1995; Käss and Wink, 1995, 1996; Doyle et al., 1997) following the early, widespread use of this gene for phylogenetic studies of land plant relationships (e.g., Chase et al., 1993). Among the conclusions that emerged, the monophyly of the Fabales (sensu Angiosperm Phylogeny Group, 2003) and the sister rela-
tionship of legumes to Polygalaceae, Surianaceae, and the rosaceous genus Quillaja Molina were very strongly supported (Doyle et al., 2000). Second, the monophyly of Leguminosae is consistently resolved although not as strongly as for the Fabales (Doyle et al., 2000; Kajita et al., 2001). Third, while monophyly of mimosoid legumes (subfamily Mimosoideae) is well supported by the rbcL data (Doyle et al., 2000), a more extensive sampling of the subfamily suggested certain mimo-
soid genera, Dinizia Ducke and Piptadeniastrium Brenan, were unresolved with respect to related caesalpinoid outgroups (Luckow et al., 2000). Fourth, the subfamily Caesalpinioideae (caesalpinoids) is consistently resolved as paraphyletic with respect to mimosoids and papilionoids (e.g., Polhill et al., 1981; Doyle et al., 2000; Kajita et al., 2001), although several well-supported subclades have been detected in recent studies of this subfamily; for example, the tribe Cerideae, resolved as the sister clade to the rest of the family (Doyle et al., 2000), the tribe Detarieae sensu lato (s.l.), distributed principally in tropical Africa and including approximately half of the genera in the Caesalpinioideae (Bruneau et al., 2001; Herendeen et al., 2003a), and the “Unütza” clade (Herendeen et al., 2003b). Lastly, the traditionally recognized subfamily Papilionoideae (sensu Polhill, 1981a, 1994) is consistently resolved as mono-
phyletic, albeit with only modest support (e.g., Doyle et al., 1997; Kajita et al., 2001).

The Papilionoideae has received the most attention, if only because it is the largest and most widespread of the three le-
gume subfamilies with an estimated 476 genera and 13,860 species (Lewis et al., in press). Papilionoids traditionally have been diagnosed by traits that now are considered synapomor-
phies of the subfamily. These include wood with predomi-
nantly paratracheal axial parenchyma that is usually storied; vessels with alternate layered pits and simple perforation plates; absence of bipinnate leaves; unidirectional initiation of sepal, petals, and stamens; clawed petals; and a seed testa with a hilar valve and no pleurogram (Polhill, 1981a; Tucker, 1987a, 2002; Tucker and Douglas, 1994; Chappill, 1995; Gas-
son, 2000). These many distinctions have sometimes resulted in papilionoids being ranked at the familial level (e.g., Hutchin-
on, 1964; Takhtajan, 1969). Moreover, support for the monophyly of Papilionoideae has not changed with family-

Despite insights gained into the higher-level relationships of the family from studies of the rbcL locus, and to a lesser extent the trnL-F region, many issues in legume phylogeny remain unresolved (reviewed in Wojciechowski, 2003). This is particu-
larly true for the relationships within the caesalpinoid and mimosoid subgroups and among some of the major papilion-
oid clades, the genistoids, dalbergioids, millettiods-phae-
loids, and Hologalegina (e.g., Crisp et al., 2000; Hu et al., 2000; Wojciechowski et al., 2000; Lavin et al., 2001). More variable nucleotide sequences are needed to improve the res-
olution of and support for the major clades within legumes. The most promising is the plastid gene matK, which has been shown by several recent studies on different papilionoid sub-
groups to provide excellent resolution among closely related genera (Hu et al., 2000; Lavin et al., 2001, 2003; Steele and Wojciechowski, 2003). Here we draw on these recent studies and a large number of new matK sequences as part of a more

extensive phylogenetic analysis of the family, with particular emphasis on the major clades of the Papilionoideae.

MATERIALS AND METHODS

Taxon sampling—Complete matK gene sequences from 330 taxa were included in this study, representing 235 genera of legumes as recognized by Polhill (1994) and four outgroup taxa from Fabales (Polygala, Suriana, Quil-
la) and Rosales (Vauquelinia) (Turner and Vilgalys, 1994; R. Geesink (Derris), Poissonia Baill. (Corsetia), Philenoptera Benth. (Loncho-
carpus), Calia Teran & Berlan (Sophora), and the newly described Maran-
iona (Hughes et al., 2004) were included. This study samples extensively in traditionally circumscribed tribes Aeschynomeneae (17/26 genera), Dalber-
giaeae (15/17), all genera of Amorphae (8), Robiniaeae (12), most genera of Trifolieae (6/7), and Vicieae (4/5). Representatives of only two, monogenic tribes, Mimozygantheae Burkart (Mimosoideae) and Euchresteae (Nakai) Ohashi (Papilionoideae), were not sampled for this analysis. Appropriate out-
group taxa from Polygalaceae, Surianaceae, Quillajaceae, and Rosaceae were chosen based on results of recent molecular phylogenetic studies of eurosids using rbcL-atpB-18S nuclear ribosomal DNA (Soltis et al., 2000), rbcL alone (Soltis et al., 1995; Kajita et al., 2001), matK (Steele et al., 2000), and the trnL-F region (Persson, 2001).

Sequences from 140 taxa are formally reported here for the first time, com-
plete with voucher specimen and database accession information, although a few of them have been used in part for phylogenetic analyses presented pre-
viously (Wojciechowski et al., 2000). Papers by Hu et al. (2000), Lavin et al. (2001, 2003), Luckow et al. (2003), Miller et al. (2003), McMahon and Huff-
ford (2004), Steele and Wojciechowski (2003), Riley-Hulting et al. (2004), and Thulin et al. (in press) provide sampling information for approximately 190 matK sequences from subgroups of the taxa included here and should be consulted for more details. The sources of plant material and GenBank ac-
ccession numbers for matK sequences from all taxa included in this paper are provided in the Appendix (see Supplemental Data accompanying the online version of this article).

DNA sequence data—The data presented here were gathered in our labo-
atories using similar methods. Genomic DNAs were isolated from field-col-
lected, greenhouse-grown plants, silica-dried and herbarium material using the procedure of Doyle and Doyle (1987) or using DNeasy Plant Minikits (Qia-
gen, Valencia, California, USA). Polymerase chain reaction (PCR) amplifi-
cations were performed using Taq and Platinum Taq DNA polymerases (Life Technologies, Gaithersburg, Maryland, USA) as described previously (Woj-
ciechowski et al., 1999; Lavin et al., 2000). For most of the newly sequen-
taxa, double-stranded copies of the matK gene and the flanking 3′ trnK intron region were amplified using primers trnK685F and trnK2R; typical reaction conditions were 2 min at 95°C for denaturation, followed by 35 cycles of 30 s at 95°C, 30–60 s at 55–57°C for annealing, 2 min 30 s at 72°C for primer extension, followed by a final 7 min incubation at 72°C. Amplification products were purified and then sequenced using these same primers and others listed in Table 1. DNA sequencing was performed on Applied Biosys-
tems 377 and 3100 sequencers (Applied Biosystems, Foster City, California, USA) at the University of California (DHS Sequencing Facility, Davis, Cali-
fornia, USA), Iowa State University (DNA Sequencing Facility, Ames, Iowa, USA), Northwoods DNA (Becida, Minnesota, USA), and Arizona State Uni-
iversity (DNA Laboratory, Tempe, Arizona, USA). Sequencer output files were assembled into contigs and edited using the program Sequencer 4.1 (GeneCodes, Ann Arbor, Michigan, USA) before alignment.

Primers for the PCR amplification and sequencing of the trnK/matK region from legumes (Table 1) were originally designed by one of us (M. F. Wojcie-
chowski) using published primer sequences (Steele and Vilgalys, 1994; John-
son and Soltis, 1995), which were modified based on the sole legume matK
sequence available at the time (Pisum sativum; Boyer and Mullett, 1988); various primers have been subsequently modified further to work more specifically with certain groups of taxa (e.g., Lavin et al., 2000; Riley-Hulting et al., 2004). Use of these primers generally resulted in 100% overlap in bidirectional sequencing of the plastid matK gene in legumes. Sequences given are all 5' to 3'; forward and reverse refer to direction with respect to matK coding sequence.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>trnK685F (forward)</td>
<td>GTATGCGCACTATGATCATATTGGA</td>
</tr>
<tr>
<td>matK4L (forward)</td>
<td>CCTCTGATGACTGGTAGAGAAGAT</td>
</tr>
<tr>
<td>matK1100L (forward)</td>
<td>TTTAGGCTGATAGGATCCAA</td>
</tr>
<tr>
<td>matK1932R (reverse)</td>
<td>CCAAGCCGCTTTACTAAGGG</td>
</tr>
<tr>
<td>matK832R (reverse)</td>
<td>TTGATGAAGAATAGATGCGTCAAA</td>
</tr>
<tr>
<td>trnK2R* (reverse)</td>
<td>CCGGAACACTAGCGGAGTGA</td>
</tr>
</tbody>
</table>

**Results**

Characteristics of the matK sequences—The matK gene in legumes ranges from 1476 bases (492 amino acids) in Erythrostemon gilliesii to 1545 bases (515 amino acids) in length in several dalbergioid taxa (e.g., Adesmia and Pictetia). The final matK data set includes 1674 aligned positions with 1042 potentially parsimony-informative characters (73.6%, which excludes the 244 indels and missing characters) among the 330 taxa analyzed. Of the total 552 420 characters in the data set, missing data accounted for 1.3% while indels and other excluded positions accounted for 14.6% (80 500 bases). By comparison, the complete sequence rbcL data set (Kajita et al., 2001) is 1404 aligned bases in length (with no indels) but contains fewer potentially informative characters (530 among the 319 sequences). That study, with a total of 242 sequences sampled from 194 legume genera and a similar emphasis on Papilionoideae but with fewer representatives from the other two subfamilies (24 Caesalpinoideae, 6 Mimosoideae) than the matK data set, contains sequences from a large number of taxa that are identical or very closely related to those in the matK data set. Comparative analyses of these genes (M. Lavin, P.S. Herendeen, and M.F. Wojciechowski, unpublished manuscript) show levels of sequence divergence up to ninefold lower for rbcL than for matK, and substitutions are distributed less uniformly among the three codon positions in rbcL. For example, the substitution rate at the third codon position in rbcL is 10 times that of the second position, whereas the third position in matK shows twice the rate of the second position. Previous comparisons of rates of substitution in matK vs. rbcL have yielded similar patterns of variation in other angiosperm groups (e.g., Steele and Vilgalys, 1994; Manos and Steele, 1997).

Within Fabales, pairwise distances (calculated across all sites as uncorrected p values in PAUP*) in matK sequences ranged from a maximum of nearly 17% among outgroups and caesalpinoïdes, to nearly 7% among mimosoides, 12% among caesalpinoïdes, and just over 19% among papilionoides. Within monophyletic genera for which we have more than three species sampled, pairwise distances varied between 0.8–1.9% in Astragalus and 0.9–4.3% in Sesbania. Within major papilionoid clades, pairwise distances varied from 0.1% to almost 11% within the genusoid clade, 0.3–12% in Loteae + Robineae; 0.0% (in Lens) to nearly 11% in the IRLC; 0.2–13% in dalbergioides, and 1.4–17.9% in the milletioides.

**Phylogenetic reconstruction**—Multiple heuristic searches of 1042 parsimony informative nucleotide characters, excluding indels, consistently converged on a large number of equally most parsimonious trees (maximum set saved = 10 000) with a minimum length of 8397 steps, a consistency index (CI) of 0.288 excluding uninformative characters, and a retention index (RI) of 0.791. The strict consensus tree of 5000 representative equally most parsimonious (MP) trees is highly re-

---

**Table 1. Sequences of oligonucleotide primers used for PCR amplification and sequencing of the plastid matK gene in legumes. Sequences given are all 5’ to 3’; forward and reverse refer to direction with respect to matK coding sequence.**

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>trnK685F (forward)</td>
<td>GTATGCGCACTATGATCATATTGGA</td>
</tr>
<tr>
<td>matK4L (forward)</td>
<td>CCTCTGATGACTGGTAGAGAAGAT</td>
</tr>
<tr>
<td>matK1100L (forward)</td>
<td>TTTAGGCTGATAGGATCCAA</td>
</tr>
<tr>
<td>matK1932R (reverse)</td>
<td>CCAAGCCGCTTTACTAAGGG</td>
</tr>
<tr>
<td>matK832R (reverse)</td>
<td>TTGATGAAGAATAGATGCGTCAAA</td>
</tr>
<tr>
<td>trnK2R* (reverse)</td>
<td>CCGGAACACTAGCGGAGTGA</td>
</tr>
</tbody>
</table>
Fig. 1. Phylogeny of Leguminosae based on parsimony analyses of plastid *matK* gene sequences. Phylogenetic relationships among the three subfamilies of Leguminosae (crown clade, “L”), and within subfamilies Caesalpinioideae and Mimosoideae. Outgroup lineages are indicated by bold lines. Members of Mimosoideae are indicated by gray boxes. Tree shown is strict consensus of 5000 equally most parsimonious trees (length = 58397 steps, consistency index = 0.288, retention index = 0.791) derived from heuristic search analyses of 330 *matK* sequences. Nodes designated by a diamond were not resolved in a 50% majority-rule consensus of the same set of 5000 equally most parsimonious trees. Nonparametric bootstrap proportions and Bayesian posterior probabilities from separate analyses (individual or range) are indicated above and below branches or immediately to left of appropriate node, respectively. Values are given for most nodes for which support values from both analyses were greater than 50%. *Vauquelinia* (Rosaceae) was designated as the outgroup for all analyses. Major papilionoid clades informally named here are indicated by a filled circle.

Resolved and presented in Figs. 1–5. The semi-strict consensus of the same set of parsimonious trees resolves only three nodes that are not present in the strict consensus, while in the 50% majority-rule tree a total of only seven nodes were not fully resolved. Branching order and support values for the major clades of legumes resolved by these *matK* data were very similar in the maximum parsimony and Bayesian analyses (Figs. 1–6). To illustrate the heterogeneity in estimated branch lengths, a phylogram representation of a typical Bayesian tree (sampled post burn-in) is shown in Fig. 6.

Analysis of the *matK* data confirms results from earlier studies in that the family is a monophyletic group, papilionoids and mimosoids, excluding *Dinizia* (tribe Mimosae), are monophyletic and nested within a paraphyletic Caesalpinioideae. All mimosoids and the majority of the caesalpinoid tribes Caesalpinieae and Cassieae comprise a strongly supported clade (Fig. 1) that is the sister group to papilionoids. Seven major clades and a number of minor clades within papilionoids are also highly supported (Figs. 2–5). In spite of this, relationships among certain of the clades, especially the gen-
Fig. 2. Phylogenetic relationships of the Genistoid sensu lato clade, as well as other papilionoids including many of those once placed in the tribes Swartzieae and Sophoraeae. The “core genistoids” (sensu Crisp et al., 2000) clade is indicated by bold lines. Nodes consistent with the presence of a 50-kb inversion in the plastid DNA genome are indicated by arrows. See Fig. 1 for details.

Of the 37 indel characters, 12 were synapomorphic for clades identified in the maximum parsimony and Bayesian analyses. For example, two single-amino-acid insertions, one at positions 421–423 and a second at positions 1498–1500, were synapomorphies for the papilionoid clade. Similarly, one-, two-, or three-amino-acid insertions/deletions uniquely mark each of the Sweetia-Vatairea clade (Fig. 2), New World Lupinus (Fig. 2), dalbergioid s.l. clade (Fig. 3), the genus Sesbania (Fig. 5), and the Caragana plus Hedysarum clade (Fig. 5). Inclusion of the indels as additional characters had little effect on phylogenetic relationships, based on comparison of the strict consensus topology derived from analysis of the data set containing the indel characters (data not shown) to that presented in Figs. 1–5. The exception involved Platycyamus regnellii which was resolved as sister to the clade defined by the MRCA of Apios americana and Phaseolus vulgaris (Fig. 4), in analyses that included the indel characters. Likewise, addition of the indel characters had little effect on bootstrap proportions for nodes receiving support in the 50% bootstrap consensus tree (<5% difference; data not shown).

Phylogenetic criteria for papilionoid clade nomenclature—We have used four criteria for recognizing and informally naming major clades within the Leguminosae, which are
consistent with formal node-based definitions under a system of phylogenetic nomenclature (de Queiroz and Gauthier, 1994). First, groups are resolved as monophyletic in strict consensus analyses. Second, bootstrap proportions greater than 70% support the clade of interest. Third, taxonomic sampling within the clade is diverse and/or extensive. Lastly, results are at least approximately congruent with that obtained by other studies (i.e., in showing support for clades that correspond to those informally recognized here). The following clade names are used throughout the discussion. The “Caesalpinioideae crown” clade includes all the Mimosoideae and members of tribes Caesalpinieae and Cassieae of subfamily Caesalpinioideae that form the sister group to them and is defined as the least inclusive clade that contains Ceratonia siliqua and Albizia julibrissin. The “papilionoid” clade is equivalent to the subfamily Papilionoideae and is delimited by the most recent common ancestor (MRCA) of Swartzia simplex and Vicia faba. Within papilionoids, the “Cladrastis” clade is delimited by the MRCA of Cladrastis platycarpa and Cladrastis lutea, which renders the genus Cladrastis paraphyletic with respect to Pickeringia and Styrpholobium (Fig. 2). The “genistoid s.l.” clade is delimited by the MRCA of Poecilanthe parviflora and Lupinus argenteus (Fig. 2). The “dalbergioid s.l.” clade comprises all descendants of the MRCA of Amorpha fruticosa and Pterocarpus indicus (Fig. 3). The “mirbelioid” clade is delimited by the MRCA of Daviesia latifolia and Gompholobium minus (Fig. 4). The “millettioid” clade is delimited by the MRCA of Xeroderris stuhlmannii and Phaseolus vulgaris (Fig. 4). The “robinioid” clade comprises all descendants of the MRCA of Robinia pseudoacacia and Lotus japonicus (Fig. 5). The robinioids and IRLC are sister groups and comprise “Hologalegina.”

DISCUSSION

The phylogenetic analyses of the legume matK sequences presented here achieve our main goal of reconstructing a robust molecular phylogeny for the Leguminosae, with the par-
Fig. 4. Phylogenetic relationships in Millettioide clade, Indigofereae, Hypocalycieae, Mirbelioide, and Baphioid clades; memberships are indicated by boxes. The “core Millettieae” clade (sensu Hu et al., 2000) is indicated by bold lines. Arrows are used to specify nodes with indicated support values. See Fig. 1 for details.

Fig. 5. Phylogenetic relationships in Millettioide clade, Indigofereae, Hypocalycieae, Mirbelioide, and Baphioid clades; memberships are indicated by boxes. The “core Millettieae” clade (sensu Hu et al., 2000) is indicated by bold lines. Arrows are used to specify nodes with indicated support values. See Fig. 1 for details.
by parsimony bootstrap (<50%) in our analyses, however. Regardless, the constituents and relationships of the Umtiza subclade within the Caesalpinoid crown clade detected in this study are in agreement with and further substantiate the findings of Herendeen et al. (2003a, b).

Although sampling of taxa from Mimosoideae was limited, our results generally agree well with those of Luckow et al. (2003), especially with respect to the monophyly of at least the vast majority of the genera traditionally assigned to the Mimosoideae and paraphyly of constituent tribes. With the exception of *Dinizia*, which appears more closely related to certain caesalpinoids than to mimosoids on the basis of morphological and molecular evidence, as concluded by Luckow et al. (2003), the rest of the taxa sampled from this subfamily are resolved as monophyletic with high support in our analyses (Fig. 1). Furthermore, our results clearly show *Piptadeniastrostrum* nested within the mimosoid clade, confirming recent results by Luckow et al. (2003). The mimosoid clade generally shows poorly resolved relationships or at least short internal branch lengths compared to other clades of Leguminosae (Fig. 6). This suggests either a slow down in the rate of substitution or a relatively recent diversification of most of the extant mem-
In contrast to the caesalpinioiids and mimosoids, our results have significant implications with regard to papilionoid phylogenetics. In all our analyses, papilionoids are strongly supported as monophyletic (Fig. 1) compared to previous rbcL studies where papilionoids were resolved as monophyletic but with relatively low statistical support (e.g., Kajita et al., 2001; 57% bootstrap and 62% parsimony jackknife). Similar to the findings of other studies involving broad sampling of caesalpinioioid legumes (e.g., Herendeen et al., 2003a), papilionoids are not resolved as sister to an isolated caesalpinioioid lineage, as are the mimosoids, but rather are nested among the major caesalpinioioid clades as an early branch in the legume phylogeny (Fig. 1). In marked contrast to the most recent rbcL analysis in which most major clades within papilionoids were weakly resolved (fig. 3 of Kajita et al., 2001), the matK strict consensus is very highly resolved (Figs. 2–5). Furthermore, both bootstrap proportions and Bayesian posterior probabilities for the major subclades often exceed 95%. The results presented here provide some of the best evidence to date in support of relationships among the major papilionoid subclades,
which heretofore have been largely unresolved by cladistic analyses of DNA sequences data.

Consistent with the results of Doyle et al. (1997) and Pennington et al. (2001), the \textit{matK} phylogeny resolves certain representatives of Swartziae and Sophoreae as the sister group to the rest of the subfamily. The clade that forms the sister group to all remaining papilionoids, here delimited by the MRCA of \textit{Swartzia simplex} and \textit{Myroaspernum soussanum} (Fig. 2), is unexpected in that it now includes representatives of a number of disparate lineages such as \textit{Angeleylocyxa} and \textit{Dipterygeae} (\textit{Dipteryx} and \textit{Pterodon}) that had been poorly resolved or supported in previous studies (e.g., Pennington et al., 2001). One of two subclades of this clade includes \textit{Swartzia} and recent segregate \textit{Bobgunnia}, \textit{Ateleia}, and \textit{Cyathostegia}, and corresponds to the "swartzzioid" clade of Ireland et al. (2000) and Pennington et al. (2001). The other contains \textit{Amburana}, \textit{Angeleylocyxa}, \textit{Dipterygeae}, \textit{Dussia}, \textit{Myrocarpus}, and \textit{Myroaspernum}. The resolution of this largest clade of morphologically eclectic genera as sister to the rest of Papilionoideae suggests that the swartzzioid clade of Pennington et al. (2001) could be expanded to encompass the majority of papilionoid genera that lack the 50-kb inversion in the plastid DNA genome.

With respect to the rest of the papilionoid subgroups, our sampling is much more extensive. The following seven well-supported clades resolved in this study are thus validated not only by extensive sampling, but also by the resolution of these subclades in other recent studies. These seven are the \textit{Cladrastis} clade, the genistoid s.l., the dalbergioid s.l., the millelioids, the millettioids, the robinioids, and the inverted-repeat-lacking clades, the last two of which comprise Hologalegina. Even if resolved by previous studies, relationships among these major papilionoid subclades have been heretofore resolved at best with only weak support (e.g., Hu et al., 2000; Kajita et al., 2001; Lavin et al., 2001; Pennington et al., 2001).

\textbf{The \textit{Cladrastis} clade}—The genera \textit{Cladrastis} and \textit{Styphnolobium} traditionally have been classified in Sophoreae s.s. (Polhill, 1981b) whereas \textit{Pickeringia} Nuttall has been classified in tribe Thermopsieae (Turner, 1981). These three genera form a well-supported clade in all our analyses. While a sister group relationship of \textit{Cladrastis} and \textit{Styphnolobium} has been observed in previous molecular studies (e.g., Doyle et al., 1997; Pennington et al., 2001) and is notable biogeographically because both genera exhibit East Asian–North American disjunctions, this study is the first to suggest a close relationship of these two genera with \textit{Pickeringia}. \textit{Pickeringia} is a monotypic genus restricted to the sclerophyllous chaparral vegetation of the California Floristic Province of western North America (Raven and Axelrod, 1995). The strongly supported position of this genus in the \textit{Cladrastis} clade confirms Polhill’s (1981b) initial prediction and Sousa and Rudd’s (1993) subsequent conclusion of a close relationship between these three genera based on floral (bracts at base of inflorescence) and chromosomal similarities ($n = 14$; Goldblatt, 1981; Palomino et al., 1993). This clade is also supported by results from cladistic analyses of nuclear ribosomal DNA internal transcribed spacers (nrDNA ITS) sequence data (M. F. Wojciechowski, unpublished data). This placement of \textit{Pickeringia} reveals that Thermopsieae sensu Yakovlev (Turner, 1981) is not monophyletic, contrary to molecular evidence presented previously (Crisp et al., 2000). Furthermore, the absence of quinolizidine alkaloids in \textit{Pickeringia} of the type that is characteristic of other Thermopsieae (Turner, 1981) is consistent with the \textit{matK} results. The presence of quinolizidine alkaloids, a prominent group of secondary metabolites once considered to be widely distributed in papilionoid legumes (Kinghorn and Balandrin, 1984), now appear to be of systematic significance only for the "genistoids" (see next paragraph). In a recent analysis (Kite and Pennington, 2003), the failure to detect similar alkaloids in extracts of \textit{Cladrastis} and \textit{Styphnolobium} is also in accordance with the phylogenetic position of these taxa based on \textit{trnL} sequence data (Pennington et al., 2001) and the result presented here. The disjunct distribution of the \textit{Cladrastis} clade in warm temperate to tropical regions of the Northern Hemisphere is common to many other legume groups (e.g., \textit{Gleditsia}, \textit{Gymnocladus}, \textit{Desmodium}, \textit{Lespedeza}, etc.; Schrire et al., in press).

\textbf{The genistoid s.l. clade}—The genistoids include the many genera traditionally classified in the tribes Genisteae, Thermopsieae, Euchrestaeae, Crotalarieae, Liparieae, Podalyrieae, and Sophoreae s.s. (Käss and Wink, 1997; Crisp et al., 2000). The concept of a "genistoid alliance" was circumscribed by Polhill (1981a, 1994) who brought together for the first time this group of putatively related, predominantly Southern Hemisphere tribes that have been considered relatively isolated among quinolizidine-alkaloid-accumulating papilionoid legumes. The alliance as recognized by Polhill comprises four separate lineages. One includes the predominantly Northern Hemisphere Genisteae sensu stricto (s.s.), Euchrestaeae, and Thermopsieae together with certain Sophoreae (\textit{Sophora} group). A second involves the mainly southern African Crotalarieae, Liparieae, and Podalyrieae, and now segregate tribe Hypocalypteae. A third comprises the endemic Australasian Bossiaeeae and Mirbelieae. The fourth includes the Neotropical–Australasian Bronniartieae (including the \textit{Templetonia} group). Early molecular phylogenetic analyses by Käss and Wink (1996, 1997) suggested that most species of the alliance formed a monophyletic group with some certain members of Sophoreae (i.e., some but not all species of \textit{Maackia} Rupe. & Maxim. and \textit{Sophora} L.) near the base of papilionoids. The analysis of Doyle et al. (1997) suggested the genistoids were polyphyletic and formed three clades, the largest of which approximates the genistoid clade of Käss and Wink. The monogeneric Euchrestaeae (\textit{Euchresta}) has been shown by subsequent analyses (Kajita et al., 2001) to be nested within a \textit{Sophora} s.s. clade, while Liparieae has been formally included within Podalyrieae (Schutte and van Wyk, 1998), a placement verified by analyses of \textit{rbcL} and \textit{nrDNA ITS} sequences (Käss and Wink, 1997; Kajita et al., 2001; van der Bank et al., 2002).

Crisp et al. (2000) confirmed the polyphyly of the genistoids sensu Polhill (1981a), but suggested that this name be restricted to a well-supported "core genistoids" group, from Africa and Eurasia, that comprises the majority of the tribes that made up Polhill’s genistoid alliance. This clade is strongly supported by the \textit{matK} data (Fig. 2; corresponds to the clade delimited by the MRCA of \textit{Bolusanthus speciosus} and \textit{Spattium junceum}). In addition, the \textit{matK} phylogeny corroborates results of other studies in resolving a core genistoid clade nested within a larger genistoid clade. A more inclusive group, referred here to as the genistoid s.l. clade and well supported by \textit{matK} sequence analysis (Fig. 2), includes the Bronniartieae (sensu Crisp and Weston, 1987; Thompson et al., 2001), \textit{Poeceillanth}e and \textit{Cyclolobium} of Millettieae (Hu et al., 2000, 2002), and a number of largely woody Neotropical genera of Sophoreae such as \textit{Acosmium} Schott, \textit{Bolusanthus} Harms, \textit{Brodichia}
Kunth, *Cadia* Forssk., *Diplotropis* Bentham, and most likely *Ormosia* Jackson (Kajita et al., 2001), *Dicraeopetalum* Harms, *Clathrotropis* Harms, and *Platycephylum* Harms (Pennington et al., 2001), several of which have not been sampled for *matK* sequences. The monophyly of the genistoid s.l. clade as defined here is also supported by the taxonomic distribution of quinolizidine alkaloids (e.g., *Kinghorn* and *Balandrin*, 1984; van Wyk, 2003). All taxa known to accumulate these alkaloids, with the exception of *Calia* (Kite and Pennington, 2003) and *Ormosia* (Kinghorn and Balandrin, 1984), are members of the genistoid s.l. clade as defined here. While the relationship of these particular taxa to this clade is not definitively resolved by our analyses (Figs. 2, 6), Pennington et al. (2001) did find weak support at least for the inclusion of *Osmosis* within an “expanded” genistoids. Further resolution and sampling of these taxa as well as *Holocalyx*, *Uribea*, and the vataireoids, may yet show quinolizidine alkaloids to be a non-molecular synapomorphy for an expanded genistoid clade.

**The dalbergioid s.l. clade**—The dalbergioid legumes, a mostly pantropical group of papilionoids, was originally circumscribed by a combined data approach to include 44 genera and ca. 1100 species from the tribes Aeschynomeneae, Adesmeae, subtribe Bryinae of Desmodieae, and tribe Dalbergieae except *Andira*, *Hymenolobium*, *Vatairea*, and *Vataireopsis* (Lavin et al., 2001). In addition, this clade is diagnosed apomorphically by the presence of the aescynomenoid root nodule (Sprent, 2001). Although the position of the dalbergioid clade within the Papilionoideae was not well resolved or supported in previous studies using *rbcL* and *trnL* (Kajita et al., 2001; Pennington et al., 2001), there was preliminary evidence that its sister group included the predominantly North American temperate tribe Amorpheae (Lavin et al., 2001). Our results, like those of McMahon and Hufford (2004), consistently show Amorpheae as the sole sister clade to the dalbergioid clade even if parsimony bootstrap support for this relationship is moderate (Fig. 3). Thus, the dalbergioid clade (sensu Lavin et al., 2001) is now expanded to encompass this tribe and is herein referred to collectively as the dalbergioid s.l. clade. The similarity in base chromosome number among dalbergioids and genera of Amorpheae, where *x* = 10 is apparently ancestral with derived cases of aneuploidy (e.g., *x* = 9 and *x* = 8; Goldblatt, 1981), supports this decision. Furthermore, the glandular punctate leaves and indehiscent, single-seeded pods (derived from a two-plus-ovulate ovary) of Amorpheae, once thought to indicate a relationship with the genera of Psoraleeae (e.g., *Stirton*, 1981), are found variously within dalbergioids. The predominantly tropical, woody Old World tribe Millettieae (Lavin et al., 1998; Hu et al., 2000, 2002; Kajita et al., 2001), with Indigoforae as its moderately supported sister group (Fig. 4). The predominantly tropical, woody Old World tribe Millettieae has been considered to be a transitional link from the “less advanced” elements of Dalbergieae and Sophoreae to putatively “more advanced” Old World tribes like Phaseoleae and Galegaeae (Geesink, 1981, 1984; Polhill, 1981a). These authors used the term “advanced” to indicate a high degree of fusion of stamens and keel petals, as well as the accumulation of nonprotein amino acids in seeds rather than alkaloids. Geesink (1984) went so far as to consider Millettieae the paraphyletic “stem” group from which all other “advanced” papilionoids branched. Early *rbcL* analyses (Doyle et al., 1997) suggested the polyphyly of both Millettieae and Phaseoleae, and this was subsequently confirmed using nuclear phytochrome gene sequences (Lavin et al., 1998), *trnK-matK* sequences (Hu et al., 2000), and a combined analysis of *rbcL*, *matK*, and *nrDNA* ITS data (Hu, 2000). Regardless, the emerging pattern of relationships derived from these studies and from this *matK* analysis is that most of the constituent genera of Millettieae and Phaseoleae clade fall out in two very well-supported subclades (Fig. 4). The first, previously referred to as “core Millettieae” (Hu et al., 2000), comprises the majority of Millettieae and includes the large genera *Millettia* Wight & Arn. (c.
150 spp.), Lonchocarpus Kunth (c. 130 spp.), Derris Lour. (50–60 spp.) and Tephrosia Pers. (c. 350 spp.), while the majority of Phaseoleae dominates the second. Although shown here (Fig. 4) as sister lineages to core Millettieae, relationships of the monogeneric Abreae and certain Phaseoleae such as *Galactia* and presumably the other genera of subtribes Galactineae and Dioiclineae (see Lewis et al., in press), are not yet resolved with certainty with respect to the core Millettieae clade. Additionally, certain genera classified in Millettieae, including *Xeroderris* and *Platycaamus*, have unresolved or weakly supported relationships with respect to the two main millettioid clades.

The large Phaseoleae clade resolved by *matK* includes tribes Desmodieae (except subtribe Bryinae) and Psoraleeae, in agreement with other studies (e.g., Kajita et al., 2001; Hu et al., 2002). Lackey’s (1981) subtribal classification of Phaseoleae, the largest tribe of legumes in number of genera, is not congruent with the monophyletic subclades detected in this and other analyses (e.g., Kajita et al., 2001). The delimitation of the subtribe Phaseolineae is restricted to the descendants of the MRCA of *Wajira* and *Phaseolus* (Fig. 4), and excludes other genera once included, such as *Psophocarpus* and *Ototeura*, which are now known to be more closely related to genera in *Glycininae* (Thulin et al., in press; A. Delgado-Salinas, M. Lavin, M. Thulin, and N. Weeden, unpublished data). In spite of the findings of Lee and Hymowitz (2001), *Glycininae* is also not monophyletic and includes certain genera of subtribe Phaseolineae and probably all genera of tribe Psoraleeae (Kajita et al., 2001; A. Delgado-Salinas, M. Lavin, M. Thulin, and N. Weeden, unpublished data).

In this and most previous analyses (e.g., Kajita et al., 2001), tribe Indigofereae emerges as the moderately supported sister group to the millettioid clade (Fig. 4). The genera of Indigofereae have an inflorescence of the simple racemose type (each node bearing a single flower), whereas the tribes Abreae, Desmodieae, Millettieae, Phaseoleae, and Psoraleeae, which form the millettioid clade, share an unusual type of inflorescence, the pseudoraceme (see Tucker, 1987a, b). Indeed, the pseudoracemose inflorescence is found only and in all members of the millettioid clade, rendering it the most readily morphologically distinguished of the newly circumscribed large subclades of Papilionoideae. In addition, typical chromosome numbers in Indigofereae (*n* = 8, 7, 6; Goldblatt, 1981) are different from that which predominate in the tribes comprising the millettioids (e.g., *Millettieae*, *Phaseoleae*: *n* = 11, 12; Goldblatt, 1981). Whether the millettioid clade should be expanded to include the Indigofereae is as yet undetermined, especially given its sister relationship is only moderately supported, but the lack of any nonmolecular evidence to support such a relationship argues against this.

**The robiniod clade**—Tribes Loteae (including Coronilleae; sensu Polhill, 1994) and Robinieae (sensu Lavin and Sousa, 1995) comprise an expanded robiniod clade, which is distributed primarily in the Northern Hemisphere of the New World, Europe, and Africa. Within this clade, a monophyletic *Sesbania* L. (Robinieae) is weakly supported as sister to Loteae and these collectively form the sister group to the remaining members of Robinieae (Fig. 5). This position of *Sesbania* thus renders Robinieae paraphyletic, a finding first suggested by a preliminary phylogenetic analysis of *matK* sequences (Wojciechowski et al., 2000) and confirmed with a recent study utilizing exhaustive sampling for *matK*, *trnL* intron, and nrDNA ITS sequences (Lavin et al., 2003). Allan et al. (2003) have provided molecular evidence for the monophyly of the largely Eurasian–North American Loteae and the paraphyly of the large genus *Lotus*. The robiniod clade is here expanded from that described by Lavin et al. (2003) to encompass *Sesbania* and Loteae. Multiple lines of molecular evidence strongly support the monophyly of this collective group. For example, surveys for the presence of the inverted repeat in plastid DNA genomes among papilionoid legumes first suggested Loteae was distinct from the other temperate herbaceous tribes (Lavin et al., 1990) and its sister group relationship to *Sesbania* was only more recently determined (Wojciechowski et al., 2000).

**The IRLC**—The inverted-repeat-lacking clade or IRLC (Wojciechowski et al., 1999, 2000) includes most members of Polhill’s (1981a) temperate herbaceous group. This group comprises all members of tribes Carmichaelieae, Cicereae, Hedysareae, Trifolieae, Vicieae, and Galegeae, but not Loteae (and Coronilleae), as well as at least three genera of Millettieae, *Afgekia* Craib, *Callerya* Endlicher, and *Wisteria* Nutt. The IRLC was essentially the first clade of legumes to be distinguished on the basis of a molecular synapomorphy, loss of one copy of the 25-kilobase inverted repeat in the plastid genome (Lavin et al., 1990; Liston, 1995), in addition to a number of morphological features shared by members of this group including a predominantly herbaceous habit, epulvinate compound leaves, and base chromosome numbers of *n* = 7 or *n* = 8 (Polhill, 1981a). The monophyly of the IRLC has been consistently detected in all subsequent cladistic analyses of molecular sequence data (Sanderson and Wojciechowski, 1996; Doyle et al., 1997; Käss and Wink, 1997; Hu et al., 2000, 2002; Wojciechowski et al., 2000; Kajita et al., 2001). Within the IRLC, the genera *Afgekia* (not included in this analysis), *Callerya*, and *Wisteria*, formerly classified in the tribe *Millettieae*, along with *Glycyrrhiza* L. of Galegeae, form a paraphyletic grade with respect to the remaining IRLC (Fig. 5). The “vicioid” subclade of the IRLC includes many of the particularly important agricultural genera such as *Cicer* L., *Lathyurus* L., *Lens* Mill., *Medicago* L., *Mellilotus* Mill., *Pisum* L., *Trifolium* L., and *Vicia* L. (Wojciechowski et al., 2000; Steele and Wojciechowski, 2003). The vicioids are morphologically the most distinctive subclade of the IRLC, apomorphically characterized by craspedidromous leaflets and consistently well supported as monophyletic with *Paroquetus* (Trifoliae) as the sister to all other vicioid taxa (Fig. 5). Other major clades within the IRLC include the Astragalale clade (Wojciechowski et al., 1999, 2000), defined as all descendants of the MRCA of *Astragalus americanus* and *Cistanthus punicus* (Fig. 5), and the hedysaroid clade (Sanderson and Wojciechowski, 1996; Wojciechowski et al., 2000), defined as all descendants of the MRCA of *Hedysarum boreale* and *Cara-gana arborescens* (Fig. 5).

**Hologalegina**—The robiniods and the IRLC comprise the largest of the well-marked papilionoid subclades, Hologalegina (Wojciechowski et al., 2000). This clade includes over 4800 species that make up the vast majority of legumes presently distributed in temperate regions of the world. The robiniods-IRLC dichotomy had been detected in studies with *rbcL* (Doyle et al., 1997; Kajita et al., 2001) but resolution was weakly supported. Furthermore, the placement of *Bolusanthus* (Sophoreae) as the sister group to Robinieae in those studies stands in marked disagreement with the results presented here.
and in most other studies that consistently place *Bolusanthus* within the genistoids (Lavin et al., 1998; Hu et al., 2000, 2002; Pennington et al., 2001). Hologalegina and its two subclades are very strongly supported in our *matK* analyses (Fig. 5). Relationships within each of the subclades have received considerable support from analysis of *matK* sequences, and the circumscription of several of the subclades has been the subject of recent studies (Wojciechowski et al., 2000; Allan et al., 2003; Lavin et al., 2003).

**Minor papilionoid clades**—One of the surprising and well-supported papilionoid subclades to be identified in the *trnL* analyses of Ireland et al. (2000) and Pennington et al. (2001), the vataireoid clade, also finds strong support in our analyses (i.e., the clade with *Sweetia*, *Luetzelburgia*, and *Vatairea*; Fig. 2). The close relationship of *Exostyles*, *Harleypodendron*, *Luetzelburgia*, *Sweetia*, *Vatairea*, and *Vataireopsis* suggested by Pennington et al. (2001) had not been recognized in formal classifications but was suspected on the basis of overall morphology and wood anatomical similarities (e.g., de Lima, 1990; Gasson, 1994). The inclusion of *Exostyles* and *Harleypodendron* in the vataireoid clade (Pennington et al., 2001) was particularly unexpected given the placement of these genera in the tribe Swartzieae because of their very nonpapilionoid floral morphology. As Pennington et al. (2001) point out, members of the vataireoid clade lack the ability to nodulate with symbiotic rhizobia, thus revealing the existence of more than one well-defined subroup of woody papilionoid legumes that are not known to form root nodules (Sprent, 2001).

Other well-supported clades within papilionoids with unresolved relationships to the major subclades include one comprising *Holocalyx* and *Uribea* and another represented by *Sophora* sect. *Calia* (Fig. 2). *Holocalyx* and *Uribea*, both monotypic genera of trees from Central America and tropical South America, have been treated in Sophorae although *Holocalyx* has been recently transferred to Swartzieae (Polhill, 1994). The placement of *Calia arizonica* as sister to *C. secundiflora* and distinctively separate from both *Styphnobium* (of the *Cladrastis* clade) and *Sophora* L. s.s. (of the genusitc s.l. clade) lends further credence to Sousa and Rudd’s (1993) distinction of both the North American *Calia* Berland. and predominantly Mesoamerican *Styphnobium* Schott as separate from *Sophora* s.s.

Our analyses also provide support for the baphioid clade of Pennington et al. (2001) as the sister group to the mirbelioid clade plus Hypoclypteae (Fig. 4). Represented here by two species of *Baphia*, the baphioid clade corresponds closely with the *Baphia* group of Polhill (1981b), a small group (c. 60 spp.) of largely African and southern Asian Sophorae that includes *Baphia*, *Airyantha*, *Baphiastrum*, *Bowringia*, *Dalhouisia*, and *Leucomphalos*, plus tropical African *Baphiopsis* of Swartzieae (Polhill, 1994).

**The 50-kb inversion clade**—A 50-kb inversion within the large, single-copy region of the plastid genome is a structural rearrangement that Doyle et al. (1996) suggested is a synapomorphy for a clade that includes most papilionoids (Fig. 2). Other molecular phylogenetic studies have noted that certain genera of Sophorae, Swartzieae, and Dipterygaceae known to lack the 50-kb inversion have unresolved relationships within the papilionoid clade (e.g., Doyle et al., 1997; Pennington et al., 2001). Consistent with the *trnL* sequence analysis of Pennington et al. (2001), *matK* sequences also resolve a clade consistent with the distribution of this inversion, albeit with weak support (Fig. 2). The precise placement of the MRCA of this inversion clade in our phylogeny is rendered uncertain because it is not yet known whether *Calia* has this rearrangement in its plastid genome. However, every taxon known to possess the 50-kb inversion, based on the taxa sampled by Doyle et al. (1996) and R. T. Pennington (Royal Botanic Garden, Edinburgh, Scotland, unpublished data), is nested in the clade that is sister to the *Cladrastis* clade (Fig. 2). Furthermore, the greater nodal support and presence of quinolizidine alkaloids in *Calia* (Kite and Pennington, 2003) argue for the more inclusive clade, i.e., MRCA of *Calia secundiflora* and *Vicia faba*, being marked by the presence of this inversion. To characterize this node more precisely, greater sampling of *matK* sequences of genera putatively branching from near the presently detected 50-kb inversion node (Fig. 2) is required, as are experiments designed to detect the presence or absence of the 50-kb inversion in these taxa (cf. Doyle et al., 1996).

**The canavanine-accumulating clade**—The *matK* phylogeny supports a large clade, distinguished by the ability to produce the nonprotein amino acid canavanine and defined by the MRCA of the mirbelioid clade and the IRLC. This clade is poorly supported by bootstrap proportions, but moderately supported by Bayesian posterior probabilities and its presence in the parsimony strict consensus (Fig. 4). A hypothesis for the single origin of canavanine biosynthesis was put forth by Bell (1981), who showed that canavanine, a close analog of arginine, was essentially mutually exclusive of alkaloid accumulation and restricted to 16 closely related papilionoid tribes (notwithstanding a dubious report from *Laburnum anagyroides*, Genistae; Wink and Mohamed, 2003), all of which group here in what we term the “canavanine-accumulating clade” (Fig. 4). However, at that time, the close relationship of the tribes Bossiaeae, Mirbeliaeae, and Hypoclypteae (formerly synonymized under Liparieae; Polhill, 1981c) to other tribes accumulating canavanine was suspect. This was because the original circumscriptions of these tribes suggested they were polymorphic for degrees of floral fusion and canavanine or alkaloid accumulation and thus were possibly intermediate between the canavanine- and alkaloid-accumulating tribes (see Fig. 4 or Polhill, 1981a, where this primary division of papilionoid tribes is centered on “*Tephrosieae*”). Notably, the *matK* results presented in this analysis place the mirbelioid clade as the sister group to the rest of the canavanine-accumulating clade. Once the taxonomy of the constituents of the current mirbelioid clade and its sister group was resolved (i.e., Hypoclypteae segregated from Liparieae, Bossiaeae stripped of the *Templetonia* group; see Crisp et al., 2000), it has become clear that the pattern of canavanine accumulation in papilionoid legumes is mutually exclusive of alkaloid accumulation and thus appears to have evolved only once along the branch subtending the MRCA of the mirbelioid clade and IRLC.

Pennington et al. (2001) and Kajita et al. (2001) did not resolve a group corresponding to this canavanine-accumulating clade in the detail shown in the present analysis. Most importantly, the mirbelioid clade was not resolved as a lineage within this canavanine-accumulating clade while the “baphioid clade” of Pennington et al. (2001), apomorphically diagnosed by unifoliolate leaves, was not resolved as the sole sister group to the canavanine-accumulating clade, as they are in this *matK* analysis (Fig. 4). Pennington et al. (2001) do
show a strongly supported baphioid clade nested within what would otherwise be the canavanine-accumulating clade, while resolution with \textit{rbcL} was so poor that the canavanine accumulators are represented by at least three clades nested among several alkaloid-accumulating lineages (e.g., genistoids in part, baphioids, the genus \textit{Amorpha}; fig. 3 of Kajita et al., 2001). In sum, the detection in this analysis of the sister relationship of the baphioid clade to that of the canavanine-accumulating clade is additional evidence for how informative the matK locus is for phylogenetic analysis at these hierarchical levels.

\textbf{Correspondence of the matK phylogeny with Polhill's tribal classification and its implications for legume evolution—}

The taxonomic implications of the results of this and other recent, especially molecular, phylogenetic studies of legumes are significant. Clearly, if a higher-level classification of the entire family is to reflect phylogenetic relationships (Fig. 6), important changes will be necessary. For example, of the papilionoid tribes with two genera or more circumscribed by Polhill (1994) only Amorphaeae and Dipterygeae, and perhaps Bossiaeae and Psoraleae, are monophyletic. Sophoreae and Swartzzieae are clearly polyphyletic, based on molecular and morphological data (Herendeen, 1995; Doyle et al., 1997; Pennington et al., 2001), and dispersed throughout several papilionoid subclades, both outside and inside the 50-kb inversion clade (e.g., the baphioids, \textit{Maackia}, \textit{Bolusanthus}, and \textit{Sophora} s.s. in the genistoids, or the vataireoids). The nonmonophyly of such groups of genera (Figs. 2, 4), especially those classified in Swartzzieae, reveals that many morphological features once considered distinctive (e.g., the swartzioid valvate calyx that bursts open at anthesis) are more prone to evolve independently than previously thought. Thus, Swartzzieae can no longer be viewed as a “transitional” lineage (cf. Polhill, 1994) between Caesalpinoideae and Papilionoideae if its component lineages are thoroughly interdigitated among those of Papilionoideae or portrayed as sufficiently distinct from either Papilionoideae or Caesalpinoideae to warrant separate subfamily status (e.g., Tucker, 2003b). Likewise, the large tropical tribe Millettiaeae, once considered the paraphyletic stem clade from which many of the temperate and tropical papilionoid groups in the Old World were derived (Geensik, 1984), comprises three lineages that belong to distantly related groups, the genistoids, meliittiods, and IRLC (Hu et al., 2002). Lastly, the morphologically diverse and cosmopolitan tribe Galegaeae is polyphyletic within the IRLC, which comprises the bulk of temperate legume diversity (Wojciechowski et al., 2000). Comparable examples of incongruence between current classification schemes for the other subfamilies and the results from recent molecular and morphological studies have been presented elsewhere (e.g., Caesalpinoideae: Bruneau et al., 2001; Mimosoideae: Luckow et al., 2003).

In contrast to the described examples in which one traditionally recognized tribe is now known to comprise several distinct lineages, multiple traditionally recognized tribes or subtribes are now known to form a single well-marked clade. The pantropical dalbergioid clade includes most genera of Dalbergiaeae and all of Aeschynomeneae, as well as the subtribe Bryinae of tribe Desmodieae. The genera of Dalbergiaeae are polyphyletic within this dalbergioid clade (Lavin et al., 2001). This composition of heretofore-eclectic genera is made even more notable when its sister group, the primarily North American Amorphaeae, is included (McMahon and Hufford, 2004). Similarly, the primarily north temperate Loteae plus North American Robiniae form the strongly supported sister clade to the largely temperate and herbaceous IRLC. The pantropical \textit{Sesbania}, previously of Robiniae, is weakly but consistently resolved as sister to Loteae. The finding of a close relationship of the South American genus \textit{Anarthrophyllum} to \textit{Lupinus} and related Genisteae and that of \textit{Hypocalyx} to the Australasian Mirbeliaeae plus Bossiaeae, rather than to the South African tribes such as Croalariaceae, is consistent with the distribution of quinolizidine alkaloids in these taxa (Van Wyk et al., 1993; van Wyk and Schutte, 1995).

While the new informal classification system proposed here for papilionoid legumes bears some resemblance to Polhill’s tribal classification, many genera have relationships that are novel and in some cases not predicted by previous classification schemes. This incongruence may in part be caused by the conventional perception of tribes (and subfamilies) as occupying “basal” or “primitive” vs. “derived” or “advanced” positions within the legume phylogenetic tree (e.g., Pennington et al., 2000, 2001; Kajita et al., 2001; Herendeen et al., 2003a; Luckow et al., 2003; Tucker, 2003a; Wojciechowski, 2003, to list a few recent examples where such terminology is still used). Terms like “basal” have been used in the legume literature to interpret morphologies, in addition to relationships, from an evolutionary perspective. Traditional legume classifications such as Polhill’s (1981a, 1994) invoked certain suites of morphologies that served a priori as a measure of divergence of an extant taxon from the MRCA of legumes. In contrast, the comprehensive and well-resolved \textit{matK} phylogeny reveals relationships in terms of sister clades and supports none of the previous ideas that invoked terms like “basal” and “derived.” For example, floral ontogenetic studies once suggested that radial or actinomorphic floral symmetry and suppression of floral organs, observed in taxa such as \textit{Cercis} and \textit{Gleditsia}, reflected minimal divergence from the MRCA of all legumes (e.g., Tucker, 1992; Tucker and Douglass, 1994), while the perfect, bilaterally symmetric flower of \textit{Cercis} suggested it was advanced and thus phylogenetically nested within the papilionoid line (Tucker and Douglass, 1994). Like recent studies using different phylogenetic data (Pennington et al., 2000, 2001; Tucker, 2002; Herendeen et al., 2003a), the \textit{matK} phylogeny reveals no tendency of genera with radial floral symmetry to occupy a basal position or form a paraphyletic grade in which is nested a clade(s) of taxa with bilateral floral symmetry. For example, all molecular analyses are unequivocal in supporting \textit{Cercis} in Ceridaeae as the sister clade to the rest of legumes (Bruneau et al., 2001; Kajita et al., 2001) with \textit{Ceratonia} and \textit{Gleditsia} part of the \textit{Umtiza} clade (Herendeen et al., 2003b) nested inside the caesalpinioideae crown clade from which mimosoids were derived (Fig. 1). The implications of results such as these for hypotheses of floral evolution in caesalpinoids or early branching lineages of papilionoids have been pointed out by others (e.g., Pennington et al., 2000; Bruneau et al., 2001) and need not be discussed further here. Similarly, the correlation of floral connotation, especially of the staminal filaments and keel petals, and accumulation of nonprotein amino acids in seeds marked the putative advanced tribes (e.g., Polhill, 1981a; Doyle et al., 1997). The \textit{matK} phylogeny suggests that the alternative states, flowers with free petals and filaments and accumulating alkaloids in seeds, are not a priori an indicator of little divergence from the MRCA of legumes. The genistoids produce most of the distinctive legume alkaloids, and genera with free stamens can be nested within a fused stamen clade (e.g., \textit{Adesmia}). It is
indeed likely that terms such “basal,” “primitive,” “derived,” and “advanced” carry some, if not many, unwarranted implications from previous classification schemes and will have to be abandoned in order to approach legume phylogeny and taxonomy from an open, more contemporary perspective.

LITERATURE CITED


