

# **Phylogenetic biome niche conservatism underlies the evolution of forest paleoendemic legume trees in tropical Africa**

Dario I. Ojeda<sup>1</sup>, Anaïs-Pasiphae Gorel<sup>2,3</sup>, Gilles Dauby<sup>4</sup>, Sandra Cervantes<sup>5</sup>, Aaron D. Pan<sup>6</sup>, Thomas Marcussen<sup>7</sup>, Samuel Vanden Abeele<sup>8</sup>, Arthur Boom<sup>9</sup>, Félix Forest<sup>10</sup>, Manuel de la Estrella<sup>11</sup>, Olivier J. Hardy<sup>12</sup>

<sup>1</sup>Finnish Museum of Natural History, LUOMUS, Botany and Mycology Unit, University of Helsinki, Helsinki, Finland

<sup>2</sup>Laboratory of Plant Ecology, Department of Plants and Crops, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

<sup>3</sup>Forest is Life, Gembloux Agro-Bio Tech, Université de Liège, Gembloux, Belgium

<sup>4</sup>MAP, Univ. Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier University, France

<sup>5</sup>Department of Forest Sciences, University of Helsinki, Helsinki, Finland

<sup>6</sup>Museum of Texas Tech University, Texas Tech University, USA

<sup>7</sup>Department of Biosciences, University of Oslo, Norway

<sup>8</sup>Department of Plant Sciences, University of Cambridge, UK

<sup>9</sup>Royal Museum for Central Africa, Biology Department, Section Vertebrates, Tervuren, Belgium

<sup>10</sup>Royal Botanic Gardens, Kew, Richmond, UK

<sup>11</sup>Departamento de Botánica, Ecología y Fisiología Vegetal, Facultad de Ciencias, Universidad de Córdoba, Spain

<sup>12</sup>Evolutionary Biology and Ecology, Faculté des Sciences, Université Libre de Bruxelles, Belgium

To whom correspondence may be addressed: [dario.ojedaalayon@helsinki.fi](mailto:dario.ojedaalayon@helsinki.fi)

## Summary

Niche conservatism is prevalent during the evolution of plant lineages. However, inferring biome niche lability and its impact on tropical tree species diversification is currently limited. To better understand biome niche lability and its effect on diversification rates, we analyzed an endemic lineage of African tropical trees, testing whether biome niche type (forest vs savanna) and biome lability are non-randomly distributed. We reconstructed a time-calibrated phylogeny of the *Berlinia* clade (16 genera, ca 201 species) using 140 nuclear genes, 75% of its extant species, and eight fossil calibrations. We also analyzed the phylogenetic signal of biome shifts across the group and its effect on diversification and extinction rates. We found the forest biome as the ancestral condition and a directional shift from forest to savanna (no reversals). The forest biome type is conserved in the *Berlinia* clade, but the lability to shift biomes is randomly distributed in the group. We did not find evidence that biome shift was associated with higher diversification or lower extinction rates. Our analyses identified five paleoendemic genera that have persisted solely in the forest biome since the Eocene. These paleoendemic and forest-restricted lineages are potentially more susceptible to habitat alterations of human related activities and future climate change, due to perhaps an innate limitation to adapt to new habitat types.

**Keywords:** Tropical Africa, paleoendemism, diversification rates, niche conservatism, phylogenomics, phylogenetic signal, Detarioideae.

# I. Introduction

Elucidating the distribution of biome niche conservatism and niche lability (the propensity to evolve and transcend major ecological boundaries) through the evolutionary history of plant lineages can provide useful insights into their evolutionary potential under future climate change. This knowledge is particularly relevant for tropical tree lineages, which due to their longer generation times, restricted distribution and habitat specificity might render these species more susceptible to changing climatic conditions. Previous studies at regional [1,2] and global levels [3,4] suggest that biome niche conservatism has been prevalent during angiosperm evolution. The resulting phylogenetic niche conservatism (PNC) is reflected in the consistency of the climate niche of tropical forest tree lineages along similar rainfall gradients across continents [5]. These studies indicate that PNC have played a major role shaping the current distribution of plant diversity.

Biomes capture major vegetation types defined by climate, life form and ecophysiology. Shifts among biome types are considered major transitions involving important adaptations [6]. Understanding these transitions are particularly relevant in regions where major changes in biome types have occurred in the past over wide evolutionary time frames, under predicted human disturbances, and future climate change [7]. In Africa, the tropical rainforest biome has experienced an overall decline of area since its maximum extension in the early Eocene to the Middle Miocene (55 -11 Ma) [8]. The contraction of the forest biome has been proposed as one of the reasons for the depauperate species diversity of many tropical tree lineages in Africa (e.g., palms and some legume lineages) [9] compared to their counterparts in Asia and the Neotropical regions [10,11]. On the other hand, the overall decrease of rainforest area since the Eocene in tropical Africa has also made new habitat types available (e.g., savanna biome), created opportunities for biome shifts, and changes in diversification rates. In spite of this, a recent analysis suggested that niche conservatism of biomes (forest and savanna) is prevalent in most (93%) speciation events in tropical African tree species at the genus level [12]. Indeed, evidence shows that several arborescent forest taxa contracted and expanded their distribution range with the cycles of rainforest changes (without shifting biomes), following the extension of the forest biome [5,13,14]. Thus, despite these repeated forest biome contraction/expansion

cycles and the availability of new biomes, most African tree lineages have maintained their biome across wide evolutionary time frames. It thus seems that the capacity to shift between the forest and savanna biomes is unevenly distributed across clades [12]. This suggests that some lineages could be more susceptible to higher extinction rates when the availability of these biomes is altered. However, understanding how biome niche lability is distributed among lineages and the timing of their shifts require robust time-calibrated phylogenetic reconstructions at the species level with good representation of extant taxa. We currently lack this knowledge for most tropical trees at the local and global levels, particularly for tree lineages that are endemic to the African tropical region.

In Africa, the Guineo-Congolian lowland rainforest has been identified as a region of plant paleoendemism [15,16]. Paleoendemic genera (taxa that diverged early and with a restricted distribution) have been previously suggested in the *Berlinia* clade (Subfamily Detarioideae, Leguminosae) [17], a group of tropical tree legumes that diversified in the Guineo-Congolia region. The *Berlinia* clade consists of 16 genera, nine of which are composed of at least 10 species, totaling ca. 201 species [18]. Unlike other African forest tree taxa, the *Berlinia* clade seems to have experienced an increase in diversification rates since the Miocene [17]. The most recent analyses within this clade indicate different diversification patterns, where some lineages seem to have exploited niches within the forest biome, while others were able to exploit the new biomes, such as open grasslands and woodlands, that became available as the forest cover area reduced [17,19]. However, limited representation of the group has hindered the reconstruction of the distribution of biome niche lability and whether this capacity has led to higher diversification and/or extinction rates.

To understand more in depth the distribution of biome niche conservatism and lability, we performed a phylogenomic reconstruction of all 16 genera of the *Berlinia* clade, including 75% of all extant species. We generated a time-calibrated phylogeny using eight fossils from the Eocene onwards to obtain the phylogenetic placement and the timing of biome shifts. Specifically, we aim to test the following hypotheses: 1) biome type shows phylogenetic signal suggesting that some lineages (genera) within the *Berlinia* clade have remained in their ancestral biomes despite the availability of new habitats during their evolutionary history, 2) the capacity to shift biomes (biome lability) from forest to savanna is non-randomly distributed within

the *Berlinia* clade at the genus level, 3) lineages with the capacity to shift biomes show higher diversification rates compared to lineages that retain the ancestral forest biome, and 4) lineages that retained their ancestral forest biome have higher extinction rates compared to lineages that shifted to the savanna biome.

## 2. Materials and Methods

### (a) The study system

The *Berlinia* clade comprises an estimated 201 species grouped in 16 genera [20]. The entire clade is endemic to Africa and here we included 152 species, representing all 16 genera, and 75% of the species (Table S1). We used *Paramacrolobium coeruleum* as an outgroup species in all the analyses.

### (b) DNA extraction and target enrichment

DNA was extracted from leaves (25-35 mg) obtained from herbarium specimens or silica gel dried samples using a modified CTAB protocol [21] and cleaned using the QIAquick PCR Purification Kit (Qiagen, Venlo, Netherlands). We employed a modification of the protocol for plastome capture [22] for library preparation. Hybrid enrichment was performed on pools of 48 samples per reaction according to the MYbaits v2.3.1 protocol, using 23 h of hybridization, a high stringency post-hybridization wash, and a final amplification using 15 PCR cycles. The Detarioideae v.1 bait [19] was used for this study. This bait set consists of 6,565 probes (120 bp long overlapping baits) targeting 1021 exons from 289 genes. Paired-end sequencing (2 × 150 bp) was performed on an Illumina NextSeq with reagent kit V2 at the GIGA platform (Liège, Belgium), assigning ca. 400,000 reads/sample. Orthologs were recovered with the strategy developed previously [23]. First, we assembled *de novo* reads for each species using SPAdes ver. 3.9 [24] and reduced redundancy of the clusters recovered using CD-HIT [25] (99%, threshold, word size = 10). Then we performed an all-by-all blast on all the samples and later filtered with a hit fraction cut-off of 0.5. We applied MCL [26] using an inflation value of 1.4 to reduce identified clusters in the samples. Clusters with less than 1,000 sequences were aligned with MAFFT [27] (-genafpair -maxiterate 1,000, and 0.1 minimal column occupancy) and tree

inference was generated with RAxML v. 8.2.9 [28]. For larger clusters (> 100 sequences) we used PASTA [29] with minimal column occupancy of 0.01 and trees inferred using fasttree [30]. Finally, orthologues were selected using the strict one-to-one strategy. This final step allows the identification and exclusion of paralog sequences, which is an advantage over other pipelines that only identify these paralog regions [31,32]. The matrix was inspected and formatted with AliView ver. 1.28 [33].

### **(c) Phylogenomic reconstruction with individual orthologs, concatenated and multispecies coalescent approaches**

We performed phylogenetic analyses on each separate ortholog (gene tree) and on the concatenated matrix (supermatrix) with maximum likelihood (ML) as implemented in RAxML ver. 8.2.9 [28] using the GTRCAT model with -f a flags, 1,000 bootstraps, and default settings. We also carried out a ML analysis as implemented in IQ-TREE 1.6.1 [34] with ultrafast likelihood bootstrap with 1,000 replicates. In addition, we carried out a Bayesian analysis using MrBayes 3.2.6 [35,36] on the concatenated matrix. For the Bayesian analyses we applied four chains, two runs of 50,000,000 generations with the invgamma rate of variation, and a sample frequency of 5,000. The performance of the Bayesian analysis was assessed with Tracer 1.7 [37]. The species tree estimation was performed under the coalescent model using the individual ML gene trees obtained with RAxML to infer a species tree with ASTRAL-II v. 5.5.7 [38]. Support was estimated using local posterior probability (LPP). We used *Paramacrolobium coeruleum* as outgroup taxon in all inferences. The obtained trees from all the analyses were visualized and edited with FigTree (Rambaut, 2016).

### **(d) Climate niche modeling and categorization of biomes**

To categorize biomes within the Berlinia clade, we used a database of 4,142 tree species across tropical Africa extracted from the RAINBIO project [39]. The bioclimatic group of the sequenced species from the Berlinia clade were inferred from georeferenced coordinates and predicted using niche similarity, a trait that encompasses the wide range of strategies allowing plants to persist in specific climatic conditions [12]. This modeling resulted in the classification of two biomes, which corresponded to the forest and savanna biomes. The biome assignment of

the *Berlinia* species was further corroborated with literature, descriptions from herbarium specimens, and from our own field expertise of the distribution of the species. Biomes were further corroborated with recent taxonomic revisions for all representatives in *Anthonotha* [40], *Englerodendron* [41], *Microberlinia* [42], and the monotypic genera *Icuria* [43], *Michelsonia* [42], and *Librevillea* [44,45] (Table S3).

#### **(e) Ancestral state reconstruction and biome shifts**

We mapped the two biomes at the species level to reconstruct ancestral states within the *Berlinia* clade and to infer the number of biome shifts. Biomes and ancestral state reconstructions were analyzed on the best RAxML tree recovered from the concatenated dataset. These reconstructions were performed using parsimony and likelihood methods as implemented in Mesquite ver. 2.75 [46]. The number of inferred biomes shifts from this analysis was later used to estimate the level of aggregation of biome lability (see below). We estimated the delta statistic ( $\delta$ ) as a measure of phylogenetic signal [47] for biome (forest and savanna) using the “delta” function in the R package ‘ape’ version 5.3 [48]. Delta measures the degree of phylogenetic signal between any given categorical trait and a phylogeny of the group of interest. The obtained  $\delta$ -value was then compared to the  $\delta$ -values obtained from 100 random permutations of the tips in our phylogenetic tree. A p-value less than the level of the test (0.05) means phylogenetic signal between the trait and the tree. We used this p-value to determine if biome type has a phylogenetic signal in this group. In addition, we assessed the phylogenetic signal by estimating D, a metric based on the sum of sister-clade differences in the phylogeny [49], using the R package caper. We performed 1000 permutations and compared the probability under a random ( $D = 1$ ) phylogenetic structure. Finally, we evaluated the level of aggregation of biome lability (labile vs. non labile) using the mean phylogenetic distance between the number of instances where this trait evolved in the *Berlinia* clade using the tree we obtained from the BEAST calibration. First, we first converted to an ultrametric tree using the chronos function in ape ( $\lambda = 0$ ). We tested the null hypothesis that the instances of biome lability at the genus level (six nodes for the genera *Julbernardia*, *Bikinia*, *Aphanocalyx*, *Brachystegia*, *Berlinia*, and *Isobertinia*) in the *Berlinia* clade were aggregated compared to 1000 random set of nodes using a custom-made R script.



## (f) Divergence time estimates

There are several fossils assigned to extant genera of the *Berlinia* clade and we employed a total of eight fossils to constrain the ages of internal nodes within the clade (Fig. 1 and Table S5) [50–57]. Divergence times were estimated with BEAST v. 1.10.3 [58] under a HYK substitution model with strict clock, coalescence constant-size, and a speciation Yule process, and a lognormal distribution for the calibration points. We employed eight fossils to generate priors of stem node age for four nodes in the topology (in subclades A and B, Table S5). These priors were generated using the age of the oldest fossil and the number of fossil sites that corresponded to each lineage. Here we employed a modification within the *Berlinia* clade (Table S6) of the approach previously used to calculate these priors [59]. In subclade A, we used the following parameters for the *Brachystegia* stem:  $\mu = 1.9673$ ,  $\sigma = 0.3993$ ,  $\text{offset} = 23.03$ , and for the *Aphanocalyx* fossil:  $\mu = 5.0252$ ,  $\sigma = 1.0939$ ,  $\text{offset} = 46$ . In subclade B, for *Anthonotha* and *Englerodendron* fossils, we applied  $\mu = 4.2752$ ,  $\sigma = 1.0939$ , and  $\text{offset} = 21.73$ . We ran two independent runs of 200 million generations, sampling trees and parameters every 10,000 generations, and a final burn-in of 20,000 million generations. Tracer 1.7 [37] was used to assess convergence among the chains as well as to evaluate the ESS parameter ( $\text{ESS} > 200$ ).

## (g) Diversification rates in the forest and savanna lineages

### (i) Time-dependent diversification

To assess the variation on diversification over time, we performed a time-dependent diversification analysis as implemented in RPANDA [60]. Our dated tree covered all extant genera and 76% of the species (fraction,  $f = 152/201$ ). We tested six models: 1) constant speciation rate ( $\lambda_{\text{con}}$ ) with no extinction (Yule null model), 2) exponential variation of speciation rate ( $\lambda_{\text{exp}}$ ) with no extinction, 3) linear variation of speciation rate ( $\lambda_{\text{lin}}$ ) with no extinction, 4) constant speciation and extinction rates, 5) exponential speciation rates with a constant extinction rate, and 6) linear speciation rates with a constant extinction rate [60]. These models estimate current speciation and extinction rates with their corresponding parameters by measuring their variation over time up to the crown age. The diversification rates and major



shifts were also evaluated using a Bayesian approach with the reversible MCMC method as implemented in BAMM v.2.5.0 [61,62]. Our interest was to identify lineages (at the genus level) exhibiting significant rate shifts in speciation, extinction, and net diversification, and to conduct a rate-through-time analysis of these shifts within the entire Berlinia clade. Incomplete sampling of the remaining species was accounted for by specifying genus-specific sampling fractions. Priors for BAMM were generated with BAMMtools v.2.1 [61] in the R statistical environment [63]. The BAMM analysis was run with four Markov chains for 50,000,000 generations sampling every 5,000 generations. Considering the size of the phylogeny, we set the expected number of shifts to 1. The convergence was assessed by the ESS of likelihood and the number of shifts (ESS > 200). We removed 20% of the samples from the posterior distribution sampled by BAMM and analyzed the output with BAMMtools. We computed 95% credible rate shift configurations, estimated the clade (focusing on the genus level) specific evolutionary rates by obtaining rates-through-time (RTT) plots ( $\lambda$ ,  $\mu$ , and  $r$ ), and obtained the visualization of the mean phylorate plots. In addition, we used the R package phytools ver. 1.5-1 [64] to generate lineage-through-time (LTT) plots to determine the temporal pattern of lineage accumulation in the Berlinia clade and its 16 genera. All tests were performed on the dated phylogeny obtained with BEAST.

## **(ii) Biome and trait-dependent diversification rates**

We assessed the effect of biome shifts as potential drivers of the diversification in the Berlinia clade. We used the two biomes (forest and savanna) as previously identified in the climate niche modeling. In addition, we also explored the effect of biome lability on the diversification rates. To investigate whether biome or biome lability have promoted diversification rates in the Berlinia clade, we used the BiSSE model [65,66] where extinction and speciation rates are associated with phenotypic evolution of a trait along a phylogeny. We used the dated tree obtained with BEAST and ran four independent MCMC chains for  $10 \times 10^4$  steps for each of 10 randomly chosen trees using an exponential prior for the rates. We discarded the first 25% of steps of each chain as burn-in. We also tested differences in speciation rates between the two biomes and labile vs not labile using maximum likelihood (ML) and tested for significant differences among the models with ANOVA tests. Both ML and MCMC were run as implemented in the R package diversitree [66].

### 3. Results

#### (a) Phylogenomic reconstructions and major lineages within the *Berlinia* clade

The aligned matrix of 140 nuclear DNA sequences (orthologs) consisted of 56,605 bp, of which 12,890 were variable sites and 11% were parsimony informative sites (Table S2). We recovered similar topologies using ML, Bayesian, and species tree with ASTRAL, obtaining highly resolved relationships among genera and within each of the 16 genera of the *Berlinia* clade (Fig. 1, Figs. S1-S5). We recovered three major lineages (subclades A-C), congruent with previous results on their taxonomic relationships [42] and with previous phylogenetic analyses based on a more limited representation of the group [17,19,41]. All genera within the *Berlinia* clade were recovered as monophyletic, except *Isoberlinia*, where *I. scheffleri* was found as sister to a clade comprising genus *Berlinia* sister to the remaining species of *Isoberlinia* included in this study (Fig. 1 and Figs. S1-S5).

#### (b) Biome shifts within the *Berlinia* clade

Using a comprehensive database of over 4,000 African tree species, along with the coordinate locations of the *Berlinia* clade specimens used in this study, we identified two biomes corresponding to the forest and savanna. Nearly 80% of the species within the *Berlinia* clade were classified as inhabiting the forest biome (Table S3). We inferred the forest biome as the ancestral habitat for the *Berlinia* clade, as well as all its 16 genera. Biome lability was observed in six genera (*Berlinia*, *Julbernardia*, *Bikinia*, *Brachystegia*, *Isoberlinia*, and *Aphanocalyx*), with nine independent transitions from the forest to savanna, all localized in subclades A and B (Fig. 2). No reversals from the savanna to forest biome were observed. We identified two distinct patterns of biome shifts. Firstly, one shift to savanna with subsequent diversification within the savanna lineages in *Brachystegia*, *Julbernardia* and *Isoberlinia* (excluding *I. scheffleri*). Secondly, one to three independent shifts to savanna within each genus, but without further diversification in the savanna lineages in *Bikinia*, *Aphanocalyx*, *Julbernardia*, and *Berlinia*. In *Julbernardia* we found the two previous patterns occurring at different time scales (Fig. 3). We did not infer any biome shift in the remaining 11 genera, while these represent some of the most species-rich genera (e.g., *Berlinia*, *Gilbertiodendron*, *Anthonotha*, and *Englerodendron*).

## **(c) Timing of biome shifts and its distribution among the major lineages within the Berlinia clade**

Our dating analyses indicate that the Berlinia clade represents an ancient lineage that likely originated in the Paleocene (64 Ma), where 50% of its current genera originated before the Miocene (Fig. 3 and Table S4). The earliest shifts to the savanna biome occurred in the Miocene, in *Julbernardia* at about 17 Ma, in *Brachystegia* at about 15 Ma, and in *Isoberlinia* at 10 Ma. These three early biome shifts are associated with the pattern where we inferred one shift to the savanna biome with subsequent diversification within this biome. Thus, as the forest cover decreased, it seems that shifting to the newly available niche was only exploited early in the evolution of the Berlinia clade, and limited to subclades A and B. The other shifts to the savanna biome occurred in the mid-Miocene to early Pliocene in *Bikinia*, *Julbernardia*, *Aphanocalyx*, and *Berlinia*. These more recent shifts to the savanna biome are characterized by a single shift without further diversification (Fig. 3). We found phylogenetic signal for biome type with the two metrics we employed to assess non-random distribution:  $\delta = 190.07$  (p-value = 0) and  $D = -0.34$  (p-value = 0). These results suggest that biome type is conserved in the Berlinia clade. However, our aggregation analysis suggests that the capacity to shift biomes (biome lability) is randomly distributed in the Berlinia clade at the genus level (p-value = 0.099).

## **(d) Diversification rates within the forest and savanna lineages**

### **(i) Time-dependent diversification**

Our analyses suggest an overall increase in species number within the Berlinia clade (Fig. 4A). Of the six models tested with RPANDA, we found a constant speciation rate with no extinction as the best supported model for our data set (Table 1). The Lineages Through Time (LTT) plot indicates an increase of lineages since the early Eocene, with a marked increase from the mid-Miocene to the present (Fig. 4B). A similar trend on net diversification rates was inferred with BAMM from the Miocene (Fig. 4C), due to an increased speciation rate with a constant extinction rate (Fig. S6-S9). All our time dependent diversification analyses suggest an overall increase in net diversification within the Berlinia clade, both in forest and savanna lineages,

suggesting that both groups might have exploited additional niche spaces (e.g., pollinators, soil types) within the forest and savanna biomes.

## **(ii) Biome and trait-dependent diversification rates**

To test the effect of the two biomes (forest and savanna) and the lability to shift these biomes at the genus level (labile and not labile) on diversification rates, we performed the binary state speciation and extinction (BiSSE) approach using both a ML and a MCMC method. We did not find significant differences between a constant rate and an association with the biome ( $\alpha = 0.05$ ,  $p = 0.305$ ) or the lability to shift between biomes ( $\alpha = 0.05$ ,  $p = 0.172$ ), suggesting that these two attributes do not seem to be associated with changes in diversification rates. The posterior probabilities of the MCMC of speciation and extinction rates for both traits overlap and indicate no difference with a constant rate of speciation in this clade (Fig. S10).

## **4. Discussion**

### **(a) Phylogenetic niche conservatism, but random distribution of biome niche lability**

Phylogenetic niche conservatism (i.e., the retention of ancestral ecological traits and environmental distribution following speciation) is considered a major determinant of global biodiversity patterns. Evidence at the community level [67,68] as well as large geographic scale analyses suggest a high prevalence of niche conservatism [1,2,4]. We found a similar trend within the *Berlinia* clade, as most of the genera within this group (11 out of 16) have diversified within the same biome. In contrast to previous analyses based on genus-level phylogeny [12], our analysis provides a more detailed picture of the distribution of biome lability of African tropical trees. Our dating and ancestral reconstruction analyses within the group suggests that the *Berlinia* clade likely originated in the forest biome during the Paleocene (ca. 64 Ma), with the divergence of the three main subclades (A-C) in the Eocene. All this suggests that several genera within subclades A and B had expanded former distributions, retained their ancestral biome over long evolutionary time frames, and have followed the contraction of this biome. For example, there are no extant members of *Anthonotha* and *Englerodendron* in Ethiopia or Kenya, where early Miocene fossils of these genera were found, so that these genera do not

seem to have exploited the availability of new drier habitat types in this geographic region [56,57]. Forests in East Africa were progressively disrupted starting in the early Miocene and often gradually replaced by deciduous forest and woodlands [53]. All extant species of *Anthonotha* and *Englerodendron* are currently distributed in the Guineo-Congolian region in central Africa [40,41]. Our results thus suggest that there are some lineages (genera) within the Berlinia clade that seem to inherently lack the capacity to exploit drier new available habitats, however this capability seems to be randomly distributed in this group. We also found no reversals to the forest biome once a lineage has shifted to savanna. Although our biome classifications might have missed some additional biomes (e.g. montane and seasonally dry forest) that were not captured, our main conclusions will not be affected if some of the species we included as forest or savanna and later reclassified into additional biomes.

#### **(b) Timing of biome shifts and diversification patterns in the Berlinia clade**

Previous age estimates of the Berlinia clade based on several fossil calibrations within Detarioideae range between 15.4 Ma to 48.4 Ma [20,69,70]. These previous dating analyses included a limited representation of these genera and either used a secondary calibration point or a single fossil [19]. We inferred older estimates in the Berlinia clade compared to previous studies (sometimes by a factor of three). The inclusion of several Berlinia clade fossils in the current study and a more complete representation of species increased the reliability of the age estimates of the origin of all 16 genera within this clade. The recent report of fossils unequivocally assigned to the Berlinia clade from the early Eocene further supports the age estimates in our study. For example, fossils assigned to the stem of genera in the Berlinia subclade B (*Striatopollis catatumbus*) have been reported from the early Eocene (Ypresian) 51.9 Ma [50], as well as fossils from Berlinia clade A (*Peregrinipollis nigericus*) [51]. Also, well-preserved leaf/lamina fossils have been recently reported from Ethiopia of prehistoric species of forest lineages (*Anthonotha shimaglae* and *Englerodendron mulugetanum*) from the early Miocene (21.7 Ma) [56,57]. Our dating analyses question the hypothesis that the diversification of legumes into six subfamilies started after the Cretaceous-Paleogene boundary mass extinction (66 Ma) [71], although it remains compatible with some of the clock models explored by [71], which estimated the crown age of legumes at 75-80 Ma. Hereafter, we interpret our

results assuming that fossil calibrations in the *Berlinia* clade are more reliable than previous ones, although the timing of diversification of legumes and their subfamilies remains an open question.

Our findings suggest nine independent shifts from forest to savanna, following two different patterns of biome evolution, with the earliest shifts occurring during the Miocene and followed by diversification in both savanna and forest lineages. However, we did not find evidence that biome shifts have promoted higher diversification rates in the *Berlinia* clade, similar to other findings of habitat shifts in Melastomataceae [72] and Malvaceae [73]. It thus seems that both forest and savanna lineages in the *Berlinia* clade have been able to exploit additional characteristics of the main biomes that we were not able to capture with our broad classification.

### **(c) Monotypic forest paleoendemic genera with lower net diversification rates**

Our analyses revealed four monotypic genera, as well as *Microberlinia* with two extant species that originated before the mid-Miocene (Fig. 3). These five genera are all restricted to the forest biome and most of them with limited distribution in tropical Africa (e.g., *Librevillea klaneii* is restricted to West tropical Africa, *Isobertlinia scheffleri* is endemic to Tanzania, and *Icuria dunensis* is restricted to Mozambique). Our BAMM analyses do not indicate that these lineages have experienced higher extinction rates than other lineages in the *Berlinia* clade. Overall, our results suggest that the *Berlinia* clade has experienced low extinction rates, including the paleoendemic lineages (mostly monotypic genera) we identified (Figs. S7-S9). However, these forest paleoendemic lineages seem to have lower net diversification rates than lineages that shifted to new habitats (Fig. S9). *Paramacrolobium coeruleum*, sister to the *Berlinia* clade, is also a monotypic genus native to the forest biome but with a much wider distribution than *Librevillea*, *Icuria*, *Microberlinia*, *Michelsonia*, and *Isobertlinia scheffleri*. The current distribution of *P. coeruleum* includes western, central, and eastern Africa, from Guinea in the west to Kenya and Tanzania in the east, and south to Angola [74]. It is thus plausible that the ancestral lineage of the *Berlinia* clade and *P. coeruleum* also inhabited the forest biome. We found that the diversification rates were lower before the early Miocene in the entire *Berlinia* clade (Fig. 4b), and that forest lineages (e.g., *Berlinia* and *Anthonotha*) and savanna lineages (e.g.,



*Brachystegia*) that originated after this period seem to have increased net diversification rates. Thus, our results support previous reports showing that the lowland rain forest region in tropical Africa is characterized by old lineages that have persisted in these stable environments [15]. Our analyses of this endemic African tropical tree lineage suggest an uneven distribution in the capacity of these genera to exploit the availability of biomes. Despite the long evolutionary history of the *Berlinia* clade and the availability of new habitats during the multiple instances of forest contraction and expansion since the Miocene [13], some genera have only persisted in the forest biome. This heterogeneity is also evident in the location in the phylogeny of species-rich lineages that exploited both biomes identified in this study (forest and savanna), as the overall extension of the forest contracted since the mid-Miocene [10]. Thus, our study identified lineages that are potentially more susceptible to reductions or modifications of the forest biome due to future climate change and land use alterations. For example, future land use and landcover (LULC) projections under the RCP 8.5 (“worst case scenario”) predicts a 5.6% area reduction of the Congolian rainforest in Central Africa due to cropland replacement, and an 8% due to climate change [7]. These fast modifications of the forest biome are more likely to affect these paleo-endemic taxa. Under these predictions, these forest paleoendemic genera might be more susceptible to habitat modifications. Additional fine-scale analyses at the population level within each species will provide a more detailed picture of the genetic diversity, demographic history, and evolutionary potential for these paleo-endemic taxa.

**Ethics.** This work did not require ethical approval from a human subject or animal welfare committee.

**Data accessibility.** SRA sequences are deposited in the NCBI Bioproject PRJNA472454.

**Supplementary material is available online.** Bait sequences and alignments have been deposited in Dryad Data Repository at XXX.

**Declaration of AI use.** We have not used AI-assisted technologies in creating this article.

**Author Contributions:** D.I.O.: Conceptualization; Data curation; Formal analysis; Writing – original draft; Writing – review & editing. A.P.G.: Formal analysis; Visualization; Writing – review & editing. G.D.: Formal analysis; Visualization; Writing – review & editing. S.C.: Conceptualization; Data curation; Formal analysis; Writing – review & editing. A.D.P.: Investigation; Methodology; Writing – review & editing. T.M.: Methodology; Writing – review & editing. S.V.A.: Resources; Writing – review & editing. A.B.: Resources; Writing – review & editing. F.F.: Funding acquisition; Supervision; M.E.: Resources; Writing – review & editing. O.H.: Conceptualization; Funding acquisition; Investigation; Project administration; Resources; Writing – review & editing.



**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** M.E. was funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 659152 (GLDAFRICA). The laboratory work was supported by the Fonds de la Recherche Scientifique-FNRS (F.R.S.-FNRS) under Grants No. T.0163.13, J.0292.17F, and T.0075.23F, and by the Belgian Federal Science Policy Office (BELSPO) through project AFRIFORD from the BRAIN program. APG was funded by the Special Research Fund Ghent University—BOF postdoctoral fellowship BOF20/PDO/003 and the EOS-CANOPI project (O.0026.22 grant). SVA was funded by the Belgian American Educational Foundation (BAEF) and the Wiener-Anspach Foundation (FWA).

**Acknowledgments.** We thank the staff at the Meise and Kew herbaria for their support during the visits and collection of material.

## References

1. Ringelberg JJ, Zimmermann NE, Weeks A, Lavin M, Hughes CE. 2020 Biomes as evolutionary arenas: Convergence and conservatism in the trans-continental succulent biome. *Glob. Ecol. Biogeogr.* **29**, 1100–1113. (doi:10.1111/geb.13089)
2. Segovia RA, Pennington RT, Baker TR, de Souza FC, Neves DM, Davis CC, Armesto JJ, Olivera-Filho AT, Dexter KG. 2020 Freezing and water availability structure the evolutionary diversity of trees across the Americas. *Sci. Adv.* **6**. (doi:10.1126/sciadv.aaz5373)
3. Crisp MD *et al.* 2009 Phylogenetic biome conservatism on a global scale. *Nature* **458**, 754–756. (doi:10.1038/nature07764)
4. Cássia-Silva C, Freitas CG, Alves DMCC, Bacon CD, Collevatti RG. 2019 Niche conservatism drives a global discrepancy in palm species richness between seasonally dry and moist habitats. *Glob. Ecol. Biogeogr.* **28**, 814–825. (doi:10.1111/geb.12895)
5. Hardy OJ *et al.* 2013 Comparative phylogeography of African rain forest trees: A review of genetic signatures of vegetation history in the Guineo-Congolian region. *Comptes Rendus - Geosci.* **345**, 284–296.
6. Donoghue M, Edwards E. 2014 Biome shifts and niche evolution in plants. *Annu. Rev. Ecol. Syst.* **45**, 774–787.
7. Robertson RD, De Pinto A, Cenacchi N. 2023 Assessing the future global distribution of land ecosystems as determined by climate change and cropland incursion. *Clim. Change*

**176**, 1–22. (doi:10.1007/s10584-023-03584-3)

8. Plana V. 2004 Mechanisms and tempo of evolution in the African Guineo – Congolian rainforest. *Philos. Trans. R. Soc. B Biol. Sci.* **359**, 1585–1594.
9. Schley RJ, Qin M, Vatanparast M, Malakasi P, de la Estrella M, Lewis GP, Klitgård BB. 2022 Pantropical diversification of padauk trees and relatives was influenced by biome-switching and long-distance dispersal. *J. Biogeogr.* **49**, 391–404. (doi:10.1111/jbi.14310)
10. Kissling WD, Eiserhardt WL, Baker WJ, Borchsenius F, Couvreur TLP, Balslev H, Svenning J-C. 2012 Cenozoic imprints on the phylogenetic structure of palm species assemblages worldwide. *Proc. Natl. Acad. Sci.* **109**, 7379–7384.
11. Hagen O, Skeels A, Onstein RE, Jetz W. 2021 Earth history events shaped the evolution of uneven biodiversity across tropical moist forests. *Proc. Natl. Acad. Sci.* **118**, e2026347118. (doi:10.1073/pnas.2026347118/-/DCSupplemental.y)
12. Gorel AP, Hardy OJ, Dauby G, Dexter KG, Segovia RA, Steppe K, Fayolle A. 2022 Climatic niche lability but growth form conservatism in the African woody flora. *Ecol. Lett.* **25**, 1164–1176.
13. Couvreur T *et al.* 2020 Tectonics, climate and the diversification of the tropical African florán and fauna. *Biol. Rev.*
14. Piñeiro R, Hardy OJ, Tovar C, Gopalakrishnan S, Vieira FG, Gilbert TMP. 2021 Contrasting genetic signal of recolonization after rainforest fragmentation in African trees with different dispersal abilities. *Proc. Natl. Acad. Sci. U. S. A.* **118**, 1–9. (doi:10.1073/pnas.2013979118)
15. Dagallier LPMJ *et al.* 2020 Cradles and museums of generic plant diversity across tropical Africa. *New Phytol.* **225**, 2196–2213.
16. Cai L *et al.* 2023 Climatic stability and geological history shape global centers of neo- and paleoendemism in seed plants. *Proc. Natl. Acad. Sci.* **120**, e2300981120. (doi:10.1073/pnas)
17. de la Estrella M, Cervantes S, Janssens SB, Forest F, Hardy OJ, Ojeda DI. 2020 The impact of rainforest area reduction in the Guineo-Congolian region on the tempo of diversification and habitat shifts in the *Berlinia* clade (Leguminosae). *J. Biogeogr.*
18. de la Estrella M, Forest F, Klitgård B, Lewis G, Mackinder BA, de Queiroz LP, Wieringa

- JJ, Bruneau A. 2018 A new phylogeny-based tribal classification of subfamily Detarioideae, an early branching clade of florally diverse tropical arborescent legumes. *Sci. Rep.* **8**, 6884.
19. Ojeda DI *et al.* 2019 Phylogenomic analyses reveal an exceptionally high number of evolutionary shifts in a florally diverse clade of African legumes. *Mol. Phylogenet. Evol.* **137**, 156–167.
20. de la Estrella M, Forest F, Wieringa JJ, Fougère-Danezan M, Bruneau A. 2017 Insights on the evolutionary origin of Detarioideae, a clade of ecologically dominant tropical African trees. *New Phytol.* **214**, 1722–1735.
21. Doyle JJ, Doyle JL. 1987 A rapid DNA isolation procedure for small amounts of fresh leaf tissue. *Phytochem. Bull.* **19**, 11–15.
22. Mariac C *et al.* 2014 Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol. Ecol. Resour.* **14**, 1103–1113.
23. Yang Y, Smith SA. 2014 Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: Improving accuracy and matrix occupancy for phylogenomics. *Mol. Biol. Evol.* **31**, 3081–3092.
24. Bankevich A *et al.* 2012 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477.
25. Li W, Godzik A. 2006 Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659.
26. Van Dongen S. 2000 A cluster algorithm for graphs. University of Utrecht.
27. Katoh M, Kuma M. 2002 MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066.
28. Stamatakis A. 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.
29. Mirarab S, Nguyen N, Guo S, Wang L-S, Kim J, Warnow T. 2015 PASTA: Ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. *J. Comput. Biol.* **22**, 377–386.
30. Price MN, Dehal PS, Arkin AP. 2009 Fasttree: Computing large minimum evolution trees

with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650.

31. Johnson MG, Gardner EM, Liu Y, Medina R, Goffinet B, Shaw AJ, Zerega NJC, Wickett NJ. 2016 HybPiper: Extracting coding sequence and introns for phylogenetics from High-throughput sequencing reads using target enrichment. *Appl. Plant Sci.* **4**, 1600016.

32. Vatanparast M, Powell A, Doyle J, Egan A. 2018 Targeting legume loci: a comparison of three methods for target enrichment bait design in Leguminosae phylogenomics. *Appl. Plant Sci.* **6**, e1036.

33. Larsson A. 2014 AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–3278.

34. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. 2015 IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274.

35. Huelsenbeck JP, Ronquist F. 2001 MRBAYES: Bayesian inference of phylogeny. *Bioinformatics* **17**, 754–755.

36. Ronquist F *et al.* 2012 MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542.

37. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018 Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **0**, 1–5.

38. Mirarab S, Warnow T. 2015 ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. In *Bioinformatics*, pp. i44-152.

39. Dauby G *et al.* 2016 RAINBIO: A mega-database of tropical African vascular plants distributions. *PhytoKeys* **74**, 1–18. (doi:10.3897/phytokeys.74.9723)

40. Breteler FJ. 2010 Revision of the African genus *Anthonotha* (Leguminosae, Caesalpinioideae). *Plant Ecol. Evol.* **143**, 70–99.

41. de la Estrella M, Wieringa J, Breteler F, Ojeda D. 2019 Re-evaluation of the genus *Englerodendron* (Leguminosae-Detarioideae), including *Isomacrolobium* and *Pseudomacrolobium*. *Aust. Syst. Bot.* **32**, 564–571.

42. Wieringa J. 1999 *Monopetalanthus* exit.: a systematic study of *Aphanocalyx*, *Bikinia*, *Icuria*, *Michelsonia* and *Tetraberlinia* (Leguminosae, Casalpinioideae). van de Wageningen Universitet.

- 571 43. Lubkea RA, Dolda AD, Brinkb E, Avisc AM, Wieringa JJ. 2018 A new species of tree,  
572 Icuria dunensis Icurri, of undescribed coastal forest in north eastern Mozambique. *South*  
573 *African J. Bot.* **115**, 292–293.
- 574 44. Aubréville A. 1968 *Légumineuses - Caesalpinioïdées (Leguminosae - Caesalpinioideae).*  
575 *Flore du Gabon. Volume 15.* Paris, France.
- 576 45. Wilks C, Issembé Y. 2000 *Les arbres de la Guinée Equatoriale: Guide pratique*  
577 *d'identification: région continentale.* Projet CUREF, Bata, Guinée Equatoriale.
- 578 46. Maddison WP, Maddison DR. 2015 Mesquite: a modular system for evolutionary analysis.
- 579 47. Borges R, Machado JP, Gomes C, Rocha AP, Antunes A. 2019 Measuring phylogenetic  
580 signal between categorical traits and phylogenies. *Bioinformatics* **35**, 1862–1869.  
581 (doi:10.1093/bioinformatics/bty800)
- 582 48. Paradis E, Claude J, Strimmer K. 2004 APE: analyses of phylogenetics and evolution in R  
583 language. *Bioinformatics* **20**, 289–290.
- 584 49. Fritz SA, Purvis A. 2010 Selectivity in mammalian extinction risk and threat types: A new  
585 measure of phylogenetic signal strength in binary traits. *Conserv. Biol.* **24**, 1042–1051.  
586 (doi:10.1111/j.1523-1739.2010.01455.x)
- 587 50. Romero IC, Kong S, Fowlkes CC, Jaramillo C, Urban MA, Oboh-Ikuenobe F, D'Apolito C,  
588 Punyasena SW. 2020 Improving the taxonomy of fossil pollen using convolutional neural  
589 networks and superresolution microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28496–  
590 28505. (doi:10.1073/pnas.2007324117)
- 591 51. Eisawi A, Schrank E. 2008 Upper Cretaceous to Neogene palynology of the Melut Basin,  
592 Southeast Sudan. *Palynology* **32**, 101–129.
- 593 52. Yemane K, Bonnefille R. 1987 Pollen and clay mineral assemblages of a late Miocene  
594 lacustrine sequence from the north- western Ethiopian highlands. *Palaeogeogr*  
595 *Palaeoclim. Palaeoecol* **60**, 123–141.
- 596 53. Vincens A, Tiercelin JJ, Buchet G. 2006 New Oligocene-early Miocene microflora from the  
597 southwestern Turkana Basin. Palaeoenvironmental implications in the northern Kenya  
598 Rift. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **239**, 470–486.  
599 (doi:10.1016/j.palaeo.2006.02.007)
- 600 54. Herendeen PS, Jacobs BF. 2000 Fossil legumes from the Middle Eocene (46.0 Ma)

Mahenge flora of Singida, Tanzania. *Am. J. Bot.* **87**, 1358–1366.

55. Bonnefille R. 2010 Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. *Glob. Planet. Change* **72**, 390–411.

56. Pan A, Jacobs BF, Bush RT, de la Estrella M, Grímsson F, Herendeen PS, van der Burgt XM, Currano ED. 2023 First evidence of a monodominant (Englerodendron, Amherstieae, Detarioideae, Leguminosae) tropical moist forest from the early Miocene (21.73 Ma) of Ethiopia. *PLoS One* **18**, 1–26. (doi:10.1371/journal.pone.0279491)

57. Pan AD, Jacobs BF, Currano ED, de la Estrella M, Herendeen PS, van der Burgt XM. 2023 A fossil Anthonotha (Leguminosae: Detarioideae: Amherstieae) species from the early Miocene (21.73 Ma) of Ethiopia. *Int. J. Plant Sci.* **184**, 541–548. (doi:10.1086/725429)

58. Drummond AJ, Rambaut A. 2007 "BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214.

59. Marshall CR. 2019 Using the fossil record to evaluate timetree timescales. *Front. Genet.* **10**, 1–20.

60. Morlon H, Lewitus E, Condamine FL, Manceau M, Clavel J, Drury J. 2016 RPANDA: An R package for macroevolutionary analyses on phylogenetic trees. *Methods Ecol. Evol.* **7**, 589–597.

61. Rabosky DL. 2014 Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS One* **9**.

62. Rabosky DL, Santini F, Eastman J, Smith SA, Sidlauskas B, Chang J, Alfaro ME. 2013 Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nat. Commun.* **4**, 1–8.

63. Team RC. 2018 R: A language and environment for computing.

64. Revell LJ. 2012 phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223.

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

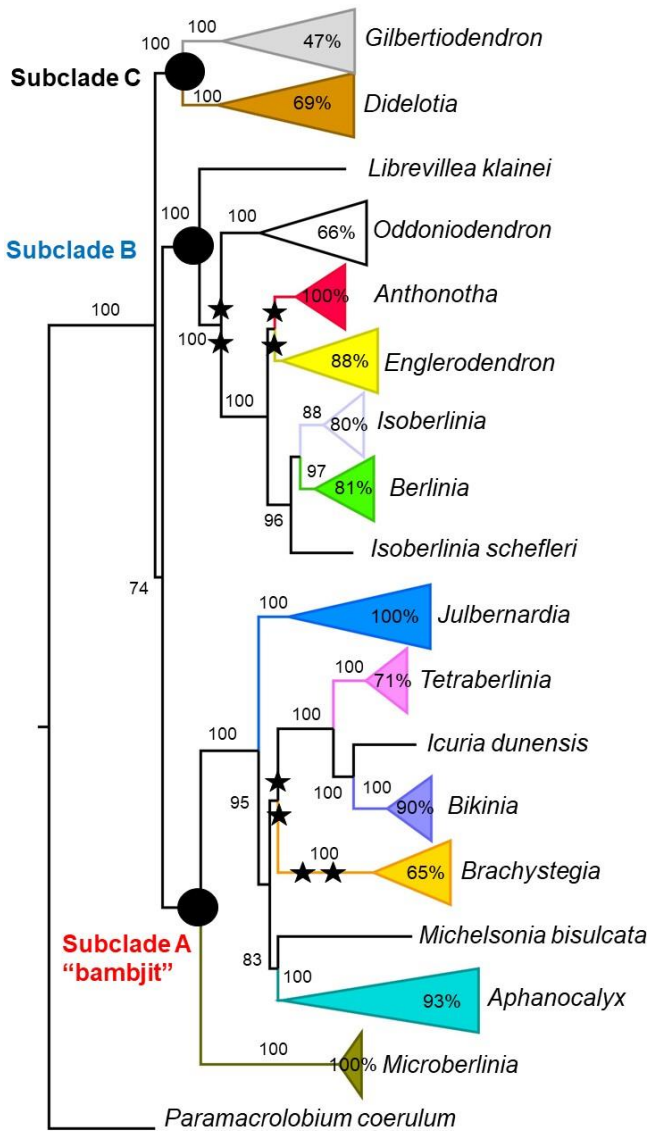
66. Fitzjohn RG, Maddison WP, Otto SP. 2009 Estimating trait-dependent speciation and extinction rates from incompletely resolved phylogenies. *Syst. Biol.* **58**, 595–611.



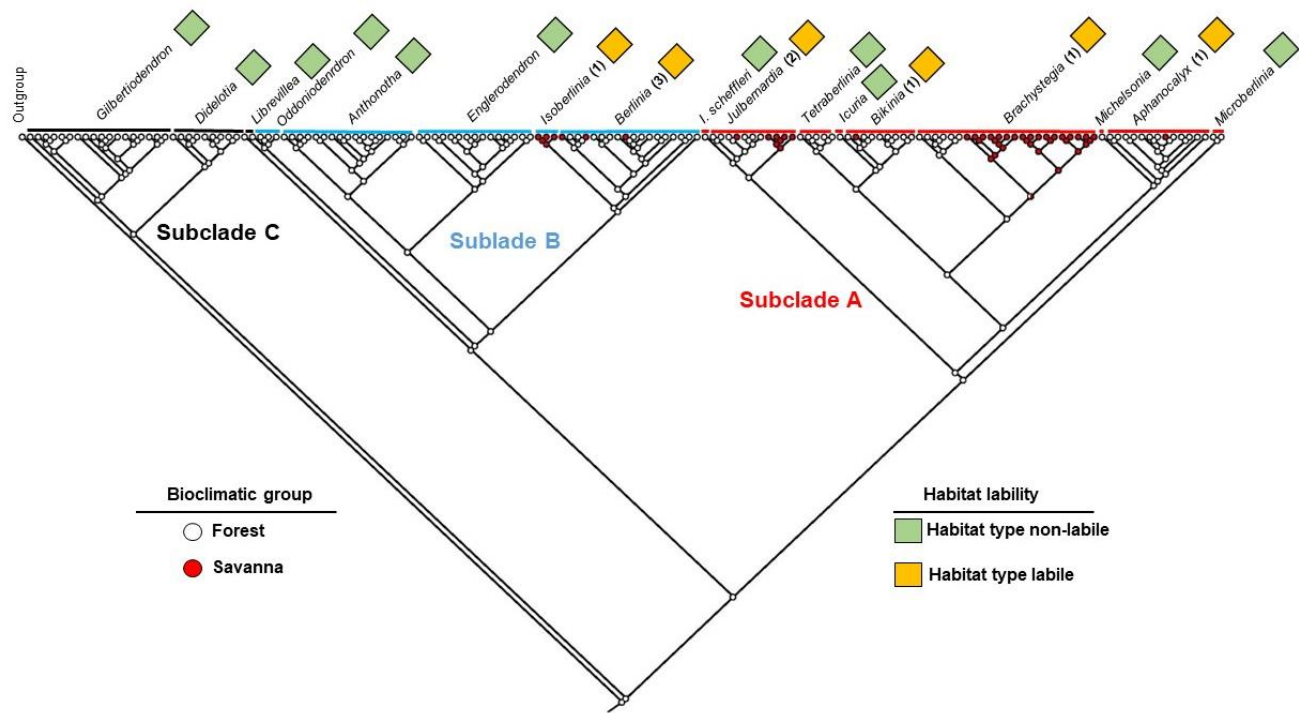
- 631 67. Silvertown J, McConway K, Gowing D, Dodd M, Fay MF, Joseph JA, Dolphin K. 2006  
632 Absence of phylogenetic signal in the niche structure of meadow plant communities. *Proc.*  
633 *R. Soc. B Biol. Sci.* **273**, 39–44. (doi:10.1098/rspb.2005.3288)
- 634 68. Losos JB, Leal M, Glor RE, De Queiroz K, Hertz PE, Rodríguez Schettino L, Chamizo  
635 Lara A, Jackman TR, Larson A. 2003 Niche lability in the evolution of a Caribbean lizard  
636 community. *Nature* **424**, 542–545. (doi:10.1038/nature01814)
- 637 69. Koenen E, de Vos JM, Atchison GW, Simon MF, Schrire BD, de Souza ER, de Queiroz  
638 LP, Hughes CE. 2013 Exploring the tempo of species diversification in legumes. *South*  
639 *African J. Bot.* **89**, 19–30.
- 640 70. Bruneau A, Mercure M, Lewis GP, Herendeen PS. 2008 Phylogenetic patterns and  
641 diversification in the caesalpinoid legumes. *Botany* **86**, 697–718.
- 642 71. Koenen EJM *et al.* 2020 Large-scale genomic sequence data resolve the deepest  
643 divergences in the legume phylogeny and support a near-simultaneous evolutionary origin  
644 of all six subfamilies. *New Phytol.* **225**. (doi:10.1111/nph.16290)
- 645 72. Veranso-Libalah MC, Kadereit G, Stone RD, Couvreur TLP. 2018 Multiple shifts to open  
646 habitats in Melastomateae (Melastomataceae) congruent with the increase of African  
647 Neogene climatic aridity. *J. Biogeogr.* **45**, 1420–1431.
- 648 73. Zizka A, Carvalho-Sobrinho JG, Pennington RT, Queiroz LP, Alcantara S, Baum DA,  
649 Bacon CD, Antonelli A. 2020 Transitions between biomes are common and directional in  
650 Bombacoideae (Malvaceae). *J. Biogeogr.* **47**, 1310–1321. (doi:10.1111/jbi.13815)
- 651 74. Science K. 2022 *Paramacrolobium coeruleum* (Taub.) J.Léonard. *Plants of the World*  
652 *Online*.



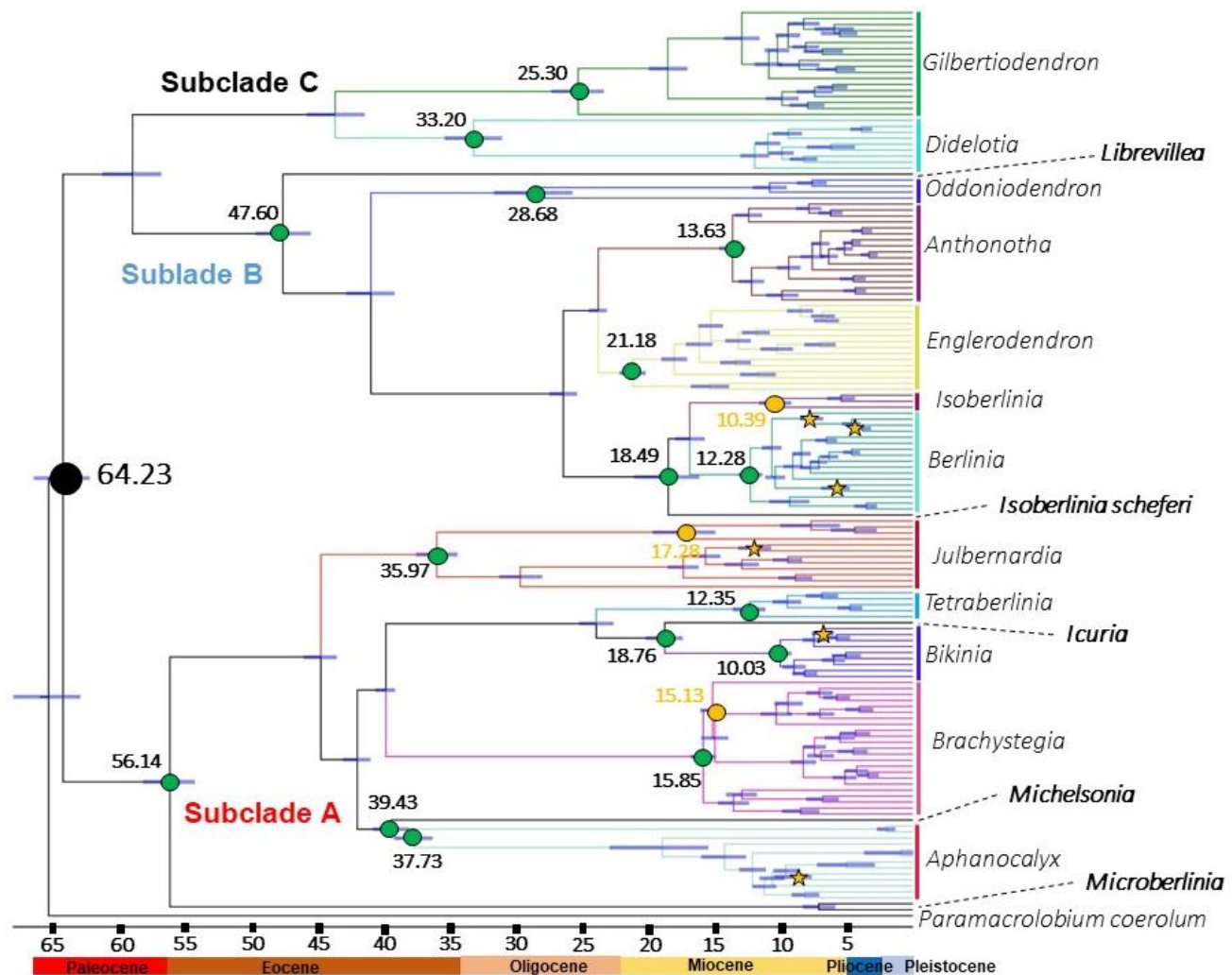
**Figures**



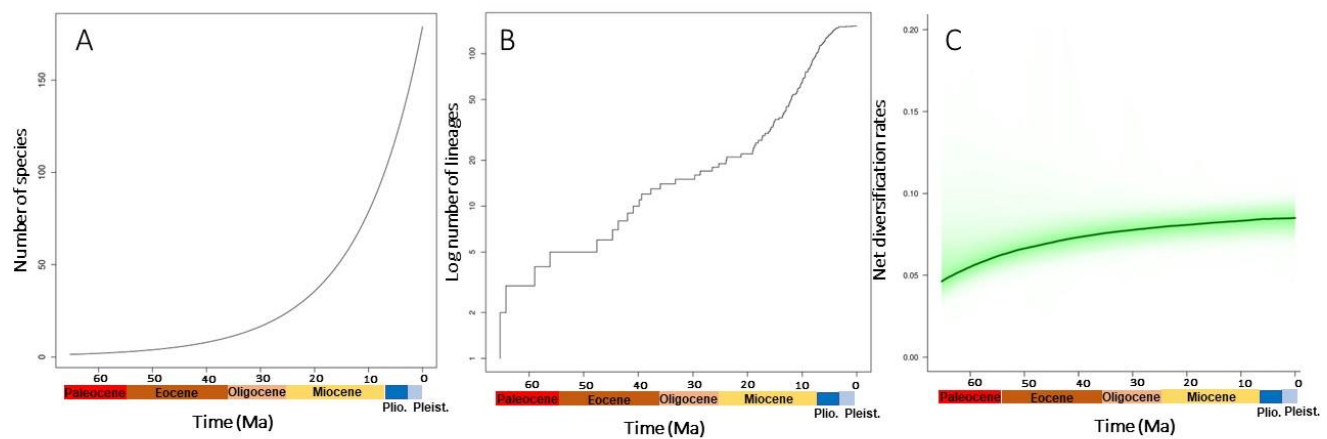
**Figure 1.** Phylogenetic relationships within the Berlinia clade based on maximum likelihood as implemented in RAxML and using the concatenated data set. The major subclades recovered within the group are marked with black circles. Percentage values of each genus represent the portion of species included. Values next to the branches represent bootstrap values. The stars indicate the locations of the eight fossil calibration points.



**Figure 2.** Distribution of the two bioclimatic groups and biome lability within the Berlinia clade at the genus level. The color of tips and nodes indicate the (ancestral) bioclimatic group. Green squares represent genera that remained in the forest biome, orange squares represent genera where shifts to the savanna biome occurred. The number in parenthesis indicates the number of transitions within each genus.



**Figure 3.** Time calibrated phylogeny of the Berlinia clade based on BEAST. The blue bars indicate 95% posterior distributions. The green circles indicate the time of the origin (crown age) of the main forest lineages (genera) and the orange circles the time of the origin of savanna lineages that diversified. Orange stars indicate shifts from forest to savanna habitats without further diversification in the savanna biome. Genera in bold indicate the paleo-endemic lineages (most monotypic).



**Figure 4.** Diversification rates of the Berlinia clade based on (A) RPANDA, (B) LTT plots and (C) BAMM

## Tables

**Table 1.** Results of the models tested with RPANDA in the Berlinia clade. Columns in bold indicate the best-fitted model for the data. LH, log-likelihood; AICc, corrected Akaike Information Criterion;  $\Delta AICc$ , change in AICc compared with the model with the lowest AICc; Models' parameters:  $\lambda_{con}$ , speciation constant;  $\lambda_{exp}$ , speciation exponential;  $\lambda_{lin}$ , speciation linear;  $\mu_0$ , no extinction;  $\mu_{con}$ , constant extinction.

Models	LH	AICc	$\Delta AICc$
<b>Time-dependent</b>			
$\lambda_{con}$ and $\mu_0$	<b>-326.45</b>	<b>654.45</b>	0.00
$\lambda_{exp}$ and $\mu_0$	-326.18	656.45	2.00
$\lambda_{lin}$ and $\mu_0$	-326.12	656.33	1.88
$\lambda_{con}$ and $\mu_{con}$	-326.54	655.11	0.66
$\lambda_{exp}$ and $\mu_{con}$	-326.18	658.53	4.08
$\lambda_{lin}$ and $\mu_{con}$	-326.25	656.58	2.13

## Supporting information

### Figures

**Figure S1.** Phylogenetic relationships in the Berlinia clade using RAxML using the individual genes separately. Values next to branches indicate 500 bootstrap values.

**Figure S2.** Phylogenetic relationships in the Berlinia clade using RAxML using the concatenated matrix of all genes. Values next to branches indicate 500 bootstrap values.

**Figure S3.** Phylogenetic relationships in the Berlinia clade using IQ-TREE using the concatenated matrix of all genes. Values next to branches indicate 1000 fast bootstrap values.

**Figure S4.** Phylogenetic relationships in the Berlinia clade using ASTRAL. Values next to branches represent posterior probabilities.

**Figure S5.** Phylogenetic relationships in the Berlinia clade using Mr. Bayes using the concatenated matrix of all genes. Values next to branches indicate Bayesian values.

**Figure S6.** Extinction, speciation, and net diversification rates of the Berlinia clade as inferred using BAMM.

**Figure S7.** Extinction rates of the Berlinia clade as inferred using BAMM.

**Figure S8.** Speciation rates of the Berlinia clade as inferred using BAMM.

**Figure S9.** Net diversification rates of the Berlinia clade as inferred using BAMM.

**Figure S10.** BiSSE analyses of the Berlinia clade.

795 **Tables**

796 **Table S1.** List of specimens used in this study.

797 **Table S2.** Statistics of the target regions captured in the Berlinia clade using the Detaroideae  
798 bait.

799 **Table S3.** Biome classifications of the Berlinia clade.

800 **Table S4.** Age estimates of the genera within the Berlinia clade.

801 **Table S5.** Fossils used in the calibration of the phylogenetic tree in the Berlinia clade.

802 **Table S6.** Strategy used to estimate the lambda distribution values for the dating analysis in the  
803 Berlinia clade.