Jian-Jun Jin ORCID iD: 0000-0002-7557-9198

Robin van Velzen ORCID iD: 0000-0002-6444-7608

Ting-Shuang Yi ORCID iD: 0000-0001-7093-9564

Research Article

Born migrators: historical biogeography of the cosmopolitan family Cannabaceae

Running title: **Biogeography of Cannabaceae**

Jian-Jun Jin[1,2,3], Mei-Qing Yang[1,4], Peter W. Fritsch[5], Robin van Velzen[6], De-Zhu Li[1,2*], and Ting-Shuang Yi[1,2*]

[1]CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[2]Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China

[3]Kunming College of Life Sciences, University of Chinese Academy of Sciences, Kunming 650201, China

[4]Baotou Medical College, Baotou 014040, Inner Mongolia, China

[5]Botanical Research Institute of Texas, Fort Worth, Texas 76107-3400, USA

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/jse.12552.

[6]Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

*Authors for correspondence. De-Zhu Li. E-mail: DZL@mail.kib.ac.cn; Ting-Shuang Yi. E-mail: tingshuangyi@mail.kib.ac.cn

Received 27 March 2019; Accepted 15 November 2019

Abstract

Dispersal scenarios have been favored over tectonic vicariance as an explanation for disjunct distributions in many plant taxa during last two decades. However, this argument has been insufficiently addressed in cosmopolitan groups showing disjunct patterns in both the temperate and tropical regions. In this study, we used the Cannabaceae, an angiosperm family distributed in tropical and temperate regions of both the New World and the Old World, to explore the role of dispersal in shaping disjunct patterns and species diversification of cosmopolitan plants. We reconstructed the phylogenetic relationships of all 10 genera and 75 species of Cannabaceae (ca. 64.1% of recognized species) based on eight DNA regions. Based on fossil calibrations, we estimated the divergence times and net diversification rates. We further inferred the ancestral geographical ranges with several models and compared the fitness of different models. The Cannabaceae and most genera were strongly supported as monophyletic except for the *Parasponia* being embedded within the *Trema*. The *Celtis* were resolved into two strongly supported clades primarily corresponding to temperate and tropical regions. We inferred that the Cannabaceae originated at ca. 93 Ma, and that subsequent rampant and widespread dispersals shaped the intercontinentally disjunct distribution of the Cannabaceae. Dispersal coincides with adaptation to drier and colder climate in the North Hemisphere, or humid and warm climate in the tropical regions, followed by rapid species

diversification. This study advances our understanding as to the formation of distribution patterns and species diversification of a plant family with tropical to temperate disjunct distributions.

Key words: ancestral geographical range analysis, Cannabaceae, dispersal, molecular dating, Northern Hemisphere, phylogeny.

1 Introduction

The concept of vicariance has dominated historical biogeography for decades, and has been widely applied to explain intercontinental distributions, especially for tropical groups disjunctly distributed among the landmasses; supposedly resulting from the sequential breakup of the supercontinent Gondwana since 180 Ma (Lomolino, 2010). However, molecular phylogenetic and biogeographical studies in the last two decades often favor dispersal scenarios over tectonic vicariance as explanation for disjunct distributions in many plant taxa, mainly because divergence times are too recent to be reasonably compatible with the estimated times of continental breakup inferred from geological data (Sanmartín & Ronquist, 2004). These studies suggest that dispersal has played a more significant role in shaping modern transoceanic angiosperm distributions than vicariance, marking a major shift in our view of the historical biogeography of these regions (Sanmartín & Ronquist, 2004; Renner et al., 2010; Bartish et al., 2011).

One widely accepted explanation for disjunction across the Northern Hemisphere is the existence of a widespread “boreotropical” flora in the Eocene and “mixed mesophytic forest” later in the Miocene, followed by extinctions and southward migrations due to climate changes during the Neogene and Quaternary (Wolfe, 1975). Recent molecular phylogenetic and biogeographical studies of extant

lineages have revealed that dispersals, especially via migration across the North Atlantic Land Bridge (NALB) and Bering Land Bridge (BLB), play a major role in disjunctions of plant genera in the North Hemisphere (Donoghue et al., 2001; Xiang & Soltis, 2001).

To better understand the role of dispersal in shaping the overall disjunct patterns in both the North Hemisphere and tropical regions, inferring the historical biogeography of cosmopolitan groups exhibiting both types of disjunctions is essential. In this study, the angiosperm family Cannabaceae is employed to address diversification history in two regions. The Cannabaceae are a good experimental group for this issue because they exhibit a wide distribution in tropical to temperate regions of the world (Yang et al., 2013). The family contains 10 genera and about 117 species (ad hoc species list described in M&M). Several genera exhibit widespread distributions: The *Celtis* L. are widely distributed in global tropical to temperate regions; The *Trema* Lour. show a pantropical disjunct distribution; The *Aphananthe* Planch. are disjunctly distributed among North America, East, South and Southeast Asia, Australia and Madagascar; and the other genera have more local distributions. Prior studies of the family have progressed in resolving phylogenetic relationships, inferring the evolution of morphological characters, and updating classification (van Velzen et al., 2006; Yang et al., 2013). However, only a few gene regions or a limited number of taxa were employed in phylogenetic analyses of this family (e.g., Yang et al., 2013; Zhang et al. 2018a). Furthermore, the origin and formation of the family's current distribution pattern have yet to be addressed.

Here we use DNA sequence data from three nuclear and five plastid DNA regions from 64% of the species of the Cannabaceae and employ fossil-calibrated

divergence time estimates and ancestral geographical range analyses to infer the biogeographical history. We use the data to examine lineage diversification in this family in the context of climatic fluctuations among different geological time intervals.

2 Material and Methods

2.1 Taxon sampling, DNA sequencing and alignment

Because previous studies have not clarified the number and circumscription of species in the Cannabaceae, we developed a working Cannabaceae species list (Table S1) compiled information from The Plant List Version 1.1 (available from www.theplantlist.org, accessed 1st January 2017), local floras, and previous studies (van Velzen et al., 2006). We sampled 75 species (64.1% of total extant species) representing all ten genera of the Cannabaceae (Tables S1, S2). We included multiple species samples for genera having wide distributions. We sampled 14 *Trema* species (93% of the total) throughout its geographical distribution, and all five *Aphananthe* species. For the largest and most widely distributed genus *Celtis*, we sampled 39 species (53.4%) to encompass the range of its biogeographical distribution and morphological variation. Because the Moraceae and Urticaceae were supported to be sister to the relatives of the Cannabaceae in previous studies (Zhang et al., 2011; Sytsma et al., 2002), *Broussonetia papyrifera* (L.) L'Hér. ex Vent. (Moraceae) and *Debregeasia saeneb* (Forsskal) Hepper & Wood (Urticaceae) were employed as the outgroup.

Five plastid (*atpB-rbcL*, *rbcL*, *rps16* intron, *trnH-psbA* and *trnL-trnF*) and three nuclear DNA regions (18S, ETS and ITS) were employed (Table S3). Sequences were initially assembled with SEQUENCHER v5.0 (GeneCodes Corporation, Ann

Arbor, Michigan, USA) and aligned with MAFFT v7.427 (Katoh & Standley, 2013). The 344 newly generated sequences in this study were deposited in GenBank (accession numbers: MN364350-MN363494, MN381740-MN381804, MN395050-MN395283), and 121 other sequences from GenBank were also used to build the final data matrix (see Table S2). The alignments, the phylogenetic trees and the dated tree have been deposited in FigShare (URL: http://dx.doi.org/10.6084/m9.figshare.10279577).

2.2 Phylogenetic analysis

Phylogenetic relationships were inferred with Bayesian inference (BI) as implemented in MRBAYES v3.2.7a (Ronquist & Huelsenbeck, 2003) and maximum likelihood (ML) as implemented in RAXML V8.2.4 (Stamatakis et al., 2008). The best-fit model of sequence evolution for each region was estimated with JMODELTEST v2.1.10 (Darriba et al., 2012) and chosen by the corrected Akaike information criterion (AICc) (see Table S4). Nuclear and chloroplast trees did not contain any well supported but conflicting clades; the plastid and nuclear data sets were thus combined in analyses.

A partitioned (Table S4) ML analysis was performed under the GTR + gamma distribution model with 1000 replicates. Maximum likelihood bootstrap (MLBS) support ≥ 70% was considered to be strongly support (Hillis & Bull, 1993). A partitioned Bayesian analysis was performed with each partition set to its optimal model. Two independent runs each consisted of 5 000 000 generations, with one cold and three incrementally heated Markov chain Monte Carlo (MCMC) chains. Trees were sampled every 1000 generations. The first 20% of the posterior trees was discarded as burn-in, as determined with TRACER v1.6 (Rambaut & Drummond,

2004). Each parameter for each run obtained a sufficient effective sample size (ESS > 500) and a caterpillar-shaped trace. All post-burn-in trees were used to construct majority-rule consensus trees. Posterior probability (PP) clade credibility values ≥ 0.95 were considered to be significant (Alfaro et al., 2003).

2.3 Molecular dating

Bayesian time estimation as implemented with BEAST v1.8.0 (Drummond & Rambaut, 2007) was performed under a lognormal relaxed molecular clock (Drummond & Rambaut, 2007) to estimate divergence times and their credibility intervals. All BEAST runs used a birth-death process tree prior with models of evolution optimized for each DNA region described above with six rate categories. Monte Carlo Markov chains (MCMC) were run for 300 000 000 generations, with parameters sampled every 1000 generations. The first 20% of the posterior trees was discarded as burn-in, as confirmed with TRACER v1.6 (Rambaut & Drummond, 2004). Each parameter obtained a sufficiently large ESS (> 500) and a caterpillar-shaped trace.

Five fossils were used for calibration with lognormal distribution constraints on stems, the standard deviation arbitrarily set to 0.5, the mean set to 1, and offset set to the estimated minimum age of each fossil. The pollen *Triorites minutipori* Muller from the Turonian of Sarawak is the oldest known *Celtis*-type pollen that is found in most living Cannabaceae genera (Muller, 1981), which is largely congruent with the previously estimated stem age of Cannabaceae (e.g., *Zhang* et al., 2018b). Furthermore, the fossil fruits of Urticaceae was dated to ca. 90 Ma (Collinson, 1989). A sister relationship between Cannabaceae and a clade comprise by Urticaceae and Moraceae (Sytsma et al., 2002; Zhang et al., 2011) implies a stem age of Cannabaceae might be earlier than 90 Ma. Therefore 89.8 Ma was conservatively

used as a constraint on the stem node of Cannabaceae. The stem node of the *Gironniera* Gaudich. was constrained to 66.0 Ma in accordance with the age of endocarp fossils of *Gironniera gonnensis* Knobl. & Mai from the upper Maastrichtian of the Eisleben Basin (Knobloch & Mai, 1986). The stem node of the *Celtis* was constrained to 56 Ma in accordance with the age of fruit and leaf fossils of *Celtis aspera* Manchester from the Palaeocene of North America and eastern Asia (Manchester et al., 2002). The stem node of the *Pteroceltis* Maxim. was constrained to 38.0 Ma in accordance with the age of fruit fossils of *Pteroceltis knowltonii* (Berry) Manchester from the Lutetian-Bartonian of Tennessee, USA (Manchester et al., 2009) and the fruit fossils of *Pteroceltis* sp. from the same age of Washington, USA (Pigg & Wehr, 2002). Finally, the stem node of the *Humulus* L. was constrained to 23.0 Ma in accordance with the age of fruit fossils of *Humulago reticulata* (Dorofeev) Doweld from the Oligocene of Antropovo, Russia (Collinson, 1989; Doweld, 2016), which is largely congruent with a *Cannabis-Humulus* divergence date of 27.8 Ma by McPartland (2018).

2.4 Diversification rate analyses

A semilogarithmic lineage-through-time (LTT) plot for the maximum clade credibility (MCC) tree of sampled Cannabaceae species was conducted with the R package ‘ape’ (Paradis et al., 2004). In the LTT plot, 1000 randomly chosen post-burn-in BEAST posterior trees were plotted to check the reliability of the LTT signal in the consensus chronogram.

Bayesian analysis of macroevolutionary mixtures (BAMM 2.5.0) (Rabosky et al., 2014a) was performed with the MCC tree for 5 000 000 generations, with trees sampled every 1000 generations. To achieve a comprehensive understanding of

diversification rates in the Cannabaceae, all extant species of the family without molecular data were added to their own genus or clade in the form of "sampling probability" in BAMM, except that the sampling probability of *Trema* is calculated as the sampling percentage of the *Trema* and the *Parasponia* Miq. species together since the *Parasponia* is embedded within the *Trema* (Table S1). The priors for the analysis were set with the setBAMMpriors function in the R package 'BAMMtools' (Rabosky et al., 2014b). A poissonRatePrior of 1.0 was used to allow shifts to occur on all branches (minCladeSizeForShift = 1). Convergence and effective sample sizes of parameters were assessed with 'coda' in R; log-likelihoods, numbers of processes, and evolutionary rate parameters all had sufficient ESS (> 500).

2.5 Biogeographical inference

Nine areas of endemism were defined for the Cannabaceae as based on species distributions, the previously defined geographic regions, and our emphasis on the large-scale intercontinental diversification of the family: Africa (A), Madagascar (B), tropical North America (including the Antilles; C), temperate and subtropical North America (D), South America (E), East Asia (F), South and Southeast Asia (G), Central-West Eurasia (H), and Australasia (including Australia, New Zealand, New Guinea, and neighboring islands in the Pacific Ocean (I). The following ancestral area pairs were excluded as based on extent species distributions and palaeogeographical information: A, B, G, or I combined with any one of C, D, or E.

Following the methods of Federman et al. (2015), we used the R package 'BioGeoBEARS v1.1.2' (Matzke, 2013) and the dispersal extinction cladogenesis model (DEC) (Ree & Smith, 2008) to fit three kinds of dispersal scenarios based on palaeogeographical and palaeoclimatic information: a terrestrial model (TR), a

terrestrial + short-distance marine model (SD), and a terrestrial + short-distance marine + long-distance marine dispersal model (LDD) (Table S5). We compared fitness of these three scenarios using AICc and AICc weights calculated from each model's log likelihood. The number of maximum areas for ancestral nodes was set at five, which is the maximum number of areas for extant observed existing species. We conducted all analyses on the maximum clade credibility (MCC) tree from the BEAST analysis with the outgroup pruned. To account for uncertainty in topology and branch lengths, we conducted a statistical test on 1,000 randomly chosen post-burn-in BEAST posterior trees with the DEC + SD model (the best-fit model).

3 Results

3.1 Phylogenetic relationships

The topologies produced by ML and BI analyses are largely congruent (Figs. S1, S2). Our phylogeny is largely congruent with that previously generated (Yang et al., 2013; Zhang et al. 2018a) but has better resolution or better support for some clades. The Cannabaceae form a clade (MLBS = 95%; BI = 1), with the family resolved into nine strongly supported clades corresponding to nine genera, except the *Trema* because the *Parasponia* are embedded within *Trema* with low support. The *Celtis*, the most species-rich genus, were resolved into two strongly supported clades corresponding to a tropical evergreen clade (MLBS = 94%; BI = 1) and a predominately temperate deciduous clade (MLBS = 100%; BI = 1) (Figs. 1, S1, S2).

3.2 Molecular dating

The stem age of the Cannabaceae was estimated to be 92.6 Ma (95% HPD: 90.5–96.4 Ma), and the crown age to be 87.4 Ma (95% HPD: 78.8–94.0 Ma) (Fig. 1; Table 1).

The divergence dates among the clades corresponding to genera (but with the *Parasponia* nested within the *Trema*) range from 87.4 Ma (95% HPD: 78.8–94.0 Ma) between the *Aphananthe* and other clades, to 25.4 Ma (95% HPD: 23.7–28.0 Ma) between *Cannabis* L. and *Humulus* (Fig. 1; Table 1). The split of the tropical and temperate clades of the *Celtis* was inferred as Oligocene to Miocene (23.6 Ma; 95% HPD: 16.4–32.4 Ma) (Fig. 1; Table 1).

3.3 Lineage diversification

The LTT plot of Cannabaceae reveals an initial relative high net diversification from ca. 87 Ma until ca. 54 Ma followed by a low rate of net diversification until ca. 20 Ma after which net diversification rate is high until the present (Fig. 2C). BAMM yielded two shifts to higher net diversification rates: one near the crown of the *Celtis* and another near the crown of the *Trema* and *Parasponia* (Figs. 2A, 2B). There is also indication of a shift within the temperate *Celtis* clade, but with relatively low probability (Fig. 2B). These shifts largely correspond to the increased net diversification rate since ca. 20 Ma observed in the Cannabaceae LTT plot (Fig. 2C).

3.4 Biogeographical patterns

The AICc score comparisons of different biogeographical models favored the DEC + SD model (Table 2). Inferences of ancestral areas were largely congruent across all analyses (Figs. 3, S3). Inferences of ancestral area conducted over the MCC tree and 1000 trees as based on the DEC + SD model were also largely congruent (Fig. 3; Tables 1, S6). Here we focus on inferences estimated from the MCC tree with the best-fit model (DEC + SD). East Asia was inferred to be the ancestral area of Cannabaceae (Figs. 3, S3). A series of dispersals were inferred to shape current disjunct distributions across global temperate to tropical regions (Table S7).

4 Discussion

4.1 Diversification through time

Our time estimates infer a late Cretaceous origin (ca. 92.6 Ma) of the Cannabaceae, and diversification of lineages corresponding to genera during the late Cretaceous to Palaeogene (Fig. 1; Table 1). East Asia was inferred to be most probable ancestral distribution region for the Cannabaceae (Figs. 3, S3; Tables 1, S6). The late Cretaceous to Palaeogene fossils of *Aphananthe*, *Celtis*, *Pteroceltis*, and *Gironniera* (Manchester, 1989; Manchester et al., 2002) imply a broad palaeogeographical distribution of the Cannabaceae in the Northern Hemisphere. Early diversification of the Cannabaceae coincides with the formation of the extensive boreotropical flora in the northern mid-latitudes during a warm phase peaking in the late Palaeocene-early Eocene Thermal Maximum (Wolfe, 1975; Zachos et al., 2001). The early branching lineages of the Cannabaceae might have experienced rapid diversification (Fig. 2C).

Climatic deterioration in the late Eocene and a drastic temperature drop at the Eocene-Oligocene boundary ca. 34 Ma resulted in the expansion of vegetation adapted to drier and colder climates in large parts of Eurasia and Northern America (Collinson, 1992; Wolfe, 1992). The boreotropical flora was disrupted, climates suitable for megathermal vegetation receded to equatorial regions, and boreotropical taxa had to adapt to drier and colder climates or retreated toward the equator to avoid extinction (Morley, 2011). The Cannabaceae successfully survived this dramatic climate change through both processes. After this transition, the diversification of the Cannabaceae lineages was somewhat asymmetric. Some lineages, such as the *Pteroceltis-Chaetachme* clade and the *Cannabis-Humulus* clade, maintained a relatively low diversification rate, whereas the *Celtis* and the *Trema* (including *Parasponia*) experienced increased net diversification rate (Fig. 2). The *Celtis* migrated and adapted to a wide variety of environments, such as drier and colder regions in the temperate and subtropical regions of the North Hemisphere, and moist and hot habitats in tropical regions, experiencing rapid speciation. The *Trema* are widely distributed in subtropical and tropical regions, even in some remote islands.

4.2 Disjunctions in the temperate and subtropical North Hemisphere

Fossil evidence indicates that the Cannabaceae were widely distributed across the middle-latitude Northern Hemisphere (Manchester, 1989; Manchester et al., 2002) by the early Cenozoic. The NALB functioned throughout the latest Cretaceous to early Eocene, and land connections may have supplied a migration route for thermophilous plant taxa between Europe and North America until the Oligocene (Tiffney & Manchester, 2001; Milne, 2006). The BLB functioned as a viable route for temperate taxa during the early to middle Cenozoic and cold-adapted plant taxa during the late Cenozoic between eastern Asia and North America (Tiffney & Manchester, 2001;

Milne, 2006). Both the NALB and BLB could be expected to have been important pathways for early migrations of the Cannabaceae across the Northern Hemisphere. Our inferred dispersals from East Asia to North America during the late Cretaceous (between nodes 73 and 9; all node ids throughout this paper correspond to Fig. 3 and Table 1), the Palaeogene (between nodes 72 and 71), and the dispersal from North America to East Asia during the Paleogene (between nodes 71 and 70; 58.5–23.6 Ma) could have been achieved via either the NALB or the BLB. The BLB can account for a much later East Asian and North American migration during the Neogene in *Humulus* (between nodes 12 and 11; 25.4–10.7 Ma) (Fig. 3; Table 1).

The rise of the Tibetan Plateau and the formation of the extreme dry climate of northern Central Asia during the late Miocene created a natural barrier for migration of many mesic taxa and played an important role in their separation into eastern and western parts (Harrison et al., 1992; An et al., 2001). Among the inferred dispersals between East Asia and Central-West Eurasia, the early dispersal such as the one between nodes 73 and 72 (71.0–61.8 Ma) were likely unaffected by this barrier because of its early age, and thus could migrate among the boreotropical forests. *Humulus lupulus* is separated by this barrier into eastern and western parts. The dispersal (25.4–10.7 Ma) of this species could have occurred either before the formation of this barrier or through long-distance dispersal. The eastern Chinese *Celtis chekiangensis* and its sister clade comprising the Central Asian/Mediterranean *Celtis australis* and *Celtis planchoniana* are widely separated geographically. The Pliocene split of these two clades is much later than the formation of the central Asian migration barrier; therefore, long-distance dispersal seems to be the only explanation for this disjunction.

4.3 Disjunctions between the Palaeotropics and Neotropics

Our inferred dates for two migrations (nodes 29 and 20, 14.4–4.92 Ma; 46 and 43; 17.0–1.79 Ma) from the Palaeotropics to the Neotropics (Fig. 3; Table 1) are too young to have been caused by the breakup of Gondwana. South America drifted westwards from Africa as the Atlantic opened between ca. 130–100 Ma (Blakey, 2008), Australia split from Antarctica ca. 35 Ma, and Antarctica split from South America ca. 35–30 Ma (McLoughlin, 2001). Both migrations from the Palaeotropics to the Neotropics have been inferred through the Northern Hemishere. However, the inferred dispersal dates are also much younger than the closure of the NALB route during the Oligocene. Species of the tropical clade of the *Celtis* and the *Trema* are evergreen trees or shrubs, and therefore the BLB route would not be a viable migration route for them during the Miocene. Therefore, transoceanic long-distance dispersals via birds or ocean currents likely account for these tropical disjunctions. Such long-distance dispersal has also been invoked for similar disjunctions in many other plant taxa (Michalak et al., 2010).

4.4 Disjunctions around the Indian Ocean Basin

The Cannabaceae have experienced at least six westward dispersals and one reverse dispersal around the Indian Ocean Basin (Fig. 3; Tables 1, S7). The Indian subcontinent separated from Madagascar during the middle-late Cretaceous (90–85 Ma), then collided with the Eurasian continent at ca. 50–25 Ma (Ali & Aitchison, 2008; van Hinsbergen et al., 2012). The “out-of-India” hypothesis thus cannot be invoked to explain the migration from Africa to South and Southeast Asia in the tropical clade of the *Celtis* (between nodes 46 and 44, 17.0–11.4 Ma) during the Miocene. Indeed, data from only a few plant taxa have been found to be consistent

with this hypothesis (Conti et al., 2002; Ducousso et al., 2004). The "boreotropical migration" hypothesis has been invoked to explain intercontinental exchange of Palaeotropical biota (Davis et al., 2002; Couvreur et al., 2011), and could explain two early Cenozoic dispersals from East Asia to Africa in *Chaetachme* (nodes 73 and *Chaetachme aristata*). Africa and Southwest Asia were connected due to the collision of the Afro-Arabian plate with the Iranian and Anatolian plates in the early to middle Miocene (Popov et al., 2004), which coincided with a warm phase peaking in the middle Miocene climatic optimum 17–15 Ma (Zachos et al., 2001). Many tropical forest taxa may have migrated between Africa and Asia via Arabia (Zhou et al., 2012). This also may be a viable route for two migrations in the tropical clade of the *Celtis* (nodes 70 and 46, 23.6–17.0 Ma; nodes 46–44, 17.0–11.4 Ma). This corridor has been inferred for the expansion of drought-adapted savanna and xerophytic shrubland vegetation after climates became distinctly cooler, drier and more seasonal from the late-middle Miocene onwards (Kürschner, 1998). Dispersals in the temperate clade of the *Celtis* (nodes 68 and 47, 8.41–2.63 Ma) and *Trema orientalis* could have occurred through this route. Dispersal of the tropical or subtropical *Aphananthe* from South and Southeast Asia to Madagascar occurred during the late Miocene (nodes 2 and 1; 9.72–6.52 Ma). Species of this genus are distributed exclusively in wet regions, and both their living and fossil species are absent from the Arabian region and Africa (Yang et al., 2017). Long-distance dispersal, widely invoked to explain disjunct patterns for many taxa in this region (Renner, 2004), may apply particularly well to *Aphananthe* species. The dispersal in the tropical clade of the *Celtis* from South and Southeast Asia to Africa (nodes 35 and 34, 8.11–4.39 Ma) may also have been achieved through overseas long-distance dispersal.

At least two dispersals in the *Celtis* tropical clade (nodes 34 and 33, 4.39–2.39 Ma; 38 and 37, 11.4–7.11 Ma) from Africa to Madagascar, and one reverse dispersal from Madagascar to Africa (between nodes 37 and *Celtis mildbraedii*, 7.11–0 Ma), occurred much later than the separation of Madagascar from continental Africa, which has been separated by the Mozambique Channel (ca. 430 km wide) from 120–116 Ma onwards (Rabinowitz et al., 1983; Ali & Aitchison, 2008). Thus, these dispersals cannot be explained by the breakup of Gondwana, and instead the migrations in the Cannabaceae between continental Africa and Madagascar were likely achieved by long-distance dispersal. Many plant taxa showing such a disjunct pattern have been explained by either transoceanic long-distance dispersal mediated by birds or wind from Africa to Madagascar (Renner, 2004; Clayton et al., 2009), or transport of vegetation mats from northeastern Mozambique and Tanzania to Madagascar by oceanic currents in the Eocene and Oligocene (extending into the early Miocene) (Ali & Huber, 2010).

Among the regions of East, South and Southeast Asia through Australasia, multiple migrations occurred during different geological times in the Cenozoic. At least five dispersals from East Asia to South and Southeast Asia and seven reverse dispersals were inferred since the late Cretaceous. Multiple migrations in the Cannabaceae occurred among South and Southeast Asia and Australasia since the Miocene. The emergence of the land masses east of Wallace's Line including Wallacea, Sulawesi, New Guinea and a series of volcanic islands along the Sunda Arc, the Banda Arc and the Halmahera Arc connecting these regions from the late Miocene onwards, offer a potential channel for dispersals among the two regions (Hall, 2009).

4.5 Migration among the Americas

The emergence of the Isthmus of Panama initiated the Great American Biotic Interchange and has been regarded as a defining event in the formation of Neotropical biodiversity of both animals and plants (Webb, 1991). However, most recent molecular biogeographical studies show that the isthmus may have been less of a requirement for plant migration (Cody et al., 2010). Among six intraspecific migrations between Tropical North America and South America in the Cannabaceae, four (*Celtis iguanaea*, *Lozanella enantiophylla*, *Trema integerrima* and *Trema micrantha*) are distributed continuously across tropical North America (including the Isthmus of Panama) and South America, and thus these species likely migrated via the isthmus. Both *Celtis pallida* and *Celtis spinosa* exhibit typical amphitropical disjunctions between North and South America. Although their migrations from South America to tropical North America each occurred later than the closure of the Isthmus of Panama (with the most latest closure at 3.5 Ma; Bacon et al., 2015), an explanation of migration via the isthmus followed by extinction from a broad area between their northern and southern disjunct regions is not as parsimonious as one of long-distance dispersal. Long-distance dispersals via birds have been proposed to explain many plant groups showing amphitropical disjunctions (Yi et al., 2015), which most probably accounts for disjunctions of these two Cannabaceae species.

4.6 Long-distance dispersal

From the inferred timing and pattern of the dispersals among the Cannabaceae, we conclude that the species of this family are born migrators, resulting in wide distribution patterns among temperate to tropical regions worldwide. Many Cannabaceae species, especially those of the *Celtis* and the *Trema*, are distributed in

many remote islands, such as the Austral Islands, Fiji, Hawaii, the Marquesas Islands, Tahiti, the North Mariana Islands, Wallis and Futuna (see detail in GBIF); all of these are far distant from any continental land mass, and some are young volcanic islands. Cannabaceae species clearly have long-distance dispersibility across broad oceans. The colorful mature drupes of most Cannabaceae species are an adaptation for bird dispersal: various species of frugivorous birds, including bulbuls and crows, have been recorded as dispersal agents for *Aphananthe*, *Celtis* and *Trema* (Soepadmo, 1977; Yoshikawa & Kikuzawa, 2009). Some recorded birds that disperse Cannabaceae seeds are migratory and have long-distance dispersal ability. Thus, bird dispersal can be inferred to be important in accounting for the wide distribution of the Cannabaceae. However, most bird migratory flyways run north and south, and the retention time of seeds in bird's digestive tract is usually short. Furthermore, fruits of the disjunctly distributed genera *Aphananthe*, *Celtis* and *Trema* lack special appendix structures of their fruits adapted to epizoochory by birds. Therefore, long-distance dispersal via birds may not explain all trans-oceanic dispersals. There is still no direct evidence of adaptation of Cannabaceae seeds for hydrochory but their hard, persistent and durable endocarp could prevent the embryo from damage in fresh and brackish water. Distributions of Cannabaceae species along sea coasts (e.g., *Celtis philippensis*) or streams (e.g., *Celtis timorensis*) (Soepadmo, 1977) suggest seed dispersed by water. Long-distance dispersal through ocean currents may also contribute to the broad distribution of the Cannabaceae. However, experiments should be carried out to detect salt tolerance of Cannabaceae seeds.

5 Conclusions

We have found that dispersal appears to have played a major role in shaping the current disjunction of the Cannabaceae. Migrations via land bridges mainly account for disjunctions in the North Hemisphere, whereas long-distance dispersal via birds and/or ocean currents best accounts for the tropical disjunct distributions. The rapid diversification of early branching lineages is consistent with the formation of an extensive boreotropical flora in the northern latitudes during the early Cenozoic. Climatic deterioration from the late Eocene significantly affected distribution patterns and diversification rates of the Cannabaceae, driving the ancestors to adapt to drier and colder climates in the North Hemisphere, or humid and warm climate after their migrations into tropical regions followed by rapid species diversification. This study advances our understanding of the formation of distribution patterns and species diversification of a plant family with tropical to temperate disjunct distributions. Future studies of plant families with similar distribution patterns could test more general biogeographical patterns and contribute to understanding the underlying process of historical diversification.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China [key international (regional) cooperative research project No. 31720103903], the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31000000), the Large-scale Scientific Facilities of the Chinese Academy of Sciences (No. 2017-LSF-GBOWS-02), the National Natural Science Foundation of China (Project No. 31270274). We thank the Botanic Garden Conservation International (BGCI) institution, the Missouri Botanical Garden, the Australian

National Herbarium for providing samples, and Freek T. Bakker, Ali Sattarian, Susanne Renner, Shelley McMahon, Nelson Zamora, Shudong Zhang, Cai Jie, Pan Li, Kyong-Sook Chung for sampling. We also thank Joseph W. Brown, Stephen A. Smith and Yan Yu for providing helpful suggestions on analyses. This study was facilitated by the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences.

References

Alfaro ME, Zoller S, Lutzoni F. 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* 20: 255–266.

Ali JR, Aitchison JC. 2008. Gondwana to Asia: plate tectonics, paleogeography and the biological connectivity of the Indian sub-continent from the Middle Jurassic through latest Eocene (166–35 Ma). *Earth-Science Reviews* 88: 145–166.

Ali JR, Huber M. 2010. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* 463: 653–656.

An Z, Kutzbach JE, Prell WL, Porter SC. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya-Tibetan plateau since Late Miocene times. *Nature* 411: 62–66.

Bacon CD, Silvestro D, Jaramillo C, Smith BT, Chakrabarty P, Antonelli A. 2015. Biological evidence supports an early and complex emergence of the Isthmus

of Panama. *Proceedings of the National Academy of Sciences USA* 112: 6110–6115.

Bartish IV, Antonelli A, Richardson JE, Swenson U. 2011. Vicariance or long-distance dispersal: historical biogeography of the pantropical subfamily Chrysophylloideae (Sapotaceae). *Journal of Biogeography* 38: 177–190.

Blakey RC. 2008. Gondwana paleogeography from assembly to breakup—a 500 My odyssey. *Geological Society of America Special Papers* 441: 1–28.

Clayton JW, Soltis PS, Soltis DE. 2009. Recent long-distance dispersal overshadows ancient biogeographical patterns in a pantropical angiosperm family (Simaroubaceae, Sapindales). *Systematic Biology* 58: 395–410.

Cody S, Richardson JE, Rull V, Ellis C, Pennington RT. 2010. The Great American Biotic Interchange revisited. *Ecography* 33: 326–332.

Collinson ME. 1989. The fossil history of the Moraceae, Urticaceae (including Cecropiaceae) and Cannabaceae. In: Crane PR, Blackmore S eds. *Evolution, systematics and fossil history of the Hamamelidae, Vol. 2: 'Higher' Hamamelidae*. Oxford: Clarendon Press. 319–339.

Collinson ME. 1992. Vegetational and floristic changes around the Eocene/Oligocene boundary in western and central Europe. In: Prothero DR, Berggren WA eds. *Eocene-Oligocene climate and biotic evolution*. Princeton: Princeton University Press. 437–450.

Conti E, Eriksson T, Schönenberger J, Sytsma KJ, Baum DA. 2002. Early Tertiary out-of-India dispersal of crypteroniaceae: evidence from phylogeny and molecular dating. *Evolution* 56: 1931–1942.

Couvreur TLP, Pirie MD, Chatrou LW, Saunders RMK, Su YCF, Richardson JE, Erkens RHJ. 2011. Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography* 38: 664–680.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772–772.

Davis CC, Bell CD, Mathews S, Donoghue MJ. 2002. Laurasian migration explains Gondwanan disjunctions: evidence from Malpighiaceae. *Proceedings of the National Academy of Sciences USA* 99: 6833–6837.

Donoghue MJ, Bell CD, Li J-H. 2001. Phylogenetic patterns in Northern Hemisphere plant geography. *International Journal of Plant Sciences* 162: S41–S52.

Doweld AB. 2016. *Humulago*, a replacement name for fossil *Humularia* Dorofeev (Cannabaceae) non extant *Humularia* Duvigneaud (Fabaceae). *Annales Botanici Fennici* 53: 403–404.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Ducousso M, Béna G, Bourgeois C, Buyck B, Eyssartier G, Vincelette M, Rabevohitra R, Randrihasipara L, Dreyfus B, Prin Y. 2004. The last common ancestor of Sarcolaenaceae and Asian dipterocarp trees was ectomycorrhizal

before the India-Madagascar separation, about 88 million years ago. *Molecular Ecology* 13: 231–236.

Federman S, Dornburg A, Downie A, Richard AF, Daly DC, Donoghue MJ. 2015. The biogeographic origin of a radiation of trees in Madagascar: implications for the assembly of a tropical forest biome. *BMC Evolutionary Biology* 15: 216.

Hall R. 2009. Southeast Asia's changing palaeogeography. *Blumea* 54: 148–161.

Harrison TM, Copeland P, Kidd WSF, Yin AN. 1992. Raising Tibet. *Science* 255: 1663–1670.

Hillis DM, Bull JJ. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42: 182–192.

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

Knobloch E, Mai DH. 1986. Monographie der Früchte und Samen in der Kreide von Mitteleuropa. *Rozpravy Ústředního ústavu Geologického* 47: 1–219.

Kürschner H. 1998. Biogeography and introduction to vegetation. In: Ghazanfar SA, Fisher M eds. *Vegetation of the Arabian Peninsula*. Dordrecht: Springer Netherlands. 63–98.

Lomolino MV. 2010. Four Darwinian themes on the origin, evolution and preservation of island life. *Journal of Biogeography* 37: 985–994.

Manchester SR, Akhmetiev MA, Kodrul TM. 2002. Leaves and fruits of *Celtis aspera* (Newberry) comb. nov. (Celtidaceae) from the Paleocene of North America and eastern Asia. *International Journal of Plant Sciences* 163: 725–736.

Manchester SR, Chen Z-D, Lu A-M, Uemura K. 2009. Eastern Asian endemic seed plant genera and their paleogeographic history throughout the Northern Hemisphere. *Journal of Systematics and Evolution* 47: 1–42.

Matzke NJ. 2013. BioGeoBEARS: BioGeography with Bayesian (and Likelihood) evolutionary analysis in R Scripts. R package, version 0.2.1, published 27 July 2013 [online]. Available from http://CRAN.R-project.org/package=BioGeoBEARS [accessed 1st September 2017].

McLoughlin S. 2001. The breakup history of Gondwana and its impact on pre-Cenozoic floristic provincialism. *Australian Journal of Botany* 49: 271–300.

Michalak I, Zhang L-B, Renner SS. 2010. Trans-Atlantic, trans-Pacific and trans-Indian Ocean dispersal in the small Gondwanan Laurales family Hernandiaceae. *Journal of Biogeography* 37: 1214–1226.

Milne RI. 2006. Northern Hemisphere plant disjunctions: a window on Tertiary land bridges and climate change? *Annals of Botany* 98: 465–472.

Morley RJ. 2011. Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests. In: Bush M, Flenley J, Gosling W eds. *Tropical rainforest responses to climatic change*. Chichester: Praxis Publishing. 1–34.

Muller J. 1981. Fossil pollen records of extant angiosperms. *The Botanical Review* 47: 1–142.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* 20: 289–290.

Pigg K, Wehr W. 2002. Tertiary flowers, fruits, and seeds of Washington State and adjacent areas. Part III. *Washington Geology* 30: 3–16.

Popov SV, Rögl F, Rozanov AY, Steininger FF, Shcherba IG, Kovac M. 2004. Lithological-Palaeogeographic maps of Paratethys. 10 maps Late Eocene to Pliocene. *Courier Forschungsinstitut Senckenberg* 250: 1–46.

Rabinowitz PD, Coffin MF, Falvey D. 1983. The separation of Madagascar and Africa. *Science* 220: 67–69.

Rabosky DL, Donnellan SC, Grundler M, Lovette IJ. 2014a. Analysis and visualization of complex macroevolutionary dynamics: an example from Australian scincid lizards. *Systematic Biology* 63: 610–627.

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

Rambaut A, Drummond AJ. 2004. Tracer, version 1.6 [online]. Available from http://tree.bio.ed.ac.uk/software/ [accessed 11th Dec 2013].

Ree RH, Smith SA. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Renner SS. 2004. Multiple Miocene Melastomataceae dispersal between Madagascar, Africa and India. *Philosophical Transactions of the Royal Society B:Biological Sciences* 359: 1485–1494.

Renner SS, Strijk JS, Strasberg D, Thébaud C. 2010. Biogeography of the Monimiaceae (Laurales): a role for East Gondwana and long-distance dispersal, but not West Gondwana. *Journal of Biogeography* 37: 1227–1238.

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *BMC Bioinformatics* 19: 1572–1574.

Sanmartín I, Ronquist F. 2004. Southern Hemisphere biogeography inferred by event-based models: plant versus animal patterns. *Systematic Biology* 53: 216–243.

Soepadmo E. 1977. Ulmaceae. In: van Steenis CGGJ ed. *Flora Malesiana series I.* Djakarta: Noordhoff-Kolff. 8: 31–76.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771.

Sytsma KJ, Morawetz J, Pires JC, Nepokroeff M, Conti E, Zjhra M, Hall JC, Chase MW. 2002. Urticalean rosids: circumscription, rosid ancestry, and phylogenetics based on *rbcL, trnL-F*, and *ndhF* sequences. *American Journal of Botany* 89: 1531–1546.

Tiffney BH, Manchester SR. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere Tertiary. *International Journal of Plant Sciences* 162: S3–S17.

van Hinsbergen DJJ, Lippert PC, Dupont-Nivet G, McQuarrie N, Doubrovine PV, Spakman W, Torsvik TH. 2012. Greater India Basin hypothesis and a two-stage Cenozoic collision between India and Asia. *Proceedings of the National Academy of Sciences USA* 109: 7659–7664.

van Velzen R, Bakker FT, Sattarian A, Van der Maesen LJG. 2006. Evolutionary relationships of Celtidaceae. In: Sattarian A ed. *Contribution to the biosystematics of Celtis* L. *(Celtidaceae) with special emphasis on the African species*. Wageningen: Wageningen University. 7–30.

Webb SD. 1991. Ecogeography and the Great American Interchange. *Paleobiology* 17: 266–280.

Wolfe JA. 1975. Some aspects of plant geography of the Northern Hemisphere during the late Cretaceous and Tertiary. *Annals of the Missouri Botanical Garden* 62: 264–279.

Wolfe JA. 1992. Climatic, floristic and vegetational changes near the Eocene-Oligocene boundary in North America. In: Prothero DR, Berggren WA eds. *Eocene-Oligocene climate and biotic evolution*. Princeton: Princeton University Press. 421–436.

Xiang Q-Y, Soltis DE. 2001. Dispersal-vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the Northern Hemisphere. *International Journal of Plant Sciences* 162: S29–S39.

Yang M-Q, Li D-Z, Wen J, Yi T-S. 2017. Phylogeny and biogeography of the amphi-Pacific genus *Aphananthe*. *PLoS ONE* 12: e0171405.

Yang M-Q, van Velzen R, Bakker FT, Sattarian A, Li D-Z, Yi T-S. 2013. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* 62: 473–485.

Yi T-S, Jin G-H, Wen J. 2015. Chloroplast capture and intra- and inter- continental biogeographic diversification in the Asian – New World disjunct plant genus *Osmorhiza* (Apiaceae). *Molecular Phylogenetics and Evolution* 85: 10–21.

Yoshikawa T, Kikuzawa K. 2009. Pre-dispersal seed predation by a granivorous bird, the masked grosbeak (*Eophona personata*), in two bird-dispersed Ulmaceae species. *Journal of Ecology & Field Biology* 32: 137–143.

Zachos J, Pagani M, Sloan L, Thomas E, Billups K. 2001. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* 292: 686–693.

Zhang H-L, Jin J-J, Moore MJ, Yi T-S, Li D-Z. 2018a. Plastome characteristics of Cannabaceae. *Plant Diversity* 40: 127-137.

Zhang Q, Onstein RE, Little SA, Sauquet H. 2018b. Estimating divergence times and ancestral breeding systems in Ficus and Moraceae. *Annals of Botany* 123: 191-204.

Zhang S-D, Soltis DE, Yang Y, Li D-Z, Yi T-S. 2011. Multi-gene analysis provides a well-supported phylogeny of Rosales. *Molecular Phylogenetics and Evolution* 60: 21–28.

Zhou L, Su YCF, Thomas DC, Saunders RMK. 2012. 'Out-of-Africa' dispersal of tropical floras during the Miocene climatic optimum: evidence from *Uvaria* (Annonaceae). *Journal of Biogeography* 39: 322–335.

Table 1 Marginal probabilities of area reconstructions from ancestral geographical range analysis of the Cannabaceae, and the mean divergence times with 95% highest probability density (HPD) values. The left and right marginal probabilities of the best split are shown. All nodes correspond to those in Fig. 3. Area designations are as follows: A, Africa; B, Madagascar; C, tropical North America (including the Antilles); D, temperate and subtropical North America; E, South America; F, East Asia; G, South and Southeast Asia; H, Central-West Eurasia; I, Australasia (including Australia, New Zealand, the island of New Guinea, and neighboring islands in the Pacific Ocean).

Node	Area reconstruction	Left marginal probability	Right marginal probability	Estimated divergence time (Ma)	95% HPD interval (Ma)
1	B|G	0.597	0.411	6.52	3.13–10.8
2	G|G	0.571	0.175	9.72	5.45–15.1
3	D|G	0.293	0.414	13.3	7.96–20.1
4	F|CDH	0.462	0.103	17.3	10.7–26.2
5	E|C	0.966	0.693	4.63	1.17–10.5
6	G|I	0.981	0.651	1.36	0–4.84
7	I|G	0.594	0.598	5.45	2.22–10.3
8	I|G	0.292	0.319	12.6	6.50–21.3
9	D|F	0.228	0.188	68.7	66.8–71.8

10	F|H	0.159	0.149	40.5	38.7–43.3
11	DFH|F	0.201	0.726	10.7	4.93–17.8
12	G|F	0.207	0.237	25.4	23.7–28.0
13	I|I	0.957	0.957	2.36	0.960–4.16
14	I|I	0.958	0.958	2.28	0.971–3.97
15	I|I	0.99	0.989	3.13	1.68–4.95
16	G|I	0.71	0.997	3.73	2.13–5.81
17	G|I	0.542	0.816	10.5	6.75–14.9
18	CE|C	0.474	0.980	0.698	0.103–1.71
19	E|C	0.48	0.576	2.92	1.31–5.07
20	D|C	0.403	0.559	4.92	2.70–8.12
21	FG|D	0.369	0.890	8.52	5.65–12.1
22	G|FGI	0.997	0.547	0.659	0.166–1.46
23	G|FGI	0.95	0.268	2.41	0.929–4.57
24	I|I	0.938	0.938	3.59	1.56–6.33
25	I|I	0.946	0.978	5.60	3.07–8.71
26	G|I	0.733	0.986	6.77	3.96–10.1
27	F|I	0.354	0.417	10.1	6.86–13.9

28	G|F	0.611	0.156	11.5	7.86–15.9
29	I|G	0.912	0.727	14.4	10.2–19.5
30	G|I	0.319	0.442	15.8	11.1–21.4
31	G|I	0.169	0.344	18.8	13.0–26.0
32	F|F	0.186	0.186	53.4	46.3–58.5
33	B|A	0.997	0.990	2.39	1.00–4.37
34	AGI|A	0.143	0.535	4.39	2.23–7.23
35	I|G	0.89	0.321	8.11	4.52–12.8
36	A|A	0.602	0.602	5.50	2.54–9.18
37	B|A	0.703	0.836	7.11	3.66–11.3
38	A|A	0.599	0.447	11.4	7.39–16.3
39	E|E	0.378	0.378	0.165	0.016–0.421
40	C|C	0.342	0.376	0.339	0.080–0.740
41	E|C	0.997	0.607	0.684	0.265–1.28
42	E|CE	0.985	0.683	0.948	0.392–1.72
43	E|CE	0.888	0.427	1.79	0.787–3.20
44	G|D	0.297	0.451	11.4	5.30–17.6
45	A|CDFH	0.827	0.120	14.4	9.81–20.1

46	G|A	0.356	0.132	17.0	11.8–23.4
47	A|F	0.529	0.475	2.63	1.08–4.75
48	G|H	0.533	0.305	7.34	4.50–11.0
49	CD|D	0.564	0.990	0.370	0.037–0.950
50	H|D	0.957	0.432	1.06	0.371–1.97
51	C|DH	0.997	0.683	1.48	0.698–2.54
52	D|CDH	0.995	0.633	2.39	1.12–3.95
53	C|CDH	0.979	0.514	3.52	2.04–5.35
54	H|H	0.959	0.959	1.37	0.493–2.54
55	F|H	0.986	0.992	2.85	1.60–4.29
56	F|F	0.975	0.975	0.331	0.031–0.875
57	F|F	0.955	0.971	0.758	0.161–1.88
58	F|F	0.968	0.968	0.550	0.143–1.11
59	F|F	0.905	0.905	0.804	0.356–1.39
60	F|F	0.937	0.987	0.982	0.512–1.58
61	F|F	0.982	0.994	1.18	0.663–1.84
62	G|F	0.812	0.997	1.39	0.804–2.15
63	F|G	0.915	0.402	1.84	1.11–2.77

64	F|F	0.86	0.818	2.42	1.53–3.57
65	F|F	0.939	0.977	2.84	1.83–4.14
66	H|F	0.715	0.995	3.42	2.25–4.92
67	CDH|F	0.745	0.943	4.61	3.05–6.66
68	G|CDFH	0.455	0.309	8.41	5.45–12.3
69	C|DFH	0.708	0.217	14.2	9.02–20.7
70	G|CDFH	0.437	0.265	23.6	16.4–32.5
71	F|D	0.216	0.180	58.5	56.7–61.3
72	H|F	0.148	0.191	61.8	57.3–66.4
73	F|F	0.167	0.172	71.0	67.3–76.2
74	F|F	0.122	0.149	87.4	78.8–94.0

Table 2 Biogeographic model fits

Results of ancestral area reconstructions based on different biogeographic models. Bolded line corresponds to the best fitting model, which represent SD events estimated under DEC, and account for 0.844 of the Akaike weights. Abbreviations:

LnL: Log likelihood

P: Number of parameters in the model

d: Estimated dispersal rate

e: Estimated extinction rate

Model	LnL	P	d	e	AICc	AICc weight
DEC (TR)	-298.57	2	0.0695	0.0307	601.31	9.67E-33
DEC (SD)	**-225.02**	**2**	**0.0514**	**0.0362**	**454.21**	**0.844**
DEC (LDD)	-226.71	2	0.0286	0.0217	457.59	0.156

Fig. 1. The maximum clade credibility (MCC) tree of Cannabaceae derived from the BEAST analysis with five fossils as stem age constraints (red circle; C1: *Triorites minutipori*, offset = 89.8 Ma; C2: *Gironniera gonnensis*, offset = 66.0 Ma; C3: *Celtis aspera*, offset = 56 Ma; C4: *Pteroceltis knowltonii*, offset = 38.0 Ma; C5: *Humulago reticulate*, offset = 23.0 Ma). Median age estimates and 95% highest posterior densities (blue node bars) (Ma) are shown for each node (see Table 1).

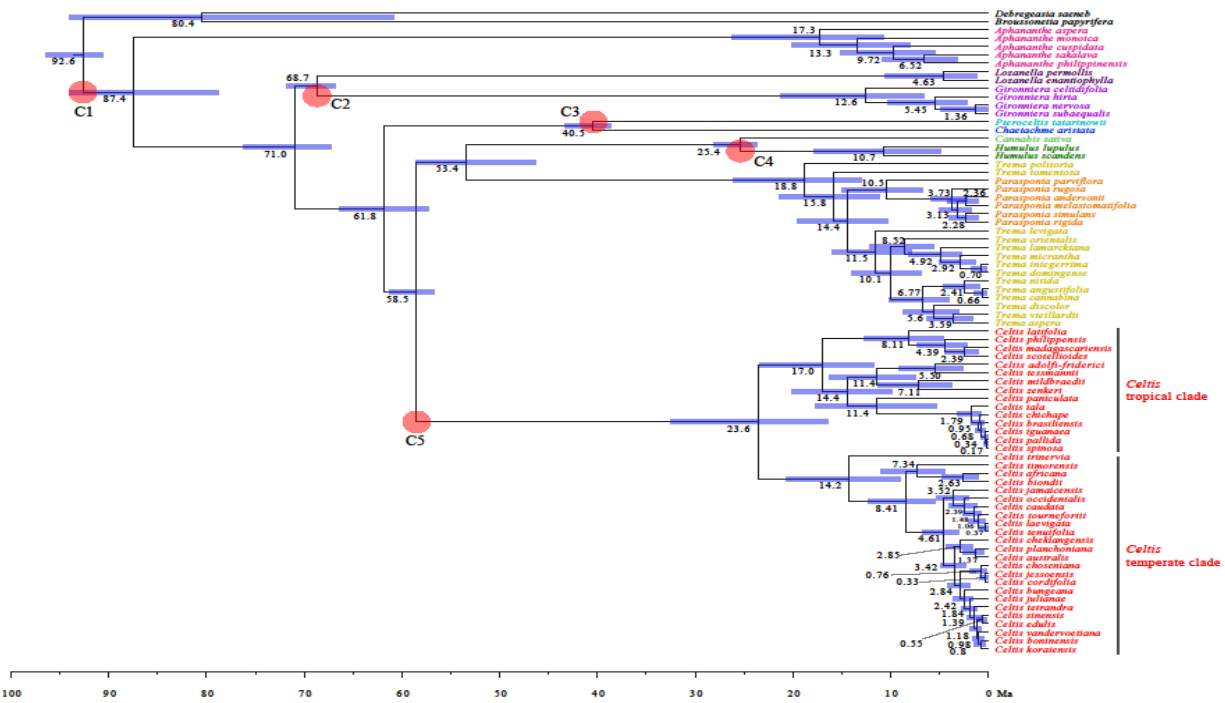

Fig. 2. Diversification analysis of Cannabaceae. **A,** The best shift phylorate plot of Cannabaceae inferred from BAMM, with colors (cool = slow, warm = fast) indicating the mean evolutionary rate across all shift configurations sampled during simulation. The circle in this plot indicates the most frequent shift along all sampled trees of the Bayesian analysis. **B,** The 95% credible set of macroevolutionary shift configurations identified as based on branch-specific marginal probabilities. For each distinct shift configuration, the locations of rate shifts are shown as red circles (rate increases). Text labels denote the posterior probability of each shift configuration. **C,** Lineage-through-time (LTT) plot for the sampled Cannabaceae species with the MCC tree (dark green) and 500 replicate trees (light green) used for the confidence level.

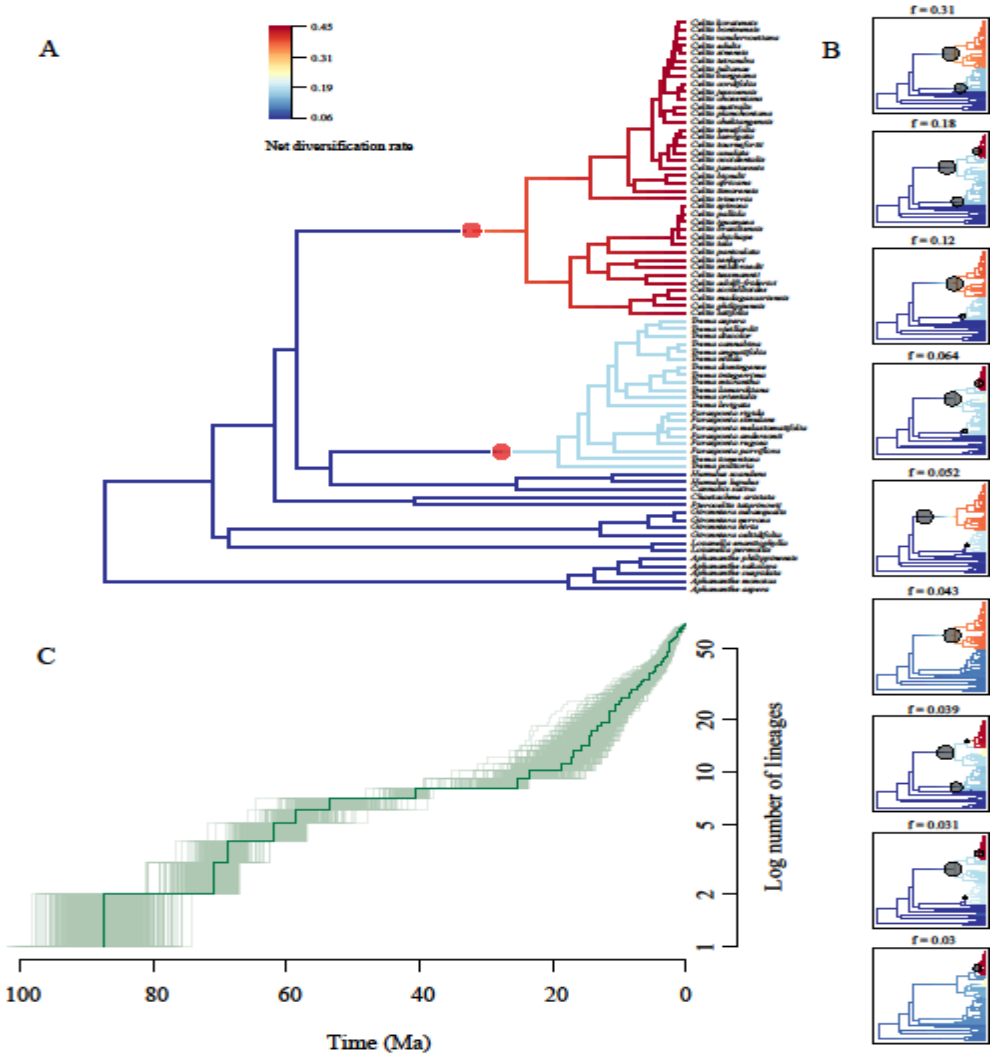

Fig. 3. Global biogeographical patterns of Cannabaceae. **A,** Ancestral area reconstruction of Cannabaceae with the DEC + SD model. Node names are shown along the tree as postorder numbers. Colors represent the nine delimited geographical areas, as indicated. The colored squares at terminal nodes represent corresponding species distributions. Squares forming two rows at an internal node represent the best split, with each row of squares depicting the inheritance scenario for the descendant lineage at the same side. Square(s) forming a single row at internal nodes indicates that both descendants inherit the same area range. The *Celtis* seedling pic in the center is by Jian-Jun Jin. **B,** Map showing nine biogeographical regions in colors as defined in this study. The base map is a Mollweide projection downloaded from https://www.your-vector-maps.com/_kepek/_kep_zip_free/wrld-bm-2-ai.zip. Area abbreviations: A, Africa; B, Madagascar; C, tropical North America (including the Antilles); D, temperate and subtropical North America; E, South America; F, East Asia; G, South and Southeast Asia; H, Central-West Eurasia; I, Australasia (including Australia, New Zealand, New Guinea, and neighboring islands in the Pacific Ocean).