Phylogeny, biogeography, and breeding system evolution in Moraceae

Qian Zhang

université PARIS-SACLAY

NNT : 2019SACLS205

Thèse de doctorat

Phylogeny, biogeography, and breeding system evolution in Moraceae

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'Université Paris-Sud

ED n°567 Sciences du végétal : du gène à l'écosystème
Spécialité de doctorat : Biologie

Thèse présentée et soutenue à Orsay, le 16/07/2019, par

Qian Zhang

Composition du Jury :

Tatiana Giraud Directrice de Recherche, CNRS (ESE)	Présidente
Mathilde Dufaÿ Professeur, Université de Montpellier (CEFE)	Rapporteur
Jean-Yves Rasplus Directeur de Recherche, INRA (CBGP)	Rapporteur
Florian Jabbour Maître de Conférences, Muséum national d'Histoire naturelle (ISYEB)	Examinateur
Jos Käfer Chargé de Recherche, Université Lyon I (ESE)	Examinateur
Hervé Sauquet Maître de Conférences, Université Paris-Sud (ESE)	Directeur de thèse

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors Dr. Hervé Sauquet and Dr. Renske Onstein, for their patience and help throughout my PhD journey. Their advice on both research as well as my career have been invaluable. I still remember when I introduced my idea, which finally became chapter III of my thesis, to Hervé at the very beginning. This idea has been elaborated and finally become a PhD thesis with help from many people I will mention below.

I would also like to extend my deepest appreciation to my thesis committee, without whose help I will not improve so much in my studies. They are Dr. Thomas Couvreur, Dr. Elliot Gardner, Dr. Damien Hinsinger, Prof. Sophie Nadot, Prof. Jacqui Shykoff and Prof. Nina Rønsted.

I am extremely grateful to my jury of defense: Prof. Mathilde Dufaÿ, Dr. Jean-Yves Rasplus, Prof. Tatiana Giraud, Dr. Florian Jabbour, Dr. Jos Käfer and Dr. Hervé Sauquet. Thank them for their participation in my defense, helpful comments to my thesis and future studies.

I deeply indebted to my coauthors as well. I am grateful to Dr. Stefan Little for his profound knowledge in fossils. I would also like to extend my sincere thanks to Dr. Elliot Gardner and Prof. Nyree Zerega, for their insightful guides on experiments and analyses in Chapter II. I must thank them again for sharing their collections of Neotropical species of Dorstenieae with me.

I also wish to thank Véronique Normand and Alodie Snirc for their great help in my experiments.

Thanks should also go to all the members in team EVA (Evolution of Angiosperms), especially Prof. Sophie Nadot, who took care of me when Hervé was not in Orsay. I wish to thank Xavier Aubriot for sharing much knowledge in taxonomy and biogeography with me. Thanks also to Xiaolei Pang, who introduced the knowledge of ITCZ to me.

I gratefully acknowledge Martine Fournier and Prof. Jacqui Shykoff's great help in the daily paper works. I would like to acknowledge Emmanuelle Jestin's excellent work in arranging the travel and accomodation for my Jury of thesis defense as well.

I had great pleasure of working with my three kind office mates Fantin Carpentier, Fanny Hartmann, Marine Duhamel and other members in team GEE (Evolutionary genetics and ecology), who brought much social life to me.

Finally, I cannot begin to express my thanks to my parents, without whose support I will never leave home so far away and have a chance to chase my dream. Thanks to Prof. Xiaoyong Chen, who helped me to apply the CSC fellowship. Thanks to my three best friends: Xiaokun Zhang, Sen Lin, Dazheng Dong, for their company, no matter how physically far away we are. Thanks also to all those who helped and accompanied with me in these years. It is a pity that the space is too limmited to list all of you, but I remember you and all your help.

ABSTRACT

Angiosperms are the most diversified clade of extant plants and are exceptionally species-rich in tropical regions. Flowers are the breeding organs of angiosperms. Flowers exhibit remarkable levels of both structural and functional diversity and for this reason have long thought to have a direct influence on the diversity and evolution of angiosperms. In this thesis, I investigated breeding system evolution and biogeographic history in the family Moraceae (40 genera, ca. 1100 species, including its largest genus, *Ficus*, ca. 750 species), which I used as a model clade to understand the origin and evolution of diversity of angiosperms.

In Chapter I, I reconstructed and calibrated a new dated phylogenetic tree for Moraceae as a whole, based on a dataset of 320 species (including 272 species sampled from 36 genera in Moraceae, and 48 species representing 8 outgroup families in Rosales), eight molecular markers (three chloroplastic, five nuclear), and twelve fossil age constraints. I then used this phylogenetic tree to reconstruct ancestral states of breeding systems in Moraceae and *Ficus* using parsimony and model-based approaches, using and comparing six configurations of breeding system data in Moraceae, differing in the number of states recognized. The crown group ages of Moraceae and *Ficus* were estimated in the Cretaceous (73.2-84.7 Ma) and in the Eocene (40.6-55.9 Ma), respectively. Dioecy was inferred as the ancestral breeding systems of Moraceae with high support by all the approaches, models and configurations, followed by several subsequent transitions to monoecy, including in *Ficus*. This result suggests that dioecy is not necessarily an evolutionary dead end. While monoecy emerged as the most likely ancestral state of *Ficus*, this result remained uncertain and sensitive to model selection.

In Chapter II, I reconstructed a dated phylogenetic tree for tribe Dorstenieae (15 genera, ca. 156 species, including *Dorstenia*, ca. 113 species), mainly distributed in tropical regions. To do so, I produced a new dataset of nuclear genomic data (102 genes) generated with a target enrichment approach (Hyb-Seq), sampling all genera (15) and 83 (53%) species in the tribe and five outgroups. This phylogenomic tree allowed me to reconstruct the biogeographic history of this group using dispersal-extinction-cladogenesis models, including models with a founder-event speciation process. The crown group ages of Dorstenieae and *Dorstenia* were estimated in the Cretaceous (65.8-79.8 Ma) and in the Cretaceous/Paleocene to Eocene period (50.8-67.3 Ma), respectively. Two long-distance dispersal events from continental Africa to South America occurred in the Cenozoic, one in *Dorstenia* and one in *Brosimum* s.l. Species of *Dorstenia* further colonized Central America during the Oligocene to Miocene (12.0-34.7 Ma).

In Chapter III, I tested the climatic niche difference (temperature and precipitation) between the two breeding systems (monoecy and gynodioecy) in *Ficus* using a new dataset of cleaned spatial occurrence records and breeding systems for 183 species. To do so, I used two comparative approaches, differing in whether phylogeny is taken into account (generalized estimating equations, GEE) or not (generalized linear models, GLM). A positive relationship between precipitation and gynodioecy was supported by GLM, but not by GEE analyses, and no relationship between temperature and breeding systems was supported by either method. Higher dispersal ability and the potential for self-fertilization may explain why monoecious species of *Ficus* have been able to colonize and survive in drier environments.

This thesis highlights the potential of phylogenetic comparative methods and

phylogenomic data to address questions of breeding system evolution and biogeography in Moraceae. More densely sampled phylogenies of *Ficus* and Moraceae as a whole will be required to confirm some of the results emerging from this thesis, such as ancestral monoecy in *Ficus*. Nevertheless, this thesis opens up several important new perspectives worth investigating in other plant clades, such as a relationship between breeding system and climatic niche.

RESUME

Les Angiospermes sont le clade le plus diversifié des plantes actuelles et sont exceptionnellement riches en espèces dans les régions tropicales. La fleur est l'organe reproducteur des Angiospermes. Il existe une diversité remarquable de fleurs tant sur le plan structurel que fonctionnel et pour cette raison, l'on pense depuis longtemps que les fleurs ont une influence directe sur la diversité et l'évolution des Angiospermes. Dans cette thèse, j'ai étudié l'évolution des systèmes sexuels et l'histoire biogéographique de la famille des Moraceae (40 genres, env. 1100 espèces, dont le genre le plus diversifié, *Ficus*, env. 750 espèces), clade modèle utilisé pour comprendre l'origine et l'évolution de la diversité chez les Angiospermes.

Dans le Chapitre I, j'ai reconstruit et calibré un nouvel arbre phylogénétique daté pour les Moraceae à partir d'un jeu de données de 320 espèces (dont 272 issues de 36 genres chez les Moraceae, et 48 espèces représentant 8 familles de groupes externes chez les Rosales), 8 marqueurs moléculaires (3 chloroplastiques, 5 nucléaires) et 12 contraintes d'âge fondées sur le registre fossile. J'ai ensuite utilisé cet arbre phylogénétique pour reconstruire les états ancestraux des systèmes sexuels chez les Moraceae et *Ficus* avec les méthodes de parcimonie et de maximum de vraisemblance, en utilisant et comparant six configurations de données de systèmes sexuels chez les Moraceae différant par le nombre d'états distingués. Les âges des groupes-couronne des Moraceae et du genre *Ficus* sont estimés au Crétacé (73.2-84.7 Ma) et à l'Eocène (40.6-55.9 Ma), respectivement. La dioécie est inférée comme l'état ancestral des systèmes sexuels chez les Moraceae, un résultat très bien soutenu par toutes les approches, modèles et configurations. Plusieurs transitions ultérieures vers la monoécie se sont ensuite produites, y compris chez *Ficus*. Ce

résultat suggère que la dioécie ne représente pas nécessairement un cul-de-sac évolutif. La monoécie semble être l'état ancestral le plus vraisemblable chez *Ficus*, mais ce résultat est peu robuste et sensible au choix du modèle.

Dans le Chapitre II, j'ai reconstruit un arbre phylogénétique daté pour la tribu des Dorstenieae (15 genres, env. 156 espèces, dont *Dorstenia*, env. 113 espèces), distribuée principalement dans les régions tropicales. Pour ce faire, j'ai produit un nouveau jeu de données génomiques nucléaires (102 gènes) à partir d'une approche d'enrichissement ciblé (Hyb-Seq) échantillonnant tous les genres (15) et 83 espèces (53%), ainsi que 5 groupes externes. Cet arbre phylogénomique m'a permis de reconstruire l'histoire biogéographique du groupe en utilisant les modèles de dispersion-extinction-cladogenèse, dont des modèles avec un processus d'événement-fondateur de dispersion. Les âges des groupes-couronne des Dorstenieae et du genre *Dorstenia* sont estimés au Crétacé (65.8-79.8 Ma) et dans la période du Crétacé/Paléocène à l'Eocène (50.8-67.3 Ma), respectivement. Deux évènements de dispersion à longue distance depuis l'Afrique continentale vers l'Amérique du Sud ont eu lieu au Cénozoïque, l'un chez *Dorstenia* et l'autre chez *Brosimum* s.l. Le genre *Dorstenia* a ensuite colonisé l'Amérique Centrale entre l'Oligocène et le Miocène (12.0-34.7 Ma).

Dans le Chapitre III, j'ai testé les différences de niche climatique (température et précipitation) entre les deux systèmes sexuels (monoécie et gynodioécie) chez *Ficus* avec un nouveau jeu de données fiables d'occurrences spatiales et de systèmes sexuels chez 183 espèces. À cette fin, j'ai utilisé deux approches comparatives, différant dans la prise en compte de la phylogénie (équations d'estimation généralisées, GEE) ou non (modèles linéaires généralisés, GLM). Une relation positive entre précipitation et gynodioécie est

soutenue par les analyses GLM, et aucune méthode ne soutient une relation entre température et système sexuel. Une meilleure capacité à se disperser et le potentiel d'autopollinisation sont deux explications possibles pour la colonisation et la survie des espèces monoïques dans des environnements plus secs.

Cette thèse démontre le potentiel des méthodes phylogénétiques comparatives et des données phylogénomiques pour répondre aux questions d'évolution des systèmes sexuels et de biogéographie chez les Moraceae. Des phylogénies plus densément échantillonnées pour *Ficus* et les Moraceae sont requises pour confirmer certains des résultats émergeant de cette thèse, tels que le caractère ancestral de la monoécie chez *Ficus*. Cependant, cette thèse ouvre plusieurs nouvelles perspectives importantes méritant d'être approfondies chez d'autres clades de plantes, telles que la relation entre système sexuel et niche climatique.

TABLE OF CONTENTS

GENERAL INTRODUCTION

Angiosperms are the dominant plant lineage in most terrestrial ecosystems, especially in tropical regions. Breeding system evolution and biogeography are two of the keys to understand the extreme diversity of angiosperms. My PhD thesis uses the mulberry family, Moraceae, as a model clade to address these topics by answering three key questions: how did breeding systems evolve in Moraceae (Chapter I)? What is the biogeographic history of Dorstenieae, the second most diversified tribe in Moraceae (Chapter II)? Do the two breeding systems in *Ficus* occupy different climatic niches (Chapter III)? This introduction will present the background of three important aspects of my thesis: 1) broader biological background, introducing current knowledge of relationship of breeding system evolution and diversity in angiosperms, and the biogeographic history of the angiosperms in three main tropical regions; 2) the plant family Moraceae and the two large subclades studied in more detail in this thesis, genus *Ficus* and tribe Dorstenieae; 3) specific methods used in my thesis, including the target enrichment approach for genomic sequencing (Chapter II); fossil-calibrated molecular dating (Chapters I and II); parametric models for reconstruction of ancestral states (Chapter I) and biogeographic history (Chapter II).

Breeding system evolution and diversity in angiosperms

The term “breeding system” has not been used consistently throughout the literature and is sometimes treated as a synonym of “reproductive system” or “mating system” (Neal and Anderson 2005; Cardoso *et al.* 2018). In this thesis, to remain consistent with previous work on *Ficus* and Moraceae, I chose the term “breeding system” to describe the arrangement of flowers of different sex at different levels, including both individual and population levels.

Angiosperms are the most diversified lineage of extant plants, accounting for approximately 96% of vascular plants (Christenhusz and Byng 2016). The diversity of their species and morphological attributes, including their flowers, have fascinated biologists for hundreds of years. Because breeding systems directly impact genetic variation and reproductive success at both intra- and interspecific levels, they are critical to understanding floral evolution and diversification. Bisexual flowers are likely ancestral in angiosperms and subsequently evolved many times independently to unisexual flowers associated with a diversity of breeding systems (Sauquet *et al.* 2017). Dioecy is rare (ca. 6% species of angiosperms) but evolved independently thousands of times during the history of angiosperms (Renner 2014). This observation has prompted multiple lines of research on dioecy evolution, including resource allocation strategies, sexual selection, genetic determination of separate sexes in angiosperms (Case and Barrett 2004; Dufay *et al.* 2014; Charlesworth 2015; Käfer *et al.* 2017; Zemp *et al.* 2018). Dioecy was once suggested to be an evolutionary dead end, following the assumption that loss of bisexuality appears to be easier than its gain and the observation that lower species richness was found in dioecious clades (Bull and Charnov 1985; Heilbuth 2000). However, this hypothesis may have been an artificial result from the sister-group comparison approach and has been challenged by two recent studies with new comparative approaches (Käfer *et al.* 2014; Sabath *et al.* 2016). Although bisexuality is predominant in angiosperms, no significant direction of breeding system transition was found in an angiosperm-wide metaanalysis, suggesting that different selective pressures and constraints have been at play in different clades for breeding system evolution in angiosperms (Goldberg *et al.* 2017).

Biogeographic history of angiosperms in the tropics

Tropical regions hold the highest terrestrial diversity on earth (Mutke and Barthlott 2005). Species richness typically decreases from tropics to poles, a pattern commonly referred to as the latitudinal gradient and observed in most clades (Wiens and Donoghue 2004), with a few exceptions (e.g., ray-finned fishes or Actinopterygii; Rabosky *et al.* 2018). Before the Middle Miocene Climatic Transition (MMCT, ca. 14 Ma), tropical climates covered a much larger area than they do now (Morley 2003). Thus, investigations on the origin, evolution and maintenance of diversity in tropical areas is critical not only to help us understand present patterns of diversity distribution, but also to predict their evolution in a changing environment. The origin of diversity of angiosperms in tropical regions, especially in the Neotropics and Africa, is one of the main questions in this thesis. Hence, I will briefly summarize below current knowledge of diversity of angiosperms in the Neotropics and Africa.

Biodiversity is not evenly distributed across tropical regions. Here, we consider the Neotropics as the area extending from central Mexico to Southern Brazil (including the Caribbean islands) (Antonelli and Sanmartín 2011). The Neotropics have been estimated to harbor the most number of vascular plant species, followed by the Indo-Pacific region (Kreft and Jetz 2007; Ulloa Ulloa *et al.* 2017). Both dispersal followed by *in situ* diversification and Andean uplift have been proposed as the main causes for extreme diversification in the Neotropical region (Hoorn *et al.* 2010). In addition, approximately ten biomes have been identified in the Neotropics (Antonelli and Sanmartín 2011; Hughes *et al.* 2013; Antonelli *et al.* 2018), suggesting high habitat diversity. On the one hand, biota exchange occurred between North and South America through the Central American land bridge during the Late Cretaceous to Eocene, and through the Panama Isthmus during the Oligocene-Miocene transition (Morley 2003; Jaramillo *et al.* 2006; Bacon *et al.* 2015).

Long-distance dispersal from continental Africa (e.g., *Ficus* section *Americanae* [Moraceae], *Begonia* [Begoniaceae], Lobeliaceae), Asia-Oceania (e.g. *Bocconia* and *Macleaya* [Papaveraceae], *Osmorhiza* [Apiaceae]) and Australia (e.g., Liliales) also contributed to the diversity of the Neotropics (Moonlight *et al.* 2015; Yi *et al.* 2015; Givnish *et al.* 2016; Knox and Li 2017; Pederneiras *et al.* 2018; Li *et al.* 2018). On the other hand, analyses of biota interchange (including invertebrates, vertebrates, ferns, and angiosperms) among different Neotropical regions suggested Amazonia as the source of Neotropical diversity (Antonelli *et al.* 2018). However, the drivers of Neotropical diversification still remain unclear at present.

Asia-Oceania consists of a heterogenous assemblage of part of Laurasia and several fragments of Gondwana, and these fragments have reached their current positions in different times from the Cretaceous to the Miocene (Hall 2009). This region has been proposed as the place of origin of angiosperms (Buerki *et al.* 2014; Coiro *et al.* 2019). However, more evidence are needed to be found for this hypothesis. Borneo and continental South-East Asia have been suggested to be the major evolutionary hotspots for Southeast Asian Biodiversity during the latest 65.5 Ma in a study examining the assembly of local biota (including invertebrates, vertebrates, and plants) (De Bruyn *et al.* 2014). Floristic exchange between Sunda and Sahul has been estimated to start approximately 33 Ma ago and proceed mainly eastwards (from Sunda including Malay Peninsula, Sumatra, Borneo to Sahul including New Guinea and other Pacific islands) (Crayn *et al.* 2015), suggesting the relationship of floras of South-East Asia and Pacific islands.

It remains obscure why continental Africa holds comparatively lower diversity than in the Neotropics and in South-East Asia. Couvreur (2015) reviewed several hypotheses for the low diversity of tropical Africa (e.g., time-to-speciation effect, area-for-speciation effect), and suggested that lower speciation rates might be one potential explanation. Although continental

Africa has been geographically isolated from the Late Cretaceous, when land bridges between continental Africa and Europe, South America, and the Indian Plate occurred, to the Middle Miocene (Morley 2003), evidence suggests that biota exchange has been continuing through long-distance dispersal. Six floristic regions have been defined in sub-Saharan Africa by geographical analyses, and these floristic regions show relationships with counterparts of other continents (Linder 2014), suggesting flora exchanges in the assembly of the African flora. Although the diversity in Africa as a whole is comparatively lower, clades such as Detarioideae (Fabaceae) harbor higher diversity in continental Africa than other tropical areas, suggesting clades of angiosperms have responded differently to historical climatic change in Africa (de la Estrella *et al.* 2017). Investigation of the biogeographic history of additional clades of angiosperms in Africa, based on solid dated phylogenies, is critical to further understand the causes of low diversity in continental Africa (Couvreur 2015).

Moraceae: an ideal model clade

General presentation of family Moraceae

The angiosperm family Moraceae is an ideal model for investigating in breeding system evolution and the origin of diversity in tropical regions. Moraceae are now classified in order Rosales, together with eight other families (APG IV 2016). Urticaceae are well supported as the sister group of Moraceae (Soltis *et al.* 2011; Zhang *et al.* 2011). Moraceae consist of seven tribes, 40 genera and around 1100 species (Clement and Weiblen 2009; Zerega and Gardner 2019). Moraceae have a pantropical geographic distribution with some species extending into temperate areas (Fig. 1). The family is characterized by milky latex, small, unisexual flowers (pistillode

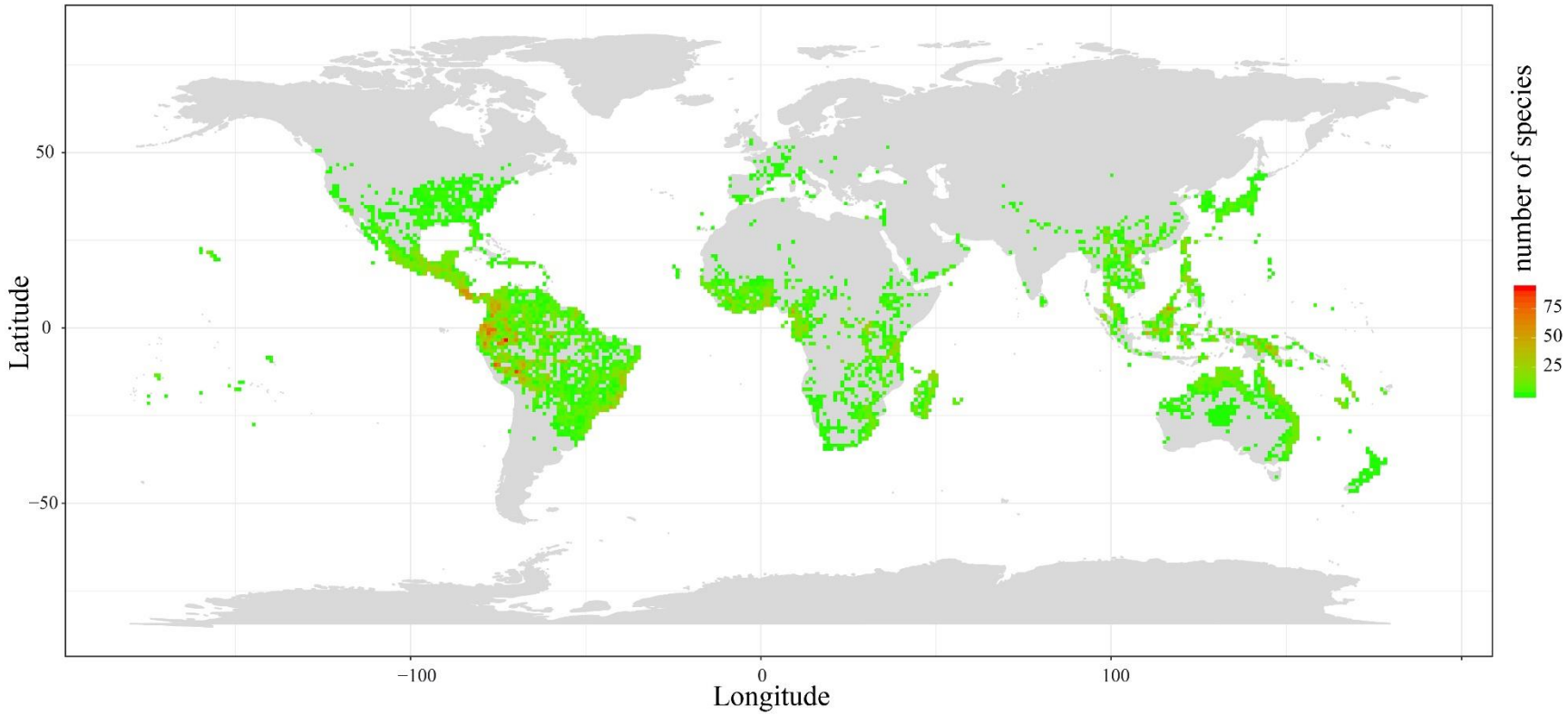


Figure 1. Geographic distribution and species richness of Moraceae. This map was compiled using a dataset of cleaned occurrence records from the Global Biodiversity Information Facility (GBIF), following the same protocol as described in Chapter III. To decrease the impact from cultivated individuals, records of widely cultivated species in Moraceae (*Artocarpus heterophyllus*, *A. altilis*, *Broussonetia papyrifera*, *Castilla elastica*, *Ficus benjamina*, *F. benghalensis*, *F. carica*, *F. elastica*, *F. lyrata*, *F. microcarpa*, *F. pumila*, *F. religiosa*, *Morus alba*, *M. nigra*) were excluded from this map.

present or not), a uniseriate perianth, and compact but variable inflorescences (Berg *et al.* 2006; Simpson 2010). The family exhibits high diversity in several key traits, including habit (tree, shrub, herb, woody liana), inflorescence architecture (from the open inflorescences of *Morus* to the closed ones of *Ficus*), pollination (by wind, e.g. *Morus*; or by insects e.g. *Ficus*) (Rohwer and Berg 1993) and breeding systems (monoecy, androdioecy, gynodioecy, dioecy) (Fig. 2) (Clement and Weiblen 2009). The seeds of Moraceae are mostly dispersed by vertebrates (Rohwer and Berg 1993; Shanahan *et al.* 2001). Dioecy has been proposed to be the ancestral breeding system of Moraceae (Datwyler and Weiblen 2004). The crown-group age of Moraceae has been estimated in the Cretaceous (72.6-110.0 Ma) (Zerega *et al.* 2005), while biogeographic analyses for Moraceae as a whole using model-based approaches have not been conducted so far. The ancestral area of Moraceae remains an enigma and may have been either in Gondwana or in Laurasia (Zerega *et al.* 2005). Except for *Ficus* (see below), biogeography analyses have been conducted so far in three genera (*Artocarpus*, *Maclura* and *Dorstenia*) in Moraceae to date. *Artocarpus* (ca. 70 spp.) has been estimated to originate in Borneo in the Eocene to Oligocene (29.8-50.8 Ma) and subsequently disperse to other Asia-Oceania areas (Williams *et al.* 2017). *Maclura* (ca. 12 spp.) has been estimated to originate in South America in the Cretaceous to Eocene (49.1-73.4 Ma), followed by dispersal to Africa and the northern hemisphere (Gardner *et al.* 2017). The particular case of *Dorstenia* is developed further below.

Moraceae harbors several species of economic value and some of them are widely cultivated. Leaves of mulberry (*Morus*) provide food for silk worms (Fig. 2 A, B, He *et al.* 2013). *Artocarpus heterophyllus* (Fig. 2 C) and *Ficus carica* (Fig. 2 I) are cultivated for their edible infructescences known as breadfruit and fig, respectively (Ghada *et al.* 2010; E W Williams *et al.* 2017). Some species of *Dorstenia* have been used in folk medicine for the treatment of infection, snake bites,

Figure 2. Representative species of the seven tribes of Moraceae. A-B. *Morus alba* female and male inflorescence (Moreae); C. *Artocarpus heterophyllus* (Artocarpeae); D-E. *Parartocarpus venenosus* female and male inflorescence (Parartocarpeae); F. *Castilla elastica* (Castilleae); G. *Maclura pomiera* syncarp (Moreae); H. *Dorstenia barteri* (Dorstenieae); I. *Ficus carica* female inflorescence (Ficeae). Photo: A,I by Qian Zhang; H by Hervé Sauquet; C from Williams et al. (2017); D, E from Zerega and Gardner (2019); G from (Gardner *et al.* 2017), reproduced with authors' permission; B, F from Wikimedia commons by Fastily and Dick Culbert, respectively.

and as anti-inflammatory in Africa and South America (Zapata-Sudo *et al.* 2010). Lastly, the fibrous bark of *Broussonetia papyrifera*, known as paper mulberry, was used to make paper in ancient China (Chung *et al.* 2017).

Genus Ficus

The most diversified genus in Moraceae is *Ficus* (ca. 750 spp.), which is known for its coevolution with its symbiotic pollinating wasps (Hymenoptera, Chalcidoidea, Agaonidae) (Cruaud *et al.* 2012). *Ficus* is characterized by very small unisexual flowers enclosed in urn-shape inflorescences called syconia (Fig. 2 I). There are three kinds of flowers in *Ficus* inflorescences: male flowers (producing pollen), female flowers (producing seeds) and gall flowers (develop wasps). Two breeding systems occur in *Ficus*, monoecy and gynodioecy, differing in the distribution of these three kinds of flowers among syconia (Fig. 3), each representing about half of the species of *Ficus* (Cook and Rasplus 2003). In monoecious *Ficus*, all three kinds of flowers co-occur in every syconium (Fig. 3). Pollinating wasps lay eggs and pollinate the host plants after entering the syconium, depending on the lengths of styles of female flowers in monoecious *Ficus*. Conversely, gynodioecious species of *Ficus* are characterized by two kinds of syconia: male and female (Fig. 3). On the one hand, functional female syconia contain only female flowers. On the other hand, functional male syconia contain both male and gall flowers, and as a result will not deliver seeds. The male pollinating wasps are wingless, and usually never leave the syconium in which they developed. When the mature winged female wasps leave the male syconium in which they developed, they have to pass the cluster of male flowers near the exit of the syconium. These pollinating wasps then look for a new syconium and struggle into it (carrying pollen from another

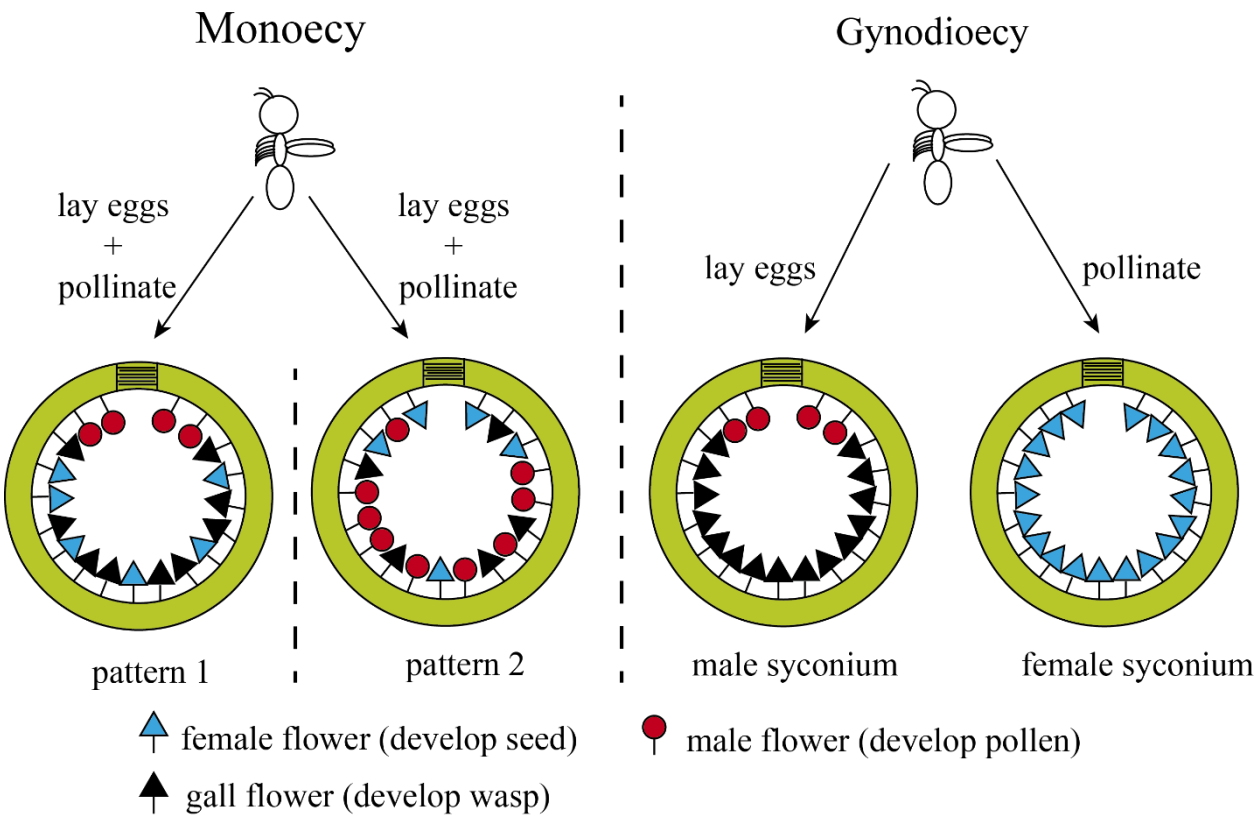


Figure 3. Diagram of syconia (enclosed, urn-shaped inflorescences) in the two breeding systems of *Ficus*, monoecy and gynodioecy.

syconium). They may lose their wings and some legs during the process. Following entry into a male syconium, they lay eggs; if they enter into a female syconium, they pollinate the female flowers (Cook and Rasplus 2003). Thus, structural gynodioecy in *Ficus* (plants are structurally either bisexual, with monoecious inflorescences, or female only) is functionally equivalent to dioecy (plants are functionally either male or female only). *Ficus* has a broad, mostly pantropical geographic distribution across all the three major tropical regions, but gynodioecious *Ficus* are remarkably absent in the Neotropics (Cruaud *et al.* 2012). Monoecy has been proposed to be the ancestral state of breeding system of *Ficus*

(Datwyler and Weiblen 2004).

Six subgenera were proposed for *Ficus* (Berg 2003). Three of them have so far been confirmed to be monophyletic, whereas subgenera *Ficus*, *Pharmacosycea*, and *Urostigma* appear to be polyphyletic in the phylogeny of Cruaud et al. (2012) based on five nuclear markers and 200 species. Pederneiras et al. reconstructed the most densely sampled phylogenetic analysis for *Ficus* to date (249 species) with nine nuclear molecular markers (Pederneiras *et al.* 2018). The topology of Pederneiras et al. (2018) was similar to the one of Cruaud et al. (2012), but with higher support for the deeper nodes. With the development of high-throughput sequencing, initial phylogenomic studies of *Ficus* have also recently been conducted based on much larger gene datasets but more limited taxon sampling (Bruun-Lund *et al.* 2017; Rasplus *et al.* 2018). Phylogenetic relationships based on complete chloroplast genomes (Bruun-Lund *et al.* 2017) differ markedly from phylogenies based on a few nuclear markers (Cruaud *et al.* 2012; Pederneiras *et al.* 2018), suggesting hybridization in the early history of *Ficus*. However, phylogenetic relationships based on nuclear RAD-seq genomic data (Rasplus *et al.* 2018) also differ from previous nuclear phylogenies (Cruaud *et al.* 2012; Pederneiras *et al.* 2018), specifically in relationships at the base of *Ficus*.

The crown-group age of *Ficus* has been estimated in the Cretaceous to the Paleocene (60.0-101.9 Ma) or in the Eocene (34.9-50.6 Ma), depending on different calibration strategies (Cruaud *et al.* 2012; Pederneiras *et al.* 2018). The ancestral area of *Ficus* has been inferred to be Eurasia (Cruaud *et al.* 2012; Pederneiras *et al.* 2018), from which *Ficus* then dispersed from the Old World to the New World twice, leading to the two Neotropical lineages (sections *Pharmacosycea* and *Americanae*) (Pederneiras *et al.* 2018). *Ficus*

subsect. *Urostigma* may have originated in the Paleocene to Eocene (40.3-60.6 Ma) in Madagascar (Chantarasuwan *et al.* 2016).

Tribe Dorstenieae

Tribe Dorstenieae includes fifteen genera and 156 species, distributed on both sides of the Atlantic (Figure 1, Chapter II). *Dorstenia*, with ca. 113 species, is the most diversified genus in Dorstenieae and the second in Moraceae (Berg and Hijman 1999; Berg 2001). New species have been discovered continuously in Africa and South America during the last decade (Mccoy and Massara 2008; dos Santos and Neto 2012; Machado and Marcelo Filho 2012; Chase *et al.* 2013; dos Santos *et al.* 2013; Leal 2014; Machado *et al.* 2014; Rzepecky 2016). *Dorstenia* has very small unisexual flowers organized and partly embedded in disk-like monoecious inflorescences, except *D. lavrani* which is dioecious (Mccoy and Massara 2008). *Dorstenia* is further characterized by the marginal appendages around the inflorescence (i.e., extensions of the flattened inflorescence receptacle), and the length of these appendages varies among species (Fig. 2 H). Inflorescence shape in *Dorstenia* ranges from square to round. Species of *Dorstenia* are diverse in habit, ranging from treelets to shrubs to herbs and the genus contains almost exclusively all of the herbaceous species known in Moraceae (Berg and Hijman 1999; Berg 2001).

Most species of *Dorstenia* are distributed in either South America or Africa (incl. Madagascar and Arabian Peninsula). *Dorstenia* has been estimated to originate in Africa in the Cretaceous (84.8-132.0 Ma) (Misiewicz and Zerega 2012). However, this estimate is older than the estimate of Moraceae as a whole (Zerega *et al.* 2005) and likely needs to be revised. Misiewicz and Zerega (2012) reconstructed the first phylogeny of *Dorstenia* using

ITS sequences, but some of the deep nodes remained weakly supported. Interestingly, their results suggested a single dispersal into the Neotropics with a subsequent re-colonization of Africa. However, sampling of both ingroup (32 out of 113 species) and outgroups (seven species represented seven genera, including two in the same tribe) was limited and the phylogeny of tribe Dorstenieae as a whole remains incompletely understood.

Methodological approaches used in this thesis

During my PhD program, I took advantage of recent methodological improvements in phylogenetics and macroevolution, including both technical advances (high-throughput sequencing) and conceptual developments in phylogenetic tree reconstruction and dating, model-based biogeographic history reconstruction, and phylogenetic comparative methods.

Hyb-seq approaches

The rapid development of high-throughput sequencing (HTS) has revolutionized phylogenetic analysis in the last decade (Drew *et al.* 2014; Matasci *et al.* 2014; Couvreur *et al.* 2019; Villaverde *et al.* 2018). The five widely used HTS approaches at present are: 1) microfluidic PCR, based on PCR amplification of targeted regions; 2) restriction enzyme-based methods, using restriction enzymes to separate genomic DNA and sequence, for instance, RAD-seq; 3) genome skimming, consisting of total genomic DNA sequencing at low depth without enrichment (typically aimed at recovering organellar genomes); 4) target enrichment approach (incl. the strategy commonly referred to as "exon capture"), enriching shotgun sequencing libraries for target genes; 5) transcriptome sequencing

(McKain *et al.* 2018). The Hyb-seq approach differs from other target enrichment strategies by not only targeting thousands of low-copy nuclear exons, but also their flanking regions, high-copy repeats, and organellar genomes simultaneously (Weitemier *et al.* 2014). One of the advantages of the target enrichment and Hyb-seq approaches is their suitability to work with degraded herbarium specimens (McKain *et al.* 2018). Herbarium collections are a treasure for biodiversity studies, when we consider the information about diversity and distribution from them, and the fact that some species are very difficult to be collected in the field (Hart *et al.* 2016).

Molecular dating

Molecular dating, which transforms the relative time from molecular branch lengths of phylogenetic trees to absolute time, using calibrations typically derived from the fossil record. Molecular dating is essential to connect the evolutionary history of a studied group of organisms with extrinsic geographic and climatic events (Sauquet 2013). Originally developed under the assumption of a strict clock (constant molecular rate) model, current molecular dating methods now allow heterogenous rates (relaxed clock models). Fossil calibrations have traditionally been applied as minimum age constraints. More recently, following the development of Bayesian approaches to molecular dating, various prior distributions (e.g., lognormal, exponential) have been implemented to describe the expected time between the fossil and calibrated node age (Ho and Phillips 2009). Further, to mitigate the impact of potential errors in calibration and the conflicts among calibration nodes, soft boundaries (allowing non-zero probability beyond both minimum and maximum age constraints) were introduced in prior distributions for calibrations (Yang and

Rannala 2005). However, even more recent molecular dating approaches (not used in this thesis), including the fossilized birth-death process and the so-called total-evidence (or tip-dating) methods treat fossils more explicitly as tips (and/or ancestors) of the phylogenetic tree, thereby avoiding the arbitrariness of prior distribution parameterization (Pyron 2011; Ronquist *et al.* 2012; Heath *et al.* 2014).

The development of molecular dating methods in the genomic era has mainly focused on handling the much larger datasets than in the Sanger sequencing era, including the challenge of adequately partitioning such datasets (Foster *et al.* 2016; Foster and Ho 2017). Although widely used software such as BEAST 2 (Bouckaert *et al.* 2014) may still be used in theory to analyze genomic data, computational times are in practice a limiting factor. New software such as MCMCTree in package PAML (Yang 2007) use approximate likelihood to solve this problem. Contrary to some early expectations, phylogenomic data so far appear to have a limited impact on divergence times, whereas fossil calibrations remain the most important factor (Dos Reis *et al.* 2016; Foster and Ho 2017). Thus, robust and well justified fossil calibration remains as critical with phylogenomic datasets as in the Sanger sequencing era (Sauquet *et al.* 2012; Ksepka *et al.* 2015).

Parametric models for ancestral state reconstruction

Ancestral state reconstruction (ASR) helps evolutionary biologists to understand the transition of traits in the evolutionary history of a clade. Parsimony approaches, which find the states that minimize the number of transition events given the states on tips of the phylogeny, were first used in ASR (Maddison *et al.* 1984). However, model-based approaches are now preferred by many, as they consider the probability of transition events

according to the branch lengths and allow to measure the relative confidence of each state at internal nodes of the phylogeny, expressed as proportional marginal likelihoods (Pagel 1999). Asymmetrical (directional) transition can also be estimated with ML approaches, allowing biologists to test more hypotheses in studies of trait evolution. Originally developed under a maximum likelihood (ML) framework, model-based approaches to ASR are also now available in a Bayesian framework. Bayesian approaches to ASR take into account uncertainty of both phylogenies (topologies and branch lengths) and model parameters (Pagel *et al.* 2004; Ronquist 2004). The reversible-jump Markov Chain Monte Carlo approach takes further advantage of the Bayesian framework to explore and visit multiple models of morphological evolution in proportion to their posterior probabilities (Pagel and Meade 2006).

Parametric models for historical biogeography

The reconstruction of ancestral distribution areas is similar to ancestral state reconstruction, and both fields have similar histories of methodological improvement. Historical biogeographic reconstruction started with parsimony-based method, for instance event-based approaches (Ronquist 2003). Inspired by the methods of ancestral state reconstruction and phylogenetic model selection, parametric models were introduced to historical biogeographic reconstruction about ten years ago, for instance the now widely used dispersal-extinction-cladogenesis (DEC) model (Ree and Smith 2008; Ree and Sanmartín 2009). Parametric biogeographic methods not only consider the topology but also the relative or absolute time, which parsimony methods do not. As in ASR models, the amount of biogeographic change (range expansion or local expansion) is proportional to

branch lengths (relative time). Absolute time links the historical biogeographic reconstruction with extrinsic geographic events (e.g., land bridges, fluctuation of sea level, climate in certain area). When multiple parametric models are considered, model selection metrics such as Akaike Information Criterion (AIC) and Bayes Factor (BF) may be used to determine the most appropriate model to explain the data.

Objectives of this thesis

This thesis focuses on breeding system evolution and the origin of diversity of angiosperm in tropical regions by using Moraceae as a model clade. My thesis is composed of three chapters, each written as a standalone manuscript for publication. Chapter I mainly focuses on breeding system evolution in Moraceae as a whole. To tackle this question, a new dated phylogenetic tree and ancestral states of breeding systems were reconstructed in Moraceae and *Ficus*. This chapter was recently published in *Annals of Botany* (Zhang *et al.* 2019a). In Chapter II, I investigated the biogeographic history of Dorstenieae, the second most species-rich tribe of Moraceae. To do so, a new dated phylogenomic tree of tribe Dorstenieae, sampled from herbarium specimens, was reconstructed and used to address the biogeographic history of the group. This chapter was recently made available as a preprint in bioRxiv and is currently being considered for recommendation by *Peer Community in Evolutionary Biology* (Zhang *et al.* 2019b). In addition, this work led to a collaboration on a smaller phylogenomic study focused on the genus *Brosimum*. While this has not yet reached the stage of a draft manuscript, I have outlined my contribution in Supplementary Data part of this thesis. Chapter III focused on characterizing niche differences of the two breeding systems in *Ficus*. To do so, I built a spatial occurrence

dataset for the genus and tested the relationship between breeding systems and climate (incl. precipitation and temperature), taking the phylogeny into account. This last chapter has not yet been submitted to any journal.

LITERATURE CITED

Antonelli A, Sanmartín I. **2011**. Why are there so many plant species in the Neotropics? *Taxon* **60**: 403–414.

Antonelli A, Zizka A, Carvalho FA, *et al.* 2018. Amazonia is the primary source of Neotropical biodiversity. *Proceedings of the National Academy of Sciences* **115**: 6034–6039.

APG IV. **2016**. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society* **181**: 1–20.

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* **112**: 6110–6115.

Berg CC. **2001**. Moreae, Artocarpeae, and *Dorstenia* (Moraceae): with introductions to the family and *Ficus* and with additions and corrections to Flora Neotropica Monograph 7. *Flora Neotropica* **83**: 1-346.

Berg CC. **2003**. Flora Malesiana precursor for the treatment of Moraceae 1: the main subdivision of *Ficus*: the subgenera. *Blumea* **48**: 166–177.

Berg CC, Corner EJH, Jarrett FM. **2006**. Moraceae-genera other than *Ficus*. *Flora Malesiana. Series I, Seed plants* **17**:1-152.

Berg CC, Hijman MEE. **1999**. The genus *Dorstenia* (Moraceae). *Ilicifolia* **2**: 1–211.

Bouckaert R, Heled J, Kühnert D, *et al.* 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* **10**: e1003537.

Bruun-Lund S, Clement WL, Kjellberg F, Rønsted N. **2017**. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Molecular Phylogenetics and Evolution* **109**: 93–104.

De Bruyn M, Stelbrink B, Morley RJ, *et al.* 2014. Borneo and Indochina are major evolutionary hotspots for Southeast Asian biodiversity. *Systematic Biology* **63**: 879–901.

Buerki S, Forest F, Alvarez N. **2014**. Proto-South-East Asia as a trigger of early angiosperm

diversification. *Botanical Journal of the Linnean Society* **174**: 326–333.

Bull JJ, Charnov EL. **1985**. On Irreversible Evolution. *Evolution* **39**: 1149–1155.

Cardoso JCF, Viana ML, Matias R, *et al.* 2018. Towards a unified terminology for angiosperm reproductive systems. *Acta Botanica Brasilica* **32**: 329–348.

Case AL, Barrett SCH. **2004**. Environmental stress and the evolution of dioecy: *Wurmbea dioica* (Colchicaceae) in Western Australia. *Evolutionary Ecology* **18**: 145–164.

Chantarasuwan B, Rønsted N, Kjellberg F, Sungkaew S, van Welzen PC. **2016**. Palaeotropical intercontinental disjunctions revisited using a dated phylogenetic hypothesis with nearly complete species level sampling of *Ficus* subsect. *Urostigma* (Moraceae). *Journal of biogeography* **43**: 384–397.

Charlesworth D. **2015**. Plant contributions to our understanding of sex chromosome evolution. *New Phytologist* **208**: 52–65.

Chase MW, Thijs KW, Kamau P, Fay MF. **2013**. *Dorstenia christenhuszii* (Moraceae), a new species from the Taita Hills, Kenya. *Phytotaxa* **81**: 45–48.

Christenhusz MJM, Byng JW. **2016**. The number of known plants species in the world and its annual increase. *Phytotaxa* **261**: 201–217.

Chung K-F, Kuo W-H, Hsu Y-H, Li Y-H, Rubite RR, Xu W-B. **2017**. Molecular recircumscription of Broussonetia (Moraceae) and the identity and taxonomic status of *B. kaempferi* var. *australis*. *Botanical Studies* **58**: 11.

Clement WL, Weiblen GD. **2009**. Morphological evolution in the mulberry family (Moraceae). *Systematic Botany* **34**: 530–552.

Coiro M, Doyle JA, Hilton J. **2019**. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist*: In press. https://doi.org/10.1111/nph.15708 .

Cook JM, Rasplus J-Y. **2003**. Mutualists with attitude: coevolving fig wasps and figs. *Trends in Ecology & Evolution* **18**: 241–248.

Couvreur TLP. **2015**. Odd man out: why are there fewer plant species in African rain forests? *Plant systematics and evolution* **301**: 1299–1313.

Couvreur TLP, Helmstetter AJ, Koenen EJM, *et al.* 2019. Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science* **9**: 1941.

Crayn DM, Costion C, Harrington MG. **2015**. The Sahul–Sunda floristic exchange: dated molecular phylogenies document Cenozoic intercontinental dispersal dynamics. *Journal of Biogeography* **42**: 11–24.

Cruaud A, Ronsted N, Chantarasuwan B, *et al.* 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Systematic Biology* **61**: 1029–1047.

Datwyler SL, Weiblen GD. **2004**. On the origin of the fig: phylogenetic relationships of Moraceae from *ndhF* sequences. *American Journal of Botany* **91**: 767–777.

Drew BT, Ruhfel BR, Smith SA, *et al.* 2014. Another look at the root of the angiosperms reveals a familiar tale. *Systematic Biology* **63**: 368–382.

Dufaÿ M, Champelovier P, Käfer J, Henry JP, Mousset S, Marais GAB. **2014**. An angiosperm-wide analysis of the gynodioecy–dioecy pathway. *Annals of Botany* **114**: 539–548.

Foster CSP, Ho SYW. **2017**. Strategies for partitioning clock models in phylogenomic dating: application to the angiosperm evolutionary timescale. *Genome biology and evolution* **9**: 2752–2763.

Foster CSP, Ho SYW, Duchêne S. **2016**. Estimating the number and assignment of clock models in analyses of multigene datasets. *Bioinformatics* **32**: 1281–1285.

Gardner EM, Sarraf P, Williams EW, Zerega NJC. **2017**. Phylogeny and biogeography of *Maclura* (Moraceae) and the origin of an anachronistic fruit. *Molecular phylogenetics and evolution* **117**: 49–59.

Ghada B, Ahmed BA, Khaled C, *et al.* 2010. Molecular evolution of chloroplast DNA in fig (*Ficus carica* L.): Footprints of sweep selection and recent expansion. *Biochemical Systematics and Ecology* **38**: 563–575.

Givnish TJ, Zuluaga A, Marques I, *et al.* 2016. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. *Cladistics* **32**: 581–605.

Goldberg EE, Otto SP, Vamosi JC, *et al.* 2017. Macroevolutionary synthesis of flowering plant sexual systems. *Evolution* **71**: 898–912.

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

Hart ML, Forrest LL, Nicholls JA, Kidner CA. **2016**. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* **65**: 1081–1092.

He N, Zhang C, Qi X, *et al.* 2013. Draft genome sequence of the mulberry tree *Morus notabilis*. *Nature communications* **4**: 2445.

Heath TA, Huelsenbeck JP, Stadler T. **2014**. The fossilized birth–death process for coherent calibration of divergence-time estimates. *Proceedings of the National Academy of Sciences* **111**: E2957–E2966.

Heilbuth JC. **2000**. Lower species richness in dioecious clades. *American Naturalist* **156**: 221–241.

Ho SYW, Phillips MJ. **2009**. Accounting for calibration uncertainty in phylogenetic estimation of

evolutionary divergence times. *Systematic Biology* **58**: 367–380.

Hoorn C, Wesselingh FP, Ter Steege H, *et al.* 2010. Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science* **330**: 927–931.

Hughes KA, Houde AE, Price AC, Rodd FH. **2013**. Mating advantage for rare males in wild guppy populations. *Nature* **503**: 108–110.

Jaramillo C, Rueda MJ, Mora G. **2006**. Cenozoic plant diversity in the Neotropics. *Science* **311**: 1893–1896.

Käfer J, de Boer HJ, Mousset S, Kool A, Dufay M, Marais GAB. **2014**. Dioecy is associated with higher diversification rates in flowering plants. *Journal of evolutionary biology* **27**: 1478–1490.

Käfer J, Marais GAB, Pannell JR. **2017**. On the rarity of dioecy in flowering plants. *Molecular Ecology* **26**: 1225–1241.

Knox EB, Li C. **2017**. The East Asian origin of the giant lobelias. *American Journal of Botany* **104**: 924–938.

Kreft H, Jetz W. **2007**. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences* **104**: 5925–5930.

Ksepka DT, Parham JF, Allman JF, *et al.* 2015. The Fossil Calibration Database—A new resource for divergence dating. *Systematic Biology* **64**: 853–859.

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 Phytologist* **214**: 1722–1735.

Leal ME. **2014**. *Dorstenia luamensis* (Moraceae), a new species from eastern Democratic Republic of Congo. *PhytoKeys*: 49–55.

Li J, Murray KG, Li P, Brown K. **2018**. Differential diversifications of South American and Eastern Asian disjunct genera *Bocconia* and *Macleaya* (Papaveraceae). *Journal of Systematics and Evolution* **56**: 25–34.

Linder HP. **2014**. The evolution of African plant diversity. *Frontiers in Ecology and Evolution* **2**: 38.

Machado AFP, Marcelo Filho DMV. **2012**. *Dorstenia romaniucii* (Moraceae), a new species from the Brazilian Atlantic Rain Forest. *Systematic Botany* **37**: 451–455.

Machado AFP, Pereira JF, Carauta JPP. **2014**. *Dorstenia triseriata* (Moraceae) a new and endangered species from Brazil. *PhytoKeys*: 31–35.

Maddison WP, Donoghue MJ, Maddison DR. **1984**. Outgroup analysis and parsimony. *Systematic Biology* **33**: 83–103.

Matasci N, Hung L-H, Yan Z, *et al.* 2014. Data access for the 1,000 Plants (1KP) project. *Gigascience* **3**: 17.

Mccoy T, Massara M. **2008**. *Dorstenia lavrani*: a dioecious new species from northern Somaliland. *Cactus and Succulent Journal* **80**: 78–82.

McKain MR, Johnson MG, Uribe-Convers S, Eaton D, Yang Y. **2018**. Practical considerations for plant phylogenomics. *Applications in plant sciences* **6**: e1038.

Misiewicz TM, Zerega NC. **2012**. Phylogeny, biogeography and character evolution of *Dorstenia* (Moraceae). *Edinburgh Journal of Botany* **69**: 413–440.

Moonlight PW, Richardson JE, Tebbitt MC, *et al.* 2015. Continental-scale diversification patterns in a megadiverse genus: the biogeography of Neotropical *Begonia*. *Journal of Biogeography* **42**: 1137–1149.

Morley RJ. **2003**. Interplate dispersal paths for megathermal angiosperms. *Perspectives in Plant Ecology, Evolution and Systematics* **6**: 5–20.

Mutke J, Barthlott W. **2005**. Patterns of vascular plant diversity at continental to global scales. *Biologiske skrifter* **55**: 521–531.

Neal PR, Anderson GJ. **2005**. Are 'mating systems''breeding systems' of inconsistent and confusing terminology in plant reproductive biology? or is it the other way around? *Plant Systematics and Evolution* **250**: 173–185.

Pagel M. **1999**. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. *Systematic Biology* **48**: 612–622.

Pagel M, Meade A. **2006**. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov Chain Monte Carlo. *The American Naturalist* **167**: 808–825.

Pagel M, Meade A, Barker D. **2004**. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* **53**: 673–684.

Pederneiras LC, Gaglioti AL, Romaniuc-Neto S, Mansano V de F. **2018**. The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). *Botanical Journal of the Linnean Society* **187**: 594–613.

Pyron RA. **2011**. Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology* **60**: 466–481.

Rabosky DL, Chang J, Title PO, *et al.* 2018. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**: 392–395.

Rasplus J-Y, Rodriguez LJ, Tollon-Cordet C, Cruaud A. **2018**. Revisiting the phylogeny of *Ficus* (Moraceae): when next generation sequencing corroborates past generation botanists. *bioRxiv*: 340463.

Ree RH, Sanmartín I. **2009**. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* **36**: 1211–1220.

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

dos Reis M, Donoghue PCJ, Yang Z. **2016**. Bayesian molecular clock dating of species divergences in the genomics era. *Nature Reviews Genetics* **17**: 71–80.

Renner SS. **2014**. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* **101**: 1588–1596.

Rohwer JG, Berg CC. **1993**. Moraceae In: *Flowering Plants Dicotyledons*. Berlin Heidelberg: Springer.

Ronquist F. **2003**. Parsimony analysis of coevolving species associations In: *Tangled trees: phylogeny, cospeciation and coevolution*. The University of Chicago Press Chicago, 22–64.

Ronquist F. **2004**. Bayesian inference of character evolution. *Trends in Ecology & Evolution* **19**: 475–481.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. **2012**. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.

Rzepecky A. **2016**. *Dorstenia horwoodii* Rzepecky sp. nov. from Nudum to Novum, a Fortyish Year Hiatus. *Cactus and Succulent Journal* **88**: 66–74.

Sabath N, Goldberg EE, Glick L, *et al.* 2016. Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. *New Phytologist* **209**: 1290–1300.

dos Santos A, Machado AFP, Mansano VDEF, Neto SR. **2013**. *Dorstenia acangatara* (Moraceae), a new and threatened species from Southeastern Brazil. *Phytotaxa* **118**: 29–34.

dos Santos A, Neto SR. **2012**. A new species of *Dorstenia* (Moraceae) from southeastern Brazil. *PhytoKeys*: 47–51.

Sauquet H. **2013**. A practical guide to molecular dating. *Comptes Rendus Palevol* **12**: 355–367.

Sauquet H, von Balthazar M, Magallón S, *et al.* 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* **8**: 16047.

Sauquet H, Ho SYW, Gandolfo MA, *et al.* 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology* **61**: 289–313.

Shanahan M, So S, Compton SG, Corlett R. **2001**. Fig-eating by vertebrate frugivores: a global

review. *Biological Reviews* **76**: 529–572.

Simpson MG. **2010**. *Plant systematics*. Cambridge: Academic press.

Soltis DE, Smith SA, Cellinese N, *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98**: 704–730.

Ulloa Ulloa C, Acevedo-Rodríguez P, Beck S, *et al.* 2017. An integrated assessment of the vascular plant species of the Americas. *Science* **358**: 1614–1617.

Villaverde T, Pokorny L, Olsson S, *et al.* 2018. Bridging the micro-and macroevolutionary levels in phylogenomics: Hyb-Seq solves relationships from populations to species and above. *New Phytologist* **220**: 636–650.

Weitemier K, Straub SCK, Cronn RC, *et al.* 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**: 1400042.

Wiens JJ, Donoghue MJ. **2004**. Historical biogeography, ecology and species richness. *Trends in Ecology Evolution* **19**: 639–644.

Williams E W, Gardner EM, Harris R, Chaveerach A, Pereira JT, Zerega NJC. **2017**. Out of Borneo: biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany* **119**: 611–627.

Williams Evelyn W, Gardner EM, Harris III R, Chaveerach A, Pereira JT, Zerega NJC. **2017**. Out of Borneo: biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany* **119**: 611–627.

Yang Z. **2007**. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**: 1586–1591.

Yang Z, Rannala B. **2005**. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212–226.

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.

Zapata-Sudo G, Mendes TCF, Kartnaller MA, *et al.* 2010. Sedative and anticonvulsant activities of methanol extract of *Dorstenia arifolia* in mice. *Journal of ethnopharmacology* **130**: 9–12.

Zemp N, Widmer A, Charlesworth D. **2018**. Has adaptation occurred in males and females since separate sexes evolved in the plant *Silene latifolia*? *Proceedings of the Royal Society B: Biological Sciences* **285**: 20172824.

Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. **2005**. Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution* **37**: 402–

416.

Zerega NC, Gardner EM. **2019**. Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. *Phytotaxa* **338**: 253–265.

Zhang Q, Onstein RE, Little SA, Sauquet H. **2019a**. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. *Annals of Botany* **123**: 191–204.

Zhang Q, Gardner E, Zerega N, Sauquet H. **2019b**. Long-distance dispersal shaped the diversity of tribe Dorstenieae (Moraceae). *bioRxiv*: 531855.

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.

CHAPTER I. ESTIMATING DIVERGENCE TIMES AND ANCESTRAL BREEDING SYSTEMS IN *FICUS* AND MORACEAE

This chapter was recently published in *Annals of Botany* in January 2019 (https://academic.oup.com/aob/article/123/1/191/5092035), in collaboration with Stefan Little, Renske Onstein and Hervé Sauquet.

ABSTRACT

- **Background and Aims** Although dioecy, which characterizes only 6% of angiosperm species, has been considered an evolutionary dead end, recent studies have demonstrated that this is not necessarily the case. Moraceae (40 genera, 1100 spp., incl. *Ficus*, 750 spp.) are particularly diverse in breeding systems (incl. monoecy, gynodioecy, androdioecy, and dioecy) and thus represent a model clade to study macroevolution of dioecy.
- **Methods** Ancestral breeding systems of *Ficus* and Moraceae were inferred. To do so, a new dated phylogenetic trees of *Ficus* and Moraceae was first reconstructed by combining a revised twelve-fossil calibration set and a densely sampled molecular dataset of eight markers and 320 species. Breeding system evolution was then reconstructed using both parsimony and model-based (maximum likelihood and Bayesian) approaches with this new timescale.

- **Key Results** The crown-group ages of *Ficus* and Moraceae were estimated in the Eocene (40.6-55.9 Ma) and Late Cretaceous (73.2-84.7 Ma), respectively. Strong support was found for ancestral dioecy in Moraceae. Although the ancestral state of *Ficus* remained particularly sensitive to model selection, the results show that monoecy and gynodioecy evolved from dioecy in Moraceae, and suggest that gynodioecy probably evolved from monoecy in *Ficus*.
- **Conclusions** Dioecy was found not to be an evolutionary dead end in Moraceae. This study provides a new time scale for the phylogeny and a new framework of breeding system evolution in *Ficus* and Moraceae.

Key words: breeding system evolution, dioecy, evolutionary dead end, molecular dating, ancestral state reconstruction, fossil calibration, *Ficus*, Moraceae

INTRODUCTION

Flowers are the reproductive structures of angiosperms. The ca. 350,000 species of angiosperms are highly diverse in floral morphological traits, including breeding systems, ranging from bisexuality (hermaphroditism) to separate sex on distinct plants (dioecy) through several presumed intermediate states between these two ends (monoecy, andromonoecy, gynomonoecy, androdioecy, gynodioecy) (Renner 2014). Although a recent study demonstrated that bisexual flowers are ancestral in angiosperms and evolved many times independently to unisexual flowers (Sauquet *et al.* 2017), the exact number and the context of these transitions remains to be characterized. Dioecy is rare in angiosperms (only ca. 6%) (Renner 2014) and has been suggested to be an evolutionary dead end (Bull and Charnov 1985; Heilbuth 2000). Transitions from hermaphroditism to dioecy have been suggested to occur through three alternative pathways: dimorphic pathway (through gynodioecy, androdioecy, or polygamodioecy) (Dufay *et al.* 2014) ; monomorphic pathway (through monoecy, andromonoecy, gynomonoecy, or polygamomonoecy) (Renner and Ricklefs 1995); and direct pathway (Barrett 2002; Goldberg *et al.* 2017). However, the view of dioecy as an evolutionary dead end has gradually changed over the last decade after it was found that transitions from dioecy to other breeding systems are possible and that the flexibility of breeding systems transitions may be caused by different selective pressures and constraints (Barrett 2013; Goldberg *et al.* 2017). Thus, it remains unclear why dioecy is in fact so rare in angiosperms (Käfer *et al.* 2017).

The mulberry family (Moraceae), consists of approximately 40 genera and 1100 species, and represents a good model system to study dioecy transitions (Clement and

Weiblen 2009). Four breeding systems are observed in the family: monoecy, androdioecy, gynodioecy, and dioecy (Datwyler and Weiblen 2004). *Ficus*, the largest genus in Moraceae, contains almost 75% of the species in the family (750 spp.). About half of the species of *Ficus* are monoecious, while the other half are gynodioecious (Cook and Rasplus 2003). In monoecious species of *Ficus,* both functionally male and female flowers coexist in the same inflorescence (syconium); female flowers vary in style length. Fig wasp pollinators (Agaonidae) lay eggs in part of these female flowers and pollinate the others (Cook and Rasplus 2003). Gynodioecious species of *Ficus* are characterized by two kinds of plants and inflorescences: female individuals bear syconia that contain only functionally female flowers (i.e., flowers that can develop seeds), and functionally male individuals bear syconia that contain both functionally male and gall flowers (female flowers that develop only wasps) (Cook and Rasplus 2003). Therefore, structurally gynodioecious species of figs are functionally dioecious.

A previous study, based on parsimony reconstruction sampling 46 *Ficus* species (but no outgroup), suggested that monoecy is ancestral in *Ficus*, and that gynodioecy originated at least once or twice within the genus (Weiblen 2000). A subsequent family-level parsimony study, sampling 83 Moraceae species (incl. 11 species of *Ficus*) suggested monoecy and dioecy to be ancestral in *Ficus* and Moraceae, respectively (Datwyler and Weiblen 2004). However, the accuracy and precision of ancestral state reconstruction depends on the reliability of the phylogenetic tree and sampling density, and important progress has been made to reconstruct phylogenetic relationships in *Ficus* and Moraceae since these two studies were published (Zerega *et al.* 2005; Clement and Weiblen 2009; Zerega *et al.* 2010; Cruaud *et al.* 2012; Williams *et al.* 2017). In addition, new probabilistic

approaches for ancestral state reconstruction, taking into account phylogenetic uncertainty and divergence times, are now available (Pagel *et al.* 2004). Parsimony, maximum likelihood (ML) and Bayesian approaches today are the most commonly used methods for reconstructing ancestral states. The accuracy of branch length estimation may be important in model-based approaches (ML, Bayesian), which rely on them for computing the likelihood of evolutionary change along the phylogeny (Pagel and Meade 2006). However, to date only few attempts have been made to estimate divergence times in Moraceae (Datwyler and Weiblen 2004; Zerega *et al.* 2005). Although more studies have addressed the timescale of *Ficus* evolution (Rønsted *et al.* 2005; Zerega *et al.* 2005; Cruaud *et al.* 2012), partly inconsistent results have been obtained, for instance with crown-group age estimates for *Ficus* ranging from 40.1 to 101.9 Mya. In addition to providing a framework to reconstruct ancestral states, estimating divergence times in *Ficus* and Moraceae is also important for other evolutionary questions, including the evaluation of the co-diversification and biogeographic history of the genus (Cruaud *et al.* 2012).

Here, we investigate two key questions: 1) what are the crown-group ages of *Ficus* and Moraceae?; and 2) what are the ancestral breeding systems of *Ficus* and Moraceae, and how many times did monoecy and dioecy evolve in *Ficus* and Moraceae? To do so, we reconstruct a new phylogenetic tree for Moraceae using a densely sampled molecular dataset, estimate divergence times using a relaxed molecular clock calibrated with a revised set of 12 fossil age constraints, and reconstruct ancestral breeding systems in *Ficus* and Moraceae with state-of-the-art model-based approaches. We estimate the age of crown group of *Ficus* and Moraceae in the Eocene and Late Cretaceous, respectively. Our results suggest that dioecy is not an evolutionary dead end in Moraceae and the transitions from

dioecy to the other breeding systems (androdioecy, gynodioecy and monoecy) occurred several times during the evolutionary history of the family. This study sheds light on the evolution of dioecy into other breeding systems and more generally improves our understanding of breeding system evolution in angiosperms.

MATERIALS AND METHODS

Taxonomic sampling and molecular dataset assembly

We selected GenBank sequences from 320 species belonging to 65 genera and eight families of Rosales, including 272 species and 36 genera (3/4 of the circa 40 genera) of Moraceae (Table S1). Our efforts include a comprehensive sample of outgroups in order to estimate accurately divergence times in the family while taking advantage of the rich fossil record of the order (see below). We used eight markers from two genomes: three chloroplast genes (*matK*, *rbcL*, *ndhF*) and five nuclear markers including two noncoding regions (ITS, ETS) and three genes (*G3pdh*, *ncpGS*, the *waxy* region). This combination of coding and non-coding markers was selected to resolve both deep- and shallow-level relationships. *MatK* and *rbcL* have been proposed as standard barcoding regions for land plants (Hollingsworth *et al.* 2009) and *ndhF* has been proved to be very useful in previous phylogenetic studies of Moraceae (Datwyler and Weiblen 2004; Zerega *et al.* 2005; Clement and Weiblen 2009). ITS, ETS, *G3pdh*, *ncpGS*, and the *waxy* region were used in the latest and most densely sampled phylogenetic study of *Ficus* (Cruaud *et al.* 2012). Considering the comparatively fast molecular rates of ITS and ETS, these markers may be difficult to align and may introduce more noise than signal at the family level. Therefore,

ITS and ETS sequences were here only used for species of *Ficus*. All sequences were aligned using Muscle v3.7 (Edgar 2004) as implemented on CIPRES (Miller *et al.* 2010); alignments were then checked and adjusted by hand. In the ITS alignment we deleted three short regions (ca. 10 bp) because of the uncertainty of gap length and position of base pairs in these regions. Combined multi-marker alignments were assembled with Mesquite v3.04 (Maddison and Maddison 2016).

Phylogenetic reconstruction

To identify and exclude problematic GenBank sequences (e.g., incorrectly identified), separate gene analyses were conducted first with RAxML v8.2.9 (Stamatakis 2014) on CIPRES (Miller *et al.* 2010), using the GTRCAT model and 100 bootstrap replicates. We then compared single-marker phylogenetic trees with each other and with the latest published phylogenies of *Ficus* (Cruaud *et al.* 2012), Moraceae (Clement and Weiblen 2009), and other families of Rosales (Wiegrefe *et al.* 1998; Potter *et al.* 2007; Zhang *et al.* 2011; Wu *et al.* 2013; Yang *et al.* 2013; Onstein *et al.* 2015; Hauenschild *et al.* 2016). A sequence was excluded when its position in one tree conflicted strongly (i.e., with high support) with those in other trees or published phylogenies. After excluding problematic sequences (*Celtis*: L12638, AY263941, AY263961, AY263925, AY263899; *Morus*: L01933) and confirming that no supported conflict remained among single-marker trees, combined phylogenetic analyses of chloroplast genes (*matK*, *rbcL*, *ndhF*), nuclear markers (ITS, ETS, *G3pdh*, *ncpGS*, the *waxy* region), and of all eight markers were conducted with maximum likelihood and Bayesian approaches using RAxML and BEAST on CIPRES.

For maximum likelihood analyses, the dataset was divided into eight partitions according to marker, each with the GTRCAT model and 1000 bootstrap replicates. All trees are presented rooted on the most external outgroup, Rosaceae, which have been shown to be the sister group of all remaining Rosales in all recent higher-level phylogenetic analyses (e.g., Wang *et al.* 2009; Soltis *et al.* 2011). The phylogenetic reconstruction by Bayesian approach was done simultaneously with divergence time estimation (see below).

Fossil calibration

The fossil record of Moraceae is comparatively poor and in need of critical revision (Collinson 1989), whereas unambiguous fossils exist for other families of Rosales (Burge and Manchester 2008; Benedict *et al.* 2011; Friis *et al.* 2011). When reliable fossils are absent or scarce for the ingroup, outgroup calibration may provide more accurate estimates than secondary calibration (Sauquet *et al.* 2012; Hipsley and Müller 2014). Therefore, we specifically designed the taxon sample of this study to include sufficient outgroup nodes to take advantage of the fossil record of Rosales. Our set of fossil age constraints includes four in Moraceae and eight distributed among the remaining families of Rosales (Table S2). To revise fossil calibrations in Rosales, we proceeded as follows (Parham *et al.* 2012; Sauquet *et al.* 2012): 1) we started from lists of calibrations used in previous molecular dating studies (Zerega *et al.* 2005; Cruaud *et al.* 2012; Magallón *et al.* 2015) and completed this list with specific reviews (Collinson 1989; Friis *et al.* 2011) and recently published fossil taxa (Manchester 1999; Calvillo-Canadell and Cevallos-Ferriz 2007; Manos *et al.* 2007; Benedict *et al.* 2011); 2) for each fossil, we critically assessed the phylogenetic

assignment based on original descriptions and subsequent reviews, and using the latest reference phylogeny for each family (Manchester 1999; Calvillo-Canadell and Cevallos-Ferriz 2007; Manos *et al.* 2007; Benedict *et al.* 2011; Friis *et al.* 2011; Jud *et al.* 2017); 3) for each fossil, we also critically revised the absolute age or age range of the fossil using the latest stratigraphy and geological time scale from the International Commission on Stratigraphy (Cohen *et al.* 2017). Because none of these fossil taxa have been included in total evidence phylogenetic analyses, our assignments here are at best 'apomorphy-based' and therefore we have been particularly conservative in both selecting our final calibrations and assigning them to clades. This implies that some taxa previously used as calibrations are typically used here to constrain the age of a more inclusive node than in previous studies (e.g., the fossil achenes we used to calibrate the stem group of *Ficus* were used to calibrated the crown group of *Ficus* in former studies; Datwyler and Weiblen 2004; Rønsted *et al.* 2005; Zerega *et al.* 2005; Cruaud *et al.* 2012).

Molecular dating analyses

We used BEAST v1.8.0 (Drummond *et al.* 2012) implemented on CIPRES (Miller *et al.* 2010) to estimate divergence times and topology simultaneously. To reduce computational burden, we used empirical base frequencies and divided the dataset into three partitions: chloroplast DNA (*matK*, *rbcL*, *ndhF*), noncoding nuclear markers (ITS, ETS), and coding nuclear genes (*G3pdh*, *ncpGS*, the *waxy* region). The substitution model for each partition was selected using the Akaike Information Criterion (AIC) with jModelTest v2.1.6 (Darriba *et al.* 2012) as implemented on CIPRES. The best substitution

model was TVM+I+G for the chloroplast (*matK*, *rbcL*, *ndhF*) and noncoding nuclear (ITS, ETS) partitions, and TPM3uf+I+G for the coding nuclear (*G3pdh*, *ncpGS*, the *waxy* region) partition.

For each calibrated node, we chose the oldest reliable fossil as a minimum age constraint. Because most fossils typically provide only minimum ages for the clade to which they can be safely attributed, a uniform prior distribution was used for each fossil calibration, using the upper (younger) boundary of the fossil oldest stratigraphic age range as the minimum bound and the maximum constraint for the root (see below) as the maximum bound. Molecular dating analyses usually require at least one maximum age constraint (Ho and Phillips 2009; Sauquet 2013). Here, we chose to set the maximum age bound for the root (i.e., the crown-group node of Rosales) to 107 Ma based on the crown-group age estimate of Rosales (98.96-106.94 Ma) in the latest, large-scale molecular dating study of angiosperms (Magallón *et al.* 2015). Although the phylogenetic relationships among sampled members of Rhamnaceae are inconsistent with previous work (Onstein *et al.* 2015; Hauenschild *et al.* 2016; see Discussion), they are unlikely to have affected divergence times because we chose conservative fossil assignments in Rhamnaceae. *Paliurus clarnensis* has been proposed to be more closely related to extant species of *Paliurus* than any other member of Rhamnaceae (Burge and Manchester 2008), thus providing a minimum age for the stem node of *Paliurus*. Because of conflict in the position of the genus, we decided to calibrate the crown group node of Rhamnaceae instead. *Coahuilanthus belindae* has been proposed to be a member of Rhamnaceae, but its position in the phylogeny remains unclear (Calvillo-Canadell and Cevallos-Ferriz 2007), therefore we used it as a minimum age constraint for the stem node of the family.

The tree prior in BEAST was set as a Birth-Death process. To produce a starting tree that conforms with the hard age constraints enforced in this analysis, we transformed the best-known RAxML tree into an ultrametric tree using the bladj function of phylocom v4.2 (Webb *et al.* 2008). Four separate runs were conducted, each with 100 million generations, sampling trees and parameters every ten thousand generations. Chain convergence was checked in Tracer v1.5 (Rambaut and Drummond 2009), with the first 10% of chain length excluded as burn-in. After confirming that chains of these four runs had converged, we combined them using LogCombiner v 1.8.0, resampling every 100 thousand generations and discarding the first 10% (i.e., 10 million) generations as burn-in. Tree statistics were summarized on the Maximum Clade Credibility (MCC) tree by TreeAnnotator v1.8.0, using median ages as node heights.

In addition to our main dating analysis, we also conducted a series of sensitivity experiments. First, we tested the impact of calibrations on the estimated topology by conducting another run keeping the maximum age constraint on the root but without internal age constraints (Sauquet *et al.* 2012). Second, we produced another run by excluding the age constraint on the stem node of *Ficus*, but maintaining all other 11 age constraints to test the influence of the former on the estimated crown-group age of *Ficus*. Last, we tested the impact of our root maximum age constraint on divergence time estimates for the ingroup. Indeed, there remains considerable uncertainty on the crown-group age of angiosperms (Doyle 2012; Magallón *et al.* 2015; Herendeen *et al.* 2017) and recent analyses that do not constrain this age consistently estimate it to be much older than the first accepted crown-group fossils (Foster *et al.* 2017). Therefore, we also re-ran our analysis using an older maximum age constraint on the root, set to 112 Ma, based on the

crown-group age of Rosales (87.82-111.78 Ma) estimated in the unconstrained analysis of Magallón *et al.* (2015), in which the crown-group age of angiosperms was estimated to be 160-256 Ma rather than 136-140 Ma (S. Magallón, pers. comm.).

Ancestral state reconstruction

The breeding system state of each species was determined from the literature, and then recorded in the PROTEUS database (Sauquet 2016). A list of all data records (each linked to an explicit reference) and the matrix are provided as Table S3. For the main analysis, we reconstructed the ancestral state with all 320 species in the dataset (i.e., including all outgroups). To characterize a reliable evolutionary scenario supported by different classification methods of breeding systems, (e.g., Anger *et al.* 2017), we analyzed six configurations of the same character and data, differing in the number of character states considered (Table 1; Fig. 1). Configuration A distinguishes among five breeding system states (as recorded and listed in Table S3). Configuration B (four states) treats androdioecy as missing data. In our dataset, there are only three androdioecious species, accounting for less than 1% of all species sampled. Configuration C (three states) is similar to Configuration B, but with gynodioecy and dioecy pooled as a single state (dioecy). In *Ficus*, gynodioecious species are functionally dioecious (Cook and Rasplus 2003), and all gynodioecious species but one (*Dryas octopetala*) in this study belong to *Ficus*. Configuration D (three states) is similar to Configuration B, but treats bisexual flowers as missing data. In Rosales, families Cannabaceae, Moraceae and Urticaceae all have unisexual flowers (Simpson 2010), thus maintaining the bisexual state is not essential to

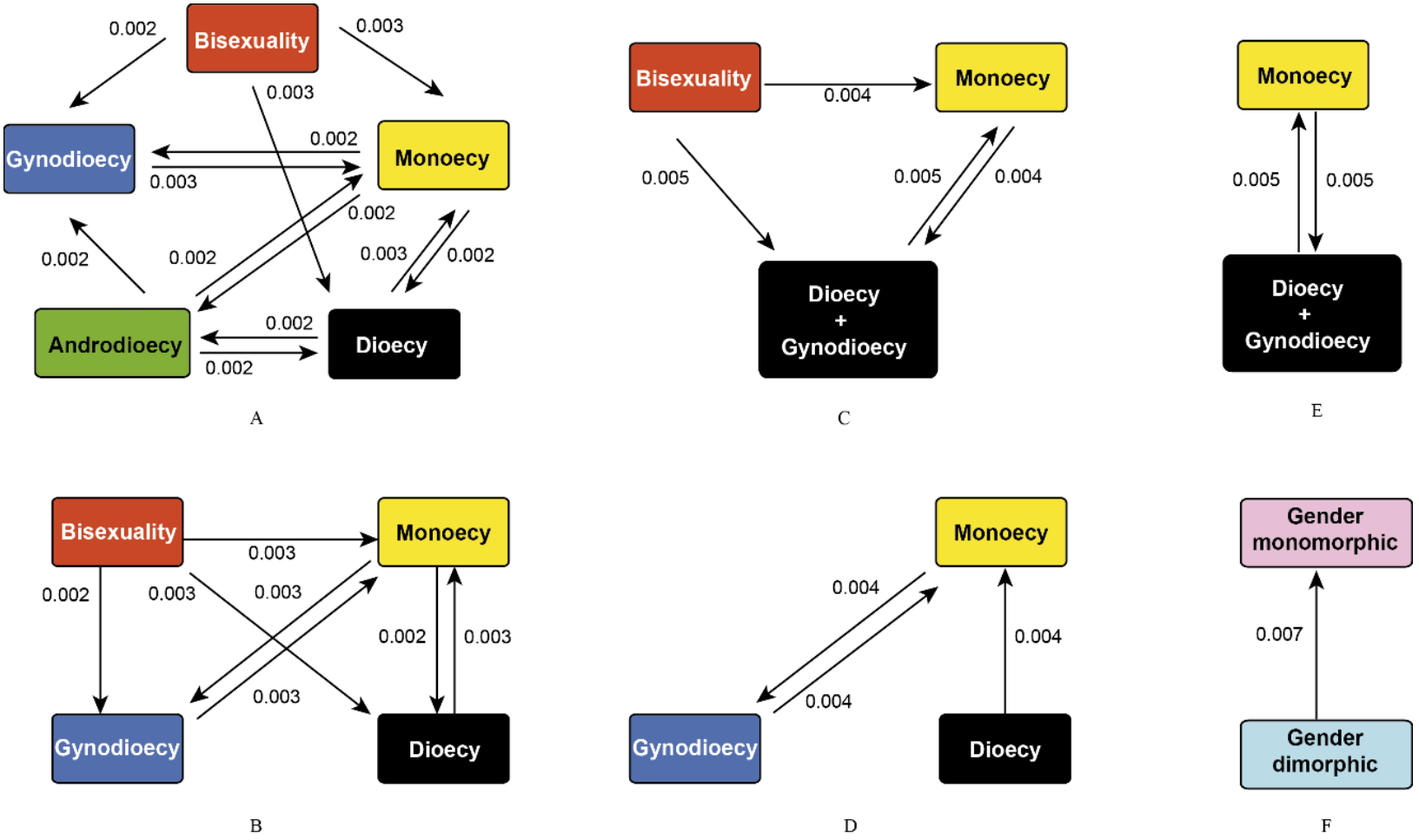


Figure 1. Schematic representation of the six configurations for the breeding system character analyzed in this study. The arrows denote transition rate parameters, as estimated in the reversible-jump Markov Chain Monte Carlo analyses, with mean rates reported. The absence of an arrow indicates a near-zero estimate for the corresponding parameter, suggesting the model does not support direct transition between the two states.

this study and could be detrimental by unnecessarily increasing the number of parameters to estimate. Configuration E (two states) combines the two transformations of Configurations C and D: gynodioecy and dioecy pooled, bisexuality treated as missing. In Configuration F (two states), we contrast gender monomorphic and gender dimorphic species. Gender monomorphic and gender dimorphic means that in one population there are one or two functional classes of sex (Lloyd 1980). Therefore bisexuality and monoecy are gender monomorphic and androdioecy, gynodioecy, and dioecy are gender dimorphic.

For each configuration, we used and compared parsimony, maximum likelihood (ML) and Bayesian approaches to reconstruct ancestral states (Sauquet *et al.* 2015; Sauquet *et al.* 2017). Parsimony analyses were conducted in Mesquite v3.04 (Maddison and Maddison 2016), ML and Bayesian analyses were conducted in BayesTraits V2 (Pagel and Meade 2013). Parsimony and ML analyses were conducted with the Maximum Clade Credibility (MCC) tree produced from the BEAST analysis, whereas Bayesian analyses were conducted with 3600 trees randomly sampled from the posterior of the BEAST analysis. To test the influence of topological uncertainty on ancestral state reconstruction, we also conducted additional Bayesian analyses with a fixed (MCC) tree.

Maximum likelihood analyses presented here explored two models for each configuration: an equal-rates model (ER, or Mk1), assuming equal transition rates among all character states (Lewis 2001), and an all-rates-different model (ARD), allowing distinct (asymmetric) transition rates among character states (Pagel 1994). The best model (equal or unequal rate) for each configuration was selected according to the Akaike Information Criterion (AIC). We calculated ΔAIC between two models for each configuration, using ΔAIC of 2 or more as a criterion for positive support of the best-fit model (Posada and

Buckley 2004).

Table 1. Summary of the six character state configurations used to reconstruct ancestral breeding systems in Moraceae.

Configuration	States	Justification
A	Bisexuality, gynodioecy, androdioecy, monoecy, dioecy	Original states recorded in species
B	Bisexuality, gynodioecy, monoecy, dioecy	Excluding state androdioecy for its low frequency in the dataset of this study
C	Bisexuality, monoecy, dioecy + gynodioecy	Based on configuration B, gynodioecy combined with dioecy: all gynodioecious species (but one) in this study belong to *Ficus*, where they are functionally dioecious
D	Gynodioecy, monoecy, dioecy	Based on configuration B, excluding state bisexuality: only distant outgroups of Moraceae include species with bisexual flowers (e.g., Rosaceae, Elaeagnaceae, Rhamnaceae)
E	Monoecy, dioecy + gynodioecy	Based on configuration B, gynodioecy combined with dioecy (as in C) and bisexuality excluded (as in D)
F	Gender monomorphic, gender dimorphic	Recognition of bisexuality and monoecy as gender monomorphic, and gynodioecy, androdioecy and dioecy as gender dimorphic (Lloyd 1980)

We also reconstructed ancestral breeding system states using both a "common" (i.e., fixed-model) Bayesian approach and a reversible-jump Markov chain Monte Carlo (rjMCMC) strategy. While both allowed us to take parameter and phylogenetic (incl. molecular dating) uncertainty into account in ancestral state reconstruction (Pagel *et al.* 2004), the rjMCMC approach allowed us to explore and visit multiple models of morphological evolution in proportion to their posterior probabilities (Pagel and Meade 2006). Both the equal-rates (ER) and the all-rates-different (ARD) model were tested in

the common Bayesian approach, and their relative fit was compared with the Bayes Factor (Kass and Raftery 1995) according to the criteria of Lodewyckx *et al.* (2011). Chain lengths were set to ten million generations (or two million generations for fixed-tree analyses) and parameters and ancestral states were sampled every 1000 generations. Chain convergence was checked in Tracer v1.5 (Rambaut and Drummond 2009), with the first 10% generations excluded as burn-in.

To investigate the impact of different taxonomic sampling strategies (with or without outgroups of *Ficus*) on ancestral state reconstruction of breeding systems in *Ficus*, we also conducted analyses with only *Ficus* species (same topology). Here, after excluding non-*Ficus* species in configuration A, we used Mesquite instead of BayesTraits for ML reconstructions because we wished to test the impact of different root states priors on ancestral state reconstruction. We applied both the equal-rates (Mk1) and the asymmetric 2-rate (AsymmMk, equivalent to ARD for a binary character) models to reconstruct the ancestral state of *Ficus*. In addition, two different root states priors (equal and equilibrium frequencies) were applied for the asymmetric 2-rate (AsymmMk) model in Mesquite (whereas BayesTraits assumes an equal frequency root state prior for all states).

RESULTS

Phylogenetic reconstruction

Chloroplast markers were insufficient to resolve relationships within *Ficus*, as most internal nodes had low support (Fig. S1A-D). We found three instances of conflict (*F. trigonata* clustered with *F. tinctoria*; *F. religiosa* clustered with *F. benghalensis*; *F. pumila*

clustered with *F. hirta*, bootstrap probability 79, 80, and 57) between topologies reconstructed from nuclear and chloroplast sequences (Fig. S1 D and E). Except these, we found no well supported conflict between chloroplast and nuclear trees (Fig. S1). We thus combined the eight markers and focus here on the results from this combined analysis (Fig. 2 and S2). Hereafter we only discuss the trees reconstructed from the whole dataset. The topologies reconstructed with maximum likelihood (ML) and Bayesian approaches were consistent (Fig. S1 M and Fig. S2). Differences were only observed in some weakly supported nodes. All families, all tribes of Moraceae except Moreae, and all genera were supported as monophyletic. Tribe Moreae was reconstructed as polyphyletic because *Streblus smithii* clustered with Maclureae (the other two species of *Streblus* sampled were instead found to be nested in Moreae; Fig. S2). In *Ficus*, subgenera *Phamacosycea*, *Urostigma*, and *Ficus* were found to be paraphyletic, and subgenus *Synoecia* nested in subgenus *Ficus*. In addition, some deep nodes (e.g., the most recent common ancestor of *F. carica* and *Urostigma*) were poorly supported in the Bayesian analysis (Fig. S2). With or without fossil age constraints, the topologies obtained with BEAST were identical and the posterior probability values of each nodes were similar as well (result not shown). Excluding the crown-group calibration of *Ficus* or using an older maximum root age constraint did not influence the topology either (Table S4).

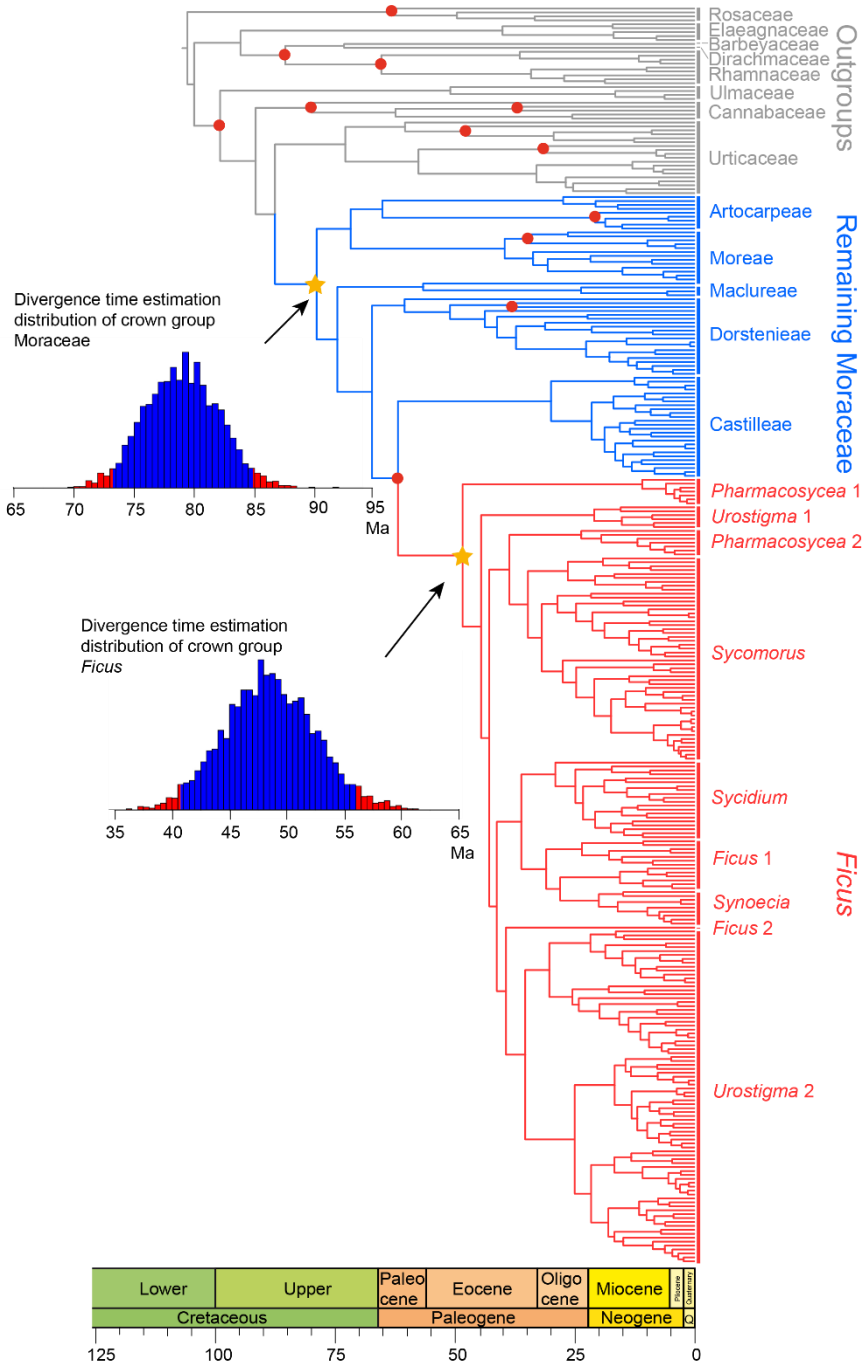


Figure 2. Dated phylogenetic tree of 320 species with family names in Rosales, tribe names in Moraceae, and section names in *Ficus*. This is the Maximum Clade Credibility (MCC) from the BEAST analysis of eight molecular markers (see text). Fossil-calibrated nodes are indicated in red. The posterior distribution of estimated ages is shown for crown-group Moraceae and *Ficus* (marked with orange star). For full details of this tree, see Fig. S2.

Molecular dating

After combination of the four separate runs, the effective sample size (ESS) of each parameter was over 100, most of them over 200. The crown-group ages of *Ficus* and Moraceae were estimated asEocene (40.6-55.9 Ma) and Late Cretaceous (73.2-84.7 Ma), respectively (Table 2; Fig. 2). The oldest tribe was Artocarpeae (64.0-68.6 Ma) and Maclureae originated most recently (8.9-41.1 Ma). In *Ficus*, subgenus *Sycomorus* was the oldest (28.0-41.1Ma) (Table 2). Excluding the *Ficus* stem-group calibration or using an older maximum root age constraint did not have strong impact on estimated ages (Table S4).

Ancestral state reconstruction

Parsimony, maximum likelihood (ML) and Bayesian analyses of the full dataset (320 species) reconstruct dioecy as ancestral for Moraceae with strong support in all six character state configurations (Table 3, Fig. 3, Fig. S3 and S4). Furthermore, we find that dioecy has evolved into monoecy at least five times in Moraceae. However, several alternative scenarios are possible due to uncertainty in the most recent common ancestor of Dorstenieae, Castilleae, and Ficeae (*Ficus*) (Fig. 3). In one possible scenario, monoecy evolved independently once in Artocarpeae (*Artocarpus*); once in Moreae (*Morus*); once in Dorstenieae (with at least two subsequent reversals); twice in Castilleae (*Perebea humilis*; and the clade of *Antiaris toxicaria* and *Mesogyne insignis*); and at least once in Ficeae (*Ficus*). In another possible scenario, monoecy evolved independently in Artocarpeae and Moreae (as above), but shares a common origin in the clade of Dorstenieae,

Castilleae, and *Ficus*, with at least two reversals (to dioecy) in Dorstenieae, and one in the ancestor of Castilleae (followed with two subsequent gains again in the tribe as above). Within *Ficus*, no matter whether gynodioecy (functional dioecy) is ancestral (and therefore intermediate) or not (see below), at least three transitions from gynodioecy to monoecy are reconstructed (within section *Sycomorus*).

Table 2. Estimated divergence times for key nodes of Moraceae. Here, we follow the tribal classification of Clement and Weiblen (2009), as shown in Fig 1.

Nodes	**Breeding systems**	**Support values (BP/PP)[1]**	**Mean age (95% HPD)[2] (Ma)**
CG[3] Rosales	Bisexuality, monoecy, androdioecy, gynodioecy, dioecy	/	105.5 (102.8-107.0)
SG[4] Moraceae	/	100/1.00	87.5 (81.7-93.3)
CG Moraceae	Monoecy, androdioecy, gynodioecy, dioecy	99/1.00	79.0 (73.2-84.7)
CG Artocarpeae	Monoecy, dioecy	99/1.00	65.6 (64.0-68.6)
CG Moreae[5]	Monoecy, dioecy	91/1.00	40.3 (34.9-46.8)
CG Maclureae	Dioecy	98/1.00	24.7 (8.9-41.1)
CG Dorstenieae	Monoecy, dioecy	98/1.00	60.5 (51.4-70.2)
CG Castilleae	Monoecy, androdioecy, dioecy	100/1.00	31.2 (18.7-47.0)
SG Ficeae	/	98/1.00	62.2 (56.0-68.6)
CG Ficeae	Monoecy, gynodioecy	99/1.00	48.5 (40.6-55.9)
CG *Pharmacosycea* 1	Monoecy	58/1.00	11.6 (5.5-19.4)
CG *Pharmacosycea* 2	Monoecy	100/1.00	24.2 (13.2-36.7)
CG *Sycomorus*	Monoecy, gynodioecy	77/1.00	38.7 (28.0-41.1)
CG *Sycidium*	Gynodioecy	100/1.00	29.2 (22.6-36.4)
CG *Synoecia*	Gynodioecy	100/1.00	20.1 (13.7-26.8)
CG *Ficus*	Gynodioecy	82/1.00	31.0 (24.7-38.0)
CG *Urostigma* 1	Monoecy	100/1.00	21.5 (12.0-31.2)
CG *Urostigma* 2	Monoecy	99/1.00	35.5 (28.0-42.3)

[1]PP: posterior probability; BP: bootstrap probability; [2]95% HPD: 95% highest posterior density; [3]CG: crown group; [4]SG: stem group; [5]*Streblus smithii* was here excluded from crown-group Moreae (see text for details).

Conversely, the ancestral breeding system of *Ficus* could not be reconstructed with confidence. First, parsimony reconstructions suggested either monoecy (configurations A, B, and D) or an equivocal state (monoecy or the combination of dioecy and gynodioecy; configurations C, E, and F) as ancestral in *Ficus* (Fig S4). Second, ML reconstructions with an equal-rate model with all configurations supported monoecy (or gender monomorphic) as ancestral in *Ficus*, whereas the unequal-rate model supported gynodioecy (or gynodioecy + dioecy or gender dimorphic) as ancestral in the genus (Table 3). The strength of support differed among different configurations but the ancestral state did not change. In the ML approach, we compared the models in each configuration using the Akaike Information Criterion (AIC). Only with configuration A was the equal-rate model selected, whereas the unequal rate model better fit the data with configurations B-E. As a result, considering best-fit models only, gynodioecy (or gynodioecy + dioecy, or gender dimorphic) was found as ancestral in *Ficus* with five out of six configurations (Table 3). Third, the Bayesian approach with reversible-jump Markov chain Monte Carlo (rjMCMC) weakly supported monoecy (or gender monomorphic) as ancestral in *Ficus*, but with a broad 95% highest posterior density (HPD; 0.00 to 1.00), suggesting high uncertainty in the estimation (Table 4). Bayesian analyses with fixed model selected the unequal rate model in 4 out of 6 configurations, two with weak support), and monoecy as ancestral in *Ficus* in 4 out of 6 configurations.

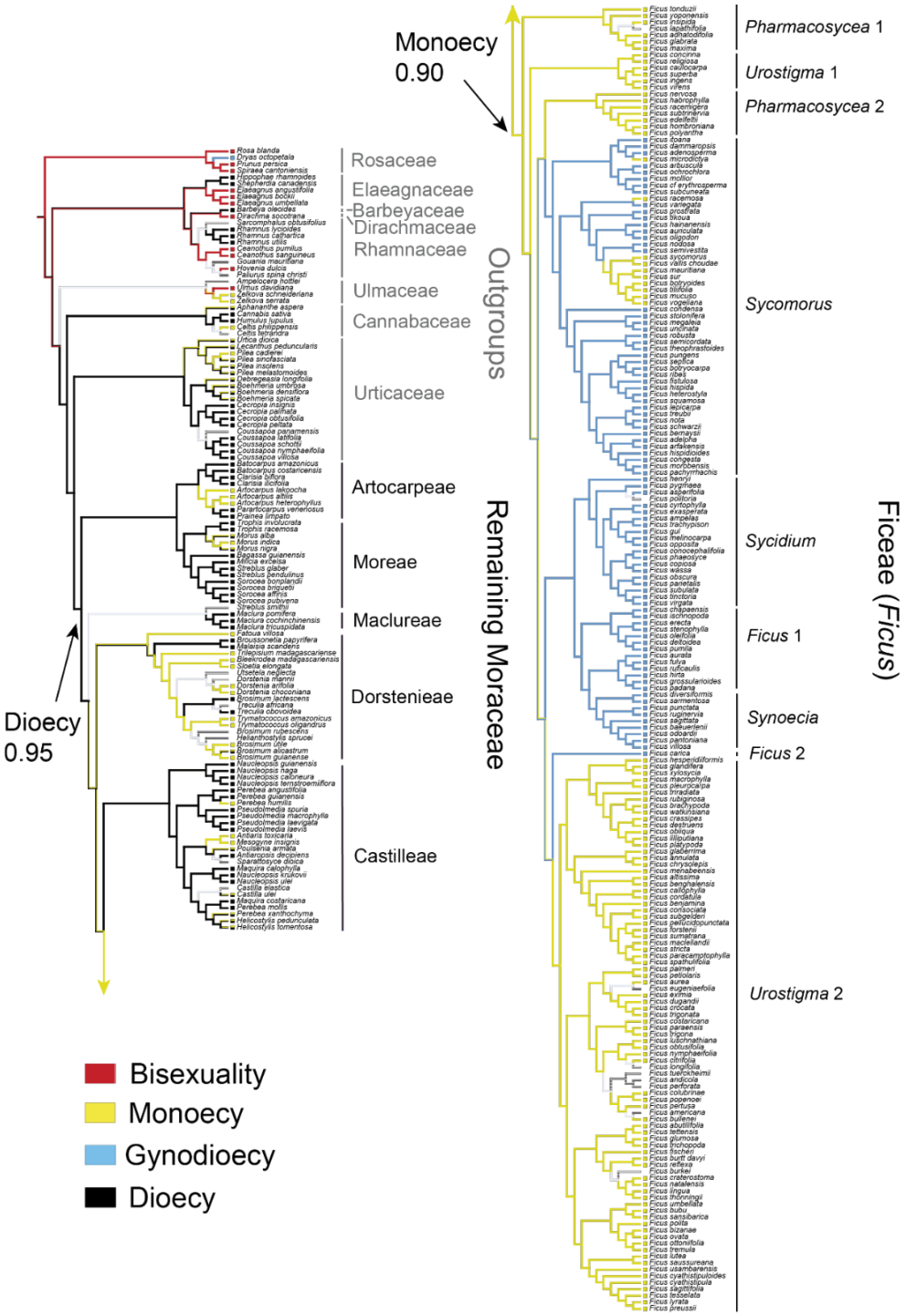


Figure 3. Ancestral state reconstruction of breeding systems in Moraceae. Here we show results from the maximum likelihood analysis of configuration A (five character states) with the equal-rate model (for other results, see Figs. S3-S5). Clade labels: family names in Rosales, tribe names in Moraceae, and section names in *Ficus*.

Table 3. Summary results of ancestral state reconstruction for the complete (320 species) dataset by maximum likelihood approach.

	CG Moraceae[1]	Prob[2]	CG *Ficus*[3]	prob	-Lh[4]	Nr of par [5]	AIC
Equal rate							
A	Dioecy	0.91	Monoecy	0.92	**121.6**	10	263.2
B	Dioecy	0.90	Monoecy	0.91	102.7	6	217.4
C	Gynodioecy + dioecy	0.95	Monoecy	0.87	91.5	3	189
D	Dioecy	0.88	Monoecy	0.89	78.9	3	163.8
E	Gynodioecy + dioecy	0.92	Monoecy	0.80	63.1	1	128.2
F	Gender dimorphic	0.89	Gender monomorphic	0.76	76.1	1	154.2
Unequal rate							
A	Androdieocy; dioecy; gynodioecy	Each ca. 0.33	Gynodioecy	0.95	124.7	20	289.4
B	Dioecy	0.99	Gynodioecy	0.98	**90.8**	12	205.6
C	Gynodioecy + dioecy	0.99	Gynodioecy + dioecy	0.84	**76.9**	6	165.8
D	Dioecy	1.00	Gynodioecy	1.00	**70.0**	6	152
E	Gynodioecy + dioecy	1.00	Gynodioecy + dioecy	0.83	61.6	2	127.2
F	Gender dimorphic	1.00	Gender dimorphic	0.84	74.6	2	153.2

[1]CG Moraceae: breeding system state of crown-group Moraceae; [2]Prob: probability of support estimated by maximum likelihood; [3]CG *Ficus*: breeding system state of crown-group *Ficus*; [4]Lh: minus loglikelihood; [5]Nr of par: number of free parameters; The best-fit model identified by AIC for each configuration is indicated in boldface.

Some of the transitions such as the direct transition between gynodioecy and dioecy in configurations A, B or D were estimated to be unlikely (i.e., transition rates were estimated to be zero) in the rjMCMC analyses (Fig 2). The 95% HPD of transition rates in each configuration broadly overlapped. In the Bayesian analysis with the unequal rate model in configuration A (Table S5), the chain did not converge. This remained true when fixing the tree (i.e. using the MCC tree) instead of using a sample of trees from the BEAST posterior.

These results suggest that the unequal rate model is overparameterized for this configuration with five character states. Indeed, twenty free parameters (transition rates) are estimated in this model. In contrast, a four-state unequal rate model requires 12 parameters, a three-state model requires six parameters, and a binary model requires only two parameters. The potential overparameterization of configuration A was also suggested by the results of reversible-jump Markov Chain Monte Carlo (rjMCMC) analyses. For each configuration, the number of free parameters in the model was estimated between one and three.

Table 4. Summary results of ancestral state reconstruction for the complete (320 species) dataset obtained with the Bayesian approach (reversible-jump MCMC).

	One MCC tree				**3600 posterior trees**			
Configuration	**CG Moraceae[1]**	**95% HPD[2]**	**CG *Ficus*[3]**	**95% HPD**	**CG Moraceae**	**95% HPD**	**CG *Ficus***	**95% HPD**
A	Dioecy	0.82-1.00	Monoecy	0.00-0.88	Dioecy	0.80-1.00	Monoecy	0-0.99
B	Dioecy	0.84-1.00	Monoecy	0.71-0.90	Dioecy	0.80-1.00	Monoecy	0.71-1.00
C	Gynodioecy + dioecy	0.86-1.00	Monoecy	0.00-0.84	Gynodioecy + dioecy	0.74-1.00	Monoecy	0.00-0.97
D	Dioecy	0.85-1.00	Monoecy	0.66-0.90	Dioecy	0.82-1.00	Monoecy	0.65-1.00
E	Gynodioecy + dioecy	0.87-1.00	Monoecy	0.00-0.85	Gynodioecy + dioecy	0.75-1.00	Monoecy	0.00-0.98
F	Gender dimorphic	0.85-1.00	Gender dimorphic	0.19-1.00	Gender dimorphic	0.76-1.00	Gender Dimorphic	0.03-1.00

[1]CG Moraceae: breeding system state of crown-group Moraceae; [2]95% HPD: 95% highest posterior density; [3]CG *Ficus*: breeding system state of crown-group *Ficus.*

Reconstructions with only species of *Ficus* also showed an uncertain result with respect to the ancestral state of the genus (Fig. S5). Both parsimony and ML with the equal-rate model strongly supported monoecy to be ancestral in *Ficus*, whereas the ML approach with the unequal rate model supported gynodioecy as ancestral in the genus. The support for ancestral gynodioecy differed with root state prior: with the equal frequencies root state prior, gynodioecy was highly supported (0.99 out of 1) whereas with the equilibrium frequencies root state prior, the probability for gynodioecy decreased sharply to 0.56, suggesting high uncertainty.

DISCUSSION

Phylogenetic relationships in Moraceae

The general topology of phylogenetic trees reconstructed in this study (both maximum likelihood and Bayesian approach) is consistent with previous studies in *Ficus* (Cruaud *et al.* 2012), Moraceae (Clement and Weiblen 2009) and Rosales (Zhang *et al.* 2011), except for relationships within Rhamnaceae, for which our results partly differ from those of two recent phylogenetic studies (Onstein *et al.* 2015; Hauenschild *et al.* 2016). This conflict may be due to the use of different molecular markers and to our sampling of this outgroup family with low density. Here we used *matK*, *rbcL*, and *ndhF*, while Hauenschild *et al.* (2016) used ITS and *trnL-F*, and Onstein *et al.* (2015) used eight molecular markers (including all the markers we used and *trnL-F*, *psbA*, *psbA-trnH*, *rpl16* and ITS). However, these differences are unlikely to affect our ancestral state reconstructions, given the distant positions of Rhamnaceae and Moraceae in the phylogeny of Rosales, and the fact that the

breeding system state of most of the species in Rhamnaceae has been scored as missing data in this study. In our phylogenetic reconstruction, outgroup relationships are consistent with previous work (Wang *et al.* 2009; Soltis *et al.* 2011; Zhang *et al.* 2011).

Within Moraceae, phylogenetic relationships among genera are consistent with recent studies (Zerega *et al.* 2005; Clement and Weiblen 2009; Zerega *et al.* 2010; Williams *et al.* 2017) except for Moreae (Fig. 2 and Fig. S2). Tribe Moreae was found to be paraphyletic due to the position of *Streblus smithii* as sister to Maclureae. This relationship is not strongly supported (posterior probability 0.76) and may be caused by too few informative sites for this species. In our dataset, *S. smithii* is only represented by the *ndhF* sequence and our separate analysis of the *ndhF* dataset, in which all three species of *Streblus* were sampled, supports the same result as our combined analysis (Fig. S1C). In the original paper where the sequence came from (Datwyler and Weiblen 2004), the species was found in an uncertain phylogenetic position. However, in a later reconstruction (Clement and Weiblen 2009) where *ndhF*, 26S, and morphological data were combined, *S. smithii* clustered with the other two species in the genus, possibly because the nuclear (26S) and morphological signal overcame a divergent *ndhF* sequence in their combined analysis.

Within *Ficus*, subgenera *Phamacosycea*, *Urostigma* and *Ficus* were found to be paraphyletic, and subgenus *Synoecia* is nested within subgenus *Ficus* and some deep nodes remain poorly supported. These results are consistent with the most recent comprehensive phylogenetic study of the genus (Cruaud *et al.* 2012), while they are in conflict with a recent phylogenomic study of *Ficus* based on full chloroplast genomes and this conflict could be caused by potential cyto-nuclear discordance (Bruun-Lund *et al.* 2017). Chloroplast markers performed poorly in *Ficus* as the informative sites they provide are

too few (Cruaud *et al.* 2012). Almost no branch length can be observed in the chloroplast tree for the shallow nodes (Fig. S1). In our analysis, *Ficus* species were represented not only by chloroplast markers but also by nuclear ones which were found informative enough in former studies (Rønsted *et al.* 2008; Cruaud *et al.* 2012). In the Bayesian analyses, 3600 trees were taken into account, therefore the uncertainty of topology was considered.

A new time scale for Moraceae diversification

The new time scale presented here for Moraceae was estimated with 12 fossil calibrations (four ingroup, eight outgroup), more than in any previous study of the family so far. Here the crown-group age of Moraceae was estimated as 73.2-84.7 Ma. This estimate is similar to that of Zerega *et al.* (2005) (72.6-110.0 Ma), but with a narrower confidence interval. Our stem-group age estimate for the family (81.7-93.3 Ma) is older than that reported by Magallón *et al.* (2015) (57.7-77.8 Ma), most likely because our increased sampling of the family and its outgroups allowed us to use more fossil calibrations in Rosales. The crown-group age of *Ficus* was here estimated as 40.6-55.9 Ma, which is similar to the age reported by Zerega *et al.* (2005) (40.1-51.0 Ma), but younger than the age found by Rønsted *et al.* (2005) (51.4-78.0 Ma) and Cruaud *et al.* (2012) (60.0-101.9 Ma). These differences may be explained by different calibration strategies. When fossil species have similar morphological traits with extant species of a clade, these fossils are often used optimistically to calibrate the crown group of this clade. This practice is problematic because it rules out the possibility that such fossils are stem relatives of the clade (Ronquist *et al.* 2012; Sauquet 2013). Here we used an apomorphy-based approach, using fossils as minimum age constraints modeled with uniform prior distributions to

calibrate the stem node of the clade with which they share apomorphies (Sauquet 2013) (Table S2). For instance, the fossil achenes attributed to *Ficus* (Collinson 1989) were here used to calibrate the stem node of *Ficus*, whereas Rønsted *et al.* (2005) and Cruaud *et al.* (2012) used the same fossils to calibrate the crown node of *Ficus.* We also used different prior distributions for fossil calibrations. Although some authors have suggested that priors such as lognormal or exponential distributions may help to improve the accuracy of divergence time estimation (Yang and Rannala 2005; Ho and Phillips 2009), the prior setting of parameters such as mean or standard deviation for these distributions is often arbitrary and the shape of these distributions implicitly assumes that the fossil diverged close to the node calibrated. Our strategy here was instead to use uniform distribution with fossil ages as the minimum bound and maximum root age as the maximum bound. The crown-group ages of tribes Artocarpeae and Dorstenieae were estimated as 64.0-68.6 Ma and 51.4-70.2 Ma, respectively, both of which are younger than the ages reported by Williams *et al.* (2017) for Artocarpeae (61.4-78.5 Ma) and Misiewicz and Zerega (2012) for *Dorstenia* (84.8-132.0 Ma).

BEAST analyses without internal calibrations showed identical topology and similar posterior support compared to our reference analysis using all calibrations, suggesting that here calibration did not influence the topology. Here, we used the maximum age estimated for crown-group Rosales by Magallón *et al.* (2015) as the maximum age constraint on the root in our analyses, but also tested the influence of older root constraints. Consistent with previous work (Sauquet *et al.* 2009; Massoni *et al.*2014; Foster *et al.* 2017), we found that alternative maximum root constraints had very little impact on estimated divergence times. In addition, the crown-group age of *Ficus* was similar with or without the stem-group *Ficus*

calibration (Table S4). These results suggest that our estimates are not an artifact of the calibration points near the nodes of interest.

Breeding system transitions in Moraceae and Ficus

It was found that increased taxon sampling density can be helpful in ancestral state reconstruction (Salisbury and Kim 2001). Ancestral states of breeding systems in Moraceae have been previously reconstructed with parsimony and a phylogenetic tree of 46 species (Weiblen 2000), 83 species (Datwyler and Weiblen 2004) and 73 species (Clement and Weiblen 2009). In this study, we reconstructed ancestral states with a more densely sampled phylogenetic tree (incl. 200 species of *Ficus* and 72 species from other genera of Moraceae), compared different model-based approaches, and took into account phylogenetic uncertainty in our Bayesian analyses. Here, we found similar results for two focal nodes (Moraceae and *Ficus*) whether using a single tree or thousands of trees sampled from the posterior of BEAST analyses, suggesting that, in our analyses, topology did not affect the estimation. In all the approaches and models, dioecy was strongly supported to be ancestral for Moraceae, while the ancestral state for *Ficus* remains unclear.

Dioecy has been reported to correlate with several ecological traits, including abiotic pollination, fleshy fruits, and woody growth form (Freeman *et al.* 1979; Renner and Ricklefs 1995; Vamosi *et al.* 2003). During the history of Moraceae, dioecy evolved into monoecy at least five times (see above; Fig. 3, Fig. S3 and S4). Although a previous study suggested no strong statistical support for a relationship between breeding system and pollination syndrome in Moraceae (Clement and Weiblen 2009), our results reinforce the general pattern that would be consistent with such a relationship: dioecious taxa tend to be

wind-pollinated, whereas monoecious taxa tend to be insect-pollinated. For instance, we infer Artocarpeae as ancestrally dioecious (Fig. 3, Fig. S3 and S4) and Datwyler and Weiblen (2004) reconstructed the clade as ancestrally anemophilous. In our reconstruction, dioecy has evolved to monoecy in *Artocarpus*, which is characterized by inflorescences with an insect pollination syndrome. This relationship may also apply to tribe Dorstenieae as a whole, most species of which are monoecious and were hypothesized to be insect pollinated according to their inflorescence structure (Berg and Hijman 1999). However, the relationship becomes less clear in Castilleae and *Ficus* (Datwyler & Weiblen 2004), partly because all species of *Ficus* are insect-pollinated (Cook and Rasplus 2003), yet only half of them are monoecious. Unfortunately, the current lack of sufficient data on actual pollination modes in Moraceae (outside *Ficus*) precludes us from further testing this potential correlation at the family level.

Although the ancestral state of *Ficus* proved to be particularly difficult to reconstruct in this study due to our exploration of various models and character state configurations, most of the evidence supports monoecy rather than gynodioecy as the ancestral state in the genus. As highlighted in recent studies of floral traits using the same methods (Sauquet *et al.* 2015, 2017), our results suggest that great caution should be exercised when interpreting results from ML analyses exploring a limited set of models. Indeed, the models with highest posterior probability identified through our reversible-jump MCMC analyses corresponded to neither the equal rates or the unequal rates models exlpored with ML. These best-fit models typically involved only one free parameter, but excluded some transitions (Fig. 1). In addition, Bayesian analyses present the advantage of taking into account phylogenetic and branch length uncertainty (Pagel *et al.* 2004). All our Bayesian analyses that took into

account phylogenetic and branch length uncertainty suggested monoecy as ancestral in *Ficus*, whether models were allowed to vary (Table 4) or were fixed (Table S5), except for configuration F. In addition, the model-averaged rates from the rjMCMC analyses of the three configurations where dioecy and gynodioecy were treated as separate states (A, B, F) were estimated to zero, suggesting that direct transitions between the two states do not occur in Moraceae. Therefore, we argue it is more probable that monoecy is ancestral in *Ficus* and represents an intermediate state between dioecy and gynodioecy in the genus (Fig. 3). From a functional point of view, it also seems more likely that the highly specialized pollination association between *Ficus* and fig wasps started in monoecious figs.

Assuming that monoecy is ancestral in *Ficus*, the phylogenetic distribution of monoecy and gynodioecy in the genus and the uncertainty remaining in the ancestral states of several deep nodes allow for various scenarios (Figs. 3, S3-S5). It is possible that gynodioecy evolved only once (after divergence of the *Pharmacosycea* 1 and *Urostigma* 1 lineages), followed by two main reversals to monoecy (*Pharmacosycea* 2 and *Urostigma* 2). Alternatively, gynodioecy may have evolved twice independently (once in *Sycomorus* and once in the ancestral lineage of *Sycidium*, *Ficus* 1, and *Synoecia*). In all scenarios, gynodioecy reverted to monoecy at least three times within *Sycomorus* (Fig. 3). The biological reasons for these fluctuations between monoecy and gynodioecy in *Ficus* remain unclear, but several hypotheses have been proposed. Gynodioecy in *Ficus* may be linked with an adaptation to seasonality (Kjellberg *et al.* 1987); reduction of non-pollinating wasps (Kerdelhué and Rasplus 1996); persistence of pollinator populations when the host fig populations are small (Kameyama *et al.* 1999); seed protection (Greeff and Compton 2002); disperser selection; de-coupling of wasp and seed size; survival from predation;

non-pollinators and predator satiation; and chronic pollinator shortages, crop asynchrony, and inbreeding depression (Harrison and Yamamura 2003). In addition, breeding system evolution in *Ficus* may have a relationship with biogeographic distribution. For instance, there appear to be no gynodioecious species of figs in South America (Cruaud *et al.* 2012). Different climates and habitats in different continents may have accelerated breeding system evolution in *Ficus*. The genus has been inferred to have originated in Eurasia, then migrated and diversified in South America during the Miocene (Cruaud *et al.* 2012). Migration to a new continent may have led to breeding system reversal from gynodioecy to monoecy in the clade *Urostigma* 2 (Fig. 3, Fig. S3 and S4) under the scenario where gynodioecy originated only once in *Ficus*.

Dioecy not an evolutionary dead end in Moraceae

Our results provide additional strong support for dioecy to be ancestral in Moraceae as a whole, regardless of the approach or model used for reconstructing ancestral states (Fig3, Table 3 and 4), consistent with previous work (Weiblen 2000; Datwyler and Weiblen 2004; Clement and Weiblen 2009). Furthermore, our extensive sample of outgroups shows that dioecy in Moraceae is ancestral to a larger clade including at least Cannabaceae, Urticaceae, and Moraceae (Fig. 3).

These results add to a growing number of studies that have challenged the notion that dioecy is an irreversible trait in flowering plants. Dioecy was once suggested to be evolutionary dead end (Bull and Charnov 1985). It is found in only approximately 6% of angiosperms and distributed widely in around 43% families of flowering plants (Renner 2014). However, “evolutionary dead end” has been used to describe different patterns of

macroevolution (Bromham *et al.* 2015), including: 1) irreversible evolution (Bull and Charnov 1985; Glémin and Muyle 2014); or 2) a state that presents more long-term disadvantages than advantages compared to other states, ultimately leading to probable extinction (Käfer *et al.* 2014; Käfer *et al.* 2017). Dioecy used to be thought as irreversible (Bull and Charnov 1985) and high extinction rates were found on dioecious clades by comparative approach (Heilbuth 2000). However, it has been known for a long time that although dioecious angiosperms are exposed to a higher risk of failure to find a sex partner, they are also advantageous in having obligate cross-fertilization (Darwin 1876). Various studies have shown that dioecious species may include individuals with other breeding systems and that dioecy in general may be more labile than previously thought (Pannell 1997; Korpelainen 1998; Ainsworth 2000; Käfer et al. 2017). These transitions could be linked to ecological processes including changing plant-pollinator relationships and environments (Case *et al.* 2008); long distance dispersal (Schaefer and Renner 2010); hybridization and polyploidy (Obbard *et al.* 2006). In addition, various studies have documented clades where dioecy is the ancestral rather than the derived state, challenging the assumption of irreversibility of dioecy evolution. For instance, dioecy has been reconstructed as ancestral in Myristicaceae, with at least four subsequent shifts to monoecy in the family (Sauquet 2003) and appears to be a transition state between monoecy and polygamy in Arecaceae (Nadot *et al.* 2016). A recent analysis of macroevolutionary dynamics of breeding system in 40 angiosperm genera supported transitions away from dioecy in *Bursera* (Burseraceae) and *Dodonaea* (Sapindaceae; Goldberg *et al.* 2017). Furthermore, two recent angiosperm-wide analyses showed that dioecy does not have a negative relationship with diversity: dioecy was shown to correlate with an increased

diversification rate (Käfer *et al.* 2014) or not to have a significant relationship with the latter (Sabath *et al.* 2016).

Conclusion

Here we reconstructed divergence times and ancestral breeding systems in Moraceae. With 12 fossil calibrations, the crown group ages of *Ficus* and Moraceae were estimated as 40.6-55.9 Ma (Eocene) and 73.2-84.7 Ma (Late Cretaceous). Dioecy was supported as the ancestral state of Moraceae and we showed that it evolved into monoecy at least five times in the family, and subsequently reversed in some clades (e.g. Artocarpeae, Dorstenieae). In *Ficus*, which represents 75% of species in Moraceae, the ancestral state was estimated to be monoecy with moderate support and at least one transition from monoecy to gynodioecy followed by at least three reversals were estimated. In future work, to investigate the exact breeding system evolutionary scenario of *Ficus*, it would be important to sample basal lineages of *Ficus* more densely and improve phylogenetic resolution of the backbone of the genus. The new time scale of Moraceae and *Ficus* we provide here will also be useful for future analyses of biogeography and co-diversification in the family. Finally, our results lend further support to the growing idea that dioecy does not always represent an evolutionary dead end, shedding light on the understanding of breeding system evolution in angiosperms more generally.

FUNDING

This work was supported by a China Scholarship Council PhD grant [grant number

201506140077] to Q.Z.

ACKNOWLEDGEMENTS

We thank Jos Käfer for his constructive review of this paper; Sophie Nadot for comments on an earlier draft; Selena Smith for comments on fossils of *Artocarpus;* Susana Magallón for discussion of divergence time estimation; and Nyree Zerega and Elliot Gardner for discussion of breeding system evolution in Moraceae.

LITERATURE CITED

Ainsworth C. 2000. Boys and girls come out to play: The molecular biology of dioecious plants. Annals of Botany **86**: 211–221**.**

Anger N, Fogliani B, Scutt CP, Gâteblé G. 2017. Dioecy in *Amborella trichopoda*: evidence for genetically based sex determination and its consequences for inferences of the breeding system in early angiosperms. *Annals of Botany* **119**: 591–597.

Barrett SCH. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics* **3**: 274–284.

Barrett SCH. 2013. The evolution of plant reproductive systems: how often are transitions irreversible? *Proceedings of the Royal Society of London B: Biological Sciences* **280**: 20130913.

Benedict JC, DeVore ML, Pigg KB. 2011. *Prunus* and *Oemleria* (Rosaceae) flowers from the Late Early Eocene Republic Flora of Northeastern Washington State, U.S.A. *International Journal of Plant Sciences* **172**: 948–958.

Berg CC, Hijman MEE. 1999. The genus *Dorstenia* (Moraceae). *Ilicifolia* **2**: 1–211.

Bromham L, Hua X, Cardillo M. 2015. Detecting macroevolutionary self-destruction from phylogenies. *Systematic Biology* **65**: 109–127.

Bruun-Lund S, Clement WL, Kjellberg F, Rønsted N. 2017. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Molecular Phylogenetics and Evolution* **109**: 93–104.

Bull JJ, Charnov EL. 1985. On Irreversible Evolution. *Evolution* **39**: 1149–1155.

Burge DO, Manchester SR. 2008. Fruit morphology, fossil history, and biogeography of *Paliurus* (Rhamnaceae). *International Journal of Plant Sciences* **169**: 1066–1085.

Calvillo-Canadell L, Cevallos-Ferriz SRS. 2007. Reproductive structures of Rhamnaceae from the Cerro del Pueblo (Late Cretaceous, Coahuila) and Coatzingo (Oligocene, Puebla) Formations, Mexico. *American Journal of Botany* **94**: 1658–1669.

Case AL, Graham SW, Macfarlane TD, Barrett SCH. 2008. A phylogenetic study of evolutionary transitions in sexual systems in Australasian *Wurmbea* (Colchicaceae). *International Journal of Plant Sciences* **169**: 141–156.

Clement WL, Weiblen GD. 2009. Morphological evolution in the mulberry family (Moraceae). *Systematic Botany* **34**: 530–552.

Cohen KM, Finney SC, Gibbard PL, Fan J-X. 2017. The ICS International Chronostratigraphic Chart (2013; updated). *Episodes* **36**: 199–204.

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' Hammelidae.* Oxford: Clarendon Press, 319–339.

Cook JM, Rasplus J-Y. **2003**. Mutualists with attitude: coevolving fig wasps and figs. *Trends in Ecology & Evolution* **18**: 241–248.

Cruaud A, Ronsted N, Chantarasuwan B, *et al.* 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Systematic Biology* **61**: 1029–1047.

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

Darwin C. **1876**. *The effects of cross and self fertilisation in the vegetable kingdom*. London: J. Murray.

Datwyler SL, Weiblen GD. **2004**. On the origin of the fig: phylogenetic relationships of Moraceae from *ndhF* sequences. *American Journal of Botany* **91**: 767–777.

Doyle JA. **2012**. Molecular and fossil evidence on the origin of angiosperms. *Annual Review of Earth and Planetary Sciences* **40**: 301–326.

Drummond AJ, Suchard MA, Xie D, Rambaut A. **2012**. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.

Dufaÿ M, Champelovier P, Käfer J, Henry JP, Mousset S, Marais GAB. **2014**. An angiosperm-wide analysis of the gynodioecy–dioecy pathway. *Annals of Botany* **114**: 539–548.

Edgar RC. **2004**. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Foster CSP, Sauquet H, van der Merwe M, McPherson H, Rossetto M, Ho SYW. **2017**. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Systematic Biology* **66**: 338–351.

Freeman DC, Harper KT, Kent Ostler W. **1979**. Ecology of plant dioecy in the intermountain region of Western North America and California. *Oecologia* **44**: 410–417.

Friis EM, Crane PR, Pedersen KR. **2011**. *Early flowers and angiosperm evolution*. Cambridge: Cambridge University Press.

Glémin S, Muyle A. **2014**. Mating systems and selection efficacy: a test using chloroplastic sequence data in angiosperms. *Journal of Evolutionary Biology* **27**: 1386–1399.

Goldberg EE, Otto SP, Vamosi JC, *et al.* 2017. Macroevolutionary synthesis of flowering plant sexual systems. *Evolution* **71**: 898–912.

Greeff JM, Compton SG. **2002**. Can seed protection lead to dioecy in *Ficus*? *Oikos* **96**: 386–388.

Harrison RD, Yamamura N. **2003**. A few more hypotheses for the evolution of dioecy in figs (*Ficus*, Moraceae). *Oikos* **100**: 628–635.

Hauenschild F, Matuszak S, Muellner-Riehl AN, Favre A. **2016**. Phylogenetic relationships

within the cosmopolitan buckthorn family (Rhamnaceae) support the resurrection of *Sarcomphalus* and the description of *Pseudoziziphus* gen. nov. *Taxon* **65**: 47–64.

Heilbuth JC. **2000**. Lower species richness in dioecious clades. *The American Naturalist* **156**: 221–241.

Herendeen PS, Friis EM, Pedersen KR, Crane PR. **2017**. Palaeobotanical redux: revisiting the age of the angiosperms. *Nature Plants* **3**: 17015.

Hipsley CA, Müller J. **2014**. Beyond fossil calibrations: realities of molecular clock practices in evolutionary biology. *Frontiers in Genetics* **5**.

Ho SYW, Phillips MJ. **2009**. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**: 367–380.

Hollingsworth PM, Forrest LL, Spouge JL, *et al.* 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 12794–12797.

Jud NA, Gandolfo MA, Iglesias A, Wilf P. **2017**. Flowering after disaster: Early Danian buckthorn (Rhamnaceae) flowers and leaves from Patagonia. *Plos One* **12**.

Käfer J, de Boer HJ, Mousset S, Kool A, Dufay M, Marais GAB. **2014**. Dioecy is associated with higher diversification rates in flowering plants. *Journal of Evolutionary Biology* **27**: 1478–1490.

Käfer J, Marais GAB, Pannell JR. **2017**. On the rarity of dioecy in flowering plants. *Molecular Ecology* **26**: 1225–1241.

Kameyama T, Harrison R, Yamamura N. **1999**. Persistence of a fig wasp population and evolution of dioecy in figs: a simulation study. *Researches on population ecology* **41**: 243–252.

Kass RE, Raftery AE. **1995**. Bayes Factors. *Journal of the American Statistical Association* **90**: 773–795.

Kerdelhué C, Rasplus J-Y. **1996**. The evolution of dioecy among *Ficus* (Moraceae): an alternative hypothesis involving non-pollinating fig wasp pressure on the fig-pollinator mutualism. *Oikos* **77**: 163–166.

Kjellberg F, Gouyon P-H, Ibrahim M, Raymond M, Valdeyron G. **1987**. The stability of the symbiosis between dioecious figs and their pollinators: a study of *Ficus carica* L. and *Blastophaga psenes* L. *Evolution* **41**: 693–704.

Korpelainen H. **1998**. Labile sex expression in plants. *Biological Reviews* **73**: 157–180.

Lewis PO. **2001**. A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology* **50**: 913–925.

Lloyd DG. **1980**. Sexual strategies in plants III. A quantitative method for describing the gender of plants. *New Zealand Journal of Botany* **18**: 103–108.

Lodewyckx T, Kim W, Lee MD, Tuerlinckx F, Kuppens P, Wagenmakers E-J. **2011**. A

tutorial on Bayes factor estimation with the product space method. *Journal of Mathematical Psychology* **55**: 331–347.

Maddison WP, Maddison DR. **2016**. Mesquite: a modular system for evolutionary analysis. Version 3.10 http://mesquiteproject.org.

Magallón S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. **2015**. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

Manchester SR. **1999**. Biogeographical relationships of North American tertiary floras. *Annals of the Missouri Botanical Garden* **86**: 472–522.

Manos PS, Soltis PS, Soltis DE, *et al.* 2007. Phylogeny of extant and fossil Juglandaceae inferred from the integration of molecular and morphological data sets. *Systematic Biology* **56**: 412–430.

Massoni J, Forest F, Sauquet H. **2014**. Increased sampling of both genes and taxa improves resolution of phylogenetic relationships within Magnoliidae, a large and early-diverging clade of angiosperms. *Molecular Phylogenetics and Evolution* **70**: 84–93.

Miller MA, Pfeiffer W, Schwartz T. **2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees In: *Gateway Computing Environments Workshop (GCE), 2010*. IEEE, 1–8.

Misiewicz TM, Zerega NC. **2012**. Phylogeny, biogeography and character evolution of *Dorstenia* (Moraceae). *Edinburgh Journal of Botany* **69**: 413–440.

Nadot S, Alapetite E, Baker WJ, Tregear JW, Barfod AS. **2016**. The palm family (Arecaceae): a microcosm of sexual system evolution. *Botanical Journal of the Linnean Society* **182**: 376–388.

Obbard DJ, Harris SA, Buggs RJA, Pannell JR. **2006**. Hybridization, polyploidy, and the evolution of sexual systems in *Mercurialis* (Euphorbiaceae). *Evolution* **60**: 1801–1815.

Onstein RE, Carter RJ, Xing YW, Richardson JE, Linder HP. **2015**. Do Mediterranean-type ecosystems have a common history?-Insights from the buckthorn family (Rhamnaceae). *Evolution* **69**: 756–771.

Pagel M. **1994**. Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London B: Biological Sciences* **255**: 37–45.

Pagel M, Meade A. **2006**. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. *The American Naturalist* **167**: 808–825.

Pagel M, Meade A. **2013**. BayesTraits V2 http://www.evolution.rdg.ac.uk/BayesTraits.html.

Pagel M, Meade A, Barker D. **2004**. Bayesian estimation of ancestral character states on phylogenies. *Systematic Biology* **53**: 673–684.

Pannell J. **1997**. Widespread functional androdioecy in *Mercurialis annua* L. (Euphorbiaceae).

Biological Journal of the Linnean Society **61**: 95–116.

Parham JF, Donoghue PCJ, Bell CJ, *et al.* 2012. Best practices for justifying fossil calibrations. *Systematic Biology* **61**: 346–359.

Posada D, Buckley TR. **2004**. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793–808.

Potter D, Eriksson T, Evans RC, *et al.* 2007. Phylogeny and classification of Rosaceae. *Plant Systematics and Evolution* **266**: 5–43.

Rambaut A, Drummond A. **2009**. Tracer v1.5 http://tree.bio.ed.ac.uk/software/tracer/.

Renner SS. **2014**. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *American Journal of Botany* **101**: 1588–1596.

Renner SS, Ricklefs RE. **1995**. Dioecy and its correlates in the flowering plants. *American Journal of Botany* **82**: 596–606.

Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP. **2012**. A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Systematic Biology* **61**: 973–999.

Rønsted N, Weiblen GD, Clement WL, Zerega NJC, Savolainen V. **2008**. Reconstructing the phylogeny of figs (*Ficus*, Moraceae) to reveal the history of the fig pollination mutualism. *Symbiosis (Rehovot)* **45**: 45.

Rønsted N, Weiblen GD, Cook JM, Salamin N, Machado CA, Savolainen V. **2005**. 60 million years of co-divergence in the fig–wasp symbiosis. *Proceedings of the Royal Society B: Biological Sciences* **272**: 2593–2599.

Sabath N, Goldberg EE, Glick L, *et al.* 2016. Dioecy does not consistently accelerate or slow lineage diversification across multiple genera of angiosperms. *New Phytologist* **209**: 1290–1300.

Salisbury BA, Kim J. **2001**. Ancestral state estimation and taxon sampling density. *Systematic Biology* **50**: 557–564.

Sauquet H. **2003**. Androecium diversity and evolution in Myristicaceae (Magnoliales), with a description of a new Malagasy genus, *Doyleanthus* gen. nov. *American Journal of Botany* **90**: 1293–1305.

Sauquet H. **2013**. A practical guide to molecular dating. *Comptes Rendus Palevol* **12**: 355–367.

Sauquet H. **2016**. PROTEUS: a database for recording morphological data and creating NEXUS matrices. Version 1.26. http://eflower.myspecies.info/proteus.

Sauquet H, von Balthazar M, Magallón S, *et al.* 2017. The ancestral flower of angiosperms and its early diversification. *Nature Communications* **8**: 16047.

Sauquet H, Carrive L, Poullain N, Sannier J, Damerval C, Nadot S. **2015**. Zygomorphy

evolved from disymmetry in Fumarioideae (Papaveraceae, Ranunculales): new evidence from an expanded molecular phylogenetic framework. *Annals of Botany* **115**: 895–914.

Sauquet H, Ho SYW, Gandolfo MA, *et al.* 2012. Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). *Systematic Biology* **61**: 289–313.

Sauquet H, Weston PH, Barker NP, Anderson CL, Cantrill DJ, Savolainen V. **2009**. Using fossils and molecular data to reveal the origins of the Cape proteas (subfamily Proteoideae). *Molecular Phylogenetics and Evolution* **51**: 31–43.

Schaefer H, Renner SS. **2010**. A three-genome phylogeny of *Momordica* (Cucurbitaceae) suggests seven returns from dioecy to monoecy and recent long-distance dispersal to Asia. *Molecular Phylogenetics and Evolution* **54**: 553–560.

Simpson MG. **2010**. *Plant systematics*. Cambridge: Academic press.

Soltis DE, Smith SA, Cellinese N, *et al.* 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* **98**: 704–730.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Vamosi JC, Otto SP, Barrett SCH. **2003**. Phylogenetic analysis of the ecological correlates of dioecy in angiosperms. *Journal of Evolutionary Biology* **16**: 1006–1018.

Wang H, Moore MJ, Soltis PS, *et al.* 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* **106**: 3853–3858.

Webb CO, Ackerly DD, Kembel SW. **2008**. Phylocom: software for the analysis of phylogenetic community structure and trait evolution. *Bioinformatics* **24**: 2098–2100.

Weiblen GD. **2000**. Phylogenetic relationships of functionally dioecious *Ficus* (Moraceae) based on ribosomal DNA sequences and morphology. *American Journal of Botany* **87**: 1342–1357.

Wiegrefe SJ, Sytsma KJ, Guries RP. **1998**. The Ulmaceae, one family or two? Evidence from chloroplast DNA restriction site mapping. *Plant Systematics and Evolution* **210**: 249–270.

Williams EW, Gardner EM, Harris III R, Chaveerach A, Pereira JT, Zerega NJC. **2017**. Out of Borneo: biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany* **119**: 611–627.

Wu Z-Y, Monro AK, Milne RI, *et al.* 2013. Molecular phylogeny of the nettle family (Urticaceae) inferred from multiple loci of three genomes and extensive generic sampling. *Molecular Phylogenetics and Evolution* **69**: 814–827.

Yang Z, Rannala B. **2005**. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Molecular Biology and Evolution* **23**: 212–226.

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.

Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. **2005**. Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution* **37**: 402–416.

Zerega NJC, Supardi N, Motley TJ. **2010**. Phylogeny and recircumscription of Artocarpeae (Moraceae) with a focus on *Artocarpus*. *Systematic Botany* **35**: 766–782.

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.

SUPPLEMENTARY INFORMATION

Figure S1. Phylogenetic trees reconstructed by maximum likelihood with different molecular marks or combinations in this study.

Figure S2. Detailed version of the dated phylogenetic tree of 320 species shown in Fig. 2.

Figure S3. Ancestral state reconstruction with 320-species dataset by parsimony approach with tip names with all six configurations (see text).

Figure S4. Ancestral state reconstruction with 320-species dataset by equal-rate maximum likelihood with configurations A to F (configuration B presented as Fig. 3).

Figure S5. Ancestral state reconstruction with only *Ficus* species in the dataset with tip names by different approaches.

Table S1. List of GenBank accession numbers for the sequences used in this study.

Table S2. detailed information for the fossil calibrations used in this study.

Table S3. Extraction of the PROTEUS database including a list of all data records and linked references, lists of characters and character states used in this study, and the matrix in ancestral state reconstruction analyses.

Table S4. A list of divergence time estimates for tribes of Moraceae.

Table S5. Summary results of ancestral state reconstruction for the complete (320 species) dataset by Bayesian approach with fixed model (equal rate or unequal rate).

CHAPTER II. LONG-DISTANCE DISPERSAL SHAPED THE DIVERSITY OF TRIBE DORSTENIEAE (MORACEAE)

This study was submitted to the preprint server bioRxiv in January 2019, in collaboration with Elliot Gardner, Nyree Zerega and Hervé Sauquet.

ABSTRACT

- **Background and Aims** The Neotropics have the highest terrestrial biodiversity on earth. Investigating the relationships between the floras of the Neotropics and other tropical areas is critical to understanding the origin and evolution of this mega-diverse region. Tribe Dorstenieae (Moraceae) has a pantropical distribution and almost equal number of species on both sides of the Atlantic. In this study, we investigate the relationship between the African and Neotropical floras using Dorstenieae (15 genera, 156 species, Moraceae) as a model clade.
- **Methods** We used a targeted enrichment strategy with herbarium samples and a nuclear bait set to assemble a dataset of 102 genes sampled from 83 (53%) species and fifteen genera (100%) of Dorstenieae, and five outgroup species. Phylogenetic relationships were reconstructed with maximum likelihood and coalescent approaches. This phylogeny was dated with a Bayesian relaxed clock model and four fossil calibrations. The biogeographic history of the group was then reconstructed with several dispersal-extinction-cladogenesis models (incl. DEC and DEC+J).

- **Key Results** The crown-group ages of Dorstenieae and *Dorstenia* were estimated in the Cretaceous (65.8-79.8 Ma) and the Paleocene (50.8-67.3 Ma), respectively. Tribe Dorstenieae as a whole appears to have originated in the joint area of continental Africa, Madagascar and Asia-Oceania area. The Neotropical species of *Dorstenia* diversified in the Eocene (29.8-44.7 Ma) and formed a clade nested within the African lineages in the genus. *Brosimum* s.l., with a crown-group age at the period of the Oligocene and Miocene (14.9-31.1 Ma), represents another Neotropical clade in Dorstenieae.
- **Conclusions** Tribe Dorstenieae originated in the joint area of continental Africa, Madagascar and Asia-Oceania area in the Cretaceous and then dispersed into Neotropics twice. Neotropical diversification after long-distance dispersal across the Atlantic is the most plausible explanation for the extant distribution pattern of Dorstenieae.

Key words: Dorstenieae, *Dorstenia*, exon capture, phylogenomics, molecular dating, Neotropical diversity, long-distance dispersal, founder-event speciation, radiation

INTRODUCTION

The Neotropical ecozone has been defined as the region from central Mexico to southern Brazil (Morrone 2014). The Neotropics hold the highest terrestrial biodiversity on earth (Antonelli and Sanmartín 2011) and harbor all major tropical biomes: lowland rain forests, seasonally dry forests, mid-elevation montane forests, savannas, high elevation grasslands and deserts (Hughes *et al.* 2012). A recent estimate of tree species based on a pantropical tree inventory database suggested that the number of tree species in the Neotropics was as many as in the Indo-Pacific region and almost triple the counterparts in continental Africa (Slik *et al.* 2015). Several hypotheses have been proposed for the origins and evolution of Neotropical biodiversity during the past two decades. These hypotheses can be coarsely classified as biotic (e.g., dispersal ability, niche conservatism) and abiotic (time, climate, mountain uplift, Antonelli & Sanmartín, 2011).

After completely splitting from continental Africa in the Cretaceous (ca. 105 Ma) (McLoughlin 2001), South America was isolated until the uplift of the Panama Isthmus (ca. 15 Ma), which connected North and South America (Montes *et al.* 2012; Bacon *et al.* 2015). Evidence of long-distance dispersals (LDD) among all of these three continents has been found with the development of molecular dating and phylogenetic approaches (Christenhusz and Chase 2013). The relationships between the floras of the Neotropics and of other continents has intrigued researchers, as investigating these relationships not only can shed light on the origin and evolution of Neotropical biodiversity, but also help to understand disjunct distributions and long-distance dispersal. Taxonomically and genetically densely sampled phylogenetic analyses represent an ideal approach to improve the understanding of the origin and evolution of

Neotropical diversity (Antonelli and Sanmartín 2011; Hughes *et al.* 2012).

The generic composition of the angiosperm tribe Dorstenieae in the Moraceae family has been under recent scrutiny. Based on taxonomic treatments and recent molecular phylogenetic studies, Dorstenieae are currently thought to consist of fifteen genera (Table 1) and approximately 156 species (Berg et al. 2001; Clement & Weiblen 2009; Zerega et al. 2010; Chung et al. 2017). The most diverse genus in Dorstenieae is *Dorstenia*, which includes 113 species (Berg and Hijman 1999; Mccoy and Massara 2008; dos Santos and Neto 2012; Machado and Marcelo Filho 2012; Chase *et al.* 2013; dos Santos *et al.* 2013; Leal 2014; Machado *et al.* 2014; Rzepecky 2016). The genera in Dorstenieae are restricted to either side of the Atlantic, except *Dorstenia*, which has almost the same number of species in South America and continental Africa (50 in the Neotropics; 62 in continental Africa, Madagascar and Arabian Peninsula; 1 in India and Sri Lanka) (Berg and Hijman 1999). In a recent phylogenetic study of Moraceae (Zhang *et al.* 2019), the relationships among the genera of Dorstenieae (except *Bosqueiopsis* and *Scyphosyce*) were reconstructed. Most of them were strongly supported, but the relationships of *Trilepisium*, *Brosimum* and *Treculia* remained unclear. The most densely sampled phylogenetic study of *Dorstenia* to date sampled 32 species (28%) and found *Dorstenia* to have originated in Africa, with three African species nested inside the Neotropical clade (Misiewicz and Zerega 2012).

Estimating absolute divergence times as accurately as possible is essential to biogeographic reconstruction to connect the evolutionary history of target taxa with past climate change and geographic events (Sauquet 2013). Previous studies have estimated the crown-group age of Dorstenieae to be at least 71 Ma (Zerega *et al.* 2005), 50.6-72.5 Ma (Gardner *et al.* 2017), or 51.4-70.2 Ma (Zhang *et al.* 2019). While these estimates overlap, the crown-group age of *Dorstenia* has remained unclear. The crown-

Table 1. Classification of Dorstenieae and the number of species for each genus.

This study	No. of species sampled in this study /No. of species	Distribution	Reference
Allaeanthus	3/4	D,E,F	(Chung *et al.* 2017)
Bleekrodea	1/3	D, F	(Rohwer and Berg 1993)
Bosqueiopsis	1/1	C	(Rohwer and Berg 1993; Berg 2001)
Broussonetia	1/4	F	(Rohwer and Berg 1993; Chung *et al.* 2017)
Brosimum	10/15	A, B	(Berg 2001)
Dorstenia	73/113	A, B, C, D, E	(Berg and Hijman 1999; Berg 2001; Mccoy and Massara 2008; dos Santos and Neto 2012; Machado and Marcelo Filho 2012; Chase *et al.* 2013; dos Santos *et al.* 2013; Leal 2014; Machado *et al.* 2014; Rzepecky 2016)
Fatoua	1/2	D, F	(Rohwer and Berg 1993)
Helianthostylis	1/2	A	(Rohwer and Berg 1993; Berg 2001)
Malaisia	1/1	F	(Wu *et al.* 2003; Clement and Weiblen 2009)
Scyphosyce	2/2	C	(Berg 1977; Rohwer and Berg 1993)
Sloetia	1/1	F	(Clement and Weiblen 2009; Tandang *et al.* 2017)
Treculia	3/3	C, D	(Rohwer and Berg 1993; Zerega *et al.* 2010)
Trilepisium	1/1	C, D	(Rohwer and Berg 1993; Berg 2001)
Trymatococcus	1/2	A	(Berg 2001)
Utsetela	1/2	C	(Berg 1977; Jongkind 1995)

Codes for distribution areas were the same as in Figure 1. A, South America; B, Central/North America; C, continental Africa; D, Madagascar; E. India and Sri Lanka; F, Southeast Asia and Oceania.

group age of *Dorstenia* has been estimated using a range of taxonomic sampling density to be 3.5-18.4 Ma (Zerega et al., 2005, with two Neotropical species included), 12.7-31.7 Ma (Zhang et al., 2018, with two neotropical and one African species included), or 84.8-132.0 Ma (Misiewicz & Zerega, 2012, with 15 neotropical and 14 African species included).

Approximately 10% of the species in Moraceae are herbaceous and all of them belong to Dorstenieae (more specifically all in *Dorstenia* and *Fatoua*; Berg, 2001). Species of Dorstenieae show diversity in pollination modes, dispersal mechanisms, and habit (Berg 2001). The species of *Dorstenia* are found in a wide variety of habitats (e.g., tropical rain forest, savannas, or crevices of cliffs) and life forms (e.g., tree, shrubs, caulescent, herbaceous; Berg & Hijman, 1999; Berg, 2001; Misiewicz & Zerega, 2012). The pantropical distribution, almost equal diversity on both sides of the Atlantic, and the diverse traits of Dorstenieae make it a good model for understanding the origin and evolution of Neotropical biodiversity and the relationships between African and Neotropical floras.

Genomic targeted enrichment approaches have been shown to be more efficient and economic than Sanger sequencing (Lemmon and Lemmon 2013; McKain *et al.* 2018) and have been widely used in phylogenetic studies in recent years (Xi *et al.* 2014; Fisher *et al.* 2016; Hart *et al.* 2016; Mitchell *et al.* 2017; Couvreur *et al.* 2019). In this study, we targeted nuclear genes as they have been suggested to hold the greatest potential for investigating the evolutionary history of angiosperms for several reasons. Firstly, nuclear genes have worked well in reconstructing more strongly supported phylogenetic relationships than organellar markers in both deep and shallow time scales (Xi *et al.* 2014; Mitchell *et al.* 2017). Secondly, nuclear genes are assumed to be unlinked, decreasing the probability of misleading phylogeny by part of the genes

used (Fisher *et al.* 2016). Lastly, nuclear genes present multiple lineage histories, contrary to plastid genes, which are usually considered to represent a single locus.

In this study, we included samples representing all fifteen genera of Dorstenieae (Table 1) to reconstruct a comprehensive dated phylogenetic tree of Dorstenieae with two primary goals: 1) to test phylogenetic relationships and monophyly of Dorstenieae genera, and relationships within *Dorstenia*; and 2) to investigate divergence times and the biogeographic history of tribe Dorstenieae and genus *Dorstenia*. This study also provided an opportunity to test the potential of the targeted enrichment strategy to resolve species-level phylogenetic relationships using herbarium material, taking advantage of a recently developed nuclear bait set for Moraceae (Gardner et al., 2016, Johnson et al., 2016).

MATERIALS AND METHODS

Specimen and sample collection

In this study, we follow the classification of Clement & Weiblen (2009) with modifications based on Zerega et al. (2010, recommending *Treculia* be transferred to Dorstenieae) and Chung et al. (2017, reinstating *Allaeanthus* in Dorstenieae). This approach recognizes fifteen genera and approximately 153 species in this tribe (Table 1). We included 83 species (93 taxa) representing all currently recognized Dorstenieae genera and 53% of the species within tribe Dorstenieae (Table S1, we also collected samples of 22 more species in Dorstenieae and they were finally excluded in the main analyses, see below). Additionally, five outgroup taxa in Moraceae were included. They represent five of the six other Moraceae tribes: Artocarpeae, Castilleae, Ficeae, Moreae,

and the newly created Parartocarpeae (Zerega and Gardner 2019). Most taxa were extracted from herbarium material sampled from the Field Museum (F), the Missouri Botanical Garden (MO), the New York Botanical Garden (NY), and the Muséum national d'Histoire naturelle (P) (Table S1). Specimens collected within the last twenty years, and reliably identified material with inflorescences or infructescences were preferred. Samples include 55 (49%) species of *Dorstenia*, representing eight out of the nine sections proposed by Berg and Hijman (1999). Section *Bazzemia*, which contains only one species from Mozambique (Berg & Hijman, 1999), was not sampled.

DNA extraction and sequencing

Whole genome DNA was extracted using a modified CTAB method (Doyle & Doyle, 1987). Extracted DNA was re-suspended in 50 µl light TE. The DNA concentration of each sample were measured using Qubit 2.0 fluorometer (Life Technologies, Carlsbad, CA, USA) following the standard protocol. For library preparation, we used 200 ng DNA when possible, but never less than 20 ng. All samples were run on an agarose gel to assess fragment size. Samples with DNA fragment lengths longer than 500 base pairs (bp) were sonicated on a Covaris M220 (Covaris, Wobum, MA, USA) for 45 seconds at 50 W peak power and a duty factor of 20%, which typically produces average fragment sizes of 550 bp. We prepared dual-indexed sequencing libraries using the KAPA Hyper Prep Library Construction Protocol (KAPA biosystems, Wilmington, MA, USA), generally following the manufacturer's protocol except that most steps were performed at one-quarter volume to save costs. Low-input or very degraded samples were not size selected. Libraries were amplified using 12 cycles of PCR, but half of the unamplified template was retained in case PCR needed to be

repeated. Products were cleaned using Solid Phase Reversible Immobilization (SPRI) and quantified using a high-sensitivity dsDNA assay on a Qubit 2.0. For samples with a concentration less than 5 μg/ml, we repeated PCR amplification with 14 cycles. Successful libraries were combined into seven pools of 13 to 14 libraries each. We hybridized the libraries to custom Moraceae probes (Gardner et al. 2016) manufacturered by Arbor Biosciences (Ann Arbor, MI, USA) as a MYbaits kit. Hybridization for 20 hours followed the manufacturer's protocol, and products were reamplified 14 cycles of PCR. Amplified products were quantified on a Qubit 2.0, and fragment sizes were determined using a High-Sensitivity DNA assay on a BioAnalyzer 2100 (Agilent Technologies, Palo Alto, Califonia, USA). When required, adapter dimer was removed using 0.7x SPRI beads. All libraries were then sequenced on a single lane of an Illumina HiSeq 2000 (2 × 100 bp, paired-end) by Genewiz (Genewiz, South Plainfield, NJ, USA).

To the 96 libraries prepared for this study, we added sequences from 14 samples prepared for other projects, including *Brosimum* (10)*, Trymatoccocus oligandrus*, *Helianthostylis sprucei*, *Allaeanthus luzonicus*, and *Malaisia scandens*, as well as five samples (*Artocarpus heterophyllus, Milicia excelsa, Parartocarpus venenosus, Ficus macrophylla*, and *Antiaropsis decipiens*) from Johnson et al. (2016) and Zerega and Gardner (Zerega and Gardner 2019). These 20 samples were all sequenced on an Illumina MiSeq with somewhat longer reads (2x300bp, v3). Finally, we used transcriptomic reads for *Broussonetia papyrifera* obtained from GenBank.

Sequence cleaning, assembly, and filtering

Sequences were assembled using HybPiper (Johnson *et al.* 2016), which uses

reference sequences to guide local *de novo* assemblies of each target gene. Because Dorstenieae are phylogenetically distant from the baits sequences, which come from Artocarpeae and Moreae (Gardner et al. 2016), a new HybPiper reference was generated using six samples sequenced here, each with at least eight million read pairs (*Dorstenia bahiensis*, *D. cayapia*, *D. erythrandra*, *D. kameruniana*, *Treculia africana*, and *Fatoua villosa*). Reads were trimmed using Trimmomatic (Bolger et al., 2014) (LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:20) and assembled with SPAdes 3.10.1 (Bankevich *et al.* 2012) using default parameters. Coding sequences were predicted with Augustus (Stanke *et al.* 2004), using *Arabidopsis thaliana* genes as a reference. A seven-way orthology search was carried out with ProteinOrtho5 (Lechner *et al.* 2011) using all CDS over 200 bp from the *de novo* assemblies in addition to the *Artocarpus* HybPiper (Johnson *et al.* 2016) reference from Kates et al. (Kates *et al.* 2018). Orthologs present in at least three taxa were included in a new seven-taxon HybPiper reference consisting of in-frame CDS. This expanded reference contained approximately 500 genes.

Assembly of all reads then proceeded as follows. We first trimmed low quality bases and adapter sequences using Trimmomatic (Bolger et al., 2014) with the following parameters: ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10 LEADING:20 TRAILING:20 SLIDINGWINDOW:4:20 MINLEN:40. To ensure that quality trimming worked as expected, we examined a subset of reads were both before and after trimming with FastQC (https://www.bioinformatics.babraham.ac.uk/ projects/fastqc/). Reference-guided assembly then proceeded with HybPiper (Johnson *et al.* 2016), using the new reference described above. Briefly, the program works as follows: reads are sorted by gene based on the reference. Local *de novo* assemblies are carried out using SPAdes (Bankevich *et al.* 2012), and coding DNA sequences (CDS)

are predicted using Exonerate (Slater and Birney 2005). When a gene is assembled into several disconnected contigs—common in degraded samples where the fragments (and therefore effective read length) are very short—HybPiper scaffolds these short contigs in the correct order based on the reference. In the event that multiple genes are assembled for a single target, HybPiper distinguishes orthologs from paralogs using a combination of alignment length and identity relative to the reference. HybPiper outputs include in-frame CDS sequences as well as "supercontigs," which contain CDS as well as any flanking non-coding sequences. For our analyses, we used only the supercontig sequences. We then filtered the sequences within loci to remove those less than 150 bp long or shorter 25% of the average length for the locus.

After filtering, the number of genes recovered for each taxon varied widely among taxa (from zero for *D. scaphigera* and 515 for *Ficus macrophylla*) and the number of taxa retrieved for each gene differed sharply among genes (from two to 101). These extreme differences would result in dataset with a high proportion of missing data, which could impact the accuracy of both phylogenetic reconstruction and divergence time estimate. Therefore, we filtered the dataset further by selecting 102 genes with high taxon occupancy and excluding taxa for which less than 30 of these genes were recovered (Table S1).

Phylogenomic reconstruction and molecular dating

Sequences were aligned with MACSE (Ranwez *et al.* 2011), gene by gene (-fs 30 -stop 50). We reconstructed phylogenetic relationships using both a concatenated maximum likelihood (ML) and a coalescent approach. *Milicia exceisa* and *Artocarpus venenosus* were specified as the most external outgroups based on recent phylogenetic

analyses of Moraceae (Clement and Weiblen 2009; Zerega *et al.* 2010; Chung *et al.* 2017; Gardner *et al.* 2017; Zhang *et al.* 2019). Sequences were concatenated into a supermatrix and partitioned according to genes using the fasta_merge.py script from HybPiper; a ML tree was reconstructed with RAxML v8.2.10 (Stamatakis 2014) as implemented on CIPRES (Miller *et al.* 2010) 1000 rapid bootstrap replicates. GTRGAMMA was chosen as the substitution model. For the coalescent approach, we first reconstructed ML trees with RAxML v8.2.10 for each gene, using with GTRGAMMA model and 200 rapid bootstrap replicates. These ML trees for each gene were used to infer a species tree using the summary coalescent approach implemented in ASTRAL-II (Mirarab and Warnow 2015). Node support was calculated based on 200 bootstrap replicates, with resampling within loci.

Because the topologies of the ML and species trees were broadly consistent, we used the ML tree to estimate divergence times with a relaxed clock model and four fossil calibrations from Zhang et al. (2018). The root age (crown-group node of Moraceae) was constrained between 73 to 85 Ma, based on the results of Zhang et al. (2019) who used a more comprehensive sample of species within and outside Moraceae along with 12 fossil age constraints. The fossil wood of *Artocarpoxylon deccanensis* Mehrotra, Prakash, and Bande (at least 64.0 Ma) (Mehrotra *et al.* 1984) was used to calibrate the split between *Artocarpus* and *Milicia* as we used *Artocarpus heterophyllus* and *Milicia excelsa* to represent Artocarpeae and Moreae, respectively. The fossil fruits *Morus tymensis* Dorofeev (at least 33.9 Ma) (Collinson 1989) used to calibrate Moreae in Zhang et al. (2018) would here provide an additional, but younger (and hence uninformative) minimum age constraint on the stem node of *Milicia*. The fossil endocarps of *Broussonetia rugosa* Chandler (Chandler 1961) were used to constrain the crown group of *Broussonetia* s.l. (incl. *Allaeanthus*, previously included

in *Broussonetia*, and *Malaisia*; Chung et al., 2017) to a minimum age of 33.9 Ma. The fossil achenes of *Ficus* (*F. lucidus* Chandler) (Chandler 1962) were used to calibrate the stem group of *Ficus* with a minimum age of 56.0 Ma. Thus, we used one secondary calibration (root) and three fossil age constraints in our analyses.

Both penalized likelihood (PL) and Bayesian relaxed clock approaches were used to estimate divergence times in Dorstenieae. PL was implemented in r8s v1.7 (Sanderson, 2003) with strict minimum and maximum age constraints as described above. For the PL approach, the best smoothing value was first determined using cross validation by testing 21 values of smoothing parameter scaling from 0.1 to 1000. The optimum value (i.e., with lowest chi-square) of 1.6 was then used as the smoothing parameter in divergence time estimation.

For the Bayesian approach, we used MCMCTree as implemented in the PAML v4.9 package (Yang 2007). MCMCTree has been used to estimate divergence times with other similar phylogenomic datasets (e.g., Foster et al., 2017), where other programs such as BEAST2 (Bouckaert *et al.* 2014) would take too long to converge. The tree topology was fixed, using the best-scoring tree from the RAxML analysis, and all calibrations were implemented as uniform priors with soft boundaries (2.5% on both sides) (Ho and Phillips 2009). We set the minimum age of the fossil and as the minimum boundary and the younger bound of the root (73 Ma) as the maximum boundary of the uniform distribution for each fossil calibration. The birth-death process was used as the tree prior. For the MCMCTree analysis, we used the supermatrix, analyzed as a single partition under the GTR substitution model.

We first ran baseml in PAML v4.9 with a strict clock model to estimate the rough mean of parameters such as the shape parameter for the overall rate and the transition/transversion rate ratio. Two steps are needed for divergence time estimation

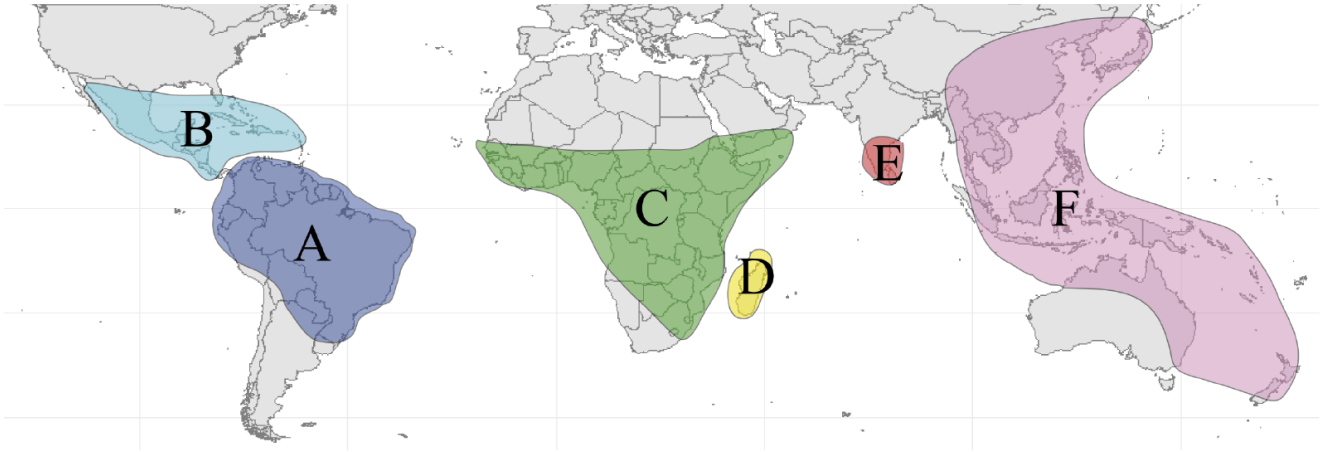
by approximate likelihood in MCMCTree. We first estimated the gradient and Hessian, and then used them to estimate the divergence times. We set the prior in both steps according to the estimates of baseml. In both steps, we ran the process for 38.5 million generations, with the first 10% of the chain length discarded as burnin, sampling a total of 10,000 generations at a frequency of once every 3,500 generations. Two independent runs with the same settings were conducted to confirm the convergence of the MCMC. To check the influence of the prior on the estimation, we used another prior setting (program defaults), followed the same steps as above, then ran the chain for 22 million generations, sampling a total of 10,000 generations at a frequency of once every 2,000 generations. After checking the convergence of the runs, we combined the results of them for each prior setting. Two additional independent runs with the two different prior settings, but without data, were also conducted to test the impact of the priors on the results. The convergence of Markov chain Monte Carlo (MCMC) was checked by reading the result log files in Tracer v1.7 (Rambaut *et al.* 2018). For the Bayesian approach, divergence was checked by confirming that the effective sample size (ESS) of parameters in the two independent runs was over 100 after removing the first 10% of chain length as burnin.

Reconstruction of biogeographic history

Distribution data for the taxa sampled were collected from monographs and revisions (Rohwer and Berg 1993; Berg and Hijman 1999; Berg 2001; Chung *et al.* 2017). We separated the distribution area of extant species of Dorstenieae into six areas (Fig. 1) following a recent study of Annonaceae, which share a similar global distribution (Couvreur *et al.* 2011): A, South America; B, North/Central America; C,

Africa; D, Madagascar; E, India and Sri Lanka; F, Southeast Asia and Oceania. We decided to combine the two areas west and east of Wallace's line (areas F and G of Couvreur et al., 2011) into a single area (area F) because all species of Dorstenieae distributed east of Wallace's line were also observed west of Wallace's line in our dataset (e.g., *Allaeanthus luzonicus*, *Fatoua villosa*).

Only Dorstenieae (93 taxa, 83 species) were kept in the biogeographic reconstruction to avoid the bias of incompletely sampled outgroups. BioGeoBEARS (Matzke 2013) in R v3.5.1 (R Core Team, 2018) was used to reconstruct the biogeographic history of the clade with the DEC and DEC+J models. Prior studies have suggested that the DEC+J model can exacerbate the bias of preferring cladogenetic events (i.e., sympatry, vicariance and founder event speciation) over anagenetic processes in the DEC model, implying that the higher likelihoods typically obtained with DEC+J do not necessarily mean a better fit to the data than DEC (Ree and Sanmartín 2018). Therefore, we employed both the DEC and the DEC+J models. In addition, we ran the analyses with the DIVA-like and BAYAREA-like models including or not including founder-event speciation (+J). All analyses were run both with a simple dispersal matrix and a time-stratified model, hereafter referred to as model 0 and model 1, respectively. Both models were inspired by Couvreur et al. (2011) and are constrained by unequal dispersal relative rates aimed at reflecting the connectivity of biogeographic areas (Fig. 1). In the time-stratified model, the constraints on dispersal probabilities vary through five time periods (0-5, 5-30, 30-45, 45-65, and 65-75 Ma) according to physical distance between areas (Fig. 1). For instance, the dispersal rates from continental Africa (C) and Madagascar (D) to India and Sri Lanka (E) were constrained to be very low during the rafting of the Indian plate 30-65 Ma. Because all sampled species of Dorstenieae currently occupy no more than two areas, the maximum number



model 0: time-constant

	A	B	C	D	E	F
A	-	1	1	0.01	0.01	1
B	1	-	1	0.01	0.01	1
C	1	1	-	1	1	1
D	0.01	0.01	1	-	1	1
E	0.01	0.01	1	1	-	1
F	1	1	1	1	1	-

model 1: time-stratified

0-5 Ma

	A	B	C	D	E	F
A	-	1	0.01	0.01	0.01	0.01
B	1	-	0.01	0.01	0.01	0.01
C	0.01	0.01	-	0.25	0.25	0.25
D	0.01	0.01	0.25	-	0.25	0.25
E	0.01	0.01	0.25	0.25	-	1
F	0.01	0.01	0.25	0.25	1	-

5-30 Ma

	A	B	C	D	E	F
A	-	0.25	0.01	0.01	0.01	0.01
B	0.25	-	0.01	0.01	0.01	0.01
C	0.01	0.01	-	0.5	0.25	0.25
D	0.01	0.01	0.5	-	0.25	0.25
E	0.01	0.01	0.25	0.25	-	1
F	0.01	0.01	0.25	0.25	1	-

30-45 Ma

	A	B	C	D	E	F
A	-	0.01	0.25	0.01	0.01	0.01
B	0.01	-	0.01	0.01	0.01	0.01
C	0.25	0.01	-	0.5	0.01	0.25
D	0.01	0.01	0.5	-	0.01	0.25
E	0.01	0.01	0.01	0.01	-	0.75
F	0.01	0.01	0.25	0.25	0.75	-

45-65 Ma

	A	B	C	D	E	F
A	-	0.25	0.5	0.01	0.01	0.01
B	0.25	-	0.75	0.01	0.01	0.75
C	0.5	0.75	-	0.5	0.01	0.25
D	0.01	0.01	0.5	-	0.01	0.25
E	0.01	0.01	0.01	0.01	-	0.25
F	0.01	0.75	0.25	0.25	0.25	-

65-75 Ma

	A	B	C	D	E	F
A	-	0.5	1	0.01	0.01	0.01
B	0.5	-	0.25	0.01	0.01	0.5
C	1	0.25	-	1	1	0.01
D	0.01	0.01	1	-	1	0.01
E	0.01	0.01	1	1	-	0.01
F	0.01	0.5	0.01	0.01	0.01	-

Figure 1. Delimitation of distribution areas of Dorstenieae and relative dispersal matrices for the time-constant (model 0) and time-stratified (model 1) models. A, South America; B, North/Central America; C, Africa; D, Madagascar; E, India and Sri Lanka; F, Southeast Asia, and Oceania. Five levels of dispersal probability, 0.01, 0.25, 0.5, 0.75 and 1, representing the probability from low to high.

of co-occurrence areas was set as three in all analyses.

RESULTS

Sequencing

Sequencing and assembly statistics appear in Table S1. Enrichment ranged from 0.34% to 84% reads on target, with the enrichment found in Artocarpeae and Moreae (58% and 84% on target, respectively) and the least efficient found in the genus *Dorstenia* (0.34% on target). The final dataset consisted of 98 taxa (89 species, spanning all of the currently recognized genera) and 102 genes. The width of the aligned supermatrix was 132,753 bp with 29.51% gaps and missing data.

Phylogenetic relationships

The maximum likelihood (ML) and coalescent analyses produced broadly identical topologies, with some weakly supported differences at shallow phylogenetic depths (Fig. 2, Figure S1a,b). Most of the nodes on the tree were strongly supported (over 90% bootstrap support , Fig. 2). *Fatoua villosa* was found to be sister to the remaining of Dorstenieae. *Malaisia scandens* and *Broussonetia papyrifera* together formed a clade that is sister to *Allaeanthus*. All genera of Dorstenieae sampled more than once were found to be monophyletic with two notable exceptions. *Trymatococcus oligandrus* and *Helianthostylis sprucei* appear to be nested in *Brosimum*, and *Scyphosyce* (two species) and *Utsetela gabonensis* were found to be nested among early-diverging lineages of *Dorstenia*. Hereafter we refer to the clade of *Brosimum*, *Trymatococcus* and

Helianthostylis as *Brosimum* s.l., and to the clade of *Dorstenia*, *Scyphosyce* and *Utsetela* as *Dorstenia* s.l. The Neotropical species of *Dorstenia* formed a clade well nested among African lineages. *Dorstenia elliptica* (from Central Africa) appears to be sister to this Neotropical *Dorstenia* clade. The Central/North American species of *Dorstenia* formed two clades nested among the South American lineages with strong support as well. Internal branch lengths were comparatively shorter in Neotropical *Dorstenia* and most of the differences between the ML and coalescent approach concentrated in this clade. Short branches were also observed in *Brosimum* s.l., the other Neotropical clade in Dorstenieae, but in this case both reconstruction approaches showed identical topologies. Six species had duplicated samples in this study, four of them were sister or close to the other samples of the same species, while sample *D. brasiliensis*-1 and *D. arifolia* were not (but still lie in the Neotropical clade). Although in similar positions on the phylogenetic trees reconstructed by two approaches, the support for the splits of these two species were low in the coalescent tree (bootstrap support less than 50%) but high in the maximum likelihood tree (bootstrap support over 90%). Several nodes were comparatively weakly supported in *Dorstenia* (Fig. 2). Our samples represented eight out of nine recognized sections in *Dorstenia*. None of them was found to be monophyletic in our analyses (Figure S1). All of the Neotropical species of Dorstenieae outside of *Dorstenia* also formed a clade (i.e., *Brosimum* s.l., incl. *Trymatococcus* and *Helianthostylis*).

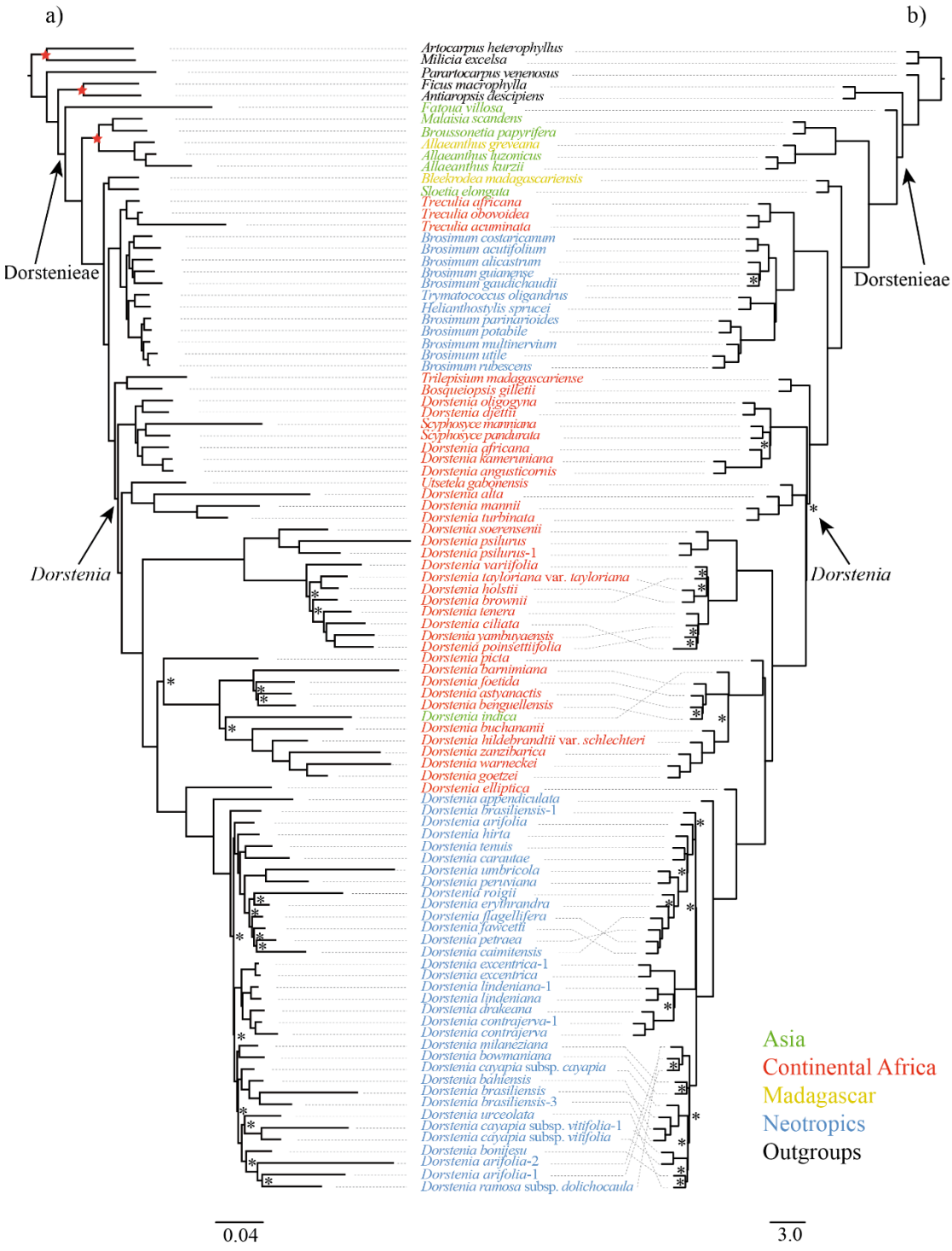


Figure 2. Maximum Likelihood phylogenomic tree (a) and ASTRAL tree (b) of Dorstenieae. Fossil-calibrated nodes are marked as red stars. Nodes with bootstrap support value (less than 90% are indicated with an asterisk. Tip names are colored by general distribution area.

Divergence times of Dorstenieae and Dorstenia

The crown-group age of tribe Dorstenieae was estimated in the Cretaceous (65.8-79.8 Ma) and that of *Dorstenia* in the Paleocene (50.8-67.3 Ma) (Table 2, Figure S2). The stem and crown-group ages of the Neotropical *Dorstenia* clade were dated in the Eocene to early Oligocene (34.6-51.8 Ma and 29.8-44.7 Ma, respectively). *Brosimum* s.l., the other Neotropical clade in Dorstenieae, was estimated to date from the late Eocene to early Miocene (stem node: 19.6-41.5 Ma; crown node: 14.9-31.1 Ma). Runs with the prior only showed different results to those with data, indicating the data had a significant impact on the posterior (results not shown). Results of the two different prior settings were similar to each other (Table 2, Figure S2a,b). Divergence times estimated with the PL approach were all compatible with those from the Bayesian approach (i.e., falling within the 95% credibility intervals) except the age of crown-group *Broussonetia* s.l., which was significantly older with PL (Table 2, Figure S2c).

Table 2. Divergence time estimates with penalized likelihood (PL with r8s) and Bayesian (MCMCTree) approaches for key nodes of Dorstenieae.

Node	**r8s (Ma)**	**MCMCTree set1 (Ma)**	**MCMCTree set2 (Ma)**
SG Dorstenieae	77.3	68.7-82.1	68.9-82.1
CG Dorstenieae	75.4	65.2-79.8	65.8-79.8
CG *Treculia*	35.3	10.9-32.7	11.2-32.4
SG *Brosimum* s.l.	40.7	19.4-42.9	19.6-41.5
CG *Brosimum* s.l.	27.9	14.7-31.7	14.9-31.1
SG *Broussonetia* s.l.	70.4	60.2-76.4	61.1-76.2
CG *Broussonetia* s.l.	52	30.5-41.8	30.2-40.7
CG *Scyphosyce*	34.7	10.8-31.5	10.7-30.9
SG *Dorstenia*	61.9	51.2-68.5	52.3-68.7
CG *Dorstenia*	61.3	49.8-67.0	50.8-67.3
SG *Dorstenia* Neo	43.4	34.3-52.1	34.6-51.8
CG *Dorstenia* Neo	33.4	29.7-44.7	29.8-44.7

SG *Dorstenia* MAm clade1	23.1	16.5-28.2	16.7-28.0
CG *Dorstenia* MAm clade1	20.2	12.6-24.9	12.8-24.6
SG *Dorstenia* MAm clade2	25.4	23.9-35.1	24.1-34.7
CG *Dorstenia* MAm clade2	13.3	12.1-28.9	12.0-28.4

CG: crown group; SG: stem group

MCMCTree set1: results of estimate in MCMCTree with the prior setting referring to the results of baseml

MCMCTree set2: results of estimate with the default prior setting in MCMCTree

Brosimum s.l. includes *Brosimum*, *Trymatococus* and *Helianthostylis*

Broussonetia s.l. includes *Broussonetia*, *Malaisia*, and *Allaeanthus*

Dorstenia Neo: Neotropical *Dorstenia* species

Dorstenia MAm Clade1: *Dorstenia* species distributed in Central and North America, from *D. erythrandra* to *D. caimitensis* in Figure 3

Dorstenia MAm Clade2: *Dorstenia* species distributed in Central and North America, from *D. excentrica* to *D. contrajerva* in Figure 3

Biogeographic history of Dorstenieae

We analyzed the dataset with both the dispersal-extinction-cladogenesis (DEC) and DEC+J (DEC with founder-event speciation) models. The time-stratified model fit the data significantly better than the time-constant model in all analyses conducted (Table S2), therefore we focus mainly on the results from the time-stratified analyses here, unless otherwise mentioned. In addition, the DEC+J model had a lower value of Akaike information criterion (AIC) than that of DEC (Table S2). The ancestral distribution area of Dorstenieae was estimated to be the combined area of continental Africa, Madagascar, and Southeast Asia and Oceania (CDF) with both the DEC and DEC+J models (Fig. 3, Fig. S3c). The ancestral area for both the stem and crown-group nodes of *Dorstenia* s.l. (incl. *Scyphosyce* and *Utsetela*) were estimated in continental Africa (C) with both models. South America (A) was found as the ancestral area of both the stem and crown-group nodes of the Neotropical *Dorstenia* clade with the DEC+J model

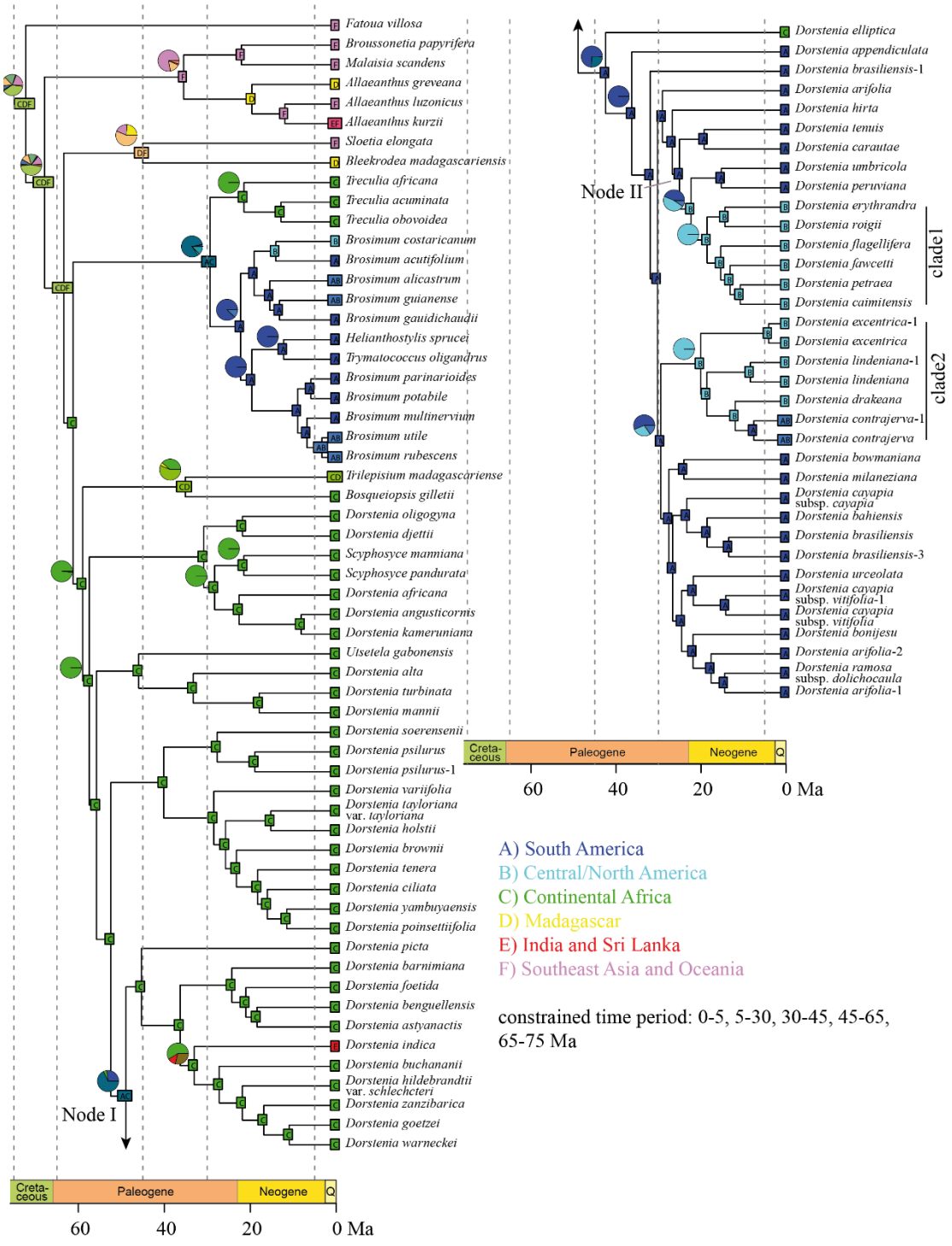


Figure 3. Biogeographic history reconstruction of Dorstenieae based on the time-stratified DEC+J model. Inferred ancestral distribution areas prior to speciation are indicated on the nodes. Pie charts for selected nodes represent the relative probability (proportional likelihoods) of alternative ancestral areas (for full details, see Figure S3d).

(Fig. 3), while the DEC model reconstructed the combined area of South America and continental Africa (AC) as ancestral for the stem-group node of this clade (Figure S3c). *Dorstenia indica*, endemic to India and Sri Lanka, was found to be nested in an African clade and diverged from its sister group in the Eocene to Oligocene (26.2-46.7 Ma). The Central/North American *Dorstenia* species clustered into two clades with strong support (Fig. 2), suggesting two independent colonizations from South America to Central/North America during the Oligocene to Miocene (Table 2). The ancestral states of the crown-group nodes of both Central/North American clades were estimated as Central/North American. The stem-group nodes of clade 1 (from *D. erythrandra* to *D. caimitensis*) and clade 2 (from *D. excentrica* to *D. contrajerva*) were estimated to be Central/North American and South American, respectively (Fig. 3). The stem and crown-group nodes of *Brosimum* s.l. were estimated in the joint area of South America and continental Africa (AC), and in South America (A), respectively, with both the DEC and DEC+J models.

The additional biogeographic models produced similar results to those of the DEC and DEC+J models, with a few exceptions (Figure S3, Table S3). Reconstruction with time-constant models showed similar results as with time-stratified models. The ancestral distribution area for the crown-group node of Dorstenieae was estimated to be Southeast Asia and Oceania (F) by all the models based on the DIVA-like model (Figure S3i-l). In the time-stratified BayArea-like models, with or without founder-event speciation, the ancestral area of the crown-group node of Dorstenieae was estimated as the joint area of continental Africa and Southeast Asia and Oceania (CF), or as Southeast Asia and Oceania (F), respectively. The same pattern was found in reconstructions with time-constant BayArea-like models (Figure S3e-h). The stem and crown-group nodes of *Dorstenia* were estimated in continental Africa (C) in all the

models. The crown-group node of Neotropical *Dorstenia* was estimated in South America (A) in all analyses, while the stem-group node was different among models. All the BayArea-like models estimated this node in continental Africa (C) (Figure S3e-h). The time-stratified DIVA-like models estimated the joint area of South America and continental Africa (AC), or South America (A) alone as the ancestral area of this node, with or without founder-event speciation respectively. The same results were found in time-constant DIVA-like models (Figure S3i-l). Lastly, we also ran all biogeographic analyses with the chronograms reconstructed with the penalized likelihood approach. The results were similar with some exceptions (results not shown). For instance, Southeast Asia and Oceania (F) was estimated as the ancestral area of the crown-group node of Dorstenieae with the time-stratified DEC+J model. The stem-group node of Neotropical *Dorstenia* was estimated as continental Africa (C) by the same model.

DISCUSSION

Success of the targeted enrichment strategy with herbarium specimens

Most (99%) of the samples in this study were from herbarium specimens (Table S1). Some of them were collected more than 40 years ago and the amount of sample collections from the herbarium was typically limited (around 3 to 20 mg) due to destructive sampling policies. Three samples (*Dorstenia aristeguietae*, *D. choconiana*, and *D. prorepens*) were filtered by HybPiper because of the low matching of reads to the reference. We excluded another fifteen taxa for the low number of genes recovered (less than 30, Table S1). The lowest amount of DNA used among the final 98 taxa retained in our dataset was 25.1 ng (*Dorstenia brasiliensis*). Samples that were

excluded ranged from being over 100 years old to 11 years old, while samples that were included were collected as long ago as 1923. These results suggest that the degradation of DNA in old herbarium specimens appears to have had little influence on this study. This may be because of the short DNA fragments (on average less than 500 bp long) needed in the library preparation, those longer than this size requiring sonication. Thus, our results suggest that the targeted enrichment strategy and HybPiper pipeline (Johnson *et al.* 2016) worked well with a broad range of ages of herbarium specimens. Sequencing herbarium samples is a valuable approach for phylogenetics, as many species can be difficult to collect. Herbarium specimens often represent reliable and accurate vouchers of species identification, and some species may be rare or have even gone extinct in the wild (Särkinen *et al.* 2012; Staats *et al.* 2013). The success of similar targeted enrichment strategies with historical specimens for phylogenetic studies has previously been highlighted in other lineages of angiosperms at various scales, including *Arabidopsis thaliana* (Brassicaceae,Staats et al., 2013), *Inga* (Fabaceae, Hart et al., 2016), and Annonaceae.

Our results also suggest that the baits, which were originally designed for tribes Artocarpeae and Moreae in the same family (Gardner *et al.* 2016), worked well in Dorstenieae. In addition to the phylogenetic markers developed from 333 inferred single-copy exons for Moraceae, we retrieved approximately another 200 genes in this study for Dorstenieae. Those assembled untargeted genes had a mean identity to at least one target gene of 84%, and a mean alignment length to at least one target gene of 87% (as a percentage of the untargeted genes). These Moraceae specific baits worked well throughout the entire Moraceae family (Zerega and Gardner 2019). Five untargeted genes were included in our final 102-gene dataset. The most genes retrieved for one taxon was 515 from the outgroup sample *Ficus macrophylla*, suggesting a high

probability that these baits would also work well to explore relationships within Ficeae (*Ficus*), the largest tribe in Moraceae (Couvreur *et al.* 2019).

Phylogenetic relationships in Dorstenieae and Dorstenia

With all the extant genera of Dorstenieae included, this is the most densely sampled phylogenetic study of Dorstenieae to date. The reconstructed relationships (Fig. 2) are generally consistent with previous work (Zerega *et al.* 2005; Clement and Weiblen 2009; Misiewicz and Zerega 2012; Chung *et al.* 2017; Zhang *et al.* 2019), but with stronger support, especially in *Brosimum* s.l. and *Dorstenia* s.l..

Prior to this study, the most densely sampled molecular phylogenetic analysis of *Dorstenia* was provided by Misiewicz and Zerega (2012), based on ITS sequences of 35 taxa (32 species) of *Dorstenia* and seven outgroup species. Our results are similar to those of Misiewicz and Zerega (2012) with respect to shallow-level relationships. However, they differ markedly at a deeper level in that the authors had found three African species (*D. variifolia*, *D. tayloriana* var. *tayloriana*, and *D. cuspidata*) to be nested in the Neotropical clade of *Dorstenia*. In our analyses, *D. variifolia* and *D. tayloriana* were also sampled and they were sister species as in Misiewicz & Zerega (2012), but we found these species nested in African clades with strong support. The difference may be explained by the root setting methods, the variation in each study of both number of genes and density of taxon sampling for both *Dorstenia* species as well as of non-*Dorstenia* species within the tribe Dorstenieae. Most differences between the ML and coalescent trees reconstructed in this study concentrated in the Neotropical *Dorstenia* clade (Fig. 2). This result and the short branches observed in this clade suggest incomplete lineage sorting in the diversification of Neotropical *Dorstenia*

species (Pamilo and Nei 1988).

Divergence times of Dorstenieae and its genera

Runs with or without data showed different results suggesting the data were informative. Runs with different prior settings showed similar estimates, suggesting that our results are robust to various assumptions on rate variation. The crown-group age of Dorstenieae was estimated in the Upper Cretaceous (65.8-79.8 Ma; Table 2), which overlaps with former studies that used fewer genes (Zerega *et al.* 2005; Gardner *et al.* 2017; Zhang *et al.* 2019). The crown-group age of *Dorstenia* was estimated in the Paleocene (50.8-67.3 Ma), which is younger than in Misiewicz and Zerega (2012). Although the estimated stem-group node of Neotropical *Dorstenia* clade in our study was younger than that in Misiewicz and Zerega (2012), the crown-group node of the same clade fell within a similar range in both studies. The difference in the stem-group node may be caused by different topologies and number of genes used to estimate the divergence time. The crown-group age of *Brosimum* s.l. and *Broussonetia* s.l. were estimated in the period from Oligocene to Miocene and the Eocene (14.9-31.1 Ma and 30.2-40.7 Ma, respectively), during which time the whole earth cooled down from Mid-Eocene Climatic Optimum and was warmer than the current climate (Zachos *et al.* 2008).

Biogeographic history of Dorstenieae

Our results suggest that the most recent common ancestor of Dorstenieae was widely distributed in the joint area of continental Africa, Madagascar and Southeast

Asia and Oceania (CDF, Fig. 3, S3) in the Cretaceous (65.8-79.8 Ma), during which time these three areas were already separated from one another (PALEOMAP project, http://www.scotese.com/). Subsequently, at least two dispersals to South America (*Brosimum* s.l. and Neotropical *Dorstenia*) occurred during the evolutionary history of Dorstenieae. In our analyses, the stem-group ages of Dorstenieae (68.9-82.1 Ma), *Brosimum* s.l. (19.6-41.5 Ma) and Neotropical *Dorstenia* (34.6-51.8 Ma) were all estimated to be younger than the separation of South America and continental Africa at ca. 105 Ma (McLoughlin, 2001), suggesting that vicariance caused by Gondwanan breakup is unlikely to have played a role in the diversification of the two Neotropical clades. Long-distance dispersal, which is an indispensable process in Neotropical flora assembly (Hughes *et al.* 2012) may instead explain the origin of the two Neotropical clades in Dorstenieae.

The dispersal (range expansion) at the origin of the Neotropical clade of *Dorstenia* is inferred to have occurred from the Paleocene to early Eocene (time period between the stem and crown-group nodes of Node I, 42.1-62.1 Ma, Fig. 3, S2, S3c), during which time angiosperms were already significantly more diverse in the Neotropics than in earlier time periods, according to palynological evidence (Jaramillo *et al.* 2006). During this period, global temperature increased, leading to the middle Eocene climatic optimum (Zachos *et al.* 2008). The increasing temperature has been suggested as one of the factors for the extension of the Neotropical region in the Eocene (Hughes *et al.* 2012). A larger Neotropical area at that time would have increased the probability of the successful colonization of species from Africa to a suitable habitat in the Neotropics. Curiously, in these reconstructions, *D. elliptica* (an African species sister to the Neotropical *Dorstenia* clade) was inferred to be the result of a back dispersal from South America to continental Africa (Fig. 3, S3c). Whether this somewhat unexpected

result is plausible or an example of pathological behavior of DEC models (Ree and Sanmartín 2018) will require further investigation.

Two independent dispersals from South America to Central/North America in *Dorstenia* were estimated from the Oligocene to Miocene (period from Node II to stem-group node of Clade 1, 16.7-30.6 Ma and period from stem to crown group of Clade 2, 12.0-34.7 Ma), during which time waves of dispersal from South to North America have been found in other lineages (Bacon *et al.* 2015). This result supports the dispersal of plants from South America to Central/North America in the Neogene (Bagley and Johnson 2014). The branch length or time between the stem and crown group of Neotropical *Dorstenia* was very short (less than 10 Ma, Table 2), suggesting the rapid divergence of the clade. The branch length between the stem and crown group of the other Neotropical Dorstenieae lineage (*Brosimum* s.l.) was also short (ca. 12 Ma, Table 2). Therefore, LDD followed by rapid diversification would explain the extant distribution pattern of Neotropical Dorstenieae species. A similar pathway was found in the pantropically distributed tribe Annoneae (Annonaceae) (Thomas *et al.* 2017; Williams *et al.* 2017).

A diversity of seed dispersal modes has been reported in Dorstenieae, including autochory by expulsion or ejection of endocarp in *Dorstenia*, *Bleekrodea*, *Fatoua*, and zoochory in *Brosimum lactescens* and *Trymatococcus amazonicus* (Berg, 2001). *Dorstenia* was suggested to be poorly adapted for LDD (Berg and Hijman 1999). In addition, small *Dorstenia* seeds of forest undergrowth species often germinate shortly after maturity, further reducing the chances of LDD (Berg 2001). Low probability of LDD may explain the single origin of Neotropical *Dorstenia* and the monophyly of the two Central/North America clades. Furthermore, our results suggest a comparatively faster succession of speciation events following establishment of *Dorstenia* in the

Neotropics, suggesting rapid speciation after LDD has been an important process in shaping the origin of Neotropical diversity.

Ree and Sanmartín (2018) recently raised methodological concerns with models including both anagenetic and cladogenetic processes, especially the DEC and DEC+J models. In the main analyses presented here, we focused on the results of DEC-based models (i.e., DEC m0, DEC m1, DEC+J m0, DEC+J m1). We also did reconstructions with DIVA-like and BayArea-like based models. The difference among DEC, DIVA-like and BayArea-like models lie mainly in the cladogenetic process they assume: DEC and DIVA models explore both sympatric speciation and vicariance processes while BayArea only explores sympatric speciation (Nicholas Joseph Matzke 2013). We compared the models with the Akaike information criterion (AIC). Whether or not founder-event speciation was allowed, led the BayArea model to rank from the best to the worst model (Table S2) in both time-stratified and time-constant analyses. AIC has been argued not to be a good criterion for models including both cladogenetic and anagenetic events due to a bias of the likelihood to favor time-independent cladogenetic processes, a problem exacerbated with the introduction of founder event-speciation in the model (Ree and Sanmartín 2018). Zero-estimate for dispersal and strong counter-intuitive unparsimonious reconstruction may be signals for this bias. These two phenomena were not observed in our reconstruction (Table S2, Figure S3). We did not rely on model selection in this study. Instead, we emphasize that all of the models we used (incl. DEC, DIVA-like and BayArea-like based models) led to very similar reconstructions (Fig. 3, Fig. S3).

Taxonomic implications

Some of the currently recognized genera of Dorstenieae and all sections of *Dorstenia* may need to be modified based on the results from our analyses. Using 102 genes (132,753 bp), we obtained the same relationships among *Malaisia*, *Broussonetia* and *Allaeanthus* as a previous study based on one chloroplast and one nuclear gene (Chung *et al.* 2017), providing additional support for the recognition of *Allaeanthus* as a separate genus.

Both *Scyphosyce* and *Utsetela* were found to be nested in *Dorstenia*. *Scyphosyce* is a genus of two species from western Africa (Berg 1977). We sampled both species of *Scyphosyce* in our analyses and found them sister to each other, suggesting their placement within *Dorstenia* was unlikely to be caused by misidentification. The basal grade of the *Dorstenia* s.l. clade was formed by species of sections *Nothodorstenia* and *Xylodorstenia* of the genus *Dorstenia* and by the genera *Scyphosyce* and *Utsetela,* all of which share woody habit and larger seeds, which are referred to by Berg and Hijman (1999) as macrosperms. This basal grade shares other characteristics as well. The inflorescences of most Dorstenieae genera are bisexual (some species of *Broussonetia, Allaeanthus*, and *Fatoua* have unisexual inflorescences). The macrospermous species commonly have only one to a few pistillate flowers per inflorescence which produce one to few large seeds per infructescence). The remaining species of *Dorstenia* are herbaceous and have several to numerous pistillate flowers per inflorescence which can produce numerous smaller seeds (Berg and Hijman 1999). *Utsetela* is a genus of two species from western Africa (Berg 1977; Jongkind 1995). Only one species of *Utsetela* was sampled in this study. Considering the comparatively long branches of *U. gabonensis*, *D. alta*, and *D. mannii*, this relationship could be caused by a long-branch attraction artefact. To test this, we excluded the three species of *Dorstenia* which

clustered with *U. gabonensis* (*D. alta*, and *D. mannii* and *D. turbinata*) and reran the phylogenetic analyses with both methods. *Utsetela gabonensis* was still in the same position (nested in *Dorstenia*) after excluding these three species (results not shown). Sampling the other species, *U. neglecta* (Jongkind, 1995), would be necessary to confirm this relationship and draw any taxonomic conclusions. Some differences among this basal grade of taxa include: tepals of individual flowers in the inflorescences of *Dorstenia* are connate, while they are free in *Scyphosyce* and *Utsetela* (Berg 1977). While all *Scyphosyce*, *Utsetela* and *Dorstenia* have drupe(let) of fruit, their receptacles are different in shape and the filaments are far more elongated in *Utsetela* then in other two genera (Berg 1977; Berg and Hijman 1999). Despite some differences, similarities in woody habit, fruit type, macrospermy, and the phylogenetic reconstruction presented here suggest that merging *Scyphosyce* and *Utsetela* into *Dorstenia* may be a reasonable taxonomic outcome to preserve the monophyly of *Dorstenia*. An alternative option would be to separate those clades into two separate genera as elaborated on below.

Nine sections were proposed by Berg and Hijman (1999) in *Dorstenia* based on inflorescence, habit and life form (i.e. geophytes, phanerophytes, hemicryptophytes) characters. Our sampled species represented all the sections except *Bazzemia*, which consists of a single species in Mozambique. None of the eight sampled sections were found to be monophyletic in this study (Figure S1) and the phylogeny can help inform future subgeneric classification of *Dorstenia*. Of particular interest to consider taxonomically are the two most basal *Dorstenia* clades that include members of the sections *Xylodorstenia* and *Nothodorstenia* as well as species from two other genera (*Scyphosyce* and *Utsetela*). Misiewicz and Zerega (2012) did not include any species of *Scyphosyce* and *Utsetela,* and they found sections *Xylodorstenia* monophyletic. This is not the case in our reconstruction (Figure S1), and is likely due to our increased

taxonomic sampling and use of many more genes, and our reconstructions from both ML and coalescent approaches strongly supported the non-monophyletic status of these sections. An alternative taxonomic solution to sinking *Scyphosyce* and *Utsetela* into *Dorstenia* is to include some *Dorstenia* species (most of section *Nothodorstenia* and at least one member of section *Xylodorstenia – D. angusticornis*) into the genus *Scyphosyce. Dorstenia africana*, *D. kameruniana*, *D. oligogyna*, *D. djettii* and *D. dorstenioides* (the former four all included in section *Nothodorstenia*) were once classified as genus *Craterogyne* (Lanjouw 1935). *Dorstenia dorstenioides*, which has been proposed as the link between sections *Xylodorstenia* and *Nothodorstenia* (Berg and Hijman 1999), was excluded in the present study due to the low number of genes represented in our main analyses (Table S1). As *Scyphyosyce* (Baillon 1875)is an older name than *Craterogyne, Scyphosyce* would take priority for the name of a new genus.

Regarding the clade containing *Utsetela,* some of the *Dorstenia* species in that clade were recently transferred to a new genus (*Maria*) established by Machado Vianna f. et al. (2013), comprising four species in *Dorstenia* section *Xylodorstenia* (*D. alta*, *D. angusticornis*, *D. scaphigera*, *D. turbinata*). *Maria* was later found to be a homonym and renamed *Hijmania* (Vianna Filho *et al.* 2016). It may be necessary to include those species into the same genus under the name *Utsetela* (Pellegrin 1928), which has priority, but until more complete taxon sampling is completed, we do not presently propose any taxonomic changes. One of the woody African macrospermous species that warrants further attention is *D. elliptica.* It was included in section *Nothodorstenia* by Berg (1978) because it had bracts resembling other members of that section. However both Misiewicz & Zerega (2012) and the present study, found *D. elliptica* to be sister to all Neotropical *Dorstenia.*

Among herbaceous species of *Dorstenia*, the presence of bracts on receptacles has

traditionally been used to distinguish among sections *Emygdioa*, *Dorstenia*, *Lecanium* on the one hand and sections *Acauloma*, *Bazzemia*, *Lomatophora*, *Kosaria* on the other hand (Berg and Hijman 1999). Although none of these sections were found to be monophyletic, the Neotropical clade contains all of the species with bracteate receptacles except *D. picta* (Figure S1). Our results suggest that traditional morphological characters for sectional delimitation within *Dorstenia* do not hold up to molecular phylogenetic scrutiny, that a close examination of alternative characters is needed, and a new intrageneric classification is warranted.

Conclusion

The targeted enrichment sequencing strategy, paired with the HybPiper pipeline, proved to be an effective approach at reconstructing phylogenetic relationships in Dorstenieae using herbarium specimens. Further molecular and morphological work will be required before solving some of the taxonomic issues highlighted in this study, such as sinking *Utsetela* and *Scyphosyce* into *Dorstenia* or separating some *Dorstenia* species into either the genera of *Utsetela* or *Scyphosyce*. Dorstenieae as a whole may have originated in the joint area of continental Africa, Madagascar and Asia-Oceania area, followed by at least two independent colonizations of South America (*Brosimum* s.l. and *Dorstenia* s.l.). Some species in these two clades further dispersed to Central/North America. The mechanical processes for the long-distance dispersal of species of Dorstenieae remains an enigma. More studies on pollination and dispersal in this tribe will be required to further elucidate the biogeographic history of this group. The development of new biogeographic models and new model selection procedures will also be essential to help to clear the biogeographic history of Dorstenieae and other

pantropically distributed lineages. The robust and most densely sampled Dorstenieae phylogeny presented here will be a valuable resource for further studies on character evolution in this fascinating tribe and will assist with future taxonomic revisionary work.

ACKNOWLEDGEMENTS

We thank the Field Museum (F), the Missouri Botanical Garden (MO), the New York Botanical Garden (NY), and the Muséum national d'Histoire naturelle (P) for sharing herbarium material; Tracy Misiewicz and Caroline Storer for assistance with DNA extraction; Charles Foster for providing helpful advice on running MCMCTree; and Santiago Ramirez-Barahona for information on fossil calibrations. This work was supported by a China Scholarship Council (CSC) PhD grant (grant No. 201506140077) and a research grant from the International Association for Plant Taxonomy (IAPT) to Q.Z. Sequencing costs were covered by Agence Nationale de la Recherche grant ANR-12-JVS7-0015-01 to H.S.

LITERATURE CITED

Antonelli A, Sanmartín I. **2011**. Why are there so many plant species in the Neotropics? *Taxon* **60**: 403–414.

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* **112**: 6110–6115.

Bagley JC, Johnson JB. **2014**. Phylogeography and biogeography of the lower Central American Neotropics: diversification between two continents and between two seas. *Biological Reviews* **89**: 767–790.

Baillon H. **1875**. Stirpes exoticae novae. *Adansonia* **11**: 292–312.

Bankevich A, Nurk S, Antipov D, *et al.* 2012. SPAdes: a new genome assembly algorithm and

its applications to single-cell sequencing. *Journal of computational biology : a journal of computational molecular cell biology* **19**: 455–477.

Berg CC. **1977**. Revisions of African Moraceae (excluding *Dorstenia*, *Ficus*, *Musanga* and *Myrianthus*). *Bulletin du Jardin botanique national de Belgique/Bulletin van de Nationale Plantentuin van Belgie* **47**: 267–407.

Berg CC. **2001**. Moreae, Artocarpeae, and *Dorstenia* (Moraceae): with introductions to the family and *Ficus* and with additions and corrections to Flora Neotropica Monograph 7. *Flora Neotropica* **83**: 1-346.

Berg CC, Hijman MEE. **1999**. The genus *Dorstenia* (Moraceae). *Ilicifolia* **2**: 1–211.

Bouckaert R, Heled J, Kühnert D, *et al.* 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLOS Computational Biology* **10**: e1003537.

Chandler MEJ. **1961**. *Flora of the Lower Headon beds of Hampshire and the Isle of Wight*. British Museum (Natural History).

Chandler MEJ. **1962**. *The Lower Tertiary Floras of Southern England: Flora of the pipe-clay series of Dorset (lower Bagshot)*. order of the Trustees of the British Museum.

Chase MW, Thijs KW, Kamau P, Fay MF. **2013**. *Dorstenia christenhuszii* (Moraceae), a new species from the Taita Hills, Kenya. *Phytotaxa* **81**: 45–48.

Christenhusz MJM, Chase MW. **2013**. Biogeographical patterns of plants in the Neotropics–dispersal rather than plate tectonics is most explanatory. *Botanical Journal of the Linnean Society* **171**: 277–286.

Chung K-F, Kuo W-H, Hsu Y-H, Li Y-H, Rubite RR, Xu W-B. **2017**. Molecular recircumscription of *Broussonetia* (Moraceae) and the identity and taxonomic status of *B. kaempferi* var. *australis*. *Botanical studies* **58**: 11.

Clement WL, Weiblen GD. **2009**. Morphological evolution in the mulberry family (Moraceae). *Systematic Botany* **34**: 530–552.

Collinson ME. **1989**. The fossil history of the Moraceae, Urticaceae (including Cecropiaceae), and Cannabaceae. *Crane, P, R,, Blackmore, S. (Ed s.), Evolution, systematics, and fossil history of the Hamamelidae* **2**: 319–339.

Couvreur TLP, Helmstetter AJ, Koenen EJM, *et al.* 2019. Phylogenomics of the Major Tropical Plant Family Annonaceae Using Targeted Enrichment of Nuclear Genes. *Frontiers in Plant Science* **9**: 1941.

Couvreur TLP, Pirie MD, Chatrou LW, *et al.* 2011. Early evolutionary history of the flowering plant family Annonaceae: steady diversification and boreotropical geodispersal. *Journal of Biogeography* **38**: 664–680.

Fisher AE, Hasenstab KM, Bell HL, Blaine E, Ingram AL, Columbus JT. **2016**. Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. *Molecular phylogenetics and evolution* **105**: 1–14.

Foster CSP, Sauquet H, van der Merwe M, McPherson H, Rossetto M, Ho SYW. **2017**. Evaluating the impact of genomic data and priors on Bayesian estimates of the angiosperm evolutionary timescale. *Systematic Biology* **66**: 338–351.

Gardner EM, Johnson MG, Ragone D, Wickett NJ, Zerega NJC. **2016**. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. *Applications in plant sciences* **4**: 1600017.

Gardner EM, Sarraf P, Williams EW, Zerega NJC. **2017**. Phylogeny and biogeography of *Maclura* (Moraceae) and the origin of an anachronistic fruit. *Molecular phylogenetics and evolution* **117**: 49–59.

Hart ML, Forrest LL, Nicholls JA, Kidner CA. **2016**. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* **65**: 1081–1092.

Ho SYW, Phillips MJ. **2009**. Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Systematic Biology* **58**: 367–380.

Hughes CE, Pennington RT, Antonelli A. **2012**. Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* **171**: 1–18.

Jaramillo C, Rueda MJ, Mora G. **2006**. Cenozoic plant diversity in the Neotropics. *Science* **311**: 1893–1896.

Johnson MG, Gardner EM, Liu Y, *et al.* 2016. HybPiper: extracting coding sequence and introns for phylogenetics from high-throughput sequencing reads using target enrichment. *Applications in plant sciences* **4**: 1600016.

Jongkind CCH. **1995**. Novitates gabonenses (24) A New Species in *Utsetela* (Moraceae). *Bulletin du Jardin botanique national de Belgique/Bulletin van de Nationale Plantentuin van Belgie* **64**: 179–181.

Kates HR, Johnson MG, Gardner EM, Zerega NJC, Wickett NJ. **2018**. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American journal of botany* **105**: 404–416.

Lanjouw J. **1935**. Studies in Moraceae I. The genera *Trymatococus* Poepp. et Endl. and *Graterogyne* Lanj. *Recueil des Travaux Botaniques Néerlandais* **32**: 276.

Leal ME. **2014**. *Dorstenia luamensis* (Moraceae), a new species from eastern Democratic Republic of Congo. *PhytoKeys* **42**: 49–55.

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. **2011**. Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**: 124.

Lemmon EM, Lemmon AR. **2013**. High-throughput genomic data in systematics and phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* **44**: 99–121.

Machado AFP, Marcelo Filho DMV. **2012**. *Dorstenia romaniucii* (Moraceae), a new species from the Brazilian Atlantic Rain Forest. *Systematic Botany* **37**: 451–455.

Machado AFP, Pereira JF, Carauta JPP. **2014**. *Dorstenia triseriata* (Moraceae) a new and

endangered species from Brazil. *PhytoKeys* **38**: 31–35.

Machado Vianna f. MD, Maia VH, de Freitas Mansano V, da Costa AF. **2013**. *Maria*,a new genus of Moraceae. *Albertoa* **38**: 289–292.

Matzke N J. **2013**. BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R scripts, CRAN: The Comprehensive R Archive Network, Berkeley, CA. *cran. r-project. org/web/packages/BioGeoBEARS/index. html*.

Matzke Nicholas Joseph. **2013**. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* **5**: 242–248.

Mccoy T, Massara M. **2008**. *Dorstenia lavrani*: a dioecious new species from northern Somaliland. *Cactus and Succulent Journal* **80**: 78–82.

McKain MR, Johnson MG, Uribe-Convers S, Eaton D, Yang Y. **2018**. Practical considerations for plant phylogenomics. *Applications in plant sciences* **6**: e1038.

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

Mehrotra RC, Prakash U, Bande MB. **1984**. Fossil woods of *Lophopetalum* and *Artocarpus* from the Deccan Intertrappean Beds of Mandla district, Madhya Pradesh, India. *Palaeobotanist* **32**: 310–320.

Miller MA, Pfeiffer W, Schwartz T. **2010**. Creating the CIPRES Science Gateway for inference of large phylogenetic trees In: *Gateway Computing Environments Workshop (GCE), 2010*. IEEE, 1–8.

Mirarab S, Warnow T. **2015**. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31**: i44–i52.

Misiewicz TM, Zerega NC. **2012**. Phylogeny, biogeography and character evolution of *Dorstenia* (Moraceae). *Edinburgh Journal of Botany* **69**: 413–440.

Mitchell N, Lewis PO, Lemmon EM, Lemmon AR, Holsinger KE. **2017**. Anchored phylogenomics improves the resolution of evolutionary relationships in the rapid radiation of *Protea* L. *American journal of botany* **104**: 102–115.

Montes C, Bayona G, Cardona A, *et al.* 2012. Arc-continent collision and orocline formation: Closing of the Central American seaway. *Journal of Geophysical Research: Solid Earth* **117**.

Morrone JJ. **2014**. Biogeographical regionalisation of the Neotropical region. *Zootaxa* **3782**: 1–110.

Pamilo P, Nei M. **1988**. Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**: 568–583.

Pellegrin F. **1928**. *Utsetela* Pellegrin, genre nouveau d'Urticacées-Artocarpées du Gabon. *Bulletin de la Société Botanique de France* **75**: 672–674.

R Core Team. **2018**. R: A Language and Environment for Statistical Computing. R foundation for statistical computing, Vienna, Austria. http://www.R-project.org/.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. **2018**. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **10**: 901–904.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. **2011**. MACSE: multiple alignment of coding sequences accounting for frameshifts and stop codons. *PloS one* **6**: e22594.

Ree RH, Sanmartín I. **2018**. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* **45**: 741–749.

Rohwer JG, Berg CC. **1993**. Moraceae In: *Flowering Plants Dicotyledons*. Berlin Heidelberg: Springer.

Rzepecky A. **2016**. *Dorstenia horwoodii* Rzepecky sp. nov. from Nudum to Novum, a Fortyish Year Hiatus. *Cactus and Succulent Journal* **88**: 66–74.

dos Santos A, Machado AFP, Mansano VDEF, Neto SR. **2013**. *Dorstenia acangatara* (Moraceae), a new and threatened species from Southeastern Brazil. *Phytotaxa* **118**: 29–34.

dos Santos A, Neto SR. **2012**. A new species of *Dorstenia* (Moraceae) from southeastern Brazil. *PhytoKeys* **12**: 47–51.

Särkinen T, Staats M, Richardson JE, Cowan RS, Bakker FT. **2012**. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PloS One* **7**: e43808.

Sauquet H. **2013**. A practical guide to molecular dating. *Comptes Rendus Palevol* **12**: 355–367.

Slater GSC, Birney E. **2005**. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* **6**: 31.

Slik JWF, Arroyo-Rodríguez V, Aiba S-I, *et al.* 2015. An estimate of the number of tropical tree species. *Proceedings of the National Academy of Sciences* **112**: 7472–7477.

Staats M, Erkens RHJ, van de Vossenberg B, *et al.* 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PLoS One* **8**: e69189.

Stamatakis A. **2014**. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stanke M, Steinkamp R, Waack S, Morgenstern B. **2004**. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic acids research* **32**: W309–W312.

Tandang DN, Tadiosa ER, Gardner EM. **2017**. *Sloetia* (Moraceae): a new generic record for the Philippines. *Telopea* **20**: 69–73.

Thomas DC, Tang CC, Saunders RMK. **2017**. Historical biogeography of *Goniothalamus* and Annonaceae tribe Annoneae: dispersal–vicariance patterns in tropical Asia and intercontinental tropical disjunctions revisited. *Journal of Biogeography* **44**: 2862–2876.

Vianna Filho MDM, Maia VH, de Freitas Mansano V, da Costa AF. **2016**. *Hijmania*, a replacement name for *Maria* (Moraceae). *Phytotaxa* **247**: 97–98.

Williams EW, Gardner EM, Harris R, Chaveerach A, Pereira JT, Zerega NJC. **2017**. Out of Borneo: biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany* **119**: 611–627.

Wu Z, Raven PH, Hong D. **2003**. Ulmaceae through Basellaceae. *Flora of China* **5**: 21-73.

Xi Z, Liu L, Rest JS, Davis CC. **2014**. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology* **63**: 919–932.

Yang Z. **2007**. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**: 1586–1591.

Zachos JC, Dickens GR, Zeebe RE. **2008**. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **451**: 279–283.

Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. **2005**. Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution* **37**: 402–416.

Zerega NC, Gardner EM. **2019**. Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. *Phytotaxa* **338**: 253–265.

Zerega NJC, Supardi N, Motley TJ. **2010**. Phylogeny and recircumscription of Artocarpeae (Moraceae) with a focus on *Artocarpus*. *Systematic Botany* **35**: 766–782.

Zhang Q, Onstein RE, Little SA, Sauquet H. **2019**. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. *Annals of Botany* **123**: 191–204.

SUPPLEMENTARY INFORMATION

Figure S1. Maximum Likelihood tree (a) and species tree (b) with bootstrap support value and tip names (including section names in *Dorstenia*).

Figure S2. Divergence time estimate by MCMCTree with two different prior settings and r8s: combined results of two independent runs with prior set1 (a) and set2 (b) by MCMCTree and r8s (c).

Figure S3. Biogeographic reconstruction with unconstrained (model 0) or constrained (model 1) models with DEC, DIVA-like and BayArea-like based models with internal nodes labelled with discrete states: a-c, DEC; d, illustrated with pie chart on internal

nodes for the result of time-stratified DEC+J model; e-h, Bayarea-like; i-l, DIVA-like model with model 0 and 1, with or without founder-event speciation process (+J), detail of DEC+J model with nodes labelled with discrete states see Figure 3.

Table S1. List of specimens collected in this study.

Table S2. AIC of biogeographic reconstruction with time-constant (model0, a) or time-stratified (model1, b) DEC, DIVA-like and BayArea-like based models.

Table S3. List of ancestral distribution area estimated for several nodes by BioGeoBEARS. a) time-constant models (model0); b) time-stratified models (model1) .

CHAPTER III. BREEDING SYSTEM EVOLUTION AND CLIMATE IN *FICUS*

This study has not been submitted to any journal yet and was conducted in collaboration with Renske Onstein and Hervé Sauquet.

ABSTRACT

- **Background and Aims** Physiological and ecological differences of breeding systems may lead to different distribution patterns. Two breeding systems, gynodioecy and monoecy, occur in *Ficus* (ca. 750 spp., Moraceae), an ecologically important genus distributed in all major tropical regions. We used *Ficus* as a model to investigate the niche difference of different breeding systems.
- **Methods** Richness maps of 183 species of *Ficus* (105 monoecy, 78 gynodioecy) were plotted. To investigate the niche difference of the two breeding systems in *Ficus*, regression was conducted with fourteen climatic variables against breeding systems with generalized linear models (GLM) and generalized estimating equations (GEE). Ancestral states of precipitation niche were reconstructed to test its relationship with the transitions of breeding systems in the history of *Ficus*.
- **Key Results** Although gynodioecious and monoecious *Ficus* overlap in their distribution, gynodioecious *Ficus* does not occur in the Neotropics and monoecious *Ficus* is distributed in drier environments. The distribution patterns of both breeding systems fit the summer positions

of the intertropical convergence zone on continents. A significantly positive relationship of precipitation and gynodioecy was supported by GLM but not GEE analyses.

- **Conclusions** Niche differences were found in *Ficus* with different breeding systems but may be the result of phylogenetic effect rather than historical correlation. This study sheds light on understanding of the origin, diversification, and distribution pattern of the two breeding systems in *Ficus*.

Key words: breeding systems, *Ficus*, monoecy, gynodioecy, generalized estimating equations (GEE), generalized linear models (GLM), climatic variables, intertropical convergence zone (ITCZ)

INTRODUCTION

The genus *Ficus* (ca. 750 spp., Moraceae) has long intrigued researchers given its taxonomic and ecological diversity and the coevolution with its wasp pollinators (Agaonidae, Cruaud et al. 2012). Two different breeding systems have been observed in *Ficus* so far: monoecy and gynodioecy, the latter is functionally similar to dioecy in this genus (hereafter referred to as dioecy, Cook and Rasplus 2003). Because of the physiological and ecological differences between breeding systems, *Ficus* species of the two breeding systems differ not only in their inflorescences (syconia), but are also expected to differ in dispersal ability, distribution and other traits (Harrison and Yamamura 2003; Nazareno *et al.* 2013; Yang *et al.* 2015).

Ficus is a typical tropical clade, but its global distribution is uneven across tropical areas (Fig. 1). Approximately 120, 105 and 510 species are known from the Neotropics, Africa (including Madagascar and adjacent Indian Ocean islands, and the Arabian Peninsula) and Asia-Oceania, respectively (Berg *et al.* 2005). Broad-scale distribution patterns also differ between breeding systems (Fig. 1). Dioecious *Ficus* does not occur in the Neotropics (Cruaud *et al.* 2012) and is mainly distributed in Asia-Oceania. Several dioecious species of *Ficus* (e.g., *F. sarmentosa*, *F. heteromorpha*, *F. erecta*, *F. coronata*, *F. tikoua* and *F. pumila*) have extended their distribution into more temperate areas in Asia (Berg *et al.* 2005; Zhao *et al.* 2014). Monoecious *Ficus* occurs across all tropical realms and also occurs in dryer areas, for example in dry parts of Atlantic Costal regions or (semi-)desert of northern Mexico (Berg 2001). However, the underlying physiological and ecological causes for these distinct distribution patterns remain debated.

Dispersal distances of *Ficus* species and their pollen are strongly influenced by dispersal abilities of their wasp pollinators. Pollinators of monoecious *Ficus* were observed in the canopy layer in both Bornean and Neotropical forests, whereas the pollinators of dioecious *Ficus* occur

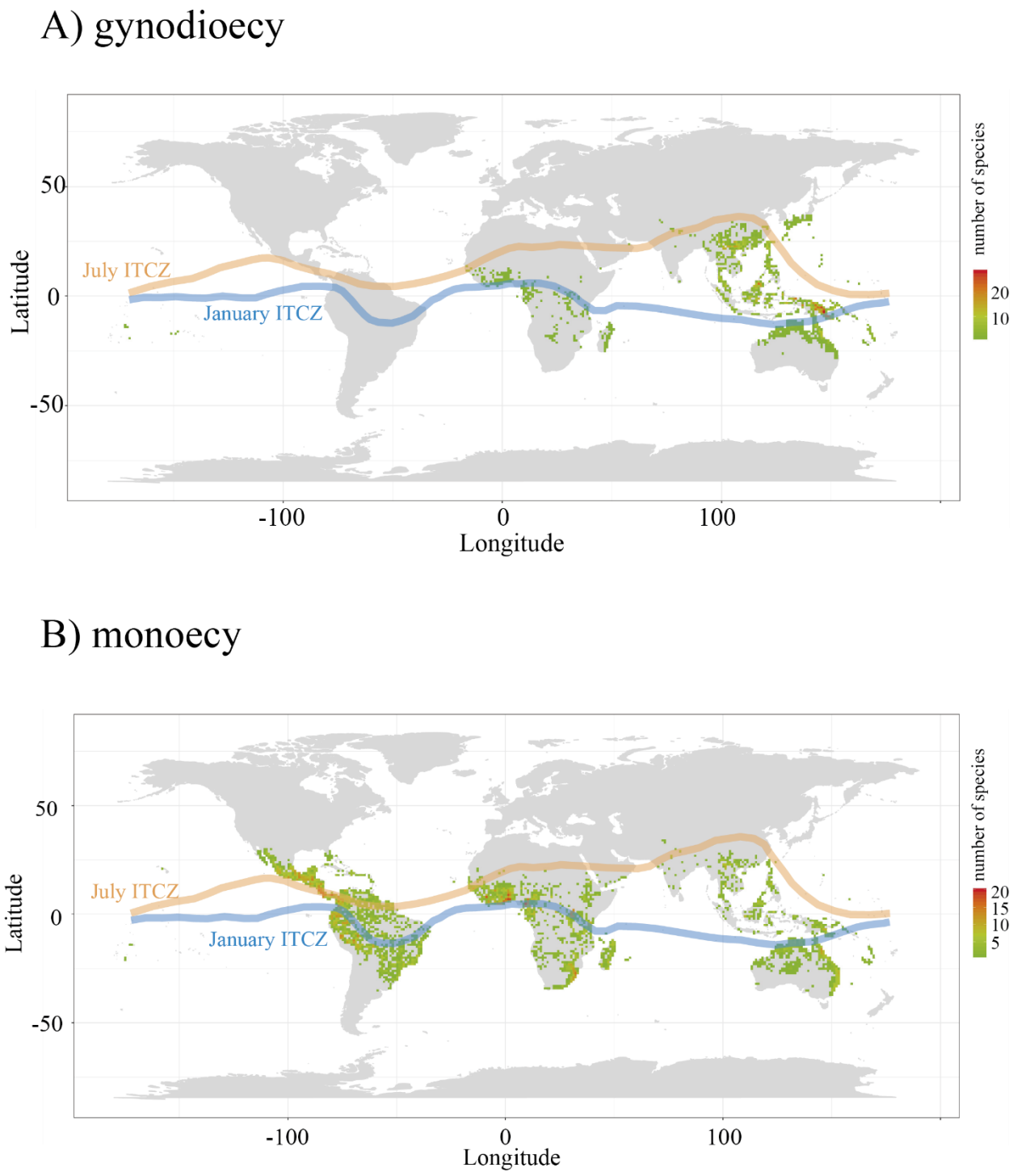


Figure 1. Distribution map and species richness of *Ficus* according to breeding system, based on cleaned occurrence records from GBIF: A) dioecy; B) monoecy.

most often in lower canopy layers (Hespenheide 1975; Harrison 2003). Considering the higher wind speed in the canopy, pollinating wasps of monoecious hosts would have longer dispersal distances than the counterpart of dioecious *Ficus*, suggesting that the pollen of monoecious *Ficus,* on average, disperse farther (Compton *et al.* 2000). The pollinators of *Ficus sycomorus* (monoecious) were found to disperse tracking the seasonal nighttime wind and the average dispersal distance was 88.6 km in Africa (160 km maximum, Ahmed et al. 2009). As a result of more frequent long-distance dispersal of pollinators (pollen), monoecious *Ficus* populations showed lower levels of genetic structure than dioecious *Ficus* populations (Nazareno *et al.* 2013). Besides pollen dispersal, the figs (syconia) are mainly eaten and dispersed by a wide range of birds, arboreal mammals and fruit bats, and while considering the diverse habits of *Ficus*, it is difficult to address differences in seed dispersal distances between breeding systems (Shanahan *et al.* 2001).

In a previous study, ancestral breeding systems were reconstructed in *Ficus* and Moraceae (Zhang *et al.* 2019). Although some uncertainty remained, depending on models and approaches, monoecy was shown to be the likely ancestral state in *Ficus*, and later evolved into dioecy once or twice, with at least three subsequent reversals to monoecy. However, the conditions in which these transitions occurred remain unclear. Understanding the ecological difference of the two breeding systems in *Ficus* will help us better understand the environmental conditions of breeding system transitions not only in *Ficus* but also in angiosperms. Phylogenetic comparative methods provide the tools to investigate the relationship between breeding systems and environmental niche (Table 1). Phylogenetic independent contrasts (Felsenstein 1985) and directional comparative methods (Lauder 1981) were first used in comparative biology to consider the phylogenetic relationship of the data when testing the correlations among traits. These two methods were later found to be very similar, as both of them estimate the evolutionary regression coefficient (Pagel 1993). The

Table 1. Estimates for interaction effects of climatic variables and breeding systems in *Ficus* from the regressions by generalized linear models (GLM) and generalized estimating equations (GEE).

Code	Variable	GLM	GLM coefficient	GEE
bio1	Annual mean temperature	not sig.	/	not sig.
bio5	Max temperature of warmest month	**	-0.022318	not sig.
bio6	Min temperature of coldest month	not sig.	/	not sig.
bio8	Mean temperature of wettest quarter	not sig.	/	not sig.
bio9	Mean temperature of driest quarter	not sig.	/	not sig.
bio10	Mean temperature of warmest quarter	not sig.	/	not sig.
bio11	Mean temperature of coldest quarter	not sig.	/	not sig.
bio12	Annual precipitation	***	0.10051	not sig.
bio13	Precipitation of wettest month	***	0.2675	not sig.
bio14	Precipitation of driest month	***	0.23656	not sig.
bio16	Precipitation of wettest quarter	***	0.17727	not sig.
bio17	Precipitation of driest quarter	***	0.13509	not sig.
bio18	Precipitation of warmest quarter	***	0.28233	not sig.
bio19	Precipitation of coldest quarter	*	0.04934	not sig.
PC1	Principal component 1	***	-0.2242	not sig.
PC2	Principal component 1	***	-0.29111	not sig.
PC1+2	Combination of principal component 1 and 2	***	-0.25431 (PC1), -0.34284 (PC2)	not sig.

Significance codes: '***' P<0.0001; '**' P<0.001; '*' P<0.01; '.' P<0.05; 'not sig.' not significant.

directional comparative approach tests the correlation by first estimating the ancestral state for each internal node. Considering the uncertainty of topology and ancestral state reconstructions, Pagel (1994) proposed a maximum likelihood approach which estimates the correlation of continuous traits without reconstructing ancestral states. Another approach based on general linear models called phylogenetic generalized least squares (PGLS) was developed for investigating the relationships among continuous traits while taking the entire phylogeny into account (Grafen 1989; Martins and Hansen 1997; Pagel 1997; Mundry 2014). However, constrained by the algorithm, the PGLS approach can only deal with continuous variables. This deficiency was solved by the development of the generalized estimating equations (GEE) approach, which can analyze both continuous and discrete data and more complex models (Paradis and Claude 2002).

Our study aims to understand how niche differences (temperature and precipitation) between

breeding systems may have influenced the global distribution of *Ficus.* Using novel approaches to study correlated evolution, we are now able to address four key questions: 1) are dioecious species of *Ficus* distributed across lower temperatures (of the coldest month) than monoecious species of *Ficus*? 2) do monoecious species of *Ficus* occur more often in drier environments than species of dioecious *Ficus*? 3) is the evolution of dioecy in *Ficus* associated with the colonization of cooler niches? 4) is the evolution of monoecy in *Ficus* associated with the colonization of dry climatic niches)? To investigate these four key questions, we assembled and cleaned a comprehensive database of *Ficus* occurrence records, extracted climatic niche variables for the species, and conducted phylogenetic comparative analyses with the GEE approach for climatic variables against the breeding systems in *Ficus*. This study sheds light on the understanding of distribution patterns and niche differences in *Ficus* according to breeding systems, and helps to understand the relationship of breeding systems and distribution in Moraceae and angiosperms more generally in a climate changing world.

MATERIALS AND METHODS

Occurrence records download and cleaning

We chose 200 species of *Ficus* (Cruaud *et al.* 2012; Zhang *et al.* 2019) to make the following tests. A total of 102,486 occurrence records were downloaded from the Global Biodiversity Information Facility (GBIF) with R package rgbif (Chamberlain and Boettiger 2017; Chamberlain *et al.* 2019) by searching by species name. We then flagged and excluded potentially erroneous coordinates (e.g. both latitude and longitude are zero; same latitude and longitude; coordinate on the sea; coordinate as the center or capital city of one country and so on) using R package

CoordinateCleaner (Zizka *et al.* 2018). In addition, we chose records later than the year 1945; of which precision was higher than 10 km; of which sources were "human observation", "observation" or "preserved specimen". Given both the presence of invasive and cultivated records in GBIF, we finally skimmed the records manually to exclude records outside the reported natural distribution area of each species by searching local floras and additional literature (Table S1). For instance, records in New Zealand were excluded as *Ficus* is naturally absent there (Gardner and Early 1996). Considering the long history and wide range of the cultivated plant *Ficus carica*, it is hard to determine the original native distribution of this species (Ghada *et al.* 2010), which was therefore excluded from our analyses. Sixteen additional species were excluded due to either the lack of GBIF data (*Ficus arfakensis*, *Ficus bullenei*, *Ficus dugandii*, and *Ficus palmeri*) or synonymy (e.g., two species included as distinct tips in our dataset, but treated as synonyms in GBIF). As a result, our final dataset consists of 183 species of Ficus with GBIF data, including 23 Neotropical species (all monoecious), 42 African species (38 monoecious, 4 dioecious), and 118 species in Asia-Oceania (44 monoecious, 74 dioecious). After all of the above steps of data cleaning, we plotted separate richness maps for each breeding system in a spatial resolution of one minute using package SpeciesGeoCodeR (Töpel *et al.* 2016). Species assignment to breeding systems follows Zhang et al. (2019). The complete dataset of cleaned occurrence records is available from supplementary data 1.

Principal component analysis and regression

Climatic data were retrieved from the WorldClim database (Fick and Hijmans 2017) at the spatial resolution of 2.5 minutes for the cleaned occurrence records obtained in the above steps. Because some of these variables were calculated from others (e.g. mean diurnal range, BIO2 in

WorldClim), we did not include them in the following analyses. The fourteen included variables and their names used in the analyses are listed in Table 1. The median of each variable of each species was calculated to conduct the principal component analysis (PCA). Considering the value of precipitation data are one order of magnitude larger than those of temperature data, we normalized climatic variables related with precipitation (bio12-19) by taking their square root before PCA. Principal components 1 and 2 (PC1 and PC2) were calculated with R package ade4 (Chessel *et al.* 2004; Dray and Dufour 2007; Dray *et al.* 2007; Bougeard and Dray 2018).

We then conducted two types of regression, differing fundamentally in whether or not phylogeny was taken into account. First, we used generalized linear models (GLM) to understand how the species' climatic niches influence breeding system variation, without taking phylogeny into account. To this end, breeding system was regressed against PC1, PC2, PC1 and 2, and for each individual climatic variable using logistic regressions by function compar.gee in R package ape (Paradis and Schliep 2018). However, the correlation between climatic niches and breeding systems may simply result from phylogenetic dependence (i.e. inherited from the ancestral species (Felsenstein 1985). Thus, we used the generalized estimating equations (GEE) approach with logistic distribution (breeding system is a discrete tratit, Paradis and Claude 2002) to test the correlation of PC1, PC2, PC 1 and 2, and each climatic variables versus breeding systems, while correcting for the phylogenetic effect. For these analyses, we used the ultrametric (dated) Maximum Clade Credibility tree from Zhang et al. (2019) but only kept the tips of the 183 *Ficus* species in the geographic dataset. The results from the two regression methods were then compared.

Ancestral state reconstruction of climatic variables

Results from the regression analyses suggested a possible relationship of precipitation with breeding system. To further investigate and visualize which transitions in breeding systems were associated with change of precipitation in the history of *Ficus*, we reconstructed the ancestral value of precipitation of the warmest quarter (bio18) with a Brownian motion model using the maximum likelihood approach implemented in the fastAnc function of R package phytools (Revell 2012). Here, we aimed to test the influence of precipitation on the distribution of *Ficus*. The environmental data in our dataset were retrieved from the coordinates of the occurrence records, suggesting they are all in the tolerant ranges of *Ficus*. Both temperature and precipitation are important to the development to both the *Ficus* and their pollinating wasps. We selected bio18 to control one factor and observe the performance of the other. When analyzing precipitation of the warmest quarter (bio18), temperature is the warmest in the period (year), hence we control the temperature as the best and are testing the relationship of precipitation with breeding system.

Additional analyses

Two additional analyses were conducted to further explore and characterize potential signal for niche difference between breeding systems in the dataset. Firstly, considering the fact that dioecious *Ficus* extended more poleward than monoecious *Ficus* only in Asia, we conducted all the analyses above with a subset of the dataset including only Asia-Oceania *Ficus* (118 species). These analyses were prompted by the observation that dioecious species of *Ficus* tend to extend to higher latitudes than monoecious species only in this region. Hereafter, we refer to this dataset was named as subsetAsia. Secondly, we built a separate species-climate dataset by extracting the

climate variables of the geographic coordinates where each species has the lowest precipitation of the warmest quarter (bio18), rather than the media value of the entire set of occurrence records of each species used in the main dataset. The rationale for this analysis stems from the fact that the median value of climatic variables may not adequately represent the climatic tolerance of the species to extreme values. Hereafter, we refer to this dataset as bio18min. For both the subsetAsia and bio18min datasets, we conducted principal component analysis, regression (GLM, GEE) and ancestral state reconstruction in the same way as described above.

RESULTS

Global distribution pattern of Ficus according to breeding systems

There were 60,432 occurrence records representing 183 species (105 monoecious, 78 gynodioecious, Table S1) in the final dataset. Most of the distribution area of monoecious *Ficus* overlapped with that of dioecious *Ficus* in Africa (incl. Madagascar) and Asia-Oceania (Fig. 1). Monoecious species of *Ficus* were also observed in comparatively drier areas of Africa and Australia (Fig. 1). In Asia, dioecious *Ficus* extends further North than monoecious *Ficus*, reaching the temperate zone, including Japan and central China (Fig. 1). Continental Southeast Asia (including southern China), northern Borneo and New Guinea hold the highest concentration of dioecious species of *Ficus* (Fig. 1A). Central America, the tropical Andes, West and Central Africa, eastern South Africa, and the Australian Wet Tropics are the hotspots of monoecious *Ficus* (Fig. 1B).

Principal component analysis and regressions

In the main analysis (median of each climatic variable), principal component 1 and 2 (PC1 and PC2) explained 45.6% and 34.4%, respectively, of the variation in climatic niche variables across all *Ficus* (Fig. 2). PC1, PC2, their combination were significantly negatively correlated with dioecy. In addition, all of the single variables related with precipitation (bio 12-19) were significantly positively correlated with dioecy in regressions using generalized linear models (GLM) analyses (Table 1), and thus negatively associated with monoecy. Conversely, most variables related to temperature were not significantly associated with breeding systems, except the maximum temperature of warmest month (bio5), which was negatively correlated with dioecy. This means that monoecious species occur in drier (and sometimes warmer) climates than dioecious *Ficus*. In the regression using the generalized estimating equations (GEE) approach, which considers the phylogenetic non-independence of the data, none of the principle components and their combination, nor any single climatic variable was significantly correlated with breeding system variation. Results were almost identical in the two additional analyses (subsetAsia and bio18min), with a few exceptions (Table S2). In GLM analyses of both the subsetAsia and bio18min dataset, the same direction of significance was found for all variables except bio19 with the bio18min dataset. However, the strength of significance was weaker for all variables of the two additional datasets than in the analyses of the main dataset. No significance was found for any variable in GEE analyses of either additional dataset, consistent with the GEE analysis of the main dataset.

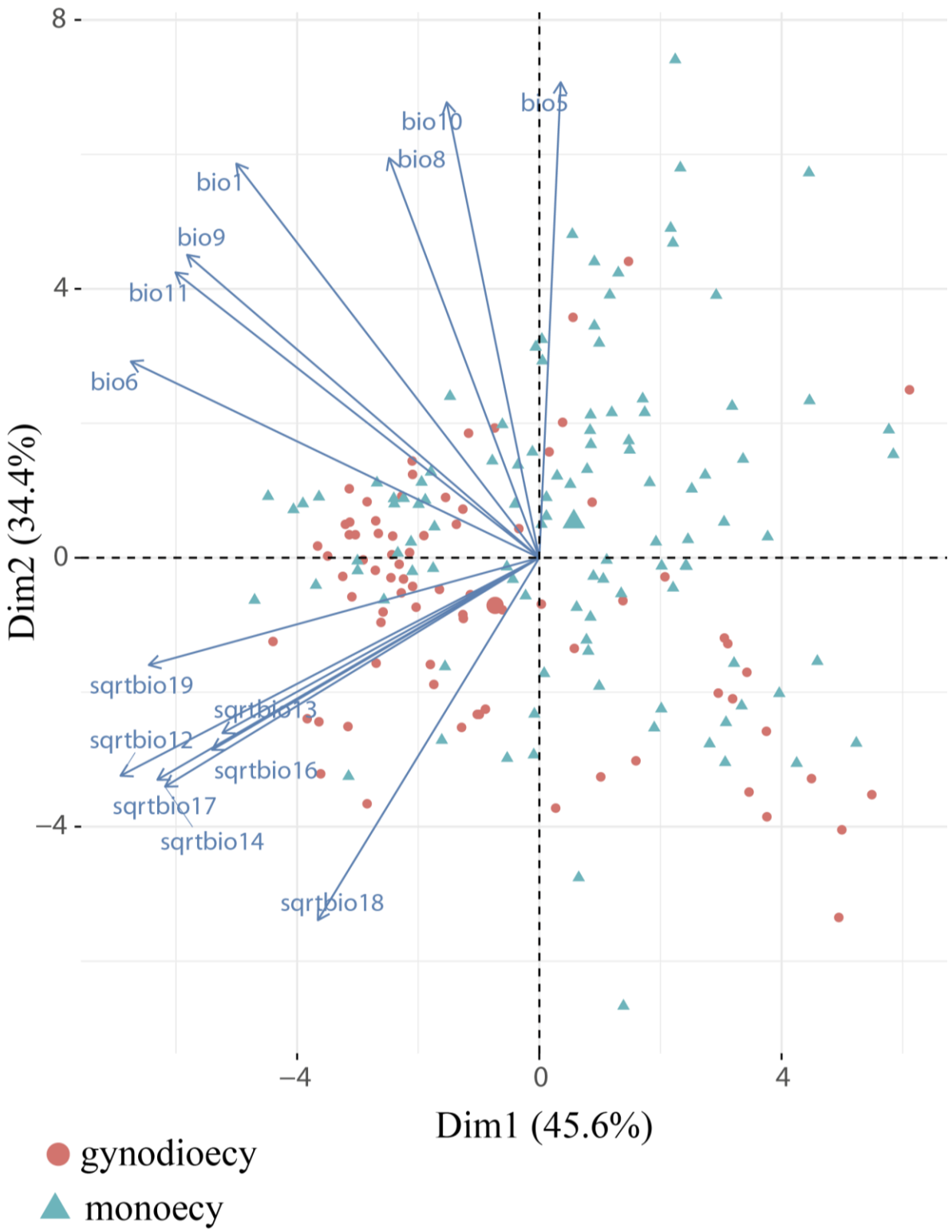


Figure 2. Principal component analysis (PCA) of climatic variables in *Ficus*. Each dot represents the median value of a species. Bioclimatic variables are described in Table 1.

Ancestral state reconstruction for precipitation

The ancestral state reconstruction of precipitation of the warmest quarter (bio18) reveals that nearly all of the drier niches in *Ficus* are occupied by monoecious species (Fig. 3), providing a visual confirmation of the relationship between breeding systems and precipitation outlined by the GLM analyses. The estimated mean value of root of the tree was 612 mm (95% confidence interval: 308- 916 mm). Sampled species in the monoecious section *Galoglychia*, which is distributed from Africa to Arabian Peninsula including Madagascar and surrounding islands (Berg 2004), showed comparatively lower values of bio18.

DISCUSSION

Different results in GLM and GEE

All the climatic variables related with precipitation (bio 12-19) were significantly correlated with dioecy in *Ficus* in regressions with generalized linear models (GLM), but not those with the generalized estimating equations (GEE) approach (Paradis and Claude 2002), which takes phylogenetic relationships into account. This difference could be a type I error for GLM to find significance with higher numbers of degrees of freedom compared with the counterpart which adapted with phylogenetic relationship in GEE (Paradis and Claude 2002).

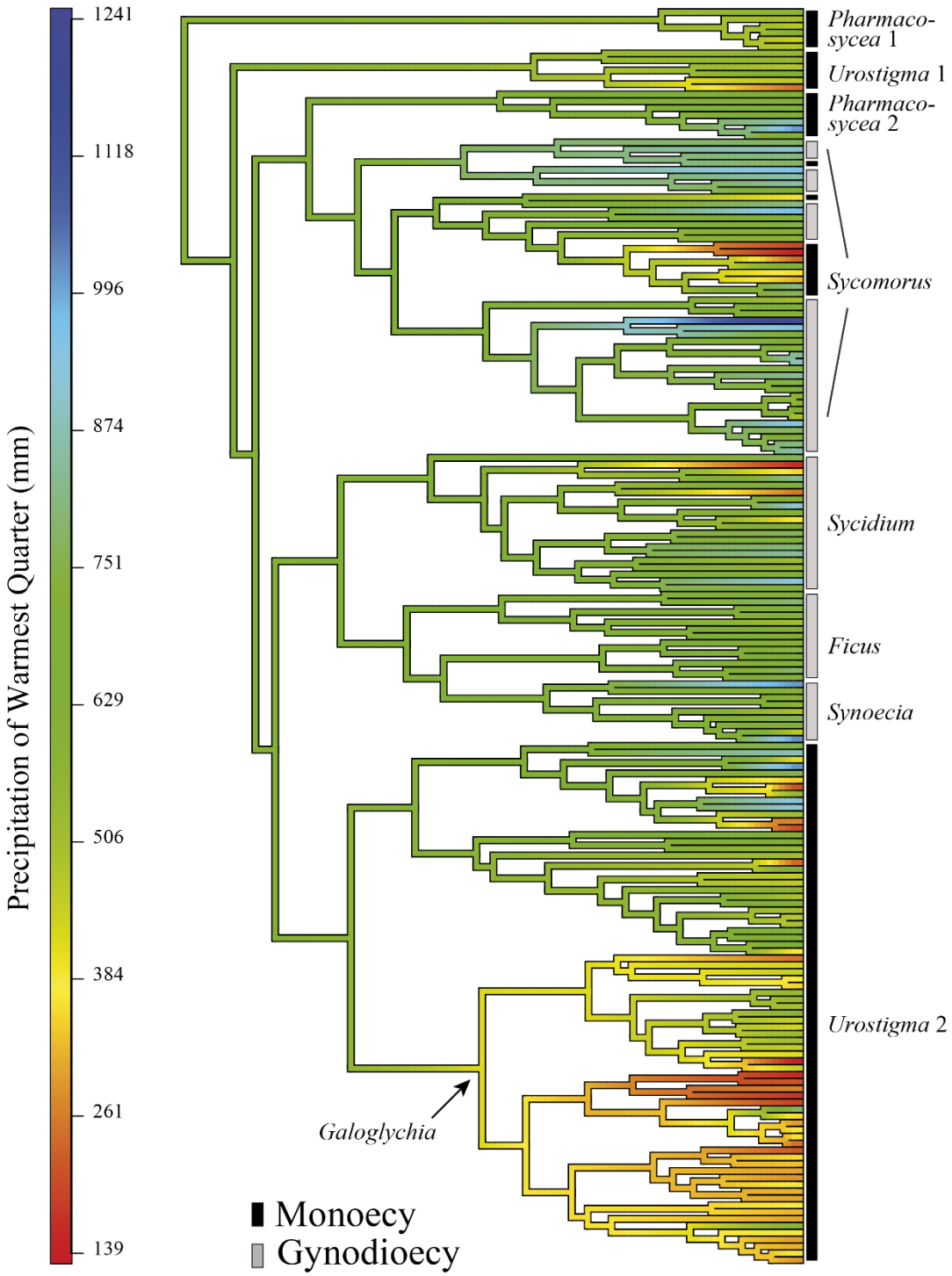


Figure 3. Ancestral state reconstruction of precipitation of the warmest quarter (bio18) in *Ficus*. Breeding systems are highlighted next to the tips. The ancestral breeding system of *Ficus* and the number of transitions remain uncertain (Zhang et al. 2019).

Temperature may not relate with the distribution of dioecious Ficus

Although temperature was suggested to be an important factor for the development of *Ficus* at its northern boundary in previous studies of several species (Peng *et al.* 2010; Chen *et al.* 2018), our results do not support any relationship between temperature and breeding systems in *Ficus* as a whole. Maximum temperature of warmest month (bio5) is the only temperature variable found to have a (negative) significant relationship with dioecy in GLM analyses, a result not supported by the GEE approach. In the GLM analysis with the subsetAsia dataset, bio5 was again the only variable showing (negatively) a significant relationship with dioecy.

Monoecious Ficus *occur in drier environments more often than dioecious* Ficus

Our richness maps of *Ficus* with two breeding systems imply that precipitation has shaped the distribution pattern within the climatic boundaries of the genus. The overlapping areas of the two breeding systems are areas with higher humidity (Fig. 1). On the African, Asian and Australian continents, this distribution pattern coincides with the summer position in both the northern and southern hemispheres of the intertropical convergence zone (ITCZ), which is identified as the maximum in time-mean precipitation (Schneider *et al.* 2014) (Fig. 1). *Ficus* is mainly distributed in peninsulas, islands, archipelagos and coastal areas in Asia-Oceania, where climates are humid. This region is also one of the distribution centers for both breeding systems. In addition, most of monoecious *Ficus* in the Neotropics is distributed in the ITCZ area (Fig. 1B).

In addition to the overlapping areas of both breeding systems, monoecious species also occupy areas with drier climates in continental Africa, Madagascar and Australia. In continental Africa, monoecious species occur in Namibia, northern South Africa, Zimbabwe and Ethiopia, which have

been suggested to be part of an arid "corridor", spreading from southwest to northeast of continental Africa (Jürgens 1997) and documented in other groups of flowering plants (Thiv *et al.* 2011; Freitas *et al.* 2018). In Madagascar, most of dioecious *Ficus* occurs in the humid forest along the east coast (Evans *et al.* 2014), while monoecious *Ficus* is distributed throughout the country (Fig. 1B) including the dry west edge. In Australia, with some exceptions, dioecious *Ficus* is distributed in the Monsoonal Tropics and aseasonal-wet areas (Greenwood 1996; Crisp *et al.* 2004). However, monoecious species are also distributed in the arid Eremean zone (Crisp *et al.* 2004) of central and western Australia. The distribution pattern of *Ficus* in Asia is different. Dioecious species of *Ficus* are distributed farther North than monoecious species (*Ficus sarmentosa* extends to 37°N in our dataset, Fig. 1). The northern boundary of dioecious *Ficus* in continental Asia coincides with the summer (July) position of ITCZ (Schneider *et al.* 2014), while monoecious species seem to be constrained in tropical areas with few records in subtropical areas (Fig. 1B). In addition to its presence in typically year-round arid regions, monoecious *Ficus* also occupies areas with temporary dry periods, where dioecious *Ficus* is absent (Fig. 1). For instance, monoecious *Ficus* occurs in the Cape Floristic Region of South Africa, the most species-rich Mediterranean-type ecosystem given its size (Linder 2014)

We hereby propose several explanations for the colonization of drier environments by monoecious *Ficus*. Firstly, the dispersal ability of dioecious *Ficus* is limited compared to that of monoecious species. Dioecious species of *Ficus* contribute more than monoecious ones to spatial genetic structure (Nazareno *et al.* 2013). Mean parent-offspring distance has been estimated to be 200 m in a population genetic structure investigation of two dioecious *Ficus* (Dev *et al.* 2011), while pollen of a monoecious species (*F. sycomorus*) was reported to be delivered as far as 160 km (Ahmed *et al.* 2009). Secondly, disadvantage of the separation of sex in dioecious *Ficus* is another

possible explanation. With both functional male and female flowers on the same individual, monoecious *Ficus* may self-fertilize (wasps from one syconium may pollinate another syconium of the same plant individual) in a highly seasonal environment (Cook and Power 1996), while dioecious *Ficus* cannot. Third, seasonal climate, for instance dry summers, may impact the production of seeds in dioecious *Ficus*. Summer is an important period for the development of both host plants and larva of pollinators (Peng *et al.* 2010). At least one crop of wasps released from male syconia of dioecious *Ficus* will serve mainly the purpose of pollination of female syconia in a period (year), rather than laying eggs in other male syconia. In dioecious *Ficus*, this crop tends to be in the summer (Zhao *et al.* 2014; but see Chen *et al.* 2015), when higher evapotranspiration implies additional constraints on water use for seed development. Therefore, dry summers may lead to reduction of seeds in the next phase of the hosts' life cycle. These three hypotheses may explain the absence of dioecious *Ficus* in Mediterranean climates, and dry climates more generally.

Conclusion

Ficus is a pan-tropically distributed genus occupying several bioregions (Crisp *et al.* 2004; Hughes *et al.* 2012; Linder 2014). Different climatic variables may played main roles in different bioregions (González-Orozco *et al.* 2013). Thus, analyses combining species from different bioregions may suffer from confounding factors. We sampled 183 species (ca. 24% of all *Ficus* species) in this study. Future more densely sampled phylogenetic trees will be required to start disentangling the main environmental factors in different bioregions. However, phylogenetic relationships in *Ficus* remain themselves incompletely understood (Cruaud et al., 2012; Bruun-Lund et al. 2017), potentially limiting our understanding of breeding system evolution in the genus

(Zhang et al. 2019). A new, highly supported phylogenetic backbone of *Ficus* was proposed in a recent study based on 600 RAD-seq loci, in which subgenus *Sycomorus* was found to be sister to the remaining of *Ficus* (Rasplus *et al.* 2018), contrary to previous work (Cruaud *et al.* 2012; Pederneiras *et al.* 2018) . If confirmed, this significant change of topology of phylogeny may affect the results of comparative analyses such as those presented in this study.

The overlapping distribution areas of monoecious and dioecious *Ficus* correspond to the July and January positions of the intertropical convergence zone (ITCZ) on continents, consistent with the idea that precipitation may have played a role in determining the general distribution pattern of *Ficus* within the climatic boundaries of the genus. A plausible positive relationship between precipitation and dioecy was found in *Ficus*, and conversely between tolerance to drier environments and monoecy in *Ficus*. The longer dispersal ability and potential for self-fertilization in monoecious *Ficus* are two hypotheses we propose to explain this pattern. Contrary to previous ideas, no significant relationship of temperature and breeding system was detected. A more densely sampled phylogenetic tree of *Ficus* and further improvements of phylogenetic comparative methods will be necessary for confirming the results obtained in this study.

ACKNOWLEDGEMENTS

This work was supported by a China Scholarship Council PhD grant [grant No. 201506140077] to Q.Z.

LITERATURE CITED

Ahmed S, Compton SG, Butlin RK, Gilmartin PM. **2009**. Wind-borne insects mediate directional pollen transfer between desert fig trees 160 kilometers apart. *Proceedings of the National Academy of Sciences* **106**: 20342–20347.

Berg CC. **2001**. Moreae, Artocarpeae, and *Dorstenia* (Moraceae): with introductions to the family and Ficus and with additions and corrections to Flora Neotropica Monograph 7. *Flora Neotropica* **83**: 1–346.

Berg CC. **2004**. Flora Malesiana precursor for the treatment of Moraceae 7: *Ficus* subgenus Urostigma. *Blumea* **49**: 463–480.

Berg CC, Corner EJH, Nooteboom HP. 2005. Moraceae (*Ficus*). *Flora Malesiana. Series I, Seed Plants* **17**: 1–70**.**

Bougeard S, Dray S. **2018**. Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software* **86**: 1–17.

Chamberlain S, Barve V, Mcglinn D, *et al.* 2019. rgbif: Interface to the Global Biodiversity Information Facility API. R package version 1.2.0, https://CRAN.R-project.org/package=rgbif.

Chamberlain SA, Boettiger C. **2017**. R Python, and Ruby clients for GBIF species occurrence data. *PeerJ* Preprints 5:e3304v1.

Chen H, Peng Y, Zhang Y, Corlett RT. **2015**. Winter cropping in *Ficus tinctoria*: an alternative strategy. *Scientific reports* **5**: 16496.

Chen H, Zhang Y, Peng Y, Corlett RT. **2018**. Latitudinal effects on phenology near the northern limit of figs in China. *Scientific reports* **8**: 4320.

Chessel D, Dufour AB, Thioulouse J. **2004**. The ade4 package-I-One-table methods. *R news* **4**: 5–10.

Compton SG, Ellwood MDF, Davis AJ, Welch K. **2000**. The flight heights of chalcid wasps (Hymenoptera, Chalcidoidea) in a lowland Bornean rain forest: fig wasps are the high fliers. *Biotropica* **32**: 515–522.

Cook JM, Power SA. **1996**. Effects of within-tree flowering asynchrony on the dynamics of seed and wasp production in an Australian fig species. *Journal of Biogeography* **23**: 487–493.

Cook JM, Rasplus J-Y. **2003**. Mutualists with attitude: coevolving fig wasps and figs. *Trends in Ecology & Evolution* **18**: 241–248.

Crisp M, Cook L, Steane D. **2004**. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present–day communities? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **359**: 1551–1571.

Cruaud A, Ronsted N, Chantarasuwan B, *et al.* 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Systematic Biology* **61**: 1029–1047.

Dev SA, Kjellberg F, Hossaert-McKey M, Borges RM. **2011**. Fine-scale population genetic structure of two dioecious Indian keystone species, *Ficus hispida* and *Ficus exasperata* (Moraceae). *Biotropica* **43**: 309–316.

Dray S, Dufour A-B. **2007**. The ade4 package: implementing the duality diagram for ecologists. *Journal of statistical software* **22**: 1–20.

Dray S, Dufour AB, Chessel D. **2007**. The ade4 package-II: Two-table and K-table methods. *R news* **7**: 47–52.

Evans M, Aubriot X, Hearn D, *et al.* 2014. Insights on the evolution of plant succulence from a remarkable radiation in Madagascar (*Euphorbia*). *Systematic Biology* **63**: 697–711.

Felsenstein J. **1985**. Phylogenies and the Comparative Method. *American Naturalist* **125**: 1–15.

Fick SE, Hijmans RJ. **2017**. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302–4315.

Freitas ES, Bauer AM, Siler CD, Broadley DG, Jackman TR. **2018**. Phylogenetic and morphological investigation of the *Mochlus afer-sundevallii* species complex (Squamata: Scincidae) across the arid corridor of sub-Saharan Africa. *Molecular phylogenetics and evolution* **127**: 280–287.

Gardner RO, Early JW. **1996**. The naturalisation of banyan figs (*Ficus* spp., Moraceae) and their pollinating wasps (Hymenoptera: Agaonidae) in New Zealand. *New Zealand Journal of Botany* **34**: 103–110.

Ghada B, Ahmed BA, Khaled C, *et al.* 2010. Molecular evolution of chloroplast DNA in fig (*Ficus carica* L.): Footprints of sweep selection and recent expansion. *Biochemical Systematics and Ecology* **38**: 563–575.

González-Orozco CE, Laffan SW, Knerr N, Miller JT. **2013**. A biogeographical regionalization of Australian *Acacia* species. *Journal of Biogeography* **40**: 2156–2166.

Grafen A. **1989**. The phylogenetic regression. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **326**: 119–157.

Greenwood DR. **1996**. Eocene monsoon forests in central Australia? *Australian Systematic Botany* **9**: 95–112.

Harrison RD. **2003**. Fig wasp dispersal and the stability of a keystone plant resource in Borneo. *Proceedings of the Royal Society of London B: Biological Sciences* **270**: S76–S79.

Harrison RD, Yamamura N. **2003**. A few more hypotheses for the evolution of dioecy in figs (*Ficus*, Moraceae). *Oikos* **100**: 628–635.

Hespenheide HA. **1975**. Selective predation by two swifts and a swallow in Central America. *Ibis* **117**: 82–

99.

Hughes CE, Pennington RT, Antonelli A. **2012**. Neotropical plant evolution: assembling the big picture. *Botanical Journal of the Linnean Society* **171**: 1–18.

Jürgens N. **1997**. Floristic biodiversity and history of African arid regions. *Biodiversity & Conservation* **6**: 495–514.

Lauder G V. **1981**. Form and function: structural analysis in evolutionary morphology. *Paleobiology* **7**: 430–442.

Linder HP. **2014**. The evolution of African plant diversity. *Frontiers in Ecology and Evolution* **2**: 38.

Martins EP, Hansen TF. **1997**. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *American Naturalist*: 646–667.

Mundry R. **2014**. Statistical issues and assumptions of phylogenetic generalized least squares In: *Modern phylogenetic comparative methods and their application in evolutionary biology*. Berlin Heidelberg: Springer.

Nazareno AG, Alzate-Marin AL, Pereira RAS. **2013**. Dioecy, more than monoecy, affects plant spatial genetic structure: the case study of *Ficus*. *Ecology and Evolution* **3**: 3495–3508.

Pagel M. **1993**. Seeking the evolutionary regression coefficient: an analysis of what comparative methods measure. *Journal of theoretical Biology* **164**: 191–205.

Pagel M. **1994**. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proceedings of the Royal Society of London B: Biological Sciences* **255**: 37–45.

Pagel M. **1997**. Inferring evolutionary processes from phylogenies. *Zoologica Scripta* **26**: 331–348.

Paradis E, Claude J. **2002**. Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology* **218**: 175–185.

Paradis E, Schliep K. **2018**. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**: 526–528.

Pederneiras LC, Gaglioti AL, Romaniuc-Neto S, Mansano V de F. **2018**. The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). *Botanical Journal of the Linnean Society* **187**: 594–613.

Peng Y-Q, Compton SG, Yang D-R. **2010**. The reproductive success of *Ficus altissima* and its pollinator in a strongly seasonal environment: Xishuangbanna, Southwestern China. *Plant ecology* **209**: 227–236.

Rasplus J-Y, Rodriguez LJ, Tollon-Cordet C, Cruaud A. **2018**. Revisiting the phylogeny of *Ficus* (Moraceae): when next generation sequencing corroborates past generation botanists. *bioRxiv*: 340463.

Revell LJ. **2012**. phytools: an R package for phylogenetic comparative biology (and other things). *Methods*

in Ecology and Evolution **3**: 217–223.

Schneider T, Bischoff T, Haug GH. **2014**. Migrations and dynamics of the intertropical convergence zone. *Nature* **513**: 45–53.

Shanahan M, So S, Compton SG, Corlett R. **2001**. Fig-eating by vertebrate frugivores: a global review. *Biological Reviews* **76**: 529–572.

Thiv M, van der Niet T, Rutschmann F, Thulin M, Brune T, Linder HP. **2011**. Old–New World and trans-African disjunctions of *Thamnosma* (Rutaceae): Intercontinental long-distance dispersal and local differentiation in the succulent biome. *American Journal of Botany* **98**: 76–87.

Töpel M, Zizka A, Silvestro D, Scharn R, Calió MF, Antonelli A. **2016**. SpeciesGeoCoder: fast categorization of species occurrences for analyses of biodiversity, biogeography, ecology, and evolution. *Systematic Biology* **66**: 145–151.

Yang LY, Machado CA, Dang XD, *et al.* 2015. The incidence and pattern of copollinator diversification in dioecious and monoecious figs. *Evolution* **69**: 294–304.

Zhang Q, Onstein RE, Little SA, Sauquet H. **2019**. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. *Annals of Botany* **123**: 191–204.

Zhao T-T, Compton SG, Yang Y-J, Wang R, Chen Y. **2014**. Phenological adaptations in *Ficus tikoua* exhibit convergence with unrelated extra-tropical fig trees. *PloS one* **9**: e114344.

Zizka A, Silvestro D, Andermann T, *et al.* 2019. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. *Methods in Ecology and Evolution* **10**: 744–751.

SUPPLEMENTARY INFORMATION

Table S1. Species list and distribution information of the dataset in this study.

Table S2. Estimates for interaction effects of climatic variables and breeding systems in *Ficus* from the regressions by generalized linear models (GLM) and generalized estimating equations (GEE) for three datasets: 1) full dataset (same as Table 1); 2) subsetAsia; and 3) bio18min.

Supplementary data 1. The cleaned occurrence records used in the analysis of this study.

CONCLUSION AND PERSPECTIVES

I investigated breeding system evolution, biogeographic history and their impacts on the diversity of angiosperms in this thesis, using the pantropically distributed family Moraceae as a model clade. My work has led to several important new results: 1) dioecy was probably the ancestral state of breeding system in Moraceae and subsequently evolved into other three breeding systems, suggesting that dioecy is not an evolutionary dead end; 2) the origin of tribe Dorstenieae was estimated to be in the Cretaceous (65.8-79.8 Ma); two independent dispersal events from Africa to South America occurred in Dorstenieae, leading to *Brosimum* s.l. and the Neotropical clade of *Dorstenia*; 3) precipitation appears to have been a factor in shaping the biogeographic distribution pattern of *Ficus* according to breeding systems.

In Chapter I, I reconstructed the ancestral state of breeding systems in Moraceae and *Ficus* with a new dated tree built for the purpose and six morphological models differing in the number of breeding system categories. While this study highlighted that dioecy was the starting point, rather than a dead end of breeding system evolution in the family, much uncertainty remained with respect to the ancestral state of *Ficus*. This uncertainty is partly the consequence of a combination of both monoecious and dioecious lineages close to the base of the *Ficus* phylogeny. However, basal relationships in *Ficus* themselves remain uncertain and it is possible that future phylogenetic work in the genus will lead to relationships that more clearly support either monoecy or gynodioecy as the ancestral state of *Ficus*. A new phylogenomic tree of 40 species of *Ficus* based on RAD-seq data has recently been proposed (Rasplus *et al.* 2018), in which, all the phylogenetic relationships of subgenera or sections were highly supported and subgenus *Sycomorus* was sister to the remaining of *Ficus.* This topology is different from ours and two other previous phylogenies of *Ficus* reconstructed from nuclear markers (Cruaud *et al.* 2012; Pederneiras *et al.* 2018), in which

section *Pharmacosycea* (subgenus *Pharmacosycea*) emerged as the sister group to the remaining of *Ficus*. Using this new phylogenomic tree, Rasplus *et al.* (2018) showed that gynodioecy may be the ancestral state in *Ficus*, contrary to the results emerging from Chapter I. Another recent phylogenomic study, based on full chloroplast genome, suggested yet another set of relationships to those reconstructed from nuclear markers, possibly as a consequence of ancient hybridization in the history of *Ficus* (Bruun-Lund *et al.* 2017). The exon capture approach (Mandel *et al.* 2014), which is similar to the Hyb-Seq method (Weitemier *et al.* 2014) we used in Chapter II, is a widely used target enrichment method. Both the Hyb-Seq and exon capture approaches are likely to lead to a considerably improved and densely sampled phylogeny of *Ficus* and Moraceae as a whole in the near future.

In Chapter II, I investigated the biogeographic history of tribe Dorstenieae with a fossil-calibrated phylogenomic tree based on a large new sequence dataset generated through a target enrichment approach (Hyb-Seq). The ancestral region of Dorstenieae was reconstructed to be the area comprising continental Africa, Madagascar and Asia-Oceania. Although this result may not seem informative enough and possibly represents an artefact from the DEC model and our limited sample of outgroups, this study led to several important results on the subsequent colonization of the Neotropics in two subclades: *Brosimum* s.l. and Neotropical *Dorstenia*. The crown group age of *Brosimum* s.l. was estimated in the Oligocene to Miocene (14.9-31.1 Ma) after a vicariance event. I also conducted separate molecular dating analyses for a subset of the dataset presented in Chapter II, focusing on *Brosimum* alone (see supplementary data). The crown group age of *Brosimum* was inferred to 18.5-29.6 Ma, consistent with the results obtained in Chapter II. Neotropical *Dorstenia* was estimated to originate from the Eocene to the Oligocene (29.8-44.7 Ma). Because Africa has been isolated from South America since ca. 105 Ma (McLoughlin 2001),

these results imply two long-distance dispersal (LDD) events from continental Africa to South America lead to explain the origin of *Brosimum* s.l. and Neotropical *Dorstenia*. The mechanism for the LDD of *Dorstenia* and *Brosimum* remain unclear, especially given that *Dorstenia* has been suggested to be poorly adapted for LDD (Berg and Hijman 1999). Further studies of dispersal related traits of *Dorstenia* and *Brosimum* will improve the understanding of the mechanism of LDD for species unlikely to do so (Onstein *et al.* 2018), a general and difficult question that has been raised in numerous other plant clades (Doyle *et al.* 2004; Barker *et al.* 2007).

The latest biogeographic study of Moraceae as a whole was provided by Zerega et al. (2005), who discussed the potential biogeographic history of Moraceae based on their phylogeny and divergence time estimates. However, parametric biogeographic approaches, such as those used in Chapter II for Dorstenieae and Pederneiras et al. (2018) for *Ficus*, have not yet been applied to Moraceae as a whole. Thus, it remains unclear whether Gondwana or Laurasia was the ancestral area of Moraceae. Reconstructing the biogeographic history of a pantropically distributed family such as Moraceae will improve the understanding of the origin of plant diversity in tropical regions. However, biogeographic model-based approaches are still relatively young and present some limitations that are currently being debated in the community. The widely used dispersal-extinction-cladogenesis (DEC) model does not explore the scenario in which speciation occurs with dispersal events. This process might be common in scenarios of long-distance dispersal to a new environment, especially in island systems (Ree and Sanmartín 2009; Matzke 2014). The introduction of the founder-event speciation process (jump process) in the DEC model, leading to the now widely used DEC+J model, offered a solution to this problem (Matzke 2013). However, a recent study demonstrated that models which includes both anagenetic and cladogenetic processes may be biased towards cladogenetic processes, and as a result over-reliance on the

founder-event speciation process to explain biogeographic distributions (Ree and Sanmartín 2018).

Most of the genera in Moraceae consist of fewer than 10 species (Rohwer and Berg 1993), and some species may be hard to collect in the field. As demonstrated in Chapter II and several recent studies (Hart *et al.* 2016; Couvreur *et al.* 2019), herbarium specimens are now a promising resource for phylogenomic work using approaches such as target enrichment, where DNA fragmentation (typical of herbarium material) is much less of an issue than traditional Sanger sequencing of PCR-amplified markers. The rapid adoption of this and other genomic approaches by the plant systematic community is likely to lead to considerable improvement of phylogeny at all levels (Johnson *et al.* 2018; Couvreur *et al.* 2019), including in Moraceae (Rasplus *et al.* 2018; Zerega and Gardner 2019). These developments may in turn result in further necessary adjustments of the current phylogenetic classification of Moraceae (Clement and Weiblen 2009; Zerega and Gardner 2019) and the circumscription of several genera. For instance, in Chapter II, I found the genera *Scyphosyce* and *Utsetela* to be nested in *Dorstenia* and *Trymatococcus* and *Helianthostylis* in *Brosimum* s.l. The latter result was also confirmed in the more densely sampled side study focusing on *Brosimum* (see supplementary data).

Species have preferentially remained in the same or similar niche/biome while colonizing new environments (Crisp *et al.* 2009), which may have played an important role in shaping the current distribution pattern of plant diversity. In Chapter III, I combined knowledge of breeding system evolution and detailed spatial data to characterize and understand differences in the geographic distribution of monoecious and gynodioecious species of *Ficus*. At the beginning of this PhD and chapter, I hypothesized that temperature was a key factor influencing the distribution of *Ficus*, based on the tendency of gynodioecious *Ficus* to occupy higher latitudes and colder environments in Asia than monoecious *Ficus* (Berg *et al.* 2005; Bain *et al.* 2014). However, none of my analyses

supported a correlation of temperature with breeding system in *Ficus*, even with a subset of the data restricted to the Asia-Oceania region. Precipitation was found to be a plausible factor in shaping the current distribution pattern of different breeding systems in *Ficus*: both breeding systems of *Ficus* are largely overlapping in tropical climates, but monoecious *Ficus* extends to drier environments than gynodioecious species. To our knowledge, this is the first time that such a correlation has been suggested and discussed.

In Chapter I, I estimated the age of crown group *Ficus* in the Eocene (40.6-55.9 Ma), during when the Eurasia region, which has been proposed as the ancestral area of *Ficus* (Pederneiras *et al.* 2018), was still separated by the Siberian and Turan Seas as Europe and Asia (Akhmetiev and Beniamovski 2009). In addition, the average global temperature was higher and tropical regions were much more widespread than they are today (Zachos *et al.* 2001; Morley 2011). The distribution areas of *Ficus* may have extended and retreated several times to track tropical climates during the history of *Ficus* (Costa *et al.* 2017). During these processes, some monoecious *Ficus* dispersed into the current drier areas, which may have been previously wetter. However, the co-occurring gynodioecious species would have been swept away during past aridification events due to their lower tolerance of dry climate. In addition, monoecious *Ficus* probably also successfully dispersed into drier areas where the genus was previously absent. To our knowledge, ours is the first study to test climatic niche difference of congeneric species with different breeding systems. Whether niche difference is a general pattern in other diversified genera with at least two breeding systems such as *Acer* (Sapindaceae, Renner *et al.* 2007) remains to be tested in more clades. Furthermore, the loci responsible for sex determination in Moraceae are still unclear. Did any sex chromosomes develop in the Moraceae genome? Are they the same in other model clades such as *Carica* (Caricaceae) (Wang *et al.* 2012), Cucurbitaceae (Boualem *et al.* 2015) and *Silene*

(Caryophyllaceae) (Bergero *et al.* 2007)? We are still far from understanding the genetic and developmental processes responsible for breeding system diversity in Moraceae, an exciting avenue for future research.

LITERATURE CITED

Akhmetiev MA, Beniamovski VN. **2009**. Paleogene floral assemblages around epicontinental seas and straits in Northern Central Eurasia: proxies for climatic and paleogeographic evolution. *Geologica Acta* **7**: 297–309.

Bain A, Chou LS, Tzeng HY, *et al.* 2014. Plasticity and diversity of the phenology of dioecious *Ficus* species in Taiwan. *Acta Oecologica-International Journal of Ecology* **57**: 124–134.

Barker NP, Weston PH, Rutschmann F, Sauquet H. **2007**. Molecular dating of the 'Gondwanan'plant family Proteaceae is only partially congruent with the timing of the break-up of Gondwana. *Journal of Biogeography* **34**: 2012–2027.

Berg CC, Corner EJH, Nooteboom HP. 2005. Moraceae (*Ficus*). *Flora Malesiana. Series I, Seed Plants* **17**: 1–70**.**

Berg CC, Hijman MEE. **1999**. The genus *Dorstenia* (Moraceae). *Ilicifolia* **2**: 1–211.

Bergero R, Forrest A, Kamau E, Charlesworth D. **2007**. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* **175**: 1945–1954.

Boualem A, Troadec C, Camps C, *et al.* 2015. A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* **350**: 688 LP – 691.

Bruun-Lund S, Clement WL, Kjellberg F, Rønsted N. **2017**. First plastid phylogenomic study reveals potential cyto-nuclear discordance in the evolutionary history of *Ficus* L. (Moraceae). *Molecular Phylogenetics and Evolution* **109**: 93–104.

Clement WL, Weiblen GD. **2009**. Morphological evolution in the mulberry family (Moraceae). *Systematic Botany* **34**: 530–552.

Costa PC, Lorenz-Lemke AP, Furini PR, Honorio Coronado EN, Kjellberg F, Pereira RAS. **2017**. The phylogeography of two disjunct Neotropical *Ficus* (Moraceae) species reveals contrasted histories between the Amazon and the Atlantic Forests. *Botanical Journal of the Linnean Society* **185**: 272–289.

Couvreur TLP, Helmstetter AJ, Koenen EJM, *et al.* 2019. Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. *Frontiers in Plant Science* **9**: 1941.

Crisp MD, Arroyo MTK, Cook LG, *et al.* 2009. Phylogenetic biome conservatism on a global scale. *Nature* **458**: 754.

Cruaud A, Ronsted N, Chantarasuwan B, *et al.* 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. *Systematic Biology* **61**: 1029–1047.

Doyle JA, Sauquet H, Scharaschkin T, Le Thomas A. **2004**. Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences* **165**: S55–S67.

Hart ML, Forrest LL, Nicholls JA, Kidner CA. **2016**. Retrieval of hundreds of nuclear loci from herbarium specimens. *Taxon* **65**: 1081–1092.

Johnson MG, Pokorny L, Dodsworth S, *et al.* 2018. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology*.

Mandel JR, Dikow RB, Funk VA, *et al.* 2014. A target enrichment method for gathering phylogenetic information from hundreds of loci: An example from the Compositae. *Applications in Plant Sciences* **2**: 1300085.

Matzke NJ. **2013**. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Frontiers of Biogeography* **5**: 242–248.

Matzke NJ. **2014**. Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. *Systematic Biology* **63**: 951–970.

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

Morley RJ. **2011**. Cretaceous and Tertiary climate change and the past distribution of megathermal rainforests In: *Tropical rainforest responses to climatic change*. Berlin Heidelberg: Springer.

Onstein RE, Baker WJ, Couvreur TLP, *et al.* 2018. To adapt or go extinct? The fate of megafaunal palm fruits under past global change. *Proc. R. Soc. B* **285**: 20180882.

Pederneiras LC, Gaglioti AL, Romaniuc-Neto S, Mansano V de F. **2018**. The role of biogeographical barriers and bridges in determining divergent lineages in *Ficus* (Moraceae). *Botanical Journal of the Linnean Society* **187**: 594–613.

Rasplus J-Y, Rodriguez LJ, Tollon-Cordet C, Cruaud A. **2018**. Revisiting the phylogeny of *Ficus* (Moraceae): when next generation sequencing corroborates past generation botanists. *bioRxiv*: 340463.

Ree RH, Sanmartín I. **2009**. Prospects and challenges for parametric models in historical biogeographical inference. *Journal of Biogeography* **36**: 1211–1220.

Ree RH, Sanmartín I. **2018**. Conceptual and statistical problems with the DEC+ J model of founder-event speciation and its comparison with DEC via model selection. *Journal of Biogeography* **45**: 741–749.

Renner SS, Beenken L, Grimm GW, Kocyan A, Ricklefs RE. **2007**. The evolution of dioecy, heterodichogamy, and labile sex expression in Acer. *Evolution* **61**: 2701–2719.

Rohwer JG, Berg CC. **1993**. Moraceae In: *Flowering Plants Dicotyledons*. Berlin Heidelberg: Springer.

Wang J, Na J-K, Yu Q, *et al.* 2012. Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proceedings of the National Academy of Sciences* **109**: 13710 – 13715.

Weitemier K, Straub SCK, Cronn RC, *et al.* 2014. Hyb-Seq: combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences* **2**: 1400042.

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.

Zerega NJC, Clement WL, Datwyler SL, Weiblen GD. **2005**. Biogeography and divergence times in the mulberry family (Moraceae). *Molecular Phylogenetics and Evolution* **37**: 402–416.

Zerega NC, Gardner EM. **2019**. Delimitation of the new tribe Parartocarpeae (Moraceae) is supported by a 333-gene phylogeny and resolves tribal level Moraceae taxonomy. *Phytotaxa* **338**: 253–265.

SUPPLEMENTARY DATA

All the supplementary data attached to this thesis are available from this online folder: https://www.dropbox.com/sh/r09q6qnpxlezpji/AACth8ARb0xr670UrMyEEtKJa?dl=0. As some of the supplementary information are too long to print in a thesis, I have selected some proper supplementary figures and tables to print here.

Summary of my contribution to the *Brosimum* phylogeny

My work on reconstructing a phylogenomic tree and the biogeographic history of tribe Dorstenieae (Chapter II) was made possible through a collaboration with Prof. Nyree Zerega and Dr. Elliot Gardner from the Chicago Botanic Garden and Northwestern University. This collaboration involved their sharing of unpublished data previously generated for species of *Brosimum* (hereafter referred to as the *Brosimum* dataset), and my own contribution to their study of the genus. The aim of the *Brosimum* paper, which will be published separately from my Chapter II on Dorstenieae, is a taxonomic revision and study of character evolution in the genus. While this study has not yet reached the stage of a draft manuscript, here I provide some details on my own contribution, which focused on estimating divergence times in *Brosimum*.

Brosimum consists of approximately fifteen species, which are restricted to the Neotropics (Rohwer and Berg 1993; Berg 2001). This dataset includes all of the 15 species of *Brosimum* and 22 outgroup species (representing 18 other genera in Moraceae and one species of Cannabaceae). The methods used to produce and analyze the *Brosimum* dataset were exactly the same in the Dorstenieae dataset (Chapter II). Specifically, to infer divergence times, I used both penalized likelihood in r8s v1.7 (Sanderson 2003) and the Bayesian approach implemented in MCMCTree

in package PAML v4.9 (Yang 2007). However, I modified the calibration strategy as follows. I calibrated the *Brosimum* phylogenetic tree with three fossil age constraints and two secondary calibrations. The three fossil age constraints were the same ones used in Chapter II. The two secondary calibrations were required to constrain the root and differ from Chapter II because of differences in outgroup sampling. The fossil wood of *Artocarpoxylon deccanensis* Mehrotra, Prakash, and Bande (at least 64.0 Ma) (Mehrotra *et al.* 1984) was used to calibrate the split of *Artocarpus heterophyllus* and *Streblus glaber*. The fossil endocarps of *Broussonetia rugosa* Chandler (Chandler 1961) were used to constrain the split of *Allaeanthus luzonicus*, *Malaisia scandens* and *Broussonetia papyrifera* to at least 33.9 Ma. The fossil achenes of *Ficus* (*F. lucidus* Chandler) (Chandler 1962) were used to constrain the split of *Ficus macrophylla* and *Antiaropsis descipiens* to a minimum age of 56.0 Ma. Except for the outgroup species *Trema orientale* (Cannabaceae), all the species in the dataset belong to Moraceae. I used the estimated crown-group age of Moraceae (73.2-84.7 Ma) and the most recent common ancestor of Moraceae and Cannabaceae (81.7-93.3 Ma) from my recent family-wide molecular dating study (Chapter I; Zhang *et al.* 2019) to provide maximum and minimum boundaries for secondary age constraints on the crown node of Moraceae and the root. The whole dataset were kept as one partition in both approaches. The best smoothing was found to be 2.6 in r8s. Two independent runs for each prior setting (default in MCMCTree and referring to the estimates from program baseml in PAML v4.9, see Chapter II) were launched with chain length of 15 million generations, sampled every 1500 generations. The first 10% were removed as burnin.

Estimates from runs with the prior alone (without data) were different from those with data, suggesting data (rather than the prior alone) are informative in producing the inferred age estimates. The stem and crown group age of *Brosimum* s.l. were estimated as boundary of the Paleogene to

Neogene (22.09-35.35 Ma) and the Oligocene to the Miocene (18.49-29.62 Ma), respectively. Estimates from two different prior settings in MCMCTree showed similar results (Fig 1, A, B). Penalized likelihood analyses (implemented in r8s) produced similar age estimates (Fig 1 C). Estimates obtained with the *Brosimum* dataset were similar to those obtained with the Dorstenieae dataset presented in Chapter II (where the stem and crown group age were 19.58-41.49 Ma and 14.87-31.15 Ma, respectively). This difference may be the result of sampling fewer outgroups and three more ingroup species of *Brosimum* in this study, compared with Chapter II, where all other genera of Dorstenieae were included.

LITERATURE CITED

Berg CC. **2001**. Moreae, Artocarpeae, and Dorstenia (Moraceae): with introductions to the family and Ficus and with additions and corrections to Flora Neotropica Monograph 7. *Flora Neotropica* **83**: 1–346.

Chandler MEJ. **1961**. *Flora of the Lower Headon beds of Hampshire and the Isle of Wight*. British Museum (Natural History).

Chandler MEJ. **1962**. *The Lower Tertiary Floras of Southern England: Flora of the pipe-clay series of Dorset (lower Bagshot)*. order of the Trustees of the British Museum.

Mehrotra RC, Prakash U, Bande MB. **1984**. Fossil woods of *Lophopetalum* and *Artocarpus* from the Deccan Intertrappean Beds of Mandla district, Madhya Pradesh, India. *Palaeobotanist* **32**: 310–320.

Rohwer JG, Berg CC. **1993**. Moraceae In: *Flowering Plants· Dicotyledons*. Springer, 438–453.

Sanderson MJ. **2003**. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**: 301–302.

Yang Z. **2007**. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution* **24**: 1586–1591.

Zhang Q, Onstein RE, Little SA, Sauquet H. **2019**. Estimating divergence times and ancestral breeding systems in *Ficus* and Moraceae. *Annals of Botany* **123**: 191–204.

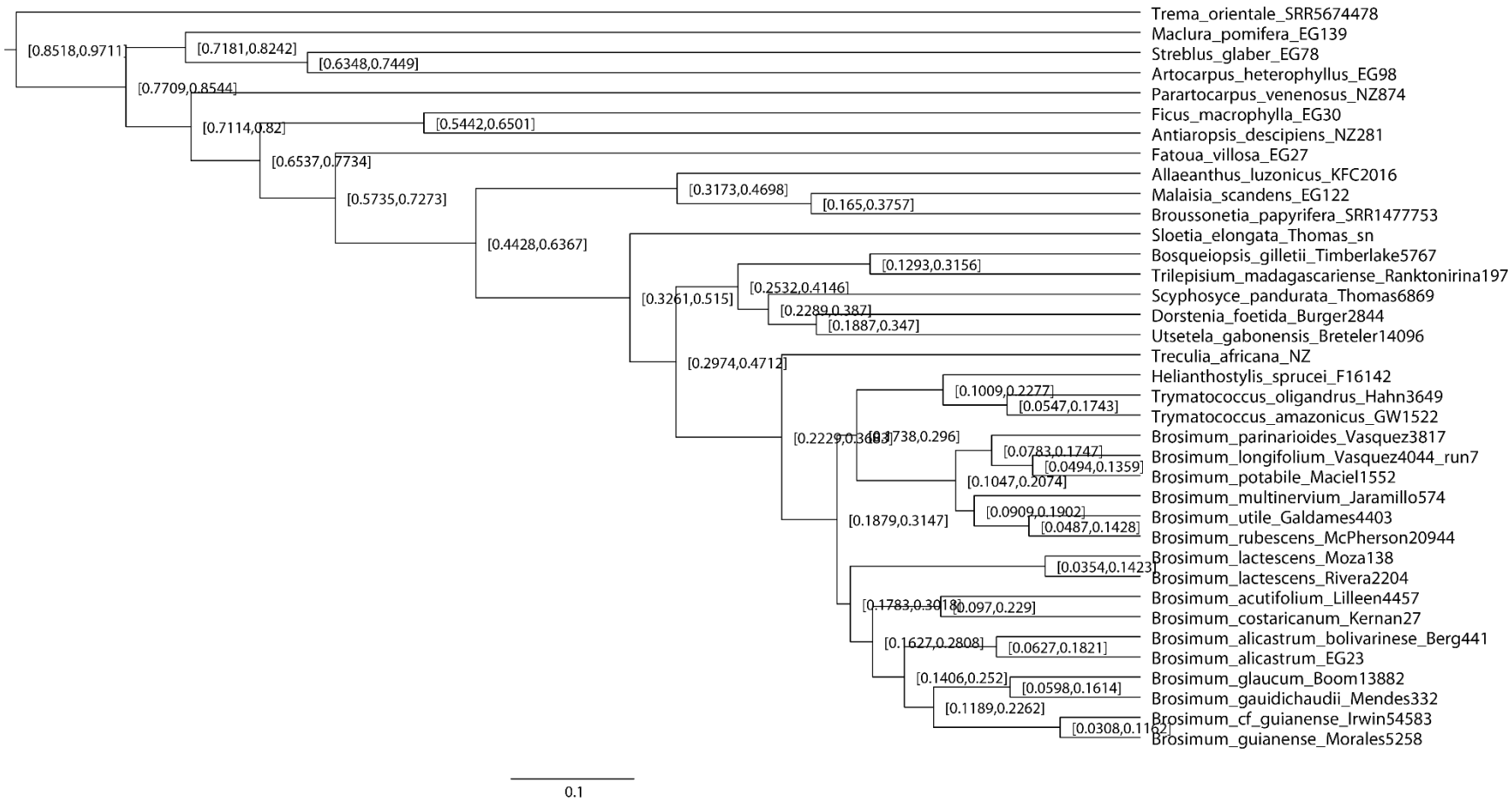


Figure 1. A. Divergence time estimates for *Brosimum* by MCMCTree with default prior setting (time unit in 100 Ma).

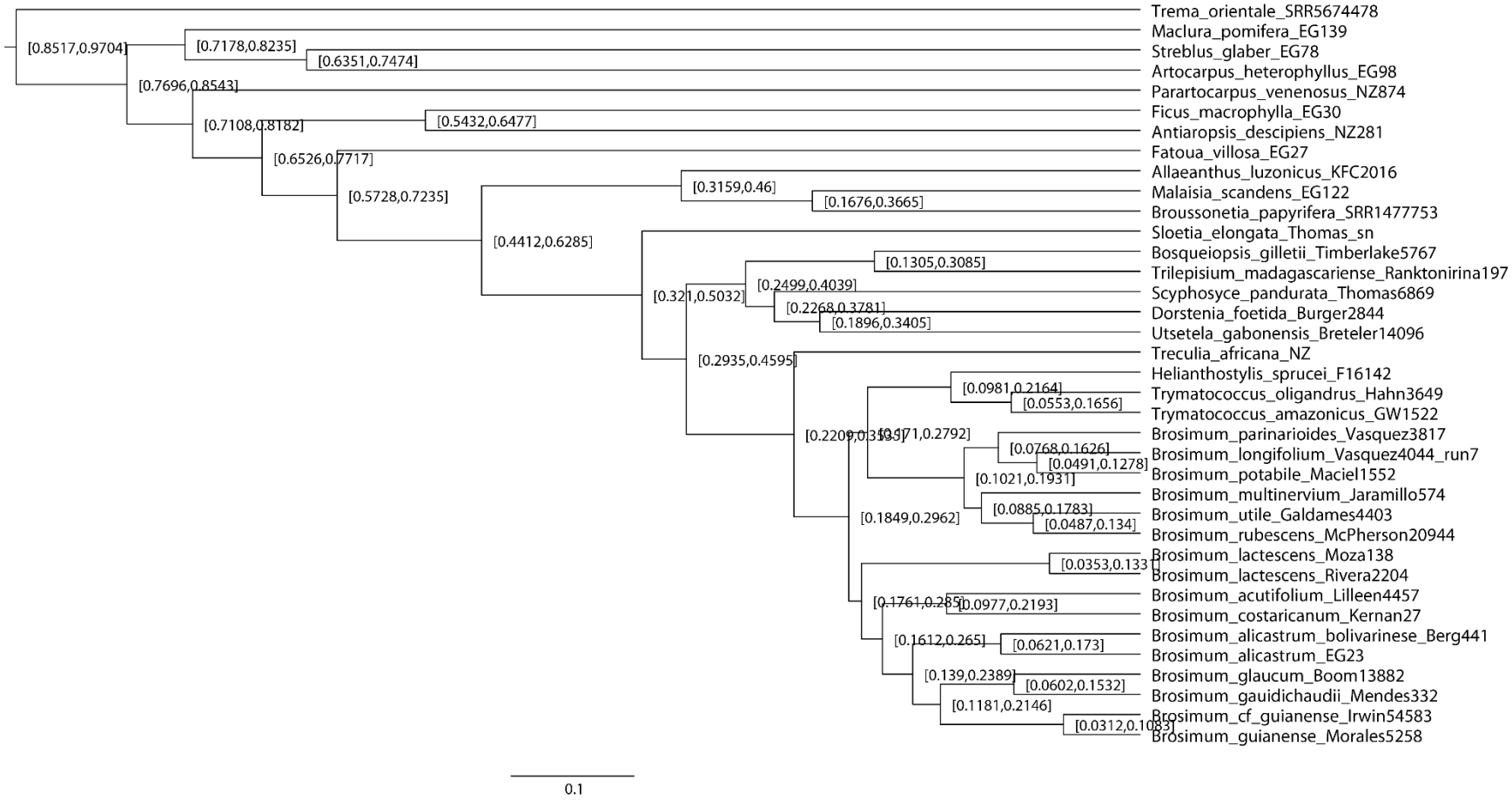


Figure 1. B. Divergence time estimates for *Brosimum* by MCMCTree with prior setting referring to the results from baseml (time unit in 100 Ma).

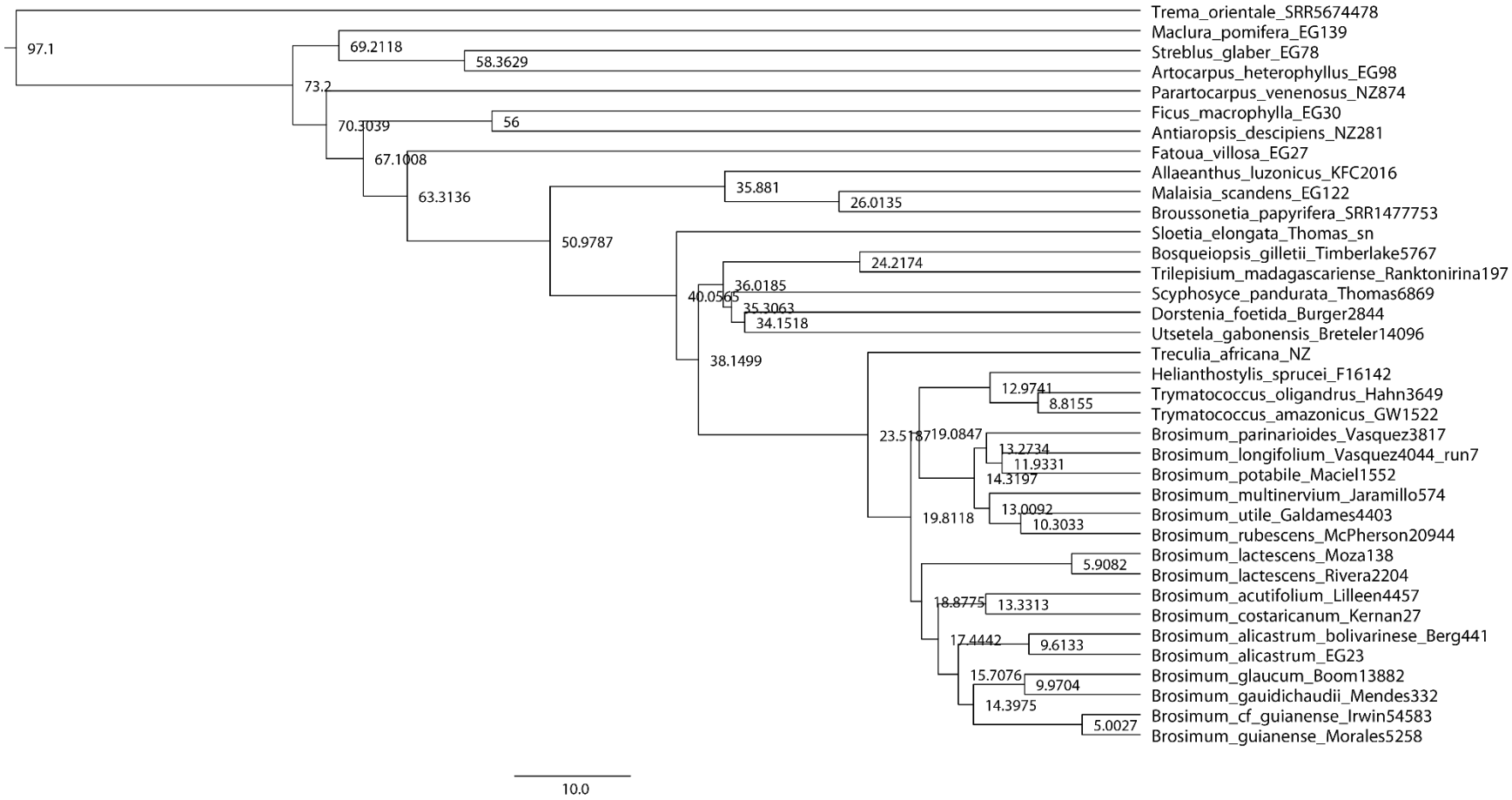


Figure 1. C. Divergence time estimation for *Brosimum* by r8s (time unit in 1 Ma).

Supplementary data for Chapter I. estimating divergence times and ancestral breeding systems in *Ficus* and moraceae

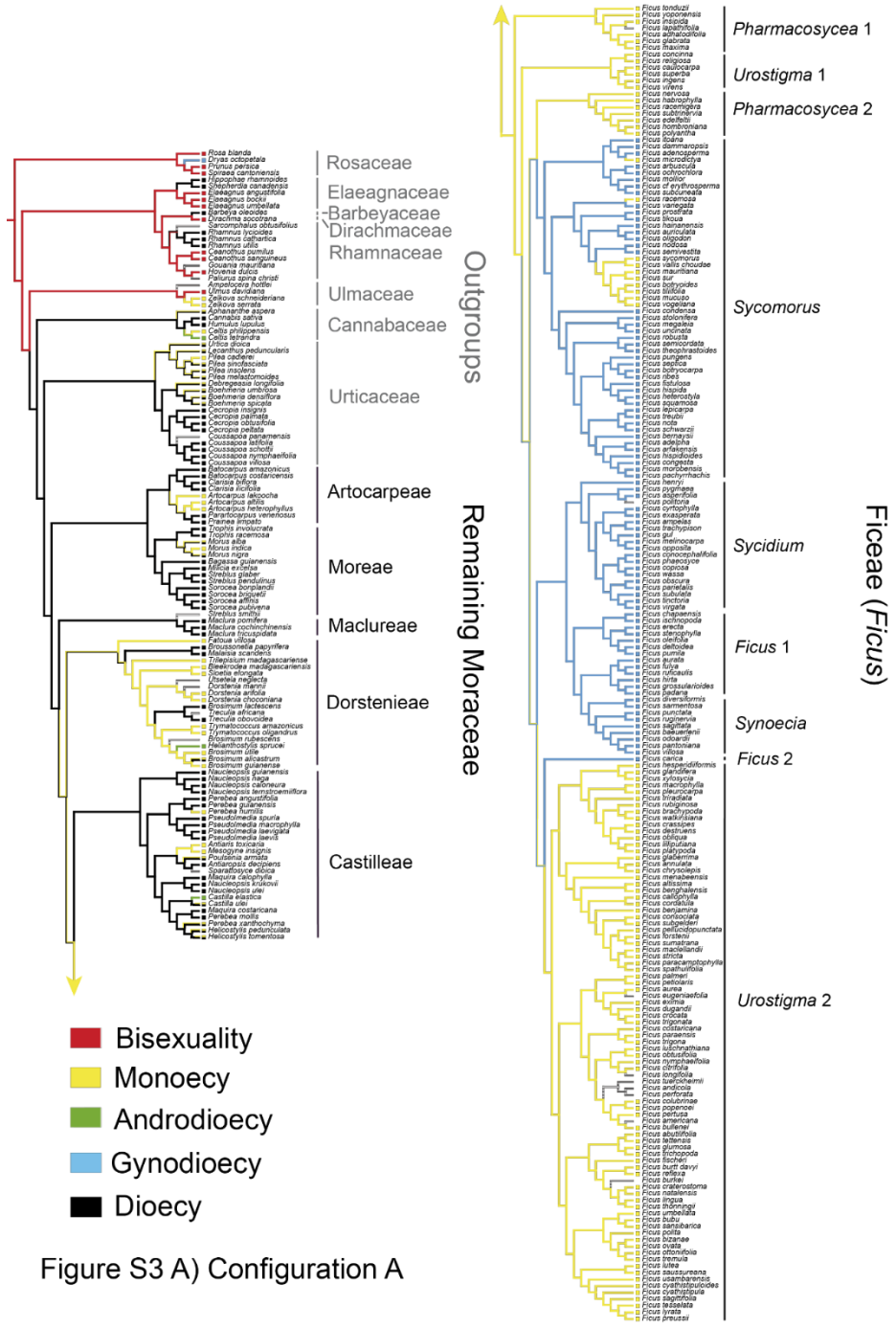


Figure S3 A) Configuration A

Figure S3. Ancestral state reconstruction with 320-species dataset by parsimony approach with tip names with all six configurations (see text).

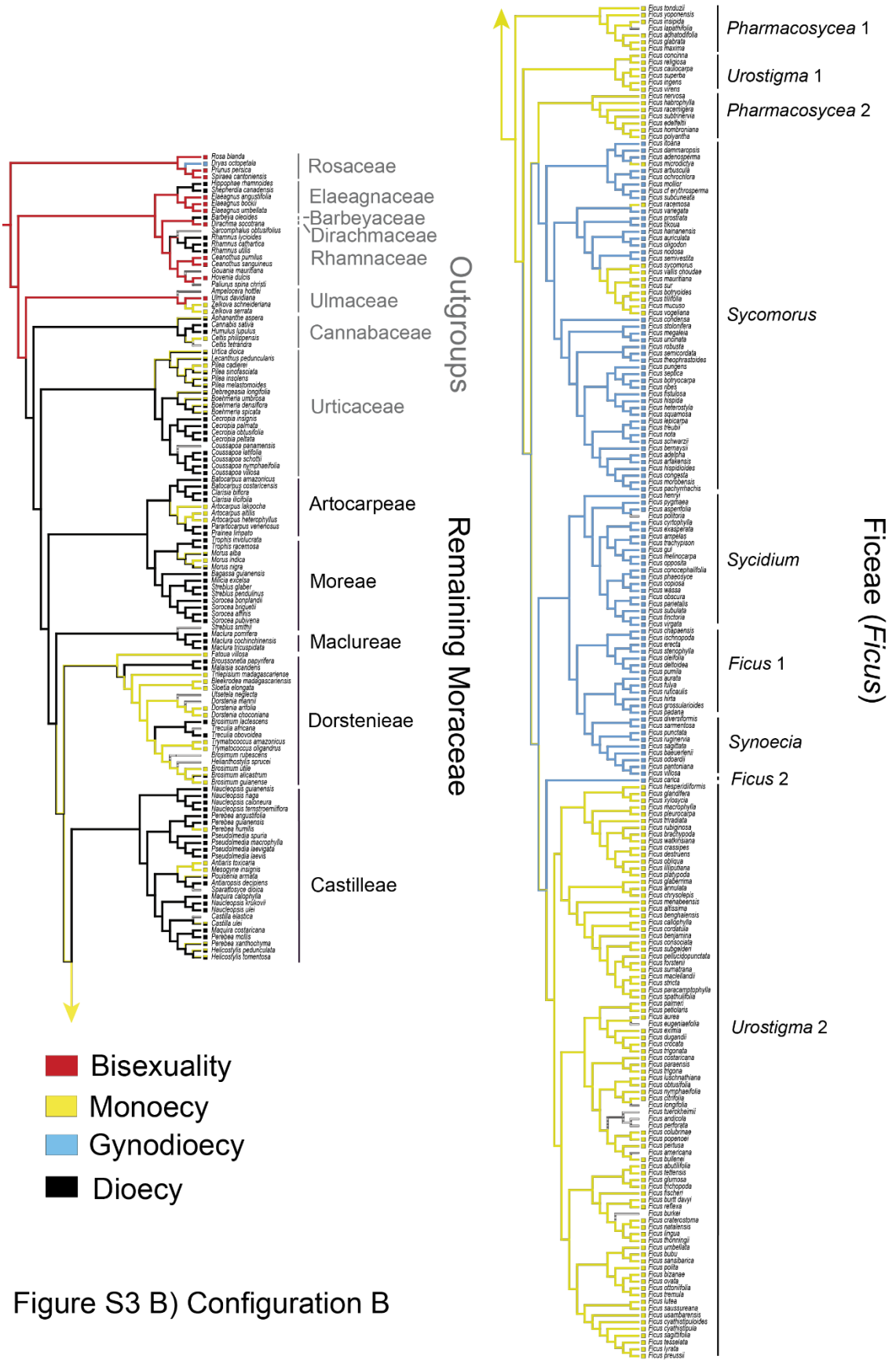


Figure S3 B) Configuration B

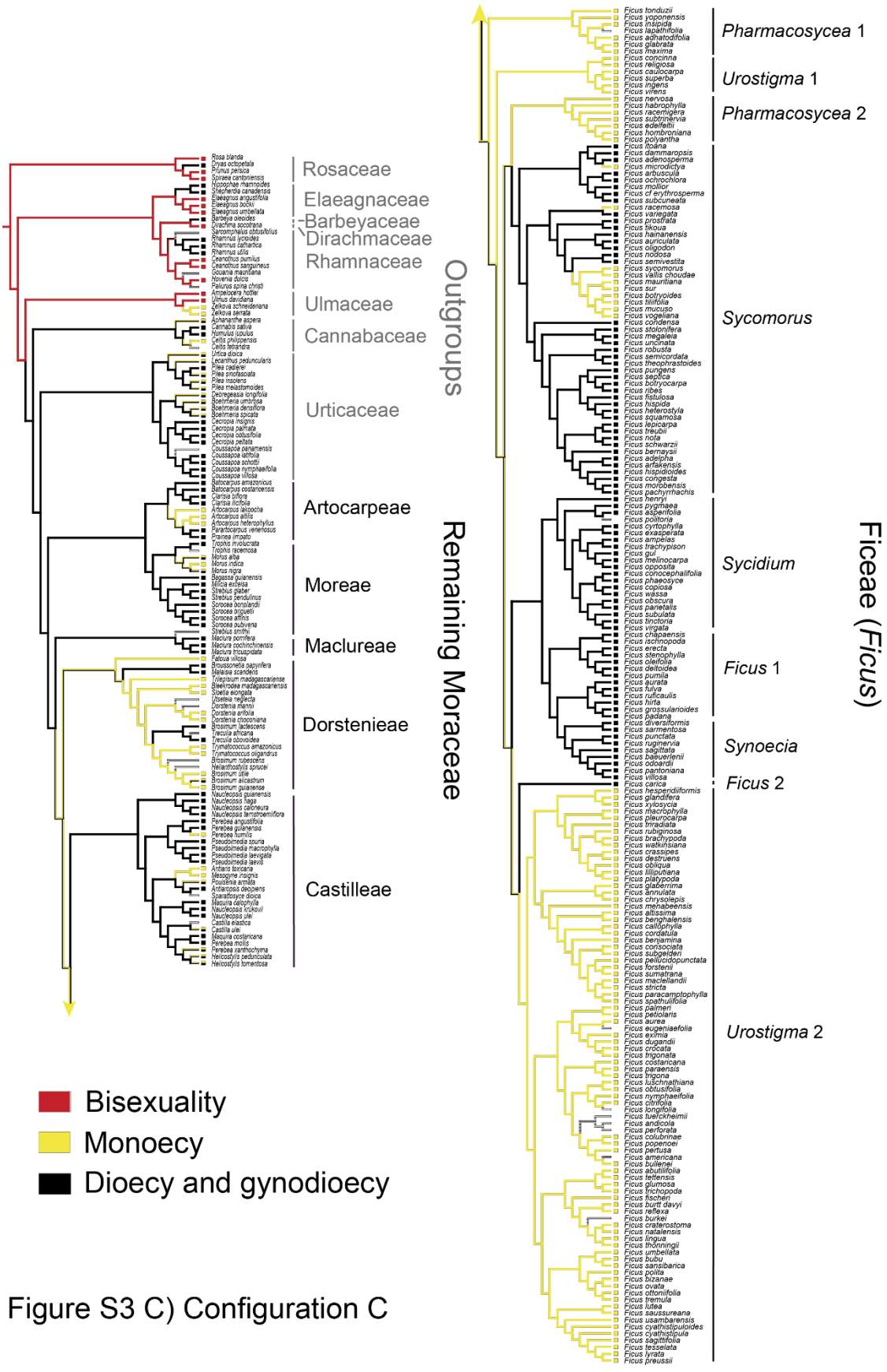


Figure S3 C) Configuration C

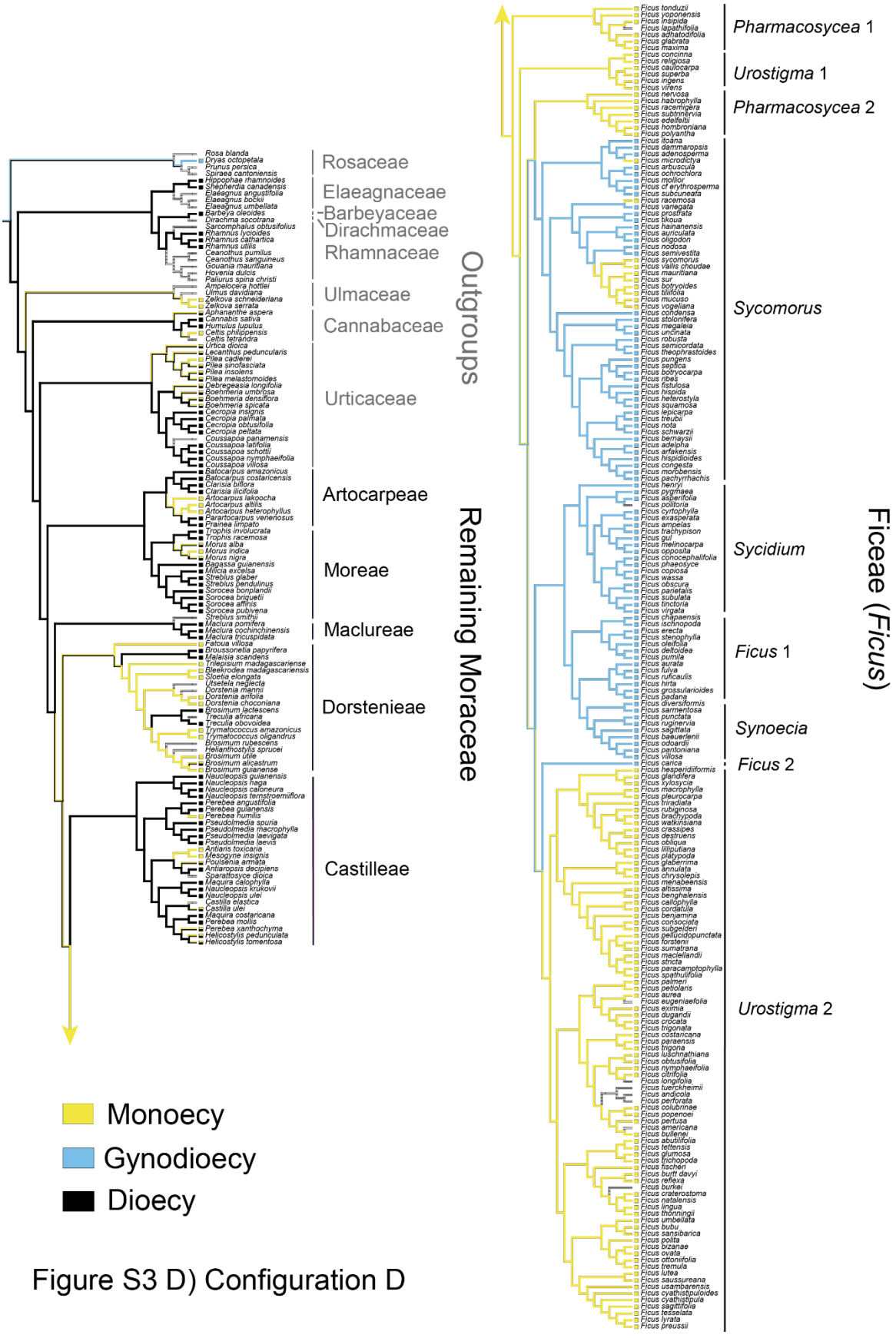


Figure S3 D) Configuration D

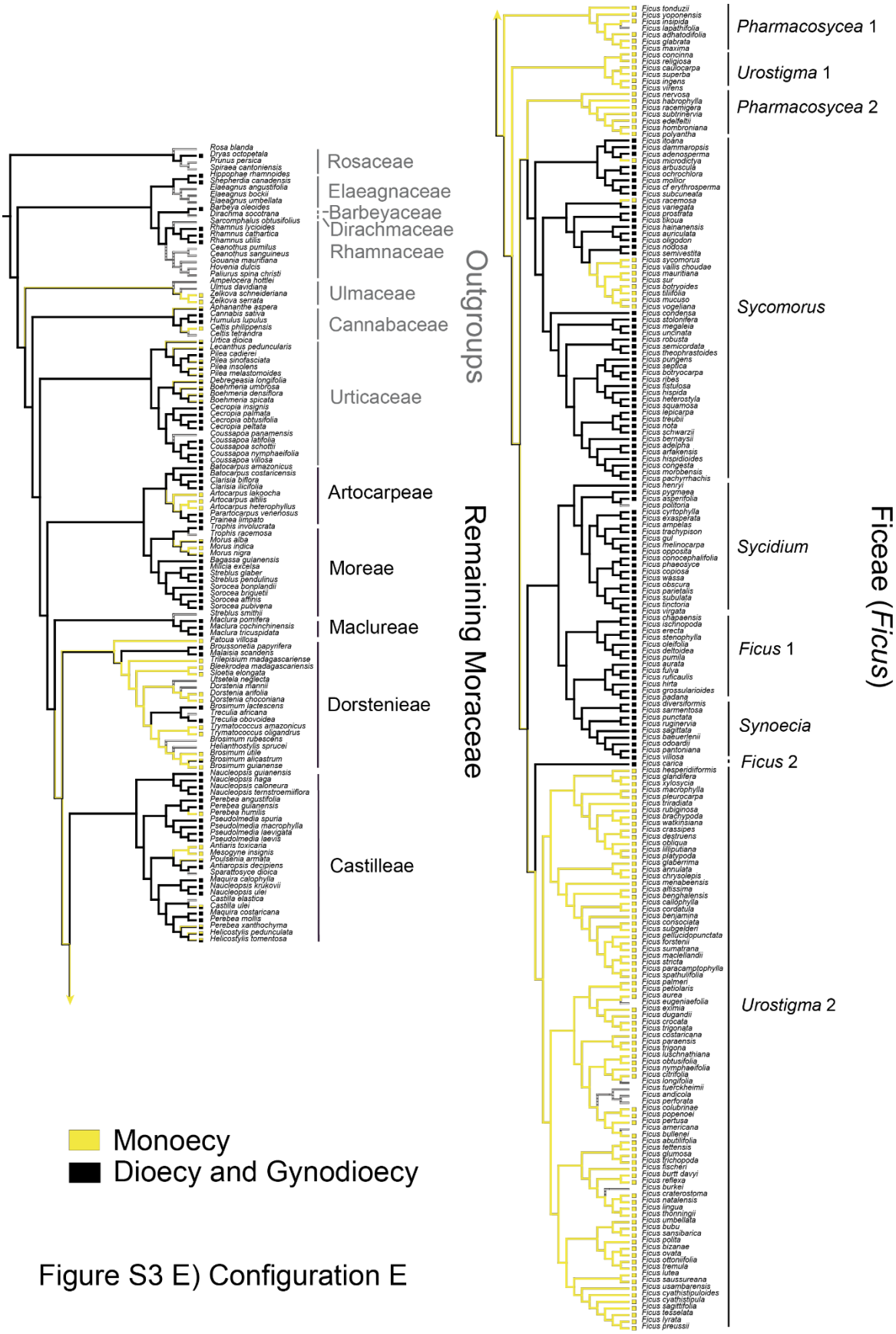


Figure S3 E) Configuration E

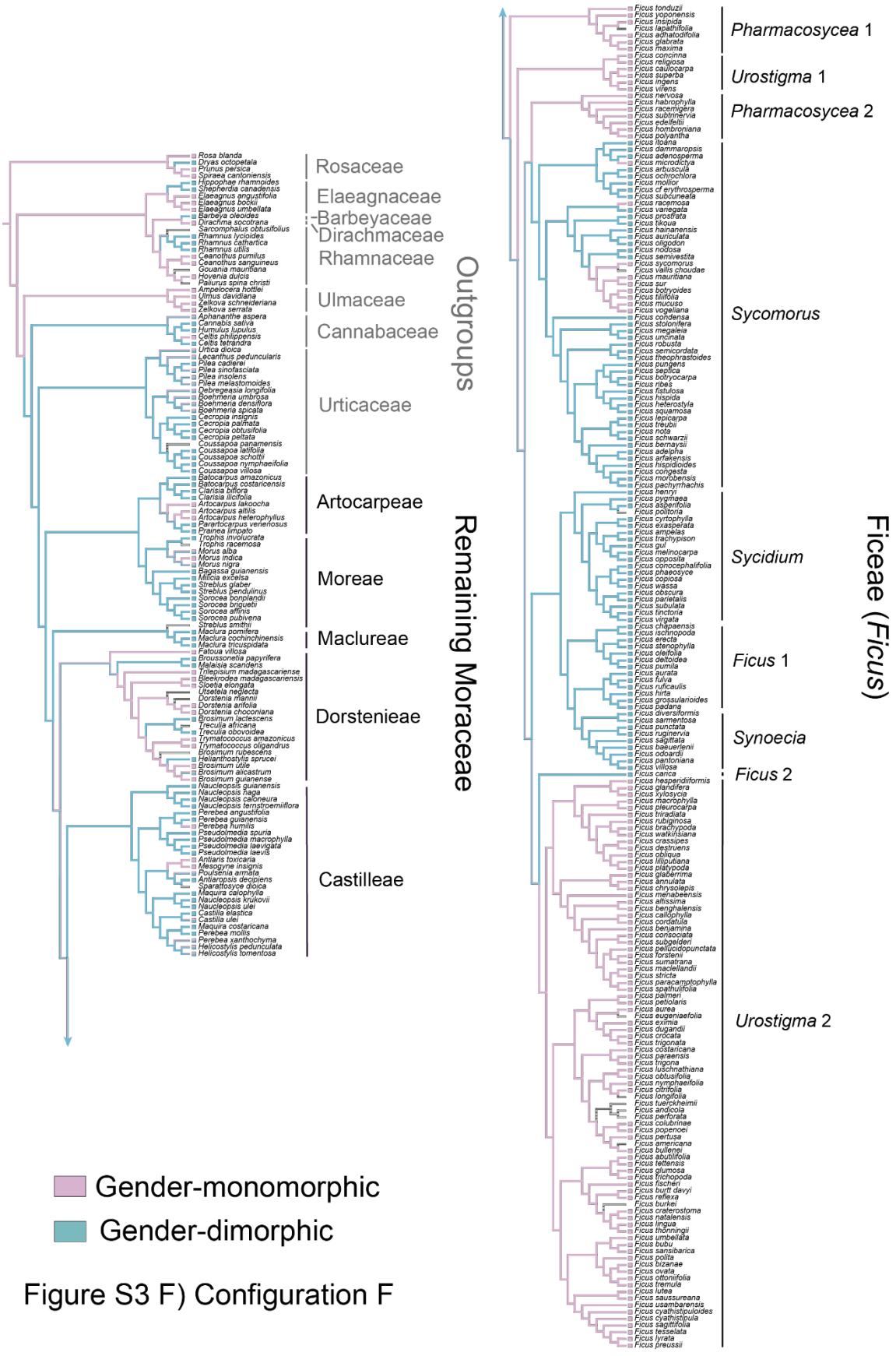


Figure S3 F) Configuration F

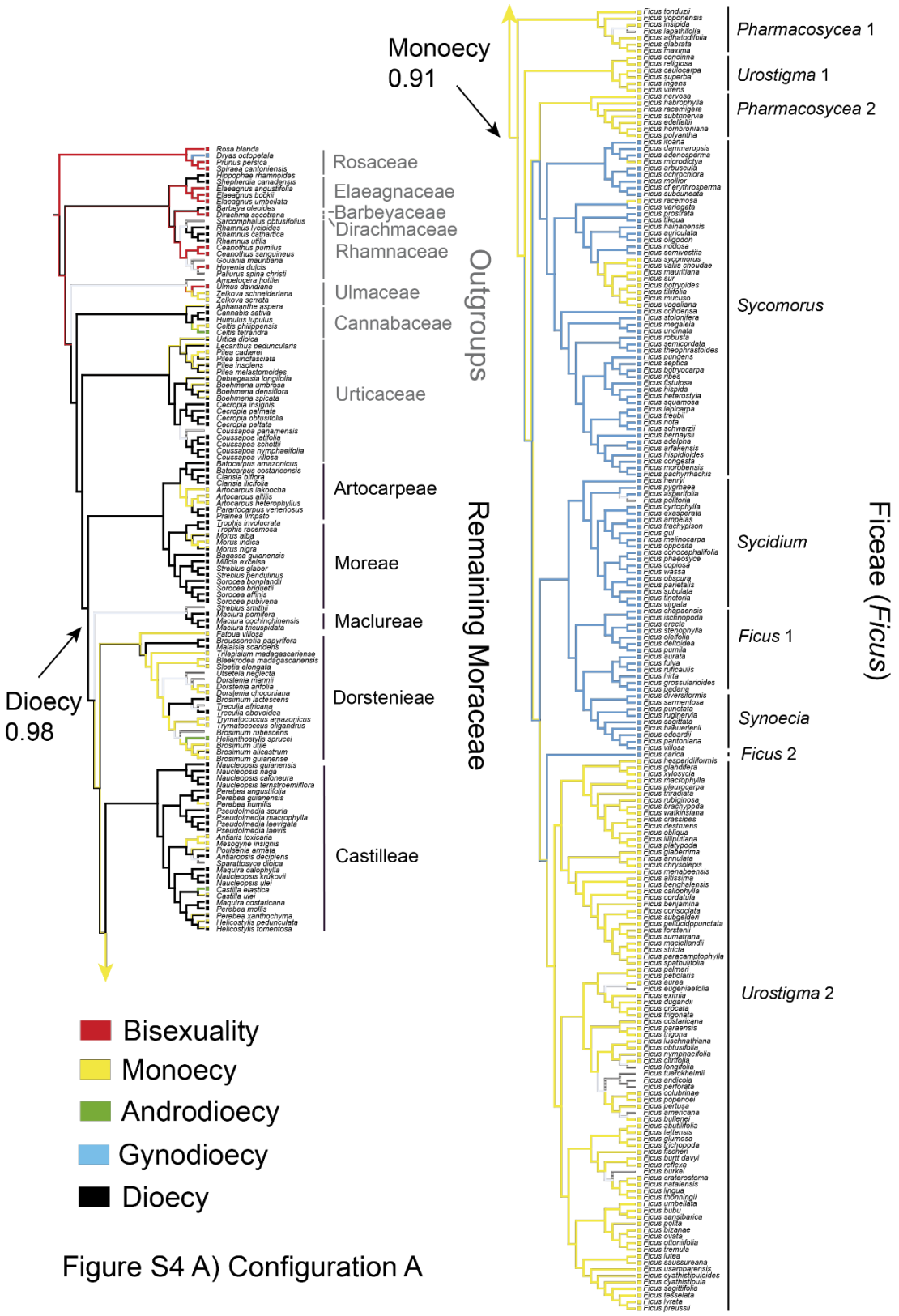


Figure S4 A) Configuration A

Figure S4. Ancestral state reconstruction with 320-species dataset by equal-rate maximum likelihood with configurations A to F (configuration B presented as Fig. 3).

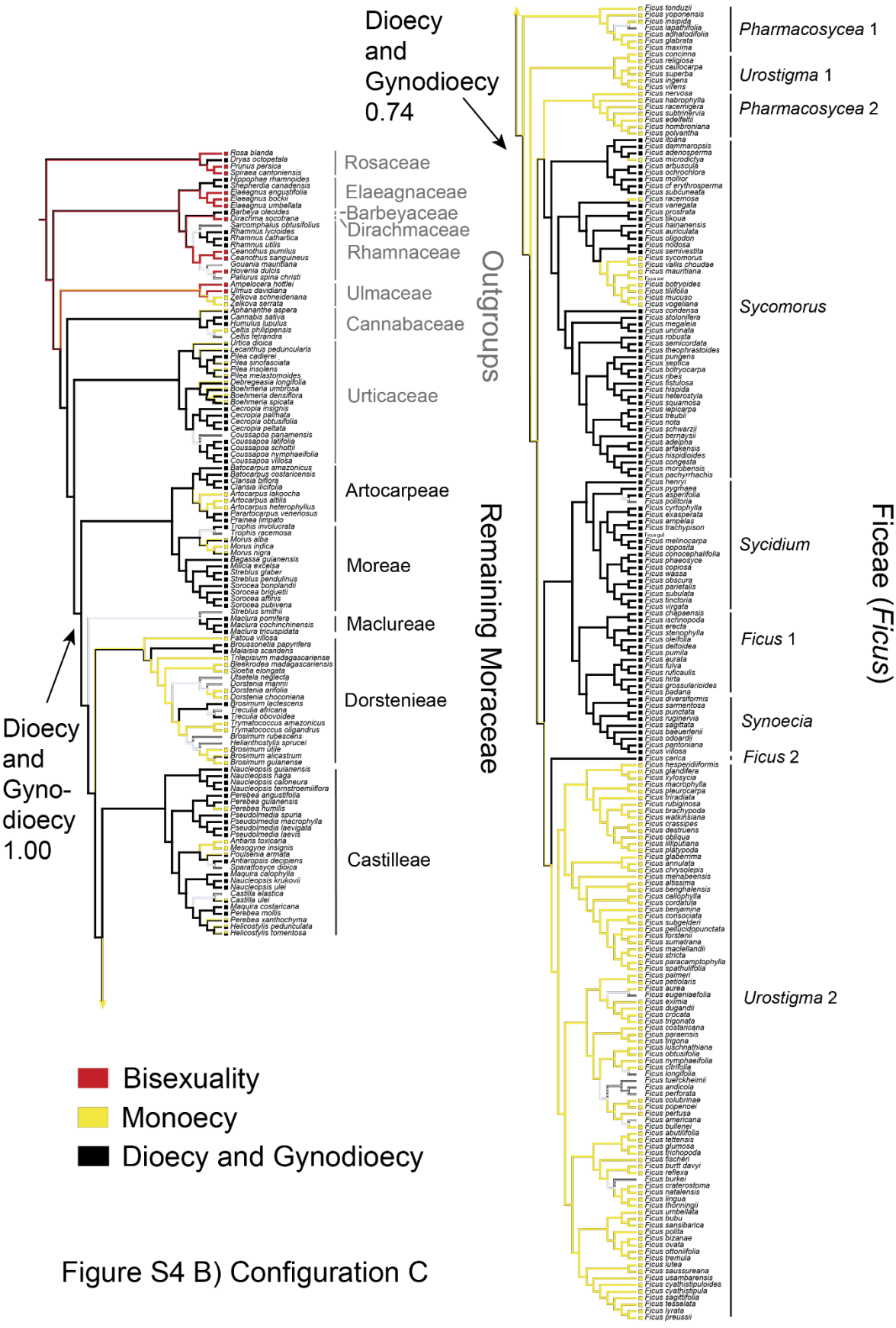


Figure S4 B) Configuration C

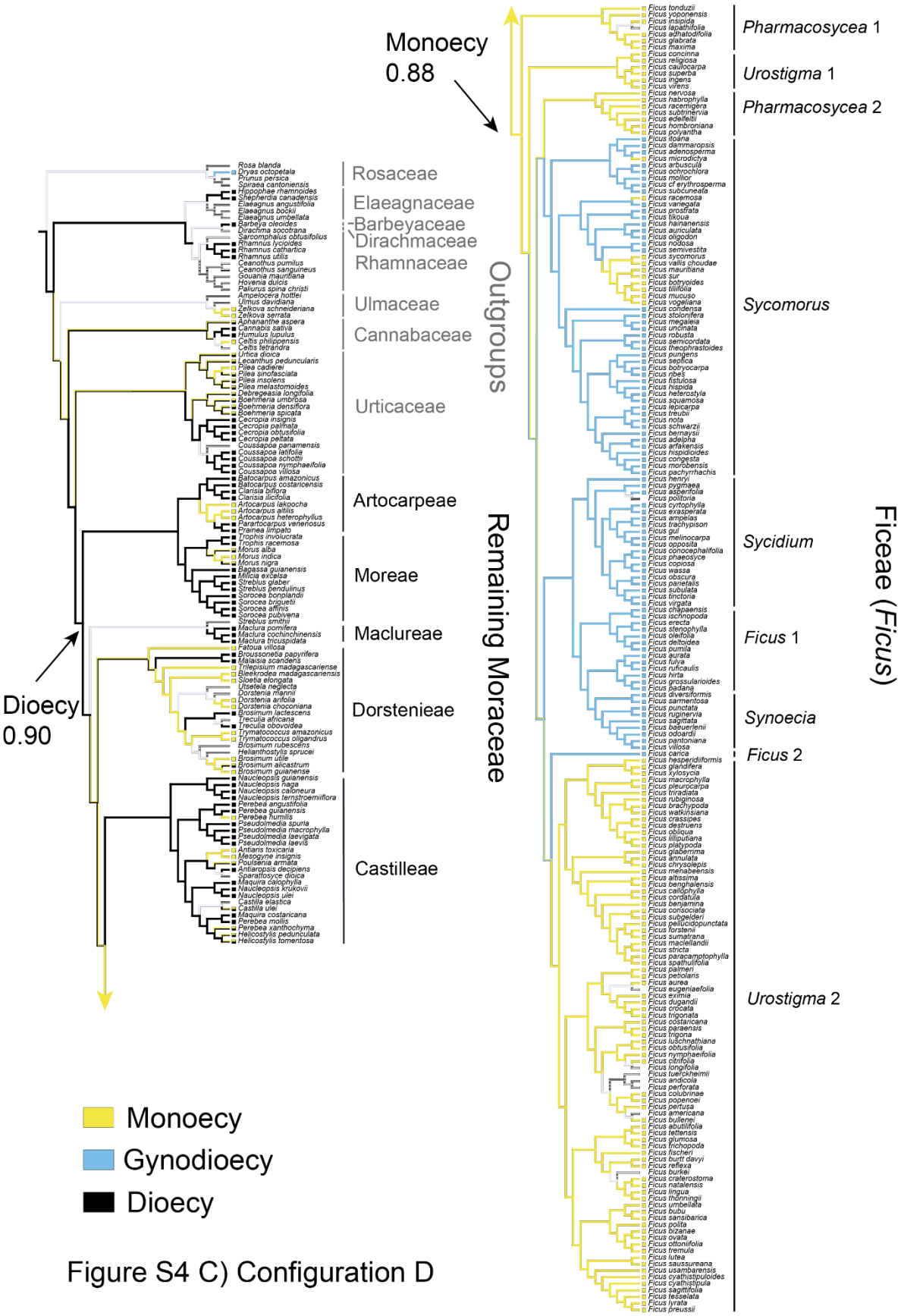


Figure S4 C) Configuration D

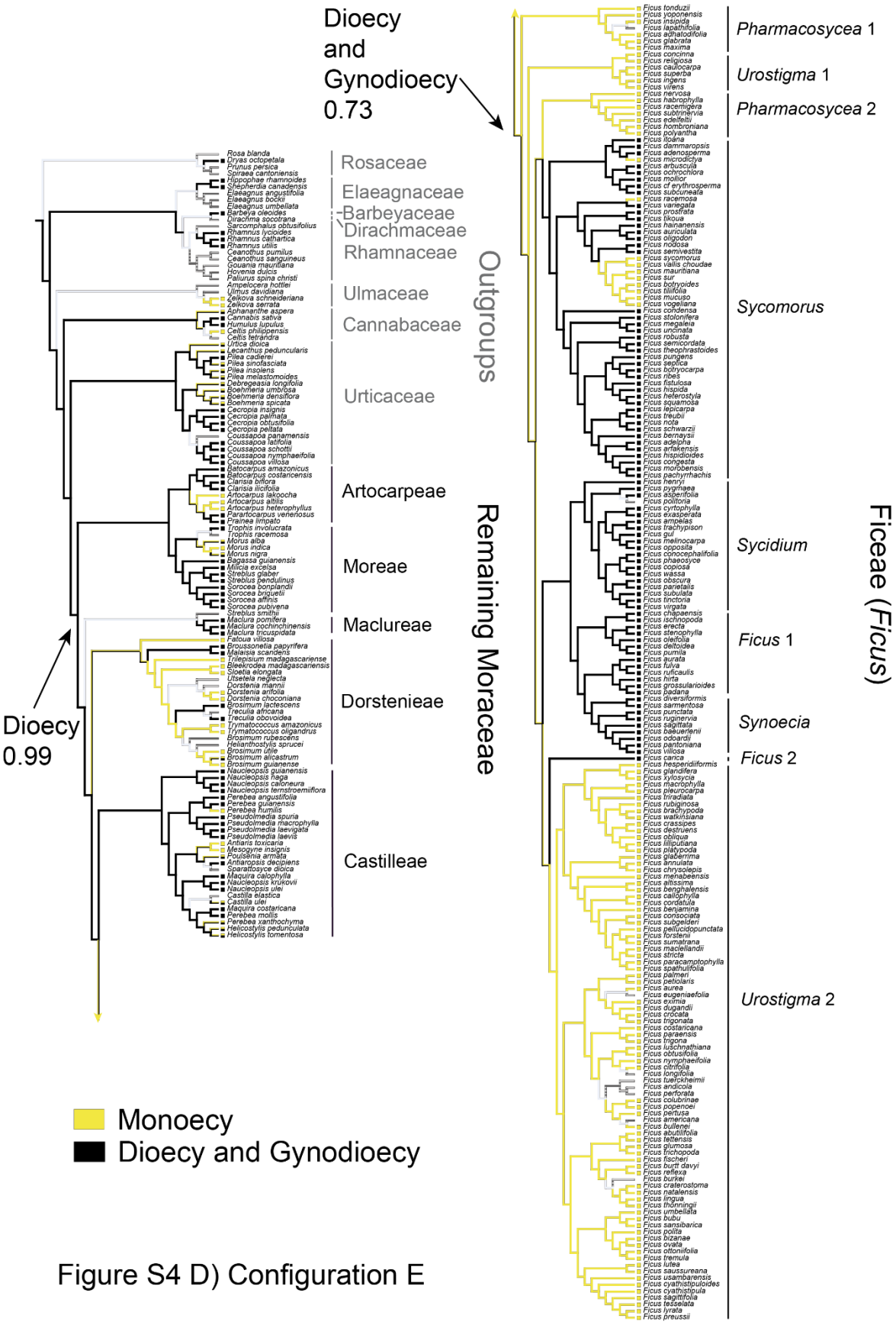


Figure S4 D) Configuration E

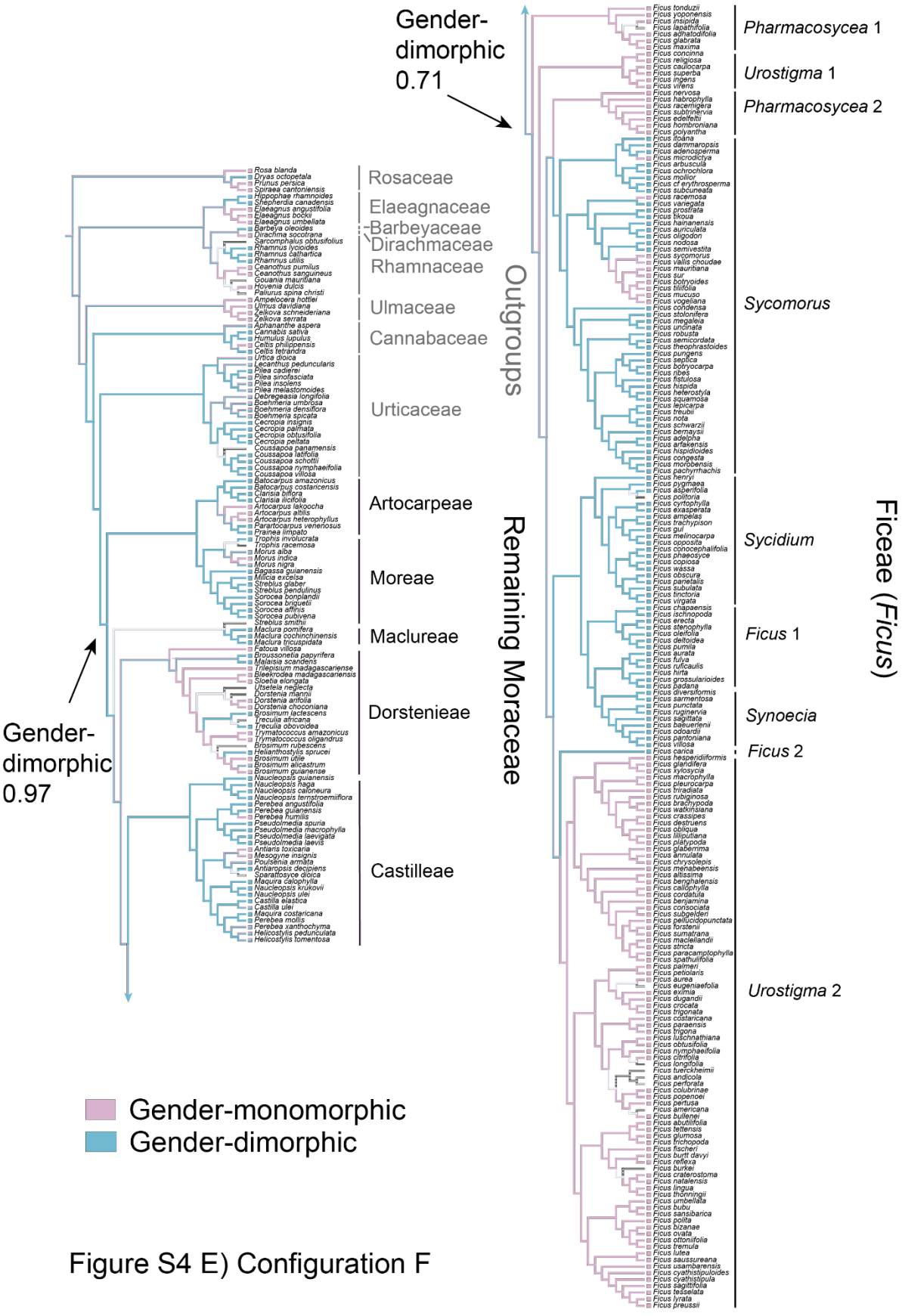
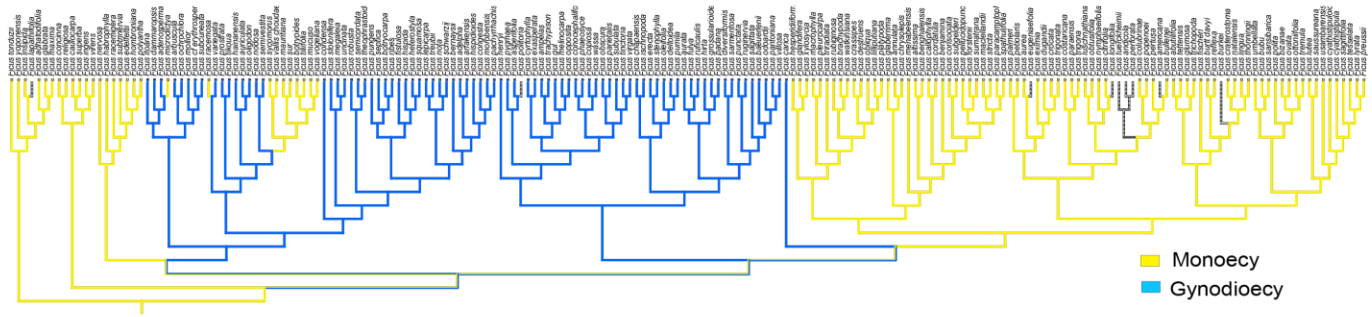
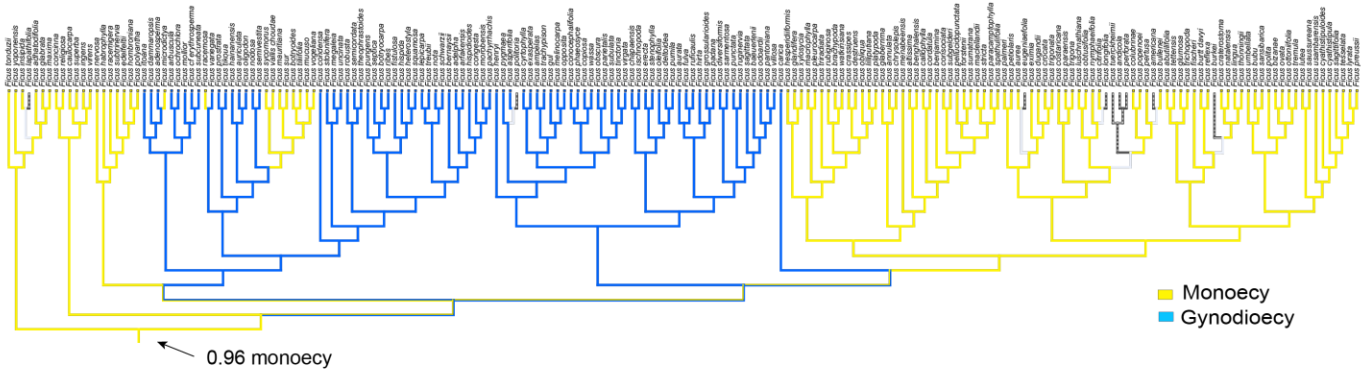


Figure S4 E) Configuration F



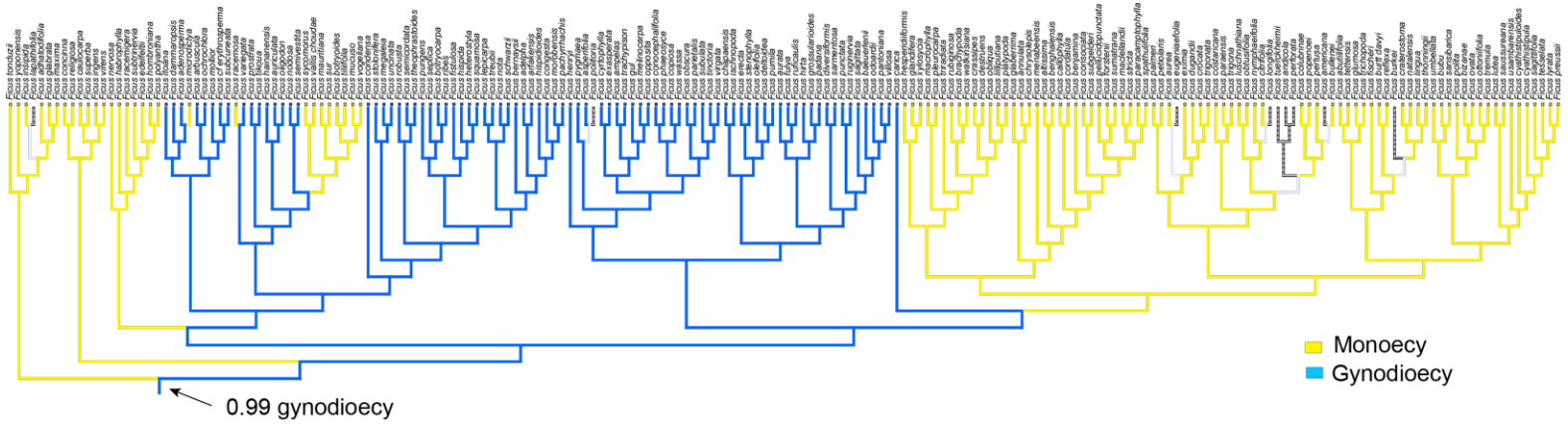
A) *Ficus* scale ancestral state reconstruction by parsimony approach



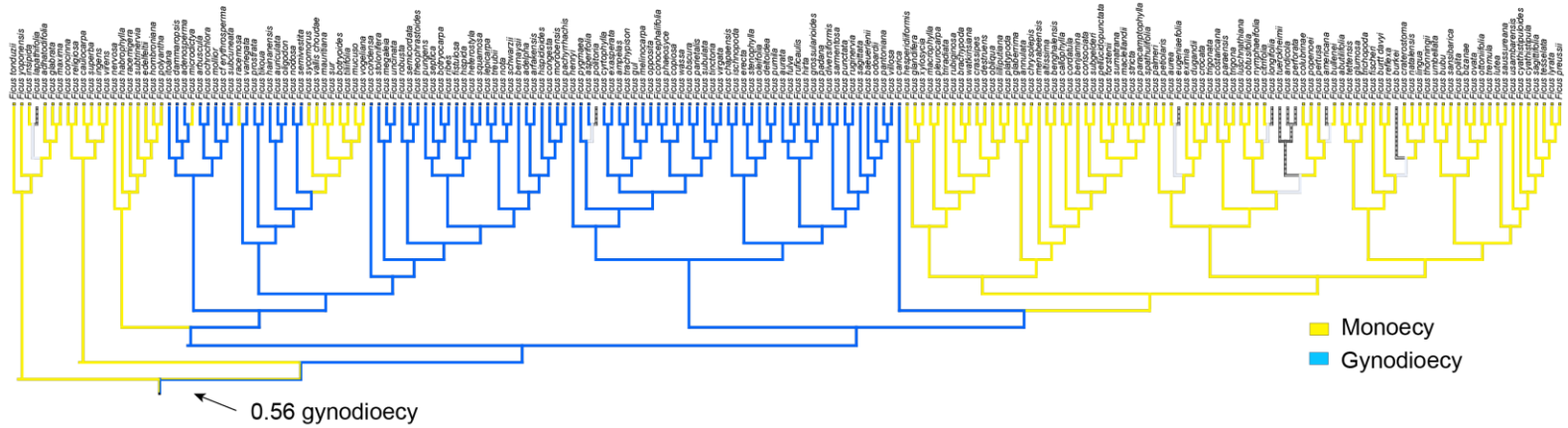
B) *Ficus* scale ancestral state reconstruction
with equal-rate model by maximum likelihood (ML) approach

Figure S5

Figure S5. Ancestral state reconstruction with only Ficus species in the dataset with tip names by different approaches: A) parsimony; B) equal rate maximum likelihood; C) unequal rate maximum likelihood with equal root state prior; D) unequal rate maximum likelihood with equilibrium root state prior.



C) *Ficus* scale ancestral state reconstruction
with unequal-rate model (equal frequencies root state prior) by maximum likelihood (ML) approch



D) *Ficus* scale ancestral state reconstruction
with unequal-rate model (equilibrium frequencies root state prior) by maximum likelihood (ML) approach

Figure S5

Table S1. GenBank accession numbers for the sequences used in this study, the hyphen symbol denotes missing data.

spp.	tribe	family	rbcL	matK	ndhF	ITS	ETS	G3pdh	ncpGS	GBSSI
Barbeya oleoides		Barbeyaceae	JF317477	JF317418	-	-	-	-	-	-
Humulus lupulus		Cannabaceae	KM360825	AY257528	AY289251	-	-	-	-	-
Celtis tetrandra		Cannabaceae	JF317479	-	JF317439	-	-	-	-	-
Celtis philippensis		Cannabaceae	KR528952	KR530552	AY289249	-	-	-	-	-
Cannabis sativa		Cannabaceae	AF500344	AF345317	AY289250	-	-	-	-	-
Aphananthe aspera		Cannabaceae	AF500339	AF345320	AF500366	-	-	-	-	-
Dirachma socotrana		Dirachmaceae	JF317482	JF317423	-	-	-	-	-	-
Shepherdia canadensis		Elaeagnaceae	U17039	KC475874	-	-	-	-	-	-
Hippophae rhamnoides		Elaeagnaceae	JF317488	JF317428	JF317448	-	-	-	-	-
Elaeagnus umbellata		Elaeagnaceae	KP088580	AY257529	-	-	-	-	-	-
Elaeagnus bockii		Elaeagnaceae	JF317484	JF317425	JF317444	-	-	-	-	-
Elaeagnus angustifolia		Elaeagnaceae	U17038	KP089052	-	-	-	-	-	-
Treculia obovoidea	Artocarpeae	Moraceae	KC628408	KC627751	-	-	-	-	-	-
Treculia africana	Artocarpeae	Moraceae	KC628540	KC627842	-	-	-	-	-	-
Prainea limpato	Artocarpeae	Moraceae	-	-	AY289296	-	-	-	-	-
Parartocarpus venenosus	Artocarpeae	Moraceae	-	-	AY289289	-	-	-	-	-
Clarisia ilicifolia	Artocarpeae	Moraceae	-	-	AY289293	-	-	-	-	-
Clarisia biflora	Artocarpeae	Moraceae	JQ592804	-	AY289292	-	-	-	-	-
Batocarpus costaricensis	Artocarpeae	Moraceae	-	-	AY289290	-	-	-	-	-
Batocarpus amazonicus	Artocarpeae	Moraceae	-	-	AY289291	-	-	-	-	-
Artocarpus lakoocha	Artocarpeae	Moraceae	KR528787	KR530413	AY289287	-	-	-	-	-
Artocarpus heterophyllus	Artocarpeae	Moraceae	KF724291	-	AY289285	-	-	-	-	-
Artocarpus altilis	Artocarpeae	Moraceae	AF500345	HM446658	AY289286	-	-	-	-	-
Sparattosyce dioica	Castilleae	Moraceae	-	-	AY289302	-	-	EU087607	-	-
Pseudolmedia spuria	Castilleae	Moraceae	HM446858	HM446734	AY289325	-	-	-	-	-
Pseudolmedia macrophylla	Castilleae	Moraceae	-	-	AY289324	-	-	-	-	-

Pseudolmedia laevis	Castilleae	Moraceae	-	-	AY289323	-	-	-	-	-
Pseudolmedia laevigata	Castilleae	Moraceae	-	-	AY289326	-	-	-	-	-
Poulsenia armata	Castilleae	Moraceae	JX987593	-	EU422993	-	-	-	-	EU084353
Perebea xanthochyma	Castilleae	Moraceae	GQ981827	GQ982060	AY289321	-	-	-	-	-
Perebea mollis	Castilleae	Moraceae	JQ625992	-	AY289322	-	-	-	-	-
Perebea humilis	Castilleae	Moraceae	-	-	AY289318	-	-	-	-	-
Perebea guianensis	Castilleae	Moraceae	-	-	AY289319	-	-	-	-	-
Perebea angustifolia	Castilleae	Moraceae	-	-	AY289320	-	-	-	-	-
Naucleopsis ulei	Castilleae	Moraceae	-	-	AY289314	-	-	-	-	-
Naucleopsis ternstroemiiflora	Castilleae	Moraceae	-	-	AY289316	-	-	-	-	-
Naucleopsis naga	Castilleae	Moraceae	-	-	AY289313	-	-	-	-	-
Naucleopsis krukovii	Castilleae	Moraceae	-	-	AY289315	-	-	-	-	-
Naucleopsis guianensis	Castilleae	Moraceae	GQ428596	-	AY289317	-	-	-	-	-
Naucleopsis caloneura	Castilleae	Moraceae	-	-	AY289312	-	-	-	-	-
Mesogyne insignis	Castilleae	Moraceae	-	-	AY289311	-	-	-	-	-
Maquira costaricana	Castilleae	Moraceae	-	-	AY289310	-	-	-	-	-
Maquira calophylla	Castilleae	Moraceae	FJ038123	FJ514665	AY289309	-	-	-	-	-
Castilla ulei	Castilleae	Moraceae	-	-	AY289305	-	-	-	-	-
Castilla elastica	Castilleae	Moraceae	AF500348	JQ588395	AY289304	-	-	EF092327	-	EU084352
Antiaropsis decipiens	Castilleae	Moraceae	-	-	AY289284	-	-	EF092326	-	-
Antiaris toxicaria	Castilleae	Moraceae	GQ436642	GQ434236	AY289303	-	-	-	-	-
Helicostylis tomentosa	Castilleae	Moraceae	FJ038122	FJ514761	AY289307	-	-	-	-	-
Helicostylis pedunculata	Castilleae	Moraceae	FJ038121	FJ514731	AY289308	-	-	-	-	-
Utsetela neglecta	Dorstenieae	Moraceae	-	-	AY289339	-	-	-	-	-
Trymatococcus oligandrus	Dorstenieae	Moraceae	JQ625978	FJ037932	AY289338	-	-	-	-	-
Trymatococcus amazonicus	Dorstenieae	Moraceae	JQ626260	JQ626558	AY289337	-	-	-	-	-
Trilepisium madagascariense	Dorstenieae	Moraceae	-	-	AY289336	-	-	-	-	-
Sloetia elongata	Dorstenieae	Moraceae	-	-	AY289280	-	-	-	-	-
Malaisia scandens	Dorstenieae	Moraceae	KM895723	-	AY289281	-	-	-	-	-
Helianthostylis sprucei	Dorstenieae	Moraceae	-	-	AY289335	-	-	-	-	-
Fatoua villosa	Dorstenieae	Moraceae	KJ773508	KF137999	AY289270	-	-	-	-	-

Dorstenia mannii	Dorstenieae	Moraceae	AF500349	-	AF500376	-	-	-	-	-
Dorstenia choconiana	Dorstenieae	Moraceae	-	-	AY289334	-	-	-	-	-
Dorstenia arifolia	Dorstenieae	Moraceae	-	-	AY289332	-	-	-	-	-
Broussonetia papyrifera	Dorstenieae	Moraceae	AF500347	AF345326	AY289269	-	-	-	-	-
Brosimum utile	Dorstenieae	Moraceae	JQ626232	-	AY289327	-	-	-	-	-
Brosimum rubescens	Dorstenieae	Moraceae	JQ625739	JQ626346	AY289330	-	-	-	-	-
Brosimum lactescens	Dorstenieae	Moraceae	JQ592792	JQ588393	AY289329	-	-	-	-	-
Brosimum guianense	Dorstenieae	Moraceae	JQ626188	GQ981948	-	-	-	-	-	-
Brosimum alicastrum	Dorstenieae	Moraceae	AF500346	GQ981947	AY289328	-	-	-	-	-
Bleekrodea madagascariensis	Dorstenieae	Moraceae	-	-	AY289268	-	-	-	-	-
Maclura tricuspidata	Maclureae	Moraceae	JF317480	JF317421	AY289272	-	-	-	-	-
Maclura pomifera	Maclureae	Moraceae	D86318	KP089143	AY289273	-	-	-	-	-
Maclura cochinchinensis	Maclureae	Moraceae	JF738991	-	AY289271	-	-	-	-	-
Trophis racemosa	Moreae	Moraceae	GQ981908	GQ982120	AY289283	-	-	-	-	-
Trophis involucrata	Moreae	Moraceae	JQ592884	JQ588436	AY289282	-	-	-	-	-
Streblus smithii	Moreae	Moraceae	-	-	AY289278	-	-	-	-	-
Streblus pendulinus	Moreae	Moraceae	AF500353	KM894939	AY289279	-	-	-	-	-
Streblus glaber	Moreae	Moraceae	-	-	AY289277	-	-	-	-	-
Sorocea pubivena	Moreae	Moraceae	-	-	AY289300	-	-	-	-	-
Sorocea briquetii	Moreae	Moraceae	-	-	AY289298	-	-	-	-	-
Sorocea bonplandii	Moreae	Moraceae	-	-	AY289299	-	-	-	-	-
Sorocea affinis	Moreae	Moraceae	GQ981880	GQ982100	AY289297	-	-	-	-	-
Morus nigra	Moreae	Moraceae	JX571868	JX495737	AY289275	-	-	-	-	-
Morus indica	Moreae	Moraceae	DQ226511	DQ226511	DQ226511	-	-	-	-	-
Morus alba	Moreae	Moraceae	D86319	AY257531	AY289274	-	-	-	-	-
Milicia excelsa	Moreae	Moraceae	JX572771	JX517997	AY289276	-	-	-	-	-
Bagassa guianensis	Moreae	Moraceae	JQ625997	FJ514656	AY289267	-	-	-	-	-
Ficus yoponensis	Ficeae	Moraceae	JQ592862	GQ981999	-	AY063594	AY063552	AY967959	-	-
Ficus xylosycia	Ficeae	Moraceae	-	-	-	AF165419	-	EF538801	-	-
Ficus watkinsiana	Ficeae	Moraceae	-	-	-	AY730118	AY730208	EF092365	EU084310	EU084367
Ficus wassa	Ficeae	Moraceae	JF738430	-	AY289348	AF165418	EF092325	DQ367635	-	DQ367655

Ficus vogeliana	Ficeae	Moraceae	-	-	-	EU091610	EU084440	EU087650	-	-
Ficus virgata	Ficeae	Moraceae	-	-	-	AF165417	AY730224	EF092404	EU084351	EU084393
Ficus virens	Ficeae	Moraceae	JQ773811	JQ773627	AY289346	AF165416	AY730150	DQ367634	-	DQ367654
Ficus villosa	Ficeae	Moraceae	-	-	-	AY730130	AY730217	EF092391	EU084340	EU084389
Ficus variegata	Ficeae	Moraceae	FJ976133	JQ773615	AY289344	AF165415	AY063539	HQ890563	EU084323	DQ367653
Ficus vallis-choudae	Ficeae	Moraceae	-	-	-	AY063574	AY063535	EF092373	EU084321	EU084373
Ficus usambarensis	Ficeae	Moraceae	-	-	-	DQ455653	DQ455677	-	-	-
Ficus uncinata	Ficeae	Moraceae	-	-	-	AY063576	AY063537	EU087669	-	-
Ficus umbellata	Ficeae	Moraceae	-	-	-	DQ455644	DQ455674	-	DQ455629	-
Ficus tuerckheimii	Ficeae	Moraceae	-	-	-	EU091608	EU084438	EU087640	-	-
Ficus triradiata	Ficeae	Moraceae	-	-	-	AY730117	AY730207	EF092364	-	-
Ficus trigonata	Ficeae	Moraceae	GQ981743	JX495719	-	EU091607	-	AY967956	-	-
Ficus trigona	Ficeae	Moraceae	GU935084	-	-	DQ455669	DQ455688	AY967973	DQ455619	EU084368
Ficus trichopoda	Ficeae	Moraceae	JX572612	JX517724	-	DQ455666	DQ455684	EU087648	-	-
Ficus treubii	Ficeae	Moraceae	-	-	-	EU091636	EU084463	EU087668	-	-
Ficus tremula	Ficeae	Moraceae	JX573114	JX970900	-	AY730111	AY730200	-	-	-
Ficus trachypison	Ficeae	Moraceae	JF739063	-	-	EU091674	EU084493	EU087688	-	-
Ficus tonduzii	Ficeae	Moraceae	JQ592861	GQ981998	-	AY730140	AY730230	EU087611	EU084297	
Ficus tinctoria	Ficeae	Moraceae	JF941560	JF953747	-	AF165413	AY730223	EF092403	-	-
Ficus tiliifolia	Ficeae	Moraceae	-	-	-	EU091609	EU084439	-	-	-
Ficus tikoua	Ficeae	Moraceae	JF317485	JF317426	-	EU091641	EU084468	EU087673	-	-
Ficus thonningii	Ficeae	Moraceae	JF265432	JF270781	-	AY730102	AY730191	EF092353	-	-
Ficus theophrastoides	Ficeae	Moraceae	-	-	-	AF165412	EU084462	-	-	-
Ficus tettensis	Ficeae	Moraceae	JX572611	JX517998	-	DQ455665	DQ455683	-	DQ455620	-
Ficus tesselata	Ficeae	Moraceae	-	-	-	DQ455662	DQ455682	EU087647	-	-
Ficus sycomorus	Ficeae	Moraceae	EU213482	JX495717	-	AY063575	AY063536	-	EU084320	-
Ficus sur	Ficeae	Moraceae	JF265438	JF270786	-	AF165411	AY063533	EU087649	EU084319	EU084372
Ficus superba	Ficeae	Moraceae	KP094195	KP093287	-	AF165410	AY730149	EF092332	DQ455631	-
Ficus sumatrana	Ficeae	Moraceae	-	-	-	EU091597	-	EU087634	-	-
Ficus subulata	Ficeae	Moraceae	KR529331	JQ773598	-	EU091677	EU084495	EU087690	-	-
Ficus subtrinervia	Ficeae	Moraceae	JF738457	-	-	AY730119	EU084411	EU087617	-	-

Ficus subgelderi	Ficeae	Moraceae	-	-	-	AY063556	AY063517	EF092336	-	-
Ficus subcuneata	Ficeae	Moraceae	JF738942	-	-	EU091620	EU084449	DQ367631	-	DQ367651
Ficus stricta	Ficeae	Moraceae	KR529326	KR530856	-	EU091595	EU084429	EU087632	-	-
Ficus stolonifera	Ficeae	Moraceae	-	-	-	EU091635	-	-	-	-
Ficus stenophylla	Ficeae	Moraceae	JQ773771	JQ773589	-	EU091640	EU084467	-	-	-
Ficus squamosa	Ficeae	Moraceae	JQ773768	JQ773586	-	EU091634	-	-	-	-
Ficus spathulifolia	Ficeae	Moraceae	-	-	-	EU091594	EU084428	EU087631	-	-
Ficus septica	Ficeae	Moraceae	JQ773766	JQ773585	AY289345	AF165409	AY730229	HQ890558	-	DQ367650
Ficus semivestita	Ficeae	Moraceae	JF738449	-	-	EU091616	EU084443	DQ367629	-	DQ367649
Ficus semicordata	Ficeae	Moraceae	KP752379	JF953744	-	EU091613	EU084441	EU087652	EU084322	
Ficus schwarzii	Ficeae	Moraceae	-	-	-	EU091633	-	-	-	-
Ficus saussureana	Ficeae	Moraceae	-	-	-	AY730090	AY730179	-	-	-
Ficus sarmentosa	Ficeae	Moraceae	JQ773753	JQ773578	-	EU091653	EU084478	EU087679	-	-
Ficus sansibarica	Ficeae	Moraceae	KF147479	KF147405	-	AY730110	AY730199	EF092359	-	-
Ficus sagittifolia	Ficeae	Moraceae	-	-	-	AY730106	AY730195	EF092356	DQ455626	-
Ficus sagittata	Ficeae	Moraceae	JQ773749	JQ773575	-	EU091652	EU084477	EU087678	EU084339	-
Ficus ruginervia	Ficeae	Moraceae	-	-	-	AF165407	EF092323	EF092393	-	-
Ficus ruficaulis	Ficeae	Moraceae	KJ688752	-	-	EU091647	-	-	EU084337	-
Ficus rubiginosa	Ficeae	Moraceae	KM895977	KM894812	-	AY063569	AY063530	EF092363	DQ455635	EU084366
Ficus robusta	Ficeae	Moraceae	-	-	-	AF165406	EU084442	-	-	DQ367648
Ficus ribes	Ficeae	Moraceae	-	-	-	EU091630	EU084458	EU087665	-	-
Ficus religiosa	Ficeae	Moraceae	KP088599	KP089073	-	AY063582	AY063543	EF092331	-	-
Ficus reflexa	Ficeae	Moraceae	-	-	-	DQ455650	-	EU087646	-	-
Ficus racemosa	Ficeae	Moraceae	KT368151	KT368151	AY289349	AF165405	-	-	EU084318	EU084371
Ficus racemigera	Ficeae	Moraceae	-	-	-	AY063587	AY063554	-	-	-
Ficus pygmaea	Ficeae	Moraceae	JX572608	JX517453	-	AY730134	AY730221	EF092399	EU084350	-
Ficus pungens	Ficeae	Moraceae	JF739141	-	-	AF165404	-	DQ367627	-	DQ367647
Ficus punctata	Ficeae	Moraceae	-	-	-	AF165403	AY063545	-	EU084343	-
Ficus pumila	Ficeae	Moraceae	AF500352	HM851109	AF500378	AY063580	AY063541	EF092390	-	EU084388
Ficus prostrata	Ficeae	Moraceae	-	-	-	EU091612	-	-	-	-
Ficus preussii	Ficeae	Moraceae	-	-	-	AY730105	AY730194	EF092355	DQ455625	-

Ficus popenoei	Ficeae	Moraceae	GQ981741	GQ981997	-	EU081761	-	AY967975	-	-
Ficus polyantha	Ficeae	Moraceae	JF738527	-	-	EU091571	-	EU087616	-	-
Ficus politoria	Ficeae	Moraceae	-	-	-	EU091671	-	EU087687	-	-
Ficus polita	Ficeae	Moraceae	JX572607	JX518117	-	DQ455642	DQ455673	-	-	-
Ficus pleurocarpa	Ficeae	Moraceae	-	-	-	AY063568	AY063529	EF538795	DQ455634	-
Ficus platypoda	Ficeae	Moraceae	-	-	-	AY730114	AY730204	EF538794	-	-
Ficus phaeosyce	Ficeae	Moraceae	-	-	-	AF165401	-	-	-	-
Ficus petiolaris	Ficeae	Moraceae	-	-	-	AY730088	AY730177	-	-	-
Ficus pertusa	Ficeae	Moraceae	JQ592858	JQ588412	-	AF165400	AY730176	AY967950	-	-
Ficus perforata	Ficeae	Moraceae	-	-	-	AY730087	AY730175	AY967951	-	-
Ficus pellucidopunctata	Ficeae	Moraceae	-	-	-	AF165399	EU084427	-	-	-
Ficus parietalis	Ficeae	Moraceae	-	-	-	AY063583	AY063544	EF092401	-	-
Ficus paraensis	Ficeae	Moraceae	-	-	-	AY730086	AY730174	AY967954	-	-
Ficus paracamptophylla	Ficeae	Moraceae	-	-	-	EU091592	EU084426	-	-	-
Ficus pantoniana	Ficeae	Moraceae	-	-	-	EU091649	-	-	-	-
Ficus palmeri	Ficeae	Moraceae	-	-	-	AY730085	AY730173	-	-	-
Ficus padana	Ficeae	Moraceae	-	-	-	AF165398	-	EF092387	-	-
Ficus pachyrrhachis	Ficeae	Moraceae	JF738400	-	-	EU091628	EU084456	DQ367626	EU084328	DQ367646
Ficus ovata	Ficeae	Moraceae	-	-	-	DQ455640	DQ455672	-	-	-
Ficus ottoniifolia	Ficeae	Moraceae	-	-	-	AY730109	AY730198	EF092358	-	-
Ficus opposita	Ficeae	Moraceae	KM895642	KM894541	-	EU091670	-	EU087686	-	-
Ficus oligodon	Ficeae	Moraceae	JQ773716	JQ773552	-	-	-	-	-	-
Ficus oleifolia	Ficeae	Moraceae	-	-	-	AY730124	EF092322	EF092382	EU084332	EU084384
Ficus odoardii	Ficeae	Moraceae	-	-	-	AF165397	-	EF092389	-	-
Ficus ochrochlora	Ficeae	Moraceae	-	-	-	AF165396	EU084448	-	-	EU084378
Ficus obtusifolia	Ficeae	Moraceae	JQ592846	GQ981996	-	AY730084	AY730172	AY967949	-	-
Ficus obscura	Ficeae	Moraceae	-	-	-	EU091676	-	EU087689	-	-
Ficus obliqua	Ficeae	Moraceae	-	KM894609	-	EF545659	EF538774	EF538793	-	-
Ficus nymphaeifolia	Ficeae	Moraceae	JQ592843	-	-	AY063566	AY063527	EU089843	-	-
Ficus nota	Ficeae	Moraceae	-	-	-	EU091626	-	EU087663	EU084327	-
Ficus nodosa	Ficeae	Moraceae	JF739106	-	-	AF165395	-	DQ367625	-	DQ367645

Ficus nervosa	Ficeae	Moraceae	KP752397	JQ773551	-	EU091570	EU084410	EU087615	-	-
Ficus natalensis	Ficeae	Moraceae	JQ773710	KF147404	-	AY730100	AY730189	EF092352	-	-
Ficus mucuso	Ficeae	Moraceae	-	-	-	AY730120	AY730210	EF092372	EU084317	-
Ficus morobensis	Ficeae	Moraceae	-	-	-	DQ367659	EU084455	DQ367624	-	DQ367644
Ficus mollior	Ficeae	Moraceae	JF738492	-	-	DQ367658	-	DQ367623	-	DQ367643
Ficus microdictya	Ficeae	Moraceae	-	-	-	AF165394	EU084447	EU087656	-	EU084377
Ficus menabeensis	Ficeae	Moraceae	-	-	-	AY730067	AY730155	-	-	-
Ficus melinocarpa	Ficeae	Moraceae	JF739039	-	-	EU091669	-	EU087685	-	-
Ficus megaleia	Ficeae	Moraceae	-	-	-	EU091625	-	EU087661	-	-
Ficus maxima	Ficeae	Moraceae	GQ981739	GQ981995	-	AY063595	AY063551	AY967958	-	-
Ficus mauritiana	Ficeae	Moraceae	-	-	-	AY063570	AY063531	EF092371	-	-
Ficus macrophylla	Ficeae	Moraceae	JX571836	JX495714	-	AY730115	AY730205	EF538792	-	-
Ficus maclellandii	Ficeae	Moraceae	JQ773704	JQ773543	-	EU091591	EU084425	EU087629	-	EU084365
Ficus lyrata	Ficeae	Moraceae	JF941548	-	-	AY730104	AY730193	-	-	-
Ficus lutea	Ficeae	Moraceae	-	-	-	AY063564	AY063525	EF092347	-	-
Ficus luschnathiana	Ficeae	Moraceae	-	-	-	AY730082	AY730170	EF092345	-	-
Ficus longifolia	Ficeae	Moraceae	-	-	-	EU091604	-	-	-	-
Ficus lingua	Ficeae	Moraceae	-	-	-	AY730099	AY730188	EF092351	-	-
Ficus lilliputiana	Ficeae	Moraceae	-	-	-	EF545657	EF538773	-	-	-
Ficus lepicarpa	Ficeae	Moraceae	-	-	-	AY730138	-	EF092376	-	-
Ficus lapathifolia	Ficeae	Moraceae	-	-	-	EU091564	EU084405	EU087609	-	-
Ficus itoana	Ficeae	Moraceae	-	-	-	AF165391	EU084446	EU087655	-	EU084376
Ficus ischnopoda	Ficeae	Moraceae	JF941545	JQ773541	-	AY730122	AY730212	EF092380	-	EU084383
Ficus insipida	Ficeae	Moraceae	GQ981738	GQ981994	AY289343	AY063592	AY063549	AY967961	EU084296	EU084354
Ficus ingens	Ficeae	Moraceae	JF265434	JF270782	-	AY730061	AY730147	EF092330	EU084303	-
Ficus hombroniana	Ficeae	Moraceae	JF739010	-	-	AF165389	-	EF092369	-	-
Ficus hispidioides	Ficeae	Moraceae	JF739068	-	-	AF165388	AY730227	DQ367622	-	DQ367642
Ficus hispida	Ficeae	Moraceae	JQ773694	JQ773498	-	EU091623	EU084454	EU087659	EU084326	
Ficus hirta	Ficeae	Moraceae	JQ773693	HQ415330	-	AY730127	EU084473	EF092386	-	-
Ficus heterostyla	Ficeae	Moraceae	-	-	-	EU091611	-	EU087651	-	-
Ficus hesperidiiformis	Ficeae	Moraceae	-	-	-	AF165387	AY730203	EF092362	-	-

Ficus henryi	Ficeae	Moraceae	-	-	-	EU091639	EU084466	EU087672	EU084331	-
Ficus hainanensis	Ficeae	Moraceae			-	EU091614	-	-	-	-
Ficus habrophylla	Ficeae	Moraceae	-	-	AY289341	EU091567	EU084408	EU087612	-	-
Ficus gul	Ficeae	Moraceae	FJ976131	-	-	AY730132	AY730219	EF092397	EU084349	-
Ficus grossularioides	Ficeae	Moraceae	KJ594707	KJ708927	-	AY063591	-	EF092385	EU084336	EU084386
Ficus glumosa	Ficeae	Moraceae	EU213479	EU214251	-	AY063562	AY063523	-	EU084316	-
Ficus glandifera	Ficeae	Moraceae	-	-	-	AY730113	AY730202	EF092361	-	-
Ficus glabrata	Ficeae	Moraceae	-	-	-	AY063593	AY063550	AY967960	-	-
Ficus glaberrima	Ficeae	Moraceae	JF941535	JQ773533	-	EU091588	-	EU087627	-	-
Ficus fulva	Ficeae	Moraceae	JQ773686	JQ773530	-	-	-	EU087675	EU084335	-
Ficus forstenii	Ficeae	Moraceae	-	-	-	EU091587	-	EU087626	-	-
Ficus fistulosa	Ficeae	Moraceae	JQ773681	KR530820	-	AY730137	AY730226	EF092375	-	EU084379
Ficus fischeri	Ficeae	Moraceae	-	-	-	DQ455649	AY730187	EF092350	DQ455623	-
Ficus eximia	Ficeae	Moraceae	-	-	-	AY730079	AY730167	EF092344	-	-
Ficus exasperata	Ficeae	Moraceae	-	-	-	EU091665	EU084489	EU087683	EU084347	EU084392
Ficus eugeniaefolia	Ficeae	Moraceae	-	-	-	AY730078	AY730166	-	-	-
Ficus erecta	Ficeae	Moraceae	JQ773677	JQ773526	-	AY730121	AY730211	EF092379	EU084330	-
Ficus edelfeltii	Ficeae	Moraceae	-	-	AY289342	AF165385	AY730209	-	-	-
Ficus dugandii	Ficeae	Moraceae	-	-	-	EU081763	-	AY967957	-	-
Ficus diversiformis	Ficeae	Moraceae	-	-	-	AY730128	AY730215	EF092392	-	-
Ficus destruens	Ficeae	Moraceae	KF496521	-	-	AF165384	EF538769	EF538790	-	-
Ficus deltoidea	Ficeae	Moraceae	-	-	-	AY063579	AY063540	EF092378	-	-
Ficus dammaropsis	Ficeae	Moraceae	-	-	-	AF165383	EU084445	DQ367621	-	DQ367641
Ficus cyrtophylla	Ficeae	Moraceae	KR529277	KR530811		EU091664	EU084488	-	EU084346	-
Ficus cyathistipuloides	Ficeae	Moraceae	-	-	-	AY063563	AY063524	EU087645	-	-
Ficus cyathistipula	Ficeae	Moraceae	-	-	-	DQ455657	DQ455679	-	-	-
Ficus crocata	Ficeae	Moraceae	JQ592827	-	-	AY730080	AY730168	EF092343	DQ455618	-
Ficus craterostoma	Ficeae	Moraceae	JX572602	JX517933	-	AY730097	AY730186	EF092349	DQ455622	-
Ficus crassipes	Ficeae	Moraceae	-	-	-	AY730112	AY730201	EF538789	-	-
Ficus costaricana	Ficeae	Moraceae	GQ981737	GQ981993	-	EU091602	EU084435	AY967952	-	-
Ficus cordatula	Ficeae	Moraceae	-	-	-	EU091584	EU084421	EU087625	EU084307	-

Ficus copiosa	Ficeae	Moraceae	JF738564	-	AY289347	AF165382	EF092324	EF092395	-	EU084390
Ficus consociata	Ficeae	Moraceae	-	-	-	AY063558	AY063519	-	-	-
Ficus conocephalifolia	Ficeae	Moraceae	JF738589	-	-	AF165381	EU084486	-	-	-
Ficus congesta	Ficeae	Moraceae	JF739104	-	-	AY730136	AY730225	DQ367620	-	DQ367640
Ficus condensa	Ficeae	Moraceae	-	-	-	AY063577	AY063538	-	EU084325	
Ficus concinna	Ficeae	Moraceae	-	-	-	AY730059	AY730145	EF092328	-	-
Ficus colubrinae	Ficeae	Moraceae	JQ592823	-	-	EU081764	-	EU089848	-	-
Ficus citrifolia	Ficeae	Moraceae	KF724292	GQ981992	-	AY730077	AY730165	AY967955	DQ455615	-
Ficus chrysolepis	Ficeae	Moraceae	-	-	-	EU091583	EU084420	EU087624	-	-
Ficus chapaensis	Ficeae	Moraceae	-	-	-	EU091638	EU084465	EU087671	-	-
Ficus cf.erythrosperma	Ficeae	Moraceae	JF738452	-	-	DQ457093	-	DQ457092	-	-
Ficus caulocarpa	Ficeae	Moraceae	JQ773663	JQ773517	-	EU091573	EU084413	EU087619	-	-
Ficus carica	Ficeae	Moraceae	KM360784	AY257530	-	EU091637	EU084464	-	-	EU084382
Ficus callophylla	Ficeae	Moraceae	-	-	-	EU091582	-	-	-	-
Ficus burtt-davyi	Ficeae	Moraceae	-	JX517875	-	DQ455647	DQ455675	EU087643	-	-
Ficus burkei	Ficeae	Moraceae	-	-	-	AY730095	AY730184	-	DQ455621	EU084369
Ficus bullenei	Ficeae	Moraceae	GQ981735	GQ981991	-	EU081758	-	AY967985	-	-
Ficus bubu	Ficeae	Moraceae	-	-	-	DQ455637	DQ455671	EU087642	DQ455628	-
Ficus brachypoda	Ficeae	Moraceae	-	-	-	EF545652	EF538768	EF538788	-	-
Ficus botryoides	Ficeae	Moraceae	-	-	-	AF165380	-	-	-	-
Ficus botryocarpa	Ficeae	Moraceae	-	-	-	AF165379	EU084452	DQ367619	-	DQ367639
Ficus bizanae	Ficeae	Moraceae	JX572600	JX518182	-	DQ455636	DQ455670	-	-	-
Ficus bernaysii	Ficeae	Moraceae	JF738935	GQ248128	-	AF165378	-	DQ367618	-	DQ367638
Ficus benjamina	Ficeae	Moraceae	AF500350	JQ773508	AF500377	AY063559	AY063520	EF092333	EU084305	EU084364
Ficus benghalensis	Ficeae	Moraceae	JX856703	GU935034	-	AY730065	AY730153	-	-	-
Ficus bauerlenii	Ficeae	Moraceae	-	-	-	AF165377	EU084474	-	-	-
Ficus auriculata	Ficeae	Moraceae	JQ773646	JQ773629	-	AF165376	FJ812281	EU087653	-	-
Ficus aurea	Ficeae	Moraceae	KJ773509	KJ772786	-	EU091598	EU084431	EU087636	-	-
Ficus aurata	Ficeae	Moraceae	-	-	-	EU091642	EU084469	-	-	-
Ficus asperifolia	Ficeae	Moraceae	-	-	-	EU091661	EU084484	EF092394	-	-
Ficus arfakensis	Ficeae	Moraceae	JF738872	-	-	DQ367657	EU084451	DQ367617	-	DQ367637

Ficus arbuscula	Ficeae	Moraceae	-	-	-	EU091617	-	-	-	EU084375
Ficus annulata	Ficeae	Moraceae	-	-	-	EU091578	EU084417	EU087622	-	-
Ficus andicola	Ficeae	Moraceae	-	-	-	AY730071	AY730159	EF092340	-	-
Ficus ampelas	Ficeae	Moraceae	JF941521	JQ773505	-	EU091659	-	-	-	-
Ficus americana	Ficeae	Moraceae	-	KJ012603	-	AY730070	AY730158	EF092339	DQ455613	-
Ficus altissima	Ficeae	Moraceae	JQ773645	JF953727	-	AY730064	AY730152	EU087621	-	EU084363
Ficus adhatodifolia	Ficeae	Moraceae	-	-	-	EU091563	EU084404	EU087608	-	-
Ficus adenosperma	Ficeae	Moraceae	KM895621	KM894523	-	AF165374	EF092321	EF092374	-	-
Ficus adelpha	Ficeae	Moraceae	-	-	-	DQ367656	EU084450	DQ367615	-	-
Ficus abutilifolia	Ficeae	Moraceae	EU213477	EU214248	-	AY730091	AY730180	EF092348	-	-
Sarcomphalus obtusifolius		Rhamnaceae	-	AY935939	AY968519	-	-	-	-	-
Rhamnus utilis		Rhamnaceae	JF317492	JF317432	JF317452	-	-	-	-	-
Rhamnus lycioides		Rhamnaceae	AJ390070	-	-	-	-	-	-	-
Rhamnus cathartica		Rhamnaceae	KM360955	AY257533	DQ851549	-	-	-	-	-
Paliurus spina-christi		Rhamnaceae	AJ390051	-	KP299601	-	-	-	-	-
Hovenia dulcis		Rhamnaceae	AJ390039	JX495724	KP299599	-	-	-	-	-
Gouania mauritiana		Rhamnaceae	JF317487	JF317427	JF317447	-	-	-	-	-
Ceanothus sanguineus		Rhamnaceae	U06795	AF049815	U78897	-	-	-	-	-
Ceanothus pumilus		Rhamnaceae	U78905	AF049841	U78902	-	-	-	-	-
Spiraea cantoniensis		Rosaceae	-	AF288127	DQ851556	-	-	-	-	-
Rosa blanda		Rosaceae	-	AB011985	DQ851551	-	-	-	-	-
Prunus persica		Rosaceae	HQ336405	HQ336405	HQ336405	-	-	-	-	-
Dryas octopetala		Rosaceae	JF317483	JF317424	JF317443	-	-	-	-	-
Zelkova serrata		Ulmaceae	AF500338	KP089378	EU002273	-	-	-	-	-
Zelkova schneideriana		Ulmaceae	KP768922	AF345328	-	-	-	-	-	-
Ulmus davidiana		Ulmaceae	KC539704	KC539635	KC539670	-	-	-	-	-
Ampelocera hottlei		Ulmaceae	AF500335	-	AF500364	-	-	-	-	-
Urtica dioica		Urticaceae	AF500361	GU266610	-	-	-	-	-	-
Pilea sinofasciata		Urticaceae	KF138224	KF138047	-	-	-	-	-	-
Pilea melastomoides		Urticaceae	KF138218	KF138041	-	-	-	-	-	-
Pilea insolens		Urticaceae	KF138215	KF138039	-	-	-	-	-	-

Pilea cadierei	Urticaceae	JF317491	JF317431	JF317451	-	-	-	-	-
Lecanthus peduncularis	Urticaceae	KF138186	KF138016	-	-	-	-	-	-
Debregeasia longifolia	Urticaceae	KF138141	KF137974	AY289252	-	-	-	-	-
Coussapoa villosa	Urticaceae	-	-	AY289261	-	-	-	-	-
Coussapoa schottii	Urticaceae	-	-	AY289260	-	-	-	-	-
Coussapoa panamensis	Urticaceae	-	-	AY289258	-	-	-	-	-
Coussapoa nymphaeifolia	Urticaceae	-	-	AY289259	-	-	-	-	-
Coussapoa latifolia	Urticaceae	-	-	AY289257	-	-	-	-	-
Cecropia peltata	Urticaceae	JQ594320	JQ589392	AY289265	-	-	-	-	-
Cecropia palmata	Urticaceae	AF061196	GU135054	AY289262	-	-	-	-	-
Cecropia obtusifolia	Urticaceae	KF138134	GQ981958	AY289263	-	-	-	-	-
Cecropia insignis	Urticaceae	GQ981692	JQ589383	AY289264	-	-	-	-	-
Boehmeria umbrosa	Urticaceae	KF138130	KF137965	-	-	-	-	-	-
Boehmeria spicata	Urticaceae	KF138127	KF137962	-	-	-	-	-	-
Boehmeria densiflora	Urticaceae	KF138114	KF137951	-	-	-	-	-	-

Table S2. Detailed information for the fossil calibrations used in this study.

MRCA	Family	Taxon	Fossil	Locality	Reference (description)	Reference (node assignment)	Oldest fossil age	Time set (Ma)	Age reference
CG Cannabaceae	Cannabaceae	*Aphananthe cretacea* Knoblock & Mai	fruits (endocarp)	Walbeck, Germany	Knobloch and Mai, 1986	Friis et al., 2011	Maastrichtian	66.0	Magallón et al., 2015
SG *Humulus*	Cannabaceae	*Humulus rolundatus* Dorofeev	fruits (endocarp)	Isakovka, USSR	Takhtajan, 1982	Collinson, 1989	Eocene and Oligocene boundary	33.9	Collinson, 1989
SG *Ficus*	Moraceae	*Ficus lucidus* Chandler	achenes	Southern England	Chandler, 1962	Collinson, 1989	Paleocene and Eocene boundary	56.0	Collinson, 1989
SG *Broussonetia*	Moraceae	*Broussonetia rugosa* Chandler	fruits (endocarp)	Southern England	Chandler, 1961	Chandler, 1961; Collinson, 1989	Eocene and Oligocene boundary	33.9	Collinson, 1989
SG *Morus*	Moraceae	*Morus tymensis* Dorofeev	fruits	USSR	Takhtajan, 1982	Collinson, 1989	Eocene and Oligocene boundary	33.9	Collinson, 1989

SG *Artocarpus* sl	Moraceae	*Artocarpoxylon deccanensis* Mehrotra, Prakash, and Bande	wood	India	Mehrotra et al., 1984	Williams et al., 2017	64-67 Ma	64.0	Hooper et al., 2010, Selena Smith, pers com
SG Rhamnaceae	Rhamnaceae	*Coahuilanthus belindae* Calvillo-Canadell & Cevallos-Ferriz	flowers	Coahuila, Mexico	Calvillo-Canadell and Cevallos-Ferriz, 2007	Calvillo-Canadell and Cevallos-Ferriz, 2007	Late Campanian	72.1	Calvillo-Canadell and Cevallos-Ferriz, 2007
CG Rhamnaceae	Rhamnaceae	*Paliurus clarnensis* Burge & Manchester	fruits	Red Gap, Oregon, USA	Burge and Manchester, 2008	Burge and Manchester, 2008	44 Ma (middle Eocene)	44.0	Burge and Manchester, 2008
CG Rosaceae	Rosaceae	*Prunus cathybrownae* Benedict, DeVore & Pigg	flowers+ fruits	Boot Hill locality, Washington, USA	Benedict et al., 2011	Benedict et al., 2011	49.42 ± 0.54 Ma	49.42	Benedict et al., 2011
SG Ulmaceae	Ulmaceae	*Ulmites ulmifolius* (Schloemer-Jäger) Kvacek	leaves	Spitsbergen, Norway	Kvacek et al., 1994	Manchester, 1999	Paleocene	56.0	Manchester, 1999
SG *Boehmeria*	Urticaceae	*Boehmeria sibirica* Dorofeev	achenes	Kireevskoe, USSR	Collinson, 1989	Collinson, 1989	Oligocene and Miocene boundary	23.03	Collinson, 1989
SG *Pilea*	Urticaceae	*Pilea lithuanica* Dorofeev	fruits	USSR	Collinson, 1989	Collinson, 1989	Oligocene and Miocene boundary	23.03	Collinson, 1989

REFERENCES

Benedict JC, DeVore ML, Pigg KB. **2011**. *Prunus* and *Oemleria* (Rosaceae) flowers from the Late Early Eocene Republic Flora of Northeastern Washington State, U.S.A. *International Journal of Plant Sciences* **172**: 948–958.

Burge DO, Manchester SR. **2008**. Fruit morphology, fossil history, and biogeography of *Paliurus* (Rhamnaceae). *International Journal of Plant Sciences* **169**: 1066–1085.

Calvillo-Canadell L, Cevallos-Ferriz SRS. **2007**. Reproductive structures of Rhamnaceae from the Cerro del Pueblo (Late Cretaceous, Coahuila) and Coatzingo (Oligocene, Puebla) Formations, Mexico. *American Journal of Botany* **94**: 1658–1669.

Chandler MEJ. **1961**. Flora of the Lower Headon beds of Hampshire and the Isle of Wight. *Bulletin of the British Museum (Natural History), Geology* **5**: 91–151.

Chandler MEJ. **1962**. *The Lower Tertiary floras of Southern England. II. Flora of the Pipe-Clay Series of Dorset (Lower Bagshot)*. London: British Museum (Natural History).

Collinson ME. **1989**. The fossil history of the Moraceae, Urticaceae (including Cecropiaceae), and Cannabaceae. *Crane, P, R,, Blackmore, S. (Ed s.), Evolution, systematics, and fossil history of the Hamamelidae* **2**: 319–339.

Friis EM, Crane PR, Pedersen KR. **2011**. *Early flowers and angiosperm evolution*. Cambridge: Cambridge University Press.

Hooper P, Widdowson M, Kelley S. **2010**. Tectonic setting and timing of the final Deccan flood basalt eruptions. *Geology* **38**: 839–842.

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

Kvacek Z, Manum SB, Boulter MC. **1994**. Angiosperms from the Palaeogene of Spitsbergen, including an unfinished work bv A. G. Nathorsl. *Palacontographica Abt. B* **232**: 103–128.

Magallón S, Gomez-Acevedo S, Sanchez-Reyes LL, Hernandez-Hernandez T. **2015**. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist* **207**: 437–453.

Manchester SR. **1999**. Biogeographical relationships of North American tertiary floras. *Annals of the Missouri Botanical Garden* **86**: 472–522.
Mehrotra RC, Prakash U, Bande MB. **1984**. Fossil woods of Lophopetalum and Artocarpus from the Deccan Intertrappean Beds of Mandla district, Madhya Pradesh, India. *Palaeobotanist* **32**: 310–320.
Williams EW, Gardner EM, Harris R, Chaveerach A, Pereira JT, Zerega NJC. **2017**. Out of Borneo: biogeography, phylogeny and divergence date estimates of *Artocarpus* (Moraceae). *Annals of Botany* **119**: 611–627.

Table S4. Divergence time estimates for tribes of Moraceae. Here we list the results of three different methods (see text): estimation with older root maximum boundary (old_root); estimation without calibration of stem-group *Ficus* (excl_*Ficus*).

	No. of species in the clade	**Old_root mean (95% HPD[1])**	**Excl_*Ficus* mean (95% HPD)**
Moraceae	272	81.1 (74.3-88.0)	78.8 (72.4-84.6)
Artocarpeae	11	65.8 (64.0-69.1)	65.6 (64.0-68.7)
Moreae[2]	13	40.3 (35.1-47.5)	40.2 (35.1-47.0)
Maclureae	3	26.1 (9.9-44.5)	24.5 (9.7-42.0)
Dorstenieae	18	61.9 (53.1-72.3)	59.9 (50.0-69.3)
Castilleae	26	32.7 (18.6-72.3)	30.8 (18.15-46.1)
Ficeae (*Ficus*)	200	49.7 (41.2-57.6)	47.8 (39.6-56.9)

[1]95% HPD: 95% highest posterior density; [2]*Streblus smithii* was here excluded from crown-group Moreae (see text for details).

Table S5. Summary results of ancestral state reconstruction for the complete (320 species) dataset by Bayesian approach with fixed model (equal rate or unequal rate).

	One MCC tree				**3600 posterior trees**			
Equal rate								
	CG Moraceae[1]	95% HPD[2]	CG *Ficus*[3]	95%HPD	CG Moraceae	95% HPD	CG *Ficus*	95%HPD
A	Dioecy	0.87-0.94	Monoecy	0.89- 0.95	Dioecy	0.81-0.99	Monoecy	0.95-0.99

B	Dioecy	0.85-0.93	Monoecy	0.86- 0.94	Dioecy	0.80-0.99	Monoecy	0.93-1.00
C	Gynodioecy + dioecy	0.92-0.97	Monoecy	0.80- 0.91	Gynodioecy + dioecy	0.79-1.00	Monoecy	0.65-1.00
D	Dioecy	0.81-0.92	Monoecy	0.83- 0.93	Dioecy	0.75-0.99	Monoecy	0.88-1.00
E	Gynodioecy + dioecy	0.86-0.96	Monoecy	0.71- 0.88	Gynodioecy + dioecy	0.72-0.99	Monoecy	0.54-1.00
F	Gender dimorphic	0.83-0.93	Gender monomorphic	0.66- 0.82	Gender dimorphic	0.71-0.98	Gender monomorphic	0.52-0.99
Unequal rate								
A	/	/	/	/	/	/	/	/
B	Dioecy	0.37-0.54	Dioecy	0.10- 1.00	Dioecy	0.70-1.00	Monoecy	0.09-0.51
C	Gynodioecy + dioecy	0.71-1.00	Gynodioecy + dioecy	0.11- 0.98	Gynodioecy + dioecy	0.52-1.00	Monoecy	0.04-1.00
D	Dioecy	0.49-0.50	Gynodioecy	0.20- 0.97	Dioecy	0.78-1.00	Monoecy	0.08-1.00
E	Gynodioecy + dioecy	0.73-1.00	Gynodioecy + dioecy	0.11-0.99	Gynodioecy + dioecy	0.58-1.00	Monoecy	0.04-1.00
F	Gender dimorphic	0.61-1.00	Gender dimorphic	0.12-1.00	Gender dimorphic	0.58-1.00	Gender dimorphic	0.04-1.00

[1]CG Moraceae: breeding system state of crown-group Moraceae; [2]95% HPD: 95% highest posterior density; [3]CG *Ficus*: breeding system state of crown-group *Ficus*.

Supplementary data for Chapter II. long-distance dispersal shaped the diversity of tribe Dorstenieae (Moraceae)

Table S1. List of specimens collected in this study.

taxa	tribe	section in Dorstenia	collection No.	No. reads	targeted reads%	gene retrieved	col date	Voucher Number
Artocarpus heterophyllus	Artocarpeae			1,072,716	57.88%	97	15/05/2014	E. Gardner 98 (SAN)
Antiaropsis descipiens	Castilleae			1,151,723	5.37%	99		N. Zerega 281 (NY)
Allaeanthus greveana	Dorsteniae		/	3,425,725	47.68%	101		O. Pascal and F. Hallé 677 (NY)
Allaeanthus kurzii	Dorsteniae		/	913,251	29.55%	100		Al Gentry and Chawalit Niyomdham 66546 (MO)
Allaeanthus luzonicus	Dorsteniae			76,606	71.24%	98		Kuo-fang Chung 2016 (HAST)
Bleekrodea madagascariensis	Dorsteniae		P06862941	10,190,180	42.29%	102	24/11/2009	MYA 375(P)
Bosqueiopsis gilletii	Dorsteniae		P06822851	8,006,493	21.83%	102	15/11/2009	J.R. Timberlake, T. Müller & F. Crawford 5767(P)
Brosimum acutifolium	Dorsteniae			377,383	26.38%	102		T. Lileen 4457
Brosimum alicastrum	Dorsteniae			235,089	64.90%	102	7/07/2013	E. Gardner 23 (CHIC)
Brosimum costaricanum	Dorsteniae			584,789	50.02%	102	31/01/1988	C. Kernan 27 (F)
Brosimum gaudichaudii	Dorsteniae			62,019	58.54%	81		S. Mendes & al. 332 (F)
Brosimum guianense	Dorsteniae			313,169	37.28%	102		J.F. Morales 5258
Brosimum multinervium	Dorsteniae			269,526	40.27%	101		Jaramillo 574
Brosimum parinarioides	Dorsteniae			360,342	43.65%	102	21/01/1983	Vasquez and Jaramillo 3817 (F)
Brosimum potabile	Dorsteniae			466,812	30.89%	102	18/01/1989	Maciel & Rosario 1552
Brosimum rubescens	Dorsteniae			624,585	42.95%	102		McPherson 20944 (F)
Brosimum utile	Dorsteniae			159,850	67.41%	100		C. Galdames 4403 (F)
Broussonetia papyrifera	Dorsteniae			51,643,997	0.90%	89		GenBank SRA reads: accession no. SRR1477753
Dorstenia africana	Dorsteniae	Nothodorstenia	/	4,238,871	34.22%	102	27/05/1993	Carvalho 5308(NY)
Dorstenia alta	Dorsteniae	Xylodorstenia	MO2506416	2,711,247	3.60%	87	24/01/2001	A Ntemi Sallu 638(MO)

Dorstenia angusticornis	Dorsteniae	Xylodorstenia	MO2506415	9,459,437	42.64%	102	08/06/1995	BO Daramola 665(MO)
Dorstenia appendiculata	Dorsteniae	Lecanium	NY 565961	11,461,715	0.34%	57	25/01/1996	Andre MA Amorim et al. 1941(NY)
Dorstenia arifolia	Dorsteniae	Dorstenia	/	10,830,558	33.04%	102	04/04/2002	AM de Carvalho 7169(NY)
Dorstenia arifolia-1	Dorsteniae	Dorstenia	V0202318F	2,191,707	19.26%	98	04/03/1980	J.P.P. Carauta 3423(F)
Dorstenia arifolia-2	Dorsteniae	Dorstenia	V0202319F	1,350,171	23.70%	66	12/04/1987	Amorim, A.M. 7(F)
Dorstenia astyanactis	Dorsteniae	Kosaria	P06822779	10,658,870	9.95%	102	23/06/2010	Haba, P.K. 677(P)
Dorstenia bahiensis	Dorsteniae	Dorstenia	P00047587	14,590,733	17.61%	102	11/07/1993	H174(P)
Dorstenia barnimiana	Dorsteniae	Acauloma	MO2506414	9,790,293	19.93%	84	1981 spring	J Lavranos & MB Bleck 19480(MO)
Dorstenia benguellensis	Dorsteniae	Kosaria	MO2506422	4,419,756	38.60%	101	05/12/1998	AS Mkeya et al 1109(MO)
Dorstenia bonijesu	Dorsteniae	Dorstenia	P00840630	4,914,477	10.77%	102	10/05/1981	V.F. Ferreirov 1729(P)
Dorstenia bowmaniana	Dorsteniae	Lecanium	V0202306F	8,048,950	27.98%	102	22/01/1971	P. Carauta 1277(F)
Dorstenia brasiliensis	Dorsteniae	Emygdioa	/	3,508,861	21.49%	89	13/07/1995	MJ Jansen-Jacobs et al 4436(NY)
Dorstenia brasiliensis-1	Dorsteniae	Emygdioa	V0228352F	5,416,062	28.00%	102	24/11/1974	L.C. Gurken 37(F)
Dorstenia brasiliensis-3	Dorsteniae	Emygdioa	V0202329F	4,612,329	1.99%	102	11/02/1993	G. Hatschbac 58895(F)
Dorstenia brownii	Dorsteniae	Lomathophora	MO2506420	3,049,365	5.53%	92	10/11/2004	MA Mwangoka et al 3606(MO)
Dorstenia buchananii	Dorsteniae	Kosaria	/	833,457	41.67%	37	04/12/1954	NC Chase 5466(F)
Dorstenia caimitensis	Dorsteniae	Emygdioa	/	2,090,362	2.24%	99	23/08/1927	WJ Eyerdam(F)
Dorstenia carautae	Dorsteniae	Lecanium	MO1491375	2,804,160	24.69%	102	05/01/1990	JM Silva 782(MO)
Dorstenia cayapia subsp. cayapia	Dorsteniae	Emygdioa	NY 585795	20,545,986	31.03%	102	27/02/1999	Andre Ma Carvalho et al 6710(NY)
Dorstenia cayapia subsp. vitifolia	Dorsteniae	Emygdioa	/	2,952,911	32.63%	101	09/10/1990	Roberto Fontes Vieira et al 554(NY)
Dorstenia cayapia subsp. vitifolia-1	Dorsteniae	Emygdioa	V0202308F	1,860,888	38.60%	68	19/04/1977	A. Krapovickas & A. Schinini 31524(F)
Dorstenia ciliata	Dorsteniae	Lomathophora	P06819570	1,219,024	13.59%	69	16/01/2000	X.M. van der Burgt 576(P)
Dorstenia contrajerva	Dorsteniae	Dorstenia	/	3,208,913	3.16%	102	24/05/1993	Ulises Chavarria 809(F)
Dorstenia contrajerva-1	Dorsteniae	Dorstenia	/	5,296,893	42.67%	102	16/10/1997	E. Martinez S. et al 28913(NY)
Dorstenia djettii	Dorsteniae	Nothodorstenia	MO2506427	4,776,155	18.56%	102	12/03/1996	J Amponsah et al 1413(MO)

Dorstenia drakeana	Dorsteniae	Dorstenia	/	6,289,082	45.96%	102	09/10/1987	SD Koch et al 87195(NY)
Dorstenia elliptica	Dorsteniae	Nothodorstenia	MO2506400	1,984,635	5.33%	95	28/01/1997	J. Nning 20(MO)
Dorstenia erythrandra	Dorsteniae	Emygdioa	/	14,844,492	22.60%	102	1927August	WJ Eyerdam 298(F)
Dorstenia excentrica	Dorsteniae	Emygdioa	/	10,039,982	5.70%	102	25/10/1981	M Nee 22374(F)
Dorstenia excentrica-1	Dorsteniae	Emygdioa	/	7,735,069	46.05%	102	10/08/1991	Liborio Lopez 6(F)
Dorstenia fawcetti	Dorsteniae	Emygdioa	/	5,545,892	14.55%	102	31/01/1980	CC Berg 990(NY)
Dorstenia flagellifera	Dorsteniae	Emygdioa	/	10,711,580	39.13%	101	1927 July	W.J. Eyerdam 196(F)
Dorstenia foetida	Dorsteniae	Kosaria	/	13,237,026	1.32%	102	25/05/1963	William Burge 2844(F)
Dorstenia goetzei	Dorsteniae	Kosaria	/	4,806,838	17.79%	98	06/08/1997	PB Phillipson 4801(MO)
Dorstenia hildebrandtii var. schlechteri	Dorsteniae	Kosaria	MO2506428	5,145,977	5.19%	100	12/04/2000	L Festo 648(MO)
Dorstenia hirta	Dorsteniae	Lecanium	NY 777756	4,311,869	35.57%	101	17/07/2000	FO Souza et al 21(NY)
Dorstenia holstii	Dorsteniae	Lomathophora	/	8,494,898	11.47%	95	08/12/2000	A Ntemi Sallu 645(MO)
Dorstenia indica	Dorsteniae	Kosaria	/	2,795,944	23.85%	95	15/12/1975	L Bernardi 16022(MO)
Dorstenia kameruniana	Dorsteniae	Nothodorstenia	/	15,998,183	9.27%	102	21/08/2003	OA Kibure & H Bofu 1045(MO)
Dorstenia lindeniana	Dorsteniae	Lecanium	/	7,149,762	37.24%	102	26/04/1966	Gayle C Jone et al 3130(F)
Dorstenia lindeniana-1	Dorsteniae	Lecanium	MO959107	4,738,696	16.39%	102	28/11/2002	D Alvarez 2721(MO)
Dorstenia mannii	Dorsteniae	Lomathophora	P06822773	3,093,559	2.12%	62	07/12/1999	Simons, E.L.A.N. & R. Westerduijn 300(P)
Dorstenia milaneziana	Dorsteniae	Lecanium	NY 95156	7,146,035	23.36%	102	17/03/1996	WW Thomas, A Amorin, J Jardim 11078(NY)
Dorstenia oligogyna	Dorsteniae	Nothodorstenia	P06822758	6,466,687	12.44%	102	31/10/1994	J.J. Wieringa 3005(P)
Dorstenia peruviana	Dorsteniae	Lecanium	MO1408784	2,200,476	23.38%	97	12/04/2002	A Fuentes 4353(MO)
Dorstenia petraea	Dorsteniae	Emygdioa	/	3,210,633	40.91%	102	16/08/1951	Grady L Webster 4073(NY)
Dorstenia picta	Dorsteniae	Lecanium	/	6,315,511	2.67%	94	05/03/1993	RE Gereau et al. 5189(MO)
Dorstenia poinsettiifolia	Dorsteniae	Lomathophora	P06822789	3,057,480	3.24%	74	08/12/1999	Simons, E.L.A.N. & R. Westerduijn 319(P)
Dorstenia psilurus var. scabra	Dorsteniae	Lomathophora	V0096384F	3,595,847	26.95%	79	between 1893 and 1911	G.A. Zenker s.n.
Dorstenia psilurus-1	Dorsteniae	Lomathophora	/	2,712,056	7.00%	86	27/04/1999	Tim Flynn et al 6522(NY)

Dorstenia ramosa subsp. dolichocaula	Dorsteniae	Dorstenia	V0181401F	6,955,051	0.93%	84	s.d.	M.D.M. Vianna Filho 2020(F)
Dorstenia roigii	Dorsteniae	Emygdioa	/	331,405	35.23%	30	07/11/1923	Ekman 17973(F)
Dorstenia soerensenii	Dorsteniae	Lomathophora	MO2506403	2,189,329	13.48%	92	21/12/1965	WJJO de Wilde 9392(MO)
Dorstenia tayloriana var. tayloriana	Dorsteniae	Lomathophora	/	8,308,220	29.65%	94	28/01/2002	MA Mwangoka & A Kalage 2673(MO)
Dorstenia tenera	Dorsteniae	Lomathophora	P06822827	3,333,699	4.96%	97	14/04/1990	F.J. Breteler 9984(P)
Dorstenia tenuis	Dorsteniae	Emygdioa	V0228351F	1,542,631	21.87%	102	14/12/1984	S.G. Tressens, E. Cabral, S. Cáceres, M. Urbani & C. Zamudio 2911(F)
Dorstenia turbinata	Dorsteniae	Xylodorstenia	MO2506406	3,233,705	4.00%	96	14/10/2002	M Cheek 11086(MO)
Dorstenia umbricola	Dorsteniae	Lecanium	NY196401	4,586,273	21.23%	84	24/02/1989	M. Alexiades, V. Pesha(NY)
Dorstenia urceolata	Dorsteniae	Lecanium	V0228354F	8,823,436	20.96%	102	16/10/1977	P.J.M. Maas & P. Carauta 3233(F)
Dorstenia variifolia	Dorsteniae	Lomathophora	MO2506425	1,620,309	32.25%	47	10/05/1987	J&J Lovett et al. 2135(MO)
Dorstenia warneckei	Dorsteniae	Kosaria	P06819719	2,582,228	1.17%	50	08/09/1984	D.W. Thomas 3660(P)
Dorstenia yambuyaensis	Dorsteniae	Lomathophora	P06819718	1,644,628	4.22%	57	22/09/1983	Mikio Kaji 24(P)
Dorstenia zanzibarica	Dorsteniae	Kosaria	MO2506423	1,249,345	19.54%	73	04/08/1986	J Lovett 895(MO)
Fatoua villosa	Dorsteniae			13,261,729	38.91%	100	10/07/2013	E. Gardner 27 (CHIC)
Helianthostylis sprucei	Dorsteniae			201,836	39.65%	101		F16142
Malaisia scandens	Dorsteniae			499,782	64.65%	102	25/05/2014	E. Gardner 122 (SAN)
Scyphosyce manniana	Dorsteniae		P06825005	531,584	17.92%	84	13/01/1987	H.F. 1220(P)
Scyphosyce pandurata	Dorsteniae		P06824977	10,995,555	29.27%	102	1987	D.W. Thomas 6869(P)
Sloetia elongata	Dorsteniae		P06889539	6,790,711	37.13%	102	30/12/1992	P. Thomas & L.E. Teo no number(P)
Treculia acuminata	Dorsteniae		P06763325	210,985	18.89%	33	21/09/1983	J.J. Floret, A.M. Louis 1351(P)
Treculia africana	Dorsteniae		/	13,473,575	38.00%	102	26/06/2013	N. Zerega 909 (SAN)
Treculia obovoidea	Dorsteniae		P06877157	10,640,372	27.93%	102	04/10/2009	Cheek, M. 15081(P)
Trilepisium madagascariense	Dorsteniae		P01060143	536,683	13.89%	39	16/09/2013	Z. Ranktonirina & Georges Be 197(P)
Trymatococcus oligandrus	Dorsteniae			319,596	37.93%	101		W. Hahn 3649
Utsetela gabonensis	Dorsteniae		P01037829	5,988,593	22.98%	102	27/09/1997	F.J. Breteler 14096(P)

Ficus macrophylla	Ficeae			2,126,710	64.98%	101	11/07/2013	E. Gardner 30 (CHIC)
Milicia excelsa	Moreae			2,721,283	83.73%	98		McPherson 16087 (US)
Parartocarpus venenosus	Parartocarpeae			92,533	28.25%	77		N. Zerega 874 (SAN)
Dorstenia aristeguietae	Dorsteniae	Lecanium	/	203,410	2.26%	excluded by HybPiper	19/10/1966	J. Steyermark 97521
Dorstenia barteri var. varteri	Dorsteniae	Lomathophora	P01037801	1,416,397	1.24%	12	17/04/2004	Tchiengue, B. 1946(P)
Dorstenia brasiliensis-2	Dorsteniae	Emygdioa	V0181423F	445,787	4.51%	15	05/05/1962	E.P. Heringer 8937(F)
Dorstenia choconiana	Dorsteniae	Lecanium	MO226270	2,125,452	0.09%	excluded by HybPiper	22/11/2000	R Rueda et al 15033(MO)
Dorstenia cuspidata var. humblotiana	Dorsteniae	Kosaria	/	1,019,056	33.58%	16	18/11/2001	PB Phillipson 5361(MO)
Dorstenia dinklagei	Dorsteniae	Lomathophora	P06778387	2,044,739	33.50%	28	15/01/1987	H.F. 1228(P)
Dorstenia dorstenioides	Dorsteniae	Xylodorstenia	MO2506410	2,716,518	0.73%	6	11/03/1976	JJFE de Wilde 8039(MO)
Dorstenia elata	Dorsteniae	Lecanium	/	1,737,743	4.74%	2	13/02/1999	R. Mello-Silva et al 1563(NY)
Dorstenia ellenbeckiana	Dorsteniae	Acauloma	MO2506411	6,536,034	72.40%	1	14/04/1974	JW Ash 2402(MO)
Dorstenia lujae	Dorsteniae	Lomathophora	P06822759	660,274	14.78%	16	09/11/1988	L.J.G. van der Maesen 5407(P)
Dorstenia nummularia	Dorsteniae	Emygdioa	/	22,093	16.07%	1	1945 August	R.A. Howard 6452(F)
Dorstenia prorepens	Dorsteniae	Lomathophora	P06777762	29,651	17.22%	excluded by HybPiper	09/04/1984	D. Thomas 3423(P)
Dorstenia ramosa subsp. dolichocaula-1	Dorsteniae	Dorstenia	V0228346F	7,151,625	33.73%	3	04/12/1960	A.P. Duarte 5839(F)
Dorstenia ramosa subsp. ramosa	Dorsteniae	Dorstenia	V0228348F	614,296	55.96%	7	25/12/1974	D. Sucre 6380(F)
Dorstenia scaphigera	Dorsteniae	Xylodorstenia	P06777132	93,849	23.63%	0	23/02/1989	Jean Louis 13797(P)
Dorstenia schliebenii	Dorsteniae	Lomathophora	/	123,855	33.02%	0	26/09/1932	HJ Schlieben 2723(F)
Dorstenia zenkeri	Dorsteniae	Lomathophora	/	304,766	26.28%	6	1902 May	J Zenker(F)
Fatoua madagascariensis	Dorsteniae		P00108207	306,061	24.44%	14	1970 March	P.Morat 3498(P)

Table S2. AIC of biogeographic reconstruction with time-constant (model0, a) or time-stratified (model1, b) DEC, DIVA-like and BayArea-like based models.

a)

model	No. parameters	LnL	AIC	d	e	j
BayArea+Jm0	3	-84.47	174.94	0.0006	0	0.0137
DEC+Jm0	3	-91.15	188.3	0.0009	0	0.0107
DIVA+Jm0	3	-92.33	190.66	0.0011	0	0.0097
DIVAm0	2	-97.65	199.3	0.0017	0	/
DECm0	2	-98.62	201.24	0.0014	0	/
BayAream0	2	-115.46	234.92	0.0012	0.0072	/

b)

model	No. parameters	LnL	AIC	d	e	j
BayArea+Jm1	3	-82.79	171.58	0.0029	0	0.0679
DIVA+Jm1	3	-83.43	172.86	0.0054	0	0.0485
DEC+Jm1	3	-84.46	174.92	0.0045	0	0.0277
DIVAm1	2	-89.1	182.2	0.0083	0.0004	/
DECm1	2	-89.98	183.96	0.0071	0.0006	/
BayAream1	2	-110.37	224.74	0.0063	0.0072	/

d: dispersal rate

e: extinction rate

j: the rate of founder-effect speciation process

Table S3. List of ancestral distribution area estimated for several nodes by BioGeoBEARS. a) time-constant models (model0); b) time-stratified models (model1)

a)

model	**CG Dorstenieae**	**SG *Brosimum* s.l.**	**CG *Brosimum* s.l.**	**SG *Dorstenia***	**CG *Dorstenia***	**SG Neo *Dorstenia***	**CG Neo *Dorstenia***	**Node I**
DEC+Jm0	CDF	C	A	C	C	C	A	C
DECm0	CDF	AC	A	C	C	AC	A	C
BayArea+Jm0	F	C	A	C	C	C	A	C
BayAream0	CF	C	A	C	C	C	A	C
DIVA+Jm0	F	AC	A	C	C	C	A	C
DIVAm0	F	AC	A	C	C	AC	A	C

b)

model	**CG Dorstenieae**	**SG *Brosimum* s.l.**	**CG *Brosimum* s.l.**	**SG *Dorstenia***	**CG *Dorstenia***	**SG Neo *Dorstenia***	**CG Neo *Dorstenia***	**Node I**
DEC+Jm1	CDF	AC	A	C	C	A	A	AC
DECm1	CDF	AC	A	C	C	AC	A	AC
BayArea+Jm1	F	C	A	C	C	C	A	C
BayAream1	CF	AC	A	C	C	C	A	C
DIVA+Jm1	F	AC	A	C	C	A	A	A
DIVAm1	F	AC	A	C	C	AC	A	AC

Codes for distribution areas were the same as in Figure 1 (A, South America; B, Central/North America; C, continental Africa; D, Madagascar; E. India and Sri Lanka; F, Southeast Asia and Oceania.)

Supplementary data for Chapter III. breeding system evoluion and climate in *Ficus*

Table S2. Estimates for interaction effects of climatic variables and breeding systems in Ficus from the regressions by generalized linear models (GLM) and generalized estimating equations (GEE) for three datasets: 1) full dataset (same as Table 1); 2) subsetAsia; and 3) bio18min.

var	GLM_M_FD	GLM_M_FD_coef	GEE_M_FD	GLM_M_Asia	GLM_M_Asia_coef	GEE_M_Asia	GLM_bio18min_FD	GLM_bio18min_FD_coef	GEE_bio18min_FD
PC1	***	-0.2242	not sig.	.	-0.13621	not sig.	*	-0.16617	not sig.
PC2	***	-0.29111	not sig.	*	-0.23273	not sig.	***	-0.33875	not sig.
PC1+2	***,***	-0.25431, -0.34284	not sig.	.,*	-0.14615,-0.24375	not sig.	***,***	-0.17919,-0.34749	not sig.
bio1 = Annual Mean Temperature	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio5 = Max Temperature of Warmest Month	**	-0.022318	not sig.	*	-0.020294	not sig.	**	-0.012796	not sig.
bio6 = Min Temperature of Coldest Month	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio8 = Mean Temperature of Wettest Quarter	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio9 = Mean Temperature of Driest Quarter	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio10 = Mean Temperature of Warmest Quarter	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio11 = Mean Temperature of Coldest Quarter	not sig.	/	not sig.	not sig.	/	not sig.	not sig.	/	not sig.
bio12 = Annual Precipitation	***	0.10051	not sig.	**	0.07439	not sig.	***	0.06578	not sig.
bio13 = Precipitation of Wettest Month	***	0.2675	not sig.	*	0.2823	not sig.	.	0.07163	not sig.
bio14 = Precipitation of Driest Month	***	0.23656	not sig.	*	0.11984	not sig.	***	0.23511	not sig.
bio16 = Precipitation of Wettest Quarter	***	0.17727	not sig.	*	0.15527	not sig.	*	0.05061	not sig.
bio17 = Precipitation of Driest Quarter	***	0.13509	not sig.	*	0.06915	not sig.	***	0.13624	not sig.
bio18 = Precipitation of Warmest Quarter	***	0.28233	not sig.	***	0.21337	not sig.	***	0.19395	not sig.
bio19 = Precipitation of Coldest Quarter	*	0.04934	not sig.	.	0.04571	not sig.	not sig.	/	not sig.

Significance codes: '***' P<0.0001; '**' P<0.001; '*' P<0.01; '.' P<0.05; 'not sig.' not significant.

GLM, generalized linear models; GEE, generalized estimating equations; M, use median for the tip value; FD, full dataset; Asia, subset dataset with Asia-Oceania *Ficus* only; bio18min, dataset assembeld from the corresponding values of each variables from coordinate of the minimum value of variable "bio18" of each species (see text).

université PARIS-SACLAY | ÉCOLE DOCTORALE Sciences du végétal : du gène à l'écosystème

Titre : Phylogénie, biogéographie et évolution des systèmes sexuels chez les Moraceae

Mots clés : *Ficus*, Dorstenieae, phylogénomique, datation moléculaire, reconstruction des états ancestraux, niche climatique

Résumé : Les Angiospermes sont le clade le plus diversifié des plantes actuelles et sont exceptionnellement riches en espèces dans les régions tropicales. Dans cette thèse, j'ai étudié l'évolution des systèmes sexuels et l'histoire biogéographique de la famille des Moraceae, clade modèle utilisé pour comprendre l'origine et l'évolution de la diversité chez les Angiospermes. Dans le Chapitre I, j'ai reconstruit et calibré un nouvel arbre phylogénétique daté pour les Moraceae. J'ai ensuite utilisé cet arbre pour reconstruire les états ancestraux des systèmes sexuels chez les Moraceae et *Ficus*. Les âges des groupes-couronne des Moraceae et du genre *Ficus* sont estimés au Crétacé et à l'Eocène, respectivement. La dioécie est inférée comme l'état ancestral des systèmes sexuels chez les Moraceae, avec plusieurs transitions ultérieures vers la monoécie, y compris chez *Ficus*. Ce résultat suggère que la dioécie ne représente pas nécessairement un cul-de-sac évolutif. Dans le Chapitre II, j'ai reconstruit un arbre phylogénétique daté pour la tribu des Dorstenieae, distribuée principalement dans les régions tropicales, à partir d'un nouveau jeu de données génomiques nucléaires produit avec une approche Hyb-Seq. L'histoire biogéographique du groupe a ensuite été reconstruite en utilisant les modèles de dispersion-extinction-cladogenèse. Les âges des groupes-couronne des Dorstenieae et du genre *Dorstenia* sont estimés au Crétacé et dans la période du Crétacé au Paléocène, respectivement. Deux évènements de dispersion à longue distance depuis l'Afrique continentale vers l'Amérique du Sud ont eu lieu au Cénozoïque (*Dorstenia* et *Brosimum* s.l.). Dans le Chapitre III, j'ai testé les différences de niche climatique (température et précipitation) entre les deux systèmes sexuels (monoécie et gynodioécie) chez *Ficus* avec un nouveau jeu de données fiables d'occurrences spatiales et de systèmes sexuels chez 183 espèces. À cette fin, j'ai utilisé deux approches comparatives : équations d'estimation généralisées (GEE) et modèles linéaires généralisés (GLM). Une relation positive entre précipitation et gynodioécie est soutenue par les analyses GLM, et aucune méthode ne soutient une relation entre température et système sexuel. Une meilleure capacité à se disperser et le potentiel d'autopollinisation sont deux explications possibles pour la colonisation et la survie des espèces monoïques dans des environnements plus secs. Cette thèse démontre le potentiel des méthodes phylogénétiques comparatives et des données phylogénomiques pour répondre aux questions d'évolution des systèmes sexuels et de biogéographie chez les Moraceae et ouvre plusieurs nouvelles perspectives importantes méritant d'être approfondies chez d'autres clades de plantes, telles que la relation entre système sexuel et niche climatique.

Title: Phylogeny, biogeography, and breeding system evolution in Moraceae

Keywords: *Ficus*, Dorstenieae, phylogenomics, molecular dating, ancestral state reconstruction, climatic niche

Abstract: Angiosperms are the most diversified clade of extant plants and are exceptionally species-rich in tropical regions. In this thesis, I investigated breeding system evolution and biogeographic history in the family Moraceae, which I used as a model clade to understand the origin and evolution of diversity of angiosperms. In Chapter I, I reconstructed and calibrated a new dated phylogenetic tree for Moraceae as a whole. I then used this tree to reconstruct ancestral states of breeding systems in Moraceae and *Ficus*. The crown group ages of Moraceae and *Ficus* were estimated in the Cretaceous and in the Eocene, respectively. Dioecy was inferred as the ancestral breeding systems of Moraceae, with several subsequent transitions to monoecy, including in *Ficus*. This result suggests that dioecy is not necessarily an evolutionary dead end. In Chapter II, I reconstructed a dated phylogenetic tree for tribe Dorstenieae, mainly distributed in tropical regions, with a new data set of nuclear genomic data generated with a Hyb-Seq approach. Biogeographic history was then reconstructed using dispersal-extinction-cladogenesis models. The crown group ages of Dorstenieae and *Dorstenia* were estimated in the Cretaceous and in the Cretaceous/Paleocene period, respectively. Two long-distance dispersal events from continental Africa to South America occurred in the Cenozoic (*Dorstenia* and *Brosimum* s.l.). In Chapter III, I tested the climatic niche difference (temperature and precipitation) between the two breeding systems (monoecy and gynodioecy) in *Ficus* using a new dataset of cleaned spatial occurrence records and breeding systems for 183 species. I used two comparative approaches: generalized estimating equations (GEE) and generalized linear models (GLM). A positive relationship between precipitation and gynodioecy was supported by GLM, but not GEE analyses, and no relationship between temperature and breeding systems was supported by either method. Higher dispersal ability and the potential for self-fertilization may explain why monoecious species of *Ficus* have been able to colonize and survive in drier environments. This thesis highlights the potential of phylogenetic comparative methods and phylogenomic data to address questions of breeding system evolution and biogeography in Moraceae, and opens up several important new perspectives worth investigating in other plant clades, such as a relationship between breeding system and climatic niche.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France