1 Parsing polyphyletic *Pueraria*: delimiting distinct evolutionary lineages through phylogeny

2
3 Ashley N. Egan^{a*}, Mohammad Vatanparast^a, William Cagle^b

4
5 ^a US National Herbarium, Department of Botany, Smithsonian Institution-NMNH, MRC 166,
6 10th and Constitution Ave, Washington DC, 20560, USA

7 ^b East Carolina University, Department of Biology, Howell Science Complex, Greenville, NC
8 27858, USA (Current address: Campbell University, School of Osteopathic Medicine, Buies
9 Creek, NC, 27506 USA)

10
11 * Corresponding author: E-mail: egana@si.edu; ashegan2@gmail.com

Abstract

Several taxonomic and phylogenetic studies have hypothesized polyphyly within *Pueraria* DC., a genus comprising 19 species (24 with varieties) including the highly invasive *Pueraria montana* var. *lobata* (Kudzu) introduced to the U.S.A. about 150 years ago. Previous efforts to investigate monophyly of the genus have been hampered by limited taxon sampling or a lack of comprehensive evolutionary context that would enable definitive taxonomic associations. This work presents a comprehensive phylogenetic investigation of *Pueraria* within the context of tribe Phaseoleae (Leguminosae). Polyphyly was found to be more extensive than previously thought, with five distinct lineages spread across the tribe and spanning over 25 mya of divergence strongly supported by two chloroplast and one nuclear marker, *AS2*, presented here as a phylogenetic marker for the first time. Our phylogenies support taxonomic revisions to rectify polyphyly within *Pueraria*, including the resurrection of *Neustanthus*, moving one species to *Teyleria*, and the creation of two new genera, *Haymondia* and *Toxicopueraria* (taxonomic revisions published elsewhere).

Key Words: Character evolution, Fabaceae, Phylogeny, Polyphyly, *Pueraria*, Taxonomy, *Neustanthus*, *Haymondia*, *Toxicopueraria*, *Teyleria*

1. Introduction

Taxonomy or classification of organisms is meant to bring order out of chaos. However, taxonomic classifications do not always reflect natural relationships. Molecular systematists have long verified the utility of phylogenetic methods for solving various taxonomic or evolutionary problems. In this study, we show an example from flowering plants where historical taxonomy does not reflect evolution, but outlines a group of phylogenetically diverse species classified under a single genus for nearly two centuries.

De Candolle (1825) described the genus *Pueraria* DC. in response to his observation that *Hedysarum tuberosum* Willd. lacked the articulated pods characteristic of most Hedysareae, creating the new genus to include *Pueraria tuberosa* (Willd.) DC. and *Pueraria wallichii* DC., a species he described based on material provided him by Dr. Nathaniel Wallich from collections in Nepal. De Candolle united these two species based on broadly shared morphological characters, such as the climbing and twining habit, woody cylindrical stems, broad trifoliolate leaves, and elongated, axillary, “almost paniculate” inflorescences. Since then, *Pueraria* has expanded to include ~20 species.

Native to Southeast Asia (Fig. 1), *Pueraria* species have played important economic roles within human history since 500 B.C. in China (Keng, 1974) and 600 A.D. in Japan (Shurtleff and Aoyagi, 1977) where they have been used to treat medical ailments and as sources of textiles, food, paper, and building materials. Today, two species have been spread around the world for various economic uses: Kudzu, *P. montana* var. *lobata* (Willd.) Sanjappa and Pradeep, was introduced from Asia to the U.S.A. about 150 years ago for soil erosion control and cattle feed, but has since become a highly invasive species (Miller and Edwards, 1983); Tropical Kudzu, *P. phaseoloides* (Roxb.) Benth., has been widely spread across the pantropics where it is

used as a cover crop and fodder (Rueda et al., 2003; Valentim and Andrade, 2005) and has become naturalized in some areas.

Several former treatments have recognized *Pueraria* as an unnatural grouping (e.g. Lackey, 1977c; van der Maesen, 1985). Based on morphological characters including the number of flowers per node, stipule type, calyx type, vexillum callosities, pod type, and seed shape, Lackey (1977c) provisionally segregated *Pueraria* into four groups (A-D; Table 1), stating that groups A-C should be kept within the genus. Seemingly contradictorily, he also stated that species in group C could be removed from *Pueraria* and allied with *Neonotonia* Lackey (1977a), that group B species were distinct enough to form a new genus, and that group D should be excluded from *Pueraria* based on morphology and the presence of canavanine (Lackey, 1977b, c). Lackey did not revise the genus.

Building on Lackey's work, Van der Maesen completed the most current revision of *Pueraria* over the course of a series of botanical treatments wherein he recognized 15 species and several varieties (van der Maesen, 1985, 1994, 2002; van der Maesen and Almeida, 1988). Van der Maesen (1985) used the same morphological characters as Lackey to delineate three sections: section *Pueraria* was defined as those species with (2-)3 flowers per node, peltate stipules, and flattened-oblong seeds and included the bulk of his recognized species; section *Schizophyllon* Baker included only *P. phaseoloides* (Roxb.) Benth., separated by its having 4-6 flowers per node, basifixed stipules, cylindrical pods, and barrel-shaped seeds; section *Breviramulae* van der Maesen included those species with flowers 4 or more per node on brachyblasts, basifixed stipules, flat pods, and flattened-ovoid seeds. Van der Maesen (1985) also perceived that *Pueraria* did not represent a natural grouping, stating "...some species of *Pueraria* are difficult to place. Anomalous in the genus, they fit nowhere else satisfactorily...".

Van der Maesen also did not revise the genus, claiming “further biosystematic research” was needed. Since van der Maesen’s and Lackey’s works, new *Pueraria* species have been described, including *P. xyszhuui* (Ohashi and Iokawa, 2006), *P. bouffordii* (Ohashi, 2005), and *P. grandiflora* B.Pan bis & Bing Liu (Pan et al., 2015).

Earlier molecular phylogenetic perspectives portray pervasive polyphyly in *Pueraria* (Cagle, 2013; Doyle et al., 2003; Lee and Hymowitz, 2001; Stefanovic et al., 2009), confirming Lackey’s (1977c) and van der Maesen’s (1985) suspicions that *Pueraria* does not represent a natural grouping. Lee and Hymowitz (2001) were the first to sample multiple *Pueraria* species within a molecular phylogeny and found evidence for four distinct lineages (*P. wallichii*, *P. stricta*, *P. phaseoloides*, and a clade comprising *P. montana* + *P. pulcherrima*) based on the chloroplast *rps16* region, but conceded that the genus needed to be the focus of a “more rigorous molecular investigation.” Other molecular works prior to our study represent useful beginnings for investigating the evolutionary history of *Pueraria*, however, all included only four to five species, each with only a single accession, and used the same DNA material extracted by Lee & Hymowitz (2001). Furthermore, the phaseoloid generic level sampling of previous molecular works was scant compared to extant generic diversity, creating uncertainty as to how *Pueraria* lineages related to other legume genera. Tribe Phaseoleae (Brons) DC. is the largest tribe in subfamily Papilionoideae. As classically circumscribed, it comprises 89 genera and ~1,600 species (Lewis et al., 2005), including a number of vital economic species, such as soybean, chickpea, and common bean, and has been historically divided into a number of subtribes. To definitively place *Pueraria* lineages within the evolutionary context of phaseoloid legumes, a critical generic-level sampling is required.

To investigate the taxonomic hypotheses suggested by Lackey, van der Maesen, and others, and to gain a more comprehensive understanding of the evolutionary history of *Pueraria*, we undertook a phylogenetic study to investigate the extent of polyphyly within *Pueraria* in the context of other phaseoloid legumes. We sampled nearly 80% of the currently recognized *Pueraria* species, placed these within a robust generic-level sampling across the phaseoleae tribe, and discuss generic-level, subtribal, and tribal relationships within Phaseoleae. We also investigated the evolutionary history and delineating power of key morphological characters deemed diagnostic of *Pueraria* species by Lackey and van der Maesen. Our phylogenetic results based on one nuclear and two chloroplast markers uncovered strong signal for an even more polyphyletic *Pueraria* than discovered previously and shed light on the diagnostic power of morphology within *Pueraria*. Our results provide the phylogenetic evidence to support earlier taxonomic revisions by Egan & Pan (2015b) in the resurrection of *Neustanthus* Benth, the movement of *Pueraria stricta* Kurz to *Teyleria* Backer, and the creation of two genera, *Haymondia* A.N.Egan & B.Pan bis, and *Toxicopueraria* A.N.Egan & B.Pan bis.

2. Materials and Methods

2.1 Taxon Sampling

To understand the evolutionary history of *Pueraria*, we sampled 20 of the 24 (79%) currently recognized taxa (Table 1). We follow the recommendations of van der Maesen (2002) in his recognition of species and their varieties and exclusion of several species due to synonymy or misidentification by excluding *P. garhwalensis* L.R.Dangwal & D.S.Rawat, *P. maclurei* (F.P.

Metcalf) F.J. Herm., and *P. stracheyi* Baker, recently determined to be a synonym of *Apios*
carnea (Wall.) Benth. ex Baker (Egan and Pan, 2015a). Van der Maesen and Lackey recognized
P. yunnanensis Franchet. as a synonym of *P. peduncularis* (Benth.) Benth., however, we are in
agreement with others (Le and Zhu, 2009) and treat it at the species level here. Because of rarity,
we were unable to collect *P. bella* Prain, known from only four collections in Myanmar and
India; *P. xyzhunii* H.Ohashi & Iokawa and *P. bouffordii* H.Ohashi were each described from a
single collection, and thus not included in our phylogeny; and *P. grandiflora* B. Pan bis & Bing
Liu was described after inception of this work (Pan et al., 2015).

Plant material was obtained from field collections performed by A.N. Egan (China,
Thailand, South Africa, Australia, and the U.S.) and from voucher specimens from a number of
herbaria around the world, including BKF, BRI, CAL, E, K, MO, NY, P, TEX and US
(herbarium codes from Index Herbariorum; (Thiers, continuously updated)). Additional DNA
samples were generously donated by Jeff J. Doyle (Cornell University). To place *Pueraria*
lineages within the evolutionary context of related taxa, we sampled widely across phaseoloid
and millettoid legumes, including 103 species across 67 genera, seven genera of which have
never before been sampled to a legume molecular phylogeny: *Arthroclianthus* Baill.,
Dysolobium (Benth.) Prain, *Eminia* Taub., *Melliniella* Harms, *Nephrodesmus* Schindl.,
Paracalyx Ali and *Pycnospora* R.Br. ex Wight & Arn. (LPWG, 2013a, b). By comparison,
Lewis *et al.* (2005) treated 89 genera in the traditionally held view of tribe Phaseoleae. Voucher
specimens, source, and DNA accession numbers are in supplementary table 1.

2.2 Molecular Markers

To estimate phylogenetic relationships, we used two chloroplast and one nuclear marker. The chloroplast *matK* gene has long been used as a molecular marker in plants, with wide application across angiosperms (e.g. as summarized in Hilu et al., 2003) and was chosen for its ongoing use in legume systematics to enable integration of these data into global efforts to build the legume tree of life (de Queiroz et al., 2015; LPWG, 2013a, b; Wojciechowski et al., 2004; Yahara et al., 2013). Short regions have been shown to amplify better when working with degraded DNA (Särkinen et al., 2012). The *matK* region amplified in this study is thus a partial version of 722 base pairs (roughly from position 1210-1932 in the full *matK* gene) chosen to enable proper amplification across samples derived from herbarium material where DNA is often degraded. The chloroplast *trnD-trnT* intergenic spacer is characterized as one of the fastest evolving chloroplast markers in angiosperms (Shaw et al., 2005) and has been shown to be useful in legumes previously (Egan and Crandall, 2008a, b; Vatanparast et al., 2011).

ASYMMETRIC LEAVES 2 (*AS2*), also known as *LBD6*, is a nuclear gene whose primary function is in the establishment of symmetry and leaf polarity where it regulates many functions including the adaxial-abaxial axis, resulting in planar leaves (Iwakawa et al., 2002). *AS2* encodes a plant-specific protein comprising an AS2/LATERAL ORGAN BOUNDARIES (AS2/LOB) domain with a cysteine repeat, a conserved single glycine residue, and a leucine-zipper-like sequence in its amino-terminal half, features that place it as a member of the *AS2/LOB* gene family (Matsumura et al., 2009). *AS2* is single copy in *Arabidopsis* (At1g65620), the genome of which contains 42 *AS2/LOB* gene family members (Matsumura et al., 2009). This project is the first to use *AS2* as a phylogenetic marker, the primers for which were designed by the first author based on a comparison of multiple legume genomes and homoeologues (A.N. Egan, unpublished data). The region of the *AS2* gene that we studied comprises the *AS2* domain

(1-293 bp) interrupted by a single intron (position 305-534) and ending with a 24 bp exon comprising the remainder of the *AS2* domain, features of which place it in subclass 1a of the *AS2/LOB* gene family (Matsumura et al., 2009). To examine whether paralogy/orthology conflation exists across our *AS2* dataset, we gathered 157 sequences of multiple *AS2/LOB* gene family members from across the *Arabidopsis*, *Populus*, *Glycine*, and *Medicago* genomes and aligned them, along with all our translated *AS2* amplicons, using MAFFT (Katoh and Standley, 2013). This alignment is presented in supplementary file 1. A majority rule consensus tree was created from 500 maximum parsimony, fast-stepwise bootstrap replicates computed in PAUP* v4.0a149 (courtesy David L. Swofford).

2.3 DNA Extraction, Amplification, and Sequencing

DNA extraction was carried out using two methods: the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA) following manufacturer's instructions or a modified CTAB method (Doyle and Doyle, 1987) using an automated DNA extraction machine (Autogen, Inc., U.S.A). The *matK* PCR protocol used the conditions of Hu *et al.* (2000) using the primers 1210F and 1932R described in Egan & Crandall (2008b). The *trnD-trnT* primers are as previously described (Shaw et al., 2005) using the same PCR conditions as *matK*. *AS2* was amplified using primers *AS2F* (5'-CAC CAT GTG CAG CAT GCA AGT TCT-3') and *AS2R* (5'-AGT TGC CCT AAG CTG GCG GAT ATG-3') under PCR conditions of 5 min at 94° followed by 35 cycles of 40 s at 94°, 1 min at 57°, 2 min at 72° and then ending with a final extension of 7 min at 72°. PCR products were purified using the QIAquick PCR Purification Kit (Qiagen, Valencia, CA) or ExoSAP-IT (Affymetrix, Santa Clara, CA). Sequencing was done on a 3130 Genetic Analyzer from Applied Biosystems using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher

Scientific, Waltham, MA). Forward and reverse sequences were edited and aligned into contigs with Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI). Sequences are deposited in Genbank under the numbers listed in supplementary table 1.

2.4 Alignments

DNA regions were aligned using MUSCLE under default parameters (Edgar, 2004). Alignments of the two coding regions, *matK* and *AS2*, were corrected to reading frame and edited by eye in SE-AL 2.0 (Rambaut, 2002). Interestingly, the intron portion of *AS2* was able to be aligned in frame at all times and was adjusted by eye. DNA marker amplification was variably successful across samples. Thus, marker datasets differ in the number of genera, species, and *Pueraria* accessions (Table 2). Of the 149 *matK* sequences included in this study, 104 are newly published, as are all *trnD-trnT* and *AS2* sequences (Supplementary Table 1).

2.5 Phylogenetic analyses

Dioclea reflexa Hook.f., *Clitoria ternatea* L. and a *Galactia* P. Browne were used as outgroups for *matK*, *trnD-trnT* and *AS2* datasets, respectively. Preliminary analyses showed strongly supported (BS > 70) incongruence between chloroplast *matK* and nuclear *AS2* in the placement of the tribe Psoraleae. This was confirmed via the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa, 1999) employed in PAUP* (Swofford, 2002) which found the *AS2* and *matK* trees to be statistically different. Because of this, each region was analyzed separately. The partition homogeneity test (Farris et al., 1994) was conducted in PAUP* v4 (Swofford, 2002) with 1000 replicates and using a heuristic search option with simple addition and no

branch swapping algorithm to test for congruence between the *matK* and *trnD-trnT* regions. The results of the partition homogeneity test indicated no incongruence between *matK* and *trnD-trnT*, and so these were combined to create a cpDNA dataset partitioned by gene.

Gaps in alignments have been shown to carry phylogenetic signal in legumes previously (Egan and Crandall, 2008b). Because the *AS2* region has a gapped, in-frame region, we ran analyses with gaps excluded and gaps coded as binary characters to test for phylogenetic signal in gaps. The binary matrix was created following simple coding method of Simmons and Ochoterena (2000) using the SeqState program (Muller, 2005). For *matK*, *trnD-trnT* and the combined dataset of these two regions (cpDNA), gaps were treated as missing in BI and ML analyses.

For phylogenetic analyses, Bayesian inference (BI) and Maximum likelihood (ML) methods were employed. Models of evolution were determined for each dataset using jModeltest (Darriba et al., 2012; Guindon and Gascuel, 2003) via the Akaike information criterion (Akaike, 1973). Bayesian inference was done using MrBayes v. 3.2.2 (Ronquist et al., 2012). For BI, the TVM+G model of evolution was used as the best-fit model for the *matK* and *AS2* data partitions whereas the TVM+I+G model was chosen for the *trnD-trnT* data partition. For the *AS2* dataset partitioned model blocks were setup accordingly for DNA sequence and binary characters. Two independent Markov Chain Monte Carlo (MCMC) analyses with four simultaneous chains and 20 million generations were run for each analysis. Trees were sampled every 1000 generations and the first 5000 trees were discarded as burn-in. The convergence of MCMC chains was visualized and confirmed with Tracer version 1.6 (Rambaut et al., 2014). All runs and parameters were checked to ensure proper mixing as evidenced by effective sample size (ESS) scores being above 100 and the standard deviation of the split frequencies having dropped below 0.01

(Ronquist et al., 2012). Bayesian analyses were performed via the CIPRES Science Gateway (Miller et al., 2010).

For *matK*, *trnD-trnT* and combined cpDNA datasets, ML analyses were performed using RAxML v. 8.0.14 (Stamatakis, 2014). Analyses were run for the best-scoring ML tree inferences under the GTR-GAMMA model for all molecular partitions. Rapid bootstrapping was performed with 1000 replications using the GTRCAT estimation to assess branch support (Stamatakis, 2006). The AS2 gap coded dataset was conducted using Garli v. 2.0.1 (Zwickl, 2006) with partitioned models as unlinked. An initial run was executed to find the best tree with 100 attachments per taxon and five replicate searches with five million generations (stopgen = 5,000,000). A subsequent bootstrap analysis was carried out in Garli with 100 replicates and a strict consensus tree was retrieved from resultant bootstrap trees using the Python package SumTrees (Sukumaran and Holder, 2010).

2.6 Divergence time analysis

Divergence time estimates were obtained using the Bayesian inference approach implemented in the package BEAST v.1.8.1 (Drummond et al., 2012) using *matK*, *trnD-trnT*, the combined cpDNA dataset, and AS2 matrices, applying the same partition delimitation and evolutionary models as those used for the MrBayes analyses. We used an uncorrelated relaxed molecular clock with a lognormal distribution of rates and a Yule speciation model (Gernhard et al., 2008; Yule, 1924). The analysis was run for 30 million generations, sampling one tree every 1,000th generation. For calibration, we applied a normal prior distribution (mean=39.7 Mya, sd = 2.0) to the root of the tree, based on the age estimate of the MRCA of *Platycamus regnellii*

Benth. and *Phaseolus vulgaris* L. (Lavin et al., 2005), a divergence date that represents the lower bound for the age of our ingroup. To summarize plausible trees and to obtain a maximum clade credibility tree, the TreeAnnotator program implemented in the BEAST package was used. Twenty-five percent (7500) of trees were excluded as burn-in from the subsequent calculations. Tracer v. 1.6 (Rambaut et al., 2014) was used to check the effective sample sizes (ESSs) of each parameter. The trees were visualized and edited using FigTree v. 1.4.2 (Rambaut, 2014).

2.7 Reconstruction of ancestral character states

Several morphological characters have been used to justify differing hypotheses concerning potential groupings within *Pueraria*. Here, we use ancestral state reconstruction to explore the evolutionary history and taxonomic value of four morphological traits: **habit** (0) erect herb (1) climbing or twining vine (2) scandent shrub (3) shrub and (4) tree; **stipule** (0) not produced below the point of insertion (basifixed) (1) produced below the point of insertion (i.e. medifixed or peltate; **number of flowers per node** (0) flowers 3 or two (1) flowers 4 or more (2) variable, but flowers 1 to 7 per node; **canavanine** (0) absent (1) present. See supplementary table 2 for complete table of coded characters. All morphological characters were coded from current descriptions or by referencing specimens in the U.S. National Herbarium. Canavanine content was based on a compilation of previous works (Polhill et al., 1994).

To reconstruct ancestral character evolution among morphological characters, MCMC reconstructions implemented in BayesTraits v. 2.0 (Pagel and Meade, 2013) were applied using ‘MultiState’ option and a random set of 145 Bayesian prior probability trees from the *matK* dataset, the dataset with the greatest taxonomic sampling. A reverse jump MCMC method was

used to integrate results over model parameter and model restrictions as recommended (Pagel and Meade, 2006). We used a hyper-exponential prior where mean of the exponential is from a uniform value of 0 - 100. All Bayesian analyses were run for character states independently with 20 million generations, sampling every 1000 generations and using a burn-in period of 100,000 generations. Based on posterior probabilities (PP) from the MrBayes analyses, the ‘AddNode’ command was applied for selected clades except for uncertain clades for which the ‘AddMRCA’ (most recent common ancestor) command was used. Convergence of the MCMC values and ESS of each parameter were inspected using Tracer v1.6. To illustrate final values of posterior probabilities for ancestral reconstruction of character evolution under MCMC for the 145 randomly chosen trees, mean values were represented on a 50% majority rule consensus Bayesian tree using pie charts at each node.

Results and Discussion

3.1 Dataset statistics

Of all four datasets, the *matK* dataset included the most genera, the highest number of *Pueraria* accessions, the most accessions overall, and the fewest gaps, and also had the highest amount of missing data, largely due to the inclusion of some full-length sequences from GenBank relative to the partial gene regions amplified in this study (Table 2). Unaligned *matK* sequences ranged from 441 bp (*Pueraria alopecuroides* E1481) to a number of sequences with 1,095 bp, with the aligned dataset of length 1128 bp. The *trnD-trnT* dataset had the greatest percentage of gapped sites, with unaligned sequences ranging from 374 bp (*Otholobium*

pubescens E1564) to 1,521 bp (*Calopogonium caeruleum* E27) and an aligned length of 2,283 bp. The combined *matK/trnD-trnT* included those accessions represented in both chloroplast marker datasets, and thus had the fewest genera, species, number of *Pueraria* accessions, and number of total accessions, had the longest aligned length but with the highest proportion of constant sites and the lowest proportion of phylogenetically informative sites (Table 2), but also presented the greatest resolution along the backbone (Fig. 2). The unaligned concatenated sequences ranged from 956 bp (*Pycnospora lutescens* E678) to 2,302 bp (*Pueraria imbricata* E1476), with an aligned length of 2,873 bp. While AS2 has the shortest aligned length of 669 bp, it had the lowest proportion of constant sites and the highest proportion of phylogenetically informative sites, with unaligned sequences ranging from 325 bp (*Pueraria lacei* E807) to several sequences of length 555 bp. Binary coding of the AS2 marker produced 81 coded gaps. AS2 with gaps coded provided better resolution than with gaps considered as missing, and for the Bayesian analyses, resolved subtribe Glycininae as a strongly supported clade (PP=1.0; Fig. 3) whereas it was broken apart into the polytomous backbone with gaps treated as missing data (data not shown). The tree produced using the nuclear AS2 marker was less resolved along the backbone of the tree but still provided insights into generic and subtribal relationships. For all markers, the results of the ML analyses were largely congruent with our results from Bayesian inference analyses. Therefore, only resultant trees from BI for the combined cpDNA, AS2 (gaps coded), *matK*, and *trnD-trnT* are illustrated here (Figs. 2-5, respectively), with bootstrap support values from ML analyses included thereon.

3.2 Polyphyly of *Pueraria*

Both morphological (Lackey, 1977c; van der Maesen, 1985) and phylogenetic studies (Cagle, 2013; Lee and Hymowitz, 2001; Stefanovic et al., 2009) have detailed the problem of polyphyly in *Pueraria*. Based on morphology, both Lackey (1977c) and van der Maesen (1985) suggested that *Pueraria* likely included unrelated lineages that could represent distinct genera, but neither took steps to revise *Pueraria* in part due to a lack of understanding of how *Pueraria* species related to other legume genera. Lee and Hymowitz (2001) were the first to confirm polyphyly within *Pueraria* via molecular means but stated that a “more rigorous molecular investigation” was needed to definitively understand its evolutionary history. Our study provides considerably increased and broader sampling both within *Pueraria* by sampling ~80% of species and within Phaseoloid legumes by sampling 58 genera for evolutionary context. Our phylogenetic analyses support previous hypotheses and findings of polyphyly in *Pueraria* across all data sets, discovering more distinct, polyphyletic lineages than previously thought. All analyses reveal five distinct clades, *P. wallichii*, *P. stricta*, *P. phaseoloides*, *P. peduncularis* + *P. yunnanensis*, and the remaining species grouped into the main clade of *Pueraria*, hereinafter called *Pueraria sensu stricto*, all once attributed to *Pueraria sensu lato*, spread across the phaseoloid legumes (Figs. 2-5) and spanning more than 25 million years of divergence (Table 3; Fig. 6).

3.2.1. Polyphyletic lineages and relation to previous hypotheses

Pueraria s.s. includes *P. tuberosa*, the type of the genus, as well as all species except *P. phaseoloides* (and its varieties), *P. stricta*, *P. peduncularis*, *P. yunnanensis*, and *P. wallichii*. Species within *Pueraria s.s.* began diverging in the late Miocene ~6.2 mya (Fig. 6). *Pueraria s.s.*

is resolved with moderate to strong support but with variable association with *Nogra* Merr. The *matK* phylogeny resolves *Pueraria s.s.* as monophyletic (PP=0.87) and sister to *Nogra* (PP=1.0; Fig. 4) from which it diverged ~7.85 mya (Fig. 6), whereas *AS2* and *trnD-trnT* support *Nogra* as nested within a strongly supported *Pueraria s.s.* clade (PP=0.99 or 1.0; Fig. 3 and Fig. 5, respectively). The uncertain association between *Nogra* and *Pueraria s.s.* was also illustrated in previous analyses based on the *rps16* region (Lee and Hymowitz, 2001). For the most part, species form cohesive groups within *Pueraria s.s.* but relationships among species are less resolved, especially for *AS2*. One group that falls out consistently among chloroplast DNA analyses is a clade comprising *P. imbricata*, *P. alopecuroides*, *P. candollei*, and *P. sikkimensis* (Figs. 2, 4, 5), species that, according to van der Maesen (1985), are united by having branching, pubescent inflorescences with stipule-like, peltate bracts on the lower portion of the inflorescence rachis.

In all analyses, the *Pueraria phaseoloides* clade is resolved as a monophyletic group with absolute support (PP=1.0; Figs. 2-5). Based on an average across divergence times estimated for each dataset, *P. phaseoloides* is estimated to have separated from *Pueraria s.s.* around 13.4 my ago (Table 3, Fig. 6), an estimate on par with that determined by Li *et al.* (2013) of 13.4 mya based only on the *matK* gene. This is in accordance with van der Maesen (1985) who separated it into its own section *Schizophyllon* based on, in part, the papery partitions separating the seeds and in line with Lackey (1977c) who also suggested that *P. phaseoloides* could comprise its own genus due to its barrel-shaped seeds and rounded pods. Most analyses resolve the *P. phaseoloides* clade as sister to *Sinodolichos* Verdc. with moderate (*matK*, PP=0.88, Fig. 4) to strong support (cpDNA, PP=0.96, Fig. 2; *AS2*, PP=0.93, Fig. 3). The *trnD-trnT* phylogeny places the *P. phaseoloides* clade in a tritomy with *Sinodolichos* and a clade comprising *Eminia*,

Pseudeminia Verdc., and *Pseudovigna* Verdc. (PP=1.0; Fig. 5), findings in support of Verdcourt's (1970) hypotheses, which suggested affinity of *P. phaseoloides* (naturalized in Africa) with *Sinodolichos* and *Pseudovigna* (Harms) Verdc. based on pollen structure, and on general appearance with a group comprising *Pseudeminia* Verdc., *Eminia* Taub. and *Pseudovigna*.

In all cases, multiple *Pueraria stricta* accessions form a monophyletic species clade that is sister to or nested within *Teyleria* Backer with strong (PP=1.0 for cpDNA, *matK*, and *trnD-trnT*; Figs. 2, 4, 5, respectively) to moderate (PP=0.88 for AS2; Fig. 3) support and set apart from *Pueraria s.s.* by ~13 my in divergence (Table 3, Fig. 6). *Pueraria stricta* is distinguished from *Pueraria s.s.* and united with *Teyleria* by presence of angular petioles and young stems, four or more flowers per node, and septate fruits (Egan and Pan, 2015b) as well as the presence of canavanine (Lackey, 1977b). Lackey (1977c) suggested that *P. stricta* be allied with his *Neonotonia* J.A. Lackey based on floral and inflorescence characters, but van der Maesen (1985) disagreed. Our phylogenetic results are more in line with Lackey's suggestions and supported by other gene phylogenies (Lee and Hymowitz, 2001).

Pueraria peduncularis and *P. yunnanensis* are sister to each other with absolute support in all analyses, are separated from *Pueraria s.s.* by ~16 my divergence (Table 3, Fig. 6), form a clade set apart from other Phaseoloid genera, and are resolved as the earliest branching lineage in subtribe Glycininae by chloroplast datasets (Figs. 2, 4, 5). AS2 places this clade in a basal polytomy within subtribe Glycininae (Fig. 3). Both Lackey (1977c) and van der Maesen (1985) separated *P. peduncularis* from the main *Pueraria* group, Lackey based on the absence of paraveinal mesophyll and van der Maesen based on morphological differences in pod and seed shape.

Pueraria wallichii accessions form a monophyletic clade with absolute support in all analyses and is the only *Pueraria* lineage placed outside the core Phaseoleae with strong support (Figs. 2-5). Given that *P. wallichii* was one of the two original species for which *Pueraria* was created, it is amazing to find that this species is a lineage that diverged from *Pueraria s.s.* over 26 mya (Table 3, Fig. 6). *Pueraria wallichii* superficially resembles *Pueraria s.s.* in its large, trifoliate leaves and potential to twine at the tips of its branches. However, this species is a shrub, sometimes straggling, has basifixed stipules, and is variable in its inflorescence type which can be paniculate in terminal racemes, or with lateral branches shortened to brachyblasts (Egan and Puttock, 2016). Lackey (1977c) suggested that *P. wallichii* was generically, if not subtribally or tribally, misplaced due to its lack of paraveinal mesophyll and presence of canavanine, a hypothesis our results support. Van der Maesen (1985) placed *P. wallichii* with *P. peduncularis* in section *Breviramulae*. That said, the affinity of *P. wallichii* is uncertain, with several associations suggested by different datasets. *Pueraria wallichii* is placed in a basal polytomy within the *matK* tree (Fig. 4) whereas *trnD-trnT* resolves *P. wallichii* as sister to *Mucuna gigantea* (Willd.) DC. with moderate support (PP=0.91; Fig. 5). The cpDNA tree resolves *P. wallichii* as sister to tribe Desmodieae with strong support (PP=0.98; Fig. 2), a clade that is then sister to Phaseoleae subtribe Kennediinae (PP=0.99). The nuclear *AS2* marker places *P. wallichii* as sister to a loosely associated clade comprising *Butea* Roxb. ex Willd. and Kennediinae with poor support (PP=0.71; Fig. 2). *Pueraria wallichii* represents an early-branching lineage of the classically circumscribed Phaseoleae tribe and is in an area of the phylogeny that has eluded resolution in a number of studies (e.g. Doyle and Doyle, 1993; Kajita et al., 2001) where phaseoleae has been found to be polyphyletic and paraphyletic with respect to tribe Millettieae

(Lewis et al., 2005). Further taxonomic, generic, or molecular sampling may help to definitively determine association of *P. wallichii*.

3.2.2. Taxonomic implications

Our molecular and morphological results illustrate pervasive polyphyly within *Pueraria* and necessitate taxonomic revisions to the genus. As illustrated across three DNA regions and corroborated by key diagnostic chemical and morphological characters, five separate lineages exist within *Pueraria* that are scattered across the phaseoloid legumes (Figs. 2-5). This work thus justifies taxonomic revisions published previously in which we described two new genera, reinstated a genus, and made one new combination in order to resolve polyphyly within *Pueraria* (Egan and Pan, 2015b). The four new *Pueraria* segregates are introduced below and henceforth referred to by their new scientific names. A full description of the taxonomy and nomenclature for each of the four segregate lineages are provided elsewhere (Egan and Pan, 2015b).

***Neustanthus* Benth.** Bentham (1852) noted distinct differences in pod shape between the then recognized species of *Pueraria*, and based on erroneous information that the pod of *P. tuberosa*, the type of *Pueraria*, was articulate, Bentham created the genus *Neustanthus* Bentham to accommodate *P. phaseoloides* and its varieties as well as *P. peduncularis*. Later, Bentham (1867) recognized that *P. tuberosa* did not have an articulated pod and merged his *Neustanthus* with *Pueraria*, stating “The genus thus established has the habit and pod of *Phaseolus* or of some allied genera, with the flowers more like those of *Dioclea*. It comprises nine species, differing often considerably from each other in the calyx, the form of the petals, or in the degree of adherence of the upper stamen, as well as in the pod, but all, I believe, most conveniently

considered as congeners.” The ‘considerable differences’ observed by Bentham were no doubt due to the fact that he included *P. peduncularis*, *P. wallichii*, and *P. phaseoloides* with other species herein referred to *Pueraria s.s.*, inadvertently creating an unnatural grouping. Our results support the recognition of *P. phaseoloides* and its varieties as a distinct evolutionary lineage and support the reinstatement of *Neustanthus* Bentham, a lineage distinguished from *Pueraria s.s.* by its basifixed stipules (Fig. 4), 4-6 flowers per node (Supplementary Fig. 1), distinct upper calyx teeth, vexillum without callosities, and more or less terete pods with 15-20 barrel-shaped seeds separated by papery partitions. This genus includes *Neustanthus phaseoloides* var. *phaseoloides* (Roxburgh) Bentham, *Neustanthus phaseoloides* var. *javanica* (Bentham) A.N.Egan & B.Pan, and *Neustanthus phaseoloides* var. *subspicata* (Bentham) A.N.Egan & B.Pan (Egan and Pan, 2015b).

***Teyleria stricta* (Kurz) A.N.Egan & B.Pan.** *Pueraria stricta* is anomalous within *Pueraria s.s.* by virtue of its basifixed stipules and presence of canavanine (Fig. 4). Our results support its separation from *Pueraria s.s.* and determine its placement in a clade with all other Glycininae species that contain canavanine. *Pueraria stricta* is nested within or sister to *Teyleria* and shares its angular petioles, quadrangular stems, small flowers along an erect rachis, monadelphous stamens, septate fruit, and sculptured seed coat. Thus, our results support the movement of *Pueraria stricta* to *Teyleria*, being henceforth recognized as *Teyleria stricta* (Kurz) A.N.Egan & B.Pan (Egan and Pan, 2015b).

***Haymondia* A.N.Egan & B.Pan.** *Pueraria wallichii*, one of the two original species to the genus, is here determined as the most evolutionarily distant segregate from *Pueraria s.s.* (Figs. 2-5). Our results corroborate Lackey’s (1977c) assertions that this species is subtribally and perhaps tribally misplaced due to the lack of paraveinal mesophyll, presence of canavanine

(Fig. 4), basifixed stipules (Fig. 4), and 4 or more flowers per node (Supplementary Fig. 1). *Pueraria wallichii* also diverges from *Pueraria s.s.* in the broad nature of the funicle and movement of the staminal column upwards upon flower maturation (Egan and Pan, 2015b). This latter feature is shared with several other genera to which *Haymondia* is loosely affiliated in our phylogenies, including *Apios*, *Mucuna*, *Cochlianthus*, and several members of tribe Desmodieae. However, given the lack of definitive association with other phaseoloid genera and the separate placement within our phylogenies, we strongly support the recognition of *Pueraria wallichii* as its own genus, hereinafter known as *Haymondia wallichii* (DC.) A.N.Egan & B.Pan (Egan and Pan, 2015b). The genus is named for the maternal grandparents of author A.N. Egan who supported her as a developing scientist.

***Toxicopueraria* A.N.Egan & B.Pan.** Like *H. wallichii*, *P. peduncularis* was thought to be misplaced within *Pueraria* by previous workers due to a lack of paraveinal mesophyll (Lackey, 1977c) and differences in pod and seed structure (van der Maesen, 1985). Our phylogenies place *P. peduncularis* and its congener, *P. yunnanensis*, at the base of subtribe Glycininae (Figs. 2-5) supported by differences in stipule attachment (Fig. 4) and an increased number of flowers per node relative to *Pueraria s.s.* (Supplementary Fig. 1). They also differ from *Pueraria s.s.* in their roots and lack of callosities on the vexillum (Egan and Pan, 2015b). As with *Haymondia*, *P. peduncularis* and *P. yunnanensis* are placed apart from other phaseoloid genera and we thus created a new genus for this sister group, hereinafter known as *Toxicopueraria* A.N.Egan & B.Pan, inclusive of *Toxicopueraria peduncularis* (Graham ex Benth) A.N.Egan & B.Pan and *Toxicopueraria yunnanensis* (Franch.) A.N.Egan & B.Pan (Egan and Pan, 2015b). The genus was named for the latin *toxicus* meaning ‘poisoned’ in

reference to its occasional use of crushed roots as a fish poison in Yunnan, China (Perry and Metzger, 1980).

3.2.3. Character evolution in support of critical delineating morphological traits

Previous taxonomic treatments focused on key morphological features, particularly the number of flowers per node, the stipule type, habit, seed and pod shape, presence of vexillum callosities, and presence or absence of canavanine, as evidence of the cohesiveness or disparity of *Pueraria* species (Lackey, 1977b; van der Maesen, 1985). Here, we focused on four morphological characters: habit, number of flowers per node, manner of stipule attachment, and the presence or absence of canavanine. These features were coded across *Pueraria* and other phaseoloid accessions (Supplementary Table 2) and used for character state reconstruction employing the *matK* tree to enable an objective viewing of the evolutionary transition and delineating power of these characters across *Pueraria* and phaseoloid legumes.

Stipule attachment. Lackey (1977c) used “stipules produced below the point of insertion” (here denoted as peltate or dorsifixed; Fig. 4) in his key to differentiate his ‘group A’ from other *Pueraria* species. This character trait strongly delineates *Pueraria s.s.* from segregate *Pueraria* lineages in our analyses (Fig. 4) and further supports the relationship of *Pueraria s.s.* with *Nogra grahamii*, a taxon that also has peltate stipules. The presence of peltate stipules suggests that three of the unsampled *Pueraria* species, *P. xyzhunii*, *P. bella*, and *P. grandiflora*, may be placed within *Pueraria s.s.* The ancestral and predominant form of stipule attachment is basifixed across the phylogeny, but with the derived state of peltate or dorsifixed stipules having arisen twice: in the ancestor of *Pueraria s.s.* and *Nogra*, and at the base of a clade comprising

Psophocarpus, *Otoptera*, and *Erythrina* wherein it was subsequently lost in *Erythrina*. However, the number of times peltate stipules have arisen across phaseoloid legumes is likely underestimated here because of incomplete sampling. For example, several, but not all, *Vigna* species are known to have peltate stipules but are not sampled here. In addition, this trait is known to have arisen, albeit rarely, in other legume clades or genera, including within *Zornia* J.F.Gmel., in which it also acts as a defining character for the genus, at the base of Aeschynomeneae subtribe Aeschynomeninae (Lavin et al., 2001), and within various singular taxa, such as *Dioclea relfexa* Hook. f. (Hooker, 1849), in which it is the only species of the genus with peltate stipules.

Canavanine. L-canavanine is a non-proteinogenic amino acid first discovered in Jack bean seeds, *Canavalia ensiformis* (L.) DC. (Kitagawa and Tomiyama, 1929). Canavanine has been widely assayed across legumes and promoted as a systematic character (Bell, 1981; Bell et al., 1978; Birdsong et al., 1960; Evans et al., 1985; Lackey, 1977b, c). Molecular phylogenies have supported this claim, resulting in the designation of the non-protein amino acid accumulating (NPAAA) clade inclusive of the inverted-repeat-lacking clade (IRLC) and the indigoferoid, millettoid, phaseoloid, and robinoid clades (Cardoso et al., 2012; de Queiroz et al., 2015; Hu et al., 2000; Hu et al., 2002; Wink and Mohamed, 2003; Wojciechowski, 2003; Wojciechowski et al., 2004). However, the presence of canavanine across the NPAAA clade is not absolute, with several losses evident, particularly within *Aeschynomeneae*, *Psoraleae*, *Amorpheae*, and *Phaseoleae* (Bell et al., 1978; Wink and Mohamed, 2003). Lackey noted the variable presence or absence of canavanine within *Pueraria s.l.* (Lackey, 1977b), suggesting that those species having canavanine (*T. stricta* and *H. wallichii*) should be excluded from the genus (Lackey, 1977c). Ancestral reconstruction of the presence of canavanine across our *matK* gene

tree (Fig. 4) confirms these findings. Within the scope of our sampling, presence of canavanine is found to be the ancestral condition with canavanine production suggested to have been lost at the base of the clade denoted here as the core Phaseoleae (after Lewis et al., 2005), subsequently regained in the ancestor of a clade comprising *Teyleria*, *Neonotonia*, *Calopogonium*, *Pachyrhizus*, and *Neorautanenia* (denoted the TNP clade here for short), and then lost again in the lineage leading to *Neorautanenia*. *Teyleria stricta* is nested along with all other canavanine-containing core-phaseoloid genera within the TNP clade. *Haymondia wallichii* is placed outside the core Phaseoleae within a basal polytomy with other genera that produce canavanine (Fig. 4). Canavanine proves to be a strong systematic character and good at delineating groups within *Pueraria s.l.* and elsewhere.

Number of flowers per node. Floral development in the context of inflorescence architecture is controlled by a number of genetic pathways, involving such genes as VEGETATIVE1 (Berbel et al., 2012), VEGETATIVE2 (Sussmilch et al., 2015), and NEPTUNE (Singer et al., 1999), among others. Genetic control of the number of flowers produced per node has been widely studied in legumes, particularly in pea, *Pisum sativum* L. (Benlloch et al., 2015; Smýkal et al., 2015) and has been touted as a heritable, phylogenetically correlated character (Lamprecht, 1947; Prenner, 2013). Both Lackey (1977c) and van der Maesen (1985) used the number of flowers per node to delineate groups with *Pueraria s.l.*, with Lackey's 'group A' and van der Maesen's section *Pueraria* (both equate to our *Pueraria s.s.* clade) distinguished from the other *Pueraria* groups as those having (2-)3 flowers per node while other groups and species have more. The ancestral condition of *Pueraria s.s.* is 3(2) flowers per node (Supplementary Fig. 1). Based on this character, all four unsampled *Pueraria* species, *P. xizhunii*, *P. bouffordii*, *P. grandiflora*, and *P. bella*, would be placed within *Pueraria s.s.* due to their having 2-3 flowers

per node. All *Pueraria s.l.* species with more than 3 flowers per node are placed apart from *Pueraria s.s.* Thus, the number of flowers per node is discerning among disparate *Pueraria s.l.* lineages. However, this character is not extremely useful as a taxonomic character outside *Pueraria*. The predominant condition across the phylogeny is 3 or fewer flowers per node, with this estimated as the ancestral condition across all nodes except at the base of *Strongylodon*, *Butea*, within Desmodieae, and three segregate *Pueraria* lineages: *Neustanthus*, *H. wallichii*, and the ancestor of *Toxicopueraria peduncularis* and *Toxicopueraria yunnanensis* (Supplementary Fig. 1).

Habit. The habit of a plant greatly impacts its evolutionary trajectory. Trees and shrubs have been shown to have a lower rate of molecular evolution as compared to herbs (Smith and Donoghue, 2008) and evolution of a climbing habit has been linked to increased diversification rates in plants (Gianoli, 2004). Van der Maesen (1985) described *T. stricta* and *H. wallichii* as shrubs, sometimes straggling, noting that this characteristic sets these species apart from other *Pueraria* lineages which are climbing vines. Tribe Phaseoleae, to which *Pueraria s.l.* belongs, was classically described as climbing herbs, rarely erect or shrubby, and very rarely trees (Bentham, 1865). Ancestral state reconstruction showed habit to be quite variable across the phylogeny (Supplementary Fig. 2), likely due in part to the nesting of tribes Psoraleae and Desmodieae within Phaseoleae, both of which do not include climbing habit. Nevertheless, a climbing habit was resolved as the ancestral condition for the clade inclusive of subtribes Phaseolinae, Glycininae, and the *Erythrina* clade. *Pueraria s.s.* lies within subtribe Glycininae and delineated as only climbing vines. *Teyleria stricta* is also within this clade, and is known to be a shrub that commonly twines at the tips (A.N. Egan, personal observation). *Haymondia wallichii*, on the other hand, less often twines at the tips, but rather forms a large shrub or even a

small tree. It is resolved as a lineage apart within the basal polytomy. Habit may yet be a good diagnostic character at the tribal and subtribal level, pending greater resolution of the backbone.

3.3. Tribal and subtribal relationships

3.3.1. Of Pueraria

A number of the taxonomic headaches that have plagued subtribal and even tribal delineations and relationships surrounding Phaseoleae have hinged on *Pueraria* – notably the troubling placement of *P. wallichii* within *Pueraria* DC. Due to a number of morphological characters deemed ancestral (and more closely associated with Diocleinae) by Lackey (1981) – woody habit, nodose to short-paniculate pseudoracemes and presence of canavanine – *Pueraria wallichii* is the one taxon that obviates placement of *Pueraria* and its segregate lineages fully within subtribe Glycininae. Instead, *Haymondia wallichii* avoids definitive subtribal associations but takes a place as an early branching lineage within the phaseoloid legumes set apart here by the clade inclusive of *Apios* and *Pueraria s.s.*

De Candolle (1825) was the first to present a working description of the traits that classically united tribe Phaseoleae as those legumes having trifoliate leaves and twining stems. That said, De Candolle classified *Pueraria*, then denoted by only two species, *P. tuberosa*, a woody vine, and *P. wallichii*, a scandent shrub, into his tribe Lotées, a group defined by foliaceous cotyledons, which *Pueraria* has. However, De Candolle did suggest that many of the twining species grouped in his tribe Lotées would likely be transferred to Phaseoleae were characteristics of germination known. Bentham (1837) moved *Pueraria* to tribe Phaseoleae

subtribe Diocleae where the genus overlaps in several characters, including peltate stipules (like *Dioclea* Kunth), nodose pseudoraceme, and woody habit, a classification that was carried through a number of later taxonomic treatments (e.g. Bentham, 1865; Hutchinson, 1964; Taubert, 1894). In several publications, Lackey noted that the understanding of *Pueraria*'s tribal and subtribal relationships were pivotal to a 'meaningful study of the entire group, because it serves as a basal point with connections from the Diocleinae...' and Glycininae (Lackey, 1981). Lackey (1981) was the first to definitively move *Pueraria* to Glycininae based on the bulk of species (his group A and our *Pueraria s.s.*) being more closely associated with Glycininae genera, sharing the not or scarcely nodose pseudoracemes, among other traits, but stating that this delineation between subtribe Glycininae and Diocleinae was not absolute. That said, all our analyses place *Pueraria s.s.* and all its segregated lineages except *P. wallichii* within subtribe Glycininae. Of interest is the placement of *Toxicopueraria*, the earliest branching lineage of subtribe Glycininae in our chloroplast datasets (Figs. 2, 4, 5). Lackey (1977c) suggested *Toxicopueraria* was generically, or even subtribally, misplaced due to the absence of paraveinal mesophyll, a feature also not found in subtribe Phaseolinae, which is phylogenetically close to *Toxicopueraria*. Although Lackey doubted the systematic value of this character at the subtribal or tribal levels (Lackey, 1977c), the intermediary placement of *Toxicopueraria* at the base of Glycininae closest to subtribe Phaseolinae provides a bridge to the evolutionary gain and loss of paraveinal mesophyll.

3.3.2. Within phaseoloid legumes

Subtribal classifications within Phaseoleae are largely resolved in our phylogenies and support previous findings. Glycininae, Phaseolinae and Kennediinae (sensu Lewis et al., 2005) are strongly supported as monophyletic groups by all analyses. Cajaninae is strongly supported as monophyletic by chloroplast analyses but includes *Apios* in the AS2 tree. A clade comprising *Erythrina*, *Psophocarpus*, and *Otoptera* is strongly resolved in all analyses and herein denoted as subtribe Erythrininae (noting the caveat that our concept of Erythrininae is far different from those before). Psoraleeae, Glycininae, Erythrininae, Cajaninae, and *Butea* comprise core Phaseoleae (sensu Lewis et al., 2005), relationships within which largely equate with previous works (de Queiroz et al., 2015; Hu et al., 2002; Kajita et al., 2001; Lavin et al., 2005; Li et al., 2013; McMahon and Sanderson, 2006; Stefanovic et al., 2009; Wojciechowski et al., 2004).

Associations among the early-branching lineages within Phaseoleae and among tribes Phaseoleae, Millettieae, Psoraleeae, and Desmodieae are more troublesome. Tribe Psoraleeae is placed outside subtribe Glycininae with strong support by our nuclear marker, AS2, a finding that differs from our chloroplast markers and all previous phylogenetic works, nearly all of which used chloroplast markers (de Queiroz et al., 2015; Hu et al., 2002; Kajita et al., 2001; Lavin et al., 2005; Li et al., 2013; McMahon and Sanderson, 2006; Stefanovic et al., 2009; Wojciechowski et al., 2004).

Furthermore, relationships between *Haymondia*, *Mucuna*, Kennediinae, Desmodieae, and the core Phaseoleae vary among DNA markers both in our analyses and in previous works. All previous analyses that sampled *Mucuna* and Desmodieae placed *Mucuna* as sister to Desmodieae (de Queiroz et al., 2015; Kajita et al., 2001; Li et al., 2013; Stefanovic et al., 2009) in agreement with our AS2 phylogeny (PP=0.79; Fig. 3). However, the *trnD-trnT* analysis places *Mucuna* as sister to *Haymondia* with moderate support (PP=0.91) and instead resolves Desmodieae as sister

to core Phaseoleae, but with weak support (PP=0.69; Fig. 5). Based on eight chloroplast genes, Stefanovic et al. (2009) determined Kennediinae as sister to core Phaseoleae with strong support, a clade then sister to Desmodieae + *Mucuna*, but with very short branch lengths. The most recent phylogenies based on chloroplast regions found support for *Shuteria* as sister to Kennediinae, with evidence suggesting a tritomy of it along with core Phaseoleae and Desmodieae + *Mucuna* (de Queiroz et al., 2015; Li et al., 2013). Relationships among these lineages differ across each of our molecular markers. Given the short branch lengths found by the eight-gene analysis of Stefanovic et al. (2009), resolution of this polytomy may be impossible due to rapid radiation, simultaneous cladogenesis (Hoelzer and Meinick, 1994), or hemiplasy resulting from ancient hybridization or incomplete lineage sorting (Avice and Robinson, 2008). Further taxon and molecular marker sampling may help to resolve relationships in these early-branching lineages.

3.3.3. Placement of newly sampled genera

Our sampling includes six genera placed for the first time within a molecular phylogenetic context in legumes. *Arthroclianthus* and *Nephrodesmus* are endemic to New Caledonia and comprise ~30 and 6 species, respectively. *Arthroclianthus* was placed in Hedysareae s.s. by Hutchinson (1964) and later moved to Desmodieae subtribe Desmodiinae Benth. by Ohashi et al. (1981), a placement supported by early molecular slot blot hybridization surveys (Bailey et al., 1997) that found association between *Arthroclianthus*, *Nephrodesmus*, and *Phyllodium* Desv. All our DNA markers support the sister relationship of *Arthroclianthus* and *Nephrodesmus* with strong support and confirm placement in Desmodieae (Figs. 2-5).

Melliniella is a mat-forming, monotypic genus from west central Africa whose author hypothesized a close relationship with *Alysicarpus*, *Pycnospora* R. Br. ex Wight & Arn. and *Pseudarthria* Wight & Arn. but disagreed with its placement in Hedysareae-Desmodiinae because of non-jointed fruits that dehisce adaxially (Harms, 1914). *Pycnospora* is also a monotypic genus found in Africa, India, Asia, and Australia noted for its black, bladderlike pods that dehisce along two valves, apically. Hutchinson (1964) erroneously placed *Melliniella* in his Lotononideae tribe based on the dehiscent, 2-valved fruit and created a new tribe, Pseudarthrieae, to accommodate *Pycnospora* and *Pseudarthria* because these taxa diverged from the classical description of Hedysareae as genera having non-articulated fruits. Our *trnD-trnT* analysis places *Pycnospora* as sister to *Melliniella* (PP=0.95; Fig. 5). All our analyses support the association of *Alysicarpus*, *Pycnospora*, and *Melliniella* as proposed by Harms (1914) and place these genera near *Desmodium* with strong support, an association promoted by Ohashi (1981).

Paracalyx includes six species originally included in *Cylista* Aiton, the type of which is now in *Rhynchosia*. Our analysis shows *Paracalyx* as sister to *Rhynchosia* with strong support (Fig. 5). However, more comprehensive sampling within *Rhynchosia*, a genus of nearly 250 species, and its close ally *Eriosema*, a genus of ~150 species, is needed to definitively determine the status and association of *Paracalyx*, especially as *Rhynchosia* is suggested herein to be paraphyletic. Our analyses support the long-held placement of *Paracalyx* in Phaseoleae subtribe Cajaninae (Bentham, 1865; Harms, 1914; Hutchinson, 1964; Lackey, 1977c; Taubert, 1894).

Eminia, a genus of 5-6 species, has long been thought to have a close association with *Pseudeminia* Verdc. and *Pseudovigna* Verdc. (Verdcourt, 1970), with Lackey (1977c) considering them potential congeners. With the exception of an erroneous placement by Hutchinson (1964) in Cajaninae, *Eminia* has been affiliated with Phaseoleae subtribe Glycininae

(Harms, 1914; Lackey, 1977c; Ohashi et al., 1981; Taubert, 1894). Our *trnD-trnT* analysis places *Eminia* as sister to *Pseudeminia* (PP=1.0; Fig. 5) within Glycininae, in support of the above.

Dysolobium (Benth.) Prain has eluded definitive associations based on morphology, at both the subtribal and generic levels. It has been variously placed within *Phaseolus* as *Phaseolus grandis* section *Dysolobium* (Bentham, 1852) or in *Vigna* (Baker, 1879) due to its bearded style, all relegating *Dysolobium* to subtribe Phaseolinae Benth. Kurz (1874) sunk the species under *Canavalia* within Diocleinae based on pod characters while Prain (1897) thought all these associations dubious. Several other treatments retained *Dysolobium* in Phaseolinae (Harms, 1914; Hutchinson, 1964; Lackey, 1981; Taubert, 1894), although Lackey (Lackey, 1977b, c) did so with reservation, stating that there was strong affinity with *Psophocarpus* based on their winged pods, curved stigmas, long, straight style hairs, and sometime hairy seeds, but with *Dysolobium* also having the habit of some *Pueraria* or *Dioclea*, the seeds of some Glycininae and the bearded style central to Phaseolinae. Unfortunately, we were only able to obtain viable sequence for our nuclear gene, *AS2*, which places *Dysolobium* outside Glycininae and Phaseolinae as a distinct lineage amidst an unresolved backbone (Fig. 3). More molecular and taxonomic sampling is needed to definitively determine the subtribal and generic associations of *Dysolobium*.

3.4. Critical differences between nuclear and cpDNA datasets

Both nuclear and chloroplast markers gave similar trees, indicating the presence of five clades of *Pueraria*. This congruence is key to our interpretation of how many separate evolutionary lineages are currently circumscribed within the genus *Pueraria*: chloroplast DNA is usually maternally inherited while nuclear is a mix of paternal and maternal, congruence across

both sets of data provides strong support for our parsing of a polyphyletic *Pueraria*. However, the SH test found that the *matK* and *AS2* gene trees were significantly different. Chloroplast and nuclear genes have been shown to give slightly different topologies (Soltis and Kuzoff, 1995; Zhang et al., 2010) in previous studies of other taxa. Our chloroplast *matK* tree strongly favors the nesting of the Psoraleeae tribe within the Glycininae subtribe (Fig. 4) whereas our nuclear marker, *AS2*, supports a large Glycininae clade that excludes Psoraleeae (Fig. 3). The evolutionary placement of Psoraleeae has been discordant in previous works. Stefanovic et al. (2009) included Psoraleeae in an eight-gene chloroplast analysis across Glycininae and found that a combined analysis of all eight chloroplast gene regions favored Psoraleeae sister to *Amphicarpaea* with moderate support, a clade then sister to the strongly supported *Glycine* + *Teramnus* sister pair. However, when the *ycf4* gene was removed from the analysis, Psoraleeae and *Teramnus* switched places, with Psoraleeae strongly supported as sister to *Glycine*. Analysis of the chloroplast-expressed nuclear gene for glutamine synthetase, *nep-GS*, showed Psoraleeae sister to *Cologania* with strong support and far removed from the clade including *Glycine* (Doyle et al., 2003). Furthermore, an analysis of seven linked nuclear genes sampled across Glycininae found considerable discord concerning the sister group to *Glycine* and the relative placement of Psoraleeae (A.N. Egan, unpublished data).

A potential explanation for this is chloroplast capture: the replacement of one plant's chloroplast genome with another (Acosta and Premoli, 2010; Tsitrone et al., 2003). Chloroplast capture is often the result of hybridization (Soltis and Kuzoff, 1995; Tsitrone et al., 2003), although research has also shown capture through horizontal gene transfer via grafting (Stegemann et al., 2012). Regardless of mechanism, what we have is a hypothesis that Psoraleeae underwent an ancient hybridization event with current or ancestral lineages within Glycininae,

potentially involving one or more progenitor genomes of the polyploid *Glycine*. Future work will attempt to look into this hypothesis further.

3.4.1. On the phylogenetic utility of AS2

The development of nuclear markers for phylogeny reconstruction in legumes has seen considerable advancement in the last decade, with researchers taking advantage of sequenced legume genomes to screen and produce several markers for phylogeny reconstruction (Choi et al., 2006; de Sousa et al., 2014; Kim et al., 2015). Several studies have focused on determining the utility of such nuclear markers across varying depths within a phylogeny (Babineau et al., 2013; Farruggia and Howard, 2011). Because shorter regions often amplify better when working with degraded DNA (Särkinen et al., 2012), we sought a nuclear marker that would provide considerable variation across a short region while amplifying across the phylogenetic breadth within legumes. Out of nearly 50 potential nuclear markers screened under these criteria, the AS2 marker amplified readily across the phaseoloid and millettoid legumes and offers considerable variation (A.N. Egan unpublished data), even surpassing the % sites that are parsimony informative as compared to our chloroplast data sets used herein (Table 2). AS2 was more resolving at the tips of the tree, supporting all the major subtribal clades as those determined by chloroplast datasets, although with some relational differences (i.e. the placement of Psoraleeae). Our gene family analysis suggests that the primers designed are paralog specific and amplified a single ortholog across all our taxa: all our AS2 amplified sequences clustered in a single clade along with At1g65620 (BS=79; Supplementary Fig. 3). It is possible that some undetectable paralogy/orthology conflation exists within our AS2 sequences due to unseen birth/death events, which could be causing discordance with chloroplast datasets or the

unresolved backbone that we see. However, considering the variable placement of Psoraleeae across a number of previous works, this may simply be another example of an underlying evolutionary phenomenon, such as ancient hybridization or incomplete lineage sorting, which could have impacted the evolution of the morphologically distinct Psoraleeae tribe. Alternatively, it is possible that, as a coding gene, *AS2* may be under positive selection, a conclusion invoked by Stefanovic et al. (2009) as a potential cause of gene tree discordance within their eight-gene dataset. Further taxonomic sampling with *AS2* and testing its application across a broader phylogenetic breadth within Fabaceae may help to elucidate these mechanisms.

Authors' contributions

All authors contributed to various aspects of this work (ordered by degree of contribution from most to least): conceived the study (ANE); aided in study design (ANE, WC, MV); obtained funds for the research (ANE); coordinated activities (ANE); obtained sample materials (ANE, WC); extracted, amplified, and sequenced DNA (WC, ANE); conducted data analyses (MV, WC, ANE); prepared figures (MV, ANE, WC); and contributed to preparation of the manuscript (ANE, MV, WC).

Acknowledgements

We express thanks to the curators of the following herbaria for access to or loans of specimens: US, K, P, BKF, GH, E, MO, NY, KUN, CDBI, HZU, and HITBC. Thanks to Xin-Fen Gao, Xu Bo, Tadashi Kajita, Tomoko Otao, Tetsukazu Yahara, Firouzeh Javadi, Luxian Liu, and Chengxin Fu for locality information and logistical support in the field and to two anonymous reviewers who helped improve the manuscript. We sincerely thank East Carolina University and

788 the Smithsonian Institution National Museum of Natural History for support. This work was
789 supported by the U.S. National Science Foundation (DEB-1352217).

790

791 **References**

792 Acosta, M.C., Premoli, A.C., 2010. Evidence of chloroplast capture in South American
793 *Nothofagus* (subgenus *Nothofagus*, *Nothofagaceae*). *Mol. Phylogen. Evol.* 54, 235-242.

794 Akaike, H., 1973. Information theory as an extension of the maximum-likelihood principle. In:
795 Petrov, B., Csake, F. (Eds.), *Second International Symposium on Information Theory*.
796 *Akademiai Kiado, Budapest*, pp. 267-281.

797 Avise, J.C., Robinson, T.J., 2008. Hemiplasy: a new term in the lexicon of phylogenetics. *Syst.*
798 *Biol.* 57, 503-507.

799 Babineau, M., Gagnon, E., Bruneau, A., 2013. Phylogenetic utility of 19 low copy nuclear genes
800 in closely related genera and species of caesalpinoid legumes. *S. Afr. J. Bot.* 89, 94-105.

801 Bailey, C.D., Doyle, J.J., Kajita, T., Nemoto, T., Ohashi, H., 1997. The chloroplast *rpl2* intron
802 and *ORF184* as phylogenetic markers in the legume tribe Desmodieae. *Syst. Bot.* 22, 133-138.

803 Baker, J.G., 1879. *Leguminosae. The flora of British India* /By J. D. Hooker assisted by various
804 botanists. Published under the authority of the secretary of state for India in council. L. Reeve &
805 Co., London.

806 Bell, E.A., 1981. Non-protein amino acids in the Leguminosae. In: Polhill, R.M., Raven, P.H.
807 (Eds.), *Advances in Legume Systematics Part 2*. Royal Botanic Gardens Kew, Richmond, pp.
808 489-499.

809 Bell, E.A., Lackey, J.A., Polhill, R.M., 1978. Systematic significance of canavanine in the
810 *Papilionoideae* (*Faboideae*). *Biochem. Syst. Ecol.* 6, 201-212.

811 Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T., Madueño, F., 2015. Genetic control of
812 inflorescence architecture in legumes. *Front. Plant Sci.* 6.

813 Bentham, G., 1837. *Commentationes de Leguminosarum Generibus*. Zollinger, Vienna.

814 Bentham, G., 1852. *Leguminosae*. In: Miquel, F.A.W. (Ed.), *Plantae Junghuhninae*. A.W.
815 Sythoff, Leiden, pp. 205-269.

816 Bentham, G., 1865. *Leguminosae*. In: Bentham, G., Hooker, J.D. (Eds.), *Genera Plantarum*.
817 Lovell Reed & Co., London, pp. 434-600.

818 Bentham, G., 1867. Notes on Pueraria, DC., correctly referred by the author to Phaseoleae. J.
819 Linn. Soc. Lond. Bot. 9, 121-125.

820 Berbel, A., Ferrándiz, C., Hecht, V., Dalmais, M., Lund, O.S., Sussmilch, F.C., Taylor, S.A.,
821 Bendahmane, A., Ellis, T.N., Beltrán, J.P., 2012. VEGETATIVE1 is essential for development
822 of the compound inflorescence in pea. Nat. Commun. 3, 797.

823 Birdsong, B.A., Alston, R., Turner, B.L., 1960. Distribution of canavanine in the family
824 Leguminosae as related to phyletic groupings. Can. J. Botany 38, 499-505.

825 Cagle, W., 2013. Parsing Polyphyletic Pueraria: Delimiting distinct evolutionary lineages
826 through phylogeny. Ph.D. Dissertation., Greenville, North Carolina, U.S.A.

827 Cardoso, D., de Queiroz, L.P., Pennington, R.T., de Lima, H.C., Fonty, É., Wojciechowski,
828 M.F., Lavin, M., 2012. Revisiting the phylogeny of papilionoid legumes: New insights from
829 comprehensively sampled early-branching lineages. Am. J. Bot. 99, 1991-2013.

830 Choi, H.-K., Luckow, M.A., Doyle, J., Cook, D.R., 2006. Development of nuclear gene-derived
831 molecular markers linked to legume genetic maps. Mol. Genet. Genomics 276, 56-70.

832 Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new
833 heuristics and parallel computing. Nat. Methods 9, 772-772.

834 De Candolle, A.P., 1825. Mémoires sur les Légumineuses. A. Belin, Paris.

835 de Queiroz, L.P., Pastore, J.F.B., Cardoso, D., Snak, C., Lima, A.L.d.C., Gagnon, E.,
836 Vatanparast, M., Holland, A.E., Egan, A.N., 2015. A multilocus phylogenetic analysis reveals
837 the monophyly of a recircumscribed papilionoid legume tribe Diocleae with well-supported
838 generic relationships. Mol. Phylogen. Evol. 90, 1-19.

839 de Sousa, F., Bertrand, Y.J., Nylander, S., Oxelman, B., Eriksson, J.S., Pfeil, B.E., 2014.
840 Phylogenetic properties of 50 nuclear loci in *Medicago* (Leguminosae) generated using
841 multiplexed sequence capture and next-generation sequencing. PloS one 9, e109704.

842 Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf
843 tissue. Phytochem. Bull. 19, 11-15.

844 Doyle, J.J., Doyle, J.L., 1993. Chloroplast DNA phylogeny of the papilionoid legume tribe
845 Phaseoleae. Syst. Bot., 309-327.

846 Doyle, J.J., Doyle, J.L., Harbison, C., 2003. Chloroplast-expressed glutamine synthetase in
847 *Glycine* and related Leguminosae: phylogeny, gene duplication, and ancient polyploidy. Syst.
848 Bot. 28, 567-577.

849 Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with
850 BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969-1973.

851 Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high
852 throughput. *Nucleic Acids Res.* 32, 1792-1797.

853 Egan, A.N., Crandall, K.A., 2008a. Divergence and diversification in North American Psoraleeae
854 (Fabaceae) due to climate change. *BMC Biol.* 6, 55.

855 Egan, A.N., Crandall, K.A., 2008b. Incorporating gaps as phylogenetic characters across eight
856 DNA regions: ramifications for North American Psoraleeae (Leguminosae). *Mol. Phylogen.
857 Evol.* 46, 532-546.

858 Egan, A.N., Pan, B., 2015a. *Pueraria stracheyi*, a new synonym to *Apios carnea* (Fabaceae).
859 *Phytotaxa* 218, 147-155.

860 Egan, A.N., Pan, B., 2015b. Resolution of polyphyly in *Pueraria* (Leguminosae,
861 Papilionoideae): The creation of two new genera, *Haymondia* and *Toxicopueraria*, the
862 resurrection of *Neustanthus*, and a new combination in *Teyleria*. *Phytotaxa* 218, 201-226.

863 Egan, A.N., Puttock, C.F., 2016. The genus *Haymondia* A.N.Egan & B.Pan bis (Fabaceae) in
864 Thailand. *Thai Forest Bull.* 44, In Press.

865 Evans, S.V., Fellows, L.E., Bell, E.A., 1985. Distribution and systematic significance of basic
866 non-protein amino acids and amines in the Tephrosieae. *Biochem. Syst. Ecol.* 13, 271-302.

867 Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, G., 1994. Testing significance of incongruence.
868 *Cladistics* 10, 315-319.

869 Farruggia, F.T., Howard, J.H., 2011. Examination of five nuclear markers for phylogenetic study
870 of Hologalegina (Leguminosae). *Brittonia* 63, 489-499.

871 Gernhard, T., Hartmann, K., Steel, M., 2008. Stochastic properties of generalised Yule models,
872 with biodiversity applications. *J. Math. Biol.* 57, 713-735.

873 Gianoli, E., 2004. Evolution of a climbing habit promotes diversification in flowering plants.
874 *Proc. R. Soc. Lond. B Biol. Sci.* 271, 2011-2015.

875 Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large
876 phylogenies by maximum likelihood. *Syst. Biol.* 52, 696-704.

877 Harms, H., 1914. Leguminosae Africanæ. VII. *Bot. Jahrb. Syst.* 51, 359-368.

878 Hilu, K.W., Borsch, T., Müller, K., Soltis, D.E., Soltis, P.S., Savolainen, V., Chase, M.W.,
879 Powell, M.P., Alice, L.A., Evans, R., 2003. Angiosperm phylogeny based on *matK* sequence
880 information. *Am. J. Bot.* 90, 1758-1776.

881 Hoelzer, G.A., Meinick, D.J., 1994. Patterns of speciation and limits to phylogenetic resolution.
882 *Trends Ecol. Evol.* 9, 104-107.

- 883 Hooker, J.D., 1849. Niger Flora. Hippolyte Bailliere, London.
- 884 Hu, J.-M., Lavin, M., Wojciechowski, M.F., Sanderson, M.J., 2000. Phylogenetic systematics of
885 the tribe millettieae (Leguminosae) based on chloroplast *trnK/matK* sequences and its
886 implications for evolutionary patterns in papilionoideae. Am. J. Bot. 87, 418-430.
- 887 Hu, J.-M., Lavin, M., Wojciechowski, M.F., Sanderson, M.J., 2002. Phylogenetic analysis of
888 nuclear ribosomal ITS/5.8 S sequences in the tribe Millettieae (Fabaceae): *Poecilanthe*-
889 *Cyclolobium*, the core Millettieae, and the *Callerya* group. Syst. Bot. 27, 722-733.
- 890 Hutchinson, J., 1964. The genera of flowering plants. Part I. The Clarendon Press, Oxford.
- 891 Iwakawa, H., Ueno, Y., Semiarti, E., Onouchi, H., Kojima, S., Tsukaya, H., Hasebe, M., Soma,
892 T., Ikezaki, M., Machida, C., 2002. The ASYMMETRIC LEAVES2 gene of *Arabidopsis*
893 *thaliana*, required for formation of a symmetric flat leaf lamina, encodes a member of a novel
894 family of proteins characterized by cysteine repeats and a leucine zipper. Plant Cell Physiol. 43,
895 467-478.
- 896 Kajita, T., Ohashi, H., Tateishi, Y., Bailey, C.D., Doyle, J.J., 2001. *rbcL* and Legume phylogeny,
897 with particular reference to Phaseoleae, Millettieae, and allies. Syst. Bot. 26, 515-536.
- 898 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
899 improvements in performance and usability. Mol. Biol. Evol. 30, 772-780.
- 900 Keng, H., 1974. Economic plants of ancient north China as mentioned in *Shih Ching* (Book of).
901 Econ. Bot. 28, 391-410.
- 902 Kim, J.-H., Lee, C., Hyung, D., Jo, Y.-J., Park, J.-S., Cook, D.R., Choi, H.-K., 2015. CSGM
903 Designer: a platform for designing cross-species intron-spanning genic markers linked with
904 genome information of legumes. Plant Methods 11, 1.
- 905 Kitagawa, M., Tomiyama, T., 1929. A new amino-compound in the jack bean and a
906 corresponding new ferment.(I). J. Biochem. 11, 265-271.
- 907 Kurz, S., 1874. Descriptions of a few Indian plants. J. As. Soc. Beng. 43, 181-189.
- 908 Lackey, J., 1977a. *Neonotonia*, a new generic name to include *Glycine wightii* (Arnott)
909 Verdcourt (Leguminosae, Papilionoideae). Phytologia 37, 209-212.
- 910 Lackey, J.A., 1977b. A revised classification of Phaseoleae (Leguminosae: Papilionoideae) and
911 its relation to canavanine distribution. Bot. J. Linn. Soc. 74, 163-178.
- 912 Lackey, J.A., 1977c. A synopsis of Phaseoleae (Leguminosae, Papilionoideae). Ph.D.
913 Dissertation, Ames, Iowa, U.S.A.

- 914 Lackey, J.A., 1981. Phaseoleae. In: Polhill, R.M., Raven, P.H. (Eds.), Advances in Legume
915 Systematics. Royal Botanic Gardens, Kew, pp. 301-321.
- 916 Lamprecht, H., 1947. The inheritance of the number of flowers per inflorescence and the origin
917 of *Pisum*, illustrated by polymeric genes. Agri Hort. Genet. 5, 16-25.
- 918 Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of
919 Leguminosae implicates a rapid diversification of lineages during the Tertiary. Syst. Biol. 54,
920 575-594.
- 921 Lavin, M., Pennington, R.T., Klitgaard, B.B., Sprent, J.I., de Lima, H.C., Gasson, P.E., 2001.
922 The dalbergioid legumes (Fabaceae): delimitation of a pantropical monophyletic clade. Am. J.
923 Bot. 88, 503-533.
- 924 Le, Z.-F., Zhu, X.-Y., 2009. *Pueraria yunnanensis* (Fabaceae) reinstated. Ann. Bot. Fenn. 46,
925 419-424.
- 926 Lee, J., Hymowitz, T., 2001. A molecular phylogenetic study of the subtribe Glycininae
927 (Leguminosae) derived from the chloroplast DNA *rps16* intron sequences. Am. J. Bot. 88, 2064-
928 2073.
- 929 Lewis, G., Schrire, B.P., MacKinder, B., Lock, M., 2005. Legumes of the World. Royal Botanic
930 Gardens, Kew.
- 931 Li, H., Wang, W., Lin, L., Zhu, X., Li, J., Zhu, X., Chen, Z., 2013. Diversification of the
932 phaseoloid legumes: effects of climate change, range expansion and habit shift. Front. Plant Sci.
933 4.
- 934 LPWG, 2013a. Legume phylogeny and classification in the 21st century: Progress, prospects and
935 lessons for other species-rich clades. Taxon 62, 217-248.
- 936 LPWG, 2013b. Towards a new classification system for legumes: Progress report from the 6th
937 International Legume Conference. S. Afr. J. Bot. 89, 3-9.
- 938 Matsumura, Y., Iwakawa, H., Machida, Y., Machida, C., 2009. Characterization of genes in the
939 ASYMMETRIC LEAVES2/LATERAL ORGAN BOUNDARIES (AS2/LOB) family in
940 *Arabidopsis thaliana*, and functional and molecular comparisons between AS2 and other family
941 members. Plant J. 58, 525-537.
- 942 McMahon, M.M., Sanderson, M.J., 2006. Phylogenetic supermatrix analysis of GenBank
943 sequences from 2228 papilionoid legumes. Syst. Biol. 55, 818-836.
- 944 Miller, J.H., Edwards, M.B., 1983. Kudzu: where did it come from? and how can we stop it?
945 South. J. Appl. For. 7, 165-169.

946 Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for
947 inference of large phylogenetic trees. Gateway Computing Environments Workshop (GCE),
948 2010. IEEE, pp. 1-8.

949 Muller, K.F., 2005. SeqState - primer design and sequence statistics for phylogenetic DNA data
950 sets. Appl. Bioinformatics 4, 65-69.

951 Ohashi, H., 2005. A new species of *Pueraria* (Leguminosae) from Guizhou, China. J. Jpn. Bot.
952 80, 9-13.

953 Ohashi, H., Iokawa, Y., 2006. A new species, *Pueraria xyszhu* (Leguminosae) from Yunnan,
954 China, with pollen stainability and pollen morphology in comparison to related species. J. Jpn.
955 Bot. 81, 26-34.

956 Ohashi, H., Polhill, R.M., Schubert, B.G., 1981. Desmodieae (Benth.) Hutch.(1964). In: Polhill,
957 R.M., Raven, P.H. (Eds.), Advances in Legume Systematics, Part 1. Royal Botanic Gardens
958 Kew, Richmond, pp. 292-300.

959 Pagel, M., Meade, A., 2006. Bayesian analysis of correlated evolution of discrete characters by
960 reversible- jump Markov chain Monte Carlo. Am. Nat. 167, 808-825.

961 Pagel, M., Meade, A., 2013. Bayes Traits v2. Available at:
962 <http://www.evolution.rdg.ac.uk/BayesTraits.html>

963 Pan, B., Liu, B., Yu, Z.-X., Yang, Y.-Q., 2015. *Pueraria grandiflora* (Fabaceae), a new species
964 from Southwest China. Phytotaxa 203, 287-291.

965 Perry, L.M., Metzger, J., 1980. Medicinal plants of East and Southeast Asia: attributed properties
966 and uses. M.I.T. Press.

967 Polhill, R.M., Bisby, F.A., Buckingham, J., Harborne, J.B., 1994. Phytochemical dictionary of
968 the Leguminosae. Chapman & Hall, London.

969 Prain, D., 1897. Noviciae Indicae XV. Some additional Leguminosae. J. As. Soc. Beng. 66, 347-
970 519.

971 Prenner, G., 2013. Papilionoid inflorescences revisited (Leguminosae-Papilionoideae). Ann. Bot.
972 112, 1567-1576.

973 Rambaut, A., 2002. Se-AL: Sequence alignment editor. Available at:
974 <http://tree.bio.ed.ac.uk/software/seal/>.

975 Rambaut, A., 2014. FigTree v1.4.2. Available at: <http://tree.bio.ed.ac.uk/software/figtree>,
976 University of Edinburgh, Edinburgh, UK

977 Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from
978 <http://beast.bio.ed.ac.uk/Tracer>.

979 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B.,
980 Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic
981 inference and model choice across a large model space. *Syst. Biol.* 61, 539-542.

982 Rueda, B., Blake, R., Nicholson, C., Fox, D., Tedeschi, L., Pell, A., Fernandes, E., Valentim, J.,
983 Carneiro, J., 2003. Production and economic potentials of cattle in pasture-based systems of the
984 western Amazon region of Brazil. *J. Anim. Sci.* 81, 2923-2937.

985 Särkinen, T., Staats, M., Richardson, J.E., Cowan, R.S., Bakker, F.T., 2012. How to open the
986 treasure chest? Optimising DNA extraction from herbarium specimens. *PloS one* 7, e43808.

987 Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W.S., Miller, J., Siripun, K.C., Winder,
988 C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: Relative utility of 21
989 noncoding chloroplast DNA sequences for phylogenetic analysis. *Am. J. Bot.* 92, 142-166.

990 Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications
991 to phylogenetic inference. *Mol. Biol. Evol.* 16, 1114-1116.

992 Shurtleff, W., Aoyagi, A., 1977. *The Book of Kudzu, a Culinary and Healing Guide*. Autumn
993 Press, Brookline, MA.

994 Simmons, M.P., Ochoterena, H., 2000. Gaps as characters in sequence-based phylogenetic
995 analysis. *Syst. Biol.* 49, 369-381.

996 Singer, S., Sollinger, J., Maki, S., Fishbach, J., Short, B., Reinke, C., Fick, J., Cox, L., McCall,
997 A., Mullen, H., 1999. Inflorescence architecture: a developmental genetics approach. *Bot. Rev.*
998 65, 385-410.

999 Smith, S.A., Donoghue, M.J., 2008. Rates of molecular evolution are linked to life history in
1000 flowering plants. *Science* 322, 86-89.

1001 Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W., Berger, J.,
1002 Greene, S.L., Nelson, M.N., Besharat, N., Vymyslický, T., Toker, C., Saxena, R.K., Roorkiwal,
1003 M., Pandey, M.K., Hu, J., Li, Y.H., Wang, L.X., Guo, Y., Qiu, L.J., Redden, R.J., Varshney,
1004 R.K., 2015. Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev.*
1005 *Plant Sci.* 34, 43-104.

1006 Soltis, D.E., Kuzoff, R.K., 1995. Discordance between nuclear and chloroplast phylogenies in
1007 the *Heuchera* group (Saxifragaceae). *Evolution* 49, 727-742.

1008 Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with
1009 thousands of taxa and mixed models. *Bioinformatics* 22, 2688-2690.

1010 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1011 large phylogenies. *Bioinformatics* 30, 1312-1313.

- 1012 Stefanovic, S., Pfeil, B.E., Palmer, J.D., Doyle, J.J., 2009. Relationships among phaseoloid
1013 legumes based on sequences from eight chloroplast regions. *Syst. Bot.* 34, 115-128.
- 1014 Stegemann, S., Keuthe, M., Greiner, S., Bock, R., 2012. Horizontal transfer of chloroplast
1015 genomes between plant species. *Proc. Natl. Acad. Sci. USA* 109, 2434-2438.
- 1016 Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing.
1017 *Bioinformatics* 26, 1569-1571.
- 1018 Sussmilch, F.C., Berbel, A., Hecht, V., Vander Schoor, J.K., Ferrándiz, C., Madueño, F., Weller,
1019 J.L., 2015. Pea VEGETATIVE2 is an FD homolog that is essential for flowering and compound
1020 inflorescence development. *Plant Cell* 27, 1046-1060.
- 1021 Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods).
1022 Sinauer Assoc., Sunderland, Mass.
- 1023 Taubert, P., 1894. Leguminosae. In: Engler, A., Prantl, K. (Eds.), *Die naturalischen*
1024 *Pflanzenfamilien*. Engelmann, Leipzig.
- 1025 Thiers, B., continuously updated. Index Herbariorum: A global directory of public herbaria and
1026 associated staff. New York Botanical Garden's Virtual Herbarium.
1027 <http://sweetgum.nybg.org/science/ih/> .
- 1028 Tsitrone, A., Kirkpatrick, M., Levin, D.A., 2003. A model for chloroplast capture. *Evolution* 57,
1029 1776-1782.
- 1030 Valentim, J., Andrade, C., 2005. Tropical kudzu (*Pueraria phaseoloides*): successful adoption in
1031 sustainable cattle production systems in the western Brazilian Amazon. *International Grasslands*
1032 *Congress*, p. 328J.
- 1033 van der Maesen, L.J.G., 1985. Revision of the genus *Pueraria* with some notes on *Teyleria*
1034 *Leguminosae*. *Wag. Ag. Un. P.* 85, 1-130.
- 1035 van der Maesen, L.J.G., 1994. *Pueraria*, the Kudzu and its relatives, an update of the taxonomy.
1036 In: Sorensen, M. (Ed.), *Proceedings of the First International Symposium on Tuberous Legumes*,
1037 Gualdeloupe, F.W.I. DSR Boghandel, Fredericksberg, Denmark, pp. 55-86.
- 1038 van der Maesen, L.J.G., 2002. *Pueraria*: botanical characteristics. In: Keung, W.M. (Ed.),
1039 *Pueraria: The genus Pueraria*. Taylor & Francis, New York City, NY, pp. 1-28.
- 1040 van der Maesen, L.J.G., Almeida, S.M., 1988. Two corrections to the nomenclature in the
1041 revision of *Pueraria* DC. *J. Bombay Nat. Hist. Soc.* 85, 233-234.
- 1042 Vatanparast, M., Takayama, K., Sousa, M.S., Tateishi, Y., Kajita, T., 2011. Origin of Hawaiian
1043 endemic species of *Canavalia* (Fabaceae) from sea-dispersed species revealed by chloroplast and
1044 nuclear DNA sequences. *J. Jpn. Bot.* 86, 15-25.

- 1045 Verdcourt, B., 1970. Studies in the Leguminosae-Papilionoideae for the 'Flora of Tropical East
1046 Africa': III. Kew Bulletin 24, 379-447.
- 1047 Wink, M., Mohamed, G.I.A., 2003. Evolution of chemical defense traits in the Leguminosae:
1048 mapping of distribution patterns of secondary metabolites on a molecular phylogeny inferred
1049 from nucleotide sequences of the *rbcL* gene. Biochem. Syst. Ecol. 31, 897-917.
- 1050 Wojciechowski, M.F., 2003. Reconstructing the phylogeny of legumes (Leguminosae): an early
1051 21st century perspective. Advances in legume systematics, part 10, 5-35.
- 1052 Wojciechowski, M.F., Lavin, M., Sanderson, M.J., 2004. A phylogeny of legumes
1053 (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported
1054 subclades within the family. Am. J. Bot. 91, 1846-1862.
- 1055 Yahara, T., Javadi, F., Onoda, Y., de Queiroz, L.P., Faith, D.P., Prado, D.E., Akasaka, M.,
1056 Kadoya, T., Ishihama, F., Davies, S., Slik, J.W.F., Yi, T., Ma, K., Bin, C., Darnaedi, D.,
1057 Pennington, R.T., Tuda, M., Shimada, M., Ito, M., Egan, A.N., Buerki, S., Raes, N., Kajita, T.,
1058 Vatanparast, M., Mimura, M., Tachida, H., Iwasa, Y., Smith, G.F., Victor, J.E., Nkonki, T.,
1059 2013. Global legume diversity assessment: Concepts, key indicators, and strategies. Taxon 62,
1060 249-266.
- 1061 Yule, G., 1924. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis.
1062 Phil. Trans. Roy. Soc. Lond. B Biol. Sci. 213, 21-87.
- 1063 Zhang, Z., Li, C., Li, J., 2010. Conflicting phylogenies of section *Macrantha* (*Acer*, *Aceroideae*,
1064 *Sapindaceae*) based on chloroplast and nuclear DNA. Syst. Bot. 35, 801-810.
- 1065 Zwickl, D., 2006. GARLI: genetic algorithm for rapid likelihood inference. See [http://www.bio.](http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html)
1066 [utexas.edu/faculty/antisense/garli/Garli.html](http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html).
1067
1068
1069

Figures Legends

Figure 1. Native (black) and introduced (green; primarily *P. montana* var. *lobata* and *P. phaseoloides*) range of *Pueraria*. Modified from van der Maesen (1985).

Figure 2. The 50% majority rule consensus tree resulting from Bayesian analysis of the cpDNA dataset (*matK/trnD-trnT*). Numbers along branches are posterior probabilities and bootstrap values from Bayesian analysis and Maximum likelihood, respectively. Values <50% were not shown. Bold lines represent clades and subclades with PP >90% for major clades.

Figure 3. The 50% majority rule consensus tree resulting from Bayesian analysis of the AS2 sequences. Numbers along branches are posterior probabilities and bootstrap values from Bayesian analysis and Maximum likelihood, respectively. Values <50% were not shown. Bold lines represent clades and subclades with PP >90% for major clades.

Figure 4. The 50% majority rule consensus tree resulting from Bayesian analysis of the *matK* sequences. Numbers along branches are posterior probabilities. Circles are ancestral state reconstruction estimates for manner of stipule attachment and presence or absence of canavanine estimated using BayesTraits.

Figure 5. The 50% majority rule consensus tree resulting from Bayesian analysis of the *trnD-trnT* sequences. Numbers along branches are posterior probabilities.

1093 **Figure 6.** BEAST chronogram based on *matK*. Numbers at nodes refer to mean age of the nodes
1094 (Mya). Blue bars represent 95% confidence intervals (CI) for the estimated mean dates. Qua:
1095 Quaternary, Plio: Pliocene, Mya: Million years ago.
1096
1097
1098

1099 **Table 1:** Current circumscription of *Pueraria* DC accepted by the authors upon initiation of this
 1100 work, largely following van der Maesen (1985).

Taxon	Distribution
<i>P. alopecuroides</i> Craib ^a	China (Yunnan), Myanmar, Thailand
<i>P. bella</i> Prain ^c	India, Myanmar
<i>P. bouffordii</i> H. Ohashi ⁺	China (Guizhou)
<i>P. calycina</i> Franch. ^a	China (Yunnan)
<i>P. candollei</i> Benth. ^a	Bangladesh, India (Assam), Myanmar, Thailand
<i>P. candollei</i> var. <i>mirifica</i> (Airy Shaw & Suvat.) Niyomdham ^{a,*}	Thailand
<i>P. edulis</i> Pamp. ^a	Bhutan, China (Yunnan), India (Manipur, Sikkim)
<i>P. grandiflora</i> Bo Pan & Bing Liu	China (Yunnan, Sichuan)
<i>P. imbricata</i> Maesen ⁺	Laos, Thailand
<i>P. lacei</i> Craib ^a	Myanmar
<i>P. montana</i> (Lour.) Merr. ^a	China, Japan, Laos, Myanmar, Philippines, Taiwan, Thailand, Vietnam
<i>P. montana</i> var. <i>lobata</i> (Willd.) Sanjappa & Pradeep ^{a,*}	SE Asia, Pacific Islands, Introduced elsewhere
<i>P. montana</i> var. <i>thomsonii</i> (Benth.) Wiersema ex D.B. Ward ^a	Bhutan, China, N India, Laos, Myanmar, Philippines, Thailand, Vietnam
<i>P. peduncularis</i> (Benth.) Benth. ^d	Bhutan, China (Yunnan, Sichuan), N India, Myanmar, Nepal, Pakistan
<i>P. phaseoloides</i> (Roxb.) Benth. ^b	S, SE, and E Asia and introduced elsewhere
<i>P. phaseoloides</i> var. <i>javanica</i> (Benth.) Baker ^b	S, SE, and E Asia and introduced elsewhere
<i>P. phaseoloides</i> var. <i>subspicata</i> (Benth.) Maesen ^{b,*}	Bhutan, NE India, Myanmar, Thailand
<i>P. pulcherrima</i> (Koord.) Koord.-Schum. ^a	Insular SE Asia
<i>P. sikkimensis</i> Prain ^a	Bhutan, NE India
<i>P. stricta</i> Kurz ^{c,^}	China (Yunnan), Myanmar, Thailand
<i>P. tuberosa</i> (Willd.) DC. ^a	India, Nepal, Pakistan
<i>P. wallichii</i> DC. ^d	Bangladesh, China (Yunnan), NE India, Myanmar, Nepal, Thailand
<i>P. xyzhuui</i> H. Ohashi & Iokawa ⁺	China (Yunnan)
<i>P. yunnanensis</i> Franchet. ^d	China (Yunnan)

1101 a-d correspond to Lackey's (1977c) groups A, B, C, and D, respectively

1102 * recognized at specific rank by Lackey 1977c

1103 ^ recognized as three species by Lackey (1977c): *P. collettii* Prain, *P. brachycarpa* Kurz, and *P.*
 1104 *stricta*

1105 + described after Lackey (1977c)

1106 **Table 2:** DNA marker dataset statistics.

DNA Marker	No. gen.	No. spp.	No. <i>Pueraria</i> acc.	No. total acc.	Aligned length	% missing sites	% gapped sites	% constant sites*	% parsimony informative*
AS2	58	100	58	147	669	8.2%	15.4%	44.1%	42.3%
<i>matK</i>	58	98	60	149	1,128	29.8%	3.0%	53.0%	31.2%
<i>trnD-trnT</i>	57	89	55	134	2,283	24.7%	29.8%	52.8%	32.9%
<i>matK/trnD-trnT</i>	39	59	37	87	2,873	16.7%	21.2%	55.2%	30.6%

1107 * statistics computed on alignments where gaps were treated as missing data. No. is number; gen.
1108 is genera; spp. is species; acc. is accessions.

1109

1110

1111

1112 **Table 3:** Summary of divergence time estimates between *Pueraria* lineages.

MRCA	Values	matK	trnDT	matK-trnDT	AS2	Avg.
<i>Pueraria s.s.</i>	Mean	26.88	27.29	27.35	24.62	26.54
vs.	SD	2.03	1.46	1.51	2.23	
<i>P. wallichii</i>	95% HPD Interval	[22.89, 30.67]	[24.45, 30.16]	[24.42, 30.32]	[19.9, 28.51]	
<i>Pueraria s.s.</i>	Mean	15.68	17.01	16.09	14.96	15.93
vs.	SD	2.08	1.82	1.75	2.26	
<i>P. yunnanensis</i>	95% HPD Interval	[11.63, 19.51]	[13.43, 20.49]	[12.63, 19.40]	[10.39, 19.11]	
<i>Pueraria s.s.</i>	Mean	14.01	11.98	11.43	14.46	12.97
vs.	SD	1.63	1.43	1.40	2.33	
<i>P. stricta</i>	95% HPD Interval	[10.79, 17.08]	[9.28, 14.79]	[8.83, 14.25]	[9.92, 18.85]	
<i>Pueraria s.s.</i>	Mean	14.36	15.21	14.16	10.03	13.44
vs.	SD	1.66	1.70	1.61	1.82	
<i>P. phaseoloides</i>	95% HPD Interval	[11.19, 17.6]	[11.93, 18.51]	[11.09, 17.36]	[6.61, 13.60]	

1113

1114

1115
1116

Pueraria montana

Teyleria stricta

Neustanthus phaseoloides

Toxicopueraria yunnanensis

Haymondia wallichii

1117
1118

1121
1122