Well-resolved phylogeny supports repeated evolution of keel flowers as a synergistic contributor to papilionoid legume diversification

Liming Cai[1,2], Domingos Cardoso[3,4], Lydia G. Tressel[1], Chaehee Lee[1,5], Bikash Shrestha[1,6], In-Su Choi[7,8], Haroldo C. de Lima[4], Luciano P. de Queiroz[9], Tracey A. Ruhlman[1], Robert K. Jansen[1], and Martin F. Wojciechowski[7]

[1] Department of Integrative Biology, University of Texas at Austin, Austin, Texas, USA
[2] Department of Biology, University of Florida, Gainesville, Florida, USA
[3] Instituto de Biologia, Universidade Federal da Bahia, Salvador, Bahia, Brazil
[4] Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil
[5] Department of Plant Sciences, University of California Davis, Davis, California, USA
[6] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[7] School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[8] Department of Biological Sciences and Biotechnology, Hannam University, Daejeon, Korea
[9] Universidade Estadual de Feira de Santana, Feira de Santana, Bahia, Brazil

Authors for correspondence:
Liming Cai
Email: cail@ufl.edu
Domingos Cardoso
Email: cardosobot@gmail.com
Martin F. Wojciechowski
Email: mfwojciechowski@asu.edu

Summary

The butterfly-shaped keel flower is a highly successful floral form in angiosperms. These flowers steer the mechanical interaction with bees and thus are hypothesized to accelerate pollinator-driven diversification. The exceptionally labile evolution of keel flowers in Papilionoideae (Fabaceae) provides a suitable system to test this hypothesis.

Using 1,456 low-copy nuclear loci, we confidently resolve the early divergence history of Papilionoideae. Constrained by this backbone phylogeny, we generated a time tree for 3,326 Fabales to evaluate the tempo and mode of diversification within a state-dependent evolutionary framework.

The first keel flowers emerged around 59.0 Ma in Papilionoideae, predating the earliest fossil by 3–4 million years. The Miocene diversification of Papilionoideae coincided with rapid evolution of keel flowers. At least six independent origins and thirty-two losses of keel flowers were identified in Papilionoideae, Cercidoideae, and Polygalaceae. However, the state-dependent diversification model was not favored.

Lack of radiation associated with keel flowers suggests that diversification within Papilionoideae was not solely driven by pollinator-mediated selection, but instead an outcome of the synergistic effects of multiple innovations including nitrogen fixation and chemical defense as well as dispersal into subtropical and temperate regions.

Keywords: Fabales, floral evolution, pollination-driven speciation, coalescent method, phylogenomics, SSE model

INTRODUCTION

Animal pollination is one of the most striking features of angiosperms (Kay *et al.*, 2006). Adaptive radiations associated with floral specialization are among the most widely discussed concepts in evolutionary biology since the era of Darwin (1862). Comparative analyses revealed several floral traits including zygomorphy that are frequently linked with species-rich clades (Sargent, 2004; O'Meara *et al.*, 2016). Further ecological investigations provided mechanistic explanation through which zygomorphy can effectively guide pollinator behavior and exclude less effective pollinators via visual and structural aids (Neal *et al.*, 1998; Sargent, 2004; Yoder *et al.,* 2020). Variations in the spatial organization of the perianth, stamen, stigma, and nectar spur can further localize pollen deposition spots, thus increasing reproductive isolation (Ree, 2005; Muchhala & Potts, 2007). Moreover, higher pollination efficiency and precision in zygomorphic flowers may bolster species persistence by reducing extinction risks (Kay *et al.*, 2006). However, additional intrinsic (e.g., chemical defenses) and extrinsic factors (e.g., biome shifts) are often associated with lineages with specialized floral syndromes. This makes it challenging to understand the relative role of key innovations like floral specialization in promoting diversification. Indeed, about half of the comparative studies published to date did not identify a significant correlation between reproductive traits and increased diversification rate (Helmstetter *et al.,* 2023). Therefore, there is a growing trend in the field to recognize the synergistic multidriver hypothesis of angiosperm radiation rather than the key innovation hypothesis (Magallón & Castillo, 2009; Augusto *et al.,* 2014; Donoghue & Sanderson, 2015). Along these lines, papilionoid legumes (subfamily Papilionoideae, Fabaceae) are especially suitable to test key innovation versus synergistic hypotheses given their highly specialized keel flower and a suite of defense traits and biome shift history known to impact legume diversification (Lewis *et al.*, 2005).

Legumes are among the most ecologically and evolutionarily successful plant families on Earth. With approximately 23,000 species and 800 genera, they have defined biomes from tropical savannas to temperate sagebrush steppes (Schrire *et al.*, 2005; Lewis *et al.*, 2005; LPWG, 2017; Legume Data Portal, https://www.legumedata.org/). Members of Fabaceae subfamilies often exhibit lineage-specific floral traits tailored to particular pollinators, such

as the showy stamens in the Mimosoid clade for bees, the night-blooming Caesalpinioideae for moths and bats, and the distinctive butterfly-like keel flowers in Papilionoideae, which are closely linked to bee pollination (Arroyo, 1981). Keel flowers are zygomorphic with one highly differentiated standard petal, two wing petals, and a keel composed of two partially to fully fused petals enclosing the reproductive organs (Westerkamp, 1997). Connation among floral components, including stamens and keel petals, is common. They are most commonly found in two species-rich clades within Fabales—Papilionoideae and Polygalaceae, although with different ontogenetic origins (Bello *et al.*, 2010). The standard serves as a visual attraction while the wings form the landing complex for pollinators (Westerkamp & Weber, 1999). The zygomorphic form and fused stamens concealed by the keel can effectively restrict the direction of entry by pollinators and aid with precise pollen placement (Neal *et al.*, 1998). Furthermore, many keel flowers exhibit primary and secondary pollination mechanisms such as valvular, pump, explosive, and brush mechanisms to prevent pollen waste and deposit pollen on particular bee body parts (Leppik, 1966; Etcheverry *et al.*, 2012; Uluer *et al.*, 2022). Loss of keel flower traits in Papilionoideae, including actinomorphy or even exposure of reproductive organs, can lead to rapid transitions to hummingbird or bat pollination (Arroyo, 1981; Neill, 1987). Moreover, several angiosperm families independently evolved the keel flower architecture, including Ranunculaceae, Geraniaceae, Balsaminaceae, and Commelinaceae (Westerkamp, 1997). They exhibit structural and functional convergence for bee pollination, although with divergent developmental pathways (Tucker, 1999, 2003; Prenner, 2004; Prenner & Klitgaard, 2008; Bello *et al.,* 2012). Therefore, the evolution and maintenance of keel flowers is considered to be an outcome of pollinator-mediated selection with the potential to promote diversification (Grant, 1949; Stebbins, 1970; Rosas-Guerrero *et al.*, 2014; Stewart *et al.*, 2022).

To rigorously assess trait-driven diversification, a system with high trait variability is essential for robust comparative hypothesis testing (Ricklefs & Renner, 1994, 2000; Donoghue & Sanderson, 2015). The remarkable diversity in floral symmetry and architecture in the Papilionoideae offers an ideal model for such investigations. Although keel flowers are considered an essential feature and a hallmark of Papilionoideae (Polhill *et*

113 *al.*, 1981), there is considerable variation in floral symmetry in early diverging clades. For
114 example, members from the tribes Sophoreae and Swartzieae are frequently actinomorphic,
115 with five petals undifferentiated or even lost to various degrees (Pennington *et al.*, 2000;
116 Cardoso *et al.*, 2012, 2013b). Such morphological variability is comparable to that found in
117 the Caesalpinioideae, suggesting an exploratory phase when a broader range of ontogenetic
118 and morphological pathways were realized (Leite *et al.*, 2015). These genera are typically
119 species-poor and more than 80% have no more than ten extant species (Plants of the
120 World Online – POWO; https://powo.science.kew.org). In contrast, the keel flower is highly
121 constrained among the large non-protein-amino-acid accumulating (NPAAA) clade (ca.
122 10,200 species), which is mostly known for economically important crops like soybeans,
123 peas, lentils, and clovers (Choi *et al.*, 2022). These general observations have prompted our
124 speculation that the evolution of keel flowers may be associated with increased
125 diversification rates in Papilionoideae.
126
127 Testing these macroevolutionary hypotheses requires a robust and densely sampled
128 phylogeny. Over the past three decades, a general phylogenetic framework for Fabaceae
129 has been developed thanks to a tremendous network of global collaboration (LPWG, 2013,
130 2017). Six distinct subfamilies are recognized by the Legume Phylogeny Working Group
131 (LPWG), with Papilionoideae as the most species-rich subfamily comprising 14,000 species
132 in 503 genera (Lewis *et al.*, 2005; LPWG, 2017). However, the LPWG's (2017) phylogeny is
133 largely based on the plastid gene *matK*, which has obvious limitations when dealing with
134 rapid radiations or complex evolutionary scenarios such as hybridization (Barrett *et al.*,
135 2014; Morales-Briones *et al.*, 2021). In Papilionoideae, studies based on *matK* and complete
136 plastomes have consistently defined 22 well-supported clades but the relationships among
137 these clades remain poorly resolved (Wojciechowski *et al.*, 2004; McMahon & Sanderson,
138 2006; Cardoso *et al.*, 2012, 2013b; Zhang *et al.*, 2020; Choi *et al.*, 2022). Most significantly,
139 Swartzieae and the ADA clade have been individually proposed as sister to all remaining
140 Papilionoideae. Given that Swartzieae contains mostly non-keel flowered taxa with
141 frequent petal losses whereas the ADA clade contains largely keel flowered ones (Torke &
142 Schaal, 2008; Cardoso *et al.*, 2015), their placement will have a strong bearing on the timing
143 and evolutionary trajectory of keel flowers. Furthermore, many of the 22 clades are

144 characterized by tropical trees that are scarce in distribution and underrepresented in
145 previous studies. These include species-poor clades such as the African *Amphimas* Pierre ex
146 Harms, the Amazonian *Aldina* Endl., and the North American *Dermatophyllum* Scheele
147 (Cardoso *et al.*, 2013). Along these lines, recent advances in legume phylogeny using multi-
148 locus nuclear datasets have promising applications in resolving complex evolutionary
149 histories (Koenen *et al.*, 2021; Zhao *et al.*, 2021) and inspired us to apply this technique to
150 address the recalcitrant relationships within Papilionoideae.
151
152 Here, we examined the primary versus synergistic role of keel flower evolution in
153 Papilionoideae evolution. We first generated a phylogenomic dataset of 1,456 low-copy
154 genes from 287 species with a focus on recalcitrant clades. Using both coalescent and
155 concatenation methods, we inferred a robust species tree and consistently resolved the
156 early divergence history of Papilionoideae. This nuclear phylogeny was subsequently used
157 as the backbone to reconstruct a 3,326-species time tree of Fabales to explore the evolution
158 of keel flowers under a broader taxonomic context in Papilionoideae, Cercidoideae, and
159 Polygalaceae. Through ancestral character reconstruction and diversification rate inference,
160 we evaluated the degree to which this rate variation could be attributed to keel flowers in
161 light of a key innovation hypothesis. We considered lack of support for this hypothesis if
162 the trait-dependent diversification model was not preferred and discussed our results in
163 the context of the ecology and biogeography of legume species to identify the processes
164 that account for their extreme diversity.
165
166 **MATERIALS AND METHODS**
167 **Taxon sampling**
168 Our taxon sampling included 279 species broadly distributed across Papilionoideae, plus
169 eight species from other legume subfamilies as outgroups (Tables S1–2). The dataset
170 comprised an aggregation of 47 newly generated transcriptomes and 32 low-pass genomic
171 sequences (Table S1) as well as published sequences from 208 species from Zhao *et al.*
172 (2021) to broaden taxon sampling (Table S2). Newly generated transcriptomes and
173 genomic sequences were mostly from morphologically and evolutionarily representative
174 genera within Papilionoideae (Cardoso *et al.*, 2012; Choi *et al.*, 2022). Collectively the

175 dataset represented 19 of the 22 well-supported clades consistently recognized in the
176 subfamily (Cardoso *et al.*, 2012; Cardoso *et al.*, 2013b; Choi *et al.*, 2022). The three clades
177 we did not include are *Amphimas*, *Aldina*, and the recently described *Cabari* (Gregório *et al.*,
178 2024). All newly generated sequences were derived from plant materials collected in the
179 field or grown in the University of Texas at Austin greenhouses (Table S1). Voucher
180 specimens were deposited in the herbaria of the Rio de Janeiro Botanic Garden (RB), State
181 University at Feira de Santana (HUEFS), and the Billie L. Turner Plant Resources Center
182 (TEX-LL). For DNA and RNA extraction, the methods described previously by Lee *et al.*
183 (2021) and Zhang *et al.* (2013) were employed, respectively.
184
185 **RNA sequencing and transcriptome assembly**
186 Total RNA samples were shipped to GENEWIZ (South Plainfield, NJ, USA) for ribosomal
187 RNA (rRNA) depletion, library preparation, and sequencing of approximately 25 million
188 paired-end reads on an Illumina HiSeq platform (Illumina, San Diego, CA). We assessed the
189 quality of raw 150 bp paired-end reads using FastQC v.0.11.5
190 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and removed filtered adapters and
191 low-quality reads using TrimGalore v.0.5.0 (https://github.com/FelixKrueger/TrimGalore).
192 The rRNA was removed from the merged trimmed reads using SortMeRNA v.2.1 (Kopylova
193 *et al.*, 2012) using eight rRNA databases for prokaryotes, archaea, and eukaryotes. Clean
194 RNA-Seq reads were assembled *de novo* using Trinity v.2.8.4 (Grabherr *et al.*, 2011), and
195 the assembled transcripts were subjected to the assessment of completeness and
196 redundancy using BUSCO v.4.1.0 (Benchmarking Universal Single-Copy Orthologs; Simão *et*
197 *al.*, 2015) with the eudicots_odb10.2019-11-20 database. The number of raw reads and
198 BUSCO evaluation results are provided in Table S1. All newly generated reads are available
199 on GenBank Sequence Read Archive under Bioproject PRJNA1120003
200 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1120003).
201
202 **Low-copy nuclear dataset assembly**
203 We used the 1,559 low-copy nuclear loci identified by Zhao *et al.* (2021) for our
204 phylogenetic reconstruction. To recover these loci in our transcriptomes, we used the
205 profile hidden Markov models implemented in HMMER v3.3.2 (Johnson *et al.,* 2010) to

206 identify orthologs. Briefly, a DNA profile database was built for each locus using the
207 function 'hmmbuild' based on the DNA alignments from Zhao *et al.* (2021). These profiles
208 were then searched against the target transcriptomes with an e-value threshold of $1e^{-50}$
209 using 'hmmsearch'. We then extracted the best hit from each species as the candidate
210 ortholog using a custom Python script (all scripts available on GitHub repository
211 https://github.com/lmcai/Papilionoideae_phylogenomics).
212
213 To recover the same set of loci from our genome sequence data, we developed a new
214 function of the PhyloHerb pipeline (Cai *et al.*, 2022) to automatically extract nuclear loci
215 from low-pass genome sequencing data. Briefly, PhyloHerb accomplishes this in three steps
216 — mapping, assembly, and scaffolding. First, it mapped reads to the reference sequence
217 using the very-sensitive mode of Bowtie v2.3.4.1 (Langmead & Salzberg, 2012). Mapped
218 reads were then assembled into contigs using SPAdes v3.8.0 under the default setting
219 (Bankevich *et al.*, 2012). The contigs were scaffolded into full-length loci based on syntenic
220 information compared to the reference using BLAST with an e-value threshold of $1e^{-40}$.
221 These functions were integrated as the '-m assemb' and '-m ortho -nuc' functions in
222 PhyloHerb v1.1. For details, see the GitHub page
223 (https://github.com/lmcai/PhyloHerb/#iv-retrieve-low-copy-nuclear-genes). Finally, for
224 quality control purposes, we also estimated the nuclear genome coverage of genome
225 sequences using Jellyfish v2.3.0 with a kmer size of 21 (Marçais & Kingsford, 2011).
226
227 **Sequence alignment, dataset cleaning, and phylogenetic inference**
228 Sequence alignment of each locus was conducted in MAFFT v7.245 (Katoh & Standley, 2013)
229 under the default settings. The aligned sequences were subjected to three rounds of
230 stringent cleaning to remove excessive missing data and potential paralogs. First, we used
231 the '-missing' function in PhyloHerb to remove sequences with >70% ambiguous sites
232 including gaps in each locus. Second, a phylogeny-based ortholog identification method
233 from Yang and Smith (2014) was used to remove potential paralogs. To accomplish this, a
234 maximum likelihood (ML) phylogeny was inferred for each gene using the DNA alignments
235 from the previous cleaning step. The best substitution model was automatically
236 determined by ModelFinder implemented in IQ-TREE v.2.2.2.6 (Minh *et al.*, 2020).

237 Branches that were more than fifteen times the mean branch length were pruned from the
238 resulting trees using a custom Python script 'prune_long_branches.py' (available on GitHub,
239 see link above). Then, a phylogeny-based ortholog identification approach using the Python
240 script 'prune_paralogs_RT.py' (Yang & Smith 2014) was applied to further remove potential
241 contaminants of paralogs. Third, all DNA sequences were realigned using the MAFFT E-INS-
242 i algorithm (--genafpair --maxiterate 1000) after removing spurious sequences identified in
243 the previous steps. Sites with more than 70% missing data were removed from the
244 alignment using the '-gt 0.7' flag in trimAL v1.5.0 (Capella-Gutiérrez *et al.*, 2009). We also
245 found that the published alignments from Zhao *et al.* (2021) were site-trimmed, thus
246 creating systematic biases for phylogenetic inference (e.g., shared gaps for sequences from
247 Zhao *et al.* (2021) compared to species sampled by us). To accommodate this, we used a
248 custom Python script to mask all sites in our sequences that were present in >60% of
249 species but absent in >90% of species from Zhao *et al.* (2021). Loci with less than 199 (70%)
250 taxa were removed from downstream phylogenomic assessments. In the end, 1,456 out of
251 the 1,559 loci passed our screening criteria (Fig. S1).
252
253 To estimate a species tree using the multispecies-coalescent model, we first inferred
254 individual gene trees for each of the 1,456 loci in IQ-TREE. The best substitution model was
255 automatically determined by ModelFinder implemented in IQ-TREE and branch support
256 was evaluated by 1000 ultrafast bootstrap replications. The species tree was inferred from
257 1,456 gene trees using ASTRAL-III v5.7.8 (Zhang *et al.*, 2017) under the default settings.
258 Meanwhile, to further mitigate the impact of artificially resolved gene tree branches with
259 low support values, we used newick-utils v1.6 (Junier & Zdobnov, 2010) to collapse gene
260 tree branches with ultrafast bootstrap support values less than 70. We then inferred a
261 species tree with these gene trees containing polytomies in ASTRAL-III.
262
263 To infer the species tree using the concatenation methods, we first further subsampled the
264 loci based on length and missing data. Any loci with alignments shorter than 900 base pairs
265 (bp; 33% loci removed) or less than 227 species (80% of the total species) were excluded
266 from the concatenation analysis, which resulted in 672 loci. The concatenated supermatrix
267 was generated by the '-m conc' function in PhyloHerb, along with a gene delineation file

input into PartitionFinder 2. The best partition scheme for ML analyses employing the general time reversible (GTR) + gamma model was selected by PartitionFinder 2 using the heuristic search algorithm 'rcluster' (Lanfear *et al.*, 2017). We then inferred an ML phylogeny of the 287 species using IQ-TREE with 100 traditional nonparametric bootstrap replications (BP) under the GTR + gamma nucleotide substitution model and linked but variable evolution rates for each partition with the '-p' flag.

Topology test

Several conflicting relationships emerged from the comparison between concatenation and coalescent analyses. To explore the origin of these conflicts and identify the best-supported phylogenetic scenarios, we conducted topology tests on six focal groups defined in Table S3. By alternating the position of the focal group and fixing the rest of the tree, we inferred an optimum phylogeny for competing scenarios in IQ-TREE. We then conducted the approximately unbiased (AU), Kishino-Hasegawa (KH), and Shimodaira-Hasegawa (SH) tests with 10,000 replicates as implemented in IQ-TREE. To compare the goodness-of-fit of alternative species trees to gene trees under the coalescent model, we conducted a pairwise likelihood ratio test (Liu *et al.*, 2019). We similarly optimized the branch length of fixed topologies with alternative positions of the focal group in MP-EST v3.0 (Liu *et al.*, 2010). We used the test2.sptree function from the R package Phybase v2.0 (Liu & Yu, 2010) to conduct the likelihood ratio test described in Liu *et al.* (2019). The likelihood ratio test statistic is given by $\tau = 2(L_{\mathrm{Tree2}} - L_{\mathrm{Tree1}})$, in which L_{Tree1} and L_{Tree2} are the log-likelihoods of the null and alternative hypotheses, respectively. The statistical significance (p-value) of the comparison is evaluated by generating a null distribution of τ via bootstrap sampling gene trees 100 times. Finally, to diagnose the causes of conflicting results, we examined the gene (gCF) and site concordance factors (sCF), as well as gene-wise likelihood scores implemented in IQ-TREE.

Bayesian divergence time estimation

To infer a Bayesian time tree, we subsampled the nuclear dataset to select the most clock-like genes using SortaDate (Smith *et al.*, 2018). Total tree length, root-to-tip variance, and topological congruence with the species tree were used as key indicators for phylogenetic

299 informativeness and stable molecular rate. We selected the 30 best genes (Table S4) for
300 dating to facilitate a reasonable running time of the Bayesian analysis, an approach
301 comparable to a recent study of Fabaceae divergence time (Koenen *et al.*, 2021). We
302 carefully selected eleven fossils to calibrate our molecular dating analyses, including four
303 fossils from representative eudicot lineages and seven well-recognized fossils within
304 Fabales (Table S5). In particular, the earliest fossil fruits and leaflets of Fabaceae from the
305 Early Paleocene (63.5 Ma) of the Denver Basin in Colorado, USA, were used to constrain the
306 stem group Fabaceae (Lyson *et al.*, 2019). Many recent fossils within Fabaceae were not
307 included because they provide redundant constraints compared to the eleven fossils we
308 selected. We used BEAST v.2.2.1 (Bouckaert *et al.*, 2014) to estimate divergence times
309 under the GTR substitution model, an uncorrelated relaxed molecular clock model, and a
310 birth-death tree prior. The topology was fixed to reflect the optimum species tree inferred
311 from the ASTRAL analysis. Fossil calibrations were set as exponential priors with the offset
312 minimum ages aligned with the fossils described in Table S5. The root node, for which we
313 used a normal prior at 128.6 Ma with a standard deviation of 1.0, truncated to minimum
314 and maximum ages of 113 Ma (the Aptian–Albian boundary) and 136 Ma (the oldest known
315 crown angiosperm fossil (Magallón *et al.*, 2015). We ran analyses under the uncorrelated
316 lognormal model for 100 million generations and confirmed convergence with Tracer
317 v1.7.1 (Rambaut *et al.*, 2018). The first 30% of the generations were discarded as burn-in
318 before summarizing median branch lengths and substitution rates with TreeAnnotator
319 from the BEAST package.
320

321 **Time tree inference for a 3,326-species *matK* dataset of Fabales**

322 To evaluate the macroevolutionary trends of floral diversification in Papilionoideae under a
323 broader taxonomic context, we generated a plastid *matK* gene dataset that included 3,106
324 species from 794 genera in Fabaceae (92% genus-level completeness), plus 220 additional
325 species from all other families in Fabales and outgroups. Sequences from Choi *et al.* (2022),
326 Legume Phylogeny Working Group (LPWG 2017), and Gregório *et al.* (2024) were
327 consolidated to generate the most comprehensive *matK* database for Fabaceae, and
328 combined with *matK* sequences from Polygalaceae, Surianaceae, and Quillajaceae available
329 on GenBank, which served as outgroups in Fabales (Table S6).

As a first step of data curation, we updated the nomenclature to reflect the most recent effort from POWO, Flora e Funga do Brasil (https://floradobrasil.jbrj.gov.br/consulta; BFG 2022), and taxonomic studies from individual clades (e.g., Hughes *et al.*, 2022). Second, we employed multiple rounds of phylogeny-guided cleaning for the *matK* sequence alignment. Initial alignment was conducted with the translation align option of MAFFT v.7.490 implemented in Geneious Prime 2023.1.2 (https://www.geneious.com/) and was manually inspected thereafter. For outgroup species, a separate alignment was generated using the L-INS-i algorithm in MAFFT to accommodate higher sequence divergence across families. These two alignments were then merged using the profile-profile alignment algorithm implemented in ClustalX (Thompson *et al.*, 2003) and the final DNA alignment contained 3,315 aligned sites. A draft ML phylogeny was inferred under the GTR+I+G model with 1000 ultrafast bootstrap replications in IQ-TREE and we filtered the *matK* matrix accordingly to remove misidentified and rogue taxa.

To infer a robust *matK* phylogeny for divergence time estimation, we used our well-supported nuclear phylogeny to constrain the topology and node ages of major clades in the *matK* dataset. This is necessary because the slow substitution rate of *matK* may result in contentious relationships and the lack of sequence divergence among close relatives may impact ancestral state estimation and divergence time estimation. As a result, relationships *between* the following clades were constrained: the ADA, Swartzieae, Cladrastis, Vataireoid, Exostyleae, Dalbergioid s.l., Genistoid s.l., Andira, Baphieae, and NPAAA. Relationships *within* each clade and *between* all subfamilies and families remained as-is. With this partially fixed topology, the ML phylogeny was inferred under the best fitting model in IQ-TREE.

Secondary calibration points of thirty nodes from the Bayesian time tree (Table S7) were used to infer the *matK* gene time tree under the penalized likelihood method implemented in TreePL v1.0 (Smith & O'Meara, 2012). These calibration points were imposed as both maximum and minimum age constraints to fix node ages. We then primed and cross-

validated our data to establish the best smoothing parameter (smooth = 10) for our dataset. Finally, we conducted a thorough search to infer the best time tree.

Testing state-dependent diversification analysis

To explore the diversification dynamics across Fabales, we performed a Bayesian analysis of macroevolutionary mixtures using the program BAMM v2.5.0 (Rabosky, 2014), which assumed clade- and time-dependent diversification scenarios to detect shifts in diversification rates. Reversible-jump Markov chain Monte Carlo was run for 20 million generations on the complete Fabales phylogeny and sampled every 1,000 generations. Posterior probabilities of rate shifts were summarized with the R package BAMMtools v.2.1.6 (Rabosky *et al.*, 2014) and visualized using a customized script from Padfield *et al.* (2024).

To explore whether the evolution of keel flowers drives pollinator-associated diversification in clades like Papilionoideae (Fabaceae) and Polygaleae (Polygalaceae), we estimated the divergence time history of Fabales under the state-dependent speciation and extinction (SSE) models. First, we coded the floral type of all species as either keel or non-keel flowers defined as above by Westerkamp (1997; Table S6). To infer the sampling ratio, we used the powoGenera function from the R package expowo (Zuanny *et al.*, 2024) to obtain the number of species for each genus based on the Plants of the World Online database (https://powo.science.kew.org; Table S8). The total species number for the keel flower clades and non-keel flower clades was enumerated to calculate the sampling ratio for these two groups. Finally, we inferred the diversification rate under the state-dependent and state-independent models using the R package hisse v2.1.8 (Beaulieu & O'Meara, 2016). We estimated rate parameters for four types of models (Table S9): (1) the Binary State Speciation and Extinction model (BiSSE), which allows state-dependent diversification (i.e., keel versus non-keel) but does not include hidden states; (2) the Hidden State Speciation and Extinction model (HiSSE), which allow both state-dependent diversification as well as hidden states; (3) CID-2 and CID-4 models, which are state-independent but allow differential diversification associated with the hidden states to serve as the null model for BiSSE and HiSSE, respectively; and (4) the trait-free, hidden

391 state-only rate model MiSSE for tip rate estimation (Maddison *et al.*, 2007; Beaulieu &
392 O'Meara, 2016; Vasconcelos *et al.*, 2022). For the HiSSE analyses, we included models with
393 the number of hidden states from two to four, which encompassed all combinations of
394 hidden states supported in the hisse package. To reconstruct the ancestral states of keel
395 flowers, we used the function 'hisse' followed by 'MarginReconHiSSE' to infer the
396 speciation and extinction rates as well as ancestral states under the standard HiSSE model
397 with two hidden states. For the MiSSE analyses, the optimum number of hidden states was
398 tested for all possible configurations ranging from 1 to 26. The best model with the lowest
399 AIC was selected for downstream analysis. We then applied the 'MarginReconMiSSE'
400 function to estimate the tip rates and evaluate the correlation between floral architecture
401 and diversification. We conducted a phylogenetic ANOVA test using the phylANOVA
402 function from the R package phytools (Revell, 2012) to compare tip rates between keel and
403 non-keel like species. Besides testing SSE model fitness using the entire Fabales phylogeny,
404 we also applied the BiSSE, HiSSE, CID-2, and CID-4 models to a subsampled phylogeny that
405 only included Papilionoideae species. Together, this resulted in a total of 11 different
406 analyses (Table S9).
407
408 **RESULTS**
409 **An efficient phylogenetic data assembly pipeline based on a curated set of loci**
410 BUSCO assessments for the assembled transcriptomes showed a high level of completeness,
411 with the percentage of complete orthologous genes ranging from 70.3% to 91.1% (Table
412 S1). The final multi-locus phylogenetic dataset included 1,456 genes with a mean length of
413 1,003 bp after trimming (Data S1). These loci on average contained 253 (85%) species and
414 the concatenated DNA dataset consisted of 941,354 sites with a gene and character
415 occupancy of 91.5% and 83.0%, respectively. None of our newly sequenced species
416 exhibited exceptional levels of missing data (<30% missing; Table S1), but three species
417 added from Zhao *et al.* (2021), which included *Pisum sativum* L., *Hesperothamnus*
418 *pentaphyllus* (Harms) Harms, and *Euchresta japonica* Hook. f. ex Regel, contained the
419 highest level of missing data (69–85% missing; Table S2).
420

421 Our comprehensive, genus-level taxon sampling represents multiple data collection
422 strategies, including high-quality reference genomes, RNA-seq, and low-pass genome
423 sequencing (i.e., genome skimming). Among these three approaches, our *in-silico* extraction
424 of phylogenetic markers from genomic sequencing data using PhyloHerb represents a cost-
425 effective approach to construct a multi-locus phylogeny, eliminating the need for target
426 enrichment library preparation. Genomic sequencing of the 32 Papilionoideae taxa yielded
427 25–30 million 150 bp paired-end reads and 1–15× nuclear coverage according to k-mer
428 distribution (Table S10; Data S2). In the final DNA matrix, the level of missing data in these
429 32 species was not statistically different from transcriptomic data (Welch two sample t-test
430 *p*-value = 0.08). Even for species with less than 2× nuclear genome coverage, at least 1,075
431 (74%) of the targeted loci were recovered (Table S10). Computationally, PhyloHerb is also
432 highly accessible. Extracting ~1,500 loci from each species on average requires 903.2s CPU
433 time and 2.9 GB peak memory including assembly and ortholog identification, which can be
434 completed on most laptops.
435
436 **Multi-locus nuclear dataset resolves contentious relationships in Papilionoideae**
437 To accommodate for incomplete lineage sorting, we used both coalescent and
438 concatenation methods for phylogenetic inference. The coalescent-based phylogeny
439 estimated by ASTRAL was largely congruent with the concatenation-based phylogeny
440 inferred by IQ-TREE and 92.6% and 89.1% of the nodes received maximum support (100
441 BP; 1.0 posterior probability, PP), respectively (Fig. 1).
442
443 Within Papilionoideae, the monophyly of 19 well-recognized clades was fully supported by
444 both coalescent and concatenation analyses (Fig. 1; Fig. S2). Relationships among these
445 clades were confidently resolved including those of recalcitrant lineages such as
446 *Dermatophyllum, Sesbania,* and the Andira clade (Fig. 1). The ADA, Swartzieae and
447 Cladrastis clades formed successive sister groups to all remaining members of
448 Papilionoideae — the 50-kb inversion clade (Fig. 1). Within the 50-kb inversion clade, the
449 Vataireoid and Exostyleae clades together formed a well-supported clade (100 BP; 1.0 PP)
450 sister to all other groups. The monophyly of the Vataireoid and Exostyleae clades was
451 strongly favored by both the AU test using the concatenated matrix and the likelihood ratio

452 test based on the multispecies coalescent (MSC) model. The alternative topology suggested
453 by Zhao *et al.* (2021), where these two groups were not monophyletic, was rejected (AU
454 test *p*-value < 0.01; MSC likelihood ratio test *p*-value <0.01; Table S3).
455
456 One major conflict in our results is the relative placement of the Dalbergioid s.l. (1,580 spp.)
457 and Genistoid s.l. (2,100 spp.) clades within the 50-kb inversion clade. A monophyletic
458 Dalbergioid s.l. + Genistoid s.l. clade was recovered but poorly supported by the
459 concatenation method (41 BP). On the other hand, the coalescent method supported the
460 monophyly of Dalbergioid s.l. plus the Andira, Baphieae, and NPAAA clades with high
461 confidence (0.98 PP; Fig. S2). The AU test on these two alternative topologies did not
462 provide definitive conclusions (*p*-value >0.05; Table S3) and the gene-wise likelihood
463 supporting each topology was approximately the same (Fig. S3). The MSC likelihood ratio
464 test, however, rejected the monophyletic Dalbergioid s.l. + Genistoid s.l. clade (*p*-value =
465 0.03). Within the Genistoid s.l. clade, tribes Ormosieae, Brongniartieae, and Leptolobieae
466 form successive sister groups to all remaining species (100 BP; 1.0 PP). *Dermatophyllum*
467 was confidently placed as sister to Brongniartieae (88 BP; 1.0 PP). The African
468 monospecific genus *Haplormosia* Harms was nested within the genistoids (98 BP; 0.90 PP).
469 Within the Dalbergioid s.l. clade, all previously identified major subclades were maximally
470 supported. Tribe Amorpheae appeared as sister to all remaining dalbergioids, and the
471 Adesmia, Dalbergia, and Pterocarpus subclades were phylogenetically successive sisters
472 (100 BP; >=0.99 PP).
473
474 The Andira, Baphieae, and NPAAA clades were each monophyletic and together formed a
475 grade comprising the remaining Papilionoideae. Within the NPAAA clade, the genus
476 *Hypocalyptus* Thunb. was sister to all other members of this clade (97 BP; 0.94 PP), while
477 the Mirbelioid, Hologalegina, and Indigofereae + Millettioid-Phaseoloid clades formed
478 successive sisters to Hypocalypteae. Within mirbelioids, both coalescent and concatenation
479 methods placed lineages from tribe Bossiaeeae as sister to a poorly supported clade
480 composed mostly of tribe Mirbelieae (<58 BP; <0.83 PP). Such low support could be
481 attributed to the two short internal branches (1.3–1.7e^{-3} in mutation units; 2.0–3.9e^{-2} in
482 coalescent units) separating Bossiaeeae, *Gompholobium* Sm, and the *Daviesia* Sm. +

483 *Viminaria* Sm. clade within Mirbelieae. These rapid radiations have rendered little
484 phylogenetic information to reconstruct this section of the phylogeny — the gene
485 concordant factor was 16.3% for the branch separating Bossiaeeae and Mirbelieae, while
486 14.0% of the genes supported the other two alternative quartet topologies at this node;
487 similarly, 16.0% of the genes supported the paraphyly of *Daviesia* + *Viminaria* and
488 *Gompholobium*, but one of the alternative topologies was supported by 17.4% of the genes
489 (Fig. 1; Data S3). Results from the site concordance factor largely reflected a similar trend
490 where each of the three quartet topologies was supported by 33% of the sites (Data S3).
491
492 The Millettioid-Phaseoloid clade within the larger NPAAA clade harbored more uncertainty.
493 The three genera *Aganope* Miq., *Austrosteenisia* R. Geesink and *Dalbergiella* Baker f. formed
494 a well-supported basal grade within the millettioids (100 BP; 1.0 PP). The concatenation
495 method confidently placed (100 BP) subtribe Clitoriinae — including *Centrosema* (DC.)
496 Benth., *Periandra* Mart. ex Benth. and *Clitoria* L. — as sister to *Dalbergiella*. In contrast, the
497 coalescent method placed Clitoriinae with a clade consisting of members from tribes
498 Desmodieae, and Psoraleeae, and the polyphyletic Phaseoleae with low confidence (0.31
499 PP). A clade comprising tribes Abreae, Diocleae, Millettieae formed the next diverging
500 group in the millettioids. The remaining millettioids included tribes Desmodieae,
501 Psoraleeae, and the polyphyletic Phaseoleae. In both coalescent and concatenation analyses,
502 *Platycyamus regnellii* Benth. and the *Butea* clade (*Spatholobus* Hassk., *Butea* Roxb. ex Willd.,
503 *Hardenbergia* Benth., and *Kennedia* Vent.) formed successive sister groups to all other
504 members within this clade (100 BP; 1.0 PP). Conflicts exist in terms of the placement of the
505 *Apios* clade (*Mucuna* Adans. and *Apios* Fabr.) and genus *Shuteria* Wight & Arn, which were
506 monophyletic (99 BP) in the concatenation analysis but paraphyletic (0.89 PP) in the
507 coalescent analysis.
508
509 **Evolution of keel flowers in Fabales**
510 The robust nuclear phylogeny was used to constrain the uncertain relationships among
511 major clades in the *matK*-based phylogeny. This broadly sampled phylogeny of Fabales
512 included 3,326 species — 3,106 Fabaceae (15.9% species diversity), 1 Quillajaceae (100%),
513 7 Surianaceae (55.6%), 206 Polygalaceae (15.4%), and 6 additional eudicots for dating

514 purposes. With fossil constraints, the time tree inferred by BEAST supported a Late
515 Cretaceous origin of crown group Fabales at 78.0 Ma (95% HPD 71.9–88.3 Ma; Fig. S4), and
516 an early Paleocene origin of crown group Fabaceae and Polygalaceae at 66.0 Ma (95% HPD
517 64.7–66.8 Ma) and 56.1 Ma (95% HPD 55.8–56.6 Ma), respectively. Within Fabaceae, we
518 inferred an explosive diversification leading to all six subfamilies within 7 million years
519 after the origin of the family. Crown group Papilionoideae was inferred to originate at 60.9
520 Ma (95% HPD 59.7–62.1 Ma; Fig. S4).
521
522 Keel flowers, which are mostly known in Papilionoideae and tribe Polygaleae
523 (Polygalaceae), have been independently lost and gained multiple times. Ancestral
524 character reconstruction suggested at least six independent origins of keel flowers in tribe
525 Polygaleae (Polygalaceae), *Cercis* L. (Cercidoideae), tribe Dipterygeae (ADA clade,
526 Papilionoideae), the MRCA of *Petaladenium* Ducke and *Dussia* Krug & Urb. ex Taub. (ADA
527 clade, Papilionoideae), *Myrospermum* Jacq. (ADA clade, Papilionoideae), and the MRCA of
528 the Cladrastis clade and the 50-kb inversion clade (Papilionoideae). In addition, there were
529 two independent losses of keel flowers in Polygalaceae and thirty in Papilionoideae
530 (summarized in Table S11; Fig. S5). These numbers, especially the losses, were strongly
531 dependent on the phylogeny. We inferred the ancestral floral architecture of
532 Papilionoideae to be non-keel-like, with the first keel flowers emerging in the common
533 ancestor of the Cladrastis and the 50-kb inversion clades around 59.0 Ma (95% HPD 57.9–
534 60.2 Ma). This is more recent than the origination of keel flowers in crown group
535 Polygalaceae, which included *Xanthophyllum* at 62.8 Ma (95% HPD 59.6–64.9 Ma), though
536 the exact order cannot be established with confidence due to overlapping confidence
537 intervals.
538
539 **State-dependent diversification**
540 The BAMM analysis inferred 18 rate shifts in Fabales and 11 of these shifts are nested
541 within Papilionoideae (Fig. 2a, c). The diversification rate increased gradually over time in
542 Fabales and nearly all rate shifts represented a significant acceleration in diversification
543 rates rather than a decrease (gray lines in Fig. 2b). With few exceptions, 9 out of the 11 rate
544 shifts in Papilionoideae took place during the Miocene. These Miocene rapid radiations

545 involved well known examples from the mimosoids, the genera *Lupinus* L. and *Astragalus* L.,
546 Psoraleeae, and Phaseolinae (clades A, H, K, L in Fig. 2a).
547
548 Although most rate shifts took place in the keel-flowered Papilionoideae, our
549 diversification rate analysis did not support the state-dependent models (Table S11). When
550 applying the state-dependent (BiSSE and HiSSE), the state-independent (CID-2 and CID-4),
551 and the completely trait-free models (MiSSE) to Fabales, MiSSE with 21 hidden states was
552 significantly better than the second-best CID-4 model, with a ΔAIC of 474.75. The
553 phylogenetic ANOVA analysis also confirmed no significant difference in the MiSSE tip rate
554 between keel and non-keel species (p-value = 0.71). This conclusion held when using a
555 subsampled phylogeny that included only Papilionoideae species. For these subsampled
556 analyses, only BiSSE, standard HiSSE with two hidden states, CID-2, and CID-4 were tested.
557 Among them, the state-independent CID-4 was significantly better than the second best
558 HiSSE model with a ΔAIC of 74.67. However, diversification parameters from both BiSSE
559 and HiSSE models suggested an orders-of-magnitude higher net diversification rate in keel-
560 flowered clades than non-keel-flowered clades. Specifically, HiSSE suggested the net
561 diversification rate to be 0.13 in keel-flowered clades and $5.3e^{-4}$ in non-keel-flowered clades;
562 BiSSE suggested it to be 0.082 in keel-flowered clades and $7.7e^{-4}$ in non-keel-flowered
563 clades.
564
565 **DISCUSSION**
566 **Insights into the early divergence relationships within papilionoids**
567 Previous phylogenetic studies focusing on the early divergences of papilionoid legumes
568 have largely utilized plastid markers (Wojciechowski *et al.*, 2004; Lavin *et al.*, 2005;
569 McMahon & Sanderson, 2006; Cardoso *et al.*, 2012, 2013b; Zhang *et al.*, 2020; Choi *et al.*,
570 2022). These studies have consistently recognized well-supported clades but struggled in
571 resolving inter-clade relationships. Based on 1,456 low-copy nuclear genes, we show that
572 the ADA, Swartzieae, and Cladrastis clades are successive sister groups to the 50-kb
573 inversion clade (Fig. 1). This result is congruent with the recent nuclear Fabaceae
574 phylogeny from Zhao *et al.* (2021) and a multi-locus phylogenetic study of the nitrogen-
575 fixing clade (Kates *et al.*, 2024) but differs from most plastid-based phylogenies where the

576 Swartzieae clade was placed as sister to all other Papilionoideae (Cardoso *et al.*, 2012;
577 Zhang *et al.*, 2020; Choi *et al.*, 2022). Our results also differ from the most recent
578 angiosperm phylogeny based on 353 nuclear genes where the ADA and Swartzieae clades
579 formed a monophyletic lineage sister to all other Papilionoideae (Zuntini *et al.*, 2024).
580
581 The 50-kb inversion clade constitutes 95% of the remaining species diversity in
582 Papilionoideae. Our well-resolved nuclear phylogeny provides several important insights
583 into relationships within this clade that were inconclusive in previous plastid sequence-
584 based studies. First, the Exostyleae and Vataireoid clades form a monophyletic group sister
585 to the remaining members of the 50-kb inversion clade (100 BP; 1.0 PP). Their close
586 relationship is morphologically supported by the denticulate, serrate, or occasionally
587 spinescent leaf margin, which is highly unusual outside the NPAAA clade in Papilionoideae
588 (Mansano *et al.*, 2004; Cardoso *et al.*, 2012, 2013a,b). Second, although conflicting
589 topologies are obtained when placing the Genistoid s.l. clade, statistically significant
590 support is only achieved for the coalescent analysis (Table S3). A monophyletic group
591 comprising the Genistoid s.l. and Dalbergioid s.l. clades was reported by Kates *et al*. (2024)
592 and Zuntini *et al.* (2024) but strongly rejected by our coalescent-based likelihood ratio test.
593 Instead, our ASTRAL phylogeny placed the Genistoid s.l. clade sister to a clade comprised of
594 the Dalbergioid s.l., Andira, Baphieae, and NPAAA clades (PP = 0.98). Within the latter clade,
595 the Dalbergioid s.l., Andira, and Baphieae clades form successive sister groups to the large
596 NPAAA clade — a relationship that has not been observed in any plastid sequence-based
597 study (Cardoso *et al.*, 2012, 2013b; Choi *et al.*, 2022).
598
599 Third, the genus *Dermatophyllum,* comprised of four North American species, has been a
600 clade recalcitrant to phylogenetic resolution in plastid-based investigations and often
601 formed a polytomy with the Dalbergioid s.l., Andira, and Genistoid s.l. clades
602 (Wojciechowski *et al.*, 2004; Cardoso *et al.*, 2012; Cardoso *et al.*, 2013b; Kates *et al.*, 2024).
603 We confidently placed *Dermatophyllum* as sister to tribe Brongniartieae within the
604 Genistoid s.l. clade (88 BP; 1.0 PP). This placement is supported by their shared niche
605 preference and the ability to synthesize a class of highly distinct alkaloids: *Dermatophyllum*
606 prefers warm temperate woodlands, chaparral, and deserts of southern United States to

central Mexico, which overlaps with many Mexican Brongniartieae especially in the genus *Brongniartia* Kunth (Pennington *et al.*, 2005; Ross & Crisp, 2005). The presence of quinolizidine alkaloids in *Dermatophyllum* further supports its placement within the Genistoid s.l. clade (Cardoso *et al.*, 2013b; Choi *et al.*, 2022). This class of alkaloids is absent in all other legumes and only sparsely distributed in eight other vascular plant families (Bisby *et al.*, 1994; Kite & Pennington, 2003; Lee *et al.*, 2013; Wink, 2013; Cely-Veloza *et al.*, 2023). We thus demonstrate a single origin of alkaloid biosynthesis in papilionoids and a chemical synapomorphy for the Genistoid s.l. clade.

Fourth, the close relationship between the African genus *Haplormosia* and the American-Australian Brongniartieae clade proposed by Cardoso *et al.* (2017) is challenged by our new result. Although several reproductive and vegetative features are shared between these two lineages — including unifoliolate leaves; single-seeded, tardily dehiscent fruits with thick, woody valves; and stamens with distinctly dimorphic anthers (Cardoso *et al.*, 2017) — the distinct biogeographic distribution makes the proposed relationship contentious. Our results instead place *Haplormosia* within a group of largely Old-World lineages, including *Camoensia* and the tribes Crotalarieae, Genisteae, Podalyrieae, and Sophoreae (98 BP; 0.90 PP). Morphologically, species from Crotalarieae, Podalyrieae, and Genistieae also exhibit unifoliolate leaves and dimorphic anthers that are alternately basifixed and dorsifixed (Schutte-Vlok & van Wyk, 2011; Le Roux & Van Wyk, 2012; du Preez *et al.*, 2021).

Finally, the NPAAA clade contains a whopping ca. 10,200 species that account for nearly 70% of species diversity in papilionoids. It is sister to the predominantly African Baphieae clade with ca. 60 species (100 BP; 1.0 PP). The NPAAA clade comprises two large, strongly supported subclades, the *Hologalegina* (Robinioid+IR-lacking clade) and Indigofereae + Millettioid-Phaseoloid clades, in agreement with all previous molecular phylogenetic studies (Wojciechowski *et al.*, 2004; Cardoso *et al.*, 2013b; LPWG, 2017; Zhao *et al.*, 2021). Within Hologalegina (ca. 5,300 spp.), the relationship between *Sesbania* and tribes Loteae and Robinieae was highly unstable with all combinations of relationships recovered using plastid data (Wojciechowski *et al.*, 2004; Farruggia *et al.*, 2018; Zhang *et al.*, 2020; Choi *et*

638 *al.*, 2022). Our results confirm *Sesbania* as sister to the warm temperate, primarily
639 northern hemisphere Loteae (100 BP; 1.0 PP), and together they form a sister group to
640 Robinieae (100 BP; 1.0 PP). The paraphyletic relationship among *Sesbania,* Loteae, and
641 Robinieae proposed by Zhao *et al.* (2021) is strongly rejected by our AU test (p-value < 0.01)
642 and MSC likelihood ratio test (p–value < 0.01). Relationships within the IR-lacking clade
643 (IRLC) align with recent studies using plastid and nuclear sequences (Choi *et al.*, 2022;
644 Zhao *et al.*, 2021). Relationships within the primarily subtropical to pantropical
645 Indigofereae + Millettioid-Phaseoloid clade (ca. 4,050 spp.) are consistent with earlier
646 findings (Hu *et al.*, 2000; Wojciechowski *et al.*, 2004; Stefanović *et al.*, 2009; LPWG, 2017;
647 Oyebanji *et al.*, 2020; Zhao *et al.*, 2021; Choi *et al.*, 2022).
648
649 **Loss and gain of keel flowers during the diversification of the papilionoids**
650 A suite of reproductive, nutritional, and defense traits has contributed to the assemblage of
651 present legume biodiversity (Lewis *et al.*, 2005). Here we focused on the floral form in
652 papilionoid legumes and more broadly in Fabales, which has been widely invoked to
653 explain the radiation of many angiosperms (Sargent, 2004; Kay & Sargent, 2009; Givnish *et*
654 *al.*, 2015). Bilaterally symmetrical flowers such as keel flowers are strongly associated with
655 bee pollination because the banner and wing petals can attract and orient the pollinators
656 while the keel petals restrict the entry to pollen (Neal *et al.*, 1998). The relative position of
657 the keel, pistil, and stamens also facilitates the precise placement of pollen to pollinators
658 and thus discourages gene flow between incipient species (Neal *et al.*, 1998). Despite being
659 a stereotyped search image for legumes, there is great lability in floral form and symmetry
660 during the early evolution of papilionoids (Fig. 2). This provides an ideal comparative
661 framework to test whether the evolution of keel flowers drives diversification (Ricklefs &
662 Renner, 1994, 2000). Extending our hypothesis testing to Fabales not only broadens the
663 investigation to another important keel lineage in Polygalaceae, but also offers
664 evolutionary context for this highly specialized floral form.
665
666 Our well-resolved nuclear gene and densely sampled plastid *matK* gene phylogenies
667 provide the essential evolutionary framework to address this question. Crown group
668 Fabaceae was estimated to originate at the K-Pg boundary at 66.0 Ma (95% HPD 64.7–66.8

669 Ma), which is in line with previous molecular-based estimates and fossil records (Lavin *et*
670 *al.*, 2005; Bruneau *et al.*, 2008; Cannon *et al.*, 2015; Koenen *et al.*, 2021; Zhao *et al.*, 2021).
671 The ancestral floral architecture of Papilionoideae is estimated to be non-keel-like, with the
672 first keel flowers in Fabaceae emerging in the common ancestor of the Cladrastis and the
673 50-kb inversion clades at 59.0 Ma (95% HPD 57.9–60.2 Ma). This estimate predates the
674 earliest known papilionoid-like flower fossil *Barnebyanthus buchananensis* Crepet &
675 Herendeen by 3 million years (Crepet & Herendeen, 1992) and closely succeeded the
676 evolution of keel flowers in crown group Polygalaceae at 62.8 Ma (95% HPD 59.6–64.9 Ma).
677 However, these estimates contradict the results of a recent study which concluded that
678 papilionoid keel flowers evolved 22.03 to 10.32 million years before they occurred in
679 crown group Polygalaceae (Uluer *et al.*, 2022).
680
681 Such a discrepancy comes mainly from the ancestral state reconstruction analysis of keel
682 flowers themselves, but also from minor differences in divergence time estimation due to
683 taxon sampling and methodology. Unlike the crown group Papilionoideae origination
684 proposed by Uluer *et al.* (2022), our results suggest a more dynamic evolutionary history of
685 keel flowers with four independent origins that postdate crown group Papilionoideae
686 (Table S11; Fig. S5). Three of these events are nested within the ADA clade (e.g.,
687 Dipterygeae and *Myrospermum*), taking place in the Neotropics during the Miocene. The
688 two oldest keel flower lineages originated during the Selandian age of the Paleocene in
689 Papilionoideae and Polygalaceae, though the exact order cannot be confidently established.
690 However, the short interval between these two old events and multiple more recent origins
691 of keel flowers in the ADA and polygalean clades may result from younger lineages
692 inheriting and further consolidating the strong relationship with pollinators already
693 established by older keel lineages (Tucker, 2002; Bello *et al.*, 2012; Uluer *et al.*, 2022).
694
695 In addition to gains, perhaps what is more intriguing is the numerous losses of keel flowers.
696 We identified more than two dozen independent losses in Papilionoideae alone (Table S10).
697 These results suggest that ancestral Fabaceae likely took a generalist pollination strategy,
698 allowing floral symmetry and petal number to be more labile (Arroyo, 1981; Bruneau *et al.*,
699 2014; Sinjushin, 2021). This is exemplified in multiple early branching papilionoids such as

Harleyodendron R.S. Cowan with rosid-like radial symmetry and lost, reduced, or poorly differentiated petals. These modifications are accompanied by expansion or shifts of pollinators to bats, passerine birds, hummingbirds, and even beetles (e.g., Arroyo, 1981; Pennington *et al.*, 2000; Lewis *et al.*, 2003). Such pronounced heterogeneity in floral architecture and pollination syndrome mirrors the early evolution of angiosperms and many clades in Gesneriaceae and Ranunculales (Damerval & Nadot, 2007; Martén-Rodríguez *et al.*, 2009; Endress, 2010). While up to 28 losses of keel flowers occurred in early branching papilionoids, only three were found within the NPAAA clade, which accounts for more than 70% of species diversity in Fabaceae. All three losses were restricted to the genus *Erythrina* L. (Millettioid-Phaseoloid clade), which is a pantropical genus of ca. 120 bird pollinated, mainly arborescent species (Bruneau, 1996, 1997). Species in *Erythrina* still retain the bilateral floral architecture but differentiate from a typical keel flower in that the stamens and style are exposed rather than protected. Their exceptionally long upper banner petal forms a pseudotubular structure enclosing the other petals, whereas the banner in a typical keel flower folds convexly upward into a boat-shape (Nesom, 2015).

The majority of the NPAAA clade, on the other hand, exhibited significant conservatism in floral form. We hypothesize this to be a joint outcome of ecological pressures imposed by pollinators and herbivores as well as developmental canalization (Bruneau *et al.*, 2014; Carvalho *et al.*, 2023; Ojeda *et al.*, 2019). Although most Fabaceae are bee-pollinated regardless of floral architecture, more reliable and specialized bee associations are developed with true keel flowers in Papilionoideae (Armbruster, 2014; Arroyo, 1981; Fenster *et al.*, 2004). The NPAAA clade has diversified largely since the Oligocene under global cooling, aridification, and expansion of grasslands (Estep *et al.*, 2014; Herbert *et al.*, 2016). The IRLC, Indigofereae, and Robinioid clades are thus strongly associated with herbaceous growth habit in arid to semi-arid temperate regions (Wojciechowski *et al.*, 2004; Schrire *et al.*, 2005, 2009). Thus adaptation to these habitats may cast additional geographical, environmental, and developmental constraints to limit the breadth of viable pollinators in the NPAAA clade. Moreover, empirical studies from various plant families have also shown that highly developed corollas may protect young buds and fruits from

731 herbivory (Carvalho *et al.*, 2023; Herrera, 2010; Gao *et al.*, 2019). Finally, developmental
732 canalization may also contribute to homeostasis in floral architecture (Mabberley & Hay,
733 1994; Leite *et al.*, 2015; Ojeda *et al.*, 2019). The modularized genetic programs for floral
734 development can produce consistent phenotypes buffered against internal and external
735 influences (i.e., canalization; Moyroud & Glover, 2017). Many ontogenetic modules for keel
736 flowers are shared between Polygalaceae and Fabaceae (Westerkamp & Weber, 1999;
737 Prenner, 2004; Bello *et al.*, 2010) and in many cases demonstrate increased canalization
738 among recently diverged angiosperms lineages (Endress, 2011). The conservation of keel
739 flower developmental pathways, including the MADS-BOX and CYC genes for floral
740 symmetry and whorl identity (Singh *et al.*, 2013; Zhang *et al.,* 2010), may further contribute
741 to the prevalence of keel flowers in derived Papilionoideae.
742
743 **A synergistic multidriver hypothesis for the diversification of Papilionoideae**
744 Evolutionary biologists have been constantly searching for key innovations or
745 environmental stimulants of angiosperm diversity (see reviews from Helmstetter *et al.*,
746 2023; Sauquet & Magallón, 2018). In particular, the direct involvement of floral traits in
747 sexual reproduction has prompted great interest in exploring its phenotypic diversity and
748 link with species diversification (Kay *et al.*, 2006). Zygomorphy (Sargent, 2004; Kay &
749 Sargent, 2009), animal pollination (Muchhala, 2006), and nectar spurs (Hodges and Arnold,
750 1995) have all been suggested to enhance speciation rates. However, a growing body of
751 empirical studies suggested that angiosperm diversification cannot be tied to a single key
752 innovation or major global event (Augusto *et al.*, 2014; Vamosi *et al.*, 2018). Instead,
753 various combinations of traits, environment, and ecology may have acted to stimulate
754 diversification in different groups (Augusto *et al.,* 2014; Magallón & Castillo, 2009). Such
755 synergistic interaction between multiple, perhaps sequential, intrinsic and extrinsic
756 contributors of diversification rate has been coined as 'synnovation' by Donoghue and
757 Sanderson (2015).
758
759 In papilionoid legumes, the contrast in species numbers between clades with and without
760 keel flowers motivated our investigation of this phenomenon using comparative methods.
761 The best-supported MiSSE model from our results was trait-free and thus indicated the

762 limited influence of floral architecture on diversification at a global scale. This lack of
763 support for SSE models is likely a conservative conclusion given our taxon sampling (ca. 15%
764 across Fabales) because larger, older, and less well-sampled phylogenies tended to support
765 trait-dependent models (Helmstetter *et al.*, 2023). The orders-of-magnitude difference in
766 net diversification rate estimated by the HiSSE model was driven by a handful keel-
767 flowered lineages including the well-characterized radiations in *Lupinus* and *Astragalus*
768 (Fig. 2 a, d). Yet even in these two genera, additional extrinsic and intrinsic factors such as
769 perennial life history and invasion of montane ecosystems have been demonstrated to
770 promote speciation (Drummond *et al.*, 2012; Azani *et al.*, 2019). Similarly, a lack of
771 correlation between nodulation ability and net diversification rate was reported by
772 Afkhami *et al.* (2018) despite that nodulating legume genera contain three times more
773 species than non-nodulating ones. These results suggest that individual traits such as keel
774 flowers and nodulation are not the sole driver of diversification in papilionoids, but rather
775 constitute a broader confluence of factors that have additively contributed to lineage
776 diversification (Donoghue & Sanderson, 2015).
777
778 Indeed, the massive diversity of Fabaceae has been hypothesized to be an outcome of a
779 suite of intrinsic traits including nitrogen fixation symbiotically with soil bacteria (Sprent,
780 2001; Werner *et al.*, 2015; Doyle, 2016; Faria *et al.*, 2022), direct chemical defense with
781 alkaloids and canavanine (Bell, 1981; Bisby *et al.*, 1994; Kursar *et al.*, 2009; Wink, 2013),
782 indirect defense via ant domatia and extrafloral nectaries (Marazzi & Sanderson, 2010;
783 Chomicki & Renner, 2015; Marazzi *et al.*, 2019), and likely variation in fruit morphology
784 and seed dispersal mechanisms that have yet to be characterized (López *et al.*, 2000;
785 Herrera *et al.*, 2019). Besides, recurrent biome shifts and rapid colonization of novel
786 ecological niches also played a significant role in promoting diversification in Fabaceae.
787 Examples include the rapid radiation of the primarily herbaceous IRLC clade in temperate
788 regions (Wojciechowski *et al.*, 2004), the dominance of mimosoids in seasonally dry forests
789 (Pennington *et al.*, 2009; Ringelberg *et al.*, 2023), and the remarkable diversification of
790 Indigofereae in tropical fire-prone savannas (Schrire *et al.,* 2009). Finally, multiple whole
791 genome duplications taking place during early evolution of Fabaceae have also been
792 proposed to buffer mass extinction events and promote genetic and species diversity

793 (Koenen *et al.*, 2021; Zhao *et al.,* 2021). Some of these events left a clear signal in our BAMM
794 analysis (e.g., rate shift in mimosoids; node A Fig. 2a) that are uncorrelated with keel
795 morphology. In genera like *Senna, Lupinus*, and *Astragalus*, intrinsic (e.g., extrafloral
796 nectaries, perennial habit) and extrinsic factors (e.g., colonizing Andes) are closely linked
797 with each other to promote diversification (Azani *et al.*, 2019; Marazzi and Sanderson,
798 2010; Drummond *et al.*, 2012). Studies of Podalyrieae in the Cape region also concluded
799 that the interaction between fire-survival strategy and soil type preference simultaneously
800 contributed to its diversity (Schnitzler *et al.*, 2011). All of this evidence points to the
801 synnovation hypothesis as a more plausible explanation for radiations in Fabaceae.
802
803 In summary, while floral architecture may have played a role in the diversification of
804 papilionoid legumes, a combination of intrinsic factors and extrinsic factors were
805 synergistically involved to drive the remarkable diversification of Papilionoideae. This may
806 begin with the accumulation of genetic variation from the paleo genome duplication that
807 occurred along the stem lineage leading to papilionoids (Koenen *et al.*, 2021), combined
808 with the origin of nodulation within the clade (Doyle, 2016) and the shifts in floral
809 architecture to bilateral symmetry. Keel flowers may facilitate speciation by enhancing the
810 persistence of species, which might in turn increased the likelihood of experiencing
811 subsequent key innovation or range shifts that promote speciation. Alternatively, keel
812 flowers may represent a phylogenetic inertia for species-rich clades rather than being the
813 primary or casual factor per se of that drove diversification (Blomberg & Garland, 2002).
814
815 **CONCLUSIONS**
816 Drawing upon insights from phylogenomics, divergence time estimation, and comparative
817 methods, we present novel insights into the early divergence history and floral evolution of
818 papilionoid legumes. With 1,456 low-copy nuclear loci, we confidently resolved the
819 recalcitrant relationships in Papilionoideae and inferred a Paleocene origin of keel flowers
820 at 59 Ma, predating the earliest fossil record by 3–4 million years. We identified at least six
821 independent origins and thirty-two losses of keel flowers across Fabales, with losses
822 predominantly taking place in the early diversification history of the papilionoids.
823 Ancestral Fabaceae thus likely displayed a more generalist pollination strategy, allowing

for greater floral flexibility over time. Although these independent gains and losses coincide with the Miocene diversification of Papilionoideae, our results do not support floral architecture as the sole driver of diversification in Fabaceae. Instead, a confluence of intrinsic and extrinsic factors including nodulation, chemical defenses, and recurrent dispersals into new biomes such as subtropical and temperate regions, likely contributed synergistically to the diversification of papilionoids.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Science Foundation to MFW (DEB-1853010) and RKJ, TAR (DEB-1853024), the Texas Ecological Laboratory Program (EcoLab) to RKJ, TAR, I-SC, and LC, the Sidney F. and Doris Blake Professorship in Systematic Botany to RKJ, the Stengl Wyer Fellowship to LC, the CNPq (Research Productivity Fellowship no. 314187/2021-9), and FAPERJ (Programa Jovem Cientista do Nosso Estado - 2022, grant no. E-26/200.153/2023) to DC, the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00341022) to I-SC, and funds from the College of Liberal Arts and Sciences at Arizona State University to MFW. We thank the TEX-LL, HUEFS, and RB herbaria for voucher deposition, the Desert Legume Program at the University of Arizona for seeds; Janet Davis for kindly allowing us to use her photo of pollinating *Baptisia australis* in Fig. 2; Alessandra Schnadelbach, Henrique Batalha Filho, and Paula Ristow for permitting us to use their labs at UFBA; and George Yatskievych (TEX/LL) for arranging a formal Material Transfer Agreement (Decree number 8772) under the SisGen Cadastro RDC6BE9, which facilitated research activities between the institutions. Comments from the Handling Editor Susana Magallón, three anonymous reviewers, and Michael J. Sanderson have greatly improved the initial manuscript.

FIGURE LEGEND

Figure 1. Phylogeny of Papilionoideae based on low-copy nuclear gene sequences and inferred from the multispecies coalescent model implemented in ASTRAL-III. Branch supports indicate posterior probability from ASTRAL and bootstrap replicates support from IQ-TREE, respectively. Only branches that are not maximally supported are labeled with conflicting topologies marked by '–'. Pie charts on each node represent gene concordance factors for the proportion of concordant (blue), discordant (green and pink), and noninformative (grey) genes. The phylogram in the left margin illustrates branch

length in mutational units. The photos (by Domingos Cardoso, except *Pickeringia* by Martin Wojciechowski, *Grazielodendron* by Claudio Nicoletti, *Baphia* by Warren McCleland (iNaturalist, https://www.inaturalist.org/photos/108404418), *Hypocalyptus* by Brian du Preez (iNaturalist, https://www.inaturalist.org/photos/60072646), *Mirbelia* by Nina Kerr (iNaturalist, https://www.inaturalist.org/photos/345474574), *Pisum* by Michela Papa (iNaturalist, https://inaturalist.ca/photos/372960709), and *Psoralea* by katrivier (iNaturalist, https://inaturalist.ca/photos/262747553) illustrate the remarkable diversity of floral architecture within the Papilionoideae.

Figure 2. Rate heterogeneity in the diversification of Fabales. (a) Time-calibrated tree of Fabales species based on plastid *matK* gene sequences. Branch colors show net diversification rates inferred from BAMM. Circles at nodes refer to clades with significant rate shifts, which approximately align with the mimosoids (A), *Swartzia*+*Bocoa*+*Candolleodendron* (B), Genisteae (C), *Lupinus* (D), NPAAA clade (E), IR-lacking clade (F), Vicioid clade (G), *Astragalus* (H), Diocleae (I), Cajaninae+Desmodieae+Phaseolinae+Erythrininae+Psoraleeae in addition to several other smaller lineages (J), Psoraleeae (K), and Phaseolinae (L). Purple lines in the inner most ring indicate the presence of keel flowers. The middle ring represents net diversification rates at the tip estimated from the MiSSE model with 21 hidden states. Warmer colors are suggestive of higher rates. Fabaceae taxonomic groups including Cercidoideae, Detarioideae, Caesalpinioideae, and Papilionoideae are identified in the outer ring. (b) Rate-through-time plot showing increased net diversification over time. The black line represents the average across the whole tree; the gray lines represent the rate through time on parts of the tree where significant rate shifts were identified. Shaded regions represent 95% confidence intervals. (c) Posterior distribution of the number of rate shifts inferred using BAMM. (d) Variation in tip-specific diversification rates for keel-flowered and non-keel-flowered lineages based on BAMM estimate. (e) MiSSE-based tip diversification rates for keel-flowered and non-keel-flowered lineages. All rates are relative and only allow us to look at relative differences between biome preferences and parts of the phylogeny. The photo by Janet Davis shows a *Bombus griseocollis* (De Geer, 1773) bee on a *Baptisia australis* (L.) R.Br. flower.

SUPPLEMENTARY INFORMATION

Supplementary figures

Figure S1 Histogram of the number of species per loci. The vertical line represents the 30% missing data threshold to select the 1,456 loci in this study.

Figure S2 Topological congruences and conflicts of Papilionoideae phylogeny inferred from IQTREE (left) and ASTRAL (right). Tangled grey lines indicate conflicting relationships. Branch supports represent nonparametric bootstrap replicates values from

897 IQ-TREE and local posterior probability from ASTRAL, respectively. Only branches that are
898 not maximally supported are labeled.
899
900 **Figure S3** Quantification of phylogenetic signal from individual genes. The two alternative
901 placements of the focal clades are illustrated with red or blue, with the red topology being
902 favored by the concatenation analysis. Genes are ordered in descending order based on
903 their support for the red topology. Differences in the log-likelihood scores (ΔlogL) between
904 these two topologies are calculated for individual genes (left) or the concatenated matrix
905 (right). Panels A-E represent five contentious relationships within Papilionoideae defined
906 in Table S3.
907
908 **Figure S4** Bayesian inference of divergence time in Fabaceae based on 30 top clock-like
909 loci. Divergence time is shown in millions of years ago (Ma) and the blue bars represent 95%
910 credibility intervals. Eleven fossil calibration points including a constraint for the root age
911 are marked by numbered stars. See Table S5 for a full list of fossils and references.
912
913 **Figure S5** Ancestral character reconstruction of keel flowers in Fabales inferred under the
914 marginal reconstruction algorithm in HiSSE. An ultrametric tree of 3,326 Fabales species is
915 inferred from *matK* sequences with branch lengths drawn in proportion to divergence time.
916 Solid and hollow branches indicate the presence and absence of keel flowers, respectively.
917 Events of keel flower gains and losses are marked by orange and magenta circles on the
918 branch, respectively. See Table S11 for age estimations of these events.
919
920 **Supplementary Tables**
921 **Table S1** Voucher information and gene assembly information of 79 newly sequenced
922 species in this study. Species name, lineage affiliation, assembly quality, SRA accession
923 number, and voucher information including collection locality are provided.
924
925 **Table S2** Species name, lineage affiliation, and number of sampled genes from 208 species
926 from Zhao *et al.* (2021).
927

Table S3 Testing alternative phylogenetic placement of Papilionoideae clades under the maximum likelihood framework for concatenated DNA sequences.

Table S4 List of top 30 clock-like genes with quantified root-to-tip branch length variation, total tree length, and topological similarity to the species tree measured by bipartition score.

Table S5 Fossil calibration points for divergence time estimation.

Table S6 Time tree, GenBank ID, and keel flower trait status for the 3,326 Fabales species from the *matK* dataset. Numbers on the phylogeny are divergence time estimation for each node based on treePL.

Table S7 Secondary calibration points for the treePL analysis. All node ages in million years ago (Ma) are derived from the BEAST time tree and are fixed in treePL. Nodes are defined by the most recent common ancestor of two species indicated in the spreadsheet.

Table S8 Genus level sampling ratio for Fabaceae. The number of species for each genus was obtained using the powoGenera function from the R package expowo based on the Plants of the World Online database (https://powo.science.kew.org). The species sampling percentage was summarized for each genus and presented along with the URL address from which the POWO data was derived.

Table S9 Taxon sampling and model parameters for the 11 diversification rate estimation analyses. We used two sets of time-calibrated phylogenies that include either all Fabales or Papilionoideae species only. The BiSSE, HiSSE, CID-2, CID-4, and MiSSE models variously allow for the rate parameters between keel/non-keel states and hidden states to differ. BiSSE, HiSSE, CID-2, and CID-4 models further varied in the number of hidden states between zero and four. The optimum number of MiSSE was determined using the MiSSEGreedy() function based on AIC scores.

Table S10 Nuclear genome sequencing coverage of 32 whole genome sequencing datasets estimated based on 21-mer frequency distribution. See Table S1 for the species name of each ID.

Table S11 Summary of independent origins and losses of keel flower in Fabales. Binary ancestral state reconstruction is inferred under the marginal reconstruction algorithm in HiSSE. The timing of these events is determined by the crown group age of involved clades. The precise age of events involving a single species cannot be determined using available data. Therefore, the age of these events is represented by a range from present to their stem group age.

Supplementary Data

Data S1 DNA alignments and maximum likelihood phylogenies inferred for the 1,456 low-copy nuclear genes used in this study. The alignments are inferred using the MAFFT-linsi algorithm. The gene trees are reconstructed using IQTREE and branch support is evaluated with 1000 ultrafast bootstrap replication. Species names and their corresponding ID are provided in sp_ID.csv. The list of 672 loci used to generate the concatenated matrix for species tree estimation is provided in G672_len900_sp227.list. Data is available at the Figshare Digital Repository (DOI: 10.6084/m9.figshare.27135984).

Data S2 Distribution of 21-mer frequency in 32 species with low-pass genome sequencing data. The 21-mer frequency is inferred from Jellyfish v2.3.0 to estimate sequencing coverage.

Data S3 Gene and site concordance factors of the Papilionoideae phylogeny. The species tree and concatenated alignment of the 672-gene dataset were used as inputs in IQ-TREE to calculate these concordance factors.

COMPETING INTERESTS

None declared.

AUTHOR CONTRIBUTIONS

Conception and experimental design, LC, DC, I-SC, RKJ, TAR, MFW. Acquisition of funds, RKJ, TAR, MFW. Field collections, I-SC, DC, HCL, LPQ, RKJ, TAR, MFW. Nucleic acid isolation, I-SC, CL, TAR. RNA isolations, TAR. Transcriptome and genome assembly, CL, LT, BS, LC. DNA alignments and phylogenetic analyses, LC, I-SC, DC, MFW. Data interpretation, LC, I-SC, DC, MFW. Production of figures and tables, LC, DC. Writing the initial draft of the manuscript, LC, DC, MFW, TAR, RKJ. All authors revised and approved the final version of the manuscript.

DATA AVAILABILITY

The raw sequence data generated and analyzed during this study are available in the NCBI GenBank repository under Bioproject PRJNA1120003. The alignment data supporting the findings of this study are available as supplementary data and have been deposited in the Figshare Digital Repository (DOI: 10.6084/m9.figshare.27135984). The multi-locus phylogeny of Papilionoideae and the dated *matK* phylogeny of Fabales are available on the Open Tree of Life (https://tree.opentreeoflife.org/curator/study/view/ot_2372). The code used for data processing and analysis is openly available on GitHub at https://github.com/lmcai/Papilionoideae_phylogenomics.

REFERENCES

Afkhami ME, Luke Mahler D, Burns JH, Weber MG, Wojciechowski MF, Sprent J, Strauss SY. 2018. Symbioses with nitrogen-fixing bacteria: nodulation and phylogenetic data across legume genera. *Ecology* **99**: 502–502.

Armbruster WS. 2014. Floral specialization and angiosperm diversity: phenotypic divergence, fitness trade-offs and realized pollination accuracy. *AoB PLANTS* **6**: plu003.

Arroyo MTK. 1981. Breeding systems and pollination biology. In: R. M. Polhill and P. H. Raven, Ed., Advances in Legume Systematics, Part 2. Richmond, UK: Royal Botanic Garden, Kew, 723–769.

Augusto L, Davies TJ, Delzon S, De Schrijver A. 2014. The enigma of the rise of angiosperms: can we untie the knot? *Ecology Letters* **17**: 1326–1338.

Azani N, Bruneau A, Wojciechowski MF, Zarre S. 2019. Miocene climate change as a driving force for multiple origins of annual species in *Astragalus* (Fabaceae, Papilionoideae). *Molecular Phylogenetics and Evolution* **137**: 210–221.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, *et al.* 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**: 455–477.

Barrett CF, Specht CD, Leebens-Mack J, Stevenson DW, Zomlefer WB, Davis JI. 2014. Resolving ancient radiations: can complete plastid gene sets elucidate deep relationships among the tropical gingers (Zingiberales)? *Annals of Botany* **113**: 119.

Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. *Systematic Biology* **65**: 583–601.

Bell EA. 1981. Non-protein amino acids in the Leguminosae. In: Advances in Legume Systematics, Part 2, eds Polhill R. M., Raven P. H. Richmond, UK: Royal Botanic Gardens, Kew, 489–499.

Bello MA, Hawkins JA, Rudall PJ. 2010. Floral ontogeny in Polygalaceae and its bearing on the homologies of keeled flowers in Fabales. *International Journal of Plant Sciences* **171**: 482–498.

Bello MA, Rudall PJ, Hawkins JA. 2012. Combined phylogenetic analyses reveal interfamilial relationships and patterns of floral evolution in the eudicot order Fabales. *Cladistics* **28**: 393–421.

Bisby FA, Buckingham J, harborne JB. 1994. *Phytochemical Dictionary of the Leguminosae*. London: Chapman and Hall.

Blomberg SP, Garland T Jr. 2002. Tempo and mode in evolution: phylogenetic inertia, adaptation and comparative methods. *Journal of Evolutionary Biology* **15**: 899–910.

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu C-H, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **10**: e1003537.

Bruneau A. 1996. Phylogenetic and biogeographical patterns in *Erythrina* (Leguminosae: Phaseoleae) as inferred from morphological and chloroplast DNA characters. *Systematic Botany* **21**: 587–605.

Bruneau A. 1997. Evolution and homology of bird pollination syndromes in *Erythrina* (Leguminosae). *American Journal of Botany* **84**: 54–71.

Bruneau A, Klitgaard BB, Prenner G, Fougère-Danezan M, Tucker SC. 2014. Floral Evolution in the Detarieae (Leguminosae): phylogenetic evidence for labile floral

development in an early-diverging legume lineage. *International Journal of Plant Sciences* **175**: 392–417.

Bruneau A, Mercure M, Lewis GP, Herendeen PS. 2008. Phylogenetic patterns and diversification in the caesalpinioid legumes. *Botany* **86**: 697–718.

Cai L, Zhang H, Davis CC. 2022. PhyloHerb: a high-throughput phylogenomic pipeline for processing genome skimming data. *Applications in Plant Sciences* **10**: e11475.

Cannon SB, McKain MR, Harkess A, Nelson MN, Dash S, Deyholos MK, Peng Y, Joyce B, Stewart CN, Rolf M. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* **32**: 193–210.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972–1973.

Cardoso D, Harris DJ, Wieringa JJ, São-Mateus WMB, Batalha-Filho H, Torke BM, Prenner G, Queiroz LP. 2017. A molecular-dated phylogeny and biogeography of the monotypic legume genus *Haplormosia*, a missing African branch of the otherwise American-Australian Brongniartieae clade. *Molecular Phylogenetics and Evolution* **107**: 431–442.

Cardoso D, Queiroz LP, Lima HC, Suganuma E, van den Berg C, Lavin M. 2013a. A molecular phylogeny of the vataireoid legumes underscores floral evolvability that is general to many early-branching papilionoid lineages. *American Journal of Botany* **100**: 403–421.

Cardoso D, Pennington RT, Queiroz LP, Boatwright JS, Van Wyk B-E, Wojciechowski MF, Lavin M. 2013b. Reconstructing the deep-branching relationships of the papilionoid legumes. *South African Journal of Botany* **89**: 58–75.

Cardoso D, Queiroz LP, Pennington RT, Lima HC, Fonty É, Wojciechowski MF, Lavin M. 2012. Revisiting the phylogeny of papilionoid legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* **99**: 1991–2013.

Cardoso D, São-Mateus WMB, Cruz DT, Zartman CE, Komura DL, Kite G, Prenner G, Wieringa JJ, Clark A, Lewis G, *et al.* 2015. Filling in the gaps of the papilionoid legume phylogeny: the enigmatic Amazonian genus *Petaladenium* is a new branch of the early-diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* **84**: 112–124.

Carvalho CS, Lima HC, Lemes MR, Zartman CE, van den Berg C, García-Dávila CR, Honorio Coronado EN, Mader M, Paredes-Villanueva K, Tysklind N, *et al.* 2023. A dated phylogeny of the Neotropical Dipterygeae clade reveals 30 million years of winged papilionate floral conservatism in the otherwise florally labile early-branching papilionoid legumes. *Botanical Journal of the Linnean Society* **202**: 449–475.

Cely-Veloza W, Kato MJ, Coy-Barrera E. 2023. Quinolizidine-type alkaloids: chemodiversity, occurrence, and bioactivity. *ACS Omega* **8**: 27862–27893.

Choi I-S, Cardoso D, Queiroz LP, Lima HC, Lee C, Ruhlman TA, Jansen RK, Wojciechowski MF. 2022. Highly resolved papilionoid legume phylogeny based on plastid phylogenomics. *Frontiers in Plant Science* **13**: 823190.

Chomicki G, Renner SS. 2015. Phylogenetics and molecular clocks reveal the repeated evolution of ant-plants after the late Miocene in Africa and the early Miocene in Australasia and the Neotropics. *New Phytologist* **207**: 411–424.

Crepet WL, Herendeen PS. 1992. Papilionoid flowers from the Early Eocene of Southeastern North America. In: Advances in Legume Systematics, Part 4, eds. Herendeen P. S., Dilcher D. L. Richmond, UK: Royal Botanic Gardens, Kew, 43–55.

Damerval C, Nadot S. 2007. Evolution of perianth and stamen characteristics with respect to floral symmetry in Ranunculales. *Annals of Botany* **100**: 631–640.

Darwin CR. 1862. On the various contrivances by which British and foreign orchids are fertilized by insects, and on the good effects of intercrossing. London: John Murray.

Donoghue MJ, Sanderson MJ. 2015. Confluence, synnovation, and depauperons in plant diversification. *New Phytologist* **207**: 260–274.

Doyle JJ. 2016. Chasing unicorns: nodulation origins and the paradox of novelty. *American Journal of Botany* **103**: 1865–1868.

Drummond CS, Eastwood RJ, Miotto STS, Hughes CE. 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovations with incomplete taxon sampling. *Systematic Biology* **61**: 443–460.

Endress PK. 2010. The evolution of floral biology in basal angiosperms. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**: 411–421.

Endress PK. 2011. Evolutionary diversification of the flowers in angiosperms. *American Journal of Botany* **98**: 370–396.

Estep MC, McKain MR, Diaz DV, Zhong J, Hodge JG, Hodkinson TR, Layton DJ, Malcomber ST, Pasquet R, Kellogg EA. 2014. Allopolyploidy, diversification, and the Miocene grassland expansion. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 15149–15154.

Etcheverry AV, Alemán MM, Figueroa-Fleming T, López-Spahr D, Gómez CA, Yáñez C, Figueroa-Castro DM, Ortega-Baes P. 2012. Pollen:ovule ratio and its relationship with other floral traits in Papilionoideae (Leguminosae): an evaluation with Argentine species. *Plant Biology* **14**: 171–178.

Faria SM, Ringelberg JJ, Gross E, Koenen EJM, Cardoso D, Ametsitsi GKD, Akomatey J, Maluk M, Tak N, Gehlot HS, *et al.* 2022. The innovation of the symbiosome has enhanced the evolutionary stability of nitrogen fixation in legumes. *New Phytologist* **235**: 2365–2377.

Farruggia FT, Lavin M, Wojciechowski MF. 2018. Phylogenetic systematics and biogeography of the pantropical genus *Sesbania* (Leguminosae). *Systematic Botany* **43**: 414–429.

Fenster CB, Armbruster WS, Wilson P, Dudash MR, Thomson JD. 2004. Pollination syndromes and floral specialization. *Annual Review of Ecology, Evolution, and Systematics* **35**: 375–403.

Gao Y, Wang C, Song B, Du F. 2019. Corolla retention after pollination facilitates the development of fertilized ovules in *Fritillaria delavayi* (Liliaceae). *Scientific Reports* **9**: 729.

Givnish TJ. 2010. Ecology of plant speciation. *Taxon* **59**: 1326–1366.

Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJ, Clements MA, Arroyo MT, Leebens-Mack J. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**: 20151553.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29**: 644.

Grant V. 1949. Pollination systems as isolating mechanisms in angiosperms. *Evolution* **3**: 82–97.

Gregório BS, Carvalho CS, Ramos G, Rocha L, Stirton CH, Lima HC, Zartman CE, Lewis GP, Torke BM, Snak C, et al. 2024. A molecular phylogeny of the early-branching Genistoid lineages of papilionoid legumes reveals a new Amazonian genus segregated from *Clathrotropis*. *Botanical Journal of the Linnean Society* **205**: 1–14.

Helmstetter AJ, Zenil-Ferguson R, Sauquet H, Otto SP, Méndez M, Vallejo-Marin M, Schönenberger J, Burgarella C, Anderson B, de Boer H, Glémin S, Käfer J. 2023. Trait-dependent diversification in angiosperms: patterns, models and data. *Ecology Letters* **26**: 640–657.

Herbert TD, Lawrence KT, Tzanova A, Peterson LC, Caballero-Gill R, Kelly CS. 2016. Late Miocene global cooling and the rise of modern ecosystems. *Nature Geoscience* **9**: 843–847.

Herendeen PS, Cardoso DBOS, Herrera F, Wing SL. 2021. Fossil papilionoids of the Bowdichia clade (Leguminosae) from the Paleogene of North America. *American Journal of Botany* **109**: 130–150.

Herrera CM. 2010. Marcescent corollas as functional structures: effects on the fecundity of two insect-pollinated plants. *Annals of botany* **106**: 659–662.

Herrera F, Carvalho MR, Wing SL, Jaramillo C, Herendeen PS. 2019. Middle to Late Paleocene Leguminosae fruits and leaves from Colombia. *Australian Systematic Botany* **32**: 385–408.

Hodges SA, Arnold ML. 1995. Spurring plant diversification: are floral nectar spurs a key innovation? *Proceedings of the Royal Society of London Series B* **262**: 343–348.

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

Hughes C, Eastwood R. 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences of the United States of America* **103**: 10334–10339.

Hughes CE, Queiroz LP, Lewis GP. 2022. Advances in Legume Systematics 14. Classification of Caesalpinioideae Part 1: new generic delimitations. *PhytoKeys* **205**: 1–2.

Johnson LS, Eddy SR, Portugaly E. 2010. Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics* **11**: 431.

Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* **26**: 1669–1670.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kay KM, Sargent RD. 2009. The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. *Annual Review of Ecology, Evolution, and Systematics* **40**: 637–656.

Kite GC, Pennington RT. 2003. Quinolizidine alkaloid status of *Styphnolobium* and *Cladrastis* (Leguminosae). *Biochemical Systematics and Ecology* **31**: 1409–1416.

Koenen EJM, Ojeda DI, Bakker FT, Wieringa JJ, Kidner C, Hardy OJ, Pennington RT, Herendeen PS, Bruneau A, Hughes CE. 2021. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous–Paleogene (K–Pg) mass extinction event. *Systematic Biology* **70**: 508–526.

Kopylova E, Noé L, Touzet H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**: 3211–3217.

Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG, Murakami ET, Drake C, McGregor R, Coley PD. 2009. The evolution of antiherbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 18073–18078.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B. 2017. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* **34**: 772–773.

Landis JB, Soltis DE, Li Z, Marx HE, Barker MS, Tank DC, Soltis PS. 2018. Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany* **105:** 348–363.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lavin M, Herendeen PS, Wojciechowski MF. 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Systematic Biology* **54**: 575–594.

Le Roux MM, Van Wyk BE. 2012. The systematic value of flower structure in *Crotalaria* and related genera of the tribe Crotalarieae (Fabaceae). *Flora* **207**: 414–426.

Lee C, Choi I-S, Cardoso D, Lima HC, Queiroz LP, Wojciechowski MF, Jansen RK, Ruhlman TA. 2021. The chicken or the egg? Plastome evolution and an independent loss of the inverted repeat in papilionoid legumes. *The Plant Journal* **107**: 861–875.

Lee ST, Cook D, Molyneux RJ, Davis TZ, Gardner DR. 2013. Alkaloid profiles of *Dermatophyllum arizonicum*, *Dermatophyllum gypsophilum*, *Dermatophyllum secundiflorum*, *Styphnolobium affine*, and *Styphnolobium japonicum* previously classified as *Sophora* species. *Biochemical Systematics and Ecology* **49**: 87–93.

Leite VG, Teixeira SP, Mansano VF, Prenner G. 2015. Floral development of the early-branching papilionoid legume *Amburana cearensis* (Leguminosae) reveals rare and novel characters. *International Journal of Plant Sciences* **176**: 94–106.

Leppik EE. 1966. Floral evolution and pollination in the Leguminosae. *Annales Botanici Fennici* **3**: 299–308.

Lewis GP, Knudsen JT, Klitgaard BB, Pennington RT. 2003. The floral scent of *Cyathostegia mathewsii* (Leguminosae, Papilionoideae) and preliminary observations on reproductive biology. *Biochemical Systematics and Ecology* **31**: 951–962.

Lewis G, Schrire B, Mackinder B, Lock M. 2005. *Legumes of the World*. Richmond, UK: Royal Botanic Gardens, Kew.

Liu L, Anderson C, Pearl D, Edwards SV. 2019. Modern phylogenomics: building phylogenetic trees using the multispecies coalescent model. In: Evolutionary Genomics. Springer, 211–239.

Liu L, Yu L. 2010. Phybase: an R package for species tree analysis. *Bioinformatics* **26**: 962–963.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evolutionary Biology* **10**: 302.

López J, Devesa JA, Ortega-Olivencia A, Ruiz T. 2000. Production and morphology of fruit and seeds in Genisteae (Fabaceae) of south-west Spain. *Botanical Journal of the Linnean Society* **132**: 97–120.

LPWG [Legume Phylogeny Working Group]. 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species–rich clades. *Taxon* **62**: 217–248.

LPWG [Legume Phylogeny Working Group]. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon* **66**: 44–77.

Lyson TR, Miller IM, Bercovici AD, Weissenburger K, Fuentes AJ, Clyde WC, Hagadorn JW, Butrim MJ, Johnson KR, Fleming RF, *et al.* 2019. Exceptional continental record of biotic recovery after the Cretaceous–Paleogene mass extinction. *Science* **366**: 977–983.

Mabberley DJ, Hay A. 1994. Homoeosis, canalization, decanalization, 'characters' and angiosperm origins. *Edinburgh Journal of Botany* **51**: 117–126.

Maddison WP, Midford PE, Otto SP. 2007. Estimating a binary character's effect on speciation and extinction. *Systematic Biology* **56**: 701–710.

Magallón S, Castillo A. 2009. Angiosperm diversification through time. *American journal of botany* **96**: 349–365.

Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

Mansano VF, Bittrich V, Azevedo Tozzi AMG, Souza AP. 2004. Composition of the Lecointea clade (Leguminosae, Papilionoideae, Swartzieae), a re-evaluation based on combined evidence from morphology and molecular data. *Taxon* **53**: 1007–1018.

Marazzi B, Gonzalez AM, Delgado-Salinas A, Luckow MA, Ringelberg JJ, Hughes CE. 2019. Extrafloral nectaries in Leguminosae: phylogenetic distribution, morphological diversity and evolution. *Australian Systematic Botany* **32**: 409–458.

Marazzi B, Sanderson MJ. 2010. Large-scale patterns of diversification in the widespread legume genus *Senna* and the evolutionary role of extrafloral nectaries. *Evolution* **64**: 3570–3592.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**: 764–770.

Maron JL, Agrawal AA, Schemske DW. 2019. Plant–herbivore coevolution and plant speciation. *Ecology* **100**: e02704.

Martén-Rodríguez S, Almarales-Castro A, Fenster CB. 2009. Evaluation of pollination syndromes in Antillean Gesneriaceae: evidence for bat, hummingbird and generalized flowers. *Journal of Ecology* **97**: 348–359.

McMahon MM, Sanderson MJ. 2006. Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Systematic Biology* **55**: 818–836.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF, Timoneda A, Yim WC, Cushman JC, Yang Y. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Systematic Biology* **70**: 219–235.

Moyroud E, Glover BJ. 2017. The evolution of diverse floral morphologies. *Current Biology* **27**: R941–R951.

Muchhala N. 2006. The pollination biology of *Burmeistera* (Campanulaceae): specialization and syndromes. *American Journal of Botany* **93**: 1081–1089.

Neal PR, Dafni A, Giurfa M. 1998. Floral symmetry and its role in plant-pollinator systems: terminology, distribution, and hypotheses. *Annual Review of Ecology, Evolution, and Systematics* **29**: 345–373.

Neill DA. 1987. Trapliners in the trees: hummingbird pollination of *Erythrina* Sect. *Erythrina* (Leguminosae: Papilionoideae). *Annals of the Missouri Botanical Garden* **74**: 27–41.

Nesom GL. 2015. Key to native and cultivated species of *Erythrina* (Fabaceae) in the USA and comments on naturalization of *Erythrina crista-galli*. *Phytoneuron* **29:** 1–8.

Ojeda DI, Koenen E, Cervantes S, de la Estrella M, Banguera-Hinestroza E, Janssens SB, Migliore J, Demenou BB, Bruneau A, Forest F, *et al.* 2019. Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. *Molecular Phylogenetics and Evolution* **137**: 156–167.

Oyebanji O, Zhang R, Chen S-Y, Yi T-S. 2020. New insights into the plastome evolution of the Millettioid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Frontiers in Plant Science* **11: 151**.

Padfield D, Kay S, Vos R, Quince C, Vos M. 2024. Macroevolutionary dynamics in micro-organisms: generalists give rise to specialists across biomes in the ubiquitous bacterial phylum Myxococcota. *Molecular Biology and Evolution* **41**: msae088.

Pennington R, Klitgaard B, Ireland H, Lavin M. 2000. New insights into floral evolution of basal Papilionoideae from molecular phylogenies. In: P.S. Herendeen and A. Bruneau (Eds). Advances in Legume Systematics 9. Royal Botanic Gardens, Kew. In: 233–248.

Pennington RT, Lavin M, Oliveira-Filho A. 2009. Woody plant diversity, evolution, and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics* **40**: 437–457.

Pennington RT, Stirton CH, Schrire BD. 2005. Tribe Sophoreae. In: Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), Legumes of the World. In: Richmond, UK: Royal Botanic Gardens, Kew, 227–249.

Polhill RM, Raven PH, Stirton CH. 1981. Evolution and systematics of the Leguminosae. In: Polhill, R.M. and Raven, P.H., Eds., Advances in Legume Systematics, Part I. Richmond, UK: Royal Botanic Garden, Kew, 1–26.

Preez B, Dreyer LL, Stirton CH, Muasya AM. 2021. A monograph of the genus *Polhillia* (Genisteae: Fabaceae). *South African Journal of Botany* **138**: 156–183.

Prenner G. 2004. Floral development in *Polygala myrtifolia* (Polygalaceae) and its similarities with Leguminosae. *Plant Systematics and Evolution* **249**: 67–76.

Prenner G, Klitgaard BB. 2008. Towards unlocking the deep nodes of Leguminosae: floral development and morphology of the enigmatic *Duparquetia orchidacea* (Leguminosae, Caesalpinioideae). *American Journal of Botany* **95**: 1349–1365.

Rabosky DL. 2014. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLOS ONE* **9**: e89543.

Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG. 2014. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* **5**: 701–707.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67**: 901–904.

Ree RH. 2005. Phylogeny and the evolution of floral diversity in *Pedicularis* (Orobanchaceae). *International Journal of Plant Sciences* **166**: 595–613.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**: 217–223.

Ricklefs RE, Renner SS. 1994. Species richness within families of flowering plants. *Evolution* **48**: 1619–1636.

Ricklefs RE, Renner SS. 2000. Evolutionary flexibility and flowering plant familial diversity: a comment on Dodd, Silvertown, and Chase. *Evolution* **54**: 1061–1065.

Ringelberg JJ, Koenen EJM, Sauter B, Aebli A, Rando JG, Iganci JR, Queiroz LP, Murphy DJ, Gaudeul M, Bruneau A, et al. 2023. Precipitation is the main axis of tropical plant phylogenetic turnover across space and time. *Science Advances* **9**: eade4954.

Rosas-Guerrero V, Aguilar R, Martén-Rodríguez S, Ashworth L, Lopezaraiza-Mikel M, Bastida JM, Quesada M. 2014. A quantitative review of pollination syndromes: do floral traits predict effective pollinators? *Ecology Letters* **17**: 388–400.

Ross JH, Crisp MD. 2005. Brongniartieae. In: Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), Legumes of the World. In: Richmond, UK: Royal Botanic Gardens, Kew, 52–259.

Sargent RD. 2004. Floral symmetry affects speciation rates in angiosperms. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**: 603–608.

Sauquet H, Magallón S. 2018. Key questions and challenges in angiosperm macroevolution. *New Phytologist* **219**: 1170–1187.

Schnitzler J, Barraclough TG, Boatwright JS, Goldblatt P, Manning JC, Powell MP, Rebelo T, Savolainen V. 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology* **60**: 343–357.

Schrire BD. 2005a. Tribe Millettieae. In: Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), Legumes of the World. In: Richmond, UK: Royal Botanic Gardens, Kew, 367–387.

Schrire BD. 2005b. Tribe Phaseoleae. In: Lewis, G., Schrire, B., Mackinder, B., Lock, M. (Eds.), Legumes of the World. In: Richmond, UK: Royal Botanic Gardens, Kew, 393–431.

Schrire BD, Lavin M, Barker NP, Forest F. 2009. Phylogeny of the tribe Indigofereae (Leguminosae–Papilionoideae): geographically structured more in succulent-rich and temperate settings than in grass-rich environments. *American Journal of Botany* **96**: 816–852.

Schrire BD, Lavin M, Lewis G. 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biologiske Skrifter* **55**: 375–422.

Schutte-Vlok AL, van Wyk B-E. 2011. A taxonomic revision of *Podalyria* (Fabaceae). *Systematic Botany* **36**: 631–660.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Sinjushin AA. 2021. Evolutionary history of the leguminous flower. *Biology Bulletin Reviews* **11**: 400–413.

Singh VK, Garg R, Jain M. 2013. A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnology Journal* **11**: 691–701.

Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. *PLOS ONE* **13**: e0197433.

Smith SA, O'Meara BC. 2012. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Sprent JI. 2001. *Nodulation in legumes*. London: Royal Botanic Gardens, Kew.

Spriggs EL, Clement WL, Sweeney PW, Madriñán S, Edwards EJ, Donoghue MJ. 2015. Temperate radiations and dying embers of a tropical past: the diversification of *Viburnum*. *New Phytologist* **207**: 340-354.

Stebbins GL. 1970. Adaptive radiation of reproductive characteristics in angiosperms, I: Pollination mechanisms. *Annual Review of Ecology and Systematics* **1**: 307–326.

Stefanović S, Pfeil BE, Palmer JD, Doyle JJ. 2009. Relationships among Phaseoloid legumes based on sequences from eight chloroplast regions. *Systematic Botany* **34**: 115–128.

Stewart AB, Diller C, Dudash MR, Fenster CB. 2022. Pollination-precision hypothesis: support from native honey bees and nectar bats. *New Phytologist* **235**: 1629–1640.

Thompson JD, Gibson TJ, Higgins DG. 2003. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* **2.3**: 2.3.1–2.3.22.

Torke BM, Schaal BA. 2008. Molecular phylogenetics of the species-rich neotropical genus *Swartzia* (Leguminosae, Papilionoideae) and related genera of the Swartzioid clade. *American Journal of Botany* **95**: 215–228.

Tucker SC. 1999. Evolutionary lability of symmetry in early floral development. *International Journal of Plant Sciences* **160**: S25–S39.

Tucker SC. 2002. Floral ontogeny of *Cercis* (Leguminosae: Caesalpinioideae: Cercideae): does it show convergence with Papilionoids? *International Journal of Plant Sciences* **163**: 75–87.

Tucker SC. 2003. Floral development in legumes. *Plant Physiology* **131**: 911–926.

Uluer DA, Forest F, Armbruster S, Hawkins JA. 2022. Reconstructing an historical pollination syndrome: keel flowers. *BMC Ecology and Evolution* **22**: 45.

Vamosi JC, Magallón S, Mayrose I, Otto SP, Sauquet H. 2018. Macroevolutionary patterns of flowering plant speciation and extinction. *Annual Review of Plant Biology* **69**: 685–706.

Van der Niet T, Peakall R, Johnson SD. 2014. Pollinator-driven ecological speciation in plants: new evidence and future perspectives. *Annals of Botany* **113**: 199–212.

Vargas P, Zardoya R. 2014. *The Tree of Life: Evolution and Classification of Living Organisms*. Sunderland MA: Sinauer Associates.

Vasconcelos T, O'Meara BC, Beaulieu JM. 2022. A flexible method for estimating tip diversification rates across a range of speciation and extinction scenarios. *Evolution* **76**: 1420–1433.

Werner GDA, Cornwell WK, Cornelissen JHC, Kiers ET. 2015. Evolutionary signals of symbiotic persistence in the legume–rhizobia mutualism. *Proceedings of the National Academy of Sciences of the United States of America* **112**: 10262–10269.

Westerkamp C. 1997. Keel blossoms: bee flowers with adaptations against bees. *Flora* **192**: 125–132.

Wink M. 2013. Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany* **89**: 164–175.

Wojciechowski MF, Lavin M, Sanderson MJ. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *American Journal of Botany* **91**: 1846–1862.

Yang Y, Smith SA. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* **31**: 3081–3092.

Yoder JB, Gomez G, Carlson CJ. 2020. Zygomorphic flowers have fewer potential pollinator species. *Biology Letters* **16**: 20200307.

Zhang J, Ruhlman TA, Mower JP, Jansen RK. 2013. Comparative analyses of two Geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biology* **13**: 228.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

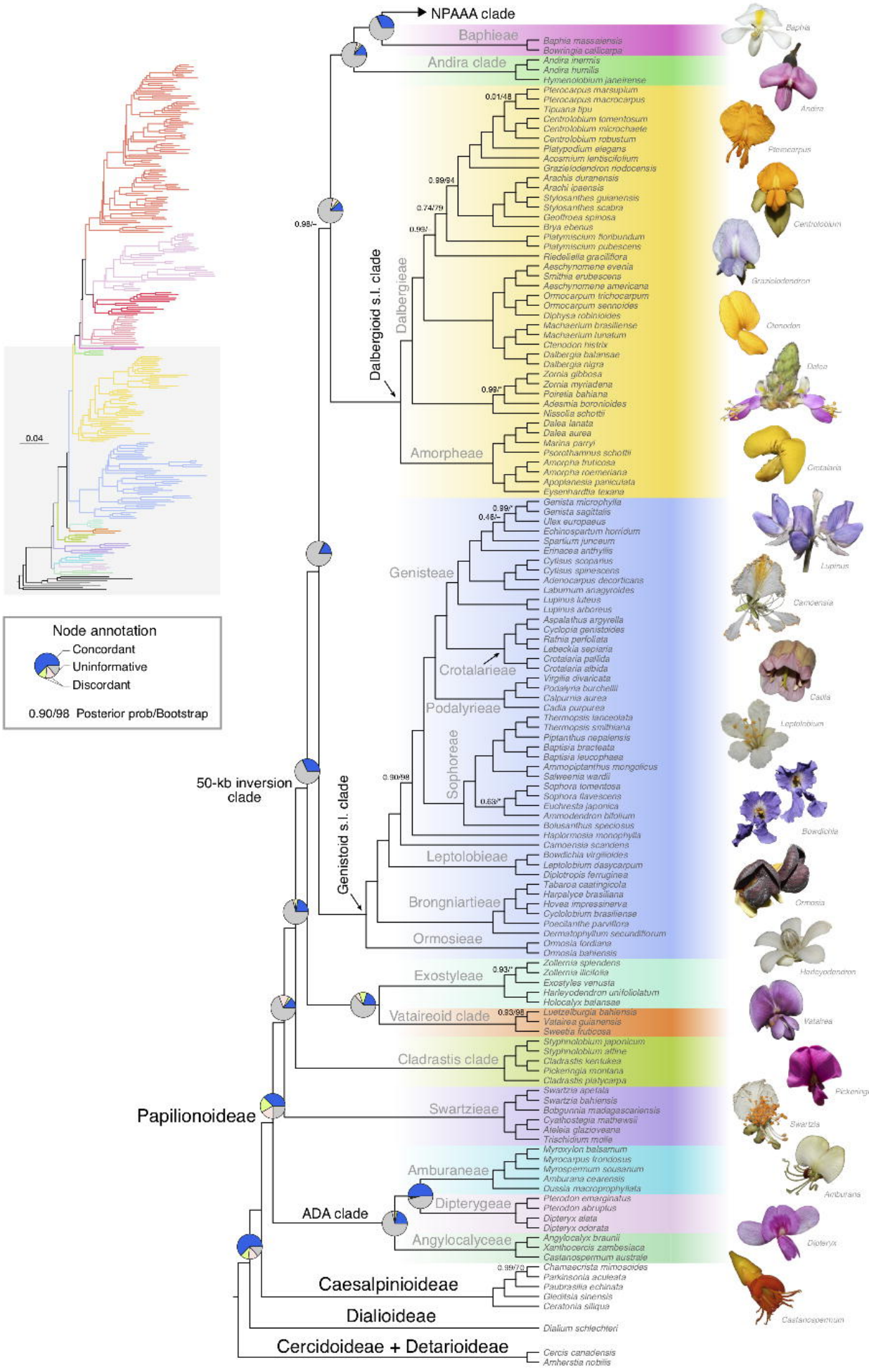
Zhang R, Wang Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, Queiroz LP, Moore MJ, Zhang SD, Chen SY. 2020. Exploration of plastid phylogenomic conflict yields new insights into the deep relationships of Leguminosae. *Systematic Biology* **69**: 613–622.

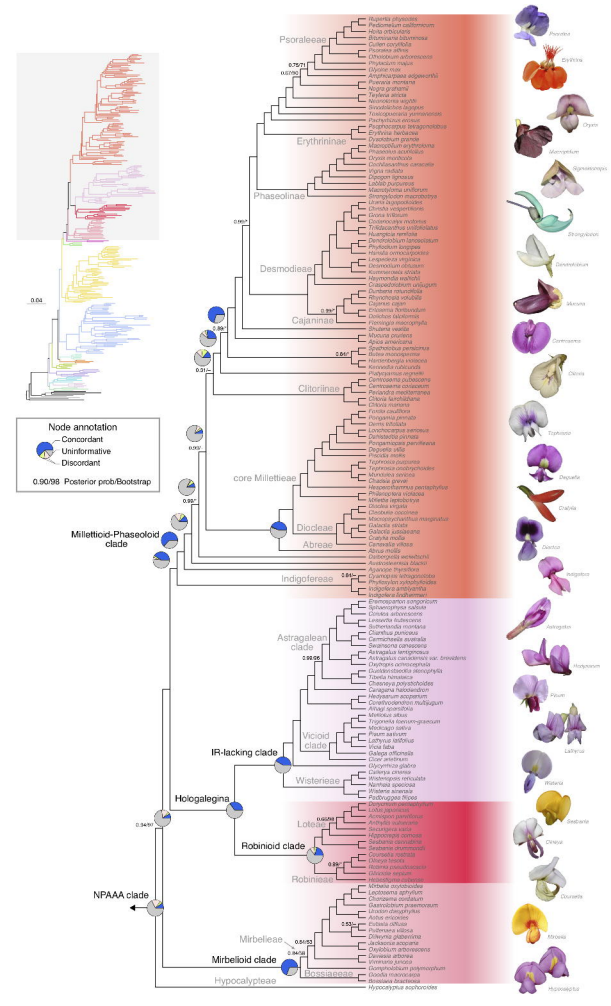
Zhang W, Kramer EM, Davis CC. 2010. Floral symmetry genes and the origin and maintenance of zygomorphy in a plant-pollinator mutualism. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 6388–6393.

Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi T-S, *et al.* 2021. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* **14**: 748–773.

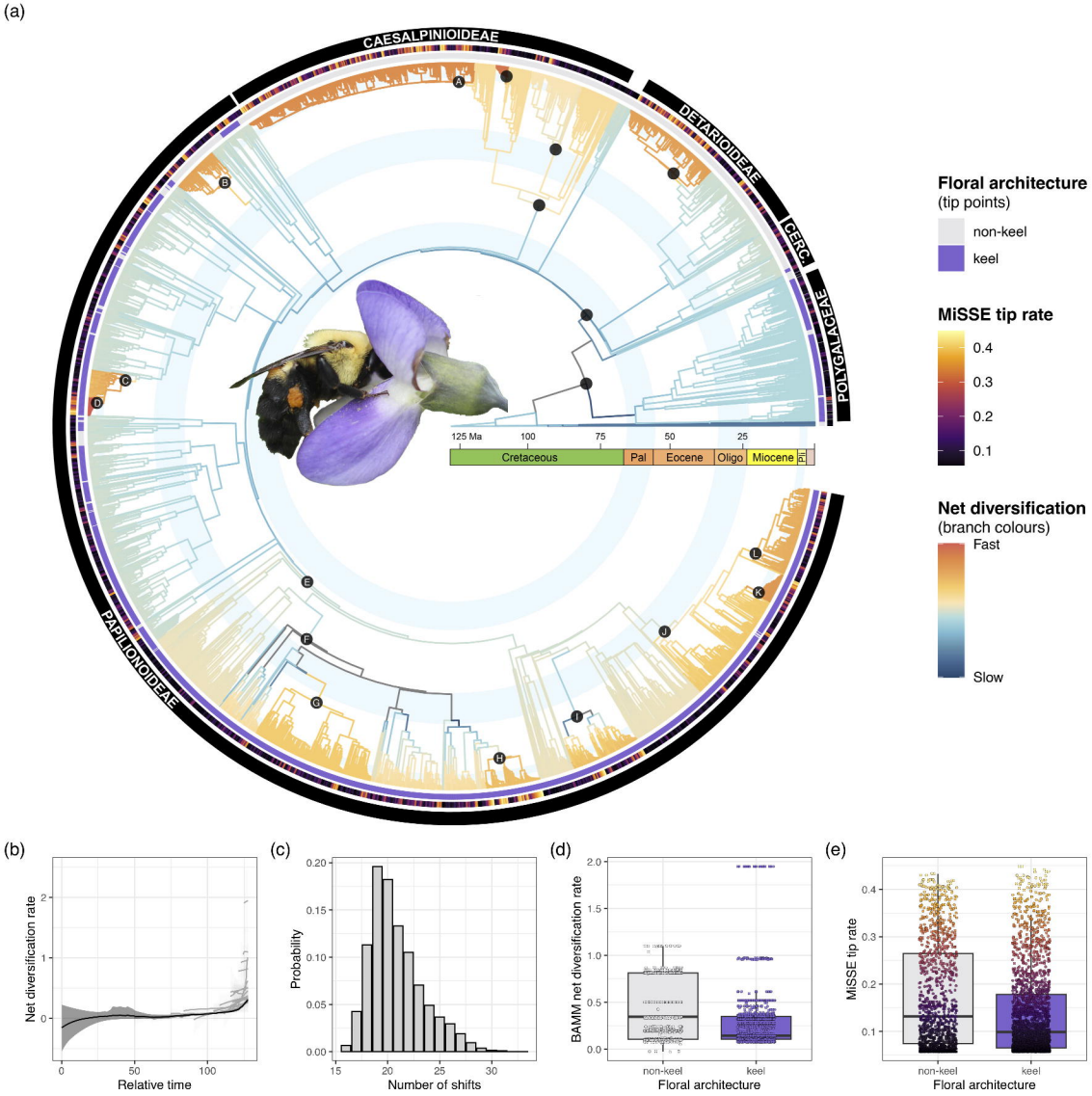
Zuanny DC, Vilela B, Moonlight PW, Särkinen TE, Cardoso D. 2024. expowo: an R package for mining global plant diversity and distribution data. *Applications in Plant Sciences* **12**: e11609.

Zuntini AR, Carruthers T, Maurin O, Bailey PC, Leempoel K, Brewer GE, Epitawalage N, Françoso E, Gallego-Paramo B, McGinnie C, *et al.* 2024. Phylogenomics and the rise of the angiosperms. *Nature* **629**: 843–850.


NPAAA clade
Baphieae
Baphia massaiensis
Bowringia callicarpa
Andira clade
Andira inermis
Andira humilis
Hymenolobium janeirense
Pterocarpus marsupium
Pterocarpus macrocarpus
Tipuana tipu
Centrolobium tomentosum
Centrolobium microchaete
Centrolobium robustum
Platypodium elegans
Acosmium lentiscifolium
Grazielodendron nodocensis
Arachis duranensis
Arachis ipaensis
Stylosanthes guianensis
Stylosanthes scabra
Geoffroea spinosa
Brya ebenus
Platymiscium floribundum
Platymiscium pubescens
Riedeliella graciliflora
Aeschynomene evenia
Smithia erubescens
Aeschynomene americana
Ormocarpum trichocarpum
Ormocarpum sennoides
Diphysa robinioides
Machaerium brasiliense
Machaerium lunatum
Ctenodon histrix
Dalbergia balansae
Dalbergia nigra
Zornia gibbosa
Zornia myriadena
Poiretia bahiana
Adesmia boronioides
Nissolia schottii
Dalea lanata
Dalea aurea
Marina parryi
Psorothamnus schottii
Amorpha fruticosa
Amorpha roemeriana
Apoplanesia paniculata
Eysenhardtia texana
Genista microphylla
Genista sagittalis
Ulex europaeus
Echinospartum horridum
Spartium junceum
Ennacea anthyllis
Cytisus scoparius
Cytisus spinescens
Adenocarpus decorticans
Laburnum anagyroides
Lupinus luteus
Lupinus arboreus
Aspalathus argyrella
Cytisus genistoides
Rafnia perfoliata
Lebeckia sepiaria
Crotalaria pallida
Crotalaria albida
Virgilia divaricata
Podalyria burchellii
Calpurnia aurea
Cadia purpurea
Thermopsis lanceolata
Thermopsis smithiana
Piptanthus nepalensis
Baptisia bracteata
Baptisia leucophaea
Ammopiptanthus mongolicus
Salweenia wardii
Sophora tomentosa
Sophora flavescens
Euchresta japonica
Ammodendron bifolium
Bolusanthus speciosus
Haplormosia monophylla
Camoensia scandens
Bowdichia virgilioides
Leptolobium dasycarpum
Diplotropis ferruginea
Tabaroa caatingicola
Harpalyce brasiliana
Hovea impressinerva
Cyclolobium brasiliense
Poecilanthe parviflora
Dermatophyllum secundiflorum
Ormosia fordiana
Ormosia bahiensis
Zollernia splendens
Zollernia ilicifolia
Exostyles venusta
Harleyodendron unifoliolatum
Holocalyx balansae
Luetzelburgia bahiensis
Vatairea guianensis
Sweetia fruticosa
Styphnolobium japonicum
Styphnolobium affine
Cladrastis kentukea
Pickeringia montana
Cladrastis platycarpa
Swartzia apetala
Swartzia bahiensis
Bobgunnia madagascariensis
Cyathostegia mathewsii
Ateleia glazioveana
Trischidium molle
Myroxylon balsamum
Myrocarpus frondosus
Myrospermum sousanum
Amburana cearensis
Dussia macroprophyllata
Pterodon emarginatus
Pterodon abruptus
Dipteryx alata
Dipteryx odorata
Angylocalyx braunii
Xanthocercis zambesiaca
Castanospermum australe
Chamaecrista mimosoides
Parkinsonia aculeata
Paubrasilia echinata
Gleditsia sinenses
Ceratonia siliqua
Dialium schlechteri
Cercis canadensis
Amherstia nobilis
Dalbergioid s.l. clade
Dalbergieae
Amorpheae
Genisteae
Crotalarieae
Podalyrieae
Sophoreae
Genistoid s.l. clade
Leptolobieae
Brongniartieae
Ormosieae
Exostyleae
Vataireoid clade
Cladrastis clade
Swartzieae
Amburaneae
Dipterygeae
Angylocalyceae
ADA clade
50-kb inversion clade
Papilionoideae
Caesalpinioideae
Dialioideae
Cercidoideae + Detarioideae
0.01/48
0.99/84
0.74/79
0.99/
0.98/–
0.99/*
0.99/*
0.46/–
0.90/98
0.63/*
0.93/*
0.93/98
0.99/79
0.04
Node annotation
Concordant
Uninformative
Discordant
0.90/98 Posterior prob/Bootstrap
Baphia
Andira
Pterocarpus
Centrolobium
Grazielodendron
Ctenodon
Dalea
Crotalaria
Lupinus
Camoensia
Cadia
Leptolobium
Bowdichia
Ormosia
Harleyodendron
Vatairea
Pickeringia
Swartzia
Amburana
Dipteryx
Castanospermum



Psoraleeae
Erythrininae
Phaseolinae
Desmodieae
Cajaninae
Clitoriinae
core Millettieae
Diocleae
Abreae
Indigoferea
Astragalean clade
Vicioid clade
IR-lacking clade
Wisterieae
Hologalegina
Loteae
Robinioid clade
Robinieae
NPAAA clade
Mirbelieae
Mirbelioid clade
Bossiaeeae
Hypocalypteae
Millettioid-Phaseoloid clade
Node annotation
Concordant
Uninformative
Discordant
0.90/98 Posterior prob/Bootstrap
0.04
0.75/71
0.02/90
0.99/*
0.99/*
0.84/*
0.31/*
0.95/*
0.99/*
0.86/*
0.99/96
0.99/96
0.94/91
0.95/98
0.99/*
0.95/*
0.94/98
0.95/93
0.94/98
Pseudoeleeae
Erythrina
Cajanus
Macrotyloma
Stigmatopogon
Desmodium
Mucuna
Cormosema
Clitoria
Tephrosia
Deguelia
Cracca
Diocleae
Astragalus
Hologalegina
Pisum
Lathyrus
Wisteria
Sesbania
Dalbergia
Dicerella
Mirbelia
Hypocalyptus


(a)
CAESALPINIOIDEAE
DETARIOIDEAE
CERC.
POLYGALACEAE
PAPILIONOIDEAE
A
B
C
D
E
F
G
H
I
J
K
L
125 Ma
100
75
50
25
Cretaceous
Pal
Eocene
Oligo
Miocene
Floral architecture
(tip points)
non-keel
keel
MiSSE tip rate
0.4
0.3
0.2
0.1
Net diversification
(branch colours)
Fast
Slow
(b)
Net diversification rate
Relative time
(c)
Probability
Number of shifts
(d)
BAMM net diversification rate
non-keel
keel
Floral architecture
(e)
MiSSE tip rate
non-keel
keel
Floral architecture