Systematic Botany (2009), 34(1): pp. 115–128

Relationships Among Phaseoloid Legumes Based on Sequences from Eight Chloroplast Regions

Saša Stefanović,[1,6] Bernard E. Pfeil,[2,6] Jeffrey D. Palmer,[3] and Jeff J. Doyle[4,5]

[1]Department of Biology, University of Toronto, Mississauga, Ontario L5L 1C6 Canada
[2]CSIRO & Centre for Plant Biodiversity Research, GPO Box 1600, Canberra ACT 2601 Australia
[3]Department of Biology, Indiana University, Bloomington, Indiana 47405 U.S.A.
[4]Department of Plant Biology, 228 Plant Science Building, Cornell University, Ithaca, New York 14853 U.S.A.
[5]Author for correspondence (jjd5@cornell.edu)
[6]These authors contributed equally to this study

Communicating Editor: Alan W. Meerow

Abstract—Generic level relationships in phaseoloid legumes have received much attention using chloroplast DNA markers. However, despite this attention not all relationships are yet well-resolved. This study includes *trnL-F* sequences from across a wide sample of phaseoloid legumes as well as seven additional chloroplast DNA loci (*rbcL, atpB, trnK/matK, rpl2, clpP, rps16, and ycf4*) analyzed separately and in combination. Together, these data provide support for many relationships generally consistent with, but only weakly supported, in earlier studies. Some major discordant phylogenetic results were found in our separate analyses; for example, *ycf4* sequences group *Glycine* and *Teramnus* with strong support; however, the combined analysis of the remaining seven loci found incongruent groupings (*Glycine* and Psoraleeae genera; *Teramnus* and *Amphicarpaea*) also with strong support. Network analysis of *ycf4* revealed that the conflicting signal (relative to the other seven loci) came from first and second codon positions. These positions also showed significant rate acceleration, together indicating that selection driving convergent molecular evolution is the likely cause of the signal in *ycf4*, rather than shared history. The major clades within the phaseoloid legumes supported by our analysis are discussed.

Keywords—Chloroplast DNA, Fabaceae, Incongruence, Leguminosae, molecular dating, Phaseoleae, Phylogeny, Psoraleeae, Desmodieae.

"Phaseoloid" legumes comprise over 100 genera and more than 2,000 species (Lewis et al. 2005), comprising many familiar and economically important members of the Leguminosae. Those primarily include pulses such as *Glycine max* (soybean), *Phaseolus* spp. (common bean, scarlet runner bean), *Vigna* spp. (cowpea, mungbean), and *Cajanus cajan* (pigeonpea), as well as some important forage plants such as *Lespedeza* and *Desmodium* (Simpson and Ogorzaly 2001). The phaseoloid subclade includes most genera classified traditionally in the tribe Phaseoleae (Lackey 1981; Polhill 1994; Lewis et al. 2005), but also includes the entire tribe Psoraleeae and most of the Desmodieae as traditionally recognized (Lewis et al. 2005). Additional genera of Phaseoleae are members of the millettioid subclade [*Ophrestia* and most members of Lackey's (1981) subtribe Diocleinae].

A number of molecular phylogenetic studies have addressed aspects of generic-level relationships in the phaseoloid group. Early studies using chloroplast DNA (cpDNA) restriction maps (Bruneau et al. 1990; Doyle and Doyle 1993) demonstrated the polyphyly of Phaseoleae and its subtribes. A series of phylogenetic analyses of the entire Leguminosae using the chloroplast gene, *rbcL*, culminated in a study that emphasized phaseoloid-millettioid taxa (Kajita et al. 2001), with 39 phaseoloid genera included. More recent comprehensive analyses of the family using chloroplast *trnK/matK* included representatives of 21 genera of this group (Wojciechowski et al. 2004). Studies addressing relationships in the core millettioid sister clade have included smaller numbers of phaseoloids (Lavin et al. 1998; Hu et al. 2000; 2002). In addition, several molecular phylogenetic studies have focused on specific groups within the phaseoloids. For example, relationships of *Glycine* and allies have been studied with chloroplast *rps16* (Lee and Hymowitz 2001) and glutamine synthetase (Doyle et al. 2003) sequences, and there have been several studies emphasizing members of the Phaseoleae subtribe Phaseolinae (*Phaseolus* and allies: Riley-Hulting et al. 2004; Thulin et al. 2004; Espert et al. 2007).

Despite the attention the group has received, obtaining consistent and well-resolved relationships among its members has proven difficult. Monophyletic groupings that correspond well to some taxa, such as the tribes Desmodieae and Psoraleeae, and Phaseoleae subtribes Phaseolinae and Cajaninae have been consistently observed. However, relationships among these well-defined groups have been largely unresolved and at best weakly supported. In addition, the placements of other genera have been more problematic, particularly those classified as Phaseoleae subtribe Glycininae, the group that includes the soybean and its allies. The only chloroplast gene sequence study that specifically addressed relationships among Glycininae, Lee and Hymowitz (2001), used Polhill's (1994) traditional circumscription of that subtribe and therefore did not include Psoraleeae, a group that other phylogenetic studies have shown to be closely related to *Glycine* (Kajita et al. 2001; Wojciechowski et al. 2004). Adding to the confusion, relationships of Glycininae genera also appear to differ markedly between chloroplast and nuclear topologies (Doyle et al. 2003).

It is important to resolve relationships among these taxa to provide a framework for understanding the evolution of organellar genomes in the group. Several deletions and rearrangements of chloroplast sequences have been characterized in phaseoloid legumes have in some cases provided characters suggesting relationships among these taxa (Bruneau et al. 1990; Doyle et al. 1995; Bailey et al. 1997). In addition, the process of gene transfer from the mitochondrial genome to the nuclear genome has also been studied in phaseoloids, notably the gene for cytochrome oxidase subunit 2 (*cox2*: Nugent and Palmer 1991; Adams et al. 1999; Daley et al. 2002). Several disparate phaseoloid legumes were shown to retain intact and expressed *cox2* genes in both mitochondria and nuclei, and both genes were found to be lost or silenced equally frequently, leading to the hypothesis that the likelihood of *cox2* inactivation is independent of its compartmental location (Adams et al. 1999). To address this hypothesis

as well as the additional questions about the fixation, redundancy and persistence of both *cox2* copies, a well-resolved and robust phylogenetic framework for this group of legumes is of great importance.

We report here results of phylogenetic studies on the multiple cpDNA sequences of phaseoloid legumes and compare them with previous taxonomic treatments. Our results provide strong support for many relationships that were either unresolved or weakly supported with smaller datasets. Additionally, we discuss in detail the origin(s) and relationships of the polyploid genus *Glycine* (soybeans) as well as diversification divergence times of phaseoloids.

Materials and Methods

***Taxon sampling*—**We initially conducted a survey of the *trnL-F* region for 79 taxa (Appendix 1), including sampling of multiple species from several genera of particular interest in understanding mitochondrial genome evolution (S. Stefanović and J. D. Palmer, unpubl. data). Based on these results we conducted a study of 33 genera, all but one shown to be monophyletic with *trnL-F*, by concatenating six to seven additional chloroplast regions to produce a supermatrix. *Pueraria*, *Lespedeza*, and *Desmodium* were the only genera for which more than one species was included in this second matrix, for a total of 36 species. Multiple representatives of *Pueraria* were kept because this genus is known to be polyphyletic (Lee and Hymowitz 2001) and of the other two genera because of their interesting mitochondrial genome evolution (Adams et al. 1999; Stefanović et al. unpubl. data).

***Molecular Techniques*—**Total genomic DNA from silica-dried or fresh material was extracted using a modified CTAB technique from Doyle and Doyle (1987) and purified using the QIAquick® purification kit (Quiagen, Valencia, California) or by ultracentrifugation in CsCl-ethidium bromide gradient (Sambrook et al. 1989). The polymerase chain reaction (PCR) was used to obtain the double-stranded DNA fragments of interest. The chloroplast (cp) genome was targeted with primers described by Taberlet et al. (1991) for the *trnL-F* region (including the *trnL* intron and *trnL-trnF* spacer), Olmstead et al. (1992) for the *rbcL* gene, Hoot et al. (1995) for the *atpB* gene, Graham and Olmstead (2000) for the *rpl2* gene (including its intron where present), Hu et al. (2000) for the *trnK/matK* region (the *trnK* intron including the *matK* gene), Lee and Hymowitz (2001) for the *rps16* intron, and Stefanović et al. (2004) for *clpP* gene (including both introns) and *ycf4*. PCR was carried out in 50 µL volumes with annealing temperatures of 50–55°C. Amplified products were separated by electrophoresis using 0.8% agarose gels, visualized with ethidium bromide, and cleaned by QIAquick® columns (Quiagen) or by polyethylene-glycol/NaCl precipitations. Cleaned products were then directly sequenced using the BigDye™ Terminator cycle sequencing kit (PE Applied Biosystem, Foster City, California) on an ABI 3100 DNA automated sequencer (PE Applied Biosystem). Sequence data were proofed, edited, and contigs assembled using Sequencher™ v.4.1 (Gene Codes Corporation, Ann Arbor, Michigan). Sequences generated in this study are submitted to GenBank (accesstion numbers EU717220-EU717531; see Appendix 1).

***Phylogenetic Analyses*—**Each of eight chloroplast regions was aligned manually using Se-Al v.2.0a11 (Rambaut 2002). Although gaps in the alignments were treated as missing data, insertions/deletions (indels) were found to be especially informative as phylogenetic characters in chloroplast data sets for one of our ingroup taxa (tribe Psoraleeae; Egan and Crandall 2008). We coded 20 indels as binary characters and appended them to the concatenated sequence matrix. Indel coding was conservative, in that complex gaps in the alignment were excluded entirely from the analyses and that single base repeats as well as indels near large complex gaps were not coded. Complex indels inferred to be homologous were identical in all but one case (indel 10) where a single substitution was inferred to have occurred after a shared 7-base pair insertion (this additional change was not reflected in the coding). Alignments are available in Nexus format from TreeBASE (study number S2150).

Parsimony analyses were conducted in PAUP* 4.0 b10 (Swofford 1998). For the 79-taxon analysis, based on the *trnL-F* sequences only, 100 random taxon addition sequences (RAS) were conducted with tree bisection and reconnection (TBR) branch swapping, and MAXTREES set to increase without limit. To assess the bootstrap support (BS), one hundred bootstrap replicates were conducted, using TBR branch swapping and MAXTREES set to 100. Because all eight sequenced regions used in this study occur in the effectively haploid chloroplast genome and their histories are linked (see Doyle 1992; Moore 1995), there is no a priori reason to believe that eight individual resulting gene trees will differ. However, their patterns of evolution might be different (e.g. differences in rates of evolution and/or base composition), leading to the incongruence among datasets (Bull et al. 1993). To account for these possibilities, we first conducted separate analyses of individual genes for the 36-taxon dataset. Parsimony settings for these 36-taxon analyses were identical to those described for the 79-taxon *trnL-F* analysis, except that for *clpP, trnK/matK, trnL-F,* and *ycf4* bootstrapping MAXTREES was set to increase without limit.

Subsequently, parsimony analyses were also conducted with a concatenated 36-taxon dataset comprising all eight regions (as well as with seven regions, excluding the incongruent *ycf4* sequences; see below). For these analyses, 1,000 RAS were run, using TBR branch swapping, with MAXTREES set to increase without limit; internal support was estimated by 1,000 bootstrap replicates, with 10 RAS each, TBR branch swapping, and MAXTREES allowed to increase without limit.

Two Bayesian analyses were conducted on the concatenated datasets: one with and one without the inclusion of *ycf4* sequences. For both of those, the data were split into three partitions containing coding, noncoding, and indel characters respectively. MrModeltest (Nylander 2004) was used to determine the best fitting model for the first two partitions among those models available in MrBayes, but using a parsimony-derived tree rather than the default NJ tree. The parsimony tree (not shown) was one of two best trees found following a 1000 RAS replicate heuristic search where a maximum of 100 trees were kept at each replicate. Both Akaike Information Criterion (AIC) and the hierarchical likelihood ratio test (hLRT) methods suggested that the GTR + I + G model was the best fit for both coding and noncoding partitions. We arbitrarily chose the JC + G model for the indel partition to reflect our uncertainty in the relative probabilities of indel events but to allow for the possibility that some indels are changing more rapidly than others. We used the coding = variable setting for the indel partition in addition to the model mentioned above because all characters in this partition were informative, along with standard format coding (0, 1 states). All shared parameters were unlinked between partitions: alpha, the rate matrix, state frequencies and the proportion of invariant sites. In all cases we used the default priors set by MrBayes.

We ran the Bayesian analyses in MrBayes version 3.1.1 (Ronquist and Huelsenbeck 2003) using 10 chains, and examined the likelihood plot to check for convergence among two replicate runs. We also examined sensitivity to model choice by using simpler models that still capture some of the most commonly observed aspects of molecular evolution. To this end, we employed: (1) the HKY + I + G model for both coding and noncoding data (alpha unlinked) and (2) the HKY + G for coding and GTR + G for noncoding data (alpha and state frequencies unlinked), with the same model for the indel characters previously used.

After initially finding discordant phylogenetic results, we also examined *ycf4* with Splits Tree 4 (Huson and Bryant 2006). We examined first and second versus third positions using uncorrected "p"-distances with Neighbor-Net (Huson and Bryant 2006) to try and localize the apparent different phylogenetic signal discovered during other analyses (see Results). We tested for selection by testing relative rate differences (Tajima's test) using the first two codon positions or the third codon position in MEGA 3.1 (Kumar et al. 2004). Mutations in many first and all second positions result in nonsynonymous changes.

***Molecular Dating*—**Examination of the Bayesian phylogenies suggested large differences in rates of molecular evolution among clades in the phylogeny. Using ML in PAUP*, we employed two tests to examine whether the lack of a molecular clock (H_a) could explain the data better than an enforced clock (H_o). In the first test, the matrix and the Bayesian consensus tree from the eight-gene analysis were loaded into PAUP* as a constraint tree. A GTR + I + G model was selected with six rate categories of the gamma distribution to provide extra rate flexibility, given that ML in PAUP* 4.0 b10 does not allow data partitioning at present. Three outgroup taxa were pruned from the tree to leave an unambiguous root placement for the molecular clock optimization (pruned taxa matched those pruned for the r8s analysis; see below). The likelihood score was determined on the Bayesian topology by optimizing branch lengths with clock vs. nonclock model settings. The second test was conducted as the first one except that four species in clade P (Fig. 4 ; taxa belonging to Phaseolinae) with the greatest distance from root to tip on visual inspection were pruned and the clock vs. nonclock test repeated. This was done to determine if the evolution of the Phaseolinae clade, which appeared to be most nonclock-like, was the only major departure from clock-like evolution. We also tested for unequal rates across loci using Tajima's relative rate test (Kumar et al. 2004) for a selected set of taxa including Phaseolinae genera.

For the seven-gene concatenated data set we used the consensus tree derived from the Bayesian analysis (as above) as inputs into r8s v1.71

(Sanderson 2003). Three outgroup taxa (*Galactia*, *Tephrosia*, and *Ophrestia*) were pruned to provide a clear root position, as required by r8s. The cross validation procedure was performed according to the r8s manual, with values of *k* ranging from -3 to +3 in increments of 0.3. The optimal smoothing parameter was found to be ~0.25, which was applied to subsequent analyses. To derive ages of nodes and an indication of variation around these estimates, the 95% credibility interval for each calibration was approximated by sampling 100 trees from the Bayesian stationary phase of the posterior distribution as r8s input, using the mean ± two standard deviations as the credibility interval (Scherson et al. 2008). The penalized likelihood (PL) method of rate smoothing (Sanderson 2003) was used to estimate dates of nodes.

A single fixed calibration point was used to derive absolute dates – node A in Fig. 4. The two calibration values applied to this node are the minimum and maximum ages found using *matK* and 12 fossil calibration points in Lavin et al. (2005). We only used *matK*-derived age estimates from that paper (and not *rbcL*) because more fossil calibration points were available for that data set (13 vs. 9 in *rbcL)* and because *matK* showed a more uniform distribution of substitutions (Lavin et al. 2005), indicating it may be less prone to homoplasy relative to information content. The standard deviation around the *matK* estimate was lower than for *rbcL* (Lavin et al. 2005).

Our use of a pre-existing calibration point is a secondary calibration and therefore needs to include the uncertainty associated with their age estimates, as well as the uncertainty in our analysis (Graur and Martin 2004). Because the Lavin et al. (2005) trees were drawn from the stable posterior distribution of a Bayesian analysis, the maximum and minimum values represent the 100% credibility interval given the assumptions of their analysis. Nodes in our analysis were profiled and we list the mean, minimum and maximum node ages found using these two fixed age regimes. Because we used the 95% credible set of trees from our phylogenetic analysis as input to r8s, our minimum value using the lower calibration and our maximum value using the upper calibration represent the 95% credibility interval of our estimate of the age, while including the uncertainty associated with using a secondary calibration (Table 2).

Results

Phylogeny of trnL-F—Parsimony analysis of 79 taxa identified 360 equally parsimonious trees of 1,468 steps each, with a consistency index (CI) of 0.61 (0.53 without autapomorphies) and a retention index (RI) of 0.81. The strict consensus tree is mostly resolved, but several clades received only weak bootstrap support (Fig. 1).

Near the root (*Indigofera*), resolution and support are weak, but two well-supported major clades are identified: a clade comprising Millettieae plus Phaseoleae subtribes Diocleinae and Ophrestiinae (Fig. 1, Clade C) and a clade containing all other Phaseoleae plus Desmodieae and Psoraleeae (Fig. 1, Clade A). Clade C (millettioid clade) has been the focus of better-sampled studies previously (Wojciechowski et al. 2004). We used it primarily as an outgroup and will not discuss it at any length here. Our analysis, however, confirms the paraphyly of *Ophrestia* and *Lonchocarpus*, two genera already shown elsewhere to be paraphyletic (Kajita et al. 2001; Hu et al. 2002). Within Clade A (phaseoloid clade), several subclades (marked in Fig. 1) are strongly supported here but their relationships relative to each other remain unresolved or only weakly supported.

Clade AA includes all representatives of four genera of Desmodieae, three with multiple species represented. *Desmodium* is strongly supported as monophyletic, whereas *Lespedeza* and *Kummerowia* are unresolved in the strict consensus tree. The grouping of the latter two genera with *Campylotropis* is consistent with the taxonomic treatment of these genera as subtribe Lespedezinae (Ohashi et al. 1981), separate from the larger subtribe Desmodiinae.

Clade V comprises Phaseoleae subtribe Kennediinae. The two multiply-sampled genera (*Kennedia* and *Hardenbergia*) were supported as monophyletic, but relationships of *Vandasia*, a monotypic segregate of *Hardenbergia* (Lackey 1981) are unresolved. Clade T includes Phaseoleae subtribe Cajaninae, within which were two strongly-supported subclades.

Clade R is dominated by Psoraleeae and by Phaseoleae subtribes Glycininae and Phaseolinae. This clade is divided into two groups, the first of which places together a monophyletic *Erythrina* (three species sampled here) with *Psophocarpus* (Clade Q). Although substantially more resolved than in the previously published *rbcL* topology (Kajita et al. 2001), the second clade (Clade B) still contains a backbone polytomy. Among the subclades found in this unresolved region are: Clade K, comprising *Pseudovigna* and one of the two species of the polyphyletic genus *Pueraria* (see Lee and Hymowitz 2001) sampled here (*P. phaseoloides*); Clade D, with *Pachyrhizus* and *Calopogonium*; and Clade P, which comprises nine species from seven genera of Phaseoleae subtribe Phaseolinae. In addition, two or more species each were included from *Glycine*, *Teramnus*, and *Amphicarpaea*, and all three genera were supported as monophyletic.

The core group of subtribe Phaseolinae (Clade P) is marked by a large cpDNA inversion (Bruneau et al. 1990), not found in *Psophocarpus*, which was classified by Lackey (1981) in that subtribe but is no longer included there (Lewis et al. 2005). Relationships within the Phaseolinae clade here included a dichotomy between a clade of several New World genera and Old World *Vigna* species, as in Thulin et al. (2004). The placement of *Dolichos lablab* with New World taxa is poorly supported here, in contrast to strong separation in their combined *trnK*-nrDNA ITS study. The closer relationship of *Strophostyles* to *Macroptilium* than to *Ramirezella*, seen in the analyses of Riley-Hulting et al. (2004), is also supported here.

Individual Analyses of Eight Chloroplast Regions—Sequences of seven additional chloroplast gene regions were obtained for a subset of 36 taxa included in the *trnL-F* study, and the *trnL-F* dataset was reduced to include the same subset. Each region was initially analyzed separately using equally weighted parsimony. Substantial topological agreement was found among these analyses (Fig. 2), particularly in identifying, with high bootstrap support, many of the clades supported in the broader *trnL-F* analysis. Of the clades potentially observable in these analyses (i.e. excluding Clade A, due to sampling and rooting issues, as well as Clade D due to the exclusion of *Calopogonium*), Clades V (Kennediinae), T (Cajaninae), and Q (*Erythrina* + *Psophocarpus*) are all identified with 99–100% BS in analyses of each of the eight regions. Clade K (*Pueraria phaseoloides* + *Pseudovigna*) also appeared in all analyses, with bootstrap support greater than 90% for all regions except *rbcL* (67%) and *rpl2* (80%). Clade AA (Desmodieae) is strongly supported as monophyletic in all analyses except *clpP*, where this grouping did not appear in the strict consensus tree. With *clpP*, as in all other analyses, the two subgroups of Desmodieae, *Desmodium* (Desmodiinae) and the three genera of subtribe Lespedezinae, are strongly supported as monophyletic (96% BS for both groups in *clpP*). Finally, the two genera of Clade E (Psoraleeae) are strongly supported (72% in *rbcL*, 95% or greater in others) in all analyses except *rpl2*.

The two "backbone" clades identified in the full *trnL-F* analysis (Clades R and B; Fig. 1) were found to be more variable in their presence and support. Clade R, which groups Clade Q (*Erythrina* + *Psophocarpus*), Clade E (Psoraleeae), Clade P (Phaseolinae) and most of the Glycininae, is resolved in all of the strict consensus trees except *ycf4*, albeit with less than 50% support in *atpB* (Fig. 2). Clade B, which includes the same taxa minus Clade Q, received strong support (greater than 80%) in all analyses except *ycf4*, while relationships within this

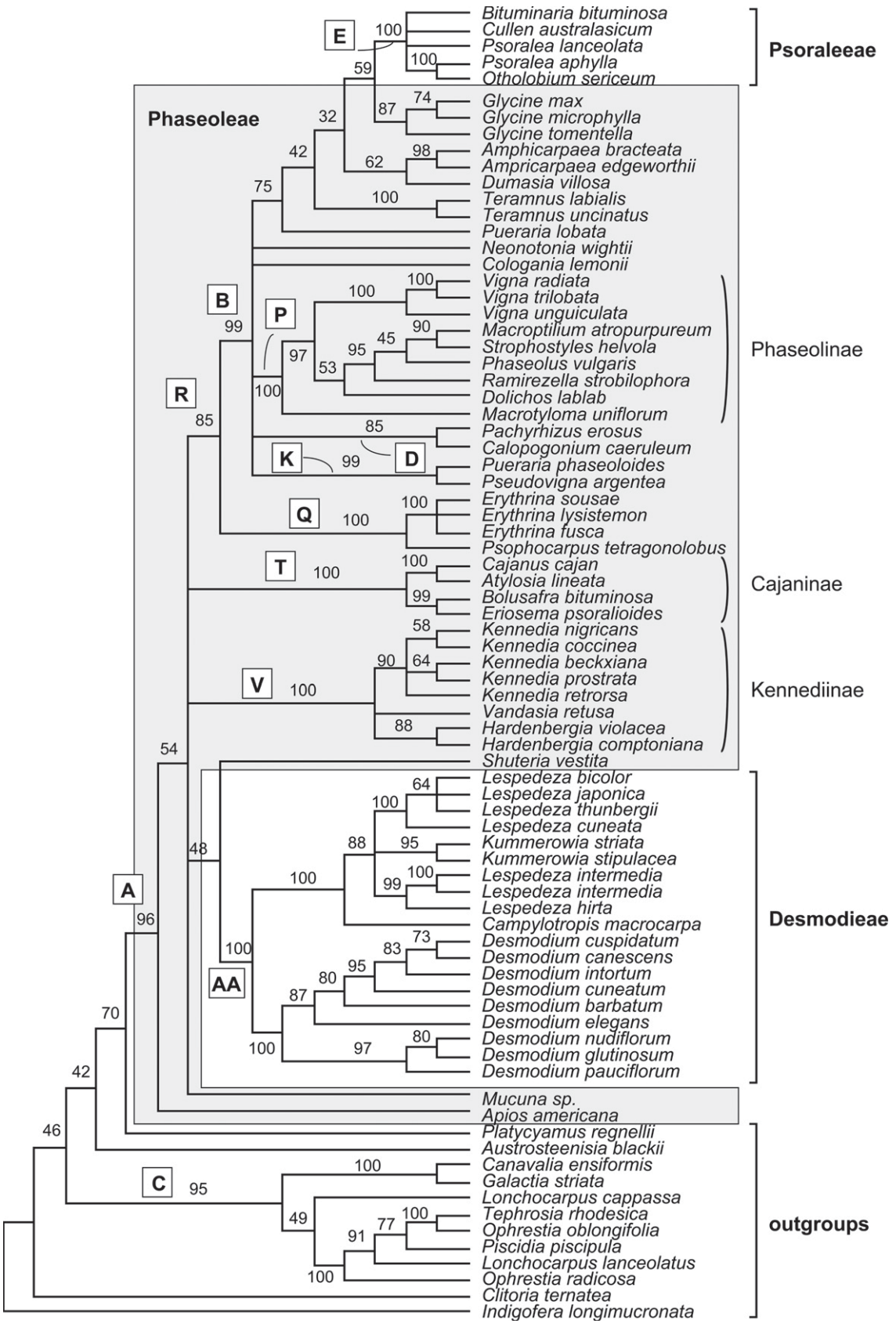

FIG. 1. The strict consensus of 360 equally parsimonious trees based on *trnL-F* sequences comprising a wide sampling of phaseoloid legume species. The tree is rooted using taxa from closely related millettioid and indigoferoid legumes as outgroups. Classification by tribe/subtribe, based on Lackey (1981), Polhill (1994), and Lewis et al. (2005), is indicated (labeled by shading and parentheses). Major clades recovered and discussed in this study are marked by bold boxed letters. Numbers indicate bootstrap support.

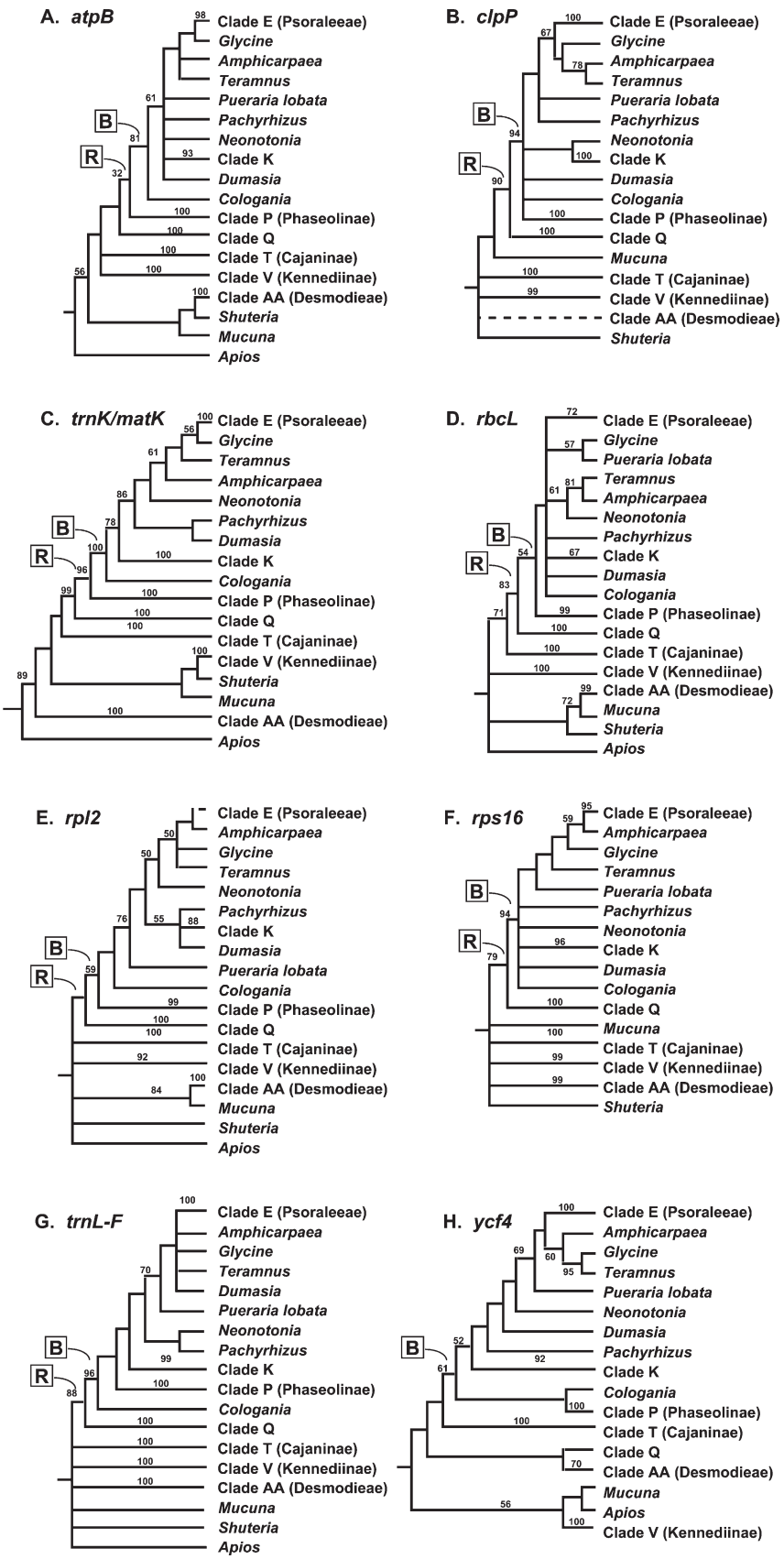

FIG. 2. Parsimony topologies of eight individually analyzed chloroplast regions. All are strict consensus topologies except for *trnK/matK* and *ycf4,* which are the single most parsimonious trees identified. Bootstrap values are indicated for nodes supported at ≥50%. Terminal units labeled as "clades" (e.g. Clade Q) or indicated with boxed letters R or B are those identified in the larger analysis of *trnL-F* (Fig. 1; see text). The lack of support for monophyly of Desmodieae in *clpP* is indicated by a dashed line (Clade AA in *trnL-F*).

clade are quite variable across analyses. For most individual regions, relatively few relationships were strongly supported, but even these were often in conflict among analyses. For example, *Teramnus* was moderately strongly supported as sister to *Amphicarpaea* with *clpP* and *rbcL*, but for *ycf4* was strongly supported as sister to *Glycine* (Fig. 2).

Of the eight analyzed cpDNA regions, *ycf4* appears to give results most in conflict with all other regions in Clade B. We therefore explored the phylogenetic contributions of this gene in greater detail. Analysis of first and second positions versus third positions in *ycf4* using Neighbor-Net (Huson and Bryant 2006) revealed that the majority of the signal that grouped *Glycine* and *Teramnus* comes from the first and second positions (Fig. 3A). Third positions alone are ambiguous with respect to the relationships among *Glycine*, *Teramnus*, *Amphicarpaea* and the Psoraleeae (Fig. 3B). Tajima's relative rate tests also show rate acceleration in the first two positions of *ycf4* from *Glycine* that can explain many differences relative to several members of Clade B (Table 1; *Neonotonia* used as the outgroup for these tests). There is also an indication that these positions in *Teramnus* may be somewhat accelerated. These positions in *Glycine* show significantly more change in all comparisons except to *Teramnus*, whereas the latter shows no significant increase in change in any comparison, although *p* values are below 0.1 in two cases (including *Glycine* vs. *Teramnus*).

Concatenated Analyses of Chloroplast Regions—Based on single gene results, concatenated analyses were run both with and without *ycf4*. Parsimony analysis of all eight regions identified a single tree (L = 9276; CI = 0.66/0.54; RI = 0.69); this tree (not shown) was hit in 998 of the 1,000 random addition TBR searches.

The paired Bayesian analyses using all eight genes and the best models for each partition (see Methods) converged quickly and produced nearly identical arithmetic means of the marginal likelihood scores (-67,651.46 and -67,653.58, respectively) after discarding trees from 100,000 generations as the burn in. The estimated clade posterior probabilities were within 1% of each other from these two analyses. The paired Bayesian analyses that excluded *ycf4* also produced similar likelihood scores (-60,699.09 and -60,698.58), other details as above. The Bayesian analysis of all eight genes identified a topology (Fig. 4) similar to the parsimony tree. The model choice sensitivity analysis found no qualitative differences between models (not shown).

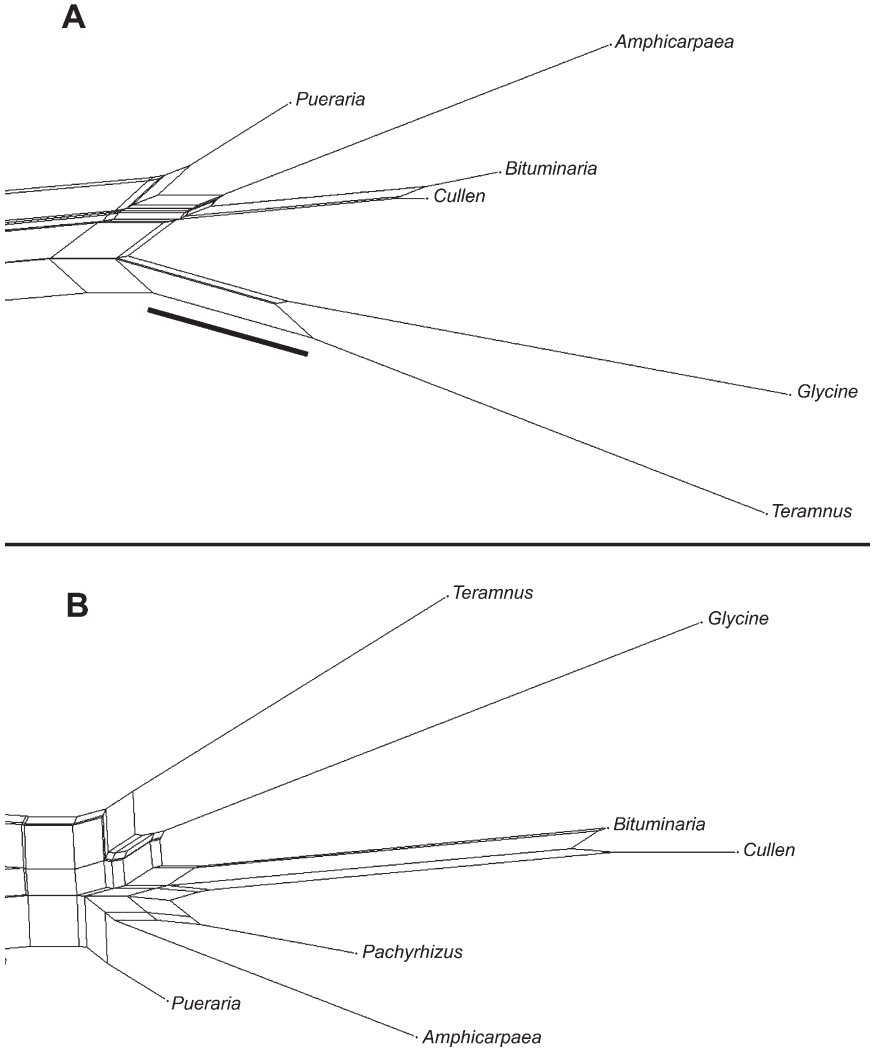

FIG. 3. Neighbor-net obtained from *ycf4* sequences. A. Network based on the first and second positions of *ycf4* showing signal that groups *Glycine* and *Termanus* (the side of the rectangle above the line); B. Network based on the of third positions of *ycf4* showing ambiguous signal and no clear support for (*Glycine* + *Teramnus*).

In the 8-region analysis, Clades AA, V, T, R, Q, B, K, P, and E are all identified with 100% BS in the parsimony analysis and with 1.0 PP in the Bayesian analysis (Fig. 4). *Mucuna* is sister to Desmodieae (Clade AA), as in the individual analyses of *rbcL*, *rpl2*, and (if *Shuteria* is not considered) of *atpB* and *trnK/matK* (Fig. 2). *Mucuna* and Desmodieae both lack the chloroplast *rpl2* intron, which is retained in *Shuteria* and other phaseoloids (Bailey et al. 1997). The position of *Apios* as sister to the remainder of the main phaseoloid group is in agreement with its placement in the comprehensive *matK* analysis of Wojciechowski et al. (2004) and not inconsistent with its weakly supported placement in the large *rbcL* analysis (Kajita et al. 2001). The placement of *Shuteria* as sister to the remaining phaseoloids is ambiguous. It received high posterior probability in the Bayesian analysis, but was found as sister to *Mucuna* plus Desmodieae in the parsimony tree (though with <50% BS). The position of Kennediinae (Clade V) relative to *Mucuna* plus Desmodieae is reversed in the two analyses, albeit with weak support in the parsimony analysis. Kennediinae was not included in Wojciechowski et al. (2004) and its placement was poorly supported in Kajita et al. (2001). Clade B contains the only other disagreement between the parsimony and Bayesian analyses, involving the placement of *Dumasia*, weakly supported by parsimony as sister to *Pachyrhizus*.

As in the broad *rbcL* analysis (Kajita et al. 2001), Cajaninae (Clade T) plus Clade R has good support here as well. Clade R splits further into two well-supported clades, labeled as B and Q (Fig. 4). The full resolution of the 8-gene analysis identified relationships within Clade B either not seen or not strongly supported in single-gene analyses, either here or in any of the previous studies (Kajita et al. 2001; Lee and Hymowitz 2001; Wojciechowski et al. 2004). The *rbcL* parsimony strict consensus tree of Kajita et al. (2001), for example, did not resolve relationships among these taxa, though it did identify Clades P and E. The *matK* analysis of Wojciechowski et al. (2004) included fewer phaseoloid taxa and had little support for relationships among them, outside of identifying these same two clades. The *rps16* parsimony strict consensus tree (Lee and Hymowitz 2001) included additional Glycininae, but did not include Psoraleeae, and its backbone was poorly supported, except for a clade that included ((*Pueraria lobata, P. pulcherrima, Nogra*) (*Amphicarpaea* (*Glycine*, *Teramnus*))), the relationships among which all had bootstrap support greater than 60%. The sister relationship between *Glycine* and *Teramnus* is strongly supported in the 8-gene analysis, although it appeared with strong support only in the *ycf4* analysis. The absence of this relationship from *rbcL* and *matK* phylogenies here is not surprising. In the Kajita et al. (2001) analysis, as here, *Amphicarpaea* and *Teramnus* were strongly supported as sister taxa, whereas in the *matK* study of Wojciechowski et al. (2004), *Glycine* and Psoraleeae (Clade E) were sisters, though with even less than our 56% BS for this relationship with *trnK/matK* (Fig. 2C). Our *rps16* tree (Fig. 2F) did not recover a *Glycine-Teramnus* sister group, unlike that of Lee and Hymowitz (2001), where the relationship had moderate support (69%). However, when the 8-gene data were reanalyzed with the two Psoraleeae (Clade E) excluded, a sister relationship of *Teramnus* and *Glycine* was recovered by parsimony, with strong (88%) bootstrap support (results not shown).

Analyses of seven regions, excluding *ycf4*, also resulted in nearly identical parsimony and Bayesian trees (Bayesian tree: Fig. 4). Parsimony analysis identified four trees (L = 9165, CI = 0.71/0.58, RI = 0.70) which differed in: (1) the placement of *Shuteria* (either as in Fig. 4 or as sister to Kennediinae); (2) the placement of *Shuteria* plus Kennediinae (either diverging immediately after *Apios* or after *Mucuna* plus Desmodieae); and (3) the placement of *Pachyrhizus, Neonotonia*, and *Dumasia* (either as in Fig. 3B or with *Pachyrhizus* and *Dumasia* as sister taxa). Notably, exclusion of *ycf4* resulted in placing *Teramnus* and *Amphicarpaea* as sisters (81% BS) and uniting *Glycine* with Psoraleeae (65% BS). Bayesian analysis recovered a tree (Fig. 4) that reflects alternative resolutions among equally parsimonious trees, but differs from all four MP trees in placing *Mucuna* plus Desmodieae between Kennediinae and *Shuteria*. As in the parsimony analysis, *Glycine* was placed with Psoraleeae (0.95 PP) and *Teramnus* was sister to *Amphicarpaea* (0.98 PP). Similar results were obtained when the third codon positions of *ycf4* are included (data not shown), although this reduced posteriors for these clades (0.86 and 0.88, respectively).

Evolutionary Rates and Dating of Nodes—Inspection of trees suggested nonclocklike behavior of many clades (Fig. 4). Likelihood ratio tests without and with a clock enforced were significantly different (-lnL = 60,057.55 vs. 60,589.20; $\chi^2 = 1063.30$; $p \ll 0.001$, n = 33, df = 31). Estimated dates of nodes based on the seven-gene concatenated data set using the penalized likelihood method are reported in Table 2.

Using these estimates, rates for branches leading to all nodes in the tree (excluding outgroups) were plotted to identify those with most divergent rates (Fig. 5). The fastest rates were those involving the taxa of Clade P (Phaseolinae), including the branch leading to the common node for this clade, as already noted by Lavin et al. (2005). The four members of this clade were removed and likelihood scores with and without a clock were again calculated. Even without Clade P, the dataset was strongly nonclocklike (-lnL = 53,692.02 vs. 54,038.35; $\chi^2 = 692.7$; $p \ll 0.001$, n = 29, df = 27). Consistent with the rate distribution (Fig. 5), removal of Phaseolinae had a much larger effect than did removal of four taxa with more average rates (*Cullen, Bituminaria, Pseudovigna*, and *Pueraria phaseoloides*); when the latter four taxa were removed, a clock was rejected with a χ^2 value of 1,052.5. Recent comparisons of the complete chloroplast genome sequences have revealed higher rates of structural and sequence change in *Phaseolus vulgaris* compared with *Glycine max* (Guo et al. 2007).

Non-clock-like behavior was not uniform across all eight chloroplast regions. Tajima's relative rate tests calculated for Phaseolinae versus either *Glycine* or *Amphicarpaea*, and an outgroup (*Clitoria* or *Galactia*) supported significantly variable rates involving all four Phaseolinae for *clpP*, *trnL-F*, *rps16*

TABLE 1. Results of Tajima's relative rate tests for comparisons of *ycf4* of *Teramnus* and *Glycine* with other members of Clade B. In each cell, results for first plus second codon positions are given first, followed by those of the third positions. *$p < 0.05$; ** $p < 0.01$; ns = not significant. Several values close to $p = 0.05$ are listed; all other nonsignificant values $p > 0.1$.

	Teramnus	*Cullen*	*Bituminaria*	*Amphicarpaea*	*Glycine*
Teramnus	-	0.059/ns	ns/ns	ns/ns	0.095/ns
Glycine	0.095/ns	**/ns	**/ns	*/ns	-

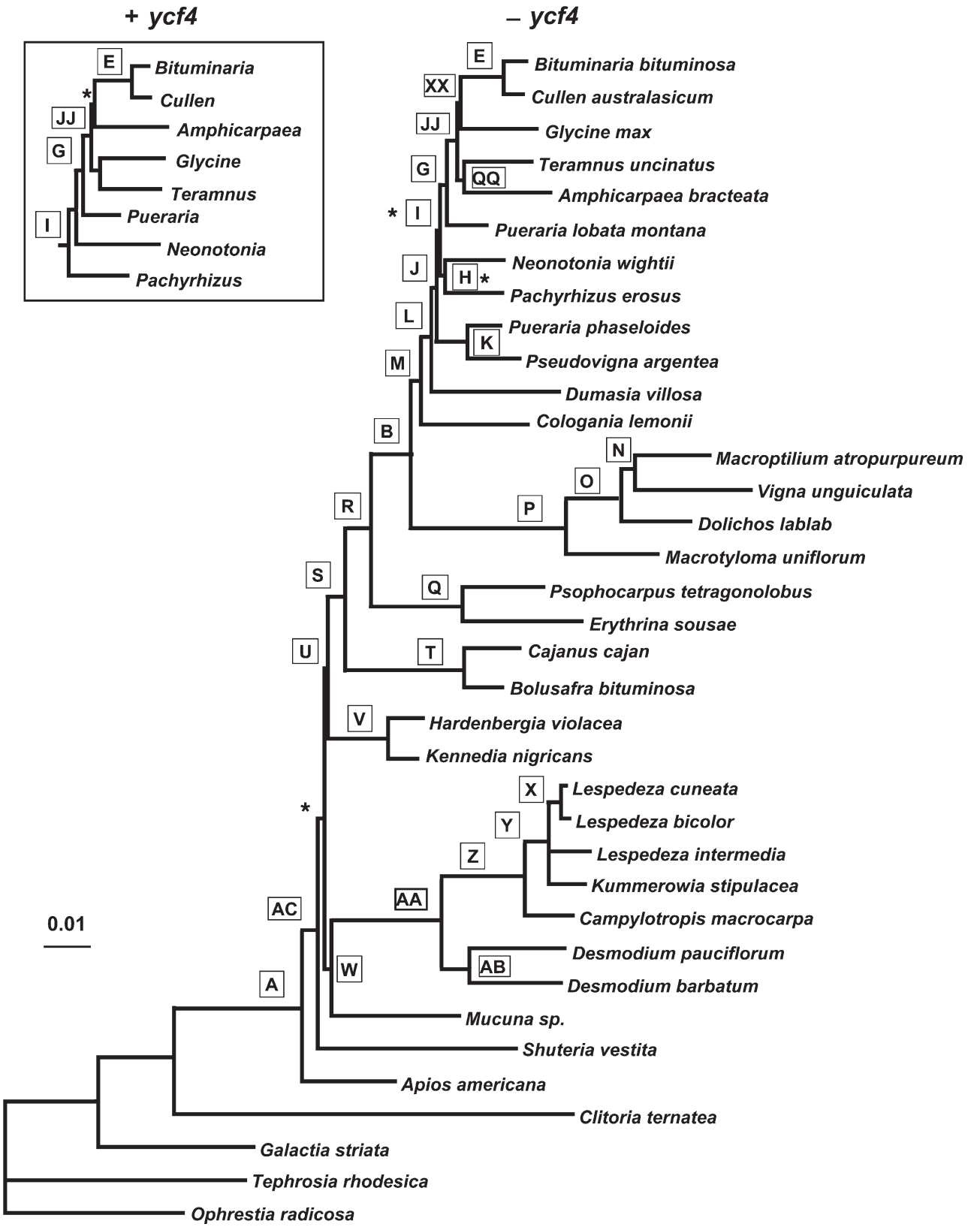

FIG. 4. Bayesian tree obtained from sequences of seven concatenated chloroplast regions (*trnL-F, rbcL, atpB, trnK/matK, rpl2, clpP,* and *rps16*, but excluding *ycf4*). Asterisk indicates branches with posterior probabilities <0.95; all other interior branches have posterior probability ≥ 0.95. Letters indicate nodes for which dates were estimated (Table 2); Node A was a fixed dating point (see text). Inset (not at the same scale) illustrates a portion of the Bayesian tree for eight concatenated chloroplast regions (including *ycf4*) showing all of the topological differences between the seven- and eight-region analyses. Otherwise, only a single significant difference in posterior probability exists between these two analyses: the clade that groups clades U and W in the seven-gene analysis has a PP of 0.89, whereas this clade has a PP of 0.95 in the eight-gene analysis.

TABLE 2. Node ages inferred using a 24.2 Ma or 32.1 Ma fixed age for node A (Fig. 3). Minimum and maximum from the 95% credible set are derived from 100 trees drawn from the stable posterior distribution using the seven-gene concatenated dataset (without *ycf4*). The calibration points represent the boundaries of the 100% credibility interval (i.e. minimum and maximum values) for ages derived using *matK* from Lavin et al. (2005). The overall 95% credibility interval is between the lowest value from the low calibration and the highest values from the high calibration. Nodes JJ and H were not present in all trees – the profile is derived from the percentage of trees containing these nodes as indicated. Ma – million years.

Node (clade)	95% credibility interval (low calibration) in Ma	95% credibility interval (high calibration) in Ma
A (Gly-Api)	**24.2 (fixed)**	**32.1 (fixed)**
B (Gly-Vig)	15.0–16.8	19.6–22.0
QQ (Amp-Ter)	7.7–10.1	10.0–13.2
JJ (Cul-Bit-Gly-Amp-Ter)	8.5–11.0 (98% of trees)	11.0–14.3 (98% of trees)
E (Cul-Bit)	2.5–3.8	3.2–5.0
XX (Cul_Bit_Gly)	8.0–10.4	10.4–13.5
G (Gly-Pue_l_m)	9.8–12.4	12.8–16.3
H (Pac-Neo)	10.9–12.2 (87% of trees)	14.2–16.1 (87% of trees)
I (Gly-Pac)	11.1–13.4	14.6–17.5
J (Gly-Pse)	11.6–13.8	15.3–18.1
K (Pse-Pue_p)	6.1–8.7	8.1–11.5
L (Gly-Dum)	12.4–14.6	16.3–19.1
M (Gly-Col)	13.7–16.0	18.0–20.9
N (Vig-Mac_a)	4.0–4.8	4.9–6.0
O (Vig-Dol)	4.5–5.3	5.6–6.7
P (Vig-Mac_u)	7.0–8.2	8.8–10.5
Q (Ery-Pso)	8.6–10.8	11.3–14.2
R (Gly-Ery)	17.6–19.5	23.0–25.5
S (Gly-Bol)	19.4–21.3	25.5–28.1
T (Bol-Caj)	5.0–6.6	6.6–8.7
U (Gly-Ken)	21.0–22.6	27.7–29.9
V (Ken-Har)	6.2–9.3	8.4–12.7
W (Muc-Les_b)	20.3–22.4	26.7–29.6
X (Les_b-Les_c)	0.5–1.1	0.6–1.5
Y (Les_b-Les_i-Kum_s)	2.4–3.3	3.1–4.3
Z (Les_b-Cam_m)	4.0–5.2	5.1–6.9
AA (Les_b-Des_b)	10.8–12.6	14.1–16.5
AB (Des_b-Des_p)	7.9–9.9	10.3–12.9
AC (Gly_m-Shu_v)	22.0–23.5	29.1–31.1

(for the single available comparison with *Vigna*) and (with the exception of the *Amphicarpaea* comparison for *Macrotyloma*) for *matK*, significant differences for the *Amphicarpaea* but not the *Glycine* comparison for *rpl2*, but generally not for *atpB* (one exception), *rbcL*, or *ycf4* (Table 3).

DISCUSSION

Evolutionary Relationships Within Phaseoloid Legumes—The concatenated analyses of up to eight chloroplast regions provide the first well-resolved and strongly supported phylogenetic hypothesis among the phaseoloid genera included in this study. Analyses of individual gene regions mostly provided only weakly supported resolution of many of these genera, as was observed in previous analyses of some of the same regions [*rbcL*: (Kajita et al. 2001); *rps16*: (Lee and Hymowitz 2001); *matK*: (Wojciechowski et al. 2004)]. Despite weak support for many relationships, individual gene trees agreed with one another in identifying nine clades, which thus appeared in the concatenated analysis, and none of these groupings were surprising. The tribes Psoraleeae (Clade E) and Desmodieae (Clade AA) are morphologically distinctive groups with long histories of taxonomic recognition (Lewis et al. 2005), more recently shown to be nested within the phaseoloid legumes (Kajita et al. 2001). Two other clades represent subtribes of Phaseoleae [Cajaninae (Clade T) and Kennediinae (Clade V)] in the system of Lackey (1981), which is modified from Bentham's (1837) classification. Clade P represents another Bentham subtribe, Phaseolinae, with the removal of *Psophocarpus*, a genus that molecular data have already shown to be distinct from other members of the subtribe (Bruneau et al. 1990). The grouping of *Psophocarpus* with *Erythrina* was previously identified in comprehensive *rbcL* studies, as was a close relationship between *Pueraria phaseoloides* and *Pseudovigna* (Kajita et al. 2001). The remaining two clades found in most or all individual analyses (Clades R and B; Fig. 2) primarily included a group of genera corresponding to Phaseoleae subtribe Glycininae, and it is within this group that concatenated analyses provided novel information. This is particularly relevant in addressing the origin or origins of the polyploid genus *Glycine*.

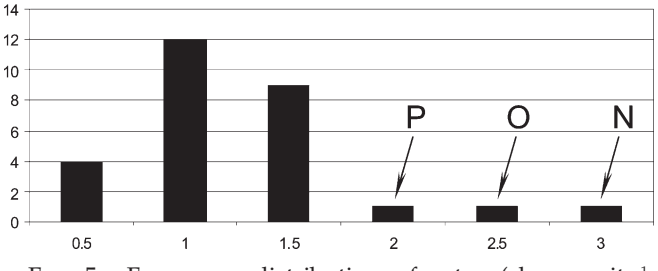

FIG. 5. Frequency distribution of rates (changes.site^{-1}. billion years^{-1}) of different branches calculated using r8s by profiling nodes from 95 trees sampled from the Bayesian posterior distribution and the 32.1 Ma calibration (24.2 Ma calibration results were qualitatively similar). The X-axis labels indicate the lower bound of each bin. The branches below nodes in the Phaseolinae (nodes names from Fig. 4) with the three largest rates are indicated.

TABLE 3. Tajima's test of relative rates for *Glycine* or *Amphicarpaea* compared to four Phaseolinae genera. The outgroup in all cases except *rpl16* was *Clitoria*; the outgroup in the *rpl16* case was *Galactia* (*Clitoria* had some missing sequence), but only the comparison to *Vigna* could be made (the other Phaseolinae had missing sequence). *$p < 0.05$; ** $p < 0.01$; ns = not significant; n/a = sequence not available.

		atpB	*clp*	*matK*	*rbcL*	*rpl2*	*rps16*	*trnL-F*	*ycf4*
Glycine	*Macrotyloma*	ns	**	*	ns	ns	n/a	**	ns
	Dolichos	**	**	**	ns	ns	n/a	**	ns
	Macroptilium	ns	**	**	ns	ns	n/a	**	ns
	Vigna	0.053	**	**	ns	ns	**	**	ns
Amphicarpaea	*Macrotyloma*	ns	**	ns	ns	0.058	n/a	**	ns
	Dolichos	ns	**	**	ns	*	n/a	**	ns
	Macroptilium	ns	**	*	ns	*	n/a	**	ns
	Vigna	ns	**	*	ns	*	**	*	ns

The comprehensive analyses of legumes at the higher (family) level using *rbcL* (Kajita et al. 2001) and *matK* (Wojciechowski et al. 2004) did not sample many Glycininae genera and neither provided much resolution among genera that were sampled. The *rps16* study of Lee and Hymowitz (2001), aimed specifically at Glycininae, did not sample the tribe Psoraleeae, which had previously been shown to be nested within Phaseoleae (Doyle et al. 1997), and even within Glycininae (Adams et al. 1999). The concatenated analysis strongly supports the monophyly of a group of genera comprising Glycininae sensu Polhill (1994), with the following emendations. First, Psoraleeae must be included, as is now widely accepted (Lewis et al. 2005). Secondly, *Shuteria* must be excluded, despite strong morphological similarities to *Dumasia* noted by Lackey (1981). The general congruence between the topology of the concatenated analysis and that of *rps16* suggests that *Mastersia*, which was sampled by Lee and Hymowitz (2001), should also be removed from Glycininae, as should some elements of the polyphyletic *Pueraria* (specifically, *P. wallichii*, also sampled by Lee and Hymowitz 2001). Genera likely to be included in Glycininae but not sampled here are *Nogra* and *Teyleria* (strongly supported as sister to *Pueraria montana* in the Lee and Hymowitz [2001] *rps16* analysis), and possibly *Phylacium* (A. N. Egan and JJD, unpublished data).

Phylogenetic Position of Glycine (Soybeans)—Relationships around *Glycine* have been particularly problematical. In his treatment of Glycininae, Lackey (1981) noted that *Glycine* was "a genus beset with taxonomic and nomenclatural difficulties, which is unfortunate, because it includes the soybean." The same uncertainty surrounds molecular phylogenetic relationships, which is also unfortunate, because it would be helpful to know the relationships of other genera to *Glycine*, given its polyploid history (Shoemaker et al. 2006).

Concatenated analyses presented here identify a strongly supported (>90% BS and 1.0 PP) clade that includes *Glycine*, *Teramnus*, *Amphicarpaea*, and Psoraleeae, with this clade sister to *Pueraria montana*, also with good support (90% BS and 1.0 PP with *ycf4*; 84%BS and 1.0 PP without *ycf4*). Thus, it is likely that *Glycine* derived its chloroplast genome from a plant bearing the chloroplast genome ancestral to the plastid genomes found in this group of plants. The closest extant relative of the *Glycine* chloroplast genome appears to Psoraleeae (Fig. 4). This result was not observed in the combined analysis of all eight regions, where instead the *Teramnus* chloroplast genome was sister to that of *Glycine* (Fig. 4; inset). However, the grouping of *Glycine* and *Teramnus* appears to be due to convergent evolution at the first and second codon positions of *ycf4*, the only region in which this result was supported (Fig. 2). Removal of *ycf4* or use of only third codon positions produced the *Glycine*-Psoraleeae sister relationship and grouped *Teramnus* with *Amphicarpaea*.

The only published nuclear gene phylogeny for these taxa does not agree with chloroplast results. The chloroplast-expressed nuclear gene for glutamine synthetase (*ncp-GS*:Doyle et al. 2003) identified *Teramnus* as sister to *Glycine* with strong support (93% BS; parsimony), and *Amphicarpaea* joined this pair with 84% BS. Although these results are similar to the combined analysis of all eight regions (Fig. 4; inset), these taxa were included in a robust clade (97% BS) that excluded Psoraleeae. The *Glycine-Amphicarpaea-Teramnus ncp-GS* clade also included *Dumasia*, but not *Neonotonia*, another major incompatibility with the chloroplast results. From work in progress on phylogenies of other nuclear genes, it appears that the relationships among these taxa are complex (A. N. Egan and J. J. Doyle, unpubl. data), likely involving introgression and lineage sorting.

The members of the *Glycine*-Psoraleeae-*Amphicarpaea*-*Teramnus* clade (Clade JJ, Fig. 4) shared a common ancestral chloroplast genome around 11 MYA, based on the 7-gene analysis (8.5–14.3 overall 95% confidence interval; Table 2). The divergence of *Glycine* and Psoraleeae (Clade XX, Fig. 4) in the 7-gene analysis is estimated at around 10.4 MYA (8.0–13.5 overall 95% confidence interval; Table 2). These dates are close to estimates of the age of duplicated regions of the soybean genome resulting from the polyploid event that led to the present $2n = 4x = 40$ chromosome complement of *Glycine* (Shoemaker et al. 2006). Two studies measured synonymous distances (K_s) of large numbers of paralogue pairs from the extensive soybean expressed sequence tag (EST) collection to identify large-scale duplication events (Blanc and Wolfe 2004; Schlueter et al. 2004). Both identified a distribution of divergence times with a large number of pairs having similar synonymous divergences, but because they used different clock calibrations their estimates ranged from under 3–5 MYA (Blanc and Wolfe 2004) to nearly 15 MYA (Schlueter et al. 2004). The latter group more recently reported a divergence date of 12.2 MYA in a study of paired genes in homoeologous chromosomal regions (Schlueter et al. 2007).

The date of divergence of paralogue pairs is a measure either of divergence of alleles in an autopolyploid (whether inherited disomically or tetrasomically), or of the divergence of the two taxa that contributed homoeologous loci to an allopolyploid. If *Glycine* paralogue pairs are younger than the divergence of *Glycine* from all of its close generic relatives, as suggested by the Blanc and Wolfe (2004) estimate, then *Glycine* cannot be an allopolyploid derived from hybridization among the ancestors of these genera. *Glycine*

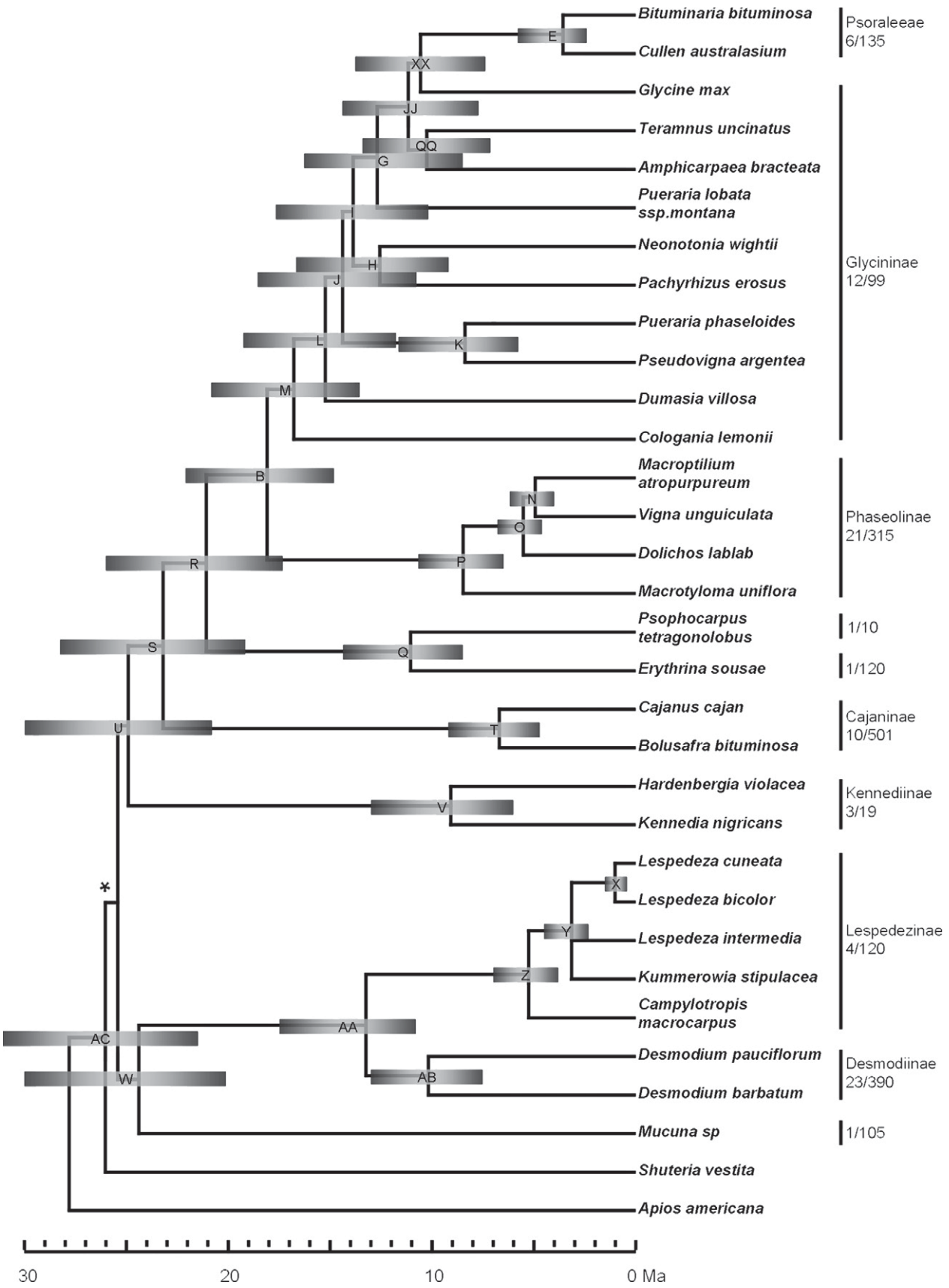

FIG. 6. Diversification and divergence times of phaseoloid clades taken from Table 2. Chronogram is estimated via penalized likelihood based on the Bayesian consensus tree obtained from the seven-gene dataset (Fig. 4; see Methods for more detail). Node labels and support follow those in Fig. 4. Bars show the overall 95% credibility interval (compare with Table 2). Scale is millions of years (Ma). Number of genera/species is given for each group.

could either be an autopolyploid or an allopolyploid produced by hybridization among extinct diploid taxa that diverged from one another more recently. If *Glycine* paralogue pairs are older than the speciation events that led to modern Glycininae, as suggested by Schlueter et al. (2004), then *Glycine* could be an allopolyploid whose genomes are derived from among these genera. Published gene trees that include putative *Glycine* homoeologues and orthologues from other genera place the two *Glycine* paralogues as sister to one another, and the two in turn as sister to the single gene from *Teramnus* (Doyle et al. 2003; Straub et al. 2006). This topology is consistent with autopolyploidy or allopolyploidy from recently diverged extinct diploid taxa (Straub et al. 2006). Additional nuclear genes are currently being investigated to address this problem (A. N. Egan and J. J. Doyle, unpubl. data).

Divergence of Core Phaseoloids—These results provide a comprehensive picture of the evolutionary divergence of one of the largest clades in the Leguminosae. Over 80 genera and more than 1,800 phaseoloid species (Lewis et al. 2005) split into two large clades early in phaseoloid history, each comprising large groups of species (Fig. 6). One of these clades split almost immediately to produce the Desmodieae, with 27 genera and around 500 species, plus *Mucuna*, with an additional 105 species. An early split in the second clade separated the Phaseoleae subtribe Cajaninae, with 10 genera and 500 species, from the remainder of the phaseoloids. This latter group in turn gave rise to the large (120 species) pantropical genus, *Erythrina*, the Phaseolinae, with over 300 species, and the Glycininae plus Psoraleeae, with over 200 species. The diversifications of most species-rich groups took place mostly within the last 15 million years, several much more recently — e.g. 300 species of Phaseolinae, 500 species of Cajaninae, 300 species of *Desmodium* and 120 of *Lespedeza*, and 135 species of Psoraleeae. Several of these lineages include major crop plants, such as several species each of *Phaseolus* and *Vigna* within Phaseolinae, *Cajanus cajan* (pigeonpea) within Cajaninae, and *Glycine max* (soybean) within the radiation that includes Psoraleeae.

Acknowledgments. Funding for this work came from National Institute of Health grant GM-35087 to JDP, National Science Foundation grant DEB-0516673 to JJD, and Natural Sciences and Engineering Council of Canada Discovery grant 326439-06 to SS. Suggestions on the manuscript and discussions with Ashley Egan are sincerely appreciated. We also thank two anonymous reviewers for their critical comments on the earlier versions of the manuscript.

Literature Cited

Adams, K.L., K. Song, P.G. Roessler, J.M. Nugent, J.L. Doyle, J.J. Doyle, and J.D. Palmer. 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes. *Proceedings of the National Academy of Sciences USA* 96: 13863–13868.

Bailey, C.D., J.J. Doyle, T. Kajita, T. Nemoto, and H. Ohashi. 1997. The chloroplast *rpl2* intron and ORF184 as phylogenetic markers in the legume tribe Desmodieae. *Systematic Botany* 22: 133–138.

Barker, N.P., B.D. Schrire, and J.-H. Kim. 2000. Generic relationships in the tribe Indigofereae (Leguminosae: Papilionoideae) based on sequence data and morphology. Pp. 311–337 in *Advances in Legume Systematics vol. 9*, ed. P.S. Herendeen, and A. Bruneau. Kew: Royal Botanic Gardens.

Bentham, G. 1837. *Commentationes de leguminosarum generibus*. Ed. Vienna: Sollingeri.

Blanc, G. and K.H. Wolfe. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *The Plant Cell* 16: 1667–1678.

Bruneau, A., J.J. Doyle, and J.D. Palmer. 1990. A chloroplast DNA inversion as a subtribal character in the Phaseoleae (Leguminosae). *Systematic Botany* 13: 378–386.

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

Daley, D.O., R. Clifton, and J. Whelan. 2002. Intracellular gene transfer: Reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome *c* oxidase. *Proceedings of the National Academy of Sciences USA* 99: 10510–10515.

Doyle, J.J. 1992. Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* 17: 144–163.

Doyle, J.J. and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Doyle, J.J. and J.L. Doyle. 1993. Chloroplast DNA phylogeny of the papilionoid legume tribe Phaseoleae. *Systematic Botany* 18: 309–327.

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.L. Doyle, and C. Harbison. 2003. Chloroplast-expressed glutamine synthetase in *Glycine* and related Leguminosae: phylogeny, gene duplication, and ancient polyploidy. *Systematic Botany* 28: 567–577.

Doyle, J.J., J.L. Doyle, and J.D. Palmer. 1995. Multiple independent losses of two genes and one intron from legume chloroplast genomes. *Systematic Botany* 20: 272–294.

Egan, A.N. and K.A. Crandall. 2008. Incorporating gaps as phylogenetic characters across eight DNA regions: ramifications for North American Psoraleeae (Leguminosae). *Molecular Phylogenetics and Evolution* 46: 532–546.

Espert, S.M., S.I. Drewes, and A.D. Burghardt. 2007. Phylogeny of *Macroptilium* (Leguminosae): morphological, biochemical and molecular evidence. *Cladistics* 23: 119–129.

Graham, S.W. and R.G. Olmstead. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712–1730.

Graur, D. and W. Martin. 2004. Reading the entrails of chickens: Molecular timescales of evolution and the illusion of precision. *Trends in Genetics* 20: 80–86.

Guo, X., S. Castillo-Ram'rez, V. Gonz‡lez, P. Bustos, J.L. Fern‡ndez-V‡zquez, R.I. Santamar'a, J. Arellano, M.A. Cevallos, and G. D‡vila. 2007. Rapid evolutionary change of common bean (*Phaseolus vulgaris* L.) plastome, and the genomic diversification of legume chloroplasts. *BMC Genomics* 8: 228.

Hoot, S.B., A. Culham, and P.R. Crane. 1995. The utility of *atpB* gene sequences in phylogenetic relationships: comparison with *rbcL* and 18S ribosomal DNA sequences in Lardizabalaceae. *Annals of the Missouri Botanical Garden* 82: 194–207.

Hu, J.-M., M. Lavin, M.F. Wojciechowski, and M.J. Sanderson. 2000. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its implications for evolutionary patterns in Papilionoideae. *American Journal of Botany* 87: 418–430.

Hu, J.-M., M. Lavin, M.F. Wojciechowski, and M.J. Sanderson. 2002. Phylogenetic analysis of nuclear ribosomal ITS/5.8S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe-Cyclolobium*, the core Millettieae, and the *Callerya* group. *Systematic Botany* 27: 722–733.

Huson, D.H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Kajita, T., H. Ohashi, Y. Tateishi, C.D. Bailey, and J.J. Doyle. 2001. *rbcL* and legume phylogeny, with particular reference to Phaseoleae, Millettieae, and allies. *Systematic Botany* 26: 515–536.

Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150–163.

Lackey, J.A. 1981. Phaseoleae. Pp. 301–327 in *Advances in Legume Systematics, Part 1*. eds. R.M. Polhill, and P.H. Raven. Kew: Royal Botanic Gardens.

Lavin, M., E. Eshbaugh, J.-M. Hu, S. Mathews, and R.A. Sharrock. 1998. Monophyletic subgroups of the tribe Millettieae (Leguminosae) as revealed by phytochrome nucleotide sequence data. *American Journal of Botany* 85: 412–433.

Lavin, M., P.S. Herendeen, and M.F. Wojciechowski. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Biology* 54: 575–594.

Lee, J. and T. Hymowitz. 2001. A molecular phylogenetic study of the subtribe Glycininae (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences. *American Journal of Botany* 88: 2064–2073.

Lewis, G., B. Schrire, B. MacKinder, and M. Lock. 2005. *Legumes of the world*. Kew: The Royal Botanic Gardens.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49: 718–726.

Nugent, J. M. and J. D. Palmer. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66: 473–481.

Nylander, J. A. A. 2004. *MrModeltest*, v. 2. Uppsala: Program distributed by the author.

Ohashi, H., R. M. Polhill, and B. G. Schubert. 1981. Desmodieae. Pp. 292–300 in *Advances in Legume Systematics Part 1*, ed. R. M. Polhill, and P. H. Raven. Kew: Royal Botanic Gardens.

Olmstead, R. G., H. J. Michaels, K. M. Scott, and J. D. Palmer. 1992. Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL*. *Annals of the Missouri Botanical Garden* 79: 249–265.

Pennington, R. T., M. Lavin, H. E. Ireland, B. Klitgaard, and J. Preston. 2001. Phylogenetic relationships of primitive papilionoid legumes based upon sequences of the chloroplast intron *trnL*. *Systematic Botany* 26: 537–556.

Polhill, R. M. 1994. Classification of the Leguminosae. Pp. xxv-xlvii in *Plants and their constituents* vol. 1, eds. F. A. Bisby, J. Buckingham, and J. B. Harborne. London: Chapman and Hall.

Rambaut, A. 2002. *Se-Al sequence alignment editor, v. 2.0a11*. Oxford: University of Oxford.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models, version 3.1.1. *Bioinformatics (Oxford, England)* 19: 1572–1574.

Riley-Hulting, E. T., A. Delgado-Salinas, and M. Lavin. 2004. Phylogenetic systematics of *Strophostyles* (Fabaceae): A North American temperate genus within a neotropical diversification. *Systematic Botany* 29: 627–653.

Sambrook, J., P. A. Fritsch, and T. Maniatis. 1989. *Molecular cloning: a laboratory manual*. Ed. 2. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.

Sanderson, M. J. 2003. r8s; inferring absolute rates of evolution and divergence times in the absence of a molecular clock. *Bioinformatics (Oxford, England)* 19: 301–302.

Scherson, R. A., R. Vidal, and M. J. Sanderson. 2008. Phylogeny, biogeography, and rates of diversification of new world *Astragalus* (Leguminosae) with an emphasis on South American radiations. *American Journal of Botany* 95: 1030–1039.

Schlueter, J. A., P. Dixon, C. Granger, D. Grant, L. Clark, J. J. Doyle, and R. C. Shoemaker. 2004. Mining the EST databases to determine evolutionary events in the legumes and grasses. *Genome* 47: 868–876.

Schlueter, J. A., J. Y. Lin, S. D. Schlueter, I. F. Vasylenko-Sanders, S. Deshpande, J. Yi, M. O'Bleness, B. A. Roe, R. T. Nelson, B. E. Scheffler, S. A. Jackson, and R. C. Shoemaker. 2007. Gene duplication and paleopolyploidy in soybean and the implications for whole genome sequencing. *BMC Genomics* 8: 330.

Simpson, B. B. and M. C. Ogorzaly. 2001. *Economic botany: plants in our world*. Third edition. New York: McGraw-Hill, Inc.

Shoemaker, R. C., J. A. Schlueter, and J. J. Doyle. 2006. Paleopolyploidy and gene duplication in soybean and other legumes. *Current Opinion in Plant Biology* 9: 104–109.

Stefanović, S., D. W. Rice, and J. D. Palmer. 2004. Long branch attraction, taxon sampling, and the earliest angiosperms: *Amborella* or monocots? *BMC Evolutionary Biology* 4: 35.

Straub, S. C. K., B. E. Pfeil, and J. J. Doyle. 2006. Testing the polyploid past of soybean using a low-copy nuclear gene–Is *Glycine* (Fabaceae: Papilionoideae) an auto- or allopolyploid? *Molecular Phylogenetics and Evolution* 39: 580–584.

Swofford, D. L. 1998. *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), v. 4.0b10*. Sunderland: Sinauer Associates.

Taberlet, P., L. Gielly, G. Pautou, and J. Bouvet. 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

Thulin, M., M. Lavin, R. Pasquet, and A. Delgado-Salinas. 2004. Phylogeny and biogeography of *Wajira* (Leguminosae): A monophyletic segregate of *Vigna* centered in the horn of Africa region. *Systematic Botany* 29: 903–920.

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.

Appendix 1. Taxa, authorities, source of plant material from which DNA was extracted, and GenBank accession numbers for sequences used in this study. Literature citations are provided for the material vouchered in previously published studies. Accession numbers are given in the following order: *trnL-F, rbcL, atpB, trnK/matK, rpl2, clpP, rps16*, and *ycf4*. Classification by tribe/subtribe follows Lackey (1981), Polhill (1994), and Lewis et al. (2005). Abbreviations of herbaria are according to Index Herbariorum. A dash indicates the sequence not available for the species. CSIRO – Commonwealth Scientific and Industrial Research Organization, Australia; DLEGP – Desert Legume Program; IUGH – Indiana University Greenhouse; SBG – Royal Botanic Garden, Sydney; SRPIS – USDA Southern Regional Plant Introduction Station.

"PHASEOLOID" CLADE: Desmodieae – Subtribe Desmodiinae - *Desmodium* Desv.: *D. barbatum* (L.) Benth., SRPIS-227476; EU717290, EU717279, EU717521, EU717420, EU717386, EU717245, EU717487, EU717455. *D. canescens* (L.) DC., *Buser 2889* (IND); EU717291, —, —, —, —, —, —, —. *D. cuneatum* Hook. & Arn., SRPIS-173341; EU717292, —, —, —, —, —, —, —. *D. cuspidatum* DC. ex Loud., *Stefanović SS-04-122* (TRTE); EU717293, —, —, —, —, —, —, —. *D. elegans* DC., (Pennington et al. 2001), AF309482, —, —, —, —, —, —, —. *D. glutinosum* Schindl., *Ellsworth 60* (IND); EU717294, —, —, —, —, —, —, —. *D. intortum* (Mill.) Urb., SRPIS-295877; EU717295, —, —, —, —, —, —, —. *D. nudiflorum* (L.) DC., *Stefanović SS-03-22* (TRTE); EU717296, —, —, —, —, —, —, —. *D. pauciflorum* (Nutt.) DC., *Stefanović SS-03-27* (TRTE); EU717297, EU717280, EU717522, EU717421, EU717387, EU717246, EU717488, EU717456. Subtribe Lespedizinae - *Campylotropis* Bunge: *C. macrocarpa* (Bunge) Rehder, *Stefanović SS-03-04* (TRTE); EU717298, EU717277, EU717519, EU717418, EU717384, EU717243, EU717485, EU717453. *Kummerowia* Schindl.: *K. stipulacea* Makino, *Stefanović SS-04-148* (TRTE); EU717299, EU717276, EU717518, EU717417, EU717383, EU717242, EU717484, EU717452. *K. striata* (Thunb.) Schindl., SRPIS-186591; EU717300, —, —, —, —, —, —, —. *Lespedeza* Michx.: *L. bicolor* Turcz., SRPIS-286476; EU717301, EU717274, EU717516, EU717415, EU717381, EU717240, EU717482, EU717450. *L. cuneata* G.Don, *Stefanović SS-03-24* (TRTE); EU717302, EU717275, EU717517, EU717416, EU717382, EU717241, EU717483, EU717451. *L. hirta* Hornem., *Stefanović SS-04-150* (TRTE); EU717303, —, —, —, —, —, —, —. *L. intermedia* (S.Watson) Britton, *Stefanović SS-03-25* (TRTE); EU717304, EU717278, EU717520, EU717419, EU717385, EU717244, EU717486, EU717454. *L. intermedia* (S.Watson) Britton, *Stefanović SS-04-152* (TRTE); EU717305, —, —, —, —, —, —, —. *L. japonica* L.H.Bailey, SRPIS-349423; EU717306, —, —, —, —, —, —, —. *L. thunbergii* Nakai, Cultivated (IUGH); EU717307, —, —, —, —, —, —, —.

Phaseoleae – Subtribe Cajaninae - *Atylosia* Wight & Arn.: *A. lineata* Wight & Arn., Doyle et al. (1995); EU717308, —, —, —, —, —, —, —. *Bolusafra* Kuntze: *B. bituminosa* Kuntze, Doyle et al. (1995); EU717309, EU717272, EU717514, EU717413, EU717362, EU717238, EU717480, EU717448. *Cajanus* DC.: *C. cajan* (L.) Millsp., *Stefanović SS-03-02* (TRTE); EU717310, EU717273, EU717515, EU717414, EU717361, EU717239, EU717481, EU717449. *Eriosema* (DC.) G.Don.: *E. psoralioides* Baill., Doyle et al. (1995); EU717311, —, —, —, —, —, —, —. Subtribe Erythrininae - *Apios* Fabr.: *A. americana* Medik., *Yatskievych & McCray 86-129* (IND); EU717312, EU717285, EU717527, EU717426, EU717392, —, —, EU717460. *Erythrina* L.: *E. sousae* Krukoff & Barneby, Doyle et al. (1995); EU717313, EU717270, EU717512, EU717411, EU717377, EU717236, EU717478, EU717446. *E. fusca* Lour., Rainforest Seed Co. 70; EU717314, —, —, —, —, —, —, —. *E. lysistemon* Hutch., Rainforest Seed Co. 22; EU717315, —, —, —, —, —, —, —. *Mucuna* Adans.: *Mucuna sp.,* Doyle et al. (1995); EU717316, EU717281, EU717523, EU717422, EU717388, EU717247, EU717489, EU717457. Subtribe Glycininae-*Amphicarpaea* Elliott ex Nutt.: *A. bracteata* (L.) Fernald, Doyle et al. (1995); EU717317, EU717257, EU717499, EU717399, EU717364, EU717223, EU717468, EU717433. *A. edgeworthii* Benth., Kajita et al. (2001), AF417042, —, —, —, —, —, —, —. *Calopogonium* Desv.: *C. caeruleum* (Benth.) Sauvalle, Doyle et al. (1995); EU717318, —, —, —, —, —, —, —. *Cologania* Kunth.: *C. lemonii* L., Doyle et al. (1995); EU717319, EU717264, EU717506, EU717405, EU717371, EU717230, EU717475, EU717440. *Dumasia* DC.: *D. villosa* DC., Doyle et al. (1995); EU717320, EU717265, EU717507, EU717406, EU717372, EU717231, EU717476, EU717441. *Glycine* Willd.: *G. max* (L.) Merr., Hu et al. (2000); EU717321, EU717256, EU717498, AF142700, EU717363, EU717222, EU717467, U26948. *G. microphylla* (Benth.) Tindale, Doyle et al. (2003); EU717322, —, —, —, —, —, —, —.

G. tomentella Hayata, G1157, CSIRO Perennial *Glycine* Germplasm Collection, AF435924, —, —, —, —, —, —, —. *Neonotonia* J.A.Lackey: *N. wightii* (Arn.) J.A.Lackey, Doyle et al. (1995); EU717323, EU717261, EU717503, EU717402, EU717368, EU717227, EU717472, EU717437. *Pachyrhizus* Rich. ex DC.: *P. erosus* (L.) Urb., Kajita et al. (2001); EU717324, EU717260, EU717502, EU717401, EU717367, EU717226, EU717471, EU717436. *Pseudovigna* (Harms) Verdc.: *P. argentea* (Willd.) Verdc., Doyle et al. (1995); EU717325, EU717262, EU717504, EU717403, EU717369, EU717228, EU717473, EU717438. *Pueraria* DC.: *P. lobata* (Willd.) Ohwi, Doyle et al. (1995); EU717326, EU717259, EU717501, —, EU717366, EU717225, EU717470, EU717435. *P. phaseoloides* Benth., Doyle et al. (2003); EU717327, EU717263, EU717505, EU717404, EU717370, EU717229, EU717474, EU717439. *Shuteria* Wight & Arn.: *S. vestita* Wight & Arn., Doyle et al. (1995); EU717328, EU717282, EU717524, EU717423, EU717389, EU717248, EU717490, —. *Teramnus* P.Browne: *T. labialis* (L.f.) Sprengl., Doyle et al. (1995); EU717329, —, —, —, —, —, —, —. *T. uncinatus* Sw., Cultivated (IUGH); EU717330, EU717258, EU717500, EU717400, EU717365, EU717224, EU717469, EU717434. Subtribe Kennediinae-*Hardenbergia* Benth.: *H. violacea* (Schneev.) Stearn, Doyle et al. (1995); EU717331, EU717284, EU717526, EU717425, EU717391, EU717250, EU717492, EU717459. *H. comptoniana* (Andrews) Benth., Chiltern Seeds 664; EU717332, —, —, —, —, —, —, —. *Kennedia* Vent.: *K. beckxiana* F.Muell., DLEG-970015; EU717333, —, —, —, —, —, —, —. *K. coccinea* Vent., DLEG-960079; EU717334, —, —, —, —, —, —, —. *K. nigricans* Lindl., DLEG-920034; EU717335, EU717283, EU717525, EU717424, EU717390, EU717249, EU717491, EU717458. *K. prostrata* R.Br., DLEG-960017; EU717336, —, —, —, —, —, —, —. *K. retrorsa* Hemsl., SBG 85258; EU717337, —, —, —, —, —, —, —. *Vandasia* Domin: *V. retusa* (Sol. ex Benth.) Domin, Doyle et al. (1995); EU717338, —, —, —, —, —, —, —. Subtribe Phaseolinae - *Dolichos* L.: *D. lablab* L., Cultivated (IUGH); EU717339, EU717267, EU717509, EU717408, EU717374, EU717233, —, EU717444. *Macroptilium* (Benth.) Urb.: *M. atropurpureum* (L.) Urb., *Stefanović SS-03-06* (TRTE); EU717340, EU717268, EU717510, EU717409, EU717375, EU717234, —, EU717443. *Macrotyloma* (Wight & Arn.) Verdc.: *M. uniflorum* (Lam.) Verdc., SRPIS-297892; EU717341, EU717269, EU717511, EU717410, EU717376, EU717235, —, EU717445. *Phaseolus* L.: *P. vulgaris* L., Doyle et al. (2003); EU717342, —, —, —, —, —, —, —. *Psophocarpus* Neck. ex DC.: *P. tetragonolobus* DC., SRPIS-477254; EU717343, EU717271, EU717513, EU717412, EU717378, EU717237, EU717479, EU717447. *Ramirezella* Rose: *R. strobilophora* Rose, Doyle et al. (1995); EU717344, —, —, —, —, —, —, —. *Strophostyles* Elliott: *S. hevola* (L.) Elliott, Doyle et al. (1995); EU717345, —, —, —, —, —, —, —. *Vigna* Savi: *V. radiata* (L.) R.Wilczek, Kajita et al. (2001); EU717346, —, —, —, —, —, —, —. *V. trilobata* (L.) Verdc., Cultivated (IUGH); EU717347, —, —, —, —, —, —, —. *V. unguiculata* (L.) Walp., Wojciechowski et al. (2004); EU717348, EU717266, EU717508, EU717407, EU717373, EU717232, EU717477, EU717442.

Psoraleae – *Bituminaria* Heist. ex Fabr.: *B. bituminosa* (L.) C.H.Stirt., Doyle et al. (1995); EU717349, EU717255, EU717497, EU717398, EU717362, EU717221, EU717466, EU717432. *Cullen* Medik.: *C. australasicum* (Schltdl.) J.W.Grimes, Adams et al. (1999); EU717350, EU717254, EU717496, EU717397, EU717361, EU717220, EU717465, EU717431. *Otholobium* C.H.Strit.: *O. sericeum* (Poir.) C.H.Stirt., Doyle et al. (1995); EU717351, —, —, —, —, —, —, —. *Psoralea* L.: *P. aphylla* L., B&T World Seeds 35039; EU717352, —, —, —, —, —, —, —. *P. lanceolata* Pursh, Doyle et al. (1995); EU717353, —, —, —, —, —, —, —.

OUTGROUPS ("millettioid" and "indigoferoid" legumes)—*Austrosteenisia* Geesink: *A. blackii* (F.Muell.) R.Geesink, (Pennington et al. 2001), AF311381, —, —, —, —, —, —, —. *Canavalia* DC.: *C. ensiformis* (L.) DC., Doyle et al. (1995); EU717354, —, —, —, —, —, —, EU717463. *Clitoria* L.: *C. ternatea* L., Doyle et al. (1995); EU717355, EU717286, EU717528, EU717427, EU717393, EU717251, —, EU717461. *Galactia* P.Browne: *G. striata* (Jacq.) Urban, Doyle et al. (1995); EU717356, EU717287, EU717529, EU717428, EU717394, —, EU717493, EU717462. *Indigofera* L.: *I. longimucronata* Baker f., (Barker et al. 2000), AF274368, —, —, —, —, —, —, —. *Lonchocarpus* Kunth: *L. capassa* Rolfe, B&T World Seeds 448908; EU717357, —, —, —, —, —, —, —. *L. lanceolatus* Benth., (Pennington et al. 2001), AF311382, —, —, —, —, —, —, —. *Ophrestia* H.M.L.Forbes: *O. oblongifolia* (E.Mey.) H.M.L.Forbes, Cultivated (IUGH); EU717358, —, —, —, —, —, —, —. *O. radicosa* (A.Rich.) Verdc., Doyle et al. (1995); EU717359, EU717289, EU717531, EU717430, EU717396, EU717253, EU717495, EU717464. *Piscidia* L.: *P. piscipula* (L.) Sarg., (Pennington et al. 2001), AF311379, —, —, —, —, —, —, —. *Platycyamus* Benth.: *P. regnellii* Benth., (Pennington et al. 2001), AF311378, —, —, —, —, —, —, —. *Tephrosia* Pers.: *T. rhodesica* Baker f., Cultivated (IUGH); EU717360, EU717288, EU717530, EU717429, EU717395, EU717252, EU717494, —.