Well-resolved phylogeny reveals multiple origins of keel flowers during the diversification of papilionoid legumes

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

[1] Department of Integrative Biology, University of Texas at Austin, Austin, Texas, USA
[2] Instituto de Biologia, Universidade Federal da Bahia, Salvador, Bahia, Brazil
[3] Instituto de Pesquisas Jardim Botânico do Rio de Janeiro, Rio de Janeiro, Brazil
[4] Department of Plant Sciences, University of California Davis, Davis, California, USA
[5] DOE Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[6] School of Life Sciences, Arizona State University, Tempe, Arizona, USA
[7] Department of Biological Sciences and Biotechnology, Hannam University, Daejeon, Korea
[8] 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 flowers are among the most successful pollination syndromes in angiosperms. These flowers steer the mechanical interaction with bees and thus are hypothesized to accelerate pollinator-driven diversification. The highly labile evolution of keel flowers in Papilionoideae (Fabaceae) provides a suitable system to test this hypothesis but a robust phylogeny is lacking.
- 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 seven independent origins and twenty-seven losses of keel flowers were identified in Papilionoideae, Polygalaceae, and Cercidoideae during this time. 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 fixing and chemical defense as well as dispersal to the subtropical and temperate regions.

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

INTRODUCTION

Biodiversity is unevenly distributed across the Tree of Life and elucidating the underlying processes is a fundamental task in biological research (Vargas & Zardoya, 2014). In plants, many speciation mechanisms have been proposed including adaptation to pollinators and arms race with herbivores (Givnish, 2010; Van der Niet *et al.*, 2014; Johnson & Schiestl, 2016; Maron *et al.*, 2019). There is also a growing consensus that these factors may act in synergy to promote diversification (Donoghue & Sanderson, 2015). Legumes (Fabaceae) are among the most ecologically and evolutionarily successful plant families on Earth. With approximately 23,000 species and 800 genera, legumes 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/). This massive diversity has been hypothesized to be an outcome of a suite of key innovations including nitrogen fixation symbiotically with soil bacteria (Sprent, 2001; Werner *et al.*, 2015; Doyle, 2016; Faria *et al.*, 2022), direct chemical defense with alkaloids and canavanine (Bell, 1981; Bisby *et al.*, 1994; Kursar *et al.*, 2009; Wink, 2013), indirect defense via ant domatia and extrafloral nectaries (Marazzi & Sanderson, 2010; Chomicki & Renner, 2015; Marazzi *et al.*, 2019), and an extraordinary range of floral architectures to enhance pollination interaction (Arroyo, 1981; Polhill *et al.*, 1981; Pennington *et al.*, 2000; Cardoso *et al.*, 2012, 2013b).

Compared to nitrogen fixation and defense chemicals, the macroevolution of floral architecture has received relatively less attention in Fabaceae. Members of Fabaceae subfamilies often exhibit lineage-specific floral traits tailored to particular pollinators, such as the showy stamens in Mimosoideae 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 bilateral flowers with two showy wings and a lower keel protecting the reproductive parts (Westerkamp, 1997). They are most commonly found in Papilionoideae and Polygalaceae, but also independently evolved several times in other angiosperm families such as Ranunculaceae, Geraniaceae, and Commelinaceae (Westerkamp, 1997). These repeated evolutions exhibit functional convergence but divergent developmental pathways (Bello *et al.,* 2012). The bilateral symmetry and fused stamens can effectively restrict the direction of entry by pollinators

82 and aid with precise pollen placement (Neal *et al.*, 1998). Furthermore, many
83 Papilionoideae flowers exhibit primary and secondary pollination mechanisms such as
84 valvular, pump, explosive, and brush mechanisms to prevent pollen waste and deposit
85 pollen on particular bee body parts (Etcheverry *et al.*, 2012; Uluer *et al.*, 2022). Therefore,
86 the evolution of keel flowers is considered to be an outcome of pollinator-mediated
87 selection (Grant, 1949; Stebbins, 1970; Rosas-Guerrero *et al.*, 2014; Stewart *et al.*, 2022).
88 There is mounting evidence from other empirical systems that pollinator-mediated
89 selection can be a powerful driver of diversification (e.g., the bat-pollinated *Burmeistera*;
90 the beak of the galea in *Pedicularis*; Ree, 2005; Muchhala, 2006), but this has yet to be
91 tested in legumes.
92
93 To rigorously assess trait-driven diversification, a system with high trait variability is
94 essential for robust comparative hypothesis testing (Ricklefs & Renner, 1994, 2000;
95 Donoghue & Sanderson, 2015). The remarkable diversity in floral symmetry and
96 architecture in the Papilionoideae offers an ideal model for such investigations. Although
97 keel flowers are considered an essential feature and a hallmark of Papilionoideae (Polhill *et*
98 *al.*, 1981), there is considerable variation in floral symmetry in early diverging
99 Papilionoideae clades. For example, members from the tribes Sophoreae and Swartzieae
100 are frequently actinomorphic, with five petals undifferentiated or even lost to various
101 degrees (Pennington *et al.*, 2000; Cardoso *et al.*, 2012, 2013b). These genera are typically
102 species-poor and more than 80% have no more than ten extant species based on records
103 from the Plants of the World Online (POWO; https://powo.science.kew.org). In contrast,
104 the keel flower is highly constrained among the large non-protein-amino-acid accumulating
105 clade (NPAAA; ca. 10,200 species), which is mostly known for economically important
106 crops like soybeans, peas, lentils, and clovers (Choi *et al.*, 2022). These general
107 observations have prompted our speculation that the evolution of keel flowers may be
108 associated with increased diversification rates in Papilionoideae.
109
110 Testing these macroevolutionary hypotheses requires a robust and densely sampled
111 phylogeny. Over the past three decades, a general phylogenetic framework for Fabaceae
112 has been developed thanks to a tremendous network of global collaboration (LPWG, 2013,

113 2017). Six distinct subfamilies are recognized by the Legume Phylogeny Working Group
114 (LPWG), with Papilionoideae as the most species-rich subfamily comprising 14,000 species
115 in 503 genera (Lewis *et al.*, 2005; LPWG, 2017). However, the LPWG phylogeny and
116 classification is largely based on two plastid genes *matK* and *rbcL*, which have obvious
117 limitations when dealing with rapid radiations or complex evolutionary scenarios such as
118 hybridization (Barrett *et al.*, 2014; Morales-Briones *et al.*, 2021). In Papilionoideae, studies
119 based on *matK*, *rbcL*, and complete plastomes have consistently defined 22 well-supported
120 clades but the relationships among these clades remain poorly resolved (Wojciechowski *et*
121 *al.*, 2004; McMahon & Sanderson, 2006; Cardoso *et al.*, 2012, 2013b; Zhang *et al.*, 2020;
122 Choi *et al.*, 2022). Most significantly, Swartzieae and the ADA clade have been individually
123 proposed as sister to all remaining Papilionoideae. Given that Swartzieae contains mostly
124 non-keel flowers with frequent petal losses whereas the ADA clade contains largely keel
125 flowers (Torke & Schaal, 2008; Cardoso *et al.*, 2015), their placement will have a strong
126 bearing on the timing and evolutionary trajectory of keel flowers. Furthermore, many of
127 the 22 clades are characterized by tropical trees that are scarce in distribution and
128 underrepresented in previous studies. These include species-poor clades such as the
129 African *Amphimas*, the Amazonian *Aldina,* and the North American *Dermatophyllum*
130 (Cardoso *et al.*, 2013). Along these lines, recent advances in legume phylogeny using multi-
131 locus nuclear datasets have promising applications in resolving complex evolutionary
132 histories (Koenen *et al.*, 2021; Zhao *et al.*, 2021) and inspired us to apply this technique to
133 address the recalcitrant relationships within Papilionoideae.
134
135 In this study, we generated a phylogenomic dataset of 1,456 low-copy genes from 287
136 species with a focus on recalcitrant clades in Papilionoideae. Using both coalescent and
137 concatenation methods, we inferred a robust species tree and consistently resolved the
138 early divergence history of Papilionoideae. This nuclear phylogeny was subsequently used
139 as the backbone to reconstruct a 3,326-species time tree of Fabales to explore the evolution
140 of keel flowers and to test whether it has promoted diversification in Fabales.
141
142 **MATERIALS AND METHODS**
143 **Taxon sampling**

144 Our taxon sampling included 279 species broadly distributed across Papilionoideae, plus
145 one Cercidoideae, one Detarioideae, one Dialioideae, and five Caesalpinioideae species as
146 outgroups (Tables S1–2). The dataset comprised an aggregation of 47 newly generated
147 transcriptomes and 32 low-pass genomic sequences (Table S1) as well as published
148 sequences from 208 species from Zhao *et al.* (2021) to broaden taxon sampling (Table S2).
149 Newly generated transcriptomes and genomic sequences were mostly from
150 morphologically and evolutionarily representative genera from the early diverging clades
151 of Papilionoideae (Cardoso *et al.*, 2012; Choi *et al.*, 2022). Collectively the dataset
152 represented 19 of the 22 major clades consistently recognized in the subfamily (Cardoso *et*
153 *al.*, 2012; Cardoso *et al.*, 2013; Choi *et al.*, 2022). The three clades we did not include are
154 *Amphimas*, *Aldina*, and the recently described *Cabari* (Gregório *et al.*, 2024). All newly
155 generated sequences were derived from plant materials collected in the field or grown in
156 the University of Texas at Austin greenhouses (Table S1). Voucher specimens were
157 deposited in the herbaria of the Rio de Janeiro Botanic Garden (RB), State University at
158 Feira de Santana (HUEFS), and the Billie L. Turner Plant Resources Center (TEX-LL). For
159 DNA and RNA extraction, the methods described previously by Lee *et al.* (2021) and Zhang
160 *et al.* (2013) were employed, respectively.
161
162 **RNA sequencing and transcriptome assembly**
163 Total RNA samples were shipped to GENEWIZ (South Plainfield, NJ, USA) for ribosomal
164 RNA (rRNA) depletion, library preparation, and sequencing of approximately 25 million
165 paired-end reads on an Illumina HiSeq platform (Illumina, San Diego, CA). We accessed the
166 quality of raw 150 bp paired-end reads using FastQC v.0.11.5
167 (www.bioinformatics.babraham.ac.uk/projects/fastqc) and removed filtered adapters and
168 low-quality reads using TrimGalore v.0.5.0
169 (https://github.com/FelixKrueger/TrimGalore). The rRNA was removed from the merged
170 trimmed reads using SortMeRNA v.2.1 (Kopylova *et al.*, 2012) with eight rRNA databases
171 for prokaryotes, archaea, and eukaryotes. Clean RNA-Seq reads were assembled *de novo*
172 using Trinity v.2.8.4 (Grabherr *et al.*, 2011), and the assembled transcripts were subjected
173 to the assessment of completeness and redundancy using BUSCO v.4.1.0 (Benchmarking
174 Universal Single-Copy Orthologs; Simão *et al.*, 2015) with the eudicots_odb10.2019-11-20

175 database. The number of raw reads and BUSCO evaluation results are provided in Table S1.
176 All newly generated reads are available on GenBank Sequence Read Archive under
177 Bioproject PRJNA1120003 (https://www.ncbi.nlm.nih.gov/sra/PRJNA1120003).
178
179 **Low-copy nuclear dataset assembly**
180 We used the 1,559 low-copy nuclear loci identified by Zhao *et al.* (2021) for our
181 phylogenetic reconstruction. To recover these loci in our transcriptomes, we used the
182 profile hidden Markov models implemented in HMMER v3.3.2 (Johnson *et al.,* 2010) to
183 identify orthologs. Briefly, a DNA profile database was built for each locus using the
184 function 'hmmbuild' based on the DNA alignments from Zhao *et al.* (2021). These profiles
185 were then searched against the target transcriptomes with an e-value threshold of 1e-50
186 using 'hmmsearch'. We then extracted the best hit from each species as the candidate
187 ortholog using a custom Python script (all scripts available on GitHub repository
188 https://github.com/lmcai/Papilionoideae_phylogenomics).
189
190 To recover the same set of loci from our genome sequence data, we developed a new
191 function of the PhyloHerb pipeline (Cai *et al.*, 2022) to automatically extract nuclear loci
192 from low-pass genome sequencing data. Briefly, PhyloHerb accomplishes this in three steps
193 — mapping, assembly, and scaffolding. First, it mapped reads to the reference sequence
194 using the very-sensitive mode of Bowtie v2.3.4.1 (Langmead & Salzberg, 2012). Mapped
195 reads were then assembled into contigs using SPAdes v3.8.0 under the default setting
196 (Bankevich *et al.*, 2012). The contigs were scaffolded into full-length loci based on syntenic
197 information compared to the reference using BLAST with an e-value threshold of 1e-40.
198 These functions were integrated as the '-m assemb' and '-m ortho -nuc' functions in
199 PhyloHerb v1.1. For details, see the GitHub page
200 (https://github.com/lmcai/PhyloHerb/#iv-retrieve-low-copy-nuclear-genes). Finally, for
201 quality control purposes, we also estimated the nuclear genome coverage of genome
202 sequences using Jellyfish v2.3.0 with a kmer size of 21 (Marçais & Kingsford, 2011).
203
204 **Sequence alignment, dataset cleaning, and phylogenetic inference**

205 Sequence alignment of each locus was implemented in MAFFT v7.245 (Katoh & Standley,
206 2013) under the default settings. The aligned sequences were subjected to three rounds of
207 stringent cleaning to remove excessive missing data and potential paralogs. First, we used
208 the '-missing' function in PhyloHerb to remove sequences with >70% ambiguous sites
209 including gaps in each locus. Second, a phylogeny-based ortholog identification method
210 from Yang and Smith (2014) was used to remove potential paralogs. To accomplish this, a
211 maximum likelihood (ML) phylogeny was inferred for each gene using the DNA alignments
212 from the previous cleaning step. The best substitution model was automatically
213 determined by ModelFinder implemented in IQ-TREE v.2.2.2.6 (Minh *et al.*, 2020).
214 Branches that were more than fifteen times the mean branch length were pruned from the
215 resulting trees using a custom Python script 'prune_long_branches.py' (available on
216 GitHub). Then, a phylogeny-based ortholog identification approach using the Python script
217 'prune_paralogs_RT.py' (Yang & Smith 2014) was applied to further remove potential
218 contaminants of paralogs. Third, all DNA sequences were realigned using the MAFFT E-INS-
219 i algorithm (--genafpair --maxiterate 1000) after removing spurious sequences identified in
220 the previous steps. Sites with more than 70% missing data were removed from the
221 alignment using the '-gt 0.7' flag in trimAL. We also found that the published alignments
222 from Zhao *et al.* (2021) were site-trimmed, thus creating systematic biases for phylogenetic
223 inference (e.g., shared gaps for sequences from Zhao *et al.* (2021) compared to species
224 sampled by us). To accommodate for this, we used a custom Python script to mask all sites
225 in our sequences that were present in >60% of species but absent in >90% of species from
226 Zhao *et al.* (2021). Loci with less than 199 (70%) taxa were removed from downstream
227 phylogenomic assessments. In the end, 1,456 out of the 1,559 loci passed our screening
228 criteria (Fig. S1).
229
230 To estimate a species tree using the multispecies-coalescent model, we first inferred
231 individual gene trees for each of the 1,456 loci in IQ-TREE. The best substitution model was
232 automatically determined by ModelFinder implemented in IQ-TREE and branch support
233 was evaluated by 1000 ultrafast bootstrap replications. The species tree was inferred from
234 1,456 gene trees using ASTRAL-III v5.7.8 (Zhang *et al.*, 2017) under the default settings.
235 Meanwhile, to further mitigate the impact of artificially resolved gene tree branches with

236 low support values, we used newick-utils v1.6 (Junier & Zdobnov, 2010) to collapse gene
237 tree branches with ultrafast bootstrap support values less than 70. We then inferred a
238 species tree with these gene trees containing polytomies in ASTRAL-III.
239
240 To infer the species tree using the concatenation methods, we first further subsampled the
241 loci based on length and missing data. Any loci with alignments shorter than 900 base pairs
242 (bp; 33% loci removed) or less than 227 species (80% of the total species) were excluded
243 from the concatenation analysis, which resulted in 672 loci. The concatenated supermatrix
244 was generated by the '-m conc' function in PhyloHerb, along with a gene delineation file
245 input into PartitionFinder 2. The best partition scheme for ML analyses employing the
246 general time reversible (GTR) + gamma model was selected by PartitionFinder 2 using the
247 heuristic search algorithm 'rcluster' (Lanfear *et al.*, 2017). We then inferred an ML
248 phylogeny of the 287 species using IQ-TREE with 100 traditional nonparametric bootstrap
249 replications (BP) under the GTR + gamma nucleotide substitution model and linked but
250 variable evolution rates for each partition with the '-p' flag.
251
252 **Topology test**
253 Several conflicting relationships emerged from the comparison between concatenation and
254 coalescent analyses. To explore the origin of these conflicts and identify the best supported
255 phylogenetic scenarios, we conducted topology tests on six focal groups defined in Table
256 S3. By alternating the position of the focal group and fixing the rest of the tree, we inferred
257 an optimum phylogeny for competing scenarios in IQ-TREE. We then conducted the
258 approximately unbiased (AU), Kishino-Hasegawa (KH), and Shimodaira-Hasegawa (SH)
259 tests with 10,000 replicates as implemented in IQ-TREE. To compare the goodness-of-fit of
260 alternative species trees to gene trees under the coalescent model, we conducted a
261 pairwise likelihood ratio test (Liu *et al.*, 2019). We similarly optimized the branch length of
262 fixed topologies with alternative positions of the focal group in MP-EST v3.0 (Liu *et al.*,
263 2010). We used the test2.sptree function from the R package Phybase v2.0 (Liu & Yu, 2010)
264 to conduct the likelihood ratio test described in Liu *et al.* (2019). The likelihood ratio test
265 statistic is given by $\tau = 2(L_{\mathrm{Tree2}} - L_{\mathrm{Tree1}})$, in which L_{Tree1} and L_{Tree2} are the log-likelihoods of
266 the null and alternative hypotheses, respectively. The statistical significance (p-value) of

267 the comparison is evaluated by generating a null distribution of τ via bootstrap sampling
268 gene trees 100 times. Finally, to diagnose the causes of conflicting results, we examined the
269 gene (gCF) and site concordance factors (sCF), as well as gene-wise likelihood scores
270 implemented in IQ-TREE.
271
272 **Bayesian divergence time estimation**
273 To infer a Bayesian time tree, we subsampled the nuclear dataset to select the most clock-
274 like genes using SortaDate (Smith *et al.*, 2018). Total tree length, root-to-tip variance, and
275 topological congruence with the species tree were used as key indicators for phylogenetic
276 informativeness and stable molecular rate. We selected the 30 best genes (Table S4) for
277 dating to facilitate a reasonable running time of the Bayesian analysis, an approach
278 comparable to a recent study of Fabaceae divergence time (Koenen *et al.*, 2021). We
279 carefully selected eleven fossils to calibrate our molecular dating analyses, including four
280 fossils from early diverging eudicot lineages and seven well-recognized fossils within
281 Fabales (Table S5). In particular, the earliest fossil fruits and leaflets of Fabaceae from the
282 Early Paleocene (63.5 Ma) of the Denver Basin in Colorado, USA, were used to constrain the
283 stem group Fabaceae (Lyson *et al.*, 2019). Many recent fossils within Fabaceae were not
284 included because they provide redundant constraints compared to the eleven fossils we
285 selected. We used BEAST v.2.2.1 (Bouckaert *et al.*, 2014) to estimate divergence times
286 under the GTR substitution model, an uncorrelated relaxed molecular clock model, and a
287 birth-death tree prior. The topology was fixed to reflect the optimum species tree inferred
288 from the ASTRAL analysis. Fossil calibrations were set as exponential priors with the offset
289 minimum ages aligned with the fossils described in Table S5. The root node, for which we
290 used a normal prior at 128.6 Ma with a standard deviation of 1.0, truncated to minimum
291 and maximum ages of 113 Ma (the Aptian–Albian boundary) and 136 Ma (the oldest known
292 crown angiosperm fossil (Magallón *et al.*, 2015). We ran analyses under the uncorrelated
293 lognormal model for 100 million generations and confirmed convergence with Tracer
294 v1.7.1 (Rambaut *et al.*, 2018). The first 30% of the generations were discarded as burn-in
295 before summarizing median branch lengths and substitution rates with TreeAnnotator
296 from the BEAST package.
297

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

299 To evaluate the macroevolutionary trends of floral diversification in Papilionoideae under a
300 broader taxonomic context, we generated a plastid *matK* gene dataset that included 3,106
301 species from 794 genera in Fabaceae (92% genus-level completeness), plus 220 additional
302 species from all other families in Fabales and outgroups. Sequences from Choi *et al.* (2022),
303 Legume Phylogeny Working Group (LPWG 2017), and Gregório *et al.* (2024) were
304 consolidated to generate the most comprehensive *matK* database for Fabaceae, and
305 combined with *matK* sequences from Polygalaceae, Surianaceae, and Quillajaceae available
306 on GenBank, which served as outgroups in Fabales (Table S6).
307
308 As a first step of data curation, we updated the nomenclature to reflect the most recent
309 effort from POWO, Flora e Funga do Brasil (https://floradobrasil.jbrj.gov.br/consulta; BFG
310 2022), and taxonomic studies from individual clades (e.g., Hughes *et al.*, 2022). Second, we
311 employed multiple rounds of phylogeny-guided cleaning for the *matK* sequence alignment.
312 Initial alignment was conducted with the translation align option of MAFFT v.7.490
313 implemented in Geneious Prime 2023.1.2 (https://www.geneious.com/) and was manually
314 inspected thereafter. For outgroup species, a separate alignment was generated using the
315 L-INS-i algorithm in MAFFT to accommodate higher sequence divergence across families.
316 These two alignments were then merged using the profile-profile alignment algorithm
317 implemented in ClustalX (Thompson *et al.*, 2003). A draft ML phylogeny was inferred under
318 the GTR+I+G model with 1000 ultrafast bootstrap replications in IQ-TREE and we filtered
319 the *matK* matrix accordingly to remove misidentified and rogue taxa.
320
321 To infer a robust *matK* phylogeny for divergence time estimation, we used our well-
322 supported nuclear phylogeny to constrain the topology and node ages of major clades in
323 the *matK* dataset. This is necessary because the slow substitution rate of *matK* may result
324 in contentious relationships and the lack of sequence divergence among close relatives may
325 impact ancestral state estimation and divergence time estimation. As a result, relationships
326 *between* the following clades were constrained: the ADA, Swartzieae, Cladrastis, Vataireoid,
327 Exostyleae, Dalbergioid s.l., Genistoid s.l., Andira, Baphieae, and NPAAA. Relationships
328 *within* each clade and *between* all subfamilies and families remained as-is. With this

329 partially fixed topology, the ML phylogeny was inferred under the best fitting model in IQ-
330 TREE.
331
332 Thirty secondary calibration points from the Bayesian time tree (Table S7) were used to
333 infer the *matK* gene time tree under the penalized likelihood method implemented in
334 TreePL v1.0 (Smith & O'Meara, 2012). These calibration points were imposed as both
335 maximum and minimum age constraints to fix node ages. We then primed and cross-
336 validated our data to establish the best smoothing parameter (smooth = 10) for our
337 dataset. Finally, we conducted a thorough search to infer the best time tree.
338
339 **Testing state-dependent diversification analysis**
340 We performed a Bayesian analysis of macroevolutionary mixtures using the program
341 BAMM v2.5.0 (Rabosky, 2014) to detect shifts in diversification rates. Reversible-jump
342 Markov chain Monte Carlo was run for 20 million generations and sampled every 1,000
343 generations. Posterior probabilities of rate shifts were summarized with the R package
344 BAMMtools v.2.1.6 (Rabosky *et al.*, 2014) and visualized using the scripts from (Padfield *et*
345 *al.*, 2024).
346
347 To explore whether the evolution of keel flowers drives pollinator-associated
348 diversification in clades like Papilionoideae (Fabaceae) and Polygaleae (Polygalaceae), we
349 estimated the divergence time history of Fabales under the state-dependent speciation and
350 extinction (SSE) models. First, we coded the floral type of all species as either keel or non-
351 keel flowers (Table S6). Here, keel flowers are defined as bilaterally symmetrical,
352 specialized flowers with petals highly differentiated into a standard, two wings, and a keel
353 composed of two partially to fully fused petals enclosing the reproductive organs.
354 Connation among floral components, including stamens and keel petals, is common. To
355 infer the sampling ratio, we used the powoGenera function from the R package expowo
356 (Zuanny *et al.*, 2024) to obtain the number of species for each genus based on the Plants of
357 the World Online (https://powo.science.kew.org) database. The total species number for
358 the keel flower clades and non-keel flower clades was enumerated to calculate the
359 sampling ratio for these two groups. Finally, we inferred the diversification rate under the

360 state-dependent and state-independent models using the R package hisse v2.1.8 (Beaulieu
361 & O'Meara, 2016). We applied our dataset to the Binary State Speciation and Extinction
362 model (BiSSE), the Hidden State Speciation and Extinction model (HiSSE), as well as the
363 state-independent models CID-2 and CID-4 that have equal parameter complexity
364 compared to BiSSE and HiSSE, respectively (Maddison *et al.*, 2007; Beaulieu & O'Meara,
365 2016). We used the function 'hisse' followed by 'MarginReconHiSSE' to infer the speciation
366 and extinction rates as well as ancestral states under the HiSSE model.
367
368 **RESULTS**
369 **An efficient phylogenetic data assembly pipeline based on a curated set of loci**
370 BUSCO assessments for the assembled transcriptomes showed a high level of
371 completeness, with the percentage of complete orthologous genes ranging from 70.3% to
372 91.1% (Table S1). The complete multi-loci dataset included 1,456 genes with a mean length
373 of 1,003 bp after trimming (Data S1). These loci on average contained 253 (85%) species
374 and the concatenated DNA dataset consisted of 941,354 sites with a gene and character
375 occupancy of 91.5% and 83.0%, respectively. None of our newly sequenced species
376 exhibited exceptional levels of missing data (<30% missing; Table S1), but three species
377 added from Zhao *et al.* (2021), which included *Pisum sativum* L., *Hesperothamnus*
378 *pentaphyllus* (Harms) Harms, and *Euchresta japonica* Hook. f. ex Regel, contained the
379 highest level of missing data (69–85% missing; Table S2).
380
381 Our comprehensive, genus-level taxon sampling represents multiple data collection
382 strategies, including high-quality reference genomes, RNA-seq, and low-pass genome
383 sequencing (i.e., genome skimming). Among these three approaches, our *in-silico* extraction
384 of phylogenetic markers from genomic sequencing data using PhyloHerb represents a cost-
385 effective approach to construct a multi-locus phylogeny, eliminating the need for target
386 enrichment library preparation. Genomic sequencing of the 32 Papilionoideae taxa yielded
387 25–30 million 150 bp paired-end reads and 1–15× nuclear coverage according to k-mer
388 distribution (Table S8; Data S2). In the final DNA matrix, the level of missing data in these
389 32 species was not statistically different from transcriptomic data (Welch two sample t-test
390 *p*-value = 0.08). Even for species with less than 2× nuclear genome coverage, at least 1,075

391 (74%) of the targeted loci were recovered (Table S8). Computationally, PhyloHerb is also
392 highly accessible. Extracting ~1,500 loci from each species on average requires 903.2s CPU
393 time and 2.9 GB peak memory including assembly and ortholog identification, which can be
394 completed on most laptops.
395
396 **Multi-locus nuclear dataset resolves contentious relationships in Papilionoideae**
397 To accommodate for incomplete lineage sorting, we used both coalescent and
398 concatenation methods for phylogenetic inference. The coalescent-based phylogeny
399 estimated by ASTRAL was largely congruent with the concatenation-based phylogeny
400 inferred by IQ-TREE and 92.6% and 89.1% of the nodes received maximum support (100
401 BP; 1.0 posterior probability, PP), respectively (Fig. 1).
402
403 Within Papilionoideae, the monophyly of these 19 clades was fully supported by both
404 coalescent and concatenation analyses (Fig. 1; Fig. S2). Relationships among these clades
405 were confidently resolved including those of recalcitrant lineages such as *Dermatophyllum,*
406 *Sesbania,* and the Andira clade (Fig. 1). Among the earliest-diverging lineages in
407 Papilionoideae, the ADA, Swartzieae and Cladrastis clades formed successive sister groups
408 to the remaining Papilionoideae (i.e., the 50-kb inversion clade, Fig. 1). Within the 50-kb
409 inversion clade, the Vataireoid and Exostyleae clades together formed a well-supported
410 clade (100 BP, 1.0 PP) sister to all other groups. The monophyly of the Vataireoid and
411 Exostyleae clades was strongly favored by both the AU test using the concatenated matrix
412 and the likelihood ratio test based on the multispecies coalescent (MSC) model. The
413 alternative topology suggested by Zhao *et al.* (2021), where these two groups were not
414 monophyletic, was rejected (AU test p-value < 0.01; MSC likelihood ratio test p-value <0.01;
415 Table S3).
416
417 One major conflict in our results is the relative placement of the Dalbergioid s.l. (1,580
418 spp.) and Genistoid s.l. (2,100 spp.) clades within the 50-kb inversion clade. A
419 monophyletic Dalbergioid s.l. + Genistoid s.l. clade was recovered but poorly supported by
420 the concatenation method (41 BP). On the other hand, the coalescent method recovered a
421 closer relationship between Dalbergioid s.l. and the Andira, Baphieae, and NPAAA clades
422

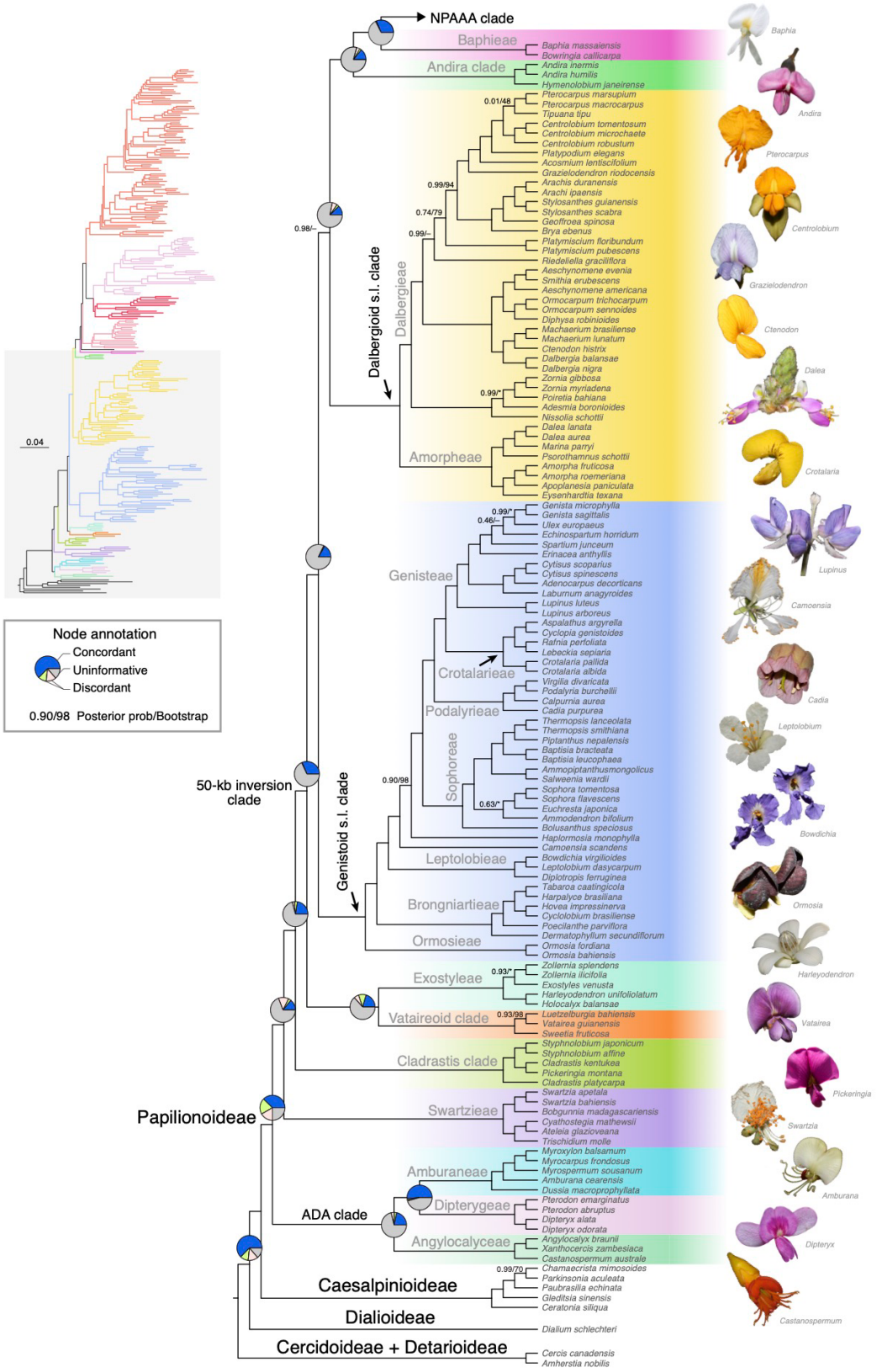
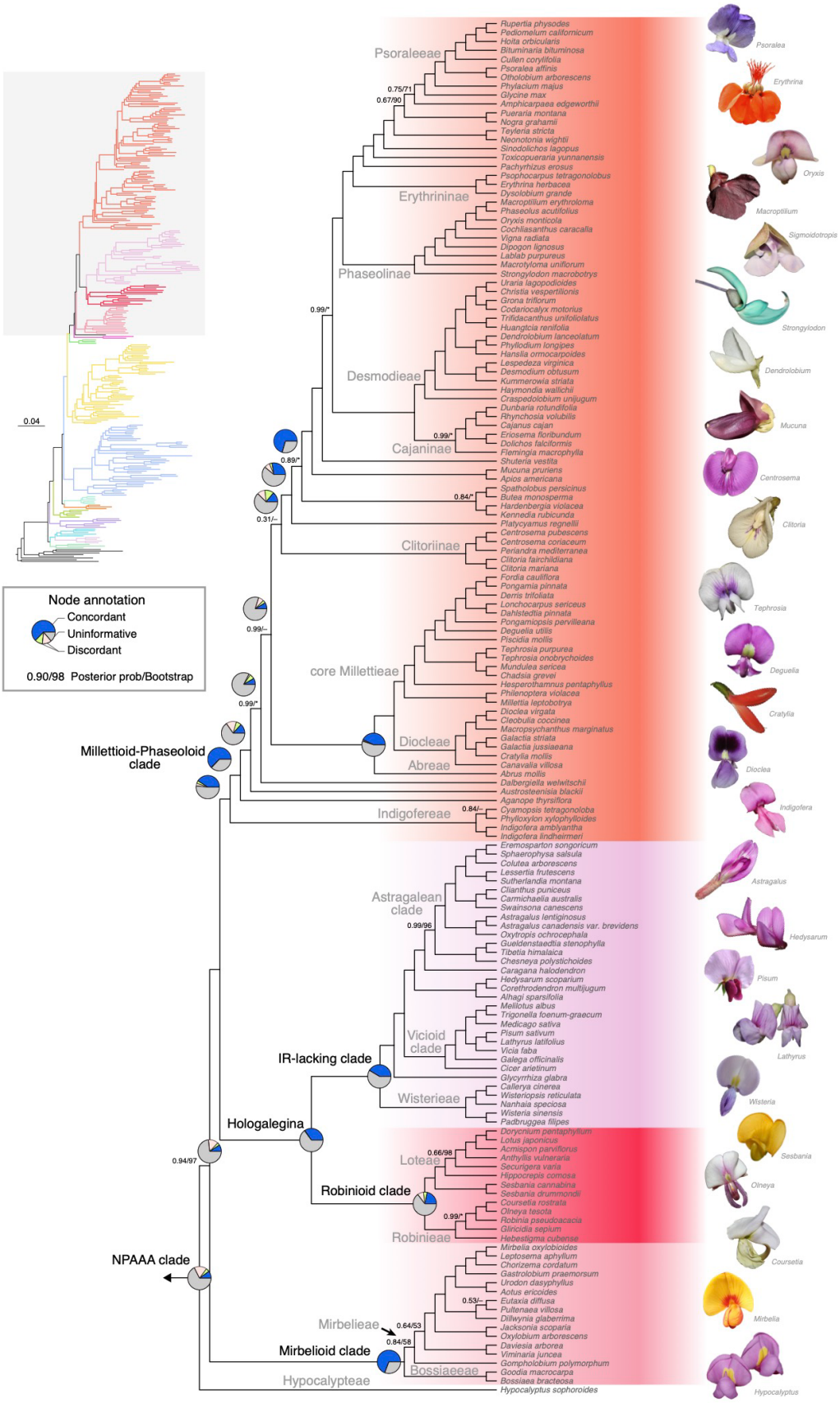


Figure 1 Phylogeny of Papilionoideae (continued on the next page).


Psoraleeae
Erythrininae
Phaseolinae
Desmodieae
Cajaninae
Clitoriinae
core Millettieae
Diocleae
Abreae
Indigofereae
Astragalean clade
Vicioid clade
Wisterieae
Loteae
Robinieae
Mirbelieae
Bossiaeeae
Hypocalyptea
Millettioid-Phaseoloid clade
IR-lacking clade
Hologalegina
Robinioid clade
Mirbelioid clade
NPAAA clade
0.04
Node annotation
Concordant
Uninformative
Discordant
0.90/98 Posterior prob/Bootstrap
Psoralea
Erythrina
Oryxis
Macroptilium
Sigmoidotropis
Strongylodon
Dendrolobium
Mucuna
Centrosema
Clitoria
Tephrosia
Deguelia
Cratylia
Dioclea
Indigofera
Astragalus
Hedysarum
Pisum
Lathyrus
Wisteria
Sesbania
Olneya
Coursetia
Mirbelia
Hypocalyptus

426 **Figure 1.** Phylogeny of Papilionoideae based on low-copy nuclear gene and inferred from
427 the multispecies coalescent model implemented in ASTRAL-III. Branch supports indicate
428 posterior probability from ASTRAL and bootstrap replicates support from IQ-TREE,
429 respectively. Only branches that are not maximally supported are labeled with conflicting
430 topologies marked by '–'. Pie charts on each node represent gene concordance factors for
431 the proportion of concordant (blue), discordant (green and pink), and noninformative
432 (grey) genes. The phylogram in the left margin illustrates branch length in mutational
433 units. The photos (by Domingos Cardoso, except *Pickeringia* by Martin Wojciechowski,
434 *Grazielodendron* by Claudio Nicoletti, *Baphia* by Chan Chee Keong (iNaturalist,
435 https://inaturalist.ca/photos/60770727), *Hypocalyptus* by Brian du Preez (iNaturalist,
436 https://www.inaturalist.org/photos/60072646), *Mirbelia* by Nina Kerr (iNaturalist,
437 https://www.inaturalist.org/photos/345474574), *Pisum* by Michela Papa (iNaturalist,
438 https://inaturalist.ca/photos/372960709), and *Psoralea* by katrivier (iNaturalist,
439 https://inaturalist.ca/photos/262747553) illustrate the remarkable diversity of floral
440 architecture within the Papilionoideae.
441

442 with high confidence (0.98 PP; Fig. S2). The AU test on these two alternative topologies did
443 not provide definitive conclusions (*p*-value >0.05; Table S3) and the gene-wise likelihood
444 supporting each topology was approximately the same (Fig. S3). The MSC likelihood ratio
445 test, however, rejected the monophyletic Dalbergioid s.l. + Genistoid s.l. clade (*p*-value =
446 0.03). Within the Genistoid s.l. clade, the genus *Ormosia* Jacks was strongly supported as
447 sister to all remaining lineages of the clade (100 BP; 1.0 PP), while *Dermatophyllum* was
448 confidently placed as an early-diverging lineage sister to Brongniartieae (88 BP; 1.0 PP).
449 The African monotypic genus *Haplormosia* Harms was nested within the genistoids (98 BP;
450 0.90 PP). Within the Dalbergioid s.l. clade, all previously identified major subclades were
451 maximally supported. The Amorpheae appeared as sister to all remaining dalbergioids, and
452 the *Adesmia*, *Dalbergia*, and *Pterocarpus* subclades were phylogenetically successive sisters
453 (100 BP; >=0.99 PP).
454

455 The Andira, Baphieae, and NPAAA clades were each monophyletic and they formed a grade
456 comprising the remaining Papilionoideae. Within the NPAAA clade, the genus *Hypocalyptus*
457 Thunb. was the earliest diverging lineage (97 BP; 0.94 PP), while the Mirbelioid,
458 Hologalegina, and Indigofereae + Millettioid-Phaseoloid clades formed successive sisters to
459 Hypocalypteae. Within mirbelioids, both coalescent and concatenation methods placed
460 lineages from tribe Bossiaeeae as the earliest diverging clade sister to a poorly supported
461 clade composed mostly of tribe Mirbelieae (<58 BP; <0.83 PP). Such low support could be

462 attributed to the two short internal branches (1.3–1.7e-3 in mutation units; 2.0–3.9e-2 in
463 coalescent units) separating Bossiaeeae and two non-monophyletic groups within
464 Mirbelieae — (*Daviesia* Sm. + *Viminaria* Sm.) and *Gompholobium* Sm. These rapid radiations
465 have rendered little phylogenetic information to reconstruct this section of the phylogeny
466 — the gene concordant factor was 16.3% for the branch separating Bossiaeeae and
467 Mirbelieae, while 14.0% of the genes supported the other two alternative quartet
468 topologies at this node; similarly, 16.0% of the genes supported the paraphyly of *Daviesia* +
469 *Viminaria* and *Gompholobium*, but one of the alternative topologies was supported by
470 17.4% of the genes (Fig. 1; Data S3). Results from the site concordance factor largely
471 reflected a similar trend where each of the three quartet topologies was supported by 33%
472 of the sites (Data S3).
473
474 The Millettioid-Phaseoloid clade within the larger NPAAA clade harbors more uncertainty.
475 The three genera *Aganope* Miq., *Austrosteenisia* R. Geesink and *Dalbergiella* Baker f. formed
476 a well-supported basal grade within the millettioids (100 BP; 1.0 PP). The concatenation
477 method confidently placed (100 BP) subtribe Clitoriinae—including *Centrosema* (DC.)
478 Benth., *Periandra* Mart. ex Benth. and *Clitoria* L.—as sister to *Dalbergiella*. In contrast, the
479 coalescent method placed Clitoriinae with a clade consisting of members from the
480 polyphyletic Phaseoleae, Desmodieae, and Psoraleeae with low confidence (0.31 PP). A
481 clade comprising Abreae, Diocleae, Millettieae formed the next diverging group in the
482 millettioids. The remaining millettioids included tribes Desmodieae, Psoraleeae, and the
483 polyphyletic Phaseoleae. In both coalescent and concatenation analyses, *Platycyamus*
484 *regnellii* Benth. was the earliest diverging lineage (100 BP; 1.0 PP), but conflicts exist in
485 terms of the placement of the *Butea* clade (genera *Spatholobus* Hassk., *Butea* Roxb. ex
486 Willd., *Hardenbergia* Benth., *Kennedia* Vent.), the Apios clade (genera *Mucuna* Adans. and
487 *Apios* Fabr.), and *Shuteria* Wight & Arn. Both coalescent and concatenation methods
488 strongly supported the *Butea* clade to be the earliest diverging among the three (100 BP;
489 1.0 PP), while the Apios clade and *Shuteria* were recovered as monophyletic (99 BP) in the
490 concatenation analysis but paraphyletic (0.89 PP) in the coalescent analysis.
491
492 **Evolution of keel flowers in Fabales**

493 The robust nuclear phylogeny was used to constrain the uncertain relationships among
494 major clades in the *matK*-based phylogeny. This broadly sampled phylogeny of Fabales
495 included 3,326 species — 3,106 Fabaceae (15.9% species diversity), 1 Quillajaceae (100%),
496 7 Surianaceae (55.6%), 206 Polygalaceae (15.4%), and 6 additional eudicots for dating
497 purposes. With fossil constraints, the time tree inferred by BEAST supported a Late
498 Cretaceous origin of crown group Fabales at 78.0 Ma (95% HPD 71.9–88.3 Ma; Fig. S4), and
499 an early Paleocene origin of crown group Fabaceae and Polygalaceae at 66.0 Ma (95% HPD
500 64.7–66.8 Ma) and 56.1 Ma (95% HPD 55.8–56.6 Ma), respectively. Within Fabaceae, we
501 inferred an explosive diversification leading to all six subfamilies within 7 million years
502 after the origin of the family. Crown group Papilionoideae was inferred to originate at 60.9
503 Ma (95% HPD 59.7–62.1 Ma; Fig. S4).
504
505 Keel flowers, which are mostly known in Papilionoideae and tribe Polygaleae
506 (Polygalaceae), have been independently lost and gained multiple times. Ancestral
507 character reconstruction suggested at least seven independent origins of keel flowers in
508 tribe Polygaleae (Polygalaceae), *Cercis* L. (Cercidoideae), *Angylocalyx* Taub. (ADA clade,
509 Papilionoideae), tribe Dipterygeae (ADA clade, Papilionoideae), the MRCA of *Petaladenium*
510 Ducke and *Dussia* Krug & Urb. ex Taub. (ADA clade, Papilionoideae), *Myrospermum* Jacq.
511 (ADA clade, Papilionoideae), and the MRCA of the Cladrastis clade and the 50-kb inversion
512 clade (Papilionoideae). In addition, there were two independent losses of keel flowers in
513 Polygalaceae and twenty-five in Papilionoideae (summarized in Table S9; Fig. S5). These
514 numbers, especially the losses, were strongly dependent on the phylogeny. We inferred the
515 ancestral floral architecture of Papilionoideae to be non-keel-like, with the first keel
516 flowers emerging in the common ancestor of the Cladrastis and the 50-kb inversion clades
517 around 59.0 Ma (95% HPD 57.9–60.2 Ma). This is more recent than the origination of keel
518 flowers in crown group Polygalaceae, which included *Xanthophyllum* at 62.8 Ma (95% HPD
519 59.6–64.9 Ma), though the exact order cannot be established with confidence due to
520 overlapping confidence intervals.
521
522 **State-dependent diversification**

523 The BAMM analysis inferred 18 rate shifts in Fabales and 11 of these shifts are nested
524 within Papilionoideae (Fig. 2a, c). The diversification rate increased gradually over time in
525 Fabales and nearly all rate shifts represented a significant acceleration in diversification
526 rates rather than a decrease (gray lines in Fig. 2b). With few exceptions, 9 out of the 11 rate
527 shifts in Papilionoideae took place during the Miocene, which marked a time of global
528 cooling and aridification (Herbert *et al.*, 2016).
529
530 Although most rate shifts took place in the keel-flowered Papilionoideae, our
531 diversification rate analysis did not support the SSE models. Among the four state-
532 dependent (BiSSE and HiSSE) and state-independent models (CID-2 and CID-4), CID-4 was
533 significantly better than the second-best HiSSE model with a ΔAIC of 168.772. However,
534 diversification parameters from both BiSSE and HiSSE models suggested an orders-of-
535 magnitude higher net diversification rate in keel-flowered clades than non-keel-flowered
536 clades. Specifically, HiSSE suggested the net diversification rate to be 0.13 in keel-flowered
537 clades and 5.5e-4 in non-keel-flowered clades; BiSSE suggested it to be 0.084 in keel-
538 flowered clades and 8.1e-04 in non-keel-flowered clades.
539
540 **DISCUSSION**
541 **Insights into the early divergence relationships within papilionoids**
542 Previous phylogenetic studies focusing on the early divergences of papilionoid legumes
543 have largely utilized plastid markers (Wojciechowski *et al.*, 2004; Lavin *et al.*, 2005;
544 McMahon & Sanderson, 2006; Cardoso *et al.*, 2012, 2013b; Zhang *et al.*, 2020; Choi *et al.*,
545 2022). These studies have consistently recognized well-supported clades but struggled in
546 resolving inter-clade relationships. Based on 1,456 low-copy nuclear genes, we show that
547 the ADA, Swartzieae, and Cladrastis clades are successive sister groups to the 50-kb
548 inversion clade (Fig. 1). This result is congruent with the recent nuclear phylogeny from
549 Zhao *et al.* (2021) but differs from most plastid-based phylogenies where the Swartzieae
550 clade was placed as sister to all other Papilionoideae (Cardoso *et al.*, 2012, 2013b; Zhang *et*
551 *al.*, 2020; Choi *et al.*, 2022).

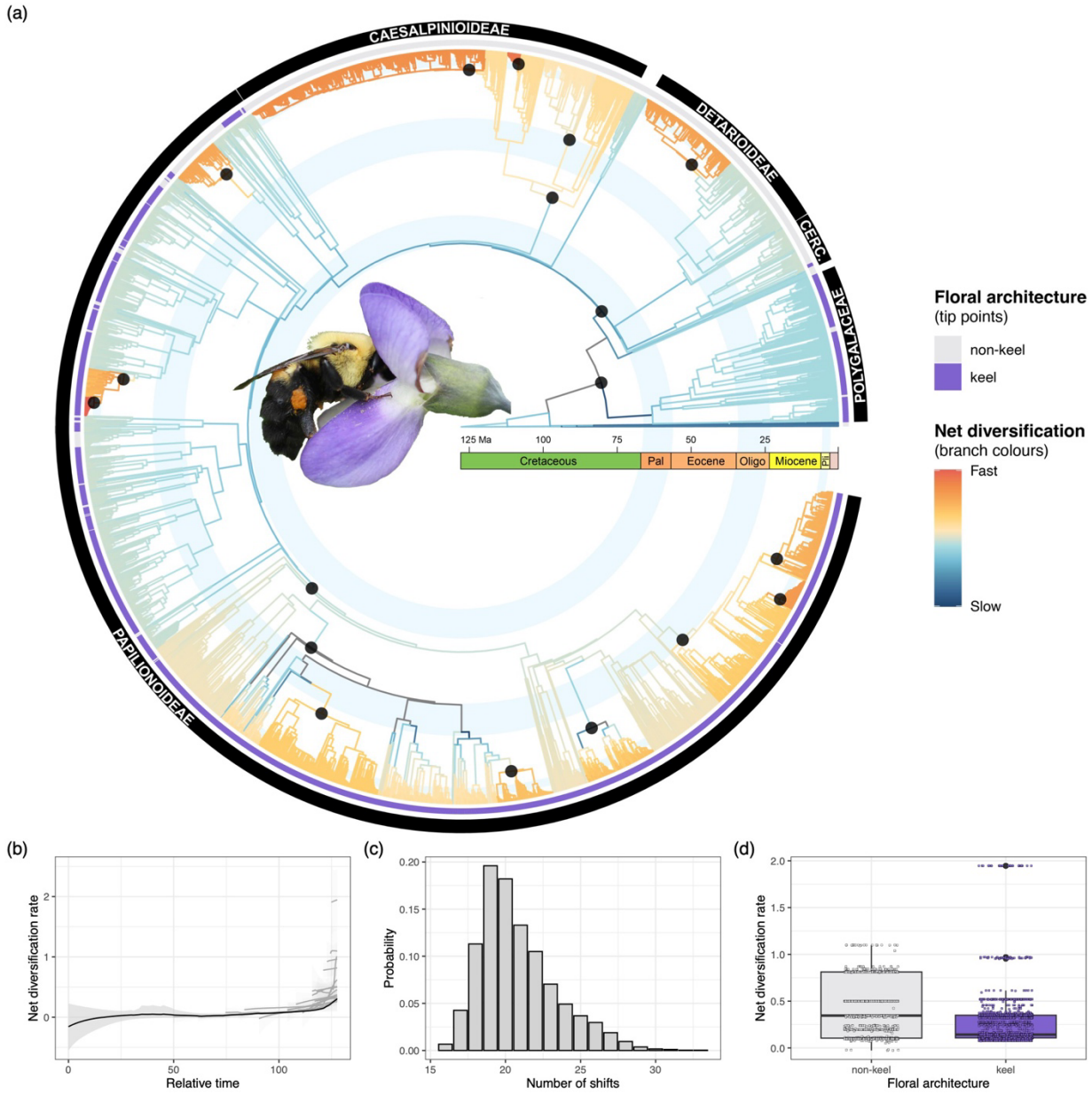


Figure 2. Rate heterogeneity in the diversification of Fabales. (a) Time-calibrated tree of Fabales species based on plastid *matK* gene sequences. Branch colors show net diversification rates inferred from BAMM and circles at nodes refer to clades with significant rate shifts. Purple lines in the outer ring indicate the presence of keel flowers. (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. 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* bee on a *Baptisia australis* flower.

565

566 The 50-kb inversion clade constitutes 95% of the remaining species diversity in
567 Papilionoideae. Our well-resolved nuclear phylogeny provides several important insights
568 into relationships within this clade that were inconclusive in previous plastid sequence-
569 based studies. First, the Exostyleae and Vataireoid clades form a monophyletic group sister
570 to the remaining members of the 50-kb inversion clade (100 BP; 1.0 PP). Their close
571 affinity is morphologically supported by the denticulate, serrate, or occasionally spinescent
572 leaf margin, which is highly unusual outside the NPAAA clade in Papilionoideae (Mansano
573 *et al.*, 2004; Cardoso *et al.*, 2012, 2013a,b). Second, although conflicting topologies are
574 obtained when placing the Genistoid s.l. clade, statistically significant support is only
575 achieved for the coalescent analysis (Table S3). Thus, based on the ASTRAL tree, the
576 Genistoid s.l. clade is sister to a clade comprised of the Dalbergioid s.l., Andira, Baphieae,
577 and NPAAA clades (PP = 0.98). Notably, the Dalbergioid s.l., Andira, and Baphieae clades
578 form successive sister groups to the large NPAAA clade — a relationship that has not been
579 observed in any plastid sequence-based study (Cardoso *et al.*, 2012, 2013b; Choi *et al.*,
580 2022).
581

582 Within the Genistoid s.l. clade, the Ormosieae, Brongniartieae, and Leptolobieae form
583 successive sister groups to all remaining species. Particularly, the genus *Dermatophyllum,*
584 comprised of four species, is a well-supported sister lineage to tribe Brongniartieae (88 BP;
585 1.0 PP). In previous studies, *Dermatophyllum* often formed a polytomy with the Dalbergioid
586 s.l., Andira, and Genistoid s.l. clades (Wojciechowski et al., 2004; Cardoso et al., 2012,
587 Cardoso et al., 2013). The close affinity between *Dermatophyllum* and Brongniartieae is
588 supported by their shared niche preference and the ability to synthesize a class of highly
589 distinct alkaloids. The preferred habitat of *Dermatophyllum* in warm temperate woodlands,
590 chaparral, and deserts of southern United States to central Mexico overlaps with many
591 Mexican Brongniartieae especially in the genus *Brongniartia* (Pennington *et al.*, 2005; Ross
592 & Crisp, 2005). The presence of quinolizidine alkaloids in *Dermatophyllum* further supports
593 its placement within the Genistoid s.l. clade (Cardoso *et al.*, 2013; Choi *et al.*, 2022). This
594 class of alkaloids is absent in all other legumes and only sparsely distributed in eight other
595 vascular plant families (Bisby *et al.*, 1994; Kite & Pennington, 2003; Lee *et al.*, 2013; Wink,

596 2013; Cely-Veloza *et al.*, 2023). Therefore, our new result provides compelling evidence for
597 a single origin of alkaloid biosynthesis in papilionoids and a chemical synapomorphy for
598 the Genistoid s.l. clade.
599
600 In addition to *Dermatophyllum*, the close relationship between the African monotypic genus
601 *Haplormosia* and the American-Australian Brongniartieae clade proposed by (Cardoso *et*
602 *al.*, 2017) is also challenged by our new result. Although several reproductive and
603 vegetative features are shared between these two lineages — including unifoliolate leaves;
604 single-seeded, tardily dehiscent fruits with thick, woody valves; and stamens with distinctly
605 dimorphic anthers (Cardoso *et al.*, 2017) — the distinct biogeographic distribution makes
606 the proposed relationship contentious. Our results instead place *Haplormosia* within a
607 group of largely Old-World lineages, including *Camoensia*, Crotalarieae, Genisteae,
608 Podalyrieae, and Sophoreae (98 BP; 0.90 PP). Morphologically, species from Crotalarieae,
609 Podalyrieae, and Genistieae also exhibit unifoliolate leaves and dimorphic anthers that are
610 alternately basifixed and dorsifixed (Schutte-Vlok & van Wyk, 2011; Le Roux & Van Wyk,
611 2012; du Preez *et al.*, 2021).
612
613 Finally, the NPAAA clade contains a whopping ca. 10,200 species that account for nearly
614 70% of species level diversity in papilionoids. It is sister to the predominantly African
615 Baphieae clade with ca. 60 species (100 BP; 1.0 PP). The NPAAA clade comprises two large,
616 strongly-supported subclades, the Hologalegina (Robinioid+IR-lacking clade) and
617 Indigofereae + Millettioid-Phaseoloid clades, in agreement with all previous molecular
618 phylogenetic studies (Wojciechowski *et al.*, 2004; Cardoso *et al.*, 2013b; LPWG, 2017; Zhao
619 *et al.*, 2021). Within Hologalegina (ca. 5,300 spp.), the pantropical genus *Sesbania* was
620 placed within the Robinioid clade. In plastid-based studies, the relationship between
621 *Sesbania*, and tribes Loteae and Robinieae is highly unstable with all combinations of
622 relationships recovered (Wojciechowski *et al.*, 2004; Farruggia *et al.*, 2018; Zhang *et al.*,
623 2020; Choi *et al.*, 2022). Our results confirm *Sesbania* as sister to the warm temperate,
624 primarily northern hemisphere Loteae (100 BP; 1.0 PP), and together they form a sister
625 group to Robinieae (100 BP; 1.0 PP). The paraphyletic relationship among *Sesbania,* Loteae,
626 and Robinieae proposed by Zhao *et al.* (2021) is strongly rejected by our AU test (*p*-value <

627 0.01) and MSC likelihood ratio test (*p*–value < 0.01). Relationships within the IR-lacking
628 clade (IRLC) align with recent studies using plastid and nuclear sequences (Choi *et al.*,
629 2022; Zhao *et al.*, 2021).
630
631 Relationships within the primarily subtropical to pantropical Indigofereae + Millettioid-
632 Phaseoloid clade (ca. 4,050 spp.) are consistent with earlier findings (Hu *et al.*, 2000;
633 Wojciechowski *et al.*, 2004; Stefanović *et al.*, 2009; LPWG, 2017; Oyebanji *et al.*, 2020; Zhao
634 *et al.*, 2021; Choi *et al.*, 2022). In the Millettieae s.l. and Phaseoleae s.l. (Schrire, 2005a,b),
635 the "core Millettieae" (Hu *et al.*, 2002) and the three traditional tribes Abreae, Desmodieae,
636 and Psoraleeae (Lewis *et al.*, 2005) are each well supported monophyletic lineages. Tribes
637 Millettieae and Phaseoleae are polyphyletic with unresolved generic-level relationships.
638
639 **Loss and gain of keel flowers during the diversification of the papilionoids**
640 A suite of reproductive, nutritional, and defense traits has contributed to the assemblage of
641 present legume biodiversity (Lewis *et al.*, 2005). Here we focused on the pollinator-driven
642 diversification in Fabales, which has been widely invoked to explain the radiation of many
643 angiosperms (Sargent, 2004; Kay & Sargent, 2009; Givnish *et al.*, 2015). Bilaterally
644 symmetrical flowers such as a keel flower are strongly associated with bees by restricting
645 the entry to and the precise placement of pollen to discourage gene flow between incipient
646 species (Neal *et al.*, 1998). Despite being a stereotyped search image for legumes, there is
647 great lability in floral form and symmetry during the early evolution history of papilionoids
648 (Fig. 2). This provides an ideal comparative framework to test whether the evolution of
649 keel flowers drives diversification (Ricklefs & Renner, 1994, 2000).
650
651 Our well-resolved nuclear gene and densely-sampled plastid *matK* gene phylogenies
652 provide the essential evolutionary framework to address this question. Crown group
653 Fabaceae was estimated to originate at the K-Pg boundary at 66.0 Ma (95% HPD 64.7–66.8
654 Ma), which is in line with previous molecular-based estimates and fossil records (Lavin *et*
655 *al.*, 2005; Bruneau *et al.*, 2008; Cannon *et al.*, 2015; Koenen *et al.*, 2021; Zhao *et al.*, 2021).
656 The ancestral floral architecture of Papilionoideae is estimated to be non-keel-like, with the
657 first keel flowers in Fabaceae emerging in the common ancestor of the Cladrastis and the

658 50-kb inversion clades at 59.0 Ma (95% HPD 57.9–60.2 Ma). This estimation predates the
659 earliest papilionoid-like flower fossil *Barnebyanthus buchananensis* by 3 million years
660 (Herendeen & Crepet, 1992) and closely succeeded the evolution of keel flowers in crown
661 group Polygalaceae at 62.8 Ma (95% HPD 59.6–64.9 Ma). However, these estimates
662 contradict the results of a recent study which concluded that papilionoid keel flowers
663 evolved 22.03 to 10.32 Ma before they occurred in Polygalaceae (Uluer *et al.*, 2022).
664
665 Such a discrepancy comes mainly from the ancestral state reconstruction analysis of keel
666 flowers themselves, but also from minor differences in divergence time estimation due to
667 taxon sampling and methodology. Unlike the crown group Papilionoideae origination
668 proposed by Uluer *et al.* (2022), our results suggest a more dynamic evolutionary history of
669 keel flowers with five independent origins that postdate crown group Papilionoideae
670 (Table S9; Fig. S5). Four of these events are nested within the ADA clade, taking place in the
671 Miocene of tropical Africa (e.g., *Angylocalyx*) and America (e.g., *Myrospermum*). The two
672 oldest keel flower lineages originated during the Selandian age of the Paleocene in
673 Papilionoideae and Polygalaceae, though the exact order cannot be confidently established.
674 However, the short interval between these two old events and multiple more recent origins
675 of keel flowers in the ADA clades may all result from the same mutualistic process where
676 younger lineages benefitted from and contributed to the strong relationship with
677 pollinators already established by older keel lineages (Tucker, 2002; Bello *et al.*, 2012;
678 Uluer *et al.*, 2022).
679
680 In addition to gains, perhaps what is more intriguing is the numerous losses of keel flowers.
681 We identified more than two dozen independent losses in Papilionoideae alone (Table S9).
682 Only one out of these 27 losses is nested within the NPAAA clade in genus *Erythrina* L.,
683 even though the NPAAA clade accounts for more than 70% of the Fabaceae species. These
684 results suggest that ancestral Fabaceae likely took a generalist pollination strategy,
685 allowing floral symmetry to be more labile. This is exemplified in multiple early branching
686 papilionoids with lost, reduced, or poorly differentiated petals. These modifications are
687 accompanied by expansion or shifts of pollinators to bats, passerine birds, hummingbirds,
688 and even beetles (e.g., Arroyo, 1981; Pennington *et al.*, 2000; Lewis *et al.*, 2003). In

689 contrast, derived lineages may have experienced stronger stabilizing selection to maintain
690 specialized floral architecture for coevolved pollinators. The NPAAA clade has diversified
691 largely during the Miocene under global cooling, aridification, and expansion of grasslands
692 (Estep *et al.*, 2014; Herbert *et al.*, 2016). The IRLC, Indigofereae, and Robinioid clades are
693 thus strongly associated with herbaceous growth habit in arid to semi-arid temperate
694 regions (Wojciechowski *et al.*, 2004; Schrire *et al.*, 2005, 2009). Adaptation to these
695 habitats may cast additional geographical, environmental, and developmental constraints
696 to limit the breadth of viable pollinators, making it advantageous to maintain ancestral
697 states rather than evolve new ones.
698
699 Finally, although the HiSSE model supports a higher net diversification rate in keel-
700 flowered lineages, the best-supported model was one where the diversification rate was
701 independent of floral architecture. The orders-of-magnitude difference in net
702 diversification rate estimated by the HiSSE model is driven by handful keel-flowered
703 lineages (Fig. 2d). Not all keel-flowered Fabaceae radiated extensively (e.g., *Cercis*), and
704 *Swartzia* (Swartzieae) diversified rapidly without keel flowers. Along these lines, Afkhami
705 *et al.* (2018) also reported a lack of correlation between the nodulation ability and net
706 diversification rate despite that nodulating legume genera contain three times more
707 species than non-nodulating ones. The lack of support for SSE models suggests that
708 individual traits including keel flowers and nodulation are not the sole driver of
709 diversification, but instead are one of several diversification-promoting traits within a
710 broader confluence of factors that have additively contributed to lineage diversification
711 (Donoghue & Sanderson, 2015). Besides innovations of key traits, recurrent biome shifts
712 and rapid colonization of novel ecological niches also played a significant role in promoting
713 diversification in Fabaceae. Examples include the rapid radiation of the primarily
714 herbaceous IRLC clade in temperate regions (Wojciechowski *et al.*, 2004), the dominance of
715 mimosoids in seasonally dry forests (Pennington *et al.*, 2009; Ringelberg *et al.*, 2023), and
716 the remarkable diversification of Indigofereae in tropical fire-prone savannas (Schrire *et*
717 *al.,* 2009). In summary, while pollinator interactions have played a role in the
718 diversification of papilionoid legumes, a combination of intrinsic factors (e.g., chemical
719 defenses and nodulation) and extrinsic factors (e.g., biome shifts and ecological

720 specialization) were synergistically involved to drive the remarkable diversification of
721 Papilionoideae.
722
723 **CONCLUSIONS**
724 Drawing insights from phylogenomics, divergence time estimation, and comparative
725 methods, we present novel insights into the early divergence history and floral evolution of
726 papilionoid legumes. With 1,456 low-copy nuclear loci, we confidently resolved the
727 recalcitrant relationships in Papilionoideae and inferred a Paleocene origin of keel flowers
728 at 59 Ma, predating the earliest fossil record by 3–4 million years. We identified at least
729 seven independent origins and 27 losses of keel flowers across Fabales, with losses
730 predominantly taking place in early diverging papilionoid lineages. Ancestral Fabaceae
731 thus likely displayed a more generalist pollination strategy, allowing for greater floral
732 flexibility over time. Although these independent gains and losses coincide with the
733 Miocene diversification of Papilionoideae, our results do not support floral architecture as
734 the sole driver of diversification in Fabaceae. Instead, a confluence of intrinsic and extrinsic
735 factors including nitrogen fixation, chemical defenses, and recurrent dispersal into new
736 biomes such as subtropical and temperate regions, likely contributed synergistically to the
737 diversification of papilionoids.
738
739 **ACKNOWLEDGEMENTS**
740 This work was supported by grants from the National Science Foundation to MFW (DEB-
741 1853010) and RKJ, TAR (DEB-1853024), the Texas Ecological Laboratory Program
742 (EcoLab) to RKJ, TAR, I-SC, and LC, the Sidney F. and Doris Blake Professorship in
743 Systematic Botany to RKJ, the Stengl Wyer Fellowship to LC, the CNPq (Research
744 Productivity Fellowship no. 314187/2021-9), and FAPERJ (Programa Jovem Cientista do
745 Nosso Estado - 2022, grant no. 200.153/2023) to DC, the National Research Foundation of
746 Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00341022) to I-SC,
747 and funds from the College of Liberal Arts and Sciences at Arizona State University to MFW.
748 We thank the TEX-LL, HUEFS, and RB herbaria for voucher deposition, the Desert Legume
749 Program at the University of Arizona for seeds; Alessandra Schnadelbach, Henrique Batalha
750 Filho, and Paula Ristow for permitting us to use their labs at UFBA; and George Yatskievych

751 (TEX/LL) for arranging a formal Material Transfer Agreement (Decree number 8772)
752 under the SisGen Cadastro RDC6BE9, which facilitated research activities between the
753 institutions.
754
755 **SUPPLEMENTARY INFORMATION**
756 **Supplementary figures**
757 **Figure S1** Histogram of the number of species per loci. The vertical line represents the
758 30% missing data threshold to select the 1,456 loci in this study.
759
760 **Figure S2** Topological congruences and conflicts of Papilionoideae phylogeny inferred
761 from IQTREE (left) and ASTRAL (right). Tangled grey lines indicate conflicting
762 relationships. Branch supports represent nonparametric bootstrap replicates values from
763 IQ-TREE and local posterior probability from ASTRAL, respectively. Only branches that are
764 not maximally supported are labeled.
765
766 **Figure S3** Quantification of phylogenetic signal from individual genes. The two alternative
767 placements of the focal clades are illustrated with red or blue, with the red topology being
768 favored by the concatenation analysis. Genes are ordered in descending order based on
769 their support for the red topology. Differences in the log-likelihood scores (ΔlogL) between
770 these two topologies are calculated for individual genes (left) or the concatenated matrix
771 (right). Panels A-E represent five contentious relationships within Papilionoideae defined
772 in Table S3.
773
774 **Figure S4** Bayesian inference of divergence time in Fabaceae based on 30 top clock-like
775 loci. Divergence time is shown in millions of years ago (Ma) and the blue bars represent
776 95% credibility intervals. Eleven fossil calibration points including a constraint for the root
777 age are marked by numbered stars. See Table S5 for a full list of fossils and references.
778
779 **Figure S5** Ancestral character reconstruction of keel flowers in Fabales inferred under the
780 marginal reconstruction algorithm in HiSSE. An ultrametric tree of 3,326 Fabales species is
781 inferred from matK sequences with branch lengths drawn in proportion to divergence

782 time. Solid and hollow branches indicate the presence and absence of keel flowers,
783 respectively. Events of keel flower gains and losses are marked by orange and magenta
784 circles on the branch, respectively. See Table S9 for age estimations of these events.
785
786 **Supplementary Tables**
787 **Table S1** Voucher information and gene assembly information of 79 newly sequenced
788 species in this study. Species name, lineage affiliation, assembly quality, SRA accession
789 number, and voucher information including collection locality are provided.
790
791 **Table S2** Species name, lineage affiliation, and number of sampled genes from 208 species
792 from Zhao et al. (2021).
793
794 **Table S3** Testing alternative phylogenetic placement of Papilionoideae clades under the
795 maximum likelihood framework for concatenated DNA sequences.
796
797 **Table S4** List of top 30 clock-like genes with quantified root-to-tip branch length variation,
798 total tree length, and topological similarity to the species tree measured by bipartition
799 score.
800
801 **Table S5** Fossil calibration points for divergence time estimation.
802
803 **Table S6** Time tree, Genbank ID, and keel flower trait status for the 3,326 Fabales species
804 from the *matK* dataset. Numbers on the phylogeny are divergence time estimation for each
805 node based on treePL.
806
807 **Table S7** Secondary calibration points for the treePL analysis. All node ages in million
808 years ago (Ma) are derived from the BEAST time tree and are fixed in treePL. Nodes are
809 defined by the most recent common ancestor of two species indicated in the spreadsheet.
810

811 **Table S8** Nuclear genome sequencing coverage of 32 whole genome sequencing datasets
812 estimated based on 21-mer frequency distribution. See Table S1 for the species name of
813 each ID.
814
815 **Table S9** Summary of independent origins and losses of keel flower in Fabales. Binary
816 ancestral state reconstruction is inferred under the marginal reconstruction algorithm in
817 HiSSE. The timing of these events is determined by the crown group age of involved clades.
818 The precise age of events involving a single species cannot be determined using available
819 data. Therefore the age of these events is represented by a range from present to their stem
820 group age.
821
822 **Supplementary Data**
823 **Data S1** DNA alignments and maximum likelihood phylogenies inferred for the 1,456 low-
824 copy nuclear genes used in this study. The alignments are inferred using the MAFFT-linsi
825 algorithm. The gene trees are reconstructed using IQTREE and branch support is evaluated
826 with 1000 ultrafast bootstrap replication. Species names and their corresponding ID are
827 provided in sp_ID.csv. The list of 672 loci used to generate the concatenated matrix for
828 species tree estimation is provided in G672_len900_sp227.list.
829
830 **Data S2** Distribution of 21-mer frequency in 32 species with genome skimming data. The
831 21-mer frequency is inferred from Jellyfish v2.3.0 to estimate sequencing coverage.
832
833 **Data S3** Gene and site concordance factors of the Papilionoideae phylogeny. The species
834 tree and concatenated alignment of the 672-gene dataset were used as inputs in IQ-TREE to
835 calculate these concordance factors.
836
837 **COMPETING INTERESTS**
838 None declared.
839
840 **AUTHOR CONTRIBUTIONS**

841 Conception and experimental design, LC, DC, I-SC, RKJ, TAR, MFW. Acquisition of funds, RKJ,
842 TAR, MFW. Field collections, I-SC, DC, HCL, LPQ, RKJ, TAR, MFW. Nucleic acid isolation, I-SC,
843 CL, TAR. RNA isolations, TAR. Transcriptome and genome assembly, CL, LT, BS, LC. DNA
844 alignments and phylogenetic analyses, LC, I-SC, DC, MFW. Data interpretation, LC, I-SC, DC,
845 MFW. Production of figures and tables, LC, DC. Writing the initial draft of the manuscript,
846 LC, DC, MFW, TAR, RKJ. All authors revised and approved the final version of the
847 manuscript.
848

849 **DATA AVAILABILITY**

850 The raw sequence data generated and analyzed during this study are available in the NCBI
851 GenBank repository under Bioproject PRJNA1120003. The alignment data supporting the
852 findings of this study are available as supplementary data and have been deposited in the
853 Figshare Digital Repository (DOI: 10.6084/m9.figshare.27135984). The code used for data
854 processing and analysis is openly available on GitHub at
855 https://github.com/lmcai/Papilionoideae_phylogenomics.
856

857 **REFERENCES**

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

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

864 **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM,**
865 **Nikolenko SI, Pham S, Prjibelski AD, *et al.* 2012**. SPAdes: a new genome assembly
866 algorithm and its applications to single-cell sequencing. *Journal of Computational Biology*
867 **19**: 455–477.

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

871 **Beaulieu JM, O'Meara BC**. **2016**. Detecting hidden diversification shifts in models of trait-
872 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, 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.

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, 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.

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, Paganucci Queiroz L, Cavalcante Lima H, 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

909 phylogeny: The enigmatic Amazonian genus *Petaladenium* is a new branch of the early-
910 diverging Amburaneae clade. *Molecular Phylogenetics and Evolution* **84**: 112–124.

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

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

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

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

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

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

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

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

934 **Farruggia FT, Lavin M, Wojciechowski MF. 2018**. Phylogenetic systematics and
935 biogeography of the pantropical genus *Sesbania* (Leguminosae). *Systematic Botany* **43**:
936 414–429.

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

938 **Givnish TJ, Spalink D, Ames M, Lyon SP, Hunter SJ, Zuluaga A, Iles WJ, Clements MA,**
939 **Arroyo MT, Leebens-Mack J. 2015**. Orchid phylogenomics and multiple drivers of their
940 extraordinary diversification. *Proceedings of the Royal Society B: Biological Sciences* **282**:
941 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.

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, Crepet WL. 1992. Paleobotanical and biogeographic history of the legumes (Leguminosae), an important component in Cenozoic and modern tropical terrestrial ecosystems. *The Paleontological Society Special Publications* **6**: 127–127.

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

Johnson SD, Schiestl FP. 2016. *Floral Mimicry*. Oxford, New York: Oxford University Press.

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.

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

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

981 **Kursar TA, Dexter KG, Lokvam J, Pennington RT, Richardson JE, Weber MG,**
982 **Murakami ET, Drake C, McGregor R, Coley PD. 2009**. The evolution of antiherbivore
983 defenses and their contribution to species coexistence in the tropical tree genus *Inga*.
984 *Proceedings of the National Academy of Sciences of the United States of America* **106**: 18073–
985 18078.

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

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

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

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

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

999 **Lee ST, Cook D, Molyneux RJ, Davis TZ, Gardner DR. 2013**. Alkaloid profiles of
1000 *Dermatophyllum arizonicum*, *Dermatophyllum gypsophilum*, *Dermatophyllum secundiflorum*,
1001 *Styphnolobium affine*, and *Styphnolobium japonicum* previously classified as *Sophora*
1002 species. *Biochemical Systematics and Ecology* **49**: 87–93.

1003 **Lewis GP, Knudsen JT, Klitgaard BB, Pennington RT. 2003**. The floral scent of
1004 *Cyathostegia mathewsii* (Leguminosae, Papilionoideae) and preliminary observations on
1005 reproductive biology. *Biochemical Systematics and Ecology* **31**: 951–962.

1006 **Lewis G, Schrire B, Mackinder B, Lock M. 2005**. *Legumes of the World*. Richmond, UK:
1007 Royal Botanic Gardens, Kew.

1008 **Liu L, Anderson C, Pearl D, Edwards SV. 2019**. Modern phylogenomics: building
1009 phylogenetic trees using the multispecies coalescent model. In: Evolutionary Genomics.
1010 Springer, 211–239.

1011 **Liu L, Yu L. 2010**. Phybase: an R package for species tree analysis. *Bioinformatics* **26**: 962–
1012 963.

1013 **Liu L, Yu L, Edwards SV. 2010**. A maximum pseudo-likelihood approach for estimating
1014 species trees under the coalescent model. *BMC Evolutionary Biology* **10**: 302.

1015 **LPWG. 2013**. Legume phylogeny and classification in the 21st century: progress, prospects
1016 and lessons for other species–rich clades. *Taxon* **62**: 217–248.

1017 **LPWG. 2017**. A new subfamily classification of the Leguminosae based on a taxonomically
1018 comprehensive phylogeny. *Taxon* **66**: 44–77.

1019 **Lyson TR, Miller IM, Bercovici AD, Weissenburger K, Fuentes AJ, Clyde WC, Hagadorn**
1020 **JW, Butrim MJ, Johnson KR, Fleming RF, *et al.* 2019**. Exceptional continental record of
1021 biotic recovery after the Cretaceous–Paleogene mass extinction. *Science* **366**: 977–983.

1022 **Maddison WP, Midford PE, Otto SP. 2007**. Estimating a binary character's effect on
1023 speciation and extinction. *Systematic Biology* **56**: 701–710.

1024 **Magallón S, Gómez-Acevedo S, Sánchez-Reyes LL, Hernández-Hernández T. 2015**. A
1025 metacalibrated time-tree documents the early rise of flowering plant phylogenetic
1026 diversity. *New Phytologist* **207**: 437–453.

1027 **Mansano VF, Bittrich V, Azevedo Tozzi AMG, Souza AP. 2004**. Composition of the
1028 Lecointea clade (Leguminosae, Papilionoideae, Swartzieae), a re-evaluation based on
1029 combined evidence from morphology and molecular data. *Taxon* **53**: 1007–1018.

1030 **Marazzi B, Gonzalez AM, Delgado-Salinas A, Luckow MA, Ringelberg JJ, Hughes CE.**
1031 **2019**. Extrafloral nectaries in Leguminosae: phylogenetic distribution, morphological
1032 diversity and evolution. *Australian Systematic Botany* **32**: 409–458.

1033 **Marazzi B, Sanderson MJ. 2010**. Large-scale patterns of diversification in the widespread
1034 legume genus Senna and the evolutionary role of extrafloral nectaries. *Evolution* **64**: 3570–
1035 3592.

1036 **Marçais G, Kingsford C. 2011**. A fast, lock-free approach for efficient parallel counting of
1037 occurrences of k-mers. *Bioinformatics* **27**: 764–770.

1038 **Maron JL, Agrawal AA, Schemske DW. 2019**. Plant–herbivore coevolution and plant
1039 speciation. *Ecology* **100**: e02704.

1040 **McMahon MM, Sanderson MJ. 2006**. Phylogenetic supermatrix analysis of GenBank
1041 sequences from 2228 papilionoid legumes. *Systematic Biology* **55**: 818–836.

1042 **Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,**
1043 **Lanfear R. 2020**. IQ-TREE 2: New models and efficient methods for phylogenetic inference
1044 in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.

1045 **Morales-Briones DF, Kadereit G, Tefarikis DT, Moore MJ, Smith SA, Brockington SF,**
1046 **Timoneda A, Yim WC, Cushman JC, Yang Y. 2021**. Disentangling sources of gene tree
1047 discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l.
1048 *Systematic Biology* **70**: 219–235.

1049 **Muchhala N. 2006**. The pollination biology of *Burmeistera* (Campanulaceae): specialization
1050 and syndromes. *American Journal of Botany* **93**: 1081–1089.

1051 **Neal PR, Dafni A, Giurfa M. 1998**. Floral symmetry and its role in plant-pollinator
1052 systems: terminology, distribution, and hypotheses. *Annual Review of Ecology, Evolution,*
1053 *and Systematics* **29**: 345–373.

1054 **Oyebanji O, Zhang R, Chen S-Y, Yi T-S. 2020**. New insights into the plastome evolution of
1055 the Millettioid/Phaseoloid Clade (Papilionoideae, Leguminosae). *Frontiers in Plant Science*
1056 **11**.

1057 **Padfield D, Kay S, Vos R, Quince C, Vos M. 2024**. Macroevolutionary dynamics in micro-
1058 organisms: generalists give rise to specialists across biomes in the ubiquitous bacterial
1059 phylum Myxococcota. *Molecular Biology and Evolution* **41**: msae088.

1060 **Pennington R, Klitgaard B, Ireland H, Lavin M. 2000**. New insights into floral evolution
1061 of basal Papilionoideae from molecular phylogenies. In: P.S. Herendeen and A. Bruneau
1062 (Eds). Advances in Legume Systematics 9. Royal Botanic Gardens, Kew. In: 233–248.

1063 **Pennington RT, Lavin M, Oliveira-Filho A. 2009**. Woody plant diversity, evolution, and
1064 ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of*
1065 *Ecology, Evolution, and Systematics* **40**: 437–457.

1066 **Pennington RT, Stirton CH, Schrire BD. 2005**. Tribe Sophoreae. In: Lewis, G., Schrire, B.,
1067 Mackinder, B., Lock,M. (Eds.), Legumes of the World. In: Richmond, UK: Royal Botanic
1068 Gardens, Kew, 227–249.

1069 **Polhill RM, Raven PH, Stirton CH. 1981**. Evolution and systematics of the Leguminosae.
1070 In: Polhill, R.M. and Raven, P.H., Eds., Advances in Legume Systematics, Part I. Richmond,
1071 UK: Royal Botanic Garden, Kew, 1–26.

1072 **Preez B, Dreyer LL, Stirton CH, Muasya AM. 2021**. A monograph of the genus *Polhillia*
1073 (Genisteae: Fabaceae). *South African Journal of Botany* **138**: 156–183.

1074 **Rabosky DL. 2014**. Automatic detection of key innovations, rate shifts, and diversity-
1075 dependence on phylogenetic trees. *PLOS ONE* **9**: e89543.

1076 **Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG.**
1077 **2014**. BAMM tools: an R package for the analysis of evolutionary dynamics on phylogenetic
1078 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.

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.

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 B, Lavin M, Lewis G, Friis I, Balslev H. 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies: 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.

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.

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

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

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.

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.

1146 **Wojciechowski MF, Lavin M, Sanderson MJ**. **2004**. A phylogeny of legumes
1147 (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported
1148 subclades within the family. *American Journal of Botany* **91**: 1846–1862.

1149 **Yang Y, Smith SA**. **2014**. Orthology inference in nonmodel organisms using transcriptomes
1150 and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics.
1151 *Molecular Biology and Evolution* **31**: 3081–3092.

1152 **Zhang J, Ruhlman TA, Mower JP, Jansen RK**. **2013**. Comparative analyses of two
1153 Geraniaceae transcriptomes using next-generation sequencing. *BMC Plant Biology* **13**: 228.

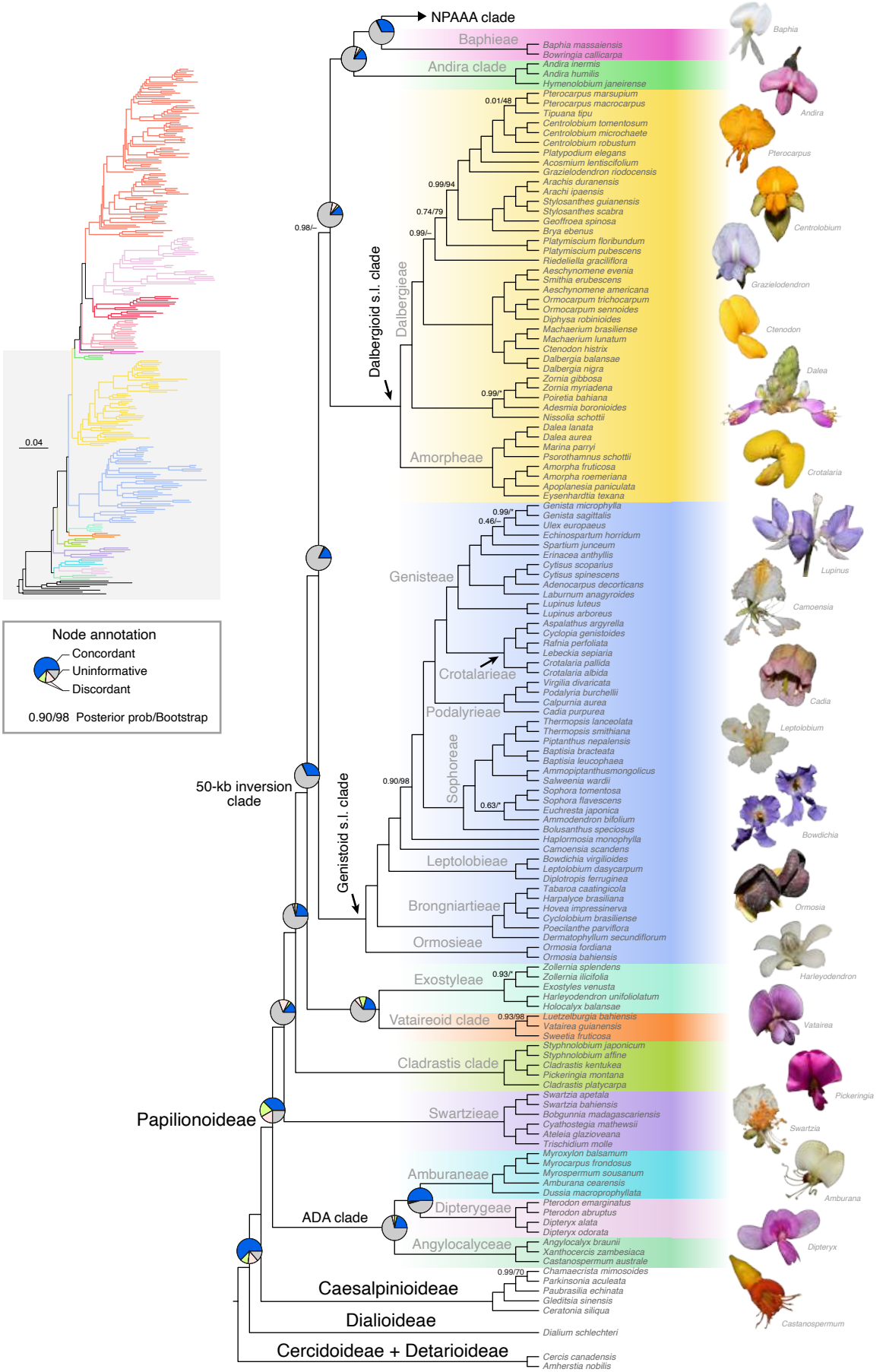
1154 **Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018**. ASTRAL-III: polynomial time species tree
1155 reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

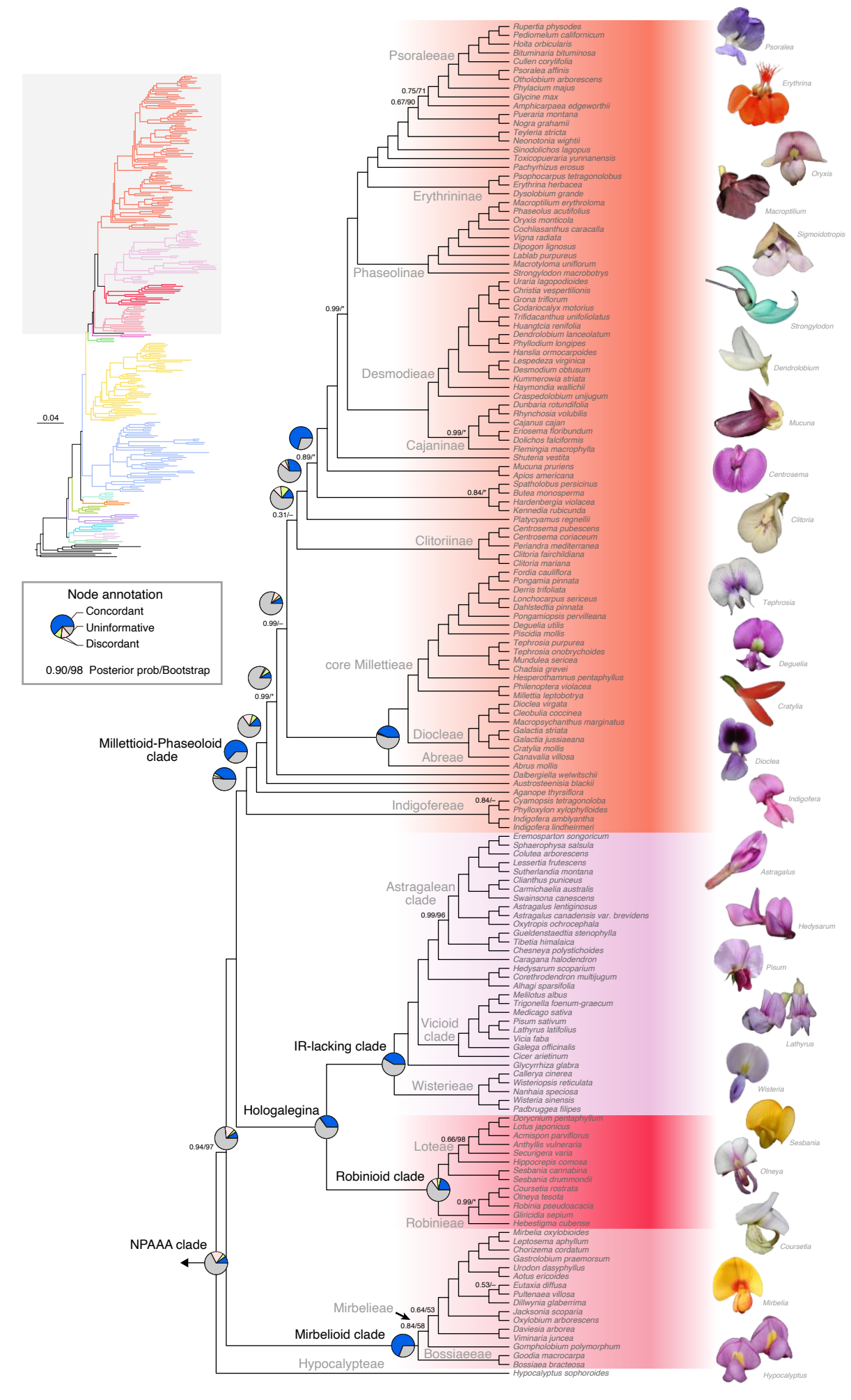
1156 **Zhang R, Wang Y-H, Jin J-J, Stull GW, Bruneau A, Cardoso D, Queiroz LP, Moore MJ,**
1157 **Zhang SD, Chen SY**. **2020**. Exploration of plastid phylogenomic conflict yields new insights
1158 into the deep relationships of Leguminosae. *Systematic Biology* **69**: 613–622.

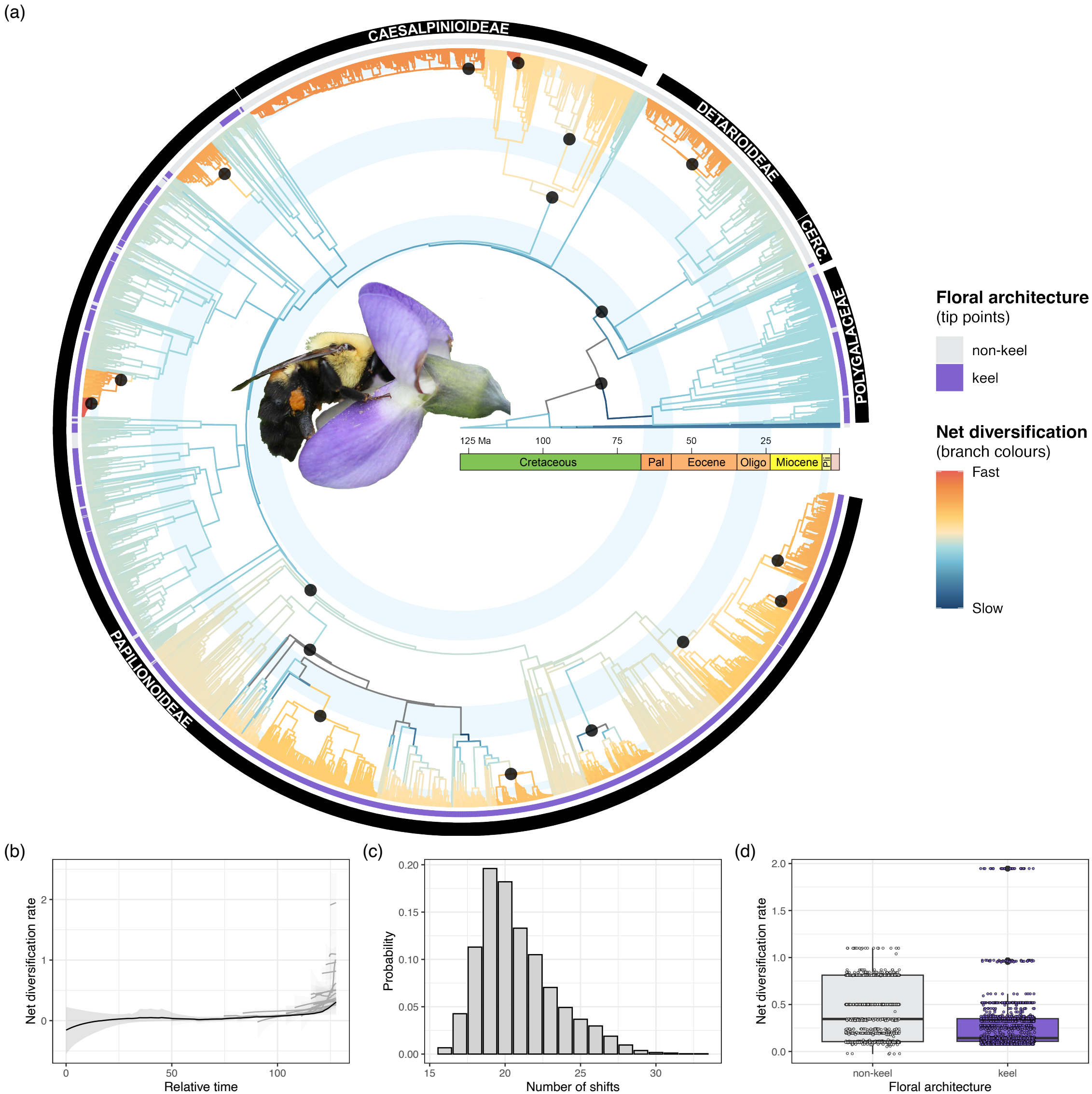
1159 **Zhao Y, Zhang R, Jiang K-W, Qi J, Hu Y, Guo J, Zhu R, Zhang T, Egan AN, Yi T-S, *et al.***
1160 **2021**. Nuclear phylotranscriptomics and phylogenomics support numerous
1161 polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing
1162 symbiosis in Fabaceae. *Molecular Plant* **14**: 748–773.

1163 **Zuanny DC, Vilela B, Moonlight PW, Särkinen TE, Cardoso D**. **2024**. expowo: an R
1164 package for mining global plant diversity and distribution data. *Applications in Plant*
1165 *Sciences*: e11609.

1166


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 riodocensis
Arachis duranensis
Arachi 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
Erinacea anthyllis
Cytisus scoparius
Cytisus spinescens
Adenocarpus decorticans
Laburnum anagyroides
Lupinus luteus
Lupinus arboreus
Aspalathus argyrella
Cyclopia 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
Ammopiptanthusmongolicus
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 sinensis
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
0.04
Node annotation
Concordant
Uninformative
Discordant
0.90/98 Posterior prob/Bootstrap
0.98/–
0.01/48
0.99/94
0.74/79
0.99/–
0.99/*
0.99/*
0.46/–
0.90/98
0.63/*
0.93/*
0.93/98
0.99/70
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
Indigofereae
Astragalean clade
Vicioid clade
Wisterieae
Loteae
Robinieae
Mirbelieae
Bossiaeeae
Hypocalypteae
Millettioid-Phaseoloid clade
IR-lacking clade
Hologalegina
Robinioid clade
Mirbelioid clade
NPAAA clade
0.04
0.75/71
0.67/90
0.99/*
0.89/*
0.31/–
0.84/*
0.99/–
0.99/*
0.99/*
0.84/–
0.99/96
0.66/98
0.99/*
0.94/97
0.64/53
0.84/58
0.53/–
Node annotation
Concordant
Uninformative
Discordant
0.90/98 Posterior prob/Bootstrap
Rupertia physodes
Pediomelum californicum
Hoita orbicularis
Bituminaria bituminosa
Cullen corylifolia
Psoralea affinis
Otholobium arborescens
Phylacium majus
Glycine max
Amphicarpaea edgeworthii
Pueraria montana
Nogra grahamii
Teyleria stricta
Neonotonia wightii
Sinodolichos lagopus
Toxicopueraria yunnanensis
Pachyrhizus erosus
Psophocarpus tetragonolobus
Erythrina herbacea
Dysolobium grande
Macroptilium erythroloma
Phaseolus acutifolius
Oryxis monticola
Cochliasanthus caracalla
Vigna radiata
Dipogon lignosus
Lablab purpureus
Macrotyloma uniflorum
Strongylodon macrobotrys
Uraria lagopodioides
Christia vespertilionis
Grona triflorum
Codariocalyx motorius
Trifidacanthus unifoliolatus
Huangtcia renifolia
Dendrolobium lanceolatum
Phyllodium longipes
Hanslia ormocarpoides
Lespedeza virginica
Desmodium obtusum
Kummerowia striata
Haymondia wallichii
Craspedolobium unijugum
Dunbaria rotundifolia
Rhynchosia volubilis
Cajanus cajan
Eriosema floribundum
Dolichos falciformis
Flemingia macrophylla
Shuteria vestita
Mucuna pruriens
Apios americana
Spatholobus persicinus
Butea monosperma
Hardenbergia violacea
Kennedia rubicunda
Platycyamus regnellii
Centrosema pubescens
Centrosema coriaceum
Periandra mediterranea
Clitoria fairchildiana
Clitoria mariana
Fordia cauliflora
Pongamia pinnata
Derris trifoliata
Lonchocarpus sericeus
Dahlstedtia pinnata
Pongamiopsis pervilleana
Deguelia utilis
Piscidia mollis
Tephrosia purpurea
Tephrosia onobrychoides
Mundulea sericea
Chadsia grevei
Hesperothamnus pentaphyllus
Philenoptera violacea
Millettia leptobotrya
Dioclea virgata
Cleobulia coccinea
Macropsychanthus marginatus
Galactia striata
Galactia jussiaeana
Cratylia mollis
Canavalia villosa
Abrus mollis
Dalbergiella welwitschii
Austrosteenisia blackii
Aganope thyrsiflora
Cyamopsis tetragonoloba
Phylloxylon xylophylloides
Indigofera amblyantha
Indigofera lindheimeri
Eremosparton songoricum
Sphaerophysa salsula
Colutea arborescens
Lessertia frutescens
Sutherlandia montana
Clianthus puniceus
Carmichaelia australis
Swainsona canescens
Astragalus lentiginosus
Astragalus canadensis var. brevidens
Oxytropis ochrocephala
Gueldenstaedtia stenophylla
Tibetia himalaica
Chesneya polystichoides
Caragana halodendron
Hedysarum scoparium
Corethrodendron multijugum
Alhagi sparsifolia
Melilotus albus
Trigonella foenum-graecum
Medicago sativa
Pisum sativum
Lathyrus latifolius
Vicia faba
Galega officinalis
Cicer arietinum
Glycyrrhiza glabra
Callerya cinerea
Wisteriopsis reticulata
Nanhaia speciosa
Wisteria sinensis
Padbruggea filipes
Dorycnium pentaphyllum
Lotus japonicus
Acmispon parviflorus
Anthyllis vulneraria
Securigera varia
Hippocrepis comosa
Sesbania cannabina
Sesbania drummondii
Coursetia rostrata
Olneya tesota
Robinia pseudoacacia
Gliricidia sepium
Hebestigma cubense
Mirbelia oxylobioides
Leptosema aphyllum
Chorizema cordatum
Gastrolobium praemorsum
Urodon dasyphyllus
Aotus ericoides
Eutaxia diffusa
Pultenaea villosa
Dillwynia glaberrima
Jacksonia scoparia
Oxylobium arborescens
Daviesia arborea
Viminaria juncea
Gompholobium polymorphum
Goodia macrocarpa
Bossiaea bracteosa
Hypocalyptus sophoroides
Psoralea
Erythrina
Oryxis
Macroptilium
Sigmoidotropis
Strongylodon
Dendrolobium
Mucuna
Centrosema
Clitoria
Tephrosia
Deguelia
Cratylia
Dioclea
Indigofera
Astragalus
Hedysarum
Pisum
Lathyrus
Wisteria
Sesbania
Olneya
Coursetia
Mirbelia
Hypocalyptus


(a)
CAESALPINIOIDEAE
DETARIOIDEAE
CERC.
POLYGALACEAE
PAPILIONOIDEAE
125 Ma
100
75
50
25
Cretaceous
Pal
Eocene
Oligo
Miocene
Pli
Floral architecture
(tip points)
non-keel
keel
Net diversification
(branch colours)
Fast
Slow
(b)
Net diversification rate
Relative time
(c)
Probability
Number of shifts
(d)
Net diversification rate
non-keel
keel
Floral architecture