



Contents lists available at ScienceDirect

South African Journal of Botany

journal homepage: www.elsevier.com/locate/sajb

Exploring the tempo of species diversification in legumes

E.J.M. Koenen^{a,1}, J.M. de Vos^{a,1,2}, G.W. Atchison^a, M.F. Simon^b, B.D. Schrire^c, E.R. de Souza^d, L.P. de Queiroz^d, C.E. Hughes^{a,*}^a Institute of Systematic Botany, University of Zurich, Zollikerstrasse 107, 8008 Zürich, Switzerland^b Embrapa Recursos Genéticos e Biotecnologia, PqEB, Caixa Postal 02372 Brasília-DF, Brasil^c Herbarium, Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AB, UK^d Universidade Estadual de Feira de Santana, Dept. de Ciências Biológicas, Feira de Santana, Bahia, Brasil

ARTICLE INFO

Available online 12 August 2013

Edited by JS Boatwright

Keywords:

Species diversification

Leguminosae

Calliandra

Indigofereae

*Lupinus**Mimosa*

Robinieae

Diversification rate shift

Speciation rate

Extinction rate

ABSTRACT

Whatever criteria are used to measure evolutionary success – species numbers, geographic range, ecological abundance, ecological and life history diversity, background diversification rates, or the presence of rapidly evolving clades – the legume family is one of the most successful lineages of flowering plants. Despite this, we still know rather little about the dynamics of lineage and species diversification across the family through the Cenozoic, or about the underlying drivers of diversification. There have been few attempts to estimate net species diversification rates or underlying speciation and extinction rates for legume clades, to test whether among-lineage variation in diversification rates deviates from null expectations, or to locate species diversification rate shifts on specific branches of the legume phylogenetic tree. In this study, time-calibrated phylogenetic trees for a set of species-rich legume clades – *Calliandra*, Indigofereae, *Lupinus*, *Mimosa* and Robinieae – and for the legume family as a whole, are used to explore how we might approach these questions. These clades are analysed using recently developed maximum likelihood and Bayesian methods to detect species diversification rate shifts and test for among-lineage variation in speciation, extinction and net diversification rates. Possible explanations for rate shifts in terms of extrinsic factors and/or intrinsic trait evolution are discussed. In addition, several methodological issues and limitations associated with these analyses are highlighted emphasizing the potential to improve our understanding of the evolutionary dynamics of legume diversification by using much more densely sampled phylogenetic trees that integrate information across broad taxonomic, geographical and temporal levels.

© 2013 SAAB. Published by Elsevier B.V. All rights reserved.

1. Introduction

The legume family (Leguminosae) is one of the most evolutionarily successful lineages of flowering plants. With c. 19,500 species and 750 genera (Lewis et al., 2005 and additions), it is the third largest plant family; it occupies a global distribution spanning all major biomes (temperate, Mediterranean, dry and wet tropical forest, savanna) (Schrire et al., 2005); it presents spectacular morphological and life history diversity, from giant rainforest trees and woody lianas, to desert shrubs, ephemeral herbs, herbaceous twining climbers, aquatics and fire-adapted savanna species (Doyle and Luckow, 2003); it shows a significantly higher than average species diversification rate over the last 60 Ma than angiosperms as a whole, despite the relatively old age of the family (Magallón and Sanderson, 2001); it forms a high proportion of overall vegetation both

in the fraction of overall species composition and abundance of individuals, especially in tropical biomes (Pennington et al., 2006, 2009); finally, it harbours the largest genus of flowering plants, *Astragalus* (Sanderson and Wojciechowski, 1996), and some of the most rapidly evolving plant clades (Hughes and Eastwood, 2006; Richardson et al., 2001; Scherson et al., 2008). Despite this, we still know rather little about the dynamics of lineage and species diversification across the family through the Cenozoic, or about the underlying drivers of diversification.

As for any large plant clade, there are numerous factors, intrinsic and extrinsic, that could be contributing to the evolutionary success of the legumes. For example, plant–animal and plant–microbe interactions (e.g. Kursar et al., 2009; Marazzi and Sanderson, 2010; McKey, 1989; Sprent, 2001), diversity of habit and other life history traits (e.g. Drummond et al., 2012), ecological adaptability, nitrogen fixation (Doyle, 2011; Sprent, 2001), diversity of sexual reproductive systems, and propensity for polyploidy and hybridization (Goldblatt, 1981; Cannon et al., 2010), have all been implicated in diversification of particular legume clades. However, no attempts have been made to estimate diversification rates across legumes, to correlate these with geography, ecology and other traits, or to compare patterns of diversification across the numerous radiations apparent within the family. Indeed, perhaps

* Corresponding author. Tel.: +41 446348441.

E-mail address: colin.hughes@systbot.uzh.ch (C.E. Hughes).¹ These authors contributed equally to this study.² Present address: Department of Ecology and Evolutionary Biology, Brown University, 80 Waterman St., Box G-W, Providence, RI 02912, USA.

surprisingly, there have been no attempts to assess the factors that might explain the extraordinary diversity of legumes, including nodulation, perhaps the most obvious candidate key evolutionary innovation in the family (Doyle, 2011).

Estimating species diversification rates and locating diversification rate shifts depend on knowledge of phylogenetic relationships, divergence time estimates and the distribution of species richness across the phylogeny. For legumes, knowledge in all three of these areas has reached a point where exploration of the dynamics of species diversification is possible. First, several family-wide phylogenies (Lavin et al., 2005; Legume Phylogeny Working Group, 2013; Simon et al., 2009; Wojciechowski et al., 2004) and species-level phylogenies for important species-rich clades (e.g. Drummond et al., 2012; Lavin et al., 2003; Schrire et al., 2009; Simon et al., 2011; Souza et al., submitted for publication) are now available. Second, the rich legume fossil record (Herendeen and Dilcher, 1992) affords exceptional opportunities for robustly cross-validated divergence time estimation using multiple, stringently selected fossil constraints (Bruneau et al., 2008; Lavin et al., 2005; Simon et al., 2009). Third, documentation of taxonomic diversity has advanced, most notably with a generic encyclopaedia of the family (Lewis et al., 2005), providing a first approximation of the spectacular variation in species richness across legume lineages, with genus sizes ranging from monospecific (192), 2–10 species (304), 11–99 species (190), 100–499 species (36), and >500 species (5 genera), presenting a classical example of the hollow curve (Scotland and Sanderson, 2004). Thus, in many respects legumes provide an ideal study group for investigating the macroevolutionary dynamics of plant diversification.

At the same time, there have been rapid advances in methods for estimating the extent of among-lineage variation in species diversification rates and detecting the phylogenetic location of shifts in rates of diversification (Stadler, 2013). Early attempts to test for differences in species diversity among lineages relied on sister group comparisons (Sanderson and Wojciechowski, 1996; Slowinski and Guyer, 1989). These methods were used in legumes to test whether the genus *Astragalus*, the largest genus of flowering plants with c. 2500 species, is in fact exceptionally species-rich (Sanderson and Wojciechowski, 1996). However, such methods have limited statistical power and cannot compare rates across a phylogeny. More sophisticated likelihood approaches test whether clade species richness is greater than expected against background diversification rates while incorporating effects of extinction rates (Magallón and Sanderson, 2001). These methods are derived from 'birth-death' models that assume constant rates of speciation and extinction among lineages, and through time (Nee et al., 1994). For example, Marazzi and Sanderson (2010) showed that the extrafloral nectary clade in the species-rich caesalpinoid genus *Senna* is more species-rich than expected and suggested that extrafloral nectaries could have acted as a key evolutionary innovation facilitating rapid species diversification in this group. Lineage Through Time (LTT) plots provide a simple graphical method to project and compare temporal trajectories of lineage diversification across whole clades or particular subclades, but again, with some notable exceptions, these have been little investigated within legumes. LTT plots compiled for the large Mirbelieae/Bossieae clade and for Podalyrieae (Crisp and Cook, 2009; Schnitzler et al., 2011) revealed repeated time-coincident antisigmoidal LTT curves indicative of either mass-extinction events, concurrent shifts to increased rates of diversification (Crisp and Cook, 2009), or high species-turnover throughout the history of clades, with similar results apparent for the genus *Prosopis* (Catalano et al., 2008), and within North American *Pediomelum* (Egan and Crandall, 2008). More recent likelihood methods and Bayesian implementations thereof relax the assumption of constant diversification across the phylogeny providing greater power to both discover possible rate shifts without any a priori hypothesis as to where they may lie on the tree (Alfaro et al., 2009; Santini et al., 2009) and to test their significance (Silvestro et al., 2011). In legumes these methods have so far only been applied to the genus *Lupinus* (Silvestro et al., 2011; Drummond et al., 2012).

In this study we test the hypothesis that there is significant among-lineage diversification rate variation across the Leguminosae. Analyses of diversification rates for a sparsely sampled higher-level phylogeny of legumes as a whole, and a series of five more densely sampled species-level phylogenies of individual legume clades are presented, in order to gain preliminary insights into the extent of among-lineage variation in diversification rates across the family. Models that allow different diversification rates in different parts of the tree are used to identify putative phylogenetic locations of diversification rate shifts. Despite the early stage of these analyses, initial ideas about the dynamics of legume diversification as well as the potential of such studies to shed light on the underlying factors that may have driven the evolutionary success of legumes are highlighted.

2. Methods and study groups

2.1. Estimating diversification rates

In this study a top-down approach using a sparsely sampled higher level legume-wide phylogeny is combined with a bottom-up approach that relies on a set of five more densely sampled species-level phylogenies of specific clades: *Calliandra*, Indigofereae, *Lupinus*, *Mimosa* and Robinieae. For each clade the following analytical approach was used: (i) previously published time-calibrated phylogenies (Drummond et al., 2012; Simon et al., 2009, 2011; Särkinen et al., 2012; Souza et al., submitted for publication) that were estimated under an uncorrelated lognormal relaxed molecular clock model in BEAST (Drummond and Rambaut, 2007) were gathered, and for Indigofereae, BEAST was run specifically for this study using the original dataset of Schrire et al. (2009); (ii) testing for constancy of birth and death rates across the phylogeny, by evaluating at each branch of the phylogeny whether the assumption of a rate shift improves the likelihood of observing the branching times in the phylogeny using MEDUSA (Modeling Stepwise Diversification Using Stepwise AIC) (Alfaro et al., 2009; Santini et al., 2009) to assess among-lineage variation in diversification rates and discover putative diversification rate shifts in one or more phylogenetic positions; (iii) estimation of diversification rates in a Bayesian framework using BayesRate (Silvestro et al., 2011) to evaluate the statistical support for differences in net diversification and underlying speciation and extinction rates among clades or tree partitions delimited by the rate shifts in the best fitting model found by MEDUSA.

Given that current phylogenies rarely contain all extant species in a clade, accounting for incomplete taxon sampling is an important issue and a potentially challenging hurdle in estimating species diversification rates (Cusimano et al., 2012; Marazzi and Sanderson, 2010), especially for sparsely sampled phylogenies or phylogenies for which taxon sampling is not representative of clade species richness (for instance when each genus of a family is sampled, but the proportion of species sampled per genus varies significantly). Objective assignment of unsampled taxa to a set of terminal clades (e.g. based on taxonomy), as implemented in MEDUSA, provides a solution to this problem (Alfaro et al., 2009; Drummond et al., 2012; Santini et al., 2009; Stadler and Bokma, 2012). The disadvantages of this approach are that in some cases there is no satisfactory objective (e.g. taxonomic) basis for assignment of unsampled taxa, and/or that the number of terminal clades needs to be reduced with the consequence of losing statistical power in the analysis. Furthermore, by restricting diversification rate analyses to backbone trees made up of a reduced number of terminal clades, it is possible that a nested subclade within one of those terminal clades is actually diversifying faster rather than the whole clade. Potential additional diversification rate shifts nested within these clades cannot be modelled. Another way of accounting for missing taxa, implemented in BayesRate, is by specifying the proportion of extant taxa sampled (Stadler, 2009; Yang and Rannala, 1997). The advantage is that all nodes in the tree are retained, but it assumes that missing taxa are located randomly across the tree, an assumption that is often

violated, e.g. when sampling is biased to include all higher-level taxa. MEDUSA analyses were run assigning unsampled species to clades for the legume-wide, *Lupinus*, and Indigofereae analyses (see Sections 2.2, 2.4, 2.5). For the diversification rate analyses of the other clades included in this study, taxon sampling is considered to be adequate for *Calliandra* (95/141) and Robinieae (66/75), and representative of clade diversity over these phylogenies (Lavin et al., 2003; Souza et al., submitted for publication). For *Mimosa* where taxon sampling was lower (261/c. 540), assignment of unsampled taxa was not possible due to the widespread non-monophyly of Barneby's (1991) morphologically defined sections and series in the molecular phylogeny of Simon et al. (2011). In addition, the possible differences between the results of either approach were investigated by executing both on the Indigofereae dataset, and ideally this should be done for other clades in the future when possible. BayesRate analyses accounted for missing species by specifying the approximate fraction of extant species sampled for each tree partition. Of course, undiscovered species and taxonomic biases could also influence diversification rate estimates, but the impacts of these on species-level phylogenies are difficult to assess.

Information criteria were used to determine how much better a model with more rate shifts explained the data. MEDUSA assesses this for diversification rate shifts in a stepwise fashion. First, for a diversification model without shifts it calculates a corrected AIC (hereafter AICc) score, which uses maximum likelihood and a penalty per model parameter and is adjusted for the number of data points available to fit models. It then searches for the best model (in terms of likelihood) that includes a single rate shift and calculates the AICc score of that one-shift-model. Subsequently, it searches for the best second shift in the one-shift-model, a third on the best two-shift-model, etc., and calculates AICc for each model. In the last step, the model with an optimal number of shifts is selected, by defining a threshold AICc required to accept additional rate shifts. By default, MEDUSA (turboMEDUSA implementation) proposes a threshold based on the number of tips in the tree, but this strategy is debated (e.g. Drummond et al., 2012), and the issue of appropriately selecting among increasingly complex diversification models is still not well understood. Instead of AICc, BayesRate uses Bayesian statistics to evaluate model fit with marginal likelihoods (rather than maximum likelihoods) obtained via integration over the entire parameter space, effectively incorporating uncertainty associated with parameter estimates and providing a natural penalty for excess parameter space. The Bayes Factor (BF), calculated as 2 times the ratio of log marginal likelihoods, is then used to compare model fit, where $|BF|$ (i.e. absolute BF) > 10 is claimed to represent "very strong support", $6 < |BF| < 10$ "strong support", $2 < |BF| < 6$ "positive support", whereas $|BF| < 2$ it is "not worth more than a bare mention" (Kass and Raftery, 1995). In our analyses, a set of primary diversification rate shifts using the default cutoff in MEDUSA as well as the additional rate shifts detected using a less stringent zero cutoff, i.e. the lowest AICc score to select the most strongly weighted rate shift model (e.g. Drummond et al., 2012) are presented. Identifying models that potentially over-fit the data at this early stage of analysis is justified by our emphasis on identifying as many potentially significant rate-shifts as possible, each of which then can be investigated with more thorough sampling in future studies. For the species-level phylogenies BayesRate analyses were executed on MCC trees for each rate shift suggested by MEDUSA, specifying the sampling fraction for each tree partition and including its stem lineage. BayesRate has the option to be run over a posterior sample of multiple trees from a Bayesian phylogenetic analysis to account for phylogenetic uncertainty, but this was beyond the scope of this exploratory study. Four BayesRate models were compared to characterize each rate shift: equal speciation and extinction rates (i.e. no shift), equal speciation but shifting extinction rates, shifting speciation but equal extinction rates, and shifting speciation and extinction rates. BF calculations were based on marginal likelihood estimation employing thermodynamic integration across 20 beta-distributed scaling classes, sampling

every 50th in a run of 100k generations after discarding 5% as burn-in. Adequacy of MCMC performance was confirmed using Tracer 1.5 (Drummond and Rambaut, 2007).

2.2. Legumes

The Fabales phylogeny of Simon et al. (2009), the most densely sampled time-calibrated phylogeny of the legume family published to date, was used as the basis for analysis of species diversification rates across the family as a whole. The original Simon et al. (2009) phylogeny included 839 terminals representing 400 genera, and was generated using BEAST with 23 fossil constraints. In order to account for very incomplete taxon sampling in that phylogeny as well as to facilitate placement of unsampled genera using the phylogenetic system of Lewis et al. (2005) and more recently published phylogenies (especially Cardoso et al., 2012), 508 tips were pruned to establish a phylogeny of 331 tips that represent a single genus, monophyletic groups of several genera, or monospecific lineages (Figs. 1 & S1; Table S1). Species-richness was assigned to each of these tips using data on species numbers for genera from Lewis et al. (2005, and updates) (Table S1). In addition to the MEDUSA analyses on the legume-wide phylogeny, a Lineages-Through-Time plot (LTT plot) for this phylogeny using 500 posterior trees from the BEAST analysis of Simon et al. (2009) was constructed (Fig. 2). Using Mesquite (Maddison and Maddison, 2011), 100 trees of 20,431 taxa (the number of spp. estimated for Fabales, see Table S1) were simulated using a time-homogeneous Yule model (i.e. under constant diversification) from 90.3 Ma to the present and an LTT of these trees plotted to allow comparison of the empirical LTT-plot against the simulated null model of constant diversification. Furthermore, net diversification rates over the whole phylogeny were estimated using a Yule model in the R package TreePar (Stadler, 2011a), with missing taxa assigned to terminal clades as in the MEDUSA analysis (Table S1) using the method of Stadler and Bokma (2012). The rate estimations were carried out with two and three partitions: one from the root until 60 Ma, one from 60 Ma to the present, or with an additional partition from 60 to 50 Ma with the third partition from 50 Ma to the present. The net diversification rates found for the 60 Ma to present partition ($r = 0.119$) and for the 60–50 Ma partition ($r = 0.100$) were used for adding the 19,592 missing taxa to the MCC tree with median node heights through simulation with the Corsim function (Cusimano et al., 2012) in the R package TreeSim (Stadler, 2011b). This function allows setting age cut-offs to restrict simulated speciation events to a certain time period, where we set the upper limit to 60 Ma, since the phylogeny is completely sampled in deep time (at least up to 60 Ma, probably up to 50 Ma). Adding these simulations to the LTT-plot allows evaluation of when undersampling leads to an underestimation of lineage accumulation. The simulation with the rate estimated from 60 Ma to the present suggests what the LTT plot might look like based on the underlying topology of the tree with a correction for missing taxa. The simulation with rate estimated between 60 and 50 Ma, including correction for missing taxa, is the projected lineage accumulation if the diversification would continue with that same rate after 50 Ma.

2.3. Calliandra

The genus *Calliandra* (Ingeae, Mimosoideae) forms a robustly-supported clade that has been recently monographed (Barneby, 1998), and comprises c. 141 species of shrubs and small trees, almost all of them restricted to the Neotropics. Species of *Calliandra* occupy a wide geographic distribution across the major tropical biomes (seasonally dry tropical forest, savanna, *campos rupestres* and less commonly in wet forests and subtropical grasslands). A recent time-calibrated phylogeny for *Calliandra* (Souza et al., submitted for publication) that includes 95 of the c. 141 species, generated using BEAST and employing one diagnostic *Calliandra* fossil pollen polyad (Caccavari and Barreda,

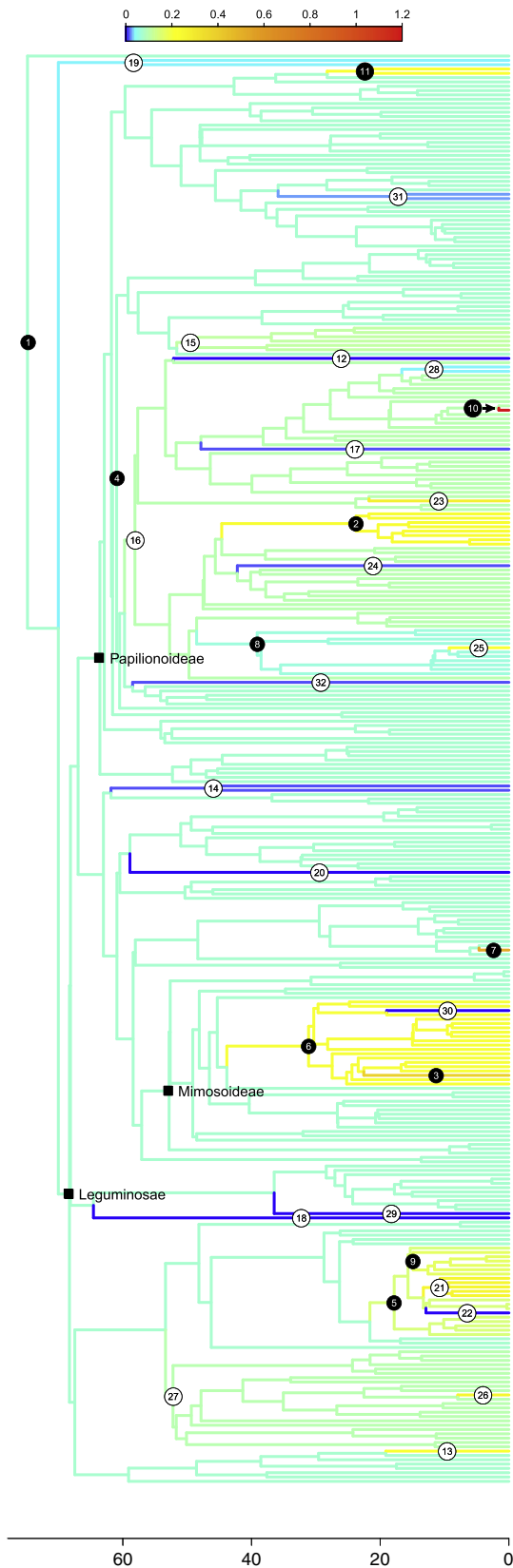


Fig. 1. Time-calibrated phylogeny of Fabales. The pruned Maximum-Clade Credibility tree of Simon et al. (2009), showing rate shifts found by MEDUSA with default settings (black node labels) and additional rate shifts found with zero cutoff settings (white node labels) (Table 1). Branches are coloured according to their net diversification rate (r) as indicated in the legend. Calculations were done on median node heights, the tree shown has the original node heights from Simon et al. (2009), since some of the branch lengths with median heights were negative. The x-axis shows divergence time in Ma.

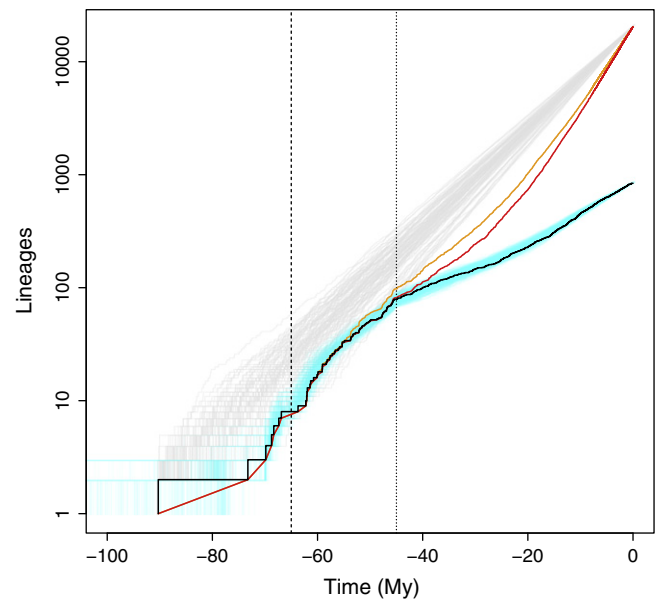


Fig. 2. Accumulation of Fabales lineages through time. Log-Lineages Through Time (LTT) plots representing 500 posterior trees from the BEAST analysis of Simon et al. (2009) shown in blue and 100 simulated Yule-model trees in grey. The black line represents the MCC tree with median node heights. The red and orange lines represent the MCC tree with missing taxa (or rather speciation events) added in by simulation from 60 Ma (see text in Section 3.1), with the Yule rate estimated from either 60 Ma to the present or between 60 and 50 Ma, respectively. The dashed line marks the K/T-boundary, the dotted line a potentially artefactual rate shift probably caused by the coincidence of 9 out of the 23 fossil constraints employed by Simon et al. (2009) at 45–46 Ma (see text in Section 3.1).

2000) and one secondary calibration from Lavin et al. (2005), was used in the diversification rate analyses here. African *Calliandra* species were here considered part of the ingroup (c.f. Souza et al., submitted for publication) while outgroup taxa (a representative set of genera from tribe Ingeae) were pruned. Taxon sampling (68% of species) includes all sections and twelve of the fourteen series and spans the geographic and ecological range of the genus and is considered representative of clade diversity (Souza et al., submitted for publication). No corrections for unsampled diversity were made in the MEDUSA analysis.

2.4. Indigoferae

The tribe Indigoferae (Papilionoideae) comprises 7 genera and 808 species (Schrire et al., 2009), with c. 750 of those assigned to *Indigofera*, the third largest genus of legumes. *Indigofera* species are herbs, shrubs and small trees and occupy a worldwide distribution in dry tropical to warm temperate forest, savannas and sclerophyllous shrubland. In order to standardize analyses across all clades investigated in this study, a new time-calibrated phylogeny was generated for Indigoferae in BEAST, using the ITS sequence alignment of Schrire et al. (2009) that includes 266 species (35% taxon sampling). To infer phylogenetic relationships for Indigoferae, four BEAST runs were employed, each using 10 million generations, and sampling every 1,000th. The analyses were calibrated based on the results of Schrire et al. (2009) by using a normal distribution at the root (the split of *Disyngestemon* and Indigoferae) with mean 45.9 Ma and standard deviation of 1.0 Ma, truncated to exclude values beyond 43.5–48.3 Ma. To facilitate convergence, the relations between *Disyngestemon*, the CRIM-clade (sensu Schrire et al., 2009) and *Indigofera* were constrained to conform to Schrire et al. (2009). Runs were combined using LogCombiner after confirming adequate MCMC performance using Tracer 1.5 and discarding 25% as burn-in. A Maximum Clade Credibility tree with median node heights, calculated using TreeAnnotator, was used for diversification

rate analyses. Because taxon sampling in this dataset for this very species-rich clade is relatively sparse (35%), two separate MEDUSA analyses were carried out, one with just the sampled taxa with no correction for unsampled species, and a second one in which unsampled species were assigned to a set of 51 terminal clades, which largely correspond to recognized sections (sensu Schrire et al., 2009; Table S2). The analyses also included the monotypic *Disyctemon*, the probable sister group of Indigofereae (Schrire et al., 2009). Unsampled species were assigned per tree partition in BayesRate analyses following Schrire et al. (2009).

2.5. *Lupinus*

The genus *Lupinus* (Genisteeae, Papilionoideae; c. 275 species) has emerged as a model system for testing species diversification rate shifts and understanding rapid species radiations in plants, and has already been the focus of four separate attempts to estimate diversification rates and locate rate shifts using different approaches based on a phylogeny that includes up to 122 species (46% taxon sampling) (Drummond et al., 2012; Hughes and Eastwood, 2006; Moore and Donoghue, 2009; Silvestro et al., 2011). Here these studies are extended by implementing a new BayesRate analysis of four tree partitions delimited on the basis of three diversification rate shifts found by a previous MEDUSA analysis (Drummond et al., 2012). Although a similar BayesRate analysis was presented by Silvestro et al. (2011), the phylogeny was more sparsely sampled and tree partitions were defined by geographic distributions of clades rather than by a best-fitting MEDUSA model. Unsampled species were assigned per tree partition in BayesRate analyses following Drummond et al. (2012).

2.6. *Mimosa*

The genus *Mimosa* (Mimoseae, Mimosoideae) is the fifth largest genus of legumes with c. 540 species and occupies an amphiatlantic distribution with more than 500 species in the Neotropics, 31 species native to Madagascar and a handful of species in Africa and the Near East (Simon et al., 2011). While the genus occurs predominantly in seasonally dry tropical forests (the Succulent Biome of Schrire et al., 2005), it has a set of important recent clades in South American savannas – *cerrado* and *campos rupestres* (Simon et al., 2009), as well as occurrences in tropical rain forests, subtropical dry forests and submontane oak–pine woodland. Taxonomic diversity is documented in detail in a comprehensive monograph of the New World species (Barneby, 1991) and a treatment of the Madagascan species (Villiers, 2002). The time-calibrated trnD–trnT phylogeny of Simon et al. (2011) included 284 terminals. In the tree used here multiple accessions of species were pruned to leave 261 species (interpreted as 259 species by Simon et al., 2011), and relied on two secondary calibrations derived from the Simon et al. (2009) legume chronogram. No correction was made for unsampled taxa in the MEDUSA analyses because of the difficulties of assigning unsampled species to clades, which in very few cases correspond to the traditional infrageneric classification of Barneby (1991).

2.7. *Robinieae*

The tribe Robinieae (Papilionoideae) is a robustly supported clade comprising 11 genera and 75 species of small trees and shrubs restricted to the New World and spanning the southern U.S.A., Mexico, Central America, the Caribbean and South America (Lavin et al., 2003). This clade has a strong predilection for seasonally dry tropical forests, with only a small number of species extending outside that biome into subtropical arid and warm temperate North America (Lavin, 2006; Lavin and Sousa, 1995). In this study, the time-calibrated phylogeny by Särkinen et al. (2012) was used, which was generated with BEAST, using the comprehensively sampled ITS data set (all 11 genera and 66 of 75 species) of Lavin et al. (2003) with crown node divergence time

estimates for Robinieae and the *Poitea–Gliricidia* clade from Simon et al. (2009) as secondary calibrations. No assignment of unsampled species was done for Robinieae.

3. Results

3.1. Legumes

While the MEDUSA analyses using the default AICc threshold detected ten diversification rate shifts, those with the lowered zero threshold found an additional 21 rate shifts, making a total of 31 (Fig. 1; Table 1). All but one of the rate shifts were nested within Leguminosae, and no rate shift was detected subtending the legumes as a whole. While some rate shifts are hard to interpret, as they are located at a deeply nested position affecting heterogeneous tree partitions (e.g. rate shifts 4, 16 & 27), other rate shifts allow for more straightforward interpretations.

The first rate shift discovered by MEDUSA (Fig. 1, Table 1, rate shift 2) indicates a 2.46-fold increase in species diversification rate in a subclade of the Temperate Herbaceous Clade comprising the Astragalean clade (Sanderson and Liston, 1995) plus the genera *Erophaca*, *Chesneya*, *Spongiocarpella*, *Gueldenstaedtia* and *Tibetia*. This includes the largest angiosperm genus, *Astragalus* with c. 2500 species. These results are in line with the findings of Sanderson and Wojciechowski (1996), in so far as neither analysis finds evidence for a rate shift associated with *Astragalus* itself, but our results differ from that study in tree topology and in assigning the rate shift to a somewhat more inclusive clade. A second notable finding is the two-step acceleration in diversification rates within mimosoid legumes, with a three-fold increase (rate shift 6) to more rapid diversification associated with a large clade that includes Mimoseae pro parte, four of the five *Acacia* segregates and the whole of tribe Ingeae, and a second nested shift (rate shift 3) to even higher rates of diversification associated with a large clade of c. 1853 species comprising the majority of genera in the tribe Ingeae plus *Acacia* s.s. These nested rate shifts within mimosoids suggest rapid diversification of the majority (c. 2700 of 3200 species) of this subfamily. Other significant rate increases include: the Neotropical genus *Tachigali* (rate shift 7), with c. 70 species of trees, centred in the tropical rain forests of the Amazon, and well-known for close associations with ants; the *Pediomelum* clade (rate shift 10), made up of three genera – *Pediomelum*, *Bituminaria* and *Cullen* – with 57 species of mainly herbs widely and disjunctly distributed in N. America, Europe, Australia, Asia and Africa, which shows the highest rate of net diversification of any clade in the legume-wide analysis, a function of the very recent estimated crown age for this clade in this phylogeny, but this diversification rate shift likely to be more accurately placed within the genus *Pediomelum*, as found in the species-level analysis of Egan and Crandall (2008); the clade comprising the genera *Dalea* and *Marina* (rate shift 11), a group of c. 203 species of herbs and shrubs of dry and arid warm temperate and subtropical montane habitats centred in Mexico, with a secondary centre of diversification in the mid-elevation Andes (Barneby, 1977). A diversification rate increase (rate shift 23) subtends the genus *Indigofera* in line with results from the densely sampled species-level analyses of tribe Indigofereae (see Sections 3.3 & 4). While the slight majority of rate shifts modelled are rate increases, a number of notable slowdowns are also apparent (Table 1). First, rather than a rate increase subtending the legumes as a whole, a significant rate slowdown (rate shift 19) is inferred in the sister group of legumes, the Surianaceae + Quillajaceae clade with 11 species. Secondly, tribe Robinieae shows a marked rate slowdown (rate shift 8; see Sections 3.6 & 4). Finally, eleven rate slowdowns (rates shifts 12, *Xeroderis*; 14, *Ceratonia* + *Acrocarpus*; 17, *Platycamus*; 18, *Duparquetia*; 20, *Pterogyne*; 22, *Librevillea*; 24, *Parochetus*; 28, *Lablab* + *Dipogon*; 29, *Baudouinia fluggeiformis* (*Baudouinia* with 6 species is non-monophyletic in this analysis); 30, *Piptadenia viridiflora* (although currently assigned to *Piptadenia*, this species has been shown to represent a distinct monospecific lineage (Jobson and Luckow, 2007));

Table 1
Species diversification rate shifts in the legume-wide analysis.

Rate shift #	AICc gain*	r	ε	Magnitude of rate increase/ slowdown (x-fold change)	Median crown age	Clade description
1	NA	6.5535e-02	8.4336e-01	Background	90.3	Fabales (19444 spp.)
2*	88.206	2.2511e-01	8.9326e-01	+2.46	37.3	Astragalean clade (sensu Sanderson & Wojciechowski, 1996)
3*	81.506	3.7594e-01	5.6816e-01	+1.89	17.0	Ingeae p.p. + <i>Acacia</i> s.s. clade (1853 spp.)
4*	60.412	6.6405e-02	8.9276e-01	+1.01	61.5	Genistoid s.l. + Baphioid + NPAAA clades (sensu Cardoso et al., 2012)
5*	56.274	1.3802e-01	9.7295e-01	+2.11	17.8	Amherstieae clade p.p. (sensu Bruneau et al., 2008)
6*	34.631	1.9856e-01	4.7682e-07	+3.03	34.1	Mimosoideae p.p.
7*	18.592	4.9982e-01	6.2188e-01	+7.63	6.6 (stem)	<i>Tachigali</i> (70 spp.)
8*	16.940	5.3344e-02	6.1626e-01	-0.58	40.0	Robinieae
9*	13.305	1.6126e-01	2.2721e-06	+1.17	14.2	<i>Humboldtia</i> , <i>Hymenostegia</i> , <i>Leonardoxa</i> , <i>Talbotiella</i> , <i>Loeserena</i> & <i>Amherstia</i>
10*	9.718	1.1562e+00	7.5477e-01	+12.63	2.3	<i>Pediomelum</i> clade
11*	7.519	2.0948e-01	4.4694e-01	+3.20	19.3	<i>Dalea</i> + <i>Marina</i> (c. 203 spp.)
12**	6.574	4.7916e-11	5.1603e-01	-5.24e-10	47.6 (stem)	<i>Xeroderris</i> (monospecific)
13	6.235	2.1528e-01	5.4490e-01	+3.28	20.0 (stem)	<i>Bauhinia</i> s.s.
14	5.510	1.6744e-02	3.3995e-07	-0.26	45.8	<i>Ceratonina</i> + <i>Acrocarpus</i> (3 spp.)
15	4.986	1.0118e-01	9.6689e-01	+1.52	39.6	Genisteeae + Crotalariaeae + Podalyrieae
16	15.176	9.1508e-02	9.3108e-01	+1.38	59.1	Indigoferoids + Millettioids + Robinioids + IRLC (sensu Cardoso et al., 2012)
17	4.529	8.6859e-03	4.7316e-01	-0.09	48.7 (stem)	<i>Platycamus</i> (2 spp.)
18**	4.404	4.7916e-11	5.1603e-01	-7.31e-10	62.1 (stem)	<i>Duparquetia</i> (monospecific)
19	3.693	3.2502e-02	1.8230e-06	-0.50	56.8	Surianaceae + Quillajaceae (11 spp.)
20**	3.584	4.7916e-11	5.1603e-01	-7.31e-10	55.3 (stem)	<i>Pterogyne</i> (monospecific)
21	3.293	2.4946e-01	1.6826e-06	+1.81	15.4	<i>Berlinia</i> clade p.p. (sensu Bruneau et al., 2008)
22**	2.635	4.7916e-11	5.1603e-01	-3.47e-10	12.3 (stem)	<i>Librevillea</i> (monospecific)
23	2.387	2.6791e-01	5.7645e-01	+2.93	21.3 (stem)	<i>Indigofera</i> incl. <i>Vaughania</i> (711 spp.)
24	1.786	1.1147e-02	4.8222e-01	-0.12	37.4 (stem)	<i>Parochetus</i> (2 spp.)
25	1.679	2.0858e-01	5.5427e-01	+3.91	13.3 (stem)	<i>Coursetia</i>
26	1.335	2.4103e-01	5.7605e-01	+2.48	10.2	<i>Copaifera</i> , <i>Sindora</i> , <i>Detarium</i> , <i>Sindoropsis</i> + <i>Tessmannia</i>
27	9.632	9.7293e-02	3.2255e-06	+1.48	42.0	resin-producing Detarieae & <i>Saraca</i> clade (sensu Bruneau et al., 2008), <i>Schotia</i> , <i>Barnebydendron</i> , + <i>Goniorhachis</i>
28	1.069	3.2830e-02	1.3594e-01	-0.36	7.5	<i>Lablab</i> + <i>Dipogon</i> (both monospecific)
29**	0.719	4.7916e-11	5.1603e-01	-7.31e-10	38.9 (stem)	<i>Baudouinia fluggeiformis</i>
30**	0.288	4.7916e-11	5.1603e-01	-2.41e-10	21.0 (stem)	<i>Piptadenia viridiflora</i>
31	0.173	2.7278e-02	2.6365e-07	-0.42	26.0	<i>Cascaronia</i> & <i>Goeffroea</i> (3 spp.)
32	0.019	1.2579e-02	4.5855e-01	-0.03	58.3 (stem)	<i>Hypocalyptus</i> (3 spp.)
Unpartitioned		6.1998e-02	9.5917e-01	Total: 17 increases/14 slowdowns		

Appropriate AICc threshold estimated by MEDUSA for tree of 331 tips is: 7.427755.

Rate shifts 2–11 marked with an * were detected using the default AICc threshold in MEDUSA, the remainder using the zero cutoff; r = net diversification rate; ε = relative extinction, rates for lineages marked ** representing a single extant species, are uninformative. Magnitudes of rate shifts: + = increase, - = slowdown.

31, *Cascaronia* + *Geoffroea*; 32, *Hypocalyptus*) are all associated with monospecific or two/three-species genera.

The accumulation of lineages through time across the Fabales tree as a whole shows a marked deviation from a constant rate null model when compared to simulated trees (Fig. 2). An initial rapid diversification followed by a slowdown around 65–60 Ma, the latter coinciding with the K/T-boundary, in turn followed by very rapid recovery, is apparent from the LTT-plot. A possible slowdown at 45 Ma seems apparent as well. The very incomplete taxon sampling means that the LTT plot is only meaningful up to when the impacts of missing taxa become pronounced, which can be evaluated by comparison with the LTT plots for simulated trees adding missing taxa using the method of Stadler and Bokma (2012) (Fig. 2). When adding missing taxa (or in fact, speciation events) with a Yule rate estimated over 60 Ma to the present (red line in Fig. 2), lineage accumulation follows the plot of the inferred trees very closely until c. 45 Ma, after which missing speciation events are increasingly added in. Projected lineage accumulation by adding speciation events with the Yule rate estimated between 60 and 50 Ma starts deviating slightly earlier but still follows the lineage accumulation from the inferred trees quite closely until 45–40 Ma. The LTT-plot therefore is not meaningful after c. 45 Ma, and should not be interpreted beyond that point. However, comparing the number of lineages at 45 Ma with the constant rate Yule trees (grey shading) suggests that an increase in net diversification

rate towards the present is necessary to reach the present-day diversity, as is also apparent from the plots with simulated speciation events. Without a more densely sampled tree, inferences of speciation and extinction rates towards the present cannot be made for legumes as a whole. The apparent slowdown in accumulation of lineages spanning the K/T-boundary and subsequent rate increase shows the type of anti-sigmoidal pattern indicative of a possible mass-extinction event (Crisp and Cook, 2009). However, with only very few lineages present in the phylogeny at this time point, stochastic effects might play a role. Furthermore, the apparent rate slowdown observed at 45 Ma (Fig. 2) is likely to be an artefact, in that nine out of 23 fossil calibrations used to constrain the legume-wide BEAST analysis have ages of 45–46 Ma (Simon et al., 2009). This means that nine speciation events were either pushed or pulled to this time point implying an increase of nine lineages on top of contemporaneous lineage accumulation in the rest of the tree.

3.2. *Calliandra*

A single rate shift subtending a robustly supported clade corresponding to species occurring in *campos rupestres* (hereafter the 'campos rupestres clade') was found using MEDUSA with either default or zero AICc cutoffs (Fig. 3A). This clade of 36 species (here represented by a sample of 27 species) is endemic to the Espinhaço Range in

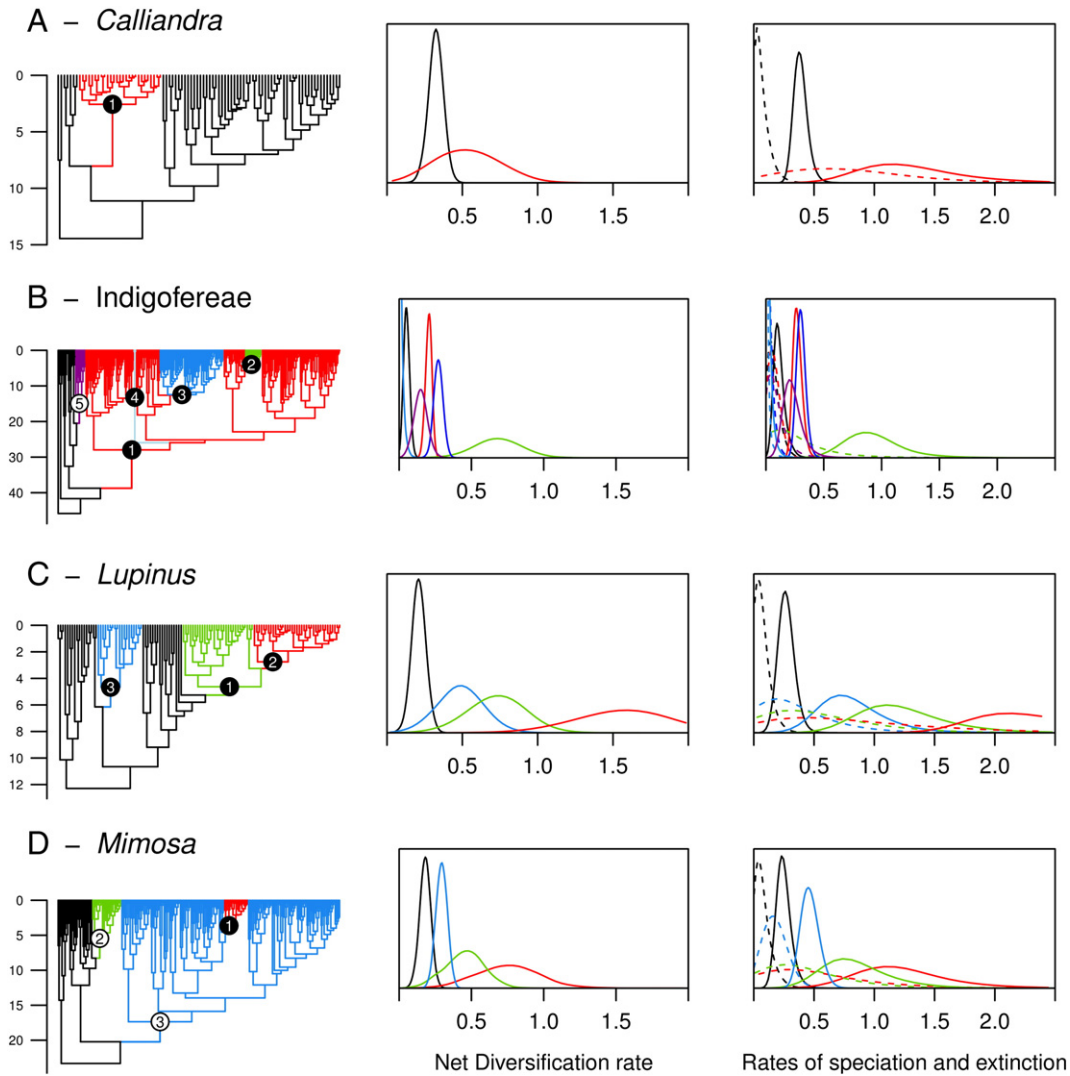


Fig. 3. Species-level time-calibrated phylogenies showing species diversification rate shifts detected by MEDUSA, and net diversification, speciation and extinction rates estimated using BayesRate for four species-rich legume clades. Diversification rate shifts are numbered in the order they were discovered by MEDUSA; black filled rate shifts are those detected under the default AICc cutoff (see Section 2.1) and unfilled rate shift symbols are those detected using the zero AICc cutoff (Table 2). Graphs are Kernel density plots of the posterior distributions of net diversification rates (middle column), and underlying speciation (solid lines) and extinction (dotted lines) rates (right-hand column) for tree clades/partitions defined by the MEDUSA rate shifts. A. the genus *Calliandra* (Mimosoideae); B. tribe Indigofereae (Papilionoideae) with rate shift 1 subtending the genus *Indigofera*; C. the genus *Lupinus* (Papilionoideae); D. the genus *Mimosa* (Mimosoideae). Time scales on phylogenies in Ma. Phylogenetic trees with terminal names and GenBank accession numbers are in the original publications: *Calliandra* (Souza et al., submitted for publication), Indigofereae (Schrire et al., 2009), *Lupinus* (Drummond et al., 2012), and *Mimosa* (Simon et al., 2011).

eastern Brazil, implying rapid in situ species diversification (radiation) of this clade in geographical isolation. BayesRate analyses to compare rates for the tree partitions delimited by this rate shift show robust support for a four-parameter model in which the *campos rupestres* clade shows both significantly higher speciation and extinction rates compared to the rest of *Calliandra*, implying a signature of high species turnover. The clade is strongly associated with higher elevation *campos rupestres* ecology, while the remainder of the genus occupies a range of other Neotropical biomes, including seasonally dry tropical forest, *cerrado* and tropical wet forest.

3.3. Indigofereae

Different numbers and phylogenetic locations of diversification rate shifts were found in the MEDUSA analysis with no correction for incomplete sampling and with unsampled species assigned to a pruned tree of 55 terminals. It is not clear whether these differences solely result from correctly assigning species richness or also from uneven pruning of terminals in different parts of the tree. This implies

that the 266 sampled species in the original phylogeny might not be representative of clade species richness, and that assignment of unsampled taxa could be critical in this case. In the latter analysis four diversification rate shifts were found using the default cutoff, and an additional rate shift using the zero cutoff (Fig. 3B). All rate shifts detected by MEDUSA were supported in the BayesRate analyses (Table 2). These rate shifts (Fig. 3B) are as follows: (1) the genus *Indigofera* as a whole; (2) a subclade comprising the 54 species that make up Sections Brachypodae, Productae and Alopecuroides that is nested within the Cape clade of Schrire et al. (2009) and which is characterized by and confined to strict Fynbos vegetation on Table Mountain sandstone substrates, while the remainder of the Cape clade species are found in karroid and renosterveld vegetation; (3) a large clade of c. 193 species nested within the Pantropical clade sensu Schrire et al. (2009) which corresponds to a reversal from predominantly herbaceous/suffrutescent species to largely woody taxa, and a shift to global, often more temperate (higher altitude) distributions with significant New World, Australian, Sino-Himalayan and Afrotropical African radiations; (4) a rate slowdown in the *Indigofera nudicaulis* lineage

Table 2
Support for species diversification rate shifts in species-level analyses of legume clades assessed using BayesRate. Preferred models, based on Bayes Factor support are indicated in bold. As no diversification rate shifts were found within the Robinieae phylogeny, no BayesRate analysis was carried out on that clade.

Clade	MEDUSA rate shift	Marginal likelihood				BayesFactor support for shift			
		Equal birth & death	Equal birth, shifting death	Equal death, shifting birth	Shifting birth & death	Equal birth & death	Equal birth, shifting death	Equal death, shifting birth	Shifting birth & death
<i>Calliandra</i>	1	−186.30	−185.96	−183.41	−182.07	0.00	0.67	5.78	8.45
Indigofereae	1 ^a	−616.29	−604.78	−605.15	−605.87	0.00	23.03	22.28	20.85
Indigofereae	1 ^c	−647.67	−638.96	−641.53	−642.09	0.00	17.42	12.28	11.16
Indigofereae	2	−576.99	−574.67	−568.67	−568.65	0.00	4.65	16.65	16.69
Indigofereae	3	−711.56	−710.65	−714.27	−715.00	0.00	1.82	−5.41	−6.87
Indigofereae	4	−547.37	−543.97	−542.13	−542.69	0.00	6.79	10.47	9.36
Indigofereae	5*	−106.45	−105.01	−107.40	−107.92	0.00	2.87	−1.90	−2.94
<i>Lupinus</i>	1	−142.18	−138.96	−130.29	−130.56	0.00	6.46	23.78	23.24
<i>Lupinus</i>	2	−80.22	−77.34	−76.42	−77.06	0.00	5.76	7.61	6.31
<i>Lupinus</i>	3	−124.11	−123.84	−121.60	−122.03	0.00	0.54	5.02	4.16
<i>Mimosa</i>	1 ^b	−598.77	−597.20	−593.85	−594.36	0.00	3.13	9.84	8.81
<i>Mimosa</i>	1 ^d	−447.39	−446.26	−443.02	−443.63	0.00	2.26	8.74	7.51
<i>Mimosa</i>	2*	−147.03	−145.04	−142.08	−142.23	0.00	3.98	9.90	9.60
<i>Mimosa</i>	3*	−513.51	−511.52	−511.94	−512.26	0.00	3.98	3.15	2.50

Rate shifts indicated with * were found only using the zero AICc cutoff in MEDUSA.

^a Rate shift tested against background rate.

^b Rate shift tested against the rest of the tree as one partition.

^c Rate shift tested against background + descendants of rate shift five, which was only recovered at zero cutoff.

^d Rate shift tested against the subtending tree partition (blue in Fig. 3D), since rate shift three was only found at zero cutoff.

(one species), although the phylogenetic position of *I. nudicaulis* in our analysis is not supported by the combined morphological and molecular data analysis by Schrire et al. (2009), hence the slowdown detected for this lineage may be an artefact of phylogenetic uncertainty, a factor not accounted for in the present analyses; (5) a rate increase subtending the genus *Microcharis* (36 species) characterized by a delicate annual habit and a unique heterochronous developmental shift where the plants are almost flowering seedlings with the seedling leaves often remaining persistent at the base of the plant. All four diversification rate increases appear to be associated primarily with increased speciation rates with support for a three-parameter equal extinction rates model (Fig. 3B, Table 2).

The Indigofereae represent one of just two cases in this study where the same nodes are represented in both the legume-wide and species-level analyses. A rate increase is detected at the same phylogenetic location in both the higher level analysis (rate shift 23) and species level analysis (Fig. 3B, rate shift 1), providing some reassurance that backbone and species-level phylogenetic approaches can provide consistent results (see also Smith et al., 2011).

3.4. *Lupinus*

The BayesRate analyses of *Lupinus* focused on the clades/tree partitions subtended by diversification rate shifts (Fig. 3C) identified by Drummond et al. (2012). The Bayes Factors found positive to very strong support for the three rate shifts considered (Fig. 3C; Table 2). These results differ slightly from a previous BayesRate analysis of *Lupinus* in finding support for three rate shifts and four tree partitions (Fig. 3C) with significantly different diversification rates, whereas three rather than four geographically defined tree partitions were found to be preferred by Bayes Factor comparison by Silvestro et al. (2011). In each case, the diversification scenario involving equal extinction rates but differential speciation rates between tree partitions received most support, in line with Drummond et al. (2012), but this was only marginally better supported than a model in which both speciation and extinction rates were different between tree partitions.

3.5. *Mimosa*

Between one and three rate shifts were found within the genus *Mimosa* (Fig. 3D). While only one of these was found using the default cutoff in MEDUSA, all three find positive to strong BF support in the

BayesRate analysis (Fig. 3D; Table 2). While the precise placement of rate shift 1 is somewhat uncertain due to poor support in that part of the *Mimosa* phylogeny, it corresponds closely to Clade O of Simon et al. (2011), a clade made up of sections Pachycarpae and Setosae (Barneby, 1991), that includes c. 55 species (Simon et al., 2010). This clade is characterized by cerrado/campos rupestres ecology and a diverse set of fire adaptations (functionally herbaceous geoxyllic subshrubs, pachycaul treelets and thickened bark/dense persistent stipules) that have apparently evolved multiple times in parallel within *Mimosa* (Simon et al., 2009). The second diversification rate increase within *Mimosa* corresponds to a subclade within Simon et al.'s (2011) Clade C, made up of North American species (Mexico, U.S.A., Central America and the Caribbean) distributed mainly in seasonally dry tropical forests with extensions into sub-montane pine-oak forests. The final diversification rate shift is a rate increase subtending a large heterogeneous clade, although we doubt the precise phylogenetic location of this shift. BayesRate analyses found positive to strong support for each rate shift (Table 2). The first and second rate shifts were inferred to be primarily shifts in speciation rate (Fig. 3D), but the small difference between Bayes Factors for these and models in which both speciation and extinction rates differed suggests that these two models cannot be robustly distinguished in these cases (Table 2). The third rate shift was associated with differential extinction rates, with lower extinction across this large clade comprising the majority of *Mimosa* species (Table 2).

3.6. Robinieae

With near-complete taxon sampling for this clade in both the legume-wide dataset (all 11 genera sampled, facilitating direct assignment of species diversity to genera) and the species-level phylogeny (all 11 genera and 66 of 75 species), these phylogenies are potentially ideal for estimating diversification rates. In the higher-level legume-wide analysis, a significant diversification rate slowdown was detected on the branch subtending the Robinieae (Fig. 1, rate shift 8), with a subsequent 3.9-fold increase (rate shift 25) associated with the genus *Coursetia*, albeit this latter shift is only detected under the zero AICc threshold and only supported by a modest AICc gain (Table 1). However, in the nearly fully sampled species-level phylogeny, this latter rate shift was not found. Indeed, in that analysis no diversification rate shifts were detected. This discrepancy could be due to differing topologies and/or branch length estimates between these two phylogenies which

were built using different DNA sequence loci. Given the much denser taxon sampling in the species-level analysis, we favour the result of constant diversification within Robinieae, pending further investigation of these two tree partitions using BayesRate.

4. Discussion

The analyses presented here represent the first family-wide exploration of variation in species diversification rates within the legumes. Despite many limitations, a number of preliminary insights into both the tempo of diversification of legume clades, and methodological issues associated with these analyses, can be drawn, and some interesting ideas for further testing can be highlighted. Overall, there is strong evidence for variation in the rate of diversification (Fig. 2), thereby rejecting a null model of constant diversification. A significantly better model fit for a partitioned model with rate shifts compared to equal rates over the whole tree model (Figs. 1 & 3; Tables 1 & 2) is also apparent in almost all cases, supporting the hypothesis that the Leguminosae have experienced significant among-lineage diversification rate variation. Several examples of differential diversification rates are identified that are worthy of further investigation to elucidate the macro-evolutionary factors that may have contributed to these potential shifts in rates of diversification across the family.

First, the antisigmoidal portion of the LTT plot (Fig. 2) spanning the 70–60 Ma period matches known trajectories for lineages proposed to show a signature of extinction (Crisp and Cook, 2009), and provides intriguing evidence suggesting that after an initial rapid diversification, Fabales could have suffered increased extinction associated with the K/T-boundary. A rapid recovery in lineage accumulation rate is seen shortly after the K/T-boundary, suggesting that legumes have successfully taken the opportunity to diversify during this major turn-over event. However, given the possibility of stochastic effects on the LTT caused by having just a few lineages present in the phylogeny at this time, the lack of any definitive legume fossils prior to the Paleocene c. 58 My (Herendeen et al., 1992; Wing et al., 2009), and uncertainties surrounding crown age estimates for the family (Lavin et al., 2005; Simon et al., 2009), this hypothesis of K/T-boundary extinction in legumes requires further investigation. Crucial to this will be a time-calibrated phylogeny with wider taxon sampling beyond Fabales across Rosids to improve the accuracy of the crown node divergence time estimate for the legumes. The LTT plot suggests that an increase in diversification rate after 45 Ma, or at least a pull-of-the-present effect observed in constant birth-death models (Stadler, 2013), is necessary to reach present-day diversity in the family, thereby refuting the well-established idea of density dependent diversification rate slowdown following the filling of ecological space (e.g. Morlon et al., 2011) in legumes as a whole, pointing instead in the direction of turnover of lineages. In addition to deviations from constant diversification revealed in the LTT plot, there is evidence for significant among-lineage variation in diversification rates across legumes for both higher level clades and within all the species-level clades investigated in detail here, except for the Robinieae, which showed a constant rate of diversification. Second, a diversification rate acceleration subtending the legumes as a whole is not found, but instead a set of rate shifts nested within the legume phylogeny apparently accounting for the high species diversity of the family. These results suggest that the high diversity of legumes is the product of many individual radiations in different clades and at different times, consistent with findings from other large clades such as angiosperms as a whole (Smith et al., 2011), palms (Baker and Couvreur, 2013), birds (Jetz et al., 2012), flies (Wiegmann et al., 2011) and ants (Moreau and Bell, 2013).

Although there is no rate shift associated with the family as a whole, a rate slowdown associated with the sister group of legumes, the Surianaceae + Quillajaceae clade with just 11 species – accounts for the massive 1800-fold discrepancy in species richness between these sister groups. Relationships among the four families of Fabales

remain uncertain due to lack of resolution in current phylogenies (Bello et al., 2009, 2012; Wang et al., 2009), and this can clearly influence where diversification rate shifts are placed in relation to the legume clade (Smith et al., 2011). However, the (Polygalaceae (Leguminosae (Quillajaceae + Surianaceae))) branching order portrayed here (Figs. 1 & S1) is the one favoured in recent combined analyses of molecular and morphological data (Bello et al., 2012). The lack of diversification rate shifts subtending many other notably diverse angiosperm families and clades, including angiosperms as a whole (Sanderson and Donoghue, 1994), with multiple nested rate shifts instead accounting for the high diversity (Magallón and Sanderson, 2001), appears to be a general pattern for species-rich plant families and higher order clades (Smith et al., 2011). Wider sampling outside Fabales across Rosids would be desirable to confirm these results.

Ecology rather than geohistory may be of primary importance in structuring legume phylogenies (Hughes et al., 2013; Lavin et al., 2004; Pennington et al., 2009). Several of the rate shifts found here appear to be closely associated with particular distinctive ecologies, and potentially with ecological (biome) shifts. For example, the primary rate shift (Fig. 3C, rate shift 1) within *Lupinus* occurred in western North America and is strongly associated with a shift from Mediterranean climate habitats to more mesic montane environments, and another rate shift (Fig. 3C, rate shift 3) is associated with a shift from lowland to upland grassland habitats in eastern South America (Drummond et al., 2012). Similarly for *Mimosa* (Fig. 3D, rate shift 1) and *Calliandra*, rate shifts are clearly associated with disturbed, fire-prone *cerrado/campos rupestres* habitats. Similarly, a shift to higher rates of species diversification within the Cape clade of *Indigofera* (Fig. 3B, rate shift 2) is associated with a subclade strictly confined to Fynbos vegetation. These results suggest that extrinsic ecological opportunities, often coupled with disturbance-prone vegetation and in some cases with recently emerged habitats, are likely to have played a key role in promoting rapid species diversification in legumes.

In contrast, the phylogenetically niche-conserved seasonally dry tropical forest (Succulent Biome sensu Schrire et al., 2005) tribe Robinieae (Lavin, 2006), appears to be diversifying at a constant and relatively slow rate, with a rate slowdown subtending the clade as a whole (see Section 3.6). The small number of rate shifts within the species-rich and predominantly seasonally dry tropical forest genus *Mimosa* (just three rate shifts in a clade of 540 species) and across large parts of the Indigoferae phylogeny, where the Succulent Biome ecology prevails (Schrire et al., 2009), provide further evidence for constant and relatively slow rates of diversification in seasonally dry tropical forest clades. Possible exceptions to this pattern are some lineages in Mexican and Central American SDTFs, such as the North American *Mimosa* subclade associated with higher rates of diversification in this study, two species-rich subclades within the genus *Bursera* (De-Nova et al., 2012), and possibly the genus *Nissolia* (Pennington et al., 2004), where diversification rate shifts in SDTF clades have been found or might be inferred. These findings of generally constant SDTF species diversification rates are in line with the emerging view of seasonally dry tropical forest as a dispersal-limited, ecologically resilient and relatively stable biome, little subject to natural disturbance, with long persistence of endemic populations and small clades of endemic species restricted to geographically isolated disjunct forest nuclei through at least much of the Neogene (Lavin et al., 2004; Pennington et al., 2006, 2009; Schrire et al., 2005).

These findings suggest that the tempos of legume species diversification may have been very different in different biomes, in line with the emerging ideas of divergent patterns of geographic phylogenetic structure in different tropical biomes, and ecology as a primary determinant of phylogenetic structure (Lavin et al., 2004; Hughes et al., 2013). A larger set of well-sampled time-calibrated phylogenies, spanning the full spectrum of biomes and especially temperate and tropical wet forests will be needed to test these ideas of species diversification rate heterogeneity across biomes. Of course it is important to bear in mind that many biome switches inferred from current

phylogenies (e.g. within Indigoferaeae, Schrire et al., 2009) do not appear to be associated with diversification rate shifts. The question as to why some biome shifts apparently prompt diversification rate shifts and others do not, is likely to depend on a clear understanding of the evolution of morphological and/or physiological traits that facilitated these ecological shifts. Trait shifts should be expected to have evolved at or immediately prior to these rate shifts, as found for *Lupinus* (Drummond et al., 2012), in line with the emerging consensus that it is the combination of both extrinsic opportunity and intrinsic innovation that best predicts radiation (Wagner et al., 2012).

In apparent contrast to the potentially widespread correlations between biome switches and diversification rate shifts in legumes, very few clearcut examples of key evolutionary innovations that coincide with diversification rate shifts have so far been documented within legumes. Examples include the evolution of perennial habit, which coincided with a shift from lowland to montane habitats and an acceleration in species diversification rate in western New World *Lupinus* (Drummond et al., 2012), and the evolution of extrafloral nectaries in a species-rich clade within the genus *Senna*, which was shown to be more species-rich than expected given the background diversification rate for *Senna* as a whole (Marazzi and Sanderson, 2010). However, it is notable that within *Mimosa*, presence of an extrafloral nectary is not associated with any diversification rate increase, indeed extrafloral nectaries occur only in a small clade that is sister to the clade of *Mimosa* where all the diversification rate shifts occur, suggesting that interpretation of extrafloral nectaries as a key evolutionary innovation may not apply consistently across legumes.

Several of the legume clades that show elevated rates of species diversification – viz. the Astragalean clade, the western New World *Lupinus* clade, mimosoids pro parte (Fig. 1, rate shift 6), and the series Setosae/Pachycarpae cerrado clade in *Mimosa* – are characterized by high levels of morphological parallelism, especially in leaf morphology and growth form, with similar morphological adaptations to extreme environmental conditions (drought, high altitude cold, fire) evolving many times in parallel (Drummond et al., 2012; Hughes and Eastwood, 2006; Sanderson and Wojciechowski, 1996; Simon et al., 2009, 2011). This suggests that the adaptability of legumes, and their ability to reinvent themselves time and again, are likely to have been an important factor contributing to their evolutionary success, as has been suggested for the angiosperms as a whole (Crepet and Niklas, 2009).

The difficulties of estimating extinction rates from phylogenies (Rabosky, 2010) and of distinguishing between increases in speciation rate which might typically characterize the early stage of an adaptive radiation from an extinction event with a constant rate of diversification were elegantly demonstrated by Crisp and Cook (2009). They showed that LTT plots for these two scenarios are very similar and that these plots result from 'broom and handle'-type tree topologies (a long stem branch subtending rapid diversification of many lineages subtended by short branch lengths). Broom and handle topologies and LTT plots for three clades of Australian legumes in the Mirbelieae + Bossieae and one African legume clade, tribe Podalyrieae, with coincident shifts in diversification in these clades, were proposed as possible examples of extinction events prompted by global cooling (Crisp and Cook, 2009). BayesRate provides an additional way to test for differences in net diversification and underlying speciation and extinction rates among clades or tree partitions delimited by rate shifts discovered by MEDUSA, and hence to potentially differentiate between the scenarios of extinction and increased diversification discussed by Crisp and Cook (2009). Of the clades analysed here, at least the *campos rupestres* clade in *Calliandra* exhibits a classical broom and handle topology (Fig. 3A), and the BayesRate analysis, when the stem lineage of this clade is included, shows robust support for a four-parameter model with shifts to both higher speciation and higher extinction rates at the base of the *campos rupestres* clade. A literal interpretation of the MEDUSA and BayesRate results would be one of on-going high extinction throughout the history of the *campos rupestres* clade (because the methods assume

that rates are constant within tree partitions, meaning that singular events such as mass-extinctions are not accounted for directly), it can, of course, not be ruled out that a single extinction event drove the phylogenetic pattern in *Calliandra*. An alternative interpretation for this clade would simply be one of recent rapid species radiation, whereby a single species persisted since divergence, with the clade only starting to diversify recently. Given the occurrence of such a high number of *Calliandra* species confined to the relatively small upland area of the Espinhaço Range, it is feasible to envisage a scenario of on-going high extinction under an ephemeral speciation model, whereby speciation is very common and very rapid, but the majority of produced species do not persist, but instead go extinct or are reabsorbed into parental forms (Rosenblum et al., 2012). This well-documented clade presents an intriguing study group for investigating these contrasting hypotheses. In contrast, the very rapid radiations documented in the New World montane clades of *Lupinus* (Drummond et al., 2012; Hughes and Eastwood, 2006), although subtended by hypothesized diversification rate increases (Fig. 3C, rate shifts 1–3), are not subtended by long branches (handles). For these *Lupinus* clades the BayesRate analyses favour a model of increased speciation rates with no signal of significant extinction, in line with previous evidence that these clades may indeed be examples of the early explosive stage of adaptive radiations (Drummond et al., 2012). However, estimation of extinction rates from phylogenies is highly prone to artefacts (Rabosky, 2010), therefore many of the results presented here would need further testing using high resolution fossil evidence, which, unfortunately is limited for legumes and most other angiosperm clades.

Given the dynamic nature of evolutionary processes, the 42 putative diversification rate shifts detected in legumes so far (Tables 1 & 2) are likely a subset of the potential rate shifts that may be expected. With more densely sampled phylogenies, more rate shifts can be anticipated, notably nested within terminal clades that include many species in the legume-wide analysis. For example, the very high absolute diversification rates estimated for two Andean clades of *Astragalus* (Scherson et al., 2008), suggest the potential for diversification rate shifts nested within this species-rich genus. Just how dynamic diversification processes have been will remain to be seen. The high number (2700) of rate shifts found across angiosperms as a whole almost certainly includes type I errors (Smith et al., 2011), while the rather small number of rate shifts for birds (25 in a clade of c. 10,000 species; Jetz et al., 2012) may be a product of the potentially over-conservative AICc cutoff currently implemented in MEDUSA. The low number of rate shifts for flies (<ten, confined to three areas in the phylogeny of a group of 152,000 named and many more unnamed species) (Wiegmann et al., 2011) is likely due to collapsing the tree to a small number of clades with assigned species numbers.

Estimating rates of species diversification is currently hampered by sparse gene and taxon sampling in phylogenies, urging caution when interpreting the phylogenetic locations of putative diversification rate shifts on particular branches, especially in parts of the tree with short branch lengths that remain weakly supported. Indeed, sparse taxon sampling and lack of resolution remain significant limitations of currently available legume phylogenies (Legume Phylogeny Working Group, 2013), particularly when it comes to estimating species diversification rates. However, in addition to topological and temporal uncertainty or limited taxon sampling in the phylogenies used, the diversification rate estimation methods employed here also have their limitations. MEDUSA gives an explicit result showing whether or not a model with different rates in parts of the tree is favoured over a model with a single rate over the whole tree. BayesRate uses a more robust statistical framework to test the significance of rate shifts with posterior estimates of all rates and Bayes Factor comparisons to evaluate model fit. Combining these two methods, as in the *Calliandra*, Indigoferaeae, *Lupinus* and *Mimosa* analyses presented here, provides a powerful approach for exploration and formulation of hypotheses. However, neither of these methods is adequate for

finding the optimal partitioning scheme of diversification rates on a tree. For example, in the legume-wide analysis, rate shift 4 shows an increase in rate of only 1.01-fold in the zero cutoff model with 32 partitions. However, in a 4-partition model, the increase is 1.24-fold for that particular shift, suggesting that the several nested rate shifts that are found through step-wise addition by MEDUSA can account for much of the increase that was found for rate shift 4 when it was originally added, to a point where that shift might not necessarily give a significantly better model fit. With BayesRate, one could test whether leaving out that particular rate shift, but keeping the other 30, might be favoured. However, using BayesRate on large phylogenies to evaluate more than just a few alternative partitioning schemes is impractical due to the large number of analyses that would be required. In most cases when rate shifts detected by MEDUSA in the species-level analyses only under the less stringent zero AICc cutoff were analysed in BayesRate, significant differences in rates between tree partitions were found. These considerations suggest that some of the zero cutoff rate shifts lead to a significantly better model fit and warrant further investigation. Furthermore, the AICc gain of the different rate shifts that are added by MEDUSA fluctuates (Table 1; e.g. compare the gain when adding rate shift 16 with that of 15, idem for rate shift 27 with several shifts before). This means that the better model fit of certain rate shifts during the step-wise addition can depend on previously added rate shifts. In the future, reversible jump MCMC (Green, 1995) or Bayesian stochastic search variable selection (BSSVS; George and McCulloch, 1993; as implemented in e.g. Drummond and Suchard, 2010) could be used to optimize and integrate over all possible partitioning schemes for rate shift models. Addition of a rate shift could then also be reversed in further MCMC iterations, as opposed to being kept fixed in further step-wise additions of shifts as in MEDUSA. Furthermore, even though modelling diversification rates with discrete shifts is an attractive and valid approach for hypothesis testing and comparisons between different partitions in a phylogenetic tree, these models might not accurately reflect macro-evolutionary processes, if in fact diversification rates vary more smoothly along lineages and within clades. The development of new models that incorporate smooth changes in diversification rate along branches and fitting those to simulated and empirical phylogenies will undoubtedly lead to further insights in how macro-evolution has shaped biodiversity.

More densely sampled phylogenies that span broad taxonomic, time, geographic and ecological ranges, and which are inferred from data from many genes, offer prospects for much more accurate and precise modelling of rate shifts across legumes and formulation of hypotheses to explain variation in species diversification rates among clades. In addition, assembly of equally densely sampled trait, geographical and ecological datasets will be needed to properly test associations between these factors and clades showing different rates of species diversification. The potential of some of the first mega-phylogenies for gaining insights into diversification has been recently demonstrated for angiosperms (Smith et al., 2011) and birds (Jetz et al., 2012), suggesting that for legumes, such analyses are within our grasp as larger and better phylogenies are developed in the near future.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.sajb.2013.07.005>.

Acknowledgements

We thank the Swiss National Science Foundation (Grant 31003A_135522 to C.E.H., and PBZHP3_147199 to J.M.de V.), the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES/PNPd no. 02697/09-2 to E.R.de S.), and the *Sistema Nacional de Pesquisa em Biodiversidade* (SISBIOTA, CNPq 563084/2010-3 and FAPESB PES0053/2011 to L.P.de Q. and E.R.de S.) for financial support, Daniele Silvestro for help with customizing BayesRate, and Matt Lavin, Peter Linder, Tommi Nyman and two reviewers for comments on an earlier version of the manuscript.

References

- Alfaro, M.E., Santini, F., Brock, C.D., Alamillo, H., Dornburg, A., Rabosky, D.L., Carnevale, G., Harmon, L.J., 2009. Lineage-specific diversification rates and high turnover in the history of jawed vertebrates. *Proceedings of the National Academy of Sciences* 106, 13410–13414.
- Baker, W.D., Couvreur, T.L.P., 2013. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. II. Diversification history and origin of regional assemblages. *Journal of Biogeography* 40, 286–298.
- Barneby, R.C., 1977. *Daleae Imagines*, an illustrated revision of *Errazurizia* Philippi, *Psorothamnus* Rydberg, *Marina* Liebmann, and *Dalea* Lucanus emend. Barneby, including all species of Leguminosae tribe Amorphaeae Borissowa ever referred to *Dalea*. *Memoirs of the New York Botanical Garden* 27, 1–892.
- Barneby, R.C., 1991. Sensitive *Censitae*: a description of the genus *Mimosa* L. (Mimosaceae) in the New World. *Memoirs of the New York Botanical Garden* 65, 1–835.
- Barneby, R.C., 1998. Silk tree, Guanacaste, Monkey's Earring. A generic system for the synandrous Mimosaceae of the Americas. Part 3 *Calliandra*. *Memoirs of the New York Botanical Garden* 74 (3), 1–223.
- Bello, M.A., Bruneau, A., Forest, F., Hawkins, J.A., 2009. Elusive relationships within order Fabales: phylogenetic analyses using matK and rbcL sequence data. *Systematic Botany* 34, 102–114.
- Bello, M.A., Rudall, P.J., Hawkins, J.A., 2012. Combined phylogenetic analyses reveal interfamilial relationships and patterns of flora evolution in the eudicot order Fabales. *Cladistics* 28, 393–421.
- Bruneau, A., Mercure, M., Lewis, G.P., Herendeen, P.S., 2008. Phylogenetic patterns and diversification in the Caesalpinioideae legumes. *Botany* 86, 697–718.
- Caccavari, M.A., Barreda, V., 2000. A new calymmate mimosoid polyad from the Miocene of Argentina. *Review of Palaeobotany and Palynology* 109, 197–203.
- Cannon, S.B., Ilut, D., Farmer, A.D., Maki, S.L., May, G.D., Singer, S.R., Doyle, J.J., 2010. Polyploidy did not predate the evolution of nodulation in all legumes. *PLoS One* 5, e11630.
- Cardoso, D., de Queiroz, L.P., Pennington, R.T., de Lima, H.C., Fonty, E., Wojciechowski, M.F., Lavin, M., 2012. Revisiting the phylogeny of Papilionoid legumes: new insights from comprehensively sampled early-branching lineages. *American Journal of Botany* 99, 1–23.
- Catalano, S.A., Vilardi, J.C., Tosto, D., Saidman, B.O., 2008. Molecular phylogeny and diversification history of *Prosopis* (Fabaceae: Mimosoideae). *Biological Journal of the Linnean Society* 93, 621–640.
- Crepet, W.L., Niklas, K.J., 2009. Darwin's second 'abominable mystery': why are there so many angiosperm species? *American Journal of Botany* 96, 366–381.
- Crisp, M.D., Cook, L.G., 2009. Explosive radiation or cryptic mass extinction? Interpreting signatures in molecular phylogenies. *Evolution* 63, 2257–2265.
- Cusimano, N., Stadler, T., Renner, S., 2012. A new method for handling missing species in diversification analysis applicable to randomly or non-randomly sampled phylogenies. *Systematic Biology* 61, 785–792.
- De-Nova, J.A., Medina, R., Montero, J.C., Weeks, A., Rosell, J.A., Olson, M.E., Eguarte, L.E., Magallón, S., 2012. Insights into the historical construction of species-rich Mesoamerican seasonally dry tropical forests: the diversification of *Bursera* (Burseraceae, Sapindales). *New Phytologist* 193, 276–287.
- Doyle, J.J., 2011. Phylogenetic perspectives on the origins of nodulation. *Molecular Plant–Microbe Interactions* 24, 1289–1295.
- Doyle, J.J., Luckow, M., 2003. The rest of the iceberg: legume diversity and evolution in a phylogenetic context. *Plant Physiology* 131, 900–910.
- Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7, 214.
- Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. *BMC Biology* 8, 114.
- Drummond, C.S., Eastwood, R.J., Miotto, S.T.S., Hughes, C.E., 2012. Multiple continental radiations and correlates of diversification in *Lupinus* (Leguminosae): testing for key innovation with incomplete taxon sampling. *Systematic Biology* 61, 443–460.
- Egan, A.N., Crandall, K.A., 2008. Divergence and diversification in North American *Psoraleae* (Fabaceae) due to climate change. *BMC Biology* 6, 55.
- George, E.I., McCulloch, R.E., 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881–889.
- Goldblatt, P., 1981. Cytology and the phylogeny of Leguminosae. In: Polhill, R.M., Raven, P.H. (Eds.), *Advances in Legume Systematics, Part 2*. Royal Botanic Gardens, Kew, pp. 427–464.
- Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82, 711–732.
- Herendeen, P.S., Dilcher, D.L., 1992. *Advances in Legume Systematics, Part 4. The Fossil Record*. Royal Botanic Gardens, Kew.
- Herendeen, P.S., Crepet, W.L., Dilcher, D.L., 1992. The fossil history of the Leguminosae: phylogenetic and biogeography implications. In: Herendeen, P.S., Dilcher, D.L. (Eds.), *Advances in Legume Systematics, Part 4. The Fossil Record*. Royal Botanic Gardens, Kew, pp. 303–316.
- Hughes, C.E., Eastwood, R., 2006. Island radiation on a continental scale: exceptional rates of plant diversification after uplift of the Andes. *Proceedings of the National Academy of Sciences* 103, 10334–10339.
- Hughes, C.E., Pennington, R.T., Antonelli, A., 2013. Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* 171, 1–18.
- Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. *Nature* 491, 444–448.
- Jobson, R.W., Luckow, M., 2007. Phylogenetic study of the genus *Piptadenia* (Mimosoideae, Leguminosae) using plastid trnL-f and trnK/matK sequence data. *Systematic Botany* 32, 569–575.

- Kass, R.E., Raftery, A.E., 1995. Bayes factors. *Journal of the American Statistical Association* 90, 773–795.
- Kursar, T., Dexter, K.G., Lokvam, J., Pennington, R.T., Richardson, J.E., Weber, M.G., Murakami, E.T., Drake, C., McGregor, R., Coley, P.D., 2009. The evolution of anti-herbivore defenses and their contribution to species coexistence in the tropical tree genus *Inga*. *Proceedings of the National Academy of Sciences* 106, 18073–18078.
- Lavin, M., 2006. Floristic and geographical stability of discontinuous seasonally dry tropical forests explains patterns of plant phylogeny and endemism. In: Pennington, R.T., Lewis, G.P., Ratter, J.A. (Eds.), *Neotropical Savannas and Seasonally Dry Forests*. CRC Press, Baton Rouge, Florida, pp. 433–447.
- Lavin, M., Sousa, M., 1995. Phylogenetic systematics and biogeography of the tribe Robinieae (Leguminosae). *Systematic Botany Monographs* 45, 1–165.
- Lavin, M., Wojciechowski, M.F., Gasson, P., Hughes, C., Wheeler, E., 2003. Phylogeny of robinoid legumes (Fabaceae) revisited: *Coursetia* and *Gliricidia* recircumscribed, and a biogeographical appraisal of the Caribbean endemics. *Systematic Botany* 28, 387–409.
- Lavin, M., Schrire, B., Lewis, G., Pennington, R.T., Delgado-Salinas, A., Thulin, M., Hughes, C., Beyra Matos, A., Wojciechowski, M.F., 2004. Metacommunity process rather than continental tectonic history better explains geographically structured phylogenies in legumes. *Philosophical Transactions of the Royal Society B* 359, 1509–1522.
- Lavin, M., Herendeen, P.S., Wojciechowski, M.F., 2005. Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the Tertiary. *Systematic Botany* 54, 575–594.
- Legume Phylogeny Working Group, 2013. Legume phylogeny and classification in the 21st century: progress, prospects and lessons. *Taxon* 62, 217–248.
- Lewis, G.P., Schrire, B.D., MacKinder, B., Lock, M., 2005. *Legumes of the World*. Royal Botanic Gardens, Kew.
- Maddison, W.P., Maddison, D.R., 2011. Mesquite: A modular system for evolutionary analysis. Version 2.75. <http://mesquiteproject.org>.
- Magallón, S., Sanderson, M.J., 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55, 1762–1780.
- Marazzi, B., Sanderson, M.J., 2010. Large-scale patterns of diversification in the widespread legume genus *Senna* and the evolutionary role of extrafloral nectaries. *Evolution* 64, 3570–3592.
- McKey, D., 1989. Interactions between ants and leguminous plants. In: Stirton, C.H., Zarucchi, J.L. (Eds.), *Advances in Legume Biology*. Monographs in Systematic Botany, 29. Missouri Botanical Garden, St. Louis.
- Moore, B.R., Donoghue, M.J., 2009. A Bayesian approach for evaluating the impact of historical events on rates of diversification. *Proceedings of the National Academy of Sciences* 106, 4307–4312.
- Moreau, C.S., Bell, C.D., 2013. Testing the museum versus cradle tropical biological diversity gradient hypothesis: phylogeny, diversification, and ancestral biogeographic range evolution of the ants. *Evolution*. <http://dx.doi.org/10.1111/evo.12105>.
- Morlon, H., Parsons, T., Plotkin, J., 2011. Inferring the dynamics of diversification: a coalescent approach. *PLoS Biology* 8, e1000493.
- Nee, S., May, R.M., Harvey, P.H., 1994. The reconstructed evolutionary process. *Philosophical Transactions of the Royal Society B* 344, 305–311.
- Pennington, R.T., Lavin, M., Prado, D.E., Pendry, C.A., Pell, S.K., Butterworth, C.A., 2004. Historical climate change and speciation: neotropical seasonally dry forest plants show patterns of both Tertiary and Quaternary diversification. *Philosophical Transactions of the Royal Society B* 359, 515–538.
- Pennington, R.T., Richardson, J.E., Lavin, M., 2006. Insights into the historical construction of species-rich biomes from dated plant phylogenies, neutral ecological theory and phylogenetic community structure. *New Phytologist* 172, 605–616.
- Pennington, R.T., Lavin, M., Oliveira-Filho, A., 2009. Woody plant diversity, evolution and ecology in the tropics: perspectives from seasonally dry tropical forests. *Annual Review of Ecology, Evolution, and Systematics* 40, 437–457.
- Rabosky, D.L., 2010. Extinction rates should not be estimated from molecular phylogenies. *Evolution* 64, 1816–1824.
- Richardson, J.E., Pennington, R.T., Pennington, T.D., Hollingsworth, P.M., 2001. Rapid diversification of a species-rich group of Neotropical rain forest trees. *Science* 293, 2242–2245.
- Rosenblum, E.B., Sarver, B.A.J., Brown, J.W., Des Roches, S., Hardwick, K.M., Hether, T.D., Eastman, J.M., Pennell, M.W., Harmon, L.J., 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evolutionary Biology* 39, 255–261.
- Sanderson, M.J., Donoghue, M.J., 1994. Shifts in diversification rate with the origin of the angiosperms. *Science* 264, 1590–1593.
- Sanderson, M.J., Liston, A., 1995. Molecular phylogenetic systematic of Galegeae, with special reference to *Astragalus*. In: Crisp, M.D., Doyles, J.J. (Eds.), *Advances in Legume Systematics 7: Phylogeny*. Royal Botanic Gardens, Kew, pp. 331–350.
- Sanderson, M.J., Wojciechowski, M.F., 1996. Diversification rates in a temperate legume clade: are there 'so many species' of *Astragalus* (Fabaceae)? *American Journal of Botany* 83, 1488–1502.
- Santini, F., Harmon, L.J., Carnevale, G., Alfaro, M.E., 2009. Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC Evolutionary Biology* 9, 194.
- Särkinen, T., Pennington, R.T., Lavin, M., Simon, M.F., Hughes, C.E., 2012. Evolutionary islands in the Andes: persistence and isolation explain high endemism in Andean dry forests. *Journal of Biogeography* 39, 884–900.
- Scherson, R.A., Vidal, R., Sanderson, M.J., 2008. Phylogeny, biogeography, and rates of diversification of New World *Astragalus* (Leguminosae) with an emphasis on South American radiations. *American Journal of Botany* 95, 1030–1039.
- Schnitzler, J., Barraclough, T.G., Boatwright, J.S., Goldblatt, P., Manning, J.C., Powell, M.P., Rebelo, T., Savolainen, V., 2011. Causes of plant diversification in the Cape biodiversity hotspot of South Africa. *Systematic Biology* 60, 343–357.
- Schrire, B.D., Lavin, M., Lewis, G.P., 2005. Global distribution patterns of the Leguminosae: insights from recent phylogenies. *Biologische Skrifter* 55, 375–422.
- Schrire, B.D., Lavin, M., Barker, N.P., Forest, F., 2009. Phylogeny of the tribe Indigoferaeae (Leguminosae–Papilionoideae): geographically structured more in succulent-rich and temperate settings than in grass-rich environments. *American Journal of Botany* 96, 816–852.
- Scotland, R.W., Sanderson, M.J., 2004. The significance of few versus many in the tree of life. *Science* 303, 643.
- Silvestro, D., Schnitzler, J., Zizka, G., 2011. A Bayesian framework to estimate diversification rates and their variation through time and space. *BMC Evolutionary Biology* 11, 311.
- Simon, M.F., Grether, R., de Queiroz, L.P., Skema, C., Pennington, R.T., Hughes, C.E., 2009. Recent assembly of the Cerrado, a neotropical plant diversity hotspot, by evolution of adaptations to fire. *Proceedings of the National Academy of Sciences* 106, 20359–20364.
- Simon, M.F., Hughes, C.E., Harris, S.A., 2010. Four new species of *Mimosa* (Leguminosae) from the central highlands of Brazil. *Systematic Botany* 35, 277–288.
- Simon, M.F., Grether, R., de Queiroz, L.P., Särkinen, T., Dutra, V., Hughes, C.E., 2011. The evolutionary history of *Mimosa* (Leguminosae): toward a phylogeny of the sensitive plants. *American Journal of Botany* 98, 1201–1221.
- Slowinski, J.B., Guyer, C., 1989. Testing the stochasticity of patterns of organismal diversity: an improved null model. *American Naturalist* 134, 907–921.
- Smith, S.A., Beaulieu, J.M., Stematakis, A., Donoghue, M.J., 2011. Understanding angiosperm diversification using small and large phylogenetic trees. *American Journal of Botany* 98, 404–414.
- Souza, E.R. de, Queiroz, L.P. de, Lewis, G.P., Forest, F., Schnadelbach, A.S., Van den Berg, C., 2013. Phylogeny of *Calliandra* (Leguminosae: Mimosoideae) based on nuclear and plastid molecular markers. *Taxon* (submitted for publication).
- Sprent, J.I., 2001. *Nodulation in Legumes*. Royal Botanic Gardens, Kew.
- Stadler, T., 2009. On incomplete sampling under birth–death models and connections to the sampling-based coalescent. *Journal of Theoretical Biology* 261, 58–66.
- Stadler, T., 2011a. Mammalian phylogeny reveals recent diversification rate shifts. *Proceedings of the National Academy of Sciences* 108, 6187–6192.
- Stadler, T., 2011b. Simulating trees on a fixed number of extant species. *Systematic Biology* 60, 676–684.
- Stadler, T., 2013. Recovering speciation and extinction dynamics based on phylogenies. *Journal of Evolutionary Biology* 26, 1203–1219.
- Stadler, T., Bokma, F., 2012. Estimating speciation and extinction rates for phylogenies of higher taxa. *Systematic Biology* 62, 220–230.
- Villiers, J.-F., 2002. Tribe Mimosaeae. In: Du Puy, D.J., Labat, J.-N., Rabevohitra, R., Villiers, J.-F., Bosser, J., Moat, J. (Eds.), *Leguminosae of Madagascar*. Royal Botanic Gardens, Kew, pp. 159–223.
- Wagner, C.E., Harmon, L.J., Seehausen, O., 2012. Ecological opportunity and sexual selection together predict adaptive radiation. *Nature* 487, 366–369.
- Wang, H., Moore, M.J., Soltis, P.S., Bell, C.D., Brockington, S.F., Alexandre, R., Davis, C.C., Latvis, M., Manchester, S.R., Soltis, D.E., 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* 106, 3853–3858.
- Wiegmann, B.M., Trautwein, M.D., Winkler, I.S., Barr, N.B., Kim, J.-W., Lambkin, C., Bertone, M.A., Cassel, B.K., Bayels, K.M., Heimberg, A.M., Wheeler, B.M., Peterson, K.J., Pape, T., Sinclair, B.J., Skevington, J.H., Blagoderov, V., Caravas, J., Kutty, S.N., Schmidt-Ott, U., Kampmeier, G.E., Thompson, F.C., Grimaldi, D.A., Beckenbach, A.T., Courtney, G.W., Friedrich, M., Meier, R., Yeates, D.K., 2011. Episodic radiations in the fly tree of life. *Proceedings of the National Academy of Sciences* 108, 5690–5695.
- Wing, S.L., Herrera, F., Jaramillo, C.A., Gómez-Navarro, C., Wilf, P., Labandeira, C.C., 2009. Late Paleocene fossils from the Cerrejón Formation, Colombia, are the earliest record of Neotropical rainforest. *Proceedings of the National Academy of Sciences* 106, 18627–18632.
- Wojciechowski, M.F., Lavin, M., Sanderson, M.J., 2004. A phylogeny of the legumes (Leguminosae) based on analysis of the plastid matK gene resolves many well-supported subclades within the family. *American Journal of Botany* 91, 1846–1862.
- Yang, Z., Rannala, B., 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. *Molecular Biology and Evolution* 14, 717–724.