Phylogeny of the Genus *Phaseolus* (Leguminosae): A Recent Diversification in an Ancient Landscape

ALFONSO DELGADO-SALINAS,1,3 RYAN BIBLER,2 and MATT LAVIN2

1Departamento de Botánica, Instituto de Biología, Universidad Nacional Autónoma de México, Apartado Postal 70-233, 04510, México, D.F. México; 2Department of Plant Sciences and Plant Pathology, Montana State University, Bozeman, Montana 59717 U.S.A.; 3Author for correspondence (adelgado@servidor.unam.mx)

**Communicating Editor: James F. Smith**

**ABSTRACT.** A combined parsimony analysis of the species of *Phaseolus* and closely related New World genera was performed with sequences from the nuclear ribosomal ITS/5.8 S and plastid *trnK* loci. Species relationships are resolved with high parsimony bootstrap support at all hierarchical levels. All species of *Phaseolus*, except five enigmatic ones, belong to one of eight clades. These eight clades show some morphological, ecological, or biogeographical distinction, and are informally recognized in a phylogenetic classification. The five enigmatic species, *Phaseolus glabellus*, *P. macrolepis*, *P. microcarpus*, *P. oaxacanus*, and *P. talamancensis* are weakly resolved as the sister clade to the Tuerckheimii group. An evolutionary rates analysis that biases for old age estimates suggests that the *Phaseolus* stem clade is the same age as the New World Phaseolinae crown clade with a maximum age of ca. 8 Ma. The *Phaseolus* crown is estimated to be no older than ca. 6 Ma, and the average age of the eight well supported crown clades within *Phaseolus* is ca. 2 Ma. The maximum age estimate of a Late Pliocene to Pleistocene diversification of *Phaseolus* post-dates the major tectonic activity in Mexico where *Phaseolus* diversity is centered.

**KEYWORDS:** evolutionary rates, informal classification, Mexico, Phaseolinae, *Phaseolus*, systematics.

The genus *Phaseolus* L. includes at least five species of domesticated beans originally cultivated in Mexico and the central Andes (Gepts 1998; Delgado-Salinas et al. 1999; Gepts et al. 1999). Because of its economic importance, the genus has been the focus of much agronomic, molecular genetic, and systematic study (e.g., Broughton et al. 2003; González-Mejía et al. 2005). Two recent studies produced somewhat conflicting classifications of *Phaseolus* species. A combined phylogenetic analysis of nrDNA ITS/5.8 S (ITS) sequences and morphological characters (Delgado-Salinas et al. 1999) resulted in a classification of approximately 50 species of *Phaseolus* into nine groups; the five domesticated species were placed in two of these. Using ITS sequence data (Gaitán et al. 2000), interbreeding information, and morphological evidence, Freytag and Debouck (2002) monographed 76 species of *Phaseolus* from Central and North America and classified these into 14 groups and two incertae sedis. The five domesticated species belonged to four of these groups.

This study is motivated by the opportunity to resolve some uncertainties remaining within the study of Delgado-Salinas et al. (1999), or inconsistencies that have emerged between the studies of Delgado-Salinas et al. (1999) and Freytag and Debouck (2002). The analysis of Delgado-Salinas et al. (1999), for example, yielded little to no clade support for certain of their species groups (e.g., the Pedicellatus and Tuerckheimii clades). Also, the relationships among their nine species groups were poorly resolved and not well supported. *Phaseolus microcarpus*, for example, was resolved without clade support as a basally branching lineage within *Phaseolus*.

Freytag and Debouck’s (2002) classification, including 22 newly described species, is partly inconsistent with that of Delgado-Salinas et al. (1999). *Phaseolus lunatus*, for example, belongs to the mostly South American Lunatus clade in Delgado-Salinas et al. (1999), but Freytag and Debouck (2002) grouped it with the *P. polystachios* and related North American and Mexican species. Also, *Phaseolus maculatus* and *P. ritensis* are placed in the well-supported monophyletic Polystachios group in Delgado-Salinas et al. (1999), but they are classified as conspecific and separate from the group containing *Phaseolus polystachios* in Freytag and Debouck (2002).

The resolution of such uncertainties and inconsistencies requires additional evidence so that a sound phylogenetic classification of *Phaseolus* species can be achieved. In this regard, an evolutionary rates analysis of the Leguminosae (Lavin et al. 2005) has revealed that *Phaseolus* and closely related genera (subtribe Phaseolinae) have the fastest rates of substitution for the chloroplast *matK* locus, thus rendering it very informative in terms of nucleotide substitution variation at the species level. We exploit this high level of variation in *matK* and flanking *trnK* non-coding intron sequences in a combined phylogenetic analysis with ITS data to produce a more highly resolved
phylogeny with which to address the above issues. In addition, because the study of evolutionary ages of such crop-containing lineages is of general interest (e.g., Gepts et al. 1999), we undertake an evolutionary rates analysis and make independent estimates for the ages derived from ITS and trnK sequences for the stem and crown clade of *Phaseolus* and its constituent lineages.

**Materials and Methods**

**Taxon Sampling.** With the goal of refining the infrageneric classification of *Phaseolus*, an effort was made to sample as many species as possible of the genus. Multiple accessions of many species were sampled to validate interspecific relationships. The extensive sampling in the study by Delgado-Salinas et al. (1999) that included 52 *Phaseolus* species represented by 115 ITS sequences is augmented in this study to a total of 70 *Phaseolus* species represented by 143 ITS and 71 trnK sequences (Appendix 1). All 14 species groups and two incertae sedis, *Phaseolus glabellus* and *P. microcarpus*, treated in Freytag and Debouck (2002) were sampled, with the exception of their monotypic sect. *Revoluti*, which includes the probably extinct *P. leptophyllus*. Of the 22 new species recognized in Freytag and Debouck (2002), all were sampled in this analysis or are considered synonyms of sampled species; these are indicated in the Discussion under each *Phaseolus* group. Only two species, *P. nilotica* and *P. leptophyllus*, were not sampled for sequence data.

The outgroups include seven other genera of New World Phaseolinae (Lackey 1981, 1983; Lewis and Delgado-Salinas 1994), which collectively form a well-supported clade with *Phaseolus* (Delgado-Salinas et al. unpublished data). A total of 14 ITS and 14 trnK sequences (Appendix 1) represent the outgroup genera *Dolichopis* Hassl., *Macrotilium* (Benth.) Urb., *Mysanthus* G.P. Lewis & A. Delgado, *Oxyrhynchus* Brandegee, *Ramirezella* Rose, *Strophostyles* Elliott, and *Vigna* Sav subgenus *Sigmoidotropis* (Piper) Verdc. *Oxytis* A. Delgado & G. P. Lewis was the only genus of New World Phaseolinae for which DNA sequences have yet to be acquired (Riley-Hulting et al. 2004). Because a phylogenetic analysis of *Vigna* sensu lato resolves all Old World *Vigna* and pantropical *Vigna* subgenus *Lasiospon* as sister to the New World Phaseolinae (Delgado-Salinas et al. unpublished data), additional outgroups included two species of Old World *Vigna*, *Vigna unguiculata* (subgenus *Vigna*) and *V. radiata* (subgenus *Ceratotropis* (Benth.) Verdc.), and two species of *Vigna* subgenus *Lasiospon* (Benth.) Verdc., *Vigna longifolia* and *V. trichocarpa*. These four species are represented by one ITS and one trnK sequence each (Appendix 1).

**Phylogenetic Data.** DNA sequences from the nuclear ribosomal 5.8 S and flanking internal transcribed spacers (the ITS region) and the chloroplast trnK intron including *matK* (the trnK locus) were analyzed because many legume studies have shown how phylogenetically informative these loci are (Delgado-Salinas et al. 1999; Riley-Hulting et al. 2004; Thulin et al. 2004; Lavin et al. 2005). PCR primers for the ITS regions are described in Delgado-Salinas et al. (1999) and those for the trnK locus are described in Riley-Hulting et al. (2004). Paralogy in the ribosomal repeats of legumes (e.g., Bailey et al. 2003; Hughes et al. 2006) is not known to cause problems in papilionoid studies. In this regard, all species of *Phaseolus* are diploid (*2n = 20 or 22; Mercado-Ruaro and Delgado-Salinas 1998) and related genera such as *Glycine* have been shown to have all ribosomal repeats localized on the short arm of one chromosome (Kollipara et al. 1997). Concerted evolution is thus potentially rapid and complete enough to render a single ribosomal repeat in *Phaseolus* and close relatives (Sanderson and Doyle 1992). For legume groups where paralogous ITS products are commonly amplified, they are readily identified as pseudogenes by the numerous small insertion-deletion regions that occur even in the 5.8 S region (e.g., Hughes et al. 2003).

DNA isolations, polymerase chain reaction (PCR) amplifications, and template purifications were performed with Qiagen Kits (i.e., QIAexpress Plant Mini Kit, Taq PCR Core Kit, QIAquick PCR Purification Kit; Qiagen, Santa Clarita, California, USA). DNA sequencing was performed on an ABI 377 sequencer at Northwoods DNA (http://www.nwdna.com/). Sequences were aligned manually with Se-Al (Rambaut 1996). Multiple alignments of the ITS region were evaluated with parsimony analyses and only results not influenced by alignment variation are reported. The ITS data set included 166 sequences and 812 sites in one alignment configuration. The trnK data set had 91 sequences by 2692 unequivocally aligned sites. The combined data matrix comprised 83 terminal taxa by 3553 sites in one alignment configuration. Missing entries amounted to 0.5% for the ITS data set and 5.1% for the trnK. Data are deposited with TreeBASE (study number S1553).

The morphological data set developed in Delgado-Salinas et al. (1999) was not used in the present study. With the addition of more terminal taxa and a reassessment of character-state assignments, multistate taxon designations became too abundant and precluded phylogenetic resolution.

**Phylogenetic Analysis.** Maximum parsimony analyses were performed with PAUP* (Swofford 2002). Heuristic searches included 100 random addition replicates, tree-bisection-reconnection, and retention of multiple parsimony trees. A maximum of 10,000 trees was allowed to accumulate because Sanderson and Doyle (1993) have shown that searches generating more than 1,000 trees yield diminishing returns in topological variation. Bootstrap analysis and partition homogeneity tests involved re-sampling with replacement (Felsenstein 1985; Sanderson 1995), where 10,000 replicates were each subjected to random addition of taxa, tree-bisection-reconnection, and invoking neither steepest descent nor retention of multiple parsimonious trees. Phylogenetic analyses were carried out independently on sequences from the ITS region and the trnK locus. Problems with the partition homogeneity test (e.g., Mason-Gamer and Kellogg 1996) include one of the data sets yielding poor resolution and the other highly resolved trees (Graham et al. 1998) or varying substitution rates between two data sets (Johnson and Whiting 2002), and thus not necessarily real data conflict. A conditional combination approach (Bull et al. 1993) was therefore taken, where conflict was assessed only among clades with bootstrap support greater than 75%.

**Evolutionary Rates Analysis.** A Bayesian phylogenetic approach was used to generate a set of phylogenetic trees with estimated branch lengths that could then be converted to time in a rates analysis. MrBayes version 3.1 (Huelsenbeck and Ronquist 2001) was used to search tree parameter space. A Metropolis-coupled markov chain monte carlo permutation of tree parameters was initiated with a random tree and four chains set at default temperatures (Huelsenbeck et al. 2001), and a nucleotide substitution model selected via the Akaikie information criterion (AIC) implemented in ModellTest (Posada and Crandall 1998) for nested models, or manually for non-nested models (Johnson and Omland 2004). AIC was performed manually using AIC = [−2(lnL)] + 2 K, where K is the number of parameters in the model (Barnham and Anderson 2002). A model having the lowest AIC value by over 2 units was considered the best fit (Barnham and
Markov chains were run for at least $5 	imes 10^6$ permutations of tree parameters, and sampled every $5 	imes 10^5$ permutations such that sampling yielded 100 parsimony trees that excluded the burn-in and autocorrelated trees. Because Bayesian posterior credibility values are often biased high (e.g., Yoshiyuki et al. 2002), we used instead the more conservative maximum parsimony bootstrap analysis to identify instances of clade conflict between the ITS and trnK analysis (cf., Douady et al. 2003).

The program r8s (Sanderson 2004) was used to estimate nucleotide substitution rates and ages of crown clades, as described in Thulin et al. (2004) and Lavin et al. (2005). Absolute rates and ages were obtained by constraining the age of the root of the *Phaseolus-Vigna* crown clade to 11 Ma. The most recent common ancestor of *Phaseolus coccineus* (a member of the New World Phaseolinae clade) and *Vigna subterranea* (a member of the Old World Vigna clade) has an estimated age of 8.0 ± 0.8 Ma, and a range of 6.4–10.4 Ma (Lavin et al. 2005). To bias our estimates towards maximum ages, we rounded the oldest age estimate for the *Phaseolus-Vigna* crown clade to 11 Ma. Means and standard deviations of substitution rates and ages of specified clades were obtained from the input of 100 Bayesian trees. Age estimates were primarily derived via the penalized likelihood (PL) method (Sanderson 2002), which was then compared to the rate constant (LF; Langley and Fitch 1974) and the highly rate variable nonparametric rate smoothing (NPRS) methods (Sanderson 2002). The ITS and trnK data sets were analyzed separately in order to make independent age estimates for each crown clade.

**RESULTS**

**Maximum Parsimony Analyses.** Analysis of the ITS data set yielded the maximum 10,000 trees, each with a length of 1875, CI = 0.457, and RI = 0.861. This data set included 400 parsimony informative sites. Analysis of the trnK data set yielded the maximum 10,000 trees, each with a length of 1082, CI = 0.687, and RI = 0.867. This data set included 368 parsimony informative sites. The trnK analysis yielded a more robust resolution at all phylogenetic levels, whereas the ITS data resolved well only the distal clades. Regardless, no clade with greater than 75% parsimony bootstrap support conflicted between the phylogenies resulting from individual ITS and trnK analyses. Analysis of the combined data set yielded the maximum 10,000 trees, each with CI = 0.564 and RI = 0.802. This data set included 729 parsimony informative sites.

The combined data analysis resolved a monophyletic *Phaseolus* with two primary lineages (clades A-B; Fig. 1) that further ramify into eight species clades and four independent species, *Phaseolus glabellus*, *P. macrolepis*, *P. microcarpus*, and *P. oaxacanus* (Fig. 1). All eight clades are resolved with over 95% bootstrap support.

**Evolutionary Rates Analysis of trnK and ITS Sequence Data.** A molecular clock was rejected for each of the analyses of the trnK (LR = 251.61, df = 87, $p < 0.00001$) and ITS (LR = 553.92, df = 159, $p < 0.00001$) data sets. In addition, the nucleotide substitution model selected using AIC for each of the trnK and ITS data sets was the general time reversible with a gamma distribution for variable sites and a proportion for invariant sites (GTR+G+I). For the trnK analysis, a mixed model invoking the general time reversible with site specific substitution rates (GTR+SS) for the coding region was used to test 200 AIC units higher than the GTR+G+I model applied to the entire sequence.

The r8s analysis of 100 trnK Bayesian trees resulted in PL rate estimates of $2.5–4.1 	imes 10^{-9}$ (Table 1) and an LF estimate of $3.6 	imes 10^{-9}$ substitutions per site per year (Fig. 2a). NPRS estimates were highly variable, $1.8–8.7 	imes 10^{-9}$ substitutions per site per year, which averaged faster than but were positively correlated with the PL estimates (Fig. 2a).

The r8s analysis of 100 ITS Bayesian trees resulted in PL estimates of $69.0–125.9 	imes 10^{-9}$ (Table 1) and an LF estimate of $83.8 	imes 10^{-9}$ substitutions per site per year (Fig. 2b). NPRS resulted in highly variable rates, $70.4–253.5 	imes 10^{-9}$ substitutions per site per year, which were uncorrelated with PL estimates (Fig. 2b).

**Age-estimation.** The PL estimated ages for the various crown clades identified in Fig. 1 are distributed between approximately 1 and 8 Ma in the trnK analysis (Fig. 2c), and 0.6 and 7 Ma for the ITS analysis (Fig. 2d). Older age estimates are derived from the trnK data (Table 1), which may be related to the closer similarity between PL and NPRS estimates (Fig. 2c). The younger ITS age estimates may be related to the greater similarity in rate estimates derived from PL and LF (Fig. 2d). The age of the *Phaseolus* stem clade is equivalent in age to the New World Phaseolinae crown clade (Table 1; Fig. 3) at 6–8 Ma. Of the eight principle crown clades within *Phaseolus*, the oldest is the Vulgaris group at ca. 4 Ma, whereas the youngest includes the Filiformis, Pedicellatus, and Polystachios groups at close to 1 Ma (Table 1). Notably, the Pedicellatus and Polystachios groups are the most species rich *Phaseolus* groups. In biasing toward maximum age estimates, the average age of the eight crown clades is only ca. 2 Ma.

**DISCUSSION**

In agreement with other molecular analyses that have sampled extensively among the New World Phaseolinae (e.g., Bruneau et al. 1995; Delgado-Salinas et al. 1993, 1999), the monophyly of *Phaseolus* is unequivocal and diagnosed not only by molecular characters but also by morphological synapomorphies. These include the tightly and laterally coiled beak of the keel petals, inflores-
FIG. 1. One of 10,000 maximum parsimony trees from a combined analysis of trnK and ITS sequences sampled from Phaseolus and outgroups. Bootstrap values greater than 75% are reported above (or below) the branch when resolved in the strict consensus. Dashed lines indicate collapsed branches in the strict consensus. Clades A and B represent the two sister clades within Phaseolus that are for the first time resolved in this study. The other eight clades represent groups recognized by Delgado-Salinas et al. (1999). Species not resolved in one of the eight species clades are shown in the box.
cences lacking swollen nodes (extrafloral nectaries), mostly persistent primary floral bracts, and foliage and reproductive parts bearing uncinate hairs (Delgado-Salinas 1985; Delgado-Salinas et al. 1999). The relationships of *Phaseolus* to other New World Phaseolinae have remained uncertain until this study. The emerging picture detected in this and a more global analysis including all Phaseolinae (Delgado-Salinas et al. unpublished data) is that *Phaseolus* is one of the early branching clades within the New World Phaseolinae radiation (Fig. 1) and never resolved as sister to a particular subset of other New World genera. The *Phaseolus* stem clade is coeval with the New World Phaseolinae crown clade (Figs. 1, 3). Specific intergeneric relationships of these New World genera are addressed elsewhere (Riley-Hulting et al. 2004; Delgado-Salinas et al. unpublished data).

**The Two Principal Clades of Phaseolus.** With the intergeneric relationships of *Phaseolus* firmly resolved with other New World Phaseolinae, an unequivocal root of the *Phaseolus* crown clade is for the first time established. Instead of *Phaseolus microcarpus* being resolved as sister to the rest of the genus (Delgado-Salinas et al. 1999), all species of *Phaseolus* belong to one of two sister clades. Clade A comprises the Pauciflorus, Pedicellatus, and Tuerckheimii groups, and the weakly resolved species (*Phaseolus glabellus, P. macrolepis, P. microcarpus*, and *P. oaxacanus*), whereas clade B comprises the Filiformis, Vulgaris, Lunatus, Leptostachys, and Polystachios groups (Fig. 1).

Clade A (Fig. 1) species are specifically and ecologically limited compared to those in Clade B. Clade A species are distributed mostly in Mexico, but also adjacent southwestern Arizona, southern New Mexico, and Texas (i.e., *P. grayanus* and *P. parvulus*), and south to northern Panama (i.e., *P. tuerckheimii*). They occur neither in South America nor on oceanic islands. The species of clade A are confined to higher elevations in oak, pine-oak, and pine forests and cloud forests (i.e., well over 1200 m), with the exception of *P. microcarpus*. Species distributions are narrower in clade A with an average elevation window of 690 ± 545 m and an average latitude window of 4° 26’ ± 5° 06’ (Delgado-Salinas unpublished data). Species of clade A flower only during the rainy season (except *P. microcarpus*, which flowers during both the dry and wet season), are sensitive to habitat disturbance, and usually do not tolerate a long frost period. No domesticated taxa occur among...
The clade A species, suggesting that the relative rarity of these species might be the cause of them not being selected for domestication. Finally, clade A species are rarely associated with infraspecific taxa. *Phaseolus parvulus*, *P. pedicellatus*, and *P. polymorphus* are the only clade A species that have traditionally included at most a few infraspecific taxa (e.g., Delgado-Salinas 1985; Freytag and Debouck 2002).

Clade B species are distributed from southeastern Canada south through eastern USA and across southern USA to southeastern California, throughout Mexico and Central America, and in the Andean region of South America. Clade B species are the only ones in the genus to inhabit islands. For example, *Phaseolus lignosus* is endemic to the Bermudas, *P. nollis* to the Galapagos, *P. lunatus* in the West Indies, and *P. lunatus, P. filiformis*, and *P. acutifolius* on several Mexican Pacific islands. They are broadly distributed elevationally throughout this range, from lowland dry and wet forests up to pine-oak and pine forests. Clade B species have a broader elevation window of 737 ± 654 m, and a broader latitude window of 6° 22’ ± 10° 21’ (Delgado-Salinas unpublished data). Species of clade B collectively flower during either the dry or rainy season, are mostly not sensitive to disturbance, and some can tolerate a long frost period (e.g., *P. coccineus, P. angustissimus*). The five main domesticated species (i.e., *P. acutifolius, P. coccineus, P. dumosus, P. lunatus*, and *P. vulgaris*) occur among the clade B species, as do other species that show features of incipient domestication (i.e., populations of *P. maculatus* and *P. polystachios* with tardily dehiscent pods bearing particularly large seeds). The commonness of clade B species in part may have facilitated discovery for domestication. Most *Phaseolus* species traditionally comprising infraspecific taxa, notably *Phaseolus acutifolius, P. coccineus, P. leptostachyus, P. lunatus, P. maculatus, P. polystachios*, and *P. vulgaris*, are from clade B. During the last decade, 16 species and 1 variety have been described from clade A, whereas 17 species and 24 varieties have been described from clade B (Debouck unpublished data; Freytag and Debouck 1996, 2002; Delgado-Salinas 2000; Torres-Gonzales et al. 2001).

### TABLE 1. Penalized likelihood estimated rates and ages from trnK and ITS/5.8 S sequences for the crown clades identified in Figs. 1 and 3. Rates are reported as substitutions per site per million years. Ages are reported in millions of years. Means, standard deviations (std), and ranges were derived from 100 Bayesian trees. The optimal smoothing parameter for both the trnK and ITS data was 10^{-0.5}. *Crown clade B in the ITS phylogeny is equivalent to the Phaseolus crown.*

<table>
<thead>
<tr>
<th>Crown clade</th>
<th>trnK Mean rate (± std)</th>
<th>ITS/5.8 S Mean rate (± std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New World Phaseolinae</td>
<td>0.00405 ± 0.00027</td>
<td>0.06902 ± 0.00942</td>
</tr>
<tr>
<td><em>Phaseolus</em></td>
<td>0.00395 ± 0.00033</td>
<td>0.09783 ± 0.01064</td>
</tr>
<tr>
<td>Clade A</td>
<td>0.00367 ± 0.00046</td>
<td>0.11556 ± 0.01166</td>
</tr>
<tr>
<td>Clade B</td>
<td>0.00328 ± 0.00047</td>
<td>0.11578 ± 0.01226</td>
</tr>
<tr>
<td>Pauiciflorus</td>
<td>0.00341 ± 0.00076</td>
<td>0.12590 ± 0.01285</td>
</tr>
<tr>
<td>Pedicellatus</td>
<td>0.00340 ± 0.00086</td>
<td>0.12256 ± 0.01245</td>
</tr>
<tr>
<td>Tuerckhiemii</td>
<td>0.00275 ± 0.00064</td>
<td>0.10867 ± 0.01155</td>
</tr>
<tr>
<td>Filiformis</td>
<td>0.00251 ± 0.00057</td>
<td>0.09915 ± 0.01086</td>
</tr>
<tr>
<td>Vulgaris</td>
<td>0.00260 ± 0.00047</td>
<td>0.10448 ± 0.01198</td>
</tr>
<tr>
<td>Leptostachyus</td>
<td>0.00303 ± 0.00056</td>
<td>0.10203 ± 0.01172</td>
</tr>
<tr>
<td>Lunatus</td>
<td>0.00311 ± 0.00064</td>
<td>0.11252 ± 0.01172</td>
</tr>
<tr>
<td>Polystachios</td>
<td>0.00320 ± 0.00080</td>
<td></td>
</tr>
</tbody>
</table>

### Rates and Ages from trnK and ITS Data

- **New World Phaseolinae**: 0.00405 ± 0.00027 (trnK), 0.06902 ± 0.00942 (ITS/5.8 S).
- **Phaseolus**: 0.00395 ± 0.00033 (trnK), 0.09783 ± 0.01064 (ITS/5.8 S).
- **Clade A**: 0.00367 ± 0.00046 (trnK), 0.11556 ± 0.01166 (ITS/5.8 S).
- **Clade B**: 0.00328 ± 0.00047 (trnK), 0.11578 ± 0.01226 (ITS/5.8 S).
- **Pauiciflorus**: 0.00341 ± 0.00076 (trnK), 0.12590 ± 0.01285 (ITS/5.8 S).
- **Pedicellatus**: 0.00340 ± 0.00086 (trnK), 0.12256 ± 0.01245 (ITS/5.8 S).
- **Tuerckhiemii**: 0.00275 ± 0.00064 (trnK), 0.10867 ± 0.01155 (ITS/5.8 S).
- **Filiformis**: 0.00251 ± 0.00057 (trnK), 0.09915 ± 0.01086 (ITS/5.8 S).
- **Vulgaris**: 0.00260 ± 0.00047 (trnK), 0.10448 ± 0.01198 (ITS/5.8 S).
- **Leptostachyus**: 0.00303 ± 0.00056 (trnK), 0.10203 ± 0.01172 (ITS/5.8 S).
- **Lunatus**: 0.00311 ± 0.00064 (trnK), 0.11252 ± 0.01172 (ITS/5.8 S).
- **Polystachios**: 0.00320 ± 0.00080 (trnK), 0.11252 ± 0.01172 (ITS/5.8 S).

**Mean Ages (± std)**

<table>
<thead>
<tr>
<th>Crown clade</th>
<th>Mean age (± std)</th>
</tr>
</thead>
<tbody>
<tr>
<td>New World Phaseolinae</td>
<td>7.7 ± 0.5</td>
</tr>
<tr>
<td><em>Phaseolus</em></td>
<td>5.8 ± 0.6</td>
</tr>
<tr>
<td>Clade A</td>
<td>4.9 ± 0.7</td>
</tr>
<tr>
<td>Clade B</td>
<td>5.0 ± 0.7</td>
</tr>
<tr>
<td>Pauiciflorus</td>
<td>3.2 ± 0.8</td>
</tr>
<tr>
<td>Pedicellatus</td>
<td>1.0 ± 0.4</td>
</tr>
<tr>
<td>Tuerckhiemii</td>
<td>3.0 ± 0.8</td>
</tr>
<tr>
<td>Filiformis</td>
<td>1.1 ± 0.5</td>
</tr>
<tr>
<td>Vulgaris</td>
<td>3.9 ± 0.7</td>
</tr>
<tr>
<td>Leptostachyus</td>
<td>3.4 ± 0.6</td>
</tr>
<tr>
<td>Lunatus</td>
<td>1.9 ± 0.5</td>
</tr>
<tr>
<td>Polystachios</td>
<td>1.4 ± 0.5</td>
</tr>
</tbody>
</table>

**Mean Age (± std)**

- **New World Phaseolinae** 6 ± 0.3
- ***Phaseolus*** 6 ± 0.3
- **Clade A** 6 ± 0.3
- **Clade B** 6 ± 0.3
- **Paiuciflorus** 6 ± 0.3
- **Pedicellatus** 6 ± 0.3
- **Tuerckhiemii** 6 ± 0.3
- **Filiformis** 6 ± 0.3
- **Vulgaris** 6 ± 0.3
- **Leptostachyus** 6 ± 0.3
- **Lunatus** 6 ± 0.3
- **Polystachios** 6 ± 0.3

**Using Informal Species Groups.** In contrast to the two primary *Phaseolus* clades that were previously undetected, the eight secondary species clades (labeled in Figs. 1, 3) are nearly the same ones resolved by Delgado-Salinas et al. (1999). The monophyly of each, as well as their interrelationships, are now well supported by bootstrap support values over 95% (Fig. 1). Excepting *P.*
Fig. 3. PL rate smoothed Bayesian consensus phylogeny derived from the trnK locus. Time lines are derived from the root that is fixed at 11 Ma. Posterior probabilities for all branches are mostly 100% with none below 90%.
P. palmeri clade. T. P. glabellus is distinguished from Phaseolus.
M. differs from AND re-
P. zimapanensis P. talamancensis Phaseolus albiflorus, P. Frey-
P. leptophyllus P. oaxacanus by Delgado-Salinas (1985). The com-
of Delgado-Salinas 1985). The petals of have a tight
B. (S

The Pauciflorus Group (sect. Minkelersia and Revolleti). This clade is diagnosed by small
globe to napiform roots, inflorescences with few flowering nodes (usually 1–3), pedicels shorter than calyx tubes bearing early caducous and often inconspicuous bracteoles, calyx lobes often longer than the tube, internal surface of the calyx always covered with uncinate hairs, often elongated petals that form a tubular structure, and fruits often with numerous small seeds not over 2.5 mm long. Although Phaseolus leptophyllus remains unsampled for DNA sequences, Delgado-Salinas (2000) included this species with the Pauciflorus group because it fits the above diagnostis. The Pauciflorus group is distributed mainly in undisturbed pine-oak forests of Mexico, and barely enters Guatemala and southwestern USA. This group comprises Phaseolus amblyosepala, P. anisophyllus (Piper) Freytag & Debouck [=P. amabilis Standl.], P. leptophyllus G. Don, P. nelsonii, P. parvulus, P. pauciflorus, Phaseolus perplexus, P. plagioclyx, P. pluriflorus, and P. tenellus.

The Tuerckheimii Group (sects. Brevilegumin, Chiapasani, Xanthotrichi). This clade is diagnosed by only molecular characters. It comprises florally diverse morphologies. The flowers of Phaseolus esquincensis, P. gladiolatus, P. hintonii, P. xanthotrichus, and P. zimapanensis have a tight lateral coil of the keel petals that spirals backward in the direction of the plane of the keel (sect. Xanthotrichi of Delgado-Salinas 1985). The petals of Phaseolus chiapasani are exceptionally large (up to 3 cm long) and turn black upon drying (sect. Chiapasani of Delgado-Salinas 1985). The flowers of the remaining species have petals predominately with distinctive bluish pigments and stigmatic regions of diverse positions on the style. This assemblage is distributed throughout Mexico and Central America except Belize. The constituent species are P. campanulatus, P. chiapasani, P. esquincensis, P. gladiolatus, P. hintonii [=P. magnio-
batus Freytag & Debouck], P. oligospermus, P. tuerckheimii, P. xanthotrichus, and P. zimapanensis.

The Pedicellatus Group (sects. Digitati and Pedicellati, and P. dasycarpus of Paniculati subsect. Volubili). This clade cannot be diagnosed by only morphological characters. Delgado-Salinas et al. (1999) included P. glabellus in this group, whereas Freytag and Debouck (2002) excluded it because characteristics otherwise not found in the Pedicellatus group, particularly the reddish corollas of P. glabellus. Phaseolus oaxacanus was included in the Pedicellatus group by both Delgado-Salinas (1985) and Freytag and Debouck (2002), and was even ranked as a variety under P. pedicellatus by Delgado-Salinas (1985). The combined morphological and ITS analysis of Delgado-Salinas et al. (1999) placed P. oaxacanus as a tentative member of the P. pedicellatus clade. Indeed, Phaseolus oaxacanus is distinguished from P. pedicellatus only by trifid inflorescence bracts, fewer ovules per ovary (2–3), paniculate inflorescences, and a more southern isolated distribution in the state of Oaxaca (Delgado-Salinas 1985). The present combined analysis is unequivocal in excluding both P. glabellus and P. oaxacanus from the Pedicellatus group. The inclusion of P. dasycarpus within this group is discussed elsewhere (Mercado-Ruaro et al. in press). This assemblage is distributed mainly in central and northern Mexico and adjacent Texas, and southern New Mexico and Arizona. This group includes Phaseolus albiflorus, P. altimontanus, P. dasycarpus, P. esperanzae, P. grayanus [=P. pyramidalis Freytag, P. palmeri Piper, P. telulensis Freytag], P. laxiflorus, P. neglectus [=P. albiolivaceus Freytag & Debouck, P. trifidus Frey-
tag], P. pedicellatus [P. purpussii Brandegee, P. scabrellus Benth. ex S. Wats.], and P. polymorphus.

Unresolved Clade A Species (sects. Bracteati and Pedicellati pro parte, and Species incertae sedis). In addition to Phaseolus glabellus, P. macrolepis, P. microcarpus, and P. oaxacanus, the central Costa Rican P. talamancensis is included here. Phaseolus macrolepis from southern Guatemala and P. talamancensis share a unique inflorescence characterized by very large floral bracts. Indeed, P. talamancensis differs from P. macrolepis only by its floral bracts that are broader than long, and these two species are the only members of Freytag and Debouck’s (2002) sect. Bracteati. The ITS sequence of P. talamancensis (AF115246) is very similar to those of P. macrolepis (DQ445752, DQ445753) but is missing nearly all of the ITS1 region such that it is resolved in a phylogenetic analysis close to the other “unresolved clade A species” but not as sister to P. macrolepis (results not shown). Freytag and Debouck (2002) placed Phaseolus oaxacanus in sect. Pedicellati, whereas P. glabellus and P. micro-
carpus were relegated to incertae sedis. However,
these three species might be shown to comprise a single clade because of subtleties in the form and texture of leaves. These three tend to have broadly ovate membranous leaflets with slightly acuminate apices, which differ slightly from all other Phaseolus species. Phaseolus oaxacanus is confined to the Sierra de Juárez in Oaxaca, P. glabellus occurs from southern Neuvo León and Tamaulipas to central Chiapas, and P. microcarpus is distributed from Durango south to Nicaragua.

The Lunatus Group (Sect. Paniculati, Subsect. Volubili Pro Parte). This clade includes the only South American radiation and oceanic island species of Phaseolus, and is generally diagnosed by falcate pods and seeds with lines radiating from the hilum along the surface of the testa. This assemblage includes endemics to the Andes (Phaseolus augusti, P. bolivianus, and P. pachyrhizoides, the Bermudas (P. lignosus), and the Galapagos (P. mollis), but has widespread species that reach the Revillagigedo Islands and the West Indies, as well as throughout Mexico, Central America, and elsewhere in South America (Freyre et al. 1996; Caicedo et al. 1999; Delgado-Salinas et al. 1999). The constituents include Phaseolus augusti, P. bolivianus, P. lignosus, P. lunatus, P. mollis, P. pachyrhizoides, and P. viridis (=P. longiplicatiff Freytag & Debouck).

The Filiformis Group (Sect. Rugosi). This clade is diagnosed by a tuberculate seed coat (Delgado-Salinas 1985) and small flowers less than 1.3 cm long. Somewhat similar rugose seeds occur in Phaseolus macvaughii and P. microcarpus (Delgado-Salinas 1985). This group is distributed from Baja California to Coahuila, and adjacent southern California. The constituent species are Phaseolus angustissimus, P. carterae, and P. filiformis.

The Vulgares Group (Sects. Acutifoli, Cocci-nei, and Phaseoli, and P. persistens of Sect. Falcati). This clade includes four of the five cultivated species, and cannot be diagnosed by morphological characters. Wide bracteoles bearing three or more nerves, however, mark all but P. acutifolius and P. parvifolius. The segregation of Phaseolus parvifolius from P. acutifolius (Freytag and Debouck 2002) appears justified (e.g., Fig. 1), but should be validated with additional genetic sampling from throughout the ranges of these two species (southwestern USA to Guatemala). Phaseolus persistens was described by Freytag and Debouck (2002) from a single specimen bearing wide bracteoles and unique short pods. They included it in their sect. Falcati close to P. leptostachyus, P. macvaughii, and P. micranthus. The combined analysis clearly shows P. persistens to be very closely related to P. vulgaris (Fig. 1), even though it is the only species in this analysis to be missing the trnK sequence. Species of this clade are distributed throughout Mexico, Central America, and Andean South America. This group comprises P. acutifolius, P. albescens, P. cocineus, P. costaricensis, P. dumasus, P. parvifolius, P. persistens, and P. vulgaris.

The Leptostachyus Group (Sect. Falcati Excluding P. persistens). This clade is diagnosed by an aneuploid chromosome number of $2n = 20$ (Mercado-Ruaro and Delgado-Salinas 1998; Delgado-Salinas et al. 1999). It barely enters southwestern USA (Arizona), but otherwise occurs throughout Mexico, Guatemala, El Salvador, Honduras, Nicaragua, and northwestern Costa Rica. This clade includes P. leptostachyus [=P. opacus Piper], P. macvaughii, and P. micranthus.

The Polystachios Group (Sects. Coriacei, and Paniculati subsects. Volubili and Lignosi). This clade is diagnosed by inflorescences of mostly panicles, callosities on the standard petal associated with the nectar guide, and pollen mostly with pseudocolpi. Debouck (1991) and Freytag and Debouck (2002) suggest a close relationship of Phaseolus lunatus to the Polystachios group, and the results here show the Lunatus group, including P. lunatus, to be sister to the Polystachios group (Fig. 1). The latter is distributed from southeastern Canada, throughout the eastern seaboard of the USA to east Texas, and throughout Mexico south to Oaxaca. This group is the most species-rich and comprises P. albinervus, P. jaliscanus [=P. scrobiculatifolius Freytag], P. juquelinesis [=P. acinciformis Freytag & Debouck], P. maculatifolius, P. maculatus [=P. venosus Piper], P. mearchallii, P. nodosus, P. novoleonenensis, P. polystachios, P. reticulatus, P. ritensis, P. rotundatus, P. salicifolius, P. simunus, P. smilacifolius, P. sonorensis, and P. xolocotzii.

Ages of the Phaseolus Clades. Given the Mexican center of diversity for Phaseolus, the geological history of this region might be relevant to the evolution of the genus. According to Nieto-Samaniego et al. (1999) and Alva-Valdivia et al. (2000), mountain building in Mexico achieved its present-day form by the Late Miocene (5 Ma) with a final major event of subduction vulcanism resulting in the modern Trans-Mexican Volcanic Belt (TMVB). Penalized likelihood age estimates (Table 1) strongly suggest that Phaseolus diversified with the formation of the modern TMVB. The age of the Phaseolus stem clade averages about 6–8 Ma, and the difference between the origin (i.e., Phaseolus stem clade) and the extant diversification (i.e., Phaseolus crown clade) could be as little as 1–2 Ma (Table 1). The approximately 2 Ma average age for the eight species clades within Phaseolus (Table 1),
however, reveals that most of the extant diversity came into existence well after the completion of tectonic activity in Mexico. The formation of such mountains as the TMVB perhaps facilitated the diversification of Phaseolus in upland regions, where Phaseolus species are today most abundant in oak, pine-oak, and pine forests.

The 11 Ma fixed age of the Phaseolinae root (Fig. 3) biases the molecular age estimates toward older ages. If this root age had been fixed at 8 Ma, the average estimated in Lavin et al. (2005) for the Phaseolinae crown clade, then all other age estimates reported here would be distinctly younger. This supports the hypothesis that the predominant modern day predilection of Phaseolus for upland oak, pine-oak, and pine forests evolved well after the formation of these upland habitats themselves.

ACKNOWLEDGEMENTS. This study was supported by the National Science Foundation (DEB-0075202), CONACYT and DGAPA (UNAM). Cindy Erickson and Tom Turley provided technical lab assistance, and L. Torres Colin and M. R. García Peña at MEXU helped with databasing and curatorial aspects of this study. Curators of the following herbaria kindly made loans available for morphological and molecular studies: CAS, F, FLAS, K, MEXU, MONT, NY, OXF, PHILA, and US. CIAT and INIFAP made their germplasm collections readily available for this study. Jim Smith, Pat Herendeen, and two anonymous referees provided valuable suggestions that improved an earlier version of this manuscript.

LITERATURE CITED


———, R. J. EASTWOOD, and C. D. BAILEY. 2006. From famine to feast? Selecting nuclear DNA sequence loci for plant


Appendix 1. Voucher specimens. The numbers following the species names are DNA accession numbers from Lavin’s lab. Phaselus nomenclature follows Freytag and Debouck (2006). GenBank accession numbers are cited for the ITS and trnL sequences used in this study.

P. carterae
P. chiapasanus
Mercado 58
Delgado-Salinas 1574
(MEXU),
(MEXU),
(TEX),
Villarreal s. n.
P. pauciflorus
1849: Guatemala,
230: U.S.A., southwest, Arizona
Native
P. coccineus
(MEXU),
P. grayanus
SYSTEMATIC BOTANY
1883:
P. pauciflorus
Estrada 15042
P. grayanus
(MEXU),
Torres-
P. laxiflorus
L. 27: Peru
Delgado-Salinas 1708
P. falcatum
189: Morelos,
P. lunatus
Hook. f. 141: Ecuador,
(Texas),
Ramı
1 AF115225.
P. grayanus
Woot. & Standl.
130: México, Chihuahua,
INIFAP-URG
10720,
P. pluriflorus
1 AF115241.
P. grayanus
1813: México, Aguascalientes,
Mercado 145 (MEXU),
P. coccineus
1 AF45979.
P. grayanus
1811: México, U.S.A.,
Ciat 1-S107,
1 AF115207.
P. grayanus
231: México, Hidalgo,
Alcantara 2405 (MEXU),
P. coccineus
1 AF45974.
P. grayanus
1857: México, Hidalgo,
Puebla, Basurto 809 (MEXU),
P. grayanus
1 AF115245.
P. grayanus
1844: México, Puebla,
Ciat 45977, 45980.
P. grayanus
1 AF115229.
P. grayanus
1 AF45979.
P. grayanus
1 AF115220.
P. grayanus
1 AF45979.
P. grayanus
1 AF115229.
P. grayanus
1 AF45977.
P. grayanus
1 AF115210.
P. grayanus
1 AF45979.
P. grayanus
1 AF115212.
P. grayanus
1 AF45979.
P. grayanus
1 AF115215.
P. grayanus
1 AF45979.
P. grayanus
1 AF115219.
P. grayanus
1 AF45979.
P. grayanus
1 AF115218.
P. grayanus
1 AF45979.
P. grayanus
1 AF115217.
P. grayanus
1 AF45979.
P. grayanus
1 AF115216.
P. grayanus
1 AF45979.
P. grayanus
1 AF115216.
P. grayanus
1 AF45979.