Syst. Biol. 54(4):575–594, 2005
Copyright © Society of Systematic Biologists
ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150590947131

Evolutionary Rates Analysis of Leguminosae Implicates a Rapid Diversification of Lineages during the Tertiary

MATT LAVIN,[1] PATRICK S. HERENDEEN,[2] AND MARTIN F. WOJCIECHOWSKI[3]

[1]*Department of Plant Sciences, Montana State University, Bozeman, Montana 59717, USA; E-mail: mlavin@montana.edu*
[2]*Department of Biological Sciences, The George Washington University, 2023 G Street NW, Washington, DC 20052, USA*
[3]*School of Life Sciences, Arizona State University, Tempe, Arizona 85287-4501, USA*

Abstract.—Tertiary macrofossils of the flowering plant family Leguminosae (legumes) were used as time constraints to estimate ages of the earliest branching clades identified in separate plastid *matK* and *rbcL* gene phylogenies. Penalized likelihood rate smoothing was performed on sets of Bayesian likelihood trees generated with the AIC-selected GTR+Γ+I substitution model. Unequivocal legume fossils dating from the Recent continuously back to about 56 million years ago were used to fix the family stem clade at 60 million years (Ma), and at 1-Ma intervals back to 70 Ma. Specific fossils that showed distinctive combinations of apomorphic traits were used to constrain the minimum age of 12 specific internal nodes. These constraints were placed on stem rather than respective crown clades in order to bias for younger age estimates. Regardless, the mean age of the legume crown clade differs by only 1.0 to 2.5 Ma from the fixed age of the legume stem clade. Additionally, the oldest caesalpinioid, mimosoid, and papilionoid crown clades show approximately the same age range of 39 to 59 Ma. These findings all point to a rapid family-wide diversification, and predict few if any legume fossils prior to the Cenozoic. The range of the *matK* substitution rate, 2.1–24.6 $\times 10^{-10}$ substitutions per site per year, is higher than that of *rbcL*, 1.6– 8.6 $\times 10^{-10}$, and is accompanied by more uniform rate variation among codon positions. The *matK* and *rbcL* substitution rates are highly correlated across the legume family. For example, both loci have the slowest substitution rates among the mimosoids and the fastest rates among the millettioid legumes. This explains why groups such as the millettioids are amenable to species-level phylogenetic analysis with these loci, whereas other legume groups are not. [Age estimation; Bayesian phylogenetics; Fabaceae; Leguminosae; *matK*; penalized likelihood rate smoothing; *rbcL*; substitution rates.]

The flowering plant family Leguminosae (Fabaceae) contains over 18,000 species distributed throughout the world in many ecological settings, from deserts of high latitudes to seasonally dry or wet tropical forests of equatorial regions (Lewis et al., 2005). Legumes diversified during the Early Tertiary (Herendeen et al., 1992) to become a ubiquitous feature of modern terrestrial biotas, similar to the timing of diversification of other prevailing terrestrial groups, such as the dominant modern families of angiosperms, polypod ferns, mammals, teleost fishes, birds, and insects (e.g., Wilf and Labandeira, 1999; Engel, 2001; Manos and Standford, 2001; Schneider et al., 2004). Second only to the grass family (Poaceae) in agricultural and economic importance, the legumes includes many species harvested as crops, or used for oils, fiber, fuel, timber, medicinals, chemicals, and horticultural varieties. Legumes play an important role in the terrestrial nitrogen cycle regardless of whether they form root nodules with symbiotic rhizobia (Sprent, 2001).

The first definitive legumes appear during the Late Paleocene, about 56 million years ago (Mya) (Herendeen, 2001; Herendeen and Wing, 2001; Wing et al., 2004). All three traditionally recognized subfamilies of legumes, the caesalpinioids, mimosoids, and papilionoids (Polhill et al., 1981), as well as other taxonomically large clades within these subfamilies (e.g., genistoids), are recorded from the fossil record soon afterward, beginning around 50 to 55 Mya (e.g., Herendeen et al., 1992). A prediction derived from the legume fossil record is that there should be little difference between the estimated age of the origin of legumes and their subsequent diversification. Now that legumes and close relatives are well sampled for molecular data (e.g., Kajita et al., 2001; Herendeen et al., 2003a; Wojciechowski et al., 2004), the goal of this article is to sample diverse legumes and immediate outgroups such that the age of the legume stem (origin) and crown (diversification) clades can be confidently estimated.

A hypothesis that posits a rapid diversification of legumes also predicts that the oldest caesalpinioid, mimosoid, and papilionoid crown clades should be equivalent in age and nearly as old as the legume crown. This finding would counter conventional wisdom that implicitly suggests that caesalpinioids harbor many ancestral traits and comprise the oldest lineages (e.g., Polhill et al., 1981). The "basal" characterization of caesalpinioids (e.g., Bruneau et al., 2001) is possibly one reason why Tucker and Douglas (1994) choose outgroups that shared radial floral symmetry with caesalpinioids (e.g., Connaraceae, Cunoniaceae, and Sapindaceae). The term "basal," however, is vague at best (e.g., Krell and Cranston, 2004). Although caesalpinioids form a paraphyletic grade from which stem the mimosoids and papilionoids (Kajita et al., 2001; Wojciechowski et al., 2004), attributing "basal" to basally branching clades would mean nothing regarding ages of extant diversifications.

With the exception of the few cursory reports in Wojciechowski (2003), no molecular age estimates have yet been detailed for crown groups within the legume family. This rate and age analysis is not only comprehensive, but incorporates multiple fossil constraints and ages estimated from multiple trees, approaches that increase the reliability of rate and age estimates (e.g., Yang and Yoder, 2003; Yoder and Yang, 2004). The rate and age analysis detailed in this study therefore can reveal not only the tempo of legume diversification, but also provide credible age estimates for crown clades with no fossil record.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

MATERIALS AND METHODS

Taxon Sampling

Only two studies have extensively sampled molecular sequence data from across legumes and close relatives (Kajita et al., 2001; Wojciechowski et al., 2004). The sequences sampled from these two studies for this analysis represent 38 of the 40 legume tribes recognized by Polhill (1994), all of the well-supported legume crown clades recently delineated by molecular phylogenetic studies (reviewed in Wojciechowski, 2003), and the designated outgroups representing the two other families of Fabales (sensu Angiosperm Phylogeny Group, 2003), Polygalaceae and Surianaceae, as well as the rosaceous genus *Quillaja*.

DNA Sequence Data

Sequence data from Wojciechowski et al. (2004) and Kajita et al. (2001) represent two plastid loci, *matK* and *rbcL*, commonly used in phylogenetic studies of plants. The *matK* data set contained 335 sequences with an aligned length of 1674 sites, 1430 of which were used in the rates analysis after gapped regions were omitted (the flanking non-coding portion of the *trnK* intron was not included). The *rbcL* data set included 241 sequences with an aligned length of 1399 sites, all of which were used in the rates analysis. The chloroplast *trnL*-F region (e.g., Pennington et al., 2001; Bruneau et al., 2001; Hu et al., 2002) was too invariable and estimates at one node broadly overlapped with those of adjacent nodes (Lavin, unpublished data). The nuclear ribosomal ITS/5.8S region (Baldwin et al., 1995) is unalignable for family-wide comparisons. DNA isolations, primer specifications, PCR amplification conditions, DNA sequencing, methods of data management, and voucher specimen information follow Wojciechowski et al. (2004) and Kajita et al. (2001). Data matrices are deposited with TreeBase (http://www.treebase.org/) study accession number S1968 for *matK* (Wojciechowski et al., 2004) and S578 for *rbcL* (Kajita et al., 2001).

Evolutionary Rates Analysis

Branch lengths were estimated using a Bayesian approach (Huelsenbeck and Ronquist, 2001; Huelsenbeck et al., 2001). The GTR+SS, GTR+Γ+I, and other nucleotide substitution models for both the *matK* and *rbcL* data sets were evaluated with Akaike information criterion (AIC) and Schwarz criterion (SC, or Bayesian information criterion), as implemented manually (e.g., Burnham and Anderson, 2002; Johnson and Omland, 2004) and in ModelTest (Posada and Crandall, 1998). Selected models were subjected to multiple runs of Metropolis-coupled Markov Chain Monte Carlo chains, each initiated with random starting trees and default substitution parameters.

Each Bayesian run comprised four chains of default temperatures and 5–20 $\times$ 10^6 generations. Tree parameters were sampled every 5 $\times$ 10^4 generations after likelihood stationarity was attained. One hundred Bayesian trees at stationarity were then systematically sampled (e.g., once every 50,000 trees) for the rates analysis in order to estimate the mean, variance, minimum, and maximum rate and age for each of 82 nodes representing the oldest crown clades within the legume family. Relative substitution rates and ages were converted to absolute rates and ages by enforcing age constraints derived from the fossil record.

A likelihood ratio test (Felsenstein, 1988; Huelsenbeck and Crandall, 1997; Posada and Crandall, 1998; Johnson and Omland, 2004) using either the GTR+Γ+I or GTR+SS nucleotide substitution models rejected a molecular clock for both the *matK* (LR = 2372, $df = 333$, $P = 0.00000$) and *rbcL* (LR = 1411, $df = 239$, $P = 0.00000$) data sets. The penalized likelihood method (Sanderson, 2002) in the program r8s (Sanderson, 2003) was thus used to estimate nucleotide substitution rates and ages of selected stem and crown clades. Via a cross-validation procedure, the penalized likelihood method finds an optimum rate-smoothing parameter for the transition of substitution rate between ancestor and descendants. The optimum level of smoothing should generally lie between the single-parameter (e.g., molecular clock; Langley and Fitch, 1974) and the parameter-rich model (e.g., typically nonparametric rate-smoothing; Sanderson, 1997).

Fossil Constraints

Age constraints derived from the legume fossil record were imposed on each of the *matK* and *rbcL* Bayesian consensus trees. The age of the root node was fixed and minimum ages were assigned to 12 internal nodes. Many fossils are potentially available, but our selection was limited to those with a unique set of apomorphic characters that could be assigned to a specific stem clade of legumes. Although either stem or crown clades may be constrained (e.g., Magallón and Sanderson, 2001), assigning fossil constraints to stem clades allowed more branch length (nucleotide substitutions) to be attributed to the minimum time frame defined by the fossil. This resulted in a faster rate of substitution and a bias toward younger age estimates. Because the objective of our study is to test the hypothesis that the oldest crown clades within legumes are indeed old relative to the legume stem clade, placing constraints on stem clades will not bias the results in favor of the hypothesis being tested. Of the 13 age constraints (A to M in Appendix 1) imposed on the *matK* phylogeny, only 9 (A, B, D, F, G, H, I, J, and K; see Appendix 1) could be placed on the *rbcL* phylogeny because of less exhaustive sampling. Absolute ages of relevant geological periods follow Berggren et al. (1996).

RESULTS

Rate-Smoothing Parameters

The optimal rate-smoothing parameter selected during cross-validation of the penalized likelihood analysis was $10^{1.5}$ for *matK* and 10^2 for *rbcL*. In both instances, these smoothing parameters resulted in a better

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

goodness of fit (estimated terminal branch lengths versus the actual terminal branch lengths) than that determined for the globally rate constant (Langley-Fitch) analysis. Nonparametric rate-smoothing consistently yielded the worst fit.

Nucleotide Substitution Estimates for the matK *Locus*

Both AIC and SC revealed the best fit of the GTR+Γ+I model compared to all other nested and non-nested models (e.g., AIC of GTR+SS = 105608 and of GTR+Γ+I = 103094 for a ΔAIC of 2514; or SC = 105677 for GTR+SS and 103158 for GTR+Γ+I for a ΔSC of 2519). Parameters estimated during the Bayesian analysis using the GTR+Γ+I and GTR+SS models (Table 1) reveal, among other things, a fairly uniform among-site rate variation. For example, the substitution rate at the 3rd codon position is only about twice that of the 2nd position.

The mean penalized likelihood (PL) substitution rate across 100 Bayesian trees for the 82 nodes identified in the *matK* phylogeny (Figs. 1 to 3, Table 2) ranges from 2.1 to 24.6 $\times$ 10^{-10} substitutions per site per year. When the 12 minimum age constraints (nodes B to M, Appendix 1) were not imposed during the PL rates analysis, a modest trend toward a faster substitution rates was observed (Fig. 4a). Three points lie distinctly above the trend line (Fig. 4a) and include two constrained nodes, K (MRCA of *Machaerium falciforme* and *Aeschynomene purpusii*; Table 2, Fig. 2) and M (MRCA of *Robinia pseudoacacia* and *Coursetia axillaris*; Table 2, Fig. 3), and one adjacent to K (node 45; Table 2, Fig. 2).

An average of 8.1 $\times$ 10^{-10} substitutions per site per year was estimated with the rate constant Langley-Fitch (LF) model (Fig. 4b), which is well within the range of estimates derived from the penalized likelihood method (Fig. 4a). This rate was increased to 14.4 $\times$ 10^{-10} substitutions per site per year when the 12 minimum fossil constraints were removed (data not shown). The nonparametric rate-smoothing (NPRS) approach resulted in highly variable rate estimates that showed no correlation with rates estimated using the penalized likelihood approach (Fig. 4b).

TABLE 1. Nucleotide substitution parameters for the GTR+Γ+I and GTR+SS models for each of the *matK* and *rbcL* loci. Parameters were estimated by Bayesian inference and sampling at likelihood stationarity (i.e., 100 trees). These models involve six substitution parameters (r), four base frequency parameters (pi), the alpha shape parameter, the proportion of invariant sites parameter [p(I)], and three relative substitution rates among codon positions, 1st (m{1}), 2nd (m{2}), and 3rd (m{3}). The GTR+Γ+I model for both loci had the best fit to the data, as revealed by either AIC or SC.

GTR+Γ+I	Mean	Variance	GTR+SS	Mean	Variance
		matK			
r(G↔T)	1.000000	0.000000	r(G↔T)	1.000000	0.000000
r(C↔T)	1.680346	0.013502	r(C↔T)	1.819096	0.005294
r(C↔G)	1.050557	0.006973	r(C↔G)	1.356160	0.004275
r(A↔T)	0.278838	0.000365	r(A↔T)	0.238177	0.000143
r(A↔G)	1.984317	0.012906	r(A↔G)	1.959152	0.005964
r(A↔C)	1.442657	0.012109	r(A↔C)	1.549988	0.005208
pi(A)	0.314941	0.000069	pi(A)	0.334913	0.000044
pi(C)	0.154121	0.000035	pi(C)	0.127895	0.000012
pi(G)	0.148623	0.000026	pi(G)	0.134290	0.000012
pi(T)	0.382314	0.000067	pi(T)	0.402901	0.000043
alpha	1.075923	0.003538	m{1}	0.860012	0.000213
p(I)	0.041595	0.000085	m{2}	0.694502	0.000173
			m{3}	1.445192	0.000210
		rbcL			
r(G↔T)	1.000000	0.000000	r(G↔T)	1.000000	0.000000
r(C↔T)	3.294871	0.079700	r(C↔T)	4.504949	0.095785
r(C↔G)	1.071123	0.011089	r(C↔G)	1.870930	0.022391
r(A↔T)	0.634750	0.002869	r(A↔T)	0.619546	0.002865
r(A↔G)	3.000933	0.056642	r(A↔G)	3.604245	0.052675
r(A↔C)	1.147267	0.015798	r(A↔C)	1.920074	0.020662
pi(A)	0.292439	0.000086	pi(A)	0.290212	0.000052
pi(C)	0.185135	0.000048	pi(C)	0.139716	0.000018
pi(G)	0.198897	0.000048	pi(G)	0.193070	0.000034
pi(T)	0.323529	0.000090	pi(T)	0.377002	0.000063
alpha	0.423598	0.000222	m{1}	0.415467	0.000256
p(I)	0.426626	0.000230	m{2}	0.251758	0.000184
			m{3}	2.334030	0.000387

Nucleotide Substitution Estimates for the rbcL *Locus*

Both AIC and SC revealed the best fit of the GTR+Γ+I model compared to all other nested and non-nested models (e.g., AIC of GTR+SS = 49008 and of GTR+Γ+I = 44614 for a ΔAIC of 4395; or SC = 49077 for GTR+SS and 44677 for GTR+Γ+I for a ΔSC of 4400). Parameters estimated during the Bayesian analysis using the GTR+Γ+I and GTR+SS models (Table 1) reveal that substitution variation at the 3rd codon position was almost 10-fold that of the 2nd position.

The mean PL substitution rate across 100 Bayesian trees for the 29 nodes identified in the *rbcL* phylogeny ranges from 1.6 to 8.6 $\times$ 10^{-10} substitutions per site per year (Table 3). If minimum fossil constraints are not imposed, a modest trend toward faster substitution rates is detected. Only two nodes show a significant deviation from the general trend (graph not shown) and include node K (MRCA of *Machaerium lunatum* and *Dalbergia hupeana*; Table 3, Fig. 2), and node G (MRCA of *Acacia farnesiana* and *Mimosa speggazzinii*; Table 3, Fig. 1). The rate constant LF model showed the same trend in estimating faster substitution rates when the minimum fossil constraints were not imposed (means of 4.8 versus 7.0 $\times$ 10^{-10} substitutions per site per year).

Comparisons of Substitution Rates Estimated for the matK *and* rbcL *Loci*

Although penalized likelihood estimates derived from the *matK* locus have a substitution rate that is on average about twice that of *rbcL*, some lineages show a similar rate between these two loci (Fig. 4c). The mimosoid stem clade (node F; Fig. 1) and the Cladrastis crown clade (node 17; Fig. 1) distinctly show the slowest mean rates for both loci (Fig. 4c). For example, the Cladrastis crown is estimated at 2.6 $\times$ 10^{-10} substitution per site per year for *matK* and 1.6 $\times$ 10^{-10} for *rbcL* (node 17, Figs. 1, 4b; Tables 2, 3). The estimated substitution rate for the inverted-repeat-loss clade (IRLC; node 75, Figs. 3, 4c) is 7.8 $\times$ 10^{-10} substitutions per site per year for the *matK* locus and 7.2 $\times$ 10^{-10} for *rbcL*. In general, the *matK* locus has a much faster rate of evolution, with the most extreme

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

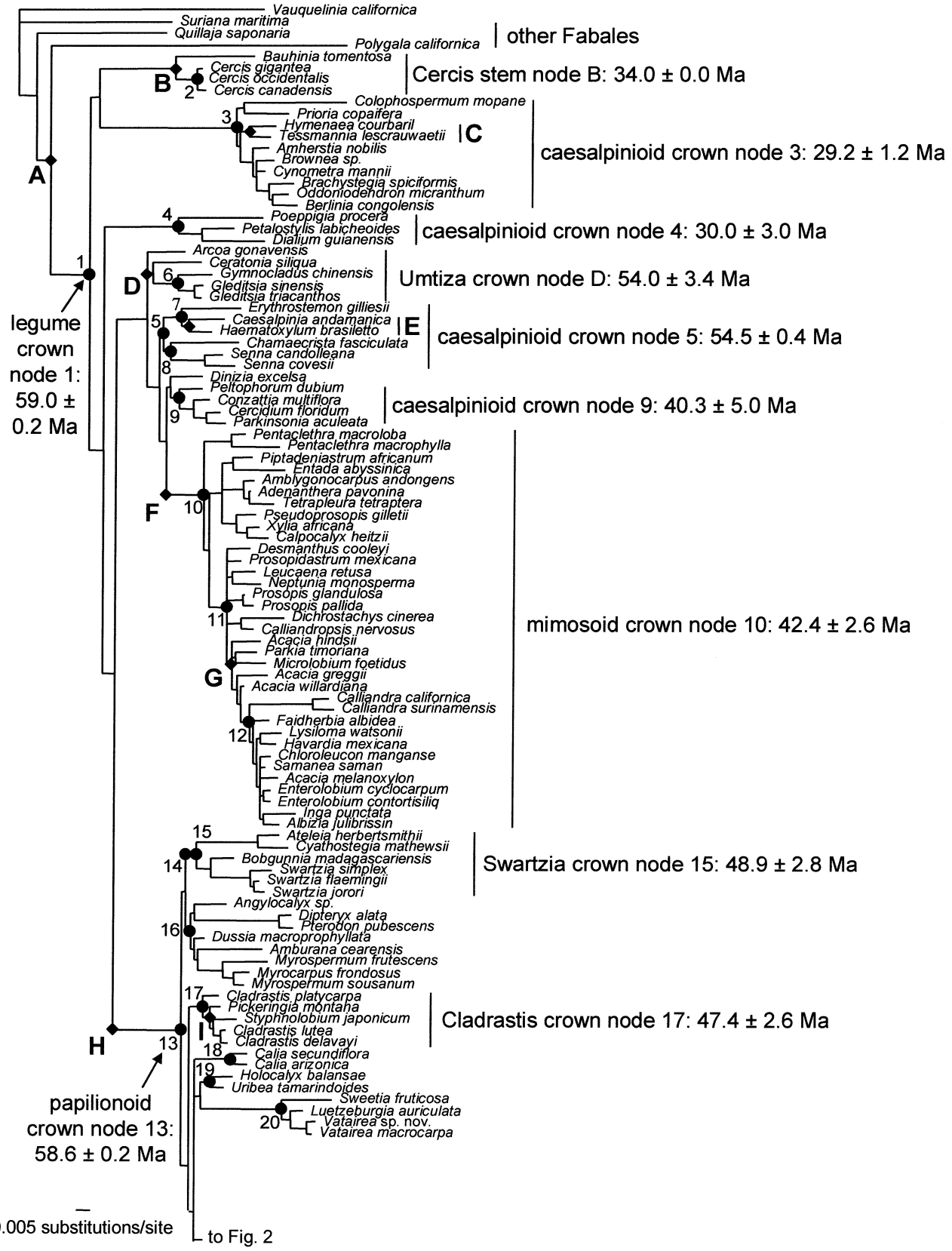

FIGURE 1. Bayesian consensus *matK* phylogeny (i.e., with averaged branch lengths) showing outgroups, caesalpinioid, mimosoid, and some of the papilionoid crown clades (e.g., the Swartzia node and the Cladrastis crown). Estimated substitution parameters for this Bayesian analysis are given in Table 1. Detailed age and rate estimates for all nodes labeled with letters or numbers are presented in Table 2. Age estimates are reported for the older crown clades. Ma = million years. The nodes labeled with diamonds are those with fixed (node A) or minimum age constraints (nodes B to M).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

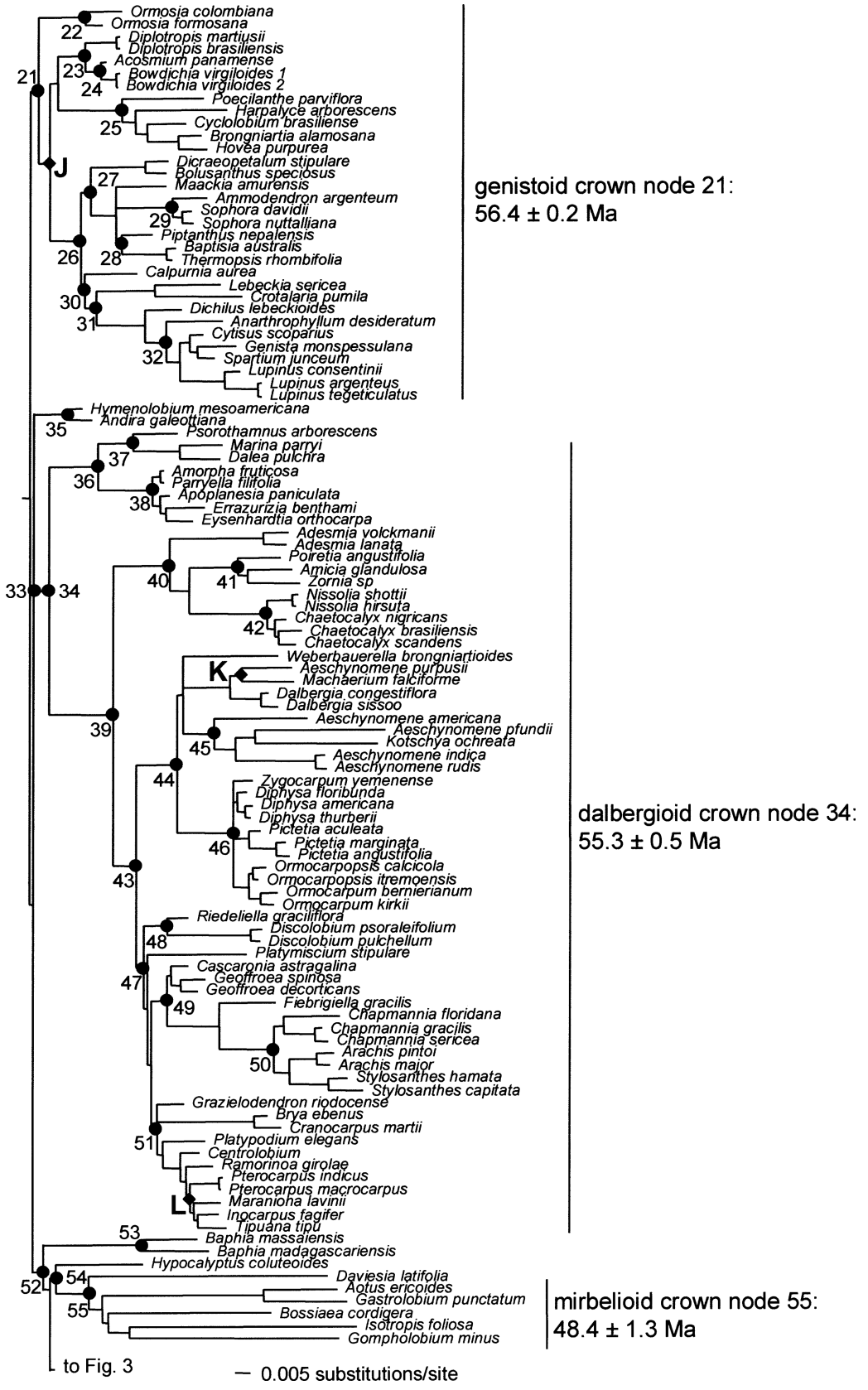

FIGURE 2. Continuation of the Bayesian consensus *matK* phylogeny from Figure 1 showing the genistoid, dalbergioid, and mirbelioid crown clades. Estimated substitution parameters for this Bayesian analysis are given in Table 1. Detailed age and rate estimates for all nodes labeled with letters or numbers are presented in Table 2. Age estimates are reported for the older crown clades. Ma = million years. The nodes labeled with diamonds are those with fixed (node A) or minimum fossil constraints (nodes B to M).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

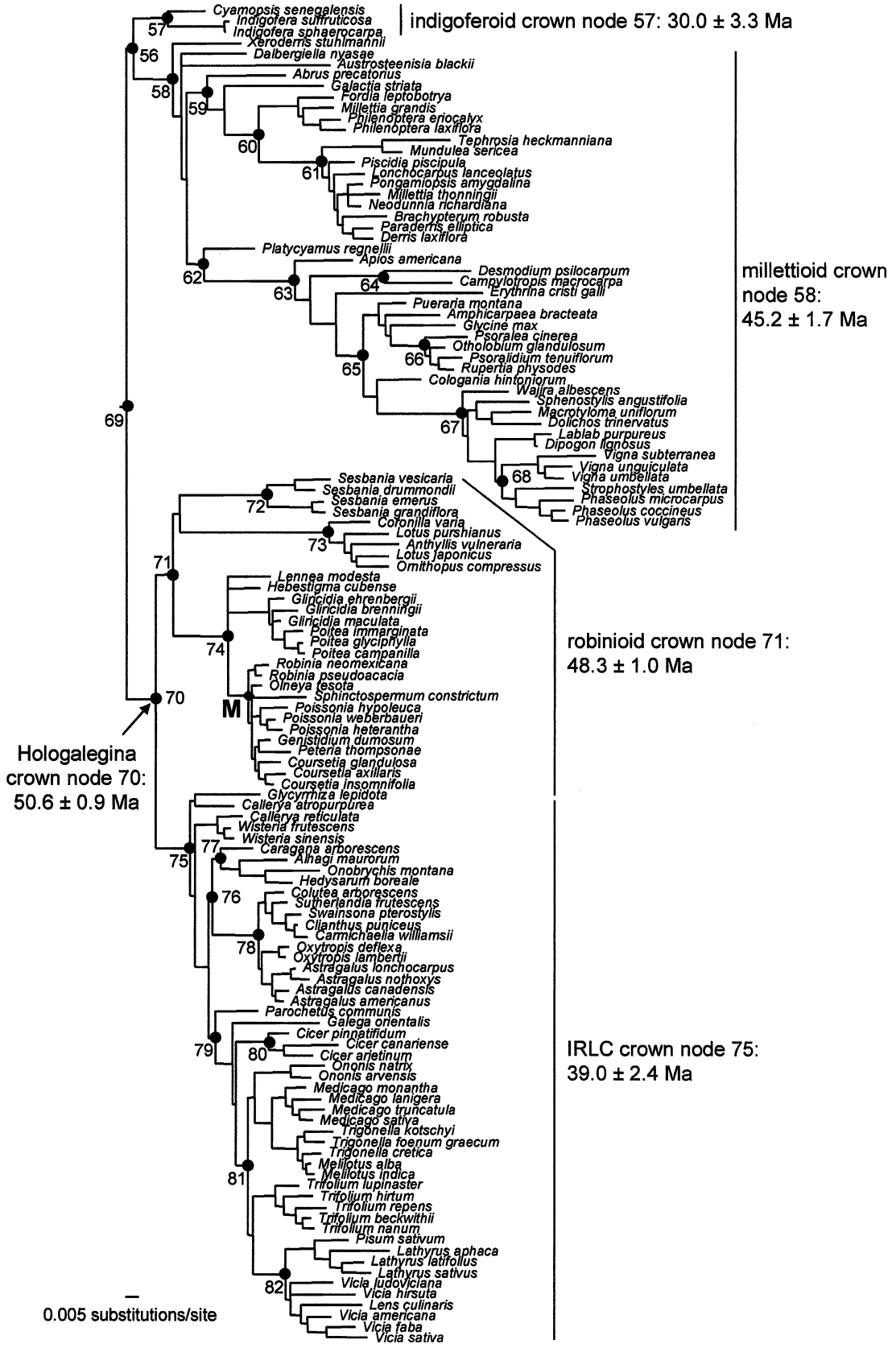

FIGURE 3. Continuation of the Bayesian consensus *matK* phylogeny from Figure 2 showing the indigoferoid, millettioid, and Hologalegina crown clades. Estimated substitution parameters for this Bayesian analysis are given in Table 1. Detailed age and rate estimates for all nodes labeled with letters or numbers are presented in Table 2. Age estimates are reported for the older crown clades. Ma = million years. The nodes labeled with diamonds are those with fixed (node A) or minimum fossil constraints (nodes B to M).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

TABLE 2. Penalized likelihood estimates of ages for fossil-constrained and unconstrained nodes across 100 *matK* Bayesian trees sampled at likelihood stationarity (for node assignment, see the Bayesian consensus, Figs. 1 to 3). The first node (A) is assigned a fixed age, whereas the next 12 nodes (B to M) are assigned minimum age constraints (indicated in parentheses) as derived from the fossil record. The letter and numeric codes for each node correspond to those shown in Figures 1 to 3. Ma = million years; SD = standard deviation; s/s = substitutions per site.

Node	*matK* node defined as the MRCA of:	Mean age (Ma)	SD (age)	Minimum (Ma)	Maximum (Ma)	Mean rate (s/s/Ma)	SD (rate)
A (60 Ma)	*Polygala californica–Cercis occidentalis*	Fixed					
B (34 Ma)	*Cercis occidentalis–Bauhinia tomentosa*	34.0	0.0	34.0	34.0	0.00084	0.00005
C (26 Ma)	*Hymenaea courbaril–Tessmannia lescrauwaetii*	26.0	0.0	26.0	26.0	0.00057	0.00007
D (34 Ma)	*Arcoa gonavensis–Gleditsia triacanthos*	54.0	3.4	39.8	57.4	0.00054	0.00009
E (45 Ma)	*Caesalpinia andamanica–Haematoxylum brasiletto*	45.0	0.0	45.0	45.0	0.00035	0.00004
F (55 Ma)	*Inga punctata–Cercidium floridum*	55.0	0.0	55.0	55.0	0.00048	0.00005
G (20 Ma)	*Acacia hindsii–Inga punctata*	30.5	3.1	25.0	37.7	0.00037	0.00006
H (55 Ma)	*Amburana cearensis–Cercidium floridum*	58.6	0.2	58.1	59.0	0.00080	0.00004
I (40 Ma)	*Styphnolobium japonicum–Pickeringia montana*	40.1	0.4	40.0	43.4	0.00024	0.00003
J (56 Ma)	*Diplotropis brasiliensis–Spartium junceum*	56.0	0.0	56.0	56.0	0.00066	0.00003
K (40 Ma)	*Machaerium falciforme–Aeschynomene purpusii*	40.0	0.0	40.0	40.0	0.00060	0.00005
L (10 Ma)	*Tipuana tipu–Pterocarpus indicus*	20.3	3.6	12.7	29.1	0.00049	0.00009
M (34 Ma)	*Robinia pseudoacacia–Coursetia axillaris*	34.0	0.0	34.0	34.0	0.00049	0.00006
1	*Hymenaea courbaril–Vicia sativa*	59.0	0.2	58.5	59.5	0.00106	0.00005
2	*Cercis occidentalis–Cercis gigantea*	5.1	2.2	1.3	13.8	0.00061	0.00007
3	*Hymenaea courbaril–Prioria copaifera*	29.2	1.2	26.9	32.6	0.00093	0.00007
4	*Petalostylis labicheoides–Poeppigia procera*	30.0	3.0	23.4	37.3	0.00093	0.00006
5	*Erythrostemon gilliesii–Senna covesii*	54.5	0.9	52.0	56.4	0.00051	0.00004
6	*Gymnocladus chinensis–Gleditsia triacanthos*	26.9	4.3	17.6	37.0	0.00042	0.00005
7	*Erythrostemon gilliesii–Haematoxylum brasiletto*	47.6	1.5	38.3	51.2	0.00046	0.00004
8	*Chamaecrista fasciculata–Senna covesii*	49.1	2.4	42.9	55.4	0.00048	0.00004
9	*Peltophorum dubium–Parkinsonia aculeata*	40.3	5.0	24.6	52.0	0.00031	0.00004
10	*Inga punctata–Pentaclethra macroloba*	42.4	2.6	34.5	47.0	0.00052	0.00005
11	*Prosopis pallida–Prosopidastrum mexicana*	33.2	3.2	22.6	40.4	0.00037	0.00006
12	*Inga punctata–Calliandra californica*	23.9	3.1	17.7	32.6	0.00051	0.00008
13	*Ateleia herbertsmithii–Albizia julibrissin*	58.6	0.2	58.1	59.0	0.00080	0.00004
14	*Swartzia simplex–Myrospermum sousanum*	55.8	1.8	47.1	57.7	0.00064	0.00008
15	*Swartzia simplex–Cyathostegia mathewsii*	48.9	2.8	40.9	54.9	0.00055	0.00004
16	*Dipteryx alata–Myrospermum sousanum*	50.8	3.8	38.4	57.1	0.00048	0.00005
17	*Styphnolobium japonicum–Cladrastis platycarpa*	47.4	2.6	40.1	53.4	0.00033	0.00003
18	*Calia secundiflora–Calia arizonica*	19.7	3.8	10.0	29.1	0.00039	0.00004
19	*Holocalyx balansae–Uribea tamarindoides*	41.3	5.4	28.8	52.1	0.00029	0.00004
20	*Sweetia fruticosa–Vatairea macrocarpa*	18.4	2.5	13.0	25.2	0.00064	0.00006
21	*Ormosia colombiana–Spartium junceum*	56.4	0.2	56.1	56.9	0.00062	0.00003
22	*Ormosia colombiana–Ormosia formosana*	20.7	4.3	12.2	32.9	0.00051	0.00005
23	*Diplotropis martiusii–Bowdichia virgiloides 1*	29.5	5.8	17.3	45.6	0.00046	0.00005
24	*Acosmium panamense–Bowdichia virgiloides 1*	12.8	4.8	4.8	25.7	0.00038	0.00008
25	*Poecilanthe parviflora–Hovea purpurea*	31.7	2.7	25.7	39.0	0.00076	0.00005
26	*Bolusanthus speciosus–Spartium junceum*	45.5	2.2	38.8	50.4	0.00068	0.00004
27	*Bolusanthus speciosus–Sophora davidii*	40.8	2.4	34.1	46.5	0.00066	0.00005
28	*Piptanthus nepalensis–Thermopsis rhombifolia*	26.5	3.4	18.4	34.5	0.00061	0.00008
29	*Ammodendron argenteum–Sophora davidii*	9.9	1.6	6.5	14.1	0.00079	0.00007
30	*Calpurnia aurea–Spartium junceum*	44.0	2.3	37.2	49.2	0.00068	0.00004
31	*Crotalaria pumila–Spartium junceum*	41.2	2.5	35.4	46.7	0.00081	0.00006
32	*Anarthrophyllum desideratum–Spartium junceum*	19.2	2.5	13.6	25.8	0.00090	0.00008
33	*Andira galeottiana–Tipuana tipu*	56.5	0.3	55.5	57.3	0.00067	0.00006
34	*Eysenhardtia orthocarpa–Tipuana tipu*	55.3	0.5	54.1	56.3	0.00089	0.00005
35	*Hymenolobium mesoamericanum–Andria galeottiana*	17.9	3.8	9.3	29.8	0.00045	0.00005
36	*Psorothamnus arborescens–Eysenhardtia orthocarpa*	36.9	3.0	29.9	44.6	0.00084	0.00007
37	*Psorothamnus arborescens–Dalea pulchra*	24.7	2.7	19.0	31.1	0.00086	0.00007
38	*Amorpha fruticosa–Eysenhardtia orthocarpa*	11.1	2.3	6.8	17.4	0.00077	0.00009
39	*Adesmia lanata–Tipuana tipu*	50.7	0.8	48.5	52.8	0.00101	0.00004
40	*Adesmia lanata–Chaetocalyx scandens*	35.3	2.3	30.1	41.3	0.00106	0.00005
41	*Poiretia angustifolia–Zornia sp*	16.1	1.9	10.8	21.4	0.00108	0.00008
42	*Nissolia hirsuta–Chaetocalyx scandens*	8.5	1.4	5.5	11.9	0.00110	0.00008
43	*Dalbergia sissoo–Tipuana tipu*	49.1	0.8	47.1	51.4	0.00090	0.00004
44	*Dalbergia sissoo–Ormocarpum kirkii*	45.6	0.8	43.9	47.3	0.00086	0.00006
45	*Aeschynomene americana–Kotschya ochreata*	36.7	2.1	31.0	40.9	0.00097	0.00005
46	*Pictetia marginata–Ormocarpum kirkii*	14.5	2.6	8.1	23.0	0.00073	0.00008
47	*Discolobium pulchellum–Tipuana tipu*	47.2	1.4	41.9	49.6	0.00074	0.00006
48	*Discolobium pulchellum–Riedeliella graciliflora*	33.6	3.6	25.0	41.1	0.00063	0.00006
49	*Cascaronia astragalina–Stylosanthes hamata*	39.8	2.3	33.7	45.0	0.00062	0.00008
50	*Chapmannia–Stylosanthes*	16.8	1.8	13.5	21.3	0.00102	0.00008
51	*Brya ebenus–Tipuana tipu*	41.9	2.3	33.4	46.5	0.00056	0.00006

(Continued on next page)

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

TABLE 2. Penalized likelihood estimates of ages for fossil-constrained and unconstrained nodes across 100 *matK* Bayesian trees sampled at likelihood stationarity (for node assignment, see the Bayesian consensus, Figs. 1 to 3). The first node (A) is assigned a fixed age whereas the next 12 nodes (B to M) are assigned minimum age constraints (indicated in parentheses) as derived from the fossil record. The letter and numeric codes for each node correspond to those shown in Figures. 1 to 3. Ma = million years; SD = standard deviation; s/s = substitutions per site. *(Continued)*

Node	*matK* node defined as the MRCA of:	Mean age (Ma)	SD (age)	Minimum (Ma)	Maximum (Ma)	Mean rate (s/s/Ma)	SD (rate)
52	*Baphia massaiensis–Vicia sativa*	55.3	0.4	54.3	56.4	0.00087	0.00006
53	*Baphia massaiensis–Baphia madagascariensis*	21.5	2.7	14.8	28.5	0.00090	0.00006
54	*Hypocalyptus coluteoides–Aotus ericoides*	54.1	1.2	51.0	56.1	0.00095	0.00008
55	*Daviesia latifolia–Aotus ericoides*	48.4	1.3	44.4	51.3	0.00126	0.00005
56	*Cyamopsis senegalensis–Phaseolus vulgaris*	52.8	1.0	50.3	55.2	0.00084	0.00005
57	*Cyamopsis senegalensis–Indigofera suffruticosa*	30.0	3.3	22.4	39.2	0.00071	0.00006
58	*Xeroderris stuhlmannii–Phaseolus vulgaris*	45.2	1.7	40.2	48.4	0.00088	0.00006
59	*Abrus precatorius–Derris laxiflora*	36.9	2.3	31.3	42.1	0.00111	0.00010
60	*Philenoptera laxiflora–Millettia thonningii*	26.1	2.0	22.0	31.0	0.00131	0.00010
61	*Mundulea sericea–Millettia thonningii*	15.0	1.6	10.9	19.4	0.00143	0.00010
62	*Platycyamus regnellii–Phaseolus vulgaris*	39.7	2.0	33.8	44.2	0.00115	0.00010
63	*Apios americana–Phaseolus vulgaris*	27.8	1.6	24.2	32.1	0.00141	0.00010
64	*Desmodium psilocarpum–Campylotropis macrocarpa*	14.2	1.6	10.5	17.7	0.00175	0.00011
65	*Pueraria montana–Phaseolus vulgaris*	19.2	1.4	15.5	22.4	0.00184	0.00011
66	*Psoralea cinerea–Rupertia physodes*	6.3	0.9	4.1	8.8	0.00185	0.00013
67	*Wajira albescens–Phaseolus vulgaris*	10.7	0.9	8.7	13.2	0.00216	0.00013
68	*Phaseolus coccineus–Vigna subterranea*	8.0	0.8	6.4	10.4	0.00246	0.00014
69	*Gliricidia maculata–Xeroderris stuhlmannii*	54.3	0.6	52.7	55.6	0.00091	0.00005
70	*Sesbania vesicaria–Vicia sativa*	50.6	0.9	47.7	52.7	0.00092	0.00005
71	*Gliricidia maculata–Sesbania vesicaria*	48.3	1.0	45.7	50.4	0.00098	0.00006
72	*Sesbania vesicaria–Sesbania grandiflora*	18.9	2.1	14.2	24.2	0.00110	0.00009
73	*Coronilla varia–Lotus japonicus*	14.4	1.3	11.7	18.6	0.00126	0.00008
74	*Gliricidia maculata–Hebestigma cubense*	38.1	1.5	31.3	40.8	0.00078	0.00007
75	*Callerya atropurpurea–Vicia sativa*	39.0	2.4	33.2	44.9	0.00078	0.00005
76	*Caragana arborescens–Astragalus americanus*	33.0	2.7	25.0	39.2	0.00080	0.00007
77	*Caragana arborescens–Hedysarum boreale*	29.3	3.0	21.1	35.4	0.00077	0.00008
78	*Colutea arborescens–Astragalus americanus*	14.8	2.0	10.3	20.9	0.00083	0.00008
79	*Parochetus communis–Vicia sativa*	32.1	2.4	26.4	37.6	0.00084	0.00007
80	*Cicer pinnatifidum–Cicer arietinum*	14.8	2.6	6.5	20.2	0.00100	0.00011
81	*Ononis natrix–Vicia sativa*	24.7	2.3	17.1	30.2	0.00108	0.00010
82	*Pisum sativum–Vicia sativa*	17.5	1.9	12.9	22.8	0.00129	0.00011

rates observed in lineages traditionally classified in the papilionoid tribe Phaseoleae (nodes 63 to 68 in Table 2; Fig. 3) with mean rates of about 20×10^{-10} substitutions per site per year or greater. The substitution rate estimated for node 65 in particular (the MRCA of *Pueraria montana* and *Phaseolus vulgaris* in Fig. 2; Tables 2, 3) is distinctly the highest for both *matK* and *rbcL* (Fig. 4c).

Estimated Ages of Old Legume Crown Clades

The ages estimated for the 28 nodes that are comparable between the *matK* and *rbcL* phylogenies show a positive linear relationship (Tables 2, 3; Fig. 5a). In both phylogenies, the estimated age of the legume crown is 59 million years (Ma) (node 1 in Tables 2, 3). The estimated ages derived from the penalized likelihood rates smoothing for older legume crown clades are mostly less than 30 Ma (Figs. 1 to 3, 6, 7). These age estimates show a strong positive relationship with the rate constant LF and the rate variable NPRS estimates, although NPRS tends to yield the oldest ages of the three methods (Fig. 5b). Imposing the 12 minimum age constraints on nodes B to M (Figs. 1 to 3) generally resulted in age estimates that are older than estimates without the fossil constraints (Fig. 5c).

The ages of the oldest caesalpinioid crown clades range from approximately 54 to 30 Ma (nodes B to E, Table 2; and nodes 2 to 9, Fig. 1). The oldest of these, nodes 5 and D (the Umtiza crown node), have estimated mean ages of about 54 Ma (Fig. 1). In contrast, the extant mimosoids trace back to a most recent common ancestor at just over 40 Ma (node 10, Fig. 1). The older papilionoid crown clades show ages comparable to the oldest extant caesalpinioid diversifications. For example, the age of the genistoid diversification is estimated at about 56 Ma (node 21, Fig. 2), that of the dalbergioids at about 55 Ma (node 34, Fig. 2), and Hologalegina at about 50 Ma (node 70, Fig. 3). Indeed, nearly all of the remaining oldest papilionoid crown clades (e.g., nodes 15, 17 in Fig. 1; node 55 in Fig. 2; node 58 in Fig. 3) are around 45 Ma or older. The ages of extant papilionoid diversifications relative to those of caesalpinioids do not change when minimum age constraints are not imposed. In this case, all of the estimated ages are younger by about 1 to 5 Ma (data not shown), the difference being greater with caesalpinioid clades because this is where most of the minimum fossil constraints are imposed (see positions of constraints B to F in Fig. 1).

Effects of Varying the Age of the Root Node

The analyses presented above are derived from fixing the age of the legume stem clade at 60 Ma. Moving this age at 1-Ma intervals back to 70 Ma has little qualitative

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

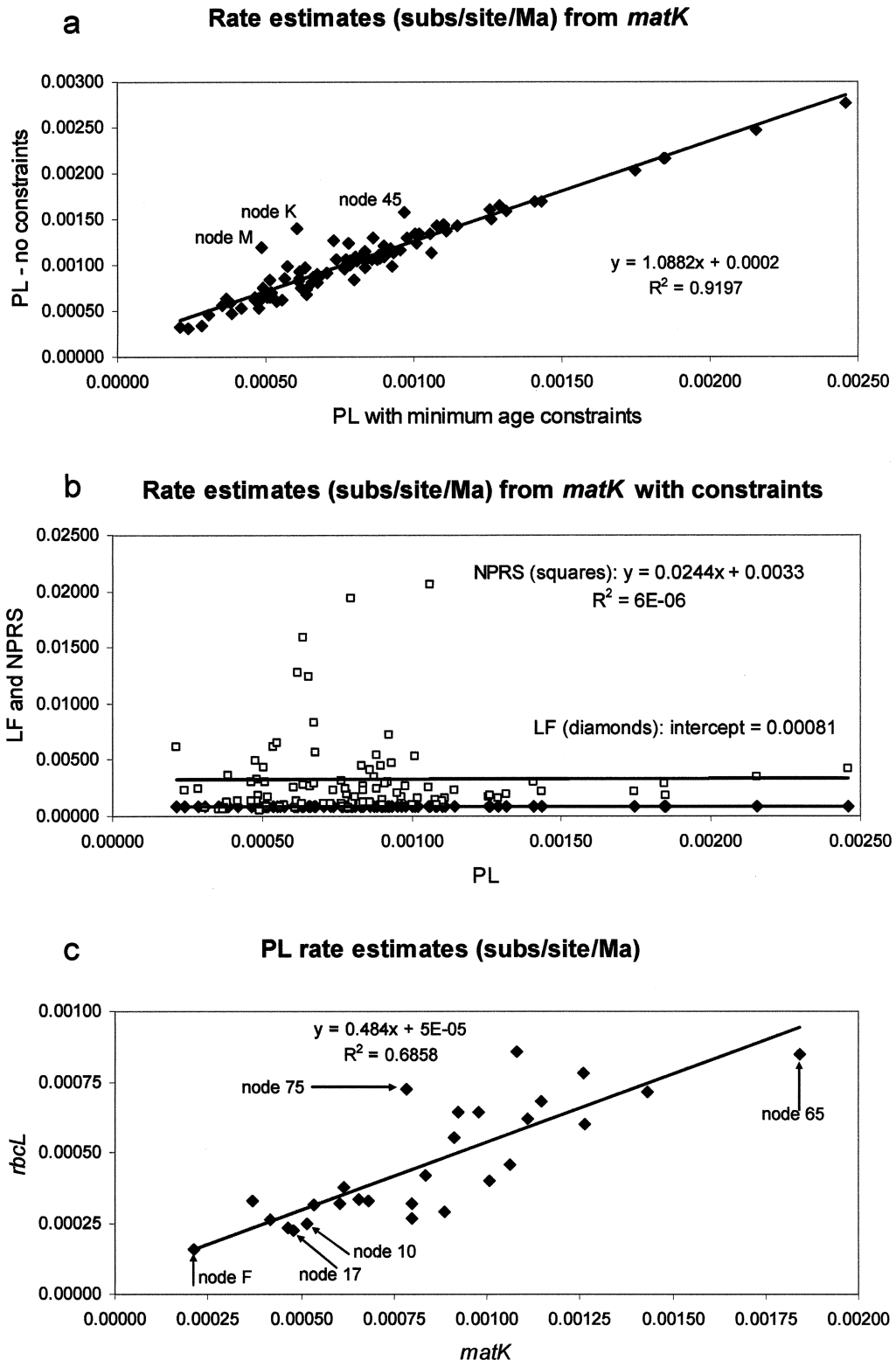

FIGURE 4. Descriptive statistical comparisons of mean rate estimates (see Tables 2 and 3) derived from the *matK* and *rbcL* data sets. (a) Rate estimates derived from *matK* using penalized likelihood (PL) for the 94 nodes (B to M, 1 to 82) listed in Table 1 compared to PL estimates for the same nodes derived from not invoking the 12 fossil-calibrated minimum age constraints; subs/site/Ma = substitution per site per million years. (b) Rate estimates derived from *matK* using PL with the 12 minimum age constraints imposed and compared to the rate constant Langley-Fitch (LF) and rate variable nonparametric rate smoothing (NPRS) methods. The optimum substitution rate estimated with LF using minimal age constraints is indicated by the line with an intercept of 0.00081 substitutions/site/Ma (diamonds). Without minimum age constraints, the rate constant estimate is 0.00144 substitutions/site/Ma. (c) Comparison of rate estimates derived from each of the *matK* and *rbcL* Bayesian consensus phylogenies for the 28 comparable nodes listed in Table 3 (not including root node A) using PL rate smoothing (see Figs. 1 and 3 for the identification of the labeled nodes).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

TABLE 3. Penalized likelihood estimates of ages for fossil-constrained and unconstrained nodes across 100 *rbcL* Bayesian trees sampled at stationarity (not shown but see parsimony consensus in Kajita et al., 2001). The legume stem clade (A) is fixed at 60 Ma, and the next eight nodes listed (B, D, F, G, H, J, K, and M) are assigned minimum age constraints (indicated in parentheses) as derived from the fossil record. These lettered and numbered nodes in the *rbcL* phylogeny correspond to a node in the *matK* phylogeny and are labeled to identify this association (see Figs. 1 to 3). Ma = million years; SD = standard deviation; s/s = substitutions per site.

rbcL Node defined as the MRCA of:	Corresponding *matK* node	Mean age (Ma)	SD (age)	Minimum (Ma)	Maximum (Ma)	Mean rate (s/s/Ma)	SD (rate)
Polygala cruciata–Cercis canadensis	A (60 Ma)	60.0					
Cercis canadensis–Bauhinia purpurea	B (34 Ma)	39.4	4.0	34.0	48.5	0.00042	0.00004
Acrocarpus sp.–Gleditsia triacanthos	D (34 Ma)	58.1	2.8	45.4	59.4	0.00031	0.00004
Erythrophleum ivorense–Albizia saman	F (55 Ma)	55.0	0.0	55.0	55.0	0.00022	0.00003
Acacia farnesiana–Mimosa speggazzinii	G (20 Ma)	29.3	3.4	25.0	36.6	0.00033	0.00004
Peltogyne confertiflora–Glycine tabacina	H (55 Ma)	58.5	0.3	57.4	59.1	0.00032	0.00003
Acosmium dasycarpum–Spartium junceum	J (56 Ma)	56.0	0.0	56.0	56.0	0.00033	0.00003
Dalbergia hupeana–Machaerium lunatum	K (40 Ma)	40.0	0.0	40.0	40.0	0.00032	0.00004
Cercis canadensis–Zenia insignis	1	59.4	0.2	58.6	59.9	0.00046	0.00003
Cercis canadensis–Cercis siliquastrum	2	14.3	3.9	7.4	27.7	0.00037	0.00005
Caesalpinia pulcherrima–Albizia saman	5	57.9	0.4	57.0	58.7	0.00023	0.00002
Gymnocladus dioica–Gleditsia triacanthos	6	30.4	5.8	18.0	46.6	0.00026	0.00004
Acacia farnesiana–Albizia saman	10	40.5	4.0	31.3	49.5	0.00025	0.00003
Xanthocercis zambesiaca–Glycine tabacina	13	57.7	0.4	56.9	58.7	0.00027	0.00004
Cladrastis sinensis–Sophora japonica	17	31.8	10.2	8.4	51.2	0.00016	0.00004
Piptanthus nepalensis–Spartium junceum	26	47.2	2.8	38.2	53.7	0.00033	0.00004
Andira inermis–Machaerium lunatum	33	57.3	0.4	56.3	58.2	0.00029	0.00004
Zornia cantoniensis–Machaerium lunatum	39	53.4	1.1	49.5	55.9	0.00040	0.00003
Goodia lotifolia–Isotropis cuneifolia	55	47.6	5.6	35.6	56.7	0.00060	0.00010
Abrus precatorius–Tephrosia grandiflora	59	36.2	3.1	25.6	42.3	0.00062	0.00005
Derris laxiflora–Tephrosia grandiflora	61	14.5	2.6	9.2	21.8	0.00072	0.00008
Apios taiwaniana–Teramnus labialis	63	35.9	2.9	29.1	42.0	0.00068	0.00006
Dipogon lignosus–Teramnus labialis	65	22.7	2.5	17.2	31.9	0.00085	0.00007
Phylloxylon perrieri–Cicer arietinum	69	54.8	1.1	51.4	56.7	0.00055	0.00004
Robinia pseudoacacia–Cicer arietinum	70	49.7	1.6	45.3	53.3	0.00064	0.00003
Robinia pseudoacacia–Sesbania sesban	71	45.4	2.6	37.9	51.6	0.00064	0.00004
Coronilla varia–Lotus corniculatus	73	24.6	2.7	18.8	31.8	0.00078	0.00005
Glycyrrhiza glabra–Cicer arietinum	75	42.4	2.1	37.0	47.2	0.00072	0.00004
Pisum sativum–Medicago sativa	81	24.1	2.4	18.2	29.6	0.00086	0.00005

effect on the results. At the 60-Ma fixed age, the estimated ages of the legume and papilionoid crown clades lay distinctly above the trend line to the upper right (Fig. 8a and b). These two age estimates, when plotted at 1-Ma intervals (Fig. 8c; the legume crown is illustrated for both *matK* and *rbcL*), are exceptional in showing successively older ages as the fixed age of the root node is moved back to 70 Ma. All other nodes for which ages are estimated have a flat trend line over the 60- to 70-Ma interval, as exemplified by those of the caesalpinioid and mimosoid crowns (Fig. 8). Regardless, the difference in mean age estimates between the legume stem and crown clades vary between 1.0 and 2.5 Ma within this 60- to 70-Ma interval.

DISCUSSION

The *matK* tree is more resolved and better supported than the *rbcL* tree, a finding consistent with the parsimony analyses of essentially these same data sets (cf. Wojciechowski et al., 2004; Kajita et al., 2001). The present study suggests this may be the result of the rate of substitution at the *matK* locus, which is more evenly distributed among the three codon positions and generally about twice as fast as that of *rbcL* (*matK* range of 2.1 to 24.6 $\times$ 10^{-10} versus *rbcL* range of 1.6 to 8.6 $\times$ 10^{-10}). These estimated rates of substitution in *matK* and *rbcL*, the rate differences between the two loci, and the among-site rate variation, are very similar to those reported for other plant groups (e.g., Albert et al., 1994; Steele and Vilgalys, 1994; Wang et al., 1999; Osaloo and Kawano, 1999; Young and dePamphilis, 2000; Xiang et al., 2000; Zhang and Renner, 2003; Neel and Cummings, 2004).

The rapid diversification of the legume family soon after its origin is revealed by both the *matK* and *rbcL* phylogenies (Figs. 6, 7). An estimated 1.0 to 2.5 Ma distinguishes the mean ages of the legume stem and crown clade. This may explain why clade support values for the monophyly of Leguminosae are sometimes lower compared to support levels for the entire Fabales crown (i.e., the MRCA of *Quillaja*, Surianaceae, Polygalaceae, and Leguminosae; sensu Angiosperm Phylogeny Group, 2003) or many of the older crown clades within legumes (Bruneau et al., 2001; Kajita et al., 2001; Wojciechowski, 2003; Wojciechowski et al., 2004).

The total duration of legume evolution was fixed at 60 Ma in our analyses, just older than the oldest definitive legume fossil at 56 Ma (Herendeen, 2001). Several of the earliest minimal time constraints are in the 55- to 56-Ma range (nodes F, H, and J, Table 2; Figs. 1, 2). Regardless, the legume crown could have been estimated closer to a 56-Ma age instead of about 59 Ma (node 1, Fig. 1). If all minimum fossil constraints are removed, the age of the legume crown is estimated at 55.9 $\pm$ 1.0 Ma, also a short time span relative to the total duration of legume

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

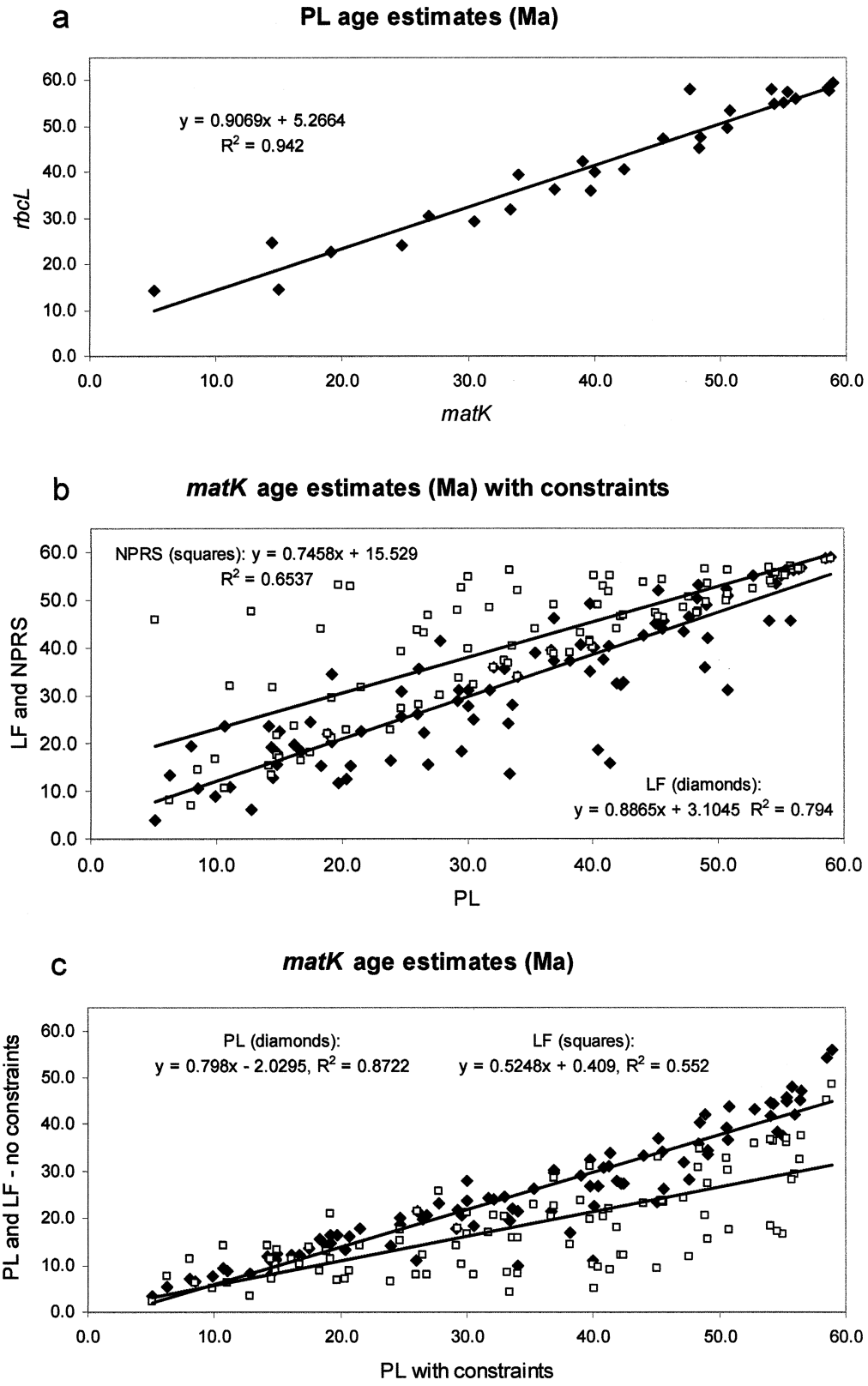

FIGURE 5. Descriptive statistical comparisons of mean age estimates derived from the *matK* and *rbcL* data sets (see Tables 2 and 3). (a) Comparison of age estimates derived from the *matK* and *rbcL* Bayesian phylogenies for the 28 comparable nodes listed in Table 3 (not including root node A) using the penalized likelihood (PL) rate smoothing approach. Ma = million years. (b) Age estimates derived from *matK* imposing minimal age constraints. PL estimates for the 94 nodes listed in Table 1 (not including root node A) are compared with those derived from rate constant Langley-Fitch (LF; diamonds) and rate variable nonparametric rate smoothing (NPRS; squares) methods. (c) Age estimates derived from *matK* with and without minimal age constraints. PL estimates using minimum age constraints for the 94 nodes listed in Table 1 (not including root node A) are compared with those derived from PL (diamonds) and LF (squares) methods without minimum age constraints.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

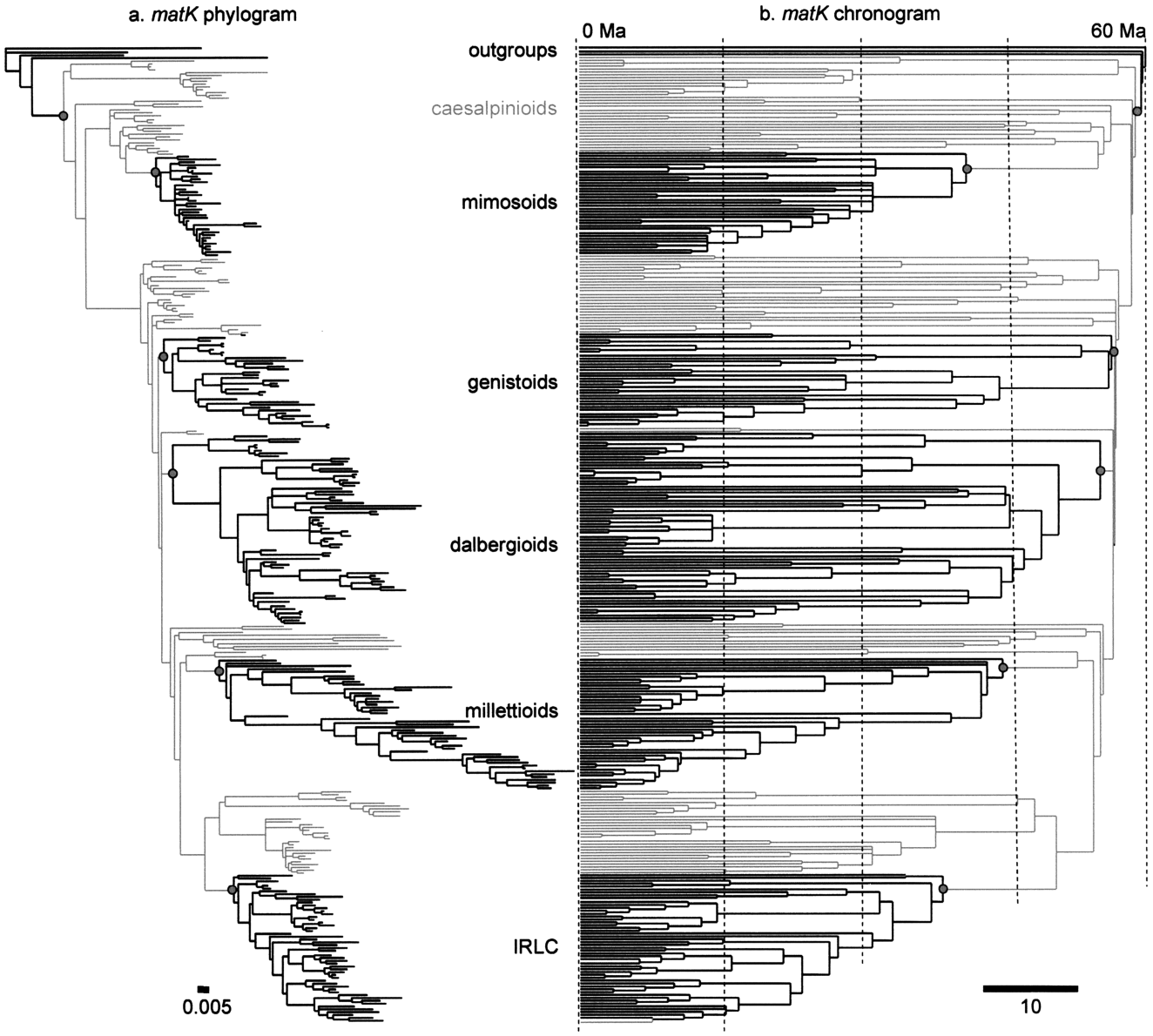

FIGURE 6. Phylogenies derived from the *matK* data set. (a) Bayesian consensus phylogram. Scale bar equals 0.005 substitutions per site. (b) Penalized likelihood rate-smoothed chronogram with the 12 minimum age constraints imposed. Scale bar equals 10 Ma. The vertical lines (b) divide the 60-Ma duration of legume evolution into 15-Ma segments. The oldest and taxonomically large crown clades are alternately shaded black and gray. The shapes of both phylogenies depict generally short internal branch lengths indicative of a rapid family-wide diversification.

evolution. Notably, increasing the fixed age of the legume stem to 70 Ma still resolves a difference on average of up to 2.5 Ma between the estimated age of the legume stem and crown (Fig. 8). These findings support the overall conclusion that an Early Tertiary legume diversification immediately followed the origin of the family.

The fossil record of legumes predicts this genetic finding of a rapid diversification of extant lineages. Many diverse legumes, especially representatives from each of the three traditionally recognized subfamilies, have continuous fossil records from the Recent back to the Late Paleocene (e.g., Herendeen and Dilcher, 1992). Moreover, this record comes from North and South America, Europe, Africa, and Asia (e.g., Axelrod, 1992; Herendeen et al., 1992). The caesalpinioids, with an estimated 161 genera and about 3000 species (Lewis et al., 2005), are most diverse in tropical regions throughout the New World, Africa, and southeast Asia. Although caesalpinioids form a basal grade within which are nested the other two subfamilies (e.g., Polhill, 1981, 1994; Herendeen et al., 2003a, 2003b), the estimated ages of the older caesalpinioid crown clades differ little from those of the mimosoids, and especially the papilionoids (compare marked crown clades in Figs. 1 to 3).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

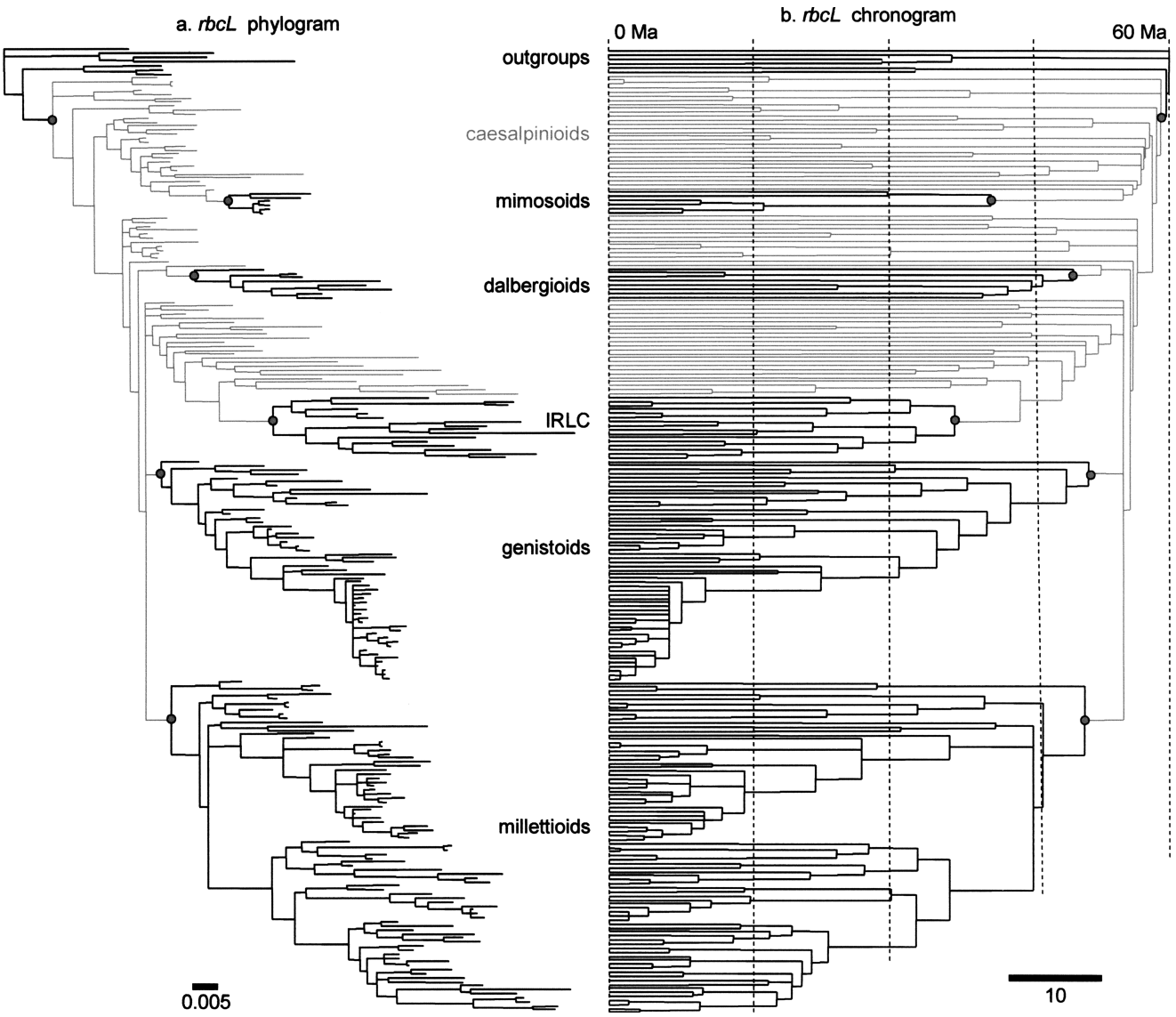

FIGURE 7. Phylogenies derived from the *rbcL* data set. (a) Bayesian consensus phylogram. Scale bar equals 0.005 substitutions per site. (b) Penalized likelihood rate-smoothed chronogram with the 12 minimum time constraints enforced. Scale bar equals 10 Ma. The vertical lines (b) divide the 60-Ma duration of legume evolution into 15-Ma segments. The shapes of both phylogenies depict generally short internal branch lengths indicative of a rapid family-wide diversification. The oldest and taxonomically large crown clades are alternately shaded black and gray, with the top gray shaded lines representing the outgroups and the following paraphyletic grade of black lines representing the caesalpinioid legumes. Otherwise, the groups are labeled. The arrows indicate the legume stem clade.

The papilionoid crown clade, and not a caesalpinioid one, is most sensitive to the effects of moving the fixed root age from 60 back to 70 Ma (Fig. 8). This implies that the oldest legume fossils are as likely to represent papilionoids as caesalpinioids. This is actually predicted by the fossil record. The oldest known legume fossils, with an age of 56 Ma (Herendeen, 2001), are fruits similar to *Diplotropis* and *Bowdichia*, representatives of the genistoid clade of papilionoids (Crisp et al., 2000). Fossil flowers assigned to *Barnebyanthus buchananensis* are close in age to these oldest legume fossils and in overall morphology to extant genistoid genera such as *Bowringia* and *Clathrotropis* (Crepet and Herendeen, 1992). Fossil leaves and fruits with affinities to the genistoid *Ormosia* come from strata at least 56 Ma (Wing et al., 2004, and unpublished).

The dalbergioids (Lavin et al., 2001), Hologalegina (Wojciechowski et al., 2000), and mirbelioids (Crisp et al., 2000), in contrast to the genistoids, lack an Early Eocene fossil record. Although these three large crown clades include papilionoids traditionally considered derived (Polhill et al., 1981; Polhill, 1994), they all have age estimates in the 50-Ma time frame or older (Figs. 2, 3). Indeed, Hologalegina contains many of the well-known temperate and herbaceous species of legumes, such as alfalfa, clovers, peas, and vetches. Remarkably, the mean age of this crown diversification is estimated at 50.6 Ma (Fig. 3, Table 2).

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

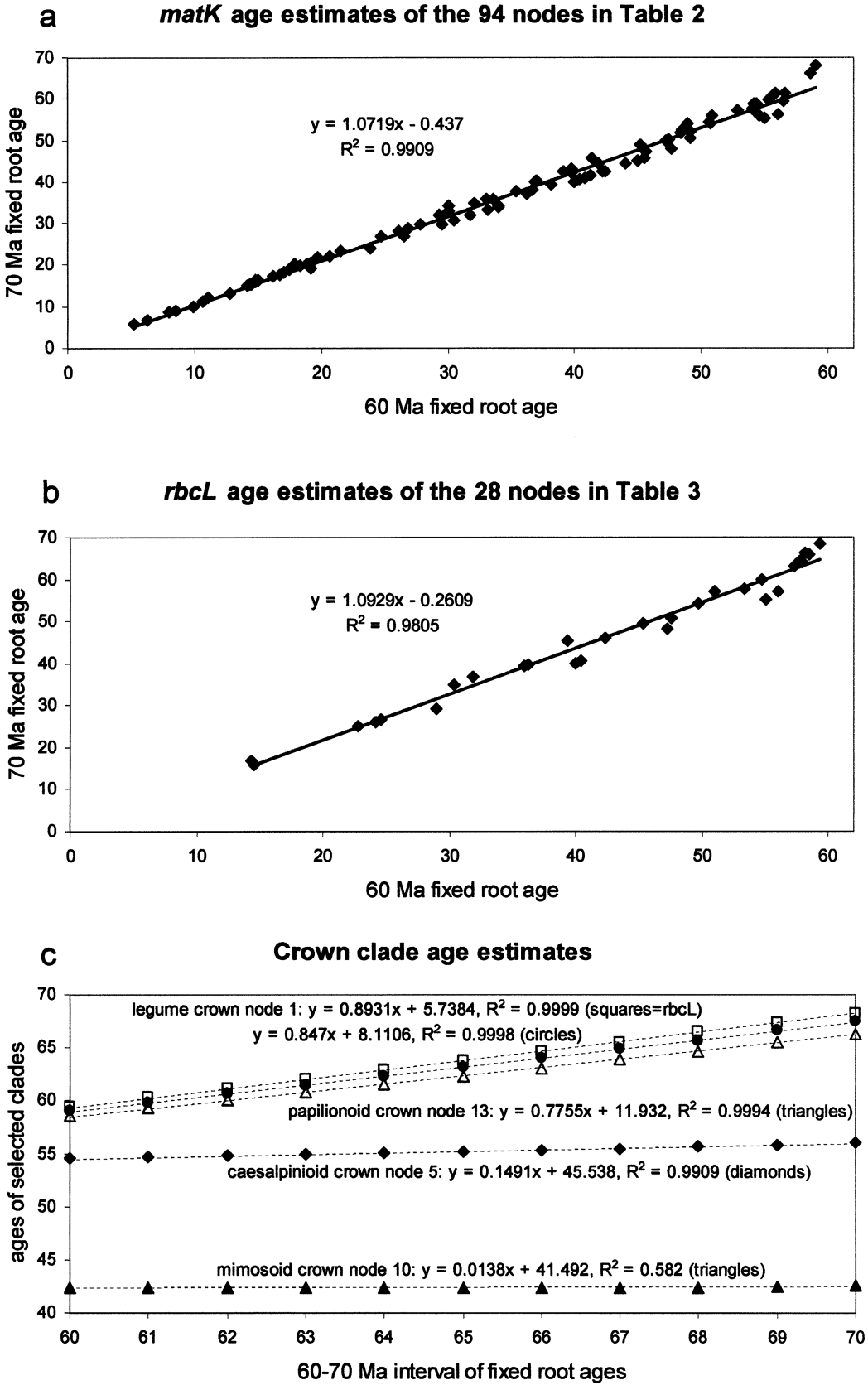

FIGURE 8. Descriptive statistical comparisons of mean age estimates when the fixed age of the legume stem clade is scaled between 60 and 70 Ma. (a) Age estimates derived from the *matK* data set for the 94 nodes identified in Table 2 (B to M, 1 to 82; not including the root node). (b) Age estimates derived from the *rbcL* data set for the 28 nodes identified in Table 3 that are shared with the *matK* phylogeny (not including root node). (c) Ages estimates derived from the *matK* data set or *rbcL* data set (where noted) for selected old crown clades.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

Subfamily Mimosoideae, with an estimated 80 genera and some 3000 species (Lewis et al., 2005), are mostly tropical to subtropical in distribution, and are abundant in arid and semiarid regions throughout the world. Of all the old legume crown clades, the mimosoids have a paucity of nucleotide substitution variation for both *matK* and *rbcL* (nodes F and 10 in Tables 2, 3; Fig. 4c). The apparent slow down in the rate of nucleotide substitution in this group may explain the findings in recent molecular systematic studies (Luckow et al., 2003; Miller et al., 2003) of many short and poorly supported branch lengths within the mimosoid crown. However, the mimosoids have a relatively long stem branch leading to the crown and accordingly a significant difference between the age of the origin and diversification of extant members the subfamily (about 15 Ma; compare nodes F and 10 in Tables 2, 3). These estimated ages derived independently from *matK* and *rbcL* are highly congruent and suggest that the lack of nucleotide variation in mimosoids is a function of both a rate slow-down and a relatively recent extant diversification, beginning about 40 Ma (see also Figs. 7b, 8b).

Numerous mimosoid leaf fossils come from the Early Tertiary, such as the 46 million year old *Acacia mahenge* fossil (Appendix 1). However, none can be assigned unequivocally to a modern subgroup. Given that no fossil constraints older than that of a 15-Ma-old fossil *Acacia* flower can be confidently assigned to an internal node within the mimosoid crown (node G, Tables 2, 3; Fig. 1), the combined molecular and fossil evidence suggests not only a relatively recent diversification of mimosoids, but also that many Early Tertiary mimosoid fossils belong to extinct lineages.

In contrast to the mimosoids, the tribe Phaseoleae (stemming from node 63 of the millettioid clade, Fig. 3) shows a very high rate of nucleotide substitution for both the *matK* and *rbcL* loci (Figs. 3, 4c, 5b; nodes 63 to 68 in Tables 2, 3). These two plastid loci contribute high levels of phylogenetic resolution and clade support in studies that sample at the genus and even species level within these clades (e.g., Kajita et al., 2001; Delgado-Salinas et al., 2003; Riley-Hulting et al., 2004; Thulin et al., 2004). That we detect much greater rates of nucleotide substitution in the Phaseoleae compared to the mimosoid crown is not an artifact of denser sampling in the former, as might be inferred from the findings of Webster et al. (2003, 2004). Both clades are approximately the same size in terms of actual numbers (i.e., approximately 3270 species of mimosoids and 2060 species of Phaseoleae; Lewis et al., 2005) and sampled numbers of species (i.e., 35 mimosoids and 25 Phaseoleae; Figs. 1, 3).

This study provides estimated ages that can serve as calibration points for rates analyses of legume clades lacking a fossil record (e.g., Schrire et al., 2003; Thulin et al., 2004; Lavin et al., 2004; Percy et al., 2004). Also, we have examined rate variation at the *matK* and *rbcL* loci across all of legumes, and have thus revealed that some groups such as the millettioids have a sufficiently fast rate of substitution so that these two plastid loci can be used in phylogenetic analyses to successfully resolve relationships even at the species level (e.g., Delgado-Salinas et al., 2003; Riley-Hulting et al., 2004; Thulin et al., 2004). In addition, the profusion of basally branching and morphologically diverse crown clades within legumes reveals that the paraphyletic grade of caesalpinioid lineages harbor neither the oldest diversifications nor some other quality of legume antiquity. Finally, the integration of fossil constraints with molecular phylogenetic data allows the identification of the rapid diversification early in the history of legume evolution. This new perspective would not have been possible with fossil data alone because as it turns out many of the oldest crown clades in legumes are unknown from the fossil record.

ACKNOWLEDGEMENTS

Anne Bruneau, Alfonso Delgado-Salinas, Colin Hughes, R. Toby Pennington, and Mats Thulin provided unpublished *matK* sequences for selected taxa. Brian Schrire and Gwilym Lewis assisted with taxon sampling, organized input files involving tree descriptions and node identifications, and provided taxonomic details taken from their unpublished Legumes of the World. Scott Wing willingly showed us legume fossils from Cerrajón, Colombia. Kathleen Pigg kindly provided information on the fossil record of Fabales other than Leguminosae. Mike Crisp, Colin Hughes, Peter Linder, Rod Page, Kathleen Pigg, and Mike Sanderson provided critical comments that improved the manuscript. Jay Rotella guided us in dealing with the intricacies of model selection. This study was supported in part by grants from the National Science Foundation (DEB-9407824 to MFW and M. J. Sanderson, and DEB-0075202 to ML, and DEB-0316375 to PH) and funds provided by Arizona State University (to MFW). Useful discussion of this study was provided at the "Deep Time" sponsored symposium on dating methods (Botany 2003, Mobile, Alabama; NSF RCN-0090283).

REFERENCES

Albert, V. A., A. Backlund, K. Bremer, M. W. Chase, J. Manhart, B. D. Mishler, and K. C. Nixon. 1994. Functional constraints and *rbcL* evidence for land plant phylogeny. Ann. Mo. Bot. Garden 81:534–567.

Angiosperm Phylogeny Group. 2003. Angiosperm Phylogeny Group. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc. 141:399–436.

Axelrod, D. I. 1992. Climatic pulses, a major factor in legume evolution. Pages 259–279 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Baldwin, B. G., M. J. Sanderson, J. M. Porter, M. F. Wojciechowski, C. S. Campbell, and M. J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. Ann. Mo. Bot. Garden 82:247–277.

Berggren, W. A., D. V. Kent, C. C. Swisher III, and M.-P. Aubry. 1996. A revised Cenozoic geochronology and chronostratigraphy. Pages 129–212 *in* Geochronology, time scales and global stratigraphic correlation (W. A. Berggren, D. V. Kent, M.-P. Aubry, and J. Hardenbol, eds.). SEPM Special Publication No 54.

Brown, R. W. 1937. Fossil legumes from Bridge Creek, Oregon. J. Wa. Acad. Sci. 27:414–418.

Brown, R. W. 1962. Paleocene flora of the Rocky Mountains and Great Plains. Geological Survey Professional Paper 375. U. S. Government Printing Office, Washington, D.C.

Bruneau, A., F. Forest, P. S. Herendeen, B. B. Klitgaard, and G. P. Lewis. 2001. Phylogenetic relationships in the Caesalpinioideae (Leguminosae) as inferred from chloroplast *trn*L intron sequences. Syst. Bot. 26:487–514.

Burnham, K. P., and D. R. Anderson. 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd edition. Springer-Verlag, New York.

Burnham, R. 1995. A new species of winged fruit from the Miocene of Ecuador: *Tipuana ecuatoriana* (Leguminosae). Am. J. Bot. 82:1599–1607.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

Crane, P. R., S. R. Manchester, and D. L. Dilcher. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinal Butte Formation (Paleocene) near Almont, North Dakota. Fieldiana Geol. New Ser. 20:1–63.

Crepet, W. L., and P. S. Herendeen. 1992. Papilionoid flowers from the early Eocene of southeastern North America. Pages 43–55 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Crepet, W. L., and D. W. Taylor. 1985. The diversification of the Leguminosae: First fossil evidence of the Mimosoideae and Papilionoideae. Science 288:1087–1089.

Crepet, W. L., and D. W. Taylor. 1986. Primitive mimosoid flowers from the Paleocene-Eocene and their systematic and evolutionary implications. Am. J. Bot. 73:548–563.

Crisp, M. D., S. Gilmore, and B.-E. van Wyk. 2000. Molecular phylogeny of the genistoid tribes of papilionoid legumes. Pages 249–276 *in* Advances in legume systematics, part 9 (P. S. Herendeen and A. Bruneau, eds.). Royal Botanic Garden, Kew, UK.

Delgado-Salinas, A., R. Bibler, and M. Lavin. 2003. Molecular phylogeny of *Phaseolus*. Abstract. The Seventeenth Biennial Meeting of The Bean Improvement Cooperative. Sacramento, California.

Dilcher, D. L., P. S. Herendeen, and F. Hueber. 1992. Fossil *Acacia* flowers with attached anther glands from Dominican Republic amber. Pages 33–42 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Doubinger, J., and P. Chotin. 1975. Etude palynologique de lignites Tertiaires du basin d'Arauco-Concepción (Chile). Rev. Esp. Micropaleont. 7:549–565.

Engel, M. S. 2001. Monophyly and extensive extinction of advanced eusocial bees: Insights from an unexpected Eocene diversity. Proc. Nat. Acad. Sci. USA 98:1661–1664.

Felsenstein, J. 1988. Phylogenies from molecular sequences: Inference and reliability. Ann. Rev. Genet. 22:521–565.

Frederiksen, N. O. 1980. Sporomorphs from the Jackson Group (Upper Eocene) and adjacent strata of Mississippi and western Alabama. U.S. Geological Survey Professional Paper 1084:1–75.

Frederiksen, N. O. 1983. Middle Eocene palynomorphs from San Diego, California. Part II. Angiosperm pollen and miscellanea. American Association Stratigraphic Palynological Contributions Series 12:32–155.

Graham, A. 1976. Studies in Neotropical paleobotany. II. The Miocene communities of Veracruz, Mexico. Ann. Mo. Botanical Garden 63:787–842.

Graham, A. 1992. The current status of the legume fossil record in the Caribbean region. Pages 161–167 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Gros, J. P. 1992. A synopsis of the fossil record of mimosoid legume wood. Pages 69–83 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Herendeen, P. S. 2001. The fossil record of the Leguminosae: Recent advances. *In* Legumes down under: The Fourth International Legume Conference, Abstracts, 34–35. Australian National University, Canberra, Australia.

Herendeen, P. S., A. Bruneau, and G. P. Lewis. 2003a. Phylogenetic relationships in caesalpinioid legumes: A preliminary analysis based on morphological and molecular data. Pages 37–62 *in* Advances in legume systematics, part 10, higher level systematics (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.

Herendeen, P. S., W. L. Crepet, and D. L. Dilcher. 1992. The fossil history of the Leguminosae: Phylogenetic and biogeographic implications. Pages 303–316 *in* Advances in legume systematics, part 4, the fossil record (P. S. Herendeen and D. L. Dilcher, eds.). Royal Botanic Gardens, Kew, UK.

Herendeen, P. S., and D. L. Dilcher. 1990a. Fossil mimosoid legumes from the Eocene and Oligocene of southeastern North America. Rev. Palaeobot. Palynol. 62:339–361.

Herendeen, P. S., and D. L. Dilcher. 1990b. *Diplotropis* (Leguminosae, Papilionoideae) from the Middle Eocene of southeastern North America. Syst. Bot. 15:526–533.

Herendeen, P. S., and D. L. Dilcher. 1991. Caesalpinia subgenus Mezoneuron (Leguminosae, Caesalpinioideae) from the Tertiary of North America. Am. J. Bot. 78:1–12.

Herendeen, P. S., and D. L. Dilcher. 1992. Advances in legume systematics, part 4. The fossil record. Royal Botanic Gardens, Kew, UK.

Herendeen, P. S., and B. F. Jacobs. 2000. Fossil legumes from the Middle Eocene (46.0 Ma) Mahenge Flora of Singida, Tanzania. Am. J. Bot. 87:1358–1366.

Herendeen, P. S., G. P. Lewis, and A. Bruneau. 2003b. Floral morphology in caesalpinioid legumes: Testing the monophyly of the "Umtiza clade." Int. J. Plant Sci. 164:S393–S407.

Herendeen, P. S., and S. Wing. 2001. Papilionoid legume fruits and leaves from the Paleocene of northwestern Wyoming. Botany 2001 Abstracts, published by Botanical Society of America (http://www.botany2001.org/).

Herendeen, P. S., and J. L. Zarucchi. 1990. Validation of *Caesalpinia* subgenus *Mezoneuron* (Desf.) Vidal and new combinations in *Caesalpinia* for two species of *Mezoneuron* from Africa. Ann. Mo. Bot. Garden 77:854–855.

Heusser, C. J. 1971. Pollen and spores of Chile. University of Arizona Press, Tucson.

Hilu, K. W., T. Borsch, K. Müller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, M. P. Powell, L. A. Alice, R. Evans, H. Sauquet, C. Neinhuis, T. A. B. Slotta, J. G. Rohwer, C. S. Campbell, and L. W. Chatrou. 2003. Angiosperm phylogeny based on *matK* sequence information. Am. J. Bot. 90:1758–1776.

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

Hueber, F. M., and J. Langenheim. 1986. Dominican amber tree had African ancestors. Geotimes 31:8–10.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Ann. Rev. Ecol. Sys. 28:437–466.

Huelsenbeck, J. P., and F. R. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754.

Huelsenbeck, J. P., R. Ronquist, R. Nielson, and J. P. Bollback. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294:2310–2314.

Hughes, C. E., G. P. Lewis, A. Daza Yamona, and C. Reynel. 2004. *Maraniona*, a new dalbergioid legume genus (Leguminosae, Papilionoideae) from Peru. Sys. Bot. 29:366–374.

Iturralde-Vinent, M. A., and R. D. E. MacPhee. 1996. Age and paleogeographical origin of Dominican amber. *Science* 273:1850–1852.

Jacobs, B. F. 2003. The plant fossil record and implications for phytogeography in tropical Africa. XVIIth AETFAT Congress Abstracts: 47. Addis Ababa University, Addis Ababa.

Johnson, J. B., and K. S. Omland. 2004. Model selection in ecology and evolution. Trends Ecol. Evol. 19:101–108.

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

Krell, F.-T., and P. S. Cranston. 2004. Which side of the tree is more basal? Sys. Entomol. 29:279–281.

Kruse, H. O. 1954. Some Eocene dicotyledonous woods from the Eden Valley, Wyoming. Ohio J. Sci. 54:243–268.

Langley, C. H., and W. Fitch. 1974. An estimation of the constancy of the rate of molecular evolution. J. Mol. Evol. 3:161–177.

Lavin, M., R. T. Pennington, B. B. Klitgaard, J. I. Sprent, H. C. De Lima, and P. E. Gasson. 2001. The dalbergioid legumes (Fabaceae): Delimitation of a monophyletic pantropical clade. Am. J. Bot. 88:503–533.

Lavin, M., B. D. Schrire, G. P. Lewis, R. T. Pennington, A. Delgado-Salinas, M. Thulin, C. E. Hughes, A. Beyra Matos, and M. F. Wojciechowski. 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. Phil. Trans. R. Soc. Biol. Sci. 359(1450):1509–1522.

Lavin, M., M. F. Wojciechowski, P. Gasson, C. E. Hughes, and E. Wheeler. 2003. Phylogeny of robinioid legumes (Fabaceae) revisited: *Coursetia* and *Gliricidia* recircumscribed, and a biogeographical appraisal of the Caribbean endemics. Sys. Bot. 28:387–409.

Leidelmeyer, P. 1966. The Paleocene and Lower Eocene pollen flora of Guyana. Leidse Geologische Medelingen 36:49–70.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

Leopold, E. B., and H. D. MacGinitie. 1972. Development and affinities of Tertiary floras in the Rocky Mountains. Pages 147–200 *in* Floristics and paleofloristics of Asia and eastern North America (A. Graham, ed.). Elsevier Publishing Company, Amsterdam.

Lewis, G., B. Schrire, B. Mackinder, and M. Lock (eds.). 2005. Page 592 *in* Legumes of the world. Royal Botanic Gardens, Kew.

Luckow, M., J. T. Miller, D. J. Murphy, and T. Livshultz. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. Pages 197–220 *in* Advances in legume systematics, part 10, higher level systematics (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.

MacGinitie, H. D. 1953. Fossil plants of the Florissant Beds, Colorado. Carnegie Institution of Washington Publications. 599:i–iii, 1–198.

Magallón, S., and M. J. Sanderson. 2001. Absolute diversification rates in angiosperms clades. Evolution 55:1762–1780.

Manchester, S. R. 2001. Update on the megafossil flora of Florissant, Colorado. Denver Museum of Nature & Science, Series 4, 1:137–161.

Manchester, S. R., K. B. Pigg, and P. R. Crane. 2004. Palaeocarpinus dakotaensis sp. nov. (Betulaceae: Coryloideae) and associated staminate catkins, pollen and leaves from the Paleocene of North Dakota. Int. J. Plant Sci. 165:1135–1148.

Manos, P. S., and A. M. Stanford. 2001. The biogeography of Fagaceae: Tracking the Tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int. J. Plant Sci. 162:S77–S93.

McClain, A. M. and S. R. Manchester. 2001. *Dipteronia* (Sapindaceae) from the Tertiary of North America and implications for the phytogeographic history of the Aceroideae. Am. J. Bot. 88:1316–1325.

McKey, D. 1994. Legumes and nitrogen: The evolutionary ecology of a nitrogen-demanding lifestyle. Pages 211–228 *in* Advances in legume systematics, part 5, the nitrogen factor (J. I. Sprent and D McKey, eds.). Royal Botanic Gardens, Kew, UK.

Miller, J. T., J. W. Grimes, D. J. Murphy, R. J. Bayer, and P. Y. Ladiges. 2003. A phylogenetic analysis of the Acacieae and Ingeae (Mimosoideae: Fabaceae) based on *trnK*, *matK*, *psbA*, *trnH*, and *trnL/trnF* sequence data. Syst. Bot. 28:558–566.

Muller, J. 1981. Fossil pollen records of extant angiosperms. Bot. Rev. 47:1–142.

Neel, M. C., and M. P. Cummings. 2004. Section-level relationships of North American *Agalinis* (Orobanchaceae) based on DNA sequence analysis of three chloroplast gene regions. BMC Evol. Biol. 4:15.

Osaloo, S. K., and S. Kawano. 1999. Molecular systematics of Trilliaceae II. Phylogenetic analyses of *Trillium* and its allies using sequences of *rbcL* and *matK* genes of cpDNA and internal transcribed spacers of 18S–26S nrDNA. Plant Species Bio. 14:75–94.

Pennington, R. T., M. Lavin, H. Ireland, B. Klitgaard, J. Preston, and J.-M. Hu. 2001. Phylogenetic relationships of basal papilionoid legumes based upon sequence of the chloroplast *trn*L intron. Syst. Bot. 26:537-556.

Percy, D. M., R. D. M. Page, and Q. C. B. Cronk. 2004. Plant-insect interactions: Double-dating associated insect and plant lineage reveals asynchrous radiations. Syst. Biol. 53:120–127.

Pigg, K. B., M. F. Wojciechowski, and M. L. DeVore. 2004. Samaras from the Late Paleocene Almont and Beicegel Creek floras of North Dakota, U.S.A., with potential affinities to *Securidaca* (Polygalaceae). Botany 2004, abstract.

Poinar, G. O., Jr. 1991. *Hymenaea protera* sp.n. (Leguminosae, Caesalpinioideae) from Dominican amber has African affinities. Experientia 47:1075–1082.

Poinar, G. O., Jr., and A. E. Brown. 2002. *Hymenaea mexicana* sp. nov. (Leguminosae: Caesalpinioideae) from Mexican amber indicates Old World connections. Bot. J. Linn. Soc. 139:125–132.

Poinar, H. N., R. J. Cano, and G. O. Poinar, Jr. 1993. DNA from an extinct plant. Nature 363:677.

Polhill, R. M. 1994. Classification of the Leguminosae. Pages xxxv–xlviii *in* Phytochemical dictionary of the Leguminosae (F. A. Bisby, J. Buckingham, and J. B. Harborne, eds.). Chapman and Hall, New York.

Polhill, R. M., P. H. Raven, and C. H. Stirton. 1981. Evolution and systematics of the Leguminosae. Pages 1–26 *in* Advances in legume systematics, part 1 (R. M. Polhill and P. H. Raven, eds.). Royal Botanic Gardens, Kew, UK.

Posada, D., and K. A. Crandall. 1998. ModelTest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

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

Sah, S. C. D., and K. Dutta. 1968. Palynostratigraphy of the Tertiary sedimentary formations of Assam. 2. Stratigraphic significance of spores and pollen in the Tertiary succession of Assam. Palaeobotanist 16:177–195.

Sanderson, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14:1218–1231.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Mol. Biol. Evol. 19:101–109.

Sanderson, M. J. 2003. r8s, version 1.6, User's Manual (April 2003). Distributed by the author (http://ginger.ucdavis.edu/r8s/). University of California, Davis.

Schneider, H., E. Schuettpelz, K. M. Pryer, R. Cranfill, S. Magallón, and R. Lupia. 2004. Ferns diversified in the shadow of angiosperms. Nature 428:553–557.

Schrire, B. D., M. Lavin, N. P. Barker, H. Cortes-Burns, I. von Senger and J.-H. Kim. 2003. Towards a phylogeny of *Indigofera* (Leguminosae-Papilionoideae): Identification of major clades and relative ages. Pages 269–302 *in* Advances in legume systematics, part 10, higher level systematics (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.

Soltis, D. E., P. S. Soltis, M. W. Chase, M. E. Mort, D. C. Albach, et al. 2000. Angiosperm phylogeny inferred from 18S rDNA, *rbcL*, and *atpB* sequences. Bot. J. Linn. Soc. 133:381–461.

Sprent, J. I. 2001. Nodulation in legumes. Royal Botanic Gardens, Kew, UK.

Steele, K. P., and R. Vilgalys. 1994. Phylogenetic analyses of Polemoniaceae using nucleotide sequences of the plastid gene *matK*. Syst. Bot. 19:126–142.

Taylor, D. W. 1990. Paleobiogeographic relationships from the Cretaceous and Early Tertiary of the North American area. Bot. Rev. 56:279–417.

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

Tucker, S. C., and A. W. Douglas. 1994. Ontogenetic evidence and phylogenetic relationships among basal taxa of legumes. Pages 11–32 *in* Advances in legume systematics, part 6, structural botany (I. K. Ferguson and S. C. Tucker, eds.). Royal Botanic Gardens, Kew, UK.

Wang, X.-R., Y. Tsumura, H. Yoshimaru, K. Nagasaka, and A. E. Szmidt. 1999. Phylogenetic relationships of Eurasian Pines (*Pinus*, Pinaceae) based on chloroplast *rbcL*, *matK*, *rpl20-rps18* spacer, and *trnV* intron sequences. Am. J. Bot. 86:1742–1753.

Webster, A. J., R. J. H. Payne, and M. Pagel. 2003. Molecular phylogenies link rates of evolution and speciation. Science 301:478.

Webster, A. J., R. J. H. Payne, and M. Pagel. 2004. Response to comments on "Molecular phylogenies link rates of evolution and speciation." Science 303:173d.

Wikström, N., V. Savolainen, and M. W. Chase. 2001. Evolution of angiosperms: Calibrating the family tree. Proc. R. Soc. B 268:2211–2220.

Wilf, P., and C. C. Labandeira. 1999. Response of plant-insect associations to Paleocene-Eocene warming. Science 284:2153–2156.

Wing, S. L., Herrera, F., and Jaramillo, C. 2004. A Paleocene flora from the Cerrajón Formation, Guajíra Peninsula, northeastern Colombia. VII International Organization of Paleobotany Conference Abstracts, pp. 146–147 (21–26 March). Museo Egidio Feruglio, Trelew, Argentina.

Wojciechowski, M. F. 2003. Reconstructing the phylogeny of legumes (Leguminosae): An early 21st century perspective. Pages 5–35 *in* Advances in legume systematics, part 10, Higher level systematics (B. B. Klitgaard and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.

Wojciechowski, M. F., M. Lavin, and M. J. Sanderson. 2004. A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. Am. J. Bot. 91:1846–1862.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

Wojciechowski, M. F., M. J. Sanderson, K. P. Steele, and A. Liston. 2000. Molecular phylogeny of the "temperate herbaceous tribes" of papilionoid legumes: A supertree approach. Pages 277–298 *in* Advances in legume systematics, part 9 (P. S. Herendeen and A. Bruneau, eds.). Royal Botanic Gardens, Kew, UK.

Wolfe J. A., and T. Tanai. 1987. Systematics, phylogeny, and distribution of *Acer* (maples) in the Cenozoic of western North America. Journal of the Faculty of Science of Hokkaido University Series IV 22:1–10.

Xiang, Q.-Y., D. E. Soltis, P. S. Soltis, S. R. Manchester, and D. J. Crawford. 2000. Timing the eastern Asian–eastern North American floristic disjunction: Molecular clock corroborates paleontological estimates. Mol. Phylogenet. Evol. 15:462–472.

Yang, Z., and A. D. Yoder. 2003. Comparison of likelihood and Bayesian methods for estimating divergence times using multiple gene loci and calibration points, with application to a radiation of cute-looking mouse legume species. Syst. Biol. 15:705–716.

Yoder, A. D., and Z. Yang. 2004. Divergence dates for Malagasy lemurs estimated from multiple gene loci: Geological and evolutionary context. Mol. Ecol. 13:757–773.

Young, N. D., and C. W. dePamphilis. 2000. Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. Mol. Biol. Evol. 17:1933–1941.

Zhang, L.-B., and S. Renner. 2003. The deepest splits in Chloranthaceae as resolved by chloroplast sequences. Int. J. Plant Sci. 164(5 Suppl.):S383–S392.

First submitted 3 June 2004; reviews returned 31 August 2004; final acceptance 18 November 2004
Associate Editor: Peter Linder

APPENDIX

APPENDIX 1: THE 13 FOSSIL CONSTRAINTS IMPOSED DURING THIS STUDY

The legume family represents an ideal group for conducting an analysis of evolutionary rates and ages because the family has a high detectability in the fossil record. A rapid Paleocene appearance means that the duration of legume evolution probably includes just the last 60 to 65 Ma. The high detectability of legumes in the fossil record is postulated by the following evidence: (1) the characteristic nitrogen-rich metabolism predisposes legumes to high biomass production (McKey, 1994); (2) legumes are colonizers of disturbed sites close to depositional environments; (3) woody legume plants often produce abundant deciduous leaves, leaflets, flowers, or fruits; (4) these deciduous parts are often coriaceous or woody; and (5) legumes are taxonomically easily recognized in the fossil record (e.g., the cross-hatchings evident on the pulvinules of leaflets or pulvinus of leaves, compound leaves, leaflets often with a smooth margin and brochidodromous venation, and pod with single placental suture bearing a characteristic alternating branching pattern of the funiculi). The relationship of this high detectability to the continuous Cenozoic fossil record that abruptly begins by the Middle to Late Paleocene is being evaluated in a separate study with mark-recapture models (Rotella and Lavin, in preparation).

(A) The Leguminosae stem clade was fixed initially at 60 Ma, but also at every one million year interval back to 70 Ma. Only this age was fixed because a fixed time span facilitates rates and age estimation in all relaxed clock methods (Sanderson, personal communication). Furthermore, the temporally continuous and spatially extensive distribution of diverse fossil legumes from the Recent back to the Late Paleocene supports a hypothesized Early to Middle Paleocene origin of the family. The legume stem clade is defined as the most recent common ancestor (MRCA) of *Polygala californica* and *Cercis occidentalis* in the *matK* phylogeny (Fig. 1) and *Polygala cruciata* and *Glycine tabacina* in the *rbcL* phylogeny (not shown, but see Kajita et al., 2001). The most important pattern derived from the study of fossil legumes during the past several decades is that all three subfamilies have an abundant fossil record in North America, Europe, and Asia, and to some degree in South America and Africa (due to relatively poor sampling). This abundant record is represented continuously by at least fossil fruits and leaves from the Recent back to the Late Paleocene (e.g., Brown, 1962; Herendeen and Dilcher, 1992; Herendeen, 2001; Jacobs, 2003; and Wing et al., 2004). Earlier reports of legume fossils from as early as the Maastrichtian are tenuous and include only sporadic pollen and wood specimens that lack definitive legume apomorphies. Such legume apomorphies would include a prolate pollen grain with narrow colpi and large operculi, or wood with vestured pits (summarized in Herendeen et al., 1992).

Wikström et al. (2001) posited a 74 to 79-Ma age (Late Cretaceous) for the legumes, and generally old estimates for other angiosperm families. These authors used only nonparametric rate smoothing, and constrained just a single node in the angiosperm phylogeny, the split between the Cucurbitales and the Fagales, which are two of three potential sister groups to Fabales (sensu Angiosperm Phylogeny Group, 2003). Implementation of a single age constraint combined with nonparametric rate-smoothing is bound to yield estimates highly biased toward older ages (Sanderson, 2002). Magallón and Sanderson (2001) used 59.9 Ma for the Fabales crown clade, which is equivalent to the legume stem. To accommodate uncertainty, we fixed the age of the legume stem at 60 Ma, but then compared results using this fixed age with those derived from fixed root ages at every one million year interval between 61 and 70 Ma.

(B) The Cercis stem clade was set to a minimum of 34 Ma (Late Eocene). This node is defined as the MRCA of *Cercis occidentalis* and *Bauhinia tomentosa* in the *matK* phylogeny (Fig. 1) and *Cercis canadensis* and *Bauhinia purpurea* in the *rbcL* phylogeny (Kajita et al., 2001). Fossils showing apomorphic traits of *Cercis* come from the Late Eocene of Sheep Rock Creek, Oregon (Herendeen and S. Manchester, unpublished data; site information provided in Wolfe and Tanai, 1987; McClain and Manchester, 2001). Fossils from this locality reveal a combination of apomorphies allowing assignment to *Cercis*, including unifoliolate leaves (i.e., with a lower pulvinus and upper pulvini), a thickish texture to the orbiculate-acuminate lamina, and a palmate primary venation combined with a pinnate secondary venation. Also found are thin-walled pods bearing a narrow non-vascularized placental wing suggestive of the fruits of *Cercis*. Additional leaf and fruit fossils also assigned to *Cercis* come from the Late Eocene Florissant Beds, Colorado (MacGinite, 1953; Manchester, 2001). Several types of legume fruits at Florissant have been identified as *Cercis*, however; and further study is required before these fossils can be definitively utilized here.

(C) The Hymenaea stem clade was set to a minimum of 34 Ma. This node is defined as the MRCA of *Hymenaea courbaril* and *Tessmannia lescrauwaetii* in the *matK* phylogeny (Fig. 1) and is undefined in the *rbcL* phylogeny. Poinar and Brown (2002) described *Hymenaea mexicana* from flowers preserved in amber with an estimated age of 22.5 to 26.0 Ma from Chiapas, Mexico. Hueber and Langenheim (1986) and Poinar (1991) place the fossil *Hymenaea* flower from La Toca Mines amber (Dominican Republic), formally named *Hymenaea protera*, as sister to the African *Hymenaea verrucosa*. Graham (1992) validated *H. protera* as bona fide *Hymenaea* with a minimum age of Late Eocene or 34 Ma. Poinar et al. (1993) further validated the generic assignment of *H. protera* with "fossil" *rbcL* sequence data. Only the *Hymenaea* crown clade, however, is well supported in their *rbcL* study. As it turns out, the *Hymenaea rbcL* sequences are small partial fragments, and contain certain sequence anomalies that need to be verified. In addition, proper outgroups were not sampled in Poinar et al. (1993), so the distinction of the *Hymenaea* stem and crown cannot be made. The above assignment, therefore, is the most precise possible given inadequate molecular sampling of *Hymenaea* and closely related genera.

(D) The Arcoa stem (or Umtiza crown; Herendeen et al., 2003a, 2003b) clade was set to a minimum of 34 Ma. This node is defined as the MRCA of *Arcoa gonavensis* and *Gleditsia triacanthos* in the *matK* phylogeny (Fig. 1) and *Acrocarpus* sp. (Genbank AF308699) and *Gleditsia triacanthos* in the *rbcL* phylogeny (Kajita et al., 2001). Fossil leaves from the Florissant Beds, Colorado (MacGinitie, 1953) referred to as "*Prosopis linearifolia*" reveal a mix of pinnate leaves

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

and a bipinnate leaf where the pinnate leaves are larger than the bipinnate one. The leaflets are distinctly linear and asymmetric. Furthermore, in the single bipinnate leaf, the terminal group of three pinnae results from a sessile terminal pinna (e.g., specimen USNM40563). These leaf characteristic are very similar to those of *Arcoa* (now endemic to southern Hispaniola). The bipinnate leaf bearing a distinctively sessile terminal pinna is also found in *Acrocarpus* and *Tetrapterocarpon*, although leaflet morphology is considerably different in the latter two genera. These three genera plus *Ceratonia*, *Gleditsia*, *Gymnocladus*, and *Umtiza* constitute the Umtiza clade (Herendeen et al., 2003b). Notably, dehiscent pods winged along both sutures, very similar to those of *Acrocarpus*, are found in the Middle Eocene Bovay Clay pit, Tennessee (Herendeen, 1992). The combination of two wings and pod dehiscence, however, is not definitive for constraining the Umtiza crown to the earlier age of 40 Ma.

(E) The Mezoneuron stem clade was set to a minimum of 45 Ma. This node is defined as the MRCA of *Caesalpinia* (subgenus *Mezoneuron*) *andamanica* and *Haematoxylum brasiletto* in the *matK* phylogeny (Fig. 1) and is undefined in the *rbcL* phylogeny (Kajita et al., 2001). Several species of *Mezoneuron* (or *Caesalpinia* subg. *Mezoneuron*) are reported from North America and England as Middle Eocene to Miocene in age, even though *Mezoneuron* is presently confined to the Paleotropics (Herendeen and Zarucchi, 1990; Herendeen and Dilcher, 1991). The apomorphic traits that allow assignment of these fossils to the Mezoneuron stem include a membranous indehiscent pod with multiple ovules, and a broad placental wing with looping to longitudinal venation.

(F) The mimosoid stem clade was set to a minimum of 55 Ma. This node is defined as the MRCA of *Inga punctata* and *Cercidium floridum* in the *matK* phylogeny (Fig. 1) and *Erythrophleum ivorense* and *Albizia saman* in the *rbcL* phylogeny (Kajita et al., 2001). The fossil flower *Protomimosoidea buchananensis* (Crepet and Taylor, 1985, 1986) comes from the Wilcox Formation of western Tennessee, which has an age estimated at the Paleocene-Eocene boundary (55 Ma). The apomorphic characters that allow placement of this fossil to the mimosoid stem include a spicate inflorescence, actinomorphic bisexual flowers, valvate petals, tubular stigmas, and 10 exserted free stamens. Crepet and Taylor (1986) suggested an affinity to the tribe Mimoseae, but this was explicitly because "the fossil flowers lack the distinctive taxonomic characters that define the other tribes of mimosoid legumes. . . ." In essence, the apomorphically diagnostic traits of these fossil flowers represent generalized mimosoid attributes and thus *Protomimosoidea buchananensis* constrains the mimosoid stem. Regardless, there are numerous other fossil leaves and fruits that collectively represent generalized mimosoid legumes from the earliest Eocene (e.g., Gros, 1992; Herendeen and Dilcher, 1990a; Herendeen and Jacobs, 2000). Notably, leaves, inflorescences, and fruits of a fossil species very similar to *Dinizia* are widespread from the southeastern USA during the Eocene (Herendeen and Dilcher, 1990a). Although *Dinizia* is no longer considered a mimosoid legume (Luckow et al., 2003), the Eocene fossils potentially referred to this genus could impose the same time constraints as the Protomimosoideae fossils. This is because the *Dinizia* stem clade is coeval with the mimosoid stem clade in the *matK* phylogeny (see Fig. 1).

(G) The Acacia stem clade was set to a minimum of 15 Ma (middle Miocene). This node is defined as the MRCA of *Acacia hindsii* and *Inga punctata* in the *matK* phylogeny (Fig. 1) and *Acacia hindsii* and *Mimosa speggazzini* in the *rbcL* phylogeny (Kajita et al., 2001). Dilcher et al. (1992) detailed the floral and leaflet morphology of specimens preserved in Dominican Republic amber from the Palo Alto mine. These mimosoid flowers come from sediments dated at the Oligocene-Miocene boundary. Dilcher et al. (1992) argue that the amber has been weathered out of older sediments and redeposited in this Oligocene-Miocene sediment. In contrast, Iturralde-Vinent and MacPhee (1996) presented evidence for the deposition of Dominican Amber during an interval 15 to 20 million years ago. Thus, the minimum age of the Acacia stem clade is 15 Ma. The assignment of the amber fossils to the *Acacia* stem (i.e., the first branching *Acacia*-containing lineage within the mimosoid crown) can be made because these fossils clearly show diagnostic apomorphic characters, such as numerous stamens with free filaments each with an anther bearing a stalked gland. Associated leaf pinnae bearing numerous small leaflets by themselves are not diagnostic, but they bear hairs identical to those on the flowers. The amber *Acacia* flowers are the only mimosoid fossils that unequivocally constrain a node nested within the mimosoid crown.

Herendeen and Jacobs (2000) described *Acacia mahengense* from Middle Eocene (46 Ma) fossil leaves in Tanzania. This taxonomic assignment was derived from morphological similarity rather than definitive apomorphies. This is problematic from the perspective of this study given that *Acacia*, as traditionally circumscribed, contains at least several unrelated mimosoid lineages (Luckow et al., 2003; Miller et al., 2003).

(H) The Papilionoideae stem clade was set to a minimum of 55 Ma (Late Paleocene). This node is defined as the MRCA of *Amburana cearensis* and *Cercidium floridum* in the *matK* phylogeny (Fig. 1) and *Peltogyne confertiflora* and *Glycine tabacina* in the *rbcL* phylogeny (Kajita et al., 2001). The flowers of *Barnebyanthus buchananensis* are the earliest generalized papilionoid fossils from the Late Paleocene to Early Eocene sediments of the Buchanan Clay Pit, western Tennessee (Crepet and Herendeen, 1992). In addition to bilateral symmetry, these fossil flowers clearly show a connate calyx with five small calyx lobes, an adaxial median sepal, a differentiated standard petal that is positioned outside the wing petals, and distinct wing and keel petals, the latter of which are not fused. Crepet and Herendeen included *Barnebyanthus* in a cladistic analysis of traits scored also for selected genera of the tribe Sophoreae and concluded that *Barnebyanthus* was sister to the genistoid genera *Bowringia* and *Clathrotropis*. Although these fossil flowers show synapomorphies associated with the basic papilionoid zygomorphic flower, they actually lack other floral synapomorphies (e.g., fused keel petals, connate filaments) that could serve to unequivocally constrain their position within the papilionoid crown clade. Thus, the fossils are consistent with several lineages within "Sophoreae" and other papilionoids with free petals and filaments. This analysis of Crepet and Herendeen illustrates that legume fossils, because they include mainly disarticulated leaves, leaflets, flowers, and fruits, generally comprise few apomorphic characters.

The minimum age of 55 Ma for the papilionoid stem clade is substantiated by a set of fossils that are suggestive of specific papilionoid groups. For example, fossils representing the lineage containing *Diplotropis* and *Bowdichia* (see node J, below) have an estimated age at around 56 Ma. Fossil leaves and fruits possibly representing *Ormosia*, are found in Colombia with an estimated age somewhere between 55 and 60 Ma (Wing et al., 2004).

(I) The Styphnolobium stem clade was set to a minimum of 40 Ma (Middle Eocene). This node is defined as the MRCA of *Styphnolobium japonicum* and *Pickeringia montana* in the *matK* phylogeny (Fig. 1) and is undefined in the *rbcL* phylogeny. Because both *Styphnolobium* and *Cladrastis* subgenus *Cladrastis* (here represented by *Cladrastis delavayi*, *C. lutea*, and *C. sinensis*) have a Middle Eocene fossil record from Tennessee (Herendeen, 1992), the stem leading to these two clades is constrained. Apomorphic traits that strongly suggest an affinity of these fossils with subgenus *Cladrastis* include a stipitate indehiscent fruit with thin pod walls and an attenuated nonplacental margin, a narrow wing along the placental suture, and seeds that are oriented longitudinally. Leaves and leaflets highly similar to those of subgenus *Cladrastis* co-occur with these same legume fruits in the Puryear Clay pit. Apomorphic traits that show an affinity of a single pod from the Miller Clay pit with *Styphnolobium* include indehiscent moniliform valves in which the constrictions between seeds narrow to less than half the diameter of the widest part of the pod. Most importantly, the seed-bearing portion is rounded in outline and longitudinally wrinkled, which suggests a diagnostic fleshy pericarp. The fossil fruit of *Cladrastis* subgenus *Platycarpus* from the Oligocene Bridge Creek flora (Brown, 1937) is superfluous to the *Cladrastis* subgenus *Cladrastis* and *Styphnolobium* fossils.

(J) The Diplotropis stem clade (essentially the Genistoid crown clade) was set to a minimum of 56 Ma, which is the oldest record for a definitive legume fossil. This node is defined as the MRCA of

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024

Diplotropis brasiliensis and *Spartium junceum* in the *matK* phylogeny (Fig. 2) and *Acosmium dasycarpum* and *Spartium junceum* in the *rbcL* phylogeny (Kajita et al., 2001). Fossil leaves and pods very similar to *Bowdichia* and *Diplotropis* are known from the Late Paleocene of western Wyoming (Herendeen and Wing, 2001), as well as the Middle Eocene of southeastern USA (Herendeen and Dilcher, 1990b). The apomorphic traits that suggest an affinity of these fossils to *Bowdichia* and *Diplotropis* include membranous pod valves, numerous seeds that are transversally oriented, a narrow wing on the placental suture, strong lateral nerves that demarcate the body (seed-containing region) from the placental and lower margins, and leaflets variably alternate to opposite (*Calpurnia* and *Maackia* of this same clade also have similar fruits). In the *matK* phylogeny (Fig. 2), the difference between the Bowdichia-Diplotropis stem and the Genistoid crown clade is trivial such that assignment of these fossils to one or the other node makes little difference with respect to rate and age estimation.

(K) The Machaerium stem clade was set to a minimum of 40 Ma. This node is defined as the MRCA of *Aeschynomene purpusii* and *Machaerium falciforme* in the *matK* phylogeny (Fig. 2) and *Dalbergia hupeana* and *Machaerium lunatum* in the *rbcL* phylogeny (Kajita et al., 2001). A series of fossil leaflets from the Puryear, Bolden, and Bovay Clay Pits (Herendeen, 1992) have numerous closely spaced craspedodromous secondary veins, a strong marginal vein, and poorly organized higher order venation that are diagnostic of a majority of species in the genus *Machaerium*. Epidermal cell structure fits the range of variation also observed in *Machaerium*.

(L) The Tipuana stem clade was set to a minimum of 10 Ma. This node is defined as the MRCA of *Tipuana tipu* and *Maraniona lavinii* in the *matK* phylogeny (Fig. 2) and is undefined in the *rbcL* phylogeny. Numerous winged fruits come from Loja and Cuenca in southern Ecuador, including *Tipuana ecuatoriana* (Burnham, 1995). This fossil fruit has a diagnostic long stipe, a unilateral wing emanating from mostly the style, parallel and distally spreading venation on wing, weak venation on stylar region of the wing, and seeds that are round in cross section (Burnham, 1995). Hughes et al. (2004) point out that the sessile fruits of *Maraniona*, which lack a pronounced wing venation, are highly similar to those of *Tipuana*, but distinctly lack the apomorphies shared between the fossil and extant species of *Tipuana*.

(M) The Robinia stem clade was set to a minimum of 34 Ma. This node is defined as the MRCA of *Robinia pseudoacacia* and *Coursetia axillaris* in the *matK* phylogeny (Fig. 3) and is undefined in the *rbcL* phylogeny. The fossil woods of *Robinia zirkelii* and the assignment of these to the Robinia stem clade are discussed in Lavin et al. (2003). This fossil wood species was widespread over North America and Europe and possesses such apomorphic traits as vestured intervessel pits, storied axial parenchyma and vessel elements, numerous thin-walled tyloses, ring-porous wood, homocellular rays, and spiral sculpturing in narrow vessels, which make assignment to the Robinia stem clade unequivocal.

Fossil Record of Fabales Other Than Leguminosae

With respect to the Fabales other than legumes (i.e., Polygalaceae, Surianaceae, and the genus *Quillaja*; sensu Angiosperm Phylogeny Group, 2003), fossil representation comes primarily from Polygalaceae, including several pollen records, fossil leaves and fruits currently under study (Kathleen Pigg, Arizona State University, personal communication). The fruits ("non-schizocarpic samaras") from the Late Paleocene Almont and Beicegel Creek (North Dakota) floras show similarities to *Securidaca* (Crane et al., 1990; Pigg et al., 2004). Polygalaceae has distinctive pollen diagnosed by more than four colporate apertures in which endoapertures may fuse. The exine stratification is obscure and surface is almost always smooth (Muller, 1981; Kathleen Pigg, Arizona State University, personal communications). Fossil pollen similar to the extant *Monnina* (Heusser, 1971) is described as *Psilastephanocolporites fissilis* from the Paleocene of Chile (Doubinger and Chotin, 1975; Muller, 1981) and from the Eocene of Guyana (Leidelmeyer, 1966). *Polygalacidites calrus* pollen is reported from the lower and upper Eocene of Assam, India (Sah and Dutta, 1968). *Securidaca bombacopsis* (Leopold and MacGinitie, 1972) is known from the Middle Eocene of North America, and pollen of *Securidaca* and a taxon similar to *Bredemeyera* are known from the upper Miocene of Mexico (Graham, 1976). Additional reports of polygalaceous pollen come from the Eocene of southeastern North America and California (Frederiksen, 1980, 1983; Taylor, 1990; Pigg et al., 2004).

These fossils are assigned to their taxonomic categories by using overall morphological similarity. Additionally, we did not sample Polygalaceae with respect to these fossils, so adding minimum constraints to nodes in the outgroups was not possible. Regardless, a minimum Late Paleocene age (Almont and Beicegel Creek; Crane et al., 1990; Manchester et al., in press; Pigg et al., 2004) assigned to the Polygalaceae stem clade (i.e., 55 Ma) does not affect the results presented because this clade is already constrained by older fixed ages (60 to 70 Ma) that are assigned to the coeval legume stem clade.

Surianaceae is now restricted to areas with subtropical to tropical coastal climates, but fossil wood with affinities to *Suriana* has been reported from the Eocene of Wyoming (Kruse, 1954). The tenuous taxonomic assignment of this fossil renders any age constraint dubious.

Potential fossil constraints may eventually come from groups related to Fabales. Wikstrom et al. (2001) assigned a 84-Ma fossil constraint to the split between Fagales and Cucurbitales (i.e., the Fagales stem clade). The three gene (Soltis et al., 2000) and *matK* (Hilu et al., 2003) phylogenies of angiosperms provide little resolution of the clades involving Fabales, Fagales, Cucurbitales, and Rosales. Fagales is sister to Cucurbitales in Soltis et al. (2000), but this larger clade is unresolved with respect to Rosales and Fabales. The Cucurbitales is sister to the other three orders in the *matK* strict consensus but with no clade support (Hilu et al., 2003). The uncertainty of the interrelationships among Fabales and related eurosid clades render assignments of the fossil constraint by Wikstrom et al. (2001) uncertain and in need of a separate study detailing the morphologies of extinct and extant eurosid taxa.

Downloaded from https://academic.oup.com/sysbio/article/54/4/575/2842908 by guest on 25 April 2024